T.C.
SELÇUK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ANA KİTLEYİ KAPSAYAN İNVARYANT GÜVEN ARALIKLARI VE SEVİYELERİNİN TAHMİNİ

Buğra SARAÇOĞLU
YÜKSEK LİSANS TEZİ
İSTATİSTİK ANABİLİM DALI
Konya, 2002

ÖZET

Yüksek Lisans Tezi

ANA KİTLEYİ KAPSAYAN İNVARYANT GÜVEN ARALIKLARI VE SEVİYELERİNİN TAHMİNİ

Buğra SARAÇOĞLU
Selçuk Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Yrd. Doç. Dr. Mehmet Fedai KAYA
2002, 55 Sayfa
Jüri: Yrd. Doç. Dr. Mehmet Fedai KAYA
Yrd. Doç. Dr. Aşır GENÇ
Yrd. Doç. Dr. Hasan KÖSE

Bu tez çalışması altı bölümden oluşmaktadır. Birinci bölümde, konuyla ilgili giriş yapılmıştır. İkinci bölümde olasılık uzayları, rasgele değişkenler ve dağılımları, üçüncü ve dördüncü bölümlerde tez çalışması için gerekli olan sıra istatistikleri, dağılımdan bağımsız istatistikler, örneklem dağılım fonksiyonu, dağılımdan bağımsız güven aralıkları ve tolerans aralıkları ile ilgili tanımlar ve temel bilgiler verilmiştir. Beşinci bölümde ise özel halde düzgün dağılım fonksiyonlarına sahip dağılımlar sınıfı için majorant vektörler ve sıra istatistikleri yardımıyla invaryant güven aralıkları oluşturulmuş ve $\alpha = 0.90$ güven seviyesi ve çeşitli n değerleri için bilgisayar programı yardımıyla majorant vektörler tespit edilmiştir. Tezin son kısmında ise iki yanlı invaryant güven aralıklarının seviyesinin tahmini problemi ele alınmış ve örneklem dağılım fonksiyonu yardımıyla bu problem çözülmüştür. Altıncı bölümde ise çalışmanın sonuçları özetlenmiştir.

Anahtar Kelimeler: Dağılımdan bağımsız istatistikler, örneklem dağılım fonksiyonu, dağılımdan bağımsız güven aralıkları, tolerans aralıkları, majorant vektörler, invaryant güven aralıkları.

ABSTRACT

Master Thesis

INVARIANT CONFIDENCE İNTERVALS CONTAINING THE MAIN MASS AND THEIR LEVEL OF ESTIMATION

Buğra SARAÇOĞLU

Selcuk University

Graduate School of Natural and Applied Sciences

Department of Statistics

Supervisor: Assist. Prof. Dr. Mehmet Fedai KAYA

2002, 55 Page

Jüri: Assist. Prof. Dr. Mehmet Fedai KAYA

Assist. Prof. Dr. Aşır GENÇ

Assist. Prof. Dr. Hasan KÖSE

The thesis consist of six chapters. The first chapter includes introduction part. In the second chapter, probability spaces, random variables and their distributions are presented. In the third and fourth chapters, some basic knowledge is given about order statistics, distribution free statistics, empirical distribution function, distribution free confidence intervals, tolerance intervals for the study. As for the fifth chapter, invariant confidence intervals are constructed by the help of order statistics and majorant vectors for distribution classes having private uniform distribution functions. The majorant vectors are determined by the aid of computer program for various n values and $\alpha = 0.90$ confidence level. In the last chapter of the thesis, estimation problem of two sided invariant confidence intervals' level is handled and this problem is solved by the aid of emprical distribution function. In the fifth chapter, the results of the statistical studies are summarized.

Key Words: Order statistics, distribution free statistics, empirical distribution function, distribution free confidence intervals, tolerance intervals, majorant vectors, invariant confidence intervals.

TEŞEKKÜR

Bu çalışma konusunu bana veren ve çalışmalarım süresince yakın ilgi ve yardımlarını hiçbir zaman esirgemeyen değerli hocam sayın Yrd.Doç.Dr. Mehmet Fedai KAYA ' ya, çalışmalarımda yardımcı olan değerli arkadaşlarım Arş. Gör. Coşkun KUŞ, Arş Gör. Naim TUĞLU, Arş. Gör. Ahmet PEKGÖR, Arş.Gör. İsmail KINACI ' ya, sayın hocam Yrd.Doç.Dr. Aşır GENÇ ' e ve manevi desteğini esirgemeyen aileme teşekkürlerimi sunarım.

İÇİNDEKİLER

Sayfa

ÇİZELGELER DİZİNİ

1.GİRİŞ

$\{\Omega, U, P\}$ bir olasılık uzayı, X bu uzayda tanımlı bir rasgele değişken olmak üzere 1853 yılında Chebyshev P. L.

$$\forall\varepsilon > 0 \text{ için } P\{|X - EX| \geq \varepsilon\} \leq \frac{\text{var}X}{\varepsilon^2}$$

eşitsizliğinin varlığını gösterdi (Kendal ve Stuart 1973). Burada EX, X rasgele değişkeninin beklenen değerini, $\text{var}X$ ise X rasgele değişkeninin varyansını gösteriyor. Bu eşitsizlik yardımıyla

$$\left(EX - \varepsilon\sqrt{\text{var}X}, EX + \varepsilon\sqrt{\text{var}X}\right)$$

aralığının X rasgele değişkeninin değerlerini kapsaması olasılığının $1-\frac{1}{\varepsilon^2}$ den büyük olduğu söylenebilir. Görüldüğü üzere Chebyshev eşitsizliği yardımıyla X rasgele değişkeninin değerlerini kapsayan bir güven aralık kurabilmek için birinci ve ikinci momentlerinin varlığına ihtiyaç duyulmaktadır. Chebyshev eşitsizliği bütün dağılımlar sınıfı için geçerli olan ve iyileştirilemeyen bir eşitsizliktir.

1919 yılında Pearson, $k \in Z$ (tamsayılar kümesi) olmak üzere $E|X - EX|^k < \infty$ koşulunun sağlanması durumunda $\forall\varepsilon > 0$ için

$$P\{|X - EX| \geq \varepsilon\} \leq \frac{E|X - EX|^k}{\varepsilon^k}$$

eşitsizliğinin varlığını ispat etti (Kendal ve Shewart 1973). Pik eşitsizliği ise $EX^2 < \infty$ olması koşulu altında $\forall\varepsilon > 0$ ve $\delta \leq \varepsilon$ için

$$P\{|X - EX| \geq \varepsilon\sqrt{\text{var}X}\} \leq \frac{1-\delta^2}{\varepsilon^2 - 2\varepsilon\delta + 1}$$

dir (Shahbazov 1973).

X sürekli bir rasgele değişken ve X rasgele değişkeninin olasılık yoğunluk fonksiyonu $m(x)$ noktasında yegane (bir tek) maksimuma sahipse o zaman

$$\forall\varepsilon > 0 \text{ için } P\{|X - m(x)| \geq \varepsilon\} \leq \frac{4}{9\cancel{\varepsilon^2}\,\varepsilon^2} E|X - \underbrace{EX}_{m(x)}|^2$$

eşitsizliği gerçeklenir. Bu eşitsizlik 1821 yılında Gaus K. F tarafından ispatlanmıştır (Kendal ve Shewart 1973).

Bu eşitsizlikten, özel halde simetrik ve tek modlu (tepeli) dağılımlar için;

$$\forall\varepsilon > 0 \text{ için } P\left\{|X - m(x)| \geq \varepsilon\sqrt{\operatorname{var} X}\right\} \leq \frac{1}{4\varepsilon^2}$$

eşitsizliği elde edilir. Bu eşitsizlik simetrik dağılımlar için çok meşhur olan $3\sigma(X) = 3\sqrt{\operatorname{var} X}$ yasasının (Cramer 1946) esaslandırılması demektir. Yani, simetrik ve tek tepeli X rasgele değişkenlerinin değerlerinin

$$\left(EX - 3\sigma(X), EX + 3\sigma(X)\right)$$

aralığında olması olasılığı 1'e yakındır. Bu kural kullanışlı olması bakımından bir çok alanda uygulamacılar tarafından kullanılır. Tek tepeli ve simetrik olmayan diğer dağılımlar için de $3\sigma(X)$ kuralı veya buna benzer bir kural var mıdır? Bu problemin çözümü 1979 yılında (Wischanski ve Petunin 1979) tarafından verilmiştir.

X sürekli ve tek tepeli bir rasgele değişken ve $\forall\varepsilon > 0$ için $EX^2 < \infty$ olmak üzere;

$$P\left\{|X - m(x)| \geq \varepsilon\sqrt{\operatorname{var} X}\right\} \leq \max\left\{\frac{4}{9\varepsilon^2}, \frac{4 - \varepsilon^2}{3\varepsilon^2}\right\}$$

eşitsizliği mevcuttur. Eğer $\varepsilon > \sqrt{\frac{8}{3}}$ olursa o taktirde

$$P\left\{|x - m(x)| \geq \varepsilon\sqrt{\operatorname{var} X}\right\} \leq \frac{4}{9\varepsilon^2} \text{ olur.}$$

1985 yılında Dharmadhikari ve Joang Dev tarafından bir tepeli dağılımlar için;

$$\tau_r = E|X - m(x)|^r, \quad s > r + 1, \quad s(s - r - 1)^r = r^r$$

olmak üzere $k > 0$ için

$$P\left\{|X - m(x)| \geq k\right\} \leq \max\left\{\frac{s\tau_r - k^r}{(s-1)k^r}, \left(\frac{r}{r+1}\right)^r \cdot \frac{\tau_r}{k^r}\right\}$$

olduğu ispatlandı.

Görüldüğü üzere bir çok uygulamada kullanılan $3\sigma(X)$ kuralı keyfi dağılımlar için tam olarak sonuçlandırılmış değildir. Yukarıda verilen Chebyshev ve

Gaus türü eşitsizlikler X rasgele değişkeninin değerleri için bu rasgele değişkenin sayısal karakteristiklerini kullanan deterministik (rasgele olmayan) güven aralıklar kurmaya imkan verir.

Şimdi de X rasgele değişkeninin değerlerini kapsayan güven aralık oluşturmak için ikinci bir yaklaşımı göz önüne alalım. $X_1, X_2, \ldots, X_n$ dağılım fonksiyonu F olan kitleden alınmış bir örneklem, f_1 ve f_2

$$\forall (u_1, u_2, \ldots, u_n) \in R^n \text{ için } f_1(u_1, u_2, \ldots, u_n) \leq f_2(u_1, u_2, \ldots, u_n)$$

özelliğine sahip iki Borel ölçülebilir fonksiyon olmak üzere;

$$\left(f_1\left(X_1, X_2, \ldots, X_n\right), f_2\left(X_1, X_2, \ldots, X_n\right)\right)$$

rasgele aralığını göz önüne alalım. X_{n+1} ; $X_1, X_2, \ldots, X_n$ örnekleminden bağımsız yeni bir rasgele değişken olmak üzere X_{n+1} in $\left(f_1\left(X_1, X_2, \ldots, X_n\right), f_2\left(X_1, X_2, \ldots, X_n\right)\right)$ rasgele aralığına düşmesi olasılığı F dağılım fonksiyonuna bağlıdır. Eğer F bilinmiyorsa ki uygulamada genellikle böyledir, o zaman önceden güvenilirliği belli olan böyle bir aralığı kurmak mümkün değildir. Ancak aşağıda görüleceği gibi sürekli dağılımlar sınıfı için bu problemin çözümü mevcuttur. 1931 yılında (Shewart 1931) kalite kontrol problemleriyle ilgili olarak tolerans aralıklar kavramını verdi. Eğer

$$F\left(f_2\left(X_1, X_2, \ldots, X_n\right)\right) - F\left(f_1\left(X_1, X_2, \ldots, X_n\right)\right)$$

rasgele değişkeninin dağılımı F den bağımsızsa ve $0 \leq \beta, \gamma \leq 1$ için

$$P\left\{ \int_{f_1(X_1, \ldots, X_n)}^{f_2(X_1, \ldots, X_n)} dF(u) \geq \beta \right\} = \gamma$$

oluyorsa o taktirde $\left(f_1\left(X_1, X_2, \ldots, X_n\right), f_2\left(X_1, X_2, \ldots, X_n\right)\right)$ rasgele aralığına tolerans aralık denir. Görüldüğü gibi $\left(f_1\left(X_1, X_2, \ldots, X_n\right), f_2\left(X_1, X_2, \ldots, X_n\right)\right)$ rasgele aralığının tolerans aralık olması demek β ve γ önceden belirlenmek üzere bu aralığın olasılık ölçüsünün β' dan büyük olması olasılığının F den bağımsız olması demektir. 1942 yılında (Wilks 1942) $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ sıra istatistikleri olmak üzere $\left(X_{(i)}, X_{(j)}\right)$, $1 \leq i < j \leq n$ aralığının tolerans aralık olduğunu gösterdi. Robbins, 1944 yılında, f_1 ve f_2 fonksiyonlarının simetrik ve sürekli olması durumunda tolerans aralıklar oluşturabilecek istatistiklerin sadece sıra istatistikleri olabileceğini gösterdi (Robbins 1944). Bu özellik istatistiksel analizde, sıra istatistiklerinin çok önemli özelliklerinden biridir.

1990 yılında Bairamov ve Petunin tarafından esas kitleyi kapsayan invaryant güven aralıkları kavramı verildi.

$G = \{G_\alpha : \alpha \in \Pi\}$ dağılım fonksiyonları $F_\alpha(u)$ olan kitlelerin sınıfını göstersin. $X_1^\alpha, X_2^\alpha, \ldots, X_n^\alpha, X_{n+1}^\alpha$, G_α kitlesinden alınmış bir örneklem f_1 ve f_2 de

$$\forall (u_1, u_2, \ldots, u_n) \in R^n \text{ için } f_1(u_1, \ldots, u_n) \leq f_2(u_1, \ldots, u_n)$$

özelliğine sahip iki Borel ölçülebilir fonksiyon olmak üzere

$$\forall \alpha \in \Pi \text{ için } P\left\{X_{n+1}^\alpha \in \left(f_1\left(X_1^\alpha, \ldots, X_n^\alpha\right), f_2\left(X_1^\alpha, \ldots, X_n^\alpha\right)\right)\right\} \geq \beta$$

olacak biçimde $\left(f_1\left(X_1^\alpha, \ldots, X_n^\alpha\right), f_2\left(X_1^\alpha, \ldots, X_n^\alpha\right)\right)$ rasgele aralığı oluşturulabilir ve bu aralığa ana kitleyi kapsayan β seviyeli invaryant güven aralık denir. G_c tüm sürekli dağılım fonksiyonlarına sahip kitlelerin sınıfı, $f_1(u_1, \ldots, u_n) \leq f_2(u_1, \ldots, u_n)$ sürekli ve simetrik fonksiyonlar ve $A = \{(u_1, \ldots, u_n) : f_1(u_1, \ldots, u_n) = f_2(u_1, \ldots, u_n)\}$ olmak üzere Lebesque ölçüsü anlamında $\mu(A) = 0$ ise $\left(f_1\left(X_1, X_2, \ldots, X_n\right), f_2\left(X_1, X_2, \ldots, X_n\right)\right)$ rasgele aralığının ana kitleyi kapsayan invaryant güven aralık olması için gerek ve yeter koşul $f_1\left(X_1, X_2, \ldots, X_n\right) = X_{(i)}$, $f_2\left(X_1, X_2, \ldots, X_n\right) = X_{(j)}$ olmasıdır (Bairamov ve Petunin 1990).

Ana kitleyi kapsayan invaryant güven aralıkları ve istatistiki hipotezlerin test edilmesi teorisi arasında sıkı bir ilişki vardır. G_c sınıfı içinde bir çok uyum iyiliği ve iki örneklem testleri invaryant güven aralıklarına bağlı olarak tasarlanmıştır.

İnvaryant güven aralıkları kavramı çok boyutlu örneklemler halinde invaryant güven kümeleri adı altında genelleştirilmiş ve gözlemlerin sınıflandırılması problemleri için hatalı sınıflandırma olasılıkları yeterince küçuk olan ve az sayıda örneklemlerde kullanılabilen testler tasarlanmıştır (Bairamov 1992, Bairamov ve Petunin 1992).

Tüm sürekli dağılımlar sınıfı F_c için en iyi özelliklere sahip invaryant güven aralıkları sıra istatistikleri yardımıyla oluşturulmasına rağmen $F = \{F_\theta : \theta \in \Theta\} \subset F_c$

sınıfı için sıra istatistiklerinden farklı olan çoğu zaman sıra istatistiklerinin lineer fonksiyonlarının da elde edilen invaryant güven aralıklarının varlığı gösterilebilir.

Konum ve ölçek parametreleri ile verilen dağılımlar ailesi için invaryant güven aralıklarına bağlı olarak tanımlanan önemli istatistiklerin sonlu ve asimptotik dağılımları bulunabilir. Bir çok durumda asimptotik dağılımların normal değil beta veya benzeri dağılımlara sahip olduğu görülür. Bu yöntemle elde edilen teoremler örneklemin belli bir parametrik aileden gelip gelmediğinin test edilmesinde kullanılır.

Konum ve ölçek parametresi içeren dağılım fonksiyonları ailesi için majorant vektörler yardımıyla invaryant güven aralıklarının oluşturulabileceği gösterilmiş ve üstel dağılım fonksiyonları ailesi için invaryant güven aralıkları oluşturulmuştur (Bairamov ve ark. 1999).

Bu çalışmada düzgün dağılım fonksiyonları ailesi için invaryant güven aralıkları oluşturulduktan sonra belli bir α seviyesi için en dar güven aralığını verecek olan majorant vektörler arasındaki ilişki bilgisayar programı yardımıyla belirlenmeye çalışılmıştır.

Çalışmanın özgün diğer kısmında ise iki yanlı invaryant güven aralıklarının seviyesi örneklem dağılım fonksiyonu yardımıyla yansız bir tahmin edici bulunmaya çalışılmış ve sonuçların doğruluğu örnekler yardımıyla gösterilmiştir.

2. TEMEL KAVRAMLAR

Bu bölümde, yapılmış olan çalışma için gerekli olan olasılık uzayları, rasgele değişkenler ve dağılımları hakkında genel bilgiler verilecektir.

2.1. Olasılık Uzayları ve Rasgele Değişkenler

İstatistik, rasgelelik içeren olaylar, süreçler ve sistemler hakkında matematiksel modeller kurmada ve bu modellerin uyumluluğuna ve bu modellerden sonuç çıkarmada gerekli bilgi ve yöntemleri ortaya koyan bir bilim dalıdır.

Tanım 2.1.1. Sonuçların kümesi belli olan ancak gerçeklendiğinde hangi sonucun ortaya çıkacağı önceden bilinmeyen işleme "olasılık deneyi" denir.

Tanım 2.1.2. Bir olasılık deneyinin tüm olabilir sonuçlarının kümesine örnek uzay denir.

Tanım 2.1.3. Örnek uzayın bir alt kümesine "olay" denir. Bir olayın gerçekleşmesi deney sonucunun bu kümenin elemanı olması demektir.

Tanım 2.1.4. Bir Ω cümlesinin alt cümlelerinden oluşan bir U sınıfı,

i. $\Omega \in U$ (2.1)

ii. $A \in U$ cümlesi için $\overline{A} \in U$ (2.2)

iii. U 'da her (A_n) dizisi için $\bigcup_{i=1}^{\infty} A_i \in U$ (2.3)

özelliklerine sahipse U sınıfına Ω 'da bir "σ - cebir" denir.

Teorem 2.1.5. U, Ω üzerinde bir σ - cebir olmak üzere,

i. $\Omega \in U$

ii. $(A_n), U$ 'da bir dizi ise $\bigcap_{i=1}^{\infty} A_i \in U$

iii. $A_i \in U$, $i=1,...,n$ $n \in N$ ise $\bigcup_{i=1}^{n} A_i \in U$

iv. $A_i \in U$, $i=1,...,n$ $n \in N$ ise $\bigcap_{i=1}^{n} A_i \in U$

v. $A, B \in U$ ise $A \setminus B \in U$

Teorem 2.1.6. $I \neq \varnothing$ bir indis cümlesi ve U_i , $(i \in I)$, σ - cebir olmak üzere, $U = \bigcap_{i \in I} U_i$ sınıfı da bir $\sigma - cebir$ 'dir.

İspat. $\forall i \in I$ için U_i , σ - cebir olsun.

i. (2.1) koşulundan $\forall i \in I$ için $\Omega \in U_i$ olduğundan $\Omega \in \bigcap_{i \in I} U_i$ olur.

ii. (2.2) koşulundan $\forall i \in I$ için $\overline{A} \in U_i$ ' dir.

$A \in \bigcap_{i \in I} U_i$ ise $\forall i \in I$ için $\overline{A} \in U_i$ ise $\overline{A} \in \bigcap_{i \in I} U_i$

iii. (2.3) koşulundan (A_n), U_i ' de bir dizi ise $\bigcup_{i=1}^{\infty} A_i \in U_i$ ' dir.

(A_n), $\bigcap_{i \in I} U_i$ 'de bir dizi olsun.

$\forall i \in I$ için $\bigcup_{n=1}^{\infty} A_n \in U_i$ ise $\bigcup_{i=1}^{\infty} A_i \in \bigcap_{i \in I} U_i$

olduğundan $\bigcap_{i \in I} U_i$, Ω 'da bir $\sigma - cebir$ 'dir.

Tanım 2.1.7. U , Ω 'da bir $\sigma - cebir$ olmak üzere,

$$P : U \to R$$
$$A \to P(A)$$

fonksiyonu,

i. $A \in U$ için $P(A) \geq 0$

ii. $P(\Omega) = 1$

iii. U 'daki ayrık cümlelerin her (A_n) dizisi için $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$

özelliklerine sahipse P 'ye U üzerinde "bir olasılık ölçüsü" denir. $P(A)$ değerine A 'nın ölçüsü veya A 'nın olasılığı denir.

Tanım 2.1.8. Ω boş olmayan bir cümle, U , Ω 'da bir $\sigma - cebir$ ve P , U üzerinde bir olasılık ölçüsü olmak üzere, (Ω, U, P) üçlüsüne "olasılık uzayı" denir.

Teorem 2.1.9. (Ω, U, P) olasılık uzayı olsun.

i. $P(\varnothing) = 0$

ii. $A_1, A_2, ..., A_n$, U'da ayrık cümleler ise $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$

iii. $P(\overline{A}) = 1 - P(A)$

iv. $A \subset B$ ise $P(A) \leq P(B)$

v. $0 \leq P(A) \leq 1$

vi. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Tanım 2.1.10. (Ω, U, P) olasılık uzayı ve

$$X : \Omega \to R$$
$$w \to X(w)$$

bir fonksiyonu, $\forall B \in \mathrm{B}$ için $X^{-1}(B) \in U$ koşulunu sağlıyorsa X'e "rasgele değişken" denir. Burada B borel kümesini B borel $\sigma - cebir$'ini göstermektedir.

Tanım 2.1.11. (Ω, U, P) olasılık uzayı ve X rasgele değişken olmak üzere,

$$F : R \to [0,1]$$
$$x \to F(x) = P(X \leq x)$$

fonksiyonuna "X rasgele değişkenin dağılım fonksiyonu" denir.

Tanım 2.1.12 X rasgele değişkenin $X(\Omega)$ değer kümesi sayılabilir olduğunda X'e kesikli rasgele değişken ve X'in belirlediği olasılık dağılımına da "kesikli dağılım" denir. X'in $x \in X(\Omega)$ değerini alması olasılığı $P(X = x) = P\{w \in \Omega : X(w) = x\}$ olmak üzere,

$$\sum_{x \in X(\Omega)} P(X = x) = P\left(\bigcup_{x \in X(\Omega)} (X = x)\right) = 1$$

dir. Kesikli X rasgele değişkenin dağılım fonksiyonu

$$F(x) = P(X \leq x)$$
$$= \sum_{\substack{a \leq x \\ a \in X(\Omega)}} P(X = a), \quad -\infty < x < \infty$$

ve F basamak fonksiyonudur.

Tanım 2.1.13. (Ω, U, P) olasılık uzayı ve X rasgele bir değişken olmak üzere,

$$f : X(\Omega) \to R$$
$$x \to f(x) = P(X = x)$$

fonksiyonuna X rasgele değişkenin "olasılık fonksiyonu" denir.

Tanım 2.1.14. Bir $f: R \to R$ fonksiyonu için

i. $f(x) \geq 0$, $x \in R$

ii. $\int_{-\infty}^{\infty} f(x)dx = 1$

özellikleri sağlanıyorsa f fonksiyonuna "olasılık yoğunluk fonksiyonu" denir.

Tanım 2.1.15. Bir X rasgele değişkenin F dağılım fonksiyonu bir f olasılık yoğunluk fonksiyonu yardımıyla,

$$F(x) = \int_{-\infty}^{x} f(x)dx \ , \ -\infty < x < \infty$$

şeklinde yazılabiliyorsa X rasgele değişkenine "mutlak sürekli veya kısaca sürekli rasgele değişken" ve f fonksiyonuna "olasılık yoğunluk fonksiyonu" denir.

Tanım 2.1.16. X, bir rasgele değişken ve $g: R \to R, \forall B \in \mathrm{B}(R)$ için $\{x : g(x) \in B\} \in \mathrm{B}(R)$ özelliğine sahip bir fonksiyon olmak üzere:

i. X kesikli ve $\sum_x |g(x)| f(x) < \infty$ olduğunda

$$E(g(X)) = \sum_x g(x) f(x)$$

ii. X sürekli ve $\int_{-\infty}^{\infty} |g(x)| f(x)dx < \infty$ olduğunda,

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)dx$$

değerine $g(X)$'in beklenen değeri denir (Öztürk 1993).

3. SIRA İSTATİSTİKLERİ VE DAĞILIMDAN BAĞIMSIZ İSTATİSTİKLER

Bu kısımda tez çalışması için gerekli olan sıra istatistikleri ve dağılımdan bağımsız istatistikler ile ilgili tanımlar ve temel bilgiler verilmiştir.

3.1. Sıra İstatistikleri

$X_1, X_2, \ldots, X_n$; $F(x)$ dağılım fonksiyonuna sahip bir örneklem olmak üzere bu örneklemin $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ olacak biçimde büyüklük sırasına göre dizilmesiyle elde edilen her bir $X_{(i)}$ rasgele değişkenine bu örneklemin i. sıra istatistiği adı verilir.

$X_{(1)} = \min\{X_1, X_2, \ldots, X_n\}$ sıra istatistiğinin dağılım fonksiyonu,

$$\begin{aligned}
F_1(x) &= P\{X_{(1)} \leq x\} \\
&= P\{\min\{X_1, X_2, \ldots, X_n\} \leq x\} \\
&= 1 - P\{\min\{X_1, X_2, \ldots, X_n\} > x\} \\
&= 1 - P\{X_1 > x, X_2 > x, \ldots, X_n > x\} \\
&= 1 - P\{X_1 > x\}P\{X_2 > x\} \ldots P\{X_n > x\} \\
&= 1 - (1 - F(x))^n
\end{aligned} \tag{3.1}$$

ve

$X_{(n)} = \text{maks}\{X_1, X_2, \ldots, X_n\}$ sıra istatistiğinin dağılım fonksiyonu,

$$\begin{aligned}
F_n(x) &= P\{X_{(n)} \leq x\} \\
&= P\{\text{maks}\{X_1, X_2, \ldots, X_n\} \leq x\} \\
&= P\{X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x\} \\
&= P\{X_1 \leq x\}P\{X_2 \leq x\} \ldots P\{X_n \leq x\} \\
&= (F(x))^n
\end{aligned} \tag{3.2}$$

şeklindedir. $1 \leq r \leq n$ olmak üzere r. sıra istatistiğinin dağılım fonksiyonu ise;

$$\begin{aligned}
F_r(x) &= P\{X_{(r)} \leq x\} \\
&= P\{X_1, X_2, \ldots, X_n \text{ 'lerden en az } r \text{ tanesi} \leq x\}
\end{aligned}$$

$$F_r(x) = P\left\{\bigcup_{i=r}^{n}(X_1, X_2, ..., X_n \text{ 'lerden tam i tanesi} \leq x)\right\}$$

olur. A_i: $\{X_1, X_2, ..., X_n$ 'lerden tam i tanesi $\leq x\}$ olması olayı olsun. A_i olayları ayrık olduğundan,

$$F_r(x) = P\left\{\bigcup_{i=r}^{n} A_i\right\} = \sum_{i=r}^{n} P\{A_i\} = \sum_{i=r}^{n} C_n^i (F(x))^i (1 - F(x))^{n-i} \tag{3.3}$$

dır. Bu ise tam olmayan beta fonksiyonudur. Yani;

$$\begin{aligned} F_r(x) &= \sum_{i=r}^{n} C_n^i (F(x))^i (1 - F(x))^{n-i} \\ &= \frac{1}{B(r, n-r+1)} \int_0^{F(x)} t^{r-1}(1-t)^{n-r}\, dt \\ &= I_{F(x)}(r, n-r+1) \end{aligned} \tag{3.4}$$

dir. Eğer X_i ' ler $f(x) = \dfrac{dF(x)}{dx}$ olacak biçimde sürekli rasgele değişkenler ise,

$$\begin{aligned} f_r(x) = \frac{dF_r(x)}{dx} &= \frac{1}{B(r, n-r+1)} \frac{d}{dx} \int_0^{F(x)} t^{r-1}(1-t)^{n-r}\, dt \\ &= \frac{1}{B(r, n-r+1)} (F(x))^{r-1}(1-F(x))^{n-r} f(x) \end{aligned} \tag{3.5}$$

olur. Burada,

$$\begin{aligned} B(r, n-r+1) &= \frac{\Gamma(r)\Gamma(n-r+1)}{\Gamma(n+1)} \\ &= \frac{(r-1)!(n-r)!}{n!} \end{aligned}$$

dir. X_i rasgele değişkenlerinin kesikli olması durumunda (3.3) eşitliğinden,

$$\begin{aligned} f_r(x) &= P\{X_{(r)} = x\} \\ &= P\{x - 1 < X_{(r)} \leq x\} \\ &= F_r(x) - F_r(x-1) \\ &= \sum_{i=r}^{n} C_n^i (F(x))^i (1-F(x))^{n-i} - \sum_{i=r}^{n} C_n^i (F(x-1))^i (1-F(x-1))^{n-i} \end{aligned} \tag{3.6}$$

şeklinde elde edilir. David 1970'te verilen bir yöntemde, sıra istatistiklerinin olasılık yoğunluk fonksiyonları ve ortak olasılık yoğunluk fonksiyonları kombinatorik yöntemler kullanılarak aşağıdaki gibi verilmiştir.

$A_r = \{x < X_{(r)} < x + \Delta x\}$ olayını göz önüne alalım.

A_r olayının gerçeklenmesi için $X_1, X_2, \ldots, X_n$ örnekleminden $r-1$ tane örneğin x den küçük, 1 tane örneğin $(x, x+\Delta x]$ aralığında ve $n-r$ tane örneğin de $x+\Delta x$ den büyük olması gerekir. Dolayısıyla A_r olayı;

$$\binom{n}{r-1}\binom{n-(r-1)}{1}\binom{n-r}{n-r} = \frac{n!}{(r-1)!(n-r)!} = \frac{1}{B(r, n-r+1)} = B$$

farklı sayıda gerçekleşebilir.

$$\begin{aligned} P\{A_r\} &= P\{x < X_{(r)} < x+\Delta x\} = F_r(x+\Delta x) - F_r(x) \\ &= BP\{X_1, \ldots, X_{r-1} \le x; X_r \in (x, x+\Delta x]; X_{r+1}, \ldots, X_n > x+\Delta x\} \\ &= B[F(x)]^{r-1}[F(x+\Delta x) - F(x)][1 - F(x+\Delta x)]^{n-r} \end{aligned} \tag{3.7}$$

elde edilir. r. sıra istatistiğinin olasılık yoğunluk fonksiyonu;

$$\lim_{\Delta x \to 0} \frac{F_r(x+\Delta x) - F_r(x)}{\Delta x} = B \lim_{\Delta x \to 0} \frac{[F(x)]^{r-1}[F(x+\Delta x) - F(x)][1 - F(x+\Delta x)]^{n-r}}{\Delta x}$$

eşitliğinde $F(x+\Delta x) - F(x) = f(x)\Delta x + o(\Delta x)$ yazılırsa;

$$\begin{aligned} \lim_{\Delta x \to 0} \frac{F_r(x+\Delta x) - F_r(x)}{\Delta x} &= B\left[\lim_{\Delta x \to 0}[F(x)]^{r-1} f(x)[1 - F(x+\Delta x)]^{n-r} + \lim_{\Delta x \to 0} \frac{o(\Delta x)}{\Delta x}\right] \\ &= B[F(x)]^{r-1}[1 - F(x)]^{n-r} f(x) \\ &= B[F(x)]^{r-1}[1 - F(x)]^{n-r} f(x) \\ &= f_r(x). \end{aligned} \tag{3.8}$$

şeklinde elde edilir.

$X_{(r)}$ ve $X_{(s)}$ $1 \le r < s \le n$ sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonu da aynı yöntemle aşağıdaki gibi bulunur.

$$A_{r,s} = \{x < X_{(r)} < x + \Delta x \text{ , } y < X_{(s)} < y + \Delta y\}$$

olayının gerçeklenmesi $X_1, X_2, \ldots, X_n$ örnekleminden $r-1$ tane örneğin x den küçük, 1 tane örneğin $(x, x+\Delta x]$ aralığında, $s-r-1$ tane örneğin $(x+\Delta x, y]$ aralığında. 1 tane gözlemin $(y, y+\Delta y]$ aralığında ve $n-s$ tane örneğin de $y+\Delta y$ den büyük olması durumunda sağlanır.

Burada mümkün olan durumların sayısı;

$$C_n^{r-1} C_{n-(r-1)}^1 C_{n-r}^{s-r-1} C_{n-s+1}^1 C_{n-s}^{n-s} = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} = \mathrm{C}$$

dir.

$$P\{x < X_{(r)} < x + \Delta x\,,\, y < X_{(s)} < y + \Delta y\} = C[F(x)]^{r-1}[F(x + \Delta x) - F(x)] \\ \times [F(y) - F(x + \Delta x)]^{s-r-1}[F(y + \Delta y) - F(y)][1 - F(y + \Delta y)]^{n-s} \tag{3.9}$$

dır. Öte yandan, $F_{r,s}(x,y) = P\{X_r < x\,,\, X_s < y\}$ olmak üzere;

$$f_{rs}(x,y) = \frac{\partial^2 F_{r,s}(x,y)}{\partial x \partial y} = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{1}{\Delta x . \Delta y} P\{x < X_{(r)} < x + \Delta x\,,\, y < X_{(s)} < y + \Delta y$$

$$F(x + \Delta x) - F(x) = f(x)\Delta x + o(\Delta x)$$

ve

$$F(y + \Delta y) - F(y) = f(y)\Delta y + o(\Delta y)$$

(3.9) ifadesinde yerine yazılırsa $x < y$ ve $r < s$ olmak üzere;

$$f_{rs}(x,y) = C[F(x)]^{r-1}[F(y) - F(x)]^{s-r-1}[1 - F(y)]^{n-s} f(y) f(x) \tag{3.10}$$

olur.

$k \leq n$ ve $1 \leq r_1 < r_2 < ... < r_k \leq n$ olmak üzere $X_{(r_1)}, X_{(r_2)}, ..., X_{(r_n)}$ sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonu da benzer şekilde;

$$\frac{n!}{(r_1 - 1)!(r_2 - r_1 - 1)!...(n - r_k)!} = D$$

olmak üzere

$$f_{r_1,...,r_k}(x_1,...,x_k) = D[F(x_1)]^{r_1-1}[F(x_2) - F(x_1)]^{r_2-r_1-1} ...[1 - F(x_k)]^{n-r_k} \\ \times f(x_1) \cdots f(x_k), \qquad x_1 \leq x_2 \leq ... \leq x_k \tag{3.11}$$

dır. $x_0 = -\infty$, $x_{k+1} = \infty$, $r_0 = 0$, $r_{k+1} = n + 1$ olarak tanımlanırsa,

$$f_{r_1,...,r_k}(x_1,...,x_k) = n!\left[\prod_{i=1}^{k} f(x_i)\right]\prod_{i=0}^{k}\left(\frac{[F(x_{i+1}) - F(x_i)]^{r_{i+1}-r_i-1}}{(r_{i+1} - r_i - 1)!}\right) \tag{3.12}$$

şekline dönüşür.

$X_{(1)}, X_{(2)}, ..., X_{(n)}$ sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonu ise;

$$f_{1,2,...,n}(x_1, x_2, ..., x_n) = \begin{cases} n! f(x_1) f(x_2) ... f(x_n) & , \quad x_1 \leq x_2 \leq ... \leq x_n \\ 0 & , \quad \text{a.h.} \end{cases} \tag{3.13}$$

olur. $X_{(r)}$ ve $X_{(s)}$ sıra istatistiklerinin dağılım fonksiyonu $F_{r,s}(x,y)$ (3.10) eşitliğinden integral alınmak suretiyle elde edilebilir. Fakat, X_i ' lerin kesikli rasgele değişken olması durumunda $r < s$ olmak üzere $F_{r,s}(x,y)$ dağılım fonksiyonu aşağıdaki gibidir.

$$F_{r,s}(x,y) = P\{\text{en az } r \text{ tane } X_i \leq x, \text{en az } s \text{ tane } X_i \leq y\} \quad x < y$$

$$= \sum_{i=r}^{n} \sum_{j=s-i}^{n-i} P\{\text{tam } i \text{ tane } X_i \leq x, \text{tam } j \text{ tane } X_i \in (x,y]\}$$

$i > s$ olması durumunda j sıfırdan başlar ve $x < y$ için

$$F_{r,s}(x,y) = \sum_{i=r}^{n} \sum_{j=maks\{0,s-i\}}^{n-i} \frac{n!}{i!\,j!\,(n-i-j)!} [F(x)]^i [F(y)-F(x)]^j [1-F(y)]^{n-i-j} \quad (3.14)$$

dır.

$x \geq y$ olması durumunda; $X_{(s)} \leq y$ olması $X_{(r)} \leq x$ olmasını gerektireceğinden

$$F_{r,s}(x,y) = F_s(y) \quad (3.15)$$

olur.

$W_{r,s} = X_{(s)} - X_{(r)}$ olmak üzere $\omega_{r,s} = y - x$ yazarak (x,y) ' den $(x,\omega_{r,s})$ ' ye dönüşüm yapılırsa ve dönüşümün jakobyeninin mutlak değeri de 1 olacağından $W_{r,s}$ rasgele değişkeninin olasılık yoğunluk fonksiyonu;

$$f(\omega_{r,s}) = A \int_{-\infty}^{\infty} [F(x)]^{r-1} [F(x+\omega_{rs}) - F(x)]^{s-r-1} [1 - F(x+\omega_{rs})]^{n-s} f(x) f(x+\omega_{rs}) dx \quad (3.16)$$

şeklinde elde edilir. Özel olarak (3.16) eşitliğinde $r = 1, s = n$ alınırsa dağılımın genişliğinin olasılık yoğunluk fonksiyonu

$$f(\omega) = n(n-1) \int_{-\infty}^{\infty} f(x)[F(x+\omega) - F(x)]^{n-2} f(x+\omega) dx \quad (3.17)$$

olarak bulunur. Dağılım fonksiyonu ise;

$$F(\omega) = n \int_{-\infty}^{\infty} f(x) \int_{0}^{\omega} (n-1) f(x+u)[F(x+u) - F(x)]^{n-2} du dx$$

$$= n \int_{-\infty}^{\infty} f(x)[F(x+u) - F(x)]^{n-1} \Big|_{u=0}^{u=\omega} dx$$

$$F(\omega) = n \int_{-\infty}^{\infty} f(x)[F(x+\omega) - F(x)]^{n-1} dx$$

olur. Özel olarak;

$$f(x)=\begin{cases}1 & , \quad 0\le x\le 1\\ 0 & , \quad a.h.\end{cases}$$

alınırsa (3.10) eşitliğinden;

$$f_{r,s}(x,y)=\begin{cases}\dfrac{n!}{(r-1)!(s-r-1)!(n-s)!}(y-x)^{s-r-1}(1-y)^{n-s} & , \quad 0\le x\le y\le 1\\ 0 & , \quad \text{a.h.}\end{cases}$$

şeklinde elde edilir.

$x>1-\omega_{rs}$ için $f(x+\omega_{rs})=0$ olacağından (3.16) ' dan ;

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

olur. $x=y(1-\omega_{rs})$ dönüşümü yapılırsa

$$\begin{aligned}f(\omega_{rs})&=\frac{n!}{(r-1)!(s-r-1)!(n-s)!}\int_0^{1} y^{r-1}(1-y)^{n-s}\omega_{rs}^{\;s-r-1}(1-\omega_{rs})^{n-s+r}\,dy\\ &=\frac{1}{B(s-r,n-s+r+1)}\omega_{rs}^{\;s-r-1}(1-\omega_{rs})^{n-s+r}\quad 0\le\omega_{rs}\le 1\end{aligned}\tag{3.18}$$

elde edilir.

3.2. Dağılımdan Bağımsız İstatistikler

$X_1,X_2,\ldots,X_n$; $F\in\mathrm{F}$ dağılım fonksiyonuna sahip bir örneklem ve $S(X_1,X_2..,X_n)$ örneklemin herhangi bir istatistiği olsun.

Tanım 3.2.1. F , tüm dağılım fonksiyonlarının sınıfı olmak üzere ,$\forall F\in\mathrm{F}$ için $S(X_1,X_2,\ldots,X_n)$ istatistiğinin dağılımı aynı ise $S(X_1,X_2,\ldots,X_n)$ istatistiğine F sınıfı için dağılımdan bağımsız istatistik denir.

Örnek olarak;

$$\mathrm{F}_1=\left\{F{:}F(x,\mu_0,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}\exp\left(-\frac{1}{2\sigma^2}(t-\mu_0)^2\right)dt\;;\;\begin{array}{l}\mu_0\in R\text{ biliniyor}\\ \sigma^2>0\text{ bilinmiyor}\end{array}\right\}$$

ve

$$\mathrm{F}_2=\left\{F{:}F(x,\mu,\sigma_0^{\,2})=\frac{1}{\sqrt{2\pi}\sigma_0}\int_{-\infty}^{x}\exp\left(-\frac{1}{2\sigma_0^{\,2}}(t-\mu)^2\right)dt\,;\;\begin{matrix}\mu\in R\text{ bilinmiyor}\\ \sigma_0^{\,2}>0\text{ biliniyor}\end{matrix}\right\}$$

dağılım fonksiyonlarının iki sınıfı olsun. $X_1,X_2,\ldots,X_n$ dağılım fonksiyonu F_1 sınıfından olan bir örneklem olmak üzere $S_1=\frac{(\overline{X}-\mu_0)\sqrt{n}}{S}$ istatistiği F_1 sınıfı için dağılımdan bağımsız bir istatistiktir ve $n-1$ serbestlik dereceli t dağılımına sahiptir.

Eğer örneklemimiz F_2 sınıfından ise bu taktirde $S_2=\frac{(n-1)S^2}{\sigma_0^2}$ istatistiği F_2 sınıfı için dağılımdan bağımsız istatistiktir ve $(n-1)$ serbestlik dereceli ki-kare dağılımına sahiptir. Burada $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ ve $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ dir.

Tanım 3.2.2. X rasgele değişkeninin dağılım fonksiyonu $F_\theta(x)$ ve

$$F_\theta(x)\in\mathrm{F}=\{F_\theta(x):F_\theta(x)=F(x-\theta);\theta\in\Theta,F\text{ biliniyor}\}\tag{3.19}$$

ise θ parametresine F sınıfı için bir konum parametresi denir. θ parametresinin X rasgele değişkeni için konum parametresi olması için gerek ve yeter koşul $X-\theta$ rasgele değişkeninin dağılımının θ dan bağımsız olmasıdır.

$$P(X-\theta\le x)=P(X\le x+\theta)=F_\theta(x+\theta)=F(x+\theta-\theta)=F(x).$$

Tanım 3.2.3. X rasgele değişkeninin dağılım fonksiyonu $F_\theta(x)$ ve

$$F_\theta(x)\in\mathrm{F}=\left\{F_\theta(x){:}F_\theta(x)=F\left(\frac{x}{\theta}\right);\theta\in\Theta,F\text{ biliniyor}\right\}\tag{3.20}$$

ise θ parametresine F sınıfı için bir ölçek parametresi denir. θ parametresinin X rasgele değişkeni için ölçek parametresi olması için gerek ve yeter koşul $\frac{X}{\theta}$ rasgele değişkeninin dağılımının θ dan bağımsız olmasıdır.

$$P\left(\frac{X}{\theta}\le x\right)=P(X\le\theta x)=F_\theta(\theta x)=F\left(\frac{\theta x}{\theta}\right)=F(x).$$

Örnek olarak,

$$\mathrm{F}=\left\{F_{\mu,\sigma}(x){:}F_{\mu,\sigma}(x)=F\left(\frac{x-\mu}{\sigma}\right)=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{x}e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}dt;\mu\in R,\sigma^2\in R^+\right\}\tag{3.21}$$

sınıfı için μ konum, σ ölçek parametresidir.

p_0. dereceli kuantilleri aynı olan dağılımlar sınıfı için dağılımdan bağımsız bir istatistik de aşağıdaki gibi verilir.

$X_1, X_2, \ldots, X_n$, $i = 1,2,\ldots,n$ için F_i dağılımına sahip bir örneklem ve $i = 1,2,\ldots,n$ için $F_i(\theta) = p_0$ olsun.

$$\Psi(t) = \begin{cases} 1 & , \quad t > 0 \\ 0 & , \quad t \leq 0 \end{cases}$$

biçiminde tanımlanan bir fonksiyon ve θ_0 bilinmek üzere

$$\Psi_i = \Psi(X_i - \theta_0) = \begin{cases} 1 & , \quad X_i > \theta \\ 0 & , \quad X_i \leq \theta_0 \end{cases}$$

fonksiyonunu tanımlayalım. $\theta = \theta_0$ olduğunda Ψ_i rasgele değişkenleri başarı olasılığı $1 - p_0$ olan bağımsız Bernoulli dağılımına sahip olur ve $B = B(\Psi_1, \Psi_2, \ldots, \Psi_n) = \sum_{i=1}^{n} \Psi_i$ istatistiğinin dağılımı n ve $1 - p_0$ parametreli binom dağılımına sahiptir (Randles ve Wolfe 1979).

Tanım 3.2.4. X ve Y aynı olasılık uzayında tanımlanmış iki rasgele değişken olmak üzere $\forall z \in R$ için $P(X \leq z) = P(Y \leq z)$ ise X ve Y rasgele değişkenlerine aynı dağılıma sahiptir denir ve $X \stackrel{d}{=} Y$ ile gösterilir. X ve Y rasgele değişkenleri aynı dağılıma sahipse aşağıdaki özellikler sağlanır.

i. $X \stackrel{d}{=} X$

ii. $X \stackrel{d}{=} Y \Leftrightarrow Y \stackrel{d}{=} X$

iii $X \stackrel{d}{=} Y$ ve $Y \stackrel{d}{=} Z \Rightarrow X \stackrel{d}{=} Z$

Teorem 3.2.5. $X_1, X_2, \ldots, X_n$ bağımsız ve her X_i rasgele değişkeni μ_i etrafında simetrik dağılıma sahip olsun. Bu takdirde

$$(X_1 - \mu_1, X_2 - \mu_2, \ldots X_n - \mu_n) \stackrel{d}{=} (\mu_1 - X_1, \mu_2 - X_2, \ldots, \mu_n - X_n)$$

dir.

İspat.

$$P(X_1-\mu_1<x_1, X_2-\mu_2<x_2,\dots,X_n-\mu_n<x_n)=P(X_1-\mu_1<x_1)\times$$
$$P(X_2-\mu_2<x_2)\dots P(X_n-\mu_{n1}<x_n)$$

diğer taraftan

$$P(X_i-\mu_i<x_i)=P(X_i<\mu_i+x_i)=1-P(X<\mu_i-X_i)=P(\mu_i-X_i<x_i)$$

yukarıda yerine yazılırsa

$$\prod_{i=1}^{n}P(X_i-\mu_i<x_i)=\prod_{i=1}^{n}P(\mu_i-X_i<x_i)\ .$$

olur.

Teorem 3.2.6. X, F dağılım fonksiyonuna sahip sürekli bir rasgele değişken olmak üzere $Y=F(X)$ rasgele değişkeni [0,1] aralığında sürekli düzgün dağılıma sahiptir.

$X_1, X_2,\dots,X_n$, F dağılımına sahip bir örneklem olmak üzere $Y_{(i)}=F\left(X_{(i)}\right)$ rasgele değişkenleri [0,1] aralığındaki sürekli düzgün dağılımdan alınmış örneklemin sıra istatistikleridir (Rohatgi 1979).

3.3. Örneklem (Empirik) Dağılım Fonksiyonu

R, reel sayılar kümesini, B' de R'de tanımlı Borel cebirini göstermek üzere (R,B) uzayında P dağılımına sahip $X_1, X_2,..,X_n$ örneklemini gözönüne alalım. $\forall B\in$B için $P(X_i\in B)=P(B)$ ile gösterilsin

$$\mathrm{I}_{X_i}(B)=\begin{cases}0 & , \quad X_i\notin B\\ 1 & , \quad X_i\in B\end{cases}$$

biçiminde tanımlanan bir fonksiyon olmak üzere,

$$E\left(\mathrm{I}_{X_i}(B)\right)=0P\{X_i\notin B\}+1P\{X_i\in B\}=P(B)$$
$$\mathrm{var}\left(\mathrm{I}_{X_i}(B)\right)=E\left(\mathrm{I}_{X_i}(B)\right)^2-\left(E\left(\mathrm{I}_{X_i}(B)\right)\right)^2=P(B)(1-P(B))$$

dir. Dolayısıyla $\mathrm{I}_{X_1}(B), \mathrm{I}_{X_2}(B),\dots,\mathrm{I}_{X_n}(B)$ rasgele değişkenleri bağımsız ve her biri Bernoulli dağılımına sahiptir.

Tanım 3.3.1. $\nu_n^*(B)$; örneklemden B kümesine düşenlerin sayısını göstermek üzere

$$P_n^*(B)=\frac{\nu_n^*(B)}{n}=\frac{1}{n}\sum_{i=1}^{n}\mathrm{I}_{X_i}(B)$$

biçiminde tanımlanan $P_n^*(B)$'ye $X_1,X_2,\ldots,X_n$ örnekleminin örneklem (empirik) dağılımı denir. $B=(-\infty,x]$ alınması durumunda

$$P_n\left((-\infty,x]\right)=F_n(x)=\begin{cases}0 & , \quad x<X_{(1)}\\ \dfrac{k}{n} & , \quad X_{(k)}\le x<X_{(k+1)}\\ 1 & , \quad x\ge X_{(\ )}\end{cases} \qquad (3.22)$$

biçiminde tanımlanan $F_n^*(x)$ fonksiyonuna $X_1,X_2,\ldots,X_n$ örnekleminin örneklem (empirik) dağılım fonksiyonu denir (Gaeusler ve Stute 1987).

Teorem 3.3.2. $X_1,X_2,\ldots,X_n$ örnekleminin örneklem dağılımı $P_n^*(B)$ olmak üzere,

$$P_n^*(B)\xrightarrow{hhhy}P(B)$$

dir (Borovkov 1984).

İspat. $\mathrm{I}_{X_1}(B),\mathrm{I}_{X_2}(B),\ldots,\mathrm{I}_{X_n}(B)$ örneklemi ve güçlü büyük sayılar yasası göz önüne alınırsa;

$$P_n^*(B)=\frac{\nu_n^*(B)}{n}=\frac{1}{n}\sum_{i=1}^{n}\mathrm{I}_{X_i}(B)\xrightarrow{hhhy}E\left(\frac{1}{n}\sum_{i=1}^{n}\mathrm{I}_{X_i}(B)\right)=P(B)$$

özel olarak $B=(-\infty,x)$ alınması durumunda

$$F_n^*(x)\xrightarrow{hhh}F(x)$$

olur.

Teorem 3.3.3. Belli sabit bir $x_0\in R$ için $F_n^*(x_0)$, $\frac{j}{n}$ değerlerini alan $(j=1,2,\ldots,n)$

$E\left(F_n^*(x)\right)=F(x)$ ve $\operatorname{var}\left(F_n^*(x)\right)=\frac{1}{n}F(x)\left(1-F(x)\right)$ olan binom dağılımına sahip bir rasgele değişkendir (Gaeusler ve Stute 1987).

İspat.

$$I(x_0)=\begin{cases}1 & , \quad x_0\ge 0\\ 0 & , \quad x_0<0\end{cases}$$

olmak üzere $j=1,2,\ldots,n$ için;

$$I(x_0 - X_j) = \begin{cases} 1 \ , & x_0 - X_j \geq 0 \\ 0 \ , & x_0 - X_j < 0 \end{cases}$$

$$= \begin{cases} 1 \ , & x_0 \geq X_j \\ 0 \ , & x_0 < X_j \end{cases}$$

fonksiyonunu göz önüne alalım. Bu durumda

$$P\{I(x_0 - X_j) = 1\} = P\{x_0 - X_j \geq 0\} = P\{X_j \leq x_0\} = F(x_0)$$

$$P\{I(x_0 - X_j) = 0\} = P\{x_0 - X_j < 0\} = P\{X_j > x_0\} = 1 - F(x_0)$$

olur. $\sum_{j=0}^{n} I(x_0 - X_j)$ rasgele değişkeni ve Tanım 1.3.1 'de verilen örneklem dağılım fonksiyonu göz önüne alınırsa

$$P\left\{F_n^*(x) = \frac{j}{n}\right\} = C_n^j \left(F(x)\right)^j \left(1 - F(x)\right)^{n-j}$$

elde edilir.

$$E(F_n^*(x)) = \sum_{j=0}^{n} \frac{j}{n} C_n^j (F(x_0))^j (1 - F(x_0))^{n-j}$$

$$= \frac{1}{n} \sum_{j=0}^{n} \frac{j}{n} \frac{n!}{j!(n-j)!} (F(x_0))^j (1 - F(x_0))^{n-j}$$

$$= \sum_{j=1}^{n} C_{n-1}^{j-1} (F(x_0))^{j-1} (1 - F(x_0))^{n-j} F(x_0)$$

$j - 1 = t$ yazılırsa;

$$E(F_n^*(x)) = F(x_0) \sum_{t=0}^{n-1} C_{n-1}^t (F(x_0))^t (1 - F(x_0))^{(n-1)-t}$$

$$= F(x_0)(F(x_0) + (1 - F(x_0)))^{n-1} = F(x_0)$$

olur. Benzer yöntemle

$$\mathrm{var}(F_n^*(x_0)) = F(x_0)(1 - F(x_0))$$

bulunur.

Sonuç 3.3.4.

$$\frac{\sqrt{n}\left(F_m^*(x_0) - F(x_0)\right)}{\sqrt{F(x)(1 - F(x))}} \xrightarrow{d} N(0,1)$$

(Randles ve Wolfe 1979).

İspat. Teorem 3.3.3. ve merkezi limit teoreminin bir sonucu olarak elde edilir.

Teorem 3.3.5. (Glivenko-Cantelli). $X_1, X_2, \ldots, X_n$ örnekleminin dağılım fonksiyonu $F(x)$ ve örneklem dağılım fonksiyonu $F_n^*(x)$ olmak üzere $\forall \varepsilon > 0$ için

$$\sup_{x \in R} \left| F_n^*(x) - F(x) \right| \xrightarrow{hhhy} 0$$

dır.

İspat. $F(x)$; $x \in R$, sürekli bir dağılım fonksiyonu olsun. $F(x)$' in sürekli ve $0 \leq F(x) \leq 1$ olmasından dolayı $\forall \varepsilon > 0$ için

$$F(z_0) = 0 \ , \ F(z_1) = \frac{1}{N} = \varepsilon \ , \ F(z_2) = \frac{2}{N} = 2\varepsilon, \ldots, F(z_N) = \frac{N}{N} = N\varepsilon$$

olacak biçimde $z_0 = -\infty$, $z_1, z_2, \ldots, z_N \in R$ noktalarını seçebiliriz. $z \in \left[z_k, z_{k+1}\right)$ olmak üzere;

$$F(z_{k+1}) - F(z_k) = \varepsilon \tag{3.23}$$

eşitliğini kullanarak aşağıdaki eşitsizlikleri elde ederiz.

$$F_n^*(z) - F(z) \leq F_n^*(z_{k+1}) - F(z_k) = F_n^*(z_{k+1}) - F(z_{k+1}) + \varepsilon \tag{3.24}$$

$$F_n^*(z) - F(z) \geq F_n^*(z_k) - F(z_{k+1}) = F_n^*(z_k) - F(z_k) - \varepsilon \tag{3.25}$$

(3.24) ve (3.25) eşitsizlikleri kullanılarak,

$$F_n^*(z_k) - F(z_k) - \varepsilon \leq F_n^*(z) - F(z) \leq F_n^*(z_{k+1}) - F(z_{k+1}) + \varepsilon \tag{3.26}$$

eşitsizliği elde edilir. Teorem 3.3.2. kullanılarak,

$$A_k = \left\{ \omega : \left| F_n^*(z_k) - F(z_k) \right| \xrightarrow{hhhy} 0 \right\}$$

olayı tanımlanırsa $k = 1,2,\ldots,n$ için $P\left(A_k\right) = 1$ olur.

$$\forall \varepsilon > 0 \text{ ve } \omega \in A_k \text{ için } \exists n_k(\omega) \ni n > n_k(\omega) \text{ için } \left| F_n^*(z_k) - F(z_k) \right| < \varepsilon$$

olur. $\bigcap_{k=0}^{N} A_k = A$ alınırsa $P(A) = 1$ ve $\max\left(n_1, n_2, \ldots, n_N\right) = n'$, $n > n'$ için

$$\left| F_n^*(z) - F(z) \right| < \varepsilon$$

olur. Yani,

$$P\left\{ \omega : \sup_{z \in R} \left| F_n^*(z) - F(z) \right| < \varepsilon \right\} = P(A) = 1$$

dir.

Tanım 3.3.6. $X_1, X_2, \ldots, X_n, F \in \mathrm{F}_c = \{F : F \text{ sürekli dağılım fonksiyonu}\}$; dağılım fonksiyonuna sahip bir örneklem ve F_n^* örneklem dağılım fonksiyonu olmak üzere

$$D_n = \sup_x \left|F_n^*(x) - F(x)\right|$$

istatistiğine Kolmogorov - Simirnov istatistiği denir (Guenther 1975).

Teorem 3.3.7. D_n istatistiği F_c sınıfı için dağılımdan bağımsız bir istatistikdir (Borovkov 1984).

İspat.

$$\begin{aligned}
P\{D_n \le t\} &= P\left\{\sup_{x \in R}\left|F_n^*(x) - F(x)\right| \le t\right\} \\
&= P\left\{\sup_{u \in [o,1]}\left|F_n^*(F^{-1}(u)) - u\right| \le t\right\} \\
&= P\left\{\sup_{u \in [o,1]}\left|\frac{1}{n}\sum_{i=1}^{n} I\{X_i \le F^{-1}(u)\} - u\right| \le t\right\} \\
&= P\left\{\sup_{u \in [o,1]}\left|\frac{1}{n}\sum_{i=1}^{n} I\{F(X_i) \le u\} - u\right| \le t\right\} \\
&= P\left\{\sup_{u \in [o,1]}\left|V_n^*(u) - u\right| \le t\right\}
\end{aligned}$$

dır. Burada V_n^*, $[0,1]$ aralığında düzgün dağılıma sahip örneklemin, örneklem dağılım fonksiyonudur. Görüldüğü üzere D_n istatistiğinin dağılımı F_c sınıfına göre dağılımdan bağımsızdır.

Teorem 3.3.8. $\lim_{n\to\infty} P\{\sqrt{n}D_n \le t\} = 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1} e^{-2i^2t^2}$ dir (Borovkov 1984).

Teorem 3.3.9. $X_1, X_2, \ldots, X_m$ ve $Y_1, Y_2, \ldots, Y_n$ aynı $F(x)$ dağılım fonksiyonuna sahip iki örneklem ve F_m^*, $X_1, X_2, \ldots, X_m$ nin örneklem dağılım fonksiyonu olmak üzere $S_j = F_m^*(Y_j)$, $j = 1,2,\ldots,n$ rasgele değişkenlerinin olasılık dağılımı $\left\{0, \frac{1}{m}, \frac{2}{m}, \ldots, \frac{m-1}{m}, 1\right\}$ noktaları üzerinde düzgün dağılıma sahiptir (Guenter 1975).

İspat .

$$S_j = F_m^*(Y_j) = \begin{cases} 0 & , \quad Y_j < X_{(1)} \\ \dfrac{i}{m} & , \quad X_{(i)} \le Y_j < X_{(i+1)} \\ 1 & , \quad Y_j \ge X_{(m)} \end{cases}$$

$$P\left\{S_j = \frac{i}{m}\right\} = P\{X_{(i)} \le Y_j < X_{(i+1)}\}, \quad i = 1,2,\ldots,m-1, \quad j = 1,2,\ldots,n$$

$$= \int_{-\infty}^{\infty} P\{X_{(i)} \le Y_j < X_{(i+1)} / Y_j = t\} f(t)dt \quad (3.3.)\text{'den}$$

$$= \int_{-\infty}^{\infty} \binom{m}{i} (F(t))^i (1-F(t))^{m-i} f(t)dt$$

$F(t) = u$ dönüşümü yapılırsa

$$P\left\{S_j = \frac{i}{m}\right\} = \frac{1}{m} \int_0^1 \frac{(m+1)!}{i!(m-i)!} u^i (1-u)^{m-i} du$$

$$= \frac{1}{m+1}$$

elde edilir. $i = 0$ ve m içinde benzer biçimde aşağıdaki gibi bulunur.

$$P\{S_j = 0\} = P\{Y_j < X_{(1)}\}$$

$$= 1 - P\{X_{(1)} \le Y_j\}$$

$$= 1 - \int_{-\infty}^{\infty} P\{X_{(1)} \le Y_j / Y_j = t\} f(t)dt$$

$$= 1 - \int_{-\infty}^{\infty} \sum_{u=1}^{m} C_m^u (F(t))^u (1-F(t))^{m-u} f(t)dt$$

$F(t) = \omega$ dönüşümü yapılırsa

$$P\{S_j = 0\} = 1 - \sum_{u=1}^{m} C_m^u \int_0^1 \omega^u (1-\omega)^{m-u} d\omega$$

$$= 1 - \sum_{u=1}^{m} \frac{1}{m+1} = \frac{1}{m+1}$$

dır.

$$P\{S_j = 1\} = P\{X_{(m)} < Y_j\} = \int_{-\infty}^{\infty} P\{X_{(m)} < Y_j / Y_j = t\} f(t)dt$$

$$= \int_{-\infty}^{\infty} (F(t))^m f(t)dt = \frac{1}{m+1}.$$

Teorem 3.3.10. $X_1, X_2, ..., X_m$ ve $Y_1, Y_2, ..., Y_n$ aynı F dağılım fonksiyonlarına sahip bağımsız iki örneklem ve F_m^* , $X_1, X_2, ..., X_m$ ' in örneklem dağılım fonksiyonu olmak üzere;

$$S_{(j)} = F_m^*\left(Y_{(j)}\right) \ , \ j = 1,2,...,n$$

rasgele değişkenlerinin olasılık dağılımı $\left\{0, \frac{1}{m}, \frac{2}{m}, ..., \frac{m-1}{m}, 1\right\}$ noktaları üzerinde negatif hipergeometrik dağılıma sahiptir (Guenter 1975).

İspat . Örneklem dağılım fonksiyonunun tanımını kullanırsak (3.22) den

$$S_{(j)} = F_m^*\left(Y_{(j)}\right) = \begin{cases} 0 & , \ Y_{(j)} < X_{(1)} \\ \dfrac{i}{m} & , \ X_{(i)} \le Y_{(j)} < X_{(i+1)} \\ 1 & , \ Y_{(j)} \ge X_{(m)} \end{cases}$$

olur.

$$P\left\{S_{(j)} = \frac{i}{m}\right\} = P\left\{X_{(i)} \le Y_{(j)} < X_{(i+1)}\right\}$$

$$= \int_{-\infty}^{\infty} P\left\{X_{(i)} \le Y_{(j)} < X_{(i+1)} / Y_{(j)} = t\right\} f_j(t)dt$$

$$= \int_{-\infty}^{\infty} \binom{m}{i} (F(t))^i (1 - F(t))^{m-i} \frac{n!}{(j-1)!(n-j)!} (F(t))^{j-1} (1 - F(t))^{n-j} f(t)dt$$

$F(t) = u$ dönüşümü yapılırsa

$$P\left\{S_{(j)} = \frac{i}{m}\right\} = \binom{m}{i} \frac{n!}{(j-1)!(n-j)!} B(i + j, m + n - i - j + 1)$$

$$= \frac{\binom{m+n-i-j}{m-i}\binom{i+j-1}{i}}{\binom{m+n}{n}}, \qquad i = 0,1,2,...,m$$

şeklinde elde edilir.

Teorem 3.3.9. Teorem 3.3.4. ve Teorem 3.3.5. ' de verilen S_j ve $S_{(j)}$ rasgele değişkenlerinin beklenen değer ve varyansları,

$$E(S_i) = \frac{1}{2}\ ,\ E\left(S_{(j)}\right) = \frac{j}{n+1}$$

$$\mathrm{var}(S_i) = \frac{1}{12} + \frac{1}{6m}\ ,\ \mathrm{var}\left(S_{(j)}\right) = \frac{j(n-j+1)(m+n+1)}{m(n+1)^2(n+2)}$$

olarak elde edilir.

4. DAĞILIMDAN BAĞIMSIZ GÜVEN ARALIKLARI VE TOLERANS ARALIKLAR

Bu kısımda kitle parametresi bilinmeyen ve belli bir dağılımlar ailesinden alınan bir örneklemin parametresini $1-\alpha$ olasılıkla içeren, alt ve üst sınırları örneklemin bir fonksiyonu olan rasgele aralıklarla ilgileneceğiz.

Tanım 4.1. $X_1, X_2, \ldots, X_n$, $F_\theta, \theta \in \Theta$ (θ bilinmiyor) dağılımına sahip bir örneklem olsun. L ve U, $\forall x \in R^n$ için $L(x) \le U(x)$ olacak biçimde reel değerli iki fonksiyon olmak üzere $\forall \theta \in \Theta$ için

$$P_\theta\left(L(X_1, X_2, \ldots, X_n) \le \theta \le U(X_1, X_2, \ldots, X_n)\right) = 1-\alpha$$

ise

$$S(X_1, X_2, \ldots, X_n) = \left(L(X_1, X_2, \ldots, X_n), U(X_1, X_2, \ldots, X_n)\right)$$

rasgele aralığına θ parametresi için $1-\alpha$ seviyeli dağılımdan bağımsız güven aralığı denir (Randles ve Wolfe 1979).

$$L(X_1, X_2, \ldots, X_n) = -\infty \text{ veya } U(X_1, X_2, \ldots, X_n) = \infty$$

olması durumunda θ için

$$S(X_1, X_2, \ldots, X_n) = \left(-\infty, U(X_1, X_2, \ldots, X_n)\right)$$

veya

$$S(X_1, X_2, \ldots, X_n) = \left(L(X_1, X_2, \ldots, X_n), \infty\right)$$

olacak biçimde tek taraflı güven aralıkları oluşturulabilir. Örnek olarak,

$$F_1 = \left\{ F\left(\frac{x-\mu}{\sigma}\right) : F\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt \,, \begin{array}{l} \sigma^2 > 0 \text{ biliniyor} \\ \mu \in R \text{ bilinmiyor} \end{array} \right\} \tag{4.1}$$

olmak üzere μ için $1-\alpha$ seviyeli dağılımdan bağımsız güven aralığı

$$S(X_1, X_2, \ldots, X_n) = \left(\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

şeklindedir.

$$F_2 = \left\{ F\left(\frac{x-\mu}{\sigma}\right) : : F\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt \,, \quad \begin{matrix} \sigma^2 > 0 \text{ bilinmiyor} \\ \mu \in R \text{ bilinmiyor} \end{matrix} \right\} \quad (4.2)$$

olmak üzere σ^2 için $1-\alpha$ seviyeli dağılımdan bağımsız güven aralığı

$$S(X_1, X_2, \ldots, X_n) = \left(\overline{X} - t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}, \overline{X} + t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}} \right) \quad (4.3)$$

dır. Burada; $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ ve $z_{\frac{\alpha}{2}}$, $t_{\frac{\alpha}{2}, n-1}$ ' de sırasıyla standart normal dağılımda ve $n-1$ serbestlik dereceli t dağılımında $1-\alpha$ olasılığına karşılık gelen noktalardır.

4.1. Bayes Güven Aralıkları

Tanım 4.1.1. $X_1, X_2, \ldots, X_n$, $F_\theta, \theta \in \Theta$ (θ bilinmiyor) dağılımına sahip bir örneklem olsun. θ parametresini tahmin etmek amacıyla kullanılan ve

$$\delta : R^n \to R$$
$$: (x_1, x_2, \ldots, x_n) \to \delta\ (x_1, x_2, \ldots, x_n)$$

biçiminde tanımlanan δ ölçülebilir fonksiyonuna karar fonksiyonu denir.

Tanım 4.1.2. θ yerine $\delta\ (x_1, x_2, \ldots, x_n) = \hat{\theta}$ almakla uğranılan kaybı temsil eden fonksiyona kayıp fonksiyonu denir ve $L(\theta, \delta)$ şeklinde gösterilir. Genellikle aşağıdaki kayıp fonksiyonları kullanılır.

$$L(\theta,\delta) = (\theta - \delta)^2 \quad , \quad L(\theta,\delta) = \frac{|\theta - \delta|}{5} \quad , \quad L(\theta,\delta) = \cos(\theta - \delta)$$

En çok kullanılan kayıp fonksiyonu $L(\theta,\delta) = (\theta - \delta)^2$ biçimindeki fonksiyondur.

Tanım 4.1.3. θ sabit olmak üzere kayıp fonksiyonunun beklenen değerine "risk fonksiyonu "denir.

$$R(\theta,\delta) = E[L(\theta,\delta)]$$

$$= \begin{cases} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} L[\theta, \delta(x_1,\ldots,x_n)] f(x_1,\theta) \ldots f(x_n,\theta) dx_1 \ldots dx_n & , \text{ sürekli durum} \\ \sum_{x_1} \cdots \sum_{x_n} L[\theta, \delta(x_1,\ldots,x_n)] f(x_1,\theta) \ldots f(x_n,\theta) & , \text{ kesikli durum} \end{cases}$$

biçiminde tanımlanır.

δ_1 ve δ_2 gibi iki ayrı karar için;

$$R(\theta,\delta_1) < R(\theta,\delta_2) \tag{4.4}$$

ilişkisi θ ' nın her değeri için gerçeklenebilir. Bu takdirde "riskin küçüğü büyüğünden daha iyidir" ölçüsüne göre δ_1 kararının δ_2 kararından daha iyi olduğu söylenebilir.

(4.4) ilişkisi $\forall\theta$ için gerçeklenmeyebilir. θ ' nın bazı değerleri için $R(\theta,\delta_1) < R(\theta,\delta_2)$ ilişkisi ve başka değerleri için de $R(\theta,\delta_1) > R(\theta,\delta_2)$ ilişkisi gerçeklenebilir. Bu takdirde δ_1 ve δ_2 ' den birinin diğerinden daha iyi olduğunu söylemek mümkün değildir.

δ ' nın alabileceği değerler için $R(\theta,\delta)$ ' ların maksimumlarının minimumunu doğuran δ_0 kararı en iyi karardır. Bu tür bir karar fonksiyonuna minimax karar fonksiyonu denir ve

$$\max_{\theta} R(\theta,\delta_0) = \min_{\delta\in D} \max_{\theta} R(\theta,\delta)$$

şeklinde belirlenir. Burada D ile δ kararlarının tamamının kümesi gösterilmektedir.

Tanım 4.1.4. $X_1, X_2, \ldots, X_n$, sürekli $F_\theta, \theta\in\Theta$ (θ bilinmiyor) dağılımına sahip bir örneklem, θ olasılık yoğunluk fonksiyonu $\pi(\theta)$ olan bir rasgele değişken olmak üzere ortalama risk;

$$r(\pi,\delta) = E[R(\theta,\delta)] = \begin{cases} \sum\limits_{\Theta} R(\theta,\delta)\pi(\theta)d\theta & , \text{ kesikli durum} \\ \int\limits_{\Theta} R(\theta,\delta)\pi(\theta)d\theta & , \text{ sürekli durum} \end{cases}$$

biçiminde verilir. $r(\pi,\delta)$ ' yı minimum yapan δ_0 kararına ise "Bayes karar fonksiyonu" veya "Bayes kararı"denir. Şimdi sürekli durumu ele alalım ve bayes karar fonksiyonunu elde edelim. $\delta_0 = \min\, r(\pi,\delta)$ olmak üzere;

$$\begin{aligned} r(\pi,\delta) &= \int\limits_{\Theta} R(\theta,\delta)\pi(\theta)d\theta \\ &= \int\limits_{\Theta}\left\{\int\limits_{-\infty}^{\infty}\cdots\int\limits_{-\infty}^{\infty}[\theta-\delta(x_1,\ldots,x_n)]^2 f(x_1,\theta)\ldots f(x_n,\theta)dx_1\ldots dx_n\right\}\pi(\theta)d\theta \\ &= \int\limits_{-\infty}^{\infty}\cdots\int\limits_{-\infty}^{\infty}\left\{\int\limits_{\Theta}[\theta-\delta(x_1,\ldots,x_n)]^2 f(x_1,\theta)\ldots f(x_n,\theta)\pi(\theta)d\theta\right\}dx_1\ldots dx_n \end{aligned} \tag{4.5}$$

$r(\pi,\delta)$ fonksiyonunu minimize etmek için (4.5) eşitliğindeki θ' nın fonksiyonu olan

$$\int_{\Theta}[\theta-\delta(x_1,\ldots,x_n)]^2 f(x_1,\theta)\ldots f(x_n,\theta)\pi(\theta)d\theta \tag{4.6}$$

ifadesini minimize etmeliyiz. (4.6) ifadesi açılırsa;

$$\begin{aligned}\int_{\Theta}[\theta-\delta(x_1,\ldots,x_n)]^2 f(x_1,\theta)\ldots f(x_n,\theta)\pi(\theta)d\theta = \\ \delta^2(x_1,\ldots,x_n)\int_{\Theta} f(x_1,\theta)\ldots f(x_n,\theta)\pi(\theta)d\theta - \\ 2\delta(x_1,\ldots,x_n)\int_{\Theta}\theta f(x_1,\theta)\ldots f(x_n,\theta)\pi(\theta)d\theta + \\ \int_{\Theta}\theta^2 f(x_1,\theta)\ldots f(x_n,\theta)\pi(\theta)d\theta\end{aligned} \tag{4.7}$$

olur. Görüldüğü gibi (4.7)' nin sağ tarafı $g(t)=at^2-2bt+c$ şeklindedir. Bu $g(t)$ fonksiyonu minimize edilirse;

$$\frac{dg(t)}{dt}=2at-2b=0 \Rightarrow t=\frac{b}{a}$$

elde edilir. Yani bayes karar fonksiyonu;

$$\delta(x_1,\ldots,x_n)=\frac{\int_{\Theta}\theta f(x_1,\theta)\ldots f(x_n,\theta)\pi(\theta)d\theta}{\int_{\Theta} f(x_1,\theta)\ldots f(x_n,\theta)\pi(\theta)d\theta}$$

biçiminde elde edilmiş olur.

Θ parametreler sınıfı $X=(X_1,X_2,\ldots,X_n)$ sürekli $f_\theta(x)$, $\theta\in\Theta$ (θ bilinmiyor) olasılık yoğunluk fonksiyonuna sahip bir rasgele vektör, θ olasılık yoğunluk fonksiyonu $\pi(\theta)$ olan bir rasgele değişken olmak üzere X ve θ ' nın ortak olasılık yoğunluk fonksiyonu;

$$f(x,\theta)=\pi(\theta)f_\theta(x)$$

X 'in marjinal dağılımı;

$$g(x)=\begin{cases}\sum_{\Theta}\pi(\theta)f_\theta(x) & , \ \theta \text{ kesikli ise} \\ \int_{\Theta}\pi(\theta)f_\theta(x) & , \ \theta \textit{ sürekli ise}\end{cases}$$

ve $X=x$ verildiğinde θ ' nın koşullu dağılımı ise;

$$h(\theta \setminus \mathrm{x}) = \frac{\pi(\theta) f_\theta(x)}{g(x)} \quad , \qquad g(x) > 0$$

dir. Buradaki $\pi(\theta)$ fonksiyonuna θ' nın prior (önsel), $h(\theta \setminus \mathrm{x})$ fonksiyonuna ise θ' nın posterior (sonsal) olasılık yoğunluk fonksiyonu denir.

Tanım 4.1.5.

$L(X) = E[h(\theta \setminus \mathrm{X})] - \mathrm{var}[h(\theta \setminus \mathrm{X})] z_{\alpha/2}$ ve $U(X) = E[h(\theta \setminus \mathrm{X})] + \mathrm{var}[h(\theta \setminus \mathrm{X})] z_{\alpha/2}$ olmak üzere θ parametresini içeren $1-\alpha$ olasılığına sahip $(L(X), U(X))$ aralığına θ için $1-\alpha$ seviyesinde bayes güven aralığı denir.

Açıktır ki; bayes güven aralığı $1-\alpha$ seviyesindeki bütün güven aralıklar arasında en dar olan aralıktır.

Örnek 4.1.6. $X_1, X_2, \ldots, X_n \sim N(\mu, 1)$, $\mu \in R$ dağılımına sahip bir örneklem ve $\mu \sim N(0,1)$ dağılımına sahip bir rasgele değişken olsun.

$$h(\mu \setminus \mathrm{x}) = \frac{f(x,\mu)}{g(x)} = \frac{\pi(\mu) f_\mu(x)}{g(x)}$$

$$g(x) = \int_{-\infty}^{\infty} f(x,\mu) d\mu = \frac{1}{(2\pi)^{(n+1)/2}} e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2} \int_{-\infty}^{\infty} e^{-\frac{n+1}{2}\left(\mu^2 - 2\mu \frac{n\bar{x}}{n+1}\right)} d\mu$$

$$= \frac{(n+1)^{-1/2}}{(2\pi)^{n/2}} e^{\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2 + \frac{n^2 \bar{x}^2}{2(n+1)}\right)}$$

olmak üzere;

$$h(\mu \setminus \mathrm{x}) = \frac{1}{\sqrt{2\pi/(n+1)}} e^{-\frac{n+1}{2}\left(\mu - \frac{n\bar{x}}{n+1}\right)^2}$$

olur ve bayes tahmini;

$$\delta(x_1, \ldots, x_n) = E[h(\mu \setminus \mathrm{x})] = \frac{n\bar{x}}{n+1}$$

şeklinde elde edilir.

Ortalama risk ise;

$$
\begin{aligned}
r(\pi,\delta) &= \int_{\mu\in R} R(\mu,\delta)\pi(\mu)d\mu \\
&= \int_{\mu\in R}\left\{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}[\mu-\delta(x_1,\ldots,x_n)]^2 f_\mu(x)dx_1\ldots dx_n\right\}\pi(\mu)d\mu \\
&= \int_{\mu\in R}\left\{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}[\mu-\delta(x_1,\ldots,x_n)]^2 f(x_1,\mu)\ldots f(x_n,\mu)dx_1\ldots dx_n\right\}\pi(\mu)d\mu \\
&= \int_{\mu\in R}\left\{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}[\mu-\frac{n\bar{x}}{n+1}]^2 f(x_1,\mu)\ldots f(x_n,\mu)dx_1\ldots dx_n\right\}\pi(\mu)d\mu \\
&= \int_{-\infty}^{\infty} E_\mu[\mu-\frac{n\bar{x}}{n+1}]^2\pi(\mu)d\mu \\
&= \int_{-\infty}^{\infty}(n+1)^{-2}(n+\mu^2)\pi(\mu)d\mu \\
&= \frac{1}{n+1}
\end{aligned}
$$

biçimindedir. O halde;

$$
h(\mu\setminus \mathrm{x}) \sim N\left(\frac{\sum_{i=1}^{n} x_i}{n+1}, \frac{1}{n+1}\right)
$$

dır ve $1-\alpha$ seviyesindeki bayes güven aralığı ise;

$$
\left(\frac{n\bar{x}}{n+1}-\frac{z_{\alpha/2}}{\sqrt{n+1}}, \frac{n\bar{x}}{n+1}+\frac{z_{\alpha/2}}{\sqrt{n+1}}\right)
$$

şeklindedir. Bilindiği gibi μ için $1-\alpha$ seviyeli dağılımdan bağımsız güven aralığı

$$
\left(\overline{X}-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X}+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)
$$

idi. Bu aralık görüldüğü gibi bayes aralığına oranla daha geniştir. Yani, bayes güven aralığının daha iyi olduğu söylenebilir.

4.2. Kuantiller İçin Dağılımdan Bağımsız Güven Aralıkları

Tanım 4.2.1 X rasgele değişkeni F dağılım fonksiyonuna sahip olsun. $0 \le p \le 1$ olmak üzere $\xi_p = \inf\{x{:}F(x) \ge p\}$ değerine F dağılım fonksiyonuna ait p. dereceden kuantil denir (David 1970).

Tanım 4.2.2 $X_1, X_2, \ldots, X_n$, F dağılım fonksiyonuna sahip bir örneklem, ξ_p F dağılım fonksiyonunun p. dereceden kuantili ve $\forall (u_1, u_2, \ldots, u_n) \in R^n$ için $f_1(u_1, u_2, \ldots, u_n) \le f_2(u_1, u_2, \ldots, u_n)$ koşuluna sahip iki borel ölçülebilir fonksiyon olmak üzere;

$$P\{\xi_p \in (f_1(X_1, X_2, \ldots, X_n), f_2(X_1, X_2, \ldots, X_n))\} = \beta = sabit$$

oluyorsa o zaman $(f_1(X_1, X_2, \ldots, X_n), f_2(X_1, X_2, \ldots, X_n))$ rasgele aralığına ξ_p kuantili için dağılımdan bağımsız güven aralığı denir.

Kuantiller için sıra istatistikleri kullanılarak dağılımdan bağımsız güven aralıkları aşağıdaki biçimde oluşturulabilir. X_i' ler sürekli rasgele değişkenler olmak üzere ;

$$\begin{aligned} P\left(X_{(r)} \le \xi_p \le X_{(s)}\right) &= P\left(X_{(r)} \le \xi_p\right) - P\left(X_{(s)} < \xi_p\right) \\ &= \sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i}. \end{aligned}$$

dır. X_i' lerin kesikli rasgele değişkenler olması durumunda ise $P\left(X \le \xi_p\right) \ge p$ ve $P\left(X < \xi_p\right) \le p$ ifadelerinden

$$P\left(X_{(r)} \le \xi_p\right) \ge I_p(r, n-r+1), \quad P\left(X_{(s)} < \xi_p\right) \le I_p(s, n-s+1)$$

olur. Dolayısıyla

$$P\left(X_{(r)} \le \xi_p \le X_{(s)}\right) \ge \sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i}$$

ve

$$P\left(X_{(r)} < \xi_p < X_{(s)}\right) \le \sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i}$$

elde edilir (David 1970).

4.3. Dağılımdan Bağımsız Tolerans Aralıkları

Tolerans aralıklar da, güven aralıkları gibi sınırları rasgele değişkenler olan aralıklardır. Güven aralıkları, kitle parametresini belli bir olasılıkla örten aralıklar olduğu halde tolerans aralıklar kitlenin kendisini belli bir olasılıkla örten aralıklardır (David 1970).

$X_1, X_2, \ldots, X_n$, $f(x)$ olasılık yoğunluk fonksiyonuna sahip bir örneklem $L(X_1, X_2, \ldots, X_n)$, $V(X_1, X_2, \ldots, X_n)$ örneklemin iki fonksiyonu ve $0 \le \beta, \gamma \le 1$ olmak üzere

$$P\left(\int_V^L f(x)dx \ge \gamma\right) = \beta$$

oluyorsa $(L(X_1, X_2, \ldots, X_n), V(X_1, X_2, \ldots, X_n))$ rasgele aralığına kitleyi en az γ olasılığı ile kapsaması olasılığı β olan tolerans aralık denir. Örnek olarak,

$$L(X_1, X_2, \ldots, X_n) = X_{(r)} \text{ ve } V(X_1, X_2, \ldots, X_n) = X_{(s)}, \quad r < s$$

alınırsa

$$P\left(\int_{X_{(r)}}^{X_{(s)}} f(x)dx \ge \gamma\right) = P\left(F\left(X_{(s)}\right) - F\left(X_{(r)}\right) \ge \gamma\right)$$

bulunur. (3.18) ve (3.4) ifadelerinden

$$P\left(\int_{X_{(r)}}^{X_{(s)}} f(x)dx \ge \gamma\right) = P\left(W_{r,s} \ge \gamma\right) = 1 - I_\gamma(s-r, n-s+r+1)$$

$$= 1 - \sum_{i=s-r}^{n} C_n^i (\gamma)^i (1-\gamma)^{n-i}$$

dır. $r = 1$, $s = n$ alınması ve $\gamma = 0{,}95$, $\beta = 0{,}90$ olması durumunda $n = 77$ bulunur. $X_{(r)} = -\infty$ ya da $X_{(s)} = \infty$ alınarak tek taraflı tolerans aralıklar oluşturulabilir. $f(x)$ kesikli olasılık fonksiyonları olması durumunda

$$P\left(\sum_{x=X_{(r)}}^{X_{(s)}} f(x) \ge \gamma\right) \ge 1 - I_\gamma(s-r, n-s+r+1)$$

dır (David 1970).

5. ANA KİTLEYİ KAPSAYAN İNVARYANT GÜVEN ARALIKLARI VE SEVİYELERİNİN TAHMİN EDİLMESİ

Bu kısımda parametrik dağılımlar ailesi için ana kitleyi kapsayan invaryant güven aralıklarıyla ilgilenilmiştir.

5.1. Parametrik Aileler İçin Ana Kitleyi Kapsayan İnvaryant Güven Aralıkları

Farzedelim ki $\mathrm{P} = \left\{P_\theta, \theta \in \Theta\right\}$ parametrik dağılım fonksiyonlarının bir ailesi, f_1 ve f_2 fonksiyonları;

$$\forall (x_1, x_2, \ldots, x_n) \in R^n \text{ için } f_1(x_1, x_2, \ldots, x_n) \leq f_2(x_1, x_2, \ldots, x_n) \tag{5.1}$$

koşuluna sahip iki borel ölçülebilir fonksiyon, $X_1, X_2, \ldots, X_n$ de $P_\theta \in \mathrm{P}$ dağılımına sahip bir örneklem olsun. Bu örneklemden bağımsız aynı dağılıma sahip bir X_{n+1} rasgele değişkenin çekildiğini düşünelim.

Tanım 5.1.1. $\beta \in R$ olmak üzere , $\forall \theta \in \Theta$ için

$$P_\theta\left\{X_{n+1} \in \left(f_1\left(X_1, X_2, \ldots, X_n\right), f_2\left(X_1, X_2, \ldots, X_n\right)\right)\right\} = \beta \tag{5.2}$$

oluyorsa

$$\left(f_1\left(X_1, X_2, \ldots, X_n\right), f_2\left(X_1, X_2, \ldots, X_n\right)\right) \equiv J\left(X_1, X_2, \ldots, X_n\right)$$

rasgele aralığına P sınıfı için β güven seviyesine sahip ana kitleyi kapsayan invaryant güven aralığı denir.

$X_{n+1}, X_{n+2}, \ldots, X_{n+m}$; aynı dağılıma sahip ve $X_1, X_2, \ldots, X_n$ örnekleminden bağımsız yeni bir örneklem olmak üzere $\forall \theta \in \Theta$ için

$$\begin{aligned} P_\theta\{X_{n+1}, X_{n+2}, \ldots, X_{n+m} &\in (f_1(X_1, X_2, \ldots, X_n), f_2(X_1, X_2, \ldots, X_n))\} \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [F_\theta(f_2(u_1, u_2, \ldots, u_n)) - F_\theta(f_1(u_1, u_2, \ldots, u_n))]^m \, dF_\theta(u_1) \ldots dF_\theta(u_n) \\ &= E_\theta[F_\theta(f_2(X_1, X_2, \ldots, X_n)) - F_\theta(f_1(X_1, X_2, \ldots, X_n))]^m \end{aligned}$$

dır.

Teorem 5.1.2.

$$S_n(X_1,X_2,\ldots,X_n,\theta)=F_\theta(f_2(X_1,X_2,\ldots,X_n)-F_\theta(f_1(X_1,X_2,\ldots,X_n)$$

ve

$$G_\theta(u)=P_\theta\{S_n(X_1,X_2,\ldots,X_n,\theta)\le u\} \tag{5.3}$$

olmak üzere $S_n(X_1,X_2,\ldots,X_n,\theta)$ istatistiğinin dağılımı $\mathrm{P}=\{P_\theta,\theta\in\Theta\}$ sınıfı için dağılımdan bağımsız istatistik ise bu durumda

$$(f_1(X_1,X_2,\ldots,X_n),f_2(X_1,X_2,\ldots,X_n))$$

rasgele aralığı P sınıfı için invaryant güven aralıktır.

İspat.

$S_n(X_1,X_2,\ldots,X_n,\theta)$; P sınıfı için dağılımdan bağımsız olduğundan;

$\forall\theta\in\Theta$ için $P\{S_n(X_1,X_2,\ldots,X_n,\theta)\le u\}=G(u)$

dır.

$$P_\theta\{X_{n+1},\ldots,X_{n+m}\in(f_1(X_1,\ldots,X_n),f_2(X_1,\ldots,X_n))\}$$

$$=E_\theta[S_n(X_1,\ldots,X_n,\theta)]^m=\int_{-\infty}^{\infty}u^m dG(u)=\alpha_m \quad , \quad m=1,2,\ldots.$$

α_m; sabit bir sayı olduğundan $(f_1(X_1,\ldots,X_n),f_2(X_1,\ldots,X_n))$ P sınıfı için ana kitleyi kapsayan α_m seviyeli invaryat güven aralığıdır.

Teorem 5.1.3. $\mathrm{P}=\{P_\theta:\theta\in\Theta\}$ regüler bir model, yani; $\dfrac{dG_\theta(x)}{dx}=g_\theta(x)$ mevcut ve sürekli bir fonksiyon ve $m(\theta)\equiv E_\theta S^m$ olmak üzere $\forall\theta\in\Theta$ için;

$$m(\theta)=\int u^m\frac{dg_\theta(x)}{d\theta}du \quad , m=1,2,\ldots$$

biçiminde türevlenebilir bir fonksiyon olsun. Şayet;

$$P_\theta\{X_{n+1},\ldots,X_{n+m}\in(f_1(X_1,\ldots,X_n),f_2(X_1,\ldots,X_n))\}=d_m \ , \ m=1,2,\ldots$$

θ ' dan bağımsız ise $S_n(X_1,\ldots,X_n,\theta)$ istatistiğinin dağılımı da θ dan bağımsızdır.

İspat.

$$P_\theta\{X_{n+1},\ldots,X_{n+m}\in(f_1(X_1,\ldots,X_n),f_2(X_1,\ldots,X_n))\}=\int u^m dG_\theta(u)=d_m, m=1,2,\ldots$$

Model regüler olduğundan $\int u^m\dfrac{dg_\theta(u)}{d\theta}=0$, $m=1,2,\ldots$ ' dır. Stone Weiserstrass teoreminden (Rudin 1959) $\frac{dg_\theta(u)}{d\theta} = 0$ olur ve bu da $\frac{dG_\theta(u)}{du} = g_\theta(u)$ fonksiyonunun θ 'dan bağımsız olduğunu gösterir.

Tanım 5.1.4. $a = (a_1, a_2, \ldots, a_n) \in R^n, b = (b_1, b_2, \ldots, b_n) \in R^n$ için;

$$D^+ = \{(u_1, u_2, \ldots, u_n) : u_1 \geq u_2 \geq \ldots \geq u_n\}$$

biçiminde tanımlanan bir küme olsun. $a_{[1]} \geq a_{[2]} \geq \ldots \geq a_{[n]}$ ve $b_{[1]} \geq b_{[2]} \geq \ldots \geq b_{[n]}$ $a \in R^n$ ve $b \in R^n$ vektörlerinin büyüklük sırasına göre dizilmiş şekillerini göstermek üzere;

$$\begin{aligned} &1.\ \sum_{i=1}^{n} a_{[i]} = \sum_{i=1}^{n} b_{[i]} \\ &2.\ \sum_{i=1}^{k} a_{[i]} \leq \sum_{i=1}^{k} b_{[i]} \qquad k = 1, 2, \ldots, n-1 \end{aligned} \tag{5.4}$$

koşulları sağlanıyorsa a ve b vektörlerine majorant vektörlerdir denir ve $a \prec b$ biçiminde gösterilir (Marshal ve Olkin. 1979).

Teorem 5.1.5. a ve b vektörlerinin birbirleriyle majorant olmaları için gerek ve yeter koşul;

$$\forall u = (u_1, \ldots, u_n) \in D^+ \text{ için} \sum_{i=1}^{n} a_i u_i \leq \sum_{i=1}^{n} b_i u_i$$

olmasıdır (Marshal ve Olkin 1979).

5.2. Konum ve Ölçek Parametresi İçeren Dağılımlar Ailesi İçin İnvaryant Güven Aralıkları

Teorem 5.1.2., Teorem 5.1.3 ve Tanım 5.1.4 kullanılarak konum ve ölçek parametreli dağılımlar ailesi için invaryant güven aralıkları oluşturulabilir.

$$\mathrm{P} = \{F_\theta(x) = F(x-\theta), \theta \in \Theta, F(x) \text{ biliniyor}\}$$

sınıfı için $X_1, X_2, \ldots, X_n$; $F_\theta \in \mathrm{P}$ dağılımına sahip bir örneklem ve bu örneklemin büyüklük sırasına göre dizilmiş hali $X_{[1]} \geq X_{[2]} \geq \ldots \geq X_{[n]}$ (burada $X_{[i]} = X_{(n-i+1)}$), $a \in R^n$ ve $b \in R^n$ vektörleri majorant vektörler olmak üzere;

$$f_1(X_1,X_2,...,X_n)=\sum_{i=1}^{n}a_iX_{[i]} \text{ ve } f_2(X_1,X_2,...,X_n)=\sum_{i=1}^{n}b_iX_{[i]} \qquad (5.5)$$

biçiminde tanımlansın. Bu durumda;

$$\begin{aligned}\overline{S_n}(X_1,...,X_n,\theta)&=F_\theta(f_2(X_1,X_2...,X_n))-F_\theta(f_1(X_1,X_2...,X_n))\\&=F_\theta\left(\sum_{i=1}^{n}b_iX_{[i]}\right)-F_\theta\left(\sum_{i=1}^{n}a_iX_{[i]}\right)\\&=F\left(\sum_{i=1}^{n}b_i\left(X_{[i]}-\theta\right)\right)-F\left(\sum_{i=1}^{n}a_i\left(X_{[i]}-\theta\right)\right)\\&=F\left(\sum_{i=1}^{n}b_i\left(X_{(n-i+1)}-\theta\right)\right)-F\left(\sum_{i=1}^{n}a_i\left(X_{(n-i+1)}-\theta\right)\right)\end{aligned} \qquad (5.6)$$

dır. $X_{(n-i+1)}-\theta$ rasgele değişkeninin dağılımı θ dan bağımsız olduğundan P sınıfı için $\overline{S_n}(X_1,...,X_n,\theta)$ istatistiğinin dağılımı bu rasgele değişkenin dağılımıyla aynıdır.

Özel olarak $a=\left(\frac{1}{n},\frac{1}{n},...,\frac{1}{n}\right)$ ve $b=\left(\frac{1}{n-1},\frac{1}{n-1},...,\frac{1}{n-1},0\right)$ alınırsa (5.6) eşitliğinden;

$$\overline{S_n}(X_1,...,X_n,\theta)=F\left(\frac{1}{n-1}\sum_{i=1}^{n-1}X_{(n-i+1)}-\theta\right)-F\left(\frac{1}{n}\sum_{i=1}^{n}X_{(n-i+1)}-\theta\right)$$

olur. Şimdi de ölçek parametresi içeren dağılımlar ailesi için $\overline{\overline{S_n}}(X_1,...,X_n,\theta)$ istatistiğinin dağılımının θ' dan bağımsızlığını araştıralım.

$$X_1,X_2,...,X_n,\ F_\theta(x)\in \mathrm{P}_1=\left\{F_\theta(x)=F\left(\frac{x}{\theta}\right),\theta\in\Theta,F(x)\text{ biliniyor}\right\}$$

dağılımına sahip bir örneklem, $f_1(X_1,X_2,...,X_n)$ ve $f_2(X_1,X_2,...,X_n)$ (5.1) eşitsizliğinde tanımlandığı gibi borel ölçülebilir iki fonksiyon olmak üzere;

$$\begin{aligned}\overline{\overline{S_n}}(X_1,...,X_n,\theta)&=F_\theta(f_2(X_1,X_2...,X_n))-F_\theta(f_1(X_1,X_2...,X_n))\\&=F_\theta\left(\sum_{i=1}^{n}b_iX_{[i]}\right)-F_\theta\left(\sum_{i=1}^{n}a_iX_{[i]}\right)\\&=F\left(\sum_{i=1}^{n}b_i\frac{X_{[i]}}{\theta}\right)-F\left(\sum_{i=1}^{n}a_i\frac{X_{[i]}}{\theta}\right)\\&=F\left(\sum_{i=1}^{n}b_i\frac{X_{(n-i+1)}}{\theta}\right)-F\left(\sum_{i=1}^{n}a_i\frac{X_{(n-i+1)}}{\theta}\right)\end{aligned} \qquad (5.7)$$

olur. $\frac{X_{(n-i+1)}}{\theta}$ rasgele değişkeninin dağılımı θ ' dan bağımsız olduğundan P_1 sınıfı için $\overline{\overline{S}}_n(X_1,\ldots,X_n,\theta)$ istatistiğinin dağılımı bu rasgele değişkenin dağılımıyla aynıdır.

$$\mathrm{P}_2=\left\{F_{\theta,\mu}(x):F_{\theta,\mu}(x)=F\left(\frac{x-\mu}{\theta}\right),\theta\in\Theta,\mu\in\Theta_1,F(x)\text{ biliniyor}\right\}$$

olmak üzere benzer biçimde;

$$\begin{aligned}S_n^*(X_1,X_2,\ldots,X_n,\theta,\mu)&=F_{\theta,\mu}\left(\sum_{i=1}^{n}b_iX_{(n-i+1)}\right)-F_{\theta,\mu}\left(\sum_{i=1}^{n}a_iX_{(n-i+1)}\right)\\&=F\left(\sum_{i=1}^{n}b_i\frac{X_{(n-i+1)}-\mu}{\theta}\right)-F\left(\sum_{i=1}^{n}a_i\frac{X_{(n-i+1)}-\mu}{\theta}\right)\end{aligned}\tag{5.8}$$

istatistiğinin dağılımı da P_2 sınıfı için aynıdır.

Üstel dağılım fonksiyonlarına sahip bir dağılımlar sınıfı için invaryant güven aralıkları aşağıdaki şekilde oluşturulmuştur.

$$X_1,X_2,\ldots,X_n;F_\theta(x)\in\mathrm{P}=\left\{F_\theta(x):F(\theta x)=1-e^{-\theta x},x\geq 0,\theta\geq 0,F(x)=1-e^{-x}\right\}$$

dağılımına sahip bir örneklem, $a\prec b$ olmak üzere

$\left(\sum_{i=1}^{n}a_iX_{[i]},\sum_{i=1}^{n}b_iX_{[i]}\right)$ rasgele aralığı göz önüne alınarak aşağıdaki iki teorem verilebilir.

Teorem 5.2.1. Θ parametreler sınıfı olmak üzere $\forall\,\theta\in\Theta$ için;

$$\begin{aligned}P_\theta\left\{X_{n+1}\in\left(\sum_{i=1}^{n}a_iX_{[i]},\sum_{i=1}^{n}b_iX_{[i]}\right)\right\}&=n!\left(\frac{1}{\prod_{j=1}^{n}\sum_{i=1}^{j}(a_i+1)}-\frac{1}{\prod_{j=1}^{n}\sum_{i=1}^{j}(b_i+1)}\right)\\&=\alpha_1(a_1,\ldots,a_n,b_1,\ldots,b_n)\equiv\alpha_1\end{aligned}\tag{5.9}$$

dır. Yani, $\left(\sum_{i=1}^{n}a_iX_{[i]},\sum_{i=1}^{n}b_iX_{[i]}\right)$ rasgele aralığı P sınıfı için α_1 seviyeli invaryant güven aralığıdır (Bairamov ve ark. 1999).

Teorem 5.2.2. $\forall\theta\in\Theta$ için;

$$P_\theta\left\{X_{n+1},X_{n+2}\ldots X_{n+m}\in\left(\sum_{i=1}^{n}a_iX_{[i]},\sum_{i=1}^{n}b_iX_{[i]}\right)\right\}=n!\sum_{k=0}^{m}\frac{(-1)^k\binom{m}{k}}{\prod_{j=1}^{n}\sum_{i=1}^{j}\{(m-k)a_i+kb_i+1\}}$$
$$=\alpha_m(a_1,\ldots,a_n,b_1,\ldots,b_n)\equiv\alpha_m \tag{5.10}$$

dir (Bairamov ve ark. 1999).

5.2.1. Düzgün dağılım fonksiyonlarına sahip bir dağılımlar sınıfı için invaryant güven aralıkları

$$X_1,X_2,\ldots,X_n;F_\theta(x)\in\mathrm{P}=\left\{F_\theta(x):F(\theta,x)=\frac{x-c}{d-c}\quad,\ c<x<d\ ,\ F(x)=x\right\}$$

dağılımına sahip bir örneklem, $a\prec b$ olmak üzere $\left(\sum_{i=1}^{n}a_iX_{[i]},\sum_{i=1}^{n}b_iX_{[i]}\right)$ rasgele aralığını göz önüne alarak aşağıdaki iki teoremi verebiliriz.

Teorem 5.2.1.1. $X_1,X_2,\ldots,X_n$ örnekleminden bağımsız yeni bir X_{n+1} rasgele değişkeninin çekildiğini düşünelim.

$$P_\theta\left\{X_{n+1}\in\left(\sum_{i=1}^{n}a_iX_{[i]},\sum_{i=1}^{n}b_iX_{[i]}\right)\right\}=\frac{1}{n+1}\sum_{i=1}^{n}(n-i+1)(b_i-a_i) \tag{5.11}$$
$$=\alpha_1(a_1,\ldots,a_n;b_1,\ldots,b_n)\equiv\alpha_1$$

dır. Yani, $\left(\sum_{i=1}^{n}a_iX_{[i]},\sum_{i=1}^{n}b_iX_{[i]}\right)$ rasgele aralığı P sınıfı için α_1 seviyeli invaryant güven aralığıdır.

İspat.

$G=\{(x_1,x_2\ldots,x_n,x_{n+1}):x_{n+1}\in(f_1(x_1,x_2\ldots,x_n),f_2(x_1,x_2\ldots,x_n)),\ c<x_n<x_{n-1}<\ldots<x_1<d\}$
olmak üzere;

$$P\{X_{n+1}\in(f_1(X_1,X_2,\ldots,X_n),f_2(X_1,X_2,\ldots,X_n))\}=P\{X_1,X_2,\ldots,X_n,X_{n+1}\in G\}$$
$$=\int_c^d\int_c^{x_1}\cdots\int_c^{x_{n-1}}\int_{f_1(x_1,\ldots,x_n)}^{f_2(x_1,\ldots,x_n)}n!dF(x_{n+1})dF(x_1)\ldots dF(x_n)$$

$$P\{X_{n+1} \in (f_1(X_1, X_2,..., X_n), f_2(X_1, X_2,..., X_n))\}$$

$$= n! \int_c^d \int_c^{x_1} ... \int_c^{x_{n-1}} [F(f_2(x_1, x_2,..., x_n)) - F(f_1(x_1, x_2,..., x_n))] dF(x_1)...dF(x_n)$$

$$P\{X_{n+1} \in (f_1(X_1, X_2,..., X_n), f_2(X_1, X_2,..., X_n))\}$$

$$= n! \int_c^d \int_c^{x_1} ... \int_c^{x_{n-1}} [F(f_2(x_1, x_2,..., x_n)) - F(f_1(x_1, x_2,..., x_n))] dF(x_1)...dF(x_n)$$

$$= n! \int_c^d \int_c^{x_1} ... \int_c^{x_{n-1}} \left[\left(\sum_{i=1}^{n} b_i \cancel{(\frac{x_i - c}{d - c})} \right) - \left(\sum_{i=1}^{n} a_i \cancel{(\frac{x_i - c}{d - c})} \right) \right] (\frac{1}{d-c})^n dx_n ... dx_2 dx_1$$?

$$= \frac{1}{n+1} \sum_{i=1}^{n} (n - i + 1)(b_i - a_i)$$

Teorem 5.2.1.2.

$X_1, X_2, ..., X_n$ örnekleminden bağımsız ve aynı dağılımdan alınan yeni bir $X_{n+1}, X_{n+2}, ..., X_{n+m}$ örneklemini ele alalım.

$$P\left\{ X_{n+1}, X_{n+2} ... X_{n+m} \in \left(\sum_{i=1}^{n} a_i X_{[i]}, \sum_{i=1}^{n} b_i X_{[i]} \right) \right\}$$

$$= \begin{cases} 2! \sum_{i_1=0}^{m} \dfrac{C_m^{i_1} (b_1 - a_1)^{i_1} (b_2 - a_2)^{m-i_1}}{(m - i_1 + 1)(m + 2)} & , \quad n = 2 \\[2ex] 3! \sum_{i_1=0}^{m} \sum_{\substack{i_2=0 \\ i_1 \neq i_2}}^{i_1} \dfrac{C_m^{i_1} C_{i_1}^{i_2} (b_1 - a_1)^{i_2} (b_2 - a_2)^{i_1 - i_2}}{(m - i_1 + 1)(m - i_2 + 2)(m + 3)} + \mathrm{M} & , \quad n = 3 \\ \vdots & \vdots \quad \vdots \\ \dfrac{n! \sum\limits_{i_1=0}^{m} \sum\limits_{\substack{i_2=0 \\ i_1 \neq i_2}}^{i_1} \sum\limits_{i_3=0}^{m-i_1} \sum\limits_{i_4=0}^{m-i_1-i_3} \sum\limits_{i_5=0}^{m-i_1-i_3-i_4} \cdots \sum\limits_{i_{n-2}=0}^{m-i_1-\sum_{k=3}^{n-3} i_k} \sum\limits_{i_{n-1}=0}^{m-i_1-\sum_{k=3}^{n-2} i_k} \mathrm{KL}}{\prod\limits_{j=1}^{n-3} (m - i_1 - \sum\limits_{s=3}^{n-j} i_s + j)(m - i_1 + n - 2)(m - i_2 + n - 1)(m + n)} + \mathrm{N} & , \quad n \geq 4 \end{cases}$$

$$= \alpha_m (a_1, ..., a_n; b_1, ..., b_n) \equiv \alpha_m \tag{5.12}$$

Burada;

$$\mathrm{K} = \left(\prod_{l=4}^{n-1} C^{i_l}_{m - i_1 - \sum_{t=3}^{l-1} i_t} \right) C_m^{i_1} C_{i_1}^{i_2} C_{m-i_1}^{i_3}$$

$$L=(b_1-a_1)^{i_2}(b_2-a_2)^{i_1-i_2}(b_n-a_n)^{m-i_1-\sum_{j=3}^{n-2}i_j}\prod_{k=3}^{n-1}(b_k-a_k)^{i_k}$$

$$M=3!\sum_{i_1=0}^{m}C_m^{i_1}\frac{(b_1-a_1)^{i_1}(b_3-a_3)^{m-i_1}}{(m-i_1+1)(m-i_1+2)(m+3)}$$

$$N=n!\sum_{i_1=0}^{m}\frac{(b_1-a_1)^{i_1}(b_n-a_n)^{m-i_1}}{(m+n)\prod_{k=1}^{n-1}(m-i_1+k)}$$

dır.

Sonuç 5.2.1.3.

$$X_1,X_2,\ldots,X_n;F(x)\in \mathrm{P}=\left\{F(x):F(x)=\frac{x-c}{d-c}\ ,\ c<x<d,\ F(x)=x\right\}$$

dağılımına sahip bir örneklem ve $a \prec b$ olmak üzere

$$n=3\ \ ,\ \ m=2\ \ ,\ \ \underline{a}=(\frac{1}{3},\frac{1}{3},\frac{1}{3})\ \ ,\ \ \underline{b}=(\frac{1}{2},\frac{1}{2},0)\qquad \text{için;}$$

$$P\left\{X_4,X_5\in\sum_{i=1}^{n}a_iX_i,\sum_{i=1}^{n}b_iX_i\right\}=3!\int_{x_1=0}^{1}\int_{x_2=0}^{x_1}\int_{x_3=0}^{x_2}\left(\frac{x_2+x_3}{2}-\frac{x_1+x_2+x_3}{3}\right)^2dx_3dx_2dx_1$$

$$=\frac{7}{360}$$

bulunur. Aynı olasılık için (5.12) eşitliği kullanılırsa;

$$P\left\{X_4,X_5\in\sum_{i=1}^{n}a_iX_i,\sum_{i=1}^{n}b_iX_i\right\}=3!\sum_{\substack{i_1=0\\ i_1\neq i_2}}^{2}\sum_{i_2=0}^{i_1}\frac{C_2^{i_1}C_{i_1}^{i_2}\left(\frac{1}{6}\right)^{i_2}\left(\frac{1}{6}\right)^{i_1-i_2}\left(-\frac{1}{3}\right)^{2-i_1}}{(3-i_1)(4-i_2)(5)}$$

$$+3!\sum_{i_1=0}^{2}C_2^{i_1}\frac{\left(\frac{1}{6}\right)^{i_1}\left(-\frac{1}{3}\right)^{2-i_1}}{(3-i_1)(4-i_2)(5)}$$

$$=\frac{7}{360}$$

elde edilir. Görüldüğü gibi her iki durumda da aynı sonuçlar çıkmıştır.

Bilindiği gibi büyük bir güven katsayısında ne kadar dar bir güven aralığı sağlanırsa tahmin o ölçüde güvenilir olur. Dolayısıyla α_1 güven katsayısının 1 değerine çok yakın seçilmesi durumunda $\left(\sum_{i=1}^{n}a_iX_{[i]},\sum_{i=1}^{n}b_iX_{[i]}\right)$ rasgele aralığının olabildiğince dar olması için $b_i\ ve\ a_i$ ' ler arasındaki fark minimum olmalıdır.

$\alpha_1 = 0,90$ seçilmesi durumunda $a_1 \geq \ldots \geq a_n$ ve $b_1 \geq \ldots \geq b_n$ olmak üzere (5.4) ' deki koşul altında;

Amaç fonksiyonu:

$$\min f = \sum_{i=1}^{n} |b_i - a_i|$$

ve kısıtları:

$$\begin{aligned}
&\sum_{i=1}^{k} a_i \leq \sum_{i=1}^{k} b_i \ , \quad k \leq n \\
&\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i \\
&\frac{1}{n+1} \sum_{i=1}^{n} (n-i+1)(b_i - a_i) = 0.90 \\
&a_i, b_i \in R \ , \quad i = 1,2,\ldots,n
\end{aligned} \tag{5.13}$$

şeklinde olan bir optimizasyon problemi karşımıza çıkar. Bu problem tümevarımsal yöntemler kullanılarak aşağıdaki şekilde çözülmüştür.

n=2 için model;

$$\begin{cases}
\min f = \sum_{i=1}^{2} |b_i - a_i| \\
b_1 \geq a_1 \\
b_1 + b_2 = a_1 + a_2 \\
\frac{1}{3} \sum_{i=1}^{2} (b_i - a_i)(3-i) = 0.90 \\
a_i, b_i \in R \ , \quad i = 1,2
\end{cases} \tag{5.14}$$

şeklindedir.

Çizelge 5.2.1.a. (5.14) modelinin çözümü;

$b_1 - a_1$	$b_2 - a_2$	$\min f$
2.7	-2.7	5.4

n=3 için model;

$$\begin{cases} \min\ f = \sum_{i=1}^{3} |b_i - a_i| \\ b_1 \geq a_1 \\ b_1 + b_2 \geq a_1 + a_2 \\ b_1 + b_2 + b_3 = a_1 + a_2 + a_3 \\ \frac{1}{4}\sum_{i=1}^{3}(b_i - a_i)(\overset{4}{\cancel{3}} - i) = 0.90 \\ a_i, b_i \in R \ , \ i = 1,2,3 \end{cases} \tag{5.15}$$

biçimindedir.

Çizelge 5.2.1.b. (5.15) modelinin çözümü;

$b_1 - a_1$	$b_2 - a_2$	$b_3 - a_3$	$\min f$
1.8	0	-1.8	3.6

$n = 4$ için model;

$$\begin{cases} \min\ f = \sum_{i=1}^{3} |b_i - a_i| \\ b_1 \geq a_1 \\ b_1 + b_2 \geq a_1 + a_2 \\ b_1 + b_2 + b_3 \geq a_1 + a_2 + a_3 \\ b_1 + b_2 + b_3 + b_4 = a_1 + a_2 + a_3 + a_4 \\ \frac{1}{5}\sum_{i=1}^{4}(b_i - a_i)(\underset{5}{\cancel{4}} - i) = 0.90 \\ a_i, b_i \in R \ , \ i = 1,2,3,4 \end{cases} \tag{5.16}$$

biçimindedir.

Çizelge 5.2.1.c. (5.16) modelinin çözümü;

$b_1 - a_1$	$b_2 - a_2$	$b_3 - a_3$	$b_4 - a_4$	$\min f$
1.5	0	0	-1.5	3

Çeşitli n değerleri için delphi programında bir simülasyon çalışması yapılmış ve sonuçlar çizelge 5.2.1.d. ' deki gibi elde edilmiştir. Program ise ek 1 ' de verilmiştir.

Çizelge 5.2.1.d. Çeşitli n değerleri için (5.13) modelinin çözümü

n	b_1-a_1	b_2-a_2	b_3-a_3	b_4-a_4	b_5-a_5	b_6-a_6	b_7-a_7	b_8-a_8	b_9-a_9	$b_{10}-a_{10}$	...	$b_{15}-a_{15}$	...	$b_{20}-a_{20}$	...	$b_{25}-a_{25}$	...	$b_{50}-a_{50}$	...	$b_{100}-a_{100}$	min f	
2	2.700	-2.700	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5.400	
3	1.800	0	-1.800	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	3.600	
4	1.500	0	0	-1.500	-	-	-	-	-	-	-	-	-	-	-	-	-			-	3.000	
5	1.350	0	0	0	-1.350	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2.700	
6	1.260	0	0	0	0	-1.260	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2.520	
7	1.200	0	0	0	0	0	-1.200	-	-	-	-	-	-	-	-	-	-	-	-	-	2.400	
8	1.157	0	0	0	0	0	0	-1.157	-	-	-	-	-	-	-	-	-	-	-	-	2.316	2,314
9	1.125	0	0	0	0	0	0	0	-1.125	-	-	-	-	-	-	-	-	-	-	-	2.250	
10	1.099	0	0	0	0	0	0	0	0	-1.099	-	-	-	-	-	-	-	-	-	-	2.216	2,198
15	1.029	0	0	0	0	0	0	0	0	0	0	-1.029	-	-	-	-	-	-	-	-	2.058	
20	0.995	0	0	0	0	0	0	0	0	0	0	0	0	-0.995	-	-	-	-	-	-	1.990	
25	0.975	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-.975	-	-	-	-	1,950	
50	0.937	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.937	-	-	1.874	
100	0.919	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.919	1.838	

5.3. İnvaryant Güven Aralıkların Seviyesinin Tahmini

$X_1, X_2, \ldots, X_n$ ve $Y_1, Y_2, \ldots, Y_m$ birbirinden bağımsız , aynı $F \in \mathrm{F}$ dağılım fonksiyonuna sahip iki örneklem ve $\left(f_1\left(X_1, X_2, \ldots X_n\right), f_2\left(X_1, X_2, \ldots X_n\right)\right)$ rasgele aralığı F sınıfı için

$$\alpha = P\{Y_1 \in (f_1(X_1, X_2, \ldots X_n), f_2(X_1, X_2, \ldots X_n))\} \tag{5.17}$$

seviyeli ana kitleyi kapsayan invaryant güven aralığı olsun. (5.17) eşitliği;

$$\begin{aligned}
\alpha &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [F(f_2(x_1, x_2, \ldots, x_n)) - F(f_1(x_1, x_2, \ldots, x_n))] dF(x_1) \ldots dF(x_n) \\
&= E[F(f_2(X_1, X_2, \ldots, X_n)) - F(f_1(X_1, X_2, \ldots, X_n))] \\
&= \int_{-\infty}^{\infty} x dP\{F(f_2(X_1, X_2, \ldots, X_n)) - F(f_1(X_1, X_2, \ldots, X_n)) \le x\}
\end{aligned}$$

şeklinde yazılabilir. Biliniyor. ki örneklem dağılım fonksiyonu teorik dağılım fonksiyonunun yansız, tutarlı tahmin edicisidir. F yerine $Y_1, Y_2, \ldots, Y_m$ örnekleminin empirik (örneklem) dağılım fonksiyonunu yazarsak α için yansız ve tutarlı bir tahmin edici bulmuş oluruz.

$$\begin{aligned}
\alpha^* &= \int_{-\infty}^{\infty} x dP\{F_m^*(f_2(X_1, X_2, \ldots, X_n)) - F_m^*(f_1(X_1, X_2, \ldots, X_n)) \le x\} \qquad (5.18) \\
&= \sum_{i=0}^{m} x_i P\{F_m^*(f_2(X_1, X_2, \ldots, X_n)) - F_m^*(f_1(X_1, X_2, \ldots, X_n)) = x_i\} \ , \quad x_0 = 0 \\
&= E[F_m^*(f_2(X_1, X_2, \ldots, X_n)) - F_m^*(f_1(X_1, X_2, \ldots, X_n))]
\end{aligned}$$

α^* tahmin edicisinin hesaplanması için öncelikle beklenen değer içerisindeki rasgele değişkenin dağılımının bulunması gerekir.

5.3.1. Tek yanlı invaryant güven aralıklarının seviyesinin tahmini

Daha önce yapılan araştırmalarda tüm sürekli dağılım fonksiyonları ailesi için tek yanlı invaryant güven aralıklarının seviyesi aşağıdaki şekilde tahmin edilmiştir.

Teorem 5.3.1.1. $f_1(X_1,X_2,...,X_n)=-\infty$ ve $f_2(X_1,X_2,...,X_n)=X_{(r)}$ olmak üzere $F_m^*(f_2(X_1,X_2,...,X_n))-F_m^*(f_1(X_1,X_2,...,X_n))$ rasgele değişkeninin dağılım fonksiyonu ;

$$P\{F_m^*(X_{(r)})\leq x\}=\begin{cases} P\{\phi\}=0 & , \quad x<0 \\ P\{X_{(r)}<Y_{(1)}\} & , \quad 0\leq x<\frac{1}{m} \\ \vdots & \quad\quad \vdots \\ P\{X_{(r)}<Y_{(k)}\} & , \quad \frac{k-1}{m}\leq x<\frac{k}{m} \\ \vdots & \quad\quad \vdots \\ P\{X_{(r)}\geq Y_{(m)}\} & , \quad x\geq 1 \end{cases}$$

olasılık fonksiyonu ise;

$$P\left\{F_m^*(X_{(r)})=\frac{k}{m}\right\}=P\{Y_{(k)}\leq X_{(r)}<Y_{(k+1)}\}$$

$$=\frac{\binom{m+n-k-r}{m-k}\binom{k+r-1}{k}}{\binom{m+n}{n}} \quad , \quad k=0,1,...,m \qquad (5.19)$$

şeklindedir.

Bu ise negatif hipergeometrik dağılımdır. Burada $Y_{(0)}=-\infty$ ve $Y_{(m+1)}=\infty$ anlamındadır. $F_m^*(X_{(r)})$ rasgele değişkeninin beklenen değer ve varyansı sırasıyla

$$E(F_m^*(X_{(r)}))=\frac{r}{n+r}$$

$$\mathrm{var}(F_m^*(X_{(r)}))=\frac{r(n-r+1)(m+n+1)}{m(n+1)^2(n+2)}$$

dır (Bairamov ve ark. 1999).

5.3.2. İki yanlı invaryant güven aralıklarının seviyesinin tahmini

$F_m^*(f_2(X_1,X_2,...,X_n))-F_m^*(f_1(X_1,X_2,...,X_n))$ rasgele değişkeninin dağılım fonksiyonu;

$$F_m^*(f_2(X_1,\dots,X_n)) - F_m^*(f_1(X_1,\dots,X_n))$$

$$= \begin{cases} 0 & , \; Y_{(i)} < f_1(X_1,\dots,X_n) < Y_{(i+1)} \,,\; Y_{(i)} < f_2(X_1,\dots,X_n) < Y_{(i+1)} \\ \dfrac{k}{m} & , \; Y_{(i)} \le f_1(X_1,\dots,X_n) < Y_{(i+1)} \,,\; Y_{(i+k)} < f_2(X_1,\dots,X_n) < Y_{(i+k+1)} \\ 1 & , \; f_1(X_1,\dots,X_n) < Y_{(1)} \;,\; f_2(X_1,\dots,X_n) > Y_{(m)} \end{cases} \tag{5.20}$$

olur. Burada $Y_{(m+1)} = \infty$, $Y_{(0)} = -\infty$, $k = 1,\dots,m$, $i = 0,\dots,m-k$ ' dır.

$$F_m^*(f_2(X_1,X_2,\dots,X_n)) - F_m^*(f_1(X_1,X_2,\dots,X_n)) \;;\; 0,\frac{1}{m},\frac{2}{m},\dots,\frac{m-1}{m},1$$

değerlerini

$$P\left\{F_m^*(f_2(X_1,X_2,\dots,X_n)) - F_m^*(f_1(X_1,X_2,\dots,X_n)) = \frac{k}{m}\right\}, \qquad k = 0,1,\dots,m$$

$$= \sum_{i=0}^{m-k} P\left\{Y_{(i)} \le f_1(X_1,\dots,X_n) < Y_{(i+1)} \,,\, Y_{(i+k)} < f_2(X_1,\dots,X_n) < Y_{(i+k+1)}\right\}$$

olasılıklarıyla alan bir rasgele değişkendir.

Teorem 5.3.2.1. $f_1(X_1,X_2,\dots,X_n) = X_{(r)}$, $f_2(X_1,X_2,\dots,X_n) = X_{(s)}$ (r<s) olmak üzere $F_m^*(X_{(s)}) - F_m^*(X_{(r)})$ rasgele değişkeninin olasılık fonksiyonu;

$$P\left\{F_m^*(X_{(s)}) - F_m^*(X_{(r)}) = \frac{k}{m}\right\}$$

$$= \sum_{i=0}^{m-k} P\left\{Y_{(i)} \le X_{(r)} < Y_{(i+1)} \,,\, Y_{(i+k)} < X_{(s)} < Y_{(i+k+1)}\right\}, \quad k = 0,\dots,m$$

$$f_{F_m^*(X_{(s)}) - F_m^*(X_{(r)})}(x)$$

$$= \begin{cases} \dfrac{n!(m+n-s)!}{(n-s)!(m+n)!} + \displaystyle\sum_{i=1}^{m-1} \dfrac{(r+i-1)!n!m!(m+n-s-i)!}{i!(r-1)!(n-s)!(n+m)!(m-i)!} + \dfrac{(m+r-1)!n!}{(r-1)!(mn)!} & , \quad x = 0 \\[2ex] \dfrac{mn!(m+n-s-1)!(s-r)}{(n-s)!(n+m)!} + \displaystyle\sum_{i=1}^{m-2} \dfrac{(s-r)(r+i-1)!n!m!(m+n-s-i-1)!}{i!(r-1)!(n-s)!(m-i-1)!(m+n)!} + \\ \dfrac{m(s-r)n!(m+r-2)!}{(r-1)!(m+n)!} & , \quad x = \dfrac{1}{m} \\[2ex] \dfrac{n!m!(m+n-s-k)!(s+k-r-1)!}{k!(s-r-1)!(n-s)!(m-k)!(m+n)!} + \displaystyle\sum_{i=1}^{m-k-1} \dfrac{rn!m!(m+n-s-i-k)!(k+s-r-1)!}{k!(s-r-1)!(n-s)!(m-k-i)!(m+n-i+1)!} + \\ \dfrac{n!m!(m+r-k-1)!(s+k-r-1)!}{k!(m-k)!(r-1)!(m+n)!} & , \quad x = \dfrac{k}{m} \\[2ex] \dfrac{n!(m+s-r-1)!}{(s-r-1)!(m+n)!} & , \quad x = 1 \end{cases} \tag{5.21}$$

şeklindedir (Kaya ve ark. 2002).

Örnek 5.3.2.2. X_1, X_2, X_3, X_4, F dağılımına sahip örneklem, Y_1, Y_2, Y_3 de aynı F dağılımına sahip X_1, X_2, X_3, X_4 örnekleminden bağımsız ve aynı dağılıma sahip başka bir örneklem olmak üzere $(X_{(r)}\ ,\ X_{(s)})$ rasgele aralığı $\alpha = P\{Y_1 \in (X_{(r)}, X_{(s)})\}$ seviyeli invaryant güven aralığı olsun. $r = 2$, $s = 3$ alınırsa;

$$\alpha = P\{Y_1 \in (X_{(2)}, X_{(3)})\} = \frac{3-2}{4+1} = \frac{1}{5}$$

olarak bulunur. Aynı olasılık (5.21) eşitliğindeki formül kullanılırsa

$$\begin{aligned}
\alpha^* &= E\left[F_m^*(X_{(3)}) - F_m^*(X_{(2)})\right] \\
&= 0\left(\frac{4}{35} + \frac{6}{35} + \frac{6}{35} + \frac{4}{35}\right) + \frac{1}{3}\left(\frac{3}{35} + \frac{4}{35} + \frac{3}{35}\right) + \frac{2}{3}\left(\frac{2}{35} + \frac{2}{35}\right) + \frac{1}{35} \\
&= \frac{1}{5} = \alpha
\end{aligned}$$

elde edilir.

Örnek 5.3.2.3. $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8$, F dağılımına sahip örneklem, Y_1, Y_2, Y_3, Y_4 F dağılımına sahip $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8$ örnekleminden bağımsız ve aynı dağılıma sahip başka bir örneklem olmak üzere $(X_{(2)}, X_{(4)})$ rasgele aralığı $\alpha = P\{Y_1 \in (X_{(2)}, X_{(4)})$ seviyeli invaryant güven aralığı olsun.

$$\alpha = P\{Y_1 \in (X_{(2)}, X_{(4)})\} = \frac{4-2}{8+1} = \frac{2}{9}$$

$$\begin{aligned}
\alpha^* &= E\left[F_m^*(X_{(4)}) - F_m^*(X_{(2)})\right] \\
&= 0\left(\frac{14}{99} + \frac{3}{11} + \frac{1}{99}\right) + \frac{1}{4}\left(\frac{14}{99} + \frac{2}{11} + \frac{8}{495}\right) + \frac{2}{4}\left(\frac{1}{11} + \frac{2}{33} + \frac{1}{55}\right) \\
&\qquad + \frac{3}{4}\left(\frac{4}{99} + \frac{8}{495}\right) + \frac{1}{99} \\
&= \frac{2}{9} = \alpha
\end{aligned}$$

olur.

Yukarıdaki örneklerden de görüldüğü üzere $F_m^*(X_{(s)}) - F_m^*(X_{(r)})$ istatistiği α için yansız bir tahmin edicidir.

6. SONUÇ VE ÖNERİLER

Bu tez çalışmasının birinci bölümünde yapılacak olan çalışmalarla ilgili bir giriş yapıldıktan sonra ikinci bölümünde olasılık uzayları, rasgele değişkenler ve dağılımları, üçüncü bölümde tez çalışması için gerekli olan sıra istatistikleri, dağılımdan bağımsız istatistikler, örneklem dağılım fonksiyonu, dördüncü bölümde ise dağılımdan bağımsız güven aralıkları ve tolerans aralıklar ile ilgili tanımlar ve temel bilgiler verilmiştir.

Tezin ana kısmını oluşturan beşinci bölümde ise araştırma sonucunda elde edilmiş orijinal sonuçlar yer almaktadır. Öncelikle parametrik aileler için invaryant güven aralıkları verilip ardından konum ve ölçek parametresi içeren dağılımlar ailesi için invaryant güven aralıkları oluşturulmuştur. Özel halde düzgün dağılım fonksiyonlarına sahip dağılımlar sınıfı

$$\mathrm{P} = \left\{ F(x) : F(x) = \frac{x-c}{d-c} \quad , \; c < x < d \right\}$$

için majorant vektörler yardımıyla α_1 seviyeli ana kitleyi kapsayan invaryant güven aralıkları oluşturulmuştur.

$X_1, X_2, \ldots, X_n \sim F \in \mathrm{P}$ dağılımına sahip bir örneklem, $a = (a_1, a_2, \ldots, a_n)$ vektörü $b = (b_1, b_2, \ldots, b_n)$ vektörü ile majorant olmak üzere Tanım 5.1.4 ve Tanım 5.1.5 ' den

$$f_1(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} a_i X_{[i]} \quad \text{ve} \quad f_2(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} b_i X_{[i]}$$

olarak alınmış ve bu örneklemden bağımsız yeni bir X_{n+1} rasgele değişkeninin $\left(\sum_{i=1}^{n} a_i X_{[i]}, \sum_{i=1}^{n} b_i X_{[i]} \right)$ aralığına düşmesi olasılığı;

$$P_\theta \left\{ X_{n+1} \in \left(\sum_{i=1}^{n} a_i X_{[i]}, \sum_{i=1}^{n} b_i X_{[i]} \right) \right\} = \frac{1}{n+1} \sum_{i=1}^{n} (n-i+1)(b_i - a_i)$$
$$= \alpha_1 (a_1, \ldots, a_n; b_1, \ldots, b_n) \equiv \alpha_1$$

şeklinde bulunmuştur. $X_1, X_2, \ldots, X_n$ örnekleminden bağımsız ve aynı dağılıma sahip yeni $X_{n+1}, X_{n+2}, \ldots, X_{n+m}$ örnekleminin bu aralığa düşmesi olasılığı ise;

$$P\left\{X_{n+1},X_{n+2}\ldots X_{n+m}\in\left(\sum_{i=1}^{n}a_iX_{[i]},\sum_{i=1}^{n}b_iX_{[i]}\right)\right\}$$

$$=\begin{cases} 2!\sum_{i_1=0}^{m}\dfrac{C_m^{i_1}(b_1-a_1)^{i_1}(b_2-a_2)^{m-i_1}}{(m-i_1+1)(m+2)} & , \; n=2 \\ 3!\sum_{i_1=0}^{m}\sum_{\substack{i_2=0\\ i_1\neq i_2}}^{i_1}\dfrac{C_m^{i_1}C_{i_1}^{i_2}(b_1-a_1)^{i_2}(b_2-a_2)^{i_1-i_2}}{(m-i_1+1)(m-i_2+2)(m+3)}+\mathrm{M} & , \; n=3 \\ \vdots & \vdots \quad \vdots \\ \dfrac{n!\sum_{i_1=0}^{m}\sum_{\substack{i_2=0\\ i_1\neq i_2}}^{i_1}\sum_{i_3=0}^{m-i_1}\sum_{i_4=0}^{m-i_1-i_3}\sum_{i_5=0}^{m-i_1-i_3-i_4}\cdots\sum_{i_{n-2}=0}^{m-i_1-\sum_{k=3}^{n-3}i_k}\sum_{i_{n-1}=0}^{m-i_1-\sum_{k=3}^{n-2}i_k}\mathrm{KL}}{\prod_{j=1}^{n-3}(m-i_1-\sum_{s=3}^{n-j}i_s+j)(m-i_1+n-2)(m-i_2+n-1)(m+n)}+\mathrm{N} & , \; n\geq 4 \end{cases}$$

$$=\alpha_m(a_1,\ldots,a_n;b_1,\ldots,b_n)\equiv\alpha_m$$

Burada;

$$\mathrm{K}=\left(\prod_{l=4}^{n-1}C^{i_l}_{m-i_1-\sum_{t=3}^{l-1}i_t}\right)C_m^{i_1}C_{i_1}^{i_2}C_{m-i_1}^{i_3}$$

$$\mathrm{L}=(b_1-a_1)^{i_2}(b_2-a_2)^{i_1-i_2}(b_n-a_n)^{m-i_1-\sum_{j=3}^{n-2}i_j}\prod_{k=3}^{n-1}(b_k-a_k)^{i_k}$$

$$\mathrm{M}=3!\sum_{i_1=0}^{m}C_m^{i_1}\frac{(b_1-a_1)^{i_1}(b_3-a_3)^{m-i_1}}{(m-i_1+1)(m-i_1+2)(m+3)}$$

$$\mathrm{N}=n!\sum_{i_1=0}^{m}\frac{(b_1-a_1)^{i_1}(b_n-a_n)^{m-i_1}}{(m+n)\prod_{k=1}^{n-1}(m-i_1+k)}$$

olarak bulunmuştur. Bu olasılıklar düzgün dağılımın parametrelerinden bağımsız olduğundan $\left(\sum_{i=1}^{n}a_iX_{[i]}\;,\;\sum_{i=1}^{n}b_iX_{[i]}\right)$ rasgele aralığının düzgün dağılım fonksiyonları ailesi için ana kitleyi kapsayan invaryant güven aralık olduğu söylenir.

Bu güven aralığının olabildiğince dar olması b_i ve a_i ' ler arasındaki farkın en küçük olması ile sağlanır. α_1=0.90 seçilerek (5.11), Tanım 5.1.4 ve Tanım 5.1.5 ' deki koşullar altında $a_1\geq\ldots\geq a_n$ ve $b_1\geq\ldots\geq b_n$ olmak üzere;

$$\begin{cases} \min f = \sum_{i=1}^{n} |b_i - a_i| \\ \sum_{i=1}^{k} a_i \le \sum_{i=1}^{k} b_i \ , \quad k \le n \\ \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i \\ \frac{1}{n+1} \sum_{i=1}^{n} (n-i+1)(b_i - a_i) = 0.90 \\ a_i, b_i \in R \ , \quad i = 1,2,\ldots,n \end{cases}$$

şeklinde bir optimizasyon modeli ile karşılaşılmış ve bu modelin çözümü delphi programı ile bulunmuş (ek 1) ve çizelge 5.2.1.d. ' deki tablo elde edilmiştir.

Son olarak tüm sürekli dağılım fonksiyonları ailesi için örneklem (empirik) dağılım fonksiyonu yardımıyla iki yanlı invaryant güven aralıklarının seviyesi için yansız bir tahmin edici bulunmuş ve doğruluğu örneklerle gösterilmiştir.

Bu tez çalışmasında majorant vektörler kesin olarak bulunamamıştır. Majorant vektörler arasındaki ilişki bilgisayar programı yardımıyla tespit edilmiştir. Majorant vektörler kesin olarak bulunduğu takdirde gücü yeterince yüksek hipotez testleri yapılabilecektir.

KAYNAKLAR

Bairamov, I. G. (1992). "Statistical criteria based on training samples in the problem of classification of observations". Teo. Prob. And Math. Statist., No: 46, 13-17

Bairamov, I. G., Petunin, Yu. I. (1992). "Statistical criteria based on training samples." Cybernetics,vol. 27, No: 3, 408-413.

Bairamov, I. G., Petunin Yu. I. (1990). "Structure of invariant confidence intervals containing the main distributed mass". Theor. Probab. Appl., vol. 35, No:1, 15-26

Bairamov, I. G., Gebizlioğlu, Ö. L., Kaya, M. F. (1999). "Distributional properties of statistics based on invariant confidence intervals". Journal of The Turkish Statistical Association, vol. 2, No:1, 15-33

Borovkov, A. A. (1984). Mathematical Statistics. Nauka, Moskow.

Cramer, H. (1946). Mathematical Methods of Statistics. Princeton University Press.

Cerit, C. (1997). Matemetik İstatistik, İstanbul.

David H. A. (1970). Order Statistics. John Wiley & Sons, Inc., New York.

Gaeusler, P., Stute, W. (1987). Seminar on Empirical Processes. Birkhauser Verlag., Boston, Basel.

Guenter, W.C. (1975). The İnverse Hipergeometric a Useful Model. Statist. Neerl. No: 29, 124-144

Kaya, M.F. (1998). Doktora Tezi, A.Ü.F.F.

Kaya, M.F., Saraçoğlu, B., Kuş, C., Pekgör, A. (2002). "Sürekli dağılımlar ailesi için ana kitleyi kapsayan iki yanlı invaryant güven aralıkların seviyesinin örneklem dağılım fonksiyonu yardımıyla tahmin edilmesi". H.Ü. İstatistik Günleri Sempozyumu

Kendal, M., Stuart, A. (1973). The Advanced Theory of Statistics. Vol. 2, Charles Grifin, London. [12.6]

Marshal, A.W., Olkin, I. (1979). Inequalities. Theory of Majorizations and It's Applications. Academic Press.

Öztürk,F. (1993). Matematiksel İstatistik, A.Ü.F.F.

Randles, R. H., Wolfe, D. A. (1979). Introduction to The Theory of Nonparametric Statistics. John Wiley & Sons, Inc., New York.

Rohatgi V. K. (1976). An Introduction to Probability Theory and Mathematical Statistics. John Wiley & Sons, Inc., New York.

Robbins, H. (1944). "On Distribution Free Tolerans Limits in Random Sampling." Ann. Math. Stat., vol.15, 70-74

Shahbazov, A. (1973). İhtimal Nazariyesi ve Riyazi Statistica. Maarif, Bakü.

Shewart, W.A. (1931). Economic Control of Quality of Manufactured Product. D. Van Nostrand, Princeton, N. J.

Wilks, S. S. (1942). "Statistical Prediction with Special Reference to The Problem of Tolerance Limits". Ann. Math. Statist., vol. 13, 400-409

Wischanski, D. F., Petunin, Yu. I. (1979). "Basis of The 3σ Rule for Unimodel Distributions." Theory Probab. Math. Statist.". No: 21, 23-35

EKLER

EK 1

Çizelge 5.2.1.d. için delphi programı

Button1Click:

```
procedure TForm1.Button1Click(Sender: TObject);
var i,j,k,n:integer;
top,a,b,alt,ust,a1,a2,a3,a4,b1,b2,b3,b4:double;
begin
n:= strtoint(edit5.text);
alt:=strtofloat(edit2.text);
ust:=strtofloat(edit3.text);
i:=0;
b:=ust;
a:=alt;
while (a<b ) do
begin
 i:=i+1;
 a:=a+0.001;
 ust:=a;
 a1:=alt;
 b1:=ust;
 b3:=0.9*(n+1);
 a3:=(n-1)*(b1-a1);
 b4:=-(b1-a1);
 a4:=b3-a3;
 top:=abs(b1-a1)+abs(b3-a3)+abs(b4-a4);
 gr.Cells[1,i]:=floattostr(b1-a1);
 gr.Cells[2,i]:=floattostr(top);
end;
end;
```

Button2Click:

```
procedure TForm1.Button2Click(Sender: TObject);
var i,j,k,n:integer;
h,top,a,b,alt,ust,a1,a2,a3,a4,b1,b2,b3,b4:double;
begin
n:= strtoint(edit5.text);
ust:=strtofloat(edit4.text);
h:=strtofloat(edit6.text);
i:=0;
b:=0;
a:=ust;
alt:=0;
while (a>b ) do
begin
 i:=i+1;
 a:=a-h;
 ust:=a;
 a1:=alt;
 b1:=ust;
 b3:=0.9*(n+1);
 a3:=(n-1)*(b1-a1);
 b4:=-(b1-a1);
 a4:=b3-a3;
 top:=abs(b1-a1)+abs(b3-a3)+abs(b4-a4);
 gr.Cells[1,i]:=floattostr(b1-a1);
 gr.Cells[2,i]:=floattostr(top);
end;
end;
```