KESİKLİ RASSAL DEĞİŞKENLER İÇİN ENTROPİ OPTİMİZASYON PRENSİPLERİ VE UYGULAMALARI

Çiğdem GİRİFTİNOĞLU

Yüksek Lisans Tezi

Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Temmuz – 2005

JÜRİ VE ENSTİTÜ ONAYI

Çiğdem Giriftinoğlu’nun **“Kesikli Rassal Değişkenler İçin Entropi Optimizasyon Prensipleri ve Uygulamaları”** başlıklı İstatistik Anabilim Dalındaki, Yüksek Lisans tezi …………….. tarihinde, aşağıdaki jüri tarafından Anadolu Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca değerlendirilerek kabul edilmiştir.

	Adı-Soyadı	İmza
Üye (Tez Danışmanı)	**:Yard. Doç. Dr. Zerrin AŞAN**	**………..**
Üye	**:Prof. Dr. Aladdin ŞAMİLOV**	**………..**
Üye	**:Yard. Doç. Dr. Veysel YILMAZ**	**………..**

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

Enstitü Müdürü

ÖZET

Yüksek Lisans Tezi

KESİKLİ RASSAL DEĞİŞKENLER İÇİN
ENTROPİ OPTİMİZASYON PRENSİPLERİ
VE
UYGULAMALARI

ÇİĞDEM GİRİFTİNOĞLU

Anadolu Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Yard. Doç. Dr. Zerrin AŞAN
2005, 93 sayfa

Bu tezde, öncelikle olasılıksal bir sistemde var olan belirsizliğin ölçümünü ifade eden entropi kavramı ve buna bağlı olarak Shannon Entropi Ölçütü ile Kullback-Leibler Entropi Ölçütü tanımlarına yer verilmiştir. Bu iki entropi ölçütünün özellikleri ile bu ölçütleri optimize etme yöntemlerinden Jaynes'in Maksimum Entropi Prensibi (MaxEnt) ve Kullback'in Minimum Çapraz Entropi Prensibi (MinxEnt) ele alınmıştır. Sözü edilen ilkelerin mantığı ve matematiği ayrıntılarıyla gösterilmiş, aralarındaki ilişki ve farklılıklara değinilmiştir. Her iki ilke, istatistiksel problemlere uygulanarak elde edilen denklem veya denklemler sistemi Yarıya Bölme Yöntemi ve Newton Yöntemi kullanılarak Visual Basic programlama dilinde hazırlanan programlarda çözülmüş ve iki prensibin sonuçları karşılaştırılarak sunulmuştur. Ayrıca Devlet İstatistik Enstitüsü'nden alınan 2003-2004 elektrik tüketimi verileri üzerinde Minimum Çapraz Entropi Prensibi kullanılarak 2004 tüketim oranları tahmin edilmiş ve gerçek değerlerle yakınlığı ortaya konulmuştur.

Anahtar Kelimeler: Shannon Entropi Ölçütü, Maksimum Entropi Prensibi, Kullback-Leibler Entropi Ölçütü, Minimum Çapraz Entropi Prensibi, Olasılık Dağılımı Tahmini

ABSTRACT

Master of Science Thesis

ENTROPY OPTIMIZATION PRINCIPLES FOR DISCRETE RANDOM VARIABLES AND APPLICATIONS

ÇİĞDEM GİRİFTİNOĞLU

Anadolu University
Graduate School of Sciences
Statistics Program

Supervisor: Assist. Prof. Dr. Zerrin AŞAN

2005, 93 pages

In this thesis, first of all the concept of entropy as an expression measuring the uncertainty in probabilistic system and depending on this concept the definitions of Shannon Entropy Measure and Kullback-Leibler Entropy Measure are given. The properties of these two entropy criterions and their optimization principles, Jaynes's Maximum Entropy Principle (MaxEnt) and Kullback's Minimum Cross-Entropy Principle (MinxEnt), are dealt. The idea and mathematical base of mentioned principles are shown in details, existing relationships and differences between these principles are also discussed. Equations which are obtained by applying both principles on statistical problems are solved via Visual Basic by using the Newton's Method and Bisection Method. Results of these two principles are compared. Furthermore, an estimation procedure is shown by using Minimum Cross-Entropy Principle on the data of 2003 rates of electric consumption taken from State Institute of Statistics. Moreover, the closeness of estimated and real rates is shown.

Keywords: Shannon's Entropy Measure, Maximum Entropy Principle, Kullback-Leibler Entropy Measure, Minimum Cross-Entropy Measure, Estimate of Probability Distribution

TEŞEKKÜR

"Kesikli Rassal Değişkenler İçin Entropi Optimizasyon Prensipleri ve Uygulamaları" başlıklı yüksek lisans tez çalışmamın oluşmasında bana destek olan ve her türlü sorunlarımla ilgilenip çözüm yolları arayan değerli tez danışmanım Yard. Doç. Dr. Zerrin AŞAN 'a teşekkür ederim.

"Entropi ve İnformasyon Teorisi", "Ölçüm Teorisi ve Olasılık" adlı dersleri ve "Entropi Optimizasyon Prensipleri" adlı seminer çalışmalarıyla bilgisini, değerli vaktini ve desteğini benden esirgemeyen, konunun belirlenmesinde olduğu kadar, çalışmanın planlanması ve sonuçlanmasında da büyük emeği geçen çok değerli hocam Prof. Dr. Aladdin ŞAMİLOV'a katkılarından dolayı yürekten teşekkür ederim.

Yüksek lisans derslerime ve tez çalışmama vakit ayırabileceğim bir çalışma ortamı sağlanmasında bana yardımcı olan ve karşılaştığım sorunlarla ilgili sürekli görüşlerinden yararlandığım İstatistik Bölüm Başkanı Prof. Dr. Ali Fuat YÜZER, Fen Fakültesi Dekan Yardımcısı Prof. Dr. Embiya AĞAOĞLU ve sevgili hocam Yard. Doç. Dr. Berna YAZICI'ya teşekkürü bir borç bilirim.

Seminer çalışmaları sırasında, bilgi ve fikirlerini benimle paylaşmaktan kaçınmayan sevgili Arş. Gör. Yeliz MERT ve Arş. Gör. Şenay YOLAÇAN'a, ayrıca bir çok konudaki sorularıma ilgi gösterip yardımlarını esirgemeyen Arş. Gör. Levent TERLEMEZ'e de teşekkürlerimi sunuyorum.

Her zaman yanımda olan, maddi ve manevi desteklerini hiçbir zaman benden esirgemeyen bu günlere ulaşmamı sağlayan aileme de sabırları ve sonsuz güvenleri için ayrıca teşekkür ederim.

Çiğdem GİRİFTİNOĞLU

Temmuz – 2005

İÇİNDEKİLER

Sayfa

ŞEKİLLER DİZİNİ

ÇİZELGELER DİZİNİ

SİMGELER ve KISALTMALAR DİZİNİ

$D(p \parallel q)$	: Nispi entropi
$D(p:q)$	: Kullback-Leibler ölçütü
DİE	: Devlet İstatistik Enstitüsü
E	: Beklenen değer
$H(p)$	: Shannon'un entropi ölçütü
$H(X)$	: X rassal değişkeninin entropisi
$H(X,Y)$	: Bileşik entropi
$H(X \mid Y)$	: Koşullu entropi
I_X	: X rassal değişkeni hakkında tam informasyon
$I_{X\,i}$	: X rassal değişkeni x_i değerini aldığındaki kısmi informasyon
$I_{X \to Y}$	: X rassal değişkeninin Y rassal değişkenine verdiği informasyon
$I_{y_j \to X}$	: Y rassal değişkeninin y_j değerini almasının X rassal değişkeni hakkında verdiği informasyon
$J(p:q)$	: Simetrik çapraz entropi ölçütü
kov	: Kovaryans
L	: Lagrange fonksiyonu
MaxEnt	: Maksimum entropi
MinxEnt	: Minimum çapraz entropi
max	: Maksimum
min	: Minimum
nat	: Natürel
Q	: Önsel olasılık dağılımı
U	: Üniform dağılım
var	: Varyans
Δ	: Sonuçları bilinen dağılım

1. GİRİŞ

1.1. Giriş

Son yıllarda termodinamikte, istatistiksel mekanikte ve informasyon teorisinde kullanılan temel bir kavram olan entropi ilk olarak termodinamiğin II. yasasında ortaya çıkmıştır. I. yasa evrendeki toplam enerjinin sabit olduğunu ve enerjinin yok edilemeyeceğini söyler. Diğer bir tanımlama ile enerjinin korunumu yasası olarak da bilinir. Bu yasaya göre enerji değişik formlarda bulunabilir ve bu enerji çeşitleri yine birbirlerine dönüştürülebilir. Birçok enerji formu kayıpsız olarak ısı enerjisine dönüşürken, ısı enerjisinin dışardan destek olmaksızın, örneğin mekanik enerjiye kayıpsız olarak dönüşümü mümkün değildir. Kayıpsız olarak enerji dönüşümü geri dönüşümlü (tersinir) süreç olarak adlandırılır. Isı enerjisi, sıcaklığı yüksek olan cisimlerden düşük olanlara doğru akar. Bu süreç geri dönüşümsüzdür (tersinmez). Bir başka deyişle dışardan yardım olmadan düşük sıcaklıktaki cisimden yüksek sıcaklıktaki cisme ısı aktarmak mümkün olmaz ve de ısı enerjisinin diğer enerji formlarına dönüşümü %100 olamaz. II. yasaya göre tüm doğal ve teknik enerji dönüşüm süreçleri geri dönüşümsüzdür ve bu süreçlerin yönü hep olasılığı yüksek olan duruma doğrudur. Enerji farklarının azaldığı ve ortadan kalktığı durum olası durumdur.

19. yüzyılda sanayi devriminin baş aktörü olan makineleri daha mükemmele ulaştırma isteği ile yapılan çalışmalar sırasında bilim adamlarının önünde beliren “Hangi tür bir makine en çok verimle çalışır?”, “Kayıplar sıfırlanabilir mi?”, “Kayıpların kaynağı nedir?” vb. sorulara yanıt bulmaya çalışılırken yapılan araştırmalar neticesinde yüzde yüzlük verimle çalışan bir makine üretmenin imkansız olduğu sonucuna varılmıştır. Çünkü ne türlü bir makine yapılırsa yapılsın makineye verilen enerji ile makineden başka bir şekle dönüştürülmüş olarak elde edilen enerji arasında sıfırlanamaz bir kayıp mevcuttur ve verilen enerjinin bir kısmı makine içi sürtünmeler vasıtasıyla ısıya dönüşmektedir. Kaybolan ısı ise hiç bir zaman enerji olarak tekrar elde edilemez. İfade edilen enerji kaybı, I. yasanın ihlali şeklinde

anlaşılmamalıdır. Kayıplardan kasıt, vardan yok olma şeklinde olmayıp, enerjinin ısı şekline dönüşüp kullanılabilir olmaktan çıkması, sistemin (makine. ortam, araç vb.) yapısına katılmasıdır. Bir süreç içinde gerekli toplam enerji sabit kaldığı halde, sürtünme ve benzeri temaslar yüzünden kullanılabilir enerji azalmaktadır ve bunun sonucu olarak yüzde yüzlük verimle çalışan bir makine yapılamamaktadır. İşte, kalitesi düşen enerji için kullanılan ölçüye "entropi" adı verilir. İş üretme yeteneğinde olmayan enerjinin evrende geri kazanımı mümkün olmayan biçimde artışı entropi ile ölçülür.

Termodinamiğin II. yasası, fiziğe geri dönüşümsüz olaylar düşüncesini getirmiştir. Bu kanuna göre fiziksel hadiselerde geri döndürülemez belirli bir eğilim vardır. Örneğin, bir bardak sıcak çay etrafına ısı vererek soğur ve hiç bir zaman çayımız verdiği ısıya kendiliğinden toplayıp eski haline gelmez. Yukarıdan serbest bırakılan bir top yerden sekip bırakıldığı yüksekliğe kadar çıkmayı başaramaz. Bir pervane ne kadar hızlı çevrilirse çevrilsin, çevirme işlemini bıraktıktan bir müddet sonra durur ve hiç bir zaman da sürtünmeye harcadığı enerjisini toparlayıp tekrar dönmeye başlamaz. Bir odaya sıkılan parfüm ilk önce yakın çevresi tarafından hissedilir, bir süre sonra karşı köşedeki kişi bile kokuyu alır, ama daha sonra koku gittikçe etkisini kaybeder ve parfüm zerrecikleri atmosferde dağılıp gider, geri dönüşsüz evrensel eğilimin etkisinde bir harekete mecbur kalır. Bütün bu sayılan süreçlerin ortak yanı; belirli bir doğrultuda, düzenden düzensizliğe, bütünden yayılmaya, kullanılır olabilirlikten kullanılamazlığa doğru yol almalarıdır.

19. yüzyılda ilk kez R.Clausius bu evrensel eğilime entropi ismini vererek matematiksel bir ifadesini oluşturmayı başarmıştır. II. yasa, kısaca entropi artışı olarak özetlenebilir. Bütün varlıkların, eninde sonunda entropisi artmaktadır. Kainattaki olayların tümü yukarıda saydığımız gibi geri dönüşümlü olmayan olaylardır. Güneş bir bardak sıcak çay gibi ısısını tüketmektedir. İçinde bulunduğumuz Samanyolu Galaksisi ve diğer galaksiler bir odaya sıkılan parfümün zerrecikleri gibi birbirlerinden hızla uzaklaşmaktadır. Böylelikle istatistiksel fizik "Evrenin entropisi sürekli olarak artmakta mıdır?" sorusunun yanıtını araştırır ve ona

göre doğa, gelişi güzelliği, ısıl eşitliği ve organizasyonun olmadığı, bileşenlerin birbirine karıştığı bir tek düzeliği tercih eder.

İstatistiksel fizikte entropi kavramının gelişimi ona yeni anlamlar kazandırmıştır. Buna göre entropi artık sadece enerjinin tüketimi esnasında kalitesinin düşmesinin bir ölçüsü değildir. İstatistiksel fizikte entropi aynı zamanda, sistemlerin düzenliliği (organize olmuşluğu) ile ilgili bir ölçü olmuştur. Buna göre doğal süreçler, termodinamik olarak meydana gelme olasılığı daha yüksek olan durumu tercih ederler. Sadece ısı değil, aynı zamanda örneğin havayı oluşturan oksijen, karbon dioksit, azot gibi moleküller de mekan içinde homojen bir biçimde birbirine karışırlar. Herhangi bir bileşeninin, bir dış etki olmaksızın yani kendiliğinden bir mekanın belli bir bölümünde birikmesi olasılığı son derece zayıftır ya da yok denecek kadar azdır.

Bu bakış açısıyla Ludwig Boltzman İstatistiksel fizikte entropiyi, olasılık kavramını da gündeme getirerek doğal ve teknik süreçlerde geçerli olan termodinamik düzensizlik denklemini ortaya koymuş, ünlü Boltzmann bağıntısını $S = k \ln w$ olarak elde etmiştir (S : Entropi, k : Boltzman sabiti, W : Termodinamik olasılık). Bir sistemde entropinin artışıyla sistemdeki düzensizliğin artacağı (organizasyonun kaybolacağı) ve tüm doğal ve fiziksel süreçlerde tersinmezlik yüzünden düzensizlik ve karmaşanın olasılığının hep en yüksek olduğu söylenebilir.

Entropi kavramının istatistiksel fizikte Boltzmann bağıntısı ile gelişiminin ardından Claude Shannon tarafından "İletişimin Matematiksel Teorisi"nde informasyon ile entropi kavramı arasında ilişki kurulmasıyla bu kavramın kullanımı hemen hemen tüm bilim dallarına yayılmıştır (Kapur ve Kesavan 1992).

İnformasyon; nesne, olay ve/ya kişilerle ilgili veri ve gerçeklerin işleme tabi tutulmuş bir formudur ve alıcı durumunda olan kişinin söz konusu sistem veya süreç hakkındaki bilgisini artırır, içinde bulunduğu belirsizliği azaltır. Ancak basit bir örnekle otobüs durağında yağmur altında bekleyenlerden birinin "yağmur yağıyor" şeklindeki iletisinin, aynı yağmurun altında bekleyen diğer insanlar için hiçbir informasyon değeri yoktur. İnformasyon bir eylem için kullanıma hazır duruma geldiğinde "bilgi" ye dönüşür. Günlük kullanımda yararlı veya işe yarar bilgi ile eş değer tuttuğumuz informasyonun anlamı fizikçi, iletişim uzmanı ve matematikçiler için biraz daha farklıdır. Onlar informasyonun ölçülme, iletilme ve depolanma boyutlarıyla uğraşır.

İletişim sürecinde informasyon, mesajlar aracılığı ile iletilir. Mesajlar resim, sözcük, nota vb. olabilir. Dr. Claude Shannon'un 1948'de hazırladığı "The Mathematical Theory of Communication" adlı kitabında anlatılan iletişim teorisi, entropi ve informasyon kavramları arasında kurulan niceliksel ilişkiye dayandırılmaktadır. İlk bakışta termodinamikte sık sık geçen entropi kavramının iletişim teorisi ile ne ilişkisi olabilir diye bir soru akla gelebilir ancak bu ilişkinin sezgisel değil tümüyle matematiksel kanıtlara dayandığı ve özellikle bilgi işlem, otomasyon ve yapay zeka vb. teknik uygulamalar için bugüne kadar çok büyük başarılara katkıda bulunduğu söylenebilir. Bu teori, tüm informasyonun (mesajların anlamsal yönü hariç) açık/kapalı, evet/hayır veya 1/0 gibi biçimlere dönüştürülebileceğini göstermektedir. Shannon'un teorisinde informasyon belirsizlikle eş tutulmaktadır. Biraz yanlış çağrışımlar yapsa da bu tespit doğrudur.

İnformasyon Miktarı = Başlangıçtaki belirsizlik-İnformasyon alındıktan sonraki belirsizlik

olarak ifade edilmiştir.

1948 yılı entropi alanındaki çalışmaların başlangıç yılı olarak kabul edilir. İnformasyon teorisinde Shannon'un, bir bilgi kaynağının belirsizliğinin ölçümü olan H niceliğini keşfetmesinin ardından matematiksel formlarının ve fiziksel değerlerinin benzerliğinden dolayı bu H niceliği de istatistiksel mekanik ve termodinamikteki S gibi entropi olarak adlandırılmıştır. Shannon, bir fiziksel sistemdeki durumların sayısı n iken ve sistemin bu durumlarda bulunma olasılıkları $p = (p_1, p_2, \ldots, p_n)$ olmak üzere entropiyi $H_n(p) = -\sum_{i=1}^{n} p_i \ln p_i$ olarak tanımlamıştır. Bunun üzerine entropi H'a açıklık getirmek için bu kavram "Shannon'un entropisi" adı ile anılmıştır (Wu 1997).

Shannon (1948)'un çalışmasını temel alan Jaynes, bir ölçek üzerinde yapılan ölçümlere dayandırılan sistemin durumlarını tanımlamak için bir olasılık dağılımı türetmenin metodunu Shannon entropi ölçütünü maksimum yapan prensibi ortaya koyarak mühendislik ve fizikteki problemlerin bir çok çeşidine bu sonuçları uygulamıştır. Bu prensip Jaynes'in Maksimum Entropi Prensibi (MaxEnt) olarak adlandırılır. Bilgi olarak sadece ortalama değerleri verilmiş olduğunda uyumlu dağılımlar genelde sonsuz sayıdadır. MaxEnt, Shannon entropi ölçütünü maksimum yapan ve ortalama değer kısıtlarıyla eşzamanlı tutarlı dağılımı seçmeyi önerir (Kapur ve Kesavan 1992).

1951'de Kullback aynı kısıtları sağlayan ve belirli bir önsel olasılık dağılımının bilinmesi durumunda bu dağılıma en yakın dağılımı seçmeyi öngören ve Kullback-Leibler ölçütünü minimize eden Minimum Çapraz Entropi Prensibi (MinxEnt) çalışmasını ortaya koymuştur.

Entropi Optimizasyon Prensipleri ve Uygulamaları çalışmasının ikinci bölümünde entropi kavramı ve bir rassal değişkenin entropisi tanımlanmıştır. Shannon entropi ölçütünün özellikleri ile bileşik ve koşullu entropi kavramlarının yanı sıra nispi entropi ve karşılıklı informasyon ele alınarak formülasyonlarına yer verilmiştir.

Çalışmanın üçüncü bölümünde Entropi Optimizasyon Prensipleri, Shannon entropi ölçütünü maksimum yapan Jaynes'in Maksimum Entropi Prensibi (MaxEnt) ile Kullback-Leibler ölçütünü minimize eden Kullback'in Minimum Çapraz Entropi Prensibi (MinxEnt) ayrıntılarıyla ele alınarak aralarındaki ilişkiye değinilmiştir.

Çalışmanın özgün yanını oluşturan dördüncü bölümde altı duruma sahip kesikli rassal değişkenlerin verilen iki kısıt ile MaxEnt dağılımı bulunmuştur. Bu dağılım bulunurken koşullu ekstremum probleminin çözümü için Langrange çarpanları yönteminden yararlanılmıştır. Bu durumda ortaya çıkan çarpanlar matematiksel olarak tam bulunamadığı için yaklaşık metotların uygulanması gereksinimi ortaya çıkmıştır. Bu nedenle de iki kısıtlı problemin nümerik çözümü ele alınmıştır. Bu amaçla önce bu denklemin çözüme sahip olduğu matematiksel analiz ve analitik geometri yöntemleriyle gösterilmiştir. Söz konusu denklemin iki yöntemle ayrı ayrı nümerik çözümü gerçekleştirilmiştir. Bu yöntemlerden birisi Yarıya Bölme Yöntemi'dir ki bu yöntemle tanım aralığının orta noktasında fonksiyonun işaretini belirlemekle denklemin çözümünün hangi alt aralıkta bulunduğu belirlenir. Daha sonra süreç devam ettirilir. Denklemin çözümü Visual Basic programlama dilinde hazırlanan bir programda istenilen hassasiyet ölçüsünde bulunmuştur ve verilen 11 farklı ortalama değere göre olasılık dağılımları hesaplanmıştır. Hazırlanan program n duruma sahip istenilen X değişkeni değerleri için de geçerlidir.

Daha sonra ise adı geçen denklemin çözümü için diğer bir metot Newton Yöntemi uygulanmıştır. Bu yöntemin Yarıya Bölme Yöntemi'nden üstünlüğü ise köke daha hızlı yaklaşması bir başka deyişle daha az sayıda iterasyon gerektirmesidir. Bu defa Newton Yöntemi ile Visual Basic programlama dilinde hazırlanan, yine keyfi n sayıda duruma sahip kesikli rassal değişken için de geçerli olan bir programda olasılık dağılımları hesaplanmıştır. Ardından verilen üç kısıt ile MaxEnt dağılımının bulunması ele alınmıştır. İki kısıt söz konusu olduğunda tek denklem ortaya çıkmasına rağmen bu problemde iki denklemden oluşan bir denklem sistemi ortaya çıkmıştır. Bu nedenle önceki problem için geçerli olan Yarıya Bölme Yöntemi kullanılmamış, denklemler sisteminin nümerik çözümü için Newton Yöntemi seçilmiş ve olasılık dağılımı elde edilmiştir.

MaxEnt prensibi ile önce incelediğimiz iki kısıtlı problem önsezi ve tecrübeye dayalı önsel bir dağılım verilmesiyle yeniden ele alınmış, MinxEnt prensibi ile bir Visual Basic programlama dilinde hazırlanan programda olasılık dağılımı hesaplanmış, MaxEnt ile bulunan dağılımla karşılaştırılmıştır. Belli bir önsel dağılımın verilmesiyle entropideki azalış gösterilmiştir.

Ayrıca DİE (Devlet İstatistik Enstitüsü)'den alınmış 2003-2004 elektrik tüketimi verileri üzerinde MinxEnt Prensibi uygulanarak 2004 tüketim oranları tahmin edilmiş ve sonuçlar 2004 gerçek oranlarla karşılaştırılarak yakınlığı gösterilmiştir.

2. ENTROPİ VE İLGİLİ KAVRAMLAR

Bu bölümde entropi kavramı açıklanarak, entropi birimi ve rassal değişkenlerin entropisi tanımlanmış, entropi fonksiyonunun özellikleri ve ekstremumu üzerinde durulmuş, Shannon entropi ölçütünün özellikleri ile bileşik, koşullu, nispi entropi ve karşılıklı informasyon kavramları ele alınarak, önemli formülasyonlara yer verilmiştir.

2.1. Entropi

Entropi tanımı en genel deyimiyle düzenden düzensizlik biçimine bir dönüşümdür. Çok eski bir kavram olan entropi ilk olarak termodinamikte 19. yüzyılın ortalarına doğru Clausius tarafından geliştirilmiştir. Daha sonra Boltzmann entropi ifadesini istatistiksel mekanikte $S = k \ln w$ ile elde etmiştir. Termodinamikteki ya da istatistiksel mekanikteki entropi bir termodinamiksel sistemde var olan düzensizlik düzeyinin bir ölçümüdür (Wu 1997). Termodinamikteki entropi bilgi sisteminin momentumu veya ortalama enerjisi hakkında bilgi verir. Eğer bu bilgi bedel, fiyat, kira gibi ekonomik değişkenler hakkında ise o zaman ekonomik entropi, coğrafi dağılım hakkında ise mekansal entropi olarak adlandırılır (Kapur ve Kesavan 1992).

Entropi Yunanca kökenli bir kelime olup sözlükteki anlamı belirsizliktir. Gerçekleştirilen çalışmayla ilgili olan “olasılıksal belirsizlik durumu” ele alındığında entropi “bir sisteme ait belirsizlik düzeyinin ölçümü” olarak ifade edilecektir. Bir başka deyişle herhangi bir olayın ortaya çıkma olasılığı, belli seviyede bu olayın gerçekleşip gerçekleşmeyeceği hakkında belirsizliğin göstergesidir. Örneğin bir para atışında Yazı ya da Tura gelip gelmeyeceği, bir zar deneyinde 6 gelip gelmeyeceği bir belirsizlik taşır. Aynı şekilde bir adayın bir seçimi kazanması, ya da maksimum pazar payına sahip olacak ürün markası da belirsizlik içerebilir.

Entropi; anlamına, türüne, değerine veya diğer herhangi bir sübjektif özelliğine bakılmaksızın, iletişim yaratan sembol, sinyal ya da sayılar dizisinin istatistiksel yapısını analiz eden Bilgi Kuramı’na dayanmaktadır. Bilgi Kuramı’nın temel prensipleri, iletişim hatlarından sinyal gönderilmesi işlevini stokastik bir süreç olarak ele alan Shannon tarafından geliştirilmiştir. Shannon'a göre bir olay hakkında bilgi edinilmesi, o olayın belirsizlik içermesi halinde söz konusu olabilir. Buna göre, ortaya çıkma olasılığı yüksek olayların meydana gelmesi fazla bilgi getirmemekte; aksine, olasılığı düşük olayların oluşması daha fazla bilgi taşımaktadır. Bu çerçevede, belli bir alternatif durumun oluşma olasılığı, o durumu oluşturan işaret, sembol ya da sayının belirsizlik derecesini temsil etmektedir. Bu nedenle, kazanılan bilgi giderilen belirsizlik miktarının dolaylı bir ölçütüdür. Shannon yukarıdaki yaklaşımla entropi kavramını, bir olayın alabileceği çeşitli alternatif durumların (değerlerin) beklenen değeri olarak matematiksel bir bağıntıyla tanımlamıştır. Bu tanıma göre de entropi belli birimlerle (bit, napier, desibel gibi) ölçülebilen niceliksel bir büyüklük olmaktadır. Böylelikle, bir rasgele sürecin entropisi hesaplanabilmekte ve söz konusu birimlerle ifade edilebilmektedir (Özkul 2001).

Sonuçlarının olasılığı ile ilişkilendirilmiş belirsizlik "olasılıksal belirsizlik" olarak adlandırılır. "Olasılıksal belirsizlik" durumunda belirsizliğin ölçütü olarak olayın gerçekleşme olasılığı dikkate alınır (Kapur ve Kesavan 1992). Bu durumda farklı olasılık dağılımlarının onlarla ilişkilendirilmiş farklı belirsizliklere sahip olduğu söylenebilir. Olasılıkların yoğunlukları arttıkça ilgili olasılık uzayının belirsizliği azalır. Örneğin kolayca görülebilir ki yazı ya da tura için (0.5,0.5) olasılık dağılımının entropisi, bir piyangoda kazanmanın (0.00001,0.99999) olasılık dağılımının entropisinden çok daha fazladır (Papoulis 1991).

Bu durum şöyle de ifade edilebilir. Bir olasılık deneyinin $a_1,a_2,\dots,a_n$ olası sonuçları, $p_1,p_2,\dots,p_n$ olasılıklarıyla

$$\sum_{i=1}^{n} p_i = 1 \text{ ve } p_i \geq 0 \quad i = 1,2,\dots,n$$

koşullarını sağlamak suretiyle ortaya çıksın.

Bu olasılık deneyinde olasılık i. sonuçta yoğunlaşmış ise ve $p_j = 0$ ise ($\forall$ $i \neq j$ için) bu durumda deneyin sonucu kesindir, belirsizlik yoktur ve yoğunluk en büyüktür. Tersi durumda, eğer deneyin tüm sonuçları eşit olasılıklara sahip ise

($p_i = \frac{1}{n}$, $i = 1,2,\ldots,n$) yoğunluk en küçük, deneyin sonuçlarının belirsizliği ise en büyük olacaktır (Kapur ve Kesavan 1992).

2.1.1. Rassal Değişkenin Entropisi

Shannon, "bir olasılık uzayının belirsizliğinin ölçüsü" olarak tanımladığı entropi için, bir deneyin sonuçlarının içerdiği ortalama bilgi miktarını önermiştir (Kapur ve Kesavan 1992).

Bu tanımdan yola çıkarak yukarıdaki olasılık deneyinde i. sonucun ortaya çıkıp çıkmama belirsizliğinin bir ölçümünü atama problemi dikkate alındığında i. sonuca karşılık gelen p_i olasılığının herhangi bir nedenle değiştiği varsayılır. Bu değişim dp_i kadar olduğunda bu değişimin ortalaması olan $\frac{dp_i}{p_i}$, belli bir ölçümün değişimini ifade eder ki bu ölçüm entropidir. Entropi h_i ile işaret edildiğinde $dh_i = -\frac{dp_i}{p_i}$ eşitliği elde edilir. Eşitlikteki negatif işaret olasılıktaki pozitif değişimin entropide negatif değişime sebep olmasındandır. Başka deyişle i. sonucun ortaya çıkma olasılığındaki artış, entropinin azalmasına neden olacaktır.

Bu $dh_i = -\frac{dp_i}{p_i}$ eşitliğinin her iki yanını 1' den p_i' ye kadar integrali alındığında;

$$\int_1^{p_i} dh_i = -\int_1^{p_i} \frac{dq_i}{q_i}$$

$$h_i(p_i) - h_i(1) = -\ln(p_i) + \ln(1)$$

$$h_i(p_i) = h_i(1) - \ln(p_i)$$

$$h_i(p_i) = -\ln(p_i) \tag{2.1}$$

formülü elde edilir. Burada $h(1) = 0$ olması, gerçekleşme olasılığı bir olan olayın entropisinin sıfır olduğunu ifade eder.

Eşitlik (2.1), olasılık deneyinin yalnız i. sonucunun entropisi olarak yorumlanır. Görüldüğü gibi entropi her p_i olasılığına karşılık gelen bir sayı, yani bir rassal değişkenin aldığı değerdir. Bu rassal değişkenin değerleri Çizelge 2.1.'deki gibidir.

Çizelge 2.1. Bir rassal değişken olarak her p_i olasılığına karşılık gelen entropi değeri

p	p_1	p_2	p_3	...	p_n
h	$-\ln p_1$	$-\ln p_2$	$-\ln p_3$	...	$-\ln p_n$

Söz konusu değişkenin matematiksel beklentisi deneyin tam entropisi olarak adlandırılır:

$$H(P) = \frac{-p_1 \ln p_1 - p_2 \ln p_2 - \ldots - p_n \ln p_n}{p_1 + p_2 + \ldots + p_n}.$$

Yukarıdaki matematiksel beklenti formülünde $p_1 + p_2 + \ldots + p_n = 1$ olduğundan,

$$H(P) = -p_1 \ln p_1 - p_2 \ln p_2 - \cdots - p_n \ln p_n$$

$$H(P) = -\sum_{i=1}^{n} p_i \ln p_i \tag{2.2}$$

formülü elde edilir.

Eşitlik (2.2), n tane duruma sahip bir deneyin entropisidir. Rassal değişken tanımı dikkate alınırsa her bir x_i durumuna karşılık gelen bir sayı atandığında (2.2) formülü X rassal değişkeninin belirsizliğinin bir ölçümü, yani entropisi olarak tanımlanır (Cover ve Thomas 1991).

Örnek olarak bir zar deneyinde ortaya çıkabilecek 6 duruma atanan sayılar x_i=10i olsun. Böylece x_1=10, x_2=20,…, x_6=60 değerlerini alan bir X rassal değişkeninden söz edilirken, aynı deneyde çift gelme olayına 1, tek gelme olayına 0 sayısı atanırsa $x_1 = x_3 = x_5 = 0$ $x_2 = x_4 = x_6 = 1$ değerlerini alan bir X rassal değişkeni olarak ele alınabilir.

Entropinin tanımını deneyin değil de X rassal değişkeninin belirsizliğinin bir ölçümü olarak ele aldığımızda (2.2) formülü,

$$H(X) = -\sum_{x \in X} P(x).\log P(x) \tag{2.3}$$

ile ifade edilir (Cover ve Thomas 1991).

Burada $P(X)$, X rassal değişkeninin olasılık küme fonksiyonudur. Bu durumda $P(x) = P\{X = x\},\ x \in X$ 'tir ve H(X), H(*P*) ile de gösterilebilir.

H(p) entropi fonksiyonu hem belirsizliğin ölçüsü hem de bilgi miktarının bir ölçüsüdür. $p_1, p_2, \ldots, p_n$ olasılıkları ile n olası sonucu olan bir deney yapılmadan önce bulunan *H(p)* entropisi deney sonuçları ile ilgili belirsizliği ölçer. Deney yapıldıktan sonra bulunan $H(p_1, p_2, \ldots, p_n)$ entropisi ise deneyden elde edilen bilgi miktarını ölçer.

Örneklem uzayının sürekli veya kesikli olması durumuna göre entropi ifadesi değişecektir. Örneklem uzayı sürekli ise, sürekli entropi fonksiyonu

$$H(f) = -k \int_{-\infty}^{\infty} f(x) \log f(x)$$

biçiminde tanımlanmaktadır. Burada *f(x)* deney sonuçlarının sürekli olasılık yoğunluk fonksiyonudur.

Örneklem uzayı kesikli ise; kesikli entropi fonksiyonu (2.3) eşitliği ile ifade edilen fonksiyon olacaktır (Kapur ve Kesavan 1992).

2.1.2. Entropi Birimi

Entropi formülünde farklı logaritmik tabanlar farklı entropi birimine neden olur (Wu 1997). Eğer logaritma tabanı e olarak alınırsa entropi birimi nat (natural kelimesinin kısaltması) olarak belirtilirken, logaritma tabanı 2 olarak alındığında ise entropi birimi bit olarak ifade edilecektir (Cover ve Thomas 1991).

Bir para deneyini ele alırsak, bu deneyde yazı ve tura olmak üzere iki durum söz konusudur ve bu durumlarda bulunma olasılıkları $P\{Yazı\} = P\{Tura\} = 1/2$ 'dir.

Dolayısıyla bu deneyin entropisi;

$$H(P) = -\frac{1}{2}\log_b^{1/2} - \frac{1}{2}\log_b^{1/2}$$

olacaktır.

Yukarıdaki eşitlikte logaritmik taban 2 kabul edildiğinde ise entropi;

$$H(P) = -\frac{1}{2}\log_2^{1/2} - \frac{1}{2}\log_2^{1/2}$$

$$H(P) = 1$$

bit olarak elde edilir.

O halde iki duruma sahip ve bu durumlarda bulunma olasılıkları eşit olan deneyin belirsizliği, entropi birimi olarak kabul edilir ve buna bit denir.

Başka türlü ifade edilmedikçe matematiksel işlemlerde kolaylık olması açısından logaritmik taban e olarak kabul edilecektir.

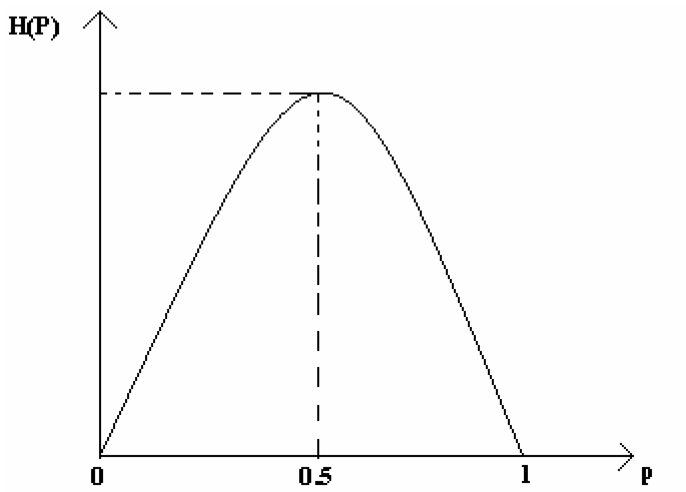
2.1.3. Entropi Fonksiyonunun Temel Özellikleri

Bir rassal değişkenin aldığı iki değerin olasılıkları eşit olmadığı yani $P(X = x_1) = p$, $P(X = x_2) = 1 - p$ olduğu durumda,

$$H(P) = -p\log p - (1-p)\log(1-p)$$

entropi fonksiyonu elde edilir.

H(P) entropi fonksiyonunun grafiği Şekil 2.1.'de gösterilmiştir. Bu grafik bize entropi fonksiyonun bazı temel özellikleri hakkında fikir verir. Entropi fonksiyonu konkav bir fonksiyondur ve p=0 ya da p=1 iken *H(P)*=0 değerini alır. Zaten bilindiği gibi olasılığın sıfır ya da bir olması durumunda rassallık dolayısıyla da belirsizlik yoktur. Benzer olarak entropi fonksiyonu maksimum değerini p=0.5'de alır ki bu durumda belirsizlik maksimumdur ve Şekil 2.1.'den de görüleceği gibi entropi fonksiyonu *X* rassal değişkeninin aldığı değerlere değil sadece olasılıklarına bağlıdır (Cover ve Thomas 1991).



Şekil 2.1. H(P) entropi fonksiyonu grafiği

Entropiyi $\log\frac{1}{P(x)}$ ’in matematiksel beklentisi olarak da tanımlamak mümkündür. Böylece *X~P(x)* ise *g(x)* rassal değişkeninin beklenen değeri,

$$E_P g(x) = \sum_{x\in\aleph} g(x)p(x)$$

olarak yazılabilir.

O halde X’in entropisi $\log\frac{1}{P(x)}$ ’ in beklenen değeri olarak da yorumlanabilir.

Bu tanımın bazı hazır sonuçlarını çıkarabiliriz;

Sonuç.1 $H(X) \geq 0$

$0 \leq p(x) \leq 1$ olması ile $\log\frac{1}{p(x)} \geq 0$ olduğu görülür ve dolayısıyla $H(X) \geq 0$ ’dır,

Sonuç.2 $H_b(X) = \left(\log_b^a\right) H_a(X)$ ’dir. **(**$\log_b^p = \log_b^a \log_a^p$)

Entropinin bu ikinci özelliği tanımdaki logaritma tabanını değiştirmeyi olanaklı kılar. Yani entropi uygun çarpanla çarpılarak başka bir tabana dönüştürülebilir (Cover ve Thomas 1991).

2.1.4. Shannon Entropi Ölçütünün Özellikleri

Mevcut verilerle elverilen maksimum entropiye (belirsizliğe) sahip olasılık dağılımı eksik bilgi üzerine dayanan çıkarımlar yapmada kullanılmalıdır. Onun maksimumundan küçük herhangi bir entropi, sağlam bir temele dayanmayan ek bilgi içerdiğinden diğer her dağılım bazı davranışlara eğilimli olacaktır. Bu yüzden bir olasılık dağılımın belirsizliğinin bir ölçümüne sahip olmak gerekir. Bu belirsizliğin ölçümü için Shannon tarafından tek bir fonksiyon tanımlanmıştır.

Her i için $p_i \geq 0$ ve $\sum_{i=1}^{n} p_i = 1$ olmak üzere $p = (p_1, p_2, \ldots, p_n)$ bir olasılık dağılımı olsun. Bu dağılım için entropinin Shannon ölçütü,

$$H_n(p) \stackrel{\Delta}{=} -\sum_{i=1}^{n} p_i \ln p_i$$

olarak verilir. Bu şekilde tanımlanan Shannon entropisi için aşağıdaki özellikler söz konusudur:

i) Entropi negatif olamaz.

$H_n(P) \geq 0$ 'dır.

Eğer $0 < p_i < 1$ ise $\ln p_i < 0$ ve $-\ln p_i > 0$ olur.

Eğer $p_i = 0$ veya $p_i = 1$ ise $p_i \ln p_i = 0$ olur.

Buradan görülüyor ki $-p_i \ln p_i \geq 0$ olacaktır. Bu da $H_n(p)$ 'nin hiçbir zaman negatif değer almayacağını gösterir (Kapur ve Kesavan 1992).

ii) $H_n(p)$ yer değiştirme özelliğine sahiptir. Yani $p_1, p_2, \ldots, p_n$ olasılıkları kendi aralarında yer değiştirse de entropi ölçümü değişmez. Rassal değişkenin aldığı değerlerin entropi üzerinde etkisi olmadığına göre bu beklenilen bir özelliktir.

iii) $H_n(p)$; 0 ile 1 arasındaki tüm p_i 'ler için $p_1, p_2, \ldots, p_n$ 'in sürekli bir fonksiyondur. Yine aynı aralıkta p_i ve $\ln p_i$ de p_i 'nin sürekli fonksiyonlarıdır. Böylece $p_i \ln p_i$ aynı aralıkta her yerde süreklidir.

Eğer $p_i = 0$ ise $p_i \ln p_i$ tanımsızdır, fakat limit alındığında;

$$\lim_{p_i \to 0} p_i \ln p_i = \lim_{p_i \to 0} \frac{\ln p_i}{1/p_i}$$

$$= \lim_{p_i \to 0} \frac{1/p_i}{-1/p_i^2}$$

$$= \lim_{p_i \to 0} (-p_i)$$

$$= 0$$

elde edilir.

Böylece $0 \ln 0 \overset{\Delta}{=} 0$ olarak tanımlanabildiğinden $p_i \ln p_i$, $p_i = 0$'da sürekli olur. Bu tanımla $p_i \to 0$ olarak fonksiyonun limiti, $p_i = 0$'da fonksiyonun aldığı değere eşittir. Bu durumda $H_n(p)$'nin 0 ile 1 arasındaki tüm p_i'ler için sürekli bir fonksiyon olduğu söylenebilir. Bu özellikle birlikte, bazı olasılık değerlerindeki küçük bir değişimin, entropide de küçük bir miktarda değişime neden olduğu söylenebilir.

iv) Olanaksız bir olayın yani olasılığı sıfır olan bir olayın ifadeye dahil edilmesiyle entropi değişmez.

$$H_{n+1}(p_1, p_2, ..., p_n, 0) = -\sum_{i=1}^{n} p_i \ln p_i - 0 \ln 0$$

$$= -\sum_{i=1}^{n} p_i \ln p_i$$

$$= H_n(p_1, p_2, ..., p_n)$$

v)

$$\Delta_1 = (1,0,...,0)$$
$$\Delta_2 = (0,1,...,0)$$
$$.$$
$$.$$
$$.$$
$$\Delta_n = (0,0,....,0)$$

şeklinde verilen n tane sonuçları bilinen dağılım her biri için $H_n(p) = 0$'dır.

vi) Yukarıdaki özellikte verilen dağılımlar dışındaki diğer dağılımlar için $H_n(p)>$ 0'dır. Yani sonuçları kesin olmayan her dağılım için entropi pozitif olacaktır.

vii) $H_n(p)$, ($p_1, p_2, \ldots, p_n$)'in konkav bir fonksiyonudur. Eğer $\varphi(p) = -p \ln p$ alınırsa $\varphi'(p) = -(1+\ln p)$ ve $\varphi''(p) = -\frac{1}{p}$ olur.

Kolayca görülebilir ki $\varphi(p)$ konkav bir fonksiyondur. Konkav fonksiyonların toplamı da konkav bir fonksiyon olduğundan $H_n(p)$, $(p_1, p_2, ..., p_n)$'in konkav bir fonksiyonudur. Lineer kısıtlara konu olan konkav bir fonksiyon için lokal maksimum aynı zamanda global maksimum da olacağından bu özellik çok kullanışlıdır (Kapur ve Kesavan 1992).

viii) Entropi fonksiyonu, bütün olasılıkların birbirine eşit olduğu durumda ($\forall$ i için $p_i=1/n$) maksimum değerini alır. Daha sonra ispatlayacağımız bu özellik $H_n(p) \le H_n(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$ ifadesi ile verilebilir.

ix) p_1 ve p_2 olasılıkları birleştirildiğinde entropi ifadesi;

$$H_{n-1}(p_1+p_2, \ldots, p_n) = -(p_1+p_2)\ln(p_1+p_2) - \sum_{i=3}^{n} p_i \ln p_i$$

$$= -(p_1+p_2)\ln(p_1+p_2) + p_1 \ln p_1 + p_2 \ln p_2 \underbrace{- \sum_{i=1}^{n} p_i \ln p_i}_{H_n(p)}$$

$$H_n(p_1, p_2, \ldots, p_n) = H_{n-1}(p_1, p_2, ..., p_n) + (p_1+p_2)\ln(p_1+p_2)$$
$$- p_1 \ln p_1 - p_2 \ln p_2$$
$$= H_{n-1}(p_1, p_2, \ldots, p_n)$$
$$+ (p_1+p_2)\left(\frac{-p_1}{p_1+p_2}\ln\frac{p_1}{p_1+p_2} - \frac{p_2}{p_1+p_2}\ln\frac{p_2}{p_1+p_2}\right)$$

$$H_n(p_1, p_2, \ldots, p_n) = H_{n-1}(p_1+p_2, \ldots, p_n) + (p_1+p_2)H_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right) \quad (2.4)$$

eşitliği elde edilir.

Entropinin pozitif olma özelliği de dikkate alınarak $H_{n-1}(p_1+p_2,p_3,\ldots,p_n) \le H_n(p_1,p_2,\ldots,p_n)$ sonucuna varılır. Yani iki sonucun birleştirilmesiyle entropi indirgenir. Bir başka deyişle iki sonucun birleştirilmesiyle belirsizlik artmaz. Eğer (2.4) eşitliğinde $H_2(p_1,p_2)$ ifadesi biliniyorsa, n=3 alınarak $H_3(p_1,p_2,p_3)$ elde edilebilirken, $n=4$ almakla da $H_4(p_1,p_2,p_3,p_4)$ bulunur. Böylelikle yineleme yoluyla herhangi bir n değeri için $H_n(p_1,p_2,\ldots,p_n)$ elde edilebilir. Bu özellik çok elverişli ve kullanışlı bir özellik olup "yineleme özelliği" olarak adlandırılır.

$H_n(p_1,p_2,\ldots,p_n)$'in elde edilmesi iki aşamada gerçekleşir. İlk aşamada $H_{n-1}(p_1+p_2,\ldots,p_n)$'e ulaşılır, ikinci aşamada ise p_1 ve p_2, p_1+p_2 olasılığına bölünerek $H_2\left(\dfrac{p_1}{p_1+p_2},\dfrac{p_2}{p_1+p_2}\right)$ entropisi elde edilir. Aslında bu özellik "dallanma prensibi" olarak da adlandırılır.

Eşitlik (2.4), bir tek dallanma söz konusu ise gerekli olan formülü ifade eder. Şimdi birden fazla dallanma durumu ele alınsın ve gerekli formül çıkarılsın.

Örneğin ilk iki sonuç ile p_3 ve p_4 sonuçlarının birleştirilmesiyle iki dallanma durumunda oluşturulacak formül aşağıdaki gibi olur.

$$H_{n-2}(p_1+p_2,p_3+p_4,\ldots,p_n) = -(p_1+p_2)\ln(p_1+p_2)-(p_3+p_4)\ln(p_3+p_4)-\sum_{i=5}^{n} p_i \ln p_i$$

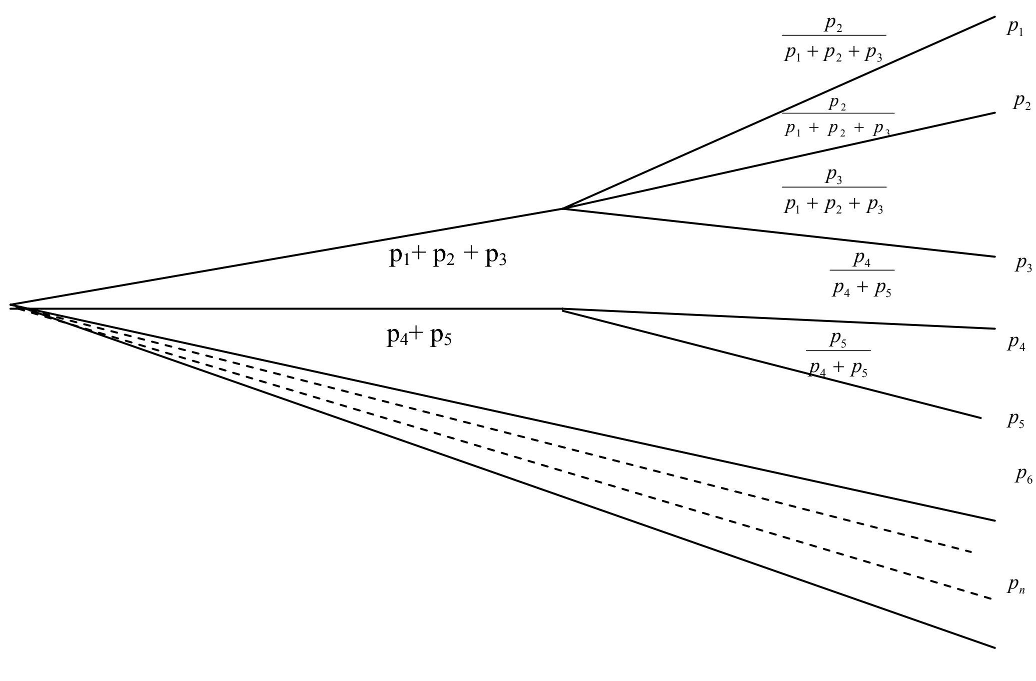
Bu formülde gerekli işlemler yapıldığında,

$$H_{n-2}(p_1+p_2,p_3+p_4,\ldots,p_n) = -(p_1+p_2)\ln\bigl(p_1+p_2\bigr)-(p_3+p_4)\ln\bigl(p_3+p_4\bigr)$$
$$+p_1\ln p_1+p_2\ln p_2+p_3\ln p_3+p_4\ln p_4\underbrace{-\sum_{i=1}^{n} p_i \ln p_i}_{H_n(P)}$$

$$H_n(p_1, p_2, ..., p_n) = H_{n-2}(p_1 + p_2, p_3 + p_4, p_5, ..., p_n)$$
$$+ (p_1 + p_2)\left(- \frac{p_1}{p_1 + p_2} \ln \frac{p_1}{p_1 + p_2} - \frac{p_2}{p_1 + p_2} \ln \frac{p_2}{p_1 + p_2} \right)$$
$$+ (p_3 + p_4)\left(- \frac{p_3}{p_3 + p_4} \ln \frac{p_3}{p_3 + p_4} - \frac{p_4}{p_3 + p_4} \ln \frac{p_4}{p_3 + p_4} \right)$$

$$\begin{aligned} H_n(p_1, p_2, ..., p_n) &= H_{n-2}(p_1 + p_2, p_3 + p_4, p_5, ..., p_6) \\ &+ (p_1 + p_2) H_2\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right) \\ &+ (p_3 + p_4) H_2\left(\frac{p_3}{p_3 + p_4}, \frac{p_4}{p_3 + p_4}\right) \end{aligned} \tag{2.5}$$

formülü elde edilir. Böylelikle aşağıdaki Şekil 2.2.' den de görüldüğü gibi istenilen kadar sonuç birleştirilebilir ve entropi değeri indirgenebilir.



Şekil 2.2. Entropinin dallanma prensibi

2.1.5. Entropi Fonksiyonunun Ekstremumu

$$H(p) = -\sum_{i=1}^{n} p_i \ln p_i \tag{2.6}$$

entropi fonksiyonu $(p_1, p_2, \ldots, p_n)$ olasılıklarına bağlı çok değişkenli bir fonksiyondur ve burada $(p_1, p_2, \ldots, p_n)$ değişkenleri arasında,

$$p_1 + p_2 + \ldots + p_n = 1 \tag{2.7}$$

şeklinde bir koşul vardır.

Fonksiyon (2.6), (2.7) kısıtı altında koşullu ekstremumunu incelemek için Lagrange Yöntemi kullanıldığında Lagrange fonksiyonu;

$$L = -\sum_{i=1}^{n} p_i \ln p_i - \lambda\left(\sum_{i=1}^{n} p_i - 1\right) \tag{2.8}$$

kurulur. Fonksiyon (2.6), (2.7) kısıtı altında ekstremuma sahipse (2.8) eşitliğinin kısmi türevleri sıfıra eşit olmalıdır. Koşullu Ekstremum'un varlık teoremine göre bu fonksiyonun kısmi türevlerinin sıfıra eşit olduğu noktalar kritik noktalardır. Ancak daha önce ispatlandı ki entropi fonksiyonu konkav bir fonksiyondur ve dolayısıyla elde edilen kritik noktalar entropi fonksiyonunu maksimize edecek değerler olacaktır.

Kritik noktaları elde etmek amacıyla kısmi türev alındığında,

$$\frac{\partial L}{\partial p_i} = -(1 + \ln p_i) - \lambda = 0$$

$$-\ln p_i - 1 - \lambda = 0$$

$$-\ln p_i = 1 + \lambda$$

$$\ln p_i = -\lambda - 1$$

$$p_i = e^{-\lambda - 1} \tag{2.9}$$

sonucuna varılır.

Böylece $p_1 = p_2 = \ldots = p_n = e^{-\lambda-1}$ değerlerini (2.7) kısıtında yerine yazmakla λ değeri aşağıdaki gibi elde edilebilir.

$$e^{-\lambda-1} + e^{-\lambda-1} + \ldots + e^{-\lambda-1} = 1$$

$$n.e^{-\lambda-1} = 1$$

$$e^{-\lambda-1} = \frac{1}{n}$$

$$-\lambda - 1 = -\ln(n)$$

$$\lambda = \ln(n) - 1$$

Bulunan λ değerinin p_i ’de yerine konulmasıyla da ;

$$p_i = e^{-\lambda-1}$$

$$= e^{-\ln(n)+1-1}$$

$$p_i = \frac{1}{n} \qquad i = 1,2,\ldots,n \tag{2.10}$$

kritik noktaları bulunmuş olur. Kritik noktaların entropi fonksiyonunda yerine yazılması ile

$$H(p) = \sum_{i=1}^{n} \frac{1}{n} \ln(\frac{1}{n}) = -n\frac{1}{n}\ln(\frac{1}{n})$$

$$= -\ln(\frac{1}{n})$$

$$H(p) = \ln(n) \tag{2.11}$$

ifadesi elde edilir. Eşitlik (2.11), (2.6) fonksiyonunun (2.7) koşulu altında maksimum değeri olarak tanımlanabilir. Bu durumda $H_n(p)$ entropi fonksiyonu maksimum değerine tüm olasılıkların eşit olduğu $U(\frac{1}{n},\frac{1}{n},\ldots,\frac{1}{n})$ ile ifade edilen kesikli düzgün dağılım halinde ulaşır.

2.2.Bileşik Entropi

Daha önceki konularda tek bir rassal değişkenin entropisini $-\sum_{x\in\chi} p(x)\log p(x)$ olarak tanımlanmıştı. Şimdi aynı olasılık uzayı üzerinde tanımlanmış iki bağımsız X ve Y rassal değişkenlerinin koşullarında tanımlanabilen bir olasılıksal sistem ele alınsın. X rassal değişkeninin aldığı değerler $\{x_1, x_2, \ldots, x_n\}$, Y rassal değişkenin aldığı değerler ise $\{y_1, y_2, \ldots, y_m\}$ gibi yazılabilir. Bu durumda X ve Y bağımsız rassal değişkenlerinin bileşik olasılık dağılımı Çizelge 2.2.'deki gibi oluşturulabilir.

Çizelge 2.2. X ve Y rassal değişkenlerinin bileşik olasılık dağılımı

	y_1	y_2	y_3	. . .	y_m	P_x
x_1	p_{11}	p_{12}	p_{13}		p_{1m}	p_1
x_2	p_{21}	p_{22}	p_{23}		p_{2m}	p_2
x_3	p_{31}	p_{32}	p_{33}		p_{3m}	p_3
. . .						
x_n	p_{n1}	p_{n2}	p_{n3}		p_{nm}	p_n
P_Y	q_1	q_2	q_3		q_m	

Burada,

$$p_i = \sum_{j=1}^{m} p_{ij} = P(X = x_i), \qquad i = 1,2,\ldots,n \quad p_1 + p_2 + \cdots + p_n = 1$$

$$q_j = \sum_{i=1}^{n} p_{ij} = P(Y = y_j), \qquad j = 1,2,\ldots,m \quad q_1 + q_2 + \cdots + q_m = 1$$

ve

$$p_i = p(x_i, y_j) = P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) = p_i q_j = p_{ij}$$

olmak üzere X ve Y rassal değişkenlerinin bileşik entropisi

$$H_{mn}(X,Y) = -\sum_{j=1}^{m}\sum_{i=1}^{n} p_{ij} \log p_{ij} \tag{2.12}$$

olarak tanımlanır. Bu iki rassal değişkenin bağımsızlık koşulunu sağlamaları nedeniyle p_{ij} yerine $p_i q_j$ yazılabilir.

Bu durumda bileşik entropi formülünden;

$$H_{mn}(X,Y) = -\sum_{j=1}^{m}\sum_{i=1}^{n} p_i q_j \log(p_i q_j)$$

$$= -\sum_{j=1}^{m}\sum_{i=1}^{n} p_i q_j (\log p_i + \log q_j)$$

$$= -\sum_{j=1}^{m}\sum_{i=1}^{n} p_i q_j \log p_i - \sum_{j=1}^{m}\sum_{i=1}^{n} p_i q_j \log q_j$$

$$= \left(-\sum_{j=1}^{m} q_j\right)\left(\sum_{i=1}^{n} p_i \log p_i\right) - \left(\sum_{i=1}^{n} p_i\right)\left(\sum_{j=1}^{m} q_j \log q_j\right)$$

$$H_{mn}(X,Y) = \sum_{j=1}^{m} q_j H(X) + \sum_{i=1}^{n} p_i H(Y)$$

eşitliği elde edilir ve $\sum_{j=1}^{m} q_j = 1$ ve $\sum_{i=1}^{n} p_i = 1$ olduğu dikkate alınırsa,

$$H_{mn}(X,Y) = H_n(X) + H_m(Y) \tag{2.13}$$

sonucuna varılır. Böylelikle entropinin "toplamsallık özelliği" ispatlanmış olur. Bir başka deyişle X ve Y rassal değişkenleri bağımsız iken bileşik entropi, bu iki rassal değişkenin entropileri toplamına eşittir. Aynı şekilde r sayıda karşılıklı bağımsız (X_1,X_2,...,X_r) vektör rassal değişkenlerinin bileşik entropisi, bunların ayrı ayrı entropileri toplamına eşittir ve aşağıdaki (2.14) eşitliğindeki gibi ifade edilir.

$$H_{mn}(X_1, X_2, \ldots, X_r) = \sum_{k=1}^{r} H(X_k) \tag{2.14}$$

2.3. Koşullu Entropi

Bir önceki konuda bahsedilen bağımsızlık koşulu sağlanmadığı takdirde koşullu entropi söz konusu olacaktır. Y rassal değişkeni için y_j değeri gözlendiğinde X rassal değişkeni hakkındaki belirsizliğin miktarını göstermek için $H(X \mid y_j)$ entropisi tanımlanır.

$$H_{mn}(X \mid y_j) = -\sum_{i=1}^{n} p(x_i \mid y_j) \log p(x_i \mid y_j) \tag{2.15}$$

Eşitlik (2.15) X rassal değişkeninin kısmi entropisi olarak ifade edilir.

Şimdi $H_{mn}(X \mid Y)$ koşullu entropisini, yani Y rassal değişkeninin tüm değerleri gerçeklenebilir olduğunda X rassal değişkeninin belirsizlik miktarı hesaplanabilir. Bu ifade ise (2.15) eşitliğinde tüm y_j 'lerin ortalaması ile aşağıdaki gibi elde edilir (Karmeshu 2003).

$$H_{mn}(X \mid Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} q(y_j)\,p(x_i \mid y_j)\log p(x_i \mid y_j)$$

bu eşitlikte koşullu olasılık tanımından $p_{ij} = p(x_i, y_j) = p(x_i \mid y_j)q(y_j)$ yazılabilir.

Bu durumda koşullu entropi,

$$H_{mn}(X \mid Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log p(x_i \mid y_j)$$

$$H_{mn}(X \mid Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}\,\log p(x_i \mid y_j) \tag{2.16}$$

şeklinde yazılabilir. Aynı şekilde,

$$H_{nm}(Y \mid X) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log p(y_j \mid x_i) \tag{2.17}$$

olarak tanımlanabilir.

Eşitlik (2.16) ve (2.17) sırasıyla X'in ve Y'nin koşullu entropisini verir.

Teorem 2.1. (Zincir Kuralı): İki rassal değişkenin bileşik entropisi, birinin koşullu entropisi ile diğerinin entropisi toplamına eşittir.

İspat: Koşullu olasılık formülü

$$P(X,Y) = P(X)P(Y \mid X)$$

dikkate alınırsa ve her iki tarafın logaritması alındığında,

$$\log P(X,Y) = \log P(X) + \log P(Y \mid X)$$

eşitliği elde edilir.

Son eşitlikte her iki yanın beklenen değeri alındığında da

$$\begin{aligned} H(X,Y) &= E\{-\log P(X) - \log P(Y \mid X)\} \\ &= E\{-\log P(X)\} + E\{-\log P(Y \mid X)\} \end{aligned}$$

$$H_{mn}(X,Y) = H_n(X) + H_{mn}(Y \mid X) \tag{2.18}$$

sonucu ile zincir kuralı ispatlanmış olur. Bu teoremin bazı sonuçları elde edilebilir.

Sonuç 1) X ve Y rassal değişkenleri bağımsız ise yani $P(Y \mid X) = P(Y)$ ise bu takdirde $H_{mn}(Y \mid X) = H_m(Y)$ olur ve (2.18) eşitliği

$$H_{mn}(X,Y) = H_n(X) + H_m(Y)$$

olarak (2.13) eşitliğine dönüşür.

Sonuç 2) Eşitlik (2.18) aynı zamanda

$$H_{mn}(X,Y) = H_m(Y) + H_{mn}(X \mid Y) \tag{2.19}$$

gibi de yazılabilir. Eşitlik (2.18) ile (2.19)'dan

$$H_n(X) + H_{mn}(Y \mid X) = H_m(Y) + H_{mn}(X \mid Y) \tag{2.20}$$

elde edilir.

Sonuç 3) X rassal değişkeni Y rassal değişkenini tamamen açıklıyorsa, bir başka deyişle $P(Y \mid X) = 1$ ise X var olduğunda Y kesin vardır ve bu durumda $H_{mn}(X,Y) = H_n(X)$ olur.

Örnek: X ve Y sistemlerinin bileşik olasılıkları Çizelge 2.3.'de verilmiş iken $H(Y \mid X)$ ve $H(X \mid Y)$ değerleri hesaplansın.

Çizelge 2.3. X ve Y değişkenlerinin bileşik olasılıkları

	y_1	y_2	y_3	P_x
x_1	0.1	0	0	0.1
x_2	0.2	0.3	0.2	0.7
x_3	0	0	0.2	0.2
P_y	0.3	0.3	0.4	1

$P(y_j \mid x_i) = \dfrac{P(x_i, y_j)}{P(x_i)}$ olduğu dikkate alınarak koşullu olasılıklar Çizelge 2.4.'deki gibi hesaplanır.

Çizelge 2.4. X ve Y değişkenlerinin koşullu olasılıkları

	y_1	y_2	y_3
x_1	$P(y_1 \mid x_1) = {}^{0.1}\!/_{0.1} = 1$	$P(y_2 \mid x_1) = {}^{0}\!/_{0.1} = 0$	$P(y_3 \mid x_1) = {}^{0}\!/_{0.1} = 0$
x_2	$P(y_1 \mid x_2) = {}^{0.2}\!/_{0.7} = \frac{2}{7}$	$P(y_2 \mid x_2) = {}^{0.3}\!/_{0.7} = \frac{3}{7}$	$P(y_3 \mid x_2) = {}^{0.2}\!/_{0.7} = \frac{2}{7}$
x_3	$P(y_1 \mid x_3) = {}^{0}\!/_{0.2} = 0$	$P(y_2 \mid x_3) = {}^{0}\!/_{0.2} = 0$	$P(y_3 \mid x_3) = {}^{0.2}\!/_{0.2} = 1$

Koşullu olasılık formülü ile

$$H(Y \mid X) = -\sum_{i=1}^{3} p_i \left\{ \sum_{j=1}^{3} \frac{P(x_i, y_j)}{P(x_i)} \log\left(\frac{P(x_i, y_j)}{P(x_i)} \right) \right\}$$

olarak hesaplanır. Burada entropi fonksiyonunu ϑ ile işaret edilirse işlemler aşağıdaki gibi olur ve

$$= -\sum_{i=1}^{3} p_i \left\{ \sum_{j=1}^{3} \vartheta\left(\frac{P(x_i, y_j)}{P(x_i)} \right) \right\}$$

$$= -\begin{Bmatrix} 0.1\big(\vartheta(1) + \vartheta(0) + \vartheta(0)\big) \\ + 0.7\left(\vartheta\left(\frac{0.2}{0.7}\right) + \vartheta\left(\frac{0.3}{0.7}\right) + \vartheta\left(\frac{0.2}{0.7}\right) \right) \\ + 0.2\big(\vartheta(0) + \vartheta(0) + \vartheta(1)\big) \end{Bmatrix}$$

$$H(Y \mid X) = -0.7\left\{ \vartheta\left(\frac{0.2}{0.7}\right) + \vartheta\left(\frac{0.3}{0.7}\right) + \vartheta\left(\frac{0.2}{0.7}\right) \right\}$$

$$= 0.3280 \text{ bit}$$

olarak koşullu entropi değeri bulunur. Benzer işlemler yaparak ta $H(X \mid Y)$ 'de elde edilebilir.

2.4.Nispi Entropi ve Karşılıklı İnformasyon

Bir X rassal değişkeninin entropisi $H(X)$, bu değişkenin hangi değer alacağının belirsizliğini ifade eden bir büyüklüktür. X rassal değişkeni hakkında önemli bilgiler elde edilmesi, $H(X)$'in yani belirsizliğin azalması anlamına gelir. Bu açıdan dikkate alındığında informasyon, belli bilgiler sayesinde $H(X)$ entropisinin değişiminin bir ölçütüdür denilebilir. Böyle olduğunda informasyon pozitif veya negatif olabilir. İnformasyonun negatif olması sistemin entropisinin artması anlamına gelir. Önemli olan sistemin belirsizliğini aşmak olduğu için informasyonun pozitif olması söz konusudur.

Yukarıda bahsedilenler şöyle de ifade edilebilir. X rassal değişkeni hakkında informasyon I_X ile işaret edilsin. Rassal değişkenin entropisi $H(X)$ iken, belli etkiler sayesinde sonraki entropisi $\breve{H}(X)$ olsun. $H(X) - \breve{H}(X) = I_X$'tir. Buradan görülür ki eğer $\breve{H}(X) > H(X)$ ise yani fiziksel sistemin belirsizliği artmış ise $I_X < 0$ olacaktır ki bu durumda negatif informasyon söz konusudur. Eğer $\breve{H}(X) < H(X)$ ise $I_X > 0$ olacaktır. Başka deyişle sistemin entropisi azalmış ise informasyon pozitiftir. Ancak $\breve{H}(X) = 0$ ise $H(X) = I_X$ olacaktır. Bu durumda bir X rassal değişkeninin belirsizliğini aşmak için gereken informasyon miktarı, sayısal olarak $H(X)$ entropisine eşittir.

Bu durumda

$$I_X = H(X) = -\sum_{i=1}^{n} p_i \log p_i \qquad (2.21)$$

eşitliği tam informasyonu yani tüm belirsizliği aşmak için gereken bilgi miktarını ifade eder. X rassal değişkeninin x_i değerini alma olasılığı $P(X=x_i)=p_i$ iken;

$$I_{x_i} = -\log p_i \qquad (2.22)$$

ile de kısmi informasyon tanımlanır.

Eşitlik (2.22)' den de görülüyor ki p_i küçük bir değere sahip olduğunda I_{x_i} büyük bir değere sahip olacaktır. Bu şu anlama geliyor ki gerçekleşme olasılığı çok küçük olan olayın içerdiği informasyon miktarı, gerçekleşme olasılığı büyük olan olayın içerdiği informasyon miktarından çok daha büyüktür.

İnformasyon kavramının tanımından ve formülünden anlaşıldığı gibi informasyon da entropi kavramı ile aynı birimle, bit ile ölçülür.

2.4.1.Nispi Entropi

Nispi entropi iki olasılık dağılımı arasındaki uzaklığın bir ölçümü olarak ifade edilebilir. İstatistikte bu terim likelihood oranının logaritmasının beklenen değeri olarak ortaya çıkar. Nispi entropi $D(p \| q)$ ile gösterilir ve gerçek dağılım p iken varsayılan q dağılımının yetersizliğinin bir ölçümüdür. Örneğin gerçek dağılım p bilinseydi, ortalama tanım uzunluğu $H(P)$ ile kodlanabilirdi. Ancak gerçek dağılım yerine varsayılan bir q dağılımı için bir kod kullanıldıysa, rassal değişkeni tanımlamak için $H(P)+D(p \| q)$ bit kadar bilgiye ihtiyaç olacaktır (Cover ve Thomas 1991).

X rassal değişkeni üzerinde tanımlanmış $p(x)$ ve $q(x)$ olasılık dağılımları E_P, p 'ye göre beklenen değeri ifade etmekte iken, iki dağılım arasındaki uzaklık

$$D(p \| q) = \sum_{x \in \aleph} p(x) \log \frac{p(x)}{q(x)} \tag{2.23}$$

$$= E_P \log \frac{p(x)}{q(x)}$$

olarak tanımlanır.

Nispi entropi negatif değer almaz ve sadece p=q halinde sıfır olur. Bununla birlikte nispi entropi dağılımlar arasındaki uzaklık olarak tanımlansa da simetri özelliği yoktur ve üçgen eşitsizliğini sağlamaz (Cover ve Thomas 1991).

Örnek: $x_1 = 1$ ve $x_2 = 0$ değerlerini alan X rassal değişken üzerinde tanımlanmış iki olasılık dağılımı p ve q olsun.

p(0)=1-r, p(1)=r; q(0)=1-s, q(1)=s ise iki dağılım arasındaki uzaklığın ölçümü olan $D(p \| q)$ nispi entropisi hesaplandığında:

$$D(p \| q) = (1-r)\log\left(\frac{1-r}{1-s}\right) + r\log\left(\frac{r}{s}\right)$$

ve

$$D(q \| p) = (1-s)\log\left(\frac{1-s}{1-r}\right) + s\log\left(\frac{s}{r}\right)$$

olarak elde edilir. Eğer $r = s$ ise $D(p \| q) = D(q \| p) = 0$ olur.

$r = \frac{1}{2}$, $s = \frac{1}{4}$ ise;

$$D(p \| q) = \frac{1}{2}\log\left(\frac{1/2}{3/4}\right) + \frac{1}{2}\log\left(\frac{1/2}{1/4}\right)$$

$$= 1 - \frac{1}{2}\log 3$$

$$= 0.2075 \text{ bit iken,}$$

$$D(q \| p) = \frac{3}{4}\log\left(\frac{3/4}{1/2}\right) + \frac{1}{4}\log\left(\frac{1/4}{1/2}\right)$$

$$= \frac{3}{4}\log 3 - 1$$

$$= 0.1887 \text{ bit}$$

olarak hesaplanır.

Genelde $D(p \| q) \neq D(q \| p)$ yazılabilir (Cover ve Thomas 1991).

2.4.2.Karşılıklı İnformasyon

Karşılıklı informasyon, bir rassal değişkenin diğer rassal değişken hakkında içerdiği informasyon miktarının bir ölçümü olarak tanımlanabilir (Cover ve Thomas 1991). Bir başka deyişle, Y rassal değişkeninin gözlenmesiyle X rassal değişkeninin entropisinde gerçekleşen azalma miktarıdır. X rassal değişkeninin belirsizliğinin aşılması için gereken bilgi miktarı $I(X) = H(X)$ kadardır. Y rassal değişkenin X hakkında içerdiği entropi (informasyon) miktarı da $H(X \mid Y)$ olacaktır. O halde X yerine Y rassal değişkeni gözlendiğinde X'de kalan entropi miktarı ya da X'i tanımlamak için gereken informasyon miktarı,

$$I(X;Y) = H(X) - H(X \mid Y) \tag{2.24}$$

olarak ifade edilir. Y'nin gözlenmesiyle, X'in entropisi $H(X)$'teki azalma, $H(X \mid Y)$ koşullu entropisi kadar olacağından (2.24) formülü kalan entropi miktarını ifade eder.

Teorem 2.2. Y rassal değişkeninin X rassal değişkenine verdiği informasyon miktarı, X rassal değişkeninin Y rassal değişkenine verdiği informasyon miktarına eşittir:

$$I_{Y \to X} = I_{X \to Y}.$$

İspat: Koşullu entropi konusunda elde edilen (2.20) eşitliği ele alındığında,

$$H(X) + H(Y \mid X) = H(Y) + H(X \mid Y) \tag{2.25}$$

$$H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) \tag{2.26}$$

$$I_{Y \to X} = I_{X \to Y}$$

olduğu görülür.

2.4.3.Karşılıklı İnformasyon Formülünün İncelenmesi

Eşitlik (2.19)'da $H(X \mid Y)$ çekilerek (2.24)'de yerine yazıldığında;

$$H(X,Y) = H(Y) + H(X \mid Y)$$

$$H(X \mid Y) = -H(Y) + H(X,Y) \tag{2.27}$$

$$I_{X \leftrightarrow Y} = H(X) + H(Y) - H(X,Y) \tag{2.28}$$

karşılıklı informasyon formülü elde edilir.

Sonuç.1 X ve Y rassal değişkenleri bağımsız ise;

$$H(X,Y) = H(X) + H(Y)$$

olduğundan

$$I_{X \leftrightarrow Y} = 0$$

sonucuna varılır.

Sonuç.2 Y rassal değişkeni X rassal değişkenini tamamen açıklıyor ise ;

$$H(X,Y) = H(X) = H(Y)$$

olacağından

$$I_{X \leftrightarrow Y} = H(X) = H(Y)$$

elde edilir.

Karşılıklı informasyon formülü beklenen değer ile de ifade edilebilir. P(x,y) bileşik olasılık fonksiyonu, P(x) ve P(y) sırasıyla X ve Y rassal değişkenlerine ait marjinal olasılık fonksiyonu olmak üzere karşılıklı informasyon formülü:

$$\begin{aligned} I_{X \leftrightarrow Y} &= H(X) + H(Y) - H(X,Y) \\ &= -\sum_x p(x)\log p(x) - \sum_y p(y)\log p(y) + \sum_{x,y} p(x,y)\log p(x,y) \\ &= E\{-\log P(X)\} + E\{-\log P(Y)\} + E\{\log P(X,Y)\} \end{aligned}$$

veya

$$I_{X \leftrightarrow Y} = E\left\{\log \frac{P(X,Y)}{P(X)P(Y)}\right\} \tag{2.29}$$

şeklinde ifade edilebilir. Buradan (2.29) eşitliği

$$I_{X \leftrightarrow Y} = \sum p(x,y)\log \frac{p(x,y)}{p(x)p(y)} \tag{2.30}$$

olarak belirtildiğinde görülüyor ki karşılıklı informasyon $I_{X \leftrightarrow Y}$, bileşik dağılım $p(x,y)$ ile çarpım dağılım $p(x).p(y)$ arasındaki nispi entropiyi ifade eder.

Eşitlik (2.30)'da $P(X = x_i, Y = y_j) = p_{ij}$, $P(X = x_i) = p_i$ $i = \overline{1,n}$, $P(Y = y_j) = q_j$ $j = \overline{1,m}$ olduğu dikkate alınırsa;

$$I_{X \leftrightarrow Y} = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log \frac{p_{ij}}{p_i q_j} \tag{2.31}$$

şeklinde ifade edilebilen eşitlikte,

$$p(x_i \mid y_j) = \frac{p_{ij}}{q_j}$$

$$p_{ij} = q_j p(x_i \mid y_j) \tag{2.32}$$

yerine yazıldığı takdirde,

$$I_{X \leftrightarrow Y} = \sum_{i=1}^{n}\sum_{j=1}^{m} q_j p(x_i \mid y_j) \log \frac{q_j p(x_i \mid y_j)}{p_i q_j}$$

$$I_{X \leftrightarrow Y} = \sum_{j=1}^{m} q_j \underbrace{\sum_{i=1}^{n} p(x_i \mid y_j) \log \frac{p(x_i \mid y_j)}{p_i}}_{I_{y_j \to X}} \tag{2.33}$$

elde edilir ki formülasyondaki $I_{y_j \to X}$ ifadesi Y rassal değişkeninin y_j değerini almasının, X rassal değişkeni hakkında içerdiği kısmi informasyonu ifade eder. Karşılıklı informasyon ile kısmi informasyon arasındaki ilişki

$$I_{X \leftrightarrow Y} = \sum_{j=1}^{m} q_j I_{y_j \to X} \tag{2.34}$$

ile gösterilir.

3. ENTROPİ OPTİMİZASYON PRENSİPLERİ

Bu bölümde Entropi Optimizasyon prensipleri; Shannon entropi ölçütünü maksimum yapan Jaynes'in maksimum entropi prensibi (MaxEnt) ile Kullback-Leibler'in çapraz entropi ölçütünün bazı özellikleri ve bu ölçütü minimize eden Kullback'in minimum çapraz entropi prensibi (MinxEnt) ayrıntılarıyla ele alınarak aralarındaki ilişkiye değinilmiştir.

3.1. Jaynes'in Maksimum Entropi Prensibi:MaxEnt

Geniş kapsamlı bir ölçek üzerinde yapılan ölçümlere dayandırılan sistemin durumlarını tanımlamak için bir olasılık dağılımı türetmenin metodu, bağımsız olarak Ingarden, Jaynes ve Kullback tarafından tespit edildiyse de başlıca öncü Jaynes'tir. Jaynes, mühendislik ve fizikteki problemlerin bir çok çeşidine bu sonuçları uygulamıştır. Bu metottan genellikle maksimum entropi prensibi (MaxEnt), maksimum belirsizlik prensibi, maksimum entropi metodu ya da Jaynes'in formülü olarak söz edilir. Maksimum belirsizlik prensibi olarak bahsetmek tercih edilse de, kullanım alanının geniş olması nedeniyle maksimum entropi prensibi deyimi kullanılır.

Bilgi olarak sadece ortalama değerleri verilmiş olduğunda uyumlu dağılımlar genelde sonsuz sayıdadır. MaxEnt, Shannon entropi ölçütünü maksimum yapan ve ortalama değer kısıtlarıyla eşzamanlı tutarlı dağılımı seçmeyi önerir. MaxEnt; her bir olasılığın negatif olmaması ve olasılıkların toplamının bir olması dışında rassal değişken hakkında hiç bir şey bilmediğimizde, üniform dağılımı bizim bilgimizin en tatmin edici temsilcisi olduğunu varsayan Laplece'ın ünlü "yetersiz sebep ilkesi"nin doğal bir uzantısıdır. Bu bölümde MaxEnt prensibinin matematiğine ayrıntılı bir şekilde değinilecektir (Kapur ve Kesavan 1992).

$x_1, x_2, \ldots, x_n$ değerlerini $p_1, p_2, \ldots, p_n$ uygun olasılıkları ile alan bir X rassal değişkeni ele alınsın. Bu rassal değişkene bağlı $g_1(X), g_2(X), \ldots, g_m(X)$ fonksiyonlarının beklenen değerleri

$$E\{g_r(x_i)\} = p_1 g_r(x_1) + p_2 g_r(x_2) + \ldots + p_n g_r(x_n) = \eta_r, \quad r = 1,2,\ldots,m \qquad (3.1)$$

şeklinde ifade edilsin. Burada η_r, X rassal değişkenine bağlı $g_r(X)$ fonksiyonun beklenen değerini gösteren bir sayı olmak üzere, matematiksel beklentiler şeklinde verilen koşullar altında $p_1, p_2, \ldots, p_n$ olasılıklarının bulunması MaxEnt prensibi ile yapılabilir. Eşitlik (3.1) ile verilen kısıtlar,

$$\sum_{i=1}^{n} p_i g_r(x_i) = \eta_r \qquad r = 1,2,...,m \tag{3.2}$$

gibi gösterilebilir. Ayrıca p_i'ler arasında

$$\sum_{i=1}^{n} p_i = 1 \tag{3.3}$$

şeklinde bir doğal ilişki de mevcuttur. Bunun yanında

$$p_i \geq 0, \qquad i = 1,2,...,n \tag{3.4}$$

eşitsizliğini Shonnon'un entropi ölçütü otomatik olarak sağlayacaktır.

Böylece $p_1, p_2, \ldots, p_n$ arasında m+1 tane ilişki bir başka deyişle, kısıt elde edilmiş olur. Eğer $m+1 < n$ ise genelde Eşitlik (3.2) ve (3.3) $p_1, p_2, \ldots, p_n$'lerin tek bir değeri için belirlenemez. Gerçektende p_i'lerin $n-m-1$ tanesine herhangi keyfi bir değer verilebilir ve (3.2) ve (3.3) eşitliklerini kullanarak kalan $m+1$ olasılık için çözülebilir. Böylece eşitlik (3.2) ve (3.3)'ün sonsuz sayıda çözümüne sahip olunur ve elbette bu çözümlerden sadece her $p_i > 0$ olanları alınır.

Sonsuz sayıda çözümler arasından makul bir seçim yapmak için yaygın fikir ve bilimsel prensiplere dayanan argümanlara başvurulur:

- Doğruyu ve yalnız doğruyu söylemek
- Verilen tüm bilgileri kullanmak ve elde edilemeyen bilgiler hakkında varsayımlarda bulunmaktan kaçınmak.

O nedenle (3.2) ve (3.3) eşitlikleri kullanılmalı fakat mümkün olan eksik bilgi hakkında varsayımlar yapılmamalıdır. Genelde bu iki eşitlik ile tutarlı olan sonsuz sayıda olasılık dağılımları vardır ve bunların her biri farklı bir belirsizlik miktarına sahiptir.

Jaynes bu eşitliklerle tutarlı olan tüm dağılımlar arasından maksimum entropiye sahip olanının seçilmesi gerektiğini önermiştir. Onun yürüttüğü mantık oldukça basittir. Eğer maksimum entropiden daha az entropisi olan bir dağılım seçilirse, entropideki bu indirgeme bilinçli ya da bilinçsiz kullanılmış olan bazı ilave edilen bilgiden gelmiş olabilir. Ancak böyle bilgiler verilmediğinden onları kullanmak doğru olmayacaktır. Böylece sadece maksimum entropiye sahip olanı kullanılmalıdır (Kapur ve Kesavan 1992).

Güdülen amaç belirsizliği indirgemek iken maksimum belirsizliğe sahip olan dağılımı seçmek bir paradoksmuş gibi görünebilir ancak öyle değildir. X rassal değişkenine bağlı verilen her bir koşul, rassal değişken hakkında belli miktarda informasyon içerir. Ancak bu koşullar X rassal değişkeni hakkında bulunan belirsizliği aşmaya yetmez. Bu belirsizlik bize verilmeyen informasyona uygun olandır ve bunun hakkında kesinlik yoktur. Bu nedenle verilmiş rassal değişkenin entropi fonksiyonunu bilinen koşullar altında maksimize etmek gerekir ki, belirsizliği yok etmek (aşmak) için gereken informasyonun maksimum miktarı bilinsin.

Matematiksel olarak yapılması gereken ise;

$$H(p) = -\sum_{i=1}^{n} p_i \ln p_i \tag{3.5}$$

entropi fonksiyonunu,

$$\sum_{i=1}^{n} p_i g_r(x_i) = \sum_{i=1}^{n} p_i g_{ri} = \eta_r \qquad r = 1,2,...,m \tag{3.6}$$

ve

$$\sum_{i=1}^{n} p_i = 1 \tag{3.7}$$

koşulları altında maksimum yapan p_i'leri bulmaktır.

Koşullu ekstremum çözümüne göre verilmiş koşulları $\lambda_1, \lambda_2, \ldots, \lambda_{m+1}$ Lagrange çarpanları ile çarparak , yardımcı L fonksiyonu

$$L \equiv -\sum_{i=1}^{n} p_i \ln p_i - \sum_{r=1}^{m} \lambda_r \left(\sum_{i=1}^{n} g_r(x_i) p_i \right) - \lambda_{m+1} \sum_{i=1}^{n} p_i \tag{3.8}$$

şeklinde kurulur. Ekstremumun varlık teoremine göre yardımcı (3.8) fonksiyonunun p_i'lere göre kısmi türevlerinin sıfıra eşit olması gerekir. Böylelikle $p_1, p_2, \ldots, p_n$ değişkenleri $\lambda_1, \lambda_2, \ldots, \lambda_{m+1}$ Lagrange çarpanlarına bağlı elde edilir. $\lambda_1, \lambda_2, \ldots, \lambda_{m+1}$ çarpanları ise (3.6) eşitliği yardımıyla bulunur.

L yardımcı fonksiyonunun p_i'lere göre kısmi türevleri;

$$L_{p_i} = -\left(\ln p_i + \frac{1}{p_i} p_i\right) - \lambda_1 g_1(x_i) - \lambda_2 g_2(x_i) - \cdots - \lambda_m g_m(x_i) - \lambda_{m+1}$$

$$= -(\ln p_i + 1) - \sum_{r=1}^{m} \lambda_r g_r(x_i) - \lambda_{m+1} \qquad i = 1,2,\ldots,n \qquad (3.9)$$

elde edilir. Eşitlik (3.9)'un sağ tarafı sıfıra eşitlenerek,

$$\ln p_i = -\sum_{r=1}^{m} \lambda_r g_r(x_i) - \lambda_{m+1} - 1$$

bulunur. Buradan

$$p_i = \exp\left\{-\sum_{r=1}^{m} \lambda_r g_r(x_i) - \lambda_{m+1} - 1\right\} \qquad (3.10)$$

veya

$$p_i = \underbrace{e^{-\lambda_{m+1}-1}}_{A} \exp\left\{-\sum_{r=1}^{m} \lambda_r g_r(x_i)\right\}$$

$$p_i = A \exp\left\{-\sum_{r=1}^{m} \lambda_r g_r(x_i)\right\} \qquad (3.11)$$

ile $p_1, p_2, \ldots, p_n$ değişkenleri, $A, \lambda_1, \lambda_2, \ldots, \lambda_m$ cinsinden yazılmış olur. Bilinmeyen $A, \lambda_1, \lambda_2, \ldots, \lambda_m$ değerleri ise (3.6) ve (3.7) eşitliklerinden bulunabilir.

Eşitlik (3.11) eşitlik (3.7)' de yerine yazıldığında, bir başka deyişle $i = 1,2,\ldots,n$ değerleri vererek toplandığında

$$1 = p_1 + p_2 + p_3 + \cdots + p_n$$

olduğundan

$$1 = A\left(\exp\left\{-\sum_{r=1}^{m} \lambda_r g_r(x_1)\right\} + \exp\left\{-\sum_{r=1}^{m} \lambda_r g_r(x_2)\right\} + \cdots + \exp\left\{-\sum_{r=1}^{m} \lambda_r g_r(x_n)\right\}\right)$$

ifadesi elde edilir. Bu ifade aşağıdaki gibi

$$1 = A\left(\sum_{i=1}^{n} \exp\left\{-\sum_{r=1}^{m} \lambda_r g_r(x_i)\right\}\right)$$

veya

$$\frac{1}{A} = Z = \sum_{i=1}^{n} \exp\left\{-\sum_{r=1}^{m} \lambda_r g_r(x_i)\right\} \tag{3.12}$$

şeklinde yazılır. Eşitlik (3.12)'de λ_r 'lere göre kısmi türevler alındığında:

$$\frac{\partial Z}{\partial \lambda_r} = \left(-g_r(x_1)\right)\exp\left\{-\sum_{r=1}^{m} \lambda_r g_r(x_1)\right\} + \left(-g_r(x_2)\right)\exp\left\{-\sum_{r=1}^{m} \lambda_r g_r(x_2)\right\}$$

$$+\ldots + \left(-g_r(x_n)\right)\exp\left\{-\sum_{r=1}^{m} \lambda_r g_r(x_n)\right\};$$

$$\frac{\partial Z}{\partial \lambda_r} = -\sum_{i=1}^{n} g_r(x_i)\exp\left\{-\sum_{r=1}^{m} \lambda_r g_r(x_i)\right\}$$

bulunur. Bu eşitlikte her iki taraf A ile çarpıldığında,

$$A\frac{\partial Z}{\partial \lambda_r} = -A\sum_{i=1}^{n} g_r(x_i)\exp\left\{-\sum_{r=1}^{m} \lambda_r g_r(x_i)\right\},$$

$$A\frac{\partial Z}{\partial \lambda_r} = -\sum_{i=1}^{n} g_r(x_i)\underbrace{A\exp\left\{-\sum_{r=1}^{m} \lambda_r g_r(x_i)\right\}}_{p_i},$$

$$-A\frac{\partial Z}{\partial \lambda_r} = \underbrace{\sum_{i=1}^{n} g_r(x_i)\,p_i}_{\eta_r}$$

elde edilir.

Eşitlik (3.6) koşulları dikkate alınırsa;

$$-A\frac{\partial Z}{\partial \lambda_r} = \eta_r \qquad r = 1,2,\ldots,m \tag{3.13}$$

veya

$$-\frac{1}{Z}\frac{\partial Z}{\partial \lambda_r} = \eta_r \tag{3.14}$$

olduğu ortaya çıkar ve m tane λ_r çarpanları (3.6) yada (3.14) denklem sisteminden belirlenir (Papoulis 1991). Bu çarpanların elde edilmesi dördüncü bölümde ayrıntılarıyla ele alınmıştır.

3.2. Kullback'in Minimum Çapraz Entropi Prensibi:MinxEnt

Kullback'in minimum çapraz entropi prensibi (MinxEnt), bir entropi optimizasyon prensibidir ve daha önce de ifade edildiği gibi teorik anlamları ve uygulama alanlarının taşıdığı önem açısından MaxEnt prensipsiyle rekabet eder. Bu prensip aynı zamanda minimum doğrultulmuş sapma prensibi ya da minimum informasyon ayırma prensibi olarak da bilinir. MinxEnt'de belirsizlik kavramı MaxEnt'te olduğu gibi temel bir rol oynamamasına rağmen, MaxEnt ile ilişkisi kurulabilir. MinxEnt'de vurgulanan konu, bir olasılık dağılımı olan P nin başka bir olasılık dağılımı Q dan olan, "çapraz entropi"si ya da "doğrultulmuş sapma"sıdır. Yapılan bu çalışmada önemli olan bir uzaklık, belirlenen bir olasılık dağılımının, üniform dağılım U'dan olan uzaklığıdır. Bu uzaklığın önemli olmasının sebebi, U'nun belirsizliği en büyük dağılım olmasıdır. Bu yüzden bir dağılım U'ya ne kadar yakın olursa belirsizlik de o kadar çok olur. Böylece bu uzaklık, P dağılımının belirsizliğinin boyutu hakkında fikir verir. Daha sonra ortaya çıkacak ki, aslında bazı kısıtlamalar göz önünde tutularak bu uzaklığın küçültülmesi, aynı kısıtlamalara bağlı kalarak entropinin en yüksek seviyeye çıkartılmasıyla özdeş olacaktır (Kapur ve Kesavan 1992).

Bu prensibin içeriğine geçmeden önce iki kavramın üzerinde durmak gerekir. Birincisi Q'ya bağlı olarak P'nin olasılıksal uzaklığı ya da çapraz entropisi, ikincisi ise önsel olasılık dağılımıdır. Öncelikli olarak bu iki kavram incelenecektir.

Daha önceki konularda bilgi, olasılık dağılımına ait doğrusal momentlerin belirlenen değerleri üzerine kurulmuştu. Hiçbir kısıtlama olmadığı taktirde,

$$-\sum_{i=1}^{n} p_i \ln p_i \tag{3.15}$$

fonksiyonunun

$$\sum_{i=1}^{n} p_i = 1 \tag{3.16}$$

kısıtına bağlı olarak maksimizasyonuyla

$$p_1 = p_2 = \cdots = p_n = \frac{1}{n} \tag{3.17}$$

olasılık değerleri elde edilir.

Buradan anlaşılacağı üzere, hiçbir kısıtlamanın olmadığı durumda, tüm sonuçlar aynı olasılık değerine sahiptir. Bu "*Laplace'ın yetersiz sebep ilkesi*"dir. Bu ilke hiçbir sebebin olmaması durumunda tüm sonuçların eşit olasılıklı kabul edilmesi gerektiğini ifade eder.

Bununla birlikte hiçbir kısıtlamanın olmaması durumunda bile, tüm sonuçların eşit olasılıklı olmadığını gösteren n tane sonucun olasılıklarının, önsel olasılık dağılımını ifade eden $q_1, q_2, \ldots, q_n$ olduğuna inanmayı gerektiren çeşitli sebepler bulunmaktadır. Bu sebepler tecrübeye, önseziye yada teoriye dayalıdır. Önsezi yada tecrübeleri doğrulamak için istenen dağılımın bazı moment değerleri

$$\sum_{i=1}^{n} p_i = 1 \tag{3.18}$$

ve

$$\sum_{i=1}^{n} p_i g_r(x_i) = \eta_r \qquad r = 1,2,...,m \tag{3.19}$$

şeklinde elde edilir.

Olasılık dağılımı $q_1, q_2, \ldots, q_n$ bu kısıtlamalardan Eşitlik (3.18)'i sağlar. Fakat diğer m tane kısıtlamayı sağlayamayabilir. Olasılık dağılımının m tane kısıtlamayı da sağlaması durumunda, yapılan incelemelerin önsezisel tahmini pekiştirdiği sonucuna varılabilir. Öte yandan, bu dağılım bazı yada tüm kısıtlamaları sağlamazsa, önsezi ile araştırmalardan elde edilen bilgi arasında bir çelişki var demektir. Buradaki problem, böyle bir durumda nasıl devam edileceğini belirlemektir.

Eşitlik (3.18) ve (3.19) kısıtlamalarını sağlayan olasılık dağılımları sonsuz sayıdadır. Bunların arasından q ile çatışmayan bir dağılım seçmek gerekir. Böyle bir durumda en iyisi, (3.18) ve (3.19) kısıtlarını sağlayan ve de bir anlamda *q'* ya, yani önsezi ve tecrübelere dayalı dağılıma en yakın dağılımı seçmektir. Bunun için *p*' nin *q*' ya olan yakınlığını bilmek gerekmektedir. Aşağıdaki metrik ölçüler bu amaç için kullanılmaktadır.

$$D_1(p:q) = \left[\sum_{i=1}^{n}(p_i - q_i)^2\right]^{\frac{1}{2}} \tag{3.20}$$

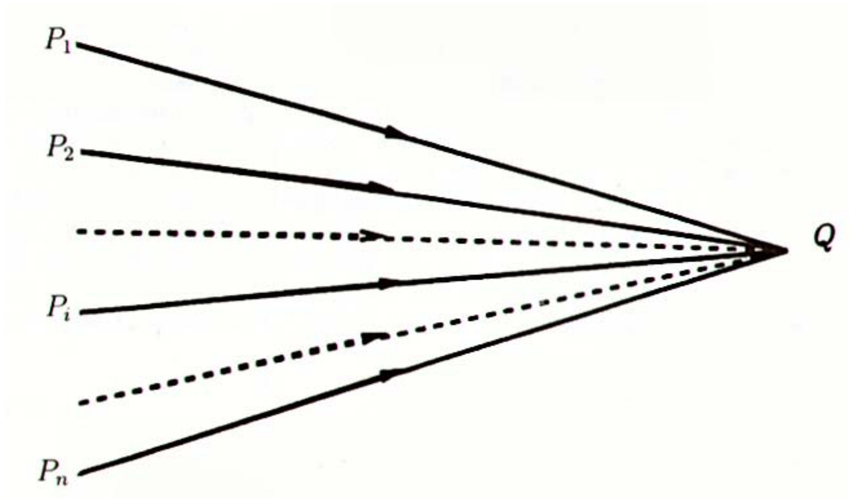
$$D_2(p:q) = \left[\sum_{i=1}^{n}(p_i - q_i)^{2m}\right]^{\frac{1}{2m}}. \tag{3.21}$$

Aslında aşağıdaki koşulları sağlayan herhangi bir diğer metrik $D(p:q)$ da seçilebilir:

1. **Negatif olmama:** $D(p:q) \geq 0$,
2. **Özdeşlik:** Ancak ve ancak p=q ise $D(p:q) = 0$,
3. **Simetri:** $D(p:q) = D(q:p)$ D(p:q) ve
4. **Üçgen eşitsizliği:** $D(p:q) + D(q:r) \geq D(p:r)$.

İlk iki koşul tüm fark ölçümlerinde önemli olmasına rağmen metrik uzaklıklara uygulanan diğer iki koşul aşağıda belirtilen sebeplerden dolayı önemli değildir:

- Eşitlik (3.18) ve de (3.19) kısıtlamalarını sağlayan her bir dağılımın önsel dağılım *q*'dan sapmasıyla yani $p_1, p_2, \ldots, p_n$'in *q*'dan sapmasıyla yani sadece tek yönlü bir farkla ilgilenilmektedir. Sonuç olarak, metrik ölçümlerin simetri koşulları, şu anki konuyla ilgisiz kalmaktadır.
- Şu an için, aynı anda sadece iki dağılımla ilgilenildiğinden 4.koşul da aranmamaktadır.



Şekil 3.1. P_1, P_2,...,P_n'lerin Q'dan doğrultulmuş sapmaları (Kapur ve Kesavan 1992)

Bu sebeplerden dolayı dört koşuldan 3 ve 4 numaralı koşullar göz önünde tutulmayacaktır. Bununla birlikte 1 ve 2 numaralı koşulları sağlayan uzaklık ölçüsüne ek olarak, metrik ölçüsü için yukarıda gerekli gösterilmemiş başka koşullara ihtiyaç duyulabilir. Bu ek ihtiyaçları bulmak için problemin tekrar incelenmesi gerekir.

Sabitleştirilmiş bir q için $D(p:q)$'yu minimum kılmak amaçlanmakta ve bu minimumlaştırma işlemi farklı p_j ler için yapıldıktan sonra global minimum elde edilmek istenmektedir.

Aynı zamanda Lagrange metodu kullanarak Eşitlik (3.18) ve (3.19)'e bağlı olarak D(p:q) mimimize edildiğinde elde edilen hiçbir olasılık sonucunun negatif olmaması istenir. Aksi taktirde, metriğin negatif olmaması kısıtlamalarını çözmek gibi ekstra bir iş daha ortaya çıkacaktır ki böyle bir durumda da minimize etme problemi gereksiz bir şekilde güçleştirilmiş olacaktır.

Minimize etme problemimizi sağlayabilmesi için $D(p:q)$ için iki tane daha koşul bulunmaktadır:

5. $D(p:q)$, $p_1, p_2, \ldots, p_n$ 'in konveks bir fonksiyonu olmalıdır.

6. $D(p:q)$, Lagrange çarpanları metodu kullanarak, kısıtlara bağlı olarak minimize edilirken elde edilen hiçbir olasılık sonucunun negatif olmaması gerekmektedir.

Yukarıda ifade edilen 1, 2, 5 ve 6 numaralı koşulları sağlayan ölçütler bulunmaktadır. Bunların en önemlisi ve en basiti 1951'de Kullback ve Leibler tarafından belirlenmiştir. Bir sonraki konuda bu ölçümün özellikleri incelenecektir.

3.2.1. Kullback-Leibler'in Çapraz Entropi Ölçütünün Özellikleri

Kullback-Leibler'in çapraz entropi ölçütü, bir önceki konuda bahsedilen dört koşulu sağlayan diğer ölçütler arasında en önemli olanıdır. Diğer ölçütler bazı terminolojiler tarafından referans gösterildiği halde, çapraz entropi ya da doğrultulmuş sapma terminolojilerine bağlı kalınacaktır. Kullback-Leibler ölçütü, MaxEnt prensibinde, Shannon ölçütü ile aynı merkezi rolü oynar. Bu önemli ölçütü ayrıntılı bir şekilde açıklamak mümkün olmakla birlikte, bunun yerine sadece önemli özelliklerinin üzerinde durulacaktır.

$p = (p_1, p_2, \ldots, p_n)$ ve $q = (q_1, q_2, \ldots, q_n)$ iki olasılık dağılımı olsun. Kullback-Leibler ölçütü,

$$D(p:q) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \tag{3.22}$$

şeklinde tanımlanır. Burada $q_i = 0$ olduğu durumlarda, bu q_i 'ye bağlı p_i 'nin de sıfır olduğu varsayılarak $0 \ln \frac{0}{0} \stackrel{\Delta}{=} 0$ eşitliği tanımlanır.

Çapraz entropi ölçütünün bazı önemli özellikleri aşağıda sıralanmıştır:

1. $D(p:q)$, $p_1, p_2, \ldots, p_n$ ve $q_1, q_2, \ldots, q_n$ 'in sürekli bir fonksiyonudur.

2. $D(p:q)$, permütasyonel olarak simetriktir; yani sonuçlar çeşitli şekillerde sınıflandırılırsa, $(p_1:q_1)$, $(p_2:q_2), \ldots, (p_n:q_n)$ şeklindeki bu çiftler kendi aralarında yer değiştirse de bu ölçümün değeri değişmez.

3. $D(p:q) \geq 0$ 'dır ve ancak ve ancak $p = q$ iken sıfıra eşit yani yok olur.

Tüm i'ler için $(1+\varepsilon_i) > 0$ olduğu durumlarda, $q_i = p_i(1+\varepsilon_i)$ olsun.

$$\sum_{i=1}^{n} q_i = 1, \ \sum_{i=1}^{n} p_i = 1$$

olduğuna göre

$$\sum_{i=1}^{n} p_i \varepsilon_i = 0 \tag{3.23}$$

eşitliğine de sahip olunur. Bu son eşitlik kullanılarak,

$$\begin{aligned} D(p:q) &= \sum_{i=1}^{n} p_i \ln \frac{p_i}{p_i(1+\varepsilon_i)} \\ &= -\sum_{i=1}^{n} p_i \ln(1+\varepsilon_i) \\ &= \sum_{i=1}^{n} p_i \left(\varepsilon_i - \ln(1+\varepsilon_i)\right) \end{aligned} \tag{3.24}$$

sonucuna ulaşılır.

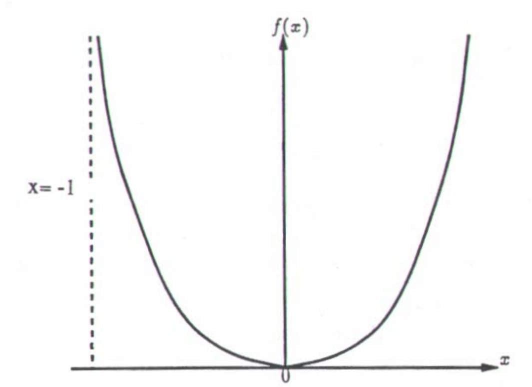
Böylece $f(x) = x - \ln(1+x)$ fonksiyonu ele alındığında;

$1 + x > 0, \ f(0) = 0$,

$$f'(x) = 1 - \frac{1}{1+x} = \frac{x}{1+x}, \ f''(x) = \frac{1}{1+x^2}, \ f'(0) = 0 \tag{3.25}$$

olduğu görülür. Aşağıdaki Şekil 3.2.'de de gösterildiği gibi $x < 0$ iken $f'(x) < 0$ ve de $x > 0$ iken $f'(x) > 0$ olduğu durumlarda $f(x)$ bir konveks fonksiyondur. Bu yüzden $x > -1$ iken $f(x) \geq 0$ ve de buna bağlı olarak $\varepsilon_i > -1$ iken $\varepsilon_i - \ln(1+\varepsilon_i) \geq 0$ 'dır.

Eşitlik (3.24) ele alındığında $D(p:q) \geq 0$ olduğu görülür. Ayrıca yalnız ve yalnız $x = 0$ ise $f(x) = 0$ ve de yalnız ve yalnız $\varepsilon_i = 0$ iken $\varepsilon_i - \ln(1+\varepsilon_i) = 0$ olmaktadır. Başka bir deyişle yalnız ve yalnız her i için $p_i = q_i$ ise, $D(p:q) = 0$ olur. Böylelikle bir önceki konu da bahsedilen 1 ve 2 koşulları sağlanmış olur.



Şekil 3.2. $f(x)$ fonksiyonu grafiği

4. Bir önceki özellikten, $D(p:q)$'nun alabileceği minimum değerin sıfır olduğu sonucu çıkarılabilir.

5. $D(p:q)$ hem p hem de q'nun konveks bir fonksiyonudur. Bu sonuç, global minimumun özelliklerini belirlemek için önem taşımaktadır. Bu özelliği ispatlamak için;

$$D(p:q) = f(p_1, p_2, \ldots, p_n; q_1, q_2, \ldots, q_n)$$

$$= \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i}$$

olsun. $D(p:q)$'nun p'nin konveks bir fonksiyonu olduğunu göstermek için f fonksiyonunun ikinci türevleri aşağıdaki gibidir:

$$\frac{\partial f}{\partial p_i} = 1 + \ln \frac{p_i}{q_i}, \qquad \frac{\partial f}{\partial q_i} = -\frac{p_i}{q_i},$$

$$\frac{\partial^2 f}{\partial p_i^{\,2}} = \frac{1}{p_i}, \qquad \frac{\partial^2 f}{\partial q_i^{\,2}} = \frac{p_i}{q_i^{\,2}},$$

$$\frac{\partial^2 f}{\partial p_i \partial p_j} = 0 \quad \text{ve} \quad \frac{\partial^2 f}{\partial q_i \partial q_j} = 0 \qquad i, j = 1,2,\ldots,n, \quad i \neq j.$$

$p_1, p_2, \ldots, p_n$'e bağlı ikinci türevlerinden oluşan Hessian matrisinin pozitif tanımlı olduğu görülür. Benzer olarak $q_1, q_2, \ldots, q_n$ ile ilgili ikinci türevlerinden oluşan Hessian matrisi de pozitif tanımlı olduğundan $D(p:q)$, hem $p_1, p_2, \ldots, p_n$'in hem de $q_1, q_2, \ldots, q_n$'in konveks bir fonksiyonudur.

6. $D(p:q)$, $p_1, p_2, \ldots, p_n$'in bir konveks fonksiyonu olduğu için, belirlenen bir q için bu fonksiyonun alabileceği maksimum değer, sonucu belli dağılımlardan birinde bulunmalıdır. Maksimum değer;

$$\max(-\ln q_1, -\ln q_2, \ldots, -\ln q_n) = \ln \frac{1}{q_{\min}} \tag{3.26}$$

olmalıdır. Burada $q_{\min} = \min(q_1, q_2, \ldots, q_n)$'dir. Aynı şekilde $D(p:q)$, q'nun bir konveks fonksiyonu olduğu için, bazı q_i değerleri yeterince küçültülerek, belirlenen bir p için alabileceği maksimum değer istenildiği kadar büyültülebilir.

7. Bir önceki bölümde bahsedilen 6. koşulun sağlanmasını garantilemek için;

$$\sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \tag{3.27}$$

fonksiyonu

$$\sum_{i=1}^{n} p_i = 1 \tag{3.28}$$

$$\sum_{i=1}^{n} p_i g_r(x_i) = \eta_r \qquad r = 1,2,\ldots,m \tag{3.29}$$

koşullarına tabi tutularak minimize edilir ve

$$p_i = q_i \exp(-\lambda_0 - \lambda_1 g_{1i} - \lambda_2 g_{2i} - \cdots - \lambda_m g_{mi}) \tag{3.30}$$

elde edilir. Böylece her $p_i > 0$ olur ve çapraz entropi ölçümü için 6 nolu koşul otomatik olarak sağlanmış olur.

Bu noktada p_i' nin üstel fonksiyonel formunun doğasına dikkat etmek gerekir. p_i aynı kısıtlamalar için q_i kere MaxEnt değeri şeklinde ortaya çıkar. Bununla birlikte, MaxEnt'in aksine burada λ_0; $\lambda_1, \lambda_2, \ldots, \lambda_m$'e ve $q_1, q_2, \ldots, q_n$'e bağlıdır.

8. $$\sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \neq \sum_{i=1}^{n} q_i \ln \frac{q_i}{p_i} \tag{3.31}$$

olduğundan, $D(p:q) \neq D(q:p)$ olur.

Bir örnek $D(p:q)$'nun simetrik olmadığını ispatlamak için yeterli olacaktır. Eğer **u** üniform dağılım ise,

$$D(p:u) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{1/n}$$

$$= \ln n + \sum_{i=1}^{n} p_i \ln p_i$$

olmasına rağmen

$$D(u:p) = \sum_{i=1}^{n} \frac{1}{n} \ln \frac{1/n}{p_i}$$

$$= -\ln n - \frac{1}{n} \sum_{i=1}^{n} \ln p_i$$

olarak elde edildiğinden bu iki eşitliğin farklı olduğu görülmektedir.

$D(p:q)$ simetrik olmadığı halde,

$$J(p:q) = D(p:q) + D(q:p)$$

$$= \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} + \sum_{i=1}^{n} q_i \ln \frac{q_i}{p_i} \qquad (3.32)$$

eşitliği simetrik bir ölçümü ifade eder ki, açıkça $J(p:q) = J(q:p)$ olduğu görülür. $J(p:q)$ simetrik çapraz entropi ölçütü ya da simetrik sapma ölçütü olarak adlandırılır.

9. Kullback-Leibler ölçütü toplanabilirlik özelliğini de sağlar. Eğer p_1 ve p_2, q_1 ve q_2 gibi iki bağımsız dağılım ise, p_{ij}'nin ortak olasılığı ifade ettiği durumda,

$$D(p_1 * p_2 : q_1 * q_2) = \sum_{j=1}^{m} \sum_{i=1}^{n} p_{1i} p_{2j} \ln \frac{p_{1i} p_{2j}}{q_{1i} q_{2j}}$$

$$= \sum_{j=1}^{m} p_{2j} \sum_{i=1}^{n} p_{1i} \ln \frac{p_{1i}}{q_{1i}} + \sum_{i=1}^{n} p_{1i} \sum_{j=1}^{m} p_{2j} \ln \frac{p_{2j}}{q_{2j}}$$

$$= D(p_1 : q_1) + D(p_2 : q_2) \qquad (3.33)$$

olarak elde edilir ve toplanabilirlik özelliğini sağladığı ispatlanmış olur.

10. Bağımsız olması çok da gerekli olmayan iki dağılım için;

$$\sum_{j=1}^{m}\sum_{i=1}^{n} p_{ij} \ln \frac{p_{ij}}{q_{ij}} = \sum_{i=1}^{n} p_{i.} \ln \frac{p_{i.}}{q_{i.}} + \sum_{i=1}^{n} p_{i.} \sum_{j=1}^{m} \frac{p_{ij}}{p_{i.}} \ln \frac{p_{ij} / p_{i.}}{q_{ij} / q_{i.}},$$

eşitliği elde edilir Burada $p_{i.} = \sum_{j=1}^{m} p_{ij}$ ve $q_{i.} = \sum_{j=1}^{m} q_{ij}$ 'dir.

11. Şimdi çok temel bir sonuç kanıtlanabilir.Bu sonuç, Kullback tarafından ileri sürülen çapraz entropinin, yani Kullback-Leibler ölçütünün minimize edilmesi ile MaxEnt arasındaki ilişkidir. Üniform dağılım **u** ile *p* arasındaki Kullback-Leibler ölçütü dikkate alındığında,

$$\begin{aligned} D(p:u) &= \sum_{i=1}^{n} p_i \ln \frac{p_i}{1/n} \\ &= \ln n - \left(-\sum_{i=1}^{n} p_i \ln p_i \right) \\ &= \ln n - S_n(p_1, p_2, ..., p_n) \end{aligned} \tag{3.34}$$

eşitliğine ulaşılır ki, buradan entropinin Shannon ölçütünün belirlenen kısıtlamalara bağlı kalınarak maksimize edilmesi ile aynı kısıtlamalara bağlı kalınarak *p*'nin çapraz entropisinin üniform dağılımdan uzaklığının minimize edilmesinin özdeş olduğu görülür.

12. Eğer *q* bir önsel olasılık dağılımı ise *p* Eşitlik (3.28) ve (3.29) kısıtlamalarına bağlı kalınarak çapraz entropiyi minimize eden olasılık dağılımıdır ve r aynı kısıtlamaları doğrulayan başka bir dağılımdır. Bu durumda Eşitlik (3.30) şu sonucu verir:

$$\begin{aligned} D(p:q) &= \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \\ &= \sum_{i=1}^{n} p_i (-\lambda_0 - \lambda_1 g_{1i} - \lambda_2 g_{2i} - \cdots - \lambda_m g_{mi}) \\ &= -\lambda_0 - \lambda_1 \eta_1 - \lambda_2 \eta_2 - \cdots - \lambda_m \eta_m \end{aligned} \tag{3.35}$$

ve

$$D(r:p)=\sum_{i=1}^{n} r_i \ln\frac{r_i}{p_i}=\sum_{i=1}^{n} r_i \ln\frac{r_i}{q_i}\frac{q_i}{p_i}$$

$$=\sum_{i=1}^{n} r_i \ln\frac{r_i}{q_i}-\sum_{i=1}^{n} r_i \ln\frac{p_i}{q_i}$$

$$=D(r:q)+\sum_{i=1}^{n} r_i(-\lambda_0-\lambda_1 g_{1i}-\lambda_2 g_{2i}-\cdots-\lambda_m g_{mi})$$

$$=D(r:q)+(-\lambda_0-\lambda_1\eta_1-\lambda_2\eta_2-\cdots-\lambda_m\eta_m)$$

$$=D(r:q)-D(p:q) \tag{3.36}$$

eşitliği elde edilir ki, böylece

$$D(r:p)+D(p:q)=D(r:q) \tag{3.37}$$

sonucuna varılır.

Bu sonuç, $D(p:q)\leq D(r:q)$ eşitsizliğini ispatlar. Dahası, eşit olma durumu sadece ve sadece $r=p$ olduğunda söz konusu olabilir. Bu yüzden, bu özel durum için, bir dejenere olmuş üçgen eşitsizliği geçerlidir.

3.2.2. Kullback'in MinxEnt Prensibinin Biçimselleşmesi

MinxEnt prensibi, MaxEnt ile aynı çizgiler üzerinde geliştirilmiştir. Bu prensibi MaxEnt'den ayıran özellik, Sahannon ölçütünü maksimize etmek yerine Kullback-Leibler ölçütünü minimize etmekle ilgileniliyor olmasıdır. Her iki durumda da kısıtlamalar aynıdır. MinxEnt prensibinin formülasyonu, elde edilen olasılık dağılımının çözümünün analitik şeklini göstermek için ve de minimize etme probleminin böyle bir çözümünün her zaman var olduğundan emin olmak için önem taşımaktadır. Bunun için

$$\sum_{i=1}^{n} p_i \ln\frac{p_i}{q_i} \tag{3.43}$$

Kullback-Leibler ölçütü

$$\sum_{i=1}^{n} p_i=1 \tag{3.44}$$

ve

$$\sum_{i=1}^{n} p_i g_r(x_i) = \eta_r \qquad\qquad r = 1,2,...,m \tag{3.45}$$

kısıtlarına bağlı olarak minimize edildiğinde,

$$p_i = q_i \exp(-\lambda_0 - \lambda_1 g_{1i} - \lambda_2 g_{2i} - \cdots - \lambda_m g_{mi}) \tag{3.46}$$

sonucu elde edilir. Burada $\lambda_0, \lambda_1, \ldots, \lambda_m$ değerleri, kısıtlar kullanılarak belirlenebilir. Bu aşamada, bu sonuç, daha önce MaxEnt için elde edilen olasılık dağılımının analitik şekliyle karşılaştırılabilir. Eşitlik (3.46)'da, q_i'nin çarpım faktörü hariç, sonucun tamamı MaxEnt'le aynıdır. Aynı zamanda, λ değerleri, $\eta_1, \eta_2, \ldots, \eta_m$'in yanı sıra $q_1, q_2, \ldots, q_n$'e de bağlı olacaktır. Bu sayede, her iki optimizasyon problemleri için bilgisayar algoritmaları geliştirilmesi oldukça kolaylaşmıştır. Aynı zamanda şuna da dikkat etmek gerekir ki, MinxEnt prensibi olasılık dağılımlarının üstel ailesi olarak ortaya çıkar. Bu da, MaxEnt'de olduğu gibi, dağılımların üstel olmayan olasılık ailelerine yol açan olasılıklı sistemlerin bu yaklaşım için uygun olmadığı anlamına gelir. Neyse ki, pratikte bu sorun yaratan bir kısıtlama değildir, fakat istisnaların olabileceği de unutulmamalıdır.

p_i için belirlenen (3.46) eşitliğini (3.44) ve (3.45) kısıtlarında yerine yazmakla,

$$\exp(\lambda_0) = \sum_{i=1}^{n} q_i \exp(-\lambda_1 g_{1i} - \lambda_2 g_{2i} - \cdots - \lambda_m g_{mi}), \tag{3.47}$$

ve

$$\eta_r \exp(\lambda_0) = \sum_{i=1}^{n} q_i g_{ri} \exp(-\lambda_1 g_{1i} - \lambda_2 g_{2i} - \cdots - \lambda_m g_{mi}),$$

elde edilir. Buradan

$$\eta_r = \frac{\sum_{i=1}^{n} q_i g_{ri} \exp(-\lambda_1 g_{1i} - \lambda_2 g_{2i} - \cdots - \lambda_m g_{mi})}{\sum_{i=1}^{n} q_i \exp(-\lambda_1 g_{1i} - \lambda_2 g_{2i} - \cdots - \lambda_m g_{mi})} \tag{3.48}$$

eşitliği çıkarılabilir. Eşitlik (3.47)'den anlaşıldığı gibi λ_0; $\lambda_1, \lambda_2, \ldots, \lambda_m$'in bir fonksiyonudur ve (3.48) eşitliğinden de $\lambda_1, \lambda_2, \ldots, \lambda_m$ değerlerinin $\eta_1, \eta_2, \ldots, \eta_m$'in birer fonksiyonu olduğu anlaşılır.

Eşitlik (3.47)'de diferansiyel almakla;

$$\exp(\lambda_0)\frac{\partial \lambda_0}{\partial \lambda_r} = -\sum_{i=1}^{n} g_{ri} q_i \exp\left(-\sum_{j=1}^{m} \lambda_j g_{ji}\right), \tag{3.49}$$

elde edilir ve burada (3.45) ve (3.46) eşitlikleri kullanılarak,

$$\begin{aligned} \frac{\partial \lambda_0}{\partial \lambda_r} &= -\sum_{i=1}^{n} p_i g_{ri} = -\sum_{i=1}^{n} p_i g_r(x_i) \\ &= -E\left[g_r(X)\right] = -\eta_r, \qquad r = 1,2,\ldots,m \end{aligned} \tag{3.50}$$

sonucuna varılır.

Eşitlik (3.49)'da tekrar λ_r'ye göre diferansiyel alırsak,

$$\exp(\lambda_0)\frac{\partial^2 \lambda_0}{\partial \lambda_r^2} + \exp(\lambda_0)\left(\frac{\partial \lambda_0}{\partial \lambda_r}\right)^2 = \sum_{i=1}^{n} g_{ri}^2 q_i \exp\left(-\sum_{j=1}^{m} \lambda_j g_{ji}\right),$$

veya

$$\begin{aligned} \frac{\partial^2 \lambda_0}{\partial \lambda_r^2} + \left(\frac{\partial \lambda_0}{\partial \lambda_r}\right)^2 &= \sum_{i=1}^{n} p_i g_{ri}^2 = \sum_{i=1}^{n} p_i g_r^2(x_i), \\ &= E[g_r^2(X)] \end{aligned}$$

elde edilir ve son eşitlikte $\frac{\partial \lambda_0}{\partial \lambda_r} = -E\left[g_r(X)\right]$ yazmakla da nihayet

$$\begin{aligned} \frac{\partial^2 \lambda_0}{\partial \lambda_r^2} &= E\left[g_r^2(X)\right] - \left\{E\left[g_r(X)\right]\right\}^2 \\ &= \operatorname{var}\left[g_r(X)\right] \end{aligned} \tag{3.51}$$

eşitliğine varılır. Yine (3.49)'da λ_s'ye göre diferansiyel alındığında,

$$\exp(\lambda_0)\frac{\partial^2 \lambda_0}{\partial \lambda_r \partial \lambda_s} + \exp(\lambda_0)\frac{\partial \lambda_0}{\partial \lambda_r}\frac{\partial \lambda_0}{\partial \lambda_s} = \sum_{i=1}^{n} g_{ri} g_{si} q_i \exp\left(-\sum_{j=1}^{m} \lambda_j g_{ji}\right),$$

$$\frac{\partial^2 \lambda_0}{\partial \lambda_r \partial \lambda_s} + \frac{\partial \lambda_0}{\partial \lambda_r}\frac{\partial \lambda_0}{\partial \lambda_s} = \sum_{i=1}^{n} p_i g_{ri} g_{si} = \sum_{i=1}^{n} p_i g_r(x_i) g_s(x_i)$$

$$= E\left[g_r(X) g_s(X)\right]$$

olarak bulunur ve $\dfrac{\partial \lambda_0}{\partial \lambda_r} = E\left[g_r(X)\right]$ ile $\dfrac{\partial \lambda_0}{\partial \lambda_s} = E\left[g_s(X)\right]$ eşitlikleri yerine yazıldığında da

$$\frac{\partial^2 \lambda_0}{\partial \lambda_r \partial \lambda_s} = E\left[g_r(x) g_s(x)\right] - E\left[g_r(x)\right] E\left[g_s(x)\right]$$

$$= kov\left[g_r(x), g_s(x)\right] \tag{3.52}$$

eşitliği elde edilir.

Böylece özetle (3.50), (3.51) ve (3.52),

$$\frac{\partial \lambda_0}{\partial \lambda_r} = -E\left[g_r(X)\right],$$

$$\frac{\partial^2 \lambda_0}{\partial \lambda_r} = \mathrm{var}\left[g_r(X)\right] \tag{3.53}$$

ve

$$\frac{\partial^2 \lambda_0}{\partial \lambda_r \partial \lambda_s} = kov\left[g_r(X), g_s(X)\right]$$

eşitliklerinden görülmektedir λ_0; $\lambda_1, \lambda_2, \ldots, \lambda_m$'in bir fonksiyonu olarak ele alınırsa, ona uygun Hessian matrisi;

$$\begin{pmatrix} \dfrac{\partial^2 \lambda_0}{\partial \lambda_1^2} & \dfrac{\partial^2 \lambda_0}{\partial \lambda_1 \partial \lambda_2} & \cdots & \dfrac{\partial^2 \lambda_0}{\partial \lambda_1 \partial \lambda_m} \\ \dfrac{\partial^2 \lambda_0}{\partial \lambda_2 \partial \lambda_1} & \dfrac{\partial^2 \lambda_0}{\partial \lambda_2^2} & \cdots & \dfrac{\partial^2 \lambda_0}{\partial \lambda_2 \partial \lambda_m} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 \lambda_0}{\partial \lambda_m \partial \lambda_1} & \dfrac{\partial^2 \lambda_0}{\partial \lambda_m \partial \lambda_2} & \cdots & \dfrac{\partial^2 \lambda_0}{\partial \lambda_m^2} \end{pmatrix} \tag{3.54}$$

gibi olur ve (3.54)'ün yardımıyla da tekrar aşağıdaki gibi yazılabilir.

$$\begin{pmatrix} \operatorname{var}[g_1(X)] & kov[g_1(X), g_2(X)] & \cdots & kov[g_1(X), g_m(X)] \\ kov[g_2(X), g_1(X)] & \operatorname{var}[g_2(X)] & \cdots & kov[g_2(X), g_m(X)] \\ \vdots & \vdots & \ddots & \vdots \\ kov[g_m(X), g_1(X)] & kov[g_m(X), g_2(X)] & \cdots & \operatorname{var}[g_m(X)] \end{pmatrix}. \tag{3.55}$$

Bununla birlikte varyans-kovaryans matrisi daima pozitif tanımlıdır ve aynı şekilde λ_0'ın $\lambda_1, \lambda_2, \ldots, \lambda_m$'ye bağlı ikinci derece kısmi türevlerinin Hessian matrisi de pozitif tanımlı olacaktır. Buna göre λ_0'ın, $\lambda_1, \lambda_2, \ldots, \lambda_m$'nin konveks bir fonksiyonu olduğu kanıtlanmış olur. D(p:q)'nun minimum değeri;

$$D_{\min} = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i}$$

ölçütünde (3.46) eşitliği dikkate alınırsa,

$$= \sum_{i=1}^{n} p_i (-\lambda_0 - \lambda_1 g_{1i} - \lambda_2 g_{2i} - \cdots - \lambda_m g_{mi})$$

elde edilir, burada da (3.45) kısıtı göz önüne alınırsa,

$$D_{\min} = -\lambda_0 - \lambda_1 \eta_1 - \lambda_2 \eta_2 - \cdots - \lambda_m \eta_m \tag{3.56}$$

eşitliğine ulaşılır. Bu eşitlikte η_r'ye göre türev alındığında,

$$\begin{aligned} \frac{\partial D_{\min}}{\partial \eta_r} &= -\sum_{s=1}^{m} \frac{\partial \lambda_0}{\partial \lambda_s} \frac{\partial \lambda_s}{\partial \eta_r} - \lambda_r - \sum_{s=1}^{m} \eta_s \frac{\partial \lambda_s}{\partial \eta_r} \\ &= -\lambda_r \end{aligned} \tag{3.57}$$

eşitliği bulunur ve bu eşitlik üzerinden de η_r'ye ve η_s'ye göre türevler ayrı ayrı alındığında da

$$\frac{\partial^2 D_{\min}}{\partial \eta_r^2} = -\frac{\partial \lambda_r}{\partial \eta_r} \qquad \text{ile} \qquad \frac{\partial^2 D_{\min}}{\partial \eta_r \partial \eta_s} = -\frac{\partial \lambda_r}{\partial \eta_s} \tag{3.58}$$

elde edilir. Böylece $\eta_1, \eta_2, \ldots, \eta_m$'ye bağlı olarak, $D_{\min}$'in ikinci derece kısmi türevlerinin Hessian matrisi aşağıdaki gibidir:

$$H = \begin{pmatrix} -\partial\lambda_1 / \partial\eta_1 & -\partial\lambda_1 / \partial\eta_2 & \cdots & -\partial\lambda_1 / \partial\eta_m \\ -\partial\lambda_2 / \partial\eta_1 & -\partial\lambda_2 / \partial\eta_2 & \cdots & -\partial\lambda_2 / \partial\eta_m \\ \vdots & \vdots & \ddots & \vdots \\ -\partial\lambda_m / \partial\eta_1 & -\partial\lambda_m / \partial\eta_2 & \cdots & -\partial\lambda_m / \partial\eta_m \end{pmatrix} \tag{3.59}$$

$D_{\min}$'in Hessian matrisinin pozitif tanımlı olduğunu gösterebilmek için H^{-1} matrisi ele alınır. H^{-1} matrisi aşağıdaki gibidir,

$$H^{-1} = \begin{pmatrix} -\partial\eta_1/\partial\lambda_1 & -\partial\eta_1/\partial\lambda_2 & \cdots & -\partial\eta_1/\partial\lambda_m \\ -\partial\eta_2/\partial\lambda_1 & -\partial\eta_2/\partial\lambda_2 & \cdots & -\partial\eta_2/\partial\lambda_m \\ \vdots & \vdots & \ddots & \vdots \\ -\partial\eta_m/\partial\lambda_1 & -\partial\eta_m/\partial\lambda_2 & \cdots & -\partial\eta_m/\partial\lambda_m \end{pmatrix} \tag{3.60}$$

çünkü yukarıdaki bu iki matrisin çarpımı birim matristir.

Buradan (3.48) eşitliği dikkate alınarak λ_s'ye göre türev alınırsa,

$$\eta_r = \frac{\sum_{i=1}^{n} q_i g_{ri} \exp(-\lambda_1 g_{1i} - \lambda_2 g_{2i} - \cdots - \lambda_m g_{mi})}{\sum_{i=1}^{n} q_i \exp(-\lambda_1 g_{1i} - \lambda_2 g_{2i} - \cdots - \lambda_m g_{mi})} \tag{3.48}$$

$$\begin{aligned} \frac{\partial\eta_r}{\partial\lambda_s} &= -\sum_{i=1}^{n} p_i \sum_{j=1}^{n} p_j g_r(x_i) g_s(x_i) + \sum_{i=1}^{n} p_i g_r(x_i) \sum_{j=1}^{n} p_j g_s(x_i) \\ &= -\{E[g_r(X) g_s(X)] - E[g_r(X)]E[g_s(X)]\} \\ &= -kov[g_r(X), g_s(X)] \end{aligned} \tag{3.62}$$

$$\frac{\partial\eta_r}{\partial\lambda_r} = -\operatorname{var}[g_r(X)] \tag{3.63}$$

elde edilir.

Eşitlik (3.59)'daki Hessian matrisinin tersi olan H^{-1} matrisinin varyans-kovaryans matrisi olduğu, bunun içinde pozitif tanımlı olduğu görülür. Böylece pozitif tanımlı bir matrisin tersi de pozitif tanımlı olacağından, H matrisi pozitif tanımlanır ve de dolayısıyla $D_{\min}$'in $\eta_1, \eta_2, \ldots, \eta_m$'nin konveks bir fonksiyonu olduğu ortaya çıkar.

Özetle, MinxEnt ile MaxEnt'i şu şekilde karşılaştırabiliriz:

1. MaxEnt'de $p_i = \exp(-\lambda_0 - \lambda_1 g_{1i} - \lambda_2 g_{2i} - \cdots - \lambda_m g_{mi})$ iken

   MinxEnt'de $p_i = q_i \exp(-\lambda_0 - \lambda_1 g_{1i} - \lambda_2 g_{2i} - \cdots - \lambda_m g_{mi})$'dir.

2. Hem MinxEnt'de hem de MaxEnt'de λ_0; $\lambda_1, \lambda_2, \ldots, \lambda_m$'in konveks bir fonksiyonudur.
3. MaxEnt'de H_{max}, $\eta_1, \eta_2, \ldots, \eta_m$'nin konkav bir fonksiyonu iken, MinxEnt'de D_{min}, $\eta_1, \eta_2, \ldots, \eta_m$'nin konveks bir fonksiyonudur.
4. MaxEnt'de $H_{\max} / \partial\eta_r = \lambda_r$ iken, MinxEnt'de $\partial D_{\min} / \partial\eta_r = -\lambda_r$'dir.

3.3. MinxEnt Prensibi ile MaxEnt Prensibi Arasındaki İlişki

Kullback'in minimum çapraz entropi prensibine göre, bir önsel dağılım q için, verilen kısıtlamaları sağlayan ve de aynı zamanda q'ya en yakın olan p dağılımı seçilmelidir. Bu demektir ki;

$$D(p:q) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \tag{3.38}$$

Kullback-Leibler ölçütü

$$\sum_{i=1}^{n} p_i = 1 \tag{3.39}$$

ve

$$\sum_{i=1}^{n} p_i g_r(x_i) = \eta_r \qquad r = 1,2,...,m \tag{3.40}$$

kısıtları altında her i için $p_i \geq 0$ olarak minimize edilir.

Eğer q verilmezse, q'nun yerine maksimum entropiye sahip olan dağılım seçilir. Hiçbir kısıtlama yok iken **u** önsel dağılım olarak kabul edilir ve

$$D(p:u) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{1/n} \tag{3.41}$$

eşitliği minimize edilir. Başka bir deyişle,

$$-\sum_{i=1}^{n} p_i \ln p_i \tag{3.42}$$

ifadesi maksimize edilir.

Böylece, bu bakış açısıyla, Jaynes'in MaxEnt prensibi, Kullback'in MinxEnt prensibinin özel bir durumu olarak kabul edilebilir.

Bununla birlikte, iki prensibin kavramsal temelleri arasındaki farka dikkat edilmelidir. MaxEnt prensibi, bir belirsizlik ölçüsü olan Shannon'ın entropi ölçütünün maksimize edilmesi üzerine kurulmuştur. Başka bir deyişle MaxEnt'e göre, bize verilen tüm bilgilere bağlı kalarak, bize verilmeyen bilginin belirsizliği maksimize edilir. Buradaki temel kavram belirsizliktir.

Öte yandan, MinxEnt'de, belirsizlik kavramı üzerinde durulmaz. En önemli kavram, bir dağılımın diğer bir dağılıma olan olasılık uzaklığı ya da doğrultulmuş sapmasıdır. Burada amaç, verilen kısıtları sağlayan dağılımlar arasından, Q'ya en yakın olan P dağılımını bulmaktır.

Bu noktada, MaxEnt ve MinxEnt prensiplerini birbirinden ayıran özellik yinelenirse MinxEnt'de, verilen bir dağılımdan olasılık uzaklığı minimize edilirken, MaxEnt'de, belirsizlik maksimize edilmektedir.

Bununla birlikte, önsel olasılık dağılımı en belirsiz dağılım olarak karşımıza çıkan bir üniform dağılım ise, bu iki prensip arasında yakın bir ilişki vardır. Verilen kısıtlara göre, en belirsiz dağılıma en yakın olmak, aynı kısıtlamalara bağlı olarak en büyük belirsizliğe sahip olmak demektir.

Daha önce belirtildiği gibi MinxEnt'in kavramsal avantajı, MaxEnt prensibine temel oluşturan ve çok daha zor bir kavram olan belirsizlik yerine , uzaklık ya da uyuşmazlık ile ilgileniyor olmasıdır.

MinxEnt'in MaxEnt' göre bir avantajı da sürekli değişken durumunda ortaya çıkar. Shannon ölçütü, koordinat dönüşümleri için sabit değildir, ve de MaxEnt'i kullanırken olduğu gibi, farklı koordinat sistemleri üzerinde, farklı değerler maksimize edilir. Öte yandan, Kullback-Leibler ölçütü, koordinat dönüşümlerinde sabittir ve de burada, farklı koordinat sistemleri üzerinde, aynı değer maksimize edilir.

Özel bir durum olarak, x'in 0'la 1 arasında değişen sürekli bir rassal değişken olduğunu ve bunun dışında başka bir bilgi olmadığı varsayıldığında MaxEnt'e göre $f(x)=1$ elde edilir, böylelikle x'in tüm değerleri eşit olasılıklıdır. Bununla birlikte, eğer $y=x^2$'nin dağılımı bulunursa, yoğunluk fonksiyonu

$$g(y) = 1/2y^{-1/2}, \qquad 0 \le y \le 1,$$

şeklinde elde edilir ki, bu yüzden y nin hiçbir değeri eşit olasılıklı olamaz. Bununla birlikte x^2 hakkında x'in 0'la 1 arasında değişen rassal bir değişken olduğu dışında hiçbir bilgi yoktur; ancak MaxEnt prensibinin x^2 dağılımına doğrudan uygulanmasıyla x^{2}'nin üniform dağılıma sahip olduğu bulunmalıdır. Bu durum bizi bir çelişkiye götürür ki bu çelişki sürekli değişken durumu için MaxEnt prensibinin kullanımını engellemektedir.

MinxEnt için ise benzeri bir çelişki bulunmamaktadır. Eğer, bazı kısıtlamalar altında x için iki dağılım birbirlerine en yakın durumda ise, bunlara uygun olan ilgili y dağılımları da aynı kısıtlamalar altında birbirlerine en yakın olacaklardır.

MaxEnt için biraz önce bahsedilen çelişki, kesikli değişkenler durumu için söz konusu değildir. Eğer rassal değişken x, $1,2,\ldots,n$ değerlerini alırsa ve bunun dışında başka bir bilgi yoksa, olasılık dağılımı (1/n, 1/n,…,1/n) şeklindedir. Bu durumda x^2, $1^2, 2^2, \ldots, n^2$ değerlerini alır ve de eğer bunun dışında başka bir bilgi yoksa olasılık dağılımı yine de (1/n, 1/n,…,1/n) şeklindedir.

Sürekli değişken durumu için, MaxEnt prensibi kullanıldığında, doğal kısıtlama dışında hiçbir kısıtlamanın olmadığı durumda, değişkenin üniform dağılımı olduğunu varsaymak gibi bir yöntem izlenebilir ve de diğer kısıtlamalar olduğunda bu dağılımın nasıl etkilendiği MaxEnt prensibi kullanılarak incelenebilir. Buna alternatif olarak da, MaxEnt kullanımında, üniform dağılıma göre çapraz entropiyi minimize ettiğimiz söylenebilir.

Bununla birlikte pratikte, çeşitli alanlarda MaxEnt'in başarısı şunu gösterir ki, sürekli değişken durumunda bile, doğal kısıt dışında hiçbir kısıtlamanın olmadığında üniform dağılıma sahip değişken düzgün bir şekilde tanımlanabilir.

4. ENTROPI OPTIMIZASYON PRENSIPLERININ UYGULAMALARI

Çalismanin özgün yanini olusturan bu bölümde entropi optimizasyon prensip leri Jaynes'in maksimum entropi prensibi (MaxEnt) ile Kullback'in Minimum çapraz entropi prensibi (MinxEnt) iki ve üç sayida kisit altinda alti duruma sahip kesikli rassal degisken üzerinde ele alinarak problem kosullu ekstremum problemine dönüstürülmüs, çözümü için ise lagrange çarpanlari yöntemi kullanilmistir. Lagrange çarpanlarinin bulunmasinda ise Yariya Bölme Yöntemi ve Newton Yöntemi uygulanmis, her iki yöntemin esaslari ve isleyisleri anlatilmistir. Jaynes'in maksimum entropi prensibi (MaxEnt) ile Kullback'in Minimum çapraz entropi prensibi temel alinarak Visual Basic programlama dilinde hazirlanan programlarda elde edilen sonuçlarla birlikte verilmistir. Ayrica DIE'nden alinmis 2003-2004 elektrik tüketimi verileri üzerinde MinxEnt Prensibi uygulanarak 2004 tüketim oranlari tahmin edilmis ve sonuçlar 2004 gerçek oranlarla karsilastirilarak yorumlanmistir.

4.1. Verilen Iki Kisit ile MaxEnt Dagilimi

Bir X rassal degiskeni $1,2,\ldots,n$ degerlerini alsin ve bu degerlere uygun olasiliklar $p_1, p_2,\ldots, p_n$ ile gösterilsin. Tayin edilmis bir h ortalamasi oldugunda maksimum entropili olasilik dagilimi bulunsun ($1 \le h \le n$).

$$H(p) = -\sum_{i=1}^{n} p_i \ln p_i \tag{4.1}$$

entropi fonksiyonunu,

$$E\{g(x)\} = \sum_{i=1}^{n} g(x_i) p_i = h \tag{4.2}$$

$$\sum_{i=1}^{n} p_i = 1 \tag{4.3}$$

ve $p_i \ge 0,\ i = 1,2,\ldots,n$ kisitlari altinda maksimize edilmelidir.

Esitlik (4.2)'de $g(x_i) = i\ (x_1 = 1, x_2 = 2,\ldots, x_n = n)$ oldugu dikkate alinirsa,

$$E\{g(x)\} = \sum_{i=1}^{n} i p_i = h$$

şeklinde yazılabilir.

Şu halde $H(p) = -\sum_{i=1}^{n} p_i \ln p_i$ entropi fonksiyonunu,

$$p_1 + p_2 + p_3 + \cdots + p_n = 1 \tag{4.3}$$

ve

$$E\{g(x)\} = p_1 + 2p_2 + 3p_3 + \cdots + np_n = \eta \tag{4.4}$$

kısıtları altında maksimum yapan p_i 'leri bulmakla, problem çözülmüş olacaktır.

Üçüncü bölümde belirtildiği gibi yardımcı L fonksiyonunu aşağıdaki gibi yazılabilir:

$$L \equiv -\sum_{i=1}^{n} p_i \ln p_i - \lambda\left(\sum_{i=1}^{n} p_i - 1\right) - \mu\left(\sum_{i=1}^{n} ip_i - \eta\right). \tag{4.5}$$

Bu fonksiyonun p_i'lere göre kısmi türevleri sıfıra eşitlenerek p_i'ler, λ ve μ Lagrange çarpanlarına bağlı elde edilir:

$$\frac{\partial L}{\partial p_i} = -\left(\ln p_i - \frac{1}{p_i} p_i\right) - \lambda - \mu i = 0$$

buradan

$$\ln p_i = -\mu i - \lambda - 1$$

$$p_i = \exp\{-\mu i - \lambda - 1\}$$

veya

$$p_i = e^{-\mu i} e^{-\lambda - 1}$$

$$p_i = e^{-\mu i} A$$

elde edilir. Burada $A = e^{-\lambda - 1}$'dir. Bu son eşitlikte $e^{-\mu}$, w ile işaret edilsin. Bu durumda,

$$p_i = w^i A, \qquad i = 1, 2, \ldots, n \tag{4.6}$$

eşitliğine ulaşılır.

Eşitlik (4.6), p_i'lerin toplamının bir olması kısıtı olan (4.3) eşitliğinde yerine yazıldığında,

$$\sum_{i=1}^{n} w^i A = 1$$

ve

$$A\sum_{i=1}^{n} w^i = 1$$

elde edilir. Buradan,

$$A = \frac{1}{Z} = \frac{1}{\sum_{i=1}^{n} w^i} \tag{4.7}$$

sonucuna varılır. Eşitlik (4.7) ile ifade edilen A'nın değerini eşitlik (4.6)'da yerine yazdığımızda ise,

$$p_i = \frac{w^i}{\sum_{i=1}^{n} w^i}, \qquad i = 1,2,\ldots,n \tag{4.8}$$

eşitlikleri ile p_i'ler tek bir değişken w'ya bağlı ifade edilmiş olur. Böylece bu p_i'leri (4.4) eşitliğinde yerine yazmakla da,

$$\frac{\sum_{i=1}^{n} i w^i}{\sum_{i=1}^{n} w^i} = \eta \qquad i = 1,2,\ldots,n \tag{4.9}$$

elde edilmiş olur. Eşitlik (4.9)'dan w değerinin bulunması ile p_i olasılık değerlerine ulaşılır.

4.2. İki Kısıtlı Problemin Nümerik Çözümü

Araştırma Problemi 4.2. Bu problem bir zar deneyine uygulandığında, zarın bir çok kez atılmasıyla zarın üzerindeki noktaların ortalamasının η olduğu farz edilerek MaxEnt ile altı yüzün olasılıkları p_i'ler belirlenir.

Bir önceki konuda ifade edilen n tane değere sahip X rassal değişkenini, n=6 durumunda ele almak yeterli olacaktır. O halde zar deneyinde verilmiş bir ortalama değer η var iken her bir yüzün olasılığı;

$$H(p) = -\sum_{i=1}^{6} p_i \ln p_i$$

entropi fonksiyonunu

$$E\{g(x)\} = p_1 + 2p_2 + 3p_3 + 4p_4 + 5p_5 + 6p_6 = \eta$$

ve

$$p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1$$

kısıtları altında maksimize etmekle bulunur.

Eşitlik (4.8) ve (4.9) dikkate alındığında;

$$p_i = \frac{w^i}{\sum_{i=1}^{6} w^i} \text{ ile } \frac{\sum_{i=1}^{n} i w^i}{\sum_{i=1}^{n} w^i} = \eta \qquad i = 1,2,\ldots,6$$

eşitlikleri elde edilir.

Buradan;

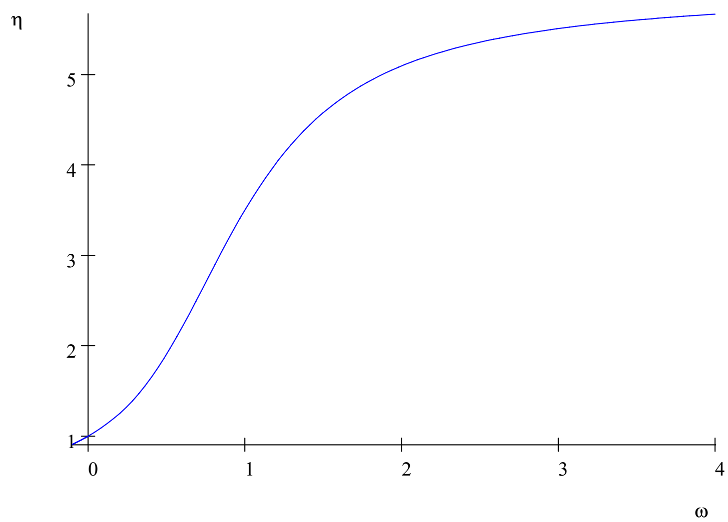
$$p_i = \frac{w^i}{w + w^2 + w^3 + w^4 + w^5 + w^6} \qquad i = 1,2,\ldots,6 \qquad (4.10)$$

$$f(w) = \frac{w + 2w^2 + \ldots + 6w^6}{w + w^2 + w^3 + w^4 + w^5 + w^6} = \eta \qquad (4.11)$$

ifadelerine ulaşılır. Eşitlik (4.11)'deki *f(w)* fonksiyonunun grafiği Şekil 4.1.'de gösterilmiştir. *f(w)* fonksiyonun da *w* parantezine alınarak sadeleştirme yapıldığında,

$$f(w) = \frac{1 + 2w + \ldots + 5w^4 + 6w^5}{1 + w + w^2 + w^3 + w^4 + w^5}$$

elde edilir ve son eşitlikte $f(0) = 1$ ile $\lim_{w \to \infty} f(w) = 6$ noktalarıyla grafik çizilebilir.



Şekil 4.1. $f(w)$ fonksiyonunun grafiği

p_i'lerin grafiğini çizebilmek için de yine her p_i değeri için:

$$p_1 = \frac{w}{w+w^2+w^3+w^4+w^5+w^6} = \frac{1}{1+w+w^2+w^3+w^4+w^5}$$

- $w=0$ iken $p_1 = 1$, $w=1$ iken $p_1 = \frac{1}{6}$, $w \to \infty$ iken $p_1 \to 0$,

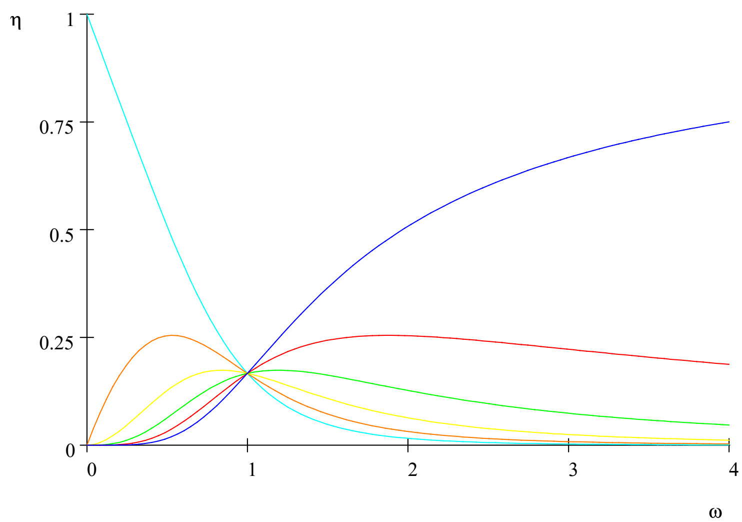
$$p_2 = \frac{w^2}{w+w^2+w^3+w^4+w^5+w^6} = \frac{w}{1+w+w^2+w^3+w^4+w^5} \leq w$$

- $w=0$ iken $p_2 = 0$, $w=1$ iken $p_2 = \frac{1}{6}$, $w \to \infty$ iken $p_2 \to 0$,

$$p_3 = \frac{w^3}{w+w^2+w^3+w^4+w^5+w^6} = \frac{w^2}{1+w+w^2+w^3+w^4+w^5} \leq w^2$$

- $w=0$ iken $p_3 = 0$, $w=1$ iken $p_3 = \frac{1}{6}$, $w \to \infty$ iken $p_3 \to 0$,

ve aynı şekilde p_4, p_5 ve p_6 için de bu noktalar belirlenerek çizilen p_i'lerin grafiği Şekil 4.2.'de gösterilmiştir. Şekil 4.2.'den görüldüğü gibi $w=1$ olması halinde p_i değerlerinin hepsi birbirine eşit ve değeri $1/6$ olacaktır.



Şekil 4.2. p_i olasılıklarının grafiği

Eşitlik (4.11)'de bazı düzenlemeler yapıldığında;

$$g(w) = (1-\eta)w + (2-\eta)w^2 + (3-\eta)w^3 + (4-\eta)w^4 + (5-\eta)w^5 + (6-\eta)w^6 = 0$$

denklemi elde edilir. Bu denklemin çözümü ile yani w'nın bulunması ile p_i'lere dolayısıyla da olasılık dağılımına ulaşılır.

4.2.1. *g(w)=0* Denkleminin Yarıya Bölme Yöntemi ile Çözümü

B. Bolzano ve A.L. Cauchy Teoremi kullanılarak $g(w) = 0$ denklemi çözülebilir.

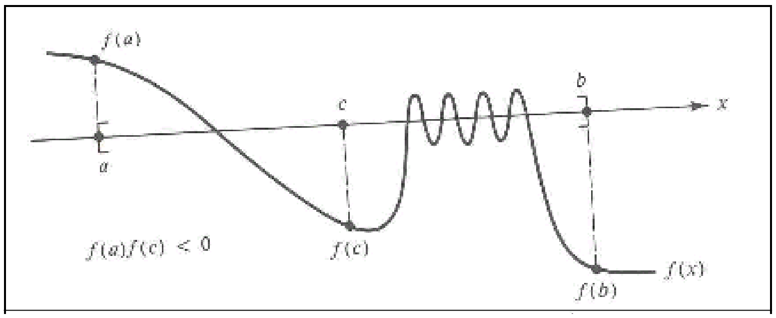
$$g(w) = (1-\eta)w + (2-\eta)w^2 + (3-\eta)w^3 + (4-\eta)w^4 + (5-\eta)w^5 + (6-\eta)w^6 = 0$$

denkleminde katsayılara dikkat edilirse $(1\text{-}\eta),(2-\eta),\ldots,(6-\eta)$ şeklindedir, yani katsayılar artandır. Aynı zaman da $1 \le \eta \le 6$ olduğu için $(1\text{-}\eta) \le 0$ ve $(6\text{-}\eta) \ge 0$'dır. Bir başka deyişle katsayıların işaretinde sadece bir kez değişim vardır. Bu nedenle de $g(w) = 0$ denkleminin *İşaretlerin Descartes Kuralı* ile sadece bir tane pozitif reel kökü vardır.

Bolzano ve Cauchy teoremine göre de, herhangi bir [a,b] aralığında *f(x)* fonksiyonu sürekli ise ve [a,b] aralığının uç noktalarında zıt işaretli değerlere sahip ise $(\mathrm{f(a)} < 0, \mathrm{f(b)} > 0 \text{ veya } \mathrm{f(a)} > 0, \mathrm{f(b)} < 0)$ o zaman (a,b) aralığında en az bir c noktası vardır ki, $f(c)=0$ ’dır. Aşağıdaki Çizelge 4.1.’de yarıya bölme yönteminin grafik üzerinde algoritması gösterilmiştir.

Çizelge 4.1. Yarıya Bölme Yöntemi grafik gösterimi ve algoritması

YARIYA BÖLME YÖNTEMİ (Bisection)



< 0 ise c=(a+b)/2

< 0 ise b=c

> 0 ise a=c

= 0 ise a veya b den biri yada herikisi de kök olabilir.

Bu durumda $g(0)=0$, $g(1)=n\left(\frac{1}{2}(n+1)-\eta\right)=6\left(\frac{1}{2}(7)-\eta\right)$ ve

$\underset{w\to\infty}{g(w)} > 0 = +\infty$ olduğundan,

i) Eğer $\eta > \frac{1}{2}(n+1)$ ise yani $\eta > 3.5$ iken; $g(1)<0$ olur ve bu durumda teoreme göre kök $w_0 > 1$ olacaktır. Bir başka deyişle $g(w)$ fonksiyonunun işaretinin pozitif olacağı yani değişeceği aralık 1’den büyük olan aralık olacaktır.

ii) Eğer $\eta < \frac{1}{2}(n+1)$ ise yani $\eta < 3.5$ iken; $g(1)>0$ olur ki bu durumda da kök $w_0 < 1$ olur. Aynı şekilde bu kez $g(w)$ fonksiyonunun işaretinin negatif olacağı, değişeceği aralık 1’den küçük olan $0 < w_0 < 1$ aralığı olacaktır.

Herhangi bir η değeri için $g(w)=0$ denkleminin w_0 köküne, dolayısıyla da bu η ortalama değerine uygun p_i ’lere ulaşılabilinir.

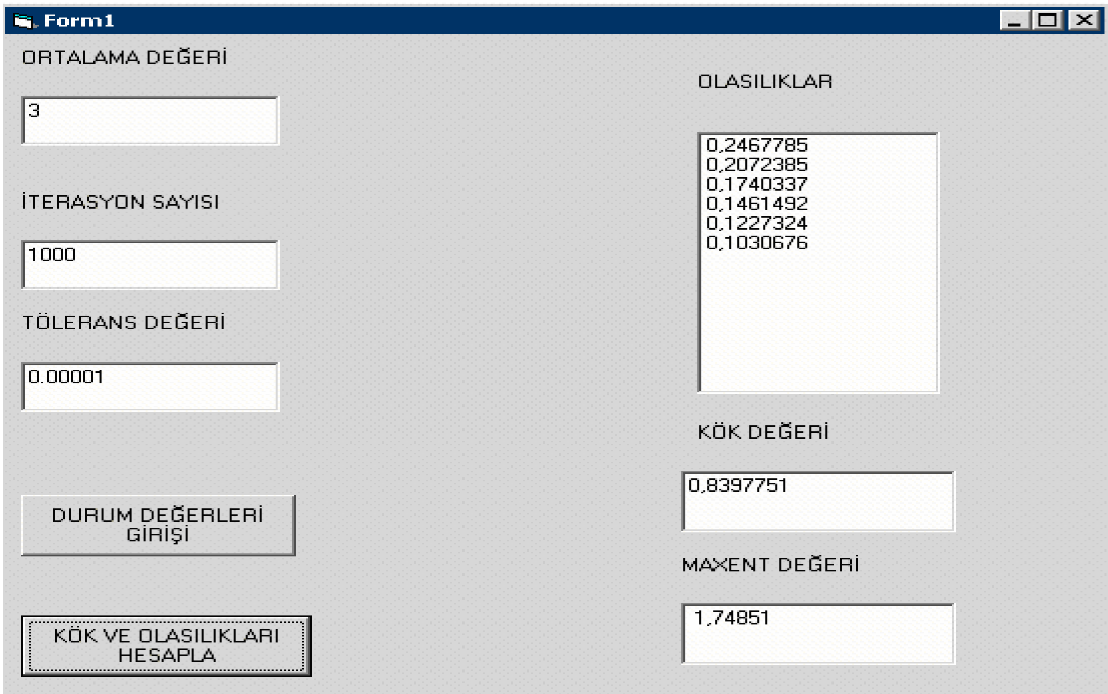
Örneğin $\eta = 1.5$ olduğunda, g(1)=6$\left(3.5-1.5\right)=12>0$ ’dır. O halde kök $0<w_0<1$ aralığındadır. Yarıya bölme yöntemi ile köke yaklaşıldığında;

- $0<w_0<1$ aralığının ortasına $w=0.5$ değerine bakılır, fonksiyonun $w=0.5$'deki değeri $g(0.5)=0.39$ olduğundan kök $0<w_0<0.5$ arasındadır.
- $0<w_0<0.5$ aralığının ortasına, fonksiyonun $w=0.25$'deki değerine bakılır, $g(0.25)=-0.56$ olduğundan kök $0.25<w_0<0.5$ arasındadır.
- $0.25<w_0<0.5$ aralığının ortasına fonksiyonun $w=0.37$'deki değerine bakılır, $g(0.37)=0.04$ olduğundan kök $0.25<w_0<0.37$ arasındadır.
- $0.25<w_0<0.37$ aralığının ortasına fonksiyonun $w=0.31$'deki değerine bakılır, $g(0.31)=-0.02$ olduğundan kök $0.31<w_0<0.37$ arasındadır.
- $0.31<w_0<0.37$ aralığının ortasına fonksiyonun $w=0.34$'deki değerine bakılır, $g(0.34)=0.003$ olduğundan kök $0.31<w_0<0.34$ arasındadır.
- $g(0.32)=-0.01$ ve $g(0.33)=-0.007$ değerlerine göre de $w_0=0.33$ yaklaşık kök değeri elde edilir.

Böylece $w_0=0.33$ sonucu (4.10) eşitliğinde $i=1,2,\dots,6$ değerleri için yerine konulduğunda da $\eta=1.5$ ortalama değer kısıtına uygun $p_1=0.6708$, $p_2=0.2213$, $p_3=0.0730$, $p_4=0.0241$, $p_5=0.0079$ ve $p_6=0.0026$ olasılık değerleri elde edilir.

Ek-1’de kodları verilen Visual Basic programlama dilinde yarıya bölme yöntemi ile hazırlanmış bir programda $\eta=3$ için bilgisayar çıktısı Çizelge 4.2.’de ve η’nın on bir tane değeri için hesaplanan olasılıklarla bu olasılıklar dahilinde maksimum entropi değeri $H_{\max}$, aşağıdaki Çizelge 4.3.’de gösterilmiştir.

Çizelge 4.2. Yarıya Bölme Yöntemi ile olasılık dağılımı hesaplayan program çıktısı



Çizelge 4.3. Verilen ortalama değerler için olasılık dağılımları

η	p_1	p_2	p_3	p_4	p_5	p_6	H_{max}
1.0	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.00000
1.5	0.6637	0.2238	0.0755	0.0255	0.0086	0.0029	0.95356
2.0	0.4781	0.2548	0.1357	0.0723	0.0385	0.0205	1.36724
2.5	0.3475	0.2398	0.1654	0.1142	0.0788	0.0544	1.61373
3.0	0.2468	0.2072	0.1740	0.1461	0.1227	0.1031	1.74843
3.5	0.1666	0.1666	0.1666	0.1666	0.1666	0.1666	1.79176
4.0	0.1031	0.1227	0.1461	0.1740	0.2072	0.2468	1.74843
4.5	0.0544	0.0788	0.1142	0.1654	0.2398	0.3475	1.61373
5.0	0.0205	0.0385	0.0723	0.1357	0.2548	0.4781	1.36724
5.5	0.0029	0.0086	0.0255	0.0755	0.2238	0.6637	0.95356
6.0	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.00000

Çizelge 4.3., $n = 6$ olduğunda η'nın farklı değerleri için elde edilen MaxEnt olasılık dağılımlarını verir. Son kolon ise η'nın farklı değerleri için entropinin maksimum değerini gösterir.

Böylece $n = 6$ ve $\frac{1}{2}(n+1) = 3.5$ iken;

- $\eta = 3.5$ olduğunda, MaxEnt dağılımı üniform dur,
- $\eta > 3.5$ olduğunda, MaxEnt dağılımı artan bir geometrik dizidir,
- $\eta < 3.5$ olduğunda, MaxEnt dağılımı azalan bir geometrik dizidir.

4.2.2. *g(w)=0* Denkleminin Newton Yöntemi ile Çözümü

f(x) fonksiyonunun aranılan ξ kökü [a,b] aralığında olsun. Bu aralığın bir ucundan yola çıkarak, örneğin b'den Taylor Formülü aşağıdaki gibi yazılır:

$$f(\xi) = 0 = f(b) + f'(b)(\xi - b) + \frac{1}{2}(\xi - b)^2 f''(c) . \qquad \xi < c < b \qquad (4.12)$$

Kalan kısım atılarak yaklaşık olarak aşağıdaki formül alındığında

$$f(b) + f'(b)(\xi - b) = 0$$

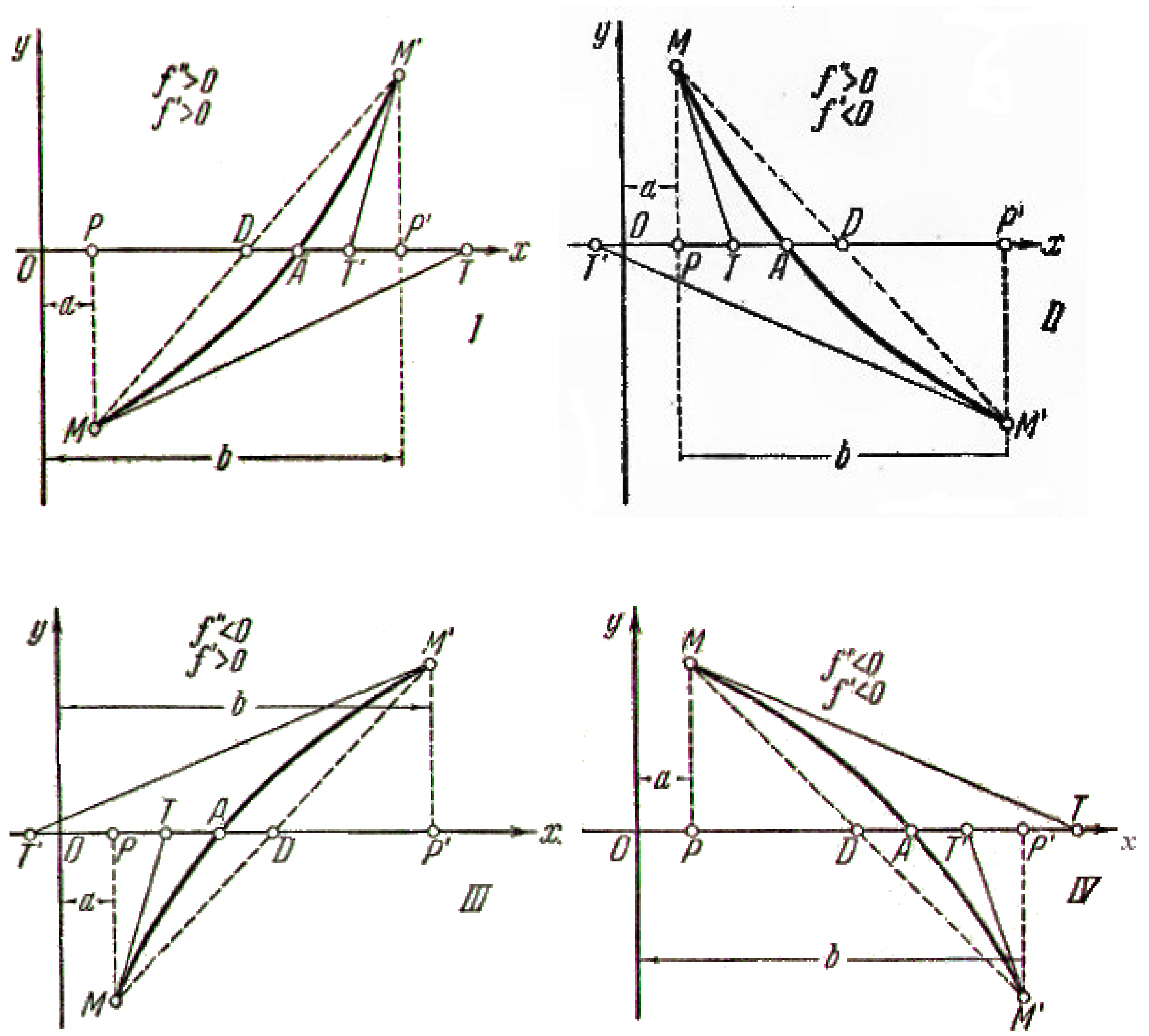
eşitliği elde edilir ki buradan,

$$\xi = b - \frac{f(b)}{f'(b)}$$

bulunur. Böylece ξ kökünün yaklaşık değeri için aşağıdaki eşitlik önerilir:

$$x_1' = b - \frac{f(b)}{f'(b)} . \qquad (4.13)$$

Bu değerin bulunmasını aşağıdaki Şekil 4.3. yardımıyla geometrik şekilde de yorumlamak mümkündür.



Şekil 4.3. Newton Yöntemi'nde başlangıç noktasını belirleme

Yukarıdaki Şekil 4.3.'de $y = f(x)$ eğri yayının apsisi b olan M' noktasında teğeti ele alınsın. Bu teğetin denklemi;

$$y - f(b) = f'(b)(x - b)$$

olarak yazılır. Burada $y = 0$ alınırsa, T' teğetinin *ox* ekseni ile kesişme noktasının *x* apsisi bulunur ki, bu da (4.13) eşitliği ile çakışmaktadır. Bu yöntem Newton Yöntemi olarak adlandırılır. Ancak (4.13) eşitliğindeki x_1' kökünün nerede bulunduğu sorusu ortaya çıkmaktadır. Şekil 4.3.'de görüldüğü gibi teğetin *ox* ekseni ile kesiştiği x_1'

noktası aralığın dışında da bulunabilir. Gösterilebilir ki *f(b)*'nin değeri $f''(x)$ değeri ile aynı işarete sahipse (I. ve IV. Durumlar) x_1', $\xi < x_1' < b$ aralığında yer alır. Gerçekten de *f(b)* ile $f'(b)$ aynı işaretlidir ve $x_1' < b$ olduğu (4.13) eşitliğinden de görülmektedir.

Diğer taraftan (4.12) ve (4.13) eşitliklerinden

$$\xi - x_1' = \xi - b + \frac{f(b)}{f'(b)} = -\frac{1}{2}\frac{f''(c)}{f'(b)}(\xi - b)^2 \tag{4.14}$$

eşitliği elde edilir. Ancak $f''(x)$ söz konusu hallerde $f'(x)$ ile aynı işaretlidir ve bu nedenle $\xi < x_1'$ olur. Nihayet aranan kökün $\xi < x_1' < b$ aralığında olduğu ispatlanmış olur. Benzer yöntemle a noktasından yola çıkarak apsisi a olan *M* noktasından teğet çizildiğinde, (4.13) eşitliği yerine,

$$x_1' = a - \frac{f(a)}{f'(a)} \tag{4.15}$$

eşitliğini alırız. Değerin bu formülle hesaplanması önceki gibidir. Yani, *f(a)*'nın işareti $f''(x)$ ile aynı ise (II. ve III. Durumlar) o zaman $a < x_1' < \xi$ aralığında yer alır.

Böylece verilen dört durumun her biri için hangi değerden başlayarak kökün daha başarılı bulunacağı gösterilmiş olur.

I ve IV hallerinde ardışık uygulamalar kökler için azalan bir dizi $b > x_1' > x_2' > \cdots > x_n' > x_{n+1}' > \xi$,

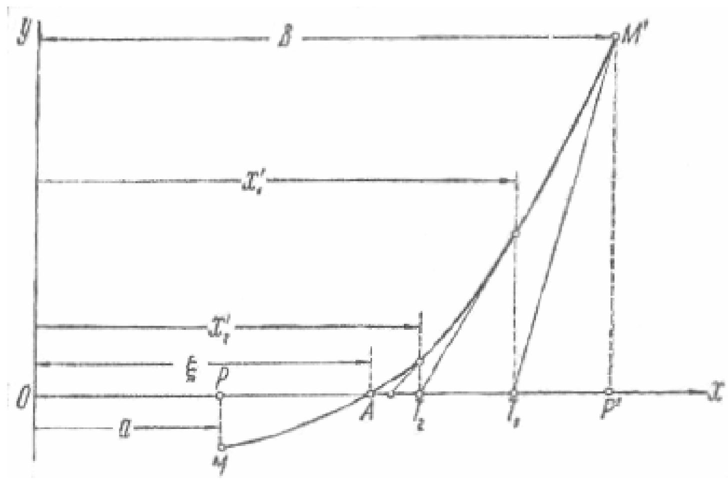
II ve III hallerinde ise artan bir dizi $a < x_1' < x_2' < \cdots < x_n' < x_{n+1}' < \xi$ ortaya çıkarır.

Genellikle ξ köküne,

$$x_{n+1}' = x_n' - \frac{f(x_n')}{f'(x_n')} \tag{4.16}$$

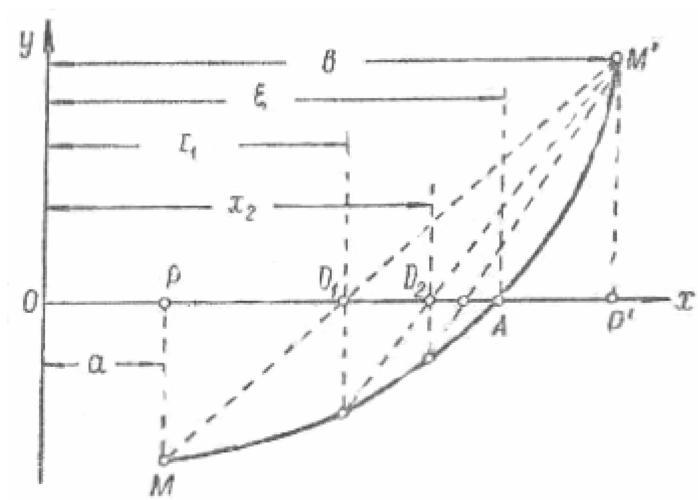
formülüyle ulaşılır.

Aşağıdaki Şekil 4.4. A noktasına teğetlerin *ox* ekseni ile kesiştiği $T_1, T_2, \ldots$ noktalarından yaklaştığını gösterir.



Şekil 4.4. b noktasından teğetlerle A noktasına yaklaşma

Her seferinde köke bir önceki yaklaşık kökten ulaşıldığı için Newton Metodu iteratif bir yöntemdir. Şekil 4.5. ise A noktasına kirişlerin *ox* ekseni ile kesiştiği $D_1, D_2, \ldots$ noktalarından yaklaştığını gösterir.



Şekil 4.5. a noktasından kirişlerle A noktasına yaklaşma

Hesaplanan yaklaşık kökün kesinlik derecesi,

$$|x_n' - \xi| \leq \frac{|f(x_n')|}{m} \tag{4.17}$$

formülü ile hesaplanır. Burada *m*, $|f'(x_n)| \geq m$ 'dir. Yani $|f'(x_n)|$ 'in verilen aralıktaki mutlak değerce en küçük değeridir.

Newton yöntemi,

$$g(w) = (1-\eta)w + (2-\eta)w^2 + (3-\eta)w^3 + (4-\eta)w^4 + (5-\eta)w^5 + (6-\eta)w^6 = 0$$

denkleminde $\eta = 1.5$ için uygulandığında $g(w)$ fonksiyonu,

$$g(w) = (-0.5)w + (0.5)w^2 + (1.5)w^3 + (2.5)w^4 + (3.5)w^5 + (4.5)w^6 = 0$$

olarak belirlenir.

$g(1) = 12 > 0$ olduğundan kök $0 < w < 1$ aralığındadır. O halde $a = 0$ ve $b = 1$ olur. $g(w)$ fonksiyonunun türevinde,

$$g'(w) = -0.5 + w + 4.5w^2 + 10w^3 + 17.5w^4 + 27w^5$$

w=0 almakla,

$$|g'(0)| = m = 0.5$$

olarak (4.17) formülündeki *m* elde edilir. $g''(w)$ fonksiyonunu,

$$g''(w) = 1 + 9w + 30w^2 + 70w^3 + 162w^5 > 0$$

ve $g(1)$ ile aynı işaretli olduğundan köke b'den yani 1'den yaklaşmaya karar verilir.

- Ve (4.16) formülünde $x_0' = 1$ yazılır, o zaman

$$x_1' = 1 - \frac{g(1)}{g'(1)} = 1 - \frac{12}{59.5} = 0.798319$$

elde edilir. Elde edilen x_1' değeri g fonksiyonunda yerine yazılarak,

$$g(x_1') = 2.127415$$

bulunur. Elde edilen x_1' değeri (4.17) formülünde yerine yazmakla da,

$$|x_1' - \xi| \le \frac{2.12}{0.5}$$

$$|x_1' - \xi| \le 4.24$$

kökün kesinlik derecesi bulunmuş olur. Ardışık işlem sırası aşağıdaki gibidir:

- $x_2' = 0.79 - \dfrac{g(0.79)}{g'(0.79)} = 0.79 - \dfrac{2.12}{14.57} = 0.554495$

$$g(x_2') = g(0.55) = 0.682226$$

$$|x_2' - \xi| \le \frac{0.68}{0.5}$$

$$|x_2' - \xi| \le 1.36$$

- $x_3' = 0.55 - \dfrac{g(0.55)}{g'(0.55)} = 0.55 - \dfrac{0.68}{6.20} = 0.444320$

$$g(x_3') = g(0.44) = 0.188652$$

$$|x_3' - \xi| \le \frac{0.18}{0.5}$$

$$|x_3' - \xi| \le 0.37$$

- $x_4' = 0.44 - \dfrac{g(0.44)}{g'(0.44)} = 0.44 - \dfrac{0.18}{2.76} = 0.415319$

$$g(x_4') = g(0.41) = 0.115000$$

$$|x_4' - \xi| \le \frac{0.11}{0.5}$$

$$|x_4' - \xi| \le 0.23$$

- $x_5' = 0.41 - \dfrac{g(0.41)}{g'(0.41)} = 0.41 - \dfrac{0.11}{2.16} = 0.356759$

$$g(x_5') = g(0.35) = 0.0147332$$

$$|x_5' - \xi| \le \frac{0.014}{0.5}$$

$$|x_5' - \xi| \le 0.029$$

- $x_6' = 0.35 - \dfrac{g(0.35)}{g'(0.35)} = 0.35 - \dfrac{0.014}{1.2} = 0.337725$

$$g(x_6') = g(0.33) = 0.0004$$

$$|x_6' - \xi| \le \frac{0.0004}{0.5}$$

$$|x_6' - \xi| \le 0.0008$$

- $x_7' = 0.33 - \dfrac{g(0.33)}{g'(0.33)} = 0.33 - \dfrac{0.00049}{1.0821} = 0.337241$

$$g(x_7') = g(0.33724) = 12 * 10^{-7}$$

$$|x_7' - \xi| \leq \frac{12 * 10^{-7}}{0.5}$$
$$|x_7' - \xi| \leq 0.000002$$

x_7' değerinin ξ ile farkının çok küçük bir değer olması nedeniyle $w = x_7' = 0.337242$ yaklaşık kök değeridir denilebilir. Bu w değeri ile

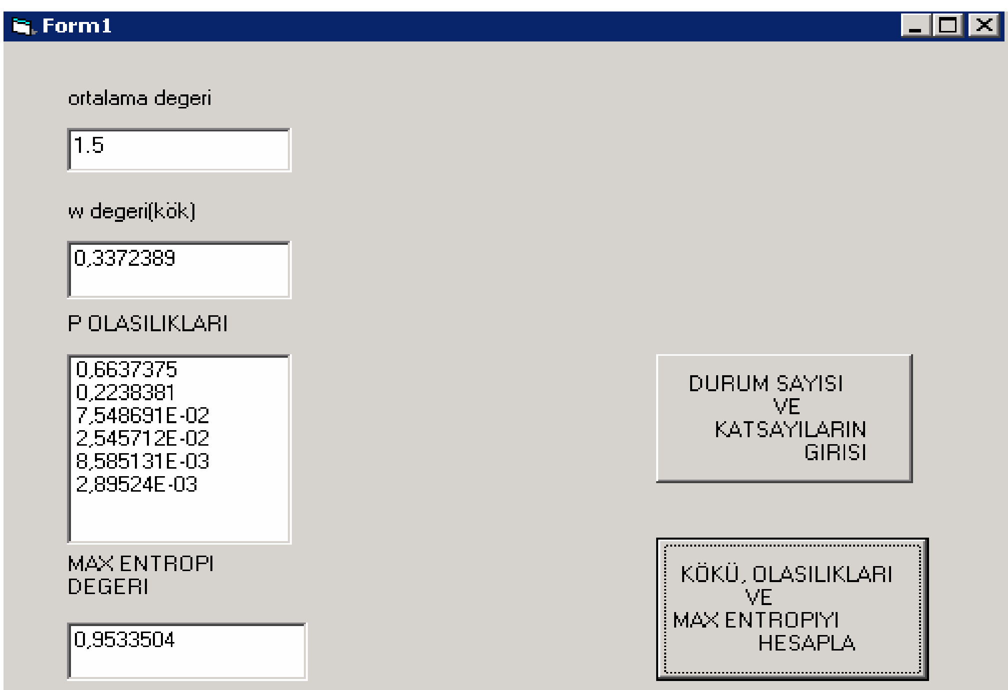
$$p_i = \frac{w^i}{w + w^2 + w^3 + w^4 + w^5 + w^6} \qquad i = 1,2,\ldots,6$$

eşitliğinden elde edilen olasılık değerleri,

$$p_1 = 0.6637, \; p_2 = 0.2238, \; p_3 = 0.0755$$
$$p_4 = 0.0255, \; p_5 = 0.0086, \; p_6 = 0.0029$$

olarak bulunur ki, bu değerler Çizelge 4.3.'teki $\eta = 1.5$ satırı ile aynıdır. Ek-2'de kodları verilen Visual Basic programlama dilinde Newton Yöntemi kullanarak hazırlanan bir programda $\eta = 1.5$ için elde edilen program çıktısı Çizelge 4.4.'deki gibidir.

Çizelge 4.4. Newton Yöntemi ile MaxEnt dağılımı hesaplayan program çıktısı



Bu çıktıdan da görüldüğü gibi sonuçlar yine Çizelge 4.3.'teki $\eta = 1.5$ olduğu satır ile çakışmaktadır.

4.3. Verilen Üç Kısıt ile MaxEnt Dağılımı

Bir önceki konuda iki kısıta sahip bir zar deneyi ele alınmış ve MaxEnt olasılık dağılımı hesaplanmıştı. Şimdi ise aynı deneyde üç kısıt söz konusu olduğunda MaxEnt olasılık dağılımının hesaplanmasına ilişkin problem aşağıdaki gibidir.

Araştırma Problemi 4.3. $H(p) = -\sum_{i=1}^{6} p_i \ln p_i$ entropi fonksiyonunun

$$\sum_{i=1}^{6} p_i = 1, \ \sum_{i=1}^{6} ip_i = 4, \ \sum_{i=1}^{6} i^2 p_i = 19$$

kısıtları altında maksimumunu veren olasılık dağılımını bulmakla problem çözülmüş olacaktır.

$$L = \sum_{i=1}^{6} p_i \ln p_i + \lambda_1 \left(\sum_{i=1}^{6} i^2 p_i \right) + \lambda_2 \left(\sum_{i=1}^{6} ip_i \right) + \lambda_3 \left(\sum_{i=1}^{6} p_i \right)$$

şeklinde Lagrange fonksiyonu yazılır. Bu fonksiyonun p_i'lere göre türevi alınarak sıfıra eşitlenmesiyle

$$L_{p_i} = \left(\ln p_i + \frac{1}{p_i} p_i \right) + i^2 \lambda_1 + i\lambda_2 + \lambda_3 = 0$$

$$\ln p_i = -i^2 \lambda_1 - i\lambda_2 - \lambda_3 - 1$$

$$p_i = \exp\left\{ - i^2 \lambda_1 - i\lambda_2 - \lambda_3 - 1 \right\}$$

$$p_i = e^{-i^2 \lambda_1} . e^{-i\lambda_2} . e^{-\lambda_3 - 1}$$

elde edilir. Yukarıdaki son eşitlikte de $e^{-\lambda_1} = w_1$, $e^{-\lambda_2} = w_2$, $e^{-\lambda_3 - 1} = A$ olarak ifade edilirse p_i'ler aşağıdaki gibi sıralanabilir:

$$p_1 = w_1 . w_2 . A ,$$

$$p_2 = w_1^4 . w_2^2 . A ,$$

$$p_3 = w_1^9 . w_2^3 . A ,$$

$$p_4 = w_1^{16}.w_2^4.A\,,$$

$$p_5 = w_1^{25}.w_2^5.A\,,$$

$$p_6 = w_1^{36}.w_2^6.A\,.$$

Bu p_i değerleri birinci kısıtta yerine yazıldığında

$$A\left(w_1.w_2 + w_1^4.w_2^2 + w_1^9.w_2^3 + w_1^{16}.w_2^4 + w_1^{25}.w_2^5 + w_1^{36}.w_2^6\right) = 1$$

$$A = \frac{1}{w_1.w_2 + w_1^4.w_2^2 + w_1^9.w_2^3 + w_1^{16}.w_2^4 + w_1^{25}.w_2^5 + w_1^{36}.w_2^6}$$

olarak A, w_1 ve w_2 cinsinden ifade edilmiş olur. Böylelikle

$$p_i = \frac{w_1^{i^2}.w_2^i}{w_1.w_2 + w_1^4.w_2^2 + w_1^9.w_2^3 + w_1^{16}.w_2^4 + w_1^{25}.w_2^5 + w_1^{36}.w_2^6}$$

şeklinde de p_i, w_1 ve w_2 cinsinden ifade edilmiş olur. Elde edilen p_i değerleri ikinci ve üçüncü kısıtta yerine yazıldığında

$$w_1.w_2.A + 2w_1^4.w_2^2.A + 3w_1^9.w_2^3.A + 4w_1^{16}.w_2^4.A + 5w_1^{25}.w_2^5.A + 6.w_1^{36}.w_2^6.A = 4$$

ile

$$w_1.w_2.A + 4w_1^4.w_2^2.A + 9w_1^9.w_2^3.A + 16w_1^{16}.w_2^4.A + 25w_1^{25}.w_2^5.A + 36.w_1^{36}.w_2^6.A = 19$$

eşitliği elde edilir. A yerine w_1 ve w_2 cinsinden ifade edilmiş değeri yazıldığında ve gerekli düzenlemeler yapıldığında ise

$$(1-4)w_1.w_2 + (2-4)w_1^4.w_2^2 + (3-4)w_1^9.w_2^3 + \cdots + (6-4).w_1^{36}.w_2^6 = 0$$
$$(1-19)w_1.w_2 + (4-19)w_1^4.w_2^2 + (9-19)w_1^9.w_2^3 + \cdots + (36-19).w_1^{36}.w_2^6 = 0$$

şeklinde iki bilinmeyenli iki denklem elde edilir. Bu iki denklem aşağıdaki gibi de ifade edilebilir:

$$f_1(w_1, w_2) = \sum_{i=1}^{6}(i - \eta_1)w_1^{(i^2)}w_2^i$$

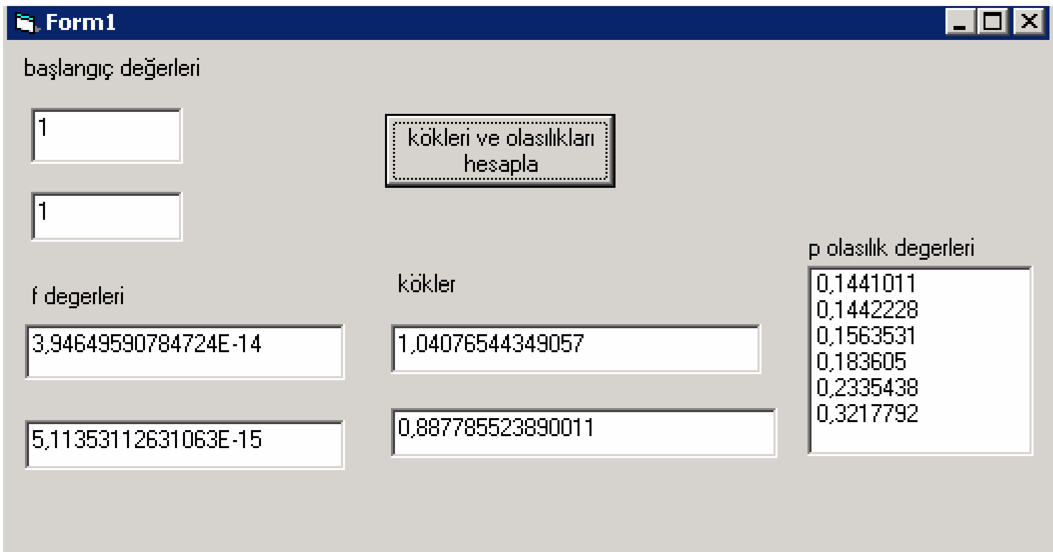
ve

$$f_2(w_1, w_2) = \sum_{i=1}^{6}(i^2 - \eta_2)w_1^{(i^2)}w_2^i\,.$$

Bu iki denklemi sağlayan w_1 ve w_2 bulunduğunda olasılık dağılımı elde edilmiş olur. Ek-3'te kısıt sayısı üç olduğunda Newton Yöntemi ile problemi çözen

Visual Basic programlama dilinde hazırlanan program kodları verilmiştir. Programda başlangıç değerler girildiğinde Newton Yöntemi ile bulunan sonuçlar aşağıdaki gibidir.

Çizelge 4.5. Newton Yöntemi ile (üç kısıt altında) MaxEnt dağılımı hesaplayan program çıktısı



4.4. Önsel Dağılım Q'nun Verilmesi ile Entropideki Azalış

Önsel dağılım verildiğinde entropide indirgeme gerçekleşir. Bu durumu ortaya koyabilmek için aşağıdaki problem MinxEnt ve MaxEnt Prensipleri ile ayrı ayrı incelenerek entropi değerlerindeki değişime dikkat edilebilir.

Araştırma Problemi 4.4. Önsel bir olasılık dağılımı Q, (0.05, 0.10, 0.15, 0.20, 0.22, 0.28) şeklinde verilmişken ve altı yüze sahip bir zarın atışları sonrasında gelen yüzlerin ortalaması 4.5 olduğunda minimum çapraz entropi olasılık dağılımı bulunsun ve Q önsel dağılımının verilmesiyle entropideki azalışı gösterilsin.

Çözüm 4.3. $D(p:q) = \sum_{i=1}^{6} p_i \frac{p_i}{q_i}$ fonksiyonunu,

$$\sum_{i=1}^{n} p_i = 1 \qquad \text{ve} \qquad \sum_{i=1}^{n} i p_i = 4.5$$

kısıtları altında minimize etmekle

$$p_i = \frac{q_i w^i}{\sum_{i=1}^{n} q_i w^i}, \qquad i = 1,2,\ldots,6$$

elde edilir. Bu eşitlik $\sum_{i=1}^{n} ip_i = 4.5$ kısıtında yerine yazmakla da;

$$p_i = \frac{q_i w^i}{q_1 w + q_2 w^2 + q_3 w^3 + q_4 w^4 + q_5 w^5 + q_6 w^6} \qquad i = 1,2,\ldots,6$$

$$f(w) = \frac{q_1 w + 2q_2 w^2 + \ldots + 6q_6 w^6}{q_1 w + q_2 w^2 + q_3 w^3 + q_4 w^4 + q_5 w^5 + q_6 w^6} = 4.5$$

yukarıdaki fonksiyonlara ulaşılır. $f(w)$ üzerinde düzenlemeler yapıldığında,

$$g(w) = q_1(1-\eta)w + q_2(2-\eta)w^2 + q_3(3-\eta)w^3 + q_4(4-\eta)w^4 + q_5(5-\eta)w^5 + q_6(6-\eta)w^6 = 0$$

denklemi elde edilir. Denklemin çözümünü sağlayan w değerinin Ek-4'te kodları verilen Visual Basic programlama dilinde hazırlanan programda $\eta = 4.5$ vermekle elde edilen olasılık dağılımı ve program çıktısı aşağıdaki Çizelge 4.6.'daki gibidir:

Çizelge 4.6. Newton Yöntemi ile MinxEnt dağılımı hesaplayan program çıktısı

MinxEnt dağılımı P_1={0.035, 0.078, 0.131, 0.192, 0.234, 0.330} şeklindedir. Ve bu dağılımın entropisi $H(P_1) = 1.606$ olarak bulunur.

Eğer Q dağılımı verilmemiş ise Laplace'ın yetersiz sebep ilkesi kullanılmalı ve Q dağılımı yerine üniform dağılım U alınarak,

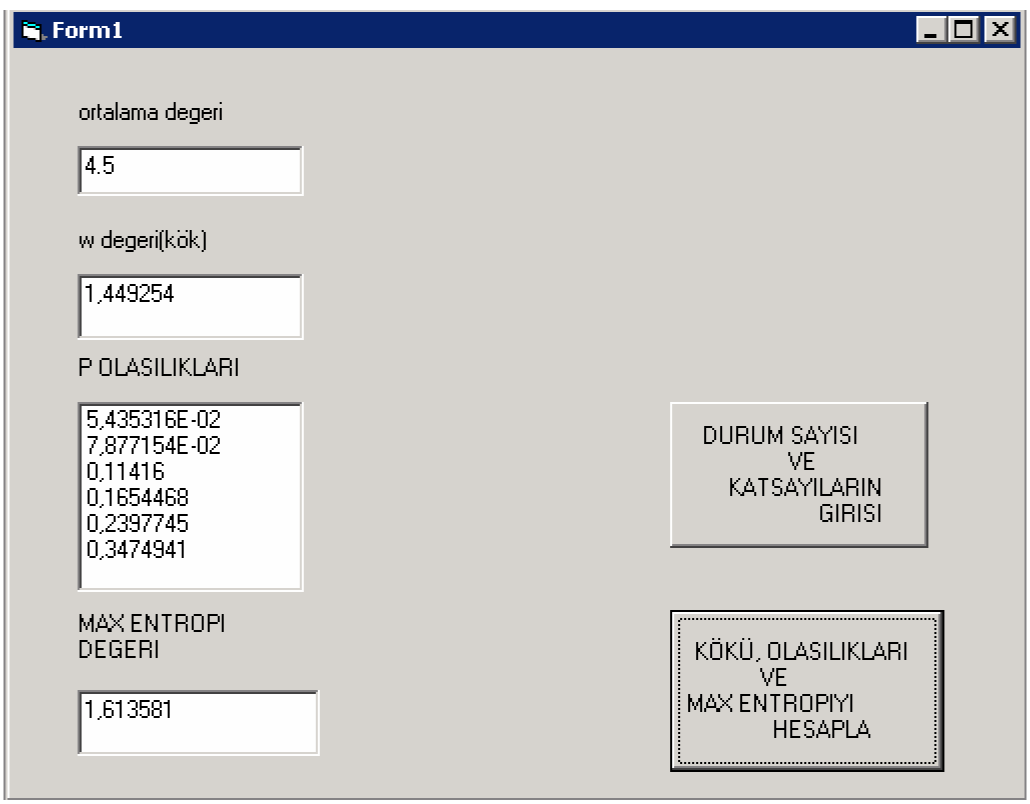
$$\sum_{i=1}^{6} p_i \ln \frac{p_i}{1/6}$$

minimize edilmeli ya da MaxEnt prensibi ile,

$$-\sum_{i=1}^{6} p_i \ln p_i$$

Shonnon entropi ölçütü verilen kısıtlar altında maksimize edilmelidir. Ek-2'de kodları verilen bilgisayar programında $\eta = 4.5$ almakla elde edilen bilgisayar çıktısı Çizelge 4.7.'de verilmiştir.

Çizelge 4.7. Newton Yöntemi ile MaxEnt dağılımı hesaplayan program çıktısı



MaxEnt dağılımıP_2={ 0.0543, 0.0788, 0.1142, 0.1654, 0.2378, 0.3475} şeklinde bulunur ve bu dağılımın entropisi de $H(P_2) = 1.613$ olarak elde edilir.

Böylece 4.5 ortalamaya sahip tüm dağılımlardan P_2 maksimum entropiye sahiptir ve aynı ortalamayla önsel dağılıma sahip herhangi bir diğer dağılım P_1 iken onun entropisi, P_2'nin entropisine eşit ya da daha küçük olmalıdır.

4.5. Elektrik Tüketim Oranlarının MinxEnt Prensibi ile Bulunması ve Gerçek Oranlarla Karşılaştırılması

Elektrik Enerjisinin Önemi

Bütün dünya ülkelerinde olduğu gibi ülkemizin de ekonomik ve sosyal kalkınmasının vazgeçilmez girdilerinden birisi olan enerjinin ihtiyacımızı karşılayacak şekilde zamanında ve güvenilir olarak temin edilmesi büyük önem arz etmektedir. Görev başına gelen hükümetlerin en önemli hedefi, dengeli ve kalıcı bir ekonomik büyümeyi sağlamaktır. Enerji tüketimi ile ekonomik faaliyetler arasında yakın bir ilişki mevcuttur. Ülkelerin milli hasılaları arttıkça, enerji üretimleri de buna paralel olarak artmaktadır ki bu da bize enerjinin, önemli üretim faktörleri arasında anahtar rol oynadığını göstermektedir. Ayrıca ülkelerin gelişmişlik düzeylerini belirleyen ölçütler arasına ülkelerde kişi başına kullanılan enerji miktarı da eklenmiştir.

Dünyada sanayi devrimi meydana geldikten sonra, ekonomik gelişme ile enerji ve özellikle elektrik enerjisi kullanımının artmasından dolayı, günümüzde elektrik enerjisi, ekonomik ve sosyal hayatın çok önemli bir unsuru haline gelmiştir. Elektrik enerjisi sektörü, ekonominin diğer sektörlerine büyük miktarlarda girdi veren bir sektördür ve diğer sektörlerdeki gelişmelerden etkilenmenin yanı sıra diğer sektörlerin gelişmesine yardımcı olur. Tarım ve sanayi sektörlerinin makineleşmesiyle alt yapının ve ulaşımın gelişmesi, elektrik enerjisi kullanan dayanıklı tüketim mallarının türü ve sayısının önemli ölçüde artması sonucu daha fazla elektrik enerjisi talep edilmektedir. Elektrik enerjisinin yetersizliği bir ülkede sadece bazı bölgelerin karanlıkta kalması demek değildir. Bunun yanı sıra üretimin düşmesi, işsizliğin artması, tüm sanayi sektörünün felce uğraması, yatırımların gerilemesi ve kısaca milli gelirin azalması anlamına gelir.

İktisadi mal olarak elektrik enerjisinin iki özelliği vardır:

i) Elektrik enerjisinin arzı talebine bağlıdır. Bu sebeple talebi aşan bir üretim söz konusu değildir. Depo edilememesi üretildiği an tüketilmesi gerekmektedir.

ii) Elektrik enerjisi saf bir enerji türü olması nedeniyle diğer türlere rahatlıkla dönüştürülebilir. Aynı zamanda kolay, temiz, sessiz ve ucuz bir şekilde üretilmekte ve istenildiği kadar küçük parçalara bölünebilmekte ve birleştirilebilmektedir. Bu sebeple kullanım alanı çok geniştir (Direm 1998).

Gelişmenin ön koşulu olan enerjinin, özellikle de elektrik enerjisinin zamanında, güvenilir ve ekonomik şekilde ve çevre faktörünü de dikkate alarak üretilmesi son derece önemlidir. Gelecek dönem için hazırlık yaparken elektrik enerjisi konusunda ülkenin içinde bulunduğu durumu bütün çıplaklığı ile ortaya koymak ve mevcut sıkıntıların belirlenmesini müteakip çözüm yollarını aramak, dünya uygulamalarını da dikkate alarak sektörün yeniden yapılanmasını sağlamak son derece önem arz etmektedir (Yiğitgüden 1999).

Araştırma Problemi 4.5.

Firmalar gelecek planlarını yaparken önlerine hedefler koyar ve bu hedefler genellikle satış hasılatı ve kar gibi firmanın devamlılığını sağlamaktaki en önemli etkenler olurlar. Bunlar genellikle üst yönetimin belirlediği hedeflerdir. Firmaların kapasitesi ise yapılacak olan yeni yatırımlara ya da tasfiye edilecek birimlere göre daha somut bir şekilde hesaplanabilmektedir. Bu kısıtlar altında belirlenen hedeflere ulaşmak için firmanın hangi ürün çeşidinden ne miktarda üretmesi gerektiğinin tahmini istenebilir

Günümüzde de en önemli enerji kaynağı olan elektrik enerjisi üretimi ve dağıtımını gerçekleştiren müesseseler de sözü edilen hedefleri belirlemek ve bu doğrultuda ilerlemek zorundadır. Özellikle Elektrik Dağıtım Müesseseleri, üretim yapan firmaların hangi üründen hangi oranda üretmesi gerektiğini bilmek istemesi gibi arzı talebine bağlı bir ürün olan elektriğin, kullanım alanlarına göre tahmini tüketim oranlarını belirlemesi gerekebilir.

Bu bölümde uygulama olarak DİE'den alınan ve kullanım alanlarına göre 2003-2004 elektrik tüketim oranlarını içeren veriler ele alınmıştır. Problem ise yukarıda belirtilen probleme benzer olarak ortalama birim fiyat tahmin edildiğinde her bir kullanım alanına düşen tüketim oranının belirlenmesidir. Kullanım alanları sekiz tane olup her birine düşen kullanım miktarı, oranı ve elde edilen TL. cinsinden değeri aşağıdaki Çizelge 4.8.'de gösterilmiştir.

Çizelge 4.8. Kullanım alanlarına göre 2003-2004 II. dönem elektrik tüketim oranları

	2003			2004		
	II. Dönem			II. Dönem		
	(Nisan, Mayıs, Haziran)			(Nisan, Mayıs, Haziran)		
Kullanım alanları	**Miktar**		**Değer**	**Miktar**		**Değer**
	Gwh	**%**	**TL.**	**Gwh**	**%**	**TL.**
Toplam	26 861	100.00	2.521.589.153	27 735	100.00	2.729.901.228
Resmi Daire	1 078	4.01	159 599 025	1 235	4.45	174 841 720
Sanayi+Otoprodüktör	13 363	49.75	897 028 250	13 226	47.69	883 636 139
Ticarethaneler	2 836	10.56	459 703 729	3 403	12.27	553 892 236
Meskenler	5 899	21.96	801 992 555	6 202	22.36	814 497 857
Tarımsal Sulama	423	1.57	47 500 555	538	1.94	62 426 557
Şantiyeler	335	1.25	52 117 136	322	1.16	52 547 468
Sokak Aydınlatması	913	3.40	1 826 028	818	2.95	64 359 278
Diğer+EÜAŞ Direkt Satışı	2 014	7.50	101 821 875	1 991	7.18	123 699 973

Problemin Kullback'in minimum çapraz entropi prensibi ile çözümü için önsel olasılık dağılımı, X rassal değişkeninin değerleri ve ortalama kısıt değeri belirlenmelidir. Çizelge 4.8.'deki verilerden 2003 elektrik tüketim oranları, önsel olasılık dağılımı olarak alınır ve 2004 Toplam değerin toplam miktara bölünmesiyle ise ortalama birim fiyat olarak E(x) birinci moment kısıtı verilir. Her bir kullanım alanının birim fiyatı (1 Gwh başına düşen fiyat) X rassal değişkeninin aldığı değerler olarak hesaplanır. Bu hesaplama 2003 yılı verilerinde her bir kullanım yerine karşılık gelen değer kolonu, miktar kolonuna bölünerek yapılır. Böylelikle her bir kullanım yeri için elektrik birim fiyatı elde edilmiş olur. Elde edilen bu veriler doğrultusunda oluşturulan X rassal değişkeninin sıralı değerleri ve önsel olasılık değerleri ile E(X) ortalama değer kısıtı Çizelge 4.9.'da gösterilmiştir.

Çizelge 4.9. X ve q önsel olasılık değerleri

Kullanım alanları	X değerleri (birim fiyatlar)	Q önsel dağılım
Diğer+EÜAŞ Direkt Satışı	50 557	0.0750
Sanayi+Otoprodüktör	67 127	0.4975
Sokak Aydınlatması	2000	0.0340
Tarımsal Sulama	112 294	0.0157
Meskenler	135 953	0.2196
Resmi Daire	148 051	0.0401
Ticarethaneler	162 095	0.1056
Şantiyeler	155 573	0.0125
	E(x)=98 428	1.0000

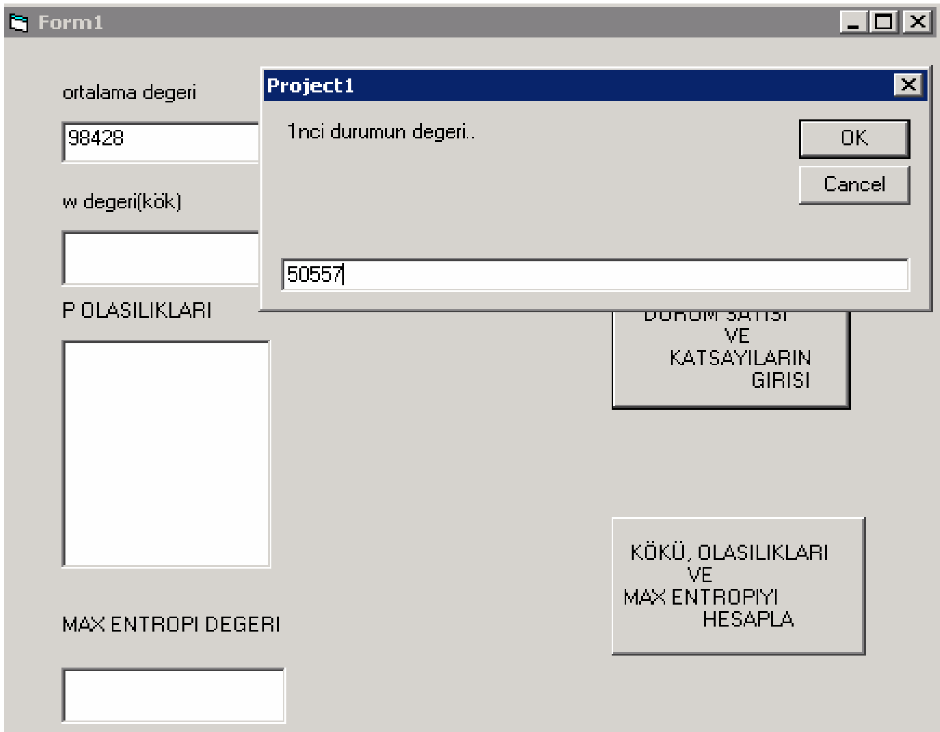
Yukarıdaki Çizelge 4.9.'da gösterilen E(X) değeri 2004 dönemine ait toplam değeri toplam miktara bölerek aşağıdaki gibi elde edilir.

$$E(X) = \frac{2729901228}{27735}$$

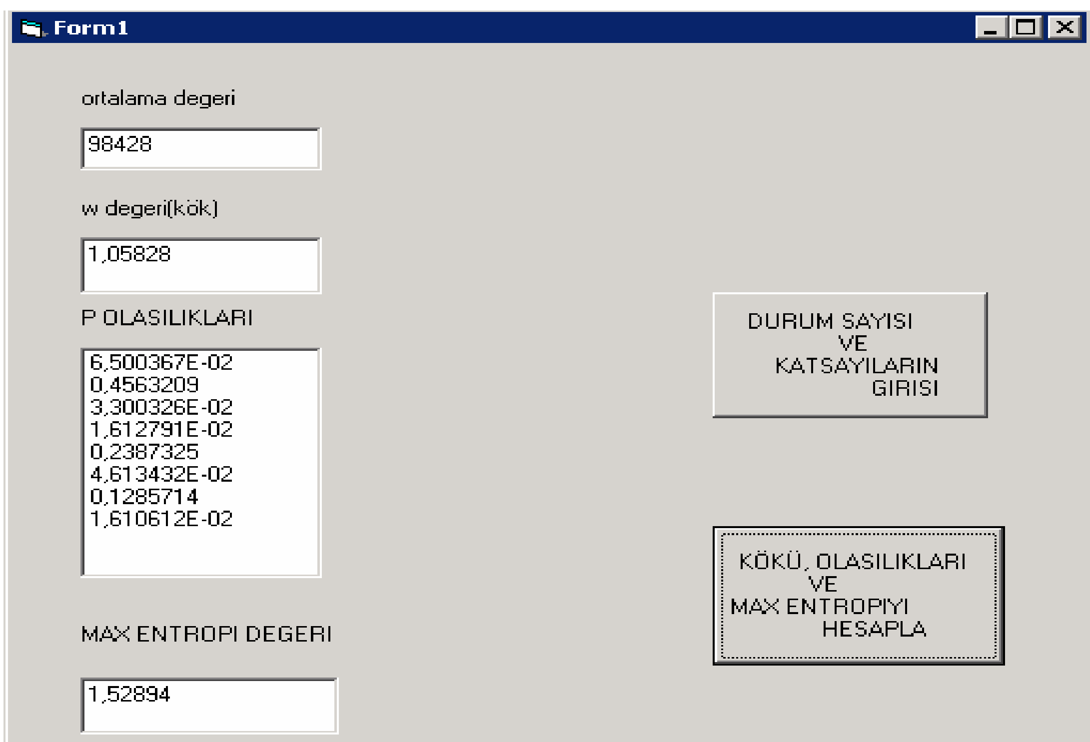
$$E(X) = 98428 \text{ TL. (ortalama birim fiyat)}$$

Ek-4'de kodları verilen MinxEnt prensibi ile çözümü elde eden programda, X değerleri girişi Çizelge 4.10.'da, kısıt olarak E(X) değeri verilmesi ve sonucunda elde edilen tahmini tüketim oranlarını gösteren bilgisayar çıktısı Çizelge 4.11.'de verilmiştir.

Çizelge 4.10. MinxEnt dağılımı hesaplayan programda X değerleri girişi



Çizelge 4.11. 2004 Tüketim oranları için MinxEnt dağılımı hesaplayan program çıktısı



MinxEnt Prensibi ile bulunan 2004 tahmini tüketim oranları ile 2004 gerçek tüketim oranları aşağıdaki Çizelge 4.12.'de karşılaştırılmıştır. Gerçek değerlerle tahmini değerlerin birbirine olan yakınlığının ortaya konması ile metodun ne ölçüde doğru çalıştığı test edilmiştir ve tatminkar sonuçlara ulaşılmıştır.

Çizelge 4.12. 2004 Gerçek tüketim oranları ile MinxEnt tahmini oranları

Kullanım alanları	2004 Elek. Tük. Oranları	2004 MinxEnt Tük. Oranları
Diğer+EÜAŞ Direkt Satışı	0.0718	**0.0650**
Sanayi+Otoprodüktör	0.4769	**0.4563**
Sokak Aydınlatması	0.0295	**0.0330**
Tarımsal Sulama	0.0194	**0.0161**
Meskenler	0.2236	**0.2387**
Resmi Daire	0.0445	**0.0461**
Ticarethaneler	0.1227	**0.1285**
Şantiyeler	0.0116	**0.0161**

5. TARTIŞMA VE SONUÇ

Entropi, sözlükteki anlamı ile belirsizlik demektir. Belirsizlik ise her yerde ve her zaman kullanılan bir kavram olup bir sistemi tanımlamak için gerekli olan toplam informasyondan daha az bir informasyona sahip olunduğunda ortaya çıkar. Belirsizlik ve informasyon öyle yakın iki kavramdır ki bir deney hakkında edinilen informasyon yok edilen belirsizlik miktarına eşittir. Gerçekleştirilen çalışmayla ilgili olan "olasılıksal belirsizlik durumu" ele alındığında entropi "bir sisteme ait belirsizlik düzeyinin ölçümü" olarak ifade edilebilir. Olasılık teorisi sadece gelecekteki olayın beklentisi ile ilişkili olan belirsizlik ile ilgilenir. Bir başka deyişle herhangi bir olayın baş gösterme olasılığı, belli düzeyde bu olayın gerçekleşip gerçekleşmeyeceği hakkında belirsizliğin göstergesidir. Shannon, "bir olasılık uzayının belirsizliğinin ölçüsü" olarak tanımladığı entropi için, bir deneyin sonuçlarının içerdiği ortalama bilgi miktarını önermiştir.

Shannon entropisi denilen bu ölçüt Jaynes tarafından, ölçümlere dayandırılan sistemin durumlarını tanımlamak için olasılık dağılımı türetmekte kullanılmıştır. Jaynes, bilgi olarak verilen ortalama değerlerle uyumlu sonsuz dağılım arasından bu ölçütü maksimize eden dağılımın seçilmesini önermiştir. Belirsizliği indirgemek amaçlanırken maksimum belirsizliğe sahip olan dağılımın seçilmesinde yürüttüğü mantık şöyledir. Maksimum entropiden daha az entropisi olan bir dağılım seçilirse, entropideki bu indirgeme bilinçli ya da bilinçsiz kullanılmış olan bazı ilave edilen bilgiden gelmiş olabilir. Ancak böyle bilgiler verilmediğinden onları kullanmak doğru olmayacaktır. Böylece sadece maksimum entropiye sahip olanı kullanılmalıdır. Maksimum entropi prensibi olarak bilinen bu entropi maksimizasyonunu başka şekilde de açıklamak mümkündür. X rassal değişkenine bağlı verilen her bir koşul, bu değişken hakkında belli miktarda informasyon içermekte ancak belirsizliğini aşmaya yetmemektedir. Belirsizliği aşmak için gereken informasyonun maksimum miktarının bilinmesi için verilmeyen informasyona uygun olan bu belirsizlik maksimize edilmektedir.

MaxEnt prensibinde bilgi, olasılık dağılımına ait doğrusal momentlerin belirlenen değerleri üzerine kurulur ve hiçbir kısıtlamanın olmadığı durumda belirsizliği en büyük olan üniform dağılımın seçilmesini önerir. Oysa hiçbir kısıtlamanın olmadığı durumda dahi olasılıkların eşit olmadığını gösteren bazı sebepler olabilir. Bu sebepler tecrübeye, önseziye yada teoriye dayalıdır. Bir başka deyişle 'n' tane durumun olasılıklarının, önsel olasılık dağılımını ifade eden $q_1, q_2, \ldots, q_n$ olduğuna inanmayı gerektiren sebepler bulunmaktadır. Böyle bir bilginin varlığında MaxEnt yetersiz kalmakta ve bu bilgi kullanılamamaktadır. Kullback tarafından geliştirilen minimum çapraz entropi prensibi ise aynı kısıtları sağlayan dağılımlar arasından verilen önsel olasılık dağılıma en yakın dağılımın seçilmesini önerir ve Kullback-Leibler ölçütü olarak bilinen bir uzaklık ölçütünü minimize ederek yakınlığı sağlar.

Bu bilgiler ışığında ele alınan sonlu sayıda duruma sahip bir sistemde iki ve üç sayıda kısıtlara tabi tutulan araştırma problemleri incelendiğinde, olasılık dağılımlarını elde edebilmek için Newton Yöntemi ile Yarıya Bölme Yöntemi'nden yararlanılmıştır. İki yöntem içinde algoritmalar geliştirilerek, programlar hazırlanmıştır. Bu programlar da gerçekleştirilen çözümler sonucunda MaxEnt'in MinxEnt prensibinin özel bir hali olduğu anlaşılmıştır. MaxEnt üniform dağılıma (belirsizliği en büyük dağılım) en yakın dağılımı seçerken, MinxEnt özel bir önsel olasılık dağılımı bilinmesi halinde bu dağılıma en yakın olanını dikkate almaktadır. Bu açıdan 2003 birim fiyatları üzerinde 2003 tüketim oranları önsel olasılık dağılımı olarak hesaba katılarak ve 2004 ortalama birim fiyat kısıtı göz önüne alınarak 2004 tüketim oranları MinxEnt prensibi tahmin edilmiştir. Birim fiyatların değişmiş olması yöntemin daha hassas sonuçlara ulaşmasını engellemiş olduğu için, akla gelecek soru bu ürün çeşitlerinin fiyatlarının önceden nasıl tahmin edileceğidir. Bu soruya da enflasyonist ortamdan kaynaklanan fiyat değişimleri, geçmiş yıllardaki verilere göre tahmin edilen fiyat ya da aynı sektördeki diğer firmaların öngördüğü fiyatların baz alınabileceği söylenebilir. Verdiği duyarlı sonuçlar sayesinde 2005 tüketim oranları da tahmin edilerek çalışma devam ettirilebilir. Ayrıca kısıt sayısını arttırabilecek algoritmalar geliştirilmesi ile de çalışma ilerletilebilir.

KAYNAKLAR

ATALAY, K.D., *Entropi maksimizasyonu ve bir uygulama*, Yüksek Lisans Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, Türkiye (2000)

COVER, T.M. ve THOMAS, J.A., *Elements of Information Theory*, A Wiley-Interscience Publication, New York, USA, (1991)

DİREM, B., *Genelleştirilmiş En Küçük Kareler Yöntemi ve Bir Uygulama*, Yüksek Lisans Tezi, Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, Eskişehir, Türkiye (1998)

FIHTENGOLTS, G.M., *Kurs differentialnogo integralnogo ischislenia*, NAUKA, Moskva, Rusia (1969)

GOLAN, A., JUDGE, G. ve MILLER, D., *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, John Wiley & Sons, New York, USA, (1996)

KARMESHU ve PAL. N.R., *Entropy Measures, Maximum Entropy* Principle *and Emerging Applications*, Springer, New York, USA, (2003)

KULLBACK, S., *Information Theory and Statistics*, Dover Publications, New York, USA, (1968)

KAPUR, J.N. ve KESAVAN, H.K., *Entropy Optimization Principles with Applications*, Academic Press, New York, USA, (1992)

ÖZKUL, S.,*Su Kalitesi Gözlem Ağlarının Entropi Yöntemi ile Değerlendirilmesi*, Turk J Engin Environ Sci, 25, 435-452 (2001)

PAPOULİS, A., *Probability, Random variables, and stochastic processes*, McGraw-Hill, Singapore, USA (1991)

ŞAMİLOV, A., *Comparision of Classical Estimate Methods Based On MaximumEntropy Distribution*, Ordered statistical data: Approximation, Bounds and characterizations (BAYRAMOĞLU İ.), İzmir, Türkiye,(2005)

ŞAMİLOV, A., *Entropi İnformasyon Teorisi*, Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, Yayınlanmış Ders Notları, Eskişehir, (2004).

YİĞİTGÜDEN, H.Y., *Türkiye'de Elektrik Enerjisi Sektöründe Özelleştirme Politikaları Ve Çalışmaları*, Midas Basım-Yayın, İstanbul, Türkiye, (1999)

WU, N.,*The Maximum Entropy Method*, Springer, Berlin, Germany (1997)

EKLER

EK-1

Yarıya Bölme Yöntemi ile MaxEnt Dağılımı Hesaplayan Program Kodları

```
Dim m As Single
Dim w As Single
Dim a As Single
Dim b As Single
Private c(1 To 100) As Integer
Dim t As Single
Dim no As Integer
Dim tol As Single
Dim i As Single
Dim k As Integer
Dim p(1 To 16) As Single
Dim s As Single, n As Integer

Private Sub Command1_Click()
no = Val(Text2.Text)
tol = Val(Text3.Text)

        k = 1: i = 0
        m = Val(Text1.Text)
                Do While (k = 1)
                        If fdegeri(i) > 0 Then
                        b = i
                        k = 0
                Else
                        a = i
                        End If
                        i = i + 1

                Loop
        i = 1
        Do While (i < no)
                w = (a + b) / 2
                If fdegeri(w) = 0 Or (b - a) / 2 < tol Then
                Exit Do
        Else
                i = i + 1
                If fdegeri(w) > 0 Then
                b = w
        Else
                a = w
```

```
            End If
            End If
        Loop

Text4.Text = w
        For j = 1 To n
            t = t + w ^ j
        Next j


        For j = 1 To n
            p(j) = (w ^ j) / t
            List1.AddItem p(j)
        Next j

        For j = 1 To n
            If p(j) <> 0 Then
            s = s + (p(j) * Log(p(j)))
            End If
        Next j
Text5.Text = s
End Sub

Private Sub Command2_Click()
        n = Val(InputBox("durumlarin sayisi.."))

        For j = 1 To n
            c(j) = InputBox(Str(j) + "nci durumun degeri..")
        Next j
End Sub

Public Function fdegeri(w As Single) As Double
        fdegeri = 0
        For j = 1 To n
            fdegeri = fdegeri + (c(j) - m) * w ^ j
        Next j
End Function
```

EK-2

Newton Yöntemi ile MaxEnt Dağılımı Hesaplayan Program Kodları

```
Dim x(1 To 100) As Single
Dim w As Single
Dim q As Single
Dim i As Single
Dim dx As Double
Dim b As Single
Dim a As Single
Dim n As Integer
Dim m As Single
Dim c(1 To 100) As Integer
Dim p(1 To 100) As Single
Dim s As Single
Dim h As Single

        Public Function ftrev(q As Single) As Double
        fturev = 0
                For j = 1 To n
                        ftrev = ftrev + (j * (c(j) - m)) * (q ^ (j - 1))
                Next j
        End Function

        Public Function fdegeri(w As Single) As Double
        fdegeri = 0
                For j = 1 To n
                        fdegeri = fdegeri + (c(j) - m) * w ^ j
                Next j
        End Function

Private Sub Command2_Click()
        n = Val(InputBox("durumlarin sayisi.."))
                For j = 1 To n
                        c(j) = InputBox(Str(j) + "nci durumun degeri..")
                Next j
End Sub

Private Sub Command1_Click()
        m = Val(Text1.Text)
        k = 1: i = 0
                Do While (k = 1)
                        If fdegeri(i) > 0 Then
                        b = i
                        k = 0
```

```
                Else
                        a = i
                        End If
                        i = i + 1
                Loop
        i = 1
        x(i) = b
        dx = 1
                Do While (dx > 0.00001)
                        w = x(i)
                        q = x(i)
                        x(i + 1) = x(i) - ((fdegeri(w)) / (ftrev(q)))
                        w = x(i + 1)
                If ftrev(a) < 0 Then
                        h = (-1) * ftrev(a)
                        dx = fdegeri(w) / h
                Else
                        dx = fdegeri(w) / (ftrev(a))
                End If
                        i = i + 1
                Loop

Text3.Text = w
                For j = 1 To n
                        t = t + w ^ j
                Next j

                For j = 1 To n
                        p(j) = (w ^ j) / t
                        List1.AddItem p(j)
                Next j

                For j = 1 To n
                        If p(j) <> 0 Then
                        s = s + (p(j) * Log(p(j)))
                        End If
                Next j
        If s < 0 Then s = (-1) * s
        Text2.Text = s
End Sub
```

EK-3

Newton Yöntemi ile (üç kısıt altında) MaxEnt Dağılımı Hesaplayan Program Kodları

```
Dim det As Double
Dim y1 As Double
Dim y2 As Double
Dim b As Double
Dim c As Double
Dim d As Double
Dim e As Double
Dim s As Integer
Dim p (1 To 20) As Single
Dim A As Single
        Private Sub Command1_Click()
                b = Text1.Text
                c = Text2.Text
                d = 0.2: e = 0.2
        Do While (s < 10)
                s = s + 1
                d = a1(b, c)
                e = a2(b, c)

                det = (a11(b, c) * a22(b, c)) - (a12(b, c) * a21(b, c))
                y1 = ((a22(b, c) / det) * a1(b, c)) - ((a12(b, c) / det) * a2(b, c))
                y2 = (-(a21(b, c) / det) * a1(b, c)) + ((a11(b, c) / det) * a2(b, c))
                b = b - y1
                c = c - y2
        Loop
Text5.Text = b
Text6.Text = c
Text4.Text = e
Text3.Text = d

         For j = 1 To 6
A = (b * c + (b ^ 4) * (c ^ 2) + (b ^ 9) * (c ^ 3) + (b ^ 25) * (c ^ 5) + (b ^ 36) * (c ^
6))
p(j) = b ^ (j ^ 2) * c ^ j / A
List1.AddItem p(j)
Next j
End Sub

Function a1(b As Double, c As Double) As Double
a1 = (-3) * (b * c) + (-2) * (b ^ 4) * (c ^ 2) + (-1) * (b ^ 9) * (c ^ 3) + (b ^ 25) * (c ^ 5)
+ (2) * (b ^ 36) * (c ^ 6)
```

```
End Function

Function a2(b As Double, c As Double) As Double
a2 = (-18) * (b * c) + (-15) * (b ^ 4) * (c ^ 2) + (-10) * (b ^ 9) * (c ^ 3) + (-3) * (b ^
16) * (c ^ 4) + (6) * (b ^ 25) * (c ^ 5) + (17) * (b ^ 36) * (c ^ 6)
End Function

Function a11(b As Double, c As Double) As Double
a11 = (-3) * (c) + (-8) * (b ^ 3) * (c ^ 2) + (-9) * (b ^ 8) * (c ^ 3) + (25) * (b ^ 24) * (c
^ 5) + (72) * (b ^ 35) * (c ^ 6)
End Function

Function a12(b As Double, c As Double) As Double
a12 = (-3) * (b) + (-4) * (b ^ 4) * (c) + (-3) * (b ^ 9) * (c ^ 2) + (5) * (b ^ 25) * (c ^ 4)
+ (12) * (b ^ 36) * (c ^ 5)
End Function

Function a21(b As Double, c As Double) As Double
a21 = (-18) * (c) + (-60) * (b ^ 3) * (c ^ 2) + (-90) * (b ^ 8) * (c ^ 3) + (-48) * (b ^ 15)
* (c ^ 4) + (150) * (b ^ 24) * (c ^ 5) + (17 * 36) * (b ^ 35) * (c ^ 6)
End Function

Function a22(b As Double, c As Double) As Double
a22 = (-18) * (b) + (-30) * (b ^ 4) * (c) + (-30) * (b ^ 9) * (c ^ 2) + (-12) * (b ^ 16) *
(c ^ 3) + (30) * (b ^ 25) * (c ^ 4) + (17 * 6) * (b ^ 36) * (c ^ 5)
End Function
```

EK-4

Newton Yöntemi ile MinxEnt Dağılımı Hesaplayan Program Kodları

```
Dim x(1 To 100) As Single
Dim w As Single
Dim q As Single
Dim i As Single
Dim j As Integer
Dim dx As Double
Dim b As Single
Dim a As Single
Dim n As Integer
Dim m As Single
Dim c(1 To 100) As Long
Dim p(1 To 100) As Single
Dim s As Single
Dim h As Single
Dim l(1 To 100) As Single

    Public Function ftrev(q As Single) As Double
        fturev = 0
        For j = 1 To n
            ftrev = ftrev + (l(j) * j * (c(j) - m)) * (q ^ (j - 1))
        Next j
    End Function
    Public Function fdegeri(w As Single) As Double
        fdegeri = 0
        For j = 1 To n
            fdegeri = fdegeri + ((c(j) - m) * w ^ j * l(j))
        Next j
    End Function

Private Sub Command2_Click()
    n = Val(InputBox("durumlarin sayisi.."))
        For j = 1 To n
            c(j) = InputBox(Str(j) + "nci durumun degeri..")
        Next j
        For j = 1 To n
            l(j) = InputBox(Str(j) + "nci önsel olasılık degeri..")
        Next j
End Sub

Private Sub Command1_Click()
    m = Val(Text1.Text)
        k = 1: i = 0
```

```
        Do While (k = 1)
                If fdegeri(i) > 0 Then
                b = i
                k = 0
        Else
                a = i
                End If
                i = i + 1
        Loop
i = 1
x(i) = b
dx = 1

        Do While (dx > 0.00001)
                w = x(i)
                q = x(i)
                x(i + 1) = x(i) - ((fdegeri(w)) / (ftrev(q)))
                w = x(i + 1)
        If ftrev(a) < 0 Then
                h = (-1) * ftrev(a)
                dx = fdegeri(w) / h
        Else
                dx = fdegeri(w) / (ftrev(a))
        End If
        i = i + 1
        Loop

Text3.Text = w
                For j = 1 To n
                        t = t + (l(j) * (w ^ j))
                Next j
                For j = 1 To n
                        p(j) = (l(j) * (w ^ j)) / t
                        List1.AddItem p(j)
                Next j

                For j = 1 To n
                        If p(j) <> 0 Then
                        s = s + (p(j) * Log(p(j)))
                        End If
                Next j
        If s < 0 Then s = (-1) * s
Text2.Text = s
End Sub
```