

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ENTROPİNİN SIRA İSTATİSTİKLERİNE
UYGULANMASI

Hayat KILIÇ

YÜKSEK LİSANS TEZİ

İstatistik Anabilim Dalı

İstatistik Programı

Danışman

Doç. Dr. Atıf EVREN

Aralık, 2018

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ENTROPİNİN SIRA İSTATİSTİKLERİNE UYGULANMASI

Hayat KILIÇ tarafından hazırlanan tez çalışması çalışması 27.12.2018 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, İstatistik Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Doç. Dr. Atif EVREN

Yıldız Teknik Üniversitesi

Danışman

Jüri Üyeleri

Doç. Dr. Atif EVREN, Danışman

Yıldız Teknik Üniversitesi

Dr. Öğr. Üyesi Doğan YILDIZ, Üye

Yıldız Teknik Üniversitesi

Dr. Öğr. Üyesi Erhan USTAOĞLU, Üye

Marmara Üniversitesi

Danışmanım Doç. Dr. Atıf EVREN sorumluluğunda tarafımca hazırlanan Entropinin Sıra İstatistiklerine Uygulanması başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim

Hayat KILIÇ

İmza



Ağabeyim Hüseyin KILIÇ'a



TEŐEKKÜR

Yüksek lisans tez çalışmamın oluşmasında bana destek olan değerli tez danışmanım Doç. Dr. Atıf Ahmet EVREN'e teşekkür ederim.

Hayat

KILIÇ



İÇİNDEKİLER

SİMGE LİSTESİ.....	10
KISALTMA LİSTESİ	11
ŞEKİL LİSTESİ.....	12
ABSTRACT.....	14
1 Giriş	1
1.1 Literatür Özeti.....	1
1.2 Tezin Amacı.....	3
1.3 Hipotez.....	3
2 Sıra İstatistikleri.....	4
2.1 Sıra İstatistikleri Tanımı.....	4
2.2 En Büyük Sıra İstatistiğinin Dağılım Fonksiyonu	4
2.3 En Büyük Sıra İstatistiğinin Olasılık Yoğunluk Fonksiyonu.....	5
2.4 En Küçük Sıra İstatistiğinin Dağılım Fonksiyonu	5
2.5 En Küçük Sıra İstatistiğinin Olasılık Yoğunluk Fonksiyonu	6
2.6 r. Sıra İstatistiğinin Dağılım Fonksiyonu	6
2.7 r. Sıra İstatistiğinin Olasılık Yoğunluk Fonksiyonu.....	7
2.8 İki Sıra İstatistiğinin Ortak Olasılık Fonksiyonu	7
2.9 En Büyük ve En Küçük Sıra İstatistiğinin Ortak Olasılık Yoğunluk Fonksiyonu	8
2.10 Aralığın Olasılık Yoğunluk Fonksiyonu.....	8
2.11 Medyan.....	9
2.12 Medyanın Olasılık Yoğunluk Fonksiyonu.....	10
2.13 Kartiller	11
2.14 Birinci Kartil (Q_1) İçin Olasılık Yoğunluk Fonksiyonu.....	11

2.15 İkinci Kartil (Q_3) için Olasılık Yoğunluk Fonksiyonu	12
2.16 Sıra İstatistiklerinin Uniform Dağılımı	12
2.17 Maksimum ve Minimum Sıra İstatistiklerinin Olasılık Yoğunluk Fonksiyonu .	13
2.18 $u_{r:n}$ ve $u_{s:n}$ ' in Ortak Olasılık Dağılım Fonksiyonu	13
2.19 R. Sıra İstatistiğinin α . Dereceden Momenti	13
2.20 r. Sıra İstatistiğinin Beklenen Değeri ve Varyansı.....	14
2.22 Korelasyon	15
2.23 Standart Uniform Dağılım için Aralığın (R_n) Dağılımı	15
2.24 Uniform Dağılım İçin Medyanın Dağılımı	16
2.25 Sıra İstatistiklerinin Üstel Dağılımı	16
2.26 R. Sıra İstatistiği İçin Olasılık Yoğunluk Fonksiyonu	17
2.27 En Küçük Sıra İstatistiğinin Olasılık Yoğunluk Fonksiyonu.....	17
2.28 En Küçük Sıra İstatistiğinin Birikimli Olasılık Fonksiyonu	17
2.29 En Büyük Sıra İstatistiğinin Olasılık Yoğunluk Fonksiyonu.....	17
2.30 En Büyük Sıra İstatistiğinin Birikimli Olasılık Fonksiyonu	18
2.31 $X_{i:n}$ ve $X_{j:n}$ ' İn Ortak Olasılık Yoğunluk Fonksiyonu	18
2.32 $X_{1:n}$ ve $X_{n:n}$ ' İn Ortak Olasılık Yoğunluk Fonksiyonu	18
2.33 N Tane Sıra İstatistiğinin Ortak Dağılımı	18
2.34 i. Sıra İstatistiğinin K. Dereceden Momenti	18
2.35 Beklenen Değer.....	19
2.36 Kovaryans	21
2.37 Korelasyon	22
2.38 Standart Üstel Dağılım İçin Range	22
2.39 Standart Üstel Dağılım İçin Medyan	22
2.40 Sıra İstatistiklerinin Lojistik Dağılımı	23

2.41 i. Sıra İstatistiğinin Olasılık Yoğunluk Fonksiyonu.....	23
2.42 En Küçük Sıra İstatistiğinin Olasılık Yoğunluk Fonksiyonu.....	23
2.43 En Küçük Sıra İstatistiğinin Birikimli Olasılık Fonksiyonu	24
2.44 En büyük Sıra İstatistiğinin Olasılık Yoğunluk Fonksiyonu	24
2.45 En Büyük Sıra İstatistiğinin Birikimli Olasılık Fonksiyonu.....	24
2.46 i. ve j. Sıra İstatistiklerinin Ortak Olasılık Yoğunluk Fonksiyonu	24
2.47 Moment	24
2.48 Beklenen Değer ve Varyans.....	25
2.49 Lojistik Dağılım için Range	26
2.50 Medyan.....	27
2.51 Sıra İstatistiklerinin Normal Dağılımı.....	28
3 Bilgi Teorisi ve Entropi.....	30
3.1 Entropi Tanım	30
3.2 Entropinin Fiziksel Anlamı	31
3.3 Entropinin İstatistiksel Anlamı	32
3.4 Shannon Entropisi	33
3.5 Shannon Entropisinin Özellikleri.....	34
3.6 Maksimum Entropi İlkesi.....	34
3.7 Renyi Entropisi	35
3.8 Renyi Entropisinin Özellikleri	36
3.9 Havrda-Charvat ve Tsallis Entropileri	36
3.10 Diferansiyel Entropiler.....	36
3.11 Uniform Dağılımın Shannon Entropisi	37
3.12 Üstel Dağılımın Shannon Entropisi	37
3.13 Normal Dağılımın Shannon Entropisi.....	37
3.14 Teorem	38

3.15 Olasılık İntegral Dönüşümü	38
3.16 Üstel Dağılım	39
3.17 Uniform Dağılım.....	41
4 Uygulama.....	43
5 Sonuç ve Tartışma	53
Kaynakça	55
Tezden Üretilmiş Yayınlar.....	57



SİMGE LİSTESİ

$\beta(a,b)$	Beta fonksiyonu
μ	Konum parametresi
$F(x)$	Olasılık dağılım fonksiyonu
$f(x)$	Olasılık yoğunluk fonksiyonu
σ	Ölçek parametresi
$I,(a,b)$	Tamamlanmamış beta fonksiyonu
s	Standart sapma
\bar{x}	X değerlerinin ortalaması

KISALTMA LİSTESİ

CDF	Birikimli Dağılım Fonksiyonu
CORR	Korelasyon
COV	Kovaryans
OYF	Olasılık Yoğunluk Fonksiyonu
PDF	Olasılık Yoğunluk Fonksiyonu



ŞEKİL LİSTESİ

Şekil 4. 1 Sıra değerine karşı entropi	43
Şekil 4. 2 Sıra değerine karşı s.sapma.....	43
Şekil 4. 3 Std.sapmaya karşı entropi	44
Şekil 4. 4 Sıra değerine karşı entropi	44
Şekil 4. 5 Sıra değerine karşı s.sapma.....	44
Şekil 4. 6 Std.sapmaya karşı entropi	45
Şekil 4. 7 Sıra değerine karşı entropi	45
Şekil 4. 8 Sıra değerine karşı s.sapma.....	46
Şekil 4. 9 Std.sapmaya karşı entropi	46
Şekil 4. 10 Sıra değerine karşı entropi	47
Şekil 4. 11 Sıra değerine karşı s.sapma	47
Şekil 4. 12 Std.sapmaya karşı entropi.....	47
Şekil 4. 13 Sıra değerine karşı entropi	48
Şekil 4. 14 Sıra değerine karşı s.sapma	48
Şekil 4. 15 Std.sapmaya karşı entropi.....	48
Şekil 4. 16 Sıra değerine karşı entropi	49
Şekil 4. 17 Sıra değerine karşı s.sapma	49
Şekil 4. 18 Std.sapmaya karşı entropi.....	49
Şekil 4. 19 Sıra değerine karşı entropi	50
Şekil 4. 20 Sıra değerine karşı s.sapma	50
Şekil 4. 21 Std.sapmaya karşı entropi.....	50
Şekil 4. 22 Sıra değerine karşı entropi	51
Şekil 4. 23 Sıra değerine karşı s.sapma	51
Şekil 4. 24 Std.sapmaya karşı entropi.....	52
Şekil 4. 25 Sıra değerine karşı s.sapma	52
Şekil 4. 26 Sıra değerine karşı entropi	52
Şekil 4. 27 Std.sapmaya karşı entropi.....	49

Entropinin Sıra İstatistiklerine Uygulanması

Hayat KILIÇ

İstatistik Anabilim Dalı

Yüksek Lisans Tezi

Danışman: Doç. Dr. Atif EVREN

Sıra istatistikleri meteorolojik, tarımsal ya da finansal verinin modellenmesinde önemlidir. Genellikle aşırı olayların modellenmesinde kullanılmaktadırlar. Bir dağılımın minimumu, birinci dörde bölü, medyanı, üçüncü dörde bölü ve maksimumu sıra istatistiklerine örnek olarak verilebilirler.

Bir istatistiksel deneyin entropisi örnekleme (veya gözleme) gidilmeden önceki belirsizliğin ölçüsüdür. Öyle ki entropi ve örnekten gelen bilgi miktarı birbirleri ile yakından ilişkili iki kavramdır. Çeşitli entropi ölçüleri Esteban ve Morales (1995) ve Pardo (2006) tarafından özetlenmiştir. Bunların arasında Shannon, Rényi ve Tsallis entropilerinin yakın zamandaki çalışmalarda yaygın bir biçimde kullanıldıklarını vurgulamak gerekir.

Jaynes (2007) tarafından gündeme getirilen maksimum entropi ilkesi özellikle önsel dağılım hakkında az ya da hiçbir bilgi olmadığı durumlarda Bayesçi modellemede kullanılmaktadır.

Uniform, üstel ve normal dağılımlar, olasılık dağılımları hakkında bazı varsayımların geçerli olması durumunda, maksimum entropiyi veren üç dağılımdır. Bunun yanı sıra sınırlandırılma açısından bu üç dağılım birbirlerinden tamamiyle farklı öz niteliklere sahiptirler.

Anahtar Kelimeler: Sınırlı dağılımlar, Sıra istatistikleri, Shannon entropisi, Belirsizlik



ABSTRACT

Entropy of Order Statistics and Application

Hayat KILIÇ

Department of Statistics
Master's Thesis

Advisor: Doç. Dr. Atif EVREN

As well as being three maximum entropy distributions, uniform, exponential and normal distributions have different properties in terms of censorization. Uniform distribution is censored from two sides. Exponential distribution is censored from below, whereas the normal distribution is uncensored. In this study, Shannon entropies of order statistics from uniform, exponential and normal distributions are considered. It has been found that entropies of order statistics are some functions of the entropies of parent distributions. They are also functions of sample size, and order of the statistic. Entropy estimates are then compared with standard deviations of order statistics. It has been detected that for order statistics of censored distributions, like uniform or exponential distribution; as the two measures of uncertainty; standard deviation, and entropy do not convey the same information, i.e., they are not positively correlated. This is probably because the lowest and/or highest order statistics do not show high variability due to censorization. For uncensored distributions like the normal distribution, entropy and standard deviation are positively correlated as expected a priori.

Therefore in case of censorization using entropy statistics may not be appropriate to measure uncertainty or variability of order statistics.

]

YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

1.1 Literatür Özeti

Sıra istatistikleri tanımı ilk kez 1942 yılında Wilks tarafından ortaya atılmış olmasına rağmen konu olarak sıra istatistiği daha eskiye dayanır. x_1, x_2, \dots, x_n , birbirinden bağımsız olan ve aynı $F(x)$ c.d.f'na sahip rasgele değişkenler olsun. Bu rasgele değişkenler büyüklüklerine göre artacak şekilde sıralansın. Bu durumda $x_{1n} \leq x_{2n} \leq \dots \leq x_{rn}$ x sıralanmış rastgele değişkenleri, $x_{r:n}$, r' inci sıra istatistiğine karşılık gelecek şekilde "sıra istatistikleri" olarak adlandırılırlar.

($r = 1, 2, \dots, n$) Genellikle aşırı olayların modellenmesinde kullanılmaktadırlar. Bir dağılımın minimumu, birinci dörde bölüneni, medyanı, üçüncü dörde bölüneni ve maksimumu sıra istatistiklerine örnek olarak verilebilirler.

Termodinamik'teki entropi kavramını olasılık ile harmanlayan ilk bilim insanlarından biri L. Boltzman'dır. Bir gazın sahip olduğu serbestlik derecesi, gazın belirsizliğidir diyen Boltzman, fizikte entropiyi ve olasılık kavramlarını tekrar gündeme getirerek, termodinamikteki düzensizlik denklemini ortaya atmıştır. Bu denklem;

$S = k \ln W$ şeklinde ifade edilir ve buradaki $k = 1,3806505 \times 10^{-23}$ J/K Boltzmann sabiti olarak isimlendirilir.

S : Entropi

k : Boltzman değişmezi

W : Termodinamik olasılık

Avusturyalı fizikçi Ludwig Boltzmann, ne yazık ki, yıllar boyu birçok bilim adamının eleştirilerine karşı mücadele etmiş ve buluşu dikkate alınmamıştır ancak buluşu günümüzde evrenin en temel yasalarından biri olarak bahsedilmektedir.

Entropi çeşitleri olarak;

Shannon Entropisi

Renyi Entropisi

Havrda-Charvat ve Tsallis Entropileri

Bileşik Entropi

Koşullu Entropi

Görelî Entropi

Bulanık Entropi

Yukarıda saydıklarımız örnek olarak verilebilir. Ancak biz sadece en çok bilinenler üzerinde yani ilk üçü üzerinde duracağız.

Suyun üç hali olan katı, sıvı ve gaz hallerinin her biri farklı entropilere sahiptir. Buzun molekülleri daha düzenli ve aralarında daha az boşluk olduğundan en az entropiye sahiptir. Sıvı hal, katı hale oranla nispeten molekülleri arasındaki boşluk daha fazla olduğu için entropisi daha yüksek ve molekülleri arasındaki boşluk en fazla olan yani en yüksek entropiye sahip olan gaz halidir.

Bilgi teorisi ve entropi kavramları birbirleriyle yakından ilişkili iki kavramdır.

Bilgi teorisi konu olarak haberleşmeyi konu almakla birlikte matematik ve istatistik gibi bilimlere de içinde barındırır. Shannon'un "A mathematical Theory of Communication" isimindeki makalesi, bu teoremin çıkışı olarak ele alınır. Bu çalışmada iyi bir iletişimin nasıl olacağı konusunda araştırmalar Shannon, iletişimin matematiksel modelini kurdu ve böylece en önemli teoremlerden biri olan bilgi teorisinin temellerini attı.

Entropi, bir rasgele değişkenin değerini tahmin ederken belirsizliği nicelikselleştirir. Örnek verecek olursak; yazı ve tura atma deneyi ve zar atma deneyini karşılaştıracak olursak; bu iki deneyin de sağladığı bilgi birbirinden farklıdır. Yazı-tura atma deneyinin olasılıkları birbirine eşit olan iki sonucu varken zar atma deneyinin eşit olasılıkları birbirine eşit olan altı tane sonucu vardır. Diğer bir ifadeyle yazı-tura atma deneyinde daha az bilgi olduğu için entropisi daha yüksektir.

Uniform, üstel ve normal dağılımlar, olasılık dağılımları hakkında bazı varsayımların geçerli olması durumunda, maksimum entropiyi veren üç dağılımdır. Bunun yanı sıra sınırlandırılma açısından bu üç dağılım birbirlerinden tamamiyle farklı öz niteliklere sahiptirler. Uygulamada sınırlandırmanın seçilen dağılımın entropisi ve standart sapması arasındaki ilişki üzerinde durulmuştur.

1.2 Tezin Amacı

Bu tezde sıra istatistiklerinin tanımına istatistik bilimindeki kullanım alanlarına ve önemine değinilmiş ayrıca belirsizliğin nicelleştirilmesinde önemli rol oynayan entropi kavramı hakkında bilgi verilmiştir. Bilgi Teorisi ve entropi arasındaki ilişki, entropinin fizikteki ve istatistikteki anlamlarına değinilmiş ve entropinin türleri verilmiştir. Son olarak ise konu ile ilgili uygulama yapılmış ve seçili üç dağılımın (üstel, normal, uniform) sıra istatistiklerinin entropileri bulunmuş.

1.3 Hipotez

Sıra istatistiği ve entropi kavramları günümüzde istatistik alanında sıkça kullanılan iki kavramdır. Son yıllarda entropi ile ilgili çalışmalar bilim dünyasındaki önemli konulardan biridir. Bu çalışmamda sıra istatistiklerinin entropileri bulunup; sınırlılığın, entropi ve standart sapma arasındaki ilişkiye etkisi gösterilmiştir.

2.1 Sıra İstatistikleri Tanımı

x_1, x_2, \dots, x_n , birbirinden bağımsız ve aynı $F(x)$ c.d.f' na sahip rasgele değişkenler olsun. Bu rasgele değişkenler büyüklüklerine göre artacak şekilde sıralansın. Bu durumda $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$ sıralanmış rastgele değişkenleri, $x_{r:n}$, r'inci sıra istatistiğine karşılık gelecek şekilde "sıra istatistikleri" olarak adlandırılırlar ($r = 1, 2, \dots, n$) (David , 1970 , Gibbons , 1971, Arnold vd., 1992) . Burada x_1, x_2, \dots, x_n , rasgele değişkenleri birbirinden bağımsız ve aynı $F(x)$ dağılım fonksiyonuna sahip iken, $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$ sıra istatistikleri, aralarındaki eşitsizlik ilişkileri nedeniyle bağımlı olacaktır.

2.2 En Büyük Sıra İstatistiğinin Dağılım Fonksiyonu

x_1, x_2, \dots, x_n , rasgele değişkenleri, $f(x)$ olasılık yoğunluk fonksiyonu ve $F(x)$ birikimli dağılım fonksiyonuna sahip mutlak sürekli bir kitleden gelen rasgele bir örneklem olsun. $r = 1, 2, \dots, n$ olmak üzere , $x_{r:n}$, ile tanımlanan r'inci sıra istatistiği için c.d.f ; $F_n(x)$ ile gösterilsin . Böylece, $x_{r:n} = \max \{x_1, x_1, \dots, x_n\}$ olarak tanımlanan ve $F_n(x)$ ile gösterilen en büyük sıra istatistiğinin dağılım fonksiyonu;

$$\begin{aligned}
 F_n(x) &= P\{x_{n:n} \leq x\} = P\{\max \{x_1, x_1, \dots, x_n\} \leq x\} \\
 &= P\left\{\sum_{i=1}^n x_i \leq x\right\} \\
 &P\{x_1 \leq x, x_2 \leq x, \dots, x_n \leq x\} \\
 &P\{x_1 \leq x\} P\{x_2 \leq x\} \dots P\{x_n \leq x\} \\
 &F(x_1)F(x_2) \dots F(x_n) \\
 &= \{F(x)\}^n, -\infty < x < \infty
 \end{aligned} \tag{2.1}$$

şeklinde elde edilir.

2.3 En Büyük Sıra İstatistiğinin Olasılık Yoğunluk Fonksiyonu

x_1, x_2, \dots, x_n , rasgele değişkenleri, $f(x)$ olasılık yoğunluk fonksiyonu ve $F(x)$ birikimli dağılım fonksiyonuna sahip mutlak sürekli bir kitleden gelen rasgele bir örneklem olsun.

$f_n(x)$ ile gösterilen en büyük sıra istatistiğinin olasılık yoğunluk fonksiyonu;

$$\begin{aligned}
 P(X_{(n)} \in [x, x + \varepsilon]) &= P(x_1, x_2, \dots, x_n \text{ 'lerden biri} \in [x, x + \varepsilon] \text{ ve diğerleri} < x) \\
 &= \sum_{i=1}^n P(X_i \in [x, x + \varepsilon] \text{ ve diğerleri} < x) \\
 &= nP(X_1 \in [x, x + \varepsilon] \text{ ve diğerleri} < x) \\
 &= nP(X_1 \in [x, x + \varepsilon])P(X_2 < x) \dots P(X_n < x) \\
 &= nf(x)\varepsilon F(x)^{n-1} \\
 f_{(n)}(x) &= nf(x)F(x)^{n-1} \tag{2.2}
 \end{aligned}$$

elde edilir.

2.4 En Küçük Sıra İstatistiğinin Dağılım Fonksiyonu

Benzer şekilde $x_{1,n} = \min\{x_1, x_2, \dots, x_n\}$ olarak tanımlanan ve $F_1(x)$ ile gösterilen en küçük sıra istatistiğinin dağılım fonksiyonu,

$$\begin{aligned}
 F_1(x) &= P\{x_{1,n} \leq x\} = P\{\min\{x_1, x_2, \dots, x_n\} \leq x\} \\
 &= 1 - P\left\{\sum_{i=1}^n x_i > x\right\} \\
 &= 1 - P\{x_1 > x, x_2 > x, \dots, x_n > x\} \\
 &= 1 - P\{x_1 > x\}P\{x_2 > x\} \dots P\{x_n > x\} \\
 &= 1 - [1 - P\{x_1 \leq x\}][1 - P\{x_2 \leq x\}] \dots [1 - P\{x_n \leq x\}] \\
 &= 1 - [1 - F(x)]^n, -\infty < x < \infty \tag{2.3}
 \end{aligned}$$

şeklinde elde edilir.

2.5 En Küçük Sıra İstatistiğinin Olasılık Yoğunluk Fonksiyonu

x_1, x_2, \dots, x_n , rasgele değişkenleri, $f(x)$ olasılık yoğunluk fonksiyonu ve $F(x)$ birikimli dağılım fonksiyonuna sahip mutlak sürekli bir kitleden gelen rasgele bir örneklem olsun. $f_1(x)$ ile gösterilen en küçük sıra istatistiğinin olasılık yoğunluk fonksiyonu;

$$\begin{aligned} P(X_{(1)} \in [x, x + \varepsilon]) &= P(x_1, x_2, \dots, x_n \text{ 'lerden biri} \in [x, x + \varepsilon] \text{ ve diğerleri} > x) \\ &= \sum_{i=1}^n P(X_i \in [x, x + \varepsilon] \text{ ve diğerleri} < x) \\ &= nP(X_1 \in [x, x + \varepsilon] \text{ ve diğerleri} > x) \\ &= nP(X_1 \in [x, x + \varepsilon]) P(\text{diğerleri} > x) \\ &= nP(X_1 \in [x, x + \varepsilon]) P(X_2 > x) \dots P(X_n > x) \\ &= nf(x) \varepsilon (1 - F(x))^{n-1} \\ &= f_{(1)}(x) = nf(x) (1 - F(x))^{n-1} \end{aligned} \tag{2.4}$$

elde edilir.

2.6 r. Sıra İstatistiğinin Dağılım Fonksiyonu

$1 \leq r \leq n$ olmak üzere, $x_{r:n}$ olarak gösterilen r. sıra istatistiğinin dağılım fonksiyonu $F_r(x)$ aşağıdaki gibi elde edilir;

$$\begin{aligned} F_n(x) &= P\{x_{r:n} \leq x\} \\ &= P\{x_1, x_2, \dots, x_n \text{ 'lerden en az } r \text{ sayısı } x \text{ 'ten küçük veya eşit}\} \\ &= P\{r \text{ tane } x_i \leq x \text{ veya } (r+1) \text{ tane } x_i \leq x \text{ veya } \dots \text{ n tane } x_i \leq x\} \\ &= \sum_{i=r}^n P\{x_1, x_2, \dots, x_n, \text{ 'lerden tam olarak } i \text{ tanesi } x \text{ 'ten küçük veya eşit}\} \\ &= \sum_{i=r}^n \binom{n}{i} [F(x)]^i [1 - F(x)]^{n-i}, -\infty < x < \infty \end{aligned} \tag{2.5}$$

Binom toplamları ve tamamlanmamış beta fonksiyonu arasındaki ilişkiye göre;

$$= \sum_{i=1}^n \binom{n}{i} [F(x)]^i [1-F(x)]^{n-i} = I_{F(x)}(r, n-r+1)$$

$$I_p(a, b) = \frac{1}{B(a, b)} \int_0^p t^{a-1} (1-t)^{b-1} dt, \dots 0 \leq p \leq 1$$

$$B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt$$

$$F_r(x) = I_{F(x)}(r, n-r+1)$$

2.7 r. Sıra İstatistiğinin Olasılık Yoğunluk Fonksiyonu

x_1, x_2, \dots, x_n , birbirinden bağımsız ve aynı $F(x)$ kümülatif dağılım fonksiyonuna ve $f(x)$ olasılık yoğunluk fonksiyonuna sahip rasgele değişkenler olsun.

$$P(X_{(r)} \in [x, x + \varepsilon]) = P(x_1, x_2, \dots, x_n \text{ 'lerden biri} \in [x, x + \varepsilon] \text{ ve } r-1 \text{ tanesi} < x)$$

$$= \sum_{i=1}^n P(X_i \in [x, x + \varepsilon] \text{ ve } r-1 \text{ tanesi} < x)$$

$$= nP(X_1 \in [x, x + \varepsilon] \text{ ve } r-1 \text{ tanesi} < x)$$

$$= nP(X_1 \in [x, x + \varepsilon]) \binom{n-1}{r-1} P(X < x)^{r-1} P(X > x)^{n-r}$$

$$f_{(r)}(x) = nf(x) \binom{n-1}{r-1} F(x)^{r-1} (1-F(x))^{n-r}$$

$$= \frac{n!}{(r-1)!(n-r)!} (F(x))^{r-1} (1-F(x))^{n-r} f(x), -\infty < x < \infty \quad (2.6)$$

2.8 İki Sıra İstatistiğinin Ortak Olasılık Fonksiyonu

$X_{r:n}$ ve $X_{s:n}$ iki sıra istatistiği ve $r < s$ olsun;

$$F_{r,s:n}(x_r, x_s) = P(X_{r:n} \leq x_r, X_{s:n} \leq x_s)$$

= P (En az r tane X_i , x_r 'den küçük veya eşit ve en az s tane X_i , x_s 'den küçük veya eşit)

$$= \sum_{j=s}^n \sum_{i=r}^j P(\text{Tam olarak } r \text{ tane } X_i, x_r' \text{ den küçük veya eşit ve tam olarak } s \text{ tane } X_i, x_s' \text{ den küçük veya eşit})$$

$$\begin{aligned}
&= \sum_{j=s}^n \sum_{i=r}^j \frac{n!}{i!(j-i)!(n-j)!} F^i(x_r) [F(x_s) - F(x_r)]^{j-i} [1 - F(x_s)]^{n-j} \\
&= \sum_{j=n}^n \sum_{i=r}^j \frac{n!}{i!(j-i)!(n-j)!} p_1^i (p_2 - p_1)^{j-i} (1 - p_2)^{n-j} = \int_0^{p_1} \int_r^{p_2} \frac{n!}{(r-i)!(s-r-1)!(n-s)} t_1^{r-1} (t_2 - t_1)^{s-r-1} (1-t_2)^{n-s} dt_2 dt_1
\end{aligned}$$

eşitliğinden yararlanarak iki sıra istatistiği için dağılım fonksiyonu;

$$F_{r,s}(x_r, x_s) = \int_0^{F(x_r)} \int_{t_1}^{F(x_s)} \frac{n!}{(r-i)!(s-r-1)!(n-s)} t_1^{r-1} (t_2 - t_1)^{s-r-1} (1-t_2)^{n-s} dt_2 dt_1$$

$-\infty < x < \infty$ aralığında, bilinen adıyla tamamlanmamış bivariate beta fonksiyonu elde edilmiş olur, kesikli ve sürekli değişkenler için.

Eğer $F(x)$ fonksiyonu sürekli ise $x_{r:n}$ ve $x_{r:s}$ 'in olasılık yoğunluk fonksiyonu;

$$\begin{aligned}
f_{r,s,n}(x_r, x_s) &= \frac{d^2}{dx_r dx_s} F_{r,s,n}(x_r, x_s) \\
\frac{d^2}{dx_r dx_s} &= \left[\int_0^{F(x_r)} \int_{t_1}^{F(x_s)} \frac{n!}{(r-1)!(s-r-1)!(n-s)} t_1^{r-1} (t_2 - t_1)^{s-r-1} (1-t_2)^{n-s} dt_2 dt_1 \right] \\
&= \frac{1}{B(r, s-r, n-s+1)} f(x_r) f(x_s) F^{r-1}(x_r) [F(x_s) - F(x_r)]^{s-r-1} [1 - F(x_s)]^{n-s}
\end{aligned}$$

$-\infty < x_r < x_s < \infty$ aralığı için elde edilir.

$$[B(r, s-r, n-s+1)]^{-1} = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \quad (2.7)$$

2.9 En Büyük ve En Küçük Sıra İstatistiğinin Ortak Olasılık Yoğunluk Fonksiyonu

Yukarıdaki eşitlikten yararlanarak $r=1$ ve $s=n$ alınarak;

$$f_{1,n,n}(x_1, x_n) = n(n-1) f(x_1) f(x_n) [F(x_n) - F(x_1)]^{n-2} \quad -\infty < x_r < x_n < \infty \quad (2.8)$$

bulunur.

2.10 Aralığın Olasılık Yoğunluk Fonksiyonu

$X_{r:n}$ ve $X_{s:n}$ 'in ortak olasılık yoğunluk fonksiyonunun,

$$f_{r,s,n}(x_r, x_s) = \frac{1}{B(r, s-r, n-s+1)} f(x_r) f(x_s) F^{r-1}(x_r) [F(x_s) - F(x_r)]^{s-r-1} [1 - F(x_s)]^{n-s}$$

olduğunu göstermiştik. Buradan hareketle,

$W_{rs} = X_{s:n} - X_{r:n}$ ve $w_{rs} = x_s - x_r$ tanımlansın w_{rs} ve X_r 'nin ortak olasılık yoğunluk fonksiyonu;

$$f_{w_{rs}}(w_{rs}) = \frac{1}{B(r, s-r, n-s+1)} f(x_r) f(x_s + w_{rs}) F^{r-1}(x_r) [F(x_r + w_{rs}) - F(x_r)]^{s-r-1} [1 - F(x_r + w_{rs})]^{n-s}$$

bulunur.

w_{rs} 'nin marjinal yoğunluk fonksiyonu;

$$f_{w_{rs}}(w_{rs}) = \int_{-\infty}^{\infty} \frac{1}{B(r, s-r, n-s+1)} f(x_r) f(x_s + w_{rs}) F^{r-1}(x_r) [F(x_r + w_{rs}) - F(x_r)]^{s-r-1} [1 - F(x_r + w_{rs})]^{n-s} dx_r, \quad (2.9)$$

elde edilir.

Yukarıdaki denklemde (2.9) $r=1$ ve $s=n$ (örneklem hacmi) seçildiğinde aralığın olasılık yoğunluk fonksiyonu;

$$f_w(w) = n(n-1) \int_{-\infty}^{\infty} f(x_r) f(x_r + w) [F(x_r + w) - F(x_r)]^{n-2} dx_r, \quad (2.10)$$

Aralığın olasılık dağılım fonksiyonu ise;

$$\begin{aligned} F_w(w) &= \int_0^w \int_{-\infty}^{\infty} n(n-1) f(x_r) f(z+x_r) [F(z+x_r) - F(x_r)]^{n-2} dx_r dz \\ &= \int_{-\infty}^{\infty} n(n-1) f(x_r) \int_0^w f(z+x_r) [F(z+x_r) - F(x_r)]^{n-2} dz dx_r \\ &= \int_{-\infty}^{\infty} n(n-1) f(x) \left(\frac{1}{n-1} [F(z+x_r) - F(x_r)]^{n-1} \Big|_0^w \right) dx \\ &= \int_{-\infty}^{\infty} n f(x) [F(w+x_r) - F(x_r)]^{n-1} dx, \end{aligned} \quad (2.11)$$

elde edilir.

2.11 Medyan

X_1, X_2, \dots, X_n , gerçel değerli herhangi rasgele değişkenler olsun. Bu rasgele değişkenler küçükten büyüğe doğru sıralandığında ortadaki terime medyan denir.

$n=2m+1$ yani tek olduğu durumlarda medyan $(m+1)$. terim olur.

$n=2m$ yani çift olduğu durumlarda;

$$\text{medyan} = \frac{1}{2} \{X_m + X_{m+1}\}$$

$$\text{Medyan} = \begin{cases} X_{(n+1)/2:n}, & n \text{ tek} \\ \frac{X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n}}{2}, & n \text{ çift} \end{cases} \text{ şeklinde bulunur.}$$

2.12 Medyanın Olasılık Yoğunluk Fonksiyonu

Örneklem hacmi (n)'in tek olduğu durumu göz önüne alalım;

$$\bar{X}_n = X_{\frac{(n+1)}{2}} \text{ diye tanımlansın,}$$

$$\begin{aligned} f_{\bar{X}_n}(x) &= \frac{n!}{\left(\frac{n+1}{2}-1\right)!\left(n-\frac{n+1}{2}\right)!} [F(x)]^{\frac{n+1}{2}-1} [1-F(x)]^{n-\frac{n+1}{2}} f(x) \\ &= \frac{n!}{\left[\left(\frac{n-1}{2}\right)!\right]^2} [F(x)(1-F(x))]^{\frac{n-1}{2}} f(x), \quad -\infty < x < \infty \end{aligned} \quad (2.12)$$

Örneklem hacmi (n)'in çift olduğu durumu göz önüne alalım;

$$\bar{X}_n = \frac{X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n}}{2} \text{ şeklinde tanımlanır.}$$

$X_{\frac{n}{2}:n}$ ve $X_{\frac{n}{2}+1:n}$ 'in ortak olasılık yoğunluk fonksiyonundan;

$$\begin{aligned} f_{X_{\frac{n}{2}:n}, X_{\frac{n}{2}+1:n}}(x_1, x_2) &= C_{x_1, x_2} [F(x_1)]^{\frac{n}{2}-1} [F(x_1) - F(x_2)]^{\frac{n}{2}+1-\frac{n}{2}-1} [1-F(x_2)]^{n-\left(\frac{n}{2}+1\right)} f(x_1) f(x_2) \\ &= \frac{n!}{\left[\left(\frac{n}{2}-1\right)!\right]^2} [F(x_1)(1-F(x_2))]^{\frac{n}{2}-1} f(x_1) f(x_2), \quad -\infty < x_1 < x_2 < \infty \end{aligned} \quad (2.13)$$

$$C_{x_1, x_2} = \frac{n!}{\left(\frac{n}{2}-1\right)!\left(\frac{n}{2}+1-\frac{n}{2}-1\right)!\left(n-\left(\frac{n}{2}+1\right)\right)!}$$

(2.13) yazılabildiğini göstermiştik.

Yukarıdaki denklemde $x_2 = 2x - x_1$ dönüşümü uygulandığında;

$$f_{X_{\frac{n}{2}}, \bar{X}_n}(x_1, x_2) = C_{x_1, x} [F(x_1)]^{\frac{n}{2}-1} [1-F(2x-x_1)]^{\frac{n}{2}-1} [f(2x-x_1)F(x_1)]^{n-n} f(x_1)f(2x_2-x_1)$$

$$= \frac{n!}{\left[\left(\frac{n}{2}-1\right)!\right]^2} [F(x_1)]^{\frac{n}{2}-1} [1-F(2x_2-x_1)]^{\frac{n}{2}-1} f(x_1)f(2x_2-x_1), \quad -\infty < x_1 < x < \infty$$

$$C_{x_1, x} = \frac{2n!}{\left(\frac{n}{2}-1\right)!(n-\frac{n}{2}-1)!(n-n)!}$$

$$f_{\bar{X}_n}(x) = \frac{n!}{\left[\left(\frac{n}{2}-1\right)!\right]^2} \int_{-\infty}^x [F(x_1)]^{\frac{n}{2}-1} [1-F(2x_2-x_1)]^{\frac{n}{2}-1} f(x_1)f(2x_2-x_1)dx_1, \quad -\infty < x < \infty \quad (2.14)$$

olasılık yoğunluk fonksiyonu elde edilir.

2.13 Kartiller

X_1, X_2, \dots, X_n , gerçel değerli herhangi rasgele değişkenler olsun. Bu rasgele değişkenleri küçükten büyüğe doğru sıraladığımızda dört eşit parçaya ayıran üç değere kartiller adı verilir. Diğer bir adıyla birinci dörde bölen ve üçüncü dörde bölen aşağıdaki gibi tanımlanır.

$$Q_1 = \begin{cases} X_{\frac{n+1}{4}:n}, & n \text{ tek} \\ \frac{X_{\frac{n}{4}:n} + X_{\frac{n+1}{4}:n}}{2}, & n \text{ çift} \end{cases}$$

$$Q_3 = \begin{cases} X_{\frac{3n+1}{4}:n}, & n \text{ tek} \\ \frac{X_{\frac{3n}{4}:n} + X_{\frac{3n+1}{4}:n}}{2}, & n \text{ çift} \end{cases}$$

2.14 Birinci Kartil (Q_1) İçin Olasılık Yoğunluk Fonksiyonu

Örnekleme hacmi (n)'in tek olduğu durumu göz önüne alalım;

$$\bar{X}_n = X_{\frac{(n+1)}{4}:n} \text{ olsun;}$$

$$f_{\bar{X}_n}(x) = \frac{n!}{\left(\frac{n+1}{4}-1\right)!\left(n-\frac{n+1}{4}\right)!} [F(x_1)]^{\frac{n+1}{4}-1} [1-F(x)]^{n-\frac{n+1}{4}} f(x) \quad (2.15)$$

bulunur.

2.15 İkinci Kartil (Q_3) için Olasılık Yoğunluk Fonksiyonu

$$\bar{X}_n = X_{\frac{(3n+1)}{4};n}$$

$$f_{\bar{X}_n}(x) = \frac{n!}{\left(\frac{3n+1}{4}-1\right)! \left(n-\frac{3n+1}{4}\right)!} [F(x)]^{\frac{3n+1}{4}-1} [1-F(x)]^{n-\frac{3n+1}{4}} f(x) \quad (2.16)$$

olarak bulunur.

2.16 Sıra İstatistiklerinin Uniform Dağılımı

x_1, x_2, \dots, x_n , birbirinden bağımsız $F_x(x) = F(x)$ ortak dağılım fonksiyonu ve $f_x(x) = f(x)$ ortak yoğunluk fonksiyonuna sahip n boyutlu örneklemden rasgele çekilmiş sürekli değişkenler olsun $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, X_1, X_2, \dots, X_n , 'in sıralı değerleri olarak gösterilsin.

$u_1, u_2, \dots, u_n; U[0,1]$ 'de bağımsız değişkenler ve $u_{(r)}, r=1, 2, \dots, n$ $r=1, 2, \dots, n$ bu değişkenlerin sıra istatistikleri olsun.

$$\begin{aligned} F_{r:n}(x) &= P(X_{r:n} \leq x) \\ &= \sum_{r=1}^n \binom{n}{r} [F(x)]^r [1-F(x)]^{n-r}, \quad -\infty < x < \infty \end{aligned}$$

olduğunu önceki bölümde göstermiştik.

$$\sum_{r=1}^n \binom{n}{r} p^r [1-p]^{n-r} = \int_0^p \frac{n!}{(r-1)!(n-r)!} t^{r-1} (1-t)^{n-r} dt, \quad 0 < p < 1 \text{ eşitliğini kullanarak;}$$

$$\begin{aligned} F_{r:n}(x) &= \int_0^{F(x)} \frac{n!}{(r-1)!(n-r)!} t^{r-1} (1-t)^{n-r} dt \\ &= I_{F(x)}(r, n-r+1), \quad -\infty < x < \infty \end{aligned} \quad (2.17)$$

tamamlanmamış Beta fonksiyonu elde edilir.

Yukarıdaki (2.17)denkleminden yararlanarak, $u_{r:n}$ 'in olasılık yoğunluk fonksiyonu;

$$\begin{aligned} F_{r:n}(u) &= \sum_{r=1}^n \binom{n}{r} u^r [1-u]^{n-r} \\ &= \int_0^u \frac{n!}{(r-1)!(n-r)!} t^{r-1} [1-t]^{n-r} dt, \quad 0 < u < 1 \end{aligned} \quad (2.18)$$

elde edilir.

2.17 Maksimum ve Minimum Sıra İstatistiklerinin Olasılık Yoğunluk Fonksiyonu

Yukarıdaki (2.18) denkleminde yola çıkarak;

$$f_1(u) = n(1-u)^{n-1} \quad (2.19)$$

$$f_n(u) = nu^{n-1} \quad (2.20)$$

olarak bulunur.

2.18 $u_{r:n}$ ve $u_{s:n}$ 'in Ortak Olasılık Dağılım Fonksiyonu

$$f_{r,s:n}(u_r, u_s) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} u_r^{r-1} (u_s - u_r)^{s-r-1} (1-u_s)^{n-s}, \quad 0 \leq u_r < u_s \leq 1 \quad (2.21)$$

2.19 R. Sıra İstatistiğinin α . Dereceden Momenti

$$\begin{aligned} E(U_{r:n})^\alpha &= \int_0^1 x^\alpha f(x) dx \\ &= \frac{n!}{(r-1)!(n-1)!} \int_0^1 x^\alpha [F(x)]^{r-1} [1-F(x)]^{n-r} f(x) dx \\ &= \frac{n!}{(r-1)!(n-1)!} \int_0^1 x^\alpha x^{r-1} dx \\ &= \frac{n!}{(r-1)!(n-1)!} \beta(\alpha+r, n-r+1) \\ &= \frac{n! \Gamma(\alpha+r) \Gamma(n-r+1)}{(r-1)! \Gamma(n+\alpha+1)!} \\ &= \frac{n! \Gamma(\alpha+r)}{(r-1)! \Gamma(n+\alpha+1)!} \\ &= \frac{n!(\alpha+r-1)!}{(r-1)!(n+\alpha)!} \end{aligned} \quad (2.22)$$

olarak bulunur ve ayrıca;

$$\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

$$\Gamma(n) = (n-1)! \quad n=1,2,\dots$$

Ayrıca α . momentin $= \frac{\beta(r+\alpha, n-r+1)}{\beta(r, n-r+1)}$ değerine eşit olduğu da görülmektedir.

2.20 r. Sıra İstatistiğinin Beklenen Değeri ve Varyansı

Özel olarak (2.22) denkleminde $\alpha = 1$ alındığında;

$$\begin{aligned}
 E(U_{r:n})^2 &= \frac{n!\Gamma(r+1)}{(r-1)!\Gamma(n+2)} \\
 &= \frac{n!r!}{(r-1)!(n+1)!} \\
 &= \frac{r}{n+1}, 1 \leq r \leq n
 \end{aligned} \tag{2.23}$$

$\alpha = 2$ için;

$$\begin{aligned}
 E(U_{r:n})^2 &= \frac{n!\Gamma(r+2)}{(r-1)!\Gamma(n+3)} \\
 &= \frac{n!(r+1)!}{(r-1)!(n+2)!} \\
 &= \frac{r(r+1)}{(n+1)(n+2)}, 1 \leq r \leq n
 \end{aligned} \tag{2.24}$$

$$\begin{aligned}
 \text{Var}(U_{r:n}) &= E(U_{r:n})^2 - [E(U_{r:n})]^2 \\
 &= \frac{r(r+1)}{(n+1)(n+2)} - \left[\frac{r}{n+1} \right]^2 \\
 &= \frac{r(n-r+1)}{(n+1)^2(n+2)} - 1 \leq r \leq n
 \end{aligned} \tag{2.25}$$

2.21 Kovaryans

$$\begin{aligned}
 E(U_{r:n}^{m_r} U_{s:n}^{m_s}) &= \int_0^1 \int_0^1 U_{r:n}^{m_r} U_{s:n}^{m_s} f(U_r, U_s) dU_r dU_s \\
 &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \beta(r+m_r, s-r) \beta(s+m_r+m_s, n-s+1) \\
 &= \frac{n!(r+m_r-1)!(s+m_r+m_s-1)!}{(n+m_r+m_s)!(r-1)!(s+m_r-1)!}
 \end{aligned} \tag{2.26}$$

(2.26) denkleminde $m_r = m_s = 1$ alınırsa,

$$E(U_{r:n} U_{s:n}) = \frac{n!}{(n+2)!(r-1)!} \frac{r!}{s!} \frac{(s+1)!}{s!}$$

$$\frac{r(s+1)}{(n+1)(n+2)}, \quad 1 \leq r < s \leq n \quad (2.27)$$

$$Cov(U_{r:n} U_{s:n}) = E(U_{r:n} U_{s:n}) - E(U_{r:n})E(U_{s:n})$$

$$\begin{aligned} & \frac{r(s+1)}{(n+1)(n+2)} - \frac{r}{(n+1)} \frac{s}{(n+1)} \\ & \frac{r(n-s+1)}{(n+1)^2(n+2)} \end{aligned} \quad (2.28)$$

2.22 Korelasyon

$$\begin{aligned} Corr(U_{r:n} U_{s:n}) &= \frac{Cov(U_{r:n} U_{s:n})}{\sqrt{Var(U_{r:n})Var(U_{s:n})}} \\ &= \frac{r(n-s+1)}{(n+1)^2(n+2)} \frac{(n+1)^2(n+2)}{\sqrt{r(n-r+1)+s(n-s+1)}} \\ &= \left\{ \frac{r(n-s+1)}{s(n-r+1)} \right\}^{\frac{1}{2}} \end{aligned} \quad (2.29)$$

2.23 Standart Uniform Dağılım için Aralığın (R_n) Dağılımı

$$R_n = X_{n:n} - X_{1:n}$$

$X_{1:n}$: x değerlerini alır.

$X_{n:n}$: y değerlerini alır.

R_n : w değerlerini alır.

$$R_n = w = y - x$$

$$\begin{aligned} f_{X_{1:n}, X_{n:n}}(x, y) &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} [F(x)]^{r-1} [F(y) - F(x)]^{s-r-1} [1 - F(y)]^{n-s} f(x) f(y) \\ &= \frac{n!}{0!(n-2)!0!} [F(x)]^0 [F(w+x) - F(x)]^{n-2} [1 - F(w+x)]^0 f(x) f(w+x) \\ &= n(n-1) f(x) f(w+x) [F(w+x) - F(x)]^{n-2} \end{aligned} \quad (2.30)$$

$f(x)=1$ ve $F(x)=u$ olduğundan (2.30) denklemi;

$$f_{X_{1:n}, X_{n:n}}(x, w+x) = n(n-1)w^{n-2}, u \in [0, 1-w] \quad (2.31)$$

Aralığın (R_n) olasılık yoğunluk fonksiyonu standart uniform dağılım için;

$$\begin{aligned}
f(w) &= \int_0^{1-w} f_{X_{1:n}, X_{n:n}}(x, w+x) du \\
&= \int_0^{1-w} n(n-1)w^{n-2} du \\
&= n(n-1)w^{n-2} \int_0^{1-w} du \\
&= n - (n-1)w^{n-2}(1-w)
\end{aligned} \tag{2.32}$$

şeklinde bulunur.

Aralığın (R_n) birikimli dağılım fonksiyonu;

$$\begin{aligned}
F_{R_n}(r) &= \int_0^r n(n-1)w^{n-2}(1-w)dw \\
&= nr^{n-1} - (n-1)r^n, \quad 0 < r < 1
\end{aligned} \tag{2.33}$$

şeklinde bulunur.

2.24 Uniform Dağılım İçin Medyanın Dağılımı

Örneklem hacmi (n) 'in tek olduğu durumu göz önüne alalım;

$$f_{\bar{u}_n}(u) = \frac{n!}{\left[\left(\frac{n-1}{2}\right)!\right]^2} u^{\frac{n-1}{2}} \left[1-u\right]^{\frac{n-1}{2}}, \quad 0 \leq u \leq 1 \tag{2.34}$$

şeklinde bulunur.

Örneklem hacmi (n)'in çift olduğu durumu göz önüne alalım;

$$\begin{aligned}
f_{\bar{u}_n}(u) &= \frac{n!}{\left[\left(\frac{n}{2}-1\right)!\right]^2} \left[1-t\right]^{\frac{n}{2}-1} \frac{u^{\frac{n}{2}-1}}{\frac{n}{2}} \Bigg|_0^1 \\
&= \frac{2n!}{\left[\left(\frac{n}{2}-1\right)!\right]^2} \frac{2}{n} \left[1-t\right]^{\frac{n}{2}-1}
\end{aligned} \tag{2.35}$$

şeklinde bulunur.

2.25 Sıra İstatistiklerinin Üstel Dağılımı

X rassal değişkeni $\lambda > 0$ için üstel dağılıma sahiptir eğer olasılık yoğunluk fonksiyonu aşağıdaki gibiyse;

$$f(x, \lambda) = \lambda e^{-\lambda x}, \quad x > 0 \text{ ve}$$

$X \sim \text{Exp}(\lambda)$ şeklinde ifade edilir.

Bu çalışmada standart üstel dağılım ($\lambda = 1$) olduğu durumla

İlgileneceğiz x_1, x_2, \dots, x_n , hepsi birbirlerinden bağımsız ve aynı standart üstel

dağılıma sahip o.y.f; $f(x) = e^{-x}$, $x \geq 0$

olan rassal değişkenler olsun. $x_{1:n}, x_{2:n}, \dots, x_{n:n}$, X_1, X_2, \dots, X_n 'in sıralı değerleri olarak gösterilsin.

$$f_{r:n}(x) = \frac{n!}{(r-1)!(n-r)!} (F(x))^{r-1} (1-F(x))^{n-r} f(x), \quad -\infty < x < \infty$$

denklemden hareketle aşağıdaki denklemler elde edilir.

2.26 R. Sıra İstatistiği İçin Olasılık Yoğunluk Fonksiyonu

$$f_{r:n}(x) = \frac{n!}{(r-1)!(n-r)!} (1-e^{-x})^{r-1} e^{-(n-r+1)x}, \quad 0 \leq x \leq \infty \quad (2.36)$$

şeklinde yazılabilir.

2.27 En Küçük Sıra İstatistiğinin Olasılık Yoğunluk Fonksiyonu

$$\begin{aligned} f_{1:n}(x) &= \frac{n!}{(r-1)!(n-r)!} (1-e^{-x})^{1-1} e^{-(n-1+1)x} \\ &= \frac{n!}{(n-1)!} e^{-nx} \\ &= ne^{-nx}, \quad 0 \leq x \leq \infty \end{aligned} \quad (2.37)$$

2.28 En Küçük Sıra İstatistiğinin Birikimli Olasılık Fonksiyonu

$$\begin{aligned} f_{1:n}(x) &= 1 - [1 - F(X)]^n \\ &= 1 - [1 - (1 - e^{-x})]^n \\ &= 1 - [e^{-x}]^n \\ &= 1 - e^{-nx} \end{aligned} \quad (2.38)$$

şeklinde bulunur.

2.29 En Büyük Sıra İstatistiğinin Olasılık Yoğunluk Fonksiyonu

$$f_{n:n}(x) = \frac{n!}{(n-1)!(n-n)!} (1-e^{-x})^{n-1} e^{-(n-n+1)x}$$

$$= ne^{-x}(1-e^{-x})^{n-1}, \quad 0 \leq x < \infty \quad (2.39)$$

2.30 En Büyük Sıra İstatistiğinin Birikimli Olasılık Fonksiyonu

$$\begin{aligned} F_{n:n}(x) &= F(x)^n \\ &= (1-e^{-x})^n \\ &= \left(1 - \frac{ne^{-x}}{n}\right)^n \\ &= \exp(-ne^{-x}) \end{aligned} \quad (2.40)$$

2.31 $X_{i:n}$ ve $X_{j:n}$ 'in Ortak Olasılık Yoğunluk Fonksiyonu

$$\begin{aligned} f_{i,j:n}(x,y) &= \frac{n!}{(i-1)!(j-i-1)!(n-i)!} [F(x)]^{i-1} f(x) [F(y)-F(x)]^{j-i-1} f(y) [1-F(y)]^{n-j} \\ &= \frac{n!}{(i-1)!(j-i-1)!(n-i)!} (1-e^{-x})^{i-1} (e^{-y}-e^{-x})^{j-i-1} e^{-x} - e^{-(n-j+1)}, \quad 0 \leq x < y < \infty \end{aligned} \quad (2.41)$$

2.32 $X_{1:n}$ ve $X_{n:n}$ 'in Ortak Olasılık Yoğunluk Fonksiyonu

$$f_{1,n:n}(x,y) = n(n-1)e^{-(x+y)} [e^{-y} - e^{-x}]^{n-2} \quad (2.42)$$

şeklinde yazılabilir.

2.33 N Tane Sıra İstatistiğinin Ortak Dağılımı

$$\begin{aligned} f_{1,2,\dots,n:n}(x_1, x_2, \dots, x_n) &= n! f(x_1) f(x_2) \dots f(x_n) \\ &= n! \prod_{i=1}^n f(x_i) \\ &= n! \prod_{i=1}^n e^{-x_i} \\ &= n! e^{-\sum_{i=1}^n x_i}, \quad 0 \leq x_1 < x_2 < \dots < x_n < \infty \end{aligned} \quad (2.43)$$

olarak bulunur.

2.34 i. Sıra İstatistiğinin K. Dereceden Momenti

$$E(X_{i:n}^k) = \int_0^{\infty} x^k f_{i:n}(x) dx$$

$$\begin{aligned}
&= \frac{n!}{(i-1)!(n-i)!} \int_0^{\infty} x^k [F(x)]^{i-1} [1-F(x)]^{n-i} f(x) dx \\
&= \frac{n!}{(i-1)!(n-i)!} \int_0^{\infty} x^k [1-e^{-x}]^{i-1} [1-e^{-x}]^{n-i} e^{-x} dx \\
&= \frac{n!}{(i-1)!(n-i)!} \int_0^{\infty} x^k [1-e^{-x}]^{i-1} e^{-(n-i+1)x} dx \\
&= \frac{n!}{(i-1)!(n-i)!} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \int_0^{\infty} x^k e^{-(n-i+1)x} dx
\end{aligned}$$

Ayrıca şunu biliyoruz ki;

$$\int_0^{\infty} x^k e^{-x} dx = (n-i+j+1)^{-(k+1)} \int_0^{\infty} t^k e^{-t} dt = \frac{\Gamma(k+1)}{(n-i+j+1)^{(k+1)}}$$

şeklinde gösterilebilir. Buradan hareketle moment,

$$E(X_{i:n}^k) = \frac{n!}{(i-1)!(n-i)!} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \frac{\Gamma(k+1)}{(n-i+j+1)^{(k+1)}}, \quad i=1,2,\dots,k \geq 1 \quad (2.44)$$

ifade edilir.

2.35 Beklenen Değer

Teorem 2.1

$X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ standart üstel dağılıma sahip bir sıra istatistiği

olsun. Bu durumda rasgele değişkenler $Z_1, Z_2, \dots, Z_n, Z_i = (n-i+1)(X_{(i)} - X_{(i-1)})$,

$X_{(0)} \equiv 0$ bağımsız ve standart normal dağılıma sahip.

İspat:

$X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ 'in ortak olasılık yoğunluk fonksiyonunun;

$$f_{1,2,\dots,n:n}(x_1, x_2, \dots, x_n) = n! e^{-\sum_{i=1}^n x_i}, \quad 0 \leq x_1 < x_2 < \dots < x_n < \infty$$

Teoremi göz önüne aldığımızda;

$$X_{(1)} = \frac{Z_1}{n}, X_{(2)} = \frac{Z_1}{n} + \frac{Z_2}{n-1}, \dots, X_{(n)} = \frac{Z_1}{n} + \frac{Z_2}{n} + \dots + Z_n$$

şeklinde yazabiliriz.

Jakobiyen matrisi ise;

$$J(Z_1, Z_2, \dots, Z_n) = \begin{pmatrix} \frac{\partial X_{(1)}}{\partial Z_1} & \frac{\partial X_{(1)}}{\partial Z_2} & \dots & \frac{\partial X_{(1)}}{\partial Z_n} \\ \frac{\partial X_{(1)}}{\partial Z_1} & \frac{\partial X_{(2)}}{\partial Z_2} & \dots & \frac{\partial X_{(2)}}{\partial Z_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial X_{(n)}}{\partial Z_1} & \frac{\partial X_{(n)}}{\partial Z_2} & \dots & \frac{\partial X_{(n)}}{\partial Z_n} \end{pmatrix} = \begin{pmatrix} \frac{1}{n} & 0 & \dots & 0 \\ \frac{1}{n} & \frac{1}{n-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n-1} & \dots & 0 \end{pmatrix} = \frac{1}{n!}$$

şeklinde yazılabilir.

$$\sum_{i=1}^n x_i = \sum_{i=1}^n z_i \text{ olsun,}$$

Z_1, Z_2, \dots, Z_n 'in ortak olasılık yoğunluk fonksiyonu;

$$f_{Z_1, Z_2, \dots, Z_n}(Z_1, Z_2, \dots, Z_n) = |J| f_{Z_1, Z_2, \dots, Z_n}(Z_1, Z_2, \dots, Z_n)$$

$$= |J| n! f(Z_1) f(Z_2) \dots f(Z_n)$$

$$= |J| n! \prod_{i=1}^n e^{-z_i}$$

$$= |J| n! e^{-\sum_{i=1}^n z_i}$$

$$= \frac{1}{n!} n! e^{-\sum_{i=1}^n z_i}$$

$$= e^{-\sum_{i=1}^n z_i}, 0 \leq Z_1 < Z_2 < \dots < Z_n < \infty$$

Bu nedenle Z_i bağımsız ve standart üstel dağılım a sahiptir.

Yukarıdaki teoremden, eğer Z_1, Z_2, \dots, Z_n n tane bağımsız ve aynı dağılıma sahip standart üstel dağılım dan seçilmiş sıra istatistiği olsun. Bu durumda;

$$(Z_{1:n}, Z_{2:n}, \dots, Z_{n:n}) = \left(\frac{w_1}{n}, \frac{w_1}{n} + \frac{w_2}{n-1}, \dots, \frac{w_1}{n} + \frac{w_2}{n-1} + \dots + \frac{w_{n-1}}{2} + w_n \right)$$

w_1, w_2, \dots, w_n : bağımsız ve standart üstel dağılıma sahip.

$k=1$ alınarak;

$$\begin{aligned}
\mu_{i:n} &= E(Z_{i:n}) \\
&= E(Z_{(1:n)}, Z_{(2:n)}, \dots, Z_{(n:n)}) \\
&= E\left(\frac{w_1}{n} + \frac{w_2}{n-1} + \dots + \frac{w_i}{n-i+1}\right) \\
&= \frac{E(w_1)}{n} + \frac{E(w_2)}{n-1} + \dots + \frac{E(w_i)}{n-i+1} \\
&= \frac{1}{n} + \frac{1}{n-1} + \dots + \frac{1}{n-i+1} \\
&= \sum_{j=1}^i \frac{1}{n-i+1}, \quad 1 \leq i \leq n
\end{aligned} \tag{2.45}$$

Eğer w_i standart üstel dağılıma sahip ise $E(w_i) = Var(w_i) = 1$ olur. Bununla;

$$\begin{aligned}
Var(Z_{i:n}) &= Var\left(\frac{w_1}{n} + \frac{w_2}{n-1} + \dots + \frac{w_i}{n-i+1}\right) \\
&= \left(\frac{Var(w_1)}{n} + \frac{Var(w_2)}{n-1} + \dots + \frac{Var(w_i)}{n-i+1}\right) \\
&= \frac{1}{n^2} + \frac{1}{(n-1)^2} + \dots + \frac{1}{(n-i+1)^2} \\
&= \sum_{j=1}^i \frac{1}{(n-i+1)^2}, \quad 1 \leq i \leq n
\end{aligned} \tag{2.46}$$

bulunur.

2.36 Kovaryans

$$\begin{aligned} \text{Cov}(Z_{i:n}, Z_{s:n}) &= \sum_{j=1}^r \text{Var}\left(\frac{w_i}{n-i+1}\right) \\ &= \sum_{j=1}^r \frac{1}{(n-i+1)^2}, \quad r \leq s \end{aligned}$$

2.37 Korelasyon

$$\text{Cor}(Z_{r:n}, Z_{s:n}) = \frac{\sqrt{\sum_{i=1}^r \frac{1}{(n-i+1)^2}}}{\sqrt{\sum_{i=1}^s \frac{1}{(n-i+1)^2}}}, \quad 1 \leq r \leq s \quad (2.48)$$

2.38 Standart Üstel Dağılım İçin Range

$$R_n = X_{(n)} - X_{(1)}$$

$$X_{(i)} = 0 \quad \text{olsun;}$$

$$P_r(R_n < r) = P_r(n-1 < r)$$

$$\left(\int_0^r f(x) dx \right)^{n-1}$$

$$\left(\int_0^r e^{-x} dx \right)^{n-1}$$

$$= (1 - e^{-r})^{n-1} \quad (2.49)$$

Yukarıdaki (2.49) denkleminin türevini alarak;

$$f_{R_n}(r) = (n-1)e^{-r}(1-e^{-r})^{n-2}, \quad 0 \leq r < \infty \quad (2.50)$$

R_n 'in olasılık yoğunluk fonksiyonunu elde etmiş oluruz.

2.39 Standart Üstel Dağılım İçin Medyan

$$\text{Medyan} = \begin{cases} X_{(n+1)/2:n}, & n \text{ tek} \\ \frac{X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n}}{2}, & n \text{ çift} \end{cases}$$

Örneklem hacmi (n) 'in tek olduğu durumu göz önüne alalım;

$$\bar{X}_n = X_{\frac{n+1}{2}:n} \quad \text{diye tanımlansın,}$$

Standart üstel dağılım için medyanın olasılık yoğunluk fonksiyonu;

$$f_{\bar{X}_n}(x) = \frac{n!}{\left[\left(\frac{(n-1)}{2}\right)!\right]^2} \left[(1-e^{-x})(1-(1-e^{-x}))\right]^{\frac{n-1}{2}} e^{-x}, \quad 0 \leq x < \infty$$

$$= \frac{n!}{\left[\left(\frac{(n-1)}{2}\right)!\right]^2} (1-e^{-2x})^{\frac{n-1}{2}} e^{-x} \quad (2.51)$$

şeklinde bulunur

2.40 Sıra İstatistiklerinin Lojistik Dağılımı

x_1, x_2, \dots, x_n , hepsi birbirlerinden bağımsız ve aynı zamanda standart lojistik dağılıma sahip; olasılık yoğunluk fonksiyonu;

$$f(x) = \frac{e^{-x}}{(1+e^{-x})^2}, \quad -\infty < x < \infty$$

$$F(x) = \frac{1}{(1+e^{-x})}, \quad -\infty < x < \infty$$

olan rassal değişkenler olsun.

$X_{1:n}, X_{2:n}, \dots, X_{n:n}$, X_1, X_2, \dots, X_n 'in sıralı değerleri olarak gösterilsin.

2.41 i. Sıra İstatistiğinin Olasılık Yoğunluk Fonksiyonu

$$f_{i:n}(x) = \frac{n!}{(n-1)!(n-1)!} \left[F(x)^{i-1}\right] \left[1-F(x)\right]^{i-1} f(x)$$

denklemden hareketle i. sıra istatistiğinin o.y.f;

$$= \frac{n!}{(n-1)!(n-1)!} \left[(1+e^{-x})^{-1}\right] \left[1-(1+e^{-x})^{-1}\right]^{n-1} e^{-x} (1+e^{-x})^{-2}$$

$$= \frac{n!}{(n-1)!(n-1)!} \left[(1+e^{-x})^{-1}\right]^{i-1} \left[e^{-y} (1+e^{-x})^{-1}\right]^{n-i} e^{-x} (1+e^{-x})^{-2}, \quad -\infty < x < \infty \quad (2.52)$$

2.42 En Küçük Sıra İstatistiğinin Olasılık Yoğunluk Fonksiyonu

$$f_{1:n}(x) = \frac{n!}{(n-1)!(n-1)!} \left[(1+e^{-x})^{-1}\right]^{1-1} \left[1-(1+e^{-x})^{-1}\right]^{n-1} e^{-x} (1+e^{-x})^{-2}$$

$$\begin{aligned}
&= \frac{n!}{(n-1)!} \left[e^{-x}(1+e^{-x})^{-1} \right]^{n-1} e^{-x}(1+e^{-x})^{-2} \\
&= n \left[e^{-x}(1+e^{-x})^{-1} \right]^{n-1} e^{-x}(1+e^{-x})^{-2}, \quad -\infty < x < \infty
\end{aligned} \tag{2.53}$$

2.43 En Küçük Sıra İstatistiğinin Birikimli Olasılık Fonksiyonu

$$\begin{aligned}
F_{1:n}(x) &= 1 - [1 - F(x)]^n \\
&= 1 - \left[1 - (1 + e^{-x})^{-1} \right]^n \\
&= 1 - \left[e^{-nx}(1 + e^{-x})^{-n} \right] \\
&= \left[\frac{1}{1 + e^{-x}} \right]^n, \quad -\infty < x < \infty
\end{aligned} \tag{2.54}$$

2.44 En büyük Sıra İstatistiğinin Olasılık Yoğunluk Fonksiyonu

$$\begin{aligned}
f_{n:n}(x) &= \frac{n!}{(n-1)!(n-n)!} \left[(1 + e^{-x})^{-1} \right]^{n-1} \left[1 - (1 + e^{-x})^{-1} \right]^{n-n} e^{-x}(1 + e^{-x})^{-2} \\
&= \frac{n!}{(n-1)!} \left[(1 + e^{-x})^{-1} \right]^{n-1} e^{-x}(1 + e^{-x})^{-2} \\
&= n \left[(1 + e^{-x})^{-1} \right]^{n-1} e^{-x}(1 + e^{-x})^{-2}, \quad -\infty < x < \infty
\end{aligned} \tag{2.55}$$

2.45 En Büyük Sıra İstatistiğinin Birikimli Olasılık Fonksiyonu

$$\begin{aligned}
F_{n:n}(x) &= [F(x)]^n \\
&= \left[(1 + e^{-x})^{-1} \right]^n \\
&= (1 + e^{-x})^{-n}, \quad -\infty < x < \infty
\end{aligned} \tag{2.56}$$

2.46 i. ve j. Sıra İstatistiklerinin Ortak Olasılık Yoğunluk Fonksiyonu

$$\begin{aligned}
f_{i,j:n}(x, y) &= \frac{n!}{(i-1)!(j-i-1)!(n-i)!} \left[F(x) \right]^{i-1} \left[F(x_j) - F(x_i) \right]^{j-i-1} \left[1 - F(x_j) \right]^{n-j} f(x_i) f(x_j) \\
&= \frac{n!}{(i-1)!(j-i-1)!(n-i)!} \left[(1 + e^{-x_i})^{-1} \right]^{i-1} \left[(1 + e^{-x_j})^{-1} - (1 + e^{-x_i})^{-1} \right]^{j-i-1} \left[1 - (1 + e^{-x_j})^{-1} \right]^{n-j} e^{-x_i}(1 + e^{-x_i})^{-2} e^{-x_j}(1 + e^{-x_j})^{-2} \\
&(1 \leq i < j \leq n)
\end{aligned} \tag{2.57}$$

2.47 Moment

$$\begin{aligned}
M_{i:n}(t) &= E[e^{tX_{i:n}}] \\
&= \frac{1}{B(i, n-i+1)} \int_{-\infty}^{\infty} \frac{e^{(n-i+1)x+tx}}{(1+e^{-x})^{n+1}} dx \\
&= \frac{B(i+t, n-i+1-t)}{B(i, n-i+1)} \\
&= \frac{\Gamma(i+t) \Gamma(n-i+1-t)}{\Gamma(i) \Gamma(n-i+1)}, 1 \leq i \leq n
\end{aligned} \tag{2.58}$$

$$\begin{aligned}
K_{i:n}(t) &= \log M_{i:n}(t) \\
&= \log \left[\frac{\Gamma(i+t) \Gamma(n-i+1-t)}{\Gamma(i) \Gamma(n-i+1)} \right] \\
&= \log \Gamma(i+t) + \log \Gamma(n-i+1-t) - \log \Gamma(i) - \log \Gamma(n-i+1), 1 \leq i \leq n
\end{aligned} \tag{2.59}$$

yukarıdaki (2.59) eşitliğinden $X_{i:n}$ 'in m. dereceden kümülan fonksiyonu;

$$\begin{aligned}
K_{i:n}^{(m)} &= \frac{d^m}{dt^m} K_{i:n}(t) \Big|_{t=0} \\
&= \frac{d^m}{dt^m} [\log \Gamma(i+t) + \log \Gamma(n-i+1-t) - \log \Gamma(i) - \log \Gamma(n-i+1)] \\
&= \frac{d^m}{dt^m} \log \Gamma(i+t) \Big|_{t=0} + \frac{d^m}{dt^m} \log \Gamma(n-i+1-t) \Big|_{t=0} \\
&= \frac{d^m}{dt^m} \log \Gamma(i) + \frac{d^m}{dt^m} \log \Gamma(n-i+1) \\
&= \phi^{(m-1)}(i) + (-1)^m \phi^{(m-1)}(n-i-1)
\end{aligned} \tag{2.60}$$

$$\phi^{(m)}(.) = \frac{d^m}{dt^m} \log \Gamma(i) = \log \Gamma(.) \frac{\Gamma'(.)}{\Gamma(.)} \text{ m. dereceden polygamma fonksiyonuna denk}$$

gelmektedir.

2.48 Beklenen Değer ve Varyans

$$\begin{aligned}
\mu_{i:n} &= K_{i:n}^{(1)} \\
&= \phi^{(1-1)}(i) + (-1)^1 \phi^{(1-1)}(n-i-1) \\
&= \phi(i) - \phi(n-i-1) \\
\sigma_{i:n} &= K_{i:n}^{(2)} \\
&= \phi^{(2-1)}(i) + (-1)^2 \phi^{(2-1)}(n-i-1)
\end{aligned} \tag{2.61}$$

$$\phi^{(1)}(i) + \phi^{(1)}(n-i-1) \quad (2.62)$$

ϕ ve $\phi^{(1)}$ sırasıyla di-gamma ve tri-gamma fonksiyonlarına karşılık gelmektedir.

2.49 Lojistik Dağılım için Range

$$R_n = X_{n:n} - X_{1:n}$$

$$F_{R_n}(r) = p(R_n \leq r)$$

$$= n \int_{-\infty}^{\infty} [F(x+r) - F(x)]^{n-1} f(x) dx, \quad 0 \leq r < \infty$$

$$[F(x+r) - F(x)]^{n-1} = \sum_{i=0}^{n-1} (-1)^i \binom{n-1}{i} [F(x+r)]^{n-1-i} [F(x)]^i$$

Bundan dolayı;

$$F_{R_n}(r) = p(R_n \leq r)$$

$$= n \sum_{i=0}^{n-1} (-1)^i \binom{n-1}{i} \int_{-\infty}^{\infty} [F(x+r)]^{n-1-i} [F(x)]^i f(x) dx$$

$$= n \sum_{i=0}^{n-1} (-1)^i \binom{n-1}{i} \int_{-\infty}^{\infty} \frac{e^{-x}}{(1+e^{-r}e^{-x})^{n-1-i} (1+e^{-x})^{i+2}} dx$$

$$= \frac{1}{(1+e^{-r}e^{-x})} \text{ dersek;}$$

$$F_{R_n}(r) = p(R_n \leq r)$$

$$= n \sum_{i=0}^{n-1} (-1)^i \binom{n-1}{i} e^{-(i+1)} \Delta_{i:n}(r), \quad 0 \leq r < \infty \quad (2.63)$$

$$m = e^{-r} - 1$$

$$= \Delta_{i:n}(r) = \int_0^1 s^{n-1} (1+ms)^{-i-2} ds$$

$$= \frac{1}{(-m)^n} \left[(-1)^{i+1} \binom{n-1}{i+1} kn(1+m) + \sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k} \frac{1}{(k-i-1)} \{(1+m)^{k-i-1} - 1\} \right]$$

$$= \frac{1}{(1-e^{-r})^n} \left[(-1)^i \binom{n-1}{i+1} r + \sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k} \frac{1}{(k-i-1)} \{(1+m)^{-(k-i-1)r} - 1\} \right]$$

$i > 2$ ise $\binom{n-1}{i+1} \rightarrow 0$ ile gösterilir.

Bulduğumuz $\Delta_{i:n}(r)$ denklemini (2.63) denkleminde yerine yazarsak ;

Kümülatif olasılık fonksiyonu;

$$F_{R_n}(r) = p(R_n \leq r)$$

$$= \frac{n}{(1-e^{-r})^n} \sum_{i=0}^{n-1} (-1)^i \binom{n-1}{i} \left[(-1)^i \binom{n-1}{i+1} r e^{-(i+1)r} + \sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k} \frac{1}{(k-i-1)} \{e^{-kr} - e^{-(i+1)r}\} \right], 0 \leq r < \infty \quad (2.64)$$

bulunur ve bulmuş olduğumuz bu denklemin r'ye göre türevi alınırsa;

$$= f_{R_n}(r) \frac{n^2 e^{-r}}{(1-e^{-r})^{n+1}} \sum_{k=0}^{n-1} (-1)^i \binom{n-1}{i} \left[(-1)^i \binom{n-1}{i+1} r e^{-(i+1)r} \right]$$

$$+ \sum (-1)^k \binom{n-1}{k} \frac{1}{(k-i-1)} \{e^{-kr} - e^{-(i+1)r}\}$$

$$+ \frac{n}{(1-e^{-r})^n} \sum_{i=0}^{n-1} (-1)^i \binom{n-1}{i} \left[(-1)^i \binom{n-1}{i} - e^{-(i+1)r} \{1-(i+1)r\} \right]$$

$$- \sum (-1)^k \binom{n-1}{k} \frac{1}{(k-i-1)} \{k e^{-kr} - (i+1) e^{-(i+1)r}\}, 0 \leq r < \infty \quad (2.65)$$

$$i > n-2 \text{ ise } \binom{n-1}{i+1} \rightarrow 0$$

range için olasılık yoğunluk fonksiyonu elde edilmiş olur.

2.50 Medyan

$M = X_{\frac{n+1}{2}:n}$ tanımlansın;

$$f_M(x) = \frac{n!}{\left(\frac{n+1}{2}-1\right)! \left(n-\frac{n+1}{2}\right)!} \left[F(x)^{\frac{n+1}{2}} \right] \left[(1-F(x))^{n-\frac{n+1}{2}} f(x) \right]$$

$$= \frac{n!}{\left[\left(\frac{n+1}{2}-1\right)!\right]^2} \left[F(x) (1-F(x))^{\frac{n+1}{2}} \right] f(x)$$

f(x) ve F(x) yukarıdaki denklemde yerine yazılırsa;

$$f_{\bar{x}_n}(x) = \frac{n!}{\left[\left(\frac{n-1}{2}\right)!\right]^2} e^{-x} (1+e^{-x})^{-2} \left[(1+e^{-x})^{-1} \{1-(1+e^{-x})^{-1}\} \right]^{\frac{n-1}{2}}$$

$$= \frac{n!}{\left[\left(\frac{(n-1)}{2} \right)! \right]^2} e^{-x(1+e^{-x})^{-2}} \left[e^{-x(1+e^{-x})^{-2}} \right]^{\frac{n-1}{2}} \quad (2.66)$$

2.51 Sıra İstatistiklerinin Normal Dağılımı

Standart normal dağılımın genel o.y.f;

$$f(x) = \frac{e^{-(x-\mu)/(2\sigma^2)}}{\sigma\sqrt{2\pi}}$$

μ :konum parametresi

σ :ölçek parametresi

olarak tanımlanır ancak biz bu bölümde sadece standart normal dağılım ile ilgileneceğiz. $\mu = 0$ ve $\sigma = 1$ alınarak standart normal dağılımın olasılık yoğunluk fonksiyonu ve birikimli olasılık dağılım fonksiyonu;

$$f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

$$F(x) = \int_{-\infty}^0 \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

olarak yazılır.

X_1, X_2, \dots, X_n , $N(0,1)$ dağılımına sahip bağımsız rassal değişkenler olmak üzere

$X_{1:n} \leq X_{2:n} \leq \dots X_{n:n}$ normal dağılımına sahip sıra istatistiğine karşılık gelsin.

$$E(X_{k:n}^k) = \frac{n!}{(k-1)!(n-k)!} x^r \phi^{k-1}(x)(1-\phi(x))^{n-k} \phi(x) dx \quad (2.67)$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2 / 2)$$

$$\phi(x) = \int_{-\infty}^x \phi(t) dt$$

(2.67) denklemini çözebilmek için birtakım nümerik metodlar vardır.

Normal dağılımda; lojistik, uniform ve üstel dağılımdakinin aksine sadece küçük n değerleri için aşikar denklemler elde edilebilir.

n=3 için beklenen değer;

$$E[X_{3:3}] = 3 \int_{-\infty}^{\infty} x \phi^2(x) \phi(x) dx$$

$$= 3 \int_{-\infty}^{\infty} \phi^2(x) d\varphi(x)$$

$$= 6 \int_{-\infty}^{\infty} \phi(x) \varphi^2(x) dx$$

$$I(a) = \int_{-\infty}^{\infty} \phi(ax) \varphi^2(x) dx$$

$$I(0) = \frac{1}{2} \int_{-\infty}^{\infty} \phi^2(x) dx = \frac{1}{4\pi} \int_{-\infty}^{\infty} \exp(-x^2) dx$$

$$I'(a) = \int_{-\infty}^{\infty} x \phi(ax) \varphi^2(x) dx = -\frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} x \exp\{-x^2(a^2 + 2)\} dx = 0$$

$$I(a) = \frac{1}{4\sqrt{\pi}}$$

$$\int_{-\infty}^{\infty} \phi(ax) \varphi^2(x) dx = I(1) \frac{1}{4\sqrt{\pi}}$$

$$E[X_{3;3}] = 6I(1) = \frac{3}{2\sqrt{\pi}}$$

şeklinde bulunur.

$N(0,1)$ dağılımının simetri özelliğinden dolayı ;

$$E[X_{1;3}]^2 = E[X_{3;3}]^2$$

$$Var(X_{1;3}) = Var(X_{3;3})$$

olduğunu biliyoruz.

$$E[X_{3;3}]^2 = 3 \int_{-\infty}^{\infty} x^2 \phi^2(x) \varphi(x) dx = -3 \int_{-\infty}^{\infty} x \phi^2(x) d\varphi(x)$$

$$= 3 \int_{-\infty}^{\infty} \varphi(x) d(x\phi^2(x)) = 3 \int_{-\infty}^{\infty} \varphi(x) \phi^2(x) dx + 6 \int_{-\infty}^{\infty} x \phi^2(x) \phi(x) dx$$

$$= \int_{-\infty}^{\infty} d(\phi^3(x)) + \frac{3}{\pi} \int_{-\infty}^{\infty} x \exp(-x^2) \phi(x) dx$$

$$= \frac{3}{2\pi} \int_{-\infty}^{\infty} \Phi(x) d \exp(-x^2)$$

$$= 1 + \frac{3}{2\pi} \int_{-\infty}^{\infty} \exp(-x^2) \varphi(x) dx = 1 + \frac{3}{(2\pi)^{3/2}} \int_{-\infty}^{\infty} \exp(-3x^2/2) dx$$

$$= 1 + \frac{\sqrt{3}}{2\pi}$$

$Var(X_{3,3}) = E(X_{3,3})^2 - E([X_{3,3}])^2$ eşitliğinden hareketle;

$$= 1 + \frac{\sqrt{3}}{2\pi} - \frac{9}{4\pi}$$

olarak bulunur.



3

Bilgi Teorisi ve Entropi

3.1 Entropi Tanım

Bir fiziksel sistemdeki belirsizliğin ya da rasgeleliğin ölçüsü, entropi olarak adlandırılır. Entropi kavramını ilk kez istatistikte kullanan kişi Shannon olmuştur.

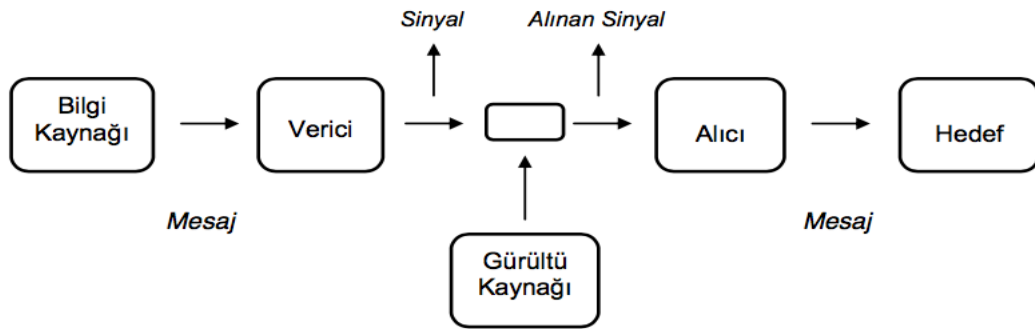
Entropinin günlük hayatımızda birçok uygulama alanı vardır. Aslında termodinamik kavramının olduğu her yerde entropi kavramı da bulunur. Entropi bir özelliktir, bu yüzden her nesne veya her şey entropiye sahiptir.

Evren iki veya daha fazla şey arasındaki ısı değişimi ile varlığını sürdürmektedir. Örneğin; güneşin dünyaya ve diğer gezegenlere ısı aktarması güneş sisteminin düzenli bir şekilde işlemesine sebep olmaktadır. Isı değişimine dünyadan daha basit bir örnek verecek olursak; kullandığımız taşıtlar, herhangi kullandığımız

makineler veya suyun kaynaması vs. Entropinin fiziksel ve istatistiksel kavramlarına daha detaylı olarak değineceğiz.

Bilgi Teorisi:

Shannon-Weaver modeli; Enformasyon(Bilgi)Teorisi olarak da isimlendirilmektedir. Bu modeli, 1949 yılında Claude Elwood Shannon (1916-2001) ve Warren Weaver (1894-1978) birlikte bulmuşlardır. Bilgi kuramının amacı herhangi bir bilgiyi elde etmek, aktarmak, işlemek ve saklamak gibi basamaklara ilişkin kuralları incelemektir.



Şekil 3. 1 Shannon'a göre genel olarak bir iletişim sistemi

Kaynak; seçtiği mesajı bir araç yardımıyla iletir. Bu araç iletici veya göndericidir.

Verici, kaynak tarafından üretilen mesajı, iletilebilecek elemandır.

Sinyal, mesajın kanalda iletilebilme biçimidir.

Kanal, sinyalin vericiden alıcıya iletildiği ortamdır.

Alıcı, sinyali yeniden mesaj sekline dönüştürerek vericinin yaptığı işlemin tersini yapan elemandır.

Hedef, mesajın ulaştığı son noktadır.

Shannon bilgi kuramını iletişim alanı için geliştirmek isterken bilgi teorisi ve entropi kavramları pek çok bilim dalında uygulamalı olarak kullanım yerleri bulmuştur.

3.2 Entropinin Fiziksel Anlamı

Termodinamik'teki entropi kavramının olasılık ile harmanlayan ilk bilim insanlarından biri L. Boltzman'dır. Bir gazın sahip olduğu serbestlik derecesi, gazın belirsizliğidir diyen Boltzman, fizikte entropiyi ve olasılık kavramlarını tekrar

gündeme getirerek, termodinamikteki düzensizlik denklemini ortaya atmıştır. Bu denklem;

$S = k \ln w$ şeklinde ifade edilir ve buradaki $k=1,3806505 \times 10^{-23} \text{ J/K}$ Boltzmann sabiti olarak isimlendirilir.

S : Entropi

k : Boltzman değışmezi

W : Termodinamik olasılık

Avusturyalı fizikçi Ludwig Boltzmann, ne yazık ki, yıllar boyu birçok bilim adamının eleştirilerine karşı mücadele etmiş ve buluşu dikkate alınmamıştır ancak buluşu günümüzde evrenin en temel yasalarından biri olarak bahsedilmektedir.

Suyun üç hali olan katı, sıvı ve gaz hallerinin her biri farklı entropilere sahiptir. Buzun molekülleri daha düzenli ve aralarında daha az boşluk olduğundan en az entropiye sahiptir. Sıvı hal, katı hale oranla nispeten molekülleri arasındaki boşluk daha fazla olduğu için entropisi daha yüksek ve molekülleri arasındaki boşluk en fazla olan yani en yüksek entropiye sahip olan gaz halidir.

3.3 Entropinin İstatistiksel Anlamı

Örneğin bir zar ve para atma deneyini kıyaslayacak olursak, zar atma deneyinin sonuçları her bir yüz için 1/6 iken para atma deneyinin sonuçları 1/2 olur. Yani para atma deneyinde zar atma deneyine oranla daha fazla bilgi olduğu için entropisi daha düşüktür.

Entropi kavramını daha somut bir şekilde örnekleyecek olursak;

1.kitaplıkta özdeş (aynı) 4 matematik kitabı olsun.

2.kitaplıkta 3 özdeş (aynı) matematik ve 1 fen kitabı olsun.

3.kitaplıkta 2 özdeş matematik ve 2 özdeş fen kitabı olsun.

Kitaplıktaki kitapların farklı dizilişlerinin sayısı diler bir adla permütasyonları aşağıdaki gibi olacaktır.

1. kitaplık: MMMM

2. kitaplık: MMMF, MMFM, MFMM, SMMM

3. kitaplık: MMFF, MFMF, FMFM, FFMM, FMMF, MFFM

Gördüğümüz gibi 1.kitaplıkta sadece bir farklı yolla kitaplar dizilebilirken, ikinci kitaplık için bu sayı 4 ve 3. kitaplık için ise 6 farklı yolla kitapların dizilebileceğini bulduk.

Buradan çıkan bazı sonuçlar;

Entropi ve bilgi birbirinin tersidir.

Eğer kitapların farklı dizilişlerinin sayısı fazla ise entropi yüksek aksine kitapların olası diziliş sayıları az ise entropi düşüktür.

Yukarıda diziliş(permutasyon) ile entropi arasındaki ilişkiyi vermeye çalıştım, birde olayı olasılık ve entropi ilişkisi yönünden inceleyelim.

1.kitaplıktan geri konulmak şartıyla seçilen 4 kitabında matematik kitabı olma olasılığı 1'dir.

2.kitaplıktan geri konulmak şartıyla seçilen 4 kitabın da istediğimiz dizilişlerden biri şeklinde olma olasılığı; $p=0,75*0,75*0,75*0,25$ yani 0,105 olarak bulunur.

3.kitaplıktan geri konulmak şartıyla seçilen 4 kitabında istediğimiz dizilişlerde olma olasılığı; $p=0,5*0,5*0,5*0,5$ yani 0,0625 olarak bulunur.

Entropi çeşitleri olarak;

Shannon Entropisi

Renyi Entropisi

Havrda-Charvat ve Tsallis Entropileri

Bileşik Entropi

Koşullu Entropi

Görelî Entropi

Bulanık Entropi

örnek olarak verilebilir. Ancak biz sadece en çok bilinenler üzerinde yani ilk ucu üzerinde duracağız.

3.4 Shannon Entropisi

X kesikli rasgele değişken ve x_1, x_2, \dots, x_n , sırasıyla p_1, p_2, \dots, p_n olasılık değerlerini alsın.

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

X kesikli rasgele değişkeninin Shannon entropi değeri;

$$H(X) = -\sum p(x_i) \log_2 p(x_i) = -E[\log_2 x_i] = -E(I_x) \quad (3.1)$$

şeklinde hesaplanır.

X sürekli rasgele değişken ise;

$$\forall x \in (-\infty, \infty) f(x) \geq 0 \text{ ve } \int_{-\infty}^{\infty} f(x) = 1$$

X sürekli rasgele değişkeninin entropi değeri;

$$H(X) = -\int p(x_i) \log_2 p(x) dx = -E[\log_2 f(x)] = -E(I_x) \quad (3.2)$$

şeklinde hesaplanır.

Her iki durum için de X'ten elde edilen bilgi;

$$I_x = -\log_2 p_i \text{ ya da } I_x = -\log_2 f(x)$$

şeklinde gösterilir. Burada $E(I_x), I_x$ 'in beklenen değerini gösterir ki entropinin bir adı da ortalama belirsizliktir.

H(x) entropi fonksiyonu en büyük değerini $p_i = \frac{1}{n}$ (bütün i değerleri için) olduğu zaman alır.

3.5 Shannon Entropisinin Özellikleri

Non-negatif bir fonksiyondur. $H(p) \geq 0$

$H(p_1, p_2, \dots, p_n)$ bütün p_k değerleri için süreklidir.

$H(p_1, p_2, \dots, p_n)$ bütün p_k değerleri için simetriktir.

E_k olaylarının olasılık değerleri eşit olduğunda maksimum entropi gerçekleşmiş olur.

$P(X = X_i) = \frac{1}{K} i = 1, 2, \dots, K$ ve bu duruma karşılık gelen maksimum entropi

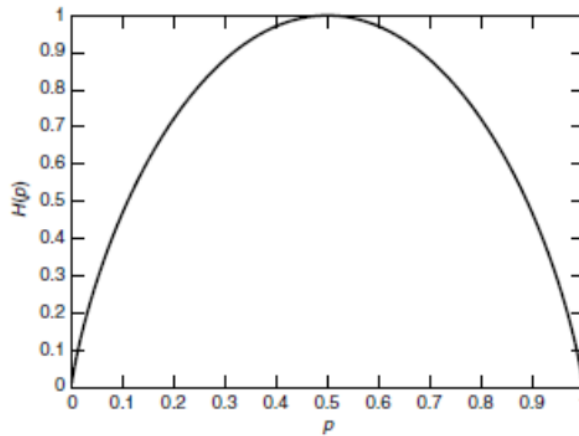
$$P(X) = \sum_{i=1}^K \frac{1}{K} \log \left(\frac{1}{K} \right) = \log K \text{ değerine eşit olacaktır.}$$

$$H(0,1) = H(1,0) = 0$$

3.6 Maksimum Entropi İlkesi

Kesikli hallerde bütün durumlar eş olasılıklı ($p_i = \frac{1}{K} i = 1, 2, \dots, K$) ise, entropi maksimumdur. Maksimum entropi düşüncesi ilk olarak Jaynes (1957) tarafından

geliştirilmiştir. Bu ilke Laplace'ın ünlü yetersiz akıl yürütme ilkesi (principle of insufficient reasoning) ile özdeştir. Genellikle Bayesçi bağlamda kullanılmaktadır. Posterior dağılımların belirlenmesinde, prior dağılımlar hakkında bilgi olmadığı ya da az bilgi olduğu hallerde, bu ilke ile bazı olasılık dağılımları prior dağılım olarak belirlenmektedir. Sürekli hallerde bunlar temelde uniform, üstel ve normal dağılımlar olmaktadır. Maksimum entropi varsayımı altında ve olasılık dağılımlarına belirli kısıtlamaların getirilmesi halinde, bu üç dağılımın nasıl elde edilebileceği konusunda Reza(1994)'ya bakılabilir.



Şekil 3. 2 Maksimum entropi grafiği

Yukarıdaki grafikten hareketle;

Fonksiyonun pozitif değerler aldığı ve konkav bir fonksiyon olduğu görülür.

Olasılıkların birbirine eşit ve 0.5 olduğu durumda maksimum değerini alır

Olasılıklar 1 ve 0 olduğunda entropi değeri 0'dır.

3.7 Renyi Entropisi

Alfred Renyi 1961 yılında Renyi entropisini tanımlamıştır. Renyi entropisi de biyolojik çeşitliliğin derecesinin ortaya konmasında kullanılmaktadır. Ayrıca Kuantum bilgi ölçüsü olarak kullanımı da önemlidir. Renyi entropisi Shannon entropisinin $\alpha \rightarrow 1$ yaklaşırken özel bir halidir. Yani Renyi entropileri özel bir durum olarak Shannon entropilerini içerir.

$$\lim_{\alpha \rightarrow 1} H_{\alpha}(X) = \lim_{\alpha \rightarrow 1} \frac{1}{1 - \alpha} \log \sum_{k=1}^N P_k^{\alpha} = \frac{\lim_{\alpha \rightarrow 1} \sum_{k=1}^N \log P_k P_k^{\alpha} / \sum_{k=1}^N P_k^{\alpha-1}}{\lim_{\alpha \rightarrow 1} \sum_{k=1}^N P_k^{\alpha-1}} = H_s(X)$$

3.8 Renyi Entropisinin Özellikleri

$H_\alpha(x)$ non negatiftir : $H_\alpha(x) \geq 0$

$$H_\alpha(0,1) = H_\alpha(0,1)$$

$\alpha \leq 1$ için Renyi entropisi konkav ancak $\alpha > 1$ için Renyi entropisi ne tamamen konkav ne de tamamen konvektir.

$H_\alpha(x)$ sınırlı, sürekli ve α değerleri için artmayan bir fonksiyondur.

Renyi entropileri farklı α değerleri için korelasyonludur.

3.9 Havrda-Charvat ve Tsallis Entropileri

Shannon entropisinin bir diğer genellemesi Constantino Tsallis(1988)'a aittir. Her ne kadar Havrda-Charvat(1967) tarafından benzer bir formülasyon daha eski tarihten bile Tsallis entropisi daha popüler olmuştur. Aslında Havrda-Charvat ve Tsallis entropilerinin formülasyonları arasında küçük bir fark vardır. Havrda-Charvat entropisi şu şekilde tanımlanmıştır:

$$H_{HC} = \frac{2^{\alpha-1}}{2^{\alpha-1}-1} \left[1 - \sum_{i=1}^K p_i^\alpha \right] \alpha > 0 \text{ ve } \alpha \neq 1 \quad (3.3)$$

Tsallis entropisi aynı zamanda β -sınıfı entropi olarak da bilinmektedir. Tsallis entropisi

$$H_T = \frac{1 - \sum_{i=1}^K p_i^\alpha}{\alpha - 1}, \alpha > 0 \text{ ve } \alpha \neq 1 \quad (3.4)$$

şeklinde tanımlanır. L'Hospital Kuralına göre

$$\lim_{\alpha \rightarrow 1} \frac{1 - \sum_{i=1}^K p_i^\alpha}{\alpha - 1} = \lim_{\alpha \rightarrow 1} \frac{\frac{d}{d\alpha}(1 - \sum_{i=1}^K p_i^\alpha)}{\frac{d}{d\alpha}(\alpha - 1)} = \lim_{\alpha \rightarrow 1} \frac{-\sum_{i=1}^K p_i^\alpha \ln p_i}{1} = -\sum_{i=1}^K p_i \ln p_i \quad (3.5)$$

Başka bir deyişle Tsallis entropisi $\alpha \rightarrow 1$ için Shannon entropisine yaklaşmaktadır. Çeşitli entropi ölçülerinin asimptotik özellikleri Pardo(2006) ve Zhang(2013) tarafından verilmektedir. Yine kesikli hallerde bu üç entropi ölçüsünün asimptotik normalliği üzerine karşılaştırmalı bir simülasyon çalışması için Evren ve Ustaoglu(2015)'e bakılabilir.

3.10 Diferansiyel Entropiler

Sürekli bir dağılımın entropisi diferansiyel entropi olarak adlandırılır. Sürekli hallerde verilen tanımlayıcı formüllerdeki toplama sembolleri integral sembolleri ile yer değiştirmektedir. Kesikli hallerden farklı olarak sürekli haller için

hesaplanan diferansiyel entropilerden bir kısmının sıfır ya da negatif olabileceğini, çünkü olasılık yoğunluklarının birden büyük olabileceğini not etmek gerekir.

3.11 Uniform Dağılımın Shannon Entropisi

Uniform dağılımın olasılık yoğunluk ve dağılım fonksiyonları aşağıdaki gibidir:

$$f(x) = \frac{1}{b-a} \quad a \leq x \leq b$$

$$F(x) = \int_{-\infty}^x f(t)dt = \frac{x-a}{b-a} \quad a \leq x \leq b$$

Bu dağılımın Shannon entropisi de şu şekilde bulunur:

$$H_{SH}(x) = -E(\ln(f(x))) = \ln(b-a)$$

Bu dağılımın varyansı ise $\sigma^2 = \frac{(b-a)^2}{12}$ olur.

Bu sebepten,

$$(b-a) = 2\sqrt{3}\sigma$$

$$H_{SH}(x) = \ln(2\sqrt{3}\sigma) \quad \text{bulunur.}$$

3.12 Üstel Dağılımın Shannon Entropisi

Üstel dağılımın olasılık yoğunluk ve dağılım fonksiyonları

$$f(x) = \lambda \exp(-\lambda x) \quad x \geq 0$$

$$F(x) = \int_{-\infty}^x f(t)dt = 1 - \exp(-\lambda x) \quad x \geq 0$$

şeklindedir. Shannon entropisi;

$$H_{SH}(x) = -E(\ln(f(x))) = 1 - \ln(\lambda)$$

(3.8) Ve bu dağılımın varyansı $\sigma^2 = \frac{1}{\lambda^2}$ olur. Böylelikle;

$$H_{SH}(x) = 1 + \ln(\sigma)$$

bulunur.

3.13 Normal Dağılımın Shannon Entropisi

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad -\infty < x < \infty$$

$$F(x) = \int_{-\infty}^x f(t)dt = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} dt$$

$$H_{SH}(x) = -E(\ln(f(x))) = \ln(\sigma\sqrt{2e\pi})$$

Uniform, üstel ve normal dağılımların Shannon entropileri incelendiğinde ortak bir özellik olarak her üçünün de standart sapmalarının logaritmik bir fonksiyonu oldukları görülmektedir.

3.14 Teorem

X olasılık yoğunluk fonksiyonu f olan sürekli bir rastlantı değişkeni olsun.

$Y = g(X)$ rastlantı değişkeni X 'in tanım aralığında monoton (artan ya da azalan) bir fonksiyon ise Y için o.y.f ;

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dx}{dy} \right| \quad (3.11)$$

yardımı ile bulunur (Mood ve ark., s200)

3.15 Olasılık İntegral Dönüşümü

Özel olarak Y, X 'in dağılım fonksiyonu olarak seçilsin.

Bu durumda $Y = F(X)$, dolayısıyla

$$\frac{dy}{dx} = f_X(x) \text{ ve } f_Y(y) = f_X(x) \left| \frac{1}{\frac{dy}{dx}} \right| = 1$$

Buradan $U = F(Y) \sim Uniform(0; 1)$.

Benzer şekilde $V = (1 - F(Y)) \sim Uniform(0; 1)$

$W = g(U) = \ln(U)$ olsun.

$$f_W(w) = f_U(g^{-1}(w)) \left| \frac{du}{dw} \right| ; f_U(u) = 1 \quad 0 < u < 1$$

$$f_U(g^{-1}(w)) = 1; U = \exp(W) \quad \frac{du}{dw} = \exp(W)$$

$$f_W(w) = \exp(w) \quad -\infty < w < 0$$

$$E(W) = \int_{-\infty}^0 w \exp(w) dw = \int_{-\infty}^0 x \exp(x) dx$$

Parçalara ayırarak integral alma tekniğine göre

$$u = x \Rightarrow du = dx; dv = \exp(x) dx \Rightarrow v = \exp(x)$$

$$\int_a^b uv dx = uv \Big|_a^b - \int_a^b v du$$

$$\int_{-\infty}^0 x \exp(x) dx = x \exp(x) \Big|_{-\infty}^0 - \int_{-\infty}^0 \exp(x) dx = -1.$$

Dolayısıyla;

$$E(\ln(F(Y))) = E(\ln(1 - (F(Y)))) = -1$$

ve böylelikle;

$$H_{SH}(y_\alpha) = -\ln(\alpha) - \ln\binom{n}{\alpha} + (n-1) + H_{SH}(y)$$

Uniform dağılım için;

$$H_{SH}(y_\alpha) = -\ln(\alpha) - \ln\binom{n}{\alpha} + (n-1) + \ln(b-a)$$

Üstel dağılım için;

$$H_{SH}(y_\alpha) = -\ln(\alpha) - \ln\binom{n}{\alpha} + n - \ln(\lambda)$$

Normal dağılım için

$$H_{SH}(y_\alpha) = -\ln(\alpha) - \ln\binom{n}{\alpha} + (n-1) + \ln(\sigma\sqrt{2\pi e})$$

olarak elde edilir.

3.16 Üstel Dağılım

$f(x) = \lambda e^{-\lambda x}$ ve $F(x) = 1 - e^{-\lambda x}$, $x > 0$, $\lambda > 0$ olmak üzere,

üstel dağılımın Shannon entropisi;

$$-E[\ln f(x)] = -E[\ln \lambda - \lambda x]$$

$$= -\ln \lambda + \lambda E(x) = -\ln \lambda + \lambda \frac{1}{\lambda} = 1 - \ln \lambda \quad (3.16)$$

bulunur. Sıra istatistiğinin ise

$$f_{y_\alpha}(y) = \alpha \binom{n}{\alpha} [F(y)]^{\alpha-1} [1-F(y)]^{n-\alpha} f(y) \quad (3.17)$$

olduğunu biliyoruz.

(3.17) denkleminin logaritması alınır;

$$\begin{aligned} \ln f_{y_\alpha}(y) &= \ln \left(\alpha \binom{n}{\alpha} [1 - e^{-\lambda y}]^{\alpha-1} [e^{-\lambda y}]^{n-\alpha} \lambda e^{-\lambda y} \right) \\ &= \ln \left(\alpha \binom{n}{\alpha} + (\alpha-1) \ln[1 - e^{-\lambda y}] + (n-\alpha) \ln(e^{-\lambda y}) + \ln \lambda e^{-\lambda y} \right) \end{aligned} \quad (3.18)$$

$$-E[\ln f_{y_\alpha}(y)] = -\ln \left(\alpha \binom{n}{\alpha} - (\alpha-1) E[\ln(1 - e^{-\lambda y})] - (n-\alpha) E[\ln(e^{-\lambda y})] - E[\ln \lambda e^{-\lambda y}] \right) \quad (3.19)$$

(3.19) denkleminde;

$$-E[\ln f_{y_\alpha}(y)] = 1 - \ln \lambda \text{ yazılırsa;}$$

$$-E[\ln f_{y_\alpha}(y)] = -\ln \left(\alpha \binom{n}{\alpha} - (\alpha-1) E[\ln(1 - e^{-\lambda y})] - (n-\alpha) E[-\lambda y] + (1 - \ln \lambda) \right)$$

$$-E[\ln f_{y_\alpha}(y)] = -\ln \left(\alpha \binom{n}{\alpha} - (\alpha-1) E[\ln(1 - e^{-\lambda y})] - (n-\alpha)(-\lambda) E[y] + (1 - \ln \lambda) \right) \quad (3.20)$$

ve (3.20) denkleminde $E[y] = \frac{1}{y}$ yazılırsa ;

$$\begin{aligned}
 -E[\ln f_{y_\alpha}(y)] &= -\ln\left(\alpha \binom{n}{\alpha}\right) - (\alpha - 1)E[\ln(1 - e^{-\lambda y})] - (n - \alpha)(-\lambda)\left(\frac{1}{y}\right) + (1 - \ln \lambda) \\
 -E[\ln f_{y_\alpha}(y)] &= -\ln\left(\alpha \binom{n}{\alpha}\right) - (\alpha - 1)E[\ln(1 - e^{-\lambda y})] + (n - \alpha) + (1 - \ln \lambda)
 \end{aligned} \tag{3.21}$$

sonucu elde edilir.

(3.21) denkleminde $E[\ln(1 - e^{-\lambda y})]$ ifadesini bulmak için;

$\ln(1 - e^{-\lambda y})$ ifadesini Taylor serisine açalım;

$$f(y) = f(a) + f'(a)(x - a) + f''(a)\frac{(x - a)^2}{2!}$$

$$f'(y) = \frac{\lambda e^{-\lambda y}}{(1 - e^{-\lambda y})}$$

$$f''(y) = \frac{\lambda e^{-\lambda y}}{(1 - e^{-\lambda y})^2}$$

$$E[f(y)] = f(a) + f'(a)E(x - a) + f''(a)\frac{E(x - a)^2}{2!}$$

$y = \frac{1}{\lambda}$ ortalama değer etrafında açarsak;

$$E[f(y)] = f(\mu) + (a)\frac{\text{Var}(x)f''(\mu)}{2}$$

$$f(\mu) = \ln[1 - e^{-\lambda \frac{1}{\lambda}}] = \ln[1 - e^{-1}] = \ln[1 - \frac{1}{e}] = \ln(e - 1) - \ln e$$

$$f(\mu) = \ln(e - 1) - 1$$

$$f''(y) = \frac{-\lambda^2 e^{-\lambda y}}{(1 - e^{-\lambda y})^2} \text{ olduğundan,}$$

$$f''(y) = \frac{\lambda^2 e^{-\lambda \frac{1}{\lambda}}}{(1 - e^{-\lambda \frac{1}{\lambda}})^2} = \frac{-\lambda^2 e^{-1}}{(1 - e^{-1})^2}$$

$$E[\ln(1 - e^{-\lambda y})] \approx \ln(e - 1) - 1 + \frac{\text{Var}(x)}{2} - \frac{\lambda^2 e^{-1}}{(1 - e^{-1})^2}$$

$$\approx \ln(e - 1) - 1 - \frac{e^{-1}}{2(1 - e^{-1})^2}$$

$$\left[1 - \frac{1}{e}\right]^2 = \frac{(e-1)^2}{e^2} \text{ olduğundan;}$$

$$\approx \ln(e-1) - 1 - \frac{e}{(e-1)^2}$$

Bu durumda sıra istatistiklerinin entropisi;

$$-E[\ln f_{y_\alpha}(y)] = -\ln\left(\alpha \binom{n}{\alpha}\right) - (\alpha-1)\left[\ln(e-1) - 1 - \frac{e}{(e-1)^2}\right] + (n-\alpha) + (1-\ln \lambda) \quad (3.22)$$

şeklinde yaklaşık değeri bulunur.

3.17 Uniform Dağılım

Sıra istatistiklerinin o.y.f.;

$$f_{y_\alpha}(y) = \frac{n!}{(\alpha-1)!(n-\alpha)!} [F(y)]^{\alpha-1} [1-F(y)]^{n-\alpha} f(y)$$

$$f_{y_\alpha}(y) = \alpha \binom{n}{\alpha} [F(y)]^{\alpha-1} [1-F(y)]^{n-\alpha} f(y)$$

şeklinde yazıldığını biliyoruz. Bu eşitliğin ln'i alınırsa;

$$-\ln f_{y_\alpha}(y) = \ln\left(\alpha \binom{n}{\alpha}\right) + (\alpha-1)\ln F(y) + (n-\alpha)\ln[1-F(y)] + \ln f(y) \quad (3.23)$$

$$f(y) = 1, \quad 0 < y < 1$$

$$F(y) = y, \quad 0 < y < 1$$

$$\ln F(y) = \ln y \quad (3.24)$$

$$\ln[1-F(y)] = \ln(1-y) \quad (3.25)$$

(3.24) ve (3.25) eşitlikleri (3.23) denkleminde yerine konulursa;

$$\begin{aligned} -E[\ln f_{y_\alpha}(y)] &= -\ln\left(\alpha \binom{n}{\alpha}\right) - (\alpha-1)E[\ln F(y)] - (n-\alpha)E[\ln(1-F(y))] - E[\ln f(y)] \\ &= -\ln\left(\alpha \binom{n}{\alpha}\right) - (\alpha-1)E[\ln y] - (n-\alpha)E[\ln(1-y)] - E[\ln 1] \end{aligned} \quad (3.26)$$

$$\ln f(y) = \ln 1 = 0 \quad (3.27)$$

$$E[\ln F(y)] = 0 \quad (3.28)$$

$$E[\ln y] = \int_0^1 \ln y dy \quad v = u, \quad du = \frac{1}{y} dy$$

$$\begin{aligned}
\int_0^1 u dv &= uv \Big|_0^1 - \int_0^1 v du \\
&= (\ln y) y \Big|_0^1 - \int_0^1 y \frac{1}{y} dy \\
&= 1 \ln 1 - 0 \ln 0 - \int_0^1 dy = 1
\end{aligned}$$

$$E[\ln y] = -1 \tag{3.29}$$

$$E[\ln(1 - F(y))] = E[\ln(1 - y)]$$

$$\int_0^1 \ln(1 - y) dy \quad x = 1 - y, \quad dx = -dy$$

$$\int_0^1 \ln x dx = -(x \ln x - x) \Big|_0^1 = 1$$

$$E[\ln(1 - y)] = 1 \tag{3.30}$$

Uniform dağılım için sıra istatistiklerinin Shannon entropisi;

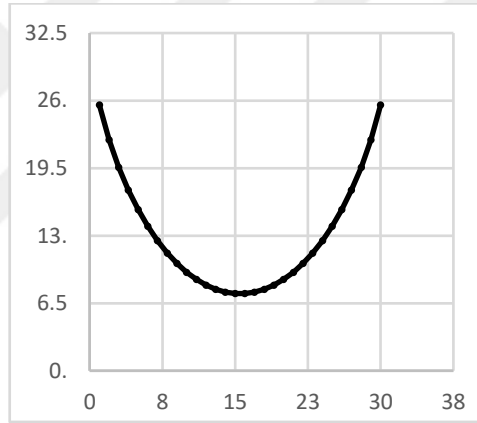
$$-E[\ln f_{y_\alpha}(y)] = -\ln\left(\alpha \binom{n}{\alpha}\right) + (\alpha - 1) - (n - \alpha) \text{ bulunur.} \tag{3.31}$$

4

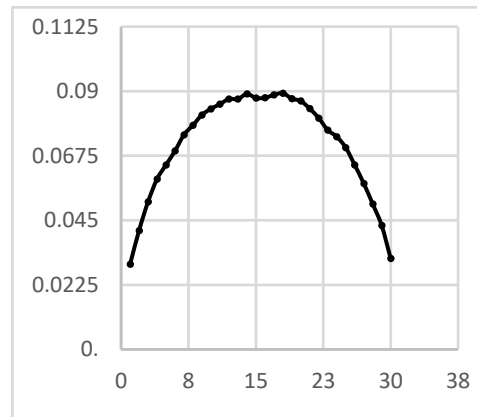
Uygulama

Uniform, üstel ve normal dağılımlardan $n=30$, $n=50$ ve $n=100$ birimlik rassal örnekler çekilmiş ve bu işlemler 1000 kere tekrarlanmıştır. Simülasyonlar Microsoft Excel yardımı ile gerçekleştirilmiştir. Daha sonra sıra istatistiklerinin Shannon entropileri hesaplanmıştır. Bu değerler sıra istatistiklerinin entropileri ile kıyaslanmıştır. Yapılan 9 simülasyon çalışmasının sonuçları aşağıdaki grafikler yardımı ile özetlenmiştir.

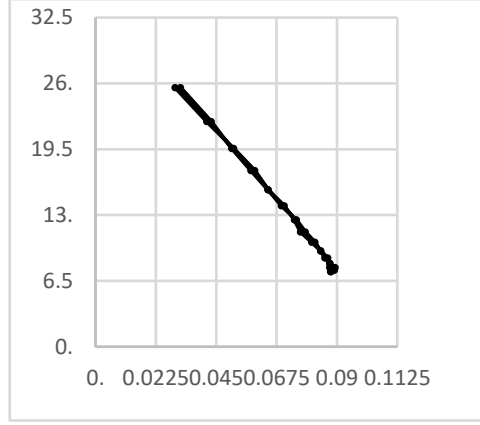
a) Dağılım: Uniform, örnek büyüklüğü=30, simülasyon sayısı=1000



Şekil 4. 1 Sıra değerine karşı entropi

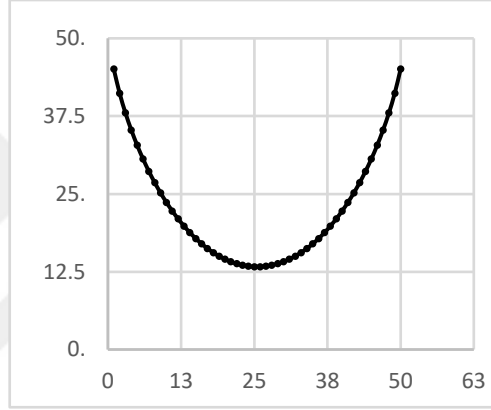


Şekil 4. 2 Sıra değerine karşı s.sapma

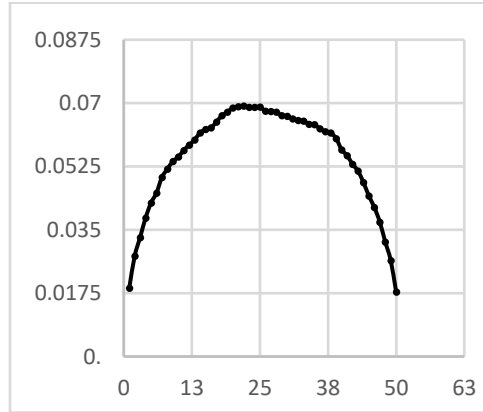


Şekil 4. 3 Std.sapmaya karşı entropi

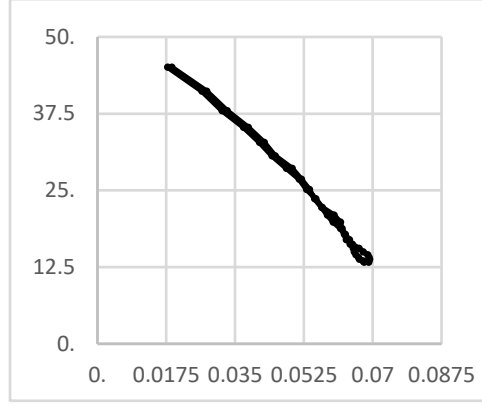
b) Dağılım: Uniform, örnek büyüklüğü=50, simülasyon sayısı=1000



Şekil 4. 4 Sıra değerine karşı entropi

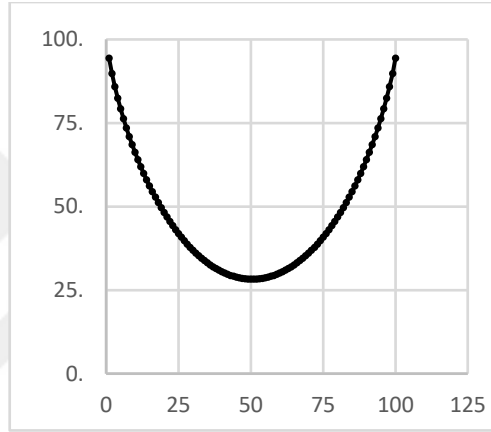


Şekil 4. 5 Sıra değerine karşı s.sapma

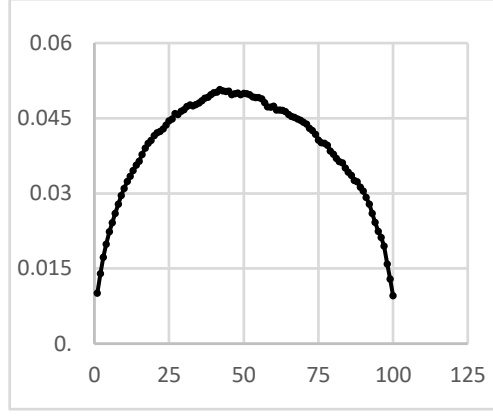


Şekil 4. 6 Std.sapmaya karşı entropi

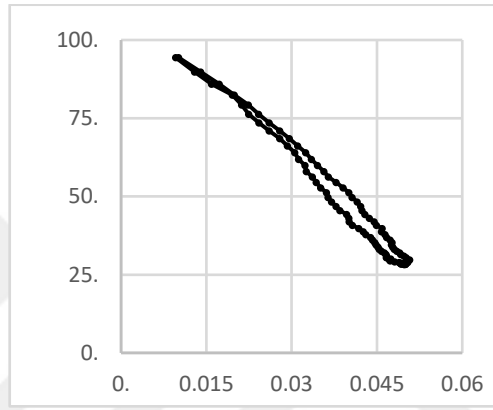
c) Dağılım: Uniform, örnek büyüklüğü=100, simülasyon sayısı=1000



Şekil 4. 7 Sıra değerine karşı entropi

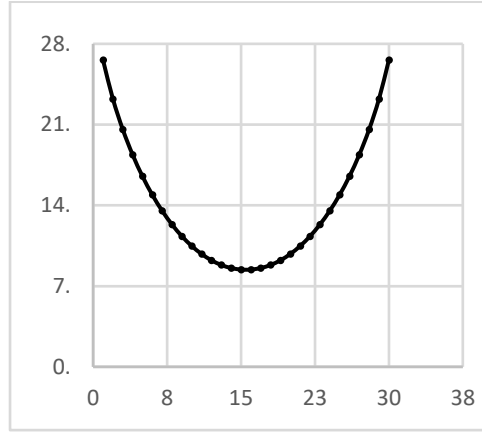


Şekil 4. 8 Sıra değerine karşı s.sapma

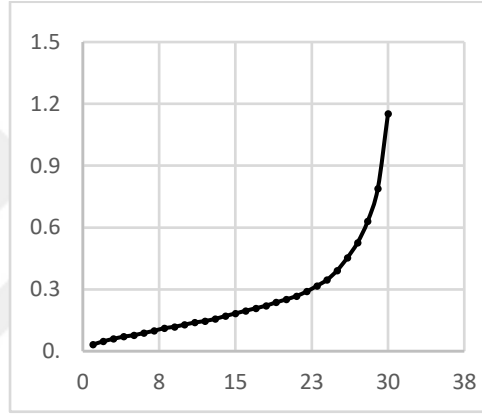


Şekil 4. 9 Std.sapmaya karşı entropi

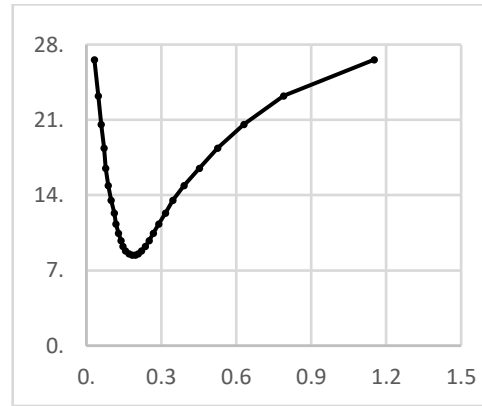
d) Dağılım: Üstel, örnek büyüklüğü=30, simülasyon sayısı=1000



Şekil 4.10 Sıra değerine karşı entropi

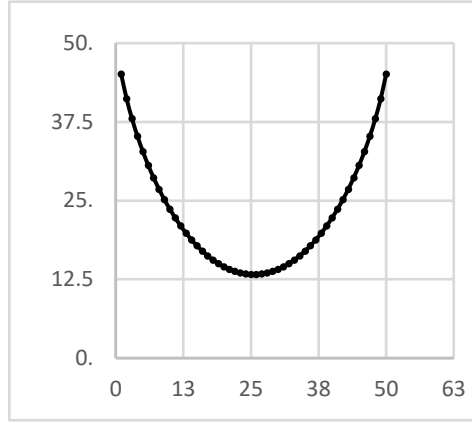


Şekil 4.11 Sıra değerine karşı s.sapma

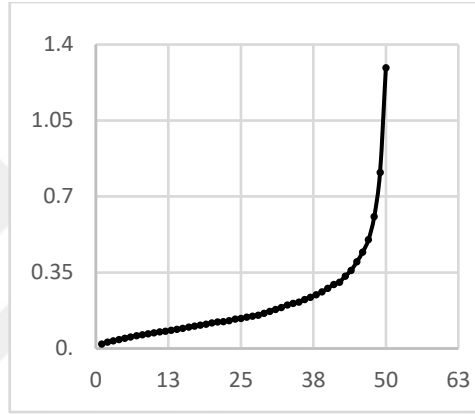


Şekil 4.12 Std.sapmaya karşı entropi

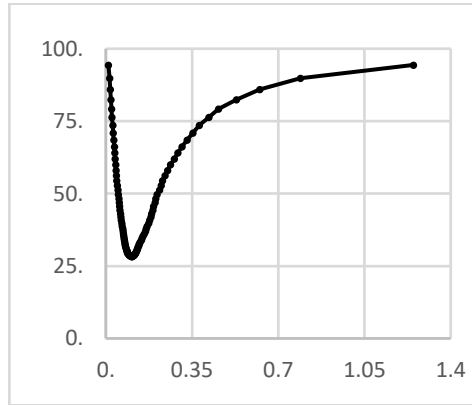
e) Dağılım: Üstel; örnek büyüklüğü=50, simülasyon sayısı=1000



Şekil 4.13 Sıra değerine karşı entropi

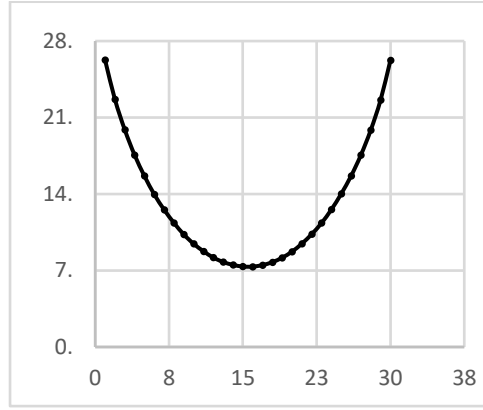


Şekil 4.14 Sıra değerine karşı s.sapma

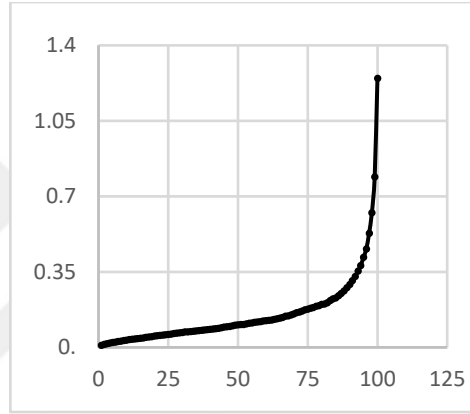


Şekil 4.15 Std.sapmaya karşı entropi

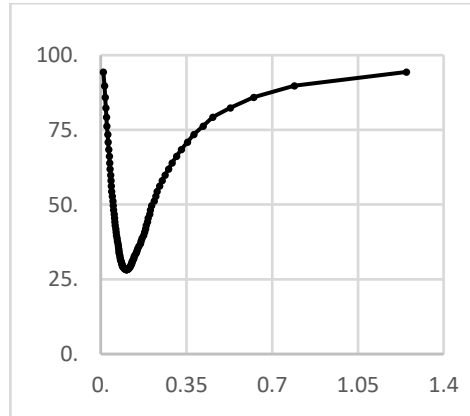
f) Dağılım: Üstel; örnek büyüklüğü=100, simülasyon sayısı=1000



Şekil 4. 16 Sıra değerine karşı entropi

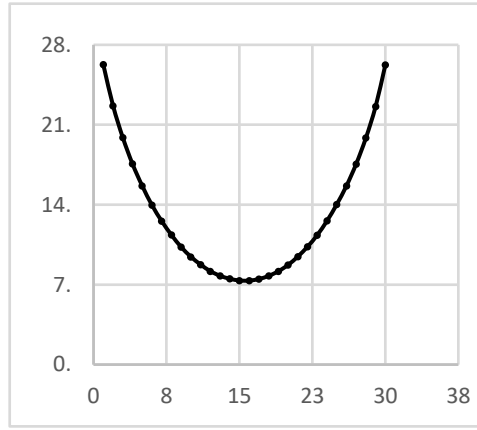


Şekil 4. 17 Sıra değerine karşı s.sapma

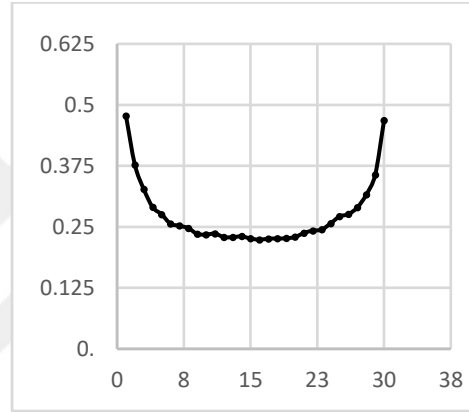


Şekil 4. 18 Std.sapmaya karşı entropi

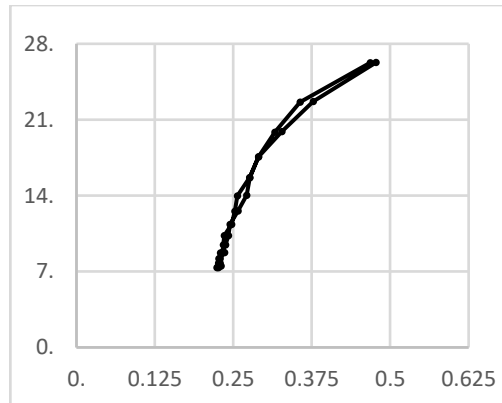
g) Dağılım: Normal; örnek büyüklüğü=30, simülasyon sayısı=1000



Şekil 4. 19 Sıra değerine karşı entropi

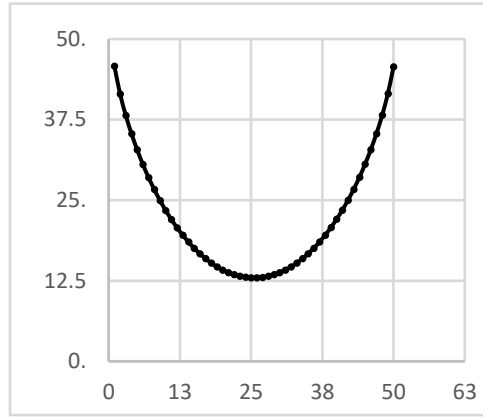


Şekil 4. 20 Sıra değerine karşı s.sapma

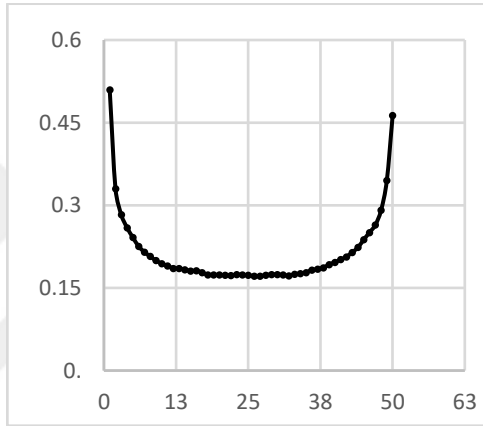


Şekil 4. 21 Std.sapmaya karşı entropi

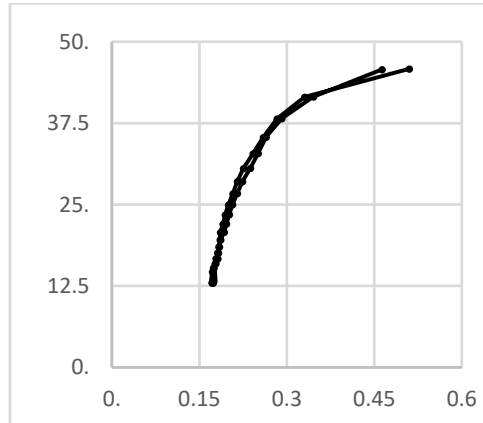
h) Dağılım: Normal; örnek büyüklüğü=50, simülasyon sayısı=1000



Şekil 4.22 Sıra değerine karşı entropi

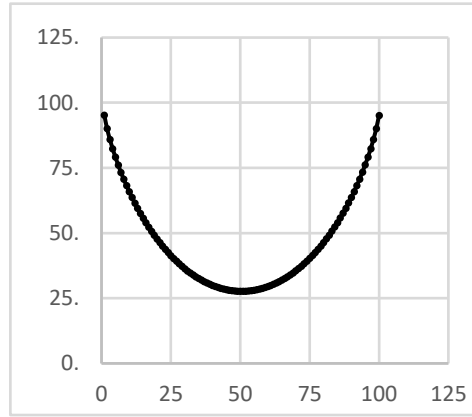


Şekil 4.23 Sıra değerine karşı s.sapma

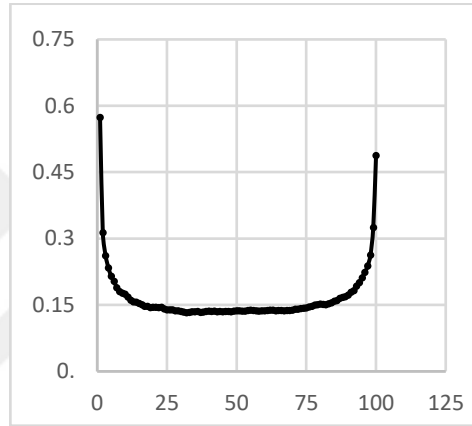


Şekil 4.24 Std.sapmaya karşı entropi

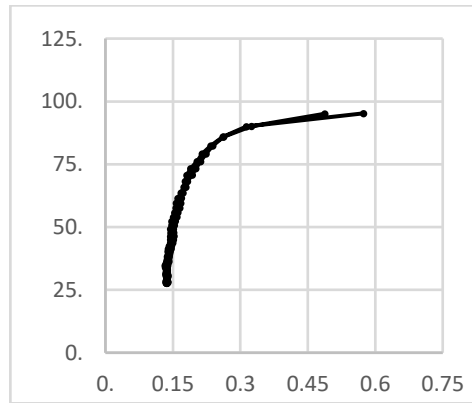
i) Dağılım: Normal; örnek büyüklüğü=100, simülasyon sayısı=1000



Şekil 4. 25 Sıra değerine karşı entropi



Şekil 4. 26 Sıra değerine karşı s.sapma



Şekil 4. 27 Std.sapmaya karşı entropi

Nominal ve ordinal dağılımlar için standart sapma hesaplanamaz. Bu gibi durumlarda değişkenliği ölçmenin tek yolu sıklıklara dayalı ölçüler kullanmaktır. Bu bağlamda çeşitli entropi formülü nominal ve ordinal dağılımlardaki değişkenliği ölçmek için önerilmiştir.

Sürekli verideki değişkenliğin ölçülmesinde varyans, standart sapma gibi klasik ölçünün yanısıra çeşitli entropi ölçüleri de önerilmiştir. Bununla birlikte kesikli hallerden farklı olarak bazı beklenmeyen durumlar ortaya çıkmaktadır. Uniform dağılımda olduğu gibi diferansiyel entropi sıfır olabilmektedir. Bazı dağılımlar için negatif entropi değerleri elde edilmektedir. Bu gibi durumlarda istatistiksel değişkenliğin analizinde entropi ölçülerinin kullanımı yararsız ve hatta yanlış yönlendirici olabilmektedir.

Uygulamada uniform, üstel ve normal dağılımların sıra istatistiklerinin Shannon entropileri ele alınmıştır. Genel olarak Shannon entropileri standart sapmaların logaritmik fonksiyonlarıdır. Bu bakımdan ön sezisel olarak standart sapmalar ile Shannon entropileri arasında pozitif korelasyon olması doğaldır. Uygulamada yapılan dokuz simülasyonda bu bağlamda pozitif korelasyonlar beklenmiştir. Gerçekleştirilen dokuz simülasyondan elde edilen bulgular Şekil 4.1'den Şekil 4.27'ye kadar özetlenmektedir.

Yapılan incelemede sınırlandırmanın standart sapma ile entropi arasında görmeyi umduğumuz pozitif korelasyonu bozucu etkisi olduğu ortaya çıkmıştır. Uniform dağılım her iki yönden sınırlandırılmış bir dağılımdır. Sıra istatistiklerinin standart sapmaları ile entropileri arasında bu sebepten pozitif korelasyonlar gözlenmemiştir. Üstel dağılım için durum bir miktar düzelmiştir. Bunun sebebi tahminen üstel dağılımın tek taraftan (soldan) sınırlandırılmış olmasıdır. Şekil 4.12, Şekil 4.15 ve Şekil 4.18'e bakıldığında sınırlandırmanın olduğu sol tarafta sıra istatistiklerinin entropisi ile standart sapması arasında ters yönlü, sağ tarafa gittikçe ise (sınırlandırmanın olmaması nedeni ile) pozitif yönlü bir ilişki olduğu gözlenmektedir. Bu kanı normal dağılımın sıra istatistiklerinin entropileri ile

standart sapmaları kıyaslandığında pekişmektedir. Çünkü normal dağılım sınırlandırılmamış bir dağılımdır ve dolayısıyla uç noktalardaki en küçük ve en büyük sıra istatistikleri için hesaplanan entropiler ile standart sapmalar pozitif korelasyonludur. Bu durum Şekil 4.21, Şekil 4.24 ve Şekil 4.27'de açıkça gözlenebilmektedir. Normal dağılım gibi sınırlı olmayan dağılımlar için entropi ve standart sapma önsel olarak beklendiği gibi pozitif korelasyonludur. Dolayısıyla sınırlandırma durumunda entropi istatistiklerinin belirsizliğin ya da değişkenliğin ölçümünde kullanılması uygun olmamaktadır.



- [1] D. Morales ve M. D. Esteban, "A summary on entropy statistics," *Kybernetika*, vol. 31, no.4, pp. 337-346, 1995.
- [2] A. Evren ve E. Ustaoglu, "On Asymptotic Normality of Entropy Measures," *International Journal of Applied Science and Technology*, vol. 5, no.5, pp. 31-38, 2015.
- [3] A. Golan, "Information and Entropy Econometrics," Now Publishers Inc., vol.2 no.1-2, pp.1-145, 2006.
- [4] J. Havrda ve F. Charvat, "Quantification Method in Classification Processes: Concept of Structural α -entropy," *Kibernetika*, vol.3, no.1, pp.30-35, 1967.
- [5] E.T. Jaynes, "Information Theory and Statistical Mechanics," *The Physical Review*, vol. 108, no.2, pp.171-190, 1957.
- [6] E.T. Jaynes, *Probability Theory: The Logic of Science*. UK: Cambridge University Press, 2005.
- [7] S. Kullback, *Information Theory and Statistics*. NY: Dover Publications, 1996.
- [8] A. M. Mood, F. A. Graybill, D. Boes, *Introduction to the Theory of Statistics*, third ed. New York: McGraw-Hill International Editions, 1974.
- [9] L. Pardo, *Statistical Inference Measures Based on Divergence Measures*, New York: Chapman&Hall/CRC, 2006.
- [10] A. Rényi, *On Measures of Entropy and Information*, California: University of California Press, 1961.
- [11] F. M. Reza, *An Introduction to Information Theory*, New York: Dover Publications, 1994.
- [12] S. M. Ross, *Simulation*, Third edition, Academic Press, 2002.
- [13] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *Journal of Statistical Physics*, vol.52, no.1-2, pp.479, 1988.
- [14] A. Ullah, "Entropy, Divergence and Distance Measures with Econometric Applications," *Journal of Statistical Planning and Inference*, vol.49, no.1, pp.137-162, 1996.
- [15] Z. Xing, "Asymptotic Normality of Entropy Estimators," *Doktora Tezi, Matematik Bölümü*, University of North Carolina, Charlotte, 2013.
- [16] M. Cover ve A. Thomas, *Elements of Information Theory*, 2006.
- [17] S. Shanti ve N. Balakrishnan, "Estimation of the Mean and Standard Deviation of the Logistic Distribution Based on Multiply Type-II Censored Samples," *A Journal of Theoretical and Applied Statistics*, vol.27, no.1-2, pp.127-142, 1995.
- [18] O. Omondi, "Order Statistics of Uniform, Logistic and Exponential

- Distributions,” Yüksek Lisans Tezi, İstatistik Bölümü, University of Nairobi, Nairobi, 2014.
- [19] B. Gümüş “Sıra İstatistiklerinin Denizcilik Sektöründe Uygulanması,” Yüksek Lisans Tezi, İstatistik Bölümü, Yıldız Teknik Üniversitesi, İstanbul, 2016.
- [20] İ. Değirmenci, “Entropi Ölçüleri ve Maksimum Entropi Prensibi,” Yüksek Lisans Tezi, İstatistik Bölümü, Hacettepe Üniversitesi, İstanbul, 2011.
- [21] P. Bromiley, N. Thacker, E. Bouhova-Thacker, “Shannon Entropy, Renyi Entropy, and Information,2010, <http://www.tina-vision.net/docs/memos/2004-004.pdf>
- [22] M. Ahsanullah ve V. Nevzorov, ve M. Shakil, “An Introduction to Order Statistics,” Atlantis Press, 2013.
- [23] M. Ahsanullah ve M.Z. Raqab, Bounds and Characterizations of Record Statistics, New York: Nova Science Publishers, 2009.



İletişim Bilgisi: hayatkiliç25@gmail.com

Makaleler

1. H. Kılıç ve A. Evren, “Comparing Shannon entropies and standard deviations of the order statistics for uncensored, semicensored, and censored distributions,” International Journal of Engineering and Applied Sciences, vol. 5, no. 10, pp.106-111, 2018.

