

ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

N ÇAPLI RASGELE BİR ÖRNEKLEM İÇİN BELİRLİ BİR ARALIKTA
DEĞER ALAN MAXİMUM GÖZLEM SAYILARININ DAĞILIMI

M. Bahar BAŞKIR

İSTATİSTİK ANABİLİM DALI

ANKARA

2004

Her hakkı saklıdır

Yard. Do. Dr. İhsan KARABULUT danışmanlığında , M. Bahar BAŞKIR tarafından hazırlanan bu çalışma 23/12/2004 tarihinde aşağıdaki jüri tarafından İstatistik Anabilim Dalı' nda Yüksek Lisans Tezi olarak kabul edilmiştir.

Başkan: Prof. Dr. Hazma GAMGAM



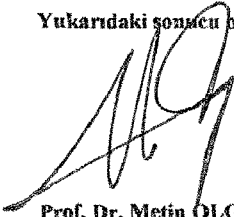
Yard. Do. Dr. Levent ÖZBEK



Yard. Do. Dr. İhsan KARABULUT



Yukarıdaki sonucu onaylıyorum.



Prof. Dr. Metin ÖLGÜN

Enstitü Müdürü

ÖZET

Yüksek Lisans Tezi

N ÇAPLI RASGELE BİR ÖRNEKLEM İÇİN BELİRLİ BİR ARALIKTA DEĞER ALAN MAXİMUM GÖZLEM SAYILARININ DAĞILIMI

M. Bahar BAŞKIR

Ankara Üniversitesi
Fen Fakültesi
İstatistik Bölümü

Danışman: Yard. Doç. Dr. İhsan KARABULUT

Günümüzde bilimsel çalışmalarda özellikle biyolojide kullanılan rasgele örneklem çapları oldukça büyük olabilmektedir. Bu gözlemler içinde kimi özelliklere sahip olanların belirli bir dağılım varsayımı altında hangi olasılıklarla hangi aralıklarda ve hangi sayıda gözlenebileceği soruları önem kazanmaktadır. Örneğin, genetikte DNA dizilişlerinde belirli bir dizilimin araştırılması gibi... Söz konusu rasgele örneklem alınıdığı dağılım $(0,1)$ aralığında düzgün dağılım olduğunda yukarıdaki sorulara yanıtlar kısmen alınmıştır. Eğer dağılım düzgün değilse bu sorulara nasıl yanıt verilebileceği üzerinde durulmuştur ve bu alandaki çalışmalardan bazıları gözden geçirilmiştir.

2004, 68 sayfa

ANAHTAR KELİMELER: k ' lı topluluk (k -aggregate), sıra istatistikleri, ardışık sıra istatistikleri arasındaki farklar (Spacings), bölünmüş farklar, yer değiştirilebilirlik (exchangeability).

ABSTRACT

Master Thesis

DISTRIBUTION OF THE MAXIMUM SIZE OF THE CLUSTER FOR THE RANDOM SAMPLE OF SIZE N

M. Bahar BAŞKIR

Ankara University
Graduate School of Natural and Applied Sciences
Department of Statistics

Supervisor: Yard. Doç. Dr. İhsan KARABULUT

The size of random sample used in scientific studies, particularly in biology, can be big. Questions like "in which interval with what probability" and "how many observations" come into prominence under particular distribution assumption for the observations, having some properties. For instance, research of a particular permutation in DNA arrangements in genetics. The questions above have been solved when the distribution from which the random sample, pointed at issue, had taken, was uniformly (0,1) distributed. If the distribution is not uniformly distributed, "how can such questions be solved" is researched and studied.

2004, 68 pages

Key Words: *k*-aggregate, order statistics, spacings, divided differences, exchangeability.

TEŞEKKÜR

Bana araştırma olanağı sağlayan ve çalışmamın her safhasında bana yardımcı olan, önerileri ile beni yönlendiren danışman hocam, Sayın Yard. Doç. Dr. İhsan KARABULUT (Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü)' a teşekkürlerimi sunarım.

Ayrıca, hiçbir zaman benden desteğini esirgemeyen sevgili aileme sonsuz teşekkürler.

M. Bahar BAŞKIR
Ankara, Aralık 2004

İÇİNDEKİLER

ÖZET.....	i
ABSTRACT.....	ii
TEŞEKKÜR.....	iii
SİMGELER DİZİNİ.....	v
ŞEKİL DİZİNİ.....	vi
ÇİZELGE DİZİNİ.....	vii
1. GİRİŞ.....	1
2. TEMEL TANIM VE TEOREMLER	3
2.1. Sıra İstatistikleri ve Ardışık İki Sıra İstatistiği Arasındaki Fark (Spacing).....	3
2.1.1. Sıra İstatistiklerinin Dağılımları	3
2.1.2. İki veya Daha Fazla Sıra İstatistiklerinin Ortak Dağılımları.....	8
2.1.3. İki Sıra İstatistiği Arasındaki Farkın Dağılımı.....	10
2.1.4. Ardışık Sıra İstatistikleri Arasındaki Farkların (Spacings) Dağılımı.....	12
2.1.4.1. Ardışık Sıra İstatistikleri Arasındaki Farkların (Spacings) Genelleştirilmiş Dağılımı.....	13
2.1.4.2. Düzgün Dağılım Durumunda Ardışık Sıra İstatistikleri Arasındaki Farkların (Spacings) Dağılımı.....	15
2.1.4.3. Üstel Dağılım Durumunda Ardışık Sıra İstatistikleri Arasındaki Farkların (Spacings) Dağılımı.....	18
2.2. Yer Değiştirilebilirlik (Exchangeability).....	19
2.3. Laplace Dönüşümü.....	20
2.4. Bölünmüş Farklar.....	21
3. PROBLEM VE ÇÖZÜMÜNE YAKLAŞIMLAR.....	26
4. UYGULAMA ÖRNEKLERİ.....	57
4.1. Bir Mod Tahmin Edicisinin Dağılımı.....	57
4.2. Genetikte Uygulama.....	61
4.3. Diğer Uygulamalar.....	62
5. TARTIŞMA VE SONUÇ.....	65
KAYNAKLAR.....	66
ÖZGEÇMİŞ.....	68

SİMGELER DİZİNİ

λ	lambda
$X_{(i)}$	i nci sıra istatistiği ($i=1,2,\dots,n$)
CMV	Stomegalovirüs
DNA	Deoksiribonükleik asit
G	Guanin
C	Sitozin
A	Adenin
T	Timin

ŞEKİL DİZİNİ

Şekil 3.1. Sürekli durumda düzgün bir doğru boyunca k' lı topluluğun konumu..... 27



ÇİZELGE DİZİNİ

Çizelge 4.1. Aralık uzunluğuna bağlı olarak maksimum palindrom sayıları.....	62
--	----



1. GİRİŞ

$(0,1)$ aralığında düzgün dağılımlı bir yığından alınan $X_1, X_2, \dots, X_n$ n çaplı rasgele örnekleme $\lambda \in (0,1)$ olacak biçimde, belirli bir λ uzunluğundaki aralıklar boyunca değer alan gözlem sayıları ve bu sayıların maksimumlarının dağılımı üzerine değişik çalışmalar yapılmıştır. $(0,1)$ aralığında düzgün dağılımdan alınan büyük çaplı rasgele örnekleme ait gözlemlerden belirlenen uzunluktaki aralıklarda yer alanların sayısı ve bu değerleri alması olasılıkları üzerine C. Mack, M. A. (1948), Naus, J. I. (1965 ve 1966) tarafından ve küçük örneklem çapları için Seth, G.R. (1950)' in ve Lieblein (1952)' in yaptıkları çalışmalar dikkat çekici olanlardır. Huffer (1988)' in ve Huffer ve Lin (2001)' in ardışık sıra istatistikleri arasındaki farklar olarak tanımlanan rasgele değişkenlerle (spacings) oluşturulan katsayıları 0,1 veya rasyoneel olan lineer kombinasyonların ortak dağılımı ile ilgili çalışması bu olasılıkların elde edilmesinde kolaylaştırıcı bir çözüm sunmaktadır. Konuya ilişkin uygulama alanında birkaç olasılık problemi aşağıdaki gibi ifade edilebilir:

1. N tane telefonun bağlı olduğu bir santral düşünölsün. N tane telefonun herhangi birinin 1 dakikalık zaman dilimi içinde santrali arama zamanları rasgele ve herhangi bir telefonun santrali meşgul etme süresi λ birim zamandır (saniye). N telefondan en az n tanesinin λ zaman aralığında santrali araması olasılığı nedir (Naus 1965)?

2. Bir sayaç ile sayılan sinyaller λ oranlı Poisson sürecini meydana getirmektedir. t süresinden daha az uzunluklu bir zaman aralığında n tane sinyal, sayaç tarafından kaydedilmektedir. Sayacın ilk kaydına kadar geçen bekleme süresinin dağılımı nedir (Naus 1965)?

3. N oldukça büyük bir doğal sayı ve $N > n$ olacak biçimde N tane kutucuk ve n tane top olsun. n tane top rasgele olarak N tane kutucuğun içine dağıtılsın. N kutucuğun belirlenen p tanesinde toplam k top bulunsun. Bu şekilde her bir p kutucuk gruplarında k top ile elde edilen k' h topluluğun bulunması olasılığı nedir (Mack 1948)?

4. Belirlenen bir sıralanışa uygun DNA dizilim parçalarının DNA baz dizilimindeki yerlerinin dağılımı düzgün dağılımlı mıdır (Nolan ve Speed 2000)?

5. Bir yığından alınan $n = 3$ çaplı örneklerdeki gözlemlerden birinin bir sıradışı (outlier) değer olma olasılığı nedir (Seth 1950, Lieblein 1952)?

Yukarıda verilen problemler birbirine benzer sorular olup belirli bir uzunluktaki aralıkta yapılacak gözlemlerin sayısının en çoğuna (maksimumuna) ilişkindir. Yukarıdaki problemler Naus (1965)' un ifade ettiği gibi, doğru üzerinde belirli uzunluktaki aralıklarda yer alan gözlem sayılarının maksimumu ile ilgilidir.

Bu derleme çalışmasında bu tür olasılıkların nasıl cevaplandırılmaya çalışıldığı sergilenip, uygulama olarak; bir mod tahmin edicisinin dağılımını belirlemede ve moleküler genetikteki DNA dizilişlerinde düzgün dağılıma uygunluk testinin nasıl yapılacağına dair iki örnek verilecektir.

Bu çalışmada amaç, bu alanda yapılan çalışmaları kısmen de olsa kullanıcılara bir özet halinde sunmaktır.

Bundan sonraki bölümlerde konuya ilişkin olarak sırası ile önemli kavramlar verilecek, "Problem ve Çözümüne Yaklaşımlar" başlığı altında sözü geçen olasılıkların hesabında ortaya sunulan yaklaşımlar ve kesin çözüm yöntemleri, son olarak da 4 ncü Bölümde "Uygulama Örnekleri" başlığı altında bir mod tahmin edicisinin dağılımı, DNA' da yer alan kimi özel dizilişlerin düzgün dağılıma uygunluğu ve yukarıda verilen 1 nci, 2 nci ve 5 nci problemler uygulama örnekleri olarak verilecektir.

2. TEMEL TANIM VE TEOREMLER

Çalışmanın bu aşamasında, konuya hazırlık amacı ile sıra istatistikleri ve ardışık sıra istatistikleri arasındaki farklarla oluşan aralıklar (spacings), yer değiştirilebilirlik (exchangeability), Laplace dönüşümü tanımı, bölünmüş farkların tanımları ile bazı teoremler verilecektir.

2.1. Sıra İstatistikleri ve Ardışık İki Sıra İstatistiği Arasındaki Fark (Spacing)

$X_1, X_2, \dots, X_n$ rasgele değişkenleri aynı dağılımdan alınmış olsun. $X_{(i)}$, i nci sıra istatistiği olmak üzere,

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

yazılır. X_i rasgele değişkenlerinin en küçüğü $X_{(1)}$, 1 nci sıra istatistiği; ikinci en küçüğü $X_{(2)}$, 2 nci sıra istatistiği; ... ; en büyüğü $X_{(n)}$, n nci sıra istatistiğidir.

$X_{(i)}$ sıra istatistikleri bağımlı rasgele değişkenlerdir.

Bu çalışmada X_i rasgele değişkenleri istatistiksel olarak bağımsız ve aynı dağılımlı olarak ele alınacaktır. X_i rasgele değişkenlerinin bağımlı olduğu durumlar için sıra istatistiklerine *bkz. David (1969) sf:82, Bölüm 5.5.*

2.1.1. Sıra İstatistiklerinin Dağılımları

$X_1, X_2, \dots, X_n$ rasgele değişkenleri, olasılık yoğunluk fonksiyonu $f(x)$, dağılım fonksiyonu $F(x)$ olan sürekli bir dağılımdan alınmış olsun. $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ sıra istatistikleri yukarıdaki gibi tanımlanmış olsun.

$X_{(r)}$, r nci sıra istatistiğinin olasılık yoğunluk fonksiyonu $f_r(x)$, dağılım fonksiyonu $F_r(x)$ olarak gösterilsin. Buna göre r nci sıra istatistiğinin dağılım fonksiyonu,

$$P(x < X_{(r)} \leq x + \Delta x) = \frac{n!}{(r-1)!(n-r)!} \{F(x)\}^{r-1} \{1 - F(x + \Delta x)\}^{n-r} \times \{F(x + \Delta x) - F(x)\} + O((\Delta x)^2) \quad (2.1)$$

dir. Burada $O((\Delta x)^2)$ ifadesindeki $(\Delta x)^2$ terimi, $(x, x + \Delta x]$ aralığında birden fazla

X_i rasgele değişkeninin bulunması olayına karşılık gelmektedir (Arnold et. al. 1992).

(2.1) ifadesine göre, $1 \leq r \leq n$ için, $X_{(r)}$ sıra istatistiğinin olasılık yoğunluk fonksiyonu,

$$f_r(x) = \lim_{\Delta x \rightarrow 0} \left\{ \frac{P(x < X_{(r)} \leq x + \Delta x)}{\Delta x} \right\} \quad (2.2)$$

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

dir.

$X_{(r)}$ sıra istatistiğinin olasılık yoğunluk fonksiyonu, n tane sıra istatistiğin ortak olasılık yoğunluk fonksiyonu ile doğrudan türetilir (Arnold et. al. 1992). n tane sıra istatistiğin ortak olasılık yoğunluk fonksiyonu,

$$f_{1,2,\dots,n}(x_1, x_2, \dots, x_n) = n! \prod_{i=1}^n f(x_i), \quad -\infty < x_1 < x_2 < \dots < x_n < \infty \quad (2.3)$$

dir.

Örneğin, U_i rasgele değişkenleri birbirinden bağımsız ve $(0,1)$ aralığındaki düzgün dağılımdan alınmış olsun $(i = 1, 2, \dots, n)$. Bu durumda $U_{(i)}$ sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonu,

$$f_{1,2,\dots,n}(u_1, u_2, \dots, u_n) = n! \quad , \quad 0 \leq u_1 < u_2 < \dots < u_n \leq 1$$

dir.

Diğer bir örnek, $X_i, i = 1, 2, \dots, n$ rasgele değişkenleri birbirinden bağımsız ve parametresi 1 olan üstel dağılımdan alındığında $X_{(i)}$ sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonu,

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

olacaktır.

Genel olarak ise (2.2) de verilen formüle göre en küçük ve en büyük sıra istatistiklerinin olasılık yoğunluk fonksiyonları sırası ile,

$$f_1(x) = n \{1 - F(x)\}^{n-1} f(x) \quad , \quad -\infty < x < \infty \quad (2.4)$$

ve

$$f_n(x) = n \{F(x)\}^{n-1} f(x) \quad , \quad -\infty < x < \infty \quad (2.5)$$

dir.

(2.4) ve (2.5) de verilen olasılık fonksiyonlarına göre $X_{(1)}$ ve $X_{(n)}$ sıra istatistiklerinin dağılım fonksiyonları sırasıyla,

$$F_1(x) = 1 - \{1 - F(x)\}^n, \quad -\infty < x < \infty \quad (2.6)$$

ve

$$F_n(x) = \{F(x)\}^n, \quad -\infty < x < \infty \quad (2.7)$$

olarak bulunur. (2.2) formülünün $(-\infty, x]$ aralığındaki integrali $X_{(r)}$ nin dağılım fonksiyonunu vermektedir.

Öte yandan, herhangi bir $r=1,2,...,n$ için r nci sıra istatistiğinin dağılım fonksiyonu,

$$F_r(x) = P(X_{(r)} \leq x) = P(X_1, X_2, \dots, X_n \text{ en az } r \text{ tanesi } \leq x) \quad (2.8)$$

$$\begin{aligned} &= \sum_{i=r}^n P(X_1, X_2, \dots, X_n \text{ tam } i \text{ tanesi } \leq x) \\ &= \sum_{i=r}^n \binom{n}{i} \{F(x)\}^i \{1 - F(x)\}^{n-i}, \quad -\infty < x < \infty \end{aligned}$$

biçiminde başarı olasılığı $F(x)$ ve denemelerin sayısı n olan Binom olasılıklarının r den başlayarak toplamı olarak ifade edilebilir.

Binom dağılımlı rasgele değişkenlerin olasılıklarının toplamı ve tam olmayan Beta fonksiyonu arasındaki ilişki kullanılarak,

$$F_r(x) = I_p(r, n - r + 1)$$

ifade edilebilir. Burada $I_p(r, n-r+1)$ tam olmayan beta fonksiyonudur ve tam olmayan Beta fonksiyonunun genel gösterimi,

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

biçimindedir.

Aşağıdaki özdeşlik kullanılarak,

$$\sum_{i=r}^n \binom{n}{i} p^i (1-p)^{n-i} = \int_0^p \frac{n!}{(r-1)!(n-r)!} t^{r-1} (1-t)^{n-r} dt, \quad 0 < p < 1 \quad (2.9)$$

(2.8) deki $X_{(r)}$ sıra istatistiğinin dağılım fonksiyonu,

$$\begin{aligned} F_r(x) &= \int_0^{F(x)} \frac{n!}{(r-1)!(n-r)!} t^{r-1} (1-t)^{n-r} dt \\ &= I_{F(x)}(r, n-r+1) \end{aligned} \quad (2.10) \quad , \quad -\infty < x < \infty$$

Pearson (1934)' ın tam olmayan Beta fonksiyonu olarak elde edilir (Arnold *et. al.* 1992).

(2.8) de verilen ifade, $X_i, i = 1, 2, \dots, n$ rasgele değişkenlerinin dağılımları sürekli veya kesikli olduğunda geçerlidir. Ancak (2.10) daki ifade dağılım sürekli olduğunda kullanılmalıdır.

2.1.2. İki veya Daha Fazla Sıra İstatistiklerinin Ortak Dağılımları

$X_{(r)}$, r nci sıra istatistiği ve $X_{(s)}$, s nci sıra istatistiğinin $1 \leq r < s \leq n$ olacak biçimde ortak olasılık yoğunluk fonksiyonu $f_{rs}(x, y)$ olarak gösterilsin.

Ortak dağılım fonksiyonu da,

$$\begin{aligned}
 P(x < X_{(r)} \leq x + \Delta x, y < X_{(s)} \leq y + \Delta y) &= \\
 &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \{F(x)\}^{r-1} \\
 &\quad \times \{F(y) - F(x + \Delta x)\}^{s-r-1} \{1 - F(y + \Delta y)\}^{n-s} \quad (2.11) \\
 &\quad \times \{F(x + \Delta x) - F(x)\} \{F(y + \Delta y) - F(y)\} \\
 &\quad + O((\Delta x)^2 \Delta y) + O(\Delta x (\Delta y)^2)
 \end{aligned}$$

dir (Arnold *et. al.* 1992).

Böylece (2.11) kullanılarak $X_{(r)}$ ve $X_{(s)}$ sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonu,

$$\begin{aligned}
f_{rs}(x, y) &= \lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} \left\{ \frac{P(x < X_{(r)} \leq x + \Delta x, y < X_{(r)} \leq y + \Delta y)}{\Delta x \Delta y} \right\} \\
&= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \{F(x)\}^{r-1} \\
&\quad \times \{F(y) - F(x)\}^{s-r-1} \{1 - F(y)\}^{n-s} \\
&\quad \times f(x)f(y), \quad -\infty < x < y < \infty
\end{aligned} \tag{2.12}$$

olarak elde edilir.

$1 \leq r_1 \leq r_2 \leq \dots \leq r_k \leq n$, $1 \leq k \leq n$ olmak üzere, $X_{(r_1)}, X_{(r_2)}, \dots, X_{(r_k)}$ sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonu, $x_1 \leq x_2 \leq \dots \leq x_k$ için,

$$\begin{aligned}
f_{r_1, r_2, \dots, r_k}(x_1, x_2, \dots, x_k) &= \frac{n!}{(r_1-1)!(r_2-r_1-1)!\dots(n-r_k)!} \\
&\quad \times \{F(x_1)\}^{r_1-1} f(x_1) \{F(x_2) - F(x_1)\}^{r_2-r_1-1} f(x_2) \\
&\quad \times \dots \times \{1 - F(x_k)\}^{n-r_k} f(x_k)
\end{aligned} \tag{2.13}$$

dir. $x_0 = -\infty, x_{k+1} = +\infty, r_0 = 0, r_{k+1} = n+1$ alındığında (2.13) formülü,

$$n! \left[\prod_{i=1}^k f(x_i) \right] \prod_{i=0}^k \left\{ \frac{[F(x_{i+1}) - F(x_i)]^{r_{i+1} - r_i - 1}}{(r_{i+1} - r_i - 1)!} \right\} \tag{2.14}$$

biçiminde yazılır (David 1969).

n tane sıra istatistiğinin ortak olasılık yoğunluk fonksiyonu,

$$n! f(x_1) f(x_2) \dots f(x_n)$$

dir.

(2.12) formülünde $s = r + 1$ alınırsa $X_{(r)}$ ve $X_{(r+1)}$ ardışık sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonu,

$$f_{r,r+1}(x_r, x_{r+1}) = \frac{n!}{(r-1)!(n-r-1)!} \{1 - F(x_r)\}^{r-1} \{1 - F(x_{r+1})\}^{n-r-1} \\ \times f(x_r) f(x_{r+1}), \quad -\infty < x_r < x_{r+1} < \infty$$

olarak elde edilir (Arnold *et. al.* 1992).

2.1.3. İki Sıra İstatistiği Arasındaki Farkın Dağılımı

$W_{rs} = X_{(s)} - X_{(r)}$ olsun. W_{rs} rasgele değişkeninin olasılık yoğunluk fonksiyonu,

$$f(w_{rs}) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \\ \times \int_{-\infty}^{\infty} \{F(x)\}^{r-1} f(x) \{F(x + w_{rs}) - F(x)\}^{s-r-1} \quad (2.15)$$

$$\times f(x + w_{rs}) \{1 - F(x + w_{rs})\}^{n-s} dx$$

dir (David 1969).

$X_1, X_2, \dots, X_n$ rasgele değişkenleri birbirinden bağımsız ve $(0,1)$ aralığında düzgün dağılıma sahip olsun. Bu durumda, $W_{rs} = X_{(s)} - X_{(r)}$ nin dağılım fonksiyonu ve olasılık yoğunluk fonksiyonu bulunmak istensin.

$X_{(r)}$, r nci sıra istatistiği $B(r, n-r+1)$ olan beta dağılımlıdır. $X_{(r)}$ nin olasılık yoğunluk fonksiyonu,

$$f_r(x) = \begin{cases} \frac{1}{B(r, n-r+1)} x^{r-1} (1-x)^{n-r} & , \quad 0 \leq x \leq 1 \\ 0 & , \quad \text{diğer yerlerde} \end{cases}$$

dir.

(2.11)' e göre $X_{(r)}$ ve $X_{(s)}$ sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonu,

$$f_{rs}(x, y) = \begin{cases} \frac{n!}{(r-1)!(s-r-1)!(n-s)!} x^{r-1} (y-x)^{s-r-1} (1-y)^{n-s} & , \quad 0 \leq x \leq y \leq 1 \\ 0 & , \quad \text{diğer yerlerde} \end{cases}$$

dir.

$x \geq 1 - w_{rs}$ için $f(x + w_{rs}) = 0$ olduğu kullanılarak (2.15) den,

$$f(w_{rs}) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \int_0^{1-w_{rs}} x^{r-1} w_{rs}^{s-r-1} (1-x-w_{rs})^{n-s+r} dx$$

olarak elde edilir.

$x = y(1 - w_{rs})$ dönüşümü kullanılarak,

$$f(w_{rs}) = \frac{1}{B(s-r, n-s+r+1)} w_{rs}^{s-r-1} (1-w_{rs})^{n-s+r} , \quad 0 \leq w_{rs} \leq 1$$

biçiminde yazılır.

Sonuç olarak, $X_i, i=1,2,\dots,n$ $(0,1)$ aralığında düzgün dağılımlı rasgele değişkenler olduğunda W_{rs} , sadece $s-r$ ye bağlı olan, s ve r ye ayrı ayrı bağlı olmayan Beta dağılımına sahiptir (David 1969).

Teorem 2.1. $X_1, X_2, \dots, X_n$ rasgele değişkenleri birbirinden bağımsız ve parametresi θ olan üstel dağılımdan alınmış rasgele değişkenler olsun.

$X_{(i)}$, $1 \leq i \leq n$ sıra istatistiklerinden elde edilen ardışık farklar,

$$S_i = X_{(i)} - X_{(i-1)} \quad , \quad 1 \leq i \leq n$$

şeklinde ifade edilsin. Buna göre,

$$P(S_i \leq x) = 1 - e^{-\theta(n-i+1)x} , \quad 1 \leq i \leq n$$

olarak elde edilir (Sukhatme 1937).

2.1.4. Ardışık Sıra İstatistikleri Arasındaki Farkların (Spacings) Dağılımı

Bu kısımda ağırlıklı olarak, ardışık sıra istatistikleri arasındaki farkların dağılımı üzerine Pyke (1965) tarafından verilen incelemeler özetlenmektedir. Ardışık sıra istatistikleri arasındaki farkın oluşturduğu aralıkların (spacings) dağılımı geliştirilmiş, düzgün ve üstel olmak üzere üç kısımda incelenmektedir.

n tane rasgele değişkeni bileşen olarak içeren rasgele vektör ile $n+1$ tane, n tane veya $n-1$ tane fark (Spacings) oluşturulabilmektedir. Rasgele örneklemin alındığı

yığın düzgün dağılımda olduğu gibi sınırlı aralıklı değerlerin alındığı yığınlar için, üstel dağılımda olduğu gibi değer aralığının bir tarafı sınırsız, normal dağılımda olduğu gibi değer aralığının iki tarafı sınırsız olabilmektedir.

Tanım 2.1. $X_1, X_2, \dots, X_n$ birbirinden bağımsız ve aynı dağılıma sahip rasgele değişkenler olsun. Sıra istatistikleri,

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

olmak üzere (Arnold *et. al.* 1992) ardışık sıra istatistikleri arasındaki farkla tanımlanan,

$$X_{(i+m)} - X_{(i)} \quad , \quad 1 \leq i < i+m \leq n$$

bir m - fark (m - spacings) olarak ifade edilir (Miller 2003).

2.1.4.1. Ardışık Sıra İstatistikleri Arasındaki Farkların (Spacings) Genelleştirilmiş Dağılımı

$X = (X_1, X_2, \dots, X_n)$, n tane bağımsız ve aynı dağılıma sahip rasgele değişkenden oluşan örneklem ve $T = (X_{(1)}, X_{(2)}, \dots, X_{(n)})$, sıra istatistiklerinden oluşan vektör olsun. T vektörünün herhangi bir alt vektörü $T_k = (X_{(k_1)}, X_{(k_2)}, \dots, X_{(k_m)})$, $1 \leq k_1 < k_2 < \dots < k_m \leq n$ in ortak olasılık yoğunluk fonksiyonu,

$$f_{T_k}(x_1, x_2, \dots, x_m) = \begin{cases} n! \prod_{j=1}^{m+1} \frac{\{F(x_j) - F(x_{j-1})\}^{k_j - k_{j-1} - 1} f(x_j)}{(k_j - k_{j-1} - 1)!}, & x_1 < x_2 < \dots < x_m \\ 0, & \text{diğer yerlerde} \end{cases} \quad (2.16)$$

dir. Burada $x_0 = -\infty, x_{m+1} = +\infty, k_{m+1} = n+1$ ve $f(x_{m+1}) = 1$ dir. T vektörünün ortak olasılık yoğunluk fonksiyonu,

$$f_T(x_1, x_2, \dots, x_n) = \begin{cases} n! f(x_1) f(x_2) \dots f(x_n) & , \quad x_1 < x_2 < \dots < x_n \\ 0 & , \quad \text{diğer yerlerde} \end{cases} \quad (2.17)$$

dir.

$S_i = X_{(i)} - X_{(i-1)}$, $2 \leq i \leq n$ ve $S = (S_2, S_3, \dots, S_n)$ olmak üzere (2.17) deki lineer dönüşüm,

$$f_S(s_2, s_3, \dots, s_n) = \begin{cases} n! \int_{-\infty}^{\infty} \prod_{i=2}^n f(x + s_2 + s_3 + \dots + s_n) dx & , \quad s_i > 0, 2 \leq i \leq n \\ 0 & , \quad \text{diğer yerlerde} \end{cases} \quad (2.18)$$

dir.

(2.16) veya (2.18) den,

$$\begin{aligned} f_{S_i}(y) &= \int_{-\infty}^{\infty} f_{(X_{(i-1)}, X_{(i)})}(x, x+y) dy \\ &= \frac{n!}{(i-2)!(n-i)!} \int_{-\infty}^{\infty} \{F(x)\}^{i-2} \{1-F(x+y)\}^{n-i} f(x) f(x+y) dy \end{aligned} \quad (2.19)$$

elde edilir.

Ardışık sıra istatistikleri arasındaki farkların (Spacings) olasılık yoğunluk fonksiyonlarını elde etmek kolay değildir. Bu aralıklar üzerinde yapılan çalışmalarda Markov sürecinin özelliklerinden yararlanmak yardımcı olacaktır (Pyke 1965).

S_i ve S_j nin ortak olasılık yoğunluk fonksiyonu, $u, v > 0$ için,

$$f_{(S_i, S_j)}(u, v) = n! \int_{-\infty}^{\infty} \int_{x+u}^{\infty} \frac{\{F(x)\}^{i-2}}{(i-2)!} \cdot \frac{\{F(y) - F(x+u)\}^{j-i-2}}{(j-i-2)!} \quad (2.20)$$

$$\times \frac{\{1 - F(y+v)\}^{n-j}}{(n-j)!} f(x) f(x+u) f(x+u+v) dy dx$$

dir. (2.20) de $j = i + 1$ için,

$$f_{(S_i, S_{i+1})}(u, v) = n! \int_{-\infty}^{\infty} \int_{x+u}^{\infty} \frac{\{F(x)\}^{i-2}}{(i-2)!} \cdot \frac{\{1 - F(x+u+v)\}^{n-i-1}}{(n-i-1)!} \\ \times f(x) f(x+u) f(x+u+v) dy dx$$

olarak elde edilir.

2.1.4.2. Düzgün Dağılım Durumunda Ardışık Sıra İstatistikleri Arasındaki Farkların (Spacings) Dağılımı

$X = (X_1, X_2, \dots, X_n)$ bağımsız, $(0, 1)$ aralığında düzgün dağılıma sahip rasgele vektör olsun.

X rasgele vektörünün olasılık yoğunluk fonksiyonu,

$$f_X(x_1, x_2, \dots, x_n) = \begin{cases} 1 & , \quad 0 \leq x_i \leq 1, 1 \leq i \leq n \text{ için} \\ 0 & , \quad \text{diğer yerlerde} \end{cases}$$

dir.

$T = (X_{(1)}, X_{(2)}, \dots, X_{(n)})$, sıra istatistiklerinin oluşturduğu rasgele vektör olsun.

T rasgele vektörünün olasılık yoğunluk fonksiyonu,

$$f_T(x_1, x_2, \dots, x_n) = \begin{cases} n! & , \quad 0 \leq x_1 \leq x_2 \leq \dots \leq x_n \leq 1 \\ 0 & , \quad \text{diğer yerlerde} \end{cases} \quad (2.21)$$

dir. $X_{(0)} = 0$ ve $X_{(n+1)} = 1$ olmak üzere ardışık sıra istatistikleri ile oluşturulan farklar

(Spacings) $S_i = X_{(i)} - X_{(i-1)}$, $1 \leq i \leq n+1$ olmak üzere $S = (S_1, S_2, \dots, S_{n+1})$ olsun.

$S_1 + S_2 + \dots + S_{n+1} = 1$ dir. S vektörünün ortak olasılık yoğunluk fonksiyonu,

$$f_S(s_1, s_2, \dots, s_{n+1}) = \begin{cases} n! & , \quad s_i \geq 0 \text{ ve } s_1 + s_2 + \dots + s_{n+1} = 1 \\ 0 & , \quad \text{diğer yerlerde} \end{cases}$$

olmaktadır. Bu nedenle, S nin olasılık ifadesi teorik olarak, $s_1 + s_2 + \dots + s_{n+1} = 0$

derecesinin alt kümesinin miktarına bağlı olarak veya

$\{s_i \geq 0 \ (1 \leq i \leq n), s_1 + s_2 + \dots + s_n \leq 1\}$ ifadesine göre elde edilebilir.

(2.21) de verilen S vektörünün olasılık yoğunluk fonksiyonu aynı sırada herhangi bir permütasyon altında değişmeden kalabilir. Düzgün dağılımlı rasgele örneklemle oluşturulan ardışık sıra istatistikleri arasındaki farklar yer değiştirilebilen (exchangeable) rasgele değişkenlerdir.

Herhangi bir S_i aralığının (Spacing) olasılık yoğunluk fonksiyonu S_1 ' in olasılık yoğunluk fonksiyonuna ve $(S_i, S_j), (i \neq j)$ çiftinin ortak olasılık yoğunluk fonksiyonu (S_1, S_2) ' nin ortak olasılık yoğunluk fonksiyonuna eşittir. Buna göre, $x, y \geq 0$ ve $x + y \leq 1$ için,

$$F_{S_i}(x) = F_{S_j}(x) = F_{X_{(1)}}(x) = 1 - (1 - x)^n$$

ve

$$\begin{aligned} F_{S_i, S_j}(x, y) &= P(X_{(1)} \leq x, X_{(2)} - X_{(1)} \leq y) \\ &= n \int_0^x \left\{ 1 - \left(1 - \frac{y}{1-u} \right)^{n-1} \right\} (1-u)^{n-1} du \\ &= 1 - \left\{ (1-x)^n + (1-y)^n - (1-x-y)^n \right\} \end{aligned} \quad (2.22)$$

dir. Buradan S_i ve (S_i, S_j) için olasılık yoğunluk fonksiyonları,

$$f_{S_i}(x) = n(1-x)^{n-1}$$

ve

(2.23)

$$f_{S_i, S_j}(x, y) = n(n-1)(1-x-y)^{n-2}$$

olarak elde edilir.

$X_{(1)} = u$ verildiğinde $X_{(2)} - X_{(1)}$ in koşullu dağılımı, $[0, 1 - u]$ aralığında düzgün dağılıma sahip $n - 1$ tane rasgele değişkenin oluşturduğu örneklemden elde edilen ardışık sıra istatistikleri arasındaki farklarla meydana gelen aralıkların dağılımına eşittir.

2.1.4.3. Üstel Dağılım Durumunda Ardışık Sıra İstatistikleri Arasındaki Farkların (Spacings) Dağılımı

$X_i, i = 1, 2, \dots, n$ rasgele değişkeninin olasılık yoğunluk fonksiyonu f , $f(x) = \lambda e^{-\lambda x}$, $x > 0$ olarak verilsin. $X_1, X_2, \dots, X_n$ rasgele örneklemden elde edilen $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ sıra istatistikleri ile ve $X_{(0)} = 0$, $S_i = X_{(i)} - X_{(i-1)}$ ($1 \leq i \leq n$) olmak üzere tanımlanan $S = (S_1, S_2, \dots, S_n)$ vektörünün olasılık yoğunluk fonksiyonu, $s_i > 0$ ($1 \leq i \leq n$) için,

$$f_S(s_1, s_2, \dots, s_n) = n! \lambda^n \prod_{i=1}^n e^{-\lambda(s_1 + \dots + s_i)} \quad (2.24)$$

$$= n! \lambda^n e^{-\lambda \sum_{i=1}^n (n-i+1)s_i}$$

$$= \prod_{i=1}^n \lambda (n-i+1) e^{-\lambda (n-i+1)s_i}$$

dir.

Görüldüğü üzere S vektörünün olasılık yoğunluk fonksiyonu n marjinal üstel olasılık yoğunluk fonksiyonunun çarpımına eşittir. Bu durumda, $S_1, S_2, \dots, S_n$ bağımsız ve parametreleri sırası ile $\lambda n, \lambda(n-1), \dots, \lambda$ olan üstel dağılıma sahip rasgele değişkenler olmaktadır. Bundan dolayı, $\{\lambda(n-i+1)S_i : 1 \leq i \leq n\}$ olarak tanımlanan

rasgele değişkenler birbirinden bağımsız ve ortalaması 1 olan üstel dağılımdan alınmış rasgele değişkenlerdir. Bu rasgele değişkenlere normalleştirilmiş aralıklar (Spacings) denir. Bu dağılım varsayımı altında ardışık sıra istatistikleri arasındaki farklar (Spacings) birbirinden bağımsızdır.

(2.24) denklemi ve normalleştirilmiş aralıkların (Spacings) gösterilen olasılıkları esas olarak Sukhatme (1937) tarafından tanımlanmıştır (Pyke 1965).

2.2. Yer Değiştirilebilirlik (Exchangeability)

Port (1994)' de ifade edildiği hali ile yukarıda da değinilen yer değiştirilebilirlik (exchangeability) tanımı ve konu ile ilişkili bazı önemli teoremler aşağıda verilmiştir:

Tanım 2.2. $X_1, X_2, \dots, X_n$ rasgele değişkenlerinin ortak dağılım fonksiyonu F olsun. $X_1, X_2, \dots, X_n$ rasgele değişkenlerinin $n!$ tane değişik takılandırması için τ , $X_1, X_2, \dots, X_n$ rasgele değişkenlerinin $n!$ takılandırmasının herhangi birini gösteren yer değiştirmelerinin bir fonksiyonu olmak üzere, F dağılımı için her $B \in \mathcal{B}(\mathcal{R}^n)$ ve τ için, $F(\tau^{-1}(B)) = F(B)$ ise $X_1, X_2, \dots, X_n$ yer değiştirilebilir rasgele değişkenler denir.

$X_1, X_2, \dots, X_n$ rasgele değişkenleri yer değiştirilebilir (exchangeable) ise $\mathcal{R}^n$ de tanımlı $X = (X_1, X_2, \dots, X_n)$ rasgele vektörünün $n!$ tane takılandırması aynı dağılıma sahiptir denir (Port 1994).

Benzer olarak, τ permütasyonları için, τX ve X aynı dağılıma sahipse X rasgele vektörü yer değiştirilebilir (exchangeably) dağılımlıdır.

Önerme 2.1. $X_1, X_2, \dots, X_n$ birbirinden bağımsız ve aynı dağılımlı rasgele değişkenler ise $X_1, X_2, \dots, X_n$ yer değiştirilebilir (exchangeable)dir (Port 1994).

Sonuç 2.1. $X_1, X_2, \dots, X_n$ rasgele değişkenleri yer değiştirilebilir (exchangeable) olsun. g ölçülebilir bir fonksiyon olmak üzere $Y_i = g(X_i), 1 \leq i \leq n$ olsun. $Y_1, Y_2, \dots, Y_n$ yer değiştirilebilir (exchangeable) rasgele değişkenlerdir (Port 1994).

2.3. Laplace Dönüşümü

Uyan (1980)' da ifade edildiği hali ile Laplace dönüşümünün tanımı ve konuya ilişkin bazı önemli teoremler aşağıda verilmiştir:

Tanım 2.3. $f(t)$ fonksiyonu $t > 0$ için t değişkenine ait bir fonksiyon olsun. $f(t)$ fonksiyonunun Laplace dönüşümü $L[f(t)] = F(s)$ biçiminde gösterilir. " L " sembolüne Laplace dönüşümü operatörü denir. Söz konusu $f(t)$ fonksiyonunun Laplace dönüşümü,

$$L[f(t)] = F(s) = \int_0^{\infty} f(t) e^{-st} dt \quad (2.25)$$

şeklinde yazılabilir. Buradaki s parametresi reel bir değerdir.

Teorem 2.2. (Laplace dönüşümü operatörü L nin lineerlik özelliği)

f_1 ve f_2 fonksiyonlarının Laplace dönüşümleri, $L[f_1(t)] = F_1(s)$ ve $L[f_2(t)] = F_2(s)$ olsun. c_1 ve c_2 sabit değerler olmak üzere,

$$\begin{aligned} L[c_1 f_1(t) + c_2 f_2(t)] &= c_1 L[f_1(t)] + c_2 L[f_2(t)] \\ &= c_1 F_1(s) + c_2 F_2(s) \end{aligned} \quad (2.26)$$

dir.

2.4. Bölünmüş Farklar

Herhangi bir polinomiyal fonksiyon bölünmüş farklardan hareketle yeniden ifade edilebilir. Bölünmüş farklar, bir fonksiyon için polinomiyal interpolasyonun bir aracıdır. Bölünmüş farklar, aşağıda verilen Tanım 2.4. e göre elde edilmiş fonksiyonlar olarak ifade edilebilir.

Tanım 2.4. f fonksiyonu D reel sayılar kümesinde tanımlı bir fonksiyon olsun. $f(x)$ fonksiyonunda x değişkeni için $x_0, x_1, x_2, \dots, x_n$ değerleri verilmiş olsun. $x_0, x_1, x_2, \dots, x_n$ değerlerinin hepsi birbirinden farklı olsun.

$f(x)$ fonksiyonunun $x_0, x_1 \in D$ için bölünmüş farkı $f[x_0, x_1]$,

$$f[x_0, x_1] = \frac{f(x_0) - f(x_1)}{x_0 - x_1}$$

olarak tanımlanır. Benzer şekilde, x_1 ve x_2 noktaları için $f(x)$ fonksiyonunun bölünmüş farkı,

$$f[x_1, x_2] = \frac{f(x_1) - f(x_2)}{x_1 - x_2}$$

dir.

x_0, x_1 ve x_2, D kümesinden alınsın ve herbiri farklı üç nokta olsun. x_0, x_1 ve x_2 noktaları için f fonksiyonunun bölünmüş farkı,

$$f[x_0, x_1, x_2] = \frac{f[x_0, x_1] - f[x_1, x_2]}{x_0 - x_2}$$

dir. Burada $f[x_0, x_1]$ ve $f[x_1, x_2]$ de ortak noktanın x_1 olduğu görülmektedir. Benzer olarak, D kümesinden alınmış birbirinden farklı $x_0, x_1, \dots, x_n$ noktaları için bölünmüş farklar,

$$f[x_0, x_1, \dots, x_n] = \frac{f[x_0, x_1, \dots, x_{n-1}] - f[x_1, x_2, \dots, x_n]}{x_0 - x_n}$$

olarak yazılır (Milne ve Thompson 1933).

Teorem 2.3. f fonksiyonunun $x_0, x_1, \dots, x_n$ noktalarındaki bölünmüş farkı,

$$f[x_0, x_1, \dots, x_n] = \sum_{j=0}^n \frac{f(x_j)}{\prod_{\substack{i=0 \\ i \neq j}}^n (x_j - x_i)}$$

dir (Jankovic 2000).

Sonuç 2.2. $f[x_0, x_1, \dots, x_n]$ olarak tanımlanan bölünmüş farklar teoreme göre simetriktir (Jankovic 2000).

Teorem 2.4. (Newton bölünmüş fark interpolasyon polinomu)

$f(x)$ fonksiyonu bölünmüş farklarla aşağıdaki biçimde ifade edilebilir. Burada x_0 yerine x alınmaktadır.

$$\begin{aligned}
f(x) &= f(x_1) + (x - x_1)f[x_1, x_2] + (x - x_1)(x - x_2)f[x_1, x_2, x_3] + \\
&+ \dots + (x - x_1)(x - x_2) \dots (x - x_{s-1})f[x_1, x_2, \dots, x_s] + \dots \\
&+ (x - x_1)(x - x_2) \dots (x - x_{n-1})f[x_1, x_2, \dots, x_n] + \\
&+ (x - x_1)(x - x_2) \dots (x - x_n)f[x, x_1, x_2, \dots, x_n]
\end{aligned}$$

(Milne ve Thompson 1933).

Teorem 2.5. (Rolle Teoremi) $f(x)$ fonksiyonu $[a, b]$ aralığında sürekli, türevlenebilir bir fonksiyon olsun. a ve b noktaları $f(x) = 0$ olacak biçimde kök değerleri ise $f'(x) = 0$ olacak biçimde $x \in (a, b)$ vardır (Milne ve Thompson 1933).

Teorem 2.6. Newton' un genel interpolasyon formülü kalan terim $R_n(x)$ ile birlikte,

$$f(x) = f(x_1) + \sum_{s=1}^{n-1} (x - x_1)(x - x_2) \dots (x - x_s) f[x_1, x_2, \dots, x_{s+1}] + R_n(x)$$

biçimindedir ve burada kalan terim,

$$R_n(x) = (x - x_1)(x - x_2) \dots (x - x_n) f[x, x_1, x_2, \dots, x_n]$$

dir. Dolayısıyla,

$$P_{n-1}(x) = f(x_1) + \sum_{s=1}^{n-1} (x - x_1)(x - x_2) \dots (x - x_s) f[x_1, x_2, \dots, x_{s+1}]$$

$(n-1)$ nci dereceden polinom olmak üzere,

$$R_n(x) = f(x) - P_{n-1}(x)$$

biçiminde ifade edilir. $P_{n-1}(x)$ in x e göre $(n-1)$ nci türevi,

$$P_{n-1}^{(n-1)}(x) = (n-1)! f[x_1, x_2, \dots, x_n]$$

dir.

Şimdiye kadar verilen ifadeler için $f(x)$ fonksiyonu sınırsız olarak alınmıştır.

Rolle teoremine göre $R_n(x)$ in birinci türevi $R_n'(x)$ de, (a, b) deki $n-1$ nokta kaybolmaktadır. Benzer olarak kalan terimin ikinci türevi $R_n''(x)$ de, (a, b) deki $n-2$ nokta kaybolmaktadır. Rolle teoremi tekrar uygulandığında $R_n^{n-1}(\eta) = 0$ olacak şekilde (a, b) aralığında bir η noktası bulunmaktadır. Bu durumda,

$$f[x_1, x_2, \dots, x_n] = \frac{f^{(n-1)}(\eta)}{(n-1)!}$$

yazılabilir.

Bu sonuçlardan hareketle şu ifade edilebilir:

$f(x)$, reel değerli bir fonksiyon olsun. $f(x)$ fonksiyonu, $[a, b]$ aralığında tanımlı ve (a, b) de n kez türevlenebilir olsun. $[a, b]$ aralığından alınan birbirinden farklı $x_0, x_1, \dots, x_n$ noktaları için,

$$f[x_0, x_1, \dots, x_n] = \frac{f^{(n)}(\xi)}{n!}$$

olacak biçimde (a, b) aralığında bir ξ noktası bulunabilir (Milne ve Thompson 1933).

Teorem 2.7. Bölünmüş farklar sonlu integrallerle de ifade edilebilir.

$(0, 1)$ aralığında türevlenebilen bir fonksiyon f olarak tanımlansın. $[0, 1]$ aralığından alınan birbirinden farklı $x_1, x_2, \dots, x_n$ noktaları mevcut olsun. f fonksiyonunun $x_1, x_2, \dots, x_n$ noktalarındaki bölünmüş farkları,

$$f[x_1, x_2, \dots, x_n] = \int_0^1 \int_0^{t_1} \dots \int_0^{t_{n-2}} f^{(n-1)}(u_n) dt_{n-1} \dots dt_2 dt_1 \quad (2.27)$$

olarak ifade edilebilir.

Burada,

$$u_n = (1 - t_1)x_1 + (t_1 - t_2)x_2 + \dots + (t_{n-2} - t_{n-1})x_{n-1} + t_{n-1}x_n$$

ve $t_1, t_2, \dots, t_{n-1}$ birbirinden farklıdır (Milne ve Thompson 1933).

3. PROBLEM VE ÇÖZÜMÜNE YAKLAŞIMLAR

Giriş kısmındaki Mack (1948)' in tanımladığı k ' lı topluluklar ile ilgili 3 ncü problem şu şekildedir:

N çok büyük bir doğal sayı olmak üzere N sayıda kutucuk ve $n < N$ olacak biçimde n tane top olsun. n top rasgele bir şekilde N kutunun içine dağıtılsın. N tane kutuda verilen p sayıdaki ardışık kutularda k tane topun yer alması olarak tanımlanan k ' lı topluluğun olasılığı bulunmak isteniyor.

Rasgele olarak kutulara yerleştirilen k top için 1 nci top ile k ncü top arasındaki uzaklık $\leq p$ kutucuk sayısı olacaktır. Böylece k ' lı topluluk elde edilebilir, ancak k top bir başka $k+1$ ' li topluluğa ait olmayacaktır.

* Herhangi bir topun N_i nci kutucuğa düşmesi olasılığı:

$$\frac{\binom{N-1}{n-1}}{\binom{N}{n}}$$

dir.

* $k-1$ topun $p-1$ kutucuğa düşmesi ve geriye kalan $n-k$ topun $N-p$ kutucuğa düşmesi olasılığı:

$$\frac{\binom{p-1}{k-1} \binom{N-p}{n-k}}{\binom{N-1}{n-1}}$$

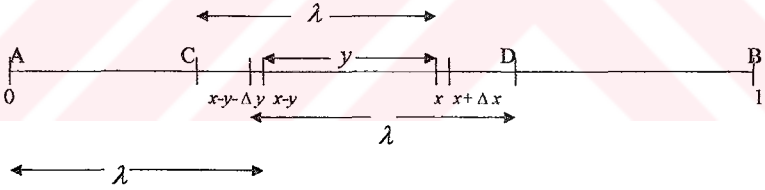
dir.

Sonuç olarak bu iki olasılığın çarpımı, N kutucuktan ardışık p tane kutucuğa k tane topun yerleştirilmesi ya da herhangi bir k ' lı topluluğun olasılığını verir:

$$\frac{\binom{p-1}{k-1} \binom{N-p}{n-k}}{\binom{N}{n}}.$$

Problem $(0,1)$ aralığındaki düzgün dağılım için ele alınsın. Düzgün dağılımdan alınan n rasgele nokta olsun.

Burada k ' lı topluluk, λ uzunluğunun içinde k noktadan oluşan kümedir.



Şekil 3.1. Sürekli durumda düzgün bir doğru boyunca k ' lı topluluğun konumu

A -başlangıç noktasından x kadar uzakta bir nokta alındığında $(x, x + \Delta x)$ gibi çok küçük bir aralığa k ncı noktanın düşmesi olasılığına bakılırsa;

$$\int_x^{x+\Delta x} 1 dt = \Delta x$$

dir.

Böylece, k ncı noktanın yerinin belirlenmesi ile k ' lı *topluluk* oluşturulmaya başlanmıştır. x noktasından y ye kadar $|CD|$ içinde kalacak şekilde C noktasına doğru öteleme yapılsın. Oluşturulan $(x - y - \Delta y, x - y)$ aralığına 1 nci noktanın düşmesi olasılığı;

$$\int_{x-y-\Delta y}^{x-y} 1 dt = \Delta y$$

dir.

Başka bir deyişle, $(0,1)$ aralığındaki düzgün dağılıma sahip n noktadan herhangi birinin $(x, x + \Delta x)$ aralığına düşmesi olasılığı Δx , $(x - y - \Delta y, x - y)$ aralığına düşmesi olasılığı Δy dir. $|CD| = 2\lambda - y$ kadardır.

Sonuç olarak, iki noktanın Δx ile Δy olasılıklarını içeren aralıkların (sırası ile $(x, x + \Delta x)$ ve $(x - y - \Delta y, x - y)$ aralıkları) arasına düşmesi olasılığı $n\Delta x(n-1)\Delta y$ dir.

İki noktanın yerinin belirlenmesinden sonra geriye kalan $k-2$ noktanın k ' lı *topluluk* tanımını gereği $(x - y, x)$ aralığına düşmesi isteniyor. $n-2$ noktadan kalan $k-2$ noktanın bu aralığa düşmesi olasılığı $C_{k-2}^{n-2} y^{k-2}$ dir.

$|CD| = 2\lambda - y$ aralığı içinde oluşturulan k ' lı topluluk $x - y - \Delta y$ ile $x + \Delta x$ aralığını kapsamaktadır. Bu nedenle, kalan noktaların k ' lı topluluk tanımı kapsamı dışında olması olasılığı da katılırsa, bir k ' lı topluluğun $(x, x + \Delta x)$ ve $(x - y - \Delta y, x - y)$ arasında bulunması olasılığı:

$$n\Delta x(n-1)\Delta y C_{k-2}^{n-2} y^{k-2} (1-2\lambda+y)^{n-k}$$

olmaktadır. Mack (1948), problem yaklaşık hesaplamalar ile çözülmektedir.

Benzer olarak, Naus (1965 ve 1966)' un doğru boyunca dizilmiş noktaların belirli bir λ uzunluklu aralıklarda yer alan gözlemlerin sayılarının maksimumuna ilişkin dağılımı üzerine çalışması bulunmaktadır.

n tane gözlem $(0,1)$ aralığındaki düzgün dağılımdan rasgele çekilmiş olsun. $P(k|n;\lambda)$, $(0,1)$ aralığında λ uzunluktaki bir aralıkta kümelenmiş gözlemlerin maksimum sayısının k veya k dan daha fazla olması olasılığını gösterebilir (Naus 1965).

$X_1, X_2, \dots, X_n$ rasgele değişkenlerinin herbiri birbirinden bağımsız ve $(0,1)$ aralığında tanımlı düzgün dağılımlı olsun. Bu rasgele örnekleme ilişkin $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonu;

$$P(X_{(1)} \leq x_1, X_{(2)} \leq x_2, \dots, X_{(n)} \leq x_n) = n! \int \int \dots \int dx_1 dx_2 \dots dx_n$$

dir.

$1 - P(k|n;\lambda)$ olasılığı $(0,1)$ aralığında yer alan λ uzunluğundaki bir aralıkta k nokta bulunmaması olasılığıdır. Başka bir deyişle, $1 - \lambda$ uzunluğundaki aralıkta $n - k$ dan daha az sayıda nokta gözlemlenmesi olasılığıdır.

$$0 \leq X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)} \leq 1$$

sıra istatistikleri, (3.1)

$$X_{(k+i-1)} - X_{(i)} > \lambda, \quad \forall i = 1, 2, \dots, n - k + 1$$

olarak tanımlansın. (3.1) de ifade edilen iki sıra istatistiği arasındaki farklardan en büyüğünün λ uzunluğundan büyük olması olasılığı, λ uzunluğu dışında kalan kısımlarda $n-k$ dan daha az sayıda noktanın bulunması olasılığı demektir.

$P(k|n; \lambda)$ olasılığının (3.1) deki ifadelerle göre hesabı çok işlem içeren integral hesaplarına dayanmaktadır. $k > n/2$ ve $\lambda \leq 1/2$ olduğu durumda $P(k|n; \lambda)$ olasılığının, $X_1, X_2, \dots, X_n$ nin λ ncı dereceden polinom olarak gösterileceği Naus (1965) tarafından ifade edilmiştir. Bu nedenle Naus (1965) hesaplamayı yaklaşımla yapmaktadır.

L tamsayı olmak üzere $P(k|n; \lambda)$ olasılığında λ yerine $1/L$ alındığında $P(k|n; \lambda)$ $1/L$ ncı dereceden polinom olarak karşımıza çıkmaktadır. $P(k|n; \lambda)$ ile $P(k|n; 1/L)$ nin katsayıları birbirinin aynısı olmaktadır. $k > n/2$ ve $\lambda \leq 1/2$ için $P(k|n; \lambda)$ ve $P(k|n; 1/L)$ yaklaşımı ile hesaplanmaktadır. $\lambda \geq 1/2$ olduğu durumda ise $P(k|n; \lambda)$ olasılığı doğrudan bu yaklaşımla bulunmaktadır (Naus 1965).

Teorem 3.1. (Naus 1965) $P(k|n; \lambda)$, $(0,1)$ aralığında λ uzunluktaki bir aralıkta kümelenmiş gözlemlerin maksimum sayısının k veya k dan daha fazla olması olasılığı aşağıdaki gibi tanımlanır:

$$\begin{aligned}
P(k|n; \lambda) &= C(k|n; \lambda) - R(k|n; \lambda) & , \lambda \geq 1/2 \text{ ve } k > (n+1)/2 \\
&= C(k|n; \lambda) & , \lambda \leq 1/2 \text{ ve } k > n/2
\end{aligned}$$

Teorem 3.1. de geçen ifadelerin daha açık bir yazımı için aşağıda verilen bazı notasyonlar kullanılacaktır:

$$b(k|n; \lambda) = \binom{n}{k} \lambda^k (1 - \lambda)^{n-k}$$

$$F_b(k|n; \lambda) = \sum_{i=0}^k b(i|n; \lambda)$$

ve

$$G_b(k|n; \lambda) = \sum_{i=k}^n b(i|n; \lambda)$$

olsun.

Bu durumda Teorem 3.1. de verilen C, H ve R sırasıyla,

$$\begin{aligned}
C(k|n; \lambda) &= (n - k + 1)b(k - 1|n; \lambda) - (n - k)b(k|n; \lambda) \\
&\quad + G_b(k|n; \lambda) + G_b(k + 1|n; \lambda) \\
&= (n - k + 1)[G_b(k - 1|n; \lambda) + G_b(k + 1|n; \lambda)] \\
&\quad - 2(n - k)G_b(k|n; \lambda)
\end{aligned}$$

ve

$$H(k|n;\lambda) = \frac{k(1-\lambda)}{\lambda} F_b(n-k|k-1;(1-\lambda)/\lambda) \\ -(n-k+1)F_b(n-k+1|k;(1-\lambda)/\lambda)$$

olmak üzere,

$$R(k|n;\lambda) = \sum_{y=k}^n b(y|n;\lambda) F_b(n-k|y;(1-\lambda)/\lambda) + H(k|n;\lambda) b(k|n;\lambda)$$

dir.

Giriş kısmında verilen 5 nci problem Seth (1950) ve Lieblein (1952) tarafından verilmiştir. Örneklem çapı küçük olduğunda örneğin $n=3$ olarak alındığında Seth (1950) ve Lieblein (1952) problemi normal dağılım ve diğer dağılımlar için ele almıştır.

Çapı 3 birim olan bir örneklem için sıra istatistikleri $X_{(1)}, X_{(2)}, X_{(3)}$ olsun. En yakın iki gözlemden elde edilen rasgele değişkenler X' ve X'' olarak gösterilsin. $X' \geq X''$ için,

$$Y_1 = \frac{X' - X''}{X_{(3)} - X_{(1)}}, Y_2 = \frac{X' - X''}{2}, Y_3 = \frac{X' + X''}{2}$$

olacak biçimde Y_i , $i=1,2,3$ rasgele değişkenleri tanımlansın.

Örnek olarak,

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

olarak alındığında Y_1 rasgele değişkeninin olasılık yoğunluk fonksiyonu,

$$f(y_1) = \frac{3\sqrt{3}}{\pi(1 - y_1 + y_1^2)} \quad , \quad 0 \leq y_1 \leq 1/2$$

olarak bulunur.

Y_2 nin olasılık yoğunluk fonksiyonu,

$$f(y_2) = \frac{4\sqrt{3}}{\pi} e^{-y_2^2} \int_{3y_2}^{\infty} e^{-\frac{1}{3}x^2} dx dy_2 \quad , \quad 0 \leq y_2 < \infty$$

dir.

$Y_3 = \frac{X' + X''}{2}$ rasgele değişkeninin olasılık yoğunluk fonksiyonu,

$$f(y_3) = \frac{6}{\pi\sqrt{2}\pi} \left[\int_0^{\infty} \int_{y_3+3y_2}^{\infty} e^{-\left(\frac{1}{2}x^2 + y_2^2 + y_3^2\right)} dx dy_2 dy_3 + \right. \\ \left. + \int_0^{\infty} \int_{-\infty}^{y_3-3y_2} e^{-\left(\frac{1}{2}x^2 + y_2^2 + y_3^2\right)} dx dy_2 dy_3 \right] \quad , \quad -\infty < y_3 < \infty$$

biçimindedir.

Probleme ardışık sıra istatistikleri arasındaki farklarla oluşan aralıkların lineer kombinasyonları olarak bakıldığında Huffer (1988)' ın çalışması şu şekilde verilmektedir:

$n-1$ çaplı $(0,1)$ aralığında düzgün dağılımlı rasgele örnekleme ele alalım. Birim aralıkta değer alan gözlemler aralığı n parçaya böler. Bu parçaların uzunlukları $S_1, S_2, \dots, S_n$ olarak gösterilmektedir. Bu uzunluklara ardışık sıra istatistikleri arasındaki farklar (spacings) adı verilmektedir. $S = (S_1, S_2, \dots, S_n)'$ rasgele sütun vektörü $S_1 + S_2 + \dots + S_n = 1$ koşuluna göre düzgün dağılıma sahiptir. $S_i, i=1, 2, \dots, n$ lerin yer değiştirilebilir (exchangeable) rasgele değişkenler olduğu 2 nci bölümde ifade edilmiştir.

$\Gamma: k \times n$ lik reel bir matris olsun. (ΓS) rasgele vektörünün dağılımı araştırılmak istensin. (ΓS) vektörünün olasılık ölçüsü $P(\Gamma)$ ile gösterilmektedir. $A, \mathfrak{R}^k$ nın herhangi bir ölçülebilir alt kümesi olmak üzere $\Pr(\Gamma S \in A) = P(\Gamma)(A)$ dir. Bazı durumlarda $P(\Gamma)$ basit terimlerin toplamı gibi gösterilebilir.

Herhangi bir $\xi \in \mathfrak{R}^k$ vektörü için,

$\Gamma_{i,\xi}: \Gamma$ matrisinin i nci sütunun ξ sütun vektörü ile yer değiştirmesi sonucu elde edilen $k \times n$ lik bir matristir.

Teorem 3.2. (Huffer 1988) $d = (d_1, d_2, \dots, d_n)'$ ve $\sum_{i=1}^n d_i = 1$ olan herhangi reel sayılar olsun. $\xi = \Gamma d$ olarak tanımlansın.

$$P(\Gamma) = \sum_{i=1}^n d_i P(\Gamma_{i,\xi})$$

dir (Huffer 1988).

Tüm i ler için $d_i \geq 0$ olduğu durumda $P(\Gamma)$ olasılığı $P(\Gamma_{i,\varepsilon})$ ların konveks kombinasyonu olmaktadır.

Teorem 3.2. Huffer (1988) tarafından aşağıdaki biçimde ispatlanmaktadır:

Teoremin ispatı sonlu farkların hesabında kullanılan bölünmüş farklar ile yapılmaktadır.

Herhangi bir f fonksiyonu için bölünmüş fark,

$$f[x_1, x_2] = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

şeklinde tanımlanmaktadır. Burada $f[x_1, x_2], f(x)$ fonksiyonunun x_1 ve x_2 değerleri için bölünmüş farkıdır.

Buna göre, $n > 2$ için bölünmüş farklar,

$$f[x_1, x_2, \dots, x_n] = \frac{f[x_2, x_3, \dots, x_n] - f[x_1, x_2, \dots, x_{n-1}]}{x_n - x_1}$$

dir. Bu tanım sadece $x_1, x_2, \dots, x_n$ değerlerinin hepsi farklı olduğunda geçerlidir.

f fonksiyonunun türevleri mevcutsa $f[x_1, x_2, \dots, x_n]$ bölünmüş farklarla tanımlanan fonksiyonların herhangi $x_i, i = 1, 2, \dots, n$ değerleri için limitleri vardır.

Aşağıda verilen (3.2) ve (3.3) eşitlikleri $f[x_1, x_2, \dots, x_n]$ in $x_1, x_2, \dots, x_n$ e göre simetrik olduğunu gösterir.

$x_1, x_2, \dots, x_n$ değerlerinin hepsi farklı olduğunda;

$$f[x_1, x_2, \dots, x_n] = \sum_{i=1}^n \left(\frac{f(x_i)}{\prod_{j \neq i} (x_i - x_j)} \right) \quad (3.2)$$

dir. f fonksiyonu sürekli ise ve $n-1$ nci türevi $f^{(n-1)}$ varsa, herhangi $x_1, x_2, \dots, x_n$ değerleri için E , $[0, 1]$ aralığında düzgün dağılımlı n tane birbirinden bağımsız rasgele değişkenin ortak olasılık yoğunluk fonksiyonu altında $S_1, S_2, \dots, S_n$ nin doğrusal kombinasyonlarının değeri olmak üzere,

$$(n-1)! f[x_1, x_2, \dots, x_n] = E f^{(n-1)} \left(\sum_{i=1}^n x_i S_i \right) \quad (3.3)$$

dir (Huffer 1988). (3.3) ün tümevarım yöntemi ile yapılan ispatında S_n ve $(S_1, S_2, \dots, S_{n-1})/(1-S_n)$ vektörünün birbirinden bağımsız olması ve S_n in olasılık yoğunluk fonksiyonunun $0 < x < 1$ için $(n-1)(1-x)^{n-2}$ ve $(S_1, S_2, \dots, S_{n-1})/(1-S_n)$ vektörünün $\mathcal{R}^{n-1}$ de düzgün dağılıma sahip olmasından yararlanır. Herhangi bir $x = (x_1, x_2, \dots, x_n)$ için $f[x] = f[x_1, x_2, \dots, x_n]$ ile gösterilsin. Eğer (3.3) de $f(u) = e^u$ alınırsa S nin Laplace dönüşümü olan,

$$E \exp(\theta S) = (n-1)! \exp[\theta] \quad , \quad \theta = (\theta_1, \theta_2, \dots, \theta_n) \quad (3.4)$$

ifadesine ulaşılır. ΓS vektörünün Laplace dönüşümü (3.4) eşitliğinden yararlanarak,

$$E \exp(\theta(\Gamma S)) = (n-1)! \exp[\theta \Gamma] \quad , \quad \theta = (\theta_1, \theta_2, \dots, \theta_k) \quad (3.5)$$

biçiminde elde edilebilir. Teorem 3.2., doğrudan doğruya (3.5) eşitliğini ve aşağıda ifade edilecek bölünmüş farklar için yardımcı teoremi takip etmektedir. Herhangi bir satır vektörü $x = (x_1, x_2, \dots, x_n)$ ve herhangi bir reel sayı u için,

$x_{i,u}$: x vektöründe i nci elemanın u ile yer değiştirmesi sonucu elde edilen vektördür.

Huffer (1988)' ın bölünmüş farklar için yardımcı teorem şu şekilde ifade edilmektedir: f fonksiyonu $n-1$ nci dereceden sürekli türeve sahip herhangi bir fonksiyon olsun.

Sabitlerin vektörü $d = (d_1, d_2, \dots, d_n)'$ ve $\sum_{i=1}^n d_i = 1$ varsayalım.

Her x vektörü için, $u = xd$ olmak üzere,

$$f[x] = \sum_{i=1}^n d_i f[x_{i,u}]$$

dir.

Bu sonuç $u, x_1, x_2, \dots, x_n$ değerlerinin hepsi birbirinden farklı olduğunda (3.2) nin kullanılması ile ispatlanabilmektedir.

(3.3) formülünden görüldüğü gibi $f[x_1, x_2, \dots, x_n]$, $x_1, x_2, \dots, x_n$ lerin sürekli fonksiyonudur, öyle ki bu yardımcı teorem tüm $x_1, x_2, \dots, x_n$ ler için geçerlidir.

Varsayalım ki $\xi = \Gamma d$ ve $u = \theta \xi$ olsun. $(\theta \Gamma)_{i,u} = \theta \Gamma_{i,\xi}$ olarak gösterilsin.

Yukarıdaki yardımcı teoreme göre,

$$\exp[\theta\Gamma] = \sum_{i=1}^n d_i \exp[\theta\Gamma_{i,\xi}]$$

olarak ifade edilir. Buradaki ifadeye göre, (3.5) formülündeki Laplace dönüşümü, benzer Laplace dönüşümlerin lineer birleşimi (kombinasyonu) biçiminde tanımlanabilmektedir.

Aşağıda verilen örnekler, düzgün dağılımlı rasgele örneklemelerle elde edilen ardışık sıra istatistikleri arasındaki farklarla tanımlanan olayların olasılıklarının bulunmasında Teorem 3.2. nin nasıl kullanılacağını göstermektedir. İlgilenilen olaylar ardışık farklar istatistiklerinin lineer kombinasyonunu içermektedir. Örnek1' de aralıkların tek bir lineer kombinasyonunun dağılımı gösterilmiştir. Sonraki iki örnekteki dağılımlar da Örnek1' deki dağılıma benzer olarak elde edilebilir (Huffer 1988).

Doğru veya çember üzerinde düzgün dağılımlı yığından çekilen rasgele bir örnekleme ait gözlemlerin kümelenmelerine ilişkin olasılıkların hesabı için çeşitli yöntemler Naus (1982)' un çalışmalarında görülmektedir (Huffer 1988). Bu ve benzeri problemler $P\{Y_i > t, 1 \leq i \leq k\}$ veya $P\{Y_i < t, 1 \leq i \leq k\}$ olasılıklarının hesabını içermektedir. Burada $(Y_1, Y_2, \dots, Y_k)' = \Gamma S$ dir. Γ , 0 ve 1 değerlerini içeren bir matristir. Glaz ve Naus (1983) doğru boyunca kümelenmelerin sayısının varyansını, i, j, k keyfi değerler olmak üzere,

$$P\left\{\sum_{i=1}^k S_i < t, \sum_{i=j}^{j+k-1} S_i < t\right\} \quad (3.6)$$

olasılıkların hesabını kullanarak elde etmiştir (Huffer 1988).

Aşağıda verilen örneklerde ilgilenilen olasılıklara Teorem 3.2. uygulanırken (3.6) olasılık hesabına basitleştirilerek sadeleştirmeler yapılmaktadır.

Örnekler:

Örnek1: (Huffer 1988) Γ , $1 \times n$ lik bir vektör olarak alınsın. ΓS nin dağılımı Teorem 3.2. yardımı ile hesaplanmak istenmektedir. $k \leq n$ ve $a_1, a_2, \dots, a_k$ reel sayılar olsun.

$\Gamma = (a_1, a_2, \dots, a_n)$, $i > k$ iken $a_i = 0$ için Teorem 3.2. uygulanacaktır.

Varsayalım ki $a_1 \neq a_2$ olsun.

Teorem 3.2.' ye göre $d_1 + d_2 = 1$ ve $d_1 a_1 + d_2 a_2 = 0$ olacak biçimde yapılabilecek seçimlerden birisi,

$$d_1 = \frac{a_2}{a_2 - a_1}$$

ve

$$d_2 = \frac{a_1}{a_1 - a_2}$$

dir. d_1 ve d_2 Teorem 3.2.' de verilen koşullara uygun olarak seçilmiştir. Böylece elde edilen ξ , $\Gamma_{i,\xi}$ ' yi daha basit hale getirecek ve sonuca ulaşmayı kolaylaştıracaktır.

Buradan,

$$\begin{aligned} P(\Gamma S > t) &= P(\Gamma) \\ &= d_1 P(\Gamma_{1,0}) + d_2 P(\Gamma_{2,0}) \end{aligned} \tag{3.7}$$

dir. $t > 0$ ve $1 \leq i \leq k$ için yukarıdaki (3.7) ifadesi,

$$P\left\{\sum_i a_i S_i > t\right\} = d_1 P\left\{\sum_{i \neq 1} a_i S_i > t\right\} + d_2 P\left\{\sum_{i \neq 2} a_i S_i > t\right\}$$

olarak yeniden ifade edilebilir.

$t > 0$ sabiti için bu olasılık,

$$Q_k(a_1, a_2, \dots, a_k) = P\left\{\sum_{i=1}^k a_i S_i > t\right\} \quad (3.8)$$

biçiminde ifade edilir.

S_i yer değiştirilebilir (exchangeable) olduğu için $Q_k(a_1, 0, a_3, \dots, a_k) = Q_{k-1}(a_1, a_3, \dots, a_k)$ v.b. biçiminde yazılabilir. Bu sebeple (3.8) deki formül,

$$Q_k(a_1, a_2, \dots, a_k) = \frac{a_2 Q_{k-1}(a_2, a_3, \dots, a_k) - a_1 Q_{k-1}(a_1, a_3, \dots, a_k)}{a_2 - a_1} \quad (3.9)$$

biçiminde de yazılabilir. $a_1, a_2, \dots, a_k$ sabitlerinin hepsi farklı olduğunda (3.9) formülü $Q_k(a_1, a_2, \dots, a_k)$ için bölünmüş farklara dönüşür. (3.9) için aşağıda verilen yardımcı teorem (Huffer 1988) daha açıklayıcı olacaktır.

$s, t, x_1, x_2, \dots, x_m$ hepsi birbirinden farklı reel sayılar olduğu varsayalım.

$x_1, x_2, \dots, x_m$ yerine x sembolü kullanılsın. f ve g fonksiyonları tüm x ler için $g(x) = xf(x)$ olsun. Buna göre,

$$g[s, t, x] = \frac{sf[s, x] - tf[t, x]}{s - t} \quad (3.10)$$

dir. (3.10), (3.2)' de bölünmüş farklar için verilen ifadenin uygulanması ile elde edilir.

$X_i, i = 1, 2, \dots, n$ $(0, 1)$ aralığındaki düzgün dağılımdan alınmış rasgele değişkenler olmak üzere,

$$P\{S_1 > x\} = (1-x)^{n-1}, \quad 0 \leq x \leq 1 \text{ için} \quad (3.11)$$

olduğu bilinmektedir. $k > 0$ için, $q_k(u) = u^{k-1} Q_1(u) = u^{k-1} P\{u S_1 > t\}$ biçiminde ifade edilir. Tanımlanan $q_k(u)$ fonksiyonu (3.11) kullanılarak,

$$q_k(u) = \begin{cases} u^{k-1} \left(1 - \frac{t}{u}\right)^{n-1}, & u > t \\ 0, & u \leq t \end{cases} \quad (3.12)$$

biçiminde yeniden ifade edilebilir.

(3.9)' da verilen eşitlikte $k=2$ alınırsa,

$$Q_2(a_1, a_2) = \frac{a_2 Q_1(a_2) - a_1 Q_1(a_1)}{a_2 - a_1} = q_2[a_1, a_2]$$

olarak hesaplanmak istenen olasılık ile bölünmüş farklarla ifade edileceği sonucuna varılır. (3.9) ve (3.10) ifadelerini kullanarak tümevarım yöntemi ile,

$$Q_k(a_1, a_2, \dots, a_k) = q_k[a_1, a_2, \dots, a_k] \quad (3.13)$$

elde edilir. (3.13) formülü $k \leq n$ için, $a_1, a_2, \dots, a_k$ birbirinden farklı olduğunda ispat edilebilmektedir. $k = n$ ve $a_1 = a_2 = \dots = a_n = t$ olduğu durumda ise bölünmüş farklar tanımlanamamaktadır. $a_1 = a_2 = \dots = a_k = 1$ olduğunda ise ardışık aralıkların

lineer kombinasyonları toplamı $\sum_{i=1}^k S_i$ dir. $\sum_{i=1}^k S_i$ ' nin beta dağılımına sahip olduğu

varsayımı altında, bölünmüş farkların tanımından kaynaklanan hesaplama zorlukları çözüme kavuşturulur.

Not1: $Z_1, Z_2, Z_3, \dots$ birbirinden bağımsız ve parametresi 1 olan üstel dağılımdan alınmış rasgele değişkenler olsun. $Z_1, Z_2, Z_3, \dots$ rasgele değişkenlerinin olasılık yoğunluk fonksiyonu,

$$f_{Z_i}(z) = e^{-z}, \quad i = 1, 2, 3, \dots \quad \text{olsun.} \quad (Z_1, Z_2, \dots, Z_n) / \sum_{i=1}^n Z_i \quad \text{ve} \quad (S_1, S_2, \dots, S_n)$$

vektörlerinin dağılımları özdeşdir. $n \rightarrow \infty$ ve k sabit alındığında $(nS_1, nS_2, \dots, nS_k)$ vektörünün dağılımı $(Z_1, Z_2, \dots, Z_k)$ vektörünün dağılımına yakınsamaktadır (Huffer 1988).

(3.12)' de t yerine x/n konulup n değeri sonsuza yaklaştırıldığında,

$$P\left\{\sum_{i=1}^k a_i Z_i > x\right\} = f_k[a_1, a_2, \dots, a_k] \quad , \quad x > 0 \quad (3.14)$$

elde edilir. Burada,

$$f_k(u) = \begin{cases} u^{k-1} e^{-x/u} & , \quad u > 0 \\ 0 & , \quad u \leq 0 \end{cases}$$

dir.

Örnek2: (Huffer 1988) $\Gamma = (\gamma_{ij})$, $2 \times n$ lik bir matris olsun. γ_{ij} , tüm i ve j ler için, 0 ve 1 değerlerine eşit olsun. Bu durumda, $P(\Gamma)$ olasılığının hesabı incelenmektedir.

γ_i : Γ matrisinin i nci sütununu gösterebilir. Γ matrisinin r , s ve t nci sütunları sırası ile $(1,0)'$, $(0,1)'$ ve $(1,1)'$ olsun. Yani,

$$\Gamma = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

olarak verilsin.

$P(\Gamma)$ olasılığı, Γ matrisinin sütunlarının değişimi (permutasyonu) altında değişmez olması sebebi ile bu olasılığın hesabı r , s ve t ye bağlıdır ve

$$P(\Gamma) = P(r, s, t)$$

biçiminde ifade edilir. $r > 0$, $s > 0$ ve $t > 0$ olsun. $\gamma_1 = (1,0)'$, $\gamma_2 = (0,1)'$ ve $\gamma_3 = (1,1)'$ olarak belirlenir. Teorem 3.2.' de verilen koşulları sağlayacak şekilde ζ sütun vektörü, $\gamma_1 + \gamma_2 - \gamma_3 = (0,0)' = 0$ ve istenilen d_i katsayıları, toplamaları 1' e eşit olacak biçimde $1 + 1 - 1 = 1$ olarak belirlenir. Bu seçim sonucunda Teorem 3.2. kullanımı,

$$P(\Gamma) = P(\Gamma_{1,0}) + P(\Gamma_{2,0}) - P(\Gamma_{3,0})$$

eşitliğini vermektedir.

Kümelenmelerin olasılıklarının hesabı: Bu türden IS tanımlamaları ile çember veya doğru üzerinde tanımlı düzgün dağılımlara sahip rasgele değişkenlerin rasgele bir örnekleminde çember veya doğru üzerinde kümelenmelerin olasılıklarının hesabı

yapılabilir. Tarama istatistiklerinin dağılımlarının bulunması için de kullanılabilir. Tarama istatistiklerinin dağılımları ile ilgili olarak geniş çapta araştırma yapılmaktadır. Bu alanda Naus (1982)' un çalışması örnek gösterilebilir (Huffer 1988).

Aşağıda verilecek olan Örnek3 kümelenmelerin olasılıklarının bulunması ile ilgilidir. Bu örnekte de yine Huffer (1988)' ın gösterimlerinden yararlanılmaktadır ve aşağıda verildiği gibidir:

A , k tane satırı ve $n-1$ tane sütunu olan bir matris olsun. Γ , A matrisine 0 sütununun eklenmesi ile elde edilen n sütunlu matris olsun ve $\Gamma = (A|0)$ ile gösterilsin. $Y = (Y_1, Y_2, \dots, Y_k)'$ ve $Y = \Gamma S$ olarak tanımlansın.

Herhangi bir $t > 0$ değeri için,

$$\{A\}_i = P\{Y_i > t, \text{ tüm } i \text{ ler için}\} \quad (3.15)$$

biçiminde gösterilsin. t ihmal edildiğinde (3.15)' deki eşitliğin sol tarafındaki ifade $\{A\}$ şeklinde yazılmaktadır. Teorem 3.2. ve sonuçları, $\{A\}$ 'nın hesabı için kullanılabilir. Aşağıdaki özellikler yardımı ile işlemler daha kolay yapılabilir:

(a) A matrisinin i nci satır elemanları j nci satır elemanlarına eşit veya daha büyükse,

i nci satır $\{A\}$ 'nın değeri değişmeksizin çıkartılabilir.

Sıra istatistikleri arasındaki farklarla oluşturulan aralıklar yer değiştirilebilir ve $A = (A|0)$ olduğu için;

(b) Sütunların yeri değiştirilebilir.

(c) Satırların yeri değiştirilebilir.

(d) 0 sütunları silinir.

Teorem 3.2.' nin uygulaması olarak Örnek1 ve Örnek2' de hesaplanmak istenen olasılıklar basit terimlerin hesaplanabilir olasılıklarının toplamı biçiminde bulunmuştur.

Ancak Örnek3' teki durum için bu terimlerin hesaplanması gerekmektedir. Bu hesaplamaları yapabilmek için ise,

$$P\left\{S_1 > t, \sum_{i=2}^{k+1} S_i > t\right\} = \sum_{j=0}^{k-1} \binom{n-1}{j} t^j (1-2t)^{n-1-j} \quad (3.16)$$

olduğunu kullanmak gerekecektir. $\{S_1 > t\}$ olayının koşullandırılması ile (3.16)'daki olasılık elde edilebilir. Ayrıca (3.16)'da verilen eşitlik, (3.6) ifadesinin bir sonucudur. Burada $(x)_+ = \max(x, 0)$ olacak biçimde negatif olmayan x değeri olarak ifade edilmektedir.

Örnek3: (Huffer 1988) Yukarıda Huffer (1988) 'ın verdiği yöntemin uygulaması için,

$$P\{S_1 + S_2 > t, S_2 + S_3 > t, S_3 + S_4 > t, S_4 + S_1 > t\} = \{A\}_t \quad (3.17)$$

olasılığı hesaplanmak istensin. Burada A matrisi,

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

olmaktadır. (S_1, S_2, S_3, S_4) , çevresi 1 birim olan çemberin üzerinde düzgün dağılıma uygun olarak konumlandırılmış dört noktanın arasındaki aralıklardır.

γ_i , A matrisinin i nci sütununu göstermek üzere ,

$$\gamma_1 = (1, 0, 0, 1)', \gamma_2 = (1, 1, 0, 0)', \gamma_3 = (0, 1, 1, 0)' \text{ ve } \gamma_4 = (0, 0, 1, 1)'$$

dir.

Teorem 3.2.' de belirtilen kořullara göre, $\gamma_1 - \gamma_2 + \gamma_3 = \gamma_4$ ve $1 - 1 + 1 = 1$ olarak ξ vektörü ve katsayılar belirlenir. Buna göre,

$$\{A\} = \{A_{1,\gamma_3}\} - \{A_{2,\gamma_4}\} + \{A_{3,\gamma_4}\}$$

dir. Bu eřitlięe bakılırsa, eřitlięin saę tarafındaki üç olayda belirtilen sütun yerine $\gamma_4 = (0,0,1,1)'$ sütun vektörü konularak iřlem yapılacaęı gösterilmektedir. Buna göre;

$$\{A\} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}.$$

Örnek3 için gereken gösterimlerin tanımlandıęı kısımda verilen (a) özellięi kullanılarak iřleme devam edilir:

$$= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

Daha sonra eřitlięe (d) özellięi uygulanır ve 0 sütunları (b) özellięinin uygulanması ile matrisin son sütunu haline getirilip silinirse,

$$= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

elde edilir. Son olarak yine (b) özellięine göre eřitlięin saę tarafındaki ilk matriste birinci sütun ile ikinci sütun yer deęiřtirilir:

$$= 2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (3.18)$$

(3.18) ifadesi kullanılarak,

$$\begin{aligned} P\{S_1 + S_2 > t, S_2 + S_3 > t, S_3 + S_4 > t, S_4 + S_1 > t\} \\ = 2P\{S_1 > t, S_2 + S_3 > t\} - P\{S_1 > t, S_2 > t\} \end{aligned} \quad (3.19)$$

olduğu görülür.

(3.19) ve (3.16) kullanılarak,

$$\begin{aligned} &= 2 \sum_{j=0}^1 \binom{n-1}{j} t^j (1-2t)_+^{n-1-j} - \binom{n-1}{0} t^0 (1-2t)_+^{n-1} \\ &= 2 \left[(1-2t)_+^{n-1} + (n-1)t(1-2t)_+^{n-2} \right] - (1-2t)_+^{n-1} \\ &= (1-2t)_+^{n-1} + 2(n-1)t(1-2t)_+^{n-2} \end{aligned}$$

şeklinde yazılır. İstenilen olasılık t ye değerler verilerek hesaplanır.

Örnek4: A matrisi,

$$\{A\} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

olsun. γ_i, A matrisinin i nci sütununu göstermek üzere ,

$$\gamma_1 = (1, 0, 0)', \gamma_2 = (1, 1, 0)', \gamma_3 = (0, 1, 1)' \text{ ve } \gamma_4 = (0, 0, 1)'$$

dir.

Teorem 3.2.' ye göre, $\gamma_1 - \gamma_2 + \gamma_3 = \gamma_4$ ve $1 - 1 + 1 = 1$ olmaktadır.

$\xi = \gamma_4 = (0, 0, 1)'$ olarak belirlenir. Buna göre,

$$\{A\} = \{A_{1,\gamma_4}\} - \{A_{2,\gamma_4}\} + \{A_{3,\gamma_4}\}$$

dir. Buradan,

$$\{A\} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Örnek3 için verilen özelliklerden (a) özelliği kullanılarak işleme devam edilir:

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Yukarıda elde edilen matrislere sırası ile (b) ve (d) özellikleri uygulandığında,

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

elde edilir. Yukarıda elde edilen ifade (3.17) formülüne göre,

$$P\{S_1 > t, S_2 + S_3 + S_4 > t\} - P\{S_1 > t, S_2 > t\} + P\{S_1 > t, S_2 + S_3 > t\}$$

olasılığına eşit olacaktır. Bu olasılığın sonucu,

$$= (1-2t)_+^{n-1} + 2(n-1)t(1-2t)_+^{n-2} + \frac{1}{2}(n-1)(n-2)t^2(1-2t)_+^{n-3}$$

dir.

Huffer (1988)' in ardışık sıra istatistikleri arasındaki farkların (Spacings) lineer kombinasyonları ile ilgili olarak yaptığı çalışmada A matrisi 0 ve 1 değerlerinden oluşan bir matris olarak ele alınmıştır. A matrisi rasyonel değerlerden oluşan bir matris olduğunda Huffer ve Lin (2001), bu çalışmayı genişleterek istenilen olasılığın hesabını yapmak için bir algoritma geliştirmiştir.

$X_1, X_2, \dots, X_n$ birbirinden bağımsız ve $(0,1)$ aralığında düzgün dağılıma sahip rasgele değişkenler ile ilgili sıra istatistikleri $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ olsun.

Ardışık sıra istatistikleri arasındaki farklar,

$$S_i = X_{(i)} - X_{(i-1)}, \quad i = 1, 2, \dots, n+1$$

olmak üzere $X_{(0)} = 0$. $X_{(n+1)} = 1$ ve $S^{(n)} = (S_1, S_2, \dots, S_{n+1})'$ olarak tanımlansın.

A , rasyonel değerler içeren herhangi bir matris ve b , rasyonel değerler içeren herhangi bir vektör ve $t > 0$ olmak üzere,

$$P\left(AS^{(n)} > tb\right) \quad (3.20)$$

olasılığının hesaplanmasında A matrisinin sütun sayısının $k < n+1$ olduğu durumda,

$$AS^{(n)} = (A|0)S^{(n)}$$

biçiminde A matrisine, k sütun sayısı $n+1$ e eşit olacak şekilde 0 sütunu veya sütunları eklenir.

(3.20)' de verilen olasılığın hesabı, $j \geq 0$ olacak biçimde pozitif tam sayı ve $\lambda > 0$ reel değerleri için,

$$R(j, \lambda) = \begin{cases} \binom{n}{j} t^j (1 - \lambda t)^{n-j} & , \quad \lambda t < 1 \\ 0 & , \quad \lambda t \geq 1 \end{cases} \quad (3.21)$$

kullanılarak elde edilir. Ancak $S^{(n)}$ vektörü yerine birbirinden bağımsız ve parametresi 1 olan üstel dağılıma sahip rasgele değişkenlerin oluşturduğu Z vektörü ile (3.21)' de verilen R fonksiyonu,

$$R(j, \lambda) = \frac{t^j e^{-\lambda t}}{j!}$$

olacaktır (Huffer ve Lin 2001).

(3.20)' de verilen olasılık,

$$Q(A, b, \lambda, p) = p! R(p, \lambda) P\left((1 - \lambda t) AS^{(n-p)} > tb\right)$$

ile gösterildiğinde ileride verilen algorithmada temel amaç,

$$Q(0, 0, \lambda, p) = p! R(p, \lambda)$$

olacak biçimde A matrisini ve b vektörünü boş küme yapacak elemanter işlemleri uygulamak ve aşağıda verilen (3.22) ile (3.23)' deki temel yinelemeleri uygulanabilir hale getirmektir. Burada boş küme, algorithmaya göre yapılan indirgemeler sonucunda

A matrisinin tek bir satır vektöründen ve b vektörünün de tek bir elemandan oluşması ile elde edilmektedir.

$A, k \times n$ lik bir matris ve $c = (c_1, c_2, \dots, c_n)'$ sabit sayıların vektörü olmak üzere,

$$\sum_{i=1}^n c_i = 1 \text{ için } \xi = Ac \text{ vektörü elde edilir. } A_{i,\xi}, A \text{ matrisinin } i \text{ nci sütunu ile } \xi$$

vektörünün yer değiştirmesi sonucu elde edilen matristir. Buna göre 3 ncü bölümde verilen Teorem 3.2. uygulandığında,

$$Q(A, b, \lambda, p) = \sum_{i=1}^n c_i Q(A_{i,\xi}, b, \lambda, p) \quad (3.22)$$

olacaktır. Yukarıdaki $A_{i,\xi}$ matrislerinin herbiri aşağıdaki özellikler sağlandığında bu matrislerin satır sayılarında azalma sağlanır.

$A = (a_{ij})$ ve $b = (b_j)$ olmak üzere, $k \geq 1$ için,

$(R_1) \ a_{ij} = 0, j > k$ için

$(R_2) \ a_{ij} = a_{j1}, j \leq k$ için

$(R_3) \ a_{11} > 0$ ve $b_1 > 0$

dir. $\delta = \frac{b_1}{a_{11}}$ olmak üzere, bu azalmanın sağlandığı durumlarda genel gösterimler,

$$Q(A, b, \lambda, p) = \sum_{i=1}^{k-1} \frac{\delta^i}{i!} Q(A_{(-i)}^*, b^* - \delta a^*, \lambda + \delta, p + i) \quad (3.23)$$

biçiminde yazılabilir (Huffer ve Lin 2001). (3.23)' de tanımlanan A^* , A matrisinin ilk satırının silinmesi ile elde edilen matris; $A_{(-i)}^*$, A^* matrisinin ilk i sütununun silinmesi ile elde edilen matris; b^* , b vektörünün ilk elemanının silinmesi ile oluşan vektör ve a^* , A matrisinin ilk sütun vektörünün birinci elemanının silinmesi ile elde edilen vektördür.

Q fonksiyonunun hesaplanması sırasında kullanılan özellikler aşağıda verilmektedir:

Herhangi bir G permütasyon matrisi için,

$$(E_1) \quad Q(A, b, \lambda, p) = Q(AG, b, \lambda, p)$$

$$(E_2) \quad Q(A, b, \lambda, p) = Q(GA, Gb, \lambda, p)$$

Köşegen elemanları pozitif olan D köşegen matrisi için,

$$(E_3) \quad Q(A, b, \lambda, p) = Q(DA, Db, \lambda, p)$$

dir. (E_1) özelliği, ardışık sıra istatistikleri arasındaki farkların yer değiştirilebilir özelliğe sahip olması ile uygulanmaktadır.

Aşağıda verilen özelliklerin herhangi biri sağlandığı zaman A matrisinin i nci satırı ve b vektörünün i nci elemanı, Q fonksiyonunun değeri değişmeksizin silinir:

$$(S_1) \quad a_{ik} \geq a_{jk}, \text{ tüm } k \text{ lar için ve } b_i \leq b_j,$$

$$(S_2) \quad a_{ik} \geq 0, \text{ tüm } k \text{ lar için ve } b_i \leq 0,$$

$$(S_3) \quad a_{ik} = 0, \text{ tüm } k \text{ lar için ve } b_i < 0,$$

$$(S_4) \quad a_{ik} + a_{jk} \leq 0, \text{ tüm } k \text{ lar için ve } b_i + b_j \geq 0,$$

$$(S_5) \quad a_{ik} \leq 0, \text{ tüm } k \text{ lar için ve } b_i \geq 0.$$

Huffer ve Lin (2001)' in ortak çalışması sonucu elde ettikleri sıra istatistikleri arasındaki farkların lineer kombinasyonlarına ilişkin algoritma aşağıdaki gibi verilebilir:

Adım 1: A matrisi B_{lm} blok matrisleri ile standart forma dönüştürülür. Standart form, B nin değeri 1 veya 1 den büyük bir tamsayı olmak üzere A matrisinin bloklara ayrılmış halidir.

Örneğin, B_{lm} blok matris ve $A, B=3$ olmak üzere,

$$A = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix}$$

biçiminde tanımlanabilir.

Adım 2: B_{lm} blok matrislerinden herhangi biri seçilir. B_{lm} blok matrisleri karesel olmak zorunda değildir. Ayrıca bu bloklar tek bir eleman da bulundurabilir. Yukarıda tanımlanan A matrisi ele alınırsa; B_{11}, B_{22}, B_{33} ana köşegen blokları hiçbir şekilde 0 elemanını içermeyecektir. Buna karşın, B_{12}, B_{13}, B_{23} bloklarının elemanlarının hepsi 0 olmalıdır.

Adım 3: Seçilen B_{lm} blok matrisinin birinci satırı A matrisine göre i nci ve j nci sütunlarda α ve β gibi birbirinden farklı iki eleman bulundursun. B_{lm} blok matrisinin ilk satırı A matrisinin r nci satırının bir parçası olmak üzere $\alpha = a_{ri}, \beta = a_{rj}$ dir.

Adım 4: $l = m$ ise, (3.22)' deki c katsayı vektörü olmak üzere

$$c_i = \frac{\beta}{\beta - \alpha}, c_j = \frac{-\alpha}{\beta - \alpha} \text{ ve } k \neq i, j \text{ için } c_k = 0 \text{ dir. Bu durumda } \sum_{i=1}^n c_i = 1 \text{ koşulu}$$

da sağlanmaktadır.

Adım 5: $l > m$ ise, $k < r$ ler için $a_{ki} = a_{kj}$ dir. A matrisinin r nci satırındaki en son, sıfır olmayan eleman s nci sütunda yer alsın ve γ ile gösterilsin. Bu durumda

(3.22)' deki c vektörü, $c_i = \frac{\gamma}{\beta - \alpha}$, $c_j = \frac{-\gamma}{\beta - \alpha}$ ve $c_s = 1$ dir. $\sum_{i=1}^n c_i = 1$ koşuluna göre $k \neq i, j, s$ için $c_k = 0$ dir.

Adım 6: $\xi = Ac$ olacak biçimde belirlenir ve $A_{i,\xi}$ matrisleri oluşturulur.

Adım 7: Elde edilen $A_{i,\xi}$ matrislerinde algoritma için verilen $(S_1), (S_2), (S_3), (S_4), (S_5), (R_1), (R_2), (R_3)$ koşulları ve $(E_1), (E_2), (E_3)$ özellikleri incelenir. $A_{i,\xi}$ matrisleri $(R_1), (R_2), (R_3)$ koşullarını sağlarsa $\delta = \frac{b_1}{a_{11}}$ elde edilir ve (3.23) uygulanabilir hale gelir.

Adım 8: $A_{i,\xi}$ matrisleri $(R_1), (R_2), (R_3)$ koşullarını sağlamazsa $A_{i,\xi}$ matrislerinden biri ele alınır ve A matrisi yerine konulur. Bu matris, Adım 1 de verildiği şekilde standart forma dönüştürülerek, matrisi boş küme haline getirecek elemanter işlemler tekrar edilir.

Örnek5: Örnek4' de verilen A matrisi Huffer ve Lin (2001)' in algoritması ile çözülmek istensin. A matrisi,

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

ele alsın. Huffer ve Lin (2001)' in algoritmasına göre ilk önce A matrisi standart forma dönüştürülür. Standart form halindeki A matrisinin köşegen blok matrisleri ve köşegenin üstünde kalan blok matrisler sırası ile,

$$B_{11} = (1, 1)', B_{22} = (1)', B_{33} = (1)'$$

ve

$$B_{12} = (0)', B_{13} = (0)', B_{23} = (0)'$$

dir. Algoritmayı uygulamak için $B_{21} = (0, 1)'$ blok matrisi ele alınsın. Algoritmanın 5 nci adımı uygulanacaktır. Sonuçta elde edilen c katsayı vektörü,

$$c' = (1, -1, 1, 0)'$$

olacaktır. $\xi = Ac$ ifadesine göre $\xi = (0, 0, 1)'$ olarak bulunur. $\xi = (0, 0, 1)'$ vektörü aşağıdaki gibi yerleştirilir:

$$A = A_{1,\xi} - A_{2,\xi} + A_{3,\xi}.$$

Elde edilen $A_{1,\xi}$, $A_{2,\xi}$, $A_{3,\xi}$ matrislerinde Huffer ve Lin (2001)' in algoritması için verilen koşullar incelenir.

$A_{1,\xi}$ matrisine (E_1) özelliği uygulandığında ,

$$A_{1,\xi} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \quad (3.24)$$

elde edilir. Bu matris algoritma için verilen $(R_1), (R_2), (R_3)$ özelliklerini sağlar. Bu durumda (3.23) formülü uygulanabilir. $\delta = b_1 / a_{11} = 1$ dir. Ayrıca (3.24)' deki matris (S_1) özelliğini sağlar. Dolayısıyla (3.24)' de verilen matris,

$$= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

biçimini alır ve (E_1) özelliğinin uygulanması ile Q fonksiyonu (3.23) formülüne göre,

$$Q((1,1,1),(1),1,0) = R(0,2) + R(1,2) + R(2,2)$$

dir. $A_{1,\varepsilon}$ matrisinde algoritmanın işleyişi benzer şekilde $A_{2,\varepsilon}, A_{3,\varepsilon}$ matrislerine de uygulanır. Sonuç olarak aranan olasılık,

$$= (1-2t)_+^{n-1} + 2(n-1)t(1-2t)_+^{n-2} + \frac{1}{2}(n-1)(n-2)t^2(1-2t)_+^{n-3}$$

olarak elde edilir.

4. UYGULAMA ÖRNEKLERİ

4.1. Bir Mod Tahmin Edicisinin Dağılımı

Bu bölümde, Huffer (1988)' in ardışık sıra istatistikleri arasındaki farkların lineer kombinasyonlarının dağılımı için kullandığı algoritma ile bir mod tahmin edicisinin dağılımının elde edilmesinin yolu gösterilmektedir.

Venter (1967), modun iki tahmin edicisini şu şekilde vermektedir.

$X_1, X_2, \dots, X_n$ rasgele değişkenleri herhangi bir dağılım fonksiyonu F ' e sahip olsun.

F sürekli bir dağılım fonksiyonu olsun. $X_1, X_2, \dots, X_n$ sıralı rasgele örnekleminin modu θ olsun.

$$V_i = X_{(i+k)} - X_{(i)} \quad , \quad i = 1, 2, \dots, n-k \quad (4.1)$$

ve

$$V_{K_n} = \min \{V_i : 1 \leq i \leq n-k\} \quad (4.2)$$

olarak tanımlansın. Mod (θ) için iki tahmin edici,

$$\theta_{1n} = \frac{1}{2} \{X_{(K_n+k)} - X_{(K_n)}\} \quad \text{ve} \quad \theta_{2n} = X_{(K_n)}$$

dir.

$F(X_{(1)})$, $F(X_{(2)})$, ..., $F(X_{(n)})$ $(0,1)$ aralığındaki düzgün dağılımdan alınmış rasgele örneklemin sıra istatistikleri olarak düşünülebilir.

$Z_1, Z_2, \dots, Z_{n+1}$ birbirinden bağımsız ve herbiri parametresi 1 olan üstel dağılıma sahip rasgele değişkenler olmak üzere,

$$U_i = Z_1 + Z_2 + \dots + Z_i, \quad i = 1, 2, \dots, n+1$$

için,

$$F(X_{(i)}) = \frac{U_i}{U_{n+1}}, \quad i = 1, 2, \dots, n$$

dir. Dolayısıyla $\frac{U_i}{U_{n+1}}, i = 1, 2, \dots, n$ $(0,1)$ aralığındaki düzgün dağılıma sahiptir.

Eğer F dağılım fonksiyonunun tersi varsa bunu $G = F^{-1}$ ile gösterilir. Bu durumda,

$$X_{(i)} = G\left(\frac{U_i}{U_{n+1}}\right), \quad i = 1, 2, \dots, n \quad (4.3)$$

olmaktadır.

$\frac{U_i}{U_{n+1}} \leq \phi_n(p) \leq \frac{U_{i+k}}{U_{n+1}}$ olacak biçimde öyle bir $\phi_n(p)$ değeri vardır ki, ortalama

değer teoremine göre,

$$\frac{G\left(\frac{U_{i+k}}{U_{n+1}}\right) - G\left(\frac{U_i}{U_{n+1}}\right)}{\frac{U_{i+k}}{U_{n+1}} - \frac{U_i}{U_{n+1}}} = G'\left(\phi_n(p)\right) \quad (4.4)$$

olmaktadır. (4.4) eşitliğinin her iki tarafı $\left(\frac{U_{i+k}}{U_{n+1}} - \frac{U_i}{U_{n+1}}\right)$ ile çarpıldığında,

$$G\left(\frac{U_{i+k}}{U_{n+1}}\right) - G\left(\frac{U_i}{U_{n+1}}\right) = \left(\frac{U_{i+k}}{U_{n+1}} - \frac{U_i}{U_{n+1}}\right) G'(\phi_n(p)) \quad (4.5)$$

elde edilir. (4.1) ve (4.3) ifadelerine göre (4.5)' de verilen eşitlik,

$$V_i = X_{(i+k)} - X_{(i)}$$

$$= \left(\frac{U_{i+k}}{U_{n+1}} - \frac{U_i}{U_{n+1}}\right) G'(\phi_n(p))$$

olarak yazılabilmektedir.

Buna göre,

$$\begin{aligned} P\left(\min(V_1, V_2, \dots, V_{n-k}) > v\right) &= P\left(\left(\frac{U_{i+k}}{U_{n+1}} - \frac{U_i}{U_{n+1}}\right) G'(\phi_n(p)) > v\right) \\ &= P\left(\frac{U_{i+k}}{U_{n+1}} - \frac{U_i}{U_{n+1}} > \frac{v}{G'(\phi_n(p))}\right) \end{aligned} \quad (4.6)$$

dir. Buradan hesaplanmak istenen olasılık,

$$P\left(\min(V_1, V_2, \dots, V_{n-k}) > v\right) = P\left(X_{(i+k)} - X_{(i)} > \frac{v}{G'(\phi_n(p))}\right)$$

dir.

Bu durumda bir mod tahmin edicisinin dağılımı için,

$$P(V_{K_n} > v) = P(\min \{V_1, V_2, \dots, V_{n-k}\} > v) \quad (4.7)$$

elde edilmek istensin.

Ardışık sıra istatistikleri arasındaki farkların (Spacings) lineer kombinasyonlarının minimumunu bulmak için 3 ncü Bölümde verilen Huffer (1988)' in ilgilendiği yöntemden yararlanılacaktır.

$n=7$, $k=3$ olarak alındığında (4.7)' deki olasılık 3 ncü Bölümde verilen Huffer (1988)' in yöntemine göre hesaplanacaktır.

Buna göre A matrisi,

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

olmaktadır. $A = (A|0)$ olduğu için 0 sütunu silinir. Buna göre,

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

dir.

Huffer (1988)' in verdiđi yöntem yukarıdaki A matrisi için uygulandığında elde edilen olasılık,

$$\begin{aligned}
 P(V_{K_n} > v) &= P(\min(V_1, V_2, V_3, V_4) > v) \\
 &= (1-2t)_+^{n-1} + 2(n-1)t(1-2t)_+^{n-2} + 4\binom{n-1}{2}t^2(1-2t)_+^{n-3} \\
 &\quad + 4\binom{n-1}{3}t^3(1-2t)_+^{n-4} + 2\binom{n-1}{4}t^4(1-2t)_+^{n-5}
 \end{aligned} \tag{4.8}$$

dir.

4.2. Genetikte Uygulama

Giriş kısmındaki Nolan ve Speed (2001)' de ilgilenilen 4 ncü problem, 3 ncü Bölümde verilen Naus (1965)' un yaklaşımı ile şu şekilde ele alınabilir:

Stomegalovirüs (CMV) ile savaşmak için bilim adamları virüsün nasıl çoğaldığına dair araştırmalar yapmaktadır. DNA' daki baz sayısı az olduğundan DNA dizileri pek çok şablon (örnek) içerir. Tamamlayıcı palindromlar bu şablon tiplerinden biridir. Tamamlayıcı palindromları oluşturan harf dizisinin (bazların) özelliđi tersten okunduđu zaman düzden okunan dizinin tamamlayıcısı olmasıdır. Örnek: GGGCATGCCC
CMV ile aynı aileden (Herpes ailesi), iki virüsün DNA eşlemeleri tamamlayıcı palindromlarla gerçekleşir.

CMV için eşleme merkezinin (replikasyon kaynağı) yeri virologlara virüslere karşı bir aşı bulmada yardımcı olabilir.

Nolan ve Speed (2000)' de aktarılan CMV' nin DNA' sı üzerinde yaptıđı araştırmada çoğalma bilgisini taşıyan 296 palindrom elde edilmiştir. CMV' nin palindromları en az

10, en fazla 18 baz çiftinden oluşmaktadır. Bu bilgiyi elde edebilmek için DNA parçalara bölünerek her bir parçanın çoğalması incelenmiştir. Bu yöntemden yola çıkarak aranan özelliği kaybetmeyecek şekilde aralık belirlenmeye çalışılmıştır. 250, 500 ve 1000 baz çifti uzunluğundaki aralıkta maksimum palindrom sayılarının olasılık dağılımı aşağıdaki şekilde elde edilmiştir.

Çizelge 4.1. Aralık uzunluğuna bağlı olarak maksimum palindrom sayıları

Aralık Uzunluğu	Maksimum palindrom sayıları							
	4	5	6	7	8	9	10	11
250	0.67	0.08	0.006	0.0003	-	-	-	-
500	-	0.61	0.12	0.12	0.001	0.0001	-	-
1000	-	-	-	0.34	0.07	0.012	0.002	0.0002

Çizelge 4.1.' de, 500 baz çiftini içeren aralıkta DNA' da en az 6 palindrom bulunması olasılığı 0.12 olarak belirlenmiştir. 250, 500 ve 1000 baz çifti içeren uzunluktaki aralıklar için olasılıklar yukarıdaki tabloda verilmiştir. Nolan ve Speed (2000), olasılık hesabı için Naus (1965)' un yaklaşımından yararlanmıştır.

4.3. Diğer Uygulamalar

Giriş kısmında yer alan 1 nci problem Naus (1965)' un yaklaşımı ile şöyle çözülebilir:

1 dakikalık süre içinde 15 tane telefonun santrali arama zamanları rasgeledir. Arama yapan bir telefonun santrali meşgul etme süresi 10 saniye olarak belirlenmiştir.

1 dakikalık süre içinde 15 telefondan 8 veya daha fazlasının aynı anda santrali arama olasılığı nedir?

Burada $i=1,2,...,8$ iken i nci arama başladıktan sonra 10 saniyeden daha az bir sürede $(7+i)$ nci arama gerçekleşmektedir. Bu durumda elde edilen olasılık Naus (1965)' un yaklaşımı ile;

$$P(8|15;1/6) = 0.037$$

dir. Dikkat edilirse $(0,1)$ aralığı 10 dilimlik parçalara ayrılmamaktadır. Arama işlemi 60 saniyelik süre içinde herhangi bir saniyede başlamaktadır. Dolayısıyla belirlenen 10 saniyelik süre 60 saniyelik süre olarak gösterilen $(0,1)$ aralığının herhangi bir kısmını göstermektedir.

Giriş kısmında verilen 2 nci problemin Naus (1965) yaklaşımı ile çözümü şu şekildedir:

Probleme göre sayaca alınan sinyaller λ oranlı Poisson sürecini oluşturmaktadır. t den daha küçük uzunluklu bir aralıkta sayaç n tane sinyali kaydetmektedir. Sayacın ilk kaydına kadar gerçekleşen bekleme sürelerinin dağılımı bulunmak isteniyor. Bunun için sayacın ilk kaydına kadarki bekleme süresi $\leq T$ olmak üzere $F_n(T, t | \lambda)$ olasılığının hesaplanması gerekmektedir. $N, (0, T)$ aralığındaki noktaların toplam sayısı olsun. Bu durumda $N, \lambda T$ ortalamalı Poisson dağılımına sahiptir. $F_n(T, t | \lambda; N)$, N 'nin bilinmesi koşuluna göre $F_n(T, t | \lambda)$ olasılığı olsun. $F_n(T, t | \lambda; N)$, T 'nin t uzunluklu alt aralığında N tane sinyalin n veya n den daha fazlası ile oluşan olasılıktır. N 'nin bilinmesi ile koşullanan Poisson dağılımı düzgün dağılıma denktir. Yani, $F_n(T, t | \lambda; N)$ olasılığı $P(n | N; t | T)$ dir. N 'nin dağılımı üzerinden bu koşullu olasılıkların toplamı,

$$F_n(T, t | \lambda) = \sum_{N=n}^{\infty} P(n | N; t | T) \frac{e^{-\lambda T} (\lambda T)^N}{N!}$$

olarak ifade edilir.

3 ncü Bölümde verilen ve Giriş kısmında 5 nci problem olarak tanımlanan Seth (1950) ve Lieblein (1952)' in $n=3$ çaplı örneklemlerden en uçta yer aldığı varsayılan örneklem bir sıradışı (outlier) değer olması olasılığı ile ilgili problemi diğer bir uygulama örneğidir. Seth (1950) ve Lieblein (1952) ilgilendikleri örneklemde sıradışı (outlier) değeri ararken örneklem çapını 3 olarak belirlemiş ve diğer uygulama örneklerinden farklı olarak örneklem alındığı yığının Normal dağılımdan alındığını varsaymıştır. Örneklem çapını küçük tutarak istenilen sıradışı değere ulaşmaya çalışmışlardır.

3 ncü Bölümde Örnek3' te verilen kümelenmelerin olasılıkları, çember üzerinde dağılan noktaların düzgün dağılımlı olup olmadığı testinde kullanılabilir. Bu tür bir uygulama; mevsimlere göre rüzgarların hareket yönlerinin düzgün dağılımlı olup olmadığının veya hastanenin acil servisine 24 saat içinde gelen hastaların gelişlerinin düzgün dağılımlı olup olmadığının testi olabilir.

5. TARTIŞMA ve SONUÇ

Bu çalışmada, n tane rasgele değişkenin $(0,1)$ aralığında düzgün dağılıma sahip bir yığından alındığında, belirlenen λ uzunluğundaki bir aralık boyunca değer alan gözlem sayıları ve bunların maksimumlarının dağılımının konu edilen farklı yöntemler yardımı ile nasıl elde edileceğinin gösterilmesi temel amaçtır.

Üçüncü bölümde verilen problemler ve bu problemlerin çözümlerine ilişkin üç farklı yaklaşım yönteminin uygulanması gösterilmektedir. Ayrıca bu bölümde ilgilenilen problem için Huffer (1988) tarafından verilen kesin çözüm yöntemi bulunmaktadır. Huffer ve Lin (2001)' in Üçüncü Bölümde gösterilen algoritması, Huffer (1988) tarafından verilen yöntemin daha geliştirilmiş biçimidir. Bu algoritmada Huffer (1988)' in yönteminde kullandığı matrisin rasyonel değerler içermesi durumu söz konusudur.

Huffer (1988)' in ardışık sıra istatistikleri arasındaki farkların (Spacings) lineer kombinasyonunun dağılımı için vermiş olduğu yöntem Dördüncü Bölümde bir mod tahmin edicisinin dağılımı için uygulanmıştır. Yine Dördüncü Bölümde Nolan ve Speed (2001)' in tanımlanan problemin genetikte uygulaması ile ilgili çalışması bulunmaktadır. Genetik üzerine yapılan çalışmada Naus (1965)' un Üçüncü Bölümde anlatılan çözüm yönteminden yararlanılmıştır.

Üçüncü Bölümde Seth (1950) ve Lieblein (1952)' in çapı 3 olan bir örnekleme sıradışı (outlier) değerın bulunması ile ilgili çalışması yer almıştır. Seth (1950) ve Lieblein (1952) bu çalışmasında problemin çözümü için verilen diğer yöntemlerden farklı olarak örneklemin alındığı yığının Normal dağılıma sahip olduğunu varsaymıştır. Ayrıca Seth (1950) ve Lieblein (1952) örnekleme çapını 3 alarak istenilen sıradışı değere ulaşmıştır. Çok büyük örneklem çaplarında sıradışı değere ulaşmak zordur. Bu nedenle Seth (1950) ve Lieblein (1952) problemi oldukça küçük örneklem çapı ile sonuca ulaştırabilmiştir.

KAYNAKLAR

- Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. 1992. A first course in order statistics. John Wiley & Sons, INC., 279 p., New York.
- David, H.A. 1969. Order statistics. John Wiley & Sons, INC., 272 p., New York.
- Glaz, J. and Naus, J. 1983. Multiple clusters on the line. Communication Statistics Theory and Methods, 12(17); 1961-1986.
- Huffer, F. 1988. Divided differences and the joint distribution of linear combinations of spacings. Journal of Applied Probability, 25; 346-354.
- Huffer, F.W. and Lin, C.T. 1997. Computing the exact distribution of the extremes of sums of consecutive spacings. Computational Statistics and Data Analysis, 26; 117-132.
- Huffer, F.W. and Lin, C.T. 2001. Computing the joint distribution of general linear combinations of spacings or exponential variates. Statistica Sinica, 11; 1141-1157.
- Jankovic, V. 2000. Divided differences. The Teaching of Mathematics, Vol. III, No. 2; 115-119.
- Lieblein, J. 1952. Properties of certain statistics involving the closest pair in a sample of three observations. Journal of Research of the National Bureau of Standards, Vol. 48, No. 3; 255-268.
- Mack, M.A. 1948. An exact formula for $Q_k(n)$, the probable number of *k-aggregates* in a random distribution of *n* points. The London, Edinburg, and Dublin Philosophical Magazine and Journal of Science, Vol. 39, No. 297; 778-790.
- Miller, E.G. 2003. A new class of entropy estimators for multi-dimensional densities. International Conference on Acoustics, Speech, and Signal Processing. Berkeley.
- Milne-Thompson, L.M. 1933. The calculus of finite differences. Macmillan, 558 p., London.
- Naus, J.I. 1965. The distribution of the size of the maximum cluster of points on a line", Journal of American Statistical Association, 60; 532-538.

- Naus, J.I. 1966. Some probabilities, expectations and variances for the size of largest clusters and smallest intervals. *Journal of American Statistical Association*, 61; 1191-1199.
- Nolan, D. and Speed, T. 2000. *Stat Labs (Mathematical Statistics Through Applications)*, Springer-Verlag New York, Inc., 282 p.
- Port, S.C. 1994. *Theoretical probability for applications*. John Wiley & Sons, INC., 894 p., New York.
- Pyke, R. 1965. Spacings. *Journal of the Royal Statistical Society, Series B*, 27; 395-436.
- Seth, G.R. 1950. On the distribution of the two closest among a set of three observations. *Annals of Mathematical Statistics*, 21; 298-301.
- Sukhatme, P.V. 1937. Tests of significance for samples of the χ^2 population with two degrees of freedom . *Annals of Eugenics*, 8; 52-56.
- Uyan, B. 1980. Çözümlü problemlerde diferansiyel denklemler, Fourier serileri, Laplace transformasyonları. Birsen Kitabevi, 373 s., İstanbul.
- Venter, J.H. 1967. On estimation of the mode. *Annals of Mathematical Statistics*, 38; 1446-1455.

ÖZGEÇMİŞ

1978 yılında İzmir’ de doğdu. İlk öğrenimini İzmir’ de, orta ve lise öğrenimini Ankara’ da tamamladı. 1997 yılında girdiği Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü’ nden 2001 yılında İstatistikçi ünvanıyla mezun oldu. Eylül 2001’ de Ankara Üniversitesi Fen Bilimleri Enstitüsü, İstatistik Ana Bilim Dalı’ nda Yüksek Lisans öğrenimine başladı.

2002 Eylül döneminden beri Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü’ nde Araştırma Görevlisi olarak görev yapmaktadır.