T.C.
İSTANBUL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

BAYESLİ MANTIKSAL ÇIKARIM ÇERÇEVESİNDE MAKSİMUM ENTROPİ OLASILIKLARININ ANALİZİ

EMİN SERHAN SÜZER

Matematik Anabilim Dalı

Matematik Programı

Danışman

Yrd.Doç.Dr. Mehmet CEVRİ

Haziran, 2016

İSTANBUL

Bu çalışma 22/06/2016 tarihinde aşağıdaki jüri tarafından Matematik Anabilim Dalı Matematik programında Yüksek Lisans Tezi olarak kabul edilmiştir.

Tez Jürisi:

İmza
Yrd. Doç. Dr. Mehmet CEVRİ
İstanbul Üniversitesi
Fen Fakültesi

İmza
Prof. Dr. Kamuran SAYGILI
İstanbul Üniversitesi
Fen Fakültesi

İmza
Prof. Dr. İsmail Müfit GİRESUNLU
İstanbul Üniversitesi
Fen Fakültesi

İmza
Doç. Dr. Ayten PEKİN
İstanbul Üniversitesi
Fen Fakültesi

İmza
Doç. Dr. Gürsel YEŞİLOT
Yıldız Teknik Üniversitesi
Fen Edebiyat Fakültesi

ÖNSÖZ

Bu çalışma boyunca yönlendirmeleri ve bilgi birikimi ile her konuda bana destek olan danışmanım Yard.Doç. Dr. Mehmet CEVRİ'ye teşekkürlerimi sunarım.

Bu çalışmayı yaparken yanımda olan ve verileri ile beni destekleyen Sütaş Genel Müdürlüğüne ve Eğitim Merkezleri Müdürü Mecit ÖZDEMİR' e teşekkür ederim.

Her zaman yanımda olan maddi ve manevi desteklerini hiçbir zaman benden esirgemeyen babam Sami SÜZER'e, annem Seval SÜZER' e çok teşekür ederim.

Haziran, 2016 Emin Serhan SÜZER

İÇİNDEKİLER

ŞEKİL LİSTESİ

TABLO LİSTESİ

SİMGE VE KISALTMA LİSTESİ

Simgeler	**Açıklama**
$\mathbf{h(X)}$	:Diferansiyel entropi
$\mathbf{H(p)}$	:Shannon entropi ölçüsü
$\mathbf{H(X,Y)}$	:Bileşik entropi
$\mathbf{H(X\|Y)}$	:Koşullu entropi
$\mathbf{L}$	:Lagrange fonksiyonu
$\mathbf{S(p)}$	: Bir A olayının p olasılığı ile meydana gelmesi halinde oluşacak sürpriz
$\chi^{\mathbf{2}}$	:Ki-kare dağılımı
$\mathbf{N(\mu,\sigma^2)}$	:μ ortalamalı, σ^2 varyanslı Normal Dağılım
$\log \mathbf{f_i}$	:f_i olasılıklarının e tabanına göre logaritması

Kısaltmalar	**Açıklama**
$\mathbf{MaxEnt}$	:Maksimum entropi
$\mathbf{H_{maks}}$	:Maksimum entropi değeri
$\mathbf{S_{maks}}$	:Maksimum belirsizlik
$\mathbf{S_{min}}$	:Minimum belirsizlik

ÖZET

YÜKSEK LİSANS TEZİ

BAYESLİ MANTIKSAL ÇIKARIM ÇERÇEVESİNDE MAKSİMUM ENTROPİ OLASILIKLARININ ANALİZİ

Emin Serhan SÜZER

İstanbul Üniversitesi

Fen Bilimleri Enstitüsü

Matematik Anabilim Dalı

Danışman: Yard.Doç. Dr. Mehmet CEVRİ

Termodinamikte önemli bir rol oynayan ve kısaca *belirsizliğin ölçüsü* olarak da adlandırılan *Entropi* kavramı farklı bakış açılarına göre bilim dünyasında çok farklı tanımlara sahiptir. Entropinin Matematik, İstatistik, Fizik, Biyoloji, Finans, Mühendislik ve Sosyal Bilimler gibi pek çok bilim alanında görüntü işlemesi, portfolyo seçimi, mal ve mülklerin fiyatlandırılması gibi çok geniş uygulamaları bulunmaktadır. Bu tez çalışmasını amacı, herhangi bir olay hakkında ön bilgileri değerlendirerek onlarla ilgili birtakım olasılık dağılımlarının atanmasını sağlayan Bayesli Mantıksal Çıkarımını kullanıp entropi fonksiyonunun maksimize edilmesini incelemek ve maksimum entropinin uygulamalarından bahsetmektir.

Bu çalışma beş ana bölümden oluşmaktadır. Birinci bölümde tezin konusu ile ilgili tarihsel süreci de içeren bir alt yapı verilmiştir.

İkinci bölüm, üç alt bölümden oluşmaktadır. Birinci alt bölümde klasik olasılık teorisi ile Bayesli olasılık teorisi irdelenerek karşılaştırılır. Bu bölümde olasılık kavramı, özellikleri, Bayesli yaklaşımı sunulmakta ve Bayesli yaklaşımının avantajlarına yer verilmektedir. Ayrıca literatürde sıkça kullanılan önsel dağılımlar hakkında bilgi verilmektedir. İkinci alt bölümde belirsizlik kavramı ve bilgi kuramına değinilerek entropi kavramı ele alınıp Entropinin temel özelliklerinden ve bazı önemli entropi ölçülerinden sunulmaktadır. Ayrıca Shannon Entropi Ölçüsü, kesikli ve sürekli dağılımlar için entropi kavramları ve bunların ayrı ayrı bileşik, koşullu entropisi tanımlarına yer verilip kesikli ve diferansiyel entropi arasındaki ilişki incelenmektedir.

Üçüncü alt bölümde ise Bayesli Mantıksal Çıkarımı çerçevesinde maksimum entropi ilkesi ayrıntılı olarak incelenmektedir.

Üçüncü bölüm, tez boyunca faydalanılan araçlardan ve uygulanan yöntemlerden oluşmaktadır.

Dördüncü bölümde, ikinci bölümde verilen Bayesli Maksimum Entropi Olasılıklarının, gerçek Dünya problemlerine uygulaması verilmektedir. Elde edilen sonuçlar yorumlanır ve ileriye dönük çalışmalar hakkında bahsedilmektedir.

Beşinci bölümde ise çalışmanın genel bir değerlendirmesi yapılmaktadır.

Haziran 2016, 111 sayfa.

Anahtar kelimeler: Bayesli Mantıksal Çıkarım, Entropi, Olasılık Dağılımları, Maksimum Entropi İlkesi.

SUMMARY

M. Sc. THESIS

ANALYSIS OF MAXIMUM ENTROPY PROBABILITIES WITHIN BAYESIAN LOGICAL INFERENCE FRAMEWORK

Emin Serhan SÜZER

İstanbul University

Institute of Graduate Studies in Science and Engineering

Department of Mathematics

Supervisor: Asst.Prof. Dr. Mehmet CEVRI

The concept of *Entropy* which plays an important role and briefly called as *the measure of uncertainty*, has quite diversified explanations in the scientific world according to different point of views. Entropy has broad applications such as image processing, portfolio selection, pricing of the goods and chattels in various scientific fields like Mathematics, Statistics, Physics, Biology, Finance, Engineering and Social Sciences. The aim of this thesis is to examine the maximizing of the entropy function and mention the applications of the maximum entropy by evaluating the prior knowledge about any event, using Bayesian Logical Inference which enables the assignment of some probability distribution concerning this event.

This study consists of five main chapters. In the first chapter, an infrastructure has been given including the historical process of the subject of the thesis.

The second chapter consists of three subsections. In the first subsection, the Classical Probability Theory and Bayesian Probability Theory is compared by scrutinizing. In this section, the concept of probability, its properties and Bayesian approach is being presented and advantages of Bayesian approach is being included. Besides this information is given about the prior distributions used frequently in literature. In the second subsection, the basic features of entropy and some important entropy measures are presented by mentioning the concept of uncertainty and information theory and by discussing the entropy concept. Furthermore, Shannon Entropy Measure, entropy concepts for the discrete and continuous distributions and by including their joint,

conditioned entropy definitions, the relation between the discrete and differential entropy is being examined. In the third subsection, maximum entropy principle is examined in detail within Bayesian Logical Inference framework.

In the third chapter, we describe the tools and applied methods we use through this thesis.

In the fourth chapter, real world application of Bayesian Maximum Entropy Probabilities given in the second chapter are presented. All of the results are interpreted and the future studies are mentioned.

In the fifth chapter, we review the study, in general.

June 2016, 111 pages.

Keywords: Bayesian Logical Inference, Entropy, Probability Distributions, The Principle of Maximum Entropy.

1. GİRİŞ

Tersinir ve tersinmez süreç olarak tanımlanan kavramlar, fen bilimlerinin önemli çalışma alanlarından birisidir. Bu süreçleri bir örnekle açıklamak gerekirse, kırılan cam bardağı örneği verilebilir. Bir cam bardağın yer çekimi etkisiyle kırılması, formunun değişmesine neden olmaktadır. Yer çekimi ile kırılan bir cam bardağın kırıldıktan sonra yer çekimi etkisine eş değer bir müdahale yapıldığında, sağlam bardak haline dönüşmesi mümkün olsaydı, tersinir süreçten bahsedilebilirdi. Bu durumun gerçekleşmesi mümkün olmaması nedeniyle bu durum *tersinmez süreç* olarak tanımlanır.

19. yüzyıl, bilimsel çalışmaların teknolojik ilerlemelere dolayısıyla sanayi devrimine neden olan yüzyıldır. Bu yüzyılda bilim adamları, bilimsel çalışmalar sonucunda termodinamik kanunlarını da ortaya çıkarmıştır. Entropi kavramı, fizikte termodinamik kanunu ile açıklanmaktadır. Termodinamikte iki temel kanun söz konusudur. Bunlardan birincisi, evrendeki tüm enerjinin sabit olduğunu, onun ne yaratılabilir ne de yok edilebilir olduğunu ileri sürer. Entropi yasası olarak bilinen ikinci kanun ise enerjinin farklı dağılımından eşit dağılımına ne ölçüde dönüştüğünü açıklamaktadır. Dolayısıyla entropi, kapalı ve tersinir bir sistemde artar.

Bilim dünyasında Entropi kavramını ilk defa kullanan Alman Fizikçi Rudolf Clausius olmuştur. *Enerjinin Korunumu İlkesi* ve *Entropi* kavramları üzerine çalışan Clausius; entropiyi kapalı bir sisteme giren ısının ve sistemin sıcaklığının bir fonksiyonu olarak tanımlamıştır. Clausius, entropinin enerji gibi korunduğunu ve sürekli arttığını belirtmiş, bunu da evrenin bir ısı ölümüne gittiğinin göstergesi olarak yorumlamıştır [1].

Termodinamiğin ikinci yasasına istatistiksel yorum getiren Ludwig Boltzmann olmuştur ve ünlü Boltzmann bağıntısını ortaya koymuştur [2].

Entropi;

$$S = k \log W \tag{1.1}$$

olarak ifade edilir. Denklem (1.1), S (Entropi) ile bir sistemin girebileceği mikroskobik durumların sayısının olasılık şeklinde ifadesi olan W arasındaki ilişkiyi gösterir. Burada k, Boltzmann sabitidir.

Uzun süre fizikçilerin çalışma alanına giren entropi, 1948 yılında *İletişim Modeli* üzerine çalışan Claude Shannon' un, John von Neumann' ın çalışmaları ile belirsizliğin ölçüsü olarak ileri sürülmüştür. Shannon Entropisi olarak da bilinen bu kavram;

$$H = -K \sum_{i=1}^{n} p_i \cdot \log p_i \tag{1.2}$$

şeklinde tanımlanmıştır. Burada K pozitif bir sayı olmak üzere, H entropiyi, p_i ise bir sistemde $i.$ aşamadaki olasılığı ifade eder.

Entropi günümüzde termodinamik, istatistiksel mekanik, istatistiksel yorum yapma, yönetim teorisi, finansman, parametre tahmini, doğrusal ve doğrusal olmayan programlama gibi çok farklı alanlarda kullanılabilen bir kavram haline gelmiştir.

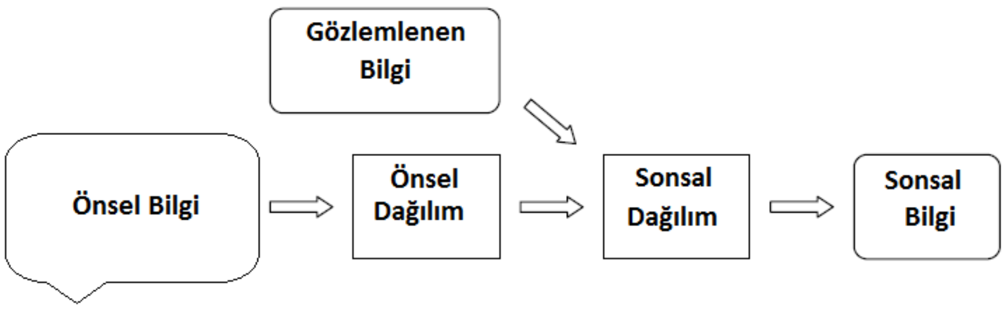
Shannon, entropi kavramını geliştirerek birçok farklı disiplinlerde kullanılmasını sağlamıştır. Maksimum Entropinin gelişmesine olanak veren Shannon Entropisi, verilen dağılımın olasılık değerlerine bağlı olarak hesaplanır. Maksimum entropi ilkesinde bu durum tersine dönmekte ve bir rastlantı değişkeninin ortalama değerlerine ilişkin bilgi verildiğinde, bu bilgi ile tutarlı olan en iyi dağılım seçilmektedir.

1957 yılında Jaynes, maksimum entropi ilkesini kullanarak istatistiksel mekaniği olasılık dağılımları açısından incelemiştir. Jaynes, bütün uygun dağılımlar arasından entropiyi maksimum yapacak olan dağılımın seçilmesi gerektiğini belirtmiştir.

Entropiyi ters problemlerin çözümü için tahmin yöntemi olarak kullanmıştır. Ters problemler, yetersiz gözlem olması durumunda bu gözlemleri kullanarak bilinmeyen parametreleri belirlemektedir. Yani, Shannon Entropisini maksimize ederek elde ettiği bu tahmin yöntemini *Maksimum Entropi* olarak adlandırmıştır.

Maksimum Entropi ilkesine göre, eldeki bilgiye bağlı kalınarak, verilmeyen bilgi maksimize edilir. Buradaki temel kavram belirsizliktir.

Bayesli yaklaşım [3], özü Bayes Teoremine dayandırılarak yapılandırılmış öznel düşünceleri yansıtan bir yaklaşımdır. Temelini 1763 yılında Thomas Bayes' in koşullu olasılık ve toplam olasılık tanımına dayalı Bayes teoreminden alan bu yaklaşım; parametre tahminlerinde, örneklem bilgisiyle bilinmeyen parametre hakkındaki öznel bilgileri seçip bir araya getirerek sonsal dağılımları oluşturur. Ayrıca elde edilen veri ve diğer bilgiler ışığında uygun olan model için tercih yapılmasını sağlar. Bayesli mantıksal çıkarımın, mühendislik, tıp, fizik, matematik gibi birçok alanda sinyal işleme ve radar gibi pek çok uygulamaları vardır [4].



Şekil 1.1: Bayesli bilgi şeması.

Genel olarak Bayesli mantıksal çıkarımı özetleyen bilgi şeması Şekil 1.1'de gösterilmiştir [5]. Burada önsel bilgiyi şu şekilde açıklamak mümkündür. Bilim adamının verileri gözlemlemeden önce parametre hakkında psikolojik durumunu ifade eden duygu, düşünce, deneyim ve daha önceki bilim adamlarının ifadelerinden elde edilen bilgi önsel bilgi olarak adlandırılmaktadır. Önsel bilgiye karşılık gelen önsel olasılık ile örneklemden olabilirlik fonksiyonu aracılığıyla gelen objektif bilginin birleştirilmesi sonucunda sonsal olasılığa ulaşılmaktadır. Sonsal bilgi sonucundaki sonsal olasılık ise veriyi gözlemledikten sonra tahmin edilecek parametre hakkında çıkarım yapmak ve karar vermek için kullanılmaktadır.

Bu tezin ana amacı, entropi fonksiyonu maksimize edilerek elde edilen maksimum entropi dağılımının Bayesli mantıksal çıkarımı çerçevesinde modellenmesini sağlamaktır.

Biz bu tezde, bir problemdeki herhangi bir belirsizliği ortadan kaldırmak için gerekli maksimum bilginin belirlenmesi işleminde verilmeyen bilgiye en uygun belirsizliklerin maksimize edilmesi ile ilgileneceğiz. Bu amaçla belirli veri kümeleri için farklı moment fonksiyonları kullanılarak Bayesli mantıksal çıkarım çerçevesinde maksimum entropi dağılımları elde edilecektir. Elde edilen dağılımlar kullanılarak gerçek veri problemlerinin maksimum entropi çerçevesinde çözümleri araştırılacaktır.

2. GENEL KISIMLAR

2.1. BAYESLİ ANALİZ

Bu bölümde Bayesli olasılık teorisi ile Frekansçı olasılık teorisi kavramları irdelenerek karşılaştırmaları yapılacaktır.

2.1.1. Olasılık ve Bayes Teoremi

Olasılık teorisi, matematiğin belirsizlik taşıyan olaylarıyla ilgilenen bir bilim dalı olup rastgele değişkenleri inceler.

Bir rastgele değişkenin gelecekteki bir gözlem sırasında alacağı değer kesin olarak bilinemez ancak değişkenin belli bir değer alma şansı belirlenebilir. Dolayısıyla gerçekleşmesi rastlantıya bağlı olan olaya *rastgele olay* denir. Örneğin, bir zar atışında seçilen sayının (1 ile 6 arasında) görülmesi bir rastgele olaydır.

Olasılık teorisinde bir rastgele olayın meydana gelme şansı, *olasılık* adı verilen bir büyüklük ile ifade edilmektedir. S, örnek uzay olmak üzere, bir A olayının gerçekleşme olasılığı $P(A)$ olduğunda, bu olasılık aşağıdaki üç koşulu sağlar [6]:

- Her A olayı için $P(A) \geq 0$
- $P(S) = 1$
- $A_1, A_2, ..., A_n$ bir S örnek uzayında sonlu veya sayılabilir sonsuzlukta ikişerli ayrık olaylar ise,

  $$P(A_1 \cup A_2 \cup ...) = P(A_1) + P(A_2) + ... + P(A_n)$$

  yazılır.

Olasılık, bilimsel karar vermede önemi bir yer tutmaktadır. Bu konuda yapılan araştırmalar, tahmin sürecinin temel noktasıdır. Aşağıda olasılık türlerinin tanımları gösterilmektedir:

Tanım 2.1. Bir deneyin, çıkabilecek sonuçları göz önüne alarak matematiksel hesaplar yaparak bir olayın olma olasılığının hesaplanmasına *teorik olasılık* denir. Yani bir deney, eşit olasılıklı N farklı sonuç verirse ve bu sonuçların M tanesi bir A olayına uygun ise, A olayının gerçekleşme olasılığı $P(A)$ ile gösterilir ve

$$P(A) = \frac{M}{N} \tag{2.1}$$

yazılır. Burada $0 \leq P(A) \leq 1$ eşitsizliği her A olayı için geçerlidir.

Örnek 2.1. Bir zar atılınca üst yüze tek sayı gelme ihtimali hesaplansın. Zar üzerinde 6 rakam vardır ve bunların 3 tanesi tek sayıdır.
O halde, $P(A) = \frac{M}{N} = \frac{3}{6} = \frac{1}{2}$ dır.

Tanım 2.2. Bir olayın olma olasılığını yapılan denemelere göre bulmaya *deneysel olasılık* denir. Bir olayın olasılığını bulmak için yapılan deneyde olayın gerçekleşme sayısının deneme sayısına oranına olayın deneysel olasılığı denir.

Örnek 2.2. Bir zar 8 kere atılsın ve üst yüze gelen sayıları not edilsin.
$2, 3, 6, 3, 4, 6, 1, 2$
Zarın üst yüzüne 1 gelme olasılığı deneysel olarak: $\frac{1}{8}$ dir.

Tanım 2.3. Bir olayın olma olasılığı hakkında kişilerin kendi görüşlerine göre verdikleri olasılık değerine *öznel olasılık* denir.

Örnek 2.3. A takımı ile B takımı arasındaki maçın sonucu için bir kişi: *"Bence A takımı* $\%80$ *kazanır."* demesi öznel olasılıktır.

Bu tanım kişisel bir tanım olup, 1763 yılında Thomas Bayes tarafından ortaya atılan Bayes Teoremine dayanmaktadır.

2.1.1.1. Koşullu Olasılık

A ve B herhangi iki olay ve $P(B) > 0$ olsun. B olayının olması koşulu ile A olayının gerçekleşmesinin koşullu olasılığı $P(A|B)$ ile ifade edilir ve

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{2.2}$$

ile verilir. Buradan,

$$P(A \cap B) = P(A|B)P(B)$$

veya A ile B olaylarının simetrik konumda olmalarını dikkate alarak, $P(A) > 0$ olması koşuluyla

$$P(A \cap B) = P(B|A)P(A)$$

yazılabilir.

Koşullu olasılık, olasılık ve istatistik biliminde önemli bir yere sahiptir. Toplam olasılık ve Bayes Teoremi, karar alma konularının temelinde yer almaktadır. Koşullu olasılık ve toplam olasılık formülünü içeren Bayes Teoremi, koşullu olasılıkların hesaplanmasında kullanılmaktadır.

Bayes Teoremi, olasılık kuramının önemli bir konularından birisi olup, bir rastgele değişken için olasılık dağılımı içinde koşullu olasılıklar ile marjinal olasılıklar arasındaki ilişkiyi gösterir. Dolayısıyla Bayes Teoremi, sonuçların (çıktıların) verilip, sebeplerin (girdilerin) bulunması yani geriye doğru analiz yapmayı sağlar.

2.1.1.2. Toplam Olasılık Formülü

$A_1, A_2, ..., A_n$' ler bir S örnek uzayının bir parçalanması ise S' deki herhangi bir B olayı için,

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i) \tag{2.3}$$

olarak hesaplanmaktadır.

2.1.1.3. Bayes Teoremi

$\forall i \in I$ için $A_i \neq \emptyset$ olmak üzere, $A_1, A_2, ..., A_n$ ile S örnek uzayının bir bölünüşü verilsin. Bir başka deyişle; $\forall\ i \neq j$ için $A_i \cap A_j = \emptyset$ ve $A_1 \cup A_2 \cup ... \cup A_n = S$ olsun. B ise Şekil 2.1'de görüldüğü gibi, $A_1, A_2, ..., A_n$ olaylarından farklı bir olay olsun. B olayı bilindiğinde herhangi bir A_n olayının olma olasılığı, Bayes Teoremine göre aşağıdaki formül ile hesaplanmaktadır [7].

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} \tag{2.4}$$

(2.3) eşitliği kullanılarak,

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum\limits_{i=1}^{n} P(A_i)P(B|A_i)} \tag{2.5}$$

elde edilir.

Bayes Teoremi formülünde bulunan her bir olasılık çarpanı farklı isimlerle adlandırılır.

- $P(A_i)$ olasılığı herhangi bir A_i için *önsel olasılık* veya *marjinal olasılık* olarak adlandırılır. Bu önseldir, çünkü B olayı hakkında önceden herhangi bir bilgiyi içermemektedir.
- $P(A_i|B)$, B olayı bilindiğinde A_i olayının koşullu olasılığı olup *sonsal olasılık* olarak adlandırılır.
- $P(B|A_i)$, A_i olayı bilindiğinde B olayının koşullu olasılığıdır ve *olabilirlik fonksiyonu* olarak adlandırılır.
- $P(B)$ terimi B olayı için *önsel olasılıktır* veya B' nin marjinal olasılığıdır ve sonsal olasılığı yani $P(A_i|B)$' yi normalize eden sabittir.

Şimdi Denklem (2.5) deki eşitliğin ispatını yapalım:

$A_1, A_2, ...A_n$ olayları ayrık ve birleşimleri örnek uzay olduğundan,
$S = A_1 \cup A_2 \cup ... \cup A_n$ ve $A_i \cap A_j = \emptyset, \quad i = 1, 2, ..., k; \quad j = 1, 2, ..., k; \quad i \neq j$ yazılır.
B olayı, $A_1 \cap B$, $A_2 \cap B$,..., $A_n \cap B$ ayrık olayların birleşimidir ve

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup ... \cup (A_n \cap B)$$

biçiminde yazılır. $A_1 \cap B$, $A_2 \cap B$,..., $A_n \cap B$ olayları ayrık olduklarından, B' nin marjinal olasılığı,

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + ... + P(A_n \cap B)$$

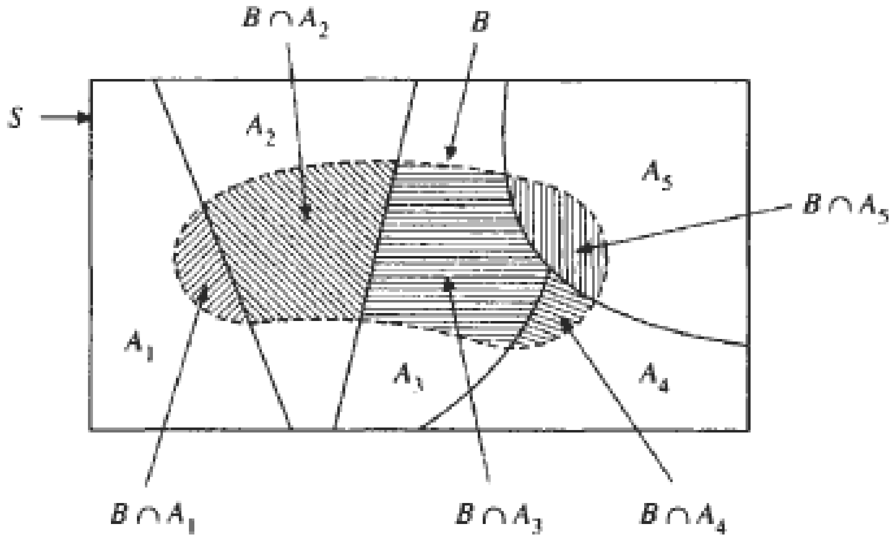
olarak ifade edilir. Denklem (2.4) kullanılırsa,

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i \cap B)}{P(A_1 \cap B) + P(A_2 \cap B) + ... + P(A_n \cap B)} \tag{2.6}$$

$$= \frac{P(A_i \cap B)}{P(A_1)P(B|A_1) + ... + P(A_n)P(B|A_n)} \quad i = 1, 2, ...n$$

$$= \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)} \quad i = 1, 2, ..., n$$

elde edilir.



Şekil 2.1: Bayes formülünü ifade eden şekil.

Aşağıdaki örnekler, Bayes Teoreminin anlaşılmasında yardımcı olacaktır.

Örnek 2.4. [7] Bir fabrikada üretilen malların $\%60$' ı $1.$ makineden, $\%30$' u $2.$ makineden ve $\%10$' u $3.$ makineden üretilmektedir. Bu makinelerin ürettikleri malların sırasıyla $\%7$, $\%4$ ve $\%1$' inin kusurlu olduğu gözlenmiştir. Rastgele seçilen bir ürünün bozuk olduğu bilindiğine göre, bunun $1.$ makine tarafından üretilme olasılığı nedir?

Çözüm:

A_i: Seçilen mal $i.$ makineden üretilmiştir. ($i = 1, 2, 3$)

B: Seçilen mal bozuktur.

$P(A_1) = 0.60$, $P(A_2) = 0.30$, $P(A_3) = 0.10$,
$P(B|A_1) = 0.07$, $P(B|A_2) = 0.04$, $P(B|A_3) = 0.01$ olmak üzere, seçilen bir ürünün bozuk olması olasılığı:

$$\begin{aligned}P(B) &= P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3)\\ &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)\\ &= (0.07).(0.60) + (0.04)(0.30) + (0.01)(0.10)\\ &= 0.042 + 0.012 + 0.001 = 0.055\end{aligned}$$

bulunur. O halde bunun 1. makineden üretilme olasılığı ise,

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(B)} = \frac{(0.60)(0.07)}{(0.056)} \approx 0.76$$

olur.

Örnek 2.5. [7] Arkeolojik kazılarda araştırma yapılan bir bölgede tarihi eser bulunması olasılığı $\%30$ ve eldeki araçların o bölgede tarihi eser olduğuna dair doğru bilgi vermesi olasılığı $\%60$ dır. Yapılan tarama sonucu bir bölgede *"tarihi eser vardır"* sinyali verildiğine göre bu bölgede gerçekten tarihi eser bulunması olasılığı nedir?

Çözüm:

A: Araştırmanın yapıldığı bölgede tarihi eser olması,
B: Eldeki araçların doğru bilgi vermesi olsun.
$P(B|A) = 0.60$, $P(B|A^c) = 0.40$ olmak üzere, *"tarihi eser vardır"* sinyali verildiğinde bu bölgede tarihi eser olması olasılığı:

$$\begin{aligned}P(A|B) &= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)}\\ &= \frac{(0.30).(0.60)}{(0.30).(0.60) + (0.70).(0.40)} = \frac{0.18}{0.46} \approx 0.39\end{aligned}$$

bulunur.

Aşağıda günümüzde bilimsel öğrenme ve karar vermede oldukça yeni olan ve güçlü bir bakış açısı sağlayan Bayesli Mantıksal Çıkarım kavramına yer verilmektedir.

2.1.2. Bayesli Yaklaşım

Hayatın büyük bir kısmı belirsizlikler içinde geçmektedir. Bu belirsizlik bilgi eksikliğine, geçmişte yaşananlara veya gerçekleşmemiş durumlara bağlı olabilir. Günlük yaşamda sezgiler ve düşünceler bilimsel amaçlar için kullanılabilir [8]. Bununla birlikte, teorik tahminlerin deneysel testleri basitçe evet veya hayır cevabını vermez. Eldeki bilginin durumu hep eksiktir. Bu eksikliğin giderilmesi için başka deneyler de yapılabilir fakat bunların ölçüm doğruluğu sınırlıdır. İstatististiksel çıkarım, eksik bilgiye dayanarak yapının teoride gerçeği çıkarma sürecidir. Bilimde, basit modellerden karmaşık modellere doğru bir süreç akışı vardır. Teorileri değiştirmek için model tahminlerinin kullanılması, ölçümler arasındaki farklarla öğrenilmektedir. Bayesli yaklaşım [9] bu farkların ve oluşumların bilimsel olarak kullanılmasında alternatif olarak ortaya çıkmıştır.

Geleneksel istatistikte bir olayın olma olasılığı, olayın geçtiği uzun dönemde göreceli frekans ile tanımlanır. Bu, yaygın olarak *Frekansçı Yaklaşım*[8] olarak adlandırılır. Bir başka deyişle Frekansçı yaklaşım, gözlenmiş veriler yardımı ile hakkında bilgi sahibi olunması istenen kitleye ait sonuçlar çıkarılmasıdır. Bu yaklaşımda olasılıklar rastgele değişkenler ile sınırlıdır. Miktarları da anlamlı bir şekilde tekrarlanan bir dizi, deney boyunca değişir. Bu bir dizi deney içinde hipotez testleri, varyans analizi, örnekleme dağılımları ve tahmin etme gibi istatistiksel analizler yani klasik yaklaşımlar bulunur [10].

Frekansçı ekolüne dahil olanlar olasılık değerlerini rastgele olaylarda meydana gelme sıklığına göre saptanması veya anakitlenin örneklemlerinin anakitleye orantısı olarak saptanması gerektiğini kabul etmektedirler. Bunlara göre yeni kanıtlar karşısında, olasılık değerinin değişme imkanı yoktur. Bu nedenle frekansçı ekolü için Bayes Teoremi sadece koşulluluklar arasındaki ilişkiyi gösterir ve bunun pratikte kullanılma gücü düşüktür. Frekansçılar, olasılığı yansız bir olay olarak görür.

Günümüzde farklı bir yaklaşım olarak, istatistiksel analizlerde Bayesli olasılık teorisi kullanılmaktadır.

Frekansçı yaklaşımcılar, 20. yüzyıla kadar Bayes Teoreminden yararlanmıştır. Son dönemlerde, Bayesli yaklaşım ön plana çıkarak önemli çözüm yöntemleri üreterek yüksek boyutlu modellerle ilgili problemlerin basitleştirilmesine olanak sağlamıştır. Bu yüzden Bayesli yaklaşım, istatistiksel analizde alternatif bakış açısı getirmesi nedeniyle 20. yüzyılın son döneminde daha üstün bir yaklaşım olarak görülmektedir. Dolayısıyla bu yaklaşım daha fazla kullanım alanına sahiptir.

Bayesli yaklaşım [11], mevcut bilinen teorilerin arıtılmasını, yeni teorilerin gelişmesini ve sürecin devam etmesini sağlamıştır. Bu süreçte tümdengelimli çıkarımın rolü, özellikle test edilebilir bir teorinin tahminler türetmesi ile ilgili olmasıdır. Yeni bir olasılık algısı olarak ortaya çıkan Bayesli yaklaşımda olasılığın matematik kuralları, sadece rastgele değişkenlerin frekanslarını hesaplamaz. Burada durum, olay veya problem 2 aşamadan oluşmaktadır. Aşamalar aşağıda gösterilmektedir [12]:

i) Birinci aşama, var olan objektif ve subjektif bilgilere göre tahmin edilmek istenen parametrelere ilişkin başlangıç dağılımından (Bilgi Öncesi Dağılımı) oluşur.

ii) İkinci aşama ise örneklemden elde edilen ek bilgilere göre parametrenin bilgi sonrası dağılımı belirlenmektedir ve bu dağılıma göre maksimum kazanç veya minimum zarar değerini veren fonksiyon bulunarak sonuca ulaşılır.

Bayesli çıkarımda bir olayın olasılığı, araştırmayı yapan kişinin araştırmaya özgü inanç düzeyi, bir başka ifade ile ön bilgisi ile denemeden elde edilen sonuçların bütünleştirilmiş halidir. Bu yaklaşım, kişinin geçmişten gelen ön bilgisi, verilerin objektif bilgisi ile birleşerek çıkarımların yapılmasına imkan veren bir yöntemdir. Dolayısıyla Bayesli yaklaşımda önsel bilgilerden önsel dağılımlara, sonsal dağılımlardan da sonsal bilgilere bir geçiş durumu söz konusudur. Bayesli yaklaşıma farklılık ve üstünlük getiren özelliklerden birisi de araştırmacının kendisidir [13].

İstatistiksel modellemelerin temel amacı olan Bayesli yaklaşım, bilinmeyen model parametrelerini tahmininde önemli faydalar sağlar. Jaynes bunu, parametre tahmininde önsel ve sonsal dağılımın, parametrenin herhangi bir ölçülebilir özelliğini değil, parametre hakkında bireylerin sahip olduğu bilgilerin düzeyini temsil etmesi olarak ifade eder. Bu yüzden Bayesli yaklaşımda bilinmeyen parametreler, birer rastgele değişkendir ve bu parametre üzerindeki bilgiler bir olasılık dağılımı yardımıyla temsil edilirler [3].

Tümevarım ve tümdengelim mantık tarafından kapsanan Bayesli analiz, bilginin mevcut durumuna dayanarak, tüm durumları inceleyip olaylara bir bütün içerisinde yaklaşır. Bayesli istatistik, olasılık ve mantık arasında bağlantı kurmaz; ön bilgiler ışığında hareket eder [4].

Olasılıkta toplam ve çarpım kuralları çok önemlidir. Bayesli olasılık tanımı ile birlikte, mantıksal çıkarımda eksik görülen bir durumu ele almak için bu 2 kural ile istenilen bulunur. O halde bu kurallar şu şekilde tanımlanır:

a)Toplam Kuralı:

Bir B olayı bilindiğinde bir A olayının olması veya olmaması olasılığı:

$$P(A|B) + P(\overline{A}|B) = 1 \tag{2.7}$$

olur. Burada; $\overline{A}$: A'nın olmama olayını temsil eder.

b)Çarpım Kuralı:

Bir C olayı bilindiğinde A ve B' nin koşullu olasılığı:

$$P(A, B|C) = P(A|C)P(B|A, C) = P(B|C)P(A|B, C) \tag{2.8}$$

yazılır. Burada, $P(A, B|C)$, ***Ortak Olasılık*** olarak adlandırılır. Denklem (2.8) de $P(A|B, C)$ yalnız bırakılırsa,

$$P(A|B, C) = \frac{P(A|C)P(B|A, C)}{P(B|C)} \tag{2.9}$$

Bayes Teoremine uygun formulasyon elde edilir. Toplam Kuralının genişletilmiş hali,

$$P((A + B)|C) = P(A|C) + P(B|C) - P(A, B|C) \tag{2.10}$$

biçiminde yazılır. Bu eşitliğin ispatını yapalım:

$$\begin{aligned}
P((A+B)|C) &= 1 - P(\overline{A+B}|C) = 1 - P(\overline{A},\overline{B}|C) \\
&= 1 - P(\overline{A}|C)P(\overline{B}|\overline{A},C) \\
&= 1 - P(\overline{A}|C)[1 - P(B|\overline{A},C)] \\
&= 1 - P(\overline{A}|C) + P(\overline{A}|C)P(B|\overline{A},C) \\
&= P(A|C) + P(\overline{A},B|C) \\
&= P(A|C) + P(B|C)P(\overline{A}|B,C) \\
&= P(A|C) + P(B|C)[1 - P(A|B,C)] \\
&= P(A|C) + P(B|C) - P(B|C)P(A|B,C)
\end{aligned}$$

olur. Denklem (2.8) kullanılırsa,

$$P((A+B)|C) = P(A|C) + P(B|C) - P(A,B|C)$$

elde edilir. Diğer yandan Denklem (2.9) baz alınırsa Bayes Teoreminin bir başka formu şu şekildedir:

$$P(H_i|D,I) = \frac{P(H_i|I)P(D|H_i,I)}{P(D|I)} \tag{2.11}$$

Bu denklemde,

. $H_i \equiv$ İlgilenilen hipotezin gerçeğini ileri süren önermedir.

. $I \equiv$ Önceden sahip olunan bilgi, inancımızdır.

. $D \equiv$ Verileri temsil eden önermedir.

. $P(D|H_i,I)$ $=$ Olabilirlik fonksiyonu olup sebep-sonuç arasındaki ilişkiyi sağlayan fonksiyondur.

. $P(H_i|I) =$ Önsel Olasılıktır.

. $P(H_i|D,I) = H_i$ için sonsal olasılıktır.

. $P(D|I) =$ Kanıt, normalizasyon sabiti veya marjinal olasılıktır.

Bu durum ile ilgili örnek aşağıda verilmiştir:

Örnek 2.6. [12] 1996 yılına ait bir gazete makalesinde, Toronto'da doktorların bir şirkette e-mail siparişi ile onaylanmamış HIV tükürük testi sattığı konusunda endişeli oldukları bildirilmiştir. Labarotuvardaki testlere göre, bu test için yanlış pozitif oranı $\%2.3$ ve yanlış negatif oranı $\%1.4$' dür. Yani bu teste dayanarak insanların $\%98.6$ 'sının aslında hastalığı vardır.

Bu örnekte, varsayalım ki yeni bir ölümcül hastalık keşfedilsin ve hastalığın bilinen bir belirtisi olmasın fakat tükürük testi yukarıdaki özellikler için mevcut olsun. Sebebi bilinmeyen hastalık kısaca UD ile gösterilsin. Şüphelenmek için bir neden olmayıp hasta UD'ye sahip olsun fakat testin pozitif olması her şeyi belirleyecektir. Gerçekten bu hastalığın hastada görülme olasılığı nedir?

Çözüm:

Burada Bayesli analiz kullanılacaktır. Bu analizin amacını göstermek için bir rastgele örneklemindeki hastalığın bir bölgede görülme sıklığı $\frac{1}{10000}$ olarak varsayılacaktır. Problemin çözümü için aşağıdaki tanımlamaları yapalım:

$H \equiv$ Hasta UD' ye sahip olsun.

$\overline{H} \equiv$ Hasta UD' ye sahip olmasın.

$D_1 \equiv$ Hasta UD için testin pozitif çıkması olsun.

$I_1 \equiv$ UD için bilinen bir neden olmaması olsun.

$P(D_1|H, I_1) = 0.986$, $P(D_1|\overline{H}, I_1) = 0.023$ ve popülasyonda UD' nin görülme sıklığı $\frac{1}{10^4}$ dür.

Başlangıç noktası olarak Bayes Teoremi yazılırsa,

$$P(H|D_1, I_1) = \frac{P(H|I_1)P(D_1|H, I_1)}{P(D_1|I_1)}$$

olur. $P(D_1|I_1)$ normalizasyon sabiti olduğu için $\sum_i P(H_i|D_1, I_1) = 1$ eşitliğini sağlar ve aşağıdaki eşitlik yazılabilir:

$$P(D_1|I_1) = P(H|I_1)P(D_1|H, I_1) + P(\overline{H}|I_1)P(D_1|\overline{H}, I_1)$$

Verilen değerler yerine yazılırsa,

$$P(H|D_1, I_1) = \frac{10^{-4} \cdot 0.986}{10^{-4} \cdot 0.986 + 0.9999 \cdot 0.023} = 0.0042$$

Böylece hastanın, hastalığa yakalanmama olasılığının $\%98.6$ olduğu görülür.

2.1.3. Frekansçı Yaklaşımla Bayesli Yaklaşımı Arasındaki Farklar

Olasılık tanımlarındaki farklılık, istatistikte Frekansçı yaklaşım ve Bayesli yaklaşım adı verilen iki farklı alanın oluşumuna neden olmuştur. Bu alanda Frekansçı yaklaşım daha fazla hakim olmasına rağmen Bayesli yaklaşımının da olasılık literatüründe önemli ölçüde etkileri vardır. Bu yaklaşımların birbirlerinden farkları aşağıda belirtilmiştir [12], [14]:

i) Alternatif bir yaklaşım olan Bayesli yaklaşım; bir teori, önerme veya bir nedenselliği, belirsiz olasılıklarla ifade etmektedir. Paradigma olasılık olup teori, önerme ve nedenselliği anlamaya, yönetmeye ve kontrol etmeye yarayan güçlü bir araçtır. Belirsizlik, Bayesli yaklaşımda ne tamamıyla kabul edilir ne de reddedilir. Bu yaklaşımda belirsiz olan ilişki, olasılığın hesaplanmasında ve karar verme sürecinde daha kesin ve net sonuçlar verir.

Frekansçı yaklaşım ise söz konusu belirsizliği, varsayımlar doğrultusunda problemde iddia edilen ilişkiye, sıklıklarına göre değerlendirerek kabul edilmesi ya da reddedilmesi yönünde karar verir.

ii) Nesnel olasılık kavramını savunan istatistikçiler Frekansçı yaklaşımı, öznel olasılık kavramını savunan istatistikçiler ise Bayesli yaklaşımı tercih etmişlerdir.

iii) Bayesli yaklaşımın, Frekansçı yaklaşımdan en önemli farkı *önsel bilginin* kullanılmasıdır. Önsel bilginin kullanımı, nesnel ve subjektif analiz tartışmalarına neden olmuştur. Subjektif analize yönelmesi nedeniyle eleştirilere maruz kalan Bayesli yaklaşım, Frekansçı yaklaşımı benimseyenler tarafından eleştirilmektedir. Bu tartışmalara ek olarak Bayesli yaklaşımda kendi disiplini içinde, nesnel ve subjektif analiz tartışmaları yapılmaktadır.

iv) Bayesli yaklaşım, doğrudan herhangi bir teorinin olasılığını ve bir modele ait parametrenin özel bir değerini hesaplamayı sağlar. Frekansçı yaklaşım ise sadece dolaylı olarak rastgele değişken istatistiğini kullanır.

v) Frekansçı yaklaşımda bir A olayının olasılığı $P(A)$, bir deneyin özdeş olarak tekrarlanmasıyla A olayının olması sıklığıdır. A önermesinde rastgele değişkenler hakkında bilgi sınırlıdır.

Bayesli yaklaşımda bir $P(A|B)$ olasılığı, bir B önermesinin veya hipotezinin doğruluğunun verilmesi durumunda A önermesinin olabilirliğinin ölçüsüdür. A, herhangi bir mantıksal önerme olabilir ve rastgele değişkenler hakkında sınırlandırma yoktur.

vi) Frekansçı yaklaşım tanımı, *aynı tekrarları* içerir. Tekrarlanan deneyler hiçbir zaman her şeyiyle aynı olmayabilir. Bayesli olasılık tanımı ise nicel sonuçlara kesin olarak ulaşmak için belirsiz olan *olasılık* terimini içerir.

vii) Frekansçı yaklaşım tümdengelim yöntemi ile paralellik gösterirken, Bayesli yaklaşım ise tümevarım yöntemiyle paralellik gösterir. Çünkü Frekansçı yaklaşım, nedensellik ilkesinin deterministik yorumuna yakın görünürken; Bayesli yaklaşım olasılık yorumuna yakındır.

viii) Bayesli yaklaşımda parametre, olasılık dağılımı olan bir rastgele değişkendir. Çıkarım; önsel olasılık ve mevcut verilerden oluşturulan son olasılık dağılımından yapılır. Burada parametreler bilinmez ve olasılıklarla ifade edilir.
Frekansçı yaklaşımda ise parametreler sabit ve süreç boyunca süreklidir. Parametre tahmini sadece eldeki veriye dayanarak hesaplanır.

viiii) Bayesli yaklaşımda amaç, denemeler yaparak en yüksek olan 1 olasılığına, yani; kesinliğe ulaşmaktır. Bu yol, doğrulamalar yapılarak ilerlenen bir yoldur.
Frekansçı yaklaşımda ise yanlışlama esas olduğundan amaç 0 olasılığına ulaşmaktır.

Tablo 2.1: Bayesli ve Frekansçı Yaklaşıma İlişkin Olasılık Süreci.

Bayesli Yaklaşım	Frekansçı Yaklaşım
Varsayımlarsız $\Downarrow$	Varsayımlarla $\Downarrow$
Deneme $\Downarrow$	Deneme $\Downarrow$
Doğrulama $\Downarrow$	Yanlışlama $\Downarrow$
$\frac{1}{2} \ll P(t,g) < 1$	$P(t,g) = 0$

Bayesli ve Frekansçı yaklaşım arasındaki farklar Tablo 2.1'de gösterilmiştir [14]. Burada *t* bir teori, *g* ise onunla ilgili gözlemlerdir.

Diğer taraftan Bayesli olasılıksal ve istatistiksel analizin avantajları ise aşağıda verilmiştir [12], [13].

2.1.4. Bayesli Mantıksal Çıkarımın Avantajları

- Bayesli yaklaşımda bilgi, belirli bir durumu ifade ettiğinden, herhangi bir bilimsel soru ve problemin cevabı için optimal bir şekilde basit ve akılcı bir yaklaşım sağlar. Bu durum geleneksel istatistiksel analiz yaklaşımına zıttır. Burada prosedür bellidir. Bu prosedür aşağıdaki gibidir:
  a) Açıkça sorunuzu ve ön bilginizi daima ifade edin.
  b)Toplam ve çarpım kurallarını uygulayın. Başlangıç noktası olarak Bayes Teoremini alın.
  Bazı problemler için, Bayesli analiz basit tanıdık bir istatistik olabilir. Bu durumda bile genellikle güçlü, yeni bir anlayış, istatistiğin yorumlanması ile ilgili bilgi sağlar.

- Hipotezin, $P(H_i|D, I)$ olasılıklarını direkt hesaplar.

- Marjinalleşme ile sıkıntılı parametreleri ortadan kaldırmak için bir yol sağlar. Bazı problemler için marjinalleşme analitik olarak yapılabilir, bazı hesapları doğrudan çözmeye izin verir.

- Ölçüm işlemi ve teorik model tahminlerinden kaynaklanan sistematik hataların etkilerini birleştirmek için bir çözüm yöntemidir.

- Parametre hakkında sahip olunan tüm bilgi modellendiği için, çıkarım yapma sürecinde bilgi kaybı olmamaktadır.
- Mevcut bütün bilgi doğru şekilde kullanıldığı zaman daha etkili sonuçlar verir.
- Bayesli mantıksal çıkarım geçmişi görerek, sorgulama yapar.
- Zaman ve maliyet tasarrufu sağlar.
- Bayesli yaklaşımda örneklem büyüklüğü için kısıtlama yoktur. Çok az sayıda veri ile parametre hakkında anlamlı çıkarımlar yapmak mümkün olup hiç veri toplamadan bile önsel dağılım yardımıyla parametre hakkında bilgi sahibi olunabilir.
- Bayesli çıkarım ile parametreler üzerindeki belirsizlikler azaltılabilir.
- Örneklerin hangi sınıfa hangi olasılıkla ait olduklarını öngörür.

2.1.5. Bayesli Veri Analizinde Önsel Olasılıkların Belirlenmesi

Bayesli veri analizinde önsel olasılıkların nasıl atandığını belirlemeden önce kesikli ve sürekli rastgele değişkenler için gerekli tanımlar aşağıda verilmiştir.
X rastgele değişkeni kesikli olsun ve $x_1, x_2, ...$ değerlerini,

$$P(X = x_i) = f(x_i); \quad i = 1, 2, ...$$

olasılıkları ile alsın. $f(x_i)$ fonksiyonuna X rastgele değişkenin *olasılık fonksiyonu* adı verilir. Gerekli olmadığı hallerde i indisi yazılmayarak $f(x)$ formunda ifade edilir. X rastgele değişkeni kesikli ise $f(x)$ olasılık fonksiyonu şu iki koşulu gerçekler:

i) $f(x) \geq 0, \quad i = 1, 2, 3, ...$ **ii)** $\sum_{i=1}^{\infty} f(x_i) = 1$

dır.

Diğer yandan X rastgele değişkeni sürekli olsun ve bir $[a, b]$ kapalı aralığı üzerinde değerler alsın. X rastgele değişkeninin bu aralıkta bulunması olasılığı, $f(x)$ olasılık yoğunluk fonksiyonu yardımıyla,

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

şeklinde tanımlanır. Ayrıca X sürekli rastgele değişken ise,

$$P(a < X < b) = P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b)$$

eşitlikleri yazılır.
Her $f(x)$ olasılık yoğunluk fonksiyonu, aşağıdaki iki koşulu gerçekler:

i)$f(x) \geq 0$ **ii)** $\int_{-\infty}^{\infty} f(x)dx = 1$

Şimdi kesikli ve sürekli rastgele değişkenler için beklenen değer ve varyans kavramlarını açıklayalım.

Tanım 2.4. X, kesikli bir rastgele değişken olsun. X'in beklenilen değeri $E(X)$ ile gösterilir ve

$$E(X) = x_1 f(x_1) + x_2 f(x_2) + ... + x_n f(x_n) = \sum_{i=1}^{n} x_i f(x_i) \tag{2.12}$$

şeklinde tanımlanır.
X rastgele değişkeni sayılabilir sonsuzluktaki $x_1, x_2, ..., x_n, ...$ sonuçlarını alıyorsa, (2.12) eşitliği,

$$E(X) = x_1 f(x_1) + x_2 f(x_2) + ... + x_n f(x_n) + ... = \sum_{i=1}^{\infty} x_i f(x_i) \tag{2.13}$$

formunu alır.

$E(X)$ ifadesine ayrıca X rastgele değişkeninin ortalaması veya kitle ortalaması denir ve genellikle μ harfi ile gösterilir. Dolayısıyla $E(X) = \mu$ yazılır.

Tanım 2.5. X bir boyutlu sürekli rastgele değişken olsun. $f(x)$, X'in olasılık yoğunluk fonksiyonu olmak üzere, X'in beklenen değeri,

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \qquad (-\infty < x < \infty) \tag{2.14}$$

biçiminde tanımlanır.

Beklenen değerin bazı özellikleri aşağıdaki gibidir [7]:

i) b bir sabit olmak üzere $E(b) = b$ dır.
ii) $E(bX) = bE(X)$ dır.
iii) X ve Y aynı örnek uzayı üzerinde tanımlanmış bağımsız rastgele değişkenler olmak üzere:

$$E(X + Y) = E(X) + E(Y)$$

iv) b ve c sabit büyüklükler ve $Y = bX + c$ olsun. O halde,

$$E(Y) = E(bX + c) = bE(X) + c$$

v) $X_1, X_2, ..., X_n$ gibi n tane rastgele değişkeni için;

$$E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n)$$

vi) X ve Y bağımsız rastgele değişkenler ise:

$$E(XY) = E(X)E(Y)$$

yazılır. Bu özellikler X rastgele değişkeninin hem kesikli, hem de sürekli olması halinde geçerlidir.

Bir rastgele değişkeninin beklenen değeri veya ortalaması, olasılık fonksiyonun merkezi hakkında bilgi verir. Fakat bir deneyden diğerine rastgele değişkenin değerlerinin dağılımı, değişimi veya yayılması ile ilgili bilgi vermez.

Dağılım, değişim veya yayılma ölçüsü olarak varyans kullanılır ve aşağıda tanımı verilmiştir.

Tanım 2.6. $f(x)$ olasılık fonksiyonuna sahip X rastgele değişkeni kesikli ise, X'in varyansı,

$$\sigma_x^2 = var(X) = E[(X-\mu)^2] = \sum_{i=1}^{n}(x_i-\mu)^2 f(x_i) \tag{2.15}$$

ile tanımlanır. Tanımdan dolayı $var(X) \geq 0$ dır. Denklem (2.15) deki varyans tanımı aşağıdaki gibi daha basit şekilde ifade edilir:

$$\begin{aligned}
\sigma_x^2 &= E[(X-\mu)^2] \\
&= E(X^2 - 2X\mu + \mu^2) \\
&= E(X^2) - 2\mu \underbrace{E(X)}_{\mu} + \underbrace{E(\mu^2)}_{\mu^2} \\
&= E(X^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - \mu^2 = E(X^2) - [E(X)]^2
\end{aligned}$$

Diğer yandan X rastlantı değişkeni sürekli ve $f(x)$ olasılık yoğunluk fonksiyonuna sahip ise X'in varyansı,

$$\sigma_x^2 = var(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx \tag{2.16}$$

ile ifade edilir. Denklem (2.15) ve (2.16) dan da görüldüğü gibi X'in varyansı, X'in kendi ortalamasından sapmasınn karesinin ortalamasıdır.

Tanım 2.7. X, μ ortalamalı kesikli veya sürekli bir rastgele değişken olsun. X'in standart sapması σ_x, varyansın kareköküdür ve

$$\sigma_x = \sqrt{var(X)} = \sqrt{E(X-\mu)^2} \tag{2.17}$$

biçiminde formülüze edilir. Varyansın bazı özellikleri aşağıda verilmiştir [7]:

b, bir sabit olmak üzere,
i) $var(bX) = b^2 var(X)$ dır.
ii) $var(X+b) = var(X)$ dır.
iii) $var(bX+c) = b^2 var(X)$ dır.
iv) X ve Y bağımsız rastgele değişkenler olmak üzere:

$$var(X+Y) = var(X) + var(Y)$$

yazılır.
v) $X_1, X_2, ..., X_n$ gibi n tane bağımsız rastgele değişkeni için genelleştirilirse;

$$var(X_1 + X_2 + ... + X_n) = var(X_1) + var(X_2) + ... + var(X_n)$$

yazılır.
X'in kesikli veya sürekli rastgele değişkeni olması halinde bu özellikler geçerlidir.

Tanım 2.8. [6] Örneklemdeki bağımsız gözlemlerin sayısından örneklemdeki gözlemlere dayanarak tahmin edilecek kitle parametrelerinin sayısının çıkarılması ile elde edilen sayıya *serbestlik derecesi* denir.

Önsel olasılık dağılımlarının belirlenmesi, Bayesli analizin en önemli noktasıdır. Genişletilmiş mantık olarak olasılık teorisi yaklaşımı benimsendiğinde, herhangi bir çıkarım sorusunun çözümüne Bayes Teoremi ile başlanır. Önsel olasılıklar, kararı verecek kişinin ilk bilgisini yansıtmalıdır. Önsel dağılım bireysel olarak seçildiği için önsel olasılığın da bireysel olasılık olduğu varsayılmaktadır [12].

Denklem (2.11) de görüldüğü gibi, önsel bilgi olan I, hipotez uzayını tanımlar ve

Bayes Teoremindeki terimleri hesaplamak için ihtiyaç duyulan bilgileri sağlar. I, bir olasılık dağılımı içinde $P(D|H_i, I)$ için kullanılacaktır. Bilginin farklı ifadeleri, farklı olasılık dağılımlarına karşılık gelir. Bu olasılık dağılımları *örnekleme dağılımları* olarak adlandırılır [12].

Literatürde sıkça kullanılan kesikli dağılımlar aşağıda verilmiştir:

2.1.5.1. Binom Dağılımı

Bu bölümde Binom dağılımında $P(D|H_i, I)$ olabilirlik fonksiyonun seçiminde, önceden sahip olunan I bilgisinin nasıl ifade edileceği gösterilecektir [12].

$I \equiv E$ önermesi birçok kez tekrarlanan olayı temsil etsin. Bu olayın Q ve $\overline{Q}$ olmak üzere iki olası sonucu bulunsun. Bu önerme, madeni paranın havaya atılması örneği ile izah edilebilir.

Q, herhangi bir olayın sonucu olup başka bir olay için gelecek sonuçtan bağımsızdır.
Boole Cebiri aksiyomlarından,

$$E = Q + \overline{Q} \tag{2.18}$$

dır. Burada, $Q + \overline{Q}$ ifadesi mantıksal toplamı belirtir.

$Q_i : i.$ olayı temsil etmek üzere n tane olayın olası sonuçları şu şekilde yazılır:

$$E_1, E_2..., E_n = (Q_1 + \overline{Q_1}), (Q_2 + \overline{Q_2}), ...(Q_n + \overline{Q_n}) \tag{2.19}$$

Eğer çarpma işlemi doğru gerçekleşirse, 2^n terimli bir toplam olacaktır. $n = 3$ için sonuç aşağıdaki gibidir:

$$\begin{aligned} E_1, E_2, E_3 = {} & Q_1, Q_2, Q_3 + Q_1, Q_2, \overline{Q_3} + Q_1, \overline{Q_2}, Q_3 + \overline{Q_1}, Q_2, Q_3 + Q_1, \overline{Q_2}, \overline{Q_3} \\ & + \overline{Q_1}, Q_2, \overline{Q_3} + \overline{Q_1}, \overline{Q_2}, Q_3 + \overline{Q_1}, \overline{Q_2}, \overline{Q_3} \end{aligned}$$

Özel olarak $Q_1, \overline{Q_2}, \overline{Q_3}$ dizisinin olasılığı, çarpım kuralı uygulamalarından elde edilir.

$$\begin{aligned} P(Q_1, \overline{Q_2}, \overline{Q_3}) &= P(Q_1|I)P(\overline{Q_2}, \overline{Q_3}|Q_1, I) \\ &= P(Q_1|I)P(\overline{Q_2}|Q_1, I)P(\overline{Q_3}|Q_1, \overline{Q_2}, I) \end{aligned}$$

I bilgisi, her bir olay için Q sonuçlarının bağımsız şekilde aynı olasılıkların atanmasını sağlar. Son eşitliğe bakılacak olursa;

$$\begin{aligned} P(Q_1, \overline{Q_2}, \overline{Q_3}) &= P(Q_1|I)P(\overline{Q_2}|I)P(\overline{Q_3}|I) \\ &= P(Q|I)P(\overline{Q}|I)P(\overline{Q}|I) \\ &= P(Q|I)P(\overline{Q}|I)^2 \end{aligned}$$

yazılır. Böylece olasılık sonuçlarının sadece Q ve $\overline{Q}$ 'ye bağlı olduğu söylenir. Daha genelleyecek olunursa n örneklemde r başarısı,

$$C_n = \frac{n!}{r!(n-r)!} = \binom{n}{r} \tag{2.20}$$

binom dağılımını elde etme imkanını verir. Böylece n olayın, r başarısının gerçekleşme olayı Q, $(n-r)$ ise $\overline{Q}$ olmak üzere, bu durum Bayesli yaklaşıma aktarılırsa;

$$P(r|n, I) = \frac{n!}{r!(n-r)!} P(Q|I)^r P(\overline{Q}|I)^{n-r} \tag{2.21}$$

elde edilir. Bu dağılıma *Binom Dağılımı* denir.
Binom dağılımının ortalaması ve varyansı sırasıyla,

$$\mu = E(X) = np \tag{2.22}$$

$$\sigma^2 = E(X^2) - [E(X)]^2 = npq \tag{2.23}$$

olarak formülüze edilir. Benzer şekilde Binom Açılımı;

$$(x+y)^n = \sum_{r=0}^{n} \frac{n!}{r!(n-r)!} x^r y^{n-r} \tag{2.24}$$

şeklinde yazılır. Önermeler cebiri E^n için değerlendirilecek olursa yani bu durum n defa gerçekleşirse;

$$E^n = (Q + \overline{Q})^n$$

olur.

Örnek 2.7. [12] I önsel bilgisi, özdeş iki paradan birini seçmeyi göstersin. A parası, hilesiz uygun bir para, B parası ise hileli para ve $P(tura) = 0.2$ olsun. D olayı, 5 atışta 3 kez tura gelmesini temsil etsin. Buna göre, A parası seçildiğinde tura gelme olasılığı nedir?

Çözüm:

$$ORAN = \frac{P(A|D,I)}{P(B|D.I)} = \frac{P(A|I)P(D|A,I)}{P(B|I)P(D|B,I)} = \frac{\frac{1}{2}P(D|A,I)}{\frac{1}{2}P(D|B,I)}$$

$P(D|A,I)$ ve $P(D|B,I)$ olası değerlendirmesi binom dağılımı ile yapılacaktır.

$$P(r|n,I) = \frac{n!}{r!(n-r)!} P(tura|A,I)^r P(yazi|A,I)^{n-r}$$

$$\begin{aligned} P(D|A,I) &= \binom{n}{r} P(tura|A,I)^r P(yazi|A,I)^{n-r} \\ &= \binom{5}{3} (0.5)^3 (0.5)^2 = 0.3125 \end{aligned}$$

$$\begin{aligned} P(D|B,I) &= \binom{n}{r} P(tura|B,I)^r P(yazi|B,I)^{n-r} \\ &= \binom{5}{3} (0.2)^3 (0.8)^2 = 0.0512 \end{aligned}$$

Bu çıkan 2 sayı oranlanırsa;

$$6.1 = \frac{P(A|D,I)}{1 - P(A|D,I)}$$

yazılır. Buradan, $P(A|D,I) = 0.86$ bulunur.

2.1.5.2. Çok Terimli Dağılım

6 yüzü olan bir zar atılınca 6 olası sonucun geldiği bilinmektedir. Eğer olası durum 2 veya daha fazla olursa binom dağılımının bir genelleştirilmesi olduğu önceki bilgiler

kullanılarak aşağıda gösterilmiştir [12]:

I önsel bilgisi, E birçok kez tekrarlanan olayı ve m olası sonuçlar önermeleri temsil etsin. $(O_1, O_2, ..., O_m)$ olayları, sonuçlar olup tek tek mantıksal olarak bağımsızdır.

$$E = O_1 + O_2 + O_3 + ... + O_m$$

olmak üzere, E olayı n defa tekrar edilirse;

$$E^n = (O_1 + O_2 + O_3 + ... + O_m)^n$$

olur. Yani;

$$O_1, \quad n_1 \quad kere \quad meydana \quad gelsin.$$

$$O_2, \quad n_2 \quad kere \quad meydana \quad gelsin.$$

$$O_3, \quad n_3 \quad kere \quad meydana \quad gelsin.$$

$$\vdots$$

$$\vdots$$

$$O_m, \quad n_m \quad kere \quad meydana \quad gelsin.$$

Buna göre,

$$P(E^n|I) = P(O_1|I)^{n_1} P(O_2|I)^{n_2} ... P(O_m|I)^{n_m}$$

elde edilir. $(O_1, O_2, ..., O_m)$ dizilerinin bağımsız bir sıra halinde aynı sayıya ait olduğunun bulunması gerekmektedir.

Binom denkleminde;

$$r! \cdot (n-r)! = n_1! \cdot n_2!$$

yazılırsa, 2 durum için

$$\frac{n!}{r!(n-r)!} = \frac{n!}{n_1! n_2!}$$

olur. Burada,

- n_1, A sayısının açılımı,

- n_2 ise $\overline{A}$ sayısının açılımı olmaktadır.

Şimdi ise olay m olası durum için genellenirse;

$$\frac{n!}{n_1!n_2!...n_m!}$$

elde edilir.

Burada $n = \sum_{i=1}^{m} n_i$ olup, $n_i \equiv O_i$ sonucu n_i kez tekrarlandığından,

$$P(n_1, n_2, ..., n_m|E^n, I) = \frac{n!}{n_1!n_2!...n_m!} \prod_{i=1}^{m} P(O_i|I)^{n_i} \tag{2.25}$$

yazılır. Oluşan bu dağılıma *Çok terimli dağılım* denir.

Bu çoklu genişleme ile karşılaştırıldığında;

$$(x_1 + x_2 + ... + x_m)^n = \sum \frac{n!}{n_1!n_2!...n_m!} x_1^{n_1} x_2^{n_2} ... x_m^{n_m} \tag{2.26}$$

eşitliği yazılır. Burada, $\sum_{i=1}^{m} n_i = n$ dir.

$(X_1, X_2, ...X_k)$ rastgele değişkeni çok terimli dağılıma sahip olsun. O halde bu dağılımın ortalaması ve varyansı sırasıyla,

$$\mu = E(X_i) = np_i \quad i = 1, 2, ..., k \tag{2.27}$$

$$Var(X_i) = np_i q_i \quad i = 1, 2, ...k \tag{2.28}$$

hesaplamaları yapılır.

2.1.5.3. Poisson Dağılımı

Poisson dağılımı, verilmiş bir zaman aralığında bir alanda ya da hacimde başarıların sayısını göstermektedir. I önsel bilgisinin olasılığı, Poisson dağılımında aşağıdaki gibi seçilmelidir:

I önsel bilgisi, $I \equiv$ Öyle bir $r \in R^+$ için herhangi bir olayın $(t, t + dt)$ zaman aralığında meydana gelme olasılığı $(t, t + dt) = rdt$ olsun. [12].

$q(t)$' nin olasılığı $(0,t)$ aralığında, E olayı ise "$(0,t+dt)$'de hesaplanmasın."

Öncelikle E olayı, A ve B önermelerinin birleşimi olmak üzere,

E olayı "$[(0,t)]$ ve $[(t,t+dt)]$ aralığında hesaplanamaz" $= A,B$ şeklinde tanımlansın.

Çarpım Kuralından,

$P(E|I) = P(A,B|I) = P(A|I)P(B|A,I)$ bilindiğinden,

$P(E|I) = q(t+dt) = q(t)(1-rdt)$ veya $\frac{dq}{dt} = -rq(t)$ yazılır.

Çözüm için başlangıç koşulları; $q(0) = 1$ ve $q(t) = e^{-rt}$ olsun.

Şimdi aşağıdaki önermeyi düşünelim:

$C \equiv (0,t)$ aralığında $0 < t_1 < t_2 < ... < t_n < t$ olmak üzere, $(dt_1, dt_2, ..., dt_n)$ küçük toleranslar ile $(t_1, t_2, ..., t_n)$ zamanlarında meydana gelen n tane sayı vardır. O halde $(2n+1)$ tane önermenin birleşimi,

$C = [(0,t_1)$ aralığında hesaplanamaz], $(dt_1$'de hesaplanir),

$[(t_1,t_2)$ aralığında hesaplanamaz] , $(dt_2$'de hesaplanir),...,

$[(t_{n-1},t_n)$ aralığında hesaplanamaz] , $(dt_n$'de hesaplanir),

$[(t_n,t)$ aralığında hesaplanamaz]

şeklinde yazılır.

Çarpım Kuralı ve farklı zaman aralıkların bağımsızlığından,

$$P(C|r,I) = e^{-rt_1}(rdt_1)... \times e^{-r(t_n-t_{n-1})}(rdt_n)e^{-r(t-t_n)} \tag{2.29}$$

yazılır. Diğer yandan $P(C|r,I) = e^{-rt}r^n dt_1...dt_n$ şeklinde de yazılabilir. Böylece verilen r' nin olasılığında, $(0,t)$ aralığında tam olarak n sayı vardır. O halde,

$$P(n|r,t,I) = e^{-rt}r^n \int_0^t dt_n ... \int_0^{t_4} dt_3 \int_0^{t_3} dt_2 \int_0^{t_2} dt_1$$

$$= e^{-rt}r^n \int_0^t dt_n .. \int_0^{t_4} dt_3 \int_0^{t_3} \frac{t_2}{1!} dt_2$$

$$= e^{-rt}r^n \int_0^t dt_n ... \int_0^{t_4} \frac{t_3^2}{2!} dt_3$$

$$= e^{-rt}\frac{(rt)^n}{n!}$$

bulunur. $rt = \lambda$ denilirse genel Poisson dağılımı,

$$P(n|r,t,I) = \frac{e^{-\lambda}\lambda^n}{n!} \tag{2.30}$$

elde edilir.

X, $0, 1, 2, ...$ değerlerini alabilen bir Poisson rastgele değişkeni olsun. X için olasılık fonksiyonu,

$$P(x|\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & x = 0, 1, 2, ... \quad ve \quad \lambda > 0 \\ 0, & \text{diğer durumlar} \end{cases}$$

olup Poisson dağılımının parametresi λ dır.

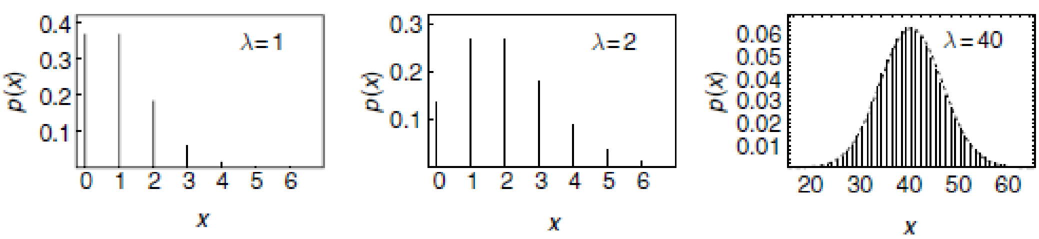
Poisson dağılımının sırasıyla ortalaması ve varyansı,

$$\mu = E(X) = \lambda \tag{2.31}$$

$$\sigma^2 = E(X^2) - [E(X)]^2 = \lambda \tag{2.32}$$

bulunur.

Şekil 2.2, Poisson dağılımının λ parametresinin artmasıyla Binom dağılımı ile olan benzerliğini gösterir.



Şekil 2.2: Poisson dağılımı.

Binom dağılımının, Poisson dağılımına yakınsaması aşağıdaki gibidir:

Binom dağılımının olasılık fonksiyonu ele alınsın.

$$P(x|n,p) = \frac{n!}{(n-x)!x!}p^x(1-p)^{n-x} \tag{2.33}$$

Kabul edelim ki n yeterince büyük p küçük olsun öyle ki np büyük değildir. $P(x|n,p)$ binom olasılığı aşağıdaki gibi hesaplanmaktadır:

$$\frac{n!}{(n-x)!x!} = \frac{n(n-1)(n-2)...(n-x-1)}{x!} \tag{2.34}$$

(2.33) eşitliği; n^x ile çarpılıp bölünerek ve $np = \lambda$ eşitliği kullanılırsa,

$$\begin{aligned} P(x|n,p) &= \frac{n(n-1)(n-2)...(n-x-1)}{n^x x!}(np)^x(1-p)^{n-x} \\ &= \frac{n(n-1)(n-2)...(n-x-1)}{n^x}\frac{\lambda^x}{x!}(1-p)^{n-x} \\ &= \frac{1(1-\frac{1}{n})(1-\frac{2}{n})...(1-\frac{(x-1)}{n})}{(1-p)^x}\frac{\lambda^x}{x!}(1-p)^n \end{aligned}$$

bulunur.

Şimdi $(1-p)^n = [(1-p)^{\frac{-1}{p}}]^{-np} = [(1-p)^{\frac{-1}{p}}]^{-\lambda}$ ve $z = -p$ denilip limitleri alındığı zaman,

$$\lim_{z \to 0}(1+z)^{\frac{1}{z}} = e$$

olduğunu bilinir. O halde

$$\lim_{p \to 0} = [(1-p)^{\frac{-1}{p}}]^{-\lambda} = e^{-\lambda}$$

yazılır. Ayrıca

$$\lim_{n \to \infty} 1(1-\frac{1}{n})(1-\frac{2}{n})...(1-\frac{(x-1)}{n}) = 1 \tag{2.35}$$

ve

$$\lim_{p \to 0}(1-p)^x = 1 \tag{2.36}$$

elde edilir. Böylece gerekli işlemler uygulandığında, Binom dağılımının

$$\lim_{n \to \infty}\lim_{p \to 0} P(x|n,p) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0,1,2,... \tag{2.37}$$

Poisson dağılımına yaklaştığı görülür.

Literatürde sıkça kullanılan sürekli dağılımlara aşağıda yer verilmiştir.

2.1.5.4. Normal Dağılım

Normal dağılım, tüm olasılık dağılımları içinde en çok kullanılan dağılımdır. Normal dağılım ayrıca *Gauss dağılımı* olarak da bilinir. Sürekli bir rastgele değişkeni için olasılık yoğunluk fonksiyonu [12]

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, -\infty < x < \infty; , -\infty < \mu < \infty; 0 < \sigma^2 < \infty \quad (2.38)$$

ise X Normal dağılıma sahiptir denir ve $X \sim N(\mu, \sigma^2)$ ile gösterilir. Burada,

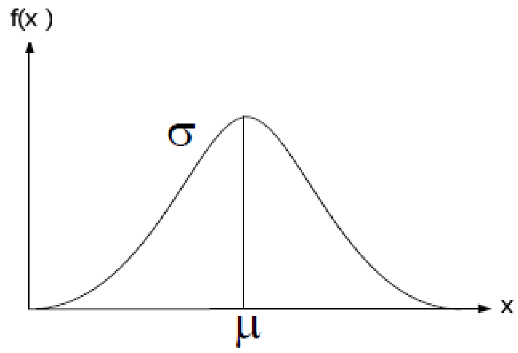
i) μ : Normal dağılımın ortalaması,

ii) σ : Normal dağılımın standart sapması,

iii) $e \cong 2,71828$,

iv) $\pi \cong 3,14159$ dır.

μ ortalamalı ve σ standart sapmalı Normal dağılıma sahip bir X rastlantı değişkeninin olasılık yoğunluk fonksiyonunun grafiği Şekil 2.3'de verilir.



Şekil 2.3: Normal dağılım.

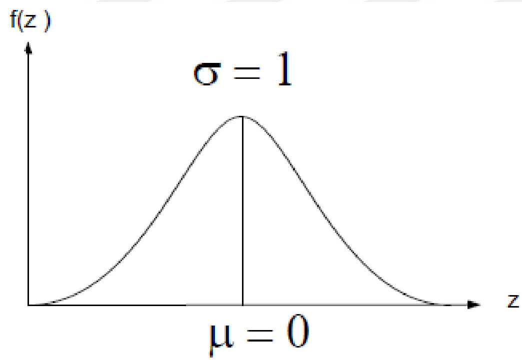
Şekil 2.3'de görüldüğü gibi Normal dağılım $x = \mu$ doğrusuna göre simetriktir. Normal dağılımın grafiği $x = \mu$ doğrusunun solunda ve sağında aynıdır. Ortalama (μ) ortadadır ve alanı iki eşit parçaya ayırır. $f(x)$ eğrisinin altındaki ve x ekseninin yukarısındaki alan 1'e eşittir [6].

2.1.5.5. Standart Normal Dağılım

Ortalaması $\mu = 0$, varyansı $\sigma^2 = 1$ ve olasılık yoğunluk fonksiyonu olan Normal dağılıma, Standart Normal dağılım denir.
$X\sim$ N(μ, σ^2) normal dağılıma sahipse, $Z=\frac{X-\mu}{\sigma}$ değişkeni, $Z\sim$ N$(0,1)$ standart normal dağılıma sahiptir ve olasılık yoğunluk fonksiyonu

$$f(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} \qquad -\infty < z < \infty \tag{2.39}$$

şeklinde gösterilir ve grafiği Şekil 2.4'de verilmiştir.



Şekil 2.4: Standart Normal dağılım.

Standart normal dağılımın özellikleri aşağıdaki gibidir [6]:

i) Eğri $f(z)$ eksenine göre simetriktir.
ii) $f'(z) = -\frac{z}{\sqrt{2\pi}}e^{\frac{-1}{2}z^2}$; $z = 0$ için $f'(z) = 0$; $f(z) = \frac{1}{\sqrt{2\pi}}$ dir.
iii) $(0, \frac{1}{\sqrt{2\pi}})$ noktası eğrinin maksimum noktasıdır.
iv) $z = -1$ ve $z = 1$ doğruları arasında eğri aşağıya doğru, diğer aralıklarda yukarıya doğru bükülür.($z = \pm 1$ büküm noktalarıdır.)
v) Eğri sola ve sağa doğru sınırsız olarak uzamakta ve $z = 0$ noktasından her iki doğrultuda uzaklaşınca z eksenine çok çabuk yaklaşır, z ekseni normal eğri için asimptottur.
vi) Eğrinin altında kalan z ekseni üzerindeki alan 1 dir. Sonuç olarak, Z rastgele değişkeninin $-\infty$ ve ∞ arasında bir değer alma olasılığı 1 dir. **i)** ve **v)** birleştirilirse

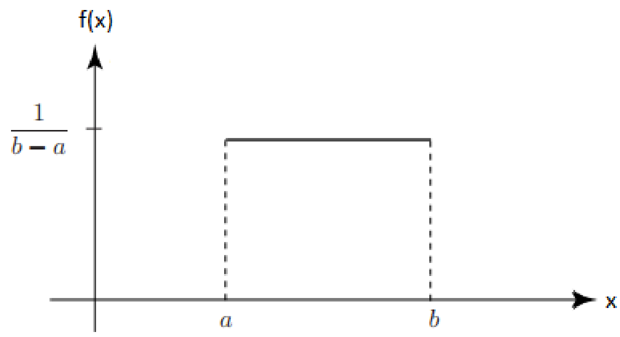
eğrinin altında $f(z)$ ekseninin sağında ve solundaki alanlardan her birinin $0,5$ olduğu görülür.

2.1.5.6. Düzgün Dağılım

X rastgele değişkeninin olasılık yoğunluk fonksiyonu,

$$f(x|a,b) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{diğer durumlar} \end{cases} \tag{2.40}$$

ise X rastgele değişkeni (a,b) açık aralığında düzgün dağılıma sahiptir ve Şekil 2.5'de grafiği verilmiştir.



Şekil 2.5: Düzgün dağılım.

Düzgün dağılımın ortalaması ve varyansı sırasıyla,

$$E(X) = \frac{a+b}{2}$$

ve

$$Var(X) = \frac{(b-a)^2}{12}$$

olarak hesaplanır.

2.1.5.7. Üstel Dağılım

Negatif olmayan değerler alan sürekli X rastgele değişkeninin olasılık yoğunluk fonksiyonu

$$f(x|\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{diğer durumlar} \end{cases} \tag{2.41}$$

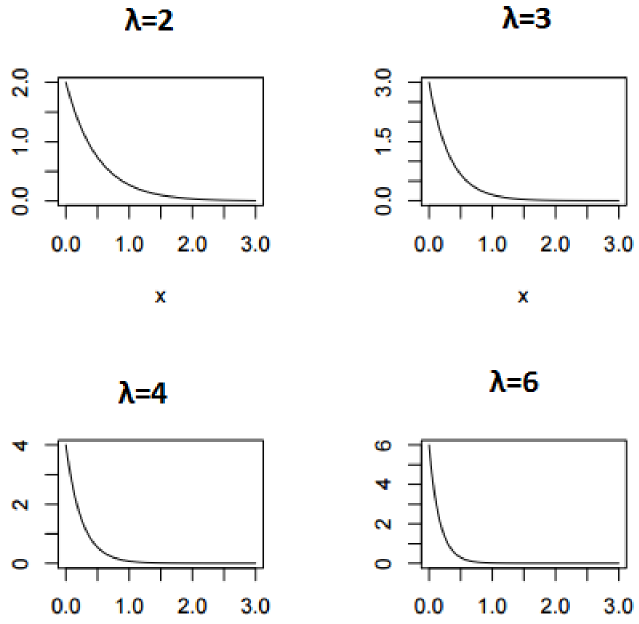
olsun. Bu taktirde X, $\lambda > 0$ parametresi ile üstel dağılıma sahiptir denir.
Üstel dağılımın sırasıyla ortalaması ve varyansı,

$$E(X) = \frac{1}{\lambda}$$

ve

$$Var(X) = \frac{1}{\lambda^2}$$

olur.
Şekil 2.6'da λ parametresinin değişimine göre dağılımın aldığı şekil görülmektedir [12].



Şekil 2.6: Üstel dağılım.

2.1.5.8. Gamma Dağılımı

X, pozitif değerler alan sürekli rastgele değişken olsun. X için olasılık yoğunluk fonksiyonu,

$$f(x|\alpha,\theta) = \begin{cases} \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{\frac{-x}{\theta}}, & x > 0 \quad \alpha,\theta > 0 \\ 0, & \text{diğer durumlar} \end{cases} \tag{2.42}$$

ise X, Gamma olasılık dağılımına sahiptir. Burada $\Gamma(\alpha)$ fonksiyonu,

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \quad \alpha > 0$$

olarak tanımlanır.

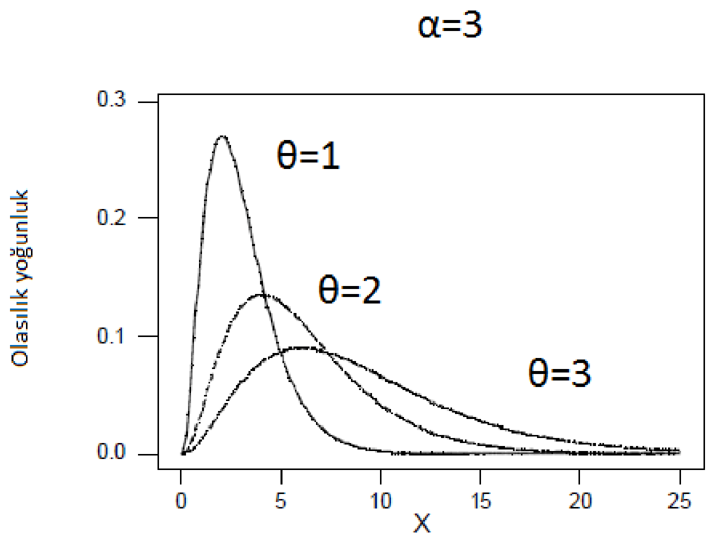
Gamma fonksiyonunun bazı özellikleri aşağıdaki gibidir:

i)$\Gamma(\alpha) = (\alpha-1)(\alpha-2)...\Gamma(1) = (\alpha-1)! \quad \alpha \in N$

ii) $\Gamma(1) = 1$

iii))$\Gamma(\frac{1}{2}) = \sqrt{\pi}$ dır.

Şekil 2.7, θ' nın değişimine göre Gamma dağılımının değişimini gösterir.



Şekil 2.7: Gamma dağılımı.

Gamma dağılımın sırasıyla ortalaması ve varyansı,

$$E(X) = \alpha\theta$$

ve

$$Var(X) = \alpha\theta^2$$

olarak hesaplanır.

2.1.5.9. χ^2 Dağılımı

$X_i \sim N(\mu_i, \sigma_i^2)$ ve X_i'ler bağımsız rastgele değişkenler olsun. $Y_i = \frac{X_i - \mu_i}{\sigma_i}(i = 1, 2, ..., n)$ rastgele değişkenleri düşünülürse, $Y_i \sim \mathrm{N}(0, 1)$ olur. Diğer yandan,

$$\sum_{i=1}^{n}(\frac{X_i - \mu_i}{\sigma_i})^2 = \sum_{i=1}^{n} Y_i^2$$

değişkeni tanımlansın. Bu rastlantı değişkeni, standart normal dağılıma sahip rastgele değişkenlerin kareleri toplamıdır.

Genel olarak X rastlantı değişkeni χ^2 dağılımına sahip ise bu rastgele değişkeninin olasılık yoğunluk fonksiyonu aşağıdaki gibidir:

$$f(x|\upsilon) = \frac{x^{\frac{\upsilon}{2}-1} e^{\frac{-x}{2}}}{2^{\frac{\upsilon}{2}} \Gamma(\frac{\upsilon}{2})} \tag{2.43}$$

Burada υ serbestlik derecesini gösterir. χ^2 sürekli bir dağılım olup, olasılık yoğunluk fonksiyonu Gamma dağılımında $\theta = 2, \alpha = \frac{\upsilon}{2}$ alınması ile bulunur.

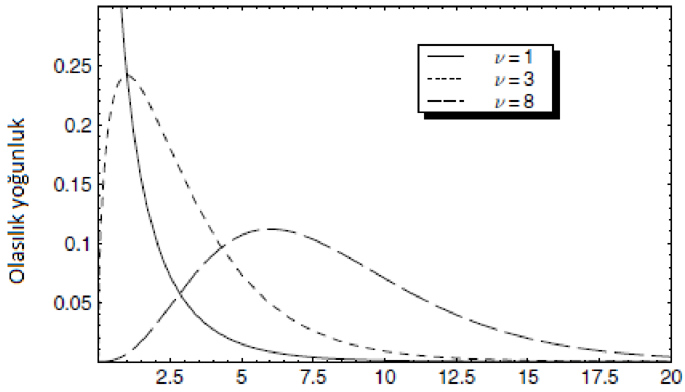
χ^2 dağılımın sırasıyla ortalaması ve varyansı,

$$E(X) = \upsilon$$

ve

$$Var(X) = 2\upsilon$$

olarak hesaplanır. Serbestlik derecesi υ değişimine göre χ^2 dağılımının alacağı şekil, Şekil 2.8'de gösterilmiştir [12].



Şekil 2.8: χ^2 dağılımı.

2.1.5.10. Student T dağılımı

T dağılımı, Z puanları üzerinde bir düzeltme ile elde edilen t puanlarının dağılımıdır. Z standart normal dağılım ve X, υ serbestlik dereceli χ^2 dağılımına sahip olsun. Eğer Z ve X bağımsız ise,

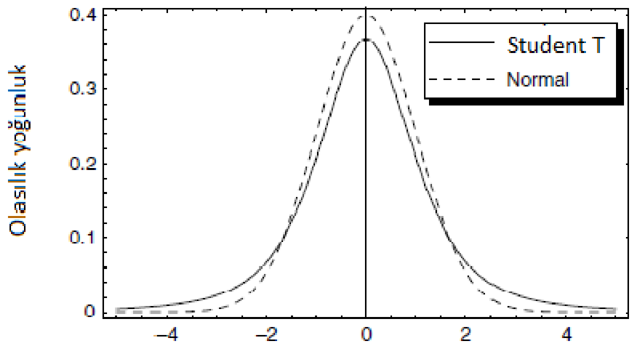
$$T = \frac{Z}{\sqrt{\frac{X}{\upsilon}}}$$

T değişkeni *Student-T* dağılımına sahiptir. Bu dağılımın olasılık yoğunluk fonksiyonu şu şekildedir:

$$f(t|\upsilon) = \frac{\Gamma[(\frac{\upsilon+1}{2})]}{\Gamma(\frac{\upsilon}{2})\sqrt{\pi\upsilon}}[1 + (\frac{t^2}{\upsilon})]^{-\frac{(\upsilon+1)}{2}}, \quad -\infty < t < \infty, \upsilon > 0 \qquad (2.44)$$

Bu formüldeki υ sembolü serbestlik derecesini, gama(Γ) sembolü ise serbestlik derecesine bağlı özel bir gama fonksiyonunu göstermektedir.

$f(t|\upsilon)$ olasılık yoğunluk fonksiyonu simetrik bir eğriye sahiptir ve serbestlik derecesi büyüdükçe, normal dağılıma yaklaşır. Bu durum Şekil 2.9'da gösterilmiştir [12].



Şekil 2.9: Student T dağılımı.

T dağılımının özellikleri aşağıdaki gibidir [6]:

i) Serbestlik derecesi denilen υ parametresi ile belirlenen sonsuz sayıda t dağılımı olup daima pozitif bir tam sayıdır. t_υ gösterimi, υ serbestlik dereceli bir t rastgele değişkenine aittir.

ii) $t\upsilon$'nin yoğunluk fonksiyonu grafiği 0 merkezli çan eğrisi şeklinde simetriktir.

iii) υ'nin değeri artarken, t_υ rastgele değişkenin varyansı azalır. Böylece υ'nin değeri artarken, t_υ ile ilgili çan eğrisi standart normal dağılım eğrisine yaklaşır.

2.1.5.11. Jeffrey'nin Önseli

X rastgele değişkeninin dağılımı, tanımlı olduğu bölge D_x ile gösterilip olasılık yoğunluk fonksiyonu $f(x,\theta)$ şeklinde tanımlansın.

Bu dağılıma sahip kitleden $X_1, X_2, ..., X_n$ bağımsız rastgele değişkenlerinin olasılık yoğunluk fonksiyonları sırasıyla, $f(x_1;\theta)$, $f(x_2;\theta)$,..., $f(x_n;\theta)$ olsun. Bu rastgele değişkenlerinin ortak olasılık yoğunluk fonksiyonu,

$$L(\theta|x_1, x_2, ..., x_n) = L(\theta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i;\theta) \tag{2.45}$$

olarak tanımlanır. Burada $x = \{x_1, x_2, ..., x_n\}$ olarak tanımlanan bir rastgele değişkeninin olabilirlik fonksiyonu, x verilmiş iken θ'yı bilinmeyen olarak kabul eden fonksiyondur. $L(\theta|\mathbf{x})$ fonksiyonuna, en çok olabilirlik fonksiyonu denir.

Jeffreys; düzgün önsel olasılıkların yanında, belirli dönüşümler altında sabit bir önsel olasılık fikrini ortaya koymuştur. Bu olasılık, en çok olabilirlik tahmin edicisini belirleyen Fisher bilgi miktarı ile belirlenir. Fisher bilgi matrisi ise

$$I(\theta) = -E\left[\frac{\partial^2 \log f(x|\theta)}{\partial \theta^2}\right] \tag{2.46}$$

eşitliği ile ifade edilir. Buna göre Jeffrey'nin önsel dağılımı; bir tahmin edicinin, en çok olabilirlik tahmin edicisinin komşuluğundaki duyarlılığı ölçen Denklem (2.46) daki Fisher bilgi matrisinin determinantının kareköküne eşittir ve

$$f(x|\theta) \propto |I(\theta)|^{\frac{1}{2}} \tag{2.47}$$

olarak yazılır.

Örnek 2.8. [15] $y_1, y_2, ..., y_n \sim$ Binom $(1, \theta)$ dağılımına göre Jeffrey'nin önsel dağılımı şu şekilde bulunur.
Önce Binom$(1, \theta)$ dağılımı yazılsın.
$P(y|\theta) \propto \theta^y(1-\theta)^{1-y} \quad \theta \in [0,1] \quad y = 0,1$ olmak üzere, fonksiyonun logaritması alınıp θ parametresine göre 2 kez türevi hesaplanıp bunun beklenen değeri alınarak $I(\theta)$ bulunur:

$$\log P(y|\theta) = y \log \theta + (1-y)\log(1-\theta)$$
$$\frac{\partial \log P(y|\theta)}{\partial \theta} = \frac{y}{\theta} - \frac{(1-y)}{(1-\theta)}$$
$$\frac{\partial^2 \log P(y|\theta)}{\partial \theta^2} = -\frac{y}{\theta^2} - \frac{(1-y)}{(1-\theta)^2}$$

yazılır. O halde Fisher bilgi miktarı,

$$\begin{aligned} I(\theta) &= -E\left[\frac{\partial^2 \log P(y|\theta)}{\partial \theta^2}\right] \\ &= \frac{E(y)}{\theta^2} + \frac{E(1-y)}{(1-\theta)^2} \\ &= \frac{1}{\theta} + \frac{1}{(1-\theta)} = \frac{1}{\theta(1-\theta)} \end{aligned}$$

şeklinde bulunur. Jeffrey önseli ise,

$$f(x|\theta) \propto |I(\theta)|^{\frac{1}{2}}$$

$$= \frac{1}{\sqrt{\theta(1-\theta)}}$$

elde edilir.

2.2. ENTROPİ KAVRAMI

Bu bölümde belirsizlik kavramı ve bilgi kuramına değinilerek entropi kavramı ele alınıp entropinin temel özelliklerinden ve bazı önemli entropi ölçülerinden bahsedilecektir.

2.2.1. Entropi

Entropi, *bilgideki belirsizliğin bir ölçüsüdür* diye tanımlanabilir [2]. Tarihsel açıdan değerlendirildiğinde, termodinamik üzerine çalışmalar yapan Clausius tarafından geliştirilmiştir. 19. yüzyılda termodinamik çalışmaları, *sistemde var olan düzensizlik düzeyinin ölçümü olarak ortalama enerjiyi* açıklama üzerine olmuştur. Ludwig Boltzmann tarafından geliştirilen entropi yasası, bu aşamada olasılık kavramını da içine almıştır. Böylece Boltzmann bağıntısı ortaya çıkarılmıştır. Claude Shannon tarafından yapılan çalışmalarda entropi, *İletişim Sürecinin* bir kavramı olarak değerlendirilmiş ve enformasyon ile entropi bir arada kullanılmaya başlanmıştır.

Enformasyon; olay ya da bireylerle ilgili gerçek verilerin işleme tutulmuş hali olup alıcı durumundaki kişinin sistem hakkında bilgisini arttırır, içinde bulunulan belirsizliği azaltır. Enformasyonun *bilgi* haline gelebilmesi için söz konusu eylemin kullanıma hazır duruma gelmesi gerekmektedir. İletişim sürecini açıklamak için kullanılan entropi kavramı, fiziksel entropi ile ilişkili olmakla birlikte bu ilişki, *matematiksel düzeyde değil, insan davranışlarını etkileyen düzeydedir*. Görüldüğü gibi, fizik disiplininin bir çalışma alanı olan entropi, 20. yüzyılın ortalarına doğru diğer disiplinlerde de kullanılmaya başlanmıştır.

2.2.2. Bilgi Kuramında Entropi

Entropinin en genel tanımı şu şekilde yapılabilir [16]: Anlamına, cinsine, değerine ya da herhangi bir öznel özelliğine bakılmaksızın, iletişim oluşturan sembol, sinyal veya sayılar dizisinin istatistiksel yapısını analiz eden bilgi kuramıdır. Bu kuramının temel amacı, bilginin elde edilmesi, aktarılması, işlenmesi ve saklanmasına ilişkin kuralları incelemektir.

Olasılıksal belirsizlik için kullanılan entropi, gönderici ile alıcı arasındaki iletişim sürecinde kaybedilen bilgi miktarının bir ölçüsüdür. Entropiyi açıklamak için kullanılan örnekler, genellikle farklı olasılık dağılımlarını farklı belirsizliklerle açıklama özelliğine sahiptir. Örneğin, yazı-tura oyunu zar atma oyunu ile karşılaştırıldığında, yazı tura oyunu daha düşük entropiye sahiptir. Bunun temel nedeni, yazı-tura oyununda eşit olasılıklı iki sonuç var iken, zar atma oyununda eşit olasılıklı altı sonuç olmasıdır.

Belirsizliğin temel ilkesi, bilgi ile entropinin ters işaretli olmasıdır. Sistemde bilginin akması halinde doğal olarak belirsizlik miktarı da azalmış olacaktır. Dolayısıyla çok sayıda bilgi elde edilince, belirsizlik entropi yardımıyla bulunmuş olacaktır [17].

Shannon tarafından geliştirilen entropi kavramı, bilginin en kısa yoldan doğru şekilde iletilmesine çözüm getirmiştir. Dolayısıyla entropi, *bilgi içeriğinin ölçüsüdür*. Bu belirsizliğin entropi olarak adlandırılmasının sebebi, termodinamikteki entropi kavramı ile aynı matematiksel gösterime sahip olmasıdır [18].

İlerleyen bölümlerde entropi hem kesikli hem de sürekli rastlantı değişkenleri için tanımlanmıştır.

2.2.3. Kesikli Durumlar İçin Entropi

Bu kısımda Shannon entropisi, bileşik entropi, koşullu entropi kavramlarına yer verilecektir.

2.2.3.1. Shannon Entropisi

Shannon' un *Matematiksel İletişim Teorisinde*, çok sayıdaki alternatif karşısında bilgi kaynağından mesajların nasıl iletileceği ve bunun en optimum yöntemle nasıl yapılacağı araştırılır. Bu araştırmada, 0 ve 1 olmak üzere iki kararlı durum vardır. Bir zamanda 1 bitlik bilgi taşıyan bir aygıt söz konusu ise, N tane aygıt toplam 2^N durum için N bitlik bilgi taşıyacaktır ve

$$\log_2 2^N = N$$

yazılabilir. Dolayısıyla sinyal sayısının logaritması, zamanla doğrusal olarak artar. Yani logaritma tabanı 2 seçildiğinde, bilgi miktarının birimi bit olmaktadır.

Shannon, meydana gelme olasılıkları $p_1, p_2, ..., p_n$ olan ve bireyin ilk olarak bunlardan hangisinin meydana geleceğini bilmediği mümkün olaylar kümesini ele almıştır. Sonuçta belirsizliği verecek bir H ölçüsüne ihtiyaç olduğunu ortaya çıkaran Shannon, bu ölçünün (K sabit olmak üzere) aşağıdaki gibi yazılabileceğini göstermiştir:

$$H(p) = -K \sum_{i=1}^{n} p_i \cdot \log_2 p_i$$

Bu ifade, Shannon Entropisi olarak adlandırılmaktadır.
Shannon'un Entropisi üzerine çalışan Ross, entropi ile olasılık arasındaki ilişkiyi; *Olasılığı düşük olan bir olayın sürpriz olarak nitelendirilmesi, bu olayın olasılığının ne kadar düşük olduğuna bağlıdır* şeklinde ifade etmiştir [19].

Shannon'un ifade ettiği bu fonksiyon aşağıdaki özelliklere sahip olmalıdır [19]:
i) Bir A olayının p olasılığı ile meydana gelmesi halinde oluşacak sürprizi $S(p)$ ile gösterilsin. Eğer, bu olayın meydana gelme olasılığı 1 ise, bu olay için sürpriz aşağıdaki gibi ifade edilecektir:

$$S(1) = 0$$

Bu olayın meydana geleceği bilindiği için gerçekleşmesi de sürpriz olmayacaktır. Bu da 0 bitlik bilgi verecektir.
ii) $p < q$ ise $S(p) < S(q)$ olacağından $S(p)$, p' nin kesin azalan bir fonksiyonudur.
iii) $S(p)$ fonksiyonu , p 'nin her değerinde sürekli olmalıdır.
iv) Bütün p_i değerleri eşit olduğunda ($\forall i, p_i = \frac{1}{n}$), S fonksiyonu n için monoton artan bir fonksiyonu olmalıdır:

$$S(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n}) = nf(\frac{1}{n})$$

v) Olasılıkları $P(A) = p$ ve $P(B) = q$ olan iki bağımsız A ve B olayı olsun. Olaylar bağımsız olduğundan $P(AB) = pq$ yazılabildiğinden, A ve B olaylarının ikisi birden meydana geldiğinde bunların sürprizini $S(pq)$ olarak ifade edilir. Önce A için ve ardından da B için meydana gelmesi ile ilgili ilave sürpriz $S(pq)-S(p)$ olacaktır. Ayrıca B olayı, A olayından bağımsız olarak meydana geleceği için bu ilave sürpriz $S(q)$ ile gösterilebilir. O halde, aşağıdaki bağıntı yazılabilir:

$$S(pq) = S(p) + S(q), \quad 0 < p \leq 1 \quad ve \quad 0 < q \leq 1$$

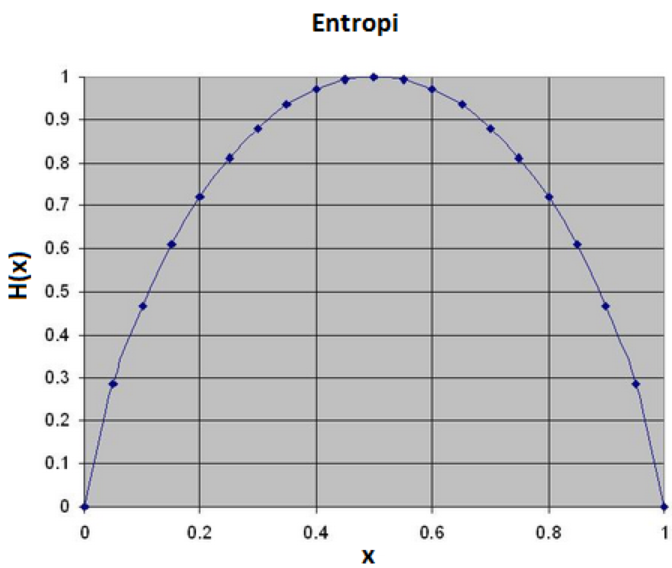
Yukarıdaki koşulların sağlanması ile,

$$S(p) = -C \log_2 p$$

yazılır. Genellikle $C = 1$ alınır. Olasılıkları $p_1, p_2, ...p_n$ olan $x_1, x_2, ..., x_n$ değerlerinden birini alabilen X rastgele değişkeni ele alınsın. X değişkeninin, x_i değerini alması halinde $(\log_2 p_i)$ de sürpriz olarak düşünülebilir. Bu sürprizin beklenen miktarı,

$$H(p) = -\sum_{i=1}^{n} p_i \cdot \log_2 p_i \tag{2.48}$$

olur. Burada $H(p)$, X rastgele değişkenin entropisi olarak tanımlanır. Tüm p_i değerleri eşit olduğunda ise $H(p)$ maksimum olur. Yani $H(p)$, X rastgele değişkeninin ortalama sürprizi olarak adlandırılır ve bu durum Şekil 2.10'da gösterilmiştir.



Şekil 2.10: Shannon Entropisi.

Entropi kavramının anlaşılması için iki örnek aşağıda verilmiştir.

Örnek 2.9. [20] X rastgele değişkeni kesikli düzgün dağılıma sahip ve X rastgele değişkeninin alacağı değerler $x_1, , x_2, ..., x_{16}$ olsun. Bu durumda X'in olasılık fonksiyonu,

$$f(x) = P(X = x) = \frac{1}{16} \qquad\qquad x = x_1, x_2, ..., x_{16}$$

dır.

O halde bu sistemin entropisi, Denklem (2.48) kullanılarak,

$$H(p) = -\sum_{i=1}^{16} \frac{1}{16} \log_2 \frac{1}{16} = 4(bit)$$

olduğu görülür.

Örnek 2.10. [20] Bir yarışa 8 at katılsın ve yarışı kazanma olasılıkları aşağıdaki gibi olsun.

$$\mathbf{X} \equiv \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} & \frac{1}{64} & \frac{1}{64} & \frac{1}{64} & \frac{1}{64} \end{pmatrix}$$

O halde bu sistemin entropisi,

$$H(X) = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{4} \log_2 \frac{1}{4} - \frac{1}{8} \log_2 \frac{1}{8} - \frac{1}{16} \log_2 \frac{1}{16} - \frac{4}{64} \log_2 \frac{1}{64} = 2(bit)$$

elde edilir.

Entropi hesabında sadece durumların olasılıkları önemlidir.

Entropi tanımında iki tabanlı logaritma yerine e tabanlı logaritma da kullanılmaktadır.

$$H(p) = -\sum_{i=1}^{n} p_i \cdot \ln(p_i) = -\sum_{i=1}^{n} p_i \cdot \ln(2^{log_2 p_i})$$

$$= -\sum_{i=1}^{n} p_i \cdot \log_2 p_i \ln(2) = -\ln(2) \sum_{i=1}^{n} p_i \cdot \log_2 p_i$$

Entropi ölçü birimi bit olmak üzere, entropi tanımında iki tabanlı logaritma yerine e tabanlı logaritma kullanılırsa $\ln(2)$ katsayısını göz önünde tutmak gerekir. Doğal logaritma ve e tabanlı üslü ifadelerin türev ve integral hesaplamaları daha kolay bulunmaktadır.

Bundan sonraki kısımda bir

$$\mathbf{X} \equiv \begin{pmatrix} x_1 & x_2 & ... & x_n \\ p_1 & p_2 & ... & p_n \end{pmatrix}$$

sisteminin entropisi,

$$S(p_1, p_2, ..., p_n) = -\sum_{i=1}^{n} p_i \cdot \ln(p_i) \tag{2.49}$$

olarak tanımlanacaktır. $S(p_1, p_2, ..., p_n)$; kesin olmayan kesikli olasılık dağılımlarını, p_i ise de verilen entropi dağılımını ifade eder.

Şimdi iki ya da daha çok sistemin bileşkesi olan sistemlerde entropi kavramını inceleyelim.

2.2.3.2. Bileşik Entropi

$$\mathbf{X} \equiv \begin{pmatrix} x_1 & x_2 & ... & x_n \\ p_1 & p_2 & ... & p_n \end{pmatrix}$$

ve

$$\mathbf{Y} \equiv \begin{pmatrix} y_1 & y_2 & ... & y_m \\ q_1 & q_2 & ... & q_m \end{pmatrix}$$

gibi iki sistem söz konusu olsun. Bu iki sistemin bileşkesi denildiğinde,

$$\{(x_i, y_j) : i = 1, 2, ..., n, \quad j = 1, 2, ..., m\}$$

kümesindeki durumlarda bulunabilen sistem anlaşılmaktadır. X ve Y gibi iki sistemin bileşkesi (X, Y) biçiminde gösterilir.

Bu sistemin (x_i, y_j) , $i = 1, 2, ..., n$, $j = 1, 2, ..., m$ durumlarında bulunma olasılıkları $P(X = x_i, Y = y_j)$ olsun. Yani,

$$P(X = x_i) = \sum_{j=1}^{m} p_{ij} = p_i \qquad (i = 1, 2, ..., n), \qquad \sum_{i=1}^{n} p_i = 1$$

$$P(Y=y_j)=\sum_{i=1}^{n}p_{ij}=q_j \qquad (j=1,2,...,m), \qquad \sum_{j=1}^{m}q_j=1$$

olmak üzere bileşik entropi,

$$H_{mn}(X,Y) \quad = -\sum_{i=1}^{n}\sum_{j=1}^{m}P(X=x_i,Y=y_j)\ln P(X=x_i,Y=y_j) \qquad (2.50)$$

olarak tanımlanır.

Teorem 2.1. [18] Bağımsız iki sistemin bileşik entropisi, sistemlerin entropi toplamlarına eşittir ve

$$H_{mn}(X,Y)=H_n(X)+H_m(Y) \qquad (2.51)$$

olarak yazılır.

Kanıt:

İki sistem birbirinden bağımsız,

$$P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j)$$

ve

$$\ln P(X=x_i,Y=y_j)=\ln P(X=x_i)+\ln P(Y=y_j)$$

olmak üzere,

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

olup X ve Y değişkenlerinin bağımsızlık koşulunu sağlamaları nedeniyle p_{ij} yerine p_iq_j yazılabilir. Bileşik entropi,

$$\begin{aligned}
H_{mn}(X,Y) &= -\sum_{j=1}^{m}\sum_{i=1}^{n}p_iq_j\ln(p_iq_j)\\
&= -\sum_{j=1}^{m}\sum_{i=1}^{n}p_iq_j(\ln p_i+\ln q_j)\\
&= -\sum_{j=1}^{m}\sum_{i=1}^{n}p_iq_j\ln p_i-\sum_{i=1}^{n}\sum_{j=1}^{m}p_iq_j\ln q_j\\
&= (-\sum_{j=1}^{m}q_j)(\sum_{i=1}^{n}p_i\ln p_i)-(\sum_{i=1}^{n}p_i)(\sum_{j=1}^{m}q_j\ln q_j)
\end{aligned}$$

yazılır. Buradan,

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

eşitliği elde edilir. $\sum_{j=1}^{m} q_j = 1$ ve $\sum_{i=1}^{n} p_i = 1$ eşitlikleri yukarıdaki denklemde yerine yazılırsa,

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

(2.51) eşitliği bulunur.
Not olarak şunu da belirtelim ki dağılımlar bağımsız olmadığı zaman bu eşitlik geçerli değildir.

Birbirinden bağımsız daha fazla değişken olması durumunda, bu dağılımların oluşturacağı bileşik dağılımın entropisi,

$$H(X_1, X_2, ..., X_p) = \sum_{i=1}^{p} H(X_i)$$

şeklinde olacaktır [17].

2.2.3.3. Koşullu Entropi

Bileşik entropi konusunda bağımsızlık koşulu sağlanmadığı zaman koşullu entropi söz konusu olacaktır. (X,Y) gibi bileşik bir sistemde, Y sisteminin y_j durumunda olduğu bilinsin. Y sisteminin y_j durumu bilindiğinde X sisteminin koşullu entropisi, $H(X|y_j)$ ile gösterilip

$$H(X|Y = y_j) \quad = -\sum_{i=1}^{n} P(X = x_i|Y = y_j) \ln P(X = x_i|Y = y_j) \qquad (2.52)$$

olarak tanımlanır. Bu eşitlik X rastgele değişkenin kısmi entropisi olarak da ifade edilir.

$H(X|Y)$ koşullu entropisi, bir başka ifadeyle Y rastgele değişkenin tüm değerleri gerçeklendiğinde X rastgele değişkeninin belirsizlik miktarı şeklinde açıklanabilir. Bu ifade ise Y sisteminin tüm y_j durumlarının ortalaması ile şu şekilde elde edilir [21].

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

Bu eşitlikte koşullu olasılık tanımından $q(y_j)p(x_i|y_j) = p_{ij} = p(x_i, y_j)$ şeklinde yazılabilir. O halde koşullu entropi,

$$\begin{aligned} H_{mn}(X|Y) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\ln p(x_i|y_j) \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}\ln p(x_i|y_j) \end{aligned} \tag{2.53}$$

şeklinde yazılabilir. Benzer şekilde, Y için koşullu entropi:

$$H(Y|X) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\ln p(y_j|x_i) \tag{2.54}$$

olarak tanımlanır.

Teorem 2.2. [22] (X, Y) bileşik entropisi için, iki rastgele değişkenin bileşik entropisi, birinin koşullu entropisi ile diğerinin entropisi toplamına eşittir. Matematiksel olarak,

$$H_{mn}(X, Y) = H_n(X) + H_{mn}(Y|X) = H_m(Y) + H_{mn}(X|Y) \tag{2.55}$$

şeklinde yazılabilir.

Kanıt:

Koşullu olasılık formülü;

$P(X, Y) = P(X)P(Y|X)$

olduğundan ve her iki tarafın doğal logaritması alınırsa,

$\ln P(X, Y) = \ln P(X) + \ln P(Y|X)$

eşitliği bulunur. Bu eşitlikte her iki tarafın beklenen değeri alındığında ise,

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

Buradan, (2.55) eşitliği

$H_{mn}(X,Y) = H_n(X) + H_{mn}(Y|X)$ elde edilir.

Bu teoremin sonucu aşağıda verilmiştir.

Sonuç 2.1. [17] Eğer (X,Y) bileşik sisteminde Y sistemi ile X sistemi bağımsız ise,

$$H_{mn}(Y|X) = \sum_{i=1}^{n} p_i H(Y|X = x_i) = -\sum_{i=1}^{n} p_i \sum_{j=1}^{m} P(Y = y_j|X = x_i) \ln P(Y = y_j|X = x_i)$$

$$= -\sum_{i=1}^{n} p_i \sum_{j=1}^{m} P(Y = y_j) \ln P(Y = y_j) = -\sum_{j=1}^{m} P(Y = y_j) \ln P(Y = y_j)$$

$$= H_m(Y)$$

olur. Bu durumda Denklem (2.55) deki bileşik entropi:

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

formunu alır.

Bu durumla ilgili örnek aşağıda verilmiştir.

Örnek 2.11. (X,Y) gibi bir sistem ile ilgili durumlar ve bu durumlarda bulunma olasılıkları aşağıdaki gibi olsun.

X / Y	x_1	x_2	$q(y)$
y_1	1/4	1/4	1/2
y_2	1/4	1/4	1/2
$p(x)$	1/2	1/2	1

Bu bileşke sistemin entropisi,

$H(X,Y) = -\frac{1}{4}\ln\frac{1}{4} - \frac{1}{4}\ln\frac{1}{4} - \frac{1}{4}\ln\frac{1}{4} - \frac{1}{4}\ln\frac{1}{4} = \ln 4 = 2\ln 2 = 2(bit)$

olup, X rastgele değişkenin entropisi,

$H(X) = -\sum p(x)\ln p(x) = -\frac{1}{2}\ln\frac{1}{2} - \frac{1}{2}\ln\frac{1}{2} = \ln 2 = 1(bit)$

Y rastgele değişkeninin entropisi,

$H(Y) = -\sum q(y)\ln q(y) = -\frac{1}{2}\ln\frac{1}{2} - \frac{1}{2}\ln\frac{1}{2} = \ln 2 = 1(bit)$

olmak üzere,

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

eşitliği sağlanır.

Bu eşitliğin her zaman sağlanmadığı aşağıdaki örnekte gösterilmiştir:

Örnek 2.12. (X,Y) gibi bir sistem ile ilgili durumlar ve bu durumlarda bulunma olasılıkları aşağıdaki gibi olsun.

Y \ X	x_1	x_2	q(y)
y_1	3/16	5/16	8/16
y_2	5/16	3/16	8/16
p(x)	8/16	8/16	1

Bu bileşke sistemin entropisi,

$H(X,Y) = -\frac{3}{16}\ln\frac{3}{16} - \frac{5}{16}\ln\frac{5}{16} - \frac{5}{16}\ln\frac{5}{16} - \frac{3}{16}\ln\frac{3}{16} = 1.34 = \frac{1.34}{\ln 2}(bit) = 1.93(bit)$

olur.

X'in entropisi,

$H(X) = -\sum p(x)\ln p(x) = -\frac{8}{16}\ln\frac{8}{16} - \frac{8}{16}\ln\frac{8}{16} = \ln 2 = 1(bit)$ ve

Y 'nin entropisi,

$H(Y) = -\sum q(y)\ln q(y) = -\frac{8}{16}\ln\frac{8}{16} - \frac{8}{16}\ln\frac{8}{16} = \ln 2 = 1(bit)$

olmak üzere,

$H(X) + H(Y) = 2\ln 2 = 2(bit) \geq 1.93(bit) = H(X,Y)$

bulunur.

Şimdi sürekli dağılımlar için entropi kavramından bahsedilecektir.

2.2.4. Sürekli Dağılımlar İçin Entropi

Bir X sürekli rastlantı değişkeninin olasılık yoğunluk fonksiyonu $f(x)$ ve bu fonksiyonun tanımlı olduğu bölge D olsun. X' in entropisi,

$$h(X) = -E[\ln f(x)] = -\int_D f(x) \ln f(x) dx \tag{2.56}$$

şeklinde tanımlanır ve diferansiyel entropi olarak adlandırılır. Diferansiyel entropi, $h(X)$ yerine $h(f)$ şeklinde de gösterilebilir [22].

Örnek 2.13. [23] $f(x)$ fonksiyonu (a, b) üzerinde düzgün dağılıma sahip olsun.

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b) \\ 0, & \text{diğer durumlar} \end{cases}$$

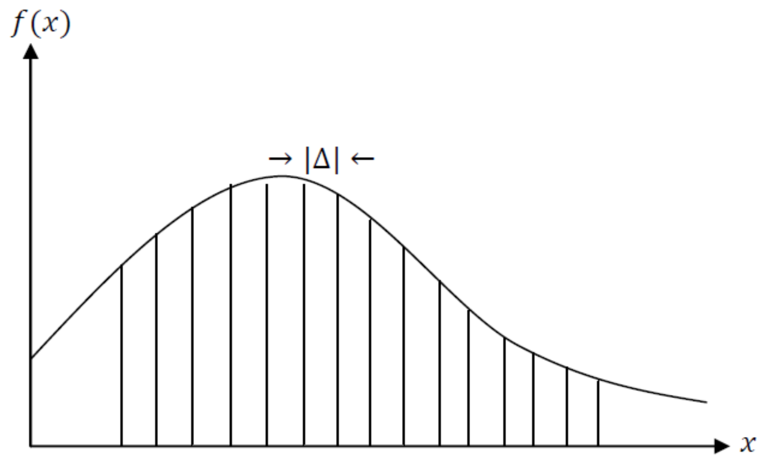
O halde, bu dağılımın sürekli entropi ile çözümü aşağıdaki gibidir:

$$\begin{aligned} h(f) &= -\int_D f(x) \ln f(x) dx \\ &= -\int_a^b \frac{1}{b-a} \ln \frac{1}{b-a} dx \\ &= -\frac{1}{b-a} \ln \frac{1}{b-a} \int_a^b dx \\ &= \ln(b-a) \end{aligned}$$

olur.

2.2.4.1. *Kesikli Entropi ile Diferansiyel Entropi İlişkisi*

Elde edilen Shannon entropi aksiyomlarının dizisinde sürekli entropi için türetme olasılığı yoktu. Bu tanımın aslı aşağıda ayrıntılı şekilde açıklanmıştır [23]:



Şekil 2.11: Kesikli Entropi ile Diferansiyel Entropiyi ifade eden şekil.

Şekil 2.11'de görüldüğü gibi $f(x)$'in altındaki alanın Δ uzunluğunda alanlara bölündüğü varsayılırsa Ortalama Değer Teoreminden,

$$f(x_i)\Delta = \int_{i\Delta}^{(i+1)\Delta} f(x)dx$$

eşitliğini sağlayan $\exists\, x_i$ değeri vardır.

Bu, Riemann integrali tanımı gereği;

$$\int_{-\infty}^{\infty} f(x)dx = \lim_{\Delta \to \infty} \sum_{i=-\infty}^{\infty} f(x_i)\Delta = 1$$

yaklaşacağını söylemektedir.

O halde, $i\Delta \leq X \leq (i+1)\Delta$ aralığındaki tüm değerler için $X^{\Delta} = x_i$ şeklinde tanımlanan bir X^{Δ} rastlantı değişkeni ele alınsın. Bu durumda olasılık değeri,

$$p_i = \int_{i\Delta}^{(i+1)\Delta} f(x)dx = f(x_i)\Delta$$

biçimindedir. Buna bağlı olarak X^{Δ}'nın entropi değeri ise,

$$
\begin{aligned}
H(X^{\Delta}) &= -\lim_{\Delta\to\infty}\sum_{i=-\infty}^{\infty} f(x_i)\Delta\ln(f(x_i)\Delta) \\
&= -\lim_{\Delta\to\infty}\sum_{i=-\infty}^{\infty}\Delta f(x_i)\ln(f(x_i)) - \lim_{\Delta\to\infty}\sum_{i=-\infty}^{\infty} f(x_i)\Delta\ln\Delta \\
&= -\int_{-\infty}^{\infty} f(x)\ln f(x)dx - \ln\Delta \qquad (2.57)
\end{aligned}
$$

olarak bulunur.

Dolayısıyla bu eşitlik, Shannon entropisinin sınırını temsil edip sürekli entropi için ifade edilen tanımla aynıdır.

Eğer $f(x)$, Riemann anlamında integrallenebilir ise, $\Delta \to 0$ iken, $-\ln(\Delta) \to -\infty$ olacağından $H(X^{\Delta}) \to \infty$ olur ve

$$
h(f) = \lim_{\mathbf{\Delta\to 0}}(H(X^{\Delta}) + \ln\Delta) = -\int_{-\infty}^{\infty} f(x)\ln f(x)dx
$$

sonucuna ulaşılır.

Kesikli rastgele değişkenler için verilen bileşik, koşullu entropi tanımları, benzer şekilde sürekli rastgele değişkenler için de ifade edilebilir.

2.2.4.2. Bileşik Entropi

Kesikli rastgele değişkenlerde olduğu gibi, bir tek sürekli rastgele değişkenler için ifade edilen sürekli entropi tanımı çok sayıda sürekli rastgele değişkenleri için de ifade edilebilir.

$f(x_1, x_2, ..., x_n)$ olasılık yoğunluk fonksiyonuna sahip $X_1, X_2, ..., X_n$ rastgele değişkenleri kümesinin sürekli entropisi,

$$
h(X_1, X_2, ..., X_n) = -\int f(x_1, x_2, ..., x_n)\ln(f(x_1, x_2, ..., x_n))dx_1 dx_2 ... dx_n \qquad (2.58)
$$

şeklinde tanımlanır [22].

2.2.4.3. Koşullu Entropi

X ve Y sürekli rastgele değişkenleri için bileşik olasılık yoğunluk fonksiyonu $f(x,y)$ olmak üzere koşullu sürekli entropi,

$$h(X|Y) = -\int f(x,y)\ln f(x|y)dxdy \tag{2.59}$$

şeklinde tanımlanır. Buradan,

$$f(x|y) = \frac{f(x,y)}{f(y)}$$

olarak yazılabildiğinden (2.59) eşitliği,

$$h(X|Y) = -\int f(x,y)\ln[\frac{f(x,y)}{f(y)}]dxdy$$

$$= h(X,Y) - h(Y)$$

şeklinde de ifade edilebilir [22].

Teorem 2.3. [24] $X_1, X_2, ..., X_n$ sürekli rastgele değişkenlerinin diferansiyel entropisi için zincir kuralı şu şekildedir:

$$h(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} h(X_i|X_1, X_2, ..., X_{i-1})$$

Kanıt:

İki rastgele değişken için bileşik entropi, denklem (2.55) eşitliğinde gösterilmişti.

$X_1, X_2, ..., X_n$, n rastgele değişkeni için bileşik entropinin elde edilmesinde tümevarım yöntemi kullanılacaktır. İlk olarak $n = 1$ için $h(X_1) = h(X_1)$ olur. Yani doğrudur. Daha sonra varsayalım ki $n-1$ rastgele değişkeni için doğru olsun. O halde, n rastgele değişkeni için diferansiyel entropi,

$$h(X_1, X_2, ..., X_n) = h(X_n|X_1, X_2, ..., X_{n-1}) + h(X_1, X_2, ..., X_{n-1})$$

$$= h(X_n|X_1, X_2, ..., X_{n-1}) + \sum_{i=1}^{n-1} h(X_i|X_1, X_2, ..., X_{i-1})$$

$$= \sum_{i=1}^{n} h(X_i|X_1, X_2, ..., X_{i-1})$$

olduğu görülür.

Sonuç 2.2. [24] Eğer $X_1, X_2, ..., X_n$ rastgele değişkenleri bağımsız ise aşağıdaki eşitlik geçerlidir:

$$h(X_1, X_2, ..., X_n) \leq \sum_{i=1}^{n} h(X_i)$$

Teorem 2.4. [24] Rastgele değişkeninde yapılan bir öteleme, diferansiyel entropinin değerini değiştirmez.

$$h(X + c) = h(X)$$

Kanıt:

$Y = X + c$ olsun. O halde, $f_Y(y) = f_X(y - c)$ ve $D_Y = \{x + c : x \in D_X\}$ olur. (2.56) eşitliğinde, $x = y - c$ yazılırsa,

$$h(X) = -\int\limits_{D_X} f_X(x) \ln f_X(x) dx$$

$$= -\int\limits_{D_Y} f_X(y - c) \ln f_X(y - c) dy$$

$$= -\int\limits_{D_Y} f_Y(y) \ln f_Y(y) dy$$

$$= h(Y) = h(X + c)$$

istenen elde edilir.

2.3. MAKSİMUM ENTROPİ İLKESİ

Bu bölümde Bayesli mantıksal çıkarımı çerçevesinde maksimum entropi ilkesi ayrıntılı olarak incelenecektir.

2.3.1. Giriş

Maksimum Entropi ilkesi, bir rastlantı değişkeninin olasılıklarının negatif olmaması ve bu olasılık değerlerinin toplamının 1 olması bilgilerinden başka rastlantı değişkeni hakkında herhangi bir bilgi olmadığı durumda bu rastlantı değişkeninin düzgün dağılımlı olması onun en çok tahmin edicisi olduğunu varsayan Laplace' ın *yetersiz neden ilkesi*' nin bir uzantısı olarak görülür [18].

Jaynes' ın entropinin maksimum edilmesi ile ilgili çalışmaları istatistiksel mekaniğin ve olasılık dağılımlarının yeniden düzenleştirilmesini ve formülüze edilmesini sağladı. Böylece istatistiksel mekanik tekrar yorumlandı ve matematiksel işlemler daha da kolaylaştırıldı [25].

Golan, Miller ve Judge ise, Jaynes' ın maksimum entropi yöntemini doğrusal regresyon üzerinden geliştirerek hata terimi üzerinden tersini alma problemlerinde uygulamışlardır.

Günümüzde maksimum entropi, pek çok bilimsel alanda geniş uygulamalara sahiptir. Özellikle görüntü işlemesi, portfolyo seçimi, emtia fiyatlandırılması gibi alanlarda kullanılmaktadır. Maksimum entropi prensibine örnek olarak; tıbbi görüntülemelerde bilgisayar ve tomografi taramasının, negatif filmlere dönüştürülmesi ve gök bilimcilerinin galaksi fotoğraflarının sonuçlarını netleştirmek için kullanmaları verilebilir [26].

Entropinin maksimize edilmesindeki temel amaç, maksimum entropi ilkesine uygun olarak var olan sonsuz sayıdaki dağılımın içinden, maksimum entropiye sahip en uygun dağılımın seçilmesidir.

2.3.2. Jaynes' ın Maksimum Entropi İlkesi

Shannon entropisi bilindiği üzere verilen dağılımın olasılık değerlerine bağlı olarak hesaplanmaktadır. Jaynes, Shannon' un ortaya çıkardığı entropi kavramını ele almıştır. Böylece eldeki problem yeniden ifade edilmiş ve *Maksimum Entropi İlkesi* ortaya çıkarılmıştır. Bu ilkeye göre, bir rastgele değişkeninin ortalama değerleriyle ilgili bilgi verildiğinde, bu bilgi ile tutarlı olan en iyi dağılım seçilir [27].

Tanım 2.9. [25] Shannon entropisinin, rastgele değişkenlerden elde edilen moment kısıtlarına göre maksimize eden yönteme *Maksimum Entropi yöntemi*, bu yöntemden elde edilen dağılıma ise *Maksimum Entropi (MaxEnt) dağılımı* denir.

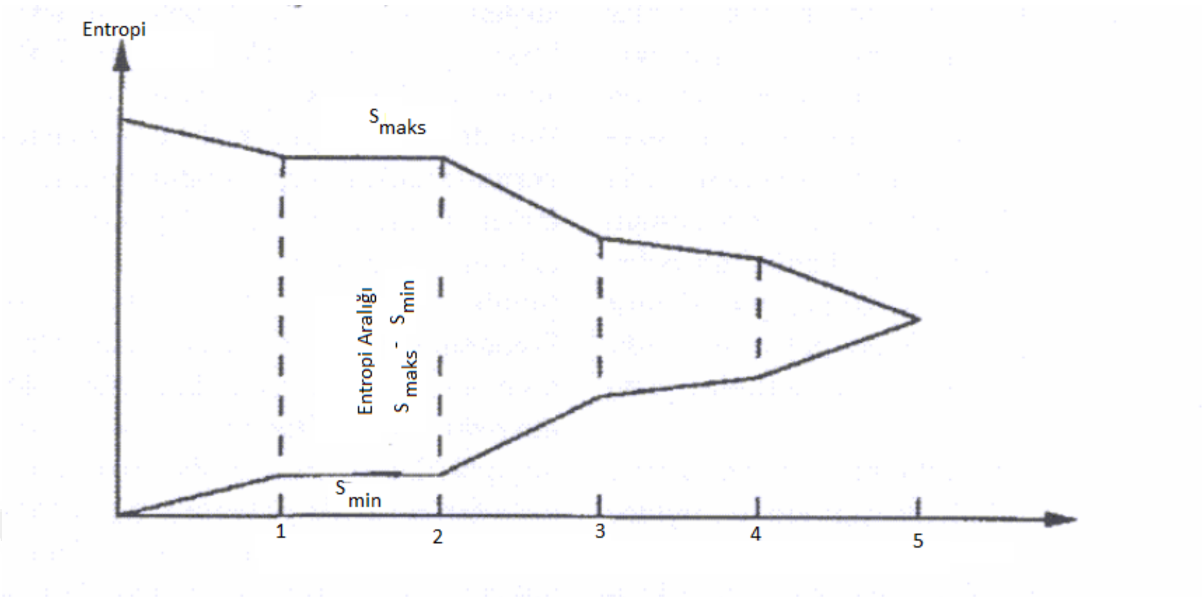
Verilen ortalama değeri ile sonsuz sayıda dağılım bulunabileceğinden, bir rastgele değişkenine ait yalnızca ortalama değerinin bilinmesi, tek başına dağılımın belirlenmesi için yeterli değildir [27]. Bu nedenle Jaynes, Claude Shannon'un belirsiz entropi olasılık dağılımları ile ön bilgiler ışığında yapılan testin olasılık dağılımları için uygun olup olmayacağını belirleyen test edilebilir bilginin nasıl birleşeceğini göstermiştir. Dolayısıyla maksimum entropi ilkesinde amaç: tek olasılık dağılımına ulaşmaktır.

Jaynes bu aşamada, bütün dağılımlar içinde maksimum entropiye sahip olan en uygun dağılımın seçilmesi gerektiğine dikkat çekmiştir. Çünkü entropi değeri, yalnızca bilinen bilgiye eklenen ek bilgiyle azaltılabilir [25].

MaxEnt metodu, eldeki bilginin dışında yeni hiçbir bilgi kullanılmadan, verilen kısıtlar altında entropi değerinin maksimize edilmesi ile ilişkilidir. Bu ilke Jaynes' ın örneği ile aşağıda açıklanmaktadır:

Üzerinde siyah örtü olan masaya, beyaz noktalı siyah bir zar atıldığı varsayılsın. Masanın üzerindeki fotoğraf makinesi, zarın her atılışında fotoğrafını çekmektedir. Masa örtüsü ve zar siyah olduğundan fotoğrafta sadece zarın beyaz noktaları görülmektedir. Binlerce defa atılan zar, film karelerinin sırası değiştirilmeden görüntülendiğinde karede düzgün bir lekeden başka birşey gözükmeyecektir [28].

Yukarıdaki deney analiz edilecek olursa aslında, fotoğrafta sadece zar üzerindeki noktalar gözükeceği için zarın kaç yüzünün olduğu bilinemeyecektir. Zarın altıdan fazla yüzünün olabilme ihtimali vardır. Ancak ortalama nokta sayısı 3.5 ise zarın altı yüzlü olma olasılığı artmış olacaktır. Dolayısıyla, elde edilen bilgi bir ölçüde artmıştır. Bu bilgiye ulaşılmasından sonra n yüzlü zarın her atılışında $p_1, p_2, ...p_n$ olasılıkları var iken, altı yüzlü zarın her atılışında $p_1, p_2, ...p_6$ olacak şekilde olasılık sayısı 6' ya düşeceğinden, bilgi artmış, belirsizlik de azalmıştır. Bu durumda, her aşamada tüm kısıtlarla uyumlu olarak S_{maks} ile gösterilen bir maksimum belirsizlik ve $S_{\min}$ ile gösterilen bir minimum belirsizlik olacaktır [18].



Şekil 2.12: Maksimum entropi temsili şekli.

Şekil 2.12' de görüldüğü gibi, başlangıçta S_{maks} en yüksek ve $S_{\min}$ de en düşük seviyededir. Her aşamada S_{maks} azalmakta ve $S_{\min}$ ise artmaktadır. Sonunda bu iki belirsizlik bir yerde buluşur. Bu buluşma noktasında tek bir durum söz konusudur. Jaynes, belirsizliği maksimum kılan düzgün dağılımın kullanılmasını önermiştir. Düzgün dağılım aşağıdaki gibi birkaç farklı şekilde ifade edilebilir [18]:

i) $x_1, x_2, ..., x_n$ değerlerini sırasıyla $p_1, p_2, ..., p_n$ olasılıklarıyla alan X rastlantı değişkeni var olsun. Bu durumda, $g_1(X), g_2(X), ..., g_m(X)$ fonksiyonlarının beklenen değerleri,

$$\sum_{i=1}^{n} p_i g_r(x_i) = a_r \quad (r = 1, 2, ..., m), \quad \sum_{i=1}^{n} p_i = 1 \tag{2.60}$$

biçiminde bulunur. Buradan anlaşıldığı gibi Denklem (2.60) ı sağlayan olasılık uzayları arasında maksimum belirsizliği veren dağılımın düzgün dağılım olduğu görülür.

ii) Hakkında oldukça az bilgi içeren dağılım, düzgün dağılımdır.

iii) En fazla rastlantı içeren dağılım, düzgün dağılımdır.

iv) Maksimum entropiden en az sapma gösteren dağılım, düzgün dağılımdır.

Bu şekilde Jaynes, düzgün dağılım ile maksimum entropiyi şu şekilde ifade etmiştir:

$$P(n_1, n_2, ..., n_6|N, p_1, p_2, ...p_6) = \underbrace{\frac{N!}{n_1!n_2!...n_6!}}_{W} \underbrace{p_1^{n_1} \times p_2^{n_2} \times ... \times p_6^{n_6}}_{P} \tag{2.61}$$

$$= W.P$$

Burada $N = n_1 + n_2 + n_3 + n_4 + n_5 + n_6$ dır.

Tanım 2.10. Faktöriyel fonksiyonu çok hızlı büyüyen bir fonksiyon olup n çok büyük olmamasına rağmen $n!$ kısa sürede hesaplanamaz. Bu durumda Stirling yaklaşımı olarak bilinen aşağıdaki formülden yararlanılır:

$$\ln(N!) \approx N \ln(N) - N \tag{2.62}$$

Şimdi bu eşitliği gösterelim. Öncelikle,
$\frac{d}{dx}(\ln(x)) = \frac{1}{x}$ ve $\frac{d}{dx}(x) = 1$ olduğu bilinmektedir. Buna göre,
$\frac{d}{dx}(x.\ln(x)) = \ln(x) + 1$ şeklinde yazılır. O halde,

$$\frac{d}{dx}(x.\ln(x) - x) = \ln(x) + 1 - 1 = \ln(x) \tag{2.63}$$

olur. Her iki taraftan integral alınırsa,

$$\int \ln(x)dx = x.\ln(x) - x \tag{2.64}$$

eşitliği elde edilir. Buradan,

$$\int_1^N \ln(x)dx = [x.\ln(x) - x]_1^N \tag{2.65}$$

$$= N.\ln(N) - N - \ln(1) + 1$$

$$= N.\ln(N) - N + 1$$

yazılır. O halde,

$$\int_1^N \ln(x)dx \approx N.\ln(N) - N \tag{2.66}$$

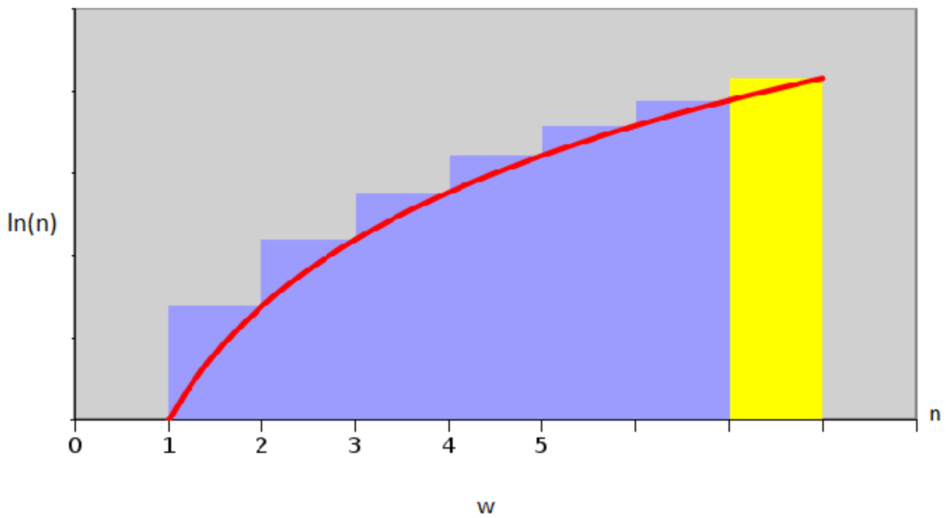
bulunur. Şekil 2.13'de Stirling yaklaşımının temsili grafiği verilmiştir. Bu şekilde w eksenindeki uzaklıklar 1 olduğu için eğrinin altında kalan alanının,

$$\int_1^n \ln(x)dx = \ln(1).1 + \ln(2).1 + ... + \ln(n).1$$

$$= \ln(1) + \ln(2) + ... + \ln(n)$$

$$= \ln[1.2.3....n] = \ln(n!)$$

olduğu görülür.



Şekil 2.13: Stirling yaklaşımı.

O halde,

$$\int_1^n \ln(x)dx = n.\ln(n) - n = \ln(n!) \tag{2.67}$$

yazılır. Buradan,

$$\ln(N!) \approx N\ln(N) - N$$

elde edilir.

Bu formül, (2.61) denkleminde gerekli düzenlemeleri yapmak için kullanılacaktır.

Şimdi, $n_i = Np_i$ ve $\sum_{i=1}^{6} p_i = 1$ olmak üzere, Stirling yaklaşımı kullanılıp, (2.61) denkleminde gerekli düzenlemeler yapılırsa,

$$\begin{aligned}
\ln W &= \ln N! - \sum_{i=1}^{6} \ln[n_i!] \\
&= N \ln N - N - \sum_{i=1}^{6} (n_i \ln n_i - n_i) \\
&= N \ln N - N - \sum_{i=1}^{6} Np_i \ln Np_i + \sum_{i=1}^{6} Np_i \\
&= N \ln N - N - \sum_{i=1}^{6} Np_i \ln(Np_i) + \sum_{i=1}^{6} Np_i \\
&= N \ln N - N - (\sum_{i=1}^{6} Np_i \ln N + Np_i \ln p_i) + \sum_{i=1}^{6} Np_i \\
&= N \ln N - N - N(\sum_{i=1}^{6} p_i \ln p_i + \ln N) + N \\
&= -N \sum_{i=1}^{6} p_i \ln p_i = NS \qquad (2.68)
\end{aligned}$$

elde edilir.

2.3.3. Genelleştirilmiş Maksimum Entropi İlkesi

Yukarıda p_i değerleri $i = 1, 2, ..., 6$ olduğundaki durum incelendi. Şimdi durum sayısı $i = 1, 2, ..., M$ için genellenirse, Denklem (2.61),

$$P(n_1, n_2, ..., n_M | N, p_1, p_2, ...p_M) = \frac{N!}{n_1! n_2! ... n_M!} m_1^{n_1} \times m_2^{n_2} \times ... \times m_M^{n_M} \qquad (2.69)$$

formuna gelir. Bu eşitlikte her iki tarafın doğal logaritma fonksiyonu alınır ve Stirling yaklaşımı kullanılırsa,

$$\ln P(n_1, n_2, ..., n_M | N, p_1, p_2, ...p_M) = \sum_{i=1}^{M} n_i \ln[m_i] + \ln[N!] - \sum_{i=1}^{M} \ln[n_i!]$$

$$= \sum_{i=1}^{M} n_i \ln[m_i] - N \sum_{i=1}^{M} p_i \ln p_i \tag{2.70}$$

bulunur. $n_i = Np_i$ eşitliği Denklem (2.70) de yerine yazılırsa,

$$\frac{1}{N} \ln P(n_1, n_2, ..., n_M | N, p_1, p_2, ...p_M) = \sum_{i=1}^{M} p_i \ln[m_i] - \sum_{i=1}^{M} p_i \ln[p_i]$$

$$= -\sum_{i=1}^{M} p_i ln[\frac{p_i}{m_i}] = S \tag{2.71}$$

elde edilir. Bu eşitlik genelleştirilmiş entropi, *Shannon-Jaynes entropisi* olarak adlandırılır [12].

2.3.4. Maksimum Entropi İlkesinin Analitik Çözümü

Maksimum entropi fonksiyonuna ait problemlerin çözümünde Lagrange çarpanları yöntemi kullanılır. Lagrange yöntemi, koşullu ekstremum yöntemi olup herhangi bir fonksiyonun minimum veya maksimum noktalarının bulunmasında kullanılır. Yöntem aşağıda açıklanmıştır [29]:

$u = f(x_1, x_2, ..., x_n)$ ve $g_k(x_1, x_2, ..., x_n)$ $(k = 1, 2, ..., m; \quad m < n)$, R^n uzayının bir D bölgesinde diferansiyellenebilen fonksiyonlar olsun. Ayrıca $g_k(x_1, x_2, ..., x_n)$ fonksiyonu D' de fonksiyonel bağlı olmadığı kabul edilsin.

$f(x_1, x_2, ..., x_n)$ fonksiyonunu,

$$g_k(x_1, x_2, ..., x_n) = 0 \qquad (k = 1, 2, ..., m; m < n) \tag{2.72}$$

koşulları altında optimize etmek için Lagrange fonksiyonu:

$$L(x_1, x_2, ..., x_n, \lambda_1, ..., \lambda_m) = f(x_1, x_2, ..., x_n) + \sum_{k=1}^{m} \lambda_k g_k(x_1, x_2, ..., x_n) \tag{2.73}$$

olarak tanımlanır. Burada $\lambda_1, \lambda_2, ..., \lambda_m$ sabitler olup *Lagrange çarpanları* olarak adlandırılır.

$f(x_1, x_2, ..., x_n)$ fonksiyonunun Denklem (2.72) koşulları altındaki ekstremumları,

$(n+m)$ değişkenli Lagrange fonksiyonunun serbest ekstremumları arasından aranır. Buna göre, aranan $x_1, x_2, ..., x_n, \lambda_1, ..., \lambda_m$ değerleri,

$$\frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} + \lambda_1 \frac{\partial g_1}{\partial x_1} + \lambda_2 \frac{\partial g_2}{\partial x_1} + ... + \lambda_m \frac{\partial g_m}{\partial x_1} = 0$$

$$\frac{\partial L}{\partial x_2} = \frac{\partial f}{\partial x_2} + \lambda_1 \frac{\partial g_1}{\partial x_2} + \lambda_2 \frac{\partial g_2}{\partial x_2} + ... + \lambda_m \frac{\partial g_m}{\partial x_2} = 0$$

$$\vdots$$

$$\vdots$$

$$\frac{\partial L}{\partial x_n} = \frac{\partial f}{\partial x_n} + \lambda_1 \frac{\partial g_1}{\partial x_n} + \lambda_2 \frac{\partial g_2}{\partial x_n} + ... + \lambda_m \frac{\partial g_m}{\partial x_n} = 0$$

ve

$$\frac{\partial L}{\partial \lambda_1} = g_1(x_1, x_2, ..., x_n) = 0$$

$$\frac{\partial L}{\partial \lambda_2} = g_2(x_1, x_2, ..., x_n) = 0$$

$$\vdots$$

$$\vdots$$

$$\frac{\partial L}{\partial \lambda_m} = g_m(x_1, x_2, ..., x_n) = 0$$

denklem sistemlerinin birlikte çözülmesiyle elde edilir.
$(x_1, x_2, ..., x_n, \lambda_1, ..., \lambda_m)$ için $\frac{\partial^2 L}{\partial x_i^2} < 0$ ise, $f(x_1, x_2, ..., x_n)$ fonksiyonu $(x_1, x_2, ..., x_n)$ noktasında bir yerel maksimuma, $\frac{\partial^2 L}{\partial x_i^2} > 0$ ise, bir yerel minimuma sahiptir.

Aşağıda maksimum entropi için bu durum incelenmektedir [12].
Varsayalım; Y_i, $i = 1, ...M$' e kadar farklı olasılıklara sahip olsun. $P(Y_i|I)$ değerine ulaşılmak istenmektedir. Eğer S, $P(Y_i|I)$' nın entropisini temsil ediyor ise maksimum entropi problemi için verilen durum şu şekildedir:

$$dS = \frac{\partial S}{\partial p_1} dp_1 + \frac{\partial S}{\partial p_2} dp_2 + ... + \frac{\partial S}{\partial p_M} dp_M = 0 \tag{2.74}$$

Yukarıdaki denklem, bir kısıtlama olmadan, eğer tüm kat sayılar ayrı ayrı sıfıra eşit ise,

dp_i' lerin bağımsız ve tek çözümü olduğunu açıklamaktadır.
Varsayalım, verilen sınırlandırma, R sabit olmak üzere, $\sum_{i=1}^{M} p_i^2 = R$ olsun. Sınırlandırma tekrardan yazılırsa, $C = \sum_{i=1}^{M} p_i^2 - R = 0$ ve türev alınırsa,

$$dC = \frac{\partial C}{\partial p_1}dp_1 + \frac{\partial C}{\partial p_2}dp_2 + ... + \frac{\partial C}{\partial p_M}dp_M = 0$$

$$= 2p_1dp_1 + 2p_2dp_2 + ... + 2p_Mdp_M \tag{2.75}$$

elde edilir. λ belirsizlik çarpanı olmak üzere, dS ile dC birleştirilirse, Lagrange fonksiyonu,

$$L = dS - \lambda dC = 0 \tag{2.76}$$

olarak yazılır. Denklem (2.74) ve Denklem (2.75), Denklem (2.76) da yerine yazılırsa,

$$(\frac{\partial S}{\partial p_1} - \lambda 2p_1)dp_1 + (\frac{\partial S}{\partial p_2} - \lambda 2p_2)dp_2 + ... + (\frac{\partial S}{\partial p_M} - \lambda 2p_M)dp_M = 0$$

eşitliği yazılır. Burada $\left(\frac{\partial S}{\partial p_1} - \lambda 2p_1\right) = 0$ olarak alınırsa eşitlik,

$$\left(\frac{\partial S}{\partial p_2} - \lambda 2p_2\right)dp_2 + ... + \left(\frac{\partial S}{\partial p_M} - \lambda 2p_M\right)dp_M = 0 \tag{2.77}$$

durumuna gelir. Geri kalan $M-1$ değişken olarak, dp_i' ler bağımsız olarak düşünülebilir. Çünkü (2.77) eşitliğini tahmin etmek için, katsayıları sıfıra eşit olmak zorundadır. Tüm p_i için, M eşitliği çözülebilir. Bu ise S için, yerel maksimum olduğunu göstermektedir.

2.3.5. Kesikli Maksimum Entropi Dağılımları

Bir X kesikli rastgele değişkeni, $1, 2, , ..., n$ şeklinde sonlu değerleri olsun. İlgilenilen değişkenle ilgili maksimum entropi olasılık dağılımları, verilen kısıtlara bağlı olarak bulunabilir.
Bu kısımda; öncelikle bir daha sonra iki kısıt olması durumunda maksimum entropi dağılımlarının oluşturulması araştırılacaktır.

2.3.5.1. Bir Kısıtlı Maksimum Entropi Dağılımı

$1, 2, ..., M$ değerlerini $p_1, p_2, ..., p_M$ olasılıkları ile alan X rastgele değişkeni için,

$$-\sum_{i=1}^{M} p_i ln[\frac{p_i}{m_i}] = S \tag{2.78}$$

entropi fonksiyonunu yalnızca,

$$\sum_{i=1}^{M} p_i = 1$$

kısıtı altında maksimize eden olasılık dağılımını bulmak için Lagrange fonksiyonu,

$$L = -\sum_{i=1}^{M} pi \ln[\frac{p_i}{m_i}] - \lambda\big(\sum_{i=1}^{M} p_i - 1\big) \tag{2.79}$$

şeklinde yazılabilir. Eşitliğin $p_1, p_2, ..., p_M$' lere göre aşağıdaki gibi türevi alınır ve sıfıra eşitlenirse [12],

$$d[-\sum_{i=1}^{M} pi \ln[\frac{p_i}{m_i}] - \lambda(\sum_{i=1}^{M} p_i - 1)] = 0 \tag{2.80}$$

$$d[-\sum_{i=1}^{M} pi \ln p_i + \sum_{i=1}^{M} pi \ln m_i - \lambda(\sum_{i=1}^{M} p_i - 1)] = 0$$

$$\sum_{i=1}^{M} \big(- \ln p_i - p_i \frac{\partial \ln p_i}{\partial p_i} + \ln m_i - \lambda \frac{\partial p_i}{\partial p_i}\big) dp_i = 0$$

$$\Rightarrow \sum_{i=1}^{M} (-\ln[\frac{p_i}{m_i}] - 1 - \lambda) dp_i = 0$$

bulunur. Burada λ Lagrange çarpanını bulmak için,

$$-\ln[\frac{p_i}{m_i}] - 1 - \lambda = 0 \tag{2.81}$$

yazılır. Eşitlikten p_i olasılığı,

$$p_i = m_i e^{-(1+\lambda)} \tag{2.82}$$

şeklinde ifade edilir. $\sum_{i=1}^{M} p_i = 1$ olduğu için, (2.82) eşitliği ile her iki taraftan $\sum_{i=1}^{M}$ için toplam alınırsa,

$$\sum_{i=1}^{M} m_i e^{-(1+\lambda)} = 1 = e^{-(1+\lambda)} \sum_{i=1}^{M} m_i \tag{2.83}$$

elde edilir. $\sum_{i=1}^{M} m_i = 1$ olduğu için $\lambda = -1$ olur. Böylece;

$$p_i = m_i \tag{2.84}$$

elde edilir.

a sabit olmak üzere, $m_i = a = \frac{1}{n}$ olsun. O halde, (2.84) eşitliğinden,

$$P(Y|I) = m(y) = \frac{1}{n} \tag{2.85}$$

olasılığı bulunur.

Bu ise, $m(y) = a$ sabit olduğu ve minimal sınırlandırma $\int p(y) = 1$ olduğu için $\sum_{i=1}^{M} p_i = 1$ kısıtından başka herhangi bir bilgi verilmemesi durumunda belirsizliği maksimum olan dağılımın *düzgün dağılım* manasına gelir.

2.3.5.2. İki Kısıtlı Maksimum Entropi Dağılımı

Bir μ ortalama değerinin verilmesi durumunda, X kesikli rastgele değişkeninin maksimum entropi olasılık dağılımı bulunmak istensin. Denklem (2.78) deki entropi fonksiyonu,

$$\sum_{i=1}^{M} p_i = 1$$

ve

$$\sum_{i=1}^{M} y_i p_i = \mu$$

kısıtları altında maksimize edilsin. Bu durumda Lagrange fonksiyonu,

$$L = -\sum_{i=1}^{M} pi \ln[\frac{p_i}{m_i}] - \lambda(\sum_{i=1}^{M} p_i - 1) - \lambda_1(\sum_{i=1}^{M} y_i p_i - \mu) \tag{2.86}$$

şeklinde yazılabilir. Lagrange çarpanı λ ve λ_1 olacaktır. Bu fonksiyonun p_i' lere göre aşağıdaki gibi türevi alınır ve sıfıra eşitlenirse [12]

$$\sum_{i=1}^{M}\big(-\ln[\frac{p_i}{m_i}]-p_i\frac{\partial \ln p_i}{\partial p_i}-\lambda\frac{\partial p_i}{\partial p_i}-y_i\lambda_1\frac{\partial p_i}{\partial p_i}\big)dp_i=0 \tag{2.87}$$

elde edilir. Buradan,

$$\sum_{i=1}^{M}(-\ln[\frac{p_i}{m_i}]-1-\lambda-y_i\lambda_1)dp_i=0$$

ve

$$-\ln[\frac{p_i}{m_i}]-1-\lambda-y_i\lambda_1=0 \tag{2.88}$$

yazılır. Eşitlikten p_i olasılığı,

$$p_i=m_ie^{-(1+\lambda)}e^{-\lambda_1 y_i} \tag{2.89}$$

şeklinde ifade edilir.

İlk kısıt $\sum\limits_{i=1}^{m}p_i=1$ olduğu için, (2.89) eşitliğinde kullanılırsa,

$$\sum_{i=1}^{M}p_i=e^{-(1+\lambda)}\sum_{i=1}^{M}m_ie^{-\lambda_1 y_i}=1 \tag{2.90}$$

şeklinde yazılır. Böylece,

$$e^{-(1+\lambda)}=\frac{1}{\sum\limits_{i=1}^{M}m_ie^{-\lambda_1 y_i}} \tag{2.91}$$

olur.

İkinci kısıttan ise,

$$\sum_{i=1}^{M}y_ip_i=\mu=\frac{\sum\limits_{i=1}^{M}y_im_ie^{-\lambda_1 y_i}}{\sum\limits_{i=1}^{M}m_ie^{-\lambda_1 y_i}} \tag{2.92}$$

veya,

$$\sum_{i=1}^{M}y_im_ie^{-\lambda_1 y_i}-\mu\sum_{i=1}^{M}m_ie^{-\lambda_1 y_i}=0 \tag{2.93}$$

yazılabilir.

$P(Y|I)$ sürekli olması durumunda p_i için, (2.89) eşitliği genelleştirilirse;

$$P(Y|I) = m(y)e^{-(1+\lambda)}e^{-\lambda_1 y} \tag{2.94}$$

olur.

Eğer $m(y)$ sabit ise,

$$P(Y|I) \quad \propto \quad e^{-\lambda_1 y} \tag{2.95}$$

bulunur. Normalizasyon sonucu λ_1 kolayca bulunabilir. Gerekli işlemler yapıldığında,

$$P(Y|\mu) = \frac{1}{\mu}e^{\frac{-y}{\mu}} \qquad\qquad y \geq 0 \tag{2.96}$$

elde edilir.

Sonuç olarak, bu iki kısıttan başka herhangi bir bilgi verilmemesi durumunda belirsizliği maksimum olan dağılımın *üstel dağılım* olduğu görülür.

2.3.6. Sürekli Maksimum Entropi Dağılımları

Sürekli durumlarda belirsizliği doğru ölçmek için,

$$S_c = -\int p(y)\ln[\frac{p(y)}{m(y)}]dy \tag{2.97}$$

formülü kullanılır. Burada $m(y)$ miktarı, *Lebesque miktarı* [12] olarak adlandırılır. Eğer $m(y)$ sabit ise bu eşitlik,

$$S_c = -\int p(y)\ln p(y)dy + \ln m(y)\int p(y)dy \tag{2.98}$$

yani,

$$S_c = -\int p(y)\ln p(y)dy + sabit$$

haline dönüşür. Maksimum entropi çözümünü bulmak için, (2.98) eşitliğin türevleri ile ilgilenildiğinden dolayı sabitin hiçbir etkisinin olmadığı görülür [12].

Tanım 2.11. [12] Hata fonksiyonu veya Gauss hata fonksiyonu olarak adlandırılan fonksiyon, özel bir fonksiyondur. Bu fonksiyon şu şekilde tanımlanır:

$$erf(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-u^2} du \tag{2.99}$$

Hata fonksiyonu için aşağdakiler geçerlidir:

i) Hata fonksiyonu, tek fonksiyondur. Yani, $erf(-z) = -erf(z)$ dır.

ii) $erf(\infty) = 1$ ve $erf(-\infty) = -1$ dır.

2.3.6.1. İki Kısıtlı Maksimum Entropi Dağılımı

$m(y)$ aşağıdaki gibi tanımlansın:

$$m(y) = \begin{cases} \frac{1}{(y_U - y_A)} & , \ y_A < y < y_U \\ 0 & , \text{diğer durumlar} \end{cases} \tag{2.100}$$

Sınırlandırmalar ise,

$$\int_{y_A}^{y_U} p(y)dy = 1$$

ve

$$\int_{y_A}^{y_U} (y-\mu)^2 p(y)dy = \sigma^2$$

olsun.

Entropi fonksiyonu,

$$H(p(y)) = -\int p(y) \ln p(y) dy$$

yukarıda verilen kısıtlar altında maksimize edilip çözülürse Lagrange fonksiyonundan,

$$L = -\int p(y) \ln p(y) dy - \lambda[\int_{y_A}^{y_U} p(y)dy - 1] - \lambda_1[\int_{y_A}^{y_U} (y-\mu)^2 p(y)dy - \sigma^2] \tag{2.101}$$

yazılır. Buradan p'ye göre türev alınıp sıfıra eşitlendiğinde,

$$-\ln p - 1 - \lambda - \lambda_1(y-\mu)^2 = 0 \tag{2.102}$$

veya

$$p = e^{-(1+\lambda)}e^{-\lambda_1(y-\mu)^2} \tag{2.103}$$

elde edilir.

$$1 + \lambda = \lambda_0$$

denilirse,

$$p(y) = e^{-\lambda_0}e^{-\lambda_1(y-\mu)^2} \tag{2.104}$$

bulunur.

λ_0 ve λ_1 kısıtlarına göre çözüm yapılacaktır. İlk sınırlandırmadan,

$$\int_{y_A}^{y_U} p(y)dy = 1 = e^{-\lambda_0}\int_{y_A}^{y_U} e^{-\lambda_1(y-\mu)^2}dy \tag{2.105}$$

yazılır. Hata fonksiyonu yardımıyla, (2.105) eşitliğinde gerekli işlemler yapıldığında λ_0 parametresine göre çözümü,

$$\lambda_0 = \ln[\frac{\sqrt{\pi}}{2\sqrt{\lambda_1}}] + \ln[erf\{\sqrt{\lambda_1}(y_U-\mu)\} - erf\{\sqrt{\lambda_1}(y_A-\mu)\}] \tag{2.106}$$

bulunur. Burada integrasyon y_A ve y_U sınırlarına bağlı olacaktır.

Şimdi integrasyon sınırlarına bağlı olarak koşullar şu şekilde olsun:

$$\sqrt{\lambda_1}(y_U-\mu) \gg 1 \qquad ve \qquad \sqrt{\lambda_1}(y_A-\mu) \ll -1 \tag{2.107}$$

Bu durumda

$$erf\{\sqrt{\lambda_1}(y_U-\mu)\} \approx 1 \qquad ve \qquad erf\{\sqrt{\lambda_1}(y_A-\mu)\} \approx -1$$

olur.

(2.106) eşitliğinden,

$$\lambda_0 \approx \ln[\frac{\sqrt{\pi}}{2\sqrt{\lambda_1}}] + \ln[2] = \ln[\sqrt{\frac{\pi}{\lambda_1}}] \tag{2.108}$$

bulunur. Bu eşitliğin her iki tarafının λ_1' e göre türevi alınırsa,

$$-\frac{\partial \lambda_0}{\partial \lambda_1} = -\frac{\partial \ln[\sqrt{\frac{\pi}{\lambda_1}}]}{\partial \lambda_1} = \frac{1}{2\lambda_1} = \sigma^2 \tag{2.109}$$

ve (2.108) ve (2.109) eşitlikleri birleştirilirse,

$$\lambda_1 = \frac{1}{2\sigma^2} \qquad\qquad e^{-\lambda_0} = \frac{1}{\sqrt{2\pi}\sigma}$$

elde edilir.

Sonuç olarak, entropiyi maksimize eden dağılımın

$$p(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \tag{2.110}$$

Normal Dağılım olduğu bulunur.

3. MALZEME VE YÖNTEM

Bu tez çalışmasında öncelikle olasılık ve istatistikteki temel kavramlar ile literatürde sıkça kullanılan dağılımlara yer verilmiştir. Tezin özünü oluşturan entropi ve maksimum entropi kavramları Bayesli mantıksal çıkarım çerçevesinde incelenmiştir. Bu kavramlar hakkında genel bilgiler verilmiştir.

Bu çalışma yapılırken entropi alanındaki kitaplar kütüphane ve web aracılığı ile taranmıştır. Bunların içinden konu hakkında referans alınanlar kaynaklar bölümünde belirtilmiştir. Uygulamada kullanılan ilk veri kümesi olan günlük süt üretimine ait veriler Sütaş Genel Müdürlüğünden alınmıştır. İkinci veri ise daha önce literatürde var olan Jaynes' ın zar deneyinden elde edilen veriler, oluşturulan Mathematica kodu ile Bayesli mantıksal çıkarım çerçevesinde incelenmiştir. Burada bahsedilen problemlerin uygulamaları ve sonuçları için Mathematica 9.1 Software programı kullanılmıştır. Tezin doküman haline getirilmesi için Latex programı kullanılmıştır.

4. BULGULAR

Çalışmanın bu bölümünde ilk olarak ortalama ve varyans için parametrik olmayan olasılık dağılımının gelişiminde Maksimum Entropi ve Bayesli Mantıksal Çıkarımın kullanılması söz konusu olacaktır. Bunun uygulaması olarak günlük süt üretiminin dağılımının Maksimum Entropi ve Bayesli analiz kullanılarak parametrik olmayan dağılımlarla ifade edilmesine çalışılacaktır. İkinci olarak ise Jaynes' ın zar deneyinden yararlanarak Maksimum Entropinin Bayesli yaklaşımla olan ilişkisi açıklanacaktır.

4.1. ORTALAMA VE VARYANSA AİT PARAMETRİK OLMAYAN DAĞILIMLAR İÇİN MAKSİMUM ENTROPİ VE BAYESLİ ANALİZİ

Anakitle ortalaması ve varyansı genel olarak örneklemden elde edilen örneklem ortalaması ve varyansı bulunarak tahmin edilir. Bir anakitle, normalizasyon için gerekli koşulları sağlarsa örneklem ortalamasının ve varyansının dağılımı bellidir. Fakat, normallik varsayımı olmadığı zaman birçok gerçek kitleden elde edilen veriler sınırlı olabilir. Örneğin; kitlenin ağırlık ve yükseklikleri negatif olmayan değerlerdir fakat bunlar pozitif yönde sınırsız olabilir. Bunların sonsuz sınırları, normal dağılım için uygun olmayabilir. Bu nedenle ortalama ve standart sapmanın parametrik olmayan dağılımları Bayesli mantıksal çıkarım ve maksimum entropi ilkesi baz alınarak oluşturulabilir. Bu verilerden yapılacak olan tahminde Merkezi Limit Teoremi ve Büyük Örneklem Teorisi kullanılmaktadır [30].

Bir anakitlenin ortalama ve varyans gibi parametrelerini, örneklem istatistiklerinden yararlanılarak olasılık dağılımlarını üretmek için Bayesli analiz yöntemi ve maksimum entropi ilkesi kullanılır. Bu metot, parametrik değildir ve örneklem büyüklüğünün özelliklerine veya herhangi dağılımsal varsayımlara yer vermez. Parametrelerin tahmin edilmesi, ya ortalama için bağımsız şekilde ya da ortalama ve varyans için aynı anda ortak olasılık dağılımın oluşturulması aracılığıyla gerçekleşir. Aşağıda tanıtılacak ilkeler doğrultusunda ineklerin süt verilerinin, ortalaması ve standart sapması için uygulaması gösterilecektir.

Denklem (2.9) kullanılarak aşağıdaki eşitlik elde edilir:

$$p(y|x,C) = \frac{p(x|y,C)p(y|C)}{\int\limits_a^b p(x|y,C)p(y|C)dy} \tag{4.1}$$

Burada, y ilgilenilen parametre, x gözlemlenen data, C ise herhangi bir önsel bilgidir.

Shannon'un kesikli dağılımlar için bulmuş olduğu entropi formülü Denklem (2.49) da verilmiştir. Diğer yandan sürekli olasılık dağılımları için entropi formülü,

$$S(p(x)) = -\int_a^b p(x)\ln(\frac{p(x)}{m(x)})dx \tag{4.2}$$

olup burada, $m(x) = \frac{1}{b-a}$ alınırsa (2.49) eşitliği elde edilir. (4.2) eşitliği belirli bir dağılım için bir belirsizlik ölçüsü sağlarken, entropi dağılımını belirlemeyi sağlar ve $p(x)$ üzerindeki bilinen kısıtlamaları maksimize eder.

4.1.1. Dağılımın Ortalaması

Gull ve Fielden [31] parametrik olmayan dağılımlarda ortalama ve standart sapmanın bulunmasıyla ilgili çalışmalarda bulunmuştur. Onların yapmış oldukları çalışmalarda Bayesli analiz ve maksimum entropi bu durumlar için kullanılırken, önsel dağılımlar için Gauss dağılımı, hedef dağılımı olarak ise entropi alınmıştır.

Bu uygulamada, dağılımlar sadece varsayımlara bağlı olup ilgili kitlenin momentleri vardır ve eldeki veriler sınırlıdır.

Birbirinden istatistiksel olarak bağımsız olan veriler için, tek bir veri noktasında çözüm bulunabilir ve daha büyük örneklem boyutu için genelleştirilebilir.

μ; ortalama ve x; örneklem veri noktası olsun. (4.1) eşitliğinden, $y \to \mu$ yazılsın.

Bu durumda ortalamanın sonsal olasılık fonksiyonu;

$$p(\mu|x) = \frac{p(x|\mu)p(\mu)}{\int_a^b p(x|\mu)p(\mu)d\mu} \tag{4.3}$$

elde edilir. Burada $p(x|\mu)$ olabilirlik fonksiyonunu atamak için (4.2) eşitliği ile verilen formda entropi, aşağıda verilen sınırlandırma ile maksimize edilecektir.
X rastlantı değişkeninin ortalaması,

$$E[x] = \int_a^b xp(x|\mu)dx = \mu \tag{4.4}$$

şeklinde yazılır. Lagrange fonksiyonu,

$$L = -\int_a^b p(x|\mu)\ln p(x|\mu)dx + \lambda(\int_a^b xp(x|\mu)dx - \mu) \tag{4.5}$$

olarak ifade edilir. L' nin p' ye göre türevi alınıp sıfıra eşitlendiğinde,

$$-(\ln p(x|\mu) + 1) + \lambda x = 0$$

elde edilir. Bu durumda olabilirlik fonksiyonu:

$$p(x|\mu) = e^{-1+\lambda x} \tag{4.6}$$

bulunur. Bu olasılık dağılımı normalize edilirse,

$$\int_a^b e^{-1+\lambda x}dx = 1 \tag{4.7}$$

şeklinde yazılır.

İntegral alınıp gerekli düzenlemeler yapıldığında,

$$e^{\lambda b} - e^{\lambda a} = e\lambda$$

$$e = \frac{e^{\lambda b} - e^{\lambda a}}{\lambda}$$

eşitliği elde edilir. Bulunan bu e değeri (4.6) eşitliğinde yerine yazıldığında, olabilirlik fonksiyonu:

$$p(x|\mu) = \frac{\lambda e^{\lambda x}}{e^{\lambda b} - e^{\lambda a}} \tag{4.8}$$

bulunur.

Sonsal dağılımın ise (4.3) eşitliğinden,

$$p(\mu|x) \propto p(x|\mu)p(\mu)$$

olduğu bilinmektedir. Diğer yandan önsel olasılık; düzgün dağılım olarak seçildiğinde, ortalamanın sonsal olasılık yoğunluk fonksiyonu:

$$p(\mu|x) = \frac{\lambda e^{\lambda x_1}}{e^{\lambda b} - e^{\lambda a}} . \frac{1}{b - a} \tag{4.9}$$

bulunur.

Böylece ortalama için sonsal olasılık elde edilmiş olur. Diğer yandan varyans için de aynı işlemler yapılabilir. Fakat Lagrange çarpanı kullanılarak standart sapma, analitik olarak çözülemez. Çözüm, nümerik olarak bulunmalıdır. Problemi çözmek için öncelikle sınırlar $[-1, 1]$ aralığında uygun hale getirilir. Herhangi bir gerçek problemin uygun çözümü için bu sınırlar içinde dönüşüm yapılması gerekir. λ için nümerik çözüm (4.8) eşitliğinde bulunan ifadeden ve (4.4) eşitliğindeki sınırlandırma ile elde edilir. Bu çözüm kullanılarak, $p(\mu|x)$ sonsal olasılığı elde etmek için her bir veri için olabilirlik fonksiyonu belirlenebilir ve önsel olasılıkla birleştirilebilir.

4.1.2. Ortalama ve Standart Sapmanın Ortak Dağılımı

Ortalama için söz konusu yapılan işlemler benzer şekilde iki boyutlu ortak olasılık dağılımında ortalama ve varyans içinde gerçekleşmektedir. Bu durumda çözümün nümerik olarak hesaplanması için iki Lagrange çarpanına gerek vardır. Ortalamada olduğu gibi, problemi çözmek için normalizasyondan dolayı sınırlar $[-1, 1]$ haline getirilip gerekli dönüşümler yapıldıktan sonra gerçek veri kümeleri için dağılımların değerlendirilmesi söz konusu olacaktır [30].

Bayes Teoreminden, verilen x verisi için μ ve σ' nın bileşik olasılık yoğunluk fonksiyonu,

$$p(\mu,\sigma|x)=\frac{p(x|\mu,\sigma)p(\mu,\sigma)}{\int\limits_{\sigma}\int\limits_{\mu}p(x|\mu,\sigma)p(\mu,\sigma)d\mu d\sigma} \tag{4.10}$$

ve entropi fonksiyonu,

$$S(p(x|\mu,\sigma))=-\int\limits_{a}^{b}p(x|\mu,\sigma)\ln(\frac{p(x|\mu,\sigma)}{\frac{1}{(b-a)}})dx \tag{4.11}$$

şeklinde tanımlansın. (4.11) eşitliğinin maksimize edilmesi için verilen kısıtlamalar ise,

$$\int\limits_{a}^{b}p(x|\mu,\sigma)dx=1 \tag{4.12}$$

$$\int\limits_{a}^{b}xp(x|\mu,\sigma)dx=\mu \tag{4.13}$$

$$\int\limits_{a}^{b}(x-\mu)^2p(x|\mu,\sigma)dx=\sigma^2 \tag{4.14}$$

şeklinde olsun.

Matematiksel işlemlerin daha kolay bulunması için önce (4.13) eşitliğinde gerekli düzenlemeler yapılırsa,

$$\int\limits_{a}^{b}xp(x|\mu,\sigma)dx-\mu.1=0$$

yazılır. Burada Denklem (4.12) kullanılırsa,

$$\int\limits_{a}^{b}xp(x|\mu,\sigma)dx-\int\limits_{a}^{b}\mu p(x|\mu,\sigma)dx=0$$

ve

$$\int\limits_{a}^{b}(x-\mu)p(x|\mu,\sigma)dx=0 \tag{4.15}$$

elde edilir.

Diğer yandan varyansın, $\sigma^2 = E(X^2) - \mu^2$ olduğu bilinmektedir. $E(X^2) = \upsilon^2$ denilirse,

$$\upsilon^2 = \sigma^2 + \mu^2 \tag{4.16}$$

yazılır. Bir başka şekilde,

$$\upsilon^2 = \int_a^b x^2 p(x|\mu, \sigma)dx$$

olup, (4.14) eşitliğinde gerekli düzenlemeler yapılırsa,

$$\int_a^b (x-\mu)^2 p(x|\mu, \sigma)dx - \sigma^2 = 0$$

$$\int_a^b [(x-\mu)^2 - \sigma^2] p(x|\mu, \sigma)dx = 0$$

sınırlandırmaları yazılır.

$\upsilon^2 = \int_a^b x^2 p(x|\mu, \sigma)dx = \int_a^b (\sigma^2 + \mu^2) p(x|\mu, \sigma)dx$ olduğundan, (4.14) eşitliği,

$$\int_a^b [x^2 - (\sigma^2 + \mu^2)] p(x|\mu, \sigma)dx = 0 \tag{4.17}$$

formunu alır.

Bu durumda, Lagrange fonksiyonu:

$$\begin{aligned} L = &- \int_a^b p(x|\mu, \sigma) \ln(p(x|\mu, \sigma))dx - \lambda_0 (\int_a^b p(x|\mu, \sigma)dx - 1) - \lambda_1 (\int_a^b (x-\mu) p(x|\mu, \sigma)dx) \\ &- \lambda_2 (\int_a^b (x^2 - (\sigma^2 + \mu^2)) p(x|\mu, \sigma)dx \end{aligned}$$

olarak yazılır. L' nin, p' ye göre türevi alınıp sıfıra eşitlenirse,

$$\ln p(x|\mu, \sigma) + 1 + \lambda_0 + \lambda_1(x-\mu) + \lambda_2(x^2 - (\mu^2 + \sigma^2)) = 0$$

bulunur. Buradan,

$$\ln p(x|\mu,\sigma) = -1 - \lambda_0 - \lambda_1(x-\mu) - \lambda_2(x^2 - (\mu^2+\sigma^2))$$

elde edilir. $1+\lambda_0 = \lambda_3$ denilirse, olabilirlik fonksiyonu;

$$p(x|\mu,\sigma) = e^{-\lambda_3-\lambda_1(x-\mu)-\lambda_2(x^2-(\mu^2+\sigma^2))} \tag{4.18}$$

elde edilir.

Normalizasyon şartları altında λ_3 bulunabilir. Yani, $\int_a^b p(x|\mu,\sigma)dx = 1$ olduğundan (4.18) eşitliğinde bulunan sonuç,

$$\int_a^b e^{-\lambda_3-\lambda_1(x-\mu)-\lambda_2(x^2-(\mu^2+\sigma^2))}dx = 1$$

ve

$$e^{\lambda_3} = \int_a^b e^{-\lambda_1(x-\mu)-\lambda_2(x^2-(\mu^2+\sigma^2))}dx$$

şeklinde yazılır. Her iki tarafın logaritması alındığında,

$$\lambda_3 = \ln\Big(\int_a^b e^{-\lambda_1(x-\mu)-\lambda_2(x^2-(\mu^2+\sigma^2))}dx\Big) \tag{4.19}$$

bulunur.

Ortalamanın ve standart sapmanın ortak sonsal olasılık fonksiyonu (4.10) eşitliğinden,

$$p(\mu,\sigma|x) \propto p(x|\mu,\sigma)p(\mu,\sigma)$$

olarak ifade edilebilir. Buradan hareketle çarpım kuralı uygulanırsa ve μ ile σ' nın bağımsız olması kullanılırsa parametrelerin sonsal olasılık yoğunluk fonksiyonu,

$$p(\mu,\sigma|x) \propto p(x|\mu,\sigma)p(\mu)p(\sigma) \tag{4.20}$$

yazılır.

Şimdi μ ve σ' nın önsel olasılık fonksiyonlarını oluşturup sonsal olasılığı elde edelim. Öncelikle μ' nün önsel olasılık yoğunluk fonksiyonu (μ_A, μ_U) aralığında düzgün dağılım olarak alınırsa;

$$p(\mu) = \frac{1}{\mu_U - \mu_A} \tag{4.21}$$

yazılır. Burada μ_U: μ' nün üst sınırı, μ_A ise μ' nün alt sınırıdır.
Standart sapmanın önsel olasılık yoğunluk fonksiyonu Jeffrey'nin önseli alınırsa,

$$p(\sigma) = \begin{cases} \frac{K}{\sigma}, & \sigma_A \leq \sigma \leq \sigma_U \\ 0, & \text{diğer durumlar} \end{cases}$$

olmak üzere,

$$\int_{\sigma_A}^{\sigma_U} p(\sigma) d\sigma = 1$$

$$\int_{\sigma_A}^{\sigma_U} \frac{K}{\sigma} d\sigma = 1$$

Gerekli integral işlemi yapıldığında K normalizasyon sabiti,

$$K = \frac{1}{\ln(\frac{\sigma_U}{\sigma_A})}$$

bulunur. O halde,

$$p(\sigma) = \frac{1}{\sigma \ln(\frac{\sigma_U}{\sigma_A})} \tag{4.22}$$

elde edilir. (4.18), (4.21) ve (4.22) eşitlikleri, denklem (4.20) de yerine yazılırsa sonsal olasılık yoğunluk fonksiyonu:

$$p(\mu, \sigma|x) \propto e^{-\lambda_3 - \lambda_1(x-\mu) - \lambda_2(x^2 - (\mu^2 + \sigma^2))} . \frac{1}{\mu_U - \mu_A} . \frac{1}{\sigma \ln(\frac{\sigma_U}{\sigma_A})} \tag{4.23}$$

elde edilir.

4.1.3. Uygulama

Bir önceki bölümde teoriksel olarak oluşturulan olasılık fonksiyonlarını gerçek dünya problemlerinden süt verilerine uygulayalım.

Tablo 4.1: Günlük Süt Üretimi.

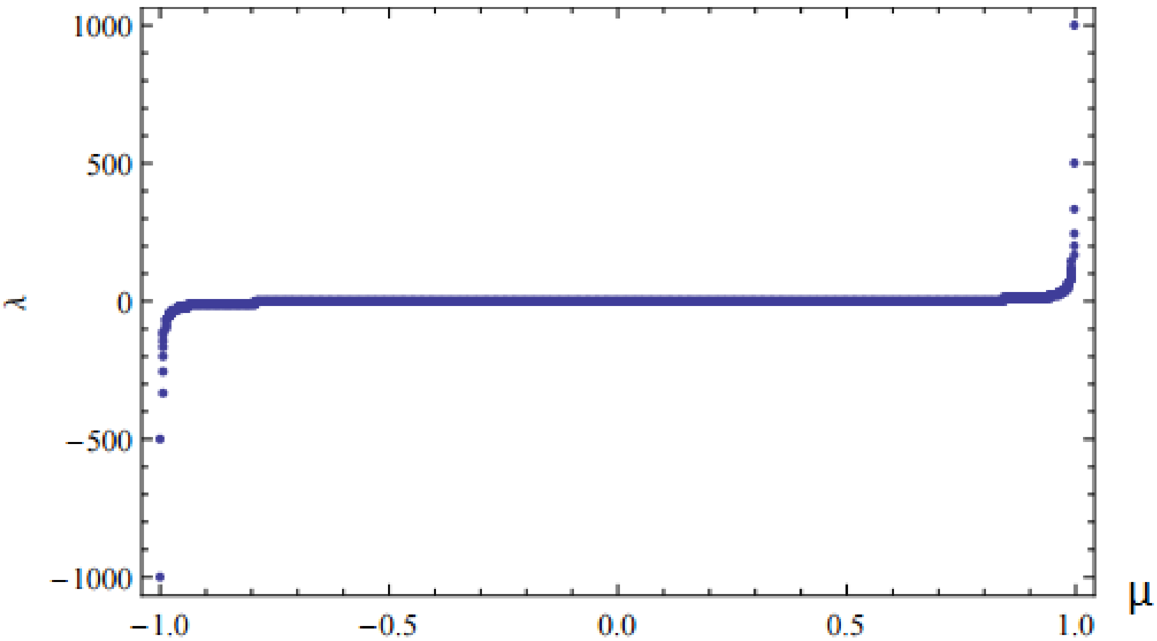
İnekler	Süt Miktarı (Kg/Gün)
1	33.1
2	34.7
3	29.2
4	25.5
5	27.8
6	39.9
7	50.3

İneklerin süt verimleri; ineğin cinsi, yaşı, kondisyonu, yediği yem, laktasyon gün sayısı, barındığı ortam, iklim şartları gibi birçok faktöre bağlı olarak değişmektedir. Sütaş Genel Müdürlüğünden alınan bilgilere dayanılarak, analizi yapılacak olan süt üretimi verilerinin, alt sınırı 0, üst sınırı 70 kilogram olduğu tespit edilmiştir. Çalışma, çiftlikten rastgele seçilen 7 inek üzerinde gerçekleştirilmiştir. Seçilen 7 ineğin günlük ürettiği süt miktarları yukarıda Tablo 4.1'de gösterilmiştir.

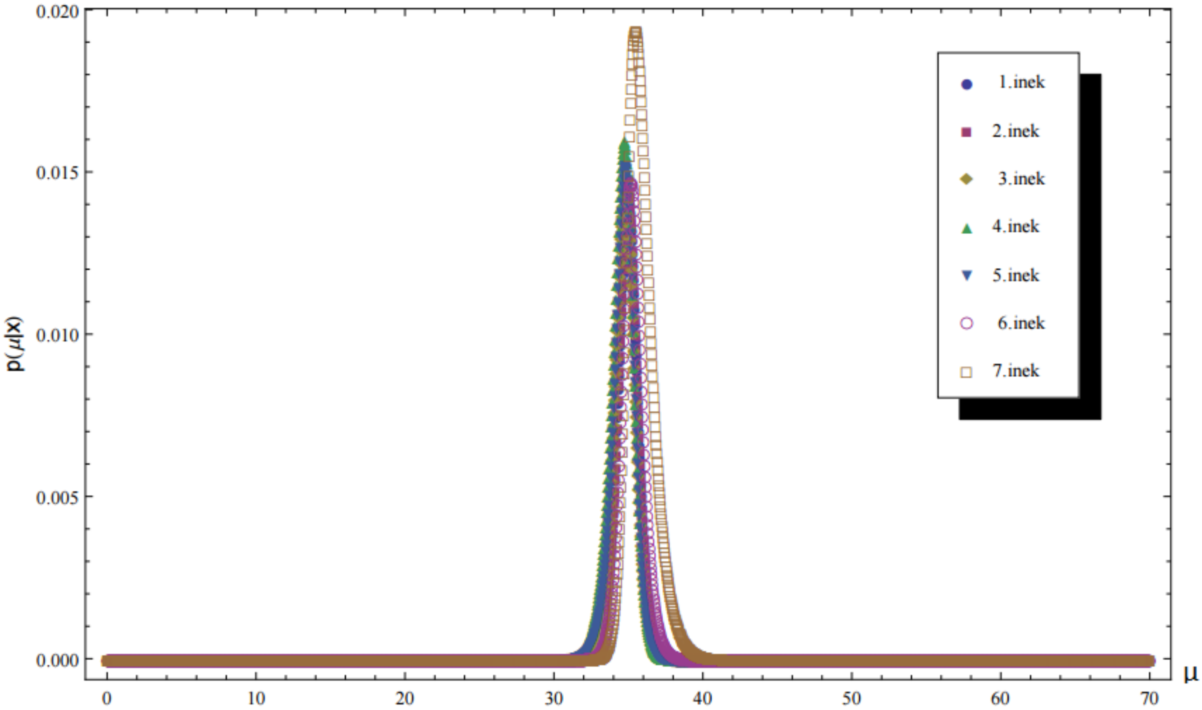
Çalışmanın amacı, ineklerin süt verimini arttırmak için yüksek süt üretimini sağlayacak yem cinsinin tespit edilmesidir. Yonca, silaj, yapılandırılmış yem kompozitlerinden hangisinin seçileceği parametrik olmayan dağılımlarla tespit edilecektir. Bunun için 7 ineğin ürettiği süt verimleri ele alınarak bu çiftlikteki ineklerin günlük ürettikleri süt miktarının ortalamasının dağılımı tahmin edilecektir.

Sütlerin ortalamasının dağılımında, Bulgular kısmında teoriksel olarak verilen uygun kısıtlar altında Lagrange çarpanı ile oluşturulan önsel, sonsal olasılık ve olabilirlik fonksiyonu baz alınmıştır. Entropi fonksiyonu Bayesli mantıksal çıkarım çerçevesinde incelenmiştir. Bu sürecin daha basit olması için yukarıdaki değerler normalize edilerek $[-1, 1]$ aralığında çözülmüştür. Bu durum herhangi bir gerçek problem söz konusu olduğunda sonsal dağılımın, olabilirlik fonksiyonu ile ilişkisini açıklar ve λ' yı belirlemek için genel sınırların ölçekli olarak çözülmesini sağlar.

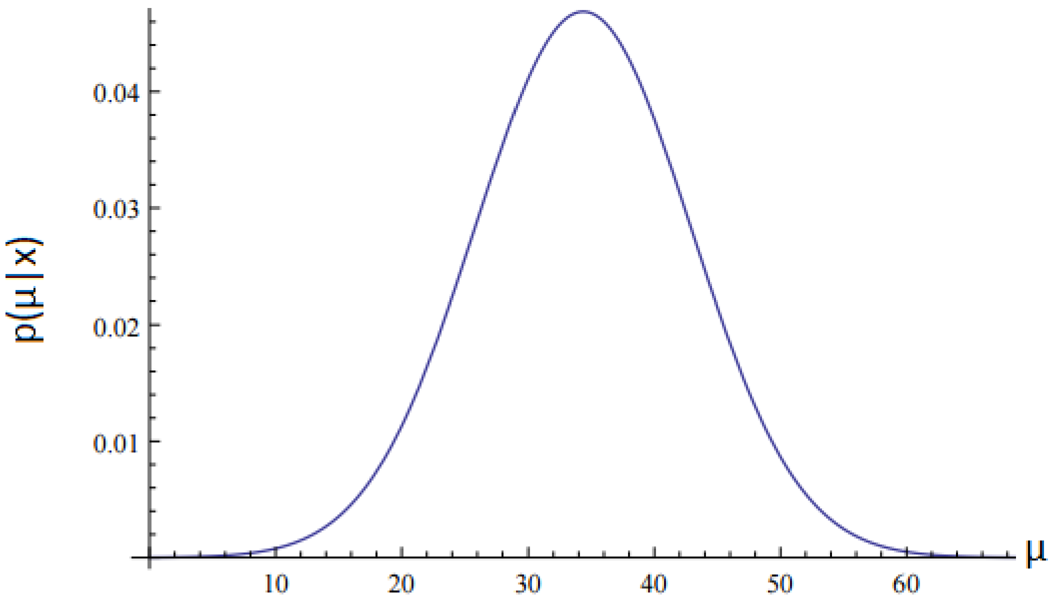
Ortalama fonksiyonu olan μ, $[-1, 1]$ aralığında normalize edilerek, λ için nümerik çözümü, Şekil 4.1'de gösterilmiştir. Diğer yandan olabilirlik fonksiyonunun hesaplanmasından sonra, ortalama için önsel olasılık olarak düzgün dağılım ele alınıp (4.9) eşitliğinde yerine yazıldığında, 7 ineğin her birisinin günlük ortalama süt üretimine ait sonsal dağılımın sonuçları Şekil 4.2'de verilmiştir. Bilgisayar simülasyon sonuçları Mathematica 9.1 yazılım programı kullanılarak elde edilmiştir.



Şekil 4.1: Normalize Edilmiş μ .



Şekil 4.2: Günlük Ortalama Süt Miktarı.



Şekil 4.3: Süt verisinin ortalamasının dağılımı.

Şekil 4.3'de görüldüğü gibi incelenen 7 süt verisinin ortalama değeri 34.3 kilogram olup, çiftliğe ait günlük ortalama süt üretiminin $\%95$ güven seviyesinde, oluşturulan güven aralığı $[26.7, 41.9]$ olarak bulunur.

4.2. BAYESLİ MAKSİMUM ENTROPİ DAĞILIMININ HESABI

Maksimum entropi ilkesi, Giriş Bölümünde bahsedilen ters problemleri çözmek için güçlü bir yöntemdir. Maksimum entropi dağılımlarının pratik olarak tahmini yapılırken entropi fonksiyonunun doğrusal olmamasından ve çok sayıda parametreden dolayı bir takım sayısal hesaplamalar gerektirmektedir. Diğer yandan, her zaman kesikli dağılımdan, sürekli dağılıma geçerken MaxEnt tahmini doğru sonucu vermeyebilir. Bu gibi durumlarda problemler, Bayesli mantıksal çıkarım çerçevesinde çözülebilir.

Entropi teoreminin ve maksimum entropi metodunun en iyi örneklerinden birisi Jaynes' ın zar deneyidir. Burada zarın birçok kez atıldığı ve ortalama olarak hangi sayının geldiği biliniyorsa, her bir yüzün ne kadar sıklıkla geldiği incelenecektir. Bu problemde zarın yüzlerinin gelme sayılarının tahmininde, altı parametre ve iki kısıt kullanılacaktır [32].

Eğer hilesiz bir zar N kez atılırsa ve ortalamasının 4.5 olduğu gözlemlenirse verilen kısıtlar altında zarın her bir yüzünün kaç defa geldiği tahmin edilecek ve entropi değeri hesaplanacaktır. Bu iki kısıt aşağıdaki gibi tanımlansın.

$\forall i$ için n_i tam sayı olmak üzere, $0 \leq n_i \leq N$ öyle ki

$$\sum_i n_i = N \tag{4.24}$$

ve

$$\sum_i i.n_i = (4.5).N \tag{4.25}$$

olsun.

Denklem (4.24) ve (4.25) de verilen kısıtlara uygun olabilecek $\{n_i\}$ dizileri Tablo 4.2'de verilmiştir.

Tablo 4.2: Zar olasılık deneyinin sayıları.

	FREKANSLAR					
Durum Sayısı	n_1	n_2	n_3	n_4	n_5	n_6
1	26	88	134	185	246	321
2	34	85	130	178	244	329
3	54	79	114	166	240	347
4	29	94	128	177	241	331
5	39	88	123	172	240	338
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Burada, n_i: i. noktanın gelme sayısıdır.

Öncelikle (4.24) eşitliğinin Tablo 4.2' deki her bir n_i için sağlandığını gösterelim:

$$26 + 88 + 134 + 185 + 246 + 321 = 1000$$

$$34 + 85 + 130 + 178 + 244 + 329 = 1000$$

$$54 + 79 + 114 + 166 + 240 + 347 = 1000$$

$$29 + 94 + 128 + 177 + 241 + 331 = 1000$$

$$39 + 88 + 123 + 172 + 240 + 338 = 1000$$

Yani N değerinin 1000, dolayısıyla $\sum_i n_i = N = 1000$ olduğu görülür.

Yine Tablo 4.2' deki değerlerin (4.25) eşitliğini sağladığını gösterelim:

$$(26 \times 1) + (88 \times 2) + (134 \times 3) + (185 \times 4) + (246 \times 5) + (321 \times 6) = 4500$$

$$(34 \times 1) + (85 \times 2) + (130 \times 3) + (178 \times 4) + (244 \times 5) + (329 \times 6) = 4500$$

$$(54 \times 1) + (79 \times 2) + (114 \times 3) + (166 \times 4) + (240 \times 5) + (347 \times 6) = 4500$$

$$(29 \times 1) + (94 \times 2) + (128 \times 3) + (177 \times 4) + (241 \times 5) + (331 \times 6) = 4500$$

$$(39 \times 1) + (88 \times 2) + (123 \times 3) + (172 \times 4) + (240 \times 5) + (338 \times 6) = 4500$$

Yani $\sum_i i.n_i$ değerinin 4500, dolayısıyla $\sum_i i.n_i = 4500 = (4.5) \times 1000 = (4.5) \times N$ olduğu görülür.

Tablo 4.2'den, her bir durumun, zarın N kez atılışında olası meydana gelecek farklı sonuçlarının sayısı,

$$W = \frac{N!}{n_1!n_2!...n_6!} \tag{4.26}$$

şeklinde bulunur.
Örneğin, Tablo 4.2' deki 1. model için gelebilecek yüzlerin sırasıyla sayısı,

$$W_1 = \frac{1000!}{26!88!...331!}$$

olur. Bu ifadenin nümerik olarak hesaplanması zordur. Bunun için Bölüm 2.3, Tanım 2.10 da verilen Stirling yaklaşımı kullanılırsa, (4.26) eşitliğinin her iki tarafından logaritması alınıp, f fonksiyonu $\sum f_i = 1$ olmak üzere, $f_i = \frac{n_i}{N}$ yani $n_i = f_i.N$ şeklinde tanımlanıp yerine yazılırsa (2.68) eşitliği elde edilir. Buradan $N \to \infty$ limit alınırsa

$$\frac{1}{N} \ln W \to -\sum_{i=1}^{6} f_i \ln f_i = -f' \ln f$$

Burada, $f' = (f_1, f_2, ..., f_6)$ ve $\ln f = (\ln f_1, \ln f_2, ..., \ln f_6)$ dır.

Şimdi maksimize edilecek fonksiyon vektörel olarak ifade edilirse,

$$S(f) = -f' \log f \tag{4.27}$$

yazılır. Bu fonksiyonu (4.24) ve (4.25) denklemlerini göz önüne alarak,

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{pmatrix}$$

ve $d = (1, 4.5)'$ olmak üzere,

$$Af = d \tag{4.28}$$

sınırlaması altında maksimize etmeye çalışacağız. Bunun için Lagrange çarpanları $\lambda = (\lambda_1, \lambda_2)'$ olmak üzere, Lagrange fonksiyonu,

$$L(\lambda) = f' \log f + \lambda'(Af - d) \tag{4.29}$$

olsun. Yani,

$$\max_{f_i}[-\sum_{i=1}^{6} f_i \log f_i - \lambda_1(\sum_{i=1}^{6} f_i - 1) - \lambda_2 \sum_{i=1}^{6}(f_i.i - 4.5)]$$

sağlayacak f_i' ler bulunacaktır. Bunun için Denklem (4.29) un f_i' ye göre türev alınıp sıfıra eşitlendiğinde,

$$\frac{\partial L}{\partial f_i} = \sum_{i=1}^{6}(\log f_i + 1 + \lambda_1 + \lambda_2.i) = 0$$

elde edilir. Buradan,

$$\log f_i = -1 - \lambda_1 - \lambda_2 i$$

ve logaritma tanımından,

$$f_i = e^{-1-\lambda_1-\lambda_2 i}, \quad (i = 1, 2, 3, 4, 5, 6) \tag{4.30}$$

bulunur. $1 + \lambda_1 = \lambda_0$ denilirse,

$$f_i = e^{-\lambda_0-\lambda_2 i}$$

elde edilir. Yani,

$$f_1 = e^{-(\lambda_0+\lambda_2)}$$

$$f_2 = e^{-(\lambda_0+2\lambda_2)}$$

$$\vdots$$

$$f_6 = e^{-(\lambda_0+6\lambda_2)}$$

yazılır. Buradan,

$$(f_1, f_2, ..., f_6)' = e^{-\begin{pmatrix} \lambda_0+\lambda_2 \\ \lambda_0+2\lambda_2 \\ \lambda_0+3\lambda_2 \\ \lambda_0+4\lambda_2 \\ \lambda_0+5\lambda_2 \\ \lambda_0+6\lambda_2 \end{pmatrix}} = e^{-\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \end{pmatrix}\begin{pmatrix} \lambda_0 \\ \lambda_2 \end{pmatrix}}$$

$$= e^{-\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{pmatrix}'\begin{pmatrix} \lambda_0 & \lambda_2 \end{pmatrix}'}$$

elde edilir. $\begin{pmatrix} \lambda_0 & \lambda_2 \end{pmatrix}' = \tilde{\lambda}$ denilirse,

$$f = e^{-A'\tilde{\lambda}} \tag{4.31}$$

çözümü bulunur.

(4.31) denkleminde bulunan f değeri (4.28) eşitliğinde yerine yazıldığında,

$$Ae^{-A'\tilde{\lambda}} = d \tag{4.32}$$

bulunur.

Lineer sistemler, lineer olmayan sistemlerden genellikle daha kolay çözülür. Bunun için, (4.32) eşitliği lineerleştirilmeye çalışılır. Bunun için Taylor serisi açılımı veya Newton yöntemi kullanılır. Denklem (4.32) de gerekli işlemler yapılırsa, lineerleştirilecek fonksiyon:

$$f(\tilde{\lambda}) = Ae^{-A'\tilde{\lambda}} - d = 0$$

olur. Bu denklemin n. iterasyonda bulunan kökü $\tilde{\lambda}_n$ olmak üzere, $(n+1)$. iterasyonda elde edilen kökü $\tilde{\lambda}_{n+1}$, Newton yöntemiyle,

$$\tilde{\lambda}_{n+1} = \tilde{\lambda}_n - \frac{f(\tilde{\lambda}_n)}{f'(\tilde{\lambda}_n)}$$
$$= \tilde{\lambda}_n + \frac{(Ae^{-A'\tilde{\lambda}_n} - d)}{A(Ae^{-A'\tilde{\lambda}_n})'}$$

elde edilir. Buradan,

$$(Ae^{-A'\tilde{\lambda}_n} - d) = (\tilde{\lambda}_{n+1} - \tilde{\lambda}_n)'A(Ae^{-A'\tilde{\lambda}_n})'$$

ve

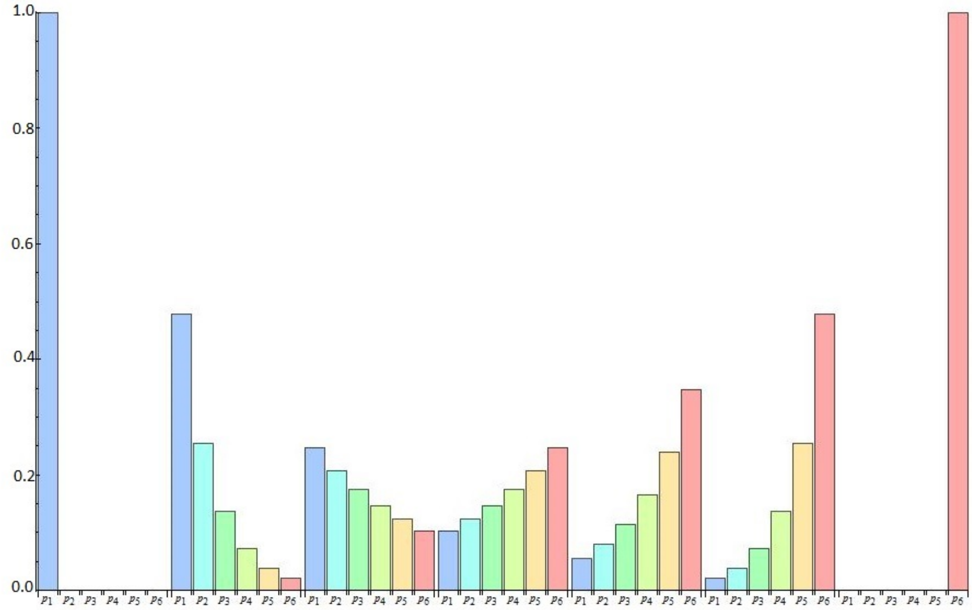
$$Ae^{-A'\tilde{\lambda}_n} - (\tilde{\lambda}_{n+1} - \tilde{\lambda}_n)'A(Ae^{-A'\tilde{\lambda}_n})' = d \tag{4.33}$$

eşitliği bulunur.

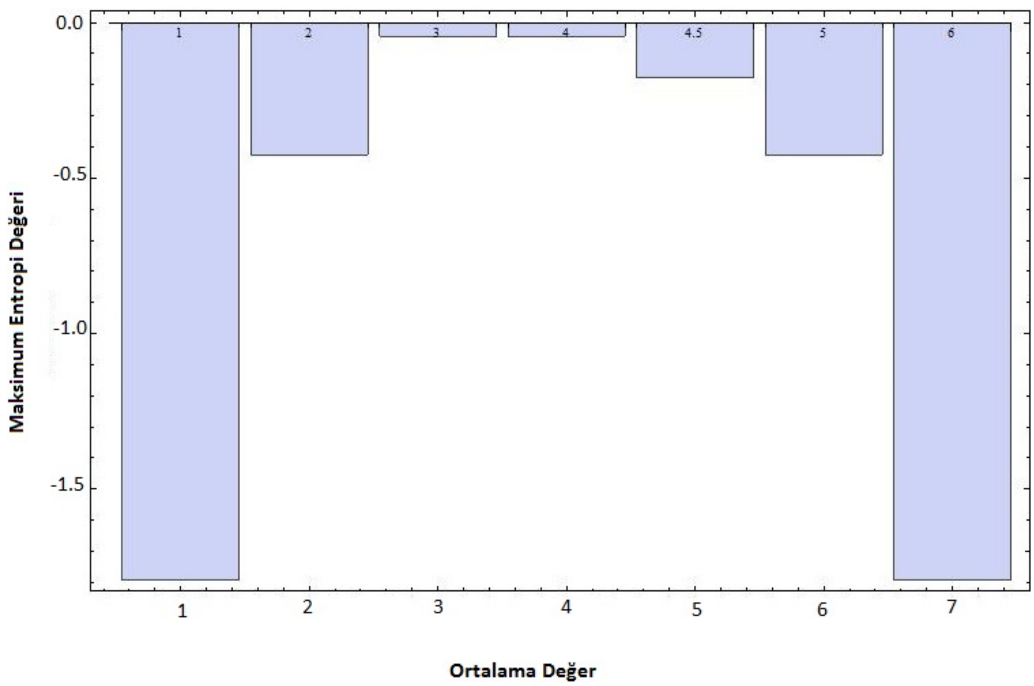
Bu denklemin çözümü için Mathematica da oluşturulan kod kullanılırsa denklemin maksimum entropi çözümü yani zarın her bir yüzünün ve ortalama değeri 4.5' un gelme olasılıkları ve maksimum entropi değerleri Tablo 4.3'de verilir. Bu tabloya göre ortalama değerin 4.5 gelmesi halinde 1 numaralı yüzün gelme olasılığı 0.054, 2 numaralı yüzün gelme olasılığı 0.078, 3 numaralı yüzün gelme olasılığı 0.114, 4 numaralı yüzün gelme olasılığı 0.165, 5 numaralı yüzün gelme olasılığı 0.239, 6 numaralı yüzün gelme olasılığı ise 0.347 dır. Bu değerler 1000 ile çarpıldığında 1 numaralı yüz yaklaşık 54 defa, 2 numaralı yüz yaklaşık 78 defa, 3 numaralı yüz yaklaşık 114 defa, 4 numaralı yüz yaklaşık 165 defa, 5 numaralı yüz yaklaşık 239 defa, 6 numaralı yüzün ise yaklaşık olarak 347 defa geldiği görülür. Ayrıca ortalama değerin 4.5 gelmesi halinde entropinin maksimum değeri ise $H_{maks} = 1.613$ bulunur.

Tablo 4.3: Zar Olasılıkları.

Ortalama(μ)	p_1	p_2	p_3	p_4	p_5	p_6	$H(p)$
1	0.999	7.673×10^{-13}	5.936×10^{-13}	5.936×10^{-13}	5.933×10^{-13}	5.931×10^{-13}	9.0363×10^{-11}
2	0.478	0.254	0.135	0.072	0.038	0.020	1.367
3	0.246	0.207	0.174	0.146	0.122	0.103	1.748
4	0.103	0.122	0.146	0.174	0.207	0.246	1.748
4.5	0.054	0.078	0.114	0.165	0.239	0.347	1.613
5	0.020	0.038	0.072	0.135	0.254	0.478	1.367
6	5.863×10^{-13}	5.863×10^{-13}	5.863×10^{-13}	5.863×10^{-13}	5.863×10^{-13}	0.999	8.257×10^{-11}

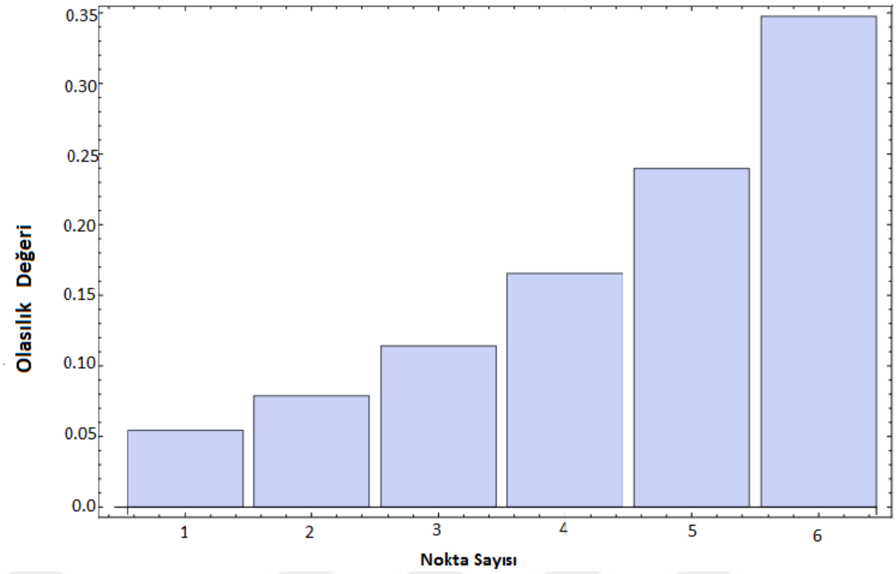


Şekil 4.4: Zar olasılık deneyi.



Şekil 4.5: Her bir zarın maksimum entropi değeri.

Tablo 4.3' deki her bir ortalamaya göre yüzlerin gelme olasılıkları Şekil 4.4'de, maksimum entropi değerleri ise Şekil 4.5'de gösterilmiştir.



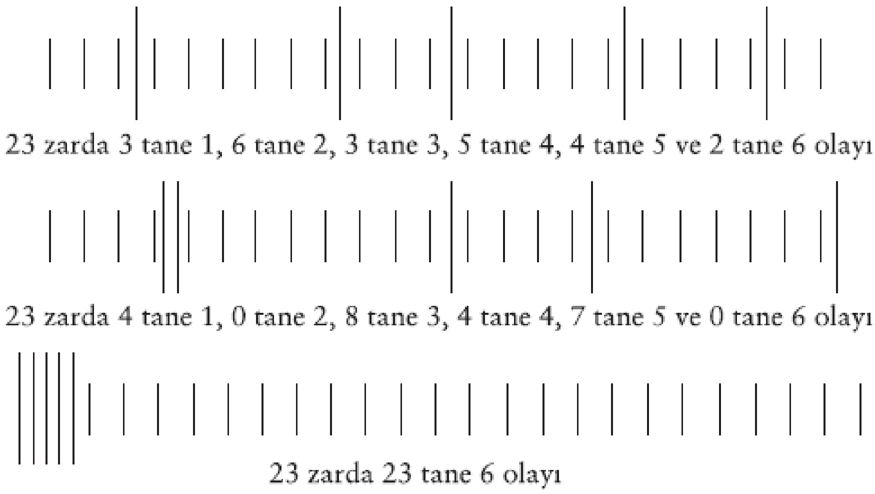
Şekil 4.6: Her bir zarın olasılık değeri şekli.

Ortalama değerin $\mu = 4.5$ olması durumundaki zarın her bir yüzünün gelme olasılıkları Şekil 4.6'da gösterilmiştir. Grafiğe göre en fazla 6, en az ise 1'in geldiği görülmektedir.

$\tilde{\lambda}$' yı hesaplarken Newton yöntemi kullanılmıştı. Fakat bu yöntem global optimizasyonu garanti etmez. Çünkü bazı yerel optimumlarda takılıp kalabilir. Bu durumda istenen verilere en uygun modelin seçilmesine yani hedef veya uygunluk fonksiyonunu optimize edecek en iyi modelin bulunmasına gereksinim duyulabilir. Bu durumda Maksimum Entropi dağılımının Bayesli yaklaşımla model seçiminin nasıl yapılabileceği verilecektir. Aşağıda, Jaynes' ın zar problemi düşünülerek; model seçimi probleminde, Maksimum Entropinin Bayesli yaklaşımla olan ilişkisi gösterilecektir.

Oluşturulacak modeller, zarın $n_1, n_2, ..., n_6$ yüzlerinin frekansları ile gözlenen belirli bir durumu gerçekleştirmeye karşılık gelir. Bir ikili dizi olarak $6 \times k$' lık bitler alınarak her model kod edilecektir. Buradaki k, N' i göstermek için gerekli olan bitlerin sayısını temsil eder. Yani, zarın kaç kez atıldığı belirtilecektir. Tek uygun modeller oluşturmak için uğraşılmaktadır. Çünkü kısıtsız model uzayının N^6 üyesi vardır ve $\{n_i\}$ ile $\sum_i n_i = N$ şeklinde tüm dağılımları içeren alt uzay, $\binom{N+5}{N}$ modellerinden oluşur. Model sayısının neden $\binom{N+5}{N}$ olduğu aşağıda gösterilmiştir:

Aşağıdaki şekilde, yan yana dizilmiş N tane ($N = 23$ olmak üzere) küçük çubuk ve bu N küçük çubuk arasına ya da çubukların en başına veya en sonuna 5 tane büyük çubuk yerleştirilmiştir.



Şekil 4.7: Zar model sayısının temsili gösterimi.

Yukarıdaki şekilde,

- En soldan başlayarak, 1. büyük çubuğun solunda kalan küçük çubuk sayısı N zarda kaç tane 1 geldiği,
- 1. büyük çubukla 2. büyük çubuk arasında kalan küçük çubuk sayısı da N zarda kaç tane 2 geldiği,
- 2. büyük çubukla 3. büyük çubuk arasında kalan küçük çubuk sayısı da N zarda kaç tane 3 geldiği,
- Benzer şekilde devam edildiğinde, 5. büyük çubukla 6. büyük çubuk arasında kalan küçük çubuk sayısı da N zarda kaç tane 6 geldiği belirtilmiştir. [33]. Buna göre,

N küçük çubuğun 5 büyük çubukla ayrımı N zarda bir olay verir ve N zarda gelebilecek her olay N küçük çubuğun 5 büyük çubukla bir ayrımını verir.

O halde, N tane küçük çubukla 5 tane büyük çubuğu kaç değişik şekilde dizileceği bulunmalıdır. Küçük çubuklara A, büyük çubuklara B denilirse, N tane A ve 5 tane B ile $N + 5$ harflik kaç değişik kelime yazılabileceği bulunmalıdır. Harfler için $N + 5$ tane yer var ve bunlardan 5 tanesi B harfine ayrılacaktır. Böyle bir seçim, $\binom{N+5}{5}$ şeklinde yapılır. O halde kombinasyon formülünden yani, $C(n, r) = \binom{n}{r} = \binom{n}{n-r}$ olduğundan,

$$\binom{N+5}{5} = \binom{N+5}{N+5-5} = \binom{N+5}{N} \tag{4.34}$$

yazılır. Oluşturulan bu modeller için eldeki veriye en uygun modelin tespiti için Bayes Teoremi kullanılırsa,

$$p(f|d) = \frac{p(f)p(d|f)}{p(d)} \tag{4.35}$$

yazılır. Buradan,

$$p(f|d) \propto p(d|f)p(f)$$

olur.

Jaynes' ın entropi teoremine göre, $p(f)$ önsel olasılığı, yani entropi olasılığı,

$$p(f) \propto e^{-f' \log f} \tag{4.36}$$

$p(d|f)$ olabilirlik fonksiyonu ise

$$p(d|f) \propto e^{-|\sum\limits_i i.f_i - 4.5|^n} e^{-|\sum\limits_i f_i - 1|^n} \tag{4.37}$$

alınıp (4.35) denkleminde yerine yazıldığında sonsal olasılık,

$$p(f|d) \propto e^{-f' \log f - |\sum\limits_i i.f_i - 4.5|^n - |\sum\limits_i f_i - 1|^n}$$

elde edilir. Her iki tarafın logaritması alındığında logaritmik sonsal olasılık fonksiyonu,

$$\log(p(f|d)) \propto -f' \log f - |\sum_i i.f_i - 4.5|^n - |\sum_i f_i - 1|^n \tag{4.38}$$

elde edilir.

5. TARTIŞMA VE SONUÇ

Bu tez çalışmasında, Entropinin Bayesli Mantıksal Çıkarım çerçevesinde maksimize edilmesine odaklanıldı. Bunun için öncelikle olasılıktaki temel kavramlar ile Bayesli yaklaşım hakkında bilgi verilerek, tezin uygulama kısmında yararlanılacak önsel olasılıklar tanıtıldı. Daha sonra entropi kavramı ele alınarak, önemli entropi ölçütlerine yer verildi. Tezin Bulgular bölümünde ilk olarak Sütaş Genel Müdürlüğünden ineklerin günde ürettikleri süt verileri alındı. Bu süt verilerinin parametrik olmayan olasılık dağılımları için ortalama ve varyansın dağılımı, Bayesli Analizi ile Maksimum Entropi çerçevesinde incelendi. Dağılımın ortalaması için eldeki verilerden 7 inek rastgele seçildi. Verilen kısıtlar altında önsel, sonsal olasılık ve olabilirlik fonksiyonu oluşturuldu. Entropi fonksiyonu, Bayesli Mantıksal Çıkarım çerçevesinde maksimize edildi. Bulunan sonuçlar ayrı ayrı grafikler halinde gösterildi. Ortalama ve standart sapma için ortak olasılık dağılım incelendiğinde ise benzer şekilde verilen kısıtlar altında önsel, sonsal olasılık ve olabilirlik fonksiyonu oluşturuldu. Maksimum Entropi, Bayesli açıdan teorik olarak sunuldu.

İstatistikte, Merkezi Limit Teoremi gereği $n \geq 30$ olduğu zaman örneklem ortalamasının yaklaşık olarak normal dağılıma yakınsadığı bilinmektedir. Elimizdeki inek verileri az sayıda olduğu için hangi dağılıma sahip olduğunu bilmek zordur. Yapılan çalışmada, Bayesli maksimum entropi parametre dağılımları, verilerin gerçek dağılımına uygun en yakın tahmini dağılımlar ürettiği görüldü. Öyle ki $n = 7$ olmasına rağmen Bayesli maksimum entropi aracılığıyla süt verileri dağılımının normal dağılıma yaklaştığı gösterildi.

Tez çalışmasında Jaynes' ın zar deneyi örneği de ele alındı. Burada, eğer bir zarı 1000 defa attığımızda gelen noktaların ortalamasının 4.5 olduğunu biliyorsak, zar üzerindeki her bir noktanın frekanslarının ne olabileceği problemi araştırıldı. Bunun için Entropi fonksiyonu oluşturularak bazı sınırlandırmalar altında Lagrange çarpanı kullanıldı ve entropi maksimize edildi. Her bir zarın olasılıkları hesaplandığında gerçek verilere en yakın verinin Tablo 4.3'den 5. veri modelinin olduğu görülüp buna da entropi fonksiyonu sayesinde ulaşıldı. Ayrıca Mathematica programı kullanılarak zarın her bir yüzünün maksimum entropi ve olasılık değerleri gösterildi. Daha sonra ise model seçimi probleminin çözümü, Bayesli Mantıksal Çıkarım çerçevesinde ele alınarak önsel, sonsal olasılık ve olabilirlik fonksiyonu teoriksel olarak oluşturuldu. Sonuç olarak, sadece zarın ortalaması bilinerek zarın her bir yüzünün ne kadar geldiğini bilmek zordur ancak bu, maksimum entropi yardımıyla bulundu. Fakat bu işlemin bazı dezavantajları olduğu aşağıda maddeler halinde verilmiştir:

- Her zaman maksimum entropi kesikli dağılım için tam olarak sürekli dağılıma yakınsayamayabilir.
- $\tilde{\lambda}$ gerçek değeri her zaman bulunmayabilir. $\tilde{\lambda}$ değerini hesaplarken Newton yöntemi kullanıldığında başlangıç değerlerine yakın değerler verilirse sonuca ulaşılabilir ama uzak değerler verilirse sonuca ulaşılamayabilir. Yani, Newton yöntemi ile yerel maksimumdan küresel maksimuma geçilemeyebilir.
- Zar probleminde gürültü ihmal edildi. Fakat gerçek dünya problemlerinin çoğunda gürültü vardır. Bu durumda ise yapılması gereken gerçek verilerle uyumlu taklidi verilerin üretilmesi söz konusu olacaktır.

Yukarıdaki problemlerin çözümü için yapılan araştırmalarda genetik algoritmalar gibi bazı algoritmaların kullanılabileceği sonucuna varıldı.

Bu tez çalışmasının sonucunda, yapılan her iki uygulamada da elde edilen sonuçlar doğrultusunda, maksimum entropi ilkesinin Bayesli Mantıksal Çıkarım çerçevesinde kullanılmasıyla bütün durumlar açısından çok iyi sonuçlar verdiği ve eldeki verilerin modellenmesinde etkili olduğu görüldü.

KAYNAKLAR

[1]. Guillen, M., 2001, *Dünyayı Değiştiren Beş Denklem* (çev.: Gürsel Tanrıöver), Ankara, Kılıçaslan Matbaacılık, TÜBİTAK Popüler Bilim Kitapları Dizisi, ISBN:975-403-206-8.

[2]. Fang, S.C., J.R. Rajasekera ve H.S.J. Tsao, 1997, *Entropy Optimization and Mathematical Programming*, ABD, Kluwer Academic Publishers, ISBN: 978-1-4613-7810-5 .

[3]. Cevri, M., 2011, *Sinyal İşlemede Bayes Tabanlı Yöntemler*, Titiz Yayınevi, ISBN: 9787-605-61770-2-6.

[4]. Cevri, M. and Ustundag D., 2014, *Performance Evaluation of Gibbs Samplingfor Bayesian Extracting Sinusolds*, Computational Problems in Engineering, In: Mastorakis, N. and Mladenov, V. (ed.), Springer V., ISBN: 978-3-319-03966-4.

[5]. O'Hagan, A., 1988, *Probability: Methods and Measurement*, Chapman and Hall, London, ISBN:978-94-010-7038-6.

[6]. Akdeniz, F., 2012, *Olasılık ve İstatistik*, Nobel Kitapevi, ISBN:978-605-397-135-1.

[7]. Cerit, C., Yuksel, M., 2011, *Olasılık Problemleri*, İstanbul.

[8]. Gürsakal, N., 1992, *Bayesgil İstatistik*, Bursa: Uludağ Üniversitesi Basımevi.

[9]. Cevri, M., Ustundag D., Gökdemir Y. and Karadağ, B., 2016, Pedigree Analysis with Bayesian Logical Inference, *Selçuk Journal of Applied Mathematics*,(Baskıda).

[10]. Cevri M. and Ustundag D., 2012, Bayesian recovery of sinusoids from noisy data with parallel tempering, *IET Signal Processing*, 6(7), 673-683.

[11]. Cevri M. and Ustundag D., 2016, Prediction the probabilities of transmission of genetic traits within Bayesian logical inference, *Acta Physica Polonica*, (Baskıda).

[12]. Gregory, P., 2005, *Bayesian Logical Data Analysis for the Physical Sciences*, Cambridge University Press, ISBN: 0 521 841 50.

[13]. Cevri, M. and Ustundag, D., 2012, *Bayesian Recovery of Sinusoids with Simulated Annealing*, Simulated Annealing-Advances, Applications and Hybridizations, In: Guerra Tsuzuki, M. S., (ed.), Intech, ISBN: 978-953-51-0710-1, 68-90.

[14]. Ekici, O., 2009, İstatistikte Bayesyen ve Klasik Yaklaşımın Kavramsal Farklılıkları, *Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 12, 89-101

[15]. Jordan, M., 2010, *Jeffreys Priors and Reference Priors*, http://www.cs.berkeley.edu/∼ jordan/course/260-spring10/lectures/lecture7.pdf.

[16]. Özkul, S., 2001, Su Kalitesi Gözlem Ağlarının Entropi Yöntemi ile Değerlendirilmesi, *Turkish Journal Engineering and Environmental Sciences*, 25(5): 435-452.

[17]. Orkan, A. L., 1992, *Bilişim Teorisi*, Marmara Üniversitesi İletişim Fakültesi Yayınları, İstanbul.

[18]. Kapur, J.N., Kesavan, H.K., 1992, *Entropy Optimization Principles With Applications*, Academic Press, New York, USA, ISBN:978-0123976703.

[19]. Ross, S., 1998, *A First Course in Probability*, 8^{th} ed., Prentice Hall, ABD, ISBN:978-0-13-603313-4.

[20]. Erkip, F., 2009, *İnformation Theory Notes*, http://www.cims.nyu.edu/∼ chou/notes/infotheory.pdf.

[21]. Karmeshu ve Pal, N.R., 2003, *Entropy Measures, Maximum Entropy Principle and Emerging Applications*, Springer, New York, USA, ISBN:978-3-540-36212-8.

[22]. Cover, T.M, ve Thomas, J.A, 1991, *Elements of İnformation Theory*, 2^{nd} ed., A Wiley-Interscience Publication, New York, USA, ISBN:978-0-471-24195-9.

[23]. Marsh, C., 2013, Introduction to Continuous Entropy, *Deparment of Computer Science Princeton University*, 1-17.

[24]. Devroye, N., 2009, *Chapter* 8*: Differential Entropy*, http://www1.ece.uic.edu/∼ devroye/course/ECE534/lectures/ch8.pdf.

[25]. Jaynes, E. T., 1957, *İnformation Theory and Statistical Mechanics*, The Physical Review, Vol. 106, No, 4, 620-630.

[26]. Howitt, R. E., and Paris, Q., 1998, Analysis of ill-posed production problems using maximum entropy, *American Journal of Agricultural Economics*, No.1, 124-128

[27]. Ben-Naim, A., 2008, *A Farewell to Entropy: Statistical Thermodynamics Based on Information*, World Scientific Publishing, ISBN:978-981-270-707-9.

[28]. Jaynes, E.T., 2002, *Probability Theory: The Logic of Science*, http://omega.albany.edu:8008/JaynesBook.html.

[29]. Ozdeger, A., 2008, *Çözümlü Analiz Problemleri*.

[30]. Price, W.J., Price, H.J., and Shafii, B., 2002, Using Bayesian Analysis and Maximum Entropy To Develop Non-Parametric Probability Distributions for the Mean and Variance, *In Proceedings of the* 22^{nd} *International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, Melville, NY, 1-17.

[31]. Gull, S.F. and Fielden J., 1984, Bayesian non-parametric statistics, *Proceedings of the Fourth Maximum Entropy Workshop*, Cambridge University Press, NY, 85-94

[32]. Pendock, N., 1993, *A Bayesian Genetic Algorıthm for Calculating Maximum Entropy Distributions*, Maximum Entropy and Bayesian Methods, In: Heidbreder G.R (ed), Kluwer Academic Publishers, USA, 187-195.

[33]. Nesin, A., 2009, *Sayma (ya da Kombinasyon Hesapları)*, http://nesinkoyleri.org/e-kutuphane/ders-notlari/sayma.pdf.

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı	Emin Serhan Süzer
Uyruğu	T.C.
Doğum yılı, Yeri	1991, Zonguldak.
Telefon	05344553515
E-mail	serhanemin@hotmail.com

Eğitim

Derece	Kurum/Anabilim Dalı/Programı	Yılı
Lisans	İ.Ü. Fen Fakültesi/Matematik Bölümü	2013
Lise	Ataköy Cumhuriyet Lisesi	2009