

EGE ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

(DOKTORA TEZİ)

ALT VE ÜST REKOR DEĞERLER İÇİN

BİR KARMA MODEL

Agâh KOZAN

Tez Danışmanı: Yrd. Doç. Dr. Halil TANIL

İstatistik Anabilim Dalı

Bilim Dalı Kodu: 406.01.01

Sunuş Tarihi: 10.06.2015

Bornova-İZMİR

2015

Agâh KOZAN tarafından Doktora tezi olarak sunulan “Alt ve Üst Rekor Değerler İçin Bir Karma Model” başlıklı bu çalışma E.Ü. Lisansüstü Eğitim ve Öğretim Yönetmeliği ile E.Ü. Fen Bilimleri Enstitüsü Eğitim ve Öğretim Yönergesi’nin ilgili hükümleri uyarınca tarafımızdan değerlendirilerek savunmaya değer bulunmuş ve 10.06.2015 tarihinde yapılan tez savunma sınavında aday oybirliği ile başarılı bulunmuştur.

Jüri Üyeleri:

İmza

Jüri Başkanı	:	Prof. Dr. Onur KÖKSOY	
Danışman	:	Yrd. Doç. Dr. Halil TANIL	
Üye	:	Doç. Dr. Güvenç ARSLAN	
Üye	:	Doç. Dr. Selma GÜRLER	
Üye	:	Doç. Dr. Sevcan DEMİR ATALAY	

EGE ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

ETİK KURALLARA UYGUNLUK BEYANI

E.Ü. Lisansüstü Eğitim ve Öğretim Yönetmeliğinin ilgili hükümleri uyarınca Doktora Tezi olarak sunduğum “Alt ve Üst Rekor Değerler için Bir Karma Model” başlıklı bu tezin kendi çalışmam olduğunu, sunduğum tüm sonuç, doküman, bilgi ve belgeleri bizzat ve bu tez çalışması kapsamında elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara atıf yaptığımı ve bunları kaynaklar listesinde usulüne uygun olarak verdiğimi, tez çalışması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını, bu tezin herhangi bir bölümünü bu üniversite veya diğer bir üniversitede başka bir tez çalışması içinde sunmadığımı, bu tezin planlanmasından yazımına kadar bütün safhalarda bilimsel etik kurallarına uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul edeceğimi beyan ederim.

10.06.2015

Agâh KOZAN

ÖZET**ALT VE ÜST REKOR DEĞERLER İÇİN
BİR KARMA MODEL**

KOZAN, Agâh

Doktora Tezi, İstatistik Anabilim Dalı

Tez Danışmanı: Yrd. Doç. Dr. Halil TANIL

Haziran 2015, 42 sayfa

Bu çalışmada, birbirinden bağımsız ve aynı sürekli dağılıma sahip rasgele değişkenler dizisi üzerinde tanımlı olan alt ve üst rekor modelleri birleştirilmiş ve elde edilen bu yeni modeldeki rasgele değişkenlere “Karma-Rekor Değerler” adı verilmiştir. Karma-rekorlar ile güncel rekorlar (current records) arasındaki ilişki üzerinde durularak, bu modelin literatürdeki yeri saptanmıştır. Ayrıca, karma-rekor sayısı ve karma-rekor zamanlarına ilişkin dağılımlar ortaya konmuştur. Son olarak, karma-rekorların küçükten büyüğe sıralanması ile ortaya çıkan ve “Sıralı-Karma-Rekorlar” adı verilen modele ilişkin ortak olasılık yoğunluk fonksiyonu elde edilmiştir.

Anahtar sözcükler: Sıra istatistikleri, rekor değerler, rekor zamanları, güncel rekorlar, rekor açıklığı, rekor kapsamı, karma-rekor değerler, karma-rekor zamanları.

ABSTRACT**A MIXED MODEL FOR LOWER AND UPPER
RECORD VALUES**

KOZAN, Ağâh

Ph.D. in Department of Statistics

Supervisor: Asst. Prof. Dr. Halil TANIL

June 2015, 42 pages

In this study, lower and upper record models, which are defined in a sequence of independent and identically distributed continuous random variables, are mixed and the random variables in this new model are named as “Mixed-Record Values”. By putting emphasis on the relation between mixed-records and current records, the place of this new model is determined. Also, some distributions about mixed-record count and times of mixed-records are presented. Finally, the new random variables, which are acquired by ordering the mixed-records in ascending order, are named as “Ordered-Mixed-Records” and joint probability density function of this model is obtained.

Keywords: Order statistics, record values, record times, current records, record range, record coverage, mixed-record values, mixed-record times.

TEŞEKKÜR

Bu çalışma süresince benden ilgisini ve desteğini esirgemeyip beni bu alanda yetiştiren değerli hocam Sayın Yrd. Doç. Dr. Halil TANIL'a, Lisans, Yüksek Lisans ve Doktora öğrenimimi gördüğüm Ege Üniversitesi İstatistik Bölümü'ndeki tüm değerli hocalarıma, hayatım boyunca her konuda beni destekleyen ve yanımda olan annem, babam ve ablama, bu yolda ilerlerken sevgi ve anlayışını eksik etmeyen çok sevdiğim eşim Elif'e sonsuz teşekkürler...

İÇİNDEKİLER

Sayfa

ÖZET

ABSTRACT ix

TEŞEKKÜR xi

ÇİZELGELER DİZİNİ xiv

1. GİRİŞ 1

2. ÖNCEKİ ÇALIŞMALAR 3

2.1 Sıra İstatistikleri 3

2.2 Rekor Değerler ve Rekor Zamanları 5

2.3 Güncel Rekor Değerler, Rekor Kapsamı ve Rekor Açıklığı 12

3. KARMA-REKOR KAVRAMI 15

3.1 Karma-Rekor İndikatörü 16

3.2 Karma-Rekor Sayısı 18

3.3 Karma-Rekor Zamanları 22

3.4 Dağılımdan-Bağımsız Bir Güven Aralığı 25

4. SIRALI-KARMA-REKOR DEĞERLER 28

5. SONUÇ VE TARTIŞMA 36

KAYNAKLAR DİZİNİ 39

ÖZGEÇMİŞ 42

EKLER

ÇİZELGELER DİZİNİ

Çizelge

Sayfa

3.2.1 Çeşitli n 'ler için $\{P(N'_n \geq k)\}_{k \leq 15}$ olasılıkları	20
3.2.2 Uniform(0,1) dağılımlı kitleden türetilen n ölçümlü örneklemeler için ortalama karma-rekor sayıları ve karma-rekor sayılarının varyansları	22

1. GİRİŞ

Uç olaylar yaşantımızda önemli bir yere sahiptir. Örneğin, insanların bir kısmı, yaratılış mizaçları gereği herhangi bir konuda karar verirken “En kötü ne olabilir?” sorusuna verecekleri cevabı dikkate alırlar. Bazı insanlarsa, iyimser mizaçları gereği, olabilecek en iyi durumları referans alarak hareket ederler. Genel olarak, bu iki farklı anlayışı birleştirerek yaşantımıza yön vermek daha gerçekçi ve daha mantıklı gözükmektedir.

Yukarıda bahsedilen tek yönlü bakış açılarının karşılığı istatistik literatüründe “Rekor Değerler” kavramı olarak karşımıza çıkmaktadır. Dağılım teorisi ilk kez Chandler (1952) tarafından ortaya konmuş olan bu kavram başlangıçta, alt rekorlar ve üst rekorlar olmak üzere iki ayrı model olarak incelenmiş, daha sonra Houchens (1984) tarafından ortaya atılan “Güncel Rekorlar” kavramı ile bu iki ayrı model kısmen birleştirilmiştir. Houchens, çalışmasında, sayılabilir sonsuz bir gözlem dizisinde, alt ve üst ayrımı yapılmaksızın kaydedilecek rekorların en küçüğü ve en büyüğünü dikkate alan bir model ortaya atmıştır. Ayrıca, O’nun bu çalışmasına dayalı olarak, istatistik literatürüne, “Rekor Kapsamı” (Record Coverage) ve “Rekor Açıklığı” (Record Range) kavramları girmiş ve günümüze kadar bu kavramlara dayalı pek çok akademik makale yazılmıştır.

Bu tez çalışmasında, Houchens’in “kısmen birleştirme” yaklaşımı bir adım öteye taşınmış, başlangıçtan itibaren alt ve üst ayrımı yapılmaksızın kaydedilen tüm rekorlardan oluşan koleksiyon dikkate alınarak “Karma-Rekorlar” adını verdiğimiz yeni bir model ortaya atılmıştır. “Başarı”, bir sistemin üst rekor kırması, “Başarısızlık” ise alt rekor kırması olarak tanımlanırsa, Houchens’in modeli o sistemi son başarı ve son başarısızlığına göre; karma-rekorlar ise, yine aynı sistemi, tüm başarı ve tüm başarısızlıklarını birlikte dikkate alacak şekilde değerlendirmek için kullanılabilir. Chandler’in üst (alt) rekorlar modeli ise, aynı sistemi, sadece tüm başarılarına (başarısızlıklarına) göre değerlendirmek için uygundur.

Bu tez çalışmasının ikinci bölümünde, bu çalışmanın konusu ile ilgili literatürde yer almış bazı modeller ve konular ayrıntılı olarak ele alınmıştır. Üçüncü bölümde karma-rekor değer kavramı tanımlanmış, karma-rekor indikatörünün tanımı verilerek karma-rekor değerler ile literatürde yer alan güncel rekor değerler arasındaki ilişki açıklanmıştır. Bu tanımdan hareketle k . gözlemin n . karma-rekor olması olasılığı elde edilmiştir. Sonrasında, karma-rekor sayısının dağılımı verilmiş ve bu istatistiğin beklenen değeri ve varyansı ortaya konmuştur. Karma-rekor sayısına ilişkin bulunan bazı sonuçların doğruluğu simülasyon yoluyla da doğrulanmıştır. Bunların yanı sıra, yeni bir gözlemin n . rekor kapsamı içine düşmesi olasılığı, dağılımdan bağımsız güven aralığı şeklinde verilmiştir. Çalışmanın dördüncü bölümünde, sıralı-karma-rekor değer modeli tanımlanmış ve ilk n adet sıralı-karma-rekor değer için ortak olasılık yoğunluk fonksiyonu için bir indirgeme bağıntısı ortaya konmuştur. Ayrıca, bu ortak olasılık yoğunluk fonksiyonu kapalı biçimde elde edilmiştir. Son olarak, beşinci bölümde ise çalışmada bulunan sonuçların önemi ve literatüre katkısı vurgulanmıştır.

2. ÖNCEKİ ÇALIŞMALAR

Sıralı rasgele değişkenlere ilişkin modeller, İstatistik literatüründe önemli bir yere sahiptir. Bu bölümde, tez konusuyla yakından ilişkili olan, sıralı rasgele değişken modelleri arasında yer alan, sıra istatistikleri, rekorlar ve rekor zamanları, güncel rekorlar, rekor kapsamı ve rekor açıklığı kavramlarına değinilecek, böylelikle tezin literatürdeki yeri belirtilmiş olacaktır.

2.1 Sıra İstatistikleri

Sıra istatistikleri pek çok alanda uygulama bulmuş bir konudur. Çünkü sıra istatistikleri yeterli istatistiklerdir ve uç değerler, açıklık, medyan gibi istatistikler sıra istatistiklerine dayalı olarak tanımlıdır. Uç değerler ele alındığında en küçük ve en büyük sıra istatistiği ile ilgilenilmektedir. Açıklık, bir başka deyişle en büyük ve en küçük sıra istatistiğinin farkı, bilindiği üzere değişkenliğin basit bir tahmin edicisi olarak ön plana çıkmakta ve özellikle kalite kontrol uygulamalarında kabul görmektedir. Medyan da benzer şekilde sıra istatistikleri ile ifade edilmekte ve konum parametresi için dayanıklı bir tahmin edici olarak kabul görmektedir. Sıra istatistiklerinin bir başka uygulama alanı ise sistem güvenilirliğidir. Bilindiği üzere, n tane bileşeninden en az k tanesi sağlam olduğunda çalışabilen sistemlere " n 'den k çıkışlı sistemler" adı verilmektedir. Bu tür sistemlerin bozulma süresi $(n - k + 1)$. sıra istatistiği ile ifade edilmektedir. Bunların yanı sıra, örneğin bir uzay mekiğinde toplanan partiküller, doğa olayları veya finansal çalışmalarda oldukça büyük veri setleri ile ilgilenilmektedir. Bunlar gibi büyük boyutlu örneklem için veri sıkıştırma aracı olarak da sıra istatistikleri kullanılabilir. Böyle bir durumda toplanan büyük hacimli örneklem yerine, sadece ilgilenilen ve yeterli sayıdaki sıra istatistiklerinden oluşan liste kullanılarak daha az veri ile çalışılır ve veri depolama alanı küçültülebilir.

Yıllar içinde sıra istatistikleri ile ilgili pek çok çalışma yapılmıştır. Sarhan and Greenberg (1962) sıra istatistikleri hakkında 1960'a kadar olan gelişmeleri sentezlemiştir. Sonrasında David (1970), Harter (1970), Galambos (1978), Balakrishnan and Cohen (1991), Arnold et al. (1992), Ahsanullah and Nevzorov (2005) gibi sıra istatistikleri üzerine pek çok çalışma literatüre geçmiştir.

Sıra istatistiklerini kısaca şu şekilde tanımlayabiliriz: $X_1, X_2, \dots, X_n$ ’ler sürekli rasgele değişkenler olsun. Bu rasgele değişkenlerin r . en küçüğü $X_{r:n}$ ile gösterilirse, $X_{1:n} < X_{2:n} < \dots < X_{n:n}$ şeklinde ifade edilen rasgele değişkenlere “Sıra İstatistikleri” denir. $1 \leq r \leq n$ olmak üzere, $X_{r:n}$ rasgele değişkenine, İstatistik literatüründe “ r . sıra istatistiği” adı verilir (Arnold et al, 1992).

Yukarıdaki bilgiler ışığında, $X_1, X_2, \dots, X_n$, birbirinden bağımsız ve aynı sürekli $F(x)$ dağılım fonksiyonu ve $f(x)$ olasılık yoğunluk fonksiyonuna sahip rasgele bir örneklem olsun. Bu durumda sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonu, $-\infty < x_1 < x_2 < \dots < x_n < \infty$ ve $n \in N^+$ olmak üzere,

$$f^{X_{1:n}, X_{2:n}, \dots, X_{n:n}}(x_1, x_2, \dots, x_n) = n! \prod_{i=1}^n f(x_i)$$

ve en küçük r adet sıra istatistiğinin ortak olasılık yoğunluk fonksiyonu ise,

$$f^{X_{1:n}, X_{2:n}, \dots, X_{r:n}}(x_1, x_2, \dots, x_r) = \frac{n!}{(n-r)!} [1 - F(x_r)]^{(n-r)} \prod_{i=1}^r f(x_i)$$

şeklinde literatürde yerini almıştır. Ayrıca, r . sıra istatistiğinin olasılık yoğunluk fonksiyonu ise, $-\infty < x < \infty$, $1 \leq r \leq n$ ve $n \in N^+$ olmak üzere,

$$f^{X_{r:n}}(x) = \frac{n!}{(r-1)!(n-r)!} F(x)^{r-1} [1 - F(x)]^{n-r} f(x)$$

şeklindedir. Öte yandan, r . ve s . sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonunun, $1 \leq r < s \leq n$, $n \in N^+$ ve $-\infty < x_r < x_s < \infty$ için,

$$\begin{aligned} f^{X_{r:n}, X_{s:n}}(x_r, x_s) \\ = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} F(x_r)^{r-1} [F(x_s) - F(x_r)]^{s-r-1} \\ \times [1 - F(x_s)]^{n-s} f(x_r) f(x_s) \end{aligned}$$

olduğu bilinmektedir (David, 1970).

2.2 Rekor Değerler ve Rekor Zamanları

Arnold et al. (1992) kitabında Rekor Değerler konusuna, “4 Temmuz 1957’de Riverside Kalifornia’da sıcaklıklar hızla yükselmışti ve insanlar merak içinde radyolarını açıp “Acaba asfaltta yumurta pişirilir mi?” “Acaba daha önce böyle yüksek bir sıcaklık yaşanmış mıydı?” sorularının cevaplarını arıyorlardı.” şeklinde bir giriş yapmıştır. İşte tam da bu ve benzeri sorular, bilim insanlarının rekorlara ve rekor değerlere dayalı istatistiklere yönelmesine sebep olmuştur. Rekor değerlerin dağılımını ortaya koyarak konuyla ilgili İstatistik literatüründeki ilk çalışmayı yaptığı kabul edilen Chandler da o zamanlar benzer soruların cevabını aramıştır. Chandler’ın (1952) bu çalışması, İstatistik literatüründe, doğrudan rekorlar üzerine yapılmış olan ilk çalışma olarak bilinmektedir. Resnick (1973) ve Shorrock’un (1972, 1973) rekorların asimptotik teorisi ile ilgili yapmış oldukları çalışmaları sayesinde 21 yıl önce ortaya atılan rekor değerler konusunda büyük gelişmeler ortaya çıkmıştır. Sonrasında, rekor değerler konusu farklı dağılımlı gözlem dizileri gibi alanlara doğru da genişlemeye başlamıştır. O günden bu yana pek çok makale ve Nevzorov (2001), Ahsanullah and Raqab (2006), Kotz and Nadaraja (2000) gibi pek çok kitap yazılmıştır. Glick’in (1978) yayınladığı çalışmasında rekor değerler konusunun ilk 25 yılını ve gelişmelerini anlatmıştır. Nagaraja’nın (1988) çalışması da benzer şekilde sonraki 10 yılı ele almıştır.

Rekor değerler modelinin tanımına göre, $X = \{X_i\}_{i \in \mathbb{N}^+}$ birbirinden bağımsız ve aynı sürekli $F(x)$ dağılım fonksiyonu ve $f(x)$ olasılık yoğunluk fonksiyonuna sahip rasgele değişkenler dizisi olmak üzere, X_j gözlemi, kendinden önceki tüm gözlemlerden büyükse bu değere “Üst Rekor Değer” adı verilir. Yani, $j = 1, 2, 3, \dots$ ve $X_{0:0} \rightarrow -\infty$ olmak üzere, $X_j > X_{j-1:j-1}$ şartını sağlayan X_j gözlemine üst rekor adı verilir. Bu durumda, üst rekor zamanları $U(n)$ ile gösterilecek olursa, bu zamanlar,

$$U(1) = 1$$

ve $n > 1$ için

$$U(n) = \min\{j > U(n-1): X_j > X_{U(n-1)}\}$$

şeklinde tanımlanır. Burada $U(1)$ başlangıç rekor değer zamanını göstermektedir ve başlangıç rekor değer her zaman için X_1 'dir. Bu tanıma göre, n . üst rekor değer marjinal olasılık yoğunluk fonksiyonu,

$$f^{X_{U(n)}}(x) = f(x) \frac{[-\ln(\bar{F}(x))]^{n-1}}{(n-1)!}, -\infty < x < \infty$$

ve ilk n adet üst rekorun ortak olasılık yoğunluk fonksiyonu ise, $R(x) = -\ln(\bar{F}(x))$ ve $r(x) = \frac{d}{dx}R(x) = \frac{f(x)}{1-F(x)}$ olmak üzere,

$$f^{X_{U(1)}, X_{U(2)}, \dots, X_{U(n)}}(x_1, x_2, \dots, x_n) = r(x_1)r(x_2) \dots r(x_{n-1})f(x_n),$$

$$-\infty < x_1 < x_2 < \dots < x_{n-1} < x_n < \infty$$

şeklinde dir. Diğer yandan, kendinden önceki tüm gözlemlerden küçük olan her değere ise “Alt Rekor Değer” adı verilir ve benzer şekilde alt rekor zamanları $L(n)$ ile gösterilirse,

$$L(1) = 1$$

ve $n > 1$ için

$$L(n) = \min\{j > L(n-1): Y_j < Y_{L(n-1)}\}$$

şeklinde tanımlanır. n . alt rekor değer marjinal olasılık yoğunluk fonksiyonu,

$$f^{Y_{L(n)}}(y) = f(y) \frac{[-\ln(F(y))]^{n-1}}{(n-1)!}, -\infty < y < \infty$$

ve ilk n adet alt rekorun ortak olasılık yoğunluk fonksiyonu ise, $H(y) = -\ln(F(y))$, $h(y) = -\frac{d}{dy}H(y) = \frac{f(y)}{F(y)}$ olmak üzere,

$$f^{Y_{L(1)}, Y_{L(2)}, \dots, Y_{L(n)}}(y_1, y_2, \dots, y_n) = h(y_1)h(y_2) \dots h(y_{n-1})f(y_n),$$

$$-\infty < y_n < y_{n-1} < \dots < y_1 < \infty$$

şeklindedir (Ahsanullah, 2004).

Rekor Değerlerin dağılımına ilişkin bazı özelliklerin literatüre kazandırılmasından sonra bu konuda yeni istatistiklerin de tanımlanması ihtiyacı doğmuştur. Sürekli dağılıma sahip $X_1, X_2, \dots, X_n$ rasgele değişkenlerinden oluşan rasgele bir örneklem ele alınsın. Wilks (1959), $X_{n:n} = \max\{X_1, X_2, \dots, X_n\}$ 'i aşan ilk yeni gözlemin kaçınıcı gözlem olduğunu gösteren $N(n)$ rasgele değişkenine ait dağılımı ortaya koymuştur. Buna göre,

$$\begin{aligned} P\{N(n) > m\} &= P\{\max\{X_{n+1}, X_{n+2}, \dots, X_{n+m}\} \leq X_{n:n}\} \\ &= \int_{-\infty}^{\infty} F^m(u) d(F^n(u)) = n \int_{-\infty}^{\infty} F^{m+n-1}(u) dF(u) \\ &= n \int_0^1 v^{m+n-1} dv = \frac{n}{n+m} \end{aligned}$$

olmaktadır. Süreklilikten dolayı birbirinden bağımsız olan $X_1, X_2, \dots, X_{n+m}$ rasgele değişkenlerinin kesin olarak birbirlerinden farklı değerler alacağını biliyoruz ve bu rasgele değişkenlerin her biri aynı dağılıma sahip olduğuna göre her bir $\{X_{n+m:n+m} = X_k\}$, $1 \leq k \leq n+m$ olayının gerçekleşmesi eşit olasılıklıdır. Bu durumda,

$$P\{X_{n+m:n+m} = X_k\} = \frac{1}{n+m}$$

olacaktır. Ayrıca,

$$\begin{aligned} P\{\max\{X_{n+1}, X_{n+2}, \dots, X_{n+m}\} \leq X_{n:n}\} &= P\{X_{n+m:n+m} = X_{n:n}\} \\ &= \sum_{k=1}^n P\{X_{n+m:n+m} = X_k\} = \frac{n}{n+m} \end{aligned}$$

olacağı açıktır. Bu sonuçlara göre n adet gözlemden sonra ikinci rekorun, bir başka deyişle başlangıç rekorundan sonraki ilk rekorun meydana gelmesi olasılığı,

$$\begin{aligned}
& P\{U(2) > n\} \\
&= P\{X_1 \text{ rasgele de\u0131i\u015fkeni } X_1, X_2, \dots, X_n \text{ arasından en b\u00fcy\u00fck de\u011ferdir}\} \\
&= \frac{1}{n}, n = 1, 2, \dots
\end{aligned}$$

olur ve ikinci rekorun n . g\u00f6zlemde meydana gelmesi olasılı\u011fı,

$$\begin{aligned}
P\{U(2) = n\} &= P\{U(2) > n - 1\} - P\{U(2) > n\} \\
&= \frac{1}{n(n-1)}, n = 2, 3, \dots
\end{aligned}$$

\u015feklinde elde edilir. Buna g\u00f6re ikinci rekorun ortalama olarak meydana gelme zamanı,

$$E(U(2)) = \sum_{n=2}^{\infty} \frac{1}{n-1} = \infty$$

olacaktır. $U(k) > U(2)$ ve $E(U(2)) = \infty$ oldu\u011funa g\u00f6re $\forall k \geq 2$ i\u00e7in $E(U(k)) = \infty$ olur. Bu sonu\u00e7tan g\u00f6r\u00fclene\u011fi \u00fczere rekor zamanları i\u00e7in beklenen bir ger\u00e7ekle\u015fme s\u00fcresinden s\u00f6z etmek m\u00fcmk\u00fcn de\u011gildir. Bu durum \u015f\u00f6yle bir olaya benzetilebilir: Talihsiz bir ki\u015finin bir kuyrukta bekledi\u011fi s\u00fcreden daha fazla bekleyece\u011fi d\u00fc\u015f\u00fcn\u00fclen bir ki\u015finin ortalama bekleme s\u00fcresi sonsuzdur. Feller, bu durumu 1966 tarihli kitabında “K\u00f6t\u00fc \u015fansın Devamlılı\u011fı” ba\u015flığı altında bir anekdot \u015feklinde anlatmı\u015f ve ispatlamı\u015ftır. \u00dcst rekor zamanları i\u00e7in elde edilen bu sonu\u00e7 alt rekorlar i\u00e7in de ge\u00e7erlidir (Arnold et al., 1992).

Tanımlanan bir ba\u015fka rekor istatisti\u011fi ise rekor indikat\u00f6r\u00fcd\u00fcr. $\{I_n\}_{n=1}^{\infty}$ \u015feklinde g\u00f6sterilen rekor indikat\u00f6rleri,

$$I_1 = 1$$

ve $n > 1$ i\u00e7in

$$I_n = \begin{cases} 1, & X_n > X_{n-1:n-1} \\ 0, & \text{di\u011fer} \end{cases}$$

şeklinde tanımlanmaktadır. Başka bir deyişle, n . rekor indikatörü, $X = \{X_i\}_{i \in \mathbb{N}^+}$ dizisindeki n . gözlem değerinin rekor değer olarak meydana gelip gelmediğini gösteren bir rasgele değişken olarak tanımlanmıştır. Rekor indikatörleri birbirinden bağımsızdır ve

$$P(I_n = i_n) = \left(\frac{1}{n}\right)^{i_n} \left(\frac{n-1}{n}\right)^{1-i_n}, i_n = 0,1$$

olasılık fonksiyonuna sahip $\left(\frac{1}{n}\right)$ parametrelili Bernoulli dağılımına uymaktadırlar (Renyi, 1962).

Tanımlanan bu rekor indikatörleri kullanılarak bir başka rekor istatistiği olan rekor zamanlarının ortak dağılımını elde etmek kolaylaşır. Buna göre, $1 < n_1 < n_2 < \dots < n_k$ için,

$$\begin{aligned} & \{U(2) = n_2, U(3) = n_3, \dots, U(k) = n_k\} \\ & \equiv \left\{ \begin{array}{l} I_2 = 0, I_3 = 0, \dots, I_{n_2-1} = 0, I_{n_2} = 1, \\ I_{n_2+1} = 0, I_{n_2+2} = 0, \dots, I_{n_3-1} = 0, I_{n_3} = 1, \\ \vdots \\ I_{n_{k-1}+1} = 0, I_{n_{k-1}+2} = 0, \dots, I_{n_k-1} = 0, I_{n_k} = 1 \end{array} \right\} \end{aligned}$$

yazılır. Bu durumda, rekor zamanlarının ortak olasılık fonksiyonu,

$$\begin{aligned} P(U(2) = n_2, U(3) = n_3, \dots, U(k) = n_k) &= \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} \dots \frac{n_2-2}{n_2-1} \frac{1}{n_2} \frac{n_2}{n_2+1} \dots \frac{1}{n_k} \\ &= [(n_1-1)(n_2-1) \dots (n_k-1)n_k]^{-1} \end{aligned}$$

şeklindedir (Arnold et al., 1992).

Bir başka rekor istatistiği ise, $X_1, X_2, \dots, X_n$ sonlu dizisindeki toplam rekor sayısını gösteren N_n rasgele değişkenidir. N_n rasgele değişkeninin dağılımı, rekor indikatörlerinin olasılık üreten fonksiyonları kullanılarak elde edilebilir, çünkü rekor sayısı ile birbirinden bağımsız rekor indikatörleri arasında $N_n = \sum_{j=1}^n I_j$ ilişkisi vardır. Bu durumda N_n rasgele değişkeninin olasılık üreten fonksiyonu,

$$G(s) = E(s^{N_n}) = \prod_{i=1}^n E(s^{I_i}) = \frac{1}{n!} \prod_{i=1}^n (s + i - 1)$$

şeklinde elde edilir. Bu durumda, n uzunluklu bir rasgele değişkenler dizisinde k tane üst rekorun ortaya çıkması olasılığı,

$$p(k) = P(N_n = k) = \frac{1}{k!} \left. \frac{d^k G(s)}{ds^k} \right|_{s=0}$$

şeklinde elde edilir. Başka bir deyişle, $P(N_n = k)$ olasılığı,

$$\left[\frac{1}{n!} \prod_{i=1}^n (s + i - 1) \right]$$

ifadesindeki s^k 'nin katsayısıdır (Arnold et al. 1992). Rekor sayısının beklenen değeri ve varyansı ise, γ Euler sabiti ($\gamma = 0.5772 \dots$) olmak üzere,

$$E(N_n) = \sum_{j=1}^n E(I_j) = \sum_{j=1}^n \frac{1}{j} \cong \ln n + \gamma$$

ve

$$Var(N_n) = \sum_{j=1}^n var(I_j) = \sum_{j=1}^n \frac{1}{j} \left(1 - \frac{1}{j}\right)$$

$$= \sum_{j=1}^n \frac{1}{j} - \sum_{j=1}^n \frac{1}{j^2}$$

$$\cong \ln n + \gamma - \frac{\pi^2}{6}$$

şeklinde elde edilir (Arnold et al., 1992). Yukarıdaki sonuçlara göre N_n rasgele değişkeni için bulunan beklenen değere bakıldığında rekorların meydana gelme sıklığının gerçekten de az olduğu görülmektedir. Örneğin, Cantelli eşitsizliği kullanıldığında,

$$P(N_{1000} \geq 20) \leq 0,036$$

elde edilir ve bu sonuç rekorların meydana gelme sıklığının ne kadar az olduğunu gözler önüne sermektedir.

Rekor indikatörlerinin ortak olasılık fonksiyonu kullanılarak $k > 1$ için $U(k)$ 'nin marjinal olasılık dağılımı elde edilirken

$$\{U(k) > n\} \equiv \{N_n < k\}$$

ve

$$\{U(k) = n\} \equiv \{N_n = k, N_{n-1} = k - 1\} \equiv \{I_n = 1, N_{n-1} = k - 1\}$$

şeklinde ilişkilerden yararlanılabilir. O hâlde, rekor indikatörleri, I_n 'lerin birbirinden bağımsız olması göz önüne alındığında $\{I_n = 1\}$ ve $\{N_{n-1} = k - 1\}$ olaylarının da birbirinden bağımsız olacağı ve bu iki olayın olasılıkları bilindiğine göre, n . gözlemin k . üst rekor olması olasılığı,

$$P(U(k) = n) = P(I_n = 1)P(N_{n-1} = k - 1)$$

olacaktır ki bu olasılık,

$$\frac{1}{n!} \prod_{i=1}^{n-1} (s + i - 1)$$

eşitliğindeki s^{k-1} 'in katsayısıdır (Arnold et al., 1992).

2.3 Güncel Rekor Değerler, Rekor Kapsamı ve Rekor Açıklığı

Rekor değerler konusu geliştikçe yeni konuların ortaya çıkmasına vesile olmuştur. Bunlardan bir tanesi de Güncel Rekor Değerler kavramıdır. Klasik rekorların yanı sıra, bazı çalışmalarda, sadece o ana kadar meydana gelmiş olan en büyük ve en küçük değer ile ilgilenilmek istenebilir. Bu ikiliye literatürde “Güncel Rekorlar” adı verilmektedir. Bu konudaki ilk çalışma, Houchens tarafından 1984’de yapılmıştır.

L'_n ve U'_n rasgele değişkenleri, bağımsız ve aynı F dağılım, f olasılık yoğunluk fonksiyonuna sahip $X = \{X_i\}_{i \in \mathbb{N}^+}$ dizisindeki, sırasıyla n . güncel alt ve üst rekor ikilisini gösterebilir. Başlangıç güncel rekor ikilisinin birbirine eşit olacağı açıktır ($L'_0 = U'_0 = X_1$). L'_n ’nin marjinal olasılık yoğunluk fonksiyonu,

$$f_{L'_n}(x) = 2^n f(x) \left\{ 1 - F(x) \sum_{j=0}^{n-1} \frac{[-\ln F(x)]^j}{j!} \right\}$$

ve U'_n ’nin marjinal olasılık yoğunluk fonksiyonu ise,

$$f_{U'_n}(x) = 2^n f(x) \left\{ 1 - [1 - F(x)] \sum_{j=0}^{n-1} \frac{[-\ln(1 - F(x))]^j}{j!} \right\}$$

şeklinde dir. L'_n ve U'_n için ortak olasılık yoğunluk fonksiyonu ise,

$$f^{L'_n, U'_n}(\ell, u) = \frac{2^n}{(n-1)!} f(\ell) f(u) (-\ln[1 - F(u) + F(\ell)])^{n-1}, \ell < u$$

şeklinde verilmiştir (Ahmadi and Balakrishnan, 2005a).

Güncel rekor ikilisine dayalı (L'_n, U'_n) aralığı, ilgili literatürde “Rekor Kapsamı” ve $R_n = U'_n - L'_n$ rasgele değişkeni ise “Rekor Açıklığı” olarak tanımlanmıştır. Görüldüğü üzere rekor açıklığı güncel rekor değerler ile oldukça ilişkilidir. $X = \{X_i\}_{i \in \mathbb{N}^+}$ dizisinde alt veya üst ayrımı olmaksızın bir rekor meydana geldiği anda yeni bir rekor açıklığının meydana geleceği açıkça görülmektedir.

Yukarıda verilen güncel rekor ikilisinin ortak olasılık yoğunluk fonksiyonunda yapılan değişken dönüşümü sayesinde n . rekor açıklığının olasılık yoğunluk fonksiyonu,

$$f_{R_n}(r) = \frac{2^n}{(n-1)!} \int_{-\infty}^{\infty} f(r+z) f(z) \left[-\ln(1 - F(r+z) + F(z))^{n-1} \right] dz$$

olacaktır ve X_i 'lerin $Uniform(0,1)$ dağıldığı varsayımı altında ilgili olasılık yoğunluk fonksiyonu,

$$f_{R_n}(r) = \frac{2^n(1-r)}{(n-1)!} [-\ln(1-r)]^{n-1}, 0 < r < 1$$

şeklinde elde edilir (Arnold et al., 1992).

Güncel rekorlar ve rekor açıklığı konularında literatürde pek çok çalışma bulunmaktadır. Örneğin, Ahmadi and Balakrishnan'ın (2005a) çalışmasına bakıldığında, kullanmış oldukları veri setine dayalı güncel rekorlar ile ilgilendikleri görülmektedir.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13
L'_n	19,0	19,0	18,4	17,4	17,4	17,4	17,2	15,6	15,6	15,6	15,6	15,3	15,3	15,3
U'_n	19,0	20,1	20,1	20,1	21,0	21,4	21,4	21,4	21,7	22,0	22,1	22,1	22,6	23,4

Yukarıdaki değerler, 1864 – 1993 yılları arasında İsviçre'nin Neurenburg şehrindeki Temmuz ayı ortalama sıcaklıklarına ilişkin veri seti kullanılarak elde edilmiştir. 1993 yılına gelindiğinde rekor kapsamının 1864'ten beri 13 kez değiştiği ve sonuçta $(L'_{13}, U'_{13}) = (15.3, 23.4)$ aralığına ulaşıldığı açıktır.

Ayrıca, Ahmadi ve Balakrishnan güncel rekorlara dayalı olarak kuantiller ve kuantil aralıkları için dağılımdan bağımsız güven aralıkları (2004) ve bazı rekor istatistikleri için güvenilirlik özellikleri (2005b) gibi konularda çalışmalar yapmışlardır. Ahmadi, Razhmkhah ve Balakrishnan (2009) güncel k -rekorlar ve bunların, dağılımdan bağımsız güven aralıkları için kullanımını incelemişlerdir.

Ahmadi ve Balakrishnan (2011) rekor kapsamına dayalı olarak sıra istatistikleri için dağılımdan bağımsız kestirim aralıkları üzerine çalışmışlardır.

3. KARMA-REKOR KAVRAMI

Bu bölümde, karma-rekorlar ve karma-rekor indikatörü tanımlanacak, karma-rekor sayısının ve karma-rekor zamanlarının dağılımları verilecek ve son olarak dağılımdan-bağımsız bir güven aralığı oluşturulacaktır.

Güncel rekorlarla yakından ilişkili olan karma-rekorlar aşağıdaki gibi tanımlanır.

Tanım 3.1

i zaman indeksi ve $X = \{X_i\}_{i \in \mathbb{N}^+}$, sayılabilir sonsuz sürekli rasgele değişkenler dizisi olmak üzere, alt veya üst ayrımı yapılmadan bu dizinin rekorları kaydedilecek olsun. Kaydedilen her bir değere “Karma-Rekor” adı verilir ve karma-rekorların zamanları, $\Delta(r)$, r . karma-rekor zamanı olmak üzere,

$$\Delta(1) = 1, \Delta(2) = 2$$

ve $r = 3, 4, 5, \dots$ için,

$$\Delta(r) = \min\{i > \Delta(r-1) : X_i < X_{1:\Delta(r-1)} \text{ veya } X_i > X_{\Delta(r-1):\Delta(r-1)}\}$$

şeklinde tanımlanır. İlk n adet karma-rekor değer ise $X_1, X_2, X_{\Delta(3)}, \dots, X_{\Delta(n)}$ şeklinde gösterilir. $\square$

Yukarıdaki tanımdan hareketle karma-rekor değerler ile güncel rekor değerler arasındaki önceden bahsedilen yakın ilişki,

$$L'_{n-1} = \min(X_1, X_2, X_{\Delta(3)}, \dots, X_{\Delta(n)})$$

ve

$$U'_{n-1} = \max(X_1, X_2, X_{\Delta(3)}, \dots, X_{\Delta(n)})$$

şeklinde ifade edilebilir. Karma-rekorlar, aynı zamanda rekor kapsamını değiştiren değerler olarak da nitelenebilir. Gözlemlenen karma-rekorların en büyüğü ve en küçüğü ile oluşturulan aralık bize rekor kapsamını, bu iki değer arasındaki uzaklık ise rekor açıklığını vermektedir.

3.1 Karma-Rekor İndikatörü

Bir gözlemin karma-rekor olup olmadığını gösteren indikatör değişken, $\xi_1 = \xi_2 = 1$ ve $i = 3, 4, \dots$ için,

$$\xi_i = \begin{cases} 1, & X_i < X_{1:i-1} \text{ veya } X_i > X_{i-1:i-1} \\ 0, & \text{diğer} \end{cases}$$

şeklinde tanımlanır. Aşağıdaki teorem, sürekli, bağımsız ve aynı dağılımlı bir rasgele değişkenler dizisinde ξ_i 'lerin dağılım özelliklerini vermektedir:

Teorem 3.1

$X = \{X_i\}_{i \in \mathbb{N}^+}$, sayılabilir sonsuz, bağımsız, aynı dağılımlı, sürekli rasgele değişkenler dizisi olmak üzere, bu dizi üzerinde tanımlı karma-rekor indikatörleri için aşağıdakiler doğrudur: $i = 3, 4, 5, \dots$ için

- i) $\xi_i, p = \frac{2}{i}$ parametrelili Bernoulli dağılımına sahiptir.
- ii) $\xi_3, \xi_4, \dots$ 'ler bağımsızdır.

İspat:

i) $i = 3, 4, 5, \dots$ ve $j_i \in \{0, 1\}$ için $\{\xi_i = j_i\}$ olayı $(i-1)! 2^{j_i} (i-2)^{1-j_i}$ farklı şekilde meydana gelir. O halde, i . karma-rekor indikatörüne ait olasılık fonksiyonu, gözlem dizisi sürekli, bağımsız ve aynı dağılımlı olduğu için

$$P\{\xi_i = j_i\} = \frac{(i-1)! 2^{j_i} (i-2)^{1-j_i}}{i!} = \frac{2^{j_i} (i-2)^{1-j_i}}{i}$$

olacaktır. Böylece, Teorem 3.1'in birinci bölümü ispatlanmış olur.

ii) $j_i \in \{0,1\}$ olmak üzere $i = 3, 4, 5, \dots, k$ için $\{\xi_3 = j_3, \xi_4 = j_4, \dots, \xi_k = j_k\}$ olayı $2(2^{j_3}(3-2)^{1-j_3})(2^{j_4}(4-2)^{1-j_4}) \dots (2^{j_k}(k-2)^{1-j_k})$ farklı şekilde meydana gelebilir. Yine, sürekli, bağımsız ve aynı dağılımlı olma varsayımından hareketle, sonuç olarak, $\{\xi_3 = j_3, \xi_4 = j_4, \dots, \xi_k = j_k\}$ olayının olasılığı,

$$\begin{aligned} P\{\xi_3 = j_3, \xi_4 = j_4, \dots, \xi_k = j_k\} \\ &= \frac{2^{j_3+1}(3-2)^{1-j_3} 2^{j_4}(4-2)^{1-j_4} \dots 2^{j_k}(k-2)^{1-j_k}}{k!} \\ &= \frac{2^{1+j_3+j_4+\dots+j_k} \prod_{i=4}^k (i-2)^{1-j_i}}{k!} \end{aligned}$$

şeklinde elde edilir. Yukarıdaki sonuç, indikatörlerin bağımsız oldukları varsayımı altında dağılımlarının Bernoulli $\left(\frac{2}{i}\right)$ olduğu bilgisi kullanılarak elde edilecek sonuç ile aynıdır. Bu durumda Teorem 3.1(ii)'nin ispatı da tamamlanmış olur. $\square$

Karma-rekor indikatörü ve dağılımı verildikten sonra, yukarıdaki bilgilere dayalı olarak k . gözlemin n . karma-rekor olması olayı ele alınır ve bu olay, $A(k, n)$ ile gösterilirse, $1 \leq n \leq k$ için,

$$A(k, n) = \{\xi_1 = 1, \xi_2 = 1, \xi_3 + \xi_4 + \dots + \xi_{k-1} = n - 3, \xi_k = 1\}$$

yazılabilir. İlk $(k-1)$ gözlem içerisindeki Karma-rekor sayısı N'_{k-1} ile gösterilirse,

$$A(k, n) = \{N'_{k-1} = n - 1, \xi_k = 1\}$$

doğrudur. Yukarıdaki eşitlikte, N'_{k-1} ve ξ_k rasgele değişkenlerinin bağımsız oldukları açıktır. Bu durumda, Teorem 3.1 kullanılarak,

$$P(A(k, n)) = \frac{2}{k} P(N'_{k-1} = n - 1)$$

elde edilir.

Çalışmanın bir sonraki bölümü, yukarıdaki eşitliğin sağında yer alan, sonlu dizideki karma-rekor sayısı ve dağılımı konusuna ayrılmıştır.

3.2 Karma-Rekor Sayısı

Sürekli, bağımsız ve aynı dağılımlı rasgele değişkenler dizisinin ilk n elemanı içerisindeki karma-rekor sayısı N'_n ile gösterilirse, $N'_n = 2 + \xi_3 + \xi_4 + \dots + \xi_n$ şeklinde tanımlanabilir. N'_n 'nin olasılık fonksiyonuna geçmeden önce, onun ispatında kullanılacak birkaç bilgi Tanım 3.2.1, Lemma 3.2.1 ve Lemma 3.2.2'de verilmiştir.

Tanım 3.2.1

W , negatif olmayan tam sayı değerleri alan bir kesikli rasgele değişken olsun. Bu rasgele değişkenin olasılık üreten fonksiyonu; $p(w)$, W 'nin olasılık fonksiyonu olmak üzere,

$$G_W(z) = E(z^W) = \sum_{w=0}^{\infty} p(w)z^w$$

şeklinde tanımlanmaktadır (Johnson et al., 2005). $\square$

Lemma 3.2.1

W , negatif olmayan tam sayı değerleri alan bir rasgele değişken olsun. Bu rasgele değişkenin k gibi bir değere eşit olması olasılığı, onun olasılık üreten fonksiyonu kullanılarak,

$$p_W(k) = P(W = k) = \frac{1}{k!} \left. \frac{d^k G_W(z)}{dz^k} \right|_{z=0}$$

şeklinde elde edilir (Johnson et al., 2005). $\square$

Lemma 3.2.2

$W_1, W_2, \dots, W_n$ rasgele değişkenleri aynı dağılımlı olma şartı aranmamakla beraber birbirinden bağımsız, negatif olmayan tamsayı değerler alan rasgele değişkenler olsun. Bu rasgele değişkenlere bağlı olarak, a_i 'ler belli sabitler olmak üzere, $S_n = \sum_{i=1}^n a_i W_i$ şeklinde bir rasgele değişken tanımlansın. Bu rasgele değişkene ilişkin olasılık üreten fonksiyon,

$$G_{S_n}(z) = E(z^{S_n}) = G_{W_1}(z^{a_1})G_{W_2}(z^{a_2}) \dots G_{W_n}(z^{a_n})$$

şeklinde elde edilir (Johnson et al., 2005). $\square$

Bu bilgilerden sonra, N'_n 'nin olasılık fonksiyonu aşağıdaki gibi verilebilir:

Teorem 3.2.1

$X = \{X_i\}_{i \in N^+}$, sayılabilir sonsuz, bağımsız, aynı dağılımlı, sürekli rasgele değişkenler dizisi olmak üzere, $n \geq 3$ ve $k = 2, 4, \dots, n$ için N'_n rasgele değişkeninin olasılık fonksiyonu aşağıdaki şekildedir:

$$P(N'_n = k) \text{ olasılığı,}$$

$$\left[\frac{2}{n!} \prod_{i=1}^{n-2} (2z + i) \right]$$

ifadesindeki z^{k-2} 'nin katsayısıdır.

İspat: Lemma 3.2.2'ye göre $N'_n - 2$ rasgele değişkeninin olasılık üreten fonksiyonu $\xi_3, \xi_4, \dots, \xi_n$ şeklindeki $(n-2)$ tane bağımsız Bernoulli rasgele değişkeninin olasılık üreten fonksiyonlarının çarpımına eşittir. ξ_i rasgele değişkeninin olasılık üreten fonksiyonu, Tanım 3.2.1 dikkate alındığında,

$$G_{\xi_i}(z) = E(z^{\xi_i}) = \frac{2z + i - 2}{i}$$

olduđuna göre $N'_n - 2$ rasgele deđiřkeninin olasılık üreten fonksiyonu

$$G_{N'_n-2}(z) = E(z^{N'_n-2}) = \frac{2}{n!} \prod_{i=1}^{n-2} (2z + i)$$

řeklinde yazılabilir. Bu sonu ve Lemma 3.2.1 kullanılarak Teoremin ispatı tamamlanmıř olur. $\square$

Yukarıdaki teoremden hareketle N'_n rasgele deđiřkeninin sađkalım fonksiyonu izelge 3.2.1'de verilmiřtir.

izelge 3.2.1 eřitli n 'ler iin $\{P(N'_n \geq k)\}_{k \leq 15}$ olasılıkları

$k \backslash n$	10	20	30	40	50	100
2	1	1	1	1	1	1
3	0,9778	0,9947	0,9977	0,9987	0,9992	0,9998
4	0,857	0,9579	0,9796	0,9879	0,9919	0,9977
5	0,5966	0,8461	0,9161	0,9462	0,9621	0,9876
6	0,2999	0,6467	0,7813	0,848	0,8866	0,9566
7	0,1019	0,4106	0,5852	0,6874	0,7529	0,889
8	0,0219	0,2119	0,3762	0,4919	0,5748	0,7761
9	0,0027	0,0881	0,2052	0,3066	0,3888	0,6256
10	0,0001	0,0294	0,0947	0,1656	0,2313	0,4603
11	0	0,0079	0,037	0,0775	0,1209	0,3072
12	0	0,0017	0,0123	0,0315	0,0556	0,1856
13	0	0,0003	0,0035	0,0112	0,0225	0,1015
14	0	3,85E-05	0,0008	0,0035	0,0081	0,0504
15	0	4,02E-06	0,0002	0,0009	0,0026	0,0227

Ayrıca, N'_n rasgele deđiřkeninin beklenen deđeri, bir bařka deyiřle n 'lik bir dizide gerekleřmesi beklenen ortalama karma-rekor sayısı Teorem 3.1(i) kullanılarak,

$$E(N'_n) = E(\xi_1 + \xi_2 + \xi_3 + \xi_4 + \dots + \xi_n) = 2 + \sum_{i=3}^n E(\xi_i) = 2 + \sum_{i=3}^n \frac{2}{i}$$

řeklinde elde edilir. N'_n rasgele deđiřkeninin varyansı ise,

$$\begin{aligned}
V(N'_n) &= V(\xi_1 + \xi_2 + \dots + \xi_n) = \sum_{i=3}^n V(\xi_i) \\
&= \sum_{i=3}^n \frac{2}{i} \left(1 - \frac{2}{i}\right) = E(N'_n) - 2 - \sum_{i=3}^n \frac{4}{i^2}
\end{aligned}$$

olarak yazılabilir.

Diğer yandan, eğer N_n , birbirinden bağımsız ve aynı dağılımlı rasgele değişkenler dizisi $X = \{X_i\}_{i \in \mathbb{N}^+}$ 'deki üst rekor sayısını gösterirse, bu rasgele değişkenin beklenen değerinin, γ , Euler Sabiti olmak üzere ($\gamma = 0,5772 \dots$),

$$E(N_n) = \sum_{i=1}^n \frac{1}{i} \cong \ln(n) + \gamma$$

olduğu bilinmektedir (Arnold et al., 1992). Bu durumda, büyük n değerleri göz önüne alınırsa, N'_n rasgele değişkeninin beklenen değeri ve varyansı için de, sırasıyla,

$$E(N'_n) = 2E(N_n) - 1 \cong 2\ln(n) + 2\gamma - 1$$

ve

$$V(N'_n) \cong 2\ln(n) + 2\gamma + 2 - \frac{2}{3}\pi^2$$

eşitlikleri yazılabilir. Yukarıda verilen ortalama karma-rekor sayısı ve karma rekor sayısının varyansına ilişkin teorik sonuçları desteklemek için, Matlab paket programı ile yazılan bir fonksiyon kullanılarak (Bkz. Ek 1) çeşitli büyüklükte örneklemeler türetilmiş ve 100000 tekrar sonrası elde edilen simülasyon sonuçları Çizelge 3.2.2'de verilmiştir.

Çizelge 3.2.2 Uniform(0,1) dağılımlı kitleden türetilen n ölçümlü örneklemeler için ortalama karma-rekor sayıları ve karma-rekor sayılarının varyansları

n	Ortalama Karma Rekor Sayısı		Karma Rekor Sayısının Varyansı	
	Teorik	Simülasyon	Teorik	Simülasyon
10	4,8579	4,8638	1,6589	1,6567
20	6,1955	6,1984	2,8108	2,8056
30	6,9900	6,9870	3,5414	3,5524
40	7,5571	7,5653	4,0761	4,0801
50	7,9984	8,0152	4,4979	4,5274
100	9,3748	9,3742	5,8348	5,7938
500	12,5856	12,5951	9,0139	9,0386
1000	13,9709	13,9571	10,3952	10,3521

3.3 Karma-Rekor Zamanları

Tanım 3.1’de yer alan karma-rekor zamanları ile henüz ilgili literatürde yer almayan rekor kapsamının ve rekor açıklığının zamanları ilişkilidir. Örneğin, $\Delta(n+1)$ ile gösterilen $(n+1)$. karma-rekorun zamanı, n . rekor kapsamının ve n . rekor açıklığının zamanlarına karşılık gelmektedir. Dolayısıyla, bu bölümde bağımsız ve aynı dağılımlı sürekli rasgele değişkenler dizisi üzerinden verilecek sonuçlar, aynı zamanda, literatürde yer alan rekor kapsamı ve rekor açıklığı modellerini tamamlayıcı niteliktedir. Şimdi, aşağıdaki teoremle, karma rekor zamanlarının ortak olasılık fonksiyonunu verelim:

Teorem 3.3.1

$X = \{X_i\}_{i \in \mathbb{N}^+}$, sayılabilir sonsuz, bağımsız, aynı dağılımlı, sürekli rasgele değişkenler dizisi olmak üzere, $3 \leq k_3 < k_4 < \dots < k_n$ için, $\Delta(3), \Delta(4), \dots, \Delta(n)$ ’lerin ortak olasılık fonksiyonu,

$$P\{\Delta(3) = k_3, \Delta(4) = k_4, \dots, \Delta(n) = k_n\} = \frac{2^{n-1}}{k_n(k_n - 1) \prod_{i=3}^n (k_i - 2)}$$

şeklindedir.

İspat: Teorem 3.1'den hareketle,

$$\{\Delta(3) = k_3, \Delta(4) = k_4, \dots, \Delta(n) = k_n\} \\ \equiv \left\{ \begin{array}{l} \xi_3 = 0, \xi_4 = 0, \dots, \xi_{k_3-1} = 0, \xi_{k_3} = 1, \\ \xi_{k_3+1} = 0, \xi_{k_3+2} = 0, \dots, \xi_{k_4-1} = 0, \xi_{k_4} = 1, \\ \vdots \\ \xi_{k_{n-1}+1} = 0, \xi_{k_{n-1}+2} = 0, \dots, \xi_{k_n-1} = 0, \xi_{k_n} = 1 \end{array} \right\}$$

yazılabilir. Buradan hareketle,

$$\begin{aligned} P\{\Delta(3) = k_3, \Delta(4) = k_4, \dots, \Delta(n) = k_n\} \\ = \left(\frac{1}{3}\right)\left(\frac{2}{4}\right)\left(\frac{3}{5}\right) \dots \left(\frac{k_3-3}{k_3-1}\right)\left(\frac{2}{k_3}\right)\left(\frac{k_3-1}{k_3+1}\right)\left(\frac{k_3}{k_3+2}\right) \dots \left(\frac{2}{k_4}\right) \dots \left(\frac{k_n-3}{k_n-1}\right)\left(\frac{2}{k_n}\right) \\ = \frac{2 \cdot 2 \cdot \dots \cdot 2}{(k_3-2)(k_4-2) \dots (k_{n-1}-2)(k_n-2)(k_n-1)k_n} \\ = \frac{2^{n-1}}{k_n(k_n-1) \prod_{i=3}^n (k_i-2)} \end{aligned}$$

elde edilir ve teoremin ispatı tamamlanmış olur. $\square$

Karma-rekor zamanlarının ortak olasılık fonksiyonu verildikten sonra, şimdi de onların tekli ve çarpım momentleri incelensin. $n = 1, 2, 3, \dots$ için, n . karma-rekor zamanının beklenen değeri, yukarıdaki teoremden hareketle,

$$\begin{aligned} E(\Delta(n)) &= 2^{n-1} \sum_{k_3=3}^{\infty} \frac{1}{(k_3-2)} \sum_{k_4=k_3+1}^{\infty} \frac{1}{(k_4-2)} \\ &\dots \sum_{k_{n-1}=k_{n-2}+1}^{\infty} \frac{1}{(k_{n-1}-2)} \sum_{k_n=k_{n-1}+1}^{\infty} \frac{1}{(k_n-1)(k_n-2)} \end{aligned}$$

şeklinde yazılabilir. $\sum_{v=u+1}^{\infty} \frac{1}{(v-1)(v-2)} = \frac{1}{u-1}$ olduğuna göre, yukarıdaki beklenen değerin

$$E(\Delta(n)) = 2^{n-1}$$

olacağı kolayca görülebilir. Diğer yandan $\alpha \in \{2, 3, \dots\}$ için $\Delta(n)$ 'in α . momentinin olmadığı, $\sum_{v=u+1}^{\infty} \frac{v^{\alpha-1}}{(v-1)(v-2)} = \infty$ bilgisi ve yine Teorem 3.3.1 kullanılarak,

$$\begin{aligned} E(\Delta(n)^\alpha) &= 2^{n-1} \sum_{k_3=3}^{\infty} \frac{1}{(k_3-2)} \sum_{k_4=k_3+1}^{\infty} \frac{1}{(k_4-2)} \\ &\quad \dots \sum_{k_{n-1}=k_{n-2}+1}^{\infty} \frac{1}{(k_{n-1}-2)} \sum_{k_n=k_{n-1}+1}^{\infty} \frac{k_n^{\alpha-1}}{(k_n-1)(k_n-2)} = \infty \end{aligned}$$

şeklinde gösterilebilir. Benzer olarak, $n \neq m$ için

$$E(\Delta(n)\Delta(m)) = \infty$$

olacağı da açıktır.

Yukarıdaki sonuçlara ilave olarak, n . karma-rekor, k . gözlemde ortaya çıktıktan sonra, $(n+1)$. karma-rekor için ortalama bekleme süresi ise, yine Teorem 3.3.1'den hareketle,

$$E(\Delta(n+1)|\Delta(n) = k) = 2k$$

olarak bulunur.

3.4 Dağılımdan-Bağımsız Bir Güven Aralığı

Dağılımdan-bağımsız (distribution-free) güven aralıkları, ilk kez Bairamov ve Petunin (1991) tarafından tanımlanmış ve parametrik olmayan istatistik yöntemler içerisinde uygulama alanı bulmuştur. Bu bölümde, daha sonra verilecek Teorem 3.4.1’de dağılımdan-bağımsız olduğu gösterilecek olan güven aralığı

$$(\min(X_1, X_2, X_{\Delta(3)}, \dots, X_{\Delta(n+1)}), \max(X_1, X_2, X_{\Delta(3)}, \dots, X_{\Delta(n+1)}))$$

olarak seçilmiştir. Önceden tanımlandığı üzere, $\Delta(n+1)$ rasgele değişkeni, $X = \{X_i\}_{i \in \mathbb{N}^+}$ dizisindeki $(n+1)$. karma-rekor zamanını göstermektedir. Hatırlanacak olursa, $L'_n = \min(X_1, X_2, X_{\Delta(3)}, \dots, X_{\Delta(n+1)})$ ve $U'_n = \max(X_1, X_2, X_{\Delta(3)}, \dots, X_{\Delta(n+1)})$ ilişkisinin olduğu daha önce vurgulanmıştı. O zaman, herhangi bir yeni gözlemin n . rekor kapsamı içine düşmesi şeklinde tanımlanan olay, $i \in \{1, 2, 3, \dots\}$ için $B = \{X_{\Delta(n+1)+i} \in (L'_n, U'_n)\}$ ile gösterilebilir. Bağımsız ve aynı dağılımlı sürekli gözlemlere dayalı olarak bu olayın olasılığını vermeden önce, onun ispatını kolaylaştıran aşağıdaki lemmayı verelim:

Lemma 3.4.1

$\psi_a(b, c) = \int_a^1 x^b (\ln x)^c dx$ olmak üzere, $a \in [0, 1] \subset \mathbb{R}$ ve $b, c \in \{0, 1, 2, \dots\}$ için

$$\psi_a(b, c) = (-1)^c c! \left(\frac{1}{(b+1)^{c+1}} - a^{b+1} \sum_{j=0}^c \frac{(-\ln a)^{c-j}}{(b+1)^{1+j} (c-j)!} \right)$$

eşitliği doğrudur (Tanıl, 2009). $\square$

Teorem 3.4.1

$X = \{X_i\}_{i \in \mathbb{N}^+}$, sayılabilir sonsuz, bağımsız, aynı dağılımlı, sürekli rasgele değişkenler dizisi olmak üzere, yeni bir gözlemin n . rekor kapsamı içine düşmesi olasılığı;

$$P\{X_{\Delta(n+1)+i} \in (L'_n, U'_n)\} = 1 - \left(\frac{2}{3}\right)^n$$

dir. $\square$

İspat: i . yeni gözlem $X_{\Delta(n+1)+i}$ 'nin, n . güncel rekor ikilisi L'_n ve U'_n 'den bağımsız olduğu açıktır. Bu durumda, $B = \{X_{\Delta(n+1)+i} \in (L'_n, U'_n)\}$ olayının olasılığı,

$$P(B) = \int_{-\infty}^{\infty} \int_{-\infty}^u (F(u) - F(\ell)) f_{L'_n, U'_n}(\ell, u) d\ell du$$

şeklinde yazılabilir. Sonrasında, L'_n ve U'_n rasgele değişkenlerinin Bölüm 2.3'de verilen ortak olasılık yoğunluk fonksiyonu yukarıdaki eşitlikte yerine yazılırsa, bu olasılık aşağıdaki şekile gelir:

$$\begin{aligned} P(B) &= \int_0^1 \int_0^y \frac{(-1)^{n-1} 2^n}{(n-1)!} (y-x) (\ln(1-y+x))^{n-1} dx dy \\ &= \frac{(-1)^{(n-1)} 2^n}{(n-1)!} \left[\int_0^1 \int_0^y y (\ln(1-y+x))^{n-1} dx dy \right. \\ &\quad \left. - \int_0^1 \int_0^y x (\ln(1-y+x))^{(n-1)} dx dy \right] \end{aligned}$$

Gösterim kolaylığı için,

$$\pi_1 = \int_0^1 \int_0^y y (\ln(1-y+x))^{n-1} dx dy$$

ve

$$\pi_2 = \int_0^1 \int_0^y x(\ln(1-y+x))^{n-1} dx dy$$

olsun. Lemma 3.4.1 dikkate alındığında,

$$\pi_1 = \int_0^1 y\psi_{1-y}(0, n-1) dy$$

ve

$$\pi_2 = \int_0^1 \psi_{1-y}(1, n-1) dy - \int_0^1 (1-y)\psi_{1-y}(0, n-1) dy$$

şeklinde elde edilir. Bu durumda,

$$\pi_1 - \pi_2 = \int_0^1 \psi_{1-y}(0, n-1) dy - \int_0^1 \psi_{1-y}(1, n-1) dy$$

olur. Böylece, yine Lemma 3.4.1 kullanılarak

$$\pi_1 - \pi_2 = (-1)^{n-1}(n-1)! \left[\frac{1}{2^n} - \frac{1}{3^n} \right]$$

şeklinde elde edilmiş olur. Sonuç olarak,

$$\begin{aligned} P(B) &= \frac{(-1)^{n-1}2^n}{(n-1)!} [\pi_1 - \pi_2] \\ &= \frac{(-1)^{n-1}2^n}{(n-1)!} \left[(-1)^{n-1}(n-1)! \left[\frac{1}{2^n} - \frac{1}{3^n} \right] \right] = 1 - \left(\frac{2}{3} \right)^n \end{aligned}$$

bulunur ve teoremin ispatı tamamlanır. $\square$

4. SIRALI-KARMA-REKOR DEĞERLER

Bu bölümde, Tanım 3.1'den hareketle $Y_i := X_{\Delta(i)}$ olmak üzere $Y_{1:n} < Y_{2:n} < \dots < Y_{n:n}$ şeklinde tanımlanan sürekli rasgele değişkenlere “Sıralı-Karma-Rekor Değerler” adı verilmiştir. Daha önce bahsedildiği üzere, $(Y_{1:n}, Y_{n:n})$ aralığı, $(n - 1)$. rekor kapsamına karşılık gelmektedir. Genel olarak, $Y_{i:n}$ 'in, rekor kapsamını veya rekor açıklığını değiştiren i . en küçük değer olduğu söylenebilir. Diğer yandan, sıralı-karma-rekorların nerelerde kullanılabileceğine dair örnekler vermek mümkündür. Örneğin, kitle ortalamasının belirli bir değerden büyük olduğu iddiasını kontrol etmek için, kaydedilen karma-rekorlardan kaçının bu değerden büyük olduğuna bakılabilir. Çünkü iddia gerçekten doğruysa, bu değerın sağındaki karma-rekor sayısının büyük olması beklenir. İlk n karma-rekor arasından kaçının μ_0 'dan büyük olduğu $K_n(\mu_0)$ ile gösterilirse, bu rasgele değişkenle, sıralı-karma-rekorlar arasındaki ilişki; $k \in \{0, 1, 2, \dots, n\}$, $Y_{0:n} := -\infty$ ve $Y_{n+1:n} := \infty$ için,

$$\{K_n(\mu_0) = k\} \equiv \{Y_{n-k:n} < \mu_0 < Y_{n-k+1:n}\}$$

şeklinde ifade edilir. Bu ilişkiden hareketle, $\{K_n(\mu_0) = k\}$ olayının olasılığını hesaplayabilmek için, sıralı-karma rekorların ortak dağılımına ihtiyaç olduğu açıktır. Yani, bu olasılık,

$$\begin{aligned} P\{K_n(\mu_0) = k\} &= P\{Y_{n-k:n} < \mu_0 < Y_{n-k+1:n}\} \\ &= \int_{-\infty}^{\mu_0} \int_{y_1}^{\mu_0} \dots \int_{y_{n-k-1}}^{\mu_0} \int_{\mu_0}^{\infty} \int_{y_{n-k+1}}^{\infty} \int_{y_{n-k+2}}^{\infty} \dots \int_{y_{n-1}}^{\infty} f^{Y_{1:n}, Y_{2:n}, \dots, Y_{n:n}}(y_1, y_2, \dots, y_n) dy_n dy_{n-1} \dots dy_1 \end{aligned}$$

şeklinde elde edilebilir.

Bu bölümde ilk olarak, $X = \{X_i\}_{i \in N^+}$ dizisi; sürekli, birbirinden bağımsız ve aynı $F(x)$ dağılım ve $f(x)$ olasılık yoğunluk fonksiyonuna sahip olduğunda sıralı-karma-rekor değerlerin ortak olasılık yoğunluk fonksiyonu için kullanılabilecek bir indirgeme bağıntısı aşağıdaki teorem ile verilmektedir.

Teorem 4.1

Sayılabilir sonsuz, bağımsız, aynı dağılımlı, sürekli rasgele değişkenler dizisi $X = \{X_i\}_{i \in N^+}$ 'de tanımlı, ilk n adet sıralı-karma-rekor değerlerin ortak olasılık yoğunluk fonksiyonu için aşağıdaki eşitlik doğrudur:

$$\begin{aligned} f^{Y_{1:n}, Y_{2:n}, \dots, Y_{n:n}}(y_1, y_2, \dots, y_n) \\ = \frac{f(y_1)}{1 - F(y_n) + F(y_2)} f^{Y_{1:n-1}, Y_{2:n-1}, \dots, Y_{n-1:n-1}}(y_2, y_3, \dots, y_n) \\ + \frac{f(y_n)}{1 - F(y_{n-1}) + F(y_1)} f^{Y_{1:n-1}, Y_{2:n-1}, \dots, Y_{n-1:n-1}}(y_1, y_2, \dots, y_{n-1}) \end{aligned}$$

$$y_1 < y_2 < \dots < y_n$$

İspat: $I_i := \left[y_i - \frac{dy_i}{2}, y_i + \frac{dy_i}{2} \right)$ ve $y_1 < y_2 < \dots < y_n$ olmak üzere, $I_1, I_2, \dots, I_n$ aralıklarını dikkate alalım. dy_i 'ler yeterince küçük pozitif reel sayılar olmak üzere, $Y_{1:n} \in I_1, \dots, Y_{n:n} \in I_n$ olayının olasılığı için,

$$P\{Y_{1:n} \in I_1, \dots, Y_{n:n} \in I_n\} \cong f^{Y_{1:n}, Y_{2:n}, \dots, Y_{n:n}}(y_1, y_2, \dots, y_n) dy_1 dy_2 \dots dy_n$$

yaklaşık eşitliği geçerlidir. Diğer yandan,

$$A_1 = \{X_{\Delta(n)} < Y_{1:n-1}\} \text{ ve } A_2 = \{X_{\Delta(n)} > Y_{n-1:n-1}\} \text{ olmak üzere,}$$

n . karma-rekorun ortaya çıkması kesin olayı $A_1 \cup A_2$ şeklinde ifade edilir. Bu ayırık iki olaya dayalı olarak $\{Y_{1:n} \in I_1, \dots, Y_{n:n} \in I_n\}$ olayının olasılığı için,

$$\begin{aligned} P\{Y_{1:n} \in I_1, \dots, Y_{n:n} \in I_n\} \\ = P\{X_{\Delta(n)} \in I_1, Y_{1:n-1} \in I_2, \dots, Y_{n-1:n-1} \in I_n\} \\ + P\{Y_{1:n-1} \in I_1, \dots, Y_{n-1:n-1} \in I_{n-1}, X_{\Delta(n)} \in I_n\} \end{aligned}$$

eşitliği yazılabilir. Benzer şekilde, yine dy_i 'ler yeterince küçük pozitif reel sayılar olmak üzere, bu eşitliğin sağındaki olasılıklar için,

$$\begin{aligned}
P\{X_{\Delta(n)} \in I_1, Y_{1:n-1} \in I_2, \dots, Y_{n-1:n-1} \in I_n\} \\
\cong \int^{X_{\Delta(n)}, Y_{1:n-1}, \dots, Y_{n-1:n-1}}(y_1, y_2, \dots, y_n) dy_1 dy_2 \dots dy_n
\end{aligned}$$

ve

$$\begin{aligned}
P\{Y_{1:n-1} \in I_1, \dots, Y_{n-1:n-1} \in I_{n-1}, X_{\Delta(n)} \in I_n\} \\
\cong \int^{Y_{1:n-1}, \dots, Y_{n-1:n-1}, X_{\Delta(n)}}(y_1, y_2, \dots, y_n) dy_1 dy_2 \dots dy_n
\end{aligned}$$

yaklaşık eşitlikleri geçerlidir. Bu eşitliklerin sağ tarafları dikkate alınırsa,

$$\begin{aligned}
& f^{Y_{1:n}, Y_{2:n}, \dots, Y_{n:n}}(y_1, y_2, \dots, y_n) \\
&= f^{X_{\Delta(n)}, Y_{1:n-1}, \dots, Y_{n-1:n-1}}(y_1, y_2, \dots, y_n) \\
&+ f^{Y_{1:n-1}, \dots, Y_{n-1:n-1}, X_{\Delta(n)}}(y_1, y_2, \dots, y_n)
\end{aligned}$$

olacağı açıktır. Teoremin ispatını tamamlamak için bu son eşitliğin sağındaki iki olasılık yoğunluk fonksiyonunun, $f^{Y_{1:n-1}, \dots, Y_{n-1:n-1}}$ cinsinden karşılıklarını bulmamız yeterlidir. İlk terim için,

$$\begin{aligned}
& f^{X_{\Delta(n)}, Y_{1:n-1}, \dots, Y_{n-1:n-1}}(y_1, y_2, \dots, y_n) \\
&= f^{Y_{1:n-1}, \dots, Y_{n-1:n-1}}(y_2, \dots, y_n) f^{X_{\Delta(n)}|Y_{1:n-1}, \dots, Y_{n-1:n-1}}(y_1|y_2, \dots, y_n)
\end{aligned}$$

yazılabilir. Bu ifade içindeki koşullu olasılık yoğunluk fonksiyonunu elde etmek için, $P\{X_{\Delta(n)} \in I_1 | Y_{1:n-1} = y_2, \dots, Y_{n-1:n-1} = y_n\}$ koşullu olasılığını ele alalım. Yeterince küçük pozitif dy_1 için,

$$\begin{aligned}
& P\{X_{\Delta(n)} \in I_1 | Y_{1:n-1} = y_2, \dots, Y_{n-1:n-1} = y_n\} \\
&\cong \int^{X_{\Delta(n)}|Y_{1:n-1}, \dots, Y_{n-1:n-1}}(y_1|y_2, \dots, y_n) dy_1
\end{aligned}$$

yaklaşık eşitliği geçerlidir.

Diğer yandan,

$$\begin{aligned}
P\{X_{\Delta(n)} \in I_1 | Y_{1:n-1} = y_2, \dots, Y_{n-1:n-1} = y_n\} \\
&= \sum_{u=1}^{\infty} P\{X_{\Delta(n-1)+1} \in (y_2, y_n), X_{\Delta(n-1)+2} \\
&\quad \in (y_2, y_n), \dots, X_{\Delta(n-1)+u-1} \in (y_2, y_n), X_u \in I_1\} \\
&= \sum_{u=1}^{\infty} (F(y_n) - F(y_2))^{u-1} P\{X_u \in I_1\} \\
&= P\{X_1 \in I_1\} \sum_{u=1}^{\infty} (F(y_n) - F(y_2))^{u-1} = \frac{P\{X_1 \in I_1\}}{1 - F(y_n) + F(y_2)}
\end{aligned}$$

ve yine yeterince küçük pozitif dy_1 için,

$$P\{X_{\Delta(n)} \in I_1 | Y_{1:n-1} = y_2, \dots, Y_{n-1:n-1} = y_n\} \cong \frac{f(y_1)dy_1}{1 - F(y_n) + F(y_2)}$$

yazılabilir. Bu iki yaklaşık eşitlik dikkate alındığında,

$$f^{X_{\Delta(n)} | Y_{1:n-1}, \dots, Y_{n-1:n-1}}(y_1 | y_2, \dots, y_n) = \frac{f(y_1)}{1 - F(y_n) + F(y_2)}$$

olacağı açıktır. Bu durumda,

$$\begin{aligned}
f^{X_{\Delta(n)}, Y_{1:n-1}, \dots, Y_{n-1:n-1}}(y_1, y_2, \dots, y_n) \\
&= f^{Y_{1:n-1}, \dots, Y_{n-1:n-1}}(y_2, \dots, y_n) \frac{f(y_1)}{1 - F(y_n) + F(y_2)}
\end{aligned}$$

elde edilir. Benzer şekilde düşünerek,

$$\begin{aligned}
f^{Y_{1:n-1}, \dots, Y_{n-1:n-1}, X_{\Delta(n)}}(y_1, y_2, \dots, y_n) \\
&= f^{Y_{1:n-1}, \dots, Y_{n-1:n-1}}(y_1, \dots, y_{n-1}) \frac{f(y_n)}{1 - F(y_{n-1}) + F(y_1)}
\end{aligned}$$

olduğu gösterilebilir. Sonuç olarak, bu son iki sonucu yerlerine yazarsak teoremin doğru olduğu görülecektir. $\square$

Yukarıda verilen indirgeme bağıntısından sonra, ilk n adet karma-rekor değerlerin ortak olasılık yoğunluk fonksiyonu, kapalı formda aşağıdaki teoremden gibidir.

Teorem 4.2

Sayılabilir sonsuz, bağımsız, aynı dağılımlı, sürekli rasgele değişkenler dizisi $X = \{X_i\}_{i \in \mathbb{N}^+}$ 'de tanımlı, ilk n adet sıralı-karma-rekor değerlerin $(Y_{1:n}, Y_{2:n}, \dots, Y_{n:n})$ ortak olasılık yoğunluk fonksiyonu,

$$f^{Y_{1:n}, Y_{2:n}, \dots, Y_{n:n}}(y_1, y_2, \dots, y_n) \\ = 2 \prod_{i=1}^n f(y_i) \sum_{i_1=n-1}^n \sum_{i_2=i_1-1}^{i_1} \dots \sum_{i_{n-2}=i_{n-3}-1}^{i_{n-3}} \prod_{j=1}^{n-2} \frac{1}{[1 - F(y_{i_j}) + F(y_{i_j-n+j+1})]} \\ y_1 < y_2 < \dots < y_n$$

şeklindedir.

İspat: Yukarıdaki teoremin ispatı matematiksel tümevarım ile yapılabilir. Buna göre, $n = 3$ için, Teorem 4.1 kullanılarak,

$$f^{Y_{1:3}, Y_{2:3}, Y_{3:3}}(y_1, y_2, y_3) \\ = \frac{f(y_1)}{1 - F(y_3) + F(y_2)} f^{Y_{1:2}, Y_{2:2}}(y_2, y_3) \\ + \frac{f(y_3)}{1 - F(y_2) + F(y_1)} f^{Y_{1:2}, Y_{2:2}}(y_1, y_2)$$

yazılabilir. $Y_{1:2} := X_{1:2}$ ve $Y_{2:2} := X_{2:2}$ olduğundan, ilk iki sıralı-karma-rekor değerlerin ortak olasılık yoğunluk fonksiyonunun

$$f^{Y_{1:2}, Y_{2:2}}(y_1, y_2) = 2f(y_1)f(y_2)$$

olacağı açıktır.

Bu durumda,

$$\begin{aligned}
 f^{Y_{1:3}, Y_{2:3}, Y_{3:3}}(y_1, y_2, y_3) \\
 &= \frac{f(y_1)}{1 - F(y_3) + F(y_2)} 2f(y_2)f(y_3) \\
 &\quad + \frac{f(y_3)}{1 - F(y_2) + F(y_1)} 2f(y_1)f(y_2)
 \end{aligned}$$

olur. O halde, ilk üç sıralı-karma-rekor değerlerin ortak olasılık yoğunluk fonksiyonu,

$$\begin{aligned}
 f^{Y_{1:3}, Y_{2:3}, Y_{3:3}}(y_1, y_2, y_3) \\
 = 2f(y_1)f(y_2)f(y_3) \left[\frac{1}{1 - F(y_3) + F(y_2)} + \frac{1}{1 - F(y_2) + F(y_1)} \right]
 \end{aligned}$$

şeklinde elde edilir ve $n = 3$ için teoremin doğru olduğu gösterilmiş olur. Şimdi, $n = r$ için teoremin doğru olduğu kabul edilsin. Yani,

$$\begin{aligned}
 f^{Y_{1:r}, Y_{2:r}, \dots, Y_{r:r}}(y_1, y_2, \dots, y_r) \\
 = 2 \prod_{i=1}^r f(y_i) \sum_{i_1=r-1}^r \sum_{i_2=i_1-1}^{i_1} \dots \sum_{i_{r-2}=i_{r-3}-1}^{i_{r-3}} \prod_{j=1}^{r-2} \frac{1}{[1 - F(y_{i_j}) + F(y_{i_j-r+j+1})]}
 \end{aligned}$$

$$y_1 < y_2 < \dots < y_r$$

doğru olsun. O halde, $n = r + 1$ için de teoremin doğru olduğunu gösterelim:

Teorem 4.1 kullanılarak,

$$\begin{aligned}
 f^{Y_{1:r+1}, Y_{2:r+1}, \dots, Y_{r+1:r+1}}(y_1, y_2, \dots, y_{r+1}) \\
 = \frac{f(y_1)}{1 - F(y_{r+1}) + F(y_2)} f^{Y_{1:r}, Y_{2:r}, \dots, Y_{r:r}}(y_2, y_3, \dots, y_{r+1}) \\
 + \frac{f(y_{r+1})}{1 - F(y_r) + F(y_1)} f^{Y_{1:r}, Y_{2:r}, \dots, Y_{r:r}}(y_1, y_2, \dots, y_r)
 \end{aligned}$$

olacağı açıktır.

Bu durumda,

$$\begin{aligned}
& f^{Y_{1:r+1}, Y_{2:r+1}, \dots, Y_{r+1:r+1}}(y_1, y_2, \dots, y_{r+1}) \\
&= \frac{f(y_1)}{1 - F(y_{r+1}) + F(y_2)} 2 \prod_{i=1}^r f(y_{i+1}) \\
&\times \sum_{i_1=r-1}^r \sum_{i_2=i_1-1}^{i_1} \dots \sum_{i_{r-2}=i_{r-3}-1}^{i_{r-3}} \prod_{j=1}^{r-2} \frac{1}{[1 - F(y_{i_j+1}) + F(y_{i_j-r+j+2})]} \\
&+ \frac{f(y_{r+1})}{1 - F(y_r) + F(y_1)} 2 \prod_{i=1}^r f(y_i) \\
&\times \sum_{i_1=r-1}^r \sum_{i_2=i_1-1}^{i_1} \dots \sum_{i_{r-2}=i_{r-3}-1}^{i_{r-3}} \prod_{j=1}^{r-2} \frac{1}{[1 - F(y_{i_j}) + F(y_{i_j-r+j+1})]}
\end{aligned}$$

olur. Buna göre, yukarıdaki kesirli ifade toplamın içine dahil edildiğinde,

$$\begin{aligned}
& f^{Y_{1:r+1}, Y_{2:r+1}, \dots, Y_{r+1:r+1}}(y_1, y_2, \dots, y_{r+1}) = 2 \prod_{i=1}^{r+1} f(y_i) \\
&\times \left[\sum_{i_0=r+1}^{r+1} \sum_{i_1=i_0-1}^{i_0} \sum_{i_2=i_1-1}^{i_1} \dots \sum_{i_{r-2}=i_{r-3}-1}^{i_{r-3}} \prod_{j=0}^{r-2} \frac{1}{[1 - F(y_{i_j}) + F(y_{i_j-r+j+1})]} \right. \\
&\left. + \sum_{i_0=r}^r \sum_{i_1=i_0-1}^{i_0} \sum_{i_2=i_1-1}^{i_1} \dots \sum_{i_{r-2}=i_{r-3}-1}^{i_{r-3}} \prod_{j=0}^{r-2} \frac{1}{[1 - F(y_{i_j}) + F(y_{i_j-r+j+1})]} \right]
\end{aligned}$$

olacaktır. Bu durumda,

$$\begin{aligned}
& f^{Y_{1:r+1}, Y_{2:r+1}, \dots, Y_{r+1:r+1}}(y_1, y_2, \dots, y_{r+1}) \\
&= 2 \prod_{i=1}^{r+1} f(y_i) \sum_{i_0=r}^{r+1} \sum_{i_1=i_0-1}^{i_0} \sum_{i_2=i_1-1}^{i_1} \dots \sum_{i_{r-2}=i_{r-3}-1}^{i_{r-3}} \prod_{j=0}^{r-2} \frac{1}{[1 - F(y_{i_j}) + F(y_{i_j-r+j+1})]}
\end{aligned}$$

olur.

Sonuç olarak,

$$f^{Y_{1:r+1}, Y_{2:r+1}, \dots, Y_{r+1:r+1}}(y_1, y_2, \dots, y_{r+1})$$

$$= 2 \prod_{i=1}^{r+1} f(y_i) \sum_{i_1=r}^{r+1} \sum_{i_2=i_1-1}^{i_1} \dots \sum_{i_{r-1}=i_{r-2}-1}^{i_{r-2}} \prod_{j=1}^{r-1} \frac{1}{[1 - F(y_{i_j}) + F(y_{i_{j-r+j+1}})]}$$

$$y_1 < y_2 < \dots < y_{r+1}$$

olacaktır. Bu sonuç, teoremden $n = r + 1$ için elde edilecek sonuçla aynı olduğu için teoremin ispatı tamamlanmış olur. $\square$

5. SONUÇ VE TARTIŞMA

Çalışmamızda birbirinden bağımsız ve aynı sürekli $F(x)$ dağılımına sahip rasgele değişkenler dizisinde, başlangıçtan itibaren alt ve üst ayrımı yapılmaksızın kaydedilen her bir rekor değere “Karma-Rekor” adı verilmiştir. Ortaya atılan bu yeni kavram literatürdeki Güncel Rekor Değerler, Rekor Kapsamı ve Rekor Açıklığı gibi konularla oldukça ilişkilidir. Tez çalışması içerisinde bu ilişkiden bahsedilmiş ve sonrasında karma-rekor kavramına ilişkin bazı dağılımsal sonuçlar elde edilmiştir. İlk olarak ilgili literatürde var olan üst rekor indikatörüne benzer şekilde “Karma-Rekor İndikatörü” tanımlanmış ve dağılımı elde edilmiştir. Rekorların sadece artan veya sadece azalan olarak ele alındığı klasik rekor değerler kavramına göre i . rekor indikatörünün dağılımı Bernoulli($\frac{1}{i}$) iken i . karma-rekor indikatörünün dağılımı ise Bernoulli($\frac{2}{i}$) şeklinde bulunmuştur. Bu sonuç, karma-rekor değer kavramında, rekorların alt veya üst olduğuna bakılmaksızın iki taraftan gerçekleşmesi durumuyla ilgilenildiği için beklenen bir sonuçtur.

Çalışma kapsamında ele alınan bir diğer karma-rekor istatistiği karma-rekor sayısıdır. Tanımlanan karma-rekor indikatörü kullanılarak n hacimli sonlu sayıda bir rasgele değişkenler dizisindeki karma-rekor sayısı tanımlanmış ve bu rasgele değişkenin olasılık fonksiyonu elde edilmiştir. Sonrasında bu sonlu dizide ortalama karma-rekor sayısı ve karma-rekor sayısının varyansı elde edilmiştir. Bulunan bu ortalama değer, klasik rekor değerler için literatürde var olan beklenen üst rekor sayısının yaklaşık olarak iki katıdır ve doğal olarak bu sonuç karma-rekor değer kavramı düşünüldüğünde beklenen bir sonuçtur.

Karma-rekor kavramı başlığı altında elde edilen bir başka istatistik ise karma-rekor zamanlarıdır. Karma-rekor kavramı tanımında yer alan karma-rekor zamanlarının ortak olasılık fonksiyonu karma-rekor indikatörü kullanılarak elde edilmiştir. Sonrasında, n . karma-rekor değerinin ortalama olarak kaçınıcı gözlemde meydana geleceği bulunmuştur. Bu sonuç literatürde var olan klasik rekor zamanlarının beklenen değeri ile kıyaslandığında oldukça önemli bir sonuç olarak karşımıza çıkmaktadır. Çünkü, n . üst rekor değerinin ortalama olarak meydana

gelme zamanı ∞ 'a giderken, n . karma-rekor değerin beklenen zamanı 2^{n-1} olarak bulunmuştur. Bu sonuç bize rekorları iki taraflı olarak ele alırsak ortalama bir meydana geliş zamanından söz etmenin mümkün hale geldiğini göstermektedir. Bunun yanı sıra, n . karma-rekor zamanının varyansı için benzer bir durumdan söz etmek maalesef mümkün olamamıştır. n . karma-rekor zamanının varyansı, n . üst rekor zamanının varyansı ile aynı şekilde ∞ 'a gitmektedir. Ayrıca, karma-rekorlar ile rekor kapsamı ve rekor açıklığı arasındaki ilişkiden dolayı, bu sonuçlar, bu modeller için de geçerli ve onları tamamlayıcı niteliktedir.

Karma-rekor kavramı başlığı altında elde edilen son sonuç ise dağılımdan bağımsız bir güven aralığıdır. Bu alt başlık altında, herhangi bir yeni gözlemin n . rekor kapsamı içine düşmesi olasılığı $1 - \left(\frac{2}{3}\right)^n$ olarak bulunmuştur.

Dördüncü bölümde, ilk n adet karma-rekor değere dayalı bir sıra istatistikleri modeli ortaya konmuştur. Bu modelde tanımlanan sıra istatistiklerine “Sıralı-Karma-Rekor Değerler” adı verilmiştir. Sıralı-karma-rekorların ortak olasılık yoğunluk fonksiyonu için ilk olarak bir indirgeme bağıntısı verilmiştir. Bu indirgeme bağıntısında, ilk n adet sıralı-karma-rekor değerin ortak olasılık yoğunluk fonksiyonunun, ilk $(n - 1)$ adet sıralı-karma-rekor değerin ortak olasılık yoğunluk fonksiyonu kullanılarak elde edilebileceği gösterilmiştir. Sonrasında, bulunan indirgeme bağıntısı kullanılarak, ilk n adet sıralı-karma-rekor değerin ortak olasılık yoğunluk fonksiyonu doğrudan yazılmıştır. Bulunan bu sonuç, tümevarım yoluyla ispatlanmıştır.

Sonuç olarak, Chandler (1952) tarafından ortaya atılan üst rekor ve alt rekor modelleri, Houchens (1984) tarafından kısmen birleştirilmiş, bu tezin dördüncü bölümünde önerilen sıralı-karma-rekor modeliyle ise bu birleştirme işlemi tamamlanmıştır. Ayrıca, sıralı-karma-rekor modelinin içerdiği bilgi, başlangıçtan itibaren kaydedilen tüm alt ve üst rekorlara dayalı olması sebebiyle, bir sistemin performans artış ve azalışlarına (başarı ve başarısızlıklarına) ilişkin özgeçmişini içermesi açısından önemlidir. Bu bilgiye dayalı yapılacak değerlendirmeler, Houchens'in modeline dayalı yapılacak değerlendirmelere göre daha ayrıntılı ve gerçekçi olacaktır, çünkü, Houchens'in modeli sadece en son başarı ve

başarısızlık bilgisini içermektedir. Bu sebeplerden dolayı, bu tez çalışmasının, klasik rekor değerler kullanılarak yapılan pek çok çalışmada olduğu gibi, yeni konuları ve farklı bakış açılarını ortaya çıkarması beklenmektedir.

KAYNAKLAR DİZİNİ

- Ahmadi, J. and Balakrishnan, N.**, 2004, Confidence intervals for quantiles in terms of record range, *Statistics & Probability Letters*, 68:395-405pp.
- Ahmadi, J. and Balakrishnan, N.**, 2005a, Distribution-free confidence intervals for quantile intervals based on current records, *Statistics and Probability Letters*, 75:190-202pp.
- Ahmadi, J. and Balakrishnan, N.**, 2005b, Preservation of some reliability properties by certain record statistics, *Statistics*, 39(4):247-354pp.
- Ahmadi, J. and Balakrishnan, N.**, 2011, Distribution-free prediction intervals for order statistics based on record coverage, *Journal of the Korean Statistical Society*, 40:181-192pp.
- Ahmadi, J., Razmkhah, M. and Balakrishnan, N.**, 2009, Current k-records and their use in distribution-free confidence intervals, *Statistics and Probability Letters*, 79:29-37pp.
- Ahsanullah, M.**, 2004, Record Values – Theory and Applications, University Press of America Inc, 314p.
- Ahsanullah, M. and Nevzorov, V.B.**, 2005, Order Statistics: Examples and Exercises, Nova Science Publishers, New York, 236p.
- Ahsanullah, M. and Raqab, M.Z.**, 2006, Bounds and Characterizations of Record Statistics, Nova Science Publishers, New York, 177p.
- Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N.**, 1992, A First Course in Order Statistics, John Wiley & Sons, New York, 279p.
- Bairamov, I.G. and Petunin, Yu.I.**, 1991, Structure of invariant confidence intervals containing the main distributed mass, *Theor. Probab. Appl.*, 35(1):15-26pp.
- Balakrishnan, N. and Cohen, A.C.**, 1991, Order Statistics and Inference: Estimation Methods, Academic, Boston, 392p.
- Chandler, K.N.**, 1952, The distribution and frequency of record values, *J.R. Statist. Soc. Ser. B*, 14:220-228pp.

KAYNAKLAR DİZİNİ (devam)

- David, H.A.**, 1970, Order Statistics, John Wiley and Sons, New York, 272p.
- Galambos, J.**, 1978, The Asymptotic Theory of Extreme Order Statistics, Wiley, New York, 368p.
- Glick, N.**, 1978, Breaking records and breaking boards, *The American Mathematical Monthly*, 85(1):2-26pp.
- Harter, H.L.**, 1970, Order Statistics and Their Use in Testing and Estimation (Vol 1 and 2), U.S. Government Printing Office, Washington, D.C., 1566p.
- Houchens, R.L.**, 1984, Record Value Theory and Inference, Ph.D. Thesis, University of California, Riverside.
- Johnson, N.L., Kemp, A.W. and Kotz, S.**, 2005, Univariate Discrete Distributions, John Wiley and Sons, Inc, Hoboken, New Jersey, 646p.
- Kotz, S. and Nadarajah, S.**, 2000, Extreme Value Distributions, Imperial College Press, London, 185p.
- Nagaraja, H.N.**, 1988, Record values and related statistics – a review, *Communications in Statistics – Theory and Methods*, 17(7):2223-2238pp.
- Nevzorov, V.B.**, 2001, Records: Mathematical Theory, Translations of Mathematical Monographs, Volume 194, American Mathematical Society, Providence, Rhode Island, 164p.
- Renyi, A.**, 1962, On the extreme elements of observations, *MTA III Oszt. Közl.*, 12:105-121pp.
- Resnick, S.I.**, 1973, Record values and maxima, *Ann. Probab.*, 1(4):650-662pp.
- Sarhan, A.E. and Greenberg, B.G.**, 1962, Contributions to Order Statistics, Wiley, New York, 482p.
- Shorrock, R.**, 1972, On record values and record times, *J. Appl. Probability*, 9:316-326pp.
- Shorrock, R.**, 1973, Record values and inter-record times, *J. Appl. Probability*, 10:543-555pp.

KAYNAKLAR DİZİNİ (devam)

- Tanil, H.**, 2009, An order statistics model based on the list of top m scores after ℓ th change, *Journal of Statistical Planning and Inference*, 139:2189-219pp.
- Wilks, S.S.**, 1959, Recurrence of extreme observations, *J. Amer. Math. Soc.*, 1(1):106-112pp.

ÖZGEÇMİŞ

1984 yılında İzmir’de doğan Agâh KOZAN, ilk, orta ve lise öğrenimini İzmir’de tamamlamıştır. 2003 yılında Ege Üniversitesi Fen Fakültesi İstatistik Bölümü’ne girmeye hak kazanmış ve 2008 yılında lisans öğrenimini bitirmiştir. Ağustos 2010’da Ege Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı’nda Yüksek Lisans öğrenimini bitirmiş ve Eylül 2010’da aynı yerde Doktora öğrenimine başlamıştır. Aralık 2008’den bu yana Ege Üniversitesi Fen Fakültesi İstatistik Bölümü’nde araştırma görevlisi olarak görev yapmaktadır.

EKLER

Ek 1 Karma-Rekor Sayısının Beklenen Deęeri ve Varyansı iin Simlasyon
Matlab Kodları

Ek 1

Karma-Rekor Sayısının Beklenen Değeri ve Varyansı için Simülasyon Matlab Kodları

Uniform(0,1) dağılımından türetilen veriler kullanılarak elde edilen Karma-rekor sayısının beklenen değeri ve varyansını teorik sonuçlar ile birlikte veren MATLAB fonksiyonu aşağıdaki gibidir. (n: Örneklem ölçümü, nn: simülasyon tekrar sayısı)

```
function karmareksay(n,nn)
tt=0;
ttt=0;
for j=1:nn
s=1;
mini=rand;
maks=mini;
for i=2:n
x=rand;
if x<mini
s=s+1;
mini=x;
elseif x>maks
s=s+1;
maks=x;
end
end
tt=tt+s;
ttt=ttt+s^2;
end
%Teorik olarak ortalama karma rekor sayısı
t=2;
for i=3:n
t=t+(2/i);
end
t
%Simülasyona dayalı ortalama karma rekor sayısı
simort=tt/nn
%Yaklaşık değer olarak ortalama rekor sayısı
yt=2*log(n)+2*(0.5772)-1;
%Teorik olarak karma rekor sayısının varyansı
v=t-2;
for i=3:n
v=v-(4/(i^2));
end
v
%Simülasyona dayalı karma rekor sayısının varyansı
simvar=(ttt-nn*simort^2)/(nn-1)
%Yaklaşık değer olarak karma rekor sayısının varyansı
yv=2*log(n)+2*(0.5772)-(2/3)*pi^2+2;
end
```