

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

KESİKLİ DAĞILIMLARDAKİ SIRA İSTATİSTİKLERİ ÜZERİNE

Fahrettin ÖZBEY

Tez Yöneticisi:

Yrd. Doç. Dr. Sinan ÇALIK

YÜKSEK LİSANS TEZİ
İSTATİSTİK ANABİLİM DALI

ELAZIĞ, 2007

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

KESİKLİ DAĞILIMLARDAKİ SIRA İSTATİSTİKLERİ ÜZERİNE

Fahrettin ÖZBEY

Yüksek Lisans Tezi
İstatistik Anabilim Dalı

Bu tez, 20 / 09 / 2007 tarihinde aşağıda belirtilen jüri tarafından oybirliği /oyçokluğu ile başarılı / başarısız olarak değerlendirilmiştir.

Danışman:

Üye:

Üye:

Üye:

Üye:

Bu tezin kabulü, Fen Bilimleri Enstitüsü Yönetim Kurulu'nun/...../..... tarih ve sayılı kararıyla onaylanmıştır.

TEŐEKKÖR

Bu alıřmanın planlanmasında ve yürütölmesinde alıřmalarım süresince benden destek ve ilgilerini esirgemeyen bilgi ve hoşgörülerinden yararlandıđım deđerli hocam sayın Yrd. Do. Dr. Sinan ALIK' a en içten teşekkür ve saygılarımı sunarım.

İÇİNDEKİLER

İçindekiler	I
Kısaltmalar	II
Özet	III
Abstract	IV
Giriş	1
1. Bölüm	
1.1. Temel Tanım Ve Teoremler.....	2
2. Bölüm	
2.1. Sıra İstatistikleri Ve Dağılımları.....	7
2.2. Kesikli Dağılımlarda Sıra İstatistikleri.....	10
2.3. Kesikli Dağılımlarda Sıra İstatistiklerinin Dağılımları.....	11
2.4. Bileşik Olasılık Kütle Fonksiyonu.....	13
3. Bölüm	
3.1. Momentler Ve Çarpım Momentleri.....	15
3.2. Kesikli Düzgün Dağılımdaki Sıra İstatistikleri.....	17
Tartışma Ve Sonuç	22
Kaynaklar.....	23
Özgeçmiş.....	24

KISALTMALAR

cdf: Kümülatif dağılım fonksiyonu

pdf: Olasılık yoğunluk fonksiyonu

pmf: Olasılık kütle fonksiyonu

d
= : Dağılım olarak aynı anlamındadır.

ÖZET

Yüksek Lisans Tezi

KESİKLİ DAĞILIMLARDAKİ SIRA İSTATİSTİKLERİ ÜZERİNE

Fahrettin ÖZBEY

Fırat Üniversitesi

Fen Bilimleri Enstitüsü

İstatistik Anabilim Dalı

Danışman: Yrd. Doç. Dr. Sinan ÇALIK

2007, Sayfa: 22

Bu çalışma dört bölümden oluşmaktadır.

Birinci bölümde, bazı temel kavramlar, teoremler verilmiştir.

İkinci bölümde, sıra istatistiklerin dağılımları için daha önce yapılan çalışmalar verilmiştir.

Üçüncü bölümde, kesikli dağılımlardaki sıra istatistiklerin m . ci momenti elde edilmiştir. Ayrıca kesikli düzgün dağılımdaki sıra istatistiklerin momentleri incelenmiştir.

Dördüncü bölümde ise, bu çalışmada elde edilen sonuçlar ve yapılabilecek çalışmalar tartışılmıştır.

Anahtar Kelimeler: Sıra istatistikleri, momentler, kesikli dağılımlar

ABSTRACT

Master Thesis

ON THE ORDER STATISTICS FORM DISCRETE DISTRIBUTIONS

Fahrettin ÖZBEY

Firat University

Graduate School of Science and Technology

Department of Statistics

Supervisor: Yrd. Doç. Dr. Sinan ÇALIK

2007, Page: 22

There are four chapter of this study. In the first chapter, some basic concepts, theorems are given.

In the second chapter, previous studies for distribution based on order statistics.

In the third chapter, the m th moments of order statistics from discrete distributions are obtained. In addition, the moments of order statistics from a discrete uniform distribution are investigated.

In the final chapter, a discussion of result are summarized and some further research topics are proposed

Keywords: Order statistics, moments, discrete distributions

GİRİŞ

Sıra istatistikleri istatistik teorisinin en önemli kavramlarından biri olup, temel istatistik yöntemlerde ve istatistiksel sonuç çıkarımında kullanılmaktadır. Aynı zamanda, teste tabi tutulan n tane ürünün yaşam zamanlarını gösterdiği için, yaşam analizi (life-time analysis) ve güvenilirlik teorisinde (reliability theory) önemli yer tutmaktadır. Sıra istatistikleri, yeterli istatistikler olduklarından örneklem hakkındaki tüm bilgiyi içerirler. Sıra istatistiklerine son yıllarda artan ilgi, bu alanın hızla gelişimini sağlamıştır. Özellikle, sıra istatistiklerine dayalı birçok istatistik, dağılımdan bağımsız (distribution-free) özelliği taşıdığı için, parametrik olmayan istatistiksel yöntemlerde geniş şekilde kullanılmaktadır.

Her bir tesadüfî değişken dağılım fonksiyonu ile karakterize edilir. Fakat pratikte dağılım fonksiyonu çoğu zaman belli olmaz. Matematiksel istatistiğin temel problemlerinden birisi tesadüfî değişkenin deneysel değerlerini kullanarak onun dağılım fonksiyonunu tahmin etmektir. Bazen dağılım fonksiyonunun analitik ifadesi belli olur, ama bu ifade birkaç bilinmeyen parametre içerir. Mesela, poisson ve üstel dağılımlar tek parametrelidir, normal ve gamma dağılımları ise iki parametrelidir. Bu parametreler moment denilen sayısal göstergeler ile ifade edilmektedir. Bilinmeyen dağılım fonksiyonu belli koşullar altında momentler yardımıyla tek olarak belirlenebilir. Ayrıca olasılık ve stokastik süreçler için önem taşıyan bazı eşitsizlikler momentler yardımıyla ifade edilmektedir.

Literatürde konumuz ile ilgili çalışmalar; kesikli dağılımlardaki sıralı istatistiklerin ilk iki momentini Khatri (1962) elde etmiştir. Balakrishnan (1986) keyfi sürekli dağılımlardaki n boyutlu bir örnekteki sıralı istatistiklerini tekli ve çarpım momentleri için mevcut olan birkaç rekürans ilişki ve eşitsizliği kesikli durumlar için vermiştir. Nagaraja (1992)' deki makalesinde kesikli sıralı istatistiklerindeki tüm gelişmeleri açık bir şekilde tartışmıştır. Arnold vd. (1992); Khatri'nin elde etmiş olduğu ilk iki momentini farklı bir yolla elde etmişlerdir. Çalık ve Güngör (2005); kesikli dağılımlardaki sıralı istatistiklerin m . ci momentini ispat etmişlerdir. Çalık ve Güngör (2004)'de bir uygulama olarak kesikli düzgün dağılımdaki sıralı istatistiklerin örnek maksimumunun beklenen değerlerini örnek boyutu $n = 15$ kadar cebirsel olarak elde etmişlerdir.

Bu çalışmanın amacı kesikli dağılımlardaki sıra istatistiklerinin m . ci momentlerini elde etmek ve bir uygulama olarak bu momentleri kullanarak kesikli düzgün dağılımdaki sıra istatistiklerin momentlerini bulmaktır.

1. BÖLÜM

1.1. TEMEL TANIM VE TEOREMLER

Tanım 1.1.1. (İhtimal)

Bir olayın meydana gelmesindeki belirsizliğin ölçülmesinde kullanılan ve belirli kurallara göre hesaplanan, değeri 0 (sıfır) ile 1 (bir) arasında değişen bir sayıdır [17].

Tanım 1.1.2 (İhtimal Dağılımı)

Tesadüfi bir değişkenin bir değişim aralığındaki belirli değerleri alabilmesi ihtimallerini, o değişkenin bir fonksiyonu olarak ifade edilmesinde kullanılan bir formül, tablo veya şekildir [17].

Tanım 1.1.3. (Kümülatif Dağılım Fonksiyonu)

X ihtimal fonksiyonu P olan bir örnek uzayı S ' de tanımlanmış bir tesadüfi değişken olsun. Herhangi bir gerçel değer x için, X ' in kümülâtif dağılım fonksiyonu gerçel sayılar kümesinde x ' e eşit veya ondan küçük değerleri alan ve S örnek uzayında X tesadüfi değişkeni ile ilişkili olan ihtimaldir. Bu fonksiyon,

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

eşitliğiyle ifade edilir [3].

Tanım 1.1.4. (Olasılık Yoğunluk Fonksiyonu)

a ve b gibi herhangi iki gerçel değerli sabit için $a \leq b$ iken,

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

koşulu sağlanıyorsa, değerleri $f(x)$ olan ve bütün gerçel sayılar için tanımlanmış bir fonksiyona, X sürekli rassal değişkeninin olasılık yoğunluk fonksiyonu denir [10].

Tanım 1.1.5. (Düzgün Dağılım)

X tesadüfi değişkeninin ihtimal fonksiyonu,

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

şeklinde tanımlanmış olsun.

X tesadüfî değişkenine, düzgün dağılmış tesadüfî değişken ve $f(x)$ ' e sürekli düzgün dağılım denir. Düzgün dağılımın a ' dan x 'e kadar olan integrali alınır, kümülâtif düzgün dağılım fonksiyonu elde edilmiş olur ve

$$F(x) = \begin{cases} 0 & , x < a \\ \frac{x-a}{b-a} & , a \leq x \leq b \end{cases}$$

şeklinde ifade edilir [3].

Tanım 1.1.6. (Bernoulli Dağılımı)

Tesadüfî deneylerde her deneme birbirinden bağımsız ve her denemedeki muhtemel sonuç sayısı iki ise, p parametrelî

$$P(x) = p^x (1-p)^{1-x} \quad , x = 0, 1$$

yoğunluk fonksiyonuna sahip ihtimal dağılımına, Bernoulli ihtimal dağılımı denir [17].

Tanım 1.1.7. (Binom İhtimal Dağılımı)

Tesadüfî deneylerde her deneme birbirinden bağımsız ve her denemedeki muhtemel sonuç sayısı iki ise, p parametrelî

$$P_b(x/n, p) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

ihtimal yoğunluk fonksiyonuna sahip ihtimal dağılımına, binom ihtimal dağılımı denir [3].

Tanım 1.1.8. (Multinom İhtimal Dağılımı)

Tesadüfî deneylerde her deneme birbirinden bağımsız ve her denemedeki muhtemel sonuç sayısı ikiden fazla ise, $P_1, P_2, \dots, P_k$ parametrelî

$$P_m(x_1, x_2, \dots, x_k; p_1, p_2, \dots, p_k; n) = \binom{n}{x_1, x_2, \dots, x_k} p_1^{x_1} \cdot p_2^{x_2} \cdot \dots \cdot p_k^{x_k}$$

ihtimal yoğunluk fonksiyonuna sahip ihtimal dağılımına, multinom ihtimal dağılımı denir [18].

Tanım 1.1.9. (Normal Dağılım)

X tesadüfî değişkeni μ ortalama ve σ standart sapma parametrelerine bağlı

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

ihtimal yoğunluk fonksiyonuna sahip ihtimal dağılımına, normal ihtimal dağılımı denir [17].

Tanım 1.1.10. (Beta Fonksiyonu)

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx$$
$$= \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)}, \alpha > 0, \beta > 0$$

olarak tanımlanmış fonksiyona beta fonksiyonu denir [12].

Tanım 1.1.11 (Stokastik Bağımsızlık)

$f_1(x_1)$ ve $f_2(x_2)$ sırasıyla X_1 ve X_2 'nin marjinal yoğunluk fonksiyonları ve f de X_1 ve X_2 'nin bileşik yoğunluk fonksiyonu olmak üzere, X_1 ve X_2 tesadüfî değişkenlerinin stokastik olarak bağımsız olması için gerek ve yeter şart $f(x_1, x_2) = f_1(x_1)f_2(x_2)$ olmasıdır [13].

Tanım 1.1.12. (Kesikli Düzgün Dağılım)

Rassal bir X değişkeninin, ancak ve ancak olasılık dağılımı aşağıdaki gibiye kesikli düzgün dağılım vardır ve adı kesikli tekdüze rassal değişken olur.

$$f(x) = \frac{1}{k} \quad x = x_1, x_2, \dots, x_k \text{ için}$$

burada $i \neq j$ iken $x_i \neq x_j$ olur [10].

Tanım 1.1.13. (Rassal Değişken)

S , bir olasılık ölçüsü olan bir örneklem uzayı ise ve X 'de S 'nin bütün öğeleri için tanımlanmış gerçek değerli bir fonksiyonsa X 'e rassal değişken denir [10].

Tanım 1.1.14 (Kesikli Rassal Değişken)

Aralığı sonlu ya da sayılabilir sonsuz olan değişkenlere kesikli rassal değişken denir [10].

Tanım 1.1.15. (Olasılık Fonksiyonu)

X kesikli bir rassal değişkense, X 'in aralığı içindeki her bir x değeri için

$$f(x) = P(X = x)$$

ile verilen fonksiyona X 'in olasılık fonksiyonu denir [10].

Tanım 1.1.16. (Ortak Olasılık Dağılımı)

X ile Y kesikli rassal değişkenler ise, X ile Y 'nin aralıklardaki her (x, y) değer çifti için

$$f(x, y) = P(X = x, Y = y)$$

ile gösterilen fonksiyona X ile Y 'nin ortak olasılık dağılımı denir [10].

Tanım 1.1.17. (Gama Fonksiyonu)

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy \quad \alpha > 0$$

olarak tanımlanmış fonksiyona gama fonksiyonu denir [10].

Tanım 1.1.18. (Tam Olmayan Beta Fonksiyonu)

$$I_p(a, b) = \int_0^p x^{a-1} (1-x)^{b-1} dx \quad a > 0, b > 0, 0 < p < 1$$

olarak tanımlanmış fonksiyona tam olmayan beta fonksiyonu denir.

Teorem 1.1.1.

Eğer X ihtimal fonksiyonu $f(x)$ ve kümülâtif dağılım fonksiyonu $F(x)$ 'e sahip sürekli bir tesadüfî değişken ise $F(x)$ 'in tanımlı olduğu bütün noktalarda

$$f(x) = F'(x)$$

eşitliği vardır [13].

Teorem 1.1.2.

$X_1, X_2, \dots, X_n$ sürekli F dağılım fonksiyonuna sahip bir örneklem olmak üzere $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ bu örneklemün sıra istatistiklerini gösterebilir.

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

olmak üzere $U_{(1)}, U_{(2)}, \dots, U_{(n)}$ $(0, 1)$ aralığındaki düzgün dağılımdan alınmış örneklemün sıra istatistikleridir [9].

Teorem 1.1.3.

Bir rassal değişken olan X 'in aralığı $x_1 < x_2 < \dots < x_n$ değerlerinden oluşuyorsa,

$$f(x_i) = F(x_i) - F(x_{i-1}) \quad \text{ve } i = 2, 3, \dots, n \text{ için}$$

$$f(x_i) = F(x_i) - F(x_{i-1})$$

olur [10].

Teorem 1.1.4.

S , dağılımı destekleyen negatif olmayan tamsayıların bir alt kümesi olsun. O zaman sağ taraftaki momentler mevcut olmak üzere

$$\mu_{r,n}^{(1)} = \sum_{x=0}^{\infty} [1 - F_r(x)] \quad (1.1.1)$$

$$\mu_{r,n}^{(2)} = 2 \sum_{x=0}^{\infty} x [1 - F_r(x)] + \mu_{r,n} \quad (1.1.2)$$

dır. [15].

2. BÖLÜM

2.1. SIRA İSTATİSTİKLERİ VE DAĞILIMLARI

$X_1, X_2, \dots, X_n$ örneklemini artan sırada $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ biçiminde sıralansın. Bu sıralanmış $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ tesadüfî değişkenlerine sıra istatistikleri denir. Burada $X_{(i)}$ n boyutlu örneklemin i . inci sıra istatistiği olarak adlandırılır. *Birinci* sıra istatistiği $X_{(1)} = \min(X_1, X_2, \dots, X_n)$ ve n . inci sıra istatistiği $X_{(n)} = \max(X_1, X_2, \dots, X_n)$ şeklindedir.

$X_1, X_2, \dots, X_n$ tesadüfî değişkenleri olasılık yoğunluk fonksiyonu $f(x)$ ve kümülâtif dağılım fonksiyonu $F(x)$ mutlak sürekli bir anakütleden tesadüfî bir örneklem olsun. $1 \leq i \leq n$ olmak üzere $X_{(i)}$ 'in (i . inci sıra istatistiğinin) olasılık fonksiyonunu

$$f_i(x) = \frac{n!}{(i-1)!(n-i)!} [F(x)]^{i-1} [1-F(x)]^{n-i} f(x), \quad -\infty < x < \infty \quad (2.1.1)$$

olarak çıkarabiliriz.

$1 \leq i \leq j \leq n$, $-\infty < x < y < \infty$ olmak üzere $X_{(i)}$ ve $X_{(j)}$ sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonu

$$f_{i,j}(x, y) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} [F(x)]^{i-1} [F(y)-F(x)]^{j-i-1} \cdot [1-F(y)] f(x) f(y) \quad (2.1.2)$$

biçimindedir.

Sıra istatistiklerinin birleşik dağılımları benzer olarak çıkarılabilirse de doğal olarak çok daha karmaşıktır. Örneğin: $X_{(i)}$ ve $X_{(j)}$ ($1 \leq i \leq j \leq n$) 'lerin bileşik cdf' si

$$F_{i,j}(x) = F_j(x_j), \quad x_i \geq x_j \quad \text{için} \\ = \sum_{s=j}^n \sum_{r=i}^s \frac{n!}{r!(s-r)!(n-s)!} [F(x_i)]^r [F(x_j)-F(x_i)]^{s-r} [1-F(x_j)]^{n-s}, \quad x_i \geq x_j \quad \text{için} \quad (2.1.3)$$

olduğu gösterilebilir. Bu ifade ister sürekli ister kesikli herhangi bir anakütle için sağlanır.

$X_{(i_1)} \leq X_{(i_2)} \leq \dots \leq X_{(i_k)}, \quad 1 \leq i_1 < i_2 < \dots < i_k \leq n, \quad (1 \leq k \leq n)$ olmak üzere $k \leq n$ sayıda sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonu

$$f_{i_1, i_2, \dots, i_k}(x_1, x_2, \dots, x_k) = \frac{n!}{(i_1-1)!(i_2-i_1-1)! \dots (n-i_k)!} [F(x_1)]^{i_1-1} [F(x_2)-F(x_1)]^{i_2-i_1-1} \dots [1-F(x_k)]^{n-i_k} f(x_1) f(x_2) \dots f(x_k), \quad x_1 \leq x_2 \leq \dots \leq x_k \quad (2.1.4)$$

olarak bulunur. Bu eşitlikte

$$\begin{aligned}x_{k+1} &= \infty, & x_0 &= -\infty \\i_{k+1} &= n, & i_0 &= 0\end{aligned}$$

şartlarının sağlanması durumunda $f_{i_1, i_2, \dots, i_k}(x_1, x_2, \dots, x_k)$ ortak olasılık yoğunluk fonksiyonu

$$f_{i_1, i_2, \dots, i_k}(x_1, x_2, \dots, x_k) = \prod_{r=1}^k f(x_r) \prod_{r=0}^k \frac{[F(x_{r+1}) - F(x_r)]^{i_{r+1} - i_r - 1}}{(i_{r+1} - i_r - 1)} \quad (2.1.5)$$

olarak yazılabilir. Buna göre n tane $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonu

$$f_{1,2,\dots,n}(x_1, x_2, \dots, x_n) = \begin{cases} n! f(x_1) f(x_2) \dots f(x_n) & , \quad x_1 \leq x_2 \leq \dots \leq x_n \\ 0 & , \quad \text{diğer yerlerde} \end{cases} \quad (2.1.6)$$

biçimindedir [9].

Şimdi denklem (2.1.6)' da tüm n sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonunu göz önüne alınarak $(X_{(1)}, X_{(2)}, \dots, X_{(i-1)})$ ve $(X_{(i+1)}, \dots, X_{(n)})$ değişkenleri üzerinden integral alınırsa $(1 \leq i \leq n)$ $X_{(i)}$ ' in marjinal yoğunluk fonksiyonunu

$$f_i(x) = n! f(x) \left[\int_{-\infty}^x \dots \int_{-\infty}^{x_1} f(x_1) \dots f(x_{i-1}) dx_1 \dots dx_{i-1} \right] \left[\int_x^{\infty} \dots \int_x^{x_{i+2}} f(x_{i+1}) \dots f(x_n) dx_{i+1} \dots dx_n \right] \quad (2.1.7)$$

olarak çıkarabiliriz. Direkt integral alınarak

$$\int_{-\infty}^x \dots \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(x_1) f(x_2) \dots f(x_{i-1}) dx_1 dx_2 \dots dx_{i-1} = \frac{[F(x)]^{i-1}}{(i-1)!} \quad (2.1.8)$$

ve

$$\int_x^{\infty} \dots \int_x^{x_{i+3}} \int_x^{x_{i+2}} f(x_{i+1}) f(x_{i+2}) \dots f(x_n) dx_{i+1} dx_{i+2} \dots dx_n = \frac{[1 - F(x)]^{n-i}}{(n-i)!} \quad (2.1.9)$$

elde edilir. (2.1.8) ve (2.1.9)' denklemleri (2.1.7)' denkleminde yerine yazılırsa $X_{(i)}$ ' in $(1 \leq i \leq n)$ olasılık yoğunluk fonksiyonu tam olarak (2.1.1) çıkarıldığı gibi aynı ifade elde edilir.

(2.1.1)' den $(i=1$ ve $i=n$ alındığında) en küçük ve en büyük sıra istatistiklerin olasılık yoğunluk fonksiyonlarını sırasıyla

$$f_1(x) = n[1 - F(x)]^{n-1}, \quad -\infty < x < \infty \quad (2.1.10)$$

ve

$$f_n(x) = n[F(x)]^{n-1} f(x), \quad -\infty < x < \infty \quad (2.1.11)$$

yazarız.

En küçük ve en büyük sıra istatistiklerin dağılım fonksiyonlarını kolayca (2.1.10) ve (2.1.11) deki olasılık yoğunluk fonksiyonlarının integralini alarak

$$F_1(x) = 1 - [1 - F(x)]^n, \quad -\infty < x < \infty \quad (2.1.12)$$

ve

$$F_n(x) = [F(x)]^n, \quad -\infty < x < \infty \quad (2.1.13)$$

çıkarırız.

Genel olarak (2.1.1)' de $X_{(i)}$ ' in olasılık yoğunluk fonksiyonunun integrali alınarak $X_{(i)}$ ' in dağılım fonksiyonu elde edilebilir. $X_{(i)}$ 'in dağılım fonksiyonunu her zaman gerçekte zorluk olmaksızın

$$\begin{aligned} F_i(x) &= (X_{(i)} \leq x) \\ &= \Pr(X_1, X_2, \dots, X_n \text{ 'lerin en az } i \text{ tanesi } x \text{ 'e eşit ve ondan küçüktür}) \\ &= \sum_{r=i}^n \Pr(X_1, X_2, \dots, X_n \text{ 'lerin tam olarak } r \text{ tanesi } x \text{ 'e eşit ve ondan küçüktür}) \\ &= \sum_{r=i}^n \binom{n}{r} [F(x)]^r [1 - F(x)]^{n-r}, \quad -\infty < x < \infty \end{aligned} \quad (2.1.14)$$

çıkarabiliriz. Böylece $X_{(i)}$ ' in $(1 \leq i \leq n)$ ' in kümülâtif dağılım fonksiyonunu basitçe n denemede başarılı olanların $F(x)$ ' li binom dağılımı (i ' den başlayarak) olasılığı olduğunu buluruz.

Biz (2.1.14)' denkleminde $i=1$ ve $i=n$ aldığımızda sırasıyla (2.1.12) ve (2.1.13)' deki ifadeleri elde edebiliriz. Bundan başka,

$$\sum_{r=i}^n \binom{n}{r} p^r (1-p)^{n-r} = \int_0^p \frac{n!}{(i-1)!(n-i)!} t^{i-1} (1-t)^{n-i} dt, \quad 0 < p < 1 \quad (2.1.15)$$

eşitliğini kullanarak biz (2.1.14)' dekine benzer olarak $X_{(i)}$ ' in kümülâtif dağılım fonksiyonunu

$$\begin{aligned} F_i(x) &= \int_0^{F(x)} \frac{n!}{(i-1)!(n-i)!} t^{i-1} (1-t)^{n-i} dt \\ &= I_{F(x)}(i, n-i+1), \quad -\infty < x < \infty \end{aligned} \quad (2.1.16)$$

olarak yazabiliriz. Burada I incomplete beta fonksiyonudur. (2.1.16)' daki $F_i(x)$ ' in ifadesi ister sürekli ister kesikli olsun herhangi bir keyfi anakütle için sağlanır. Bununla birlikte anakütle mutlak sürekli olduğu farz edilecek şartlar altında (2.1.16)' daki kümülâtif dağılım fonksiyonu ifadenin diferansiyeli alınarak $X_{(i)}$ ' in $(1 \leq i \leq n)$ ' in olasılık yoğunluk fonksiyonu tam olarak (2.1.1)' deki ifadeyi elde ederiz.

2.2 KESİKLİ DAĞILIMLARDA SIRA İSTATİSTİKLERİ

Şimdi i . inci sıra istatistiği için üç ifade elde edilecek. İlk ikisi $F_i(x)$ cdf' sine dayandırılmıştır. Sonuncusu ise çok terimli argumente dayandırılmıştır.

Yaklaşım 1. (Binom Toplamı)

$F_i(x)$ bir ifadesini sağlayan denklemin kesikli durumda $X_{(i)}$ ' in her bir mümkün x değeri için biz

$$f_i(x) = F_i(x) - F_i(x-1) \quad (2.2.1)$$

yazabiliriz. Bu nedenle biz

$$f_i(x) = \sum_{r=i}^n \binom{n}{r} \{ [F(x)]^r [1-F(x)]^{n-r} - [F(x-1)]^r [1-F(x-1)]^{n-r} \} \quad (2.2.2)$$

yazabiliriz. Herhangi bir kimse benzer olarak $F_i(x)$ için negatif binom toplam formunda kullanılabilir.

Yaklaşım 2. (Beta integral Formu)

Hesaplama, amaçları için önceki ifadeden daha iyidir. $f_i(x)$ için bu ifade bir integral formunda kesikli sıra istatistiklerinin bağımlı yapılarını çalışmak için faydalıdır. (2.1.15) ve (2.2.1) kullanılarak yapılır. Bir başka deyişle,

$$f_i(x) = i C(i, n) \int_{F(x-1)}^{F(x)} u^{i-1} (1-u)^{n-i} du \quad (2.2.3)$$

burada

$$C(i, n) = \frac{n!}{i! (n-i)!} \quad (2.2.4)$$

dır.

Yaklaşım 3. (Çoklu Argument)

Mutlak sürekli durumlarda, çoklu denemeleri gerektiren bir argument (2.1.1) ile verilen $X_{(i)}$ ' in pdf' sini elde etmek için kullanılmıştır. O fikir daima burada kullanacağız. Fakat $X_{(i)}$ ' in pmf' si için sonuç daha karma karışıktır. Herhangi bir X gözlem değeri için şu üç farklı olayı göz önüne alalım; $(X < x)$, $(X = x)$, $(X > x)$. Bu olayların olasılıkları sırasıyla $F(x-1)$, $f(x)$ ve $1-F(x)$ dir. $\{X_{(i)}\}$ olayı $i(n-i+1)$ farklı şekilde meydana gelebilir;

$r = 0, 1, \dots, i-1$ ve $s = 0, 1, \dots, n-1$ olmak üzere $(i-1-r)$ tane gözlem değeri x 'den küçük, $(n-i-s)$ tanesi x 'den büyük ve kalanları ise x değerine eşittir.

o halde

$$f_i(x) = \sum_{r=0}^{i-1} \sum_{s=0}^{n-i} \frac{n! [F(x-1)]^{i-1-r} [1-F(x)]^{n-i-s} [F(x)]^{s+r+1}}{(i-1-r)! (n-i-s)! (s+r+1)!} \quad (2.2.5)$$

olacaktır.

2.3. KESİKLİ DAĞILIMLARDA SIRA İSTATİSTİKLERİNİN DAĞILIMLARI

X , $f(k) = \Pr(X = k)$ pmf'li ve $F(k) = \Pr(X \leq k)$ kümülâtif olasılık kütle fonksiyonlu $0, 1, 2, \dots$ değerlerini alan bir kesikli tesadüfî değişken olsun.

Farz edelim ki; $X_1, X_2, \dots, X_n$ X 'de olduğu gibi aynı pmf'li n tane bağımsız ve aynı dağılmış tesadüfî değişken olsun. $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ uygun sıralı istatistikler olsun.

$f_i, X_{(i)}$ 'in pmf'si

$$\begin{aligned} f_i(x) &= \Pr(X_{(i)} = x) \\ &= F_i(x) - F_i(x-1) \\ &= \sum_{r=i}^n C(r, n) \{ [F(x)]^r [1-F(x)]^{n-r} - [F(x-1)]^r [1-F(x-1)]^{n-r} \} \end{aligned}$$

olarak yazabiliriz.

$$\sum_{r=i}^n C(r, n) \{ [F(x)]^r [1-F(x)]^{n-r} \} = \int_0^{F(x)} i C(i, n) u^{i-1} (1-u)^{n-i} du, \quad 0 < u < 1$$

Biz

$$f_i(x) = \int_{F(x-1)}^{F(x)} i C(i, n) u^{i-1} (1-u)^{n-i} du \quad (2.3.1)$$

yazabiliriz. $i = 1$ için $X_{(1)}$ 'in pmf'sini

$$f_1(x) = n \int_{F(x-1)}^{F(x)} (1-u)^{n-1} du = [\bar{F}(x-1)]^n - [\bar{F}(x)]^n, \quad \bar{F}(x) = 1 - F(x)$$

olarak yazarız. $i = 1$ için $X_{(n)}$ 'in pmf'sini

$$f_n(x) = n \int_{F(x-1)}^{F(x)} u^{n-1} du = [F(x)]^n - [F(x-1)]^n$$

yazabiliriz.

$1 \leq i \leq j \leq n$ olmak üzere $X_{(i)}$ ve $X_{(j)}$ ' in bileşik pmf' sini

$$\begin{aligned} f_{i,j}(x,y) &= \Pr(X_{(i)} = x, X_{(j)} = y) \\ &= \sum_{r=0}^{i-1} \sum_{s=0}^{n-j} \sum_{u+w \leq s-r-1} C(i-r-1, i+w, j-1-u, j+s, n) \\ &\quad \{[F(x-1)]^{i-1-r} [F(x)]^{w+1+r} \} [F(y-1) - F(y)]^{j-u-i-w-1} [F(y)]^{u+1+s} [1-F(y)]^{n-j-s} \end{aligned}$$

olarak elde edilir. Basit olarak verilen ifadeden

$$\begin{aligned} f_{i,j}(x,y) &= \Pr(X_{(i)} = x, X_{(j)} = y) = i(j-i) C(i, j, n) \\ &\quad \int_{F(x-1)}^{F(x)} \int_{F(y-1)}^{F(y)} \alpha^{i-1} (\beta - \alpha)^{j-i-1} (1-\beta)^{n-j} d\alpha d\beta \end{aligned}$$

elde edilir. $W_{(1)}$ örnek aralığı $W_{(1)} = X_{(n)} - X_{(1)}$ ile tanımlanır. Kolayca

$$\Pr(W_{(1)} = 0) = \sum_{u=0}^{\infty} [\Pr(X = u)]^n$$

gösterilir.

$k > 0$ için

$$\begin{aligned} \Pr(W_{(1)} = k) &= \sum_{u=0}^{\infty} \Pr(X_{(1)} = u, X_{(n)} = u+k) \\ &= \sum_{u=0}^{\infty} \{ [F(u+k) - F(u-1)]^n - [F(u+k) - F(u)]^n \\ &\quad + [F(u+k-1) - F(u)]^n - [F(u+k-1) - F(u-1)]^n \} \\ &= \sum_{u=0}^{\infty} n(n-1) \int_{F(u-1)}^{F(u)} \int_{F(u+k-1)}^{F(u+k)} (\beta - \alpha)^{n-2} d\alpha d\beta \end{aligned}$$

yazabiliriz. $X_{(i_1)} \leq X_{(i_2)} \leq \dots \leq X_{(i_k)}$ ' in bileşik pmf' si

$$\begin{aligned} f_{i_1, i_2, \dots, i_k}(x_1, x_2, \dots, x_k) &= C(r_1, r_2, \dots, r_k; n) \\ &\quad \cdot \int_D \prod (i_r - i_{r-1}) (u_r - u_{r-1})^{i_r - i_{r-1} - 1} (1 - u_k)^{n - i_k} du_1 du_2 \dots du_k \end{aligned} \quad (2.3.2)$$

buradan

$$r_0 = 0, \quad u_0 = 0 \quad \text{ve} \quad \int_D = \int_{F(x_1-1)}^{F(x_1)} \int_{F(x_2-2)}^{F(x_2)} \dots \int_{F(x_k-1)}^{F(x_k)}$$

dır. Burada $C(r_1, r_2, \dots, r_k; n)$ (2.4.9)'da tanımlanmıştır.

2.4. BİLEŞİK OLASILIK KÜTLE FONKSİYONU

Biz şimdi $X_{(i_1)}, X_{(i_2)}, \dots, X_{(i_k)}$ ile çok sayıda k sıra istatistiklerin bileşik pmf'ini elde ederiz. Burada $1 \leq i_1 < i_2 < \dots < i_k \leq n$ pmf'ini için uygun bir form bir integral olarak ifade edebiliriz. Bunun için tüm sıra istatistikleri ve kalan sıra istatistikleri ile ilgili toplamın bileşik pmf'ini için bir integral formu ile başlarız. Burada tartışmaya çağrı; gözlenmiş değerler ile bağlantılı sayılar arasındaki teması kesmemektir.

$x_1 \leq x_2 \leq \dots \leq x_n$, öyle ki bazı m , $1 < m < n$, $1 \leq r_1 < r_2 < \dots < r_m = n$ için $(x_1 = \dots = x_{r_1}) < (x_{r_1+1} = \dots = x_{r_2}) < \dots < (x_{r_{m-2}+1} = \dots = x_{r_{m-1}}) < (x_{r_{m-1}} < x_{r_{m-1}+1} = \dots = x_{r_m})$ olsun.

O zaman $r_0 = 0$ 'lı

$$f_{1,2,\dots,n}(x_1, x_2, \dots, x_n) = n! \prod_{s=1}^m \frac{[f(x_{r_s})]^{r_s - r_{s-1}}}{(r_s - r_{s-1})!} \quad (2.4.1)$$

$$f_{1,2,\dots,n}(x_1, x_2, \dots, x_n) = n! \int_D du_1 du_2 \dots du_n \quad (2.4.2)$$

burada

$$D = \{(u_1, u_2, \dots, u_n) : u_1 \leq u_2 \leq \dots \leq u_n, F(x_{r_s} - 1) \leq u_t \leq F(x_{r_s}), 1 \leq s \leq m, r_{s-1} \leq t \leq r_s\} \quad (2.4.3)$$

olur.

$k < n$ olduğundan $x_{i_1} \leq x_{i_2} \leq \dots \leq x_{i_k}$ için

$$f_{i_1, i_2, \dots, i_k}(x_{i_1}, x_{i_2}, \dots, x_{i_k}) = \sum f_{1,2,\dots,n}(x_1, x_2, \dots, x_n) \quad (2.4.4)$$

burada $\sum$, $x_1 \leq x_2 \leq \dots \leq x_n$ 'e göre $x_{i_1}, x_{i_2}, \dots, x_{i_k}$ 'ların dışında tüm x_r 'lerin üzerindeki toplam için geçerlidir. $f_{1,2,\dots,n}$ için (2.4.2) tasviri kullanıldığında (2.4.4)'ün sağ tarafı $n! \sum_t \int_{D_t} du_1 \dots du_n$ olarak ifade edilebilir. Burada D_t 'lerin ayrık ve (2.4.3) verilen D formundadır. Bu D_t 'lerin x_r 'ler arasında bağlantıların farklı biçimleri uygundur. Bundan başka

$$\bigcup D_t = \{(u_1, u_2, \dots, u_n) : u_1 \leq u_2 \leq \dots \leq u_n, F(x_{i_r} - 1) \leq u_{i_r} \leq F(x_{i_r}), 1 \leq r \leq k\} \quad (2.4.5)$$

olur. Böylece biz

$$f_{i_1, i_2, \dots, i_k}(x_{i_1}, x_{i_2}, \dots, x_{i_k}) = n! \int_{\bigcup D_t} du_1 du_2 \dots du_n \quad (2.4.6)$$

burada $\bigcup D_t$ (2.4.5) ile verilmiştir. $i_1, i_2, \dots, i_k$ 'ların dışında r için u_r 'lerin bütünleşmek üzere biz $i_0 = 0$ ve $u_0 = 0$ ile

$$\int_{\bigcup D_i} du_1 du_2 \dots du_n = [(n-i_k)! \prod_{r=1}^k (i_r - i_{r-1} - 1)!]^{-1} \\ \cdot \int_B [\prod_{r=1}^k (u_{i_r} - u_{i_{r-1}})^{i_r - i_{r-1} - 1}] (1 - u_{i_k})^{n-i_k} du_{i_1} du_{i_2} \dots du_{i_k}$$

yazarız.

Burada B ,

$$B = \{(u_{i_1}, u_{i_2}, \dots, u_{i_k}) : u_{i_1} \leq u_{i_2} \leq \dots \leq u_{i_k}, F(x_r - 1) \leq u_r \leq F(x_r), r = i_1, i_2, \dots, i_k\} \quad (2.4.7)$$

ile verilen k boyutta uzaydır.

Bu tartışma aşağıdaki gibi özetlenmiş olabilir.

Theorem 2.4.1

$1 \leq i_1 < i_2 < \dots < i_k \leq n$ için $X_{(i_1)}, X_{(i_2)}, \dots, X_{(i_k)}$ 'in bileşik pmf' si

$$f_{i_1, i_2, \dots, i_k}(x_{i_1}, x_{i_2}, \dots, x_{i_k}) = C(i_1, i_2, \dots, i_k : n) \\ \cdot \int_B [\prod_{r=1}^k (u_{i_r} - u_{i_{r-1}})^{i_r - i_{r-1} - 1}] (1 - u_{i_k})^{n-i_k} du_{i_1} du_{i_2} \dots du_{i_k} \quad (2.4.8)$$

ile verilmiştir. Burada $i_0 = 0, u_0 = 0$

$$C(i_1, i_2, \dots, i_k : n) = \frac{n!}{(n-i_k)! \prod_{r=1}^k (i_r - i_{r-1} - 1)!} \quad (2.4.9)$$

ve B (2.4.7) ile verilmiştir.

Yukarıdaki sonuç k türlü bir integral olarak sıra istatistiğinin bileşik pmf' sini ifade eder.

B integrasyon bölgesi, sıra istatistiği bağlanmadığı sürece alt boyutların alt uzayların bir çarpım uzayı olarak ifade edebiliriz.

Örneğin; bağlanmaz durumda B , k boyutlu bir dikdörtgendir. Bununla birlikte hatta bu durumda bileşik pmf' si sıra istatistiklerin ardışık olmadıkça mevcut değildir.

İki sıra istatistiğinin bileşik pmf' sini bulmak için başka bir yaklaşım onların bileşik cdf' sini kullanmaktır. Yani;

$$f_{i,j}(x_i, x_j) = F_{i,j}(x_i, x_j) - F_{i,j}(x_i - 1, x_j) - F_{i,j}(x_i, x_j - 1) + F_{i,j}(x_{i-1}, x_{j-1}) \quad (2.4.10)$$

ifadesi kullanılır. Burada $F_{i,j}$ ifadesi mutlak sürekli durumla aynıdır ve (2.1.3)' de verilmiştir.

3. BÖLÜM

3.1. MOMENTLER VE ÇARPIM MOMENTLERİ

$\mu_{i,n}^{(k)}$ ile $E\left(X_{(i)}^k\right)$ ($1 \leq i \leq n$), sıra istatistiklerinin tekli momentlerini tanıyalım.

Açıktır ki bu momentler sürekli durumda

$$\begin{aligned}\mu_{i,n}^{(k)} &= \int_{-\infty}^{\infty} x^k f_i(x) dx \\ &= \frac{n!}{(i-1)!(n-i)!} \int_{-\infty}^{\infty} x^k [F(x)]^{i-1} [1-F(x)]^{n-i} f(x) dx\end{aligned}\quad (3.1.1)$$

ile ve kesikli durumda

$$\mu_{i,n}^{(k)} = \sum_x x^k f_i(x), \quad 1 \leq i \leq n \quad (3.1.2)$$

ile hesaplanabilir.

Benzer olarak

$$\begin{aligned}\mu_{i,j,n}^{(k,l)} &= \iint_{x < y} x^k y^l f_{i,j}(x, y) dx dy \\ &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \cdot \\ &\quad \iint_{x < y} x^k y^l [F(x)]^{i-1} [F(y)-F(x)]^{j-i-1} [1-F(y)]^{n-j} f(x) f(y) dx dy\end{aligned}\quad (3.1.3)$$

ile ve kesikli durumda

$$\mu_{i,j,n}^{(k,l)} = \sum_{x \leq y} \sum_x x^k y^l f_{i,j}(x, y), \quad 1 \leq i \leq j \leq n \quad (3.1.4)$$

ile hesaplanabilir.

Sıra istatistiklerinin varyansı

$$Var\left(X_{(i)}\right) = \mu_{i,n}^{(2)} - \mu_{i,n}^2, \quad (1 \leq i \leq n) \quad (3.1.5)$$

olarak hesaplanabilir.

burada

$$\mu_{(i)} = \mu_{(i)}^{(1)}$$

dir.

Benzer olarak sıra istatistiklerinin kovaryansı, $\sigma_{i,j,n}$ $1 \leq i \leq j \leq n$ ile tanımlanmıştır ve

$$Cov\left(X_{(i)}, X_{(j)}\right) = \mu_{i,j,n} - \mu_{i,n} \mu_{j,n}, \quad 1 \leq i \leq j \leq n \quad (3.1.6)$$

olarak hesaplanabilir.

$X_{(i)}$ ' in Momenti

Biz $X_{(i)}$ momentini elde etmek için Teorem 1.1.2 deki $X_{(i)} \stackrel{d}{=} F^{-1}(U_{(i)})$ dönüşümünü kullanabiliriz.

$X_{(i)}$ ' in ortalamasını

$$\mu_{i,n} = iC(i,n) \int_0^1 F^{-1}(u) u^{i-1} (1-u)^{n-i} du$$

olarak ifade edebilir. Burada $C(i,n)$, (2.2.4)' de verilmiştir. Bununla birlikte $F^{-1}(u)$ kesikli dağılımların hemen hemen tümü iyi bir formda yazıldığından bu yaklaşım sık sık pratik değildir. S , negatif olmayan tamsayıların bir alt kümesi olduğu zaman bir kesikli dağılımlarda herhangi bir kimse $X_{(i)}$ ' in momentlerini direkt elde edebilmek için Teorem 1.1.4 de $F_i(x)$ cdf' si kullanılarak elde edebilir.

Birinci amacımız Teorem 1.1.4 de elde edilmiş momentlerin bir genellemesini elde etmektir. Bu nedenle aşağıdaki teoremi ifade edebiliriz.

Teorem 3.1.1

S , dağılımın destek kümesi negatif olmayan tamsayıların bir alt kümesi olsun. Sağ taraftaki momentler her zaman mevcut olmak üzere

$$\mu_{r,n}^{(m)} = \sum_{x=0}^{\infty} [(x+1)^m - x^m] (1 - F_r(x))$$

dir.

İspat:

Not edelim ki; $\mu_{r,n}^{(m)}$ mevcut $k \rightarrow \infty$ giderken $k^m P(X_{(r)} > k) \rightarrow 0$ olur. Şimdi

$k \rightarrow \infty$ giderken

$$\begin{aligned} \sum_{x=0}^{\infty} x^m f_r(x) &= \sum_{x=0}^k x^m [P(X_{(r)} > x-1) - P(X_{(r)} > x)] \\ &= \sum_{x=0}^{k-1} [(x+1)^m - x^m] P(X_{(r)} > x) - k^m P(X_{(r)} > k) \end{aligned}$$

göz önüne alalım. Biz

$$\begin{aligned} \mu_{r,n}^{(m)} &= \lim_{k \rightarrow \infty} \sum_{x=0}^{k-1} [(x+1)^m - x^m] P(X_{(r)} > x) - \lim_{k \rightarrow \infty} k^m P(X_{(r)} > k) \\ &= \sum_{x=0}^{\infty} [(x+1)^m - x^m] (1 - F_r(x)) \end{aligned}$$

elde ederiz. Böylece teorem ispat edilmiştir.

Özel olarak $m = 1$ ve $m = 2$ alarak daima

$$\mu_{r,n}^{(1)} = \sum_{x=0}^{\infty} [1 - F_r(x)]$$

ve

$$\mu_{r,n}^{(2)} = 2 \sum_{x=0}^{\infty} x [1 - F_r(x)] + \mu_{r,n}$$

yazabilir. Bu son iki ifade Teorem 1.1.4 ifadesiyle aynıdır.

Not edelim ki Teorem 3.1.1 tüm kesikli dağılımlar için sağlanır. Genelde bu momentleri analitik olarak hesaplamak kolay değildir. Bazen, örnek ekstremlerinin momentleri yazılabilmektedir. Şimdi Teorem 3.1.1 kullanılarak kesikli düzgün dağılımdaki sıra istatistiklerinin örnek maksimumunun momentlerinin nasıl elde edileceğini gösterelim.

3.2. KESİKLİ DÜZGÜN DAĞILIMDAKİ SIRA İSTATİSTİKLERİ

X , $S = \{1, 2, \dots, N\}$ destek kümesi kesikli düzgün anakütleli tesadüfî değişken olsun. O zaman bir X , $[1, N]$ kesikli düzgündür yazabiliriz. Not edelim ki $x \in S$ için pmf' si

$f(x) = \frac{1}{N}$ ve cdf' si $F(x) = \frac{x}{N}$ ile verilmiştir. Bu nedenle i . inci sıra istatistiğinin cdf' si

$$F_i(x) = \sum_{r=i}^n \binom{n}{r} \left(\frac{x}{N}\right)^r \left(1 - \frac{x}{N}\right)^{n-r}, \quad x \in S$$

ile verilir.

Herhangi bir kimse direkt olarak x ve N seçerek binomial dağılım cdf' si için tabloları kullanabilir. Örneğin $N = 10$, S de her x için $x = 10p$, $p = 0,1(0,1)1,0$ olarak ifade edilebilir. Böylece $x \in S$ için

$$F_i(x) = \sum_{r=i}^n \binom{n}{r} p^r (1-p)^{n-r}$$

binomial tablodan okunabilir ve $f_i(x)$, (2.2.1)' den elde edilmiş olabilir.

Amacımıza uygun özel toplam ve aşağıdaki yardımcı teoremle başlayalım

3.2.1. Özel Toplam

Parametrik olmayan istatistik kuramında, özelliklede sıra toplamıyla uğraşırken artı değerli ilk $N - 1$ tamsayısının kuvvet toplamlarına, yani $n = 0, 1, 2, 3, \dots$ iken

$$S(N-1, n) = 1^n + 2^n + 3^n + \dots + (N-1)^n = \sum_{x=1}^{N-1} x^n \quad (3.2.1)$$

ifadesine gerek duyarız. Aşağıdaki teorem bu toplamın bulunmasını kolaylaştırır.

Yardımcı Teorem 3.2.1

Artı değerli her N ve n tamsayısı için

$$\sum_{x=0}^{k-1} \binom{k}{n} S(N-1, n) = N^k - 1$$

dır.

Bu teoremin bir eksiği $S(N-1, n)$ toplamlarının her birini ayrı ayrı, önce $n = 0$, sonra $n = 1$, sonra $n = 2$, vb. için bulmak zorunda kalmamızdır. Sözelimi, $n = 1$ için şunu elde ederiz:

$$\binom{1}{0} S(N-1, 0) = N - 1$$

burada

$$S(N-1, 0) = 1^0 + 2^0 + \dots + (N-1)^0 = N - 1$$

Benzer biçimde $n = 2$ için de şunu buluruz:

$$\begin{aligned} \binom{2}{0} S(N-1, 0) + \binom{2}{1} S(N-1, 1) &= N^2 - 1 \\ N - 1 + 2 S(N-1, 1) &= N^2 - 1 \end{aligned}$$

burada

$$S(N-1, 1) = 1^1 + 2^1 + \dots + (N-1)^1 = \frac{1}{2}(N-1)N$$

dir.

Aynı tekniği kullanarak

$$S(N-1, 2) = \frac{1}{6}(N-1)N(2N-1)$$

ve

$$S(N-1, 3) = \frac{1}{4}(N-1)^2 N^2$$

elde edebilir.

Şimdi $S(N-1, n)$ toplamı ile kesikli düzgün dağılım sıra istatistiklerinin örnek maksimumun m . ci momenti arasındaki ilişkiyi elde edebiliriz.

İlişki 1.

Kesikli düzgün dağılımlı ve $m \geq 1$ için

$$\mu_{n,n}^{(m)} = N^m - \frac{\sum_{k=1}^m \binom{m}{k} S(N-1, n+m-k)}{N^n} \quad (3.2.2)$$

dir.

İspat:

Teorem 3.1.1 de $\mu_{r,n}^{(m)}$ ifadesini göz önüne alalım. Not edelim ki kesikli düzgün dağılım için $F_n(x) = \left(\frac{x}{N}\right)^n$, $r = n$ ve $m \geq 1$ için biz

$$\begin{aligned} \mu_{n,n}^{(m)} &= \sum_{x=0}^N [(x+1)^m - x^m] \left[1 - \left(\frac{x}{N}\right)^n\right] \\ &= \sum_{x=0}^{N-1} [(x+1)^m - x^m] \left[1 - \left(\frac{x}{N}\right)^n\right] \\ &= \sum_{x=0}^{N-1} (x+1)^m \left[1 - \left(\frac{x}{N}\right)^n\right] - \sum_{x=0}^{N-1} x^m \left[1 - \left(\frac{x}{N}\right)^n\right] \\ \mu_{n,n}^{(m)} &= N^m - \frac{1}{N^n} \left\{ \sum_{x=0}^{N-1} [(x+1)^m - x^m] x^n \right\} \end{aligned} \quad (3.2.3)$$

yazabiliriz.

(3.2.1)' de (3.2.3)' in sağ tarafındaki toplam kullanılarak biz daima

$$\sum_{x=0}^{N-1} [(x+1)^m - x^m] x^n = \sum_{k=1}^m \binom{m}{k} S(N-1, n+m-k) \quad (3.2.4)$$

eşitliğini çıkarabiliriz. Böylece (3.2.2) elde ederiz.

Alternatif İspat:

(3.1.2)' de $\mu_{r,n}^{(m)}$ ifadesini göz önüne alalım. Not edelim ki; kesikli düzgün dağılım için (2.2.1)' den

$$f_n(x) = \left(\frac{x}{N}\right)^n - \left(\frac{x-1}{N}\right)^n$$

$r = n$ $m \geq 1$ için biz, (3.1.2)' den

$$\begin{aligned}
\mu_{n,n}^{(m)} &= \sum_x x^m f_r(x) = \sum_{x=1}^N x^m \left[\left(\frac{x}{N} \right)^n - \left(\frac{x-1}{N} \right)^n \right] \\
&= \frac{1}{N^n} \left\{ \sum_{x=1}^N x^m [x^n - (x-1)^n] \right\} \\
&= \frac{1}{N^n} \left[\sum_{x=1}^N x^m x^n - \sum_{x=1}^N x^m (x-1)^n \right] \\
&= \frac{1}{N^n} \left[\sum_{x=1}^{N-1} x^m x^n + N^m N^n - \sum_{x=1}^N x^m (x-1)^n \right] \\
&= N^m - \frac{1}{N^n} \left[\sum_{x=1}^N x^m (x-1)^n - \sum_{x=1}^{N-1} x^m x^n \right] \\
&= N^m - \frac{1}{N^n} \left[\sum_{x=1}^{N-1} x^m (x-1)^n + N^m (N-1)^n - \sum_{x=1}^{N-1} x^m x^n \right] \\
&= N^m - \frac{1}{N^n} \left\{ \sum_{x=1}^{N-1} [x^m (x-1)^n - x^m x^n] + N^m (N-1)^n \right\} \\
\mu_{n,n}^{(m)} &= N^m - \frac{1}{N^n} \left\{ \sum_{x=1}^{N-1} x^m [(x-1)^n - x^n] + N^m (N-1)^n \right\} \tag{3.2.5}
\end{aligned}$$

yazabiliriz. Diğer taraftan (3.2.1)' den biz daima

$$\sum_{x=1}^{N-1} x^m [(x-1)^n - x^n] + N^m (N-1)^n = \sum_{k=1}^m \binom{m}{k} S(N-1, n+m-k)$$

eşitliğini elde ederiz. Bu son ifade (3.2.5)' de yerine yazılırsa (3.2.4) elde edilir. Böylece ispat tamamlanmış olur.

Örnek 3.2.1

$X, [1, N]$ kesikli düzgün tesadüfî değişken olduğu zaman örnek maksimumun durumu ilişki 1'den

$$\mu_{n,n} = N - \frac{S(N-1, n)}{N^n}$$

ve

$$\mu_{n,n}^{(2)} = N^2 - \frac{S(N-1, n+1) + 2S(N-1, n)}{N^n}$$

olur. (3.2.6)' in sağ tarafındaki toplam kolayca hesaplanabilir. Abramowitz and Stegun [1], birkaç n ve N değeri için tabloştürmüştür. Beyer [6], $n = 10$ kadar toplamın cebirsel ifadeleri mevcuttur.

Bunları kullanarak örneğin $n = 2$ için

$$\mu_{2,2} = \frac{(4N-1)(N+1)}{6N}$$

ve

$$\sigma_{2,2}^2 = \frac{(2N^2+1)(N^2-1)}{36N^2}$$

elde ederiz.

TARTIŞMA VE SONUÇ

Kesikli dağılımlardaki sıralı istatistiklerin m . ci momenti elde edilmiştir. Ayrıca bu momentleri kullanarak kesikli düzgün dağılımdaki sıra istatistiklerin örnek maksimumunun cebirsel olarak kolay bir şekilde (ilişki 1) elde edileceği gösterilmiştir. Tüm kesikli dağılımlarda Teorem 3.1.1.'i kullanarak momentlerin elde edilemeyeceği tespit edilmiştir. Üçüncü ve dördüncü momentleri elde etmek için momentler arasındaki ilişkileri de kullanarak ortalamaya göre momentler elde edilebilse de çok uzun ve karmaşık formüller ile karşılaşılmaktadır. Bununla beraber N ve n' nin çeşitli değerleri için programlama yapılarak momentleri sayısal olarak da elde edilebilir.

Kaynaklar

- [1] Abramowitz and Stegun, I.A., 1965, Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables. Dover, New York.
- [2] Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N., 1992, A First Course in Order Statistics, John Wiley and Sons, New York.
- [3] Aytaç, M., 1994, Matematiksel İstatistik, Bursa.
- [4] Balakrishnan, N., Cohen, A.C., 1991, Order Statistics and Inference, New York
- [5] Balakrishnan, N., 1986, Order Statistics from Discrete Distributions, Commun. Statist. Theor. Meth., 15, 3, 657-675.
- [6] Beyer, W.H., 1991, CRC Standard Probability and Statistics Tables and Formulas. CRC Pres, Boca Raton, FL.
- [7] Çalik, S. and Güngör, M., 2004, On the Expected values of the Sample Maximum of Order Statistics from Discrete Uniform Distribution, Applied Mathematic and Computation, 157, 695-700.
- [8] Çalik, S. and Güngör, M., 2005, On the moment of order statistics from discrete distributions, International Conference, Ordered Statistical Data: Approximations, Bounds and Characterizations, İzmir, June 15-18.
- [9] David, H. A., 1970, Order Statistics, New York.
- [10] Freund, J. E., 2002, Matematiksel İstatistik, İstanbul.
- [11] Gibbons, J. D., 1971, Nonparametric Statistical Inference, Tokyo.
- [12] Giri, N. C., 1975, Introduction To Probability and Statistics, New York.
- [13] Hogg, R. Y., Craig, A. T., 1978, Introduction To Mathematical Statistics, New York.
- [14] Hollander, M., Wolfe, D. A., 1973, Nonparametric Statistical Methods, New York.
- [15] Khatri, C.G., 1962, Distribution of Order Statistics for Discrete Case, Ann. Inst. Statist. Math., 14, 167-171.
- [16] Nagaraja, H. N., 1992, Order Statistics from Discrete Distributions, Statistics, 23, 45-47.
- [17] Özçelik, S., 1987, Uygulamalı Genel İstatistik, Elazığ.
- [18] Wilks, S. S., 1962, Mathematical Statistics, Tokyo.

ÖZGEÇMİŞ

Fahrettin ÖZBEY 1982 yılında Elazığ'da doğdu. İlk, orta, lise öğrenimini Elazığ'da tamamladı. 2001 yılında girdiği Fırat Üniversitesi Fen Edebiyat Fakültesi Matematik Bölümü'nden 2005 yılında mezun oldu. Fırat Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı'nda 2005 yılında başladığı yüksek lisans öğrenimine devam etmektedir.