

KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ ANABİLİM DALI

ÇOKGENSEL ALANDA İKİ DEĞİŞKENLİ SÜREKLİ DAĞILIM
FONKSİYONLARI İÇİN UYUM İYİLİĞİ TESTLERİ

YÜKSEK LİSANS TEZİ

Tuncay ULUYURT

HAZİRAN 2015
TRABZON



**KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ ANABİLİM DALI

**ÇOKGENSEL ALANDA İKİ DEĞİŞKENLİ SÜREKLİ DAĞILIM
FONKSİYONLARI İÇİN UYUM İYİLİĞİ TESTLERİ**

Tuncay ULUYURT

**Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsünde
"YÜKSEK LİSANS (İSTATİSTİK)"
Unvanı Verilmesi İçin Kabul Edilen Tezdir.**

**Tezin Enstitüye Verildiği Tarih : 08/05/2015
Tezin Savunma Tarihi : 12/06/2015**

Tez Danışmanı : Yrd. Doç. Dr. Orhan KESEMEN

Trabzon 2015

Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü
İstatistik ve Bilgisayar Bilimleri Anabilim Dalında
Tuncay ULUYURT tarafından hazırlanan

ÇOKGENSEL ALANDA İKİ DEĞİŞKENLİ SÜREKLİ DAĞILIM
FONKSİYONLARI İÇİN UYUM İYİLİĞİ TESTLERİ

başlıklı bu çalışma, Enstitü Yönetim Kurulunun 12 / 06 / 2015 gün ve 1604 sayılı
kararıyla oluşturulan jüri tarafından yapılan sınavda
YÜKSEK LİSANS TEZİ
olarak kabul edilmiştir.

Jüri Üyeleri

Başkan : Prof.Dr. Tahir HANALIOĞLU

Üye : Yrd.Doç.Dr. Halil İbrahim ŞAHİN

Üye : Yrd.Doç.Dr. Orhan KESEMEN

Prof. Dr. Sadettin KORKMAZ
Enstitü Müdürü

ÖNSÖZ

“Çokgensel Alanda İki Değişkenli Sürekli Dağılım Fonksiyonları İçin Uyum İyiliği Testleri” isimli bu tez Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü İstatistik ve Bilgisayar Bilimleri Anabilim Dalı, Yüksek Lisans Programı’nda hazırlanmıştır.

Tez çalışma süresinde değerli yardım ve katkılarıyla beni yönlendiren benden yardımlarını, desteğini, sabrını ve bilgisini esirgemeyen değerli danışman hocam Yrd. Doç. Dr. Orhan KESEMEN’e, lisans eğitimimden bu yana bilgilerinden faydalanmamı sağlayan bütün bölüm hocalarıma, araştırma görevlisi arkadaşlarıma teşekkürü bir borç bilirim.

Son olarak, tüm hayatım boyunca maddi ve manevi her zaman beni destekleyen, her adımında arkamda duran anneme ve eşime sonsuz teşekkürlerimi sunarım.

Bu tezin, bundan sonraki çalışmalara katkı sağlamasını temenni ederim.

Tuncay ULUYURT
Trabzon 2015

TEZ ETİK BEYANNAMESİ

Yüksek Lisans Tezi olarak sunduđum “Çokgensel Alanda İki Deđişkenli Sürekli Dağılım Fonksiyonları İçin Uyum İyiliđi Testleri” başlıklı bu çalışmayı baştan sona kadar danışmanım Yrd. Doç. Dr. Orhan KESEMEN’in sorumluluđunda tamamladıđımı, verileri/örnekleri kendim topladıđımı, deneyleri/analizleri ilgili laboratuvarlarda yaptıđımı/yaptırdıđımı, başka kaynaklardan aldıđım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiđimi, çalışma sürecinde bilimsel araştırma ve etik kurallara uygun olarak davrandıđımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiđimi beyan ederim. 08/05/2015

Tuncay ULUYURT

İÇİNDEKİLER

	<u>Sayfa No</u>
ÖNSÖZ	III
TEZ ETİK BEYANNAMESİ.....	IV
İÇİNDEKİLER.....	V
ÖZET	VII
SUMMARY	VIII
ŞEKİLLER DİZİNİ.....	IX
TABLolar DİZİNİ.....	XII
SEMBOLLER DİZİNİ	XIII
1. GENEL BİLGİLER.....	1
1.1. Giriş	1
1.2. Olasılık Fonksiyonları	1
1.2.1. Olasılık Yoğunluk Fonksiyonu.....	1
1.2.2. Sıklık Göstergesi (Histogram)	4
1.2.3. Birikimli Dağılım Fonksiyonu.....	5
1.2.4. Birikimli Dağılım Fonksiyonun Sayısal Yöntemlerle Bulunması.....	6
1.3. Uyum İyiliği Testleri	8
1.3.1. Klasik Uyum İyiliği Testi (Ki-Kare)	8
1.3.2. EDF Uyum İyiliği Testleri.....	13
1.4. Tek Değişkenli Uyum İyiliği Test Tablolarının Benzetimle Bulunması.....	16
1.5. İki Değişkenli Dağılımlar	20
1.5.1. Ortak Olasılık Yoğunluk Fonksiyonu.....	21
1.5.2. Ortak Birikimli Dağılım Fonksiyonu.....	21
1.6. İki Değişkenli Uyum İyiliği Testleri.....	22
1.6.1. Ki –Kare Uyum İyiliği Testi.....	22
1.6.2. EDF Uyum İyiliği Testleri.....	27
1.7. Çokgenler.....	32

1.7.1.	Çokgendeki Nokta Problemi.....	33
1.7.2.	İki Çokgenin Kesişim Çokgeni.....	34
1.8.	Üçgenleştirme	35
1.8.1.	Dışbükey Çokgen Alanların Üçgenleştirilmesi	37
1.8.1.	İçbükey Çokgen Alanların Üçgenleştirilmesi.....	38
1.8.2.	Bir Üçgenin Alt Üçgenlere Bölünmesi.....	39
1.8.3.	Tabanı Üçgen Olan Kesilmiş Prizmanın Hacim Hesabı.....	40
2.	YAPILAN ÇALIŞMALAR.....	42
2.1.	Çokgensel Alanda Dağılım Fonksiyonunun Hesaplanması	43
2.1.1.	Analitik Hesaplama	44
2.1.2.	Sayısal Hesaplama	46
2.2.	Çokgensel Alanda Uyum İyiliği Testleri.....	48
2.2.1.	Çokgensel Alanda Klasik Uyum İyiliği Testi.....	48
2.2.2.	Çokgensel Alanda EDF Uyum İyiliği Testleri	51
3.	BULGULAR VE SONUÇLAR.....	55
4.	ÖNERİLER.....	57
5.	KAYNAKLAR	58
ÖZGEÇMİŞ		

Yüksek Lisans

ÖZET

ÇOKGENSEL ALANDA İKİ DEĞİŞKENLİ SÜREKLİ DAĞILIM FONKSİYONLARI
İÇİN UYUM İYİLİĞİ TESTLERİ

Tuncay ULUYURT

Karadeniz Teknik Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik ve Bilgisayar Bilimleri Anabilim Dalı
Danışman: Yrd. Doç. Dr. Orhan KESEMEN
2015, 59 Sayfa

Bu çalışmada, iki boyutlu keyfi bir alanın sürekli iki değişkenli dağılımına göre klasik ve deneysel dağılım fonksiyonları yardımıyla uyum iyiliği testlerinin yapılabilmesi için yeni bir yaklaşım denenmiştir. İki-boyutlu keyfi bir alan çokgensel bir yaklaşımla tanımlanmıştır.

Tanım bölgesi çokgensel bir alan olan rastgele değişkenlerin dağılım fonksiyonun hesaplanması için çokgensel bölge öncelikle üçgenlere ayrılır. Üçgenleme işleminde Delaunay algoritması tercih edilmiştir. Her bir üçgen ayrı ayrı incelenerek istatistiksel işlemler gerçekleştirilir. Her bir üçgende dağılım fonksiyonu, gözlenen frekanslar, beklenen frekanslar ve deneysel dağılım fonksiyonu değerleri hesaplanması için farklı yaklaşımlar denenmiştir.

Bu yaklaşımlar yardımıyla elde edilen parametreler ile iki değişkenli Ki-kare ve Kolmogorov-Smirnov testleri başarıyla uygulanmıştır.

Anahtar Kelimeler: Ki-kare testi, Kolmogorov-Smirnov testi, İki değişkenli dağılımlar, Uyum iyiliği testleri, Çokgensel yaklaşımlar.

Master Thesis

SUMMARY

GOODNESS OF FIT TESTS FOR CONTINUOUS BIVARIATE DISTRIBUTION
FUNCTIONS IN POLYGONAL AREA

Tuncay ULUYURT

Karadeniz Technical University
The Graduate School of Natural and Applied Sciences
Statistical and Computer Science Graduate Program
Supervisor: Assist. Prof. Dr. Orhan KESEMEN
2015, 59 Pages

In this study there has been tried a new approach in order to do goodness of fit tests of an arbitrary two-dimensional area with the help of classical and empirical distribution functions based on bivariate distributions. Arbitrary two-dimensional area is defined by a polygonal approach.

Polygonal definition area is firstly divided into triangles to calculate the distribution function of the random variables. The Delaunay algorithm is preferred for triangulation operation. Statistical calculations are carried out by analyzing each of the triangles separately. For each triangular different approaches have been tried to the calculate distribution function, observed frequencies, the expected frequency and the values of empirical distribution function

With the help of parameters obtained by using these approaches, Chi-square and Kolmogorov-Smirnov tests with two variables have been carried out successfully.

Key Words: Chi-square test, Kolmogorov-Smirnov test, Bivariate distributions, Goodness of fit test, Polygonal approximation.

ŞEKİLLER DİZİNİ

	<u>Sayfa No</u>
Şekil 1. Olasılık yoğunluk fonksiyonu	2
Şekil 2. Pozitif bölgede tanımlı üstel olasılık yoğunluk fonksiyonu	3
Şekil 3. Tanımlı bir aralıkta olasılık yoğunluk fonksiyonu	4
Şekil 4. Standart normal dağılımdan oluşturulmuş 200 tane rastgele sayısının [-3,3] aralığındaki sıklıklarının gösteren histogram.....	4
Şekil 5. Bir fonksiyonun integrali.....	6
Şekil 6. İntegral alanın dilimlere ayrılması.....	7
Şekil 7. Gözlenen frekansların gösterimi.....	12
Şekil 8. Beklenen frekansların gösterimi.....	13
Şekil 9. Deneysel dağılım fonksiyonu ile teorik dağılım fonksiyonun karşılaştırılması ($x_i \leq x$)	15
Şekil 10. Deneysel dağılım fonksiyonu ile teorik dağılım fonksiyonu arasındaki örnek noktalarındaki farklar ($x_i \leq x$).....	15
Şekil 11. Deneysel dağılım fonksiyonu ile teorik dağılım fonksiyonun karşılaştırılması ($x_i < x$)	16
Şekil 12. Deneysel dağılım fonksiyonu ile teorik dağılım fonksiyonu arasındaki örnek noktalarındaki farklar ($x_i < x$).....	16
Şekil 13. İki değişkenli uygulama için örnek olasılık yoğunluk fonksiyonu.....	24
Şekil 14. İki değişkenli uygulama için gözlenen frekanslar	25
Şekil 15. İki değişkenli uygulama için beklenen frekanslar	26
Şekil 16. Deneysel dağılım fonksiyonu ($x_i \leq x, y_i \leq y$).....	30
Şekil 17. Deneysel dağılım fonksiyonu ($x_i < x, y_i < y$).....	30
Şekil 18. Teorik dağılım fonksiyonu	31
Şekil 19. Çokgen tanımlaması; (a) Dışbükey bir çokgen; (b) İçbükey bir çokgen; (c) İç içe girmiş çokgen; (d) Kenarları kesişmiş çokgen.	33

Şekil 20.	Çokgendeki nokta problemi; (a) Seçilen noktalar çokgenin içinde; (b) Seçilen noktalar çokgeni dışında	34
Şekil 21.	İki çokgenin kesişimi; (a) P ve Q çokgenleri; (b) Kesişim bölgesi	35
Şekil 22.	(a) Keyfi olarak çizilmiş çokgensel alan (b) Delaunay üçgenleştirme algoritması uygulanmış çokgensel alan	36
Şekil 23.	Çevresel çemberlerin gösterimi	37
Şekil 24.	(a) Keyfi olarak çizilmiş çokgensel alan (b) Delaunay üçgenleştirme algoritması uygulanmış çokgensel alan	37
Şekil 25.	İçbükey çokgenlerin üçgenleştirilmesi; (a) İçbükey bir çokgen; (b) İçbükey çokgenin delaunay algoritması ile üçgenleştirilmesi	38
Şekil 26.	İçbükey çokgenlerin üçgenleştirilmesindeki fazla üçgenlerin atılması; (a) Çokgen dışındaki üçgenlerin bulunması; (c) Çokgen dışındaki çokgenlerin silinmesi	39
Şekil 27.	Üçgenlerin alt üçgenlere bölünmesi; (a) $K = 1$ için üçgende bölünme gerçekleşmedi; (b) $K = 2$ için üçgen 4 alt üçgene bölündü; (c) $K = 3$ için üçgen 9 alt üçgene bölündü; (d) $K = 4$ için üçgen 16 alt üçgene bölündü.....	40
Şekil 28.	Kesilmiş üçgen prizmanın hacminin hesaplanması; (a) Taban alanı; (b) Kesilmiş üçgen prizma	41
Şekil 29.	İki boyutlu keyfi bir bölgenin (Avustralya) 20 köşeden oluşan çokgensel bir bölge ile yaklaşık temsili; (a) keyfi bölge; (b) çokgensel bölge	42
Şekil 30.	Çokgenin üçgenlere bölünmesi; (a) İçbükey çokgenin üçgenleştirilmesi; (b) Çokgen dışında kalan üçgenlerin silinmesi.....	44
Şekil 31.	Üçgensel bölgenin gösterimi; (a) Ω_j bölgesi; (b) Ω_{jL} ve Ω_{jR} üçgensel bölgelerinin belirlenmesi.	45
Şekil 32.	Üçgensel bölgedeki olasılık değerinin bulunması için alt üçgenlere bölünmesi; (a) Üçgen bölgenin 1 üçgene bölünmesi; (b) Üçgen bölgenin 4 üçgene bölünmesi; (c) Üçgen bölgenin 9 üçgene bölünmesi; (d) Üçgen bölgenin 16 üçgene bölünmesi; (e) Üçgen bölgenin 25 üçgene bölünmesi; (f) Üçgen bölgenin 36 üçgene bölünmesi;.....	47
Şekil 33.	Benzetim için kullanılan olasılık yoğunluk fonksiyonu; (a) Fonksiyonun üç boyutlu görünümü; (b) Fonksiyonun iki boyutlu renklendirilmiş görünümü	50
Şekil 34.	Avustralya bölgesinde üretilen rastgele sayılar; (a) Rastgele sayıların iki boyutlu düzlemde gösterimi; (b) Rastgele sayılarını üçgenlere göre sınıflandırılması	50

Şekil 35.	Çokgeni oluşturan üçgenlerin alt üçgenlere bölünmesi; (a) Alt üçgenlerin iki boyutlu düzlemde gösterimi; (b) Alt üçgenlerin üç boyutlu gösterimi	50
Şekil 36.	Beklenen ve gözlenen frekansların grafiksel karşılaştırılması; (a) Gözlenen frekansların üçgen sütun gösterimi; (b) Beklenen frekansların üçgen sütun gösterimi.	51
Şekil 37.	Rastgele noktalar; (a) Rastgele noktaların gösterimi; (b) Rastgele noktanın sol ve alt sınırları	53
Şekil 38.	Kesişim bölgesi; (a) Kesişim bölgesinin hesaplanması; (b) Kesişim bölgesinin üçgenleştirilmesi	53
Şekil 39.	Olasılık değerinin hesaplanması; (a) Kesişim bölgesi; (b) Olasılık sütunları	54

TABLolar DİZİNİ

	<u>Sayfa No</u>
Tablo 1. Hileli bir zarın gelme olasılıkları.....	9
Tablo 2. 100 defa atılan hileli bir zarın beklenen gözlenen frekansları.....	9
Tablo 3. Ki-kare işlem tablosu.....	10
Tablo 4. Ki-kare kritik değerler tablosu.....	17
Tablo 5. Normal dağılımdan elde edilmiş Ki-kare istatistik değerler tablosu	18
Tablo 6. Kolmogorov-Smirnov kritik değerler tablosu	19
Tablo 7. Normal dağılımdan elde edilmiş Ki-kare istatistik değerler tablosu	20
Tablo 8. İki değişkenli Ki-kare işlem tablosu	26
Tablo 9. İki değişkenli Kolmogorov-Smirnov işlem tablosu.....	31
Tablo 10. Ki-kare benzetim sonuçları.....	56
Tablo 11. Kolmogorov-Smirnov benzetim sonuçları	56

SEMBOLLER DİZİNİ

X	: Rastgele deęişken
Y	: Rastgele deęişken
$X \sim U(0,1)$: Düzgün dağılımdan $[0,1)$ aralığında rastgele bir deęişken
$X \sim N(0,1)$: Normal dağılımdan 0 ortalamalı 1 standart sapmalı bir rastgele deęişken
PDF	: Olasılık Dağılım Fonksiyonu
CDF	: Birikimli Dağılım Fonksiyonu
EDF	: Deneysel Dağılım Fonksiyonu
χ^2	: Ki-kare test istatistięi
$\chi_{v,1-\alpha}^{2*}$: Ki-kare tablo deęeri
\hat{D}	: Kolmogorov-Smirnov test istatistięi
KS^*	: Kolmogorov-Smirnov tablo deęeri
P	: Olasılık deęeri
$f(x)$: Olasılık yoğunluk fonksiyonu
$F(x)$: Birikimli dağılım fonksiyonu
$f(x, y)$: İki deęişkenli olasılık yoğunluk fonksiyonu
$F(x, y)$: İki deęişkenli birikimli dağılım fonksiyonu
Ω	: Çokgensel tanım bölgesi
Ω_j	: j . üçgenin tanım bölgesi
Ω_L	: Bir üçgenin sol tarafının tanım bölgesi
Ω_R	: Bir üçgenin sağ tarafının tanım bölgesi
p_k	: P çokgeninin köşe koordinatları
q_k	: Q çokgeninin köşe koordinatları
s_k	: Kesişim çokgeninin köşe koordinatları

1. GENEL BİLGİLER

1.1. Giriş

Rastgele sayıların veya gözlem değerlerinin uyum iyiliği testleri veri analizinde ve çıkarımsal istatistikte en temel problemlerden birisidir (Wang, 2012). Pearson Ki-kare testini temel alan klasik testler genelde ayrık veriler için kullanılırken sürekli değişkenler için ise Kolmogorov-Smirnov testi, Anderson-Darling testi, Cramér-von Mises testi gibi birçok test yaygın bir biçimde kullanılmaktadır. Ki-kare testi verilen bir aralıktaki teorik dağılımdan elde edilen beklenen frekanslar ile aynı aralıkta bulunan gözlem sayılarının farklarındaki anlamlılığı ölçmektedir. Kolmogorov-Smirnov testi, Anderson-Darling testi ve Cramér-von Mises testi gibi testler ise deneysel dağılım fonksiyonu (EDF) ile hipotez (teorik) dağılımı arasındaki tutarsızlığı ölçmek için kullanılmaktadır. Bu tür testlere EDF tabanlı testler denir.

Uyum iyiliği testi genelde tek değişkenli dağılımlar için geliştirilmiştir. Öte yandan literatürde çok değişkenli normallik testleri için de kullanıldığı çalışmalar bulunmaktadır (Krishnaiah, 1990; D'Agostino, 1986; Justel vd., 1997). Ancak çokgensel bir bölge ile sınırlandırılmış iki değişkenli dağılım fonksiyonları ile ilgili herhangi bir çalışma yapılmamıştır.

Bu çalışmada, iki boyutlu keyfi bir bölge çokgensel bir yapıda tanımlanmış ve bu çokgensel alan içerisinde uyum iyiliği testi yapılması için yöntemler geliştirilmiştir.

1.2. Olasılık Fonksiyonları

İnsanoğlunun gerçek hayatta karşılaştığı olaylardan veya durumlardan elde ettiği rastgele verileri bir kurala bağla ihtiyacı olasılık yoğunluk fonksiyonunu ve bunun dağılım fonksiyonunu doğurmuştur.

1.2.1. Olasılık Yoğunluk Fonksiyonu

Olasılık yoğunluk fonksiyonu $f(x)$, sürekli rastgele bir değişkenin (X) olasılık dağılımını belirlemek için kullanılır. X rastgele değişkenini bir aralıkta olma olasılığı

yüksekse bu deęişkenin bulanabileceęi aralıktaki $f(x)$ fonksiyonu da büyük deęerler almaktadır. Dolayısıyla bir rastgele deęişkeni a ve b aralığında olma olasılığı (1) eşitliğinde olduęu gibi olasılık yoğunluk fonksiyonun $f(x)$ $[a, b]$ aralığında integrali alınarak bulunabilir.

$$P(a \leq X \leq b) = \int_a^b f(x) dx \quad (1)$$

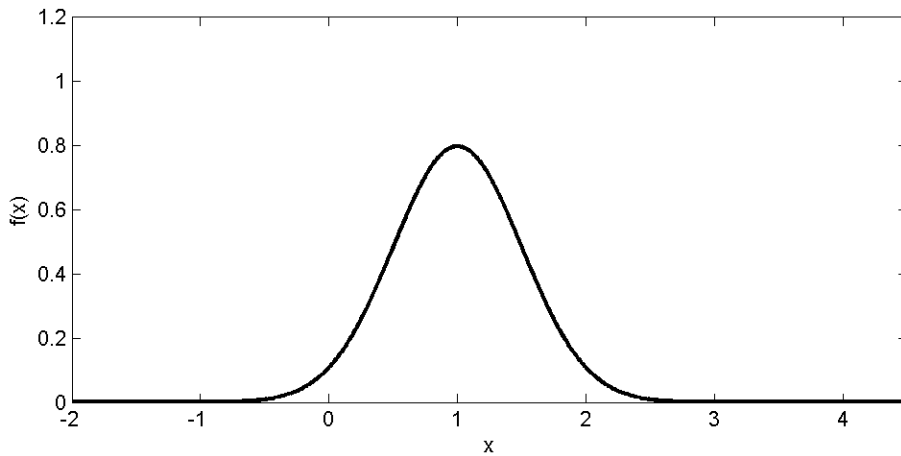
Bir fonksiyonun olasılık yoğunluk fonksiyonu olabilmesi için ařaęıdaki kořulları saęlamalıdır.

1. Olasılık yoğunluk fonksiyonu eşitlik (2)'de verildięi gibi her x deęeri için sıfırdan büyük veya eşit olmalıdır (Şekil 1).

$$f(x) \geq 0 \quad (2)$$

2. Olasılık yoğunluk fonksiyonun eksi sonsuzda artı sonsuza integrali alınırsa bire eşit olur (3). Dięer bir deyişle olasılık yoğunluk fonksiyonu ile x eksenini arasında kalan bölgenin alanı 1'e eşit olmalıdır.

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad (3)$$

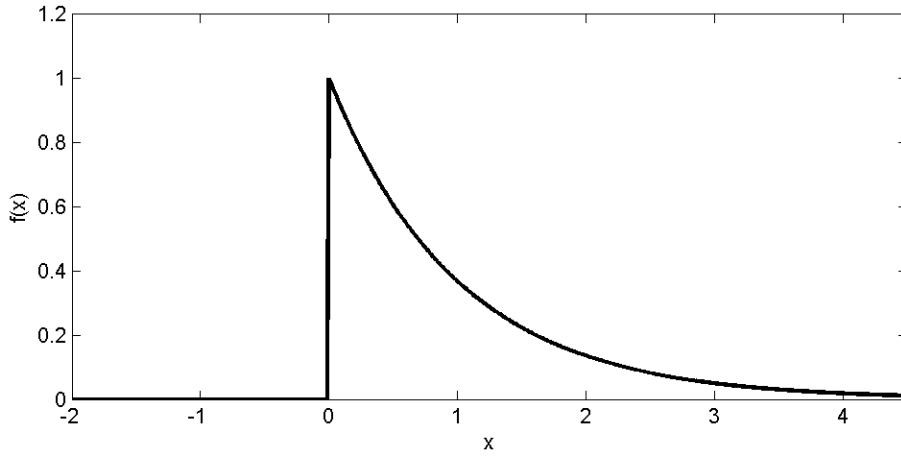


Şekil 1. Olasılık yoğunluk fonksiyonu

Olasılık yoğunluk fonksiyonunun sıfır olduğu bölgelerde X rastgele değişkenin görülme olasılığı sıfırdır. Dolayısıyla olasılık yoğunluk fonksiyonu (4) eşitliğinde olduğu gibi bazı durumlarda parçalı bir fonksiyon biçiminde yazılabilir (Şekil 2). Örnek olarak,

$$f(x) = \begin{cases} \frac{1}{\lambda} e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (4)$$

biçiminde pozitif bölgede bir fonksiyon olarak tanımlanmaktadır.

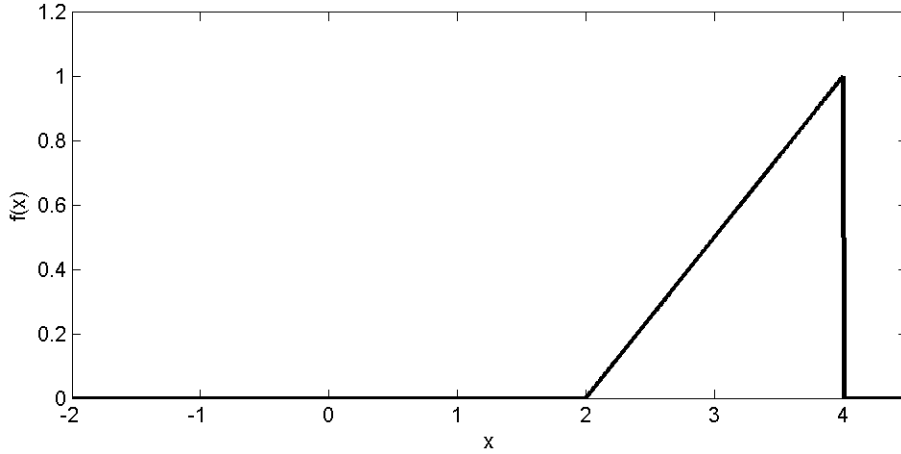


Şekil 2. Pozitif bölgede tanımlı üstel olasılık yoğunluk fonksiyonu

Öte yandan (5) eşitliğindeki örnekte olduğu gibi,

$$f(x) = \begin{cases} x/2 - 1, & 2 \leq x \leq 4 \\ 0, & \text{diğer yerlerde} \end{cases} \quad (5)$$

belli bir aralıkta da tanımlanabilir (Şekil 3).

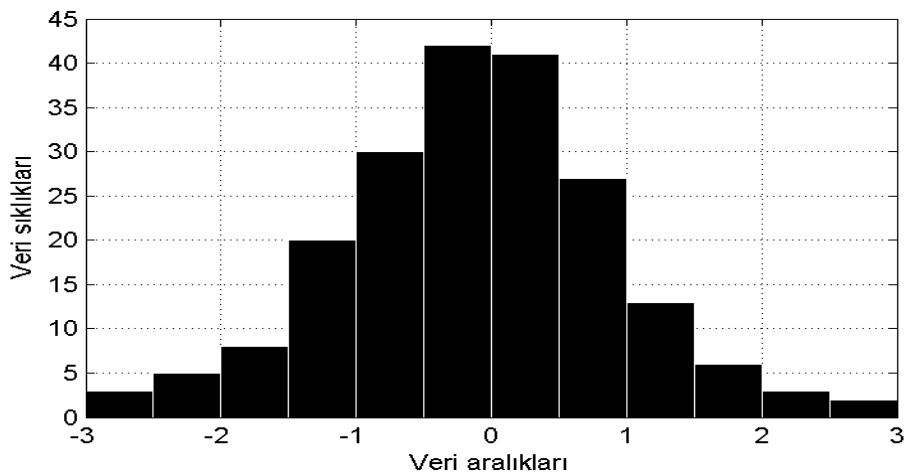


Şekil 3. Tanımlı bir aralıkta olasılık yoğunluk fonksiyonu

Gözlem değerleri, araştırmacılar tarafından gerçek hayattan toplanan verilerdir. Bu verilerin olasılık yoğunluk fonksiyonunu kestirebilmek için ilk yapılan işlem verilerin belli aralıklardaki histogramını çizmektir. Histogram olasılık yoğunluk fonksiyonun kesitimi için iyi bir yaklaşım olarak verilmektedir.

1.2.2. Sıklık Göstergesi (Histogram)

Diğer bir adıyla sıklık göstergesi olan histogram, gözlem değerlerinin eşit aralıklara sahip ardışık bölgelerde görülme miktarının grafiksel olarak gösterilmesidir. Şekil 4'de standart normal dağılımdan oluşturulmuş 200 tane rastgele sayının $[-3,3]$ aralığında bulunmuş sıklıkları gösterilmiştir.



Şekil 4. Standart normal dağılımdan oluşturulmuş 200 tane rastgele sayısının $[-3,3]$ aralığındaki sıklıklarının gösteren histogram

1.2.3. Birikimli Dağılım Fonksiyonu

Sürekli rastgele bir değişken (X) için dağılımını tanımlayacak farklı bir yaklaşım ise dağılım fonksiyonudur. Dağılım fonksiyonu, olasılık yoğunluk fonksiyonun eksi sonsuzdan x 'e kadar integralinin alınmasıyla elde edilir (6).

$$F_X(x) = \int_{-\infty}^x f(u) du, \quad -\infty < x < \infty \quad (6)$$

İntegral sınırlarından birisi değişken olduğu için integral sonucu bir değer değil bir fonksiyon olmaktadır. Diğer bir deyişle, olasılık yoğunluk fonksiyonunun x eksenini ile arasında kalan alanın eksi sonsuzdan x 'e kadar olan bölgesinin alanı o noktadaki dağılım değeri vermektedir.

Dağılım fonksiyonunun diğer bir anlamı ise X değişkeninin herhangi bir x değerinden küçük olma olasılığı olarak verilmektedir (7).

$$F_X(x) = P(X \leq x) \quad (7)$$

X rastgele değişkeninin $[a, b]$ aralığında olasılık değeri ise (8) eşitliğindeki gibi verilmektedir.

$$P(a \leq X \leq b) = F(b) - F(a) \quad (8)$$

Öte yandan X rastgele değişkeninin sürekli bir dağılım fonksiyonuna göre herhangi bir x değerini alma olasılığı sıfırdır (9).

$$P(X = x) = F(x) - F(x) = 0 \quad (9)$$

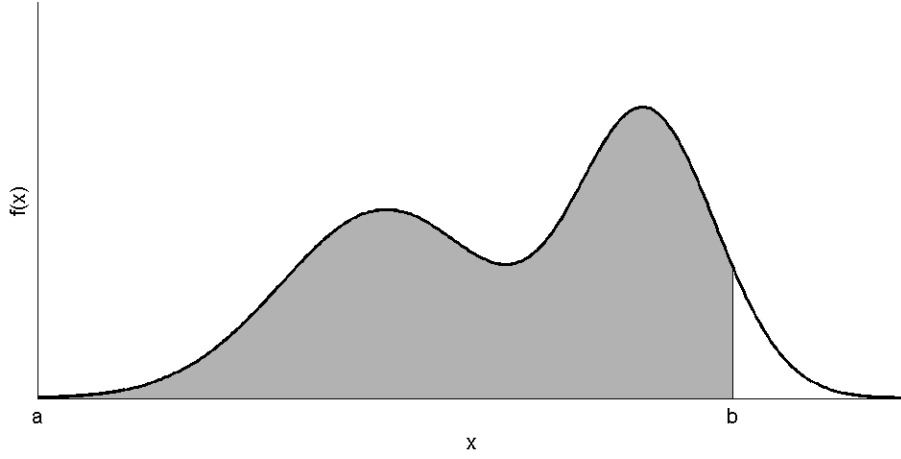
Dağılım fonksiyonunun en genel özellikleri ise fonksiyonun $[0,1]$ aralığında monoton artan bir fonksiyon olmasıdır (10).

$$F(b) \geq F(a), \quad b > a \quad (10)$$

1.2.4. Birikimli Dağılım Fonksiyonunun Sayısal Yöntemlerle Bulunması

Birikimli dağılım fonksiyonunun bulunmasında olasılık yoğunluk fonksiyonun integralinin alınması gerekir. Ancak bazı durumlarda integralinin analitik olarak hesaplanması mümkün olmayabilir. Bu durumda asimptotik veya sayısal yaklaşımlar kullanılmaktadır.

Pozitif tanımlı bir fonksiyonun ($f(x) \geq 0, \forall x$) $[a, b]$ aralığındaki integrali o fonksiyonun $[a, b]$ aralığındaki eğrisi ile x ekseninde kalan alan olduğu bilinmektedir (Şekil 5).



Şekil 5. Bir fonksiyonun integrali

Analitik olarak hesaplanamayan bir integralin sayısal değerinin bulunmasında en çok bilinen yöntemlerden birisi yamuk kuralıdır (Howard ve Musto, 2008). Bu kural bir fonksiyonun $[a, b]$ aralığında integralinin alınması için aralığı m tane eşit aralıklı dilimlere ayırır (Şekil 6). Her bir dilimin genişliği (11) eşitliğinde verilmektedir.

$$\Delta x = \frac{b - a}{m} \quad (11)$$

Her dilimin sınırları (12) eşitliğindeki gibi hesaplanabilir.

$$x_j = x_0 + j \cdot \Delta x, \quad j = 0, 1, 2, \dots, m \quad (12)$$

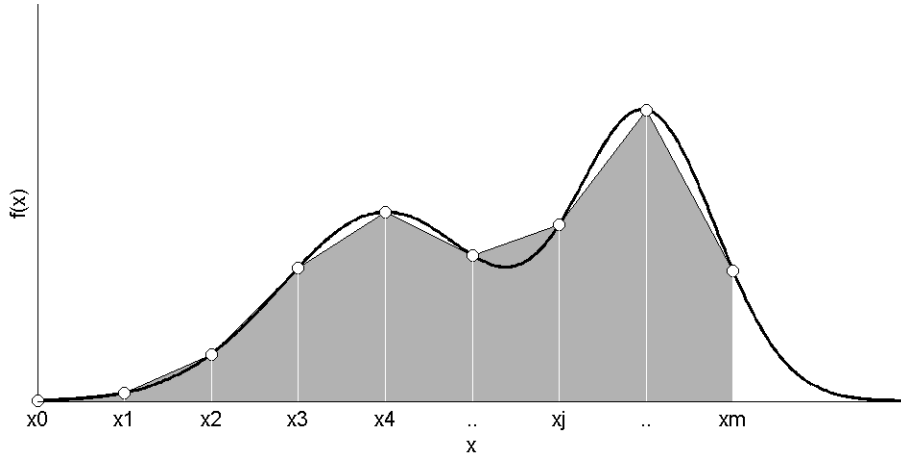
Bir dilimlerin alt ve üst sınırlarının fonksiyonu kestiği noktalar arasında düz bir çizgi çekilirse aralık yan yatmış bir yamuğa dönüşür (Şekil 6). Herhangi bir yamuğun alanı, yamuğun taban uzunluğu ile tavan uzunluğunun toplamının yarısının yamuk genişliği ile çarpılması ile bulunur. Aynı işlemi integral hesabında bir yamuk dilimine uygulanırsa,

$$I_j = \frac{f(x_j) + f(x_{j+1})}{2} \Delta x \quad (13)$$

eşitliğiyle bulunur. Tüm yamukların alanı toplanırsa,

$$F(b) - F(a) = \left[\frac{1}{2} f(x_0) + \sum_{j=1}^{m-1} f(x_j) + \frac{1}{2} f(x_m) \right] \Delta x \quad (14)$$

eşitliği elde edilir. Burada $x_0 = a$ ve $x_n = b$ olarak verilmektedir.



Şekil 6. İntegral alanın dilimlere ayrılması

a değerinden önce olasılık yoğunluk fonksiyonu tanımlı değil veya sıfıra eşitse $F(a)$ değeri sıfıra eşit olacağından (14) eşitliği,

$$F(b) = \left[\frac{1}{2} f(a) + \sum_{n=1}^{m-1} f(x_n) + \frac{1}{2} f(b) \right] \Delta x \quad (15)$$

biçimine dönüşür. Olasılık yoğunluk fonksiyonları çoğu zaman sonsuz bir aralıkta tanımlanabilir. Bu durumda dağılım fonksiyonunu hesaplamak için olasılık yoğunluk fonksiyonunun bir tolerans değerinden büyük olan aralıkları ($f(x) > \epsilon$) için sayısal yöntem uygulanabilir

1.3. Uyum İyiliği Testleri

İster bilgisayarda üretilen bir rastgele sayılar dizisi olsun ister gerçek hayattan toplanan gözlem değerleri olsun, bu değerlerin verilen bir dağılımdan gelip gelmediği araştırmak için kullanılan tüm testlere uyum iyiliği testi denir. Bu konuda ilk olarak Pearson tarafından 1900 yılında ortaya atılan Ki-kare testi olmuştur (Gibbons ve Chakraborti, 2011). Bu günümüze kadar güncelliğini korumaktadır. Basit olması ve dağılımının bilinmesi çoğu zaman tercih edilen klasik bir teknik haline gelmesini sağlamıştır. Öte yandan daha duyarlı teknikler özellikle deneysel dağılım fonksiyonuna dayalı tekniklerin gelişmesi daha sonraki yıllarda gerçeklemiştir.

Diğer yaygın uyum iyiliği testi Kolmogorov-Smirnov testidir. Bu test yöntemi daha çok normallik testlerinde kullanılmaktadır. Örneklem verilerinin tanımlanmış bir normal olasılık dağılımına uyumluluk gösterip göstermemesini test etmek için kullanılır. Örneklem verileri standardize edilerek (yani her bir veri değerinden teorik dağılım olasılığı için verilmiş ortalama çıkartılır ve sonucu verilmiş teorik yığın standart sapmasına bölünerek) elde edilen normalize veriler standart normal dağılım ile karşılaştırılır. Fakat bu türlü dönüşüm yapmanın sınamanın gücünü azalttığı ispat edilmiştir. Bu taraflılığın düzeltilmesi Lilliefors testi ortaya atılmıştır. Fakat, Anderson-Darling testi veya Shapiro-Wilk testi normallik sınaması olarak hem Lilliefors testinden hem de Kolmogorov-Smirnov tek örneklem normallik testinden daha güçlüdür. Kolmogorov-Smirnov tek örneklem testi, gözlenen ve beklenen birikimli frekans dağılışı arasındaki mutlak farklılıklar dikkate alınarak geliştirilmiştir. Örneklem hacmi küçük olduğu için Ki-kare testi uygulanamadığı durumlarda Kolmogorov-Smirnov testi kullanılabilir (URL-1, 2015).

1.3.1. Klasik Uyum İyiliği Testi (Ki-Kare)

Klasik anlamda uyum iyiliği testleri içinde en yaygın yöntemlerden birisi Ki-kare testidir. Ki-kare testi elde edilen gözlem verilerinin seçilen bir dağılıma uyup uymadığını

test etmek için kullanılan kullanışlı bir tekniktir. Bu teknik biyoloji, coğrafya ve psikoloji gibi alanlarda yaygın bir biçimde kullanılmaktadır.

Ki-karenin uygulama alanlarına örnek verilecek olursa: Hileli bir zar imal edildiğini düşünölsün. Bu zarın altı yüzünden hangisinin öste geleceđi araştırıldığında aşğıdaki tablodaki olasılıklar (p_j) bulunduđu varsayölsün.

Tablo 1. Hileli bir zarın gelme olasılıkları

j	1	2	3	4	5	6
p_j	0.07	0.14	0.28	0.25	0.18	0.08

Yapılan bir deneyde deđişik zarlar kullanılmıřtır. Bu zarlardan hangisinin üretilen hileli zar olup olmadıđı tek tek araştırılması istenmektedir. Bunun zarlardan bir tanesi seçilip 100'er defa atıldığında her sayının gelme miktarı bulunarak gözlenen frekansları (O_j) olarak kaydedilmiřtir (Tablo 2). Aynı sayöların Tablo 1'deki olasılıkları 100'le çarpılarak beklenen frekansları (E_j) bulunarak Tablo 2'de gösterilmiřtir.

Tablo 2. 100 defa atılan hileli bir zarın beklenen gözlenen frekansları

j	1	2	3	4	5	6
E_j	7	14	28	25	18	8
O_j	6	13	29	28	17	7

Tablo 2'de verilen gözlenen frekanslar ile beklenen frekanslar arasındaki farkın küçük olması uyumun yüksek olduđunu, düşük olması ise uyum düşük olduđunu gösterecektir. Bulunacak farklılara toplam bir ölçüt getirmek için farkların toplamı alınarak toplam bir hata bulunmak istendiğinde farkların bazölarını negatif bazölarının pozitif olması sađlıklı bir sonuç alınmasını imkânsız kılmaktadır. Bu durumda farkların kareleri alınarak toplandıđında,

$$SSE = \sum_j^m (O_j - E_j)^2 \quad (16)$$

eşitliği elde edilir. Bu eşitlikte elde edilen her farkın karesinin beklenen frekansla ilişkili olduğu görülmektedir. Yani beklenen frekansın yüksek olması hatanın da yüksek olmasını doğurmaktadır. Bu ilişkinin giderilmesi farklı hacimdeki verilerin aynı ölçekte değerlendirilmesi zorlaştırmaktadır. Sorunu çözmek için her hata karesi beklenen frekans ile bölünerek normalize edilmektedir (17).

$$\hat{\chi}^2 = \sum_j^m \frac{(O_j - E_j)^2}{E_j} \quad (17)$$

Eşitlik (17)'de bulunan toplam değer, hesaplanmış Ki-kare değeri veya test istatistiği olarak ifade edilmektedir. Bu eşitliklere göre Tablo 2'de verilen değerler yeniden değerlendirilerek Tablo 3'deki tablo elde edilmiştir.

Tablo 3. Ki-kare işlem tablosu

Sayı (j)	O_j	E_j	$(O_j - E_j)$	$(O_j - E_j)^2$	$\frac{(O_j - E_j)^2}{E_j}$
1	6	7	-1	1	0.1429
2	13	14	-1	1	0.0714
3	29	28	1	1	0.0357
4	28	25	3	9	0.3600
5	17	18	-1	1	0.0556
6	7	8	-1	1	0.1250

$$\hat{\chi}^2 = \mathbf{0.7906}$$

Tablo 3'deki işlemler sonucunda elde edilen $\hat{\chi}^2$ değerinin bir anlam ifade edebilmesi için bu değer bir ölçüt ile karşılaştırılması gerekmektedir. Bu ölçütü elde etmek için χ^2 rastgele değişkeninin dağılımının bilinmesi gerekir. Bu dağılımın χ^2_ν dağılımına yakınsamaktadır. Buradaki ν ifadesi serbestlik derecesi olarak bilinmektedir. ν değeri ise,

$$\nu = \text{grup veya sınıf sayısı} - \text{kategori sayısı} \quad (18)$$

olarak verilmektedir.

Testlerin dayandığı en önemli kavramlardan birisi de anlamlılık testidir. Bu testin temeli kurulan iki hipoteze dayandırılmaktadır. Bu hipotezler: sıfır hipotezi (H_0) ve karşıt hipotezdir (H_1). Hipotezler,

H_0 : Gözlem değerleri $F_0(x)$ dağılımından gelmektedir,

H_1 : Gözlem değerleri $F_0(x)$ dağılımından gelmemektedir,

biçiminde verilmektedir. Burada $F_0(x)$ gözlem değerlerinin geldiği öne sürülen dağılım fonksiyonudur. Hipotezlerin gerçekleşmesinde bir yanılma payı (α) konularak testler gerçekleştirilir.

H_0 hipotezinin gerçekleşmesi gereken χ^2_v dağılımının $(1 - \alpha)$ değerini veren $\chi^2_{v,1-\alpha}^*$ değerinin belirlenmesidir. Bu ise χ^2 tablosundan (O'Connor ve Andre, 2011) yararlanarak bulunur. Sonuç olarak (20) eşitsizliğinin doğru olması durumunda H_0 kabul edilmiş olur. Aksi durumda karşıt hipotez kabul edilmiş olur.

$$\hat{\chi}^2 < \chi^2_{v,1-\alpha}^* \quad (19)$$

Sürekli fonksiyonlarda ki-kare testi yapılabilmesi için öncelikle sınıf sayısının ve sınıf aralığının belirlenmesi gerekir. Sınıf sayısı (m) belirlenmesinde literatürde değişik çalışmalar olmasına rağmen çoğu zaman araştırmacının tercihinine bırakılır. Sınıf genişliği,

$$\Delta x = \frac{x_n - x_0}{m} \quad (20)$$

biçiminde bulunur. Burada x_0 verideki en küçük değeri, x_n ise verideki en büyük değeri göstermektedir. j . sınıf alt sınırı,

$$L_j = x_0 + (j - 1) \cdot \Delta x, \quad j = 1, 2, \dots, m \quad (21)$$

biçiminde hesaplanırken üst sınırı

$$U_j = x_0 + j \cdot \Delta x, \quad j = 1, 2, \dots, m \quad (22)$$

biçiminde hesaplanmaktadır.

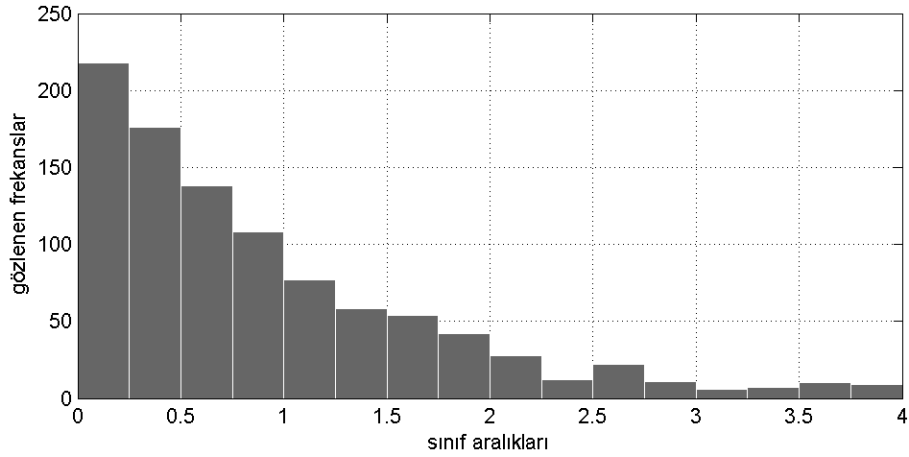
Her sınıf aralığına düşen gözlem sayısı (gözlenen frekanslar),

$$O_j = \sum_{i=1}^n \begin{cases} 1, & L_j \leq x_i < U_j \\ 0, & \text{diğer yerlerde} \end{cases}, \quad j = 1, 2, \dots, m \quad (23)$$

biçiminde hesaplanırken beklenen frekanslar,

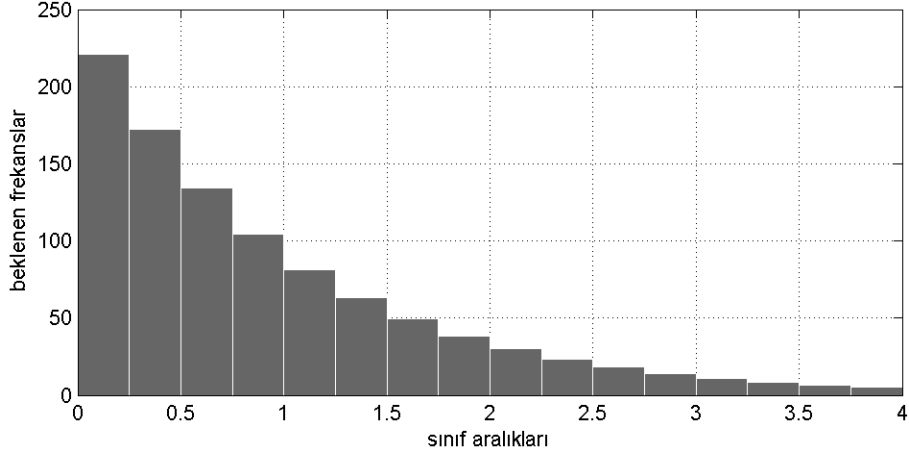
$$E_j = n[F_0(U_j) - F_0(L_j)], \quad j = 1, 2, \dots, m \quad (24)$$

biçiminde hesaplanmaktadır. Örnek olarak ortalaması $\mu = 1$ olan üstel dağılımdan rastgele 1000 tane sayı üretilmiştir. Bu sayılar $m = 16$ sınıfa bölünerek $\Delta x = 0.25$ sınıfların gözlenen frekansları bulunmuştur (Şekil 4).



Şekil 7. Gözlenen frekansların gösterimi

Aynı sınıf aralıkları kullanılarak ortalaması $\mu = 1$ olan üstel dağılımdan beklenen frekanslar hesaplanmıştır (Şekil 5).



Şekil 8. Beklenen frekansların gösterimi

1.3.2. EDF Uyum İyiliği Testleri

EDF tabanlı uyum iyiliği testleri başlangıç açısından aynı işlemleri içermesine rağmen test istatistiklerinin hesaplanmasında değişiklik göstermektedir. EDF tabanlı uyum iyiliği testlerinden en çok bilineni Kolmogorov-Smirnov testidir.

Kolmogorov-Smirnov testi ilk olarak 1933 yılında Kolmogorov tarafından geliştirilmiştir. Bu test yöntemi rastgele seçilmiş bir örneklem grubuna uygulanmıştır. Smirnov ise 1939 yılında Kolmogorov testini iki bağımsız örneklem grubu için kullanılan uyum iyiliği testine uyarlamıştır. Daha sonraları benzerliklerinden dolayı bu test yöntemi Kolmogorov-Smirnov testi olarak isimlendirilmiştir (Massey ve Frank, 1951; Sheskin, 2003).

Tek örneklem Kolmogorov-Smirnov testi için $x = \{x_1, x_2, x_3, \dots, x_n\}$ örnek değerlerinin bağımsız, tanımlı ve sürekli bir dağılımdan elde edildiği varsayılır. Bu dağılım için aşağıdaki hipotez,

$$H_0: F(x) = F_0(x), \quad \forall x, \quad (25)$$

kurulmakta ve bu hipotezin doğruluğu test edilmek istenmektedir. Burada $F_0(x)$ dağılımı bilinen bir birikimli dağılım fonksiyondur. Kurulan hipotezi doğruluğunu test etmek için öncelikle test istatistiğinin belirlenmesi gerekir. Kolmogorov-Smirnov test istatistiği \hat{D} ,

$$\hat{D} = \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)| \quad (26)$$

biçiminde bulunur. Burada $F_n(x)$ fonksiyonu deneysel birikimli dağılım fonksiyonu olarak tanımlanmakta ve

$$F_n(x) = \frac{k_n(x)}{n}, \quad \{k_n(x): x_k \leq x\} \quad (27)$$

biçiminde bulunmaktadır. (26) eşitliğinde supremum hesaplarırken her x_i noktasındaki değeri hesaplanır. x_i noktasının solunda karşılaşılan ilk değere kadar aynı deneysel birikimli dağılım fonksiyonu kullanılır.

Deneysel dağılım fonksiyonu hesaplanırken i . gözlem değerinden küçük eşit olanlar $\{x_i \leq x\}$ işleme katılır (Şekil 9, Şekil 10). Ancak bazı çalışmada sadece küçük değerler $\{x_i < x\}$ işleme katılmaktadır (Şekil 11, Şekil 12). Öte yandan en sağlam karşılaştırma ise ikisinin birlikte kullanılmasıdır. Bu durumda deneysel dağılım fonksiyonu,

$$F_n^+(x) = \frac{k_n(x)}{n}, \quad k_n(x): x_k \leq x \quad (28)$$

ve

$$F_n^-(x) = \frac{k_n(x)}{n}, \quad k_n(x): x_k < x \quad (29)$$

biçiminde iki farklı fonksiyon olarak bulunmaktadır. Test istatistikleri ise,

$$\hat{D}^+ = \sup_{x \in \mathbb{R}} |F_n^+(x) - F_0(x)|, \quad (30)$$

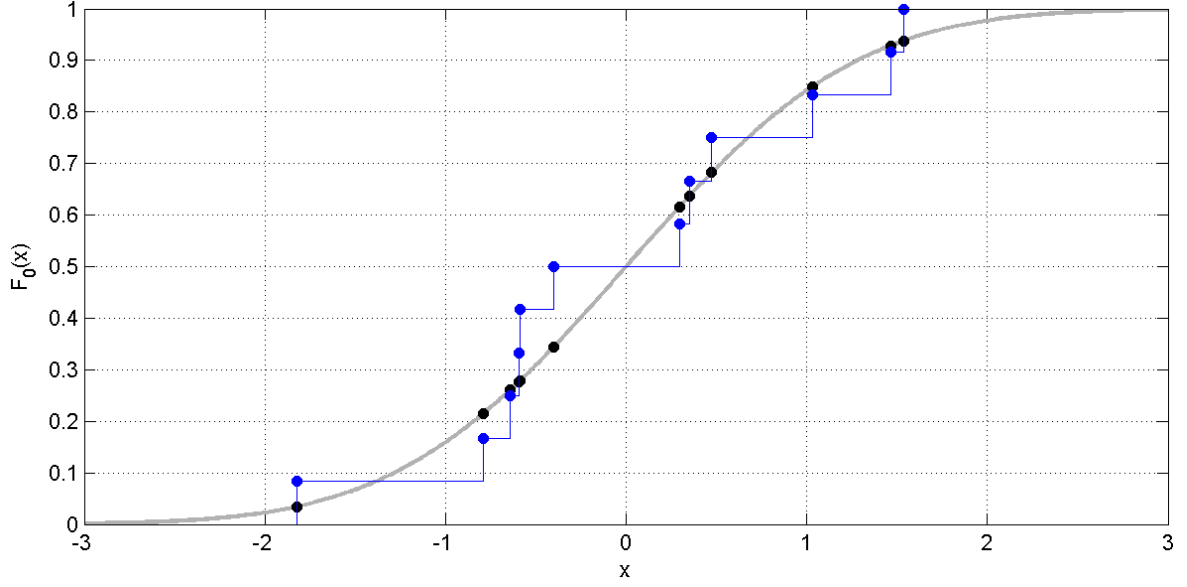
ve

$$\hat{D}^- = \sup_{x \in \mathbb{R}} |F_n^-(x) - F_0(x)|, \quad (31)$$

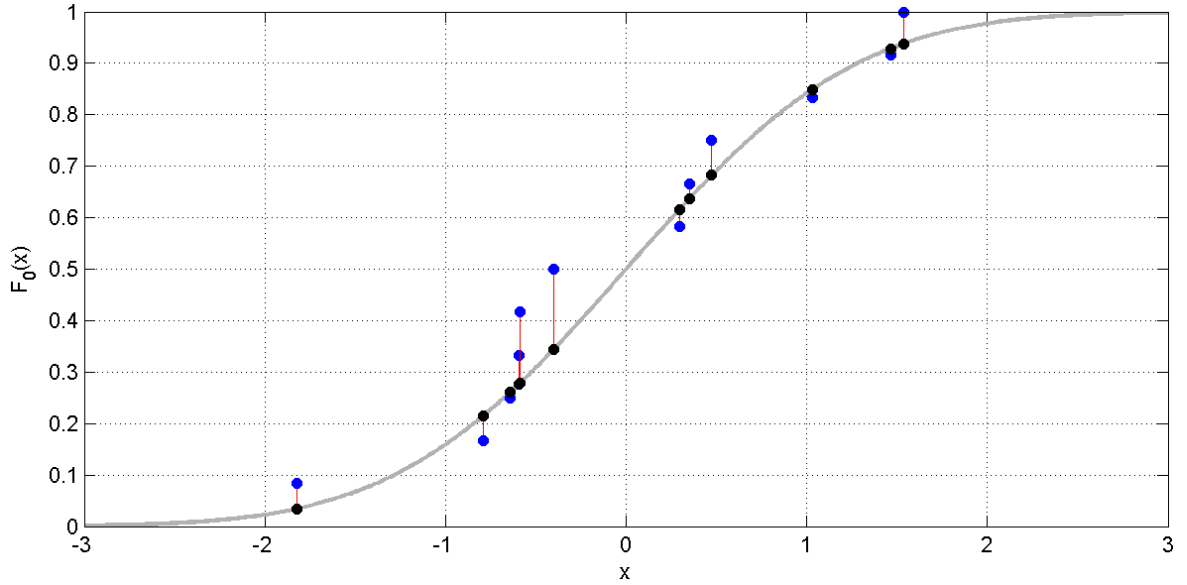
biçiminde bulunurlar. Ortak test istatistiği ise,

$$\hat{D} = \max(\hat{D}^-, \hat{D}^+) \quad (32)$$

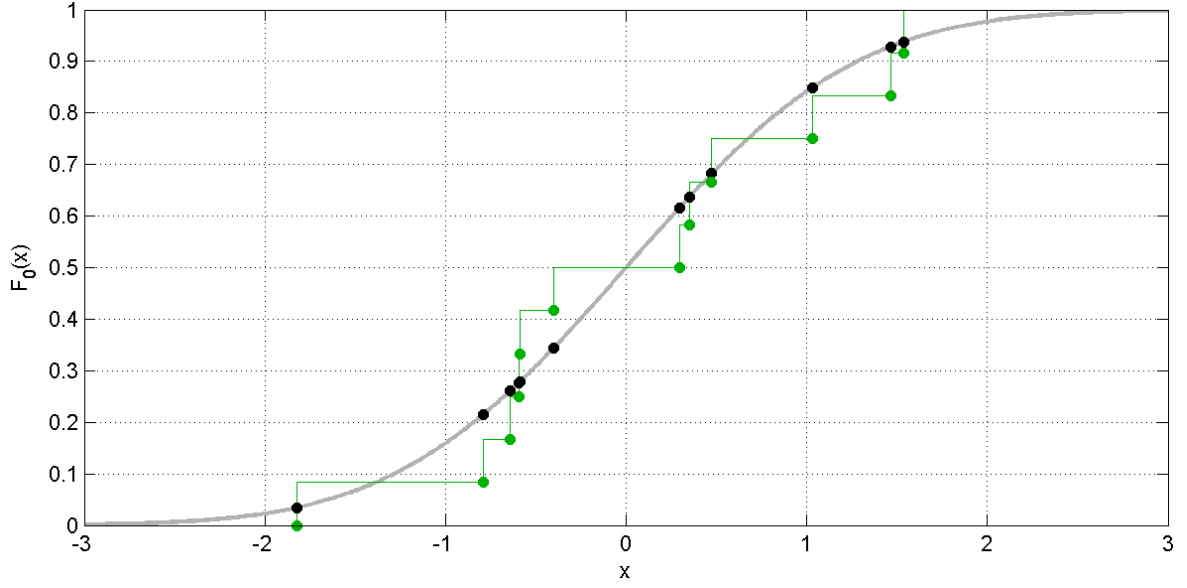
olarak bulunur.



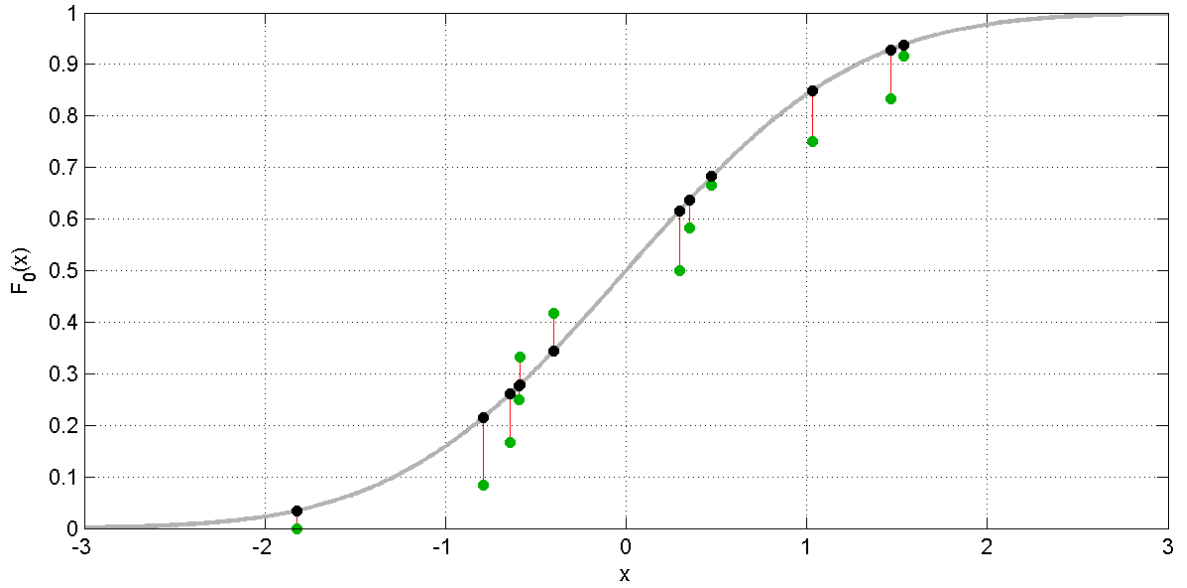
Şekil 9. Deneysel dağılım fonksiyonu ile teorik dağılım fonksiyonunun karşılaştırılması ($x_i \leq x$)



Şekil 10. Deneysel dağılım fonksiyonu ile teorik dağılım fonksiyonu arasındaki örnek noktalarındaki farklar ($x_i \leq x$)



Şekil 11. Deneysel dağılım fonksiyonu ile teorik dağılım fonksiyonunun karşılaştırılması ($x_i < x$)



Şekil 12. Deneysel dağılım fonksiyonu ile teorik dağılım fonksiyonu arasındaki örnek noktalarındaki farklar ($x_i < x$)

1.4. Tek Değişkenli Uyum İyiliği Test Tablolarının Benzetimle Bulunması

Ki-kare uyum iyiliği testinin kritik değerler Tablo 4'te verilmiştir. Bu tablo MATLAB paket programındaki `chi2inv` fonksiyonu ile elde edilmiştir. Bu tablo değerlerini normal dağılımdan elde edilmiş 1000 örnekten oluşan 10.000 örneklemlerle benzetim ile elde

edilmiş kritik tablo deęerleri Tablo 5'te verilmiřtir. İki tablo arasında %99.28 oranında bir benzerlik bulunmaktadır. Dolayısıyla herhangi bir test istatistięinin tablo deęeri bilinmezse benzetim teknięi ile belirlenebilir.

Tablo 4. Ki-kare kritik deęerler tablosu

s.d.	$\alpha = 0.100$	$\alpha = 0.050$	$\alpha = 0.025$	$\alpha = 0.010$	$\alpha = 0.005$
1,000	2,706	3,841	5,024	6,635	7,879
2,000	4,605	5,991	7,378	9,210	10,597
3,000	6,251	7,815	9,348	11,345	12,838
4,000	7,779	9,488	11,143	13,277	14,860
5,000	9,236	11,070	12,833	15,086	16,750
6,000	10,645	12,592	14,449	16,812	18,548
7,000	12,017	14,067	16,013	18,475	20,278
8,000	13,362	15,507	17,535	20,090	21,955
9,000	14,684	16,919	19,023	21,666	23,589
10,000	15,987	18,307	20,483	23,209	25,188
11,000	17,275	19,675	21,920	24,725	26,757
12,000	18,549	21,026	23,337	26,217	28,300
13,000	19,812	22,362	24,736	27,688	29,819
14,000	21,064	23,685	26,119	29,141	31,319
15,000	22,307	24,996	27,488	30,578	32,801
16,000	23,542	26,296	28,845	32,000	34,267
17,000	24,769	27,587	30,191	33,409	35,718
18,000	25,989	28,869	31,526	34,805	37,156
19,000	27,204	30,144	32,852	36,191	38,582
20,000	28,412	31,410	34,170	37,566	39,997
21,000	29,615	32,671	35,479	38,932	41,401
22,000	30,813	33,924	36,781	40,289	42,796
23,000	32,007	35,172	38,076	41,638	44,181
24,000	33,196	36,415	39,364	42,980	45,559
25,000	34,382	37,652	40,646	44,314	46,928
26,000	35,563	38,885	41,923	45,642	48,290
27,000	36,741	40,113	43,195	46,963	49,645
28,000	37,916	41,337	44,461	48,278	50,993
29,000	39,087	42,557	45,722	49,588	52,336
30,000	40,256	43,773	46,979	50,892	53,672
35,000	46,059	49,802	53,203	57,342	60,275
40,000	51,805	55,758	59,342	63,691	66,766
45,000	57,505	61,656	65,410	69,957	73,166
50,000	63,167	67,505	71,420	76,154	79,490
60,000	74,397	79,082	83,298	88,379	91,952
70,000	85,527	90,531	95,023	100,425	104,215
80,000	96,578	101,879	106,629	112,329	116,321
90,000	107,565	113,145	118,136	124,116	128,299
100,000	118,498	124,342	129,561	135,807	140,169

Tablo 5. Normal dağılımdan elde edilmiş Ki-kare istatistik değerler tablosu

s.d.	$\alpha = 0.100$	$\alpha = 0.050$	$\alpha = 0.025$	$\alpha = 0.010$	$\alpha = 0.005$
1,000	2,704	3,844	4,900	6,724	8,100
2,000	4,634	5,906	7,214	9,026	10,514
3,000	6,224	7,752	9,256	11,272	12,760
4,000	7,750	9,340	10,880	12,890	14,880
5,000	9,224	11,168	12,776	14,804	16,352
6,000	10,492	12,438	14,160	16,428	17,856
7,000	11,968	14,016	15,920	18,400	19,904
8,000	13,382	15,578	17,594	20,222	22,094
9,000	14,680	17,040	19,380	21,860	23,620
10,000	16,136	18,490	20,756	23,572	25,376
11,000	17,264	19,640	21,896	25,064	26,888
12,000	18,654	20,994	23,256	26,090	27,910
13,000	19,648	22,056	24,156	26,956	29,364
14,000	21,320	23,870	26,600	29,390	31,580
15,000	22,496	25,120	27,776	31,104	33,120
16,000	23,434	26,222	28,500	31,662	34,518
17,000	24,740	27,512	30,284	33,776	35,324
18,000	25,924	28,926	31,510	35,196	37,514
19,000	27,120	30,080	32,640	35,680	38,400
20,000	28,370	31,268	33,914	37,274	39,962
21,000	29,600	32,636	35,364	38,884	41,348
22,000	30,768	33,942	36,610	40,336	43,188
23,000	31,808	35,024	37,568	40,928	43,568
24,000	32,950	36,250	39,400	42,900	45,000
25,000	34,488	37,400	40,468	44,368	46,968
26,000	35,504	38,906	41,552	45,440	48,572
27,000	36,560	39,640	42,720	46,696	49,496
28,000	37,736	41,158	44,464	48,582	51,424
29,000	39,380	42,680	45,800	49,400	51,440
30,000	40,236	43,832	46,808	50,652	53,504
35,000	46,304	49,976	53,432	57,176	59,624
40,000	51,240	55,258	59,194	63,786	67,148
45,000	57,080	61,404	64,716	69,132	73,180
50,000	62,942	67,634	71,306	75,488	78,446
60,000	74,454	78,846	82,750	87,508	91,290
70,000	85,590	90,418	94,678	100,642	104,334
80,000	96,578	102,086	106,784	112,940	115,856
90,000	107,288	112,930	117,662	124,214	129,674
100,000	118,676	124,736	129,584	136,452	140,492

Kolmogorov-Smirnov uyum iyiliği testinin kritik değerler Tablo 6’da verilmiştir. Bu tablo MATLAB paket programından elde edilmiştir. Bu tablo değerlerini normal dağılımdan elde edilmiş 1000 örnekten oluşan 10.000 örneklemlerle benzetim ile elde edilmiş kritik tablo değerleri Tablo 7’de verilmiştir. İki tablo arasında %99.30 oranında bir

benzerlik bulunmaktadır. Dolayısıyla her hangi bir test istatistiğinin tablo değeri bilinmezse benzetim tekniği ile belirlenebilir.

Tablo 6. Kolmogorov-Smirnov kritik değerler tablosu

s.d.	$\alpha = 0.100$	$\alpha = 0.050$	$\alpha = 0.025$	$\alpha = 0.010$	$\alpha = 0.005$
1,000	0,900	0,950	0,975	0,990	0,995
2,000	0,684	0,776	0,842	0,900	0,929
3,000	0,565	0,636	0,708	0,785	0,829
4,000	0,493	0,565	0,624	0,689	0,734
5,000	0,447	0,509	0,563	0,627	0,669
6,000	0,410	0,468	0,519	0,577	0,617
7,000	0,381	0,436	0,483	0,538	0,576
8,000	0,358	0,410	0,454	0,507	0,542
9,000	0,339	0,387	0,430	0,480	0,513
10,000	0,323	0,369	0,409	0,457	0,489
11,000	0,308	0,352	0,391	0,437	0,468
12,000	0,296	0,338	0,375	0,419	0,449
13,000	0,285	0,325	0,361	0,404	0,432
14,000	0,275	0,314	0,349	0,390	0,418
15,000	0,266	0,304	0,338	0,377	0,404
16,000	0,258	0,295	0,327	0,366	0,392
17,000	0,250	0,286	0,318	0,355	0,381
18,000	0,244	0,279	0,309	0,346	0,371
19,000	0,237	0,271	0,301	0,337	0,361
20,000	0,232	0,265	0,294	0,329	0,352
21,000	0,226	0,259	0,287	0,321	0,344
22,000	0,221	0,253	0,281	0,314	0,337
23,000	0,217	0,248	0,275	0,307	0,330
24,000	0,212	0,242	0,269	0,301	0,323
25,000	0,208	0,238	0,264	0,295	0,317
26,000	0,204	0,233	0,259	0,290	0,311
27,000	0,200	0,229	0,254	0,284	0,305
28,000	0,197	0,225	0,250	0,279	0,300
29,000	0,194	0,221	0,246	0,275	0,295
30,000	0,190	0,218	0,242	0,270	0,290
35,000	0,177	0,202	0,224	0,251	0,269
40,000	0,166	0,189	0,210	0,235	0,252
45,000	0,156	0,179	0,198	0,222	0,238
50,000	0,148	0,170	0,188	0,211	0,226
60,000	0,136	0,155	0,172	0,193	0,207
70,000	0,126	0,144	0,160	0,179	0,192
80,000	0,118	0,135	0,150	0,167	0,179
90,000	0,111	0,127	0,141	0,158	0,169
100,000	0,106	0,121	0,134	0,150	0,161

Tablo 7. Normal dağılımdan elde edilmiş Kolmogorov-Smirnov istatistik değerler tablosu

s.d.	$\alpha = 0.100$	$\alpha = 0.050$	$\alpha = 0.025$	$\alpha = 0.010$	$\alpha = 0.005$
0,900	0,949	0,974	0,989	0,994	0,900
0,681	0,779	0,845	0,903	0,930	0,681
0,563	0,633	0,703	0,780	0,819	0,563
0,495	0,567	0,630	0,693	0,734	0,495
0,449	0,508	0,560	0,620	0,663	0,449
0,408	0,467	0,515	0,570	0,607	0,408
0,378	0,432	0,479	0,526	0,568	0,378
0,361	0,408	0,455	0,502	0,540	0,361
0,336	0,387	0,430	0,476	0,508	0,336
0,324	0,372	0,415	0,459	0,484	0,324
0,309	0,351	0,396	0,439	0,477	0,309
0,295	0,339	0,375	0,416	0,450	0,295
0,287	0,329	0,366	0,408	0,437	0,287
0,274	0,315	0,349	0,390	0,415	0,274
0,267	0,309	0,344	0,383	0,410	0,267
0,260	0,296	0,329	0,367	0,392	0,260
0,252	0,287	0,318	0,355	0,381	0,252
0,242	0,277	0,310	0,347	0,373	0,242
0,235	0,268	0,297	0,334	0,356	0,235
0,230	0,264	0,291	0,329	0,349	0,230
0,227	0,259	0,289	0,321	0,345	0,227
0,220	0,252	0,280	0,311	0,338	0,220
0,217	0,246	0,273	0,305	0,323	0,217
0,213	0,241	0,270	0,305	0,326	0,213
0,209	0,238	0,263	0,293	0,319	0,209
0,205	0,234	0,258	0,289	0,309	0,205
0,200	0,229	0,253	0,280	0,296	0,200
0,194	0,222	0,245	0,270	0,287	0,194
0,192	0,218	0,244	0,272	0,293	0,192
0,189	0,216	0,239	0,268	0,287	0,189
0,177	0,203	0,226	0,252	0,275	0,177
0,166	0,188	0,209	0,233	0,249	0,166
0,155	0,177	0,195	0,220	0,238	0,155
0,149	0,170	0,188	0,212	0,227	0,149
0,136	0,155	0,174	0,192	0,211	0,136
0,125	0,144	0,161	0,181	0,192	0,125
0,116	0,134	0,150	0,166	0,177	0,116
0,112	0,128	0,141	0,158	0,173	0,112
0,106	0,121	0,135	0,149	0,160	0,106

1.5. İki Değişkenli Dağılımlar

İki değişkenli dağılımlar üç boyutlu bir uzayda tanımlı fonksiyonlardır ve matematiksel olarak, $(F(x, y): (x, y) \in \mathbb{R})$ biçiminde gösterilmektedir. İki değişken

birbirinden bağımsız ise bunların birlikte oluşturduğu dağılıma ortak olasılık fonksiyonu denir.

1.5.1. Ortak Olasılık Yoğunluk Fonksiyonu

X ve Y sürekli rastgele değişkenler olmak üzere, X ve Y 'nin ortak olasılık yoğunluk fonksiyonu $(f(x, y))$ üç boyutlu uzayda bir yüzey oluşturmaktadır. Bu yüzey fonksiyonu,

$$f(x, y) \geq 0 \quad \forall x, y \quad (33)$$

koşuluna göre her (x, y) noktasında sifıra eşit veya sıfırdan büyük bir pozitif değere sahip olmalıdır.

$f(x, y)$ ortak olasılık yoğunluk fonksiyonu ile $((x, y)$ düzlemi arasında) altında kalan hacim 1'e eşit olmalıdır (Mark ve Turin, 2011; Walck, 2007). Bu ifadeye göre,

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1 \quad (34)$$

koşulunu sağlamalıdır. Şayet olasılık yoğunluk fonksiyonu dar bir dörtgen bölgede $([a, b] \times [c, d])$ tanımlı ise,

$$\int_a^b \int_c^d f(x, y) dy dx = 1 \quad (35)$$

koşulunu sağlamalıdır.

1.5.2. Ortak Birikimli Dağılım Fonksiyonu

X ve Y sürekli rastgele değişkenler olmak üzere, X ve Y 'nin ortak birikimli dağılım fonksiyonu üç boyutlu uzayda bir hacim oluşturmaktadır. X ve Y 'nin ortak birikimli dağılım fonksiyonu 36'deki eşitlikle,

$$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dy dx \quad (36)$$

biçiminde hesaplanmaktadır (Mark ve Turin, 2011; Walck, 2007). Bu eşitlik,

$$0 \leq F_{XY}(x, y) \leq 1 \quad (37)$$

koşulunu sağlamalıdır. Şayet olasılık yoğunluk fonksiyonu dar bir dörtgen bölgede $([a, b] \times [c, d])$ tanımlı ise ortak birikimli dağılım fonksiyonu,

$$F_{XY}(x, y) = \int_a^x \int_c^y f(x, y) dy dx = 1 \quad (38)$$

biçiminde hesaplanır.

1.6. İki Değişkenli Uyum İyiliği Testleri

Tek değişkenliye göre yaygın kullanımı olmasa da bazı durumlarda çözüm üretebilir. İki değişkenli bir uzayda olasılık yoğunluk fonksiyonu bir yüzey teşkil etmekte ve dağılım fonksiyonu ise bu yüzeyin altında kalan hacmi vermektedir.

1.6.1. Ki –Kare Uyum İyiliği Testi

Bu testteki istatistik sapmaların karelerinin toplamının beklenen frekans sayısına bölünmesiyle elde edilmiştir. Ki-kare uyum iyiliği testi birçok uygulamada gözlenen değerlerin tek değişkenli dağılımlara uyup uymadığını kontrol etmek için kullanılmıştır. Ancak iki boyutlu bir alanda elde edilen gözlem değerleri test edilmesi için dörtgensel bir alanla sınıflandırılması gerekmektedir (Şekil 13).

İki değişkenli sürekli fonksiyonlarda ki-kare testi yapılabilmesi için verilen sınıf sayısına göre sınıf bölgesinin belirlenmesi gerekir. Sınıf sayısı $(m_x \times m_y)$ biçimindeki iki bölen sayı tarafından belirlenir. Sınıf genişlikleri,

$$\Delta x = \frac{x_n - x_0}{m_x} \quad (39)$$

ve

$$\Delta y = \frac{y_n - y_0}{m_y} \quad (40)$$

biçiminde belirlenir. Burada x_0 ve y_0 verideki en küçük değerleri, x_n ve y_n ise verideki en büyük değeri göstermektedir. (i, j) bölgesinin alt sınırları,

$$Lx_i = x_0 + (i - 1) \cdot \Delta x, \quad i = 1, 2, \dots, m_x \quad (41)$$

ve

$$Ly_j = y_0 + (j - 1) \cdot \Delta y, \quad j = 1, 2, \dots, m_y \quad (42)$$

biçiminde hesaplanırken üst sınırları

$$Ux_i = x_0 + i \cdot \Delta x, \quad i = 1, 2, \dots, m_x \quad (43)$$

ve

$$Uy_j = y_0 + j \cdot \Delta y, \quad j = 1, 2, \dots, m_y \quad (44)$$

biçiminde hesaplanmaktadır.

Her sınıf aralığına düşen gözlem sayısı (gözlenen frekanslar),

$$O_{i,j} = \sum_{k=1}^n \begin{cases} 1, & Lx_i \leq x_k < Ux_i \text{ ve } Ly_j \leq y_k < U_j \\ 0, & \text{diğer yerlerde} \end{cases}, \quad \begin{pmatrix} i = 1, 2, \dots, m_x \\ j = 1, 2, \dots, m_y \end{pmatrix} \quad (45)$$

biçiminde hesaplanırken beklenen frekanslar,

$$E_{i,j} = n[F_0(Ux_j, Uy_j) - F_0(Lx_i, Ly_j)], \quad \begin{matrix} (i = 1, 2, \dots, m_x; \\ j = 1, 2, \dots, m_y) \end{matrix} \quad (46)$$

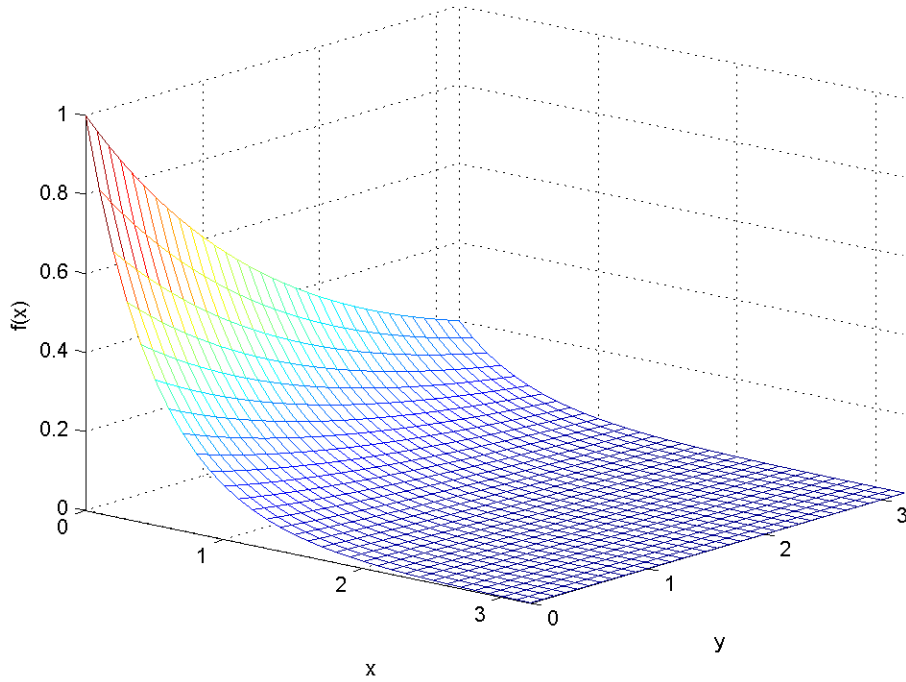
biçiminde hesaplanmaktadır. Şekil 13'te olasılık yoğunluk fonksiyonu,

$$f(x, y) = e^{(-2x-y/2)} \quad (47)$$

biçiminde örnek bir fonksiyon seçilmiştir. Bu fonksiyonun birikimli dağılım fonksiyonu,

$$\begin{aligned} F_0(x, y) &= \int_0^x \int_0^y e^{(-2x-y/2)} \\ &= (e^{-2x} - 1)(e^{-y/2} - 1) \end{aligned} \quad (48)$$

biçiminde hesaplanmaktadır.

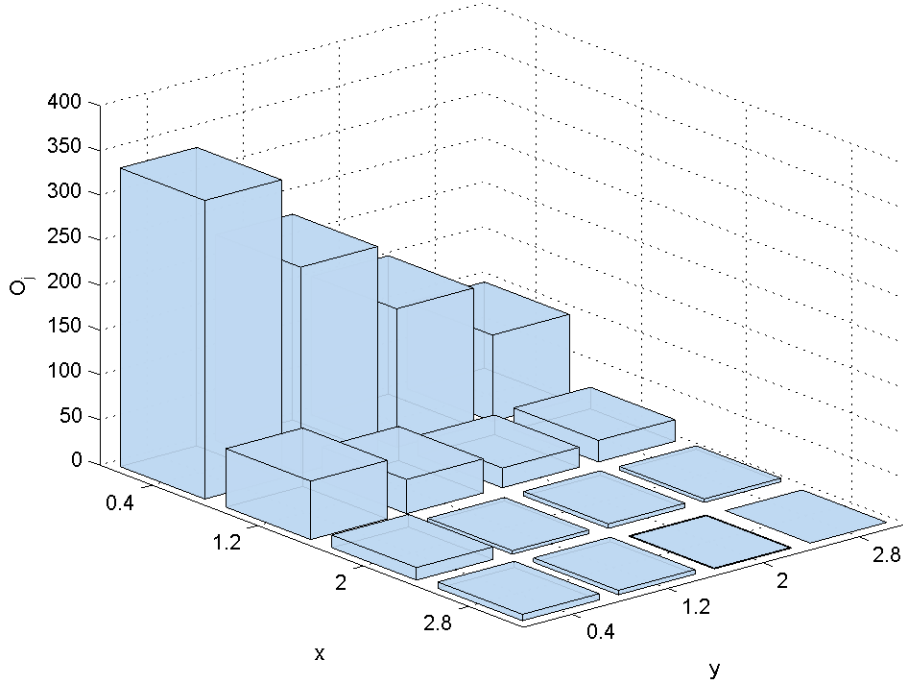


Şekil 13. İki değişkenli uygulama için örnek olasılık yoğunluk fonksiyonu

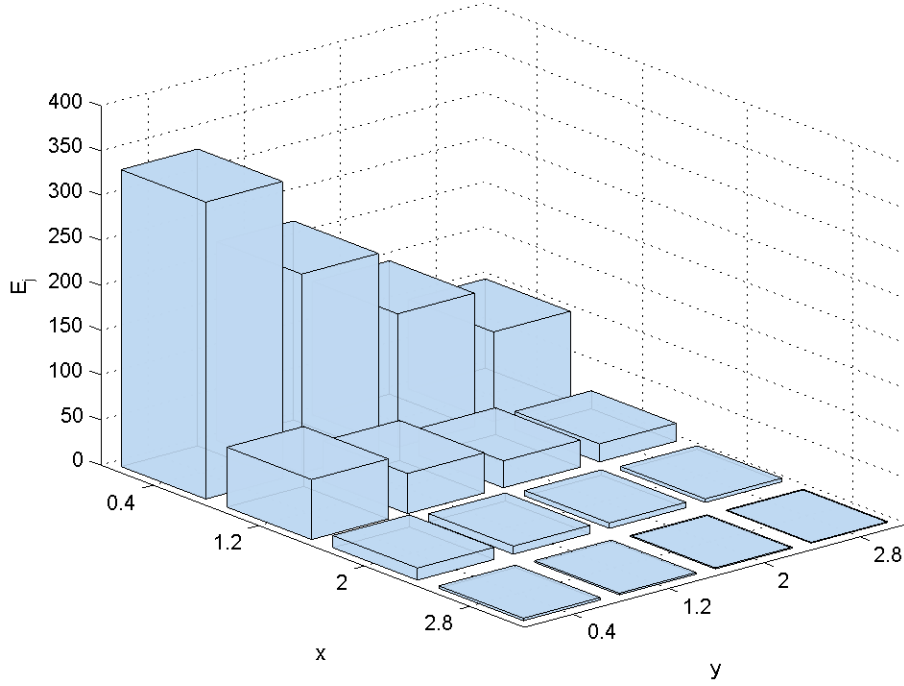
Benzetim için eşitlik (48)'deki dağılımdan rastgele 1000 tane sayı üretilmiştir. Bu sayılar $m_x = 4$ ve $m_y = 4$ bölenler yardımıyla 16 sınıfa bölünerek ($\Delta x, \Delta y = 0.8$)

sınıfların gözlenen frekansları bulunmuştur (Şekil 14). Bu gözlenen frekanslara karşılık beklenen frekanslar (46) eşitliği yardımıyla hesaplanmıştır (Şekil 15).

Hesaplama sonucunda hem gözlenen hem de beklenen frekansları 4×4 lük bir matris biçiminde verilmektedir. Veriler iki boyutlu olmasına rağmen aynı kategori içerisinde değerlendirildiğinden tek bir sütun matrisine dönüştürülürse Tablo 8'deki beklenen (E_j) ve gözlenen (O_j) frekanslar verilir. Bulunan bu frekanslar yardımıyla klasik Ki-kare testi uygulanırsa test istatistiği $\chi^2 = 20.68$ biçiminde bulunur. Toplam 16 sınıf için serbestlik derecesi $\nu = 15$ olarak alınıp $\alpha = 0.05$ anlamlılık payına göre tablo değeri $\chi^2_{\nu=15,1-\alpha=0.95} = 25.00$ olarak bulunur. Bu sonuçlara göre $\chi^2 < \chi^2_{\nu,1-\alpha}$ eşitsizliği doğrulanarak ve öne sürülen hipotez kabul edilmiştir.



Şekil 14. İki değişkenli uygulama için gözlenen frekanslar



Şekil 15. İki değişkenli uygulama için beklenen frekanslar

Tablo 8. İki değişkenli Ki-kare işlem tablosu

y_j	x_j	O_j	E_j	$(O_j - E_j)$	$(O_j - E_j)^2$	$\frac{(O_j - E_j)^2}{E_j}$
0,40	0,40	332,00	330,22	1,78	3,17	0,01
0,40	1,20	65,00	66,67	-1,67	2,79	0,04
0,40	2,00	14,00	13,46	0,54	0,29	0,02
0,40	2,80	7,00	2,72	4,28	18,34	6,75
1,20	0,40	229,00	221,35	7,65	58,48	0,26
1,20	1,20	38,00	44,69	-6,69	44,76	1,00
1,20	2,00	4,00	9,02	-5,02	25,23	2,80
1,20	2,80	5,00	1,82	3,18	10,10	5,55
2,00	0,40	154,00	148,38	5,62	31,62	0,21
2,00	1,20	22,00	29,96	-7,96	63,31	2,11
2,00	2,00	5,00	6,05	-1,05	1,10	0,18
2,00	2,80	1,00	1,22	-0,22	0,05	0,04
2,80	0,40	96,00	99,46	-3,46	11,97	0,12
2,80	1,20	24,00	20,08	3,92	15,36	0,76
2,80	2,00	4,00	4,05	-0,05	0,00	0,00
2,80	2,80	0,00	0,82	-0,82	0,67	0,82

$$\hat{\chi}^2 = 20.68$$

1.6.2. EDF Uyum İyiliği Testleri

Çok değişkenli Kolmogorov-Smirnov uyum iyiliği testi Justel, Peña ve Zamar (1997) tarafından önerilmiştir. Bu test Rosenblatt's dönüşümünü kullanarak elde edilen bir test istatistiğini kullanır. Yalnız bu istatistik iki değişkenli durum için geliştirilmiştir. Ancak istenirse N-değişkenli bir durum içinde genişletilebilir.

Kolmogorov-Smirnov test istatistiği veya benzer test istatistikleri çok değişkenli verilere uyarlanması gerekir. İki ortak birikimli dağılım fonksiyonu arasındaki farktan dolayı kolaylıkla hesaplanamaz. Bu maksimum fark,

$$\begin{aligned}
 &P(X < x \text{ ve } Y < y) \\
 &P(X < x \text{ ve } Y > y) \\
 &P(X > x \text{ ve } Y < y) \\
 &P(X > x \text{ ve } Y > y)
 \end{aligned} \tag{49}$$

olasılıklardan birisini içerebilir. Dolayısıyla test yapmadan önce bunlardan birisi tercih edilmelidir. Bu çalışmada $P(X < x \text{ ve } Y < y)$ olasılık değerleri kullanılmıştır.

Bu fonksiyonun birikimli dağılım fonksiyonu,

$$\begin{aligned}
 F_0(x, y) &= \int_0^x \int_0^y e^{(-2x-y/2)} \\
 &= (e^{-2x} - 1)(e^{-y/2} - 1)
 \end{aligned} \tag{50}$$

biçiminde bulunur.

İki değişkenli Kolmogorov-Smirnov testi için $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ örnek değerlerinin bağımsız, tanımlı ve sürekli bir dağılımdan elde edildiği varsayılır. Bu dağılım için aşağıdaki hipotez,

$$H_0: F(x, y) = F_0(x, y), \quad \forall x, y, \tag{51}$$

kurulmakta ve bu hipotezin doğruluğu test edilmek istenmektedir. Burada $F_0(x, y)$ dağılımı bilinen bir birikimli dağılım fonksiyondur. Kurulan hipotezi doğruluğunu test etmek için öncelikle test istatistiğinin belirlenmesi gerekir. Kolmogorov-Smirnov test istatistiği \widehat{D} ,

$$\widehat{D} = \sup_{(x,y) \in \mathbb{R}^2} |F_n(x, y) - F_0(x, y)|, \quad (52)$$

biçiminde bulunur. Burada $F_n(x, y)$ fonksiyonu deneysel birikimli dağılım fonksiyonu olarak tanımlanmakta ve

$$F_n(x, y) = \frac{k_n(x, y)}{n}, \quad \{k_n(x, y): x_k \leq x \wedge y_k \leq y\} \quad (53)$$

biçiminde bulunmaktadır. (52) eşitliğinde supremum hesaplarırken her x_i noktasındaki değeri hesaplanır. x_k noktasının solunda ve y_k altında karşılaşılan ilk değere kadar aynı deneysel birikimli dağılım fonksiyonu kullanılır.

Deneysel dağılım fonksiyonu hesaplanırken k . gözlem değerinden küçük eşit olanlar $\{x_k \leq x \vee y_k \leq y\}$ işleme katılır (Şekil 16). Ancak bazı çalışmada sadece küçük değerler $\{x_i < x \vee y_k < y\}$ işleme katılmaktadır (Şekil 17). Öte yandan en sağlam karşılaştırma ise ikisinin birlikte kullanılmasıdır. Bu durumda deneysel dağılım fonksiyonu,

$$F_n^+(x, y) = \frac{k_n(x, y)}{n}, \quad \{k_n(x, y): x_k \leq x \wedge y_k \leq y\} \quad (54)$$

ve

$$F_n^-(x, y) = \frac{k_n(x, y)}{n}, \quad \{k_n(x, y): x_k < x \wedge y_k < y\} \quad (55)$$

biçiminde iki farklı fonksiyon olarak bulunmaktadır. Test istatistikleri ise,

$$\widehat{D}^+ = \sup_{(x,y) \in \mathbb{R}^2} |F_n^+(x, y) - F_0(x, y)|, \quad (56)$$

ve

$$\hat{D}^- = \sup_{(x,y) \in \mathbb{R}^2} |F_n^-(x,y) - F_0(x,y)|, \quad (57)$$

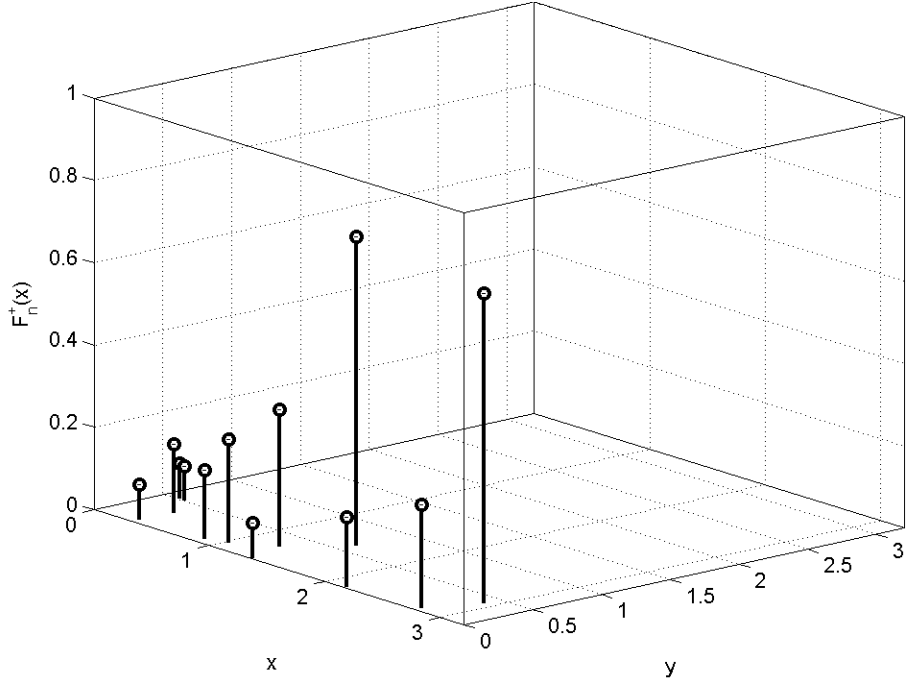
biçiminde bulunurlar. Ortak test istatistiği ise,

$$\hat{D} = \max(\hat{D}^-, \hat{D}^+) \quad (58)$$

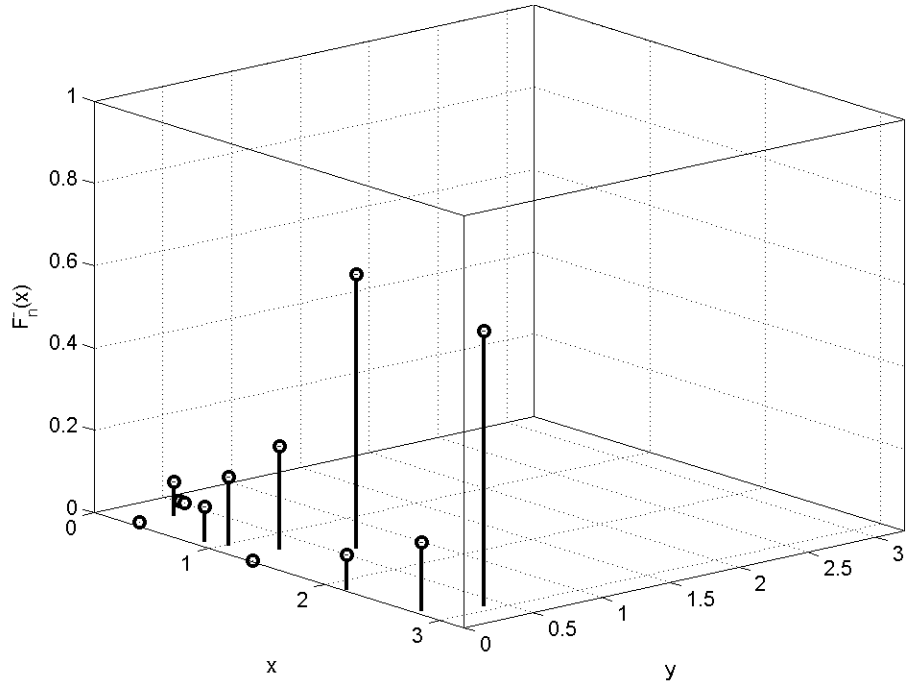
olarak bulunur. Şekil 18'de her veri noktası (x_k, y_k) için bulunan birikimli dağılım fonksiyonun değerleri gösterilmiştir.

Benzetim için bölüm 1.5.1'de verilen örnek ele alınmış ve iki değişkenli dağılımlar için Kolmogorov-Smirnov test istatistiğini bulmak için Tablo 9'deki işlem tablosu verilmiştir.

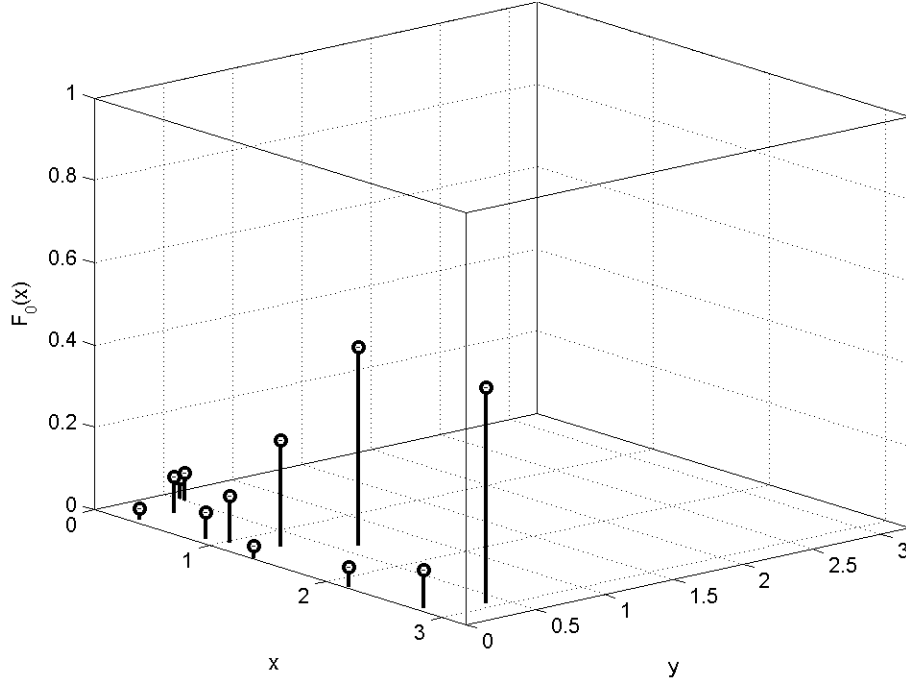
Hesaplama sonucunda hem teorik hem de deneysel dağılım değerleri bulunmuştur. Bulunan bu değerler yardımıyla Kolmogorov-Smirnov testi uygulanırsa test istatistiği $\hat{D} = 0.27$ biçiminde bulunur. Toplam 16 örnek alınıp $\alpha = 0.05$ anlamlılık payına göre tablo değeri (kritik değeri) $KS_{1-\alpha=0.95}^* = 0.34$ olarak bulunur. Bu sonuçlara göre; $\hat{D} < KS_{1-\alpha=0.95}^*$ eşitsizliği doğrulanarak ve öne sürülen hipotez kabul edilmiştir.



Şekil 16. Deneysel dağılım fonksiyonu $(x_i \leq x, y_i \leq y)$



Şekil 17. Deneysel dağılım fonksiyonu $(x_i < x, y_i < y)$



Şekil 18. Teorik dağılım fonksiyonu

Tablo 9. İki değişkenli Kolmogorov-Smirnov işlem tablosu

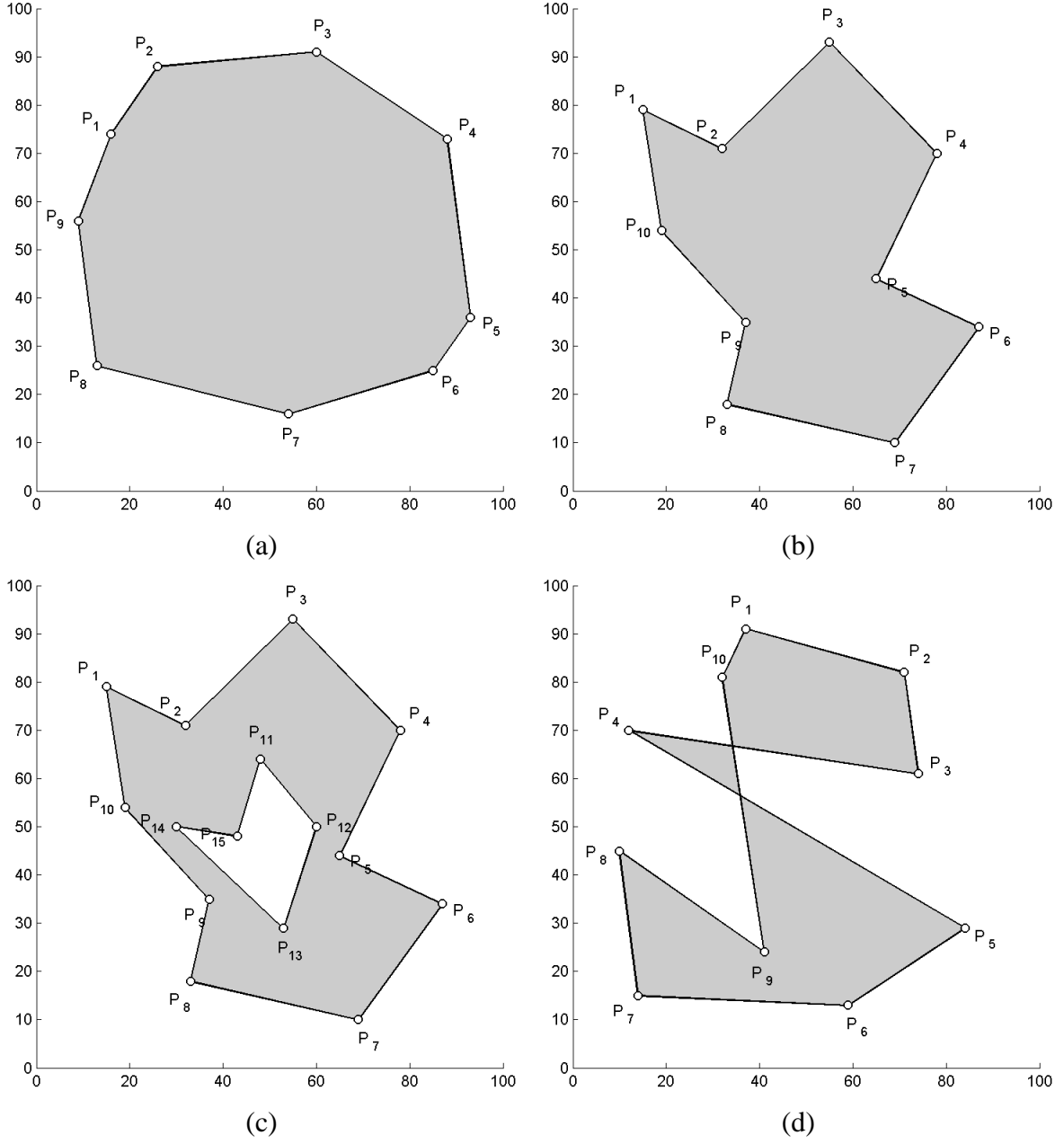
y_i	x_i	$F_n^-(x_i)$	$F_n^+(x_i)$	$F_0(x_i)$	\hat{D}^-	\hat{D}^+	$\max(\hat{D}^-, \hat{D}^+)$
0,02	1,34	0,00	0,08	0,03	0,03	0,06	0,06
0,53	0,11	0,00	0,08	0,05	0,05	0,04	0,05
0,08	0,87	0,08	0,17	0,06	0,02	0,10	0,10
0,39	2,91	0,67	0,75	0,52	0,14	0,23	0,23
0,30	0,34	0,08	0,17	0,09	0,00	0,08	0,08
0,52	0,17	0,00	0,08	0,07	0,07	0,02	0,07
0,63	1,53	0,67	0,75	0,48	0,18	0,27	0,27
0,03	2,16	0,08	0,17	0,05	0,04	0,12	0,12
0,28	1,26	0,25	0,33	0,26	0,01	0,08	0,08
0,07	0,31	0,00	0,08	0,02	0,02	0,06	0,06
0,05	2,78	0,17	0,25	0,09	0,07	0,16	0,16
0,13	1,01	0,17	0,25	0,11	0,05	0,14	0,14

$$\hat{D} = 0,27$$

1.7. Çokgenler

Bu çalışmada geometrik alanlar yardımıyla birçok çalışma gerçekleştirilmiştir. Bu çalışmalarda yapılacak tüm hesaplamalar çokgensel bölgeleri temel alarak yapılmıştır.

Çokgenler düzlem geometrisinde ve sayısal matematikte yaygın bir kullanım alanı bulunmaktadır. Çokgenler noktalardan ve doğru parçalarından oluşmaktadır. Bu doğru parçaları noktaları birleştirmek için kullanılmaktadır (Devadoss ve O'Rourke, 2011). Bir çokgen (Ω) sonlu sayıda doğru parçası tarafından sınırlandırılmış kapalı düzlem bölgesi olarak tanımlanır (Şekil 4). Buradaki her doğru parçası kenar ve her nokta ise düğüm olarak tanımlanmaktadır. Basit tanımlı bir çokgende doğrular birbirlerini kesmezler. Şekil 4(a)'da dışbükey bir çokgen gösterilmiştir. Dışbükey çokgenler tüm iç açıları 180° 'den küçük olan çokgenlerdir. Şekil 4(b)'de ise içbükey bir çokgen gösterilmiştir. İçbükey çokgenler iç açılarından herhangi biri 180° 'den büyük olan çokgenlerdir. Şekil 4(c) İç içe bir çokgen ve Şekil 4(d) ise kenarları kesişmiş bir çokgen verilmiştir. Bu iki çokgen basit çokgen olarak kabul edilmemektedir (Devadoss ve O'Rourke, 2011). Bu çalışmada basit çokgenler temel alınmıştır.

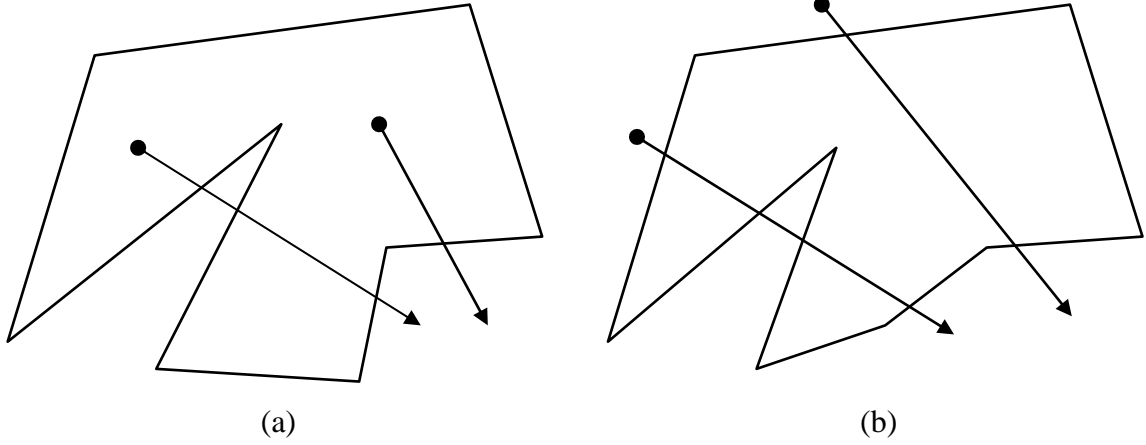


Şekil 19. Çokgen tanımlaması; (a) Dışbükey bir çokgen; (b) İçbükey bir çokgen; (c) İç içe girmiş çokgen; (d) Kenarları kesişmiş çokgen.

1.7.1. Çokgendeki Nokta Problemi

Hesaplamalı geometride en yaygın problemlerden birisi, verilen bir noktanın (R) keyfi bir çokgen içerisinde olup olmadığının bulunmasıdır. Bu çokgenler bir daire veya bir dörtgen olduğunda çözüm oldukça basittir. Ancak çeşitkenar bir çokgen olduğunda bu iş o kadar kolay yapılmıyor. Bu konuda literatürde en yaygın iki yöntem bulunmaktadır (Foley vd., 1990; Haines, 1994; Harrington, 1983; O'Rourke, 1998; Sedgewick, 1988; Weiler,

1994; Hormann ve Agathos, 2001). Bunlardan birincisi Tek-Çift yöntemidir. Bu yöntemde çokgen içerisinde olup olmadığı incelenen bir nokta (R) ile çokgen dışında bir nokta (S) arasında bir doğru çekilir. Bu doğru çokgeni kenarlarından bazılarını keser. Kestiği noktaların sayısı tek ise (R) noktası çokgenin içerisinde, eğer çift ise çokgenin dışındadır (Şekil 20).



Şekil 20. Çokgendeki nokta problemi; (a) Seçilen noktalar çokgenin içinde; (b) Seçilen noktalar çokgeni dışında

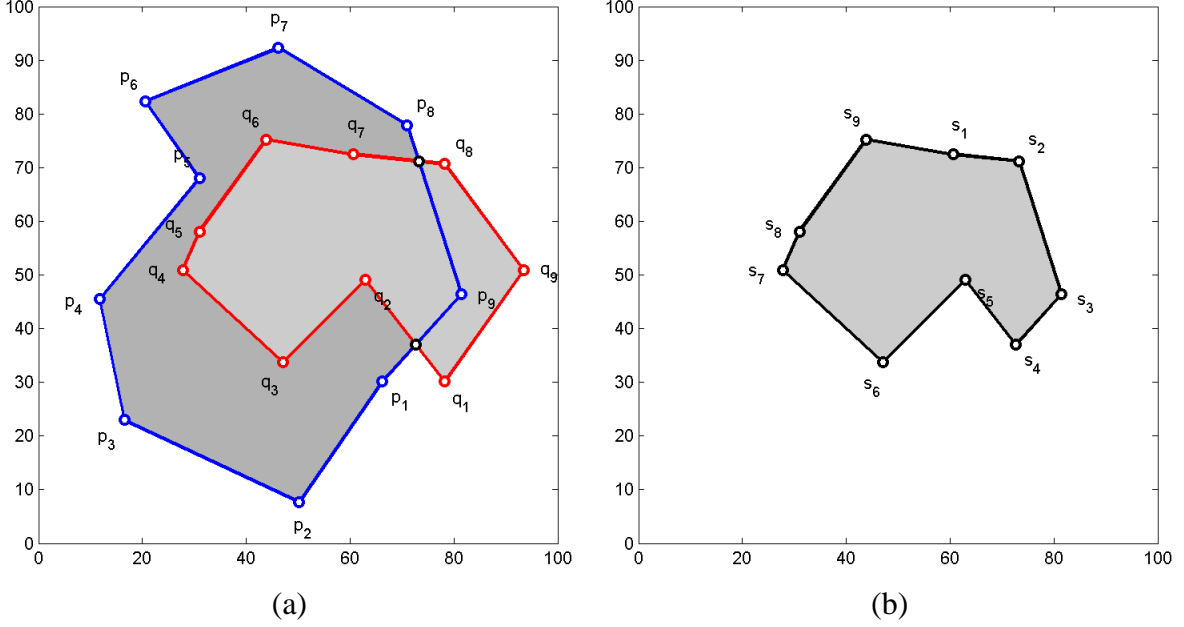
İkinci yöntem ise Artı-Eksi yöntemidir. Çokgenin noktalarının saat yönünde sıralandığını varsayalım. Eğer verilen nokta çokgenin tüm kenarlarının sağında ise nokta çokgen içerisindedir. Eğer nokta kenarların herhangi birinin solunda ise çokgenin dışındadır. Aynı şekilde çokgenin noktaları saat yönünün tersi yönde sıralanmışsa verilen nokta çokgen kenarlarının solunda olursa nokta içeride değilse dışındadır.

Bu çalışmada basitlik açısından tek çift yöntemi tercih edilmiştir.

1.7.2. İki Çokgenin Kesişim Çokgeni

Hesaplamalı geometride önemli diğer bir konu ise mantıksal operatörlerin kullanımınıdır. İki çokgenin birleşimi, kesişimi ve farkı hesaplamalı geometride, bilgisayar bilimlerinde, yöneylem araştırmaları gibi alanlarda kullanılmaktadır (Toussaint, 1985). Bu çalışmada iki çokgenin kesişimi üzerinde durulacaktır. Bu problemdeki konu verilen iki çokgenin kesişim bölgesini tanımlayan bölgenin tanımlanmasıdır (Şekil 21(a)). Bu bölge çokgensel bir bölge olarak belirleneceğinden köşe noktaları bulunmaya çalışılır (Şekil 21(b)). Literatürde iki çokgenin kesişimi ile ilgili birçok çalışma bulunmaktadır (O'Rourke

vd., 1982). Bu çalışmada iki çokgenin birleşiminde dışarıda kalan çokgen köşelerinin kırılmasıyla elde edilen noktalar kümesi birleştirilerek kesişim bölgesi bulunmuştur.



Şekil 21. İki çokgenin kesişimi; (a) P ve Q çokgenleri; (b) Kesişim bölgesi

1.8. Üçgenleştirme

Üçgenleştirme, iki boyutlu bir düzlemde N tane noktanın birbirleriyle üçgen olacak biçimde birleştirilmesiyle elde edilen bir ağdır. Bu işlem gerçekleştirilirken noktaları birleştiren doğruların birbirini kesmemesine özen gösterilir. Tüm olası üçgenlerin içerisinde dar açılı üçgenlerin sayısının en yüksek yapacak üçgenler seçilir.

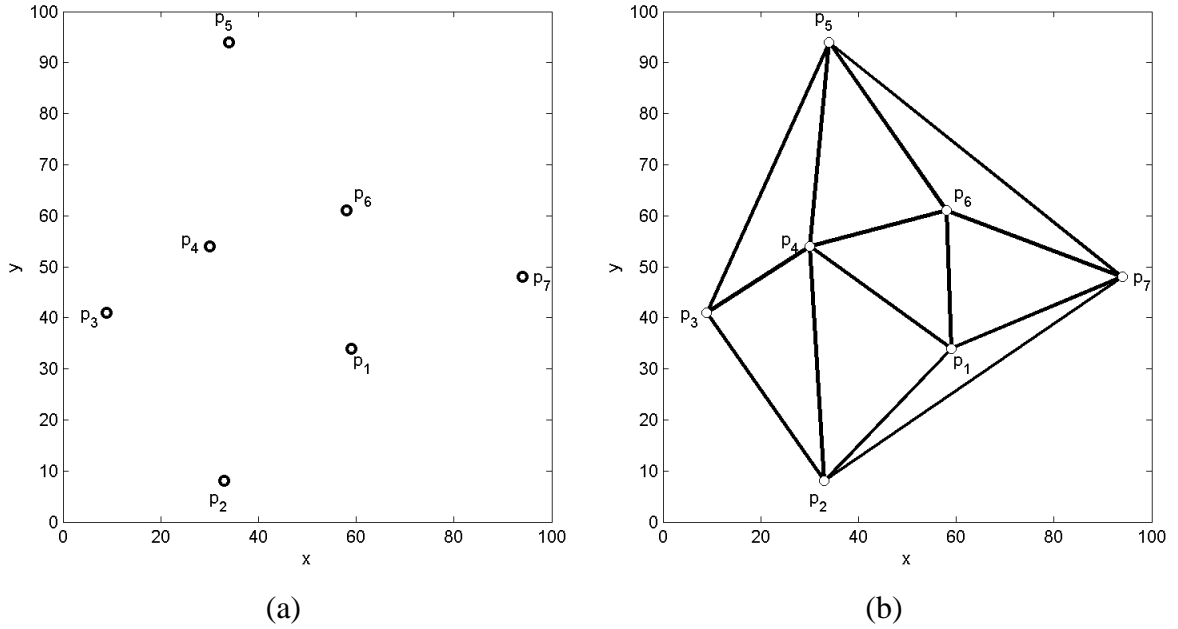
Üçgenleme algoritması, her üç noktadan bir çember geçecek ve hiçbir çevresel çemberin içinde nokta kalmayacak şekilde, bir bölgeyi üçgen parçalara bölen bir algoritmadır (Yanalak, 1997). Bu işlemi gerçekleştirilirken en basit yöntem tüm noktaların 3'lü kombinasyonları alınarak her kombinasyonda bulunan üç noktadan geçirilen çember içersinide eğer diğer noktalardan birisi varsa bu kombinasyon listeden silinir. En son kalan liste üçgenleştirme listesi olur.

Delaunay üçgenleştirme algoritması aşağıdaki belirtilen özellikleri sağlamaktadır (Yanalak, 1997; Worboys ve Duckham, 2004).

1. Algoritmanın çalışmaya başladığı ilk (başlangıç) noktası ne olursa olsun algoritma her seferinde aynı üçgenlemeyi yapmaktadır.

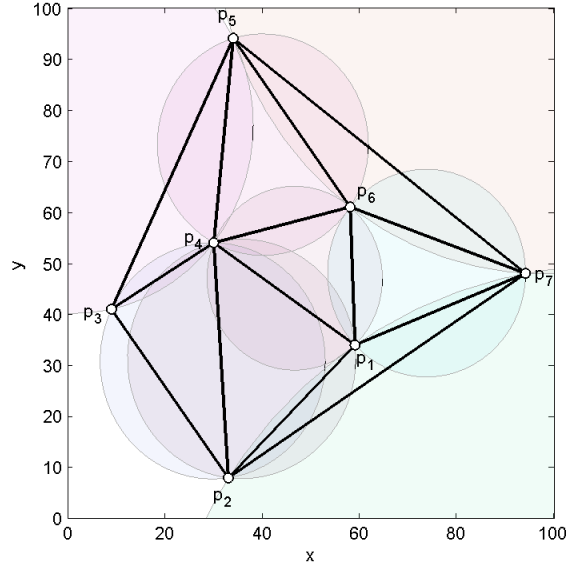
2. Oluşan üçgenlerin eş açılık özelliği vardır. Bu durum noktanın kendisinden uzakta bulunan noktalarla doğrudan üçgen kurabilmesinin önüne geçmektedir.
3. Algoritma sonucunda oluşan çevresel çemberlerin içerisinde nokta bulunmamaktadır.
4. Her bir noktayı birleştiren doğru parçaları oluşturulan üçgenlerin bir kenarını temsil etmektedir.

Şekil 22(a)'da iki boyutlu bir düzlemde rastgele seçilmiş noktalar bulunmaktadır. Bu noktalar Delaunay üçgenleştirme algoritması uyguladığında, Şekil 22(b)'deki gibi, üçgenlere bölünmüş bir alan elde edilmektedir.



Şekil 22. (a) Keyfi olarak çizilmiş çokgensel alan (b) Delaunay üçgenleştirme algoritması uygulanmış çokgensel alan

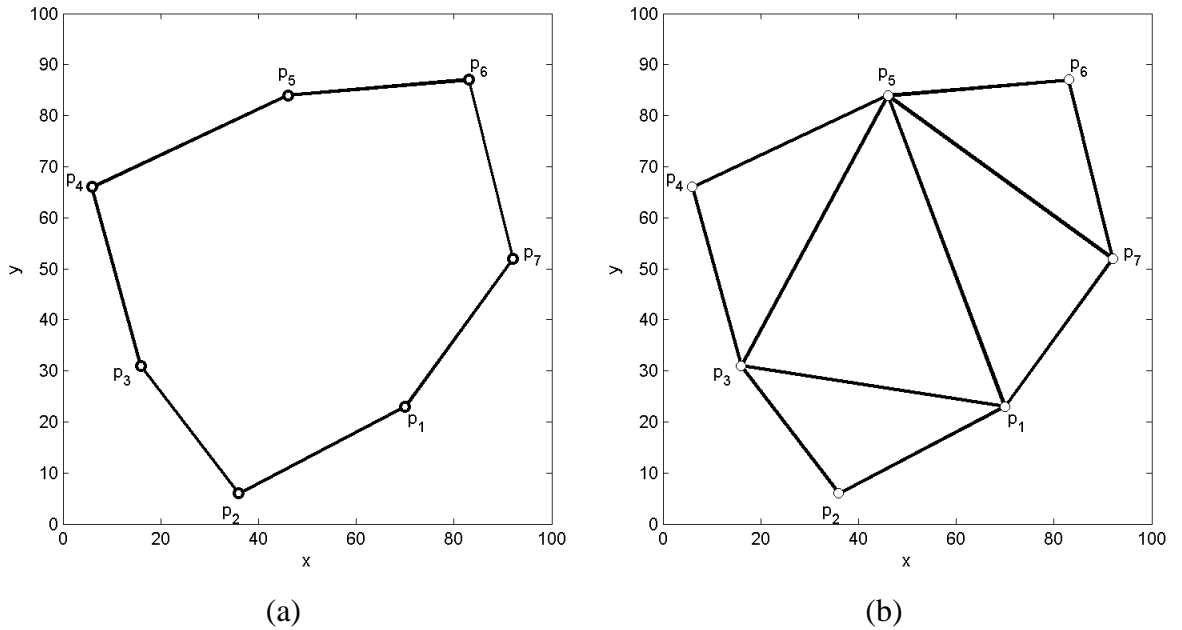
Şekil 23'te ise her üç noktadan bir çevresel çember geçecek şekilde çokgenin parçalandığını ve çizilen tüm çevresel çemberlerin içinde nokta bulunmadığı gösterilmiştir.



Şekil 23. Çevresel çemberlerin gösterimi

1.8.1. Dışbükey Çokgen Alanların Üçgenleştirilmesi

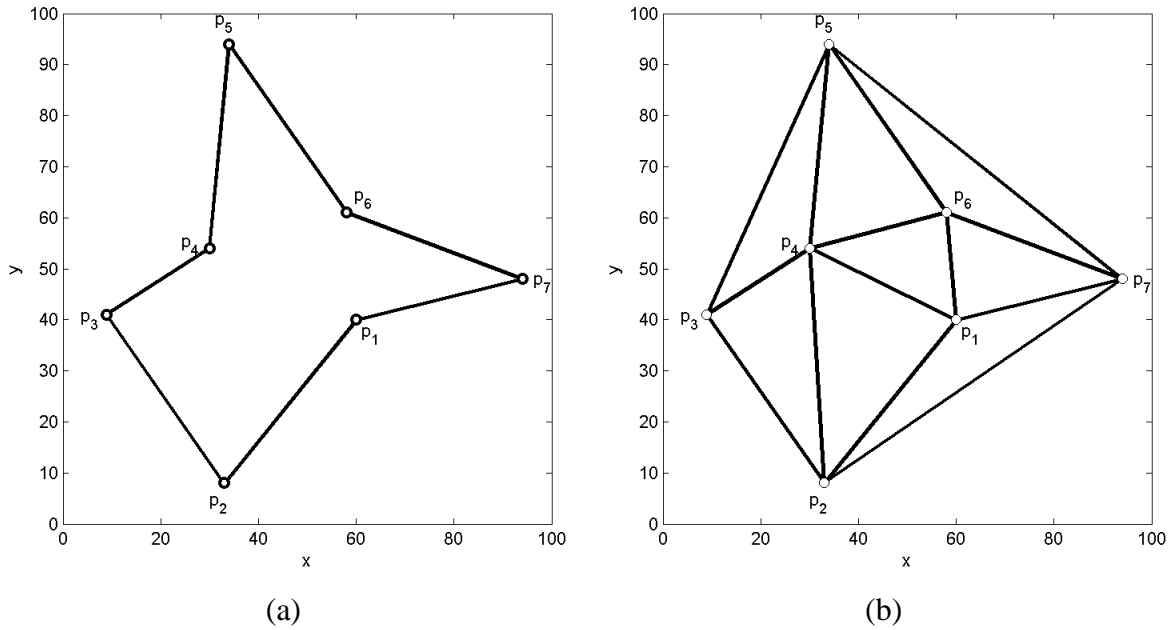
İki boyutlu bir ortamda rastgele saçılmış noktaların birleştirilmesiyle elde edilen üçgenlerin tümü dışbükey bir yapı oluşturmaktadır. Rastgele noktalar kümesi yerine çokgen oluşturulan bir noktalar kümesi verildiğine bu çokgen bölgenin üçgen parçalara ayrılması durumunda da dışbükey bir yapı oluşturmaktadır (Şekil 24).



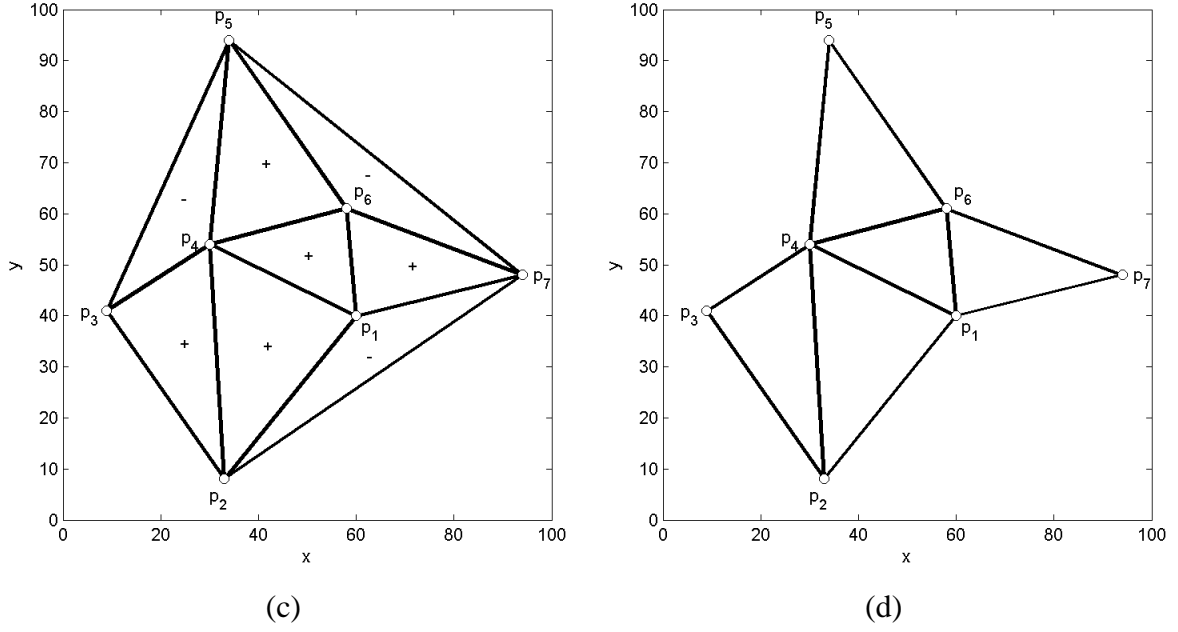
Şekil 24. (a) Keyfi olarak çizilmiş çokgensel alan (b) Delaunay üçgenleştirme algoritması uygulanmış çokgensel alan

1.8.1. İçbükey Çokgen Alanların Üçgenleştirilmesi

Dışbükey çokgenleri üçgenleştirirken üçgenlerin oluşturduğu yapı çokgeni kapsamaktadır. Oysa içbükey bir çokgen (Şekil 25(a)) üçgenleştirilirken üçgenlerin oluşturduğu yapı çokgenin dışına çıkarak dışbükey bir yapı göstermektedir (Şekil 25(b)). Bu durumda istemeyen üçgenlerin oluşması söz konusu olmaktadır. Bu sorunun çözmek için öncelikle üçgenleştirme sonucunda oluşan her bir üçgenin ağırlık merkezi bulunur (Şekil 26(a)). Ağırlık merkezleri çokgenin içinde olanlar saklanırken, çokgenin dışındakiler listeden silinmektedir (Şekil 26(b)).



Şekil 25. İçbükey çokgenlerin üçgenleştirilmesi; (a) İçbükey bir çokgen; (b) İçbükey çokgenin delaunay algoritması ile üçgenleştirilmesi

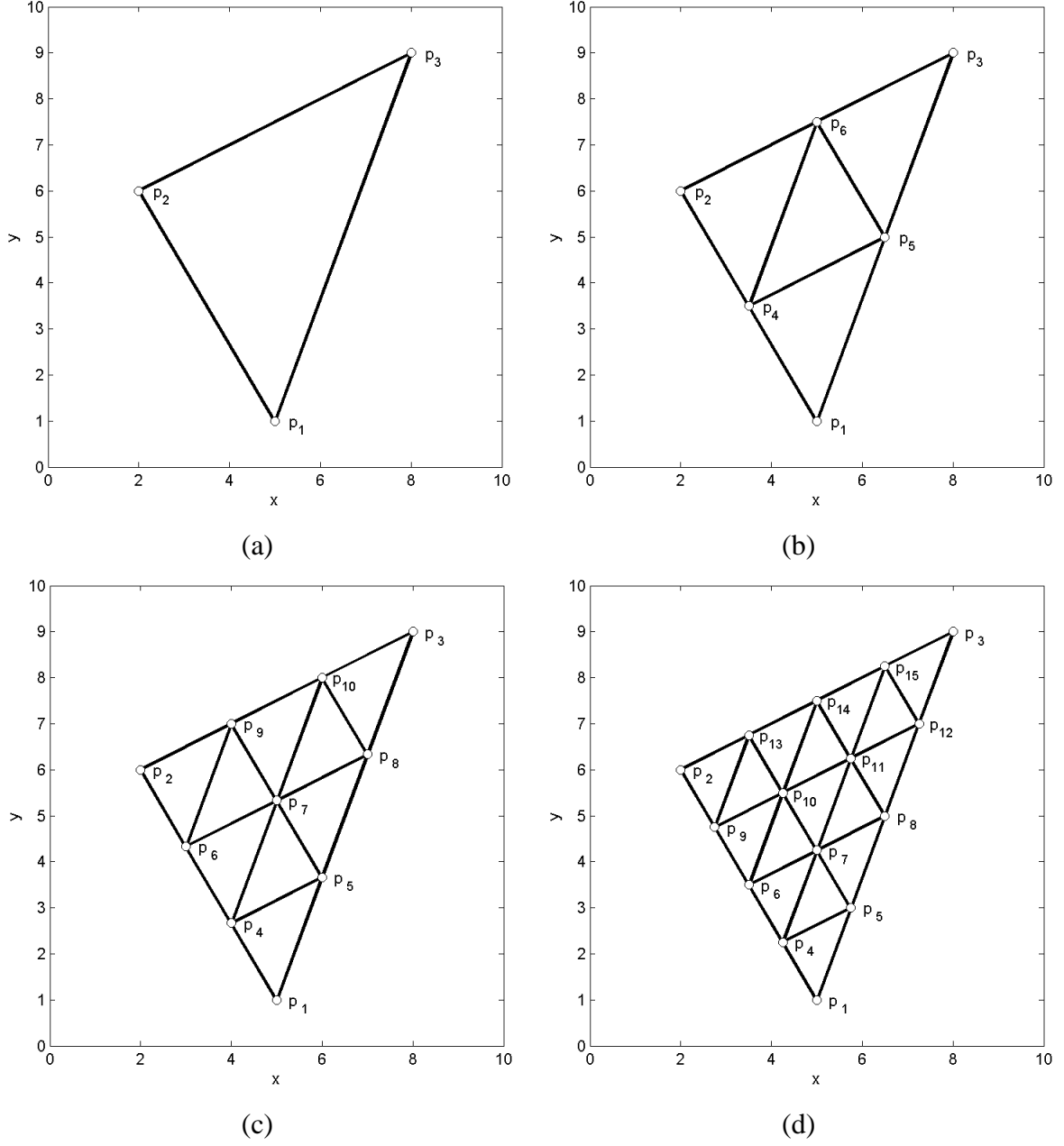


Şekil 26. İçbükey çokgenlerin üçgenleştirilmesindeki fazla üçgenlerin atılması; (a) Çokgen dışındaki üçgenlerin bulunması; (c) Çokgen dışındaki çokgenlerin silinmesi

1.8.2. Bir Üçgenin Alt Üçgenlere Bölünmesi

Çokgensel bir alan kenar noktaları yardımıyla üçgenlere ayrılmaktadır. Böylelikle çokgensel alan üzerinde yapılacak birçok işlem basit bir üçgen üzerine indirgenmektedir. Bazı durumlarda bu yeterli gelmeyebilir. Bu durumda her üçgenin tekrar küçük üçgenlere bölünmesi gerekir. Bu konudaki birçok yaklaşım (Schneider, 2002) olmasına rağmen, bu çalışmada kenarlardan eşit uzaklıkta paralel doğrularla üçgenler alt üçgenlere bölme yaklaşımı tercih edilmiştir.

Bu yaklaşıma göre her kenar eşit uzaklıkta K parçaya ayrılarak ayırım noktaları işaretlenir. Burada K değeri her kenarın bölünme sayıdır. Her kenarı K tane parçaya ayrılan kenarlardan çizilen paralel doğrular bir üçgeni K^2 tane küçük üçgene böler.



Şekil 27. Üçgenlerin alt üçgenlere bölünmesi; (a) $K = 1$ için üçgen bölünme gerçekleşmedi; (b) $K = 2$ için üçgen 4 alt üçgene bölündü; (c) $K = 3$ için üçgen 9 alt üçgene bölündü; (d) $K = 4$ için üçgen 16 alt üçgene bölündü

1.8.3. Tabanı Üçgen Olan Kesilmiş Prizmanın Hacim Hesabı

İki boyutlu bir düzlemde koordinatları bilinen N köşeden oluşan bir çokgenin alanı,

$$A = \frac{1}{2} \left| \sum_{i=1}^N (x_{i+1} y_i - x_i y_{i+1}) \right| \quad (59)$$

biçiminde verilmektedir (Preparata ve Ian, 1985). Bu eşitlikte $i = N$ olduğunda $i + 1 \rightarrow 1$ olarak bulunmaktadır. Eşitlik (28) çokgensel alan bir üçgen olduğunda da geçerlidir (Şekil 28(a)).

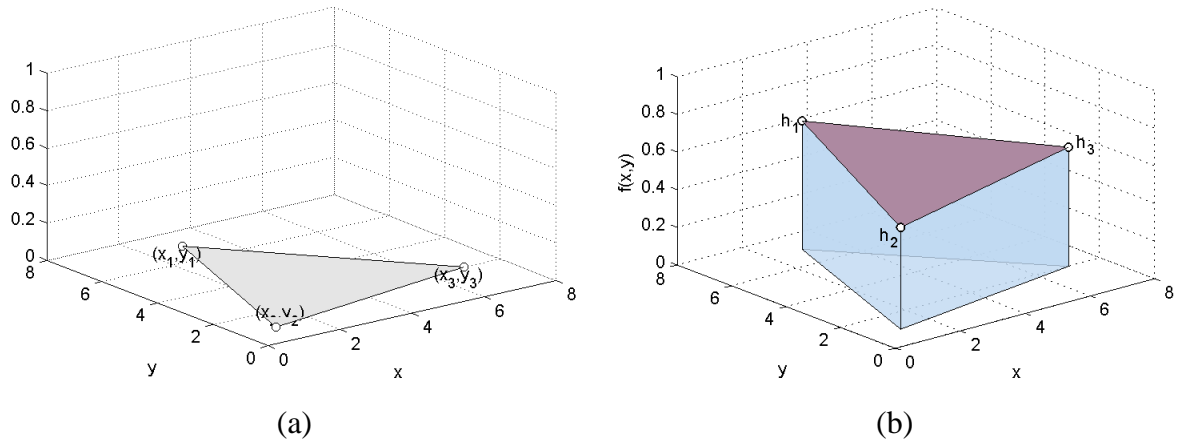
Üçgen taban alanı bilinen bir prizmanın hacmi,

$$V = A \cdot h \quad (60)$$

biçiminde verilmektedir. Burada h prizmanın yüksekliğidir. Eğer üçgen dik prizma kesilmiş yüzeye sahipse prizmanın hacmi

$$V = A \cdot \frac{h_1 + h_2 + h_3}{3} \quad (61)$$

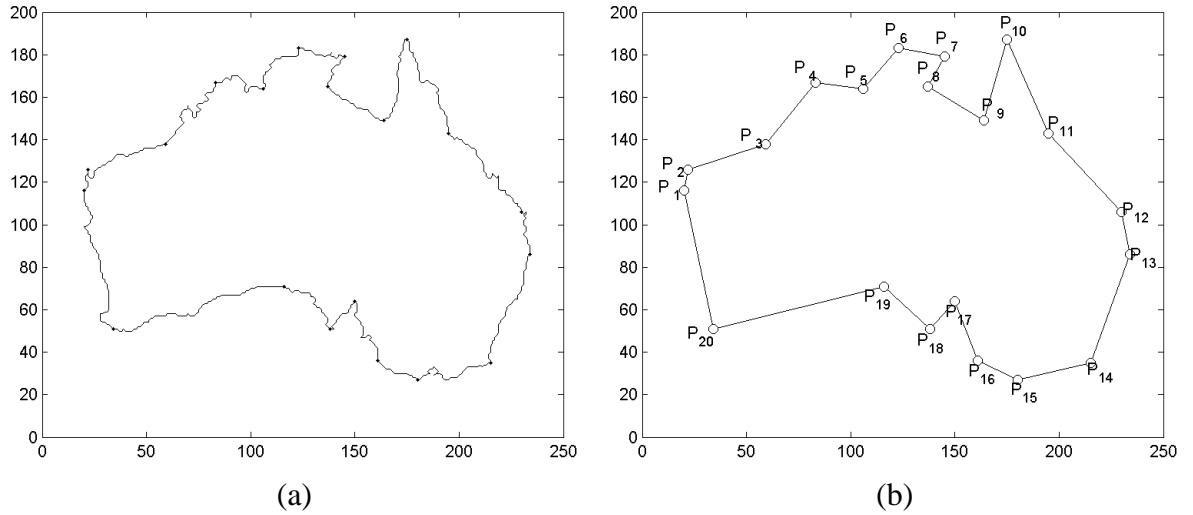
biçiminde bulunmaktadır. $\{h_1, h_2, h_3\}$ üçgen prizmanın köşe koordinatlarındaki yüksekliğidir.



Şekil 28. Kesilmiş üçgen prizmanın hacminin hesaplanması; (a) Taban alanı; (b) Kesilmiş üçgen prizma

2. YAPILAN ÇALIŞMALAR

Çokgensel alanla sınırlandırılmış bir bölgede iki değişkenli olasılık yoğunluk fonksiyonlarının uygulama alanlarına, bir şehirdeki kirlilik oranı veya suç oranı, bir ülkedeki deprem görülme sıklığının dağılımı, bir ormandaki belli bir ağaç türünün yoğunluğu (Kesemen ve Doğru, 2012), bir arazide zararlı böceklerin yayılım yoğunluğu, belirli bir bölgedeki trafik yoğunluğu, bir ülkedeki salgın hastalığın görülme konumları ve sıklıkları gibi örnekler verilebilir. Ayrıca ilgilenilen bölge dikdörtgen alan içerisinde birçok parçalı yoğunluk fonksiyonu olabilir. Bu konuda, fiziksel veya siyasi olarak bölünmüş birçok şehir örnek verilebilir (örneğin, Belfast, Beyrut, Kudüs, Mostar, ve Lefkoşe). Özellikle siyasi olarak bölünmüş şehirlerdeki veriler zamanla farklılaşmaya başlar. Bu durumda tüm şehri bir bölge gibi değerlendirmek mümkün olmayabilir. Bu örneklerin her birinde sınırları tamamen keyfi belirlenmiş bir bölgeden belirlenen olasılık yoğunluk fonksiyonu temel alınarak istatistik testleri yapılması gerekebilir. Keyfi olarak verilmiş alan (Şekil 29(a)) uygun bir çokgen yardımıyla tanımlanarak (Şekil 29(b)) işlemler gerçekleştirilebilir.



Şekil 29. İki boyutlu keyfi bir bölgenin (Avustralya) 20 köşeden oluşan çokgensel bir bölge ile yaklaşık temsili; (a) keyfi bölge; (b) çokgensel bölge

2.1. Çokgensel Alanda Dağılım Fonksiyonunun Hesaplanması

Keyfi bir alan içinde gözlenen verilerin bilindik bir geometrik şekil ile tanımlanmasında en uygun yöntem çokgensel tanımlamadır (Kesemen ve Doğru, 2012). Çokgensel bir alan içerisinde tanımlanan olasılık yoğunluk değerinden beklenen frekans sayısını bulmak için integralinin alınması gerekir. Bu ise oldukça zor olmaktadır. Önerilen yöntemde ise çokgensel alan uygun üçgenlere bölünerek (Şekil 30) hem sınıf sayısı belirlenebilir hem de bu alandaki beklenen frekans sayısı hesaplanabilmektedir.

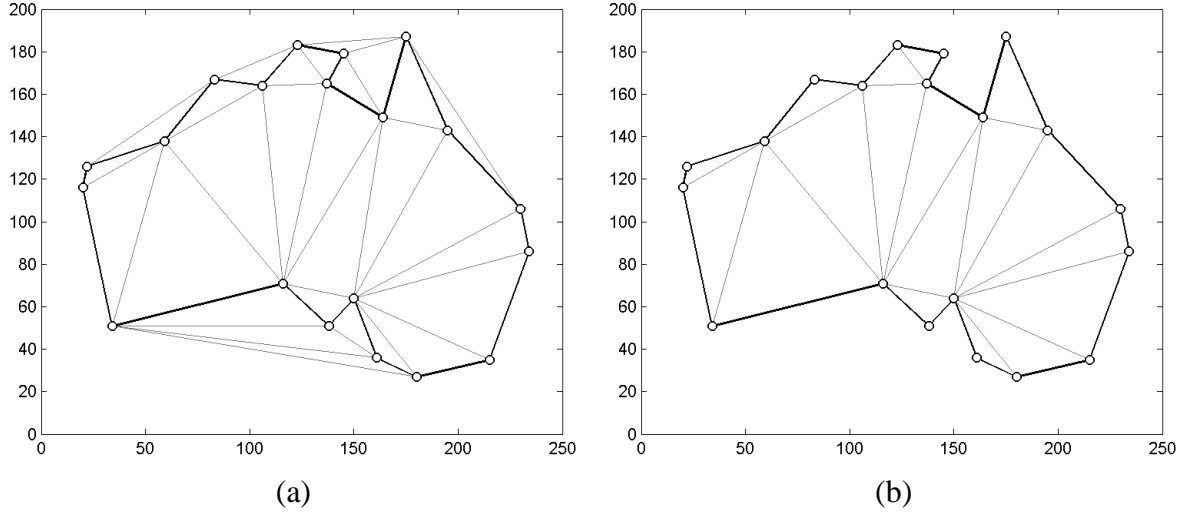
Çokgensel alanda dağılım fonksiyonunun hesaplanabilmesi için öncelikle çokgensel alan Delaunay üçgenleme algoritması yardımıyla üçgenlere bölünür. Ayrı ayrı her üçgen üzerinden integral alınarak o üçgenin olasılık değeri,

$$P((X, Y) \in \Omega_j) = \iint_{\Omega_j} f(x, y) dx dy \quad (62)$$

biçiminde bulunur. Burada Ω_j 'ler kesişimi boş küme olmak üzere; hesaplanan üçgensel bölgeyi, j ise seçilen üçgenin indisini vermektedir. Ω_j 'lerin birleşimi Ω bölgesini vereceğinden, her üçgenin olasılık değerlerinin toplamı tüm çokgenin olasılık değerinin bulunması sağlar.

$$P((X, Y) \in \Omega) = \sum_j P((X, Y) \in \Omega_j) \quad (63)$$

Olasılık değeri hesaplanırken öncelikle tercih edilen yöntem analitik hesaplamadır. Ancak çoğu durumda analitik hesaplama mümkün olmayabilir. Bu durumda sayısal yöntemler kullanılır.



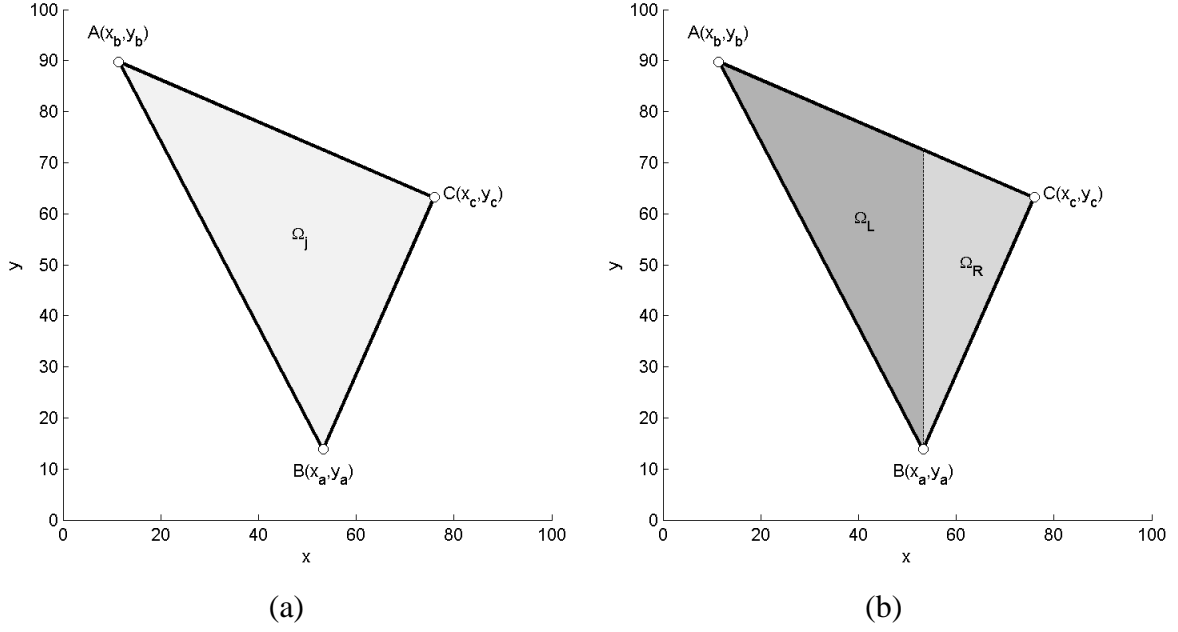
Şekil 30. Çokgenin üçgenlere bölünmesi; (a) İçbükey çokgenin üçgenleştirilmesi; (b) Çokgen dışında kalan üçgenlerin silinmesi

2.1.1. Analitik Hesaplama

Dörtgenel alanlarda iki değişkenli olasılık yoğunluk fonksiyonlarının sabit sınırlara göre integral hesabı oldukça kolay olmasına rağmen üçgenel alanda integral alma işlemi biraz karmaşıktır. Bu karmaşıklığı ortadan kaldırmak için Şekil 19a'da Ω üçgen bölgesinin köşe koordinatlarından x bileşenlerinin $\{x_a, x_b, x_c\}$ ortancası bulunur. Bu ortancadan y eksenine paralel olarak çekilen bir doğru üçgeni iki parçaya bölerek (Şekil 19b) integral alınmasını kolaylaştırmaktadır. Ω_j bölgesi iki üçgene bölünerek $(\Omega_{jL}, \Omega_{jR})$ ayrı ayrı integral hesabı yapılır ve bunların toplamı,

$$\begin{aligned}
 P((X, Y) \in \Omega_j) &= \iint_{\Omega_j} f(x, y) dx dy \\
 &= \iint_{\Omega_{jL}} f(x, y) dx dy + \iint_{\Omega_{jR}} f(x, y) dx dy
 \end{aligned} \tag{64}$$

biçiminde toplam integralle bulunabilmektedir.



Şekil 31. Üçgensel bölgenin gösterimi; (a) Ω_j bölgesi; (b) Ω_{jL} ve Ω_{jR} üçgensel bölgelerinin belirlenmesi.

Ω_{jL} üçgeninin marjinal fonksiyonu,

$$g_{jL}(x) = \begin{cases} \int_{\overline{AC}}^{\overline{AB}} f(x, y) dy, & m_{\overline{AC}}x_b + c_{\overline{AC}} \leq y_b \\ \int_{\overline{AB}}^{\overline{AC}} f(x, y) dy, & m_{\overline{AC}}x_b + c_{\overline{AC}} > y_b \end{cases} \quad (65)$$

biçiminde verilmektedir. Burada \overline{AB} , A ve B noktalarından geçen doğru denklemi; \overline{AC} , A ve C noktalarından geçen doğru denklemidir. j . üçgenin sol tanım bölgesinin hacmi (olasılık değeri) ise,

$$P((X, Y) \in \Omega_{jL}) = \int_{x_a}^{x_b} g_{jL}(x) dx \quad (1)$$

biçiminde elde edilir. Aynı şekilde Ω_{jR} üçgenin marjinal fonksiyonu,

$$g_{j_R}(x) = \begin{cases} \int_{\overline{AC}}^{\overline{BC}} f(x, y) dy, & m_{\overline{AC}}x_b + c_{\overline{AC}} \leq y_b \\ \int_{\overline{BC}}^{\overline{AC}} f(x, y) dy, & m_{\overline{AC}}x_b + c_{\overline{AC}} > y_b \end{cases} \quad (2)$$

biçiminde verilmektedir. Burada \overline{BC} , B ve C noktalarından geçen doğru denklemi; \overline{AC} , A ve C noktalarından geçen doğru denklemdir. j . üçgenin sağ tanım bölgesinin hacmi (olasılık değeri) ise,

$$P((X, Y) \in \Omega_{j_R}) = \int_{x_b}^{x_c} g_{j_R}(x) dx \quad (3)$$

biçiminde elde edilir. j . üçgenin toplam olasılığı,

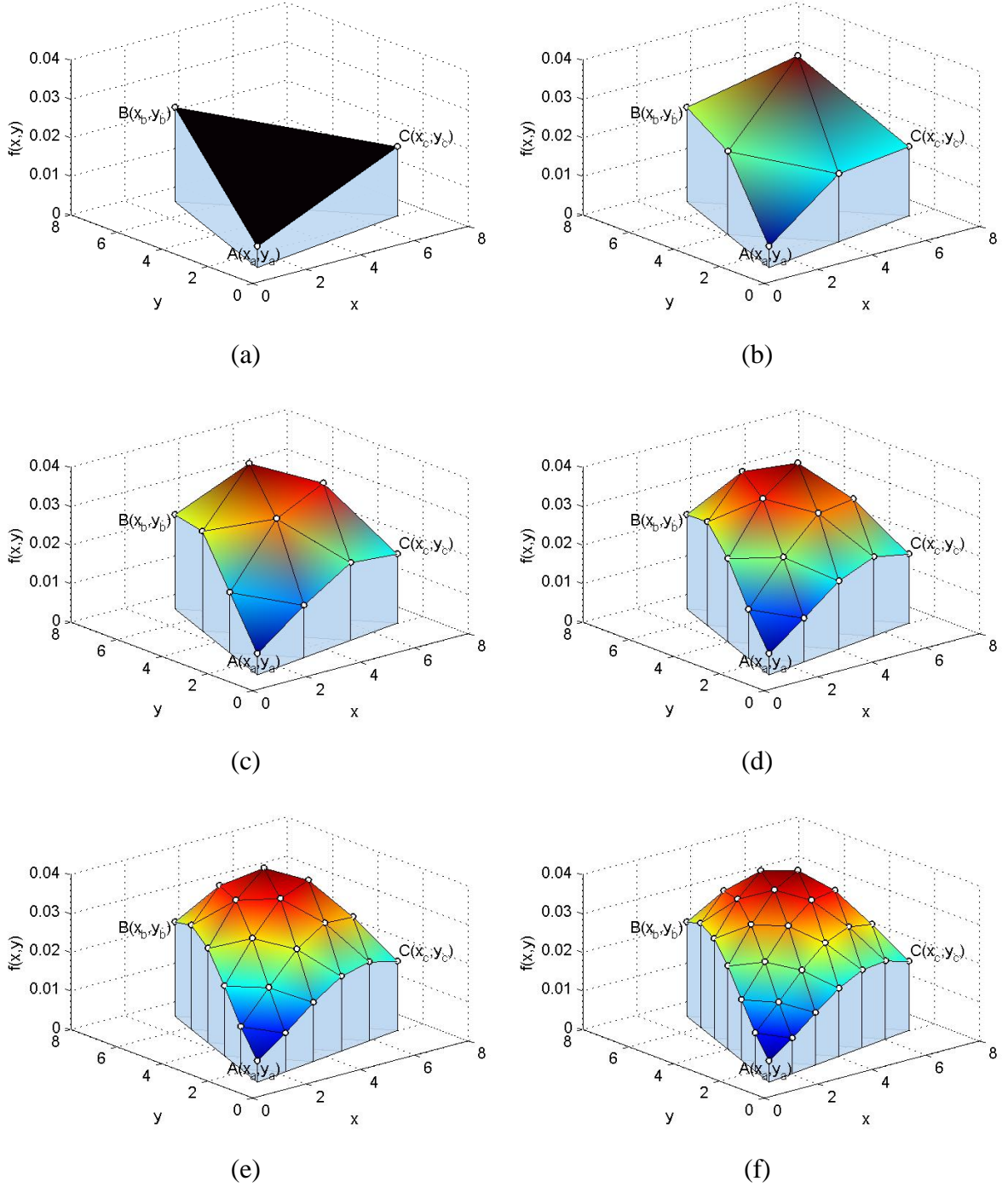
$$P((X, Y) \in \Omega_j) = P((X, Y) \in \Omega_{j_L}) + P((X, Y) \in \Omega_{j_R}) \quad (4)$$

biçiminde bulunur.

2.1.2. Sayısal Hesaplama

Üçgensel bölgede olasılık değerinin hesaplanmasında analitik yöntemler tercih edilebilir. Ancak verilen olasılık yoğunluk fonksiyonu ile bu fonksiyonun marjinal fonksiyonun integralinin alınması analitik olarak her zaman mümkün olmayabilir. Bu durumda olasılık değerinin bulunmasında sayısal yöntemlere başvurulur.

İki değişkenli bir fonksiyonun integralinin hesaplanması için bölüm 1.2.3'de verilen tek değişkenli bir fonksiyonun integralinin hesaplanmasında kullanılan yamuk kuralı tercih edilmiştir. Yamuk kuralında genelde bir eğri altındaki alan hesaplanırken burada yüzey altında kalan hacim hesaplanması gerekir. Dolayısıyla verilen üçgenin öncelikle alt üçgenlere bölünmesi gerekir. Bunun için bölüm 1.7.2'de verilen yöntem kullanılarak verilen üçgen alt üçgenlere bölünür (Şekil 32). Eğer hesaplamada daha duyarlı bir sonuç istenirse üçgen sayısı istendiği kadar artırılabilir.



Şekil 32. Üçgensel bölgedeki olasılık değerinin bulunması için alt üçgenlere bölünmesi; (a) Üçgen bölgenin 1 üçgene bölünmesi; (b) Üçgen bölgenin 4 üçgene bölünmesi; (c) Üçgen bölgenin 9 üçgene bölünmesi; (d) Üçgen bölgenin 16 üçgene bölünmesi; (e) Üçgen bölgenin 25 üçgene bölünmesi; (f) Üçgen bölgenin 36 üçgene bölünmesi;

Her alt üçgenin köşe koordinatlarının olasılık yoğunluk fonksiyonundaki değeri bulunarak alt üçgenin yükseklikleri bulunur. Bulunan bu yükseklikler kullanılarak bölüm

1.7.3'deki yöntemle hacim değeri yani olasılık değeri bulunmuş olur. Tüm alt üçgenlerin olasılık değerleri toplamı toplam olasılık değerini verecektir.

2.2. Çokgensel Alanda Uyum İyiliği Testleri

Keyfi bir alan içinde gözlenen verilerin bilindik bir geometrik şekil ile tanımlanmasında en uygun yöntem çokgensel tanımlamadır (Kesemen ve Doğru, 2012). Çokgensel bir alan içerisinde tanımlanan olasılık yoğunluk değerinden beklenen frekans sayısını bulmak için integralinin alınması gerekir. Bu ise oldukça zor olmaktadır. Önerilen yöntemde ise çokgensel alan uygun üçgenlere bölünerek (Şekil 30) hem sınıf sayısı belirlenebilir hem de bu alandaki beklenen frekans sayısı hesaplanabilmektedir. Beklenen frekans sayısı belirlenirken üçgen üzerinden integral alınmasında üçgen tepe noktasından iki parçaya bölünerek analitik integrali hesaplanabilmektedir (Uluyurt vd., 2014). Şayet üçgen içerisinde tanımlı olan olasılık yoğunluk fonksiyonun integrali alınmazsa sayısal yöntemlere başvurulur. Aynı şekilde üçgen içerisine düşen gözlenen frekans sayısı da algoritmik olarak hesaplanabilmektedir (Kesemen ve Uluyurt, 2013).

2.2.1. Çokgensel Alanda Klasik Uyum İyiliği Testi

Klasik uyum iyiliği testlerinde önceki bölümlerde olduğu gibi Ki-kare testi tercih edilmiştir. Ki-kare testi sınıflandırılmış verilerde sınıf içerisine düşen gözlenen frekans ile beklenen frekansın farkına dayalı bir yöntemdir. Gerek tek değişkenli Ki-kare uyum iyiliği testlerinde ve gerekse iki değişkenli Ki-kare uyum iyiliği testlerinde öncelikle sınıf sayısının ve sınıf bölgesinin sınırlarının belirlenmesi gerekir. Tek değişkenli Ki-kare uyum iyiliği testinde sınıf bölgesi iki sınır değer arasında kalan bölge olarak tanımlanırken, iki değişkenli bölgede sınıf bölgesi bir dikdörtgen alan olarak verilmektedir. Çokgensel uygulamalarda ise sınıf bölgesi bir üçgen yardımıyla tanımlanmaktadır. Bir rastgele değişkenin hangi sınıfa düştüğünü tek değişkenli sınıf aralığı ile dörtgensel bölgede iki değişkenli sınıf bölgelerinde büyüktür, küçüktür ilkesi belirlenmektedir. Üçgensel veya çokgensel bölgede ise iki boyutlu bir rastgele değişkenin hangi bölgeye düştüğü bölüm 1.6.1.'de teknikle belirlenmektedir. Çokgensel alanlarda Ki-kare testinin işlem adımları Algoritma 1'de verilmiştir.

Algoritma 1. Çokgensel alanlarda Ki-kare testi algoritması

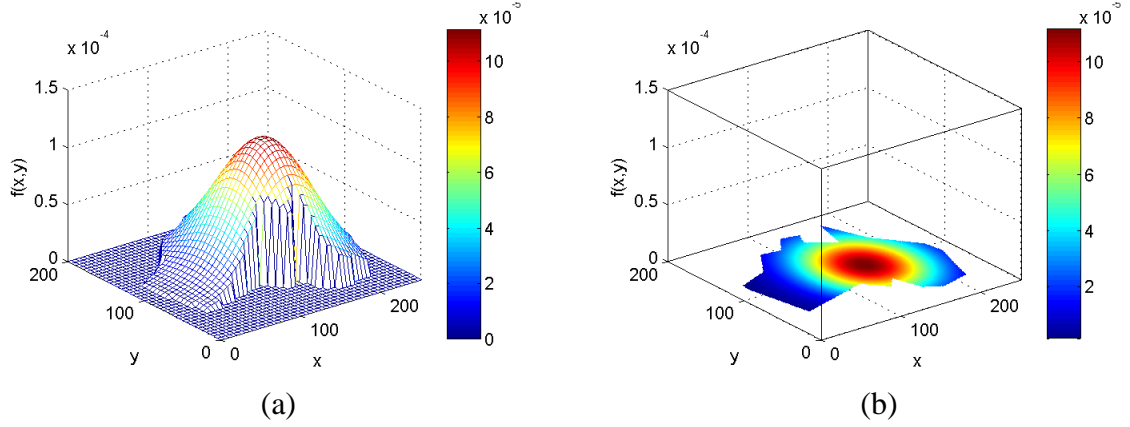
- Adım1. Çokgensel alanın sınırlarını belirle, H_0 hipotezini kur ve çokgen alan içerisinde tanımlı olasılık yoğunluk fonksiyonunu ve α güven riskini (anlamlılık seviyesini) belirle,
- Adım 2. Çokgen düğümlerini kullanarak, çokgeni delaunay algoritması yardımıyla üçgenlere böl,
- Adım 3. Eğer üçgenler test için yeterli sayıda değilse büyük üçgenleri bölüm 1.7.2'deki yöntemle alt üçgenlere böl,
- Adım 4. Elde edilen her üçgende bulunan örnek sayısını (gözlenen frekans, O_j) bölüm 1.6.1.'deki yöntemle bul,
- Adım 5. Her bir üçgenin beklenen frekansını (E_j) hesapla,
- Adım 6. χ^2 eşitliğinde gözlenen frekans ile beklenen frekansı yerine koyarak test istatistiğini hesapla,
- Adım 7. $\chi^2_{\alpha, m-1}$ tablo değeri ile hesaplanan istatistiği karşılaştır,
- Adım 8. Test istatistiği tablo değerinden küçükse H_0 hipotezini kabul et, aksi durumda reddet.
-

Bu çalışmada, Avustralya bölgesi incelenmiştir. Bu incelemeyi yapabilmek için görüntü halinde verilmiş Avustralya bölgesinin sınırları gözeysel bir yapıda elde edilmiştir (Şekil 29). Bu noktalar nokta indirgeme yöntemi kullanılarak (Douglas ve Peucker, 1973) 20 noktaya indirgenmiştir. 20 noktalı (düğümlü) bir çokgen üçgenlere bölünerek 18 üçgen elde edilmiştir (Şekil 30(b)).

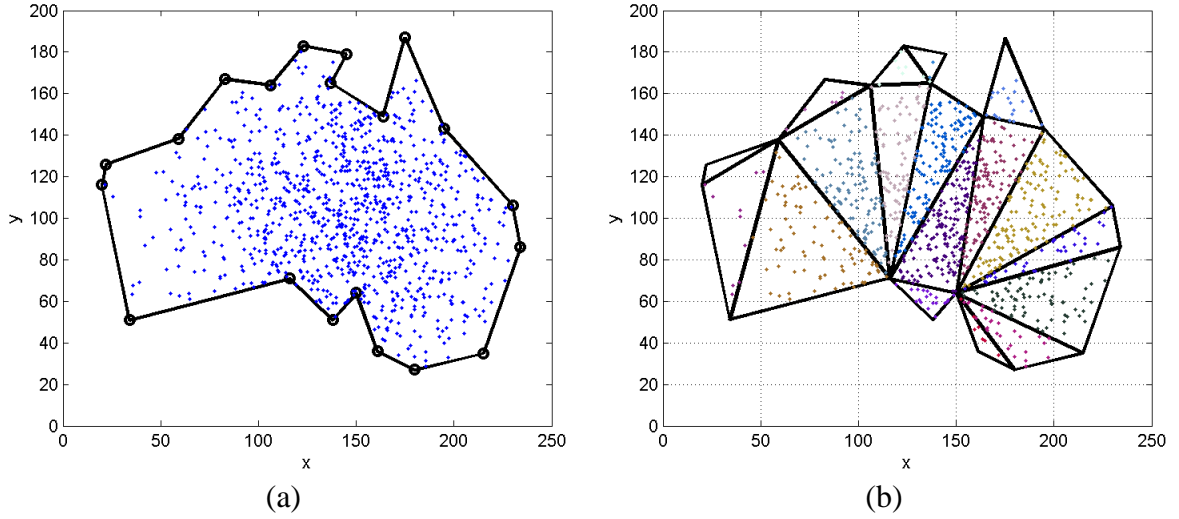
Benzetim için Avustralya bölgesinde olasılık yoğunluk fonksiyonu (Şekil 33),

$$f(x, y) = \frac{1}{9000} e^{-\frac{(x-150)^2 + (y-100)^2}{4000}} \quad (66)$$

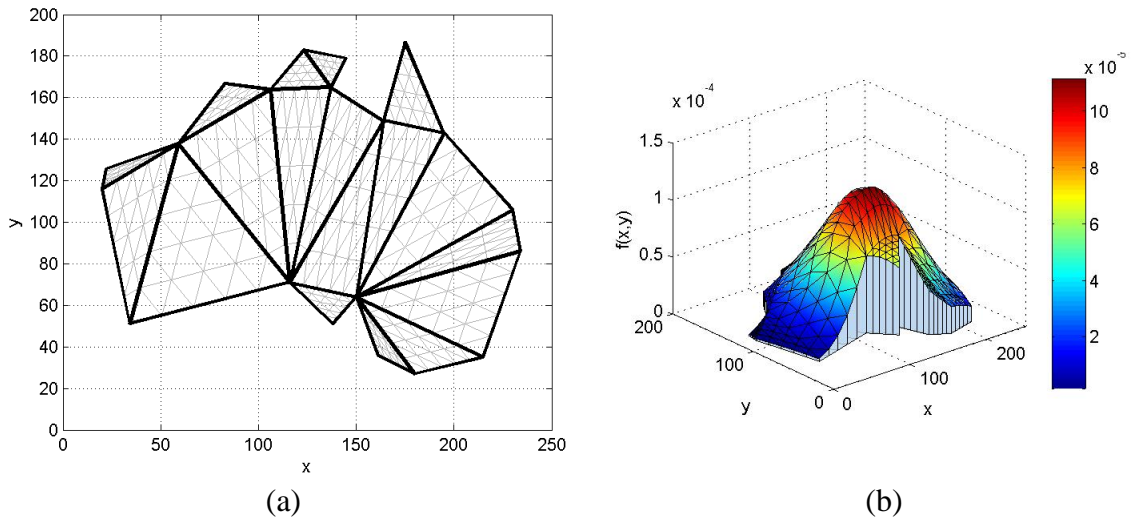
olan bir dağılımdan red-kabul yöntemiyle (Tiryaki, 2014) 1000 tane rastgele sayı üretilmiştir (Şekil 34). Bu rastgele sayılar Şekil 34(b)'de her üçgen için sınıflandırılarak her üçgenin gözlenen değerleri bulunmuştur. Olasılık değerinin hassas biçiminde hesaplanabilmesi için üçgenlerin alt üçgenlere bölünmesi Şekil 35'de gösterilmiştir. Bu alt üçgenler sayesinde her üçgenin olasılık değeri sayısal olarak hesaplanmıştır. Elde edilen bu olasılık değerleri örnek sayısı ile çarpılırsa beklenen frekanslar elde edilir (Şekil 36(b)).



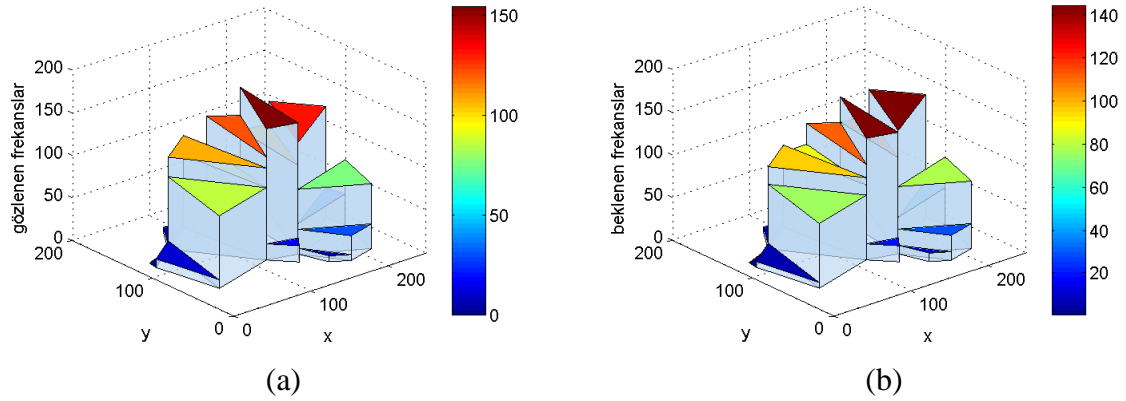
Şekil 33. Benzetim için kullanılan olasılık yoğunluk fonksiyonu; (a) Fonksiyonun üç boyutlu görünümü; (b) Fonksiyonun iki boyutlu renklendirilmiş görünümü



Şekil 34. Avustralya bölgesinde üretilen rastgele sayılar; (a) Rastgele sayıların iki boyutlu düzlemde gösterimi; (b) Rastgele sayıların üçgenlere göre sınıflandırılması



Şekil 35. Çokgeni oluşturan üçgenlerin alt üçgenlere bölünmesi; (a) Alt üçgenlerin iki boyutlu düzlemde gösterimi; (b) Alt üçgenlerin üç boyutlu gösterimi



Şekil 36. Beklenen ve gözlenen frekansların grafiksel karşılaştırılması; (a) Gözlenen frekansların üçgen sütun gösterimi; (b) Beklenen frekansların üçgen sütun gösterimi.

2.2.2. Çokgensel Alanda EDF Uyum İyiliği Testleri

EDF uyum iyiliği testleri için Kolmogorov-Smirnov testi kullanılmıştır. Test istatistiği açısından benzerlik gösterdiği için diğer EDF yöntemleri de tercih edilebilirdi. Ancak EDF yöntemleri içinde en yaygın kullanılanı Kolmogorov-Smirnov yöntemi olduğu için bu yöntem tercih edilmiştir.

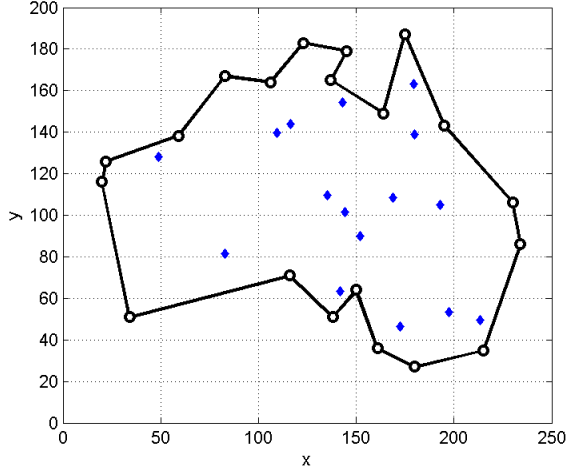
EDF tabanlı testlerinin genel yaklaşımı teorik dağılım ile deneysel dağılım arasında farkın en büyük değeri test istatistiği olarak kabul edilmiştir. Teorik dağılım fonksiyonunun bazen hesaplanamadığı durumlarda sadece örneklerin bulunduğu noktadaki değerinin bulunması yeterli olacaktır. Bunun için seçilen örnek noktasının solunda ve altında kalan bölgenin olasılık değerinin hesaplanması gerekir. Ancak olasılık değeri hesaplanırken kullanılan üçgenler seçilen noktanın solunda ve altında kalan bölgeye göre olasılık değerinin hesaplanmasında doğrudan kullanılma şansı bulunmamaktadır. Çünkü seçilen noktanın yatay ve düşey çizgileri üçgenlerinin çoğunu kesmektedir (Şekil 37(b)). Dolayısıyla çokgen $P = \{p_i = (x_i, y_i), i = 1, 2, \dots, N_p\}$ ile hesap bölgesinin $Q = \{q_i = (x_i, y_i), i = 1, 2, \dots, N_q\}$ kesişim bölgesinin $S = \{s_i = (x_i, y_i), i = 1, 2, \dots, N_s\}$ belirlenmesi gerekir. Bölüm 1.6.2’de yöntem kullanılarak kesişimi bölgesi belirlenir (Şekil 38(a)). Kesişim bölgesinin teorik dağılım fonksiyonunun belirlenmesi için kesişim bölgesi öncelikle üçgenlere ayrılır. Bu üçgenlerden kesişim bölgesi dışına düşenler atılarak elde edilen üçgenlere göre dağılım fonksiyonu hesaplanır.

Çokgensel alanlarda Kolmogorov-Smirnov testinin işlem adımları Algoritma 2’de verilmiştir.

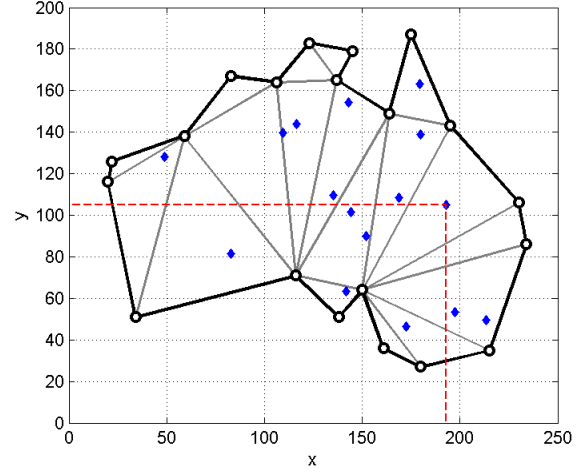
Algoritma 2. Çokgensel alanlarda EDF testleri algoritması

- Adım 1. Çokgensel alanın sınırlarını belirle, H_0 hipotezini kur ve çokgen alan içerisinde tanımlı olasılık yoğunluk fonksiyonunu ve α güven riskini (anlamlılık seviyesini) belirle,
- Adım 2. $i = 1$ ata,
- Adım 3. p_i noktasının dağılım fonksiyon değeri $F(x_i, y_i)$ değeri yani beklenen frekansı hesaplamak için $[x_{min}, x_i] \times [y_{min}, y_i]$ bölgesi ile Ω çokgen bölgesinin kesişim çokgenini $\Omega(x_k, y_k; x_i, y_i)$ bul,
- Adım 4. $\Omega(x_k, y_k; x_i, y_i)$ bölgesini delaunay algoritması ile üçgenleştir. Tüm üçgenlerin olasılık değerlerini toplayarak $F(x_i, y_i)$ hesapla,
- Adım 5. $[x_{min}, x_i] \times [y_{min}, y_i]$ bölgesindeki tüm örneklerin sayısını bul,
- Adım 6. Beklenen frekans ile gözlenen frekansın farkının mutlak değeri hesaplayarak sakla,
- Adım 7. $i < N$ ise i değerini 1 artırarak Adım 3'e git işlemlere devam et,
- Adım 8. Tüm mutlak farkların değerinden test istatistiğini hesapla,
- Adım 9. Test istatistiği ile tablo değerini karşılaştır,
- Adım 10. Test istatistiği tablo değerinden küçükse H_0 hipotezini kabul et, aksi durumda reddet.
-

Bu bölümde, bölüm 2.1.1'de verilen örnek bölge incelenmiştir. Benzetim için Avustralya bölgesinde olasılık yoğunluk fonksiyonu eşitlik (66)'da verilmiştir. Bu bölgede red-kabul yöntemiyle (Tiryaki, 2014) 16 tane rastgele sayı üretilmiştir (Şekil 37(a)). Bu rastgele sayılara göre deneysel dağılım fonksiyonu eşitlik (53) yardımıyla her nokta için ayrı ayrı hesaplanır. Her rastgele sayı için elde edilen teorik dağılım değerleri ile deneysel dağılım değerlerinin mutlak farkının en büyük değerine göre test istatistiği bulunur.

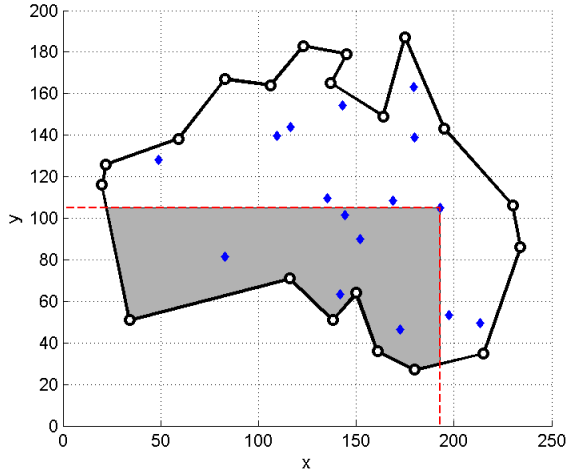


(a)

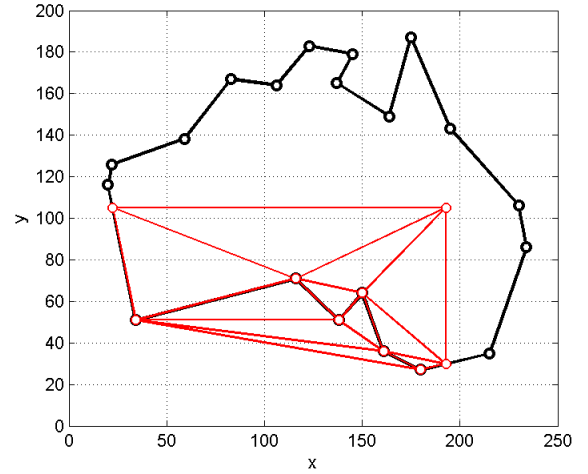


(b)

Şekil 37. Rastgele noktalar; (a) Rastgele noktaların gösterimi; (b) Rastgele noktanın sol ve alt sınırları

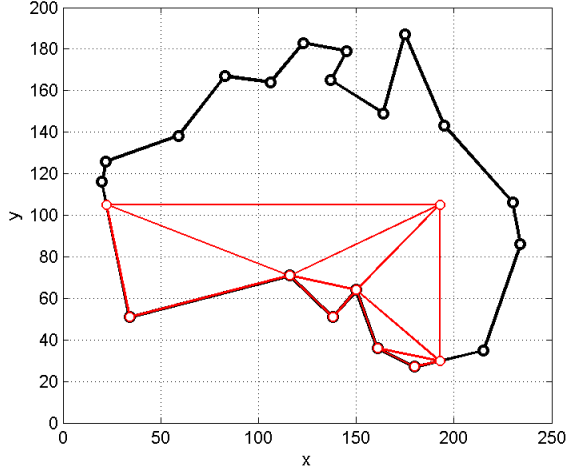


(a)

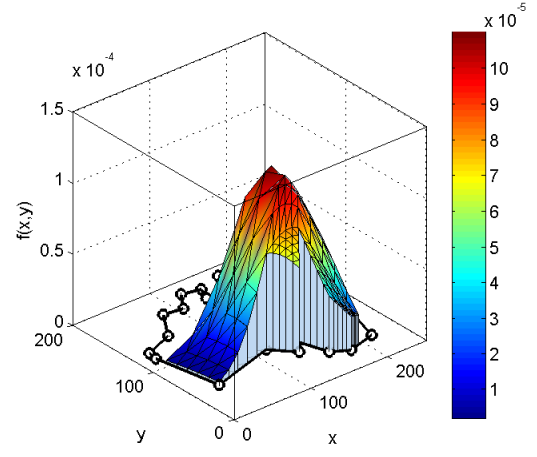


(b)

Şekil 38. Kesişim bölgesi; (a) Kesişim bölgesinin hesaplanması; (b) Kesişim bölgesinin üçgenleştirilmesi



(a)



(b)

Şekil 39. Olasılık değerinin hesaplanması; (a) Kesişim bölgesi; (b) Olasılık sütunları

3. BULGULAR VE SONUÇLAR

Şimdiye kadar iki değişkenli verilerin uyum iyiliği testlerinde dörtgenel alanlar kullanılmaktaydı. Bu çalışmada iki boyutlu keyfi bir bölge ele alınarak çokgen yaklaşım ile uyum iyiliği testleri uygulanması için yöntemler geliştirilmiştir.

Uygulanan testleri klasik uyum testleri ve EDF uyum testleri biçiminde iki ayrı kategoride incelenmiştir. Bunun nedeni her iki kategoride de birçok test bulunmasına rağmen her kategoride bir test seçilmiştir. Diğer testlerin uygulanmasında da benzer teknikler kullanılabileceğinin gösterilmesi amaçlanmıştır.

Çokgenel alanlarda uyum iyiliği testleri dörtgenel alanlarda uygulanan tekniklere dayansa da hesaplamadaki zorluklar farklı tekniklerin geliştirilmesini sağlamıştır.

Çokgenlerin üçgenlere bölünerek uyum iyiliği testlerinin yapılmasını uygulamada oldukça kolaylıklar sağlamıştır. Ki-kare testinde çokgenlerin üçgenlere bölünmesiyle beklenen ve gözlenen frekanslar hesaplanabilmektedir. Burada her bir üçgen bir sınıf olarak kabul edilir. Ki-kare istatistiği normalize edilerek hesaplandığı için farklı sınıf büyüklükleri sorun teşkil etmez. Ancak çokgen bölgesi az nokta yardımıyla tanımlanmışsa bu durumda Ki-kare için gerekli olan sınıf sayısı elde edilemeyebilir. Bu durumda çokgeni oluşturan üçgenlerden büyük olanları alt üçgenlere bölünerek işlem yapılır. Bazı durumlarda ise bazı üçgenlerdeki beklenen frekans sayısı istenen miktarda olmayabilir. Bu durumda ise her üçgen ayrı ayrı beklenen ve gözlenen frekansları hesaplanıp daha sonra yetersiz kalan üçgenler diğer üçgenlerle birleştirilir. Tercih edilen birleştirme ise komşu üçgenler arasında yapılır.

Çokgenel bölge yaklaşımı sınırlı bir alanda gerçekleştirilirken, dağılım fonksiyonu sınırsız bir alanda bulunabilir. Bu durumda belirli bir bölgeden elde edilen veriler Ki-kare testi ile değerlendirilirken seçilen bölgedeki toplam olasılık değerinin "1" olması gerekmez. Çünkü incelenen bölge üçgenlere bölündüğünde her bölge kendi içinde bağımsız davranabilmektedir.

Bölüm 2.2.1'de Avustralya bölgesi için üretilen 1000 rastgele sayı 18 sınıfa göre beklenen ve gözlenen frekanslar bulunmuştur. Yapılan bu işlem 1000 defa tekrar edildiğinde elde edilen Ki-kare istatistiğine göre serbestlik derecesi 17 olan tablo değeri karşılaştırılmış kurulan hipotezin kabul edilme sayıları Tablo 10'da verilmiştir.

Tablo 10. Ki-kare benzetim sonuçları

$1 - \alpha$	0.5	0.6	0.7	0.8	0.9	0.95	0.99
$N(H_0)$	521	624	717	810	903	955	992

Tablo 10'ya göre Ki-kare benzetiminin başarılı olduğu söylenebilir.

Tek değişkenli Kolmogorov-Smirnov uyum iyiliği testlerinde verilen veriler öncelikle sıraya dizilerek işleme katılırdı. Ancak iki boyutlu verilerin sıraya dizilmesindeki karmaşıklık sıraya dizme işlemini geçersiz kılmaktadır. Dolayısıyla işleme sıraya dizmeye gerek duyulmaksızın gerçekleştirilmiştir.

Bölüm 2.2.2'de Avustralya bölgesi için üretilen 16 rastgele sayıya göre kesişim bölgesi belirlenerek teorik dağılım değerleri hesaplanmıştır. Aynı zamanda her rastgele sayının altında ve solunda kalan rastgele sayılara göre deneysel dağılım değerleri bulunmuştur. Yapılan bu işlem 1000 defa tekrar edildiğinde elde edilen Kolmogorov-Smirnov istatistiğine göre sıfır hipotezinin kabul edilme sayıları Tablo 11'de verilmiştir.

Tablo 11. Kolmogorov-Smirnov benzetim sonuçları

$1 - \alpha$	0.8	0.9	0.95	0.98	0.99
$N(H_0)$	837	923	962	984	989

Tablo 11'ye göre Kolmogorov-Smirnov benzetiminin başarılı olduğu söylenebilir.

4. ÖNERİLER

Bu çalışmada, keyfi çokgensel bir bölgeden elde edilen verilerin bir dağılıma uyup uymadığını test etmek için sadece Ki-kare ve Kolmogorov-Smirnov uyum iyiliği testleri üzerinde durulmuştur. Geliştirilen bu yöntemler çokgensel alanlardaki veriler için istatistik biliminde bulunan diğer testlerde uyarlanabilir. Özellikle EDF tabanlı testler genelde Kolmogorov-Smirnov test istatistiğine benzer biçimde test istatistikleri elde ettiklerinden çokgensel alan için bu yöntemlerin uyarlanması sorun olmayacaktır.

Geliştirilen yöntemler basit çokgensel alanlar için geliştirilmiştir. Ancak basit çokgen dışındaki diğer çokgenler içinde uyum iyiliği testleri araştırılabilir.

Çokgen alanda integral hesabının kolay yapılabilmesi için üçgenleştirme yoluna gidilmiştir. Ancak benzer bir yöntem olarak çokgenin her (x_k, y_k) noktasından y eksenine paralel geçirilen doğrular yardımıyla çokgen paralel bölgelere bölünüp işlem gerçekleştirilebilir. Ancak bu durumda karmaşık çokgen yapılarında hesaplama güçlüklerine neden olabileceği göz önünde tutulmalıdır.

Bu çalışmada Kolmogorov-Smirnov uyum iyiliği testinde rastgele seçilen bir noktanın dağılım değeri hesaplanırken noktanın solunda ve altında kalan bölge değerlendirilmektedir. Bu noktanın altında ve sağında; üstünde ve solunda; üstünde ve sağında kalan bölgeler için de yöntem değerlendirilebilir.

Çokgensel bölge yaklaşımı ile dar bir bölgede istatistiksel işlemler yapılabilmektedir. Oysa sonsuz bir bölgede tanımlı olan keyfi parçalı olan bir dağılım fonksiyonu için geliştirilen yöntemler genelleştirilebilir.

Komşu bölgelerdeki değişimlerin bölgesel etmenlerden mi yoksa uygulanan yöntemlerden mi olduğu araştırılabilmesi için çokgensel bölge istatistik yaklaşımı kullanılabilir.

Bu çalışmada çokgensel alanda birçok istatistik yöntemin kullanılabilirliği vurgulanmıştır. Çokgensel alanlarda diğer istatistiksel yöntemlerin uygulanabilirliği araştırılabilir.

Pratikte birçok uygulama alanının olabileceği ifade edilen çokgensel bölgedeki istatistiksel yaklaşımların uygulama alanlarındaki başarısı test edilebilir. Hatta geliştirilen yöntemler pratik uygulamaları yaygınlaştırılarak iki değişkenli istatistik analizlerin önemi artırılabilir.

5. KAYNAKLAR

- D'Agostino, R. B., 1986. Goodness-of-fit-techniques, 68., CRC press.
- Devadoss, S. L. ve O'Rourke, J., 2011. Discrete and computational geometry, Princeton University Press.
- Douglas, D. H. ve Peucker, T. K., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, Cartographica: The International Journal for Geographic Information and Geovisualization, 10, 2, 112-122.
- Foley, J., van Dam, A., Feiner, S. ve Hughes, J., 1990. Computer Garphics: Rpinciples and Practice, 2nd ed., Addison-Wesley.
- Gibbons, J. D. ve Chakraborti, S., 2011. Nonparametric statistical inference, Springer, Berlin Heidelberg.
- Haines, E., 1994. Point in polygon strategies, in: P. Heckbert (Ed.), Graphic Gems IV, Akademic Press, Boston, MA.
- Harrington, S., 1983. Computer Graphics: A Programming Approach, McGraw-Hill.
- Hormann, K. ve Agathos, A., 2001. The point in polygon problem for arbitrary polygons, Computational Geometry, 20, 3, 131-144.
- Howard, W. E. ve Musto, J., 2008. Engineering Computation: An Introduction Using MATLAB and Excel, McGraw Hill.
- Justel, A., Peña, D. ve Zamar, R., 1997. A multivariate Kolmogorov-Smirnov test of goodness of fit, Statistics & Probability Letters, 35, 3, 251-259.
- Kesemen, O. ve Doğru, F. Z., 2012. Çokgen Alanlarda İki Değişkenli Birikimli Dağılım Fonksiyonun Bulunması, İstatistik Araştırma Dergisi, 09, 01, 65-71.
- Kesemen, O. ve Uluhurt, T., 2013. Bivariate Chi-Square Goodness Of Fit Test In Polygonal Areas, 8th International Statistics Congress, October, Antalya, Bildiriler Kitabı: 80-81.
- Krishnaiah, P. R., 1990. Handbook of statistics, Motilal Banarsidass Publishe.
- Mark, B. L. ve Turin, W., 2011. Probability, Random Processes, and Statistical Analysis, Cambridge University Press Textbooks.

- Massey, J. ve Frank, J., 1951. The Kolmogorov-Smirnov test for goodness of fit, Journal of the American statistical Association, 46, 253, 68-78.
- O'Rourke, J., 1998. Computational Geometry in C, 2nd ed., Cambridge University Press.
- O'Connor, P. ve Andre, K., 2011. Practical reliability engineering, John Wiley & Sons.
- O'Rourke, J., Chien, C. B., Olson, T. ve Naddor, D., 1982. A new linear algorithm for intersecting convex polygons, Computer Graphics and Image Processing, 19, 4, 384-391.
- Preparata, F. P. ve Ian, S. M., 1985. Computational Geometry An Introduction, Springer-Verlag, New York.
- Schneider, P. E. D. H., 2002. Geometric tools for computer graphics, Morgan Kaufmann.
- Sedgewick, R., 1988. Algorithms, 2nd ed., Addison-Wesley.
- Sheskin, D. J., 2003. Handbook of Parametric and Nonparametric Statistical Procedures, CRC Press.
- Tiryaki, B. K., 2014. Çokgensel Alan İçerisinde İki Değişkenli Keyfi Bir Dağılımdan Rastgele Sayı Üretimi, Yüksek Lisans, KTÜ, Fen Bilimleri Enstitüsü, Trabzon.
- Toussaint, G. T., 1985. A simple linear algorithm for intersecting convex polygons, The visual computer, 1, 2, 118-123.
- Uluyurt, T., Kesemen, O. ve Tiryaki, B. K., 2014. Bivariate Kolmogorov-Smirnov Goodness Of Fit Test On Polygonal Areas, 9th International Statistic Days Symposium, May, Side-Antalya-Turkey, Bildiriler Kitabı: 68-68.
- Walck, C., 2007. Handbook on statistical distributions for experimentalists, University of Stockholm.
- Wang, J. C., 2012. Sample distribution function based goodness-of-fit test for complex surveys, Computational Statistics & Data Analysis, 56, 3, 664-679.
- Weiler, K., 1994. An incremental angle point in polygon test, in: P. Heckbert (Ed.), Graphic Gems IV, , Academic Press, Boston, MA.
- Worboys, M. F. ve Duckham, M., 2004. GIS: a computing perspective, CRC press.
- Yanalak, M., 1997. Sayısal Arazi Modellerinden Hacim Hesaplarında En Uygun Enterpolasyon Yönteminin Araştırılması, Doktora Tezi, İTÜ , Fen Bilimleri Enstitüsü, İstanbul.
- URL-1, http://tr.wikipedia.org/wiki/Kolmogorov-Smirnov_s%C4%B1namas%C4%B1 Kolmogorov-Smirnov sınaması. 12 Ocak 2015.

ÖZGEÇMİŞ

Tuncay ULUYURT, 28 Kasım 1984 tarihinde Kayseri’de doğdu. İlköğrenimini Güventürk İlköğretim Okulu’nda, ortaöğrenimini ise Sami YANGIN Anadolu Lisesi’nde tamamladı. 2004 yılında Kayseri Erciyes Üniversitesi, Fen-Edebiyat Fakültesi, Matematik bölümüne yerleşti ve 2008 yılında bu bölümden mezun oldu. 2010 yılında Artvin Çoruh Üniversitesi Arhavi Meslek Yüksek Okulunda öğretim görevlisi olarak göreve başlamıştır. Aynı yıl Karadeniz Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik ve Bilgisayar Bilimleri Anabilim dalında tezli yüksek lisans programına başladı. Ulusal ve uluslararası birçok sempozyum ve kongrede yayınlanmış bildirileri bulunmaktadır.

Tuncay Uluyurt halen görevini devam ettirmekte, evli ve bir çocuk babasıdır.