T.C.
BİTLİS EREN ÜNİVERSİTESİ ve FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİK ANABİLİM DALI
YÜKSEK LİSANS TEZİ

PARETO DAĞILIMINDA ÖRNEKLEM SEÇİMİNİN TAHMİN EDİCİYE ETKİSİ

Seval ŞAHİN

KASIM 2017

İSTATİSTİK ANABİLİM DALI
YÜKSEK LİSANS TEZİ

PARETO DAĞILIMINDA ÖRNEKLEM SEÇİMİNİN TAHMİN EDİCİYE ETKİSİ

Hazırlayan
Seval ŞAHİN

Danışman
Yrd. Doç. Dr. Fahrettin ÖZBEY

Jüri Üyeleri
Prof. Dr. Sinan ÇALIK
Yrd. Doç. Dr. Fahrettin ÖZBEY
Yrd. Doç. Dr. Ayşe METİN KARAKAŞ

KASIM 2017

Seval ŞAHİN tarafından hazırlanan "Pareto Dağılımında Örneklem Seçiminin Tahmin Ediciye Etkisi" adlı tez çalışması 24/11/2017 tarihinde yapılan sınavla aşağıdaki jüri tarafından oybirliği/~~oyçokluğu~~ ile Bitlis Eren Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri **İmza**

Prof. Dr. Sinan ÇALIK
(Başkan)
Yrd. Doç. Dr. Fahrettin ÖZBEY
(Danışman)
Yrd. Doç. Dr. Ayşe METİN KARAKAŞ
(Üye)

Bu tezin kabulü, Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 07/12/2017 gün ve 48/03 Sayılı kararı ile onaylanmıştır.

Doç. Dr. Koray KÖKSAL
Enstitü Müdürü

ÖZET

PARETO DAĞILIMINDA ÖRNEKLEM SEÇİMİNİN TAHMİN EDİCİYE ETKİSİ

Seval ŞAHİN

Yüksek Lisans Tezi

Bitlis Eren Üniversitesi Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı
Danışman: Yrd. Doç. Dr. Fahrettin ÖZBEY
Kasım 2017, 52 sayfa

Bu çalışmada, ilk olarak belli bir dağılımdan örneklem üretme yöntemleri verildi. Daha sonra dağılımdan örneklem üretmek için yeni yöntemler araştırıldı. Sonuç olarak; eski ve yeni yöntemler kullanılarak Pareto dağılımından örneklemler üretildi. Bu örneklemler kullanılarak, maksimum olabilirlik yöntemi ile parametre tahminleri yapıldı. Bu tahmini parametreler örneklem oluşturmak için kullanılan parametrelerle karşılaştırıldı ve yeni yöntem ile oluşturulan örneklemlerden daha iyi sonuçlar elde edildi.

Anahtar Kelimeler: Pareto dağılımı, Tahmin Edici, Örneklem

ABSTRACT

THE EFFECT OF CHOOSING THE SAMPLE ON THE ESTIMATOR IN PARETO DISTRIBUTION

Seval ŞAHİN

Master Thesis

Bitlis Eren University Graduate School of Natural and Applied Sciences
Department of Statistic
Supervisor: Asst. Prof. Dr. Fahrettin ÖZBEY
November 2017, 52 pages

In this study, firstly, methods of producing samples from a certain distribution were given. After then, new methods were investigated for generate scattered samples. As a result; samples were produced from the Pareto distribution using old and new methods. The parameter estimates were made with the maximum likelihood method using these samples. These estimated parameters were compared with the parameters used to create the sample and better results were obtained in the samples created with the new method.

Keywords: Pareto distribution, Estimator, Sample

TEŞEKKÜR

Tez çalışmamın hazırlanmasında ve düzenlenmesinde yardımlarını esirgemeyen ve çalışmalarım boyunca değerli fikirlerinden ve tecrübelerinden yararlandığım ve bana göstermiş olduğu yoğun ilgi, anlayış ve iyi niyeti için öncelikle değerli danışmanım Sayın Yrd. Doç. Dr. Fahrettin ÖZBEY' e sonsuz saygı ve teşekkürlerimi sunuyorum.

Çalışmamın her safhasında destek ve bilgilerini benimle paylaşan canım ablam Yrd. Doç. Dr. Sultan ŞAHİN BAL'a, ömrüm boyunca en iyi seviyeye gelmem için maddi ve manevi desteğini benden hiçbir zaman esirgemeyen canım aileme, hayatımda manevi destekleriyle her zaman yanımda olan değerli arkadaşım Ayşe KILIÇ başta olmak üzere tüm arkadaşlarıma teşekkür ederim.

Seval ŞAHİN

İÇİNDEKİLER DİZİNİ

Sayfa

ŞEKİLLER DİZİNİ

ÇİZELGELER DİZİNİ

ÇİZELGE Sayfa

1. GİRİŞ

Örneklem oluşturma yöntemleri incelendiğinde temelde iki farklı türden örneklem oluşturula bilinir. Bunların ilki bir anakütleden çeşitli yöntemler kullanılarak elde edilen örneklemlerdir. Diğeri ise bir dağılımın olasılık (yoğunluk) fonksiyonu veya dağılım fonksiyonu için parametre atamaları yapılarak üretilen örneklemlerdir. Literatürde her iki örneklem seçimi için çeşitli yöntemler geliştirilmiştir.

Tahmin teorisinde yapılan çalışmalar incelendiğinde; genellikle ya yeni tahmin metotları geliştirilir ya da yeni dağılımlar oluşturularak var olan tahmin yöntemleri ile parametre tahminleri yapılmaktadır. Her iki tür çalışmada da ortak olan prensip bir dağılım ele alınır ve bu dağılımdan yararlanarak örneklemler oluşturulur.

Bu çalışmada ilk olarak belli bir dağılımdan örneklem üretmek için var olan yöntemler incelendi ve yeni yöntemler araştırılacaktır. Daha sonra eski ve yeni yöntemler Pareto dağılımında uygulanarak örneklemler üretilecektir. Üretilen bu örneklemler kullanarak maksimum olabilirlik yöntemi yardımıyla parametre tahminleri yapılacaktır. Bu tahmini parametreler örneklem oluşturmak için kullanılan parametrelerle karşılaştırılacaktır.

1.1. Temel Tanımlar

Tanım 1.1.1 İstatistiksel araştırmalarda incelemeye amaçladığımız birimlerin oluşturduğu topluluğa anakütle denir. Birim ise anakütleyi oluşturan en küçük parçadır. Anakütle ile ilgili bilgi toplamak istendiğinde tüm birimlerin teker teker incelenmesi gerekmektedir. Bu işleme tam sayım adı verilir.

Anakütle birim sayısı fazla ise, tüm birimlerin incelenmesi fazla zaman alabilir ve masraflı olabilir. Araştırmalar gerekli bir fayda elde etmek için yapılırlar ve yapılan her araştırma belirli bir sürede bitirilebiliyorsa ve araştırmada yapılan masraf faydayı aşıyorsa araştırma yapmanın bir anlamı kalmayacaktır. Bu nedenle örnekleme teorisi geliştirilmiştir.

Bir anakütleden, anakütle birim sayısından daha az sayıda birim seçilerek, bu birimler yardımı ile anakütle parametrelerinin tahmin edilmesi işlemlerine örnekleme denir. Parametre ise anakütleleri birbirinden ayırmaya yarayan ortalama, varyans, oran gibi ölçülerin genel ifadesidir. Anakütleden seçilen az sayıda birimin oluşturduğu topluluk ise örnek olarak adlandırılır [1].

Tanım 1.1.2 Gözlemden gözleme değişik değerler alabilen objelere, özelliklere ya da durumlara değişken denir [2].

Tanım 1.1.3 Bir kitlenin belli bir özelliğini incelemek üzere, kitleden belirli bir kurallara göre seçilen birimler topluluğuna örneklem denir. Örneklem kitle birimlerinin gözlenen bir alt kümesidir. Bir örneklemin tanımlayıcı sayısal ölçüsüne örneklem istatistiği denir [3].

Tanım.1.1.4 Varyansın incelenmesinde S^2 örneklem varyansı $(X_i - \bar{X})^2$ sapmaların ortalamasıdır. Varyansa bu nedenle ortalamaya göre ikinci moment denir ve m_2 ile de gösterilir. Veri kümesi için karşılık gelen çarpıklık ölçütü $(X_i - \bar{X})^3$ kübik sapmaların ortalamasının alınmasıyla bulunur. m_3 ile gösterilen bu ölçüye çarpıklık ölçütü denir. Yani $\bar{X}$ gözlem değerlerinin ortalaması olmak üzere

$$m_3 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^3 \tag{1.1}$$

$$m_3' = \frac{(m_3)}{S^3} \tag{1.2}$$

dir.

$X_1, X_2, \dots, X_n$ gözlem değerleri verilsin. Bu değerlerin aritmetik ortalaması $\bar{X}$ ve standart sapması S olmak üzere; X_i'lere gelen $y_i = \frac{(X_i - \bar{X})}{S}$, i= $1,2, \dots \dots n$ gözlem değerlerine standartlaştırılmış gözlem değerleri denir.

Standartlaştırılmış çarpıklık ölçütü olarak tanımlanır. $m_3^{'} > 0$ ise sağa çarpık (pozitif çarpıklık) veri kümesi, $m_3^{'} < 0$ ise sola çarpık (negatif çarpıklık) veri kümesi vardır. $m_3^{'} = 0$ ise veri değerleri simetrik dağılım gösterir [3].

Tanım 1.1.5

$$m_4 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^4 \tag{1.3}$$

olmak üzere

$$m_4^{'} = \frac{(m_4)}{S^4} \tag{1.4}$$

eşitliği standartlaştırılmış sivrilik ölçütüdür. Normal dağılım için $m_4^{'} = 3$ olduğundan sivrilik $K = m_3^{'} - 3$ biçiminde de tanımlanır. $K < 0$ için merkeze yakın yerde eğri normal dağılım eğrisine göre fazla düzdür.(basıktır.) Normal dağılım için $K = 0$ ve $K > 0$ için merkeze yakın yerde eğri normal dağılım eğrisinden daha dar ve yüksektir [3].

Tanım 1.1.6 Seri değerlerinin aritmetik ortalamadan farklarının kareli ortalamasına standart sapma denilir. Standart sapma örneklemede S, anakütlede de σ ile gösterilir. Standart sapmanın karesine varyans adı verilir. Varyans örneklemede S^2, anakütlede σ^2 ile ifade edilir [1].

Örneklem için;

$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}} \tag{1.5}$$

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{n-1} \tag{1.6}$$

Anakütle için;

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}} \tag{1.7}$$

$$\sigma^2 = \frac{\sum(X_i-\mu)^2}{N} \tag{1.8}$$

Frekans serilerinde, örneklem için;

$$s = \sqrt{\frac{\sum(X_i-\bar{X})^2 f_i}{n-1}} \tag{1.9}$$

$$s^2 = \frac{\sum[(X_i-\bar{X})^2 f_i]}{n-1} \tag{1.10}$$

Anakütle için;

$$\sigma = \sqrt{\frac{\sum(X_i-\mu)^2 f_i}{N}} \tag{1.11}$$

$$\sigma^2 = \frac{\sum(X_i-\mu)^2 f_i}{N} \tag{1.12}$$

formülleri kullanılır.

Gruplandırılmış serilerde, örneklem için;

$$s = \sqrt{\frac{\sum(m_i-\bar{X})^2 f_i}{n-1}} \tag{1.13}$$

$$s^2 = \frac{\sum[(m_i-\bar{X})^2 f_i]}{n-1} \tag{1.14}$$

Anakütle için;

$$\sigma = \sqrt{\frac{\sum(m_i-\mu)^2 f_i}{N}} \tag{1.15}$$

$$\sigma^2 = \frac{\sum(m_i-\mu)^2 f_i}{N} \tag{1.16}$$

formülleri kullanılır.

Tanım 1.1.7 $g(x) = x^k$ fonksiyonunun beklenen değerine X rasgele değişkeninin 0 'a göre k −nıncı mertebeden momenti denir. Yani

$$m_k = E(x^k) \tag{1.17}$$

dır.

X , x_i değerlerini p_i olasılıklarıyla alan kesikli rasgele değişken ise k −nıncı mertebeden moment aşağıdaki gibidir.

$$m_k = E(x^k) = \sum_i^k x_i{}^k p_i \tag{1.18}$$

$X, f(x)$ olasılık yoğunluk fonksiyonuna sahip sürekli rasgele değişken ise k −nıncı mertebeden moment aşağıdaki gibidir.

$$m_k = E(x^k) = \int_{-\infty}^{+\infty} x^k f(x)dx \tag{1.19}$$

Kolayca görülür ki;

$$E[(ax)^k] = a^k E(x^k) \tag{1.20}$$

dir. Ayrıca c keyfi bir sabit olmak üzere $E[(x-c)^k]$ şeklindeki ifadeleri de hesaplayabiliriz [3].

Tanım 1.1.8 $E[(x-c)^k]$ beklenen değerine c noktasına göre k −nıncı mertebeden moment denir.

$c - m_1 = E(x)$ alınırsa $E[x - E(x)]^k$ değerine beklenen değere (ortalamaya) göre k −nıncı mertebeden moment denir [3].

Tanım 1.1.9 Beklenen değere göre momentlere merkezi momentler denir ve $\mu_k = E[x - E(x)]^k$ ile gösterilir. $c = 0$ noktasına göre momentlere adi momentler denir. İlk üç mertebeden momentler için;

$$\mu_1 = E[x - E(x)] = E(x - m_1) = E(x) - m_1 = m_1 - m_1 = 0$$

$$\mu_2 = E[x - E(x)]^2 = E[(x - m_1)^2]$$

$$= E(x^2) - 2m_1E(x) + m_1{}^2 = m_2 - m_1{}^2$$

$$\mu_3 = E[x - E(x)]^3 = E[(x - m_1)^3]$$

$$= E(x^3) - 3m_1E(x^2) + 3m_1{}^2E(x) - m_1{}^3$$

$$= m_3 - 3m_1m_2 + 3m_1{}^2m_1 - m_1{}^3$$

$$= m_3 - 3m_1m_2 + 2m_1{}^3 \tag{1.21}$$

yazılabilir [3].

Tanım 1.1.10 İkinci mertebeden merkezi momente x rastgele değişkeninin varyansı denir. Varyans genellikle $Var(x)$, ya da σ^2 x ile gösterilir ve

$$Var(x) = \sigma^2\ x = m_2 - m_1{}^2 \tag{1.22}$$

ifadesi kullanılarak hesaplanır. σ^2 parametresi, rasgele değişkenin beklenen değeri etrafında bir dağılım ölçüsüdür [3].

Tanım 1.1.11 $E(x) = 0$ ve $Var(x) = 1$ ortalama ve varyansına sahip x rastgele değişkenine standartlaştırılmış rastgele değişken denir.
x, beklenen değeri m_1 ve varyansı σ^2 olan bir rasgele değişkense,

$$\mathrm{y} = \frac{\mathrm{x} - m_1}{\sigma} \tag{1.23}$$

Rasgele değişkeni standartlaştırılmış bir rasgele değişkendir. Yani

$$E(y) = 0 \tag{1.24}$$

$$Var(y) = 1 \tag{1.25}$$

şeklindedir [3].

Tanım 1.1.12 Standart sapmanın beklenen değere orana varyasyon(değişim) kat sayısı (D.K) denir. O halde

$$D.K = \frac{\sigma}{m_1} \text{ ya da } D.K = \frac{\sigma}{m_2}.100 \tag{1.26}$$

dır. Beklenen değer 1 e eşit olduğunda varyasyon katsayısı standart sapmaya eşit olur [3].

Tanım 1.1.13 $E[|x - E(x)|]^k$ İfadesi x rasgele ifadesinin $k -$ ıncı mertebeden mutlak merkezli momenti denir [3].

Tanım 1.1.14 x bir rasgele değişken olsun. $h > 0$ olmak üzere $|t| < h$ aralığında her değeri alan t için e^{tx} beklenen değeri x ‘in moment çıkaran fonksiyonu olarak tanımlanır.

a) x , $f(x)$ olasılık fonksiyonuna sahip kesikli bir rasgele değişken olsun. x in moment çıkaran fonksiyonu

$$m_x(t) = E(e^{tx}) = \sum_x e^{tx} f(x) \tag{1.27}$$

b) x , $f(x)$ olasılık fonksiyonuna sahip sürekli bir rasgele değişken olsun, x in moment çıkaran fonksiyonu

$$m_x(t) = E(e^{tx}) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx \tag{1.28}$$

e^{tx} fonksiyonuna MacLaurin serisine açtıktan sonra kesikli hal için moment çıkaran fonksiyonunu yeniden yazarsak

$$m_x(t) = \sum_x \left[1 + tx + \frac{t^2x^2}{2!} + \cdots + \frac{t^rx^r}{r!} + \cdots\right] f(x)$$

$$= \sum_x f(x) + \sum_x xf(x) + \frac{t^2}{2!}\sum_x x^2 f(x) + \cdots + \frac{t^r}{r!}\sum_x x^r f(x) + \cdots$$

$$= 1 + m_1 + m_2\frac{t^2}{2!} + \cdots + m_r\frac{t^r}{r!} + \cdots \tag{1.29}$$

elde edilir. Görüldüğü gibi x'in moment çıkaran fonksiyonu MacLaurin serisine açılımında $\frac{t^r}{r!}$ in katsayısı sıfıra göre $r - inci$ momenttir [3].

Tanım 1.1.15 Aynı özelliğe sahip birimler topluluğu olarak tanımlanan kitleden, belirlenen bir örnekleme yöntemiyle aynı genişlikte seçilebilecek tüm örneklemlerin oluşturduğu uzaya örnekleme uzayı adı verilir [4].

Tanım 1.1.16 Örneklemin çekim sürecinden sonra sıra tahmin ediciye gelir. Sonlu kitleden çekilen örneklemden yararlanarak kitlenin özelliklerini tahmin etmek amacıyla tanımlanan matematiksel eşitliğe tahmin edici adı verilir [4].

Tanım 1.1.17 Örnekleme uzayının her bir noktasına gerçek sayı bağlayan fonksiyona tesadüfi değişkeni denir [4].

Tanım 1.1.18 Bir tahmin edicinin örneklem uzayında aldığı değerlerin göstermiş olduğu dağılıma örneklem dağılım adı verilir [4].

Tanım 1.1.19 $A \subset \Re$ olsun. $f(x) = \|x\|$ şeklinde tanımlanan $f : A \to \Re$ fonksiyonuna tam kısım fonksiyonu adı verilir. Burada $\|x\|, x$ sayısından büyük olmayan tam sayıların en büyüğünü göstermektedir. $\|x\|$, $n \le x$ eşitsizliğini gerçekleyen n tamsayılarının en büyüğünü gösterdiğinden, p bir tamsayı olmak üzere, $p \le x < p+1$ eşitsizliğini sağlayan x reel sayıları için $\|x\| = p$ dir [6].

1.2. Pareto Dağılımı

Üstel ve gamma dağılımının bir karışımı olan Pareto dağılımı ilk olarak İsviçreli ekonomist Vilfredo Pareto tarafından ifade edilmiştir. Vilfredo İsviçre'nin gelir dağılımını modellemek için Pareto dağılımını kullandı. O zamandan beri, Pareto dağılımı ağır kuyruklu dağılım modellemesinde yaygın olarak kullanılmıştır. Benzer yaklaşımla nüfus dağılımını, deprem büyüklüklerini, orman yangını alanlarını, petrol ve doğal gaz alanlarını modellemek içinde Pareto dağılımı kullanmaktadır. Pareto dağılımının birçok uygulamasına ekonomi, biyoloji ve fizik alanlarında rastlamak mümkündür.

Pareto dağılımının olasılık yoğunluk fonksiyonu, sıfıra göre momenti, ortalamaya göre momenti, bu momentlerden faydalanarak bu dağılımların ortalaması, varyansı, çarpıklık ve basıklık değerleri verilecektir.

Pareto dağılımının olasılık yoğunluk fonksiyonu ve dağılım fonksiyonu sırasıyla aşağıdaki gibidir.

$$F(x) = 1 - \left(\frac{x}{\lambda}\right)^{-v} \tag{1.30}$$

$$f(x) = v\lambda^{v} x^{-v-1} \quad , x \geq \lambda, v > 0 \tag{1.31}$$

Pareto dağılımının sıfıra göre $r - inci$ momenti;

$$E(x^r) = \frac{v\lambda^r}{(v - r)} \tag{1.32}$$

şeklindedir.

Bu son eşitlikte $r = 1$ yazılırsa Pareto dağılımının ortalaması;

$$\mu = \frac{v\lambda}{v - 1} \tag{1.33}$$

şeklinde elde edilir.

Pareto dağılımının ortalamaya göre $r - inci$ momenti;

$$E(X - \mu)^r = \sum_{i=0}^{r} \binom{r}{i} (-1)^{r-i} \left(\frac{v\lambda}{v - 1}\right)^{r-i} \frac{v\lambda^i}{(v - i)} \tag{1.34}$$

şeklindedir. Bu son eşitlikte $r = 2$ yazılırsa Pareto dağılımının varyansı;

$$Var(X) = \frac{v\lambda^2}{(v-1)^2(v-2)}, \quad v > 2 \tag{1.35}$$

şeklinde elde edilir. (1.34)'de 3.moment hesaplanarak (1.2)'de yerine yazılırsa Pareto dağılımının çarpıklık değeri;

$$Ç = 2\frac{v+1}{v-3}\sqrt{\frac{v-2}{v}}, \quad v > 3 \tag{1.36}$$

şeklinde ifade edilir. (1.34)'de 4.moment hesaplanarak (1.4)'de yerine yazılırsa Pareto dağılımının basıklık değeri;

$$B = \frac{3(v-2)(3v^3+v+2)}{v(v-3)(v-4)} \quad ,v > 4 \tag{1.37}$$

şeklinde ifade edilir [5].

1.3. Maksimum Olabilirlik Yöntemi

Bu yönteme maksimum benzerlik yöntemi adı da verilmektedir. Elimizde farklı anakütleler ile tesadüfi olarak alınmış bir örnek olduğu düşünelim. Bu örneğin bu anakütlelerin herbirinden alınma olasılığı muhtemelen farklı ve bazı anakütlelerden alınma olasılığı diğerlerine göre daha yüksek olacaktır. Bu örnek bu anakütlelerden birinden alınmışsa, diğer anakütlelerden çok seçilmiş olduğu anakütleden alınmış olabilir. Örneğin ortalaması 75 olan bir örnek olduğu düşünülürse, bu örnek muhtemelen ortalaması 75 olan bir anakütleden alınmıştır. Diğer bir ifade ile bu örnek ortalaması 70 veya 80 olan anakütlelerden çok, ortalaması 75 olan bir anakütleden alınmış olabilir. Bu yaklaşım maksimum olabilirlik yönteminin temelini oluşturmaktadır.

Konuyu açıklayabilmek için bir örnek verilirse: İçinde siyah ve beyaz olmak üzere 5'er top bulunan torbalar olduğunu varsayılırsa, birinci torbada siyah top olmasın, ikinci torbada 1 siyah, üçüncü torbada 2 siyah, dördüncü torbada 3, beşinci torbada 4 siyah, altıncı torbada 5 siyah top olsun. Bu durumda toplam 6 torba olacaktır. Torbalardaki siyah top sayısını (i) ile ifade edilir $(i = 0,1,2,3,4,5)$. Torbalardaki siyah top oranı,

$$p_0 = \frac{0}{5} \quad p_1 = \frac{1}{5} \quad p_2 = \frac{2}{5} \cdots p_5 = \frac{5}{5} \tag{1.38}$$

yani,

$$p_i = \frac{i}{5} \tag{1.39}$$

olacaktır. Bu olayda 6 farklı p oranı olan 6 farklı durum sözkonusudur. Bunları Ω ile,

$$\Omega_0, \Omega_1, \ldots\ldots, \Omega_5 \tag{1.40}$$

olarak ifade edilir. Bu torbaların birinden iadeli seçimle dört toptan oluşan bir örnek alındığını ve bu örnekte siyah top olduğu varsayılır. Bu örnek $\Omega_0, \Omega_1, \ldots\ldots, \Omega_5$'dan alınmış olabilir. Hangisinden alınmış olabileceği benzerlik yaklaşımı ile açıklanmaktadır. Örnekteki siyah top oranı,

$$p = \frac{3}{4} = 0{,}75 \tag{1.41}$$

olacaktır. Bu örneğin bu sonucu gerçekleştirme olasılığı en yüksek torbadan alınmış olabilir. Birleşik yoğunluk fonksiyonu,

$$\begin{aligned} f(X_1, X_2, X_3, X_4, X_5, P) &= \prod f(X_i, p) \\ &= f(X_1, p) f(X_2, p) f(X_3, p) f(X_4, p) f(X_5, p) \\ &= \left(p^{X_1} q^{1-X_1}\right)\left(p^{X_2} q^{1-X_2}\right)\left(p^{X_3} q^{1-X_3}\right)\left(p^{X_4} q^{1-X_4}\right)\left(p^{X_5} q^{1-X_5}\right) \\ &= p^{\sum X_i} q^{n-\sum X_i} \end{aligned} \tag{1.42}$$

olacaktır.

Çizelge 1.1. Torbadaki siyah top sayıları için olasılıklar

P	$f(X_i, p)$
0	$(0)^3(1)^{4-3} = 0$
0,2	$(0{,}2)^3(0{,}8)^{4-3} = 0{,}0064$
0,4	$(0{,}4)^3(0{,}6)^{4-3} = 0{,}0384$
0,6	$(0{,}6)^3(0{,}4)^{4-3} = 0{,}0864$
0,8	$(0{,}8)^3(0{,}2)^{4-3} = 0{,}1024$
1,0	$(1)^3(0)^{4-3} = 0$

Torbalar için olasılıklar Çizelge 1.1’ de verilmiştir. Görüldüğü gibi en yüksek olasılık siyah top oranı 0,8 olan 5. torba yani Ω_5 için bulunmuştur. Bu nedenle bu örnek 5. torbadan alınmış olabilir. Birleşik yoğunluk fonksiyonu 0,8 olasılığı için maksimize olmuştur. Yukarıda hesaplanan olasılıklar sıralı çekiliş olasılıklarıdır. Sırasız çekilişler için bunların $C_n^X = C_4^3 = 4$ ile çarpılması gerekir. Ama bu işlem sonucu değiştirmeyecektir. Dikkat edilirse dağılım binom dağılımıdır.

Maksimum olabilirlik yönteminde benzerlik fonksiyonundan yararlanılır. Benzerlik fonksiyonu örneğin birleşik olasılık dağılım fonksiyonudur. Bu fonksiyonu ℓ ile ifade edilirse n gözlem için,

$$\ell = f(X_1, X_2, \ldots\ldots, X_n) \tag{1.43}$$

olacaktır. $X_1, X_2, \ldots\ldots, X_n$ bağımsız ve X ile aynı dağılıma sahip değişkenler olduklarından benzerlik fonksiyonu,

$$\ell = f(X_1) \cdot f(X_2) \ldots\ldots f(X_n) \tag{1.44}$$

şekline dönüştürülebilir.

Benzerlik fonksiyonu ile birleşik olasılık dağılım fonksiyonları aynı fonksiyonlar olmakla birlikte yorumları farklıdır. Her iki fonksiyonda da k sayıda parametre ve n sayıda örnek birimi yer almaktadır. Fakat X ile ifade edilen gözlemleri ile, θ ile ifade edilen parametreler farklı yorumlanmaktadır. Birleşik olasılık fonksiyonunda X ’ler $(X_1, X_2, \ldots\ldots, X_n)$ değişken, parametreler $(\theta_1, \theta_2, \ldots\ldots, \theta_k)$ ise sabittir. Benzerlik fonksiyonunda ise örnek birimleri ile tahminciler belirlendiğinden X 'ler sabit θ'lar ise değişkendir.

Bu yöntemde belirlenen benzerlik fonksiyonunun parametrelere göre birinci türevi alınıp sıfıra eşitlenir. Böylece fonksiyonu maksimize edecek formüller elde edilir. Bu formüller θ parametrelerinin tahmincileri $\hat{\theta}$'lerdir. Yapılan bu türev alma işlemi ile fonksiyon maksimize edilmiş olabileceği gibi minimize edilmiş de olabilir. Bunu belirlemek için ikinci türevler alınarak işaretlerine bakılmalıdır.

Bir fonksiyonun kendisi maksimize etmekle logaritmasını maksimize etmek arasında fark yoktur. Fonksiyonun logaritmasının alınarak bunun maksimize edilmesi uygulamada işlemleri kolaylaştıracaktır. Bu nedenle maksimizasyon işlemi logaritmik benzerlik fonksiyonu ile yapılmaktadır. Logaritmik benzerlik fonksiyonunu L ile ifade edilirse,

$$L = \ln \ell \tag{1.45}$$

olacaktır. L'nin sağlayacağı kolaylıklar verilen örneklerin incelenmesi ile daha iyi değerlendirilebilir.

Maksimum olabilirlik yöntemi genellikle tutarlı ve etkin tahminciler vermektedir. X tesadüfi değişkeninin beklenen değeri,

$$E(X) = \sum_{i=1}^{2} p_i X_i = 0(1-p) + 1p = p \tag{1.46}$$

olacaktır. Yukarıdaki açıklamada 4 birimden oluşan bir örnek alınmıştı. Örnek birim sayısını n olarak alınıp, benzerlik fonksiyonunu belirlenirse,

$$\ell = f(X_1) f(X_2) f(X_3) \ldots f(X_n)$$

$$\ell = \left[p^{X_1}(1-p)^{1-X_1}\right]\left[p^{X_2}(1-p)^{1-X_2}\right]\left[p^{X_3}(1-p)^{1-X_3}\right]\left[p^{X_4}(1-p)^{1-X_4}\right]\left[p^{X_5}(1-p)^{1-X_5}\right] \tag{1.47}$$

Çarpım halinde olan ifadelerde tabanlar aynı olduğundan üsler toplanırsa,

$$\ell = p^{X_1 + X_2 + X_3 + \ldots + X_4}(1-p)^{(1-X_1)+(1-X_2)+(1-X_3)+\ldots+(1-X_n)}$$

$$\ell = p^{\sum_{i=1}^{n} X_i}(1-p)^{n-\sum_{i=1}^{n} X_i} \tag{1.48}$$

olacaktır. Benzerlik fonksiyonunun logaritması,

$$L = \ln \ell = \left(\sum_{i=1}^{n} X_i \right) \ln p + \left(n - \sum_{i=1}^{n} X_i \right) \ln (1-p) \tag{1.49}$$

olur. p 'nin tahmincisi belirleneceğine göre logaritmik benzerlik fonksiyonu L' nin p 'ye göre kısmi türevi alınarak sıfıra eşitlenir.

$$\frac{dL}{dp} = \left(\sum_{i=1}^{n} X_i \right) \left[\frac{1}{p} \right] + \left(n - \sum_{i=1}^{n} X_i \right) \left[\frac{1}{1-p} \right] (-1)$$

$$\left(\sum_{i=1}^{n} X_i \right) \left[\frac{1}{p} \right] + \left(n - \sum_{i=1}^{n} X_i \right) \left[\frac{1}{1-p} \right] (-1) = 0 \tag{1.50}$$

Paydalar eşitlenerek içler dışlar çarpımı yapılırsa,

$$\frac{\left(\sum_{i=1}^{n} X_i \right)(1-\hat{p}) + \left(n - \sum_{i=1}^{n} X_i \right)(-\hat{p})}{\hat{p}(1-\hat{p})} = 0$$

$$\left(\sum_{i=1}^{n} X_i \right)(1-\hat{p}) + \left(n - \sum_{i=1}^{n} X_i \right)(-\hat{p}) = 0$$

$$\sum_{i=1}^{n} X_i - \hat{p} \sum_{i=1}^{n} X_i - n\,\hat{p} + \hat{p} \sum_{i=1}^{n} X_i = 0$$

$$\sum_{i=1}^{n} X_i - n\,\hat{p} = 0$$

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n} \tag{1.51}$$

bulunur. Burada n olayın tekrar sayısı olduğundan örnekte 4'tür. Siyah top sayısı üç olduğundan, $\sum_{i=1}^{n} X_i = 3$ olacaktır. Bu durumda anakütle oranı,

$$\hat{p} = \frac{3}{4} = 0{,}75 \tag{1.52}$$

olarak tahmin edilecektir [7].

Örnek 1.3.1. Normal dağılmış bir anakütlenin ortalamasının tahmincisini maksimum olabilirlik yöntemi ile belirlenirse, normal dağılım iki parametreli bir dağılımdır ve genel ifadesi,

$$f(X)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{(X_i-\mu)^2}{2\sigma^2}} \tag{1.53}$$

şeklindedir. Bu ifadeyi,

$$f(X)=(2\pi\sigma^2)^{-1/2}e^{-\left(\frac{1}{2}\right)\left(\frac{X_i-\mu}{\sigma}\right)^2} \tag{1.54}$$

şeklinde yazılabilir. Fonksiyonun logaritması alınırsa,

$$\ln f(X)=-\frac{1}{2}\ln\left(2\pi\sigma^2\right)-\left(\frac{1}{2}\right)\left(\frac{X_i-\mu}{\sigma}\right)^2 \tag{1.55}$$

benzerlik fonksiyonu,

$$\ell = f(X_1)\,f(X_2)\,f(X_3)\ldots\ldots f\left(X_n\right) \tag{1.56}$$

olduğundan bunların tek tek logaritmaları için logaritmik benzerlik fonksiyonu,

$$L=\ln\ell=\ln f(X_1)+\ln f(X_2)+\ln f\left(X_3\right)+\ldots+\ln f\left(X_n\right)$$

$$L=\sum_{i=1}^{n}\ln f(X_i) \tag{1.57}$$

olacaktır. Burada $\ln f(X)$ 'i yerine koyulursa,

$$L=\sum_{i=1}^{n}\left[-\frac{1}{2}\ln\left(2\pi\sigma^2\right)-\frac{1}{2}\left[\frac{X_i-\mu}{\sigma}\right]^2\right]$$

$$=-\frac{1}{2}\sum_{i=1}^{n}\ln\left(2\pi\sigma^2\right)-\frac{1}{2}\sum_{i=1}^{n}\left[\frac{X_i-\mu}{\sigma}\right]^2$$

$$=-\frac{1}{2}n\ln\left(2\pi\sigma^2\right)-\frac{1}{2}\cdot\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i-\mu\right)^2$$

$$=-\frac{n}{2}\ln\left(2\pi\sigma^2\right)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(X_i-\mu\right)^2 \tag{1.58}$$

Dağılımın iki parametresi μ ve σ^2 tahmin edileceğinden, L'nin μ ve σ^2'ye göre birinci türevleri alınacaktır.

$$\frac{dL}{d\mu} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(X_i - \mu)(-1)$$

$$\frac{dL}{d(\sigma^2)} = -\frac{n}{2}\cdot\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(X_i - \mu)^2 \tag{1.59}$$

Bunlar sırası ile sıfıra eşitlenirse,

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(X_i - \hat{\mu})(-1) = 0$$

$$\sum_{i=1}^{n}(X_i - \hat{\mu}) = 0$$

$$\sum_{i=1}^{n} X_i - \sum_{i=1}^{n}\hat{\mu} = 0$$

$$\sum_{i=1}^{n} X_i - n\hat{\mu} = 0$$

$$\hat{\mu} = \frac{\sum_{i=1}^{n} X_i}{n} \tag{1.60}$$

olarak belirlenir. Normal dağılmış bir anakütle ortalamasının maksimum olabilirlik yöntemi ile belirlenen tahmincisi örnek ortalamasıdır. İkinci kısmi türevi sıfıra eşitlenirse,

$$-\frac{n}{2}\cdot\frac{1}{\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}\sum_{i=1}^{n}(X_i - \hat{\mu})^2 = 0 \tag{1.61}$$

Paydaları eşitlenip, içler dışlar çarpımı yapılırsa,

$$\frac{\hat{\sigma}^2 n + \sum_{i=1}^{n}(X_i - \hat{\mu})^2}{2\hat{\sigma}^4} = 0$$

$$-\hat{\sigma}^2 n + \sum_{i=1}^{n}(X_i - \hat{\mu})^2 = 0$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \hat{\mu})^2}{n} \tag{1.62}$$

elde edilir. Buna göre normal dağılmış bir anakütlenin varyasının maksimum olabilirlik yöntemi ile elde edilen tahmincisi örnek varyansıdır. Örnek varyansının, anakütle varyansının sapmasız tahmincisi olmadığı bunun yanında örnek varyansı anakütle varyansının asimtotik etkin tahmincisidir [7].

1.4. R Programı

R dili ilk olarak Yeni Zelanda' daki Aucland Üniversitesi İstatistik Bölümü'nden Ross Ihaka ve Robert Gentleman tarafından yazılmıştır. Daha sonra dünyanın çeşitli yerlerindeki araştırmacılar R'yi geliştirmek için bir araya gelmiş ve 1997'de bu gruba "R core team" adı verilmiştir.

R diline, ilk geliştiricileri olan Ross Ihaka ve Robert Gentleman tarafından S diline atıfta bulunarak "R" ismi verilmiştir. R dilinin tasarımı önemli ölçüde Becker, Chamber ve Wilks'in geliştirdiği S dili le Sussman'ın geliştirdiği Scheme dillerinden etkilenmiştir. Görünüm açısından S diline benzeyen R, uygulama ve anlamsal yönden Scheme diline yakındır. S-PLUS yazılımının akademik ve öğretim amaçlı kullanılmasında lisans ücretlerinin pahalı bulunması nedeniyle Yeni Zelanda'lı iki adını verdikleri programlama dilini geliştirmeye karar vermişlerdir.

R dilinin ilk sürümü "R core team " tarafından 29 Şubat 2000 tarihinde yayınlanmıştır. İstatistiksel yazılım geliştirme ortamı veri manipülasyonu, hesaplama ve grafik gösterim için tasarlanmıştır. Fonksiyonel bir programlama dili olan R istatistikçiler ve matematikçiler için kod yazmayı kolaylaştıran fonksiyonlara sahiptir.

R-Project2in web sitesinde yapılan tanıma göre R, istatistiksel hesaplamalar ve grafikler için bir dil ve ortamdır. R, yaygın olarak kullanılan SPSS, SAS gibi istatistik paket programlarından farklıdır. R bir istatistik paket program değil istatistiksel yazılım geliştirme ortamıdır.

Avantajları;

1. Ücretsiz temin edilebilmesi,
2. Nesne yönelimli bir programlama dili olması,
3. Farklı amaçlar için geliştirilmiş paketler eklenerek fonksiyonelliğinin arttırılabilmesi,
4. 2-D, 3-D gelişmiş grafik araçlarına sahip olması,
5. Analizin nasıl yapılması gerektiği hakkında düşündürmesi,
6. Hızlı olmasıdır.

Dezavantajları;

1. Öğrenmesi zor bir programlama dilidir.

2. Gelişmiş veri işleme özelliklerine sahip olmasına rağmen bunların kullanılabilmesi özellikle dizi ve matris işlemlerine hakim olmayı gerektirir.
3. Çok büyük veri dosyaları ile çalışmak için uygun değildir. Birkaç yüz megabyte'dan daha büyük veri dosyaları açılmak istendiğinde yetersiz bellek sorunu meydana gelebilir.
4. Ticari bir ürün olmadığı için kullanımında karşılaşılan sorunların iletileceği müşteri destek birimi yoktur.
5. Hata yapmak kolaydır ve tespit edebilmesi zor olabilir.
6. Veriyi işlenecek hale getirmek zaman alıcı ve hataya açık bir süreçtir [8-9].

1.5. Örnekleme Yöntemleri

Örnekleme yöntemlerinin geliştirilmesi çok yeni olmasına karşın, örnekleme yöntemlerinden yararlanma çok eskidir. Bununla birlikte, örnekleme kuramı çok kısa bir süre içerisinde hızla gelişmiş ve çeşitli örnekleme yöntemleri ortaya koymuştur. Böylece örnekleme yöntemleri geniş bir uygulama alanı bulmuştur [4-10]. Çok çeşitli örnekleme yöntemleri, örnekleme hatası hakkında objektif bir kriterin sağlanıp sağlanamaması bakımından iki büyük bir gruba ayrılırlar [10].

1. Olasılıksal olmayan örnekleme yöntemleri
2. Olasılıksal örnekleme yöntemleri

1.5.1. Olasılıksal Olmayan Örnekleme Yöntemleri

Örnek birimlerinin gelişigüzel olasılıklarla seçildiği örnekleme yöntemine olasılıksal olmayan örnekleme adı verilir. Bu yöntemde örneklem birimlerinin herbirinin bir seçim olasılığı söz konusu olmadığından varyans hesaplanamaz. Yani tahminlerin örnekleme hataları ile ilgili objektif bir ölçü verilemez. Bu nedenle, olasılıksal örneklemeye göre daha az başvurulan bir yöntemdir. Tahminlerin duyarlılıkları ancak sübjektif olarak yorumlanabilir. Bilimsel araştırmalarda bu örnekleme yöntemine başvurulmamakla birlikte uygulama kolaylığı olması yönünden pek çok alanda özellikle kamuoyu yoklamalarında sık sık kullanılmaktadır [4].

Olasılıksal olmayan örnekleme yöntemlerinin belli başlı üç türü vardır [10].

1. Keyfi örnekleme
2. Dilim örneklemesi
3. Kota örneklemesi.

1.5.1.1. Keyfi Örnekleme

Keyfi örnekleme yeni bir örnekleme tekniği olmayıp esası 100 yıllarca öncesine dayanır. Fakat, keyfi örneklemenin iş problemlerine uygulanması son zamanlarda geliştirilmiştir. Keyfi örneklemede, örneği oluşturan birimler olasılık esasına göre seçilmezler, seçimi yapanın arzu ve düşüncelerine bağlı olarak seçilirler. Bu nedenle örneklemeci, yığının her yanından, yığını en iyi bir şekilde temsil edeceğini sandığı ve kendilerinden gerekli bilgilerin kolayca toplanabileceği yeterli sayıda birim seçimine çalışır. Örneğe seçilen birimlerden elde edilen bilgilerden yığına ilişkin istatistiklerin tahminsel değerleri hesaplanır. Bu seçim işi, örneklemeciden örneklemeciye değişir. Bir keyfi seçimin diğer bir keyfi seçime tercih edilmesi için objektif bir kriter yoktur. Örnek birimleri keyfi olarak seçildiğinde, elde edilen sonuçların güven sınırları hesaplayacak bir objektif yöntem yoktur. Böyle yöntemlerde örneğe seçilen bir birimin örneğe girme olasılığı bilinmemektedir [10].

1.5.1.2. Dilim Örneklemesi

Kitle çok geniş olduğundan örneklem birimlerine ulaşma maliyeti (para ve zaman yönünden) yüksek ise, keyfi örneklemenin yerine dilim örneklemesi tercih edilir. Geniş kitle bir takım dilimlere ayrılır. Bu dilimlerden kitleyi simgeleyebilecek örneklem birimlerinin çekimi yöntemine dilim örneklemesi adı verilir [4].

1.5.1.3. Kota Örneklemesi

Kitle, incelenen özellikleri yönünden farklılık gösteren bazı alt gruplara ayrılır. İncelenen özelliklerin önem dereceleriyle orantılı örneklem biriminin çekildiği örnekleme yöntemine kota örneklemesi adı verilir.

Kitle çok geniş bir alana yayılmış ve birim başına düşen maliyetler arasında önemli bir farklılık yok ise, kitle incelenecek özellikleri yönünden kolaylıkla alt gruplara ayrılabiliyorsa, kota örneklemesi diğer örnekleme yöntemlerine tercih edilir [4].

1.5.2. Olasılıksal Örnekleme

Örneklem birimlerinin belirli olasılıklarla çekildiği örnekleme yöntemine olasılıksal örnekleme adı verilir. Olasılıksal örnekleme, yığındaki her birime hesaplanabilen ve sıfır olmayan bir seçilme olasılığı tanıyan örneklemedir. Örneklem dağılımının değişim ölçüsü, bir diğer deyişle, varyans yardımıyla parametrenin içinde bulunduğu sınırlar tahmin edilir. İşte olasılıksal örneklemenin yararı bu sınırların belirlenebilmesidir [4-10].

1.5.2.1. Çok Aşamalı Örnekleme

Örneklem birimleri bazı araştırmalarda bir aşamada örnekleme çekilir. Bu yönteme tek aşamalı örnekleme adı verilir. Eğer örneklem birimleri birden çok aşamada örnekleme çekiliyor ise, aşama sayısına göre iki aşamalı, üç aşamalı, … adlarını alır. Bu yönteme genel olarak çok aşamalı örnekleme adı verilir. Bazı araştırmalarda, birinci aşamada alınan örneklem birimlerinin içerdikleri kitle birimleri birbirine benzer özellik gösterir. Bu durumda örneklem biriminin tümünü incelemek yerine bundan yine bir örneklem seçmek para, zaman ve emek yönünden tasarruf sağlayacağı gibi sonuçta örneklemin kitleyi simgeleme niteliğini etkilemez [4].

1.5.2.2. Basit Tesadüfi Örnekleme

Herbir örneklem birimine eşit seçilme olasılığı vererek (seçilen birim yerine konularak) ve seçilen birimin bir kez daha örnekleme alınmadığı yönteme basit tesadüfi örnekleme denir. Burada herbir örneklem birimine eşit seçilme olasılığı verilmesinin anlamı örneklem uzayından herbir örneklemin eşit olasılıkla seçilmesi anlamına gelir.

Basit tesadüfi örnekleme, örnekleme kuramının kurulmasında temel teşkil etmesi bakımından yeri ve önemi büyüktür [4].

1.5.2.3. Tabakalı Örnekleme

Örneklem birimlerinin herhangi bir ölçüsüne ilişkin birimden birime değişim büyük ise, bu durumda kitle, değişkenliği daha küçük alt gruplara ayrılabilir. Böylece kitle varyansı büyük iken, alt grupların varyansı daha küçük olacaktır. Bu ise, duyarlılıkta önemli bir kazanç sağlar. Kitle herbir kitle birimi bir ve yalnız bir tabakaya ait olacak ve hiçbir kitle birimi açıkta kalmayacak; tabaka içi değişim olabildiğince küçük, tabakalar arası değişim oldukça büyük kalacak şekilde alt gruplara bölünüp örneklemin herbir tabakadan ayrı ayrı ve birbirinden bağımsız olarak çekildiği örnekleme yöntemine tabakalı örnekleme adı verilir.

Herbir tabakaya basit tesadüfi örnekleme yönteminin uygulandığı tabakalı örneklemeye tabakalı tesadüfi örnekleme adı verilir [4].

1.5.2.4. Sistematik Örnekleme

Kitle birimlerinin düzgün bir şekilde sıralanabildiği varsayılırsa ilk k birimden herhangi birinin başlangıç noktası olarak alındığı ve bundan sonra gelen her k 'ıncı birimin örnekleme seçildiği yönteme sistematik örnekleme adı verilir. Sistematik örneklem çekimi son derece kolay ve kısa zamanda gerçekleştiğinden bazı araştırmalarda tercih edilir [4].

2. ÖNCEKİ ÇALIŞMALAR

2.1. Dağılımlardan Rasgele Sayı Üretilmesi

Bu bölümde düzgün dağılıma sahip olmayan rasgele değişkenlerin dağılımlarından, yani düzgün dağılımın dışındaki dağılımlardan sayı üretmek için kullanılan yöntemler ele alınacaktır.

2.1.1. Ters Dönüşüm Yöntemi

F dağılım fonksiyonuna sahip bir X rasgele değişkenin dağılımından sayı üretmek için en çok kullanılan yöntemlerden biri, F dağılım fonksiyonunun genelleştirilmiş tersi denen

$$F^{-} : (0,1) \rightarrow R$$

$$u \rightarrow F^{-}(u) = \inf \{x : F(x) \geq u\} \tag{2.1}$$

fonksiyonuna dayalı $X = F^{-}(U)$ dönüşümünü kullanmaktır. Burada U rasgele değişkeni $(0,1)$ aralığı üzerindeki düzgün dağılıma, yani $U(0,1)$ dağılıma sahiptir.

$$x \in \{F^{-}(u) : u \in (0,1)\} \text{ için } F(F^{-}(u)) \geq u \text{ ve } F^{-}(F(x)) \leq x$$

$$\{(u,x) : F^{-}(u) \leq x\} = \{(u,x) : F(x) \geq u\} \tag{2.2}$$

olmak üzere

$$P(F^{-}(U) \leq x) = P(U \leq F(x)) = F(x) \tag{2.3}$$

dır. $F^{-}(U)$ rasgele değişkeninin dağılım fonksiyonu F dir, yani $F^{-}(U)$ dönüşümü ile ortaya çıkan rasgele değişken X rasgele değişkeninin kendisidir. $X = F^{-}(U)$ dönüşümü integral dönüşümü olarak bilinmektedir. $U(0,1)$ düzgün dağılımdan üretilen sayılar integral dönüşümü sonucunda X rasgele değişkenin dağılımından üretilmiş sayılar olacaktır. Böylece herhangi bir X rasgele değişkenin dağılımından sayı üretme işlemi çözülmüş gibi görünmektedir, ancak buradaki zorluk bazı dağılımlar için F^{-} genelleştirilmiş ters fonksiyonunun açık bir ifadesinin elde edilememesidir. Örneğin $X \approx N(0,1)$ için,

$$f(x) = \frac{1}{2}\int_{-\infty}^{x} e^{\frac{z^2}{2}}\, dz \quad , x \in \Re \tag{2.4}$$

olmak üzere;

$$F^{-}(u) = F^{-1}(u) \quad , u \in (0,1) \tag{2.5}$$

dır. $F^{-1}(U)$ değerleri u ya bağlı olarak açık bir şekilde kolayca yazılıp hesaplanamamaktadır. X sürekli bir rasgele değişken olduğunda dağılımın destek kümesi üzerinde F artan bir fonksiyon olmakta ve bu durumda yukarıdaki dönüşüm $X = F^{-1}(U)$ biçimini almaktadır. Bu durumda, $X = F^{-1}(U)$ rasgele değişkenin dağılım fonksiyonu,

$$\begin{aligned} P(X \le x) &= P\left[F^{-1}(U) \le x\right] = P\left[F\left(F^{-1}(U)\right) \le F(x)\right] \\ &= P\left[U \le F(x)\right] = F(x) \end{aligned} \tag{2.6}$$

dır. F dağılım fonksiyonunun F^{-1} ters fonksiyonunun değerlerinin hesaplanabilir olması durumunda sürekli bir X rasgele değişkeninin dağılımından sayı üretmek için algoritma aşağıdaki gibidir [11].

1. $U(0,1)$ dağılımından U üretilir.
2. $X = F^{-1}(U)$ hesaplanır.

2.1.1.1. Ters Dönüşümün Tabular Yaklaşımı

Çiftler dizini $(F(x_i), x_i)$ tablo haline getirilir, burada $x_i < x_{i+1}$ dir. Sıklıkla bu durum önemli bir kurulum zamanı gerektirir. Aşağıdaki algoritma, bu çiftleri ekleyen parçalı doğrusal yaklaşımı F 'ye çevirir [12].

1. $F(X_i) \le U \le F(X_{i+1})$ dan X_i bulunur.

2. $X = \dfrac{\left[F(X_{i+1}) - U\right]X_i + \left[U - f(X_i\right]X_{i+1}}{F(X_{i+1}) - F(X_i)}$ 'e geri dönülür.

2.1.1.2. Ampirik Dağılım Fonksiyonlar

n gözlemleri $X_1, X_2, \ldots, X_n$'den doğru ancak bilinmeyen dağılıma sahip olduğu varsayılır, böylece $X_1 \le X_2 \le \ldots \le X_n$ olarak sıralanır. $F(X_i) = i/n$ yaklaşımını varsayılırsa, ampirik dağılım fonksiyonunun kullanıldığını söyleyebiliriz. $X = F^{-1}(U)$ üretirken ampirik dağılım fonksiyonunda sabit eklemenin kullanılması $1/n$ olasılık ile $X = X_i$ ayarına karşılık gelir [12].

2.1.1.3. Üstel ve Ampirik Dağılımın Birleşimi

$X_1 \le X_2 \le \ldots \le X_n$ başlangıç gözlemleri olup daha sonra ilk $n-k$ gözlemlerine bir parçalı doğrusal dağılım fonksiyonu ve k en büyük gözlemlerinin bir üstel değiştirilmiş uydurmasıdır. $F(0) = 0$ varsayılırsa ve $X_0 = 0$ tanımlanırsa dağılım fonksiyonu,

$$F(t) = \begin{cases} i/n + (t - X_i)/\left[n(X_{i+1} - X_i)\right] & \textit{for } X_i \le t \le X_{i+1}, \\ 1 - (k/n)\exp\left(-(t - X_{n-k})/\theta\right) & \textit{for } t > X_{n-k}, \end{cases} \quad i = 0,1,\ldots,n-k-1 \tag{2.7}$$

burada

$$\theta = \left(X_{n-k}/2 + \sum_{i=n-k+1}^{n} (X_i - X_{n-k}) \right) \Big/ k. \tag{2.8}$$

Birleşik dağılımın ortalaması $1 \le k \le n$ için $(X_1 + \ldots X_n)/n$ dir. Başka bir rutin hesaplama varyansının

$$\frac{1}{3n}\left[2\sum^{n-k-1} X_i^2 + \sum_{i=1}^{n-k-1} X_i X_{i+1} + X_{n-k}^2\right] + \frac{k}{n}\left[(\theta + X_{n-k})^2 + \theta^2\right] - \left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]^2 \tag{2.9}$$

olduğunu da gösterir.

Bu birleşik dağılımından ters dönüşümü yöntemi ile değişken üretmek için aşağıdaki algoritma kullanılır;

1. $(0,1)$ dağılımından U üretilir.

2. $U > 1 - k/n$ ise bu durumda $X = X_{n-k}\theta \log(n(1-U)/k)$'e geri dönülür;

3. Aksi durumda $V \leftarrow nU, I \leftarrow \lfloor V \rfloor$ kurulur ve $X = (V - I)(X_{I+1} - X_I) + X_I$' ya geri dönülür [12].

2.1.1.4. Ters Dönüşümün Fonksiyonel Yaklaşımları

F^{-1}için yada F için kapalı form ifadesi olmasa bile, yine de yardımcı tablolar kullanılmadan değerlendirilmesi kolay bir fonksiyonla F^{-1}'li yaklaştırmak mümkündür. Olasılık ve istatistikte görülen birçok fonksiyon için yaklaşık değerlerin listelerini verir [13-14]. Bir g fonksiyonu için en yakın tahminler aşağıda verilen türlerden biridir;

1. Sürekli kesir:

$$g(t) \cong Q_0(t), \tag{2.10}$$

burada

$$Q_i(t) = a_i + b_i / Q_{i+1}(t) \qquad i = 0,1,\ldots,n-1 \tag{2.11}$$

için

$$Q_n(t) = a_n(t) \tag{2.12}$$

dir.

2. Rasyonel yaklaşım:

$$g(t) \cong \frac{a_0 + a_1 t + a_2 t^2 + \ldots + a_n t^n}{1 + b_1 t + b_2 t^2 + \ldots + b_n t^n} \tag{2.13}$$

Bu yaklaşım altı ondalık basamakta göreli bir doğruluk derecesine sahiptir ve $0{,}5 < u < 1$için geçerlidir. Normalin simetrisi $U = 1 - U$ ve $X = -X$ dönüşümleri ile $0 < U < 1$'e genişletilmesini sağlar [12].

2.1.2. Kabul-Red Yöntemi

X rasgele değişkenin olasılık yoğunluk fonksiyonu f ve aldığı değerler kümesi D olsun. V rasgele değişkenin olasılık yoğunluk fonksiyonu g ve aldığı değerlerin kümesi D olmak üzere V sayıları kolayca üretilirse $a > 0$ sabiti ve $\forall x \in D$ için;

$$f(x) \leq a \cdot g(x) \tag{2.14}$$

koşulu sağlanır. Bu durumda aşağıdaki algoritma ile olasılık yoğunluk fonksiyonu f olan dağılımdan sayı üretilebilir.

1. V sayısı olasılık yoğunluk fonksiyonu g olan dağılımından üretilir.
2. V den bağımsız olarak $U \approx U(0,1)$ üretilir.

$U \cdot a \cdot g(V) \leq f(V)$ ise $X = V$ kabul edilir yani bir X sayısı üretilmiş olur, aksi durumda reddedilir yani 1. adıma geçilir (başka bir ifade ile $Y \approx U(0, a \cdot g(V))$ üretilir. $Y < f(V)$ ise X kabul edilir, aksi durumda reddedilir.)

Bu algoritma ile üretilen X sayıları için,

$$P(X \leq x) = P(V \leq x \mid Uag(V) \leq f(V))$$

$$= \frac{P\left(V \leq x, U \leq \frac{f(V)}{ag(V)}\right)}{P\left(U \leq \frac{f(V)}{ag(V)}\right)}$$

$$= \frac{\int_{-\infty}^{x} \left(\int_{0}^{\frac{f(V)}{ag(V)}} du \right) g(v)\,dv}{\int_{-\infty}^{\infty} \left(\int_{0}^{\frac{f(V)}{ag(V)}} du \right) g(v)\,dv} = \frac{\frac{1}{a}\int_{-\infty}^{x} f(v)\,dv}{\frac{1}{a}\int_{-\infty}^{\infty} f(v)\,dv} = \int_{-\infty}^{x} f(v)\,dv \tag{2.15}$$

olmak üzere, üretilen sayılar olasılık yoğunluk fonksiyonu f olan bir dağılımdandır.

Olasılık yoğunluk fonksiyonu f olan dağılımdan X sayılarını üretmek için yukarıdaki algoritma aşağıdaki gibi de yazılabilir [11].

1. Birbirinden bağımsız olarak $V \approx g$ ve $U \approx U(0,1)$ üretilir.
2. Eğer $U \leq \dfrac{f(V)}{ag(V)}$ ise $X = V$ olur ve X sayısı üretilir.

2.1.3. Ayrışım Yöntemi

Ayrışım yönteminde rasgele sayılar üretilecek olan dağılımın f olasılık yoğunluk fonksiyonu, kolayca sayılar üretebilen bazı olasılık yoğunluk fonksiyonların karması olarak yazılmaktadır. $i = 1, 2, \ldots, n$ için $p_i \geq 0, \sum_{i=1}^{n} p_i$ ve $g_1, g_2, \ldots, g_n$ 'ler olasılık yoğunluk fonksiyonları olmak üzere,

$$f(x) = \sum_{i=1}^{n} p_i \, g_i \tag{2.16}$$

olduğunda, f 'nin dağılımından sayı üretmek için algoritma aşağıdaki gibidir.

1. $i = 1, 2, \ldots, n$ sayıları sırasıyla $p_1, p_2, \ldots, p_n$ olasılıkları ile alan kesikli dağılımdan bir i sayısı üretilir.
2. Olasılık yoğunluk fonksiyonu g_i olan dağılımdan bir sayı üretilir ve bu üretilen çıkılır(alınır).

Y bir rasgele değişken, D_Y bu rasgele değişkenin aldığı değerlerin kümesi ve h, Y 'nin olasılık yoğunluk fonksiyonu olmak üzere, olasılık yoğunluk fonksiyonlarının bir $\{g_y : y \in D_y\}$ ailesi yardımıyla,

$$f(x) = \int_{D_Y} g_y(x) h(y) dy \tag{2.17}$$

biçimde yazıldığında f 'nin dağılımından sayı üretmek için algoritma aşağıdaki gibidir.

1. Olasılık yoğunluk fonksiyonu h olan dağılımdan bir Y sayısı üretilir.
2. Olasılık yoğunluk fonksiyonu g_Y olan dağılımdan bir sayı üretilir ve bu üretilen sayı alınır.

Bu algoritma bir X rasgele değişkenin başka bir Y rasgele değişkenine göre koşullu dağılımının bilinmesi ve bu koşullu dağılım ile Y 'nin dağılımından sayı üretilebilir olmasında da kullanılabilir. $Y = y$ verildiğinde X 'in koşullu olasılık yoğunluk fonksiyonu g ve Y 'nin marginal olasılık yoğunluk fonksiyonu h olmak üzere X 'in olasılık yoğunluk fonksiyonu,

$$f(x) = \int g(x \mid y) h(y)\, dy \tag{2.18}$$

dır [11].

2.1.4. Bazı Sürekli Dağılımlardan Sayı Üretme

2.1.4.1. Cauchy Dağılımı

Cauchy dağılımı $\alpha = 1$ ile simetrik sabit bir dağılım olup 1serbestlik derecesine sahip t dağılımıdır. Medyanı μ ve σ ölçeğe sahip bir Cauchy dağılımının olasılık yoğunluk fonksiyonu,

$$f(x) = \sigma / \left[\pi\left(\sigma^2 + [x-\mu]^2\right)\right], \qquad -\infty \le x \le \infty \tag{2.19}$$

dağılım fonksiyonu,

$$F(x) = \frac{1}{\pi}\arctan\left((x-\mu)/\sigma\right) + \frac{1}{2} \tag{2.20}$$

Ters çevirerek üretme:

$$X = \sigma \tan\left(\pi\left(U - \frac{1}{2}\right)\right) + \mu \tag{2.21}$$

Bu son denklem fiziksel olarak (μ, σ) merkezli bir ibrenin dönme hareketiyle X, x eksenin kesim noktası olarak ifade edilir. İki bağımsız standart normal değişkenin oranı $\mu = 0$ ve $\sigma = 1$ ile bir Cauchy değişkeni olur. Bununla birlikte, bu verimli bir üretim yöntemi değildir [12].

2.1.4.2. Weibull Dağılımı

Weibull dağılımına sahip X rasgele değişkenin olasılık yoğunluk fonksiyonu,

$$f(x)=\begin{cases}\frac{1}{\theta}c\left(\frac{x}{\theta}\right)^{c-1}e^{-\left(\frac{x}{\theta}\right)^{c-1}} & ,x>0\\ 0 & ,d\cdot y\end{cases} \tag{2.22}$$

ile verilir $(\theta>0, c>0)$. Dağılım fonksiyonu,

$$F(x)=1-e^{-\left(\frac{x}{\theta}\right)^{c}} \tag{2.23}$$

olmak üzere

$$F^{-1}(y)=\theta(-Log(y))^{\frac{1}{c}} \tag{2.24}$$

dır. Ters dönüşüm yöntemi ile bu dağılımdan sayı üretilebilir [11].

2.1.4.3. Laplace ve Üstel Kuvvet Dağılımları

Üstel kuvvet dağılım sınıfı simetrik bir dağılım biçiminin etkisini incelemek için elverişlidir. Olasılık yoğunluk fonksiyonu,

$$f(x)=\mu\cdot\exp\left(-|\mu x|^{\alpha}\right)/[2\Gamma(1+1/\alpha)] \tag{2.25}$$

için

$$-\infty<x<\infty\ ,\alpha\geq 1 \text{ve } \mu=0 \tag{2.26}$$

dır. α 'nın çeşitli değerleri için sırasıyla $\alpha\to\infty$ ve $\alpha=2$ olduğunda, bu sınıf $(-1/\mu, 1/\mu)$ 'de düzgün ve normal özel durumlar da içerir. $\alpha=1$ olduğu zaman, bu dağılımdan elde edilen bir değişken, her biri $1/\mu$ beklentisi olan iki aynı dağılımlı bağımsız üstel değişkenin farkıyla aynı dağılıma sahiptir. Bu durumda buna Laplace dağılımı denir. Üstel kuvvet dağılım $\left(E\left[X^4\right]/E^2\left[X^2\right]\right)$ nin ortalama, varyans ve basıklığı $0, \Gamma(3/\alpha)/\left[\mu^2\Gamma(1/\alpha)\right]$, ve $\Gamma(5/\alpha)\Gamma(1/\alpha)/\left[\Gamma(3/\alpha)^2\right]$ dir. Dağılımı moment eşlemesi ile uydurmak basittir. α parametresi yalnızca basıklık tarafından belirlenir ve bu nedenle basit bir arama yöntemi ile bulunabilir. Varyans daha sonra μ yi belirler. Üstel kuvvet değişkenlerini redderek üretir [15]. $\alpha=2$ ise ve Laplace $1\leq\alpha\leq 2$ ise büyüklük yoğunluğu için normali kullanılır. $\alpha=1$ veya $\alpha=2$ dönüşüm; örneğin $\alpha=1$ için:

1. U üretilir.

2. $U \le \frac{1}{2}$ ise, $X = \log(2U)/\mu$'e geri dönülür, aksi durumda $X = -\log(2-2U)/\mu$'ye geri dönülür [12].

2.1.4.4. Genelleştirilmiş Laplace Dağılımı

Genelleştirilmiş Laplace dağılımına sahip X rasgele değişkeninin olasılık yoğunluk fonksiyonu,

$$f(x) = \frac{1}{2\theta\Gamma\left(1+\frac{1}{\alpha}\right)} e^{-\left|\frac{x}{\theta}\right|^{\alpha}} \quad , -\infty < x < \infty \tag{2.27}$$

şeklindedir $(\theta > 0, \alpha > 0)$. Bu dağılımın aşağıdaki özellikleri dikkat çekicidir.

$$\alpha \to \infty \text{ için } X \approx U(-\theta, \theta)$$

$$\alpha = 2 \text{ için } X \approx N(0, \sigma^2 = \theta^2)$$

$$\alpha = 1 \text{ için Laplace dağılımı}$$

Bu dağılımdan kabul-red yöntemine göre sayı üretmek için $g(x)$ fonksiyonu aşağıdaki gibi seçilebilir [12].

$$\alpha \ge 2 \text{ için } g(x) \equiv \text{Normal dağılım}$$

$$1 \le \alpha \le 2 \text{ için } g(x) \equiv \text{Laplace dağılımı}$$

2.1.4.5. Erlang ve Gama Dağılımları

Erlang dağılımı λ ve k olmak üzere iki parametreye sahiptir. Bir Erlang rasgele değişkeni her biri $1/\lambda$ beklenti ile birlikte olan k bağımsız aynı dağılımlı üstel değişkenlerin toplamıdır. Açık üretme yöntemi,

$$X = -\sum_{i=1}^{k} \log(U_i)/\lambda \tag{2.28}$$

Hafifletmek için basitleştirilirse,

$$X = -\log\left(\prod_{i=1}^{k} U_i\right)\Big/\lambda \tag{2.29}$$

dir. Olasılık yoğunluk fonksiyonu,

$$0 \le x \text{ için } f(x) = \lambda e^{-\lambda x} (\lambda x)^{k-1} / (k-1)! \tag{2.30}$$

dağılım fonksiyonu,

$$F(x) = 1 - \sum_{j=0}^{k-1} e^{-\lambda x} (\lambda x)^j \big/ j! \tag{2.31}$$

dir. Gama dağılımı k 'nın sadece bir tam sayı olmasından ziyade herhangi bir negatif olmayan gerçek sayı olmasını sağlayarak elde edilir. Gama dağılımının olasılık yoğunluk fonksiyonu,

$$0 \le x \text{ için } f(x) = \lambda e^{-\lambda x} (\lambda x)^{k-1} \big/ \Gamma(k) \tag{2.32}$$

dır. Sırasıyla ortalama, varyansı k/λ ve k/λ^2 dir. $k \ge 1$ için mod $(k-1)/\lambda$ ve varyansı ortalamanın karesine eşit veya daha azdır. $k \le 1$ için mod 0(burada $f(x)$ sonsuzdur) ve varyansı ortalamanın karesinden daha büyüktür. Şekil parametresi k olan bir gama değişkeni üretmek için basit fakat yavaş bir algoritma tanımlanmıştır [16]. Bu algoritma

1. $m = \lfloor k \rfloor$ ve $q = k - m$
2. $Z = -\log\left(\prod_{i=1}^{m} U_i \right)$
3. $q,\ q-1$ parametreleriyle beta dağılımından W rasgele değişkeni üretilir.
4. $Y = \log U$ ve U düzgün tesadüfi bir sayı üretilir.
5. $X = (Z + WY)/\lambda$ üretilir.

k büyük olduğunda 2.adım çok sayıda düzgün rasgele değişkenleri gerektirir. $k \ge 2$ için en azından hızlı olan red'e dayalı bir algoritma önermektedir [17]. Ortalama $k = 1$ ve $\theta = \left(1 + \sqrt{4k-3}\right)/2$ ölçüsü ile Laplace dağılımı beta dağılımını sınırlar. Bu algoritma aşağıdaki gibidir.

1. $k-1$ ortalama ve θ ölçüsü ile bir X Laplace değişkeni üretilir.
2. $X < 0$ ise 1'e gidilir.
3. U üretilir.
4. $U \le T(X)$ ise burada;

$$T(x) = \left| \frac{(\theta-1)x}{\theta(k-1)} \right|^{k-1} \exp\left[-x + \left(|x - (k-1)| + (k-1)(\theta+1) \right)/\theta \right] \tag{2.33}$$

X alınsın, aksi durumda 1'e gidilir [12].

2.1.4.6. Ki-kare Dağılımı

Ki-kare dağılımına sahip bir X rasgele değişkeninin dağılımının $\Gamma\left(\alpha=\frac{r}{2},\beta=2\right)$ olduğu gözönüne alınarak, r sayısının değerine göre aşağıdaki algoritma ile dağılımdan sayı üretilebilir. Parametresi $\theta=2$ olan üstel dağılımdan $\left\|\frac{r}{2}\right\|$ tane X_i sayıları üretilir. Z standart normal dağılımdan üretilen sayı olmak üzere,

1. r çift sayı ise $X=\sum_{i=1}^{r/2} X_i$
2. r tek sayı ise $X=\sum_{i=1}^{[r/2]} X_i+Z^2$

alınır [11].

2.1.4.7. F Dağılımı

F dağılımı, istatistiksel hipotez testinde örneğin varyans analizinde önemli olan iki parametreli dağılımdır. Sırasıyla w ve v serbestlik dereceleri ile Y_1 ve Y_2 ki-kare olmak üzere,

$$X=(Y_1/w)/(Y_2/v) \tag{2.34}$$

dir. Bu durumda w ve v parametreleri ile F -dağılımlıdır. Olasılık yoğunluk fonksiyonu,

$$f(x)=\frac{(v/w)^{v/2}}{\beta(w/2,v/2)}x^{(w/2)-1}(x+v/w)^{-(w+v)/2}, \quad x\geq 0, \tag{2.35}$$

burada $\beta(a,b)$ beta fonksiyonudur. Sırasıyla ortalama, varyans ve mod ; $(v>2)$ için $v/(v-2)$, $(v>4)$ için $2v^2(v+w-2)/\left[w(v-4)(v-2)^2\right]$,ve $(v>2)$ için $v(w-2)/[w(v+2)]$ dir. F değişkeni üretmek için bir beta değişkenine dönüştürülür.

1. $a/2$ ve $b/2$ parametreleri ile beta dağılımından Z 'yi üretilir.
2. Açık bir notasyon kullanarak $F(a,b)=(b/a)Z/(1-Z)$ 'ye geri dönülür [12].

2.1.4.8. Lojistik Dağılım

Lojistik dağılıma sahip bir X rasgele değişkenin olasılık yoğunluk fonksiyonu ve dağılım fonksiyonu,

$$f(x) = \exp\left(-[x-a]/b\right) / \left[b\left(1+\exp\left(-[x-a]/b\right)\right)^2\right] \tag{2.36}$$

$$F(x) = 1/\left[1+\exp\left(-[x-a]/b\right)\right] \tag{2.37}$$

olmak üzere,

$$F^{-1}(U) = a - b\log\left(\frac{1}{U} - 1\right) \tag{2.38}$$

ile bu dağlımda sayı üretilebilir. Lojistik dağılım bir setin orta aralığı için asimtotik bir dağılımı olup bağımsız aynı dağılımlı gözlemlerinin en küçüğü ve en büyüğünün ortalaması olarak tanımlanmaktadır [12].

2.1.5. Normal Dağılımlardan Sayı Üretme

2.1.5.1. Box-Muller Yöntemi

$X \approx N(0,1)$, $Y \approx N(0,1)$ ve X, Y bağımsız rasgele değişkenler olup R ve $\theta, (X,Y)$ rasgele vektörünün kutupsal koordinatları olmak üzere,

$$\begin{aligned} X &= R\cos\theta \\ Y &= R\sin\theta \\ X^2 + Y^2 &= R^2 \end{aligned} \tag{2.39}$$

olup,

$$\begin{aligned} X &= R\cos\theta \\ Y &= R\sin\theta \end{aligned} \tag{2.40}$$

dönüşümünün Jakobien matrisi,

$$J = \begin{bmatrix} \frac{\partial X}{\partial R} & \frac{\partial X}{\partial R} \\ \frac{\partial Y}{\partial R} & \frac{\partial Y}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \cos\theta & -R\sin\theta \\ \sin\theta & R\cos\theta \end{bmatrix} \tag{2.41}$$

şeklinde olup $|J| = R$ olarak bulunur. Buna göre R,θ 'nın ortak olasılık yoğunluk fonksiyonu,

$$f(r,\theta) = \frac{1}{2\pi} r e^{-\frac{1}{2}r^2}, \quad 0 < r < \infty, 0 < 2\pi < \theta \tag{2.42}$$

dır. $D = R^2$ dönüşümü yapıldığında, D,θ 'nın ortak olasılık yoğunluk fonksiyonu,

$$f(d,\theta) = \frac{1}{2\pi}\frac{1}{2} e^{-\frac{1}{2}d}, \; 0 < d < \infty, \; 0 < 2\pi < \theta \tag{2.43}$$

olarak bulunur. D ,parametresi 2 olan üstel dağılımlı, $\theta \approx U(0,2\pi)$ olan düzgün dağılımdır ve D ile θ bağımsızdır. D ve θ 'nın dağılımlarından sayı üretilip

$$\begin{aligned} X &= \sqrt{d}\cos\theta = \sqrt{-2\ln(U_1)}\cos(2\pi U_2) \\ Y &= \sqrt{d}\sin\theta = \sqrt{-2\ln(U_2)}\sin(2\pi U_2) \end{aligned} \tag{2.44}$$

dönüşümleri yapıldığında, yani kutupsal koordinatlardan sayı üretip dik koordinat sistemine geçildiğinde; X,Y rasgele değişkenlerinin dağılımlardan sayı üretilmiş olur. Yukarıdaki dönüşüm Box-Muller dönüşümü olarak bilinir. Box-Muller dönüşümlerinde $(\sin\theta), \cos(\theta)$ trigonometrik fonksiyonları kullanıldığından ve bu fonksiyonlar periyodik fonksiyonlar olduklarından bu dönüşümlerle normal dağılımdan sayı üretmek sorun olabilir.

$X \approx N(\mu,\sigma^2)$ dağılımından sayı üretmek için $Z \approx N(0,1)$ dağılımından sayı üretilir ve $Z\sigma + \mu$ dönüşü yapılır [11].

3. MATERYAL VE YÖNTEM

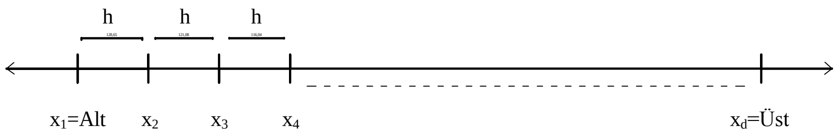
Bir olasılık yoğunluk fonksiyonundan örneklem üretmek için çeşitli yöntemler mevcuttur. Bunlardan başlıcaları önceki bölümde bahsedilen Ters dönüşüm yöntemi, Kabul-red yöntemi ve Ayrışım yöntemleridir. Bu bölüm de olasılık yoğunluk fonksiyonundan örneklem üretmek için yeni bir yöntem verilecektir. $f(x)$ olasılık yoğunluk fonksiyonundan n büyüklüğünde bir örneklem elde etmek için aşağıdaki yöntem kullanılabilir.

İlk olarak örneklemin alt ve üst sınır değerleri belirlenmelidir. Örneklemin alt ve üst sınır değerleri

$$\int_{Alt}^{Üst} f(x)dx = 0.99 \tag{3.1}$$

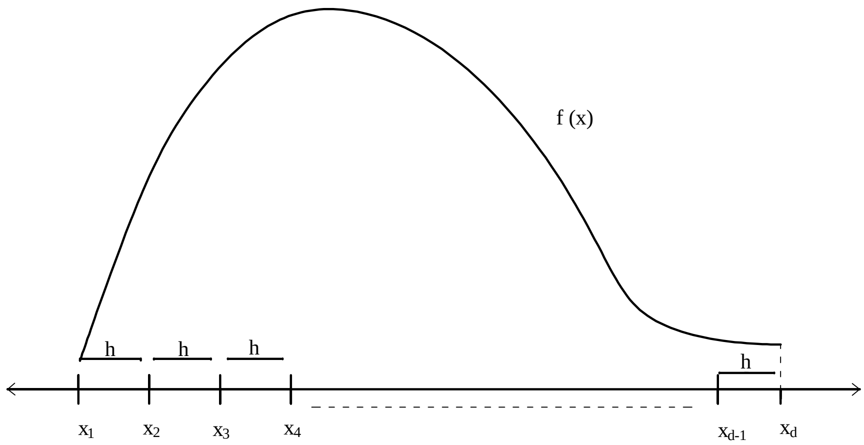
eşitliğini sağlayacak şekilde belirlenir.

$[Alt, Üst]$ kapalı aralığı d-1 tane eşit aralığa bölünürse aşağıdaki şekil elde edilir.



Şekil 3.1. Örneklem aralıkları

$[x_1, x_d]$ kapalı aralığında $f(x)$ olasılık yoğunluk fonksiyonunun temsili grafiği Şekil 3.2.'de verilmiştir.



Şekil 3.2. Olasılık yoğunluk fonksiyonunda aralıklar

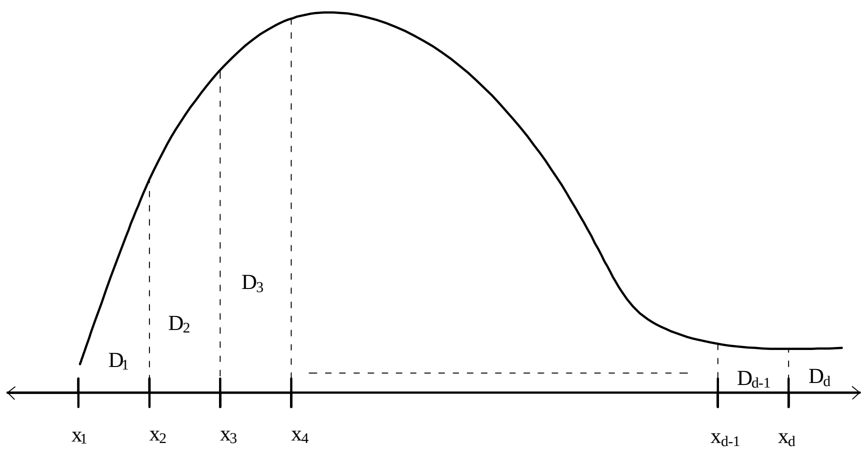
$f(x)$ olasılık yoğunluk fonksiyonunun tanım aralığı $a < x < \infty$ şeklinde olsun. $a = x_1$ ve

$$D_i = \int_{x_i}^{x_{i+1}} f(x)dx \qquad\qquad i = 1,2,......,d-1 \tag{3.2}$$

olmak üzere $i-inci$ aralıkta verilerin $\% \; D_i$ 'si bulunur. (3.1) eşitliğinden

$$\sum_{i=1}^{d-1} D_i = 0.99 \tag{3.3}$$

elde edilir. $[x_1, x_d]$ kapalı aralığında verilen $\% \; 99$ 'u bulunur. Geriye kalan $\% \; 1$ 'lik kısım ise $[x_d, \infty)$ aralığında bulunur.



Şekil 3.3. Aralık yoğunlukları

Bu durumda örneklemin seçileceği aralıklar

$[x_1, x_2) \Rightarrow 1$-inci aralık

$[x_2, x_3) \Rightarrow 2$-inci aralık

$[x_3, x_4) \Rightarrow 3$-üncü aralık

.

.

.

$[x_{d-1}, x_d) \Rightarrow (d-1)$-inci aralık

$[x_d, \infty) \Rightarrow d$-inci aralık

şeklindedir. Bu d tane aralıktan toplamda n büyüklüğünde bir örneklem oluşturulacaktır. Bunun için 1-inci aralıkta $\|n\, D_1\|$ tane tesadüfi sayı seçilir, 2-inci aralıkta $\|n\, D_2\|$ tane tesadüfi sayı seçilir ve böyle devam ederek d-inci aralıkta $n - \left\| n \sum_{i=1}^{d-1} D_i \right\|$ tane tesadüfi sayı seçilir.

Bu seçilen sayıların birleştirilmesiyle örneklem oluşturulur.

4. BULGULAR

Bu bölümde ilk olarak Pareto dağılımında parametre tahmini yapmak için maksimum olabilirlik yöntemi açıklanacaktır. Materyal ve yöntem bölümünde bahsedilen metot ele alınarak örneklem oluşturulacaktır. Daha sonra Pareto dağılımının dağılım fonksiyonundan ve parametrelerinden yararlanarak ters dönüşüm yöntemiyle örneklem elde edilecektir. Son olarak da bu iki örneklemden faydalanılarak maksimum olabilirlik yöntemi ile tahmini parametre değerleri hesaplanacak ve gerçek parametrelerle karşılaştırılıp hangi örneklem yöntemiyle daha iyi değerler elde edileceği karşılaştırılacaktır.

Pareto dağılımı iki parametreli bir dağılımdır ve genel ifadesi;

$$f(x) = \frac{v\lambda^{v}}{x^{v+1}} \tag{4.1}$$

şeklindedir. Bu ifadeyi,

$$f(x) = v\lambda^{v} x^{-(v+1)} \tag{4.2}$$

şeklide yazılabilir. Fonksiyonun logaritması alınırsa;

$$\ln f(x) = \ln v + v \ln \lambda - (v+1)\ln x \tag{4.3}$$

elde edilir. Bu ifadenin benzerlik fonksiyonu,

$$\ell = f(x_1) f(x_2) \ldots\ldots f(x_n) \tag{4.4}$$

olduğundan bunların tek tek logaritmaları için logaritmik benzerlik fonksiyonu,

$$\begin{aligned} L &= \ln \ell = \ln f(x_1) + \ln f(x_2) + \ldots\ldots + \ln f(x_n) \\ L &= \sum_{i=1}^{n} \ln f(x_i) \end{aligned} \tag{4.5}$$

şeklindedir. Burada Pareto fonksiyonu yerine yazılırsa,

$$\begin{aligned} L &= \sum_{i=1}^{n} \left[\ln v + v \ln \lambda + (v+1)\ln x_i\right] \\ &= n \ln v + n v \ln \lambda - (v+1)\sum_{i=1}^{n} \ln(x_i) \end{aligned} \tag{4.6}$$

ifadesi elde edilir. Pareto dağılımının v ve λ parametreleri tahmin edileceğinden, L'nin v ve λ'ya göre birinci türevleri alınacaktır.

$$\frac{dL}{dv} = \frac{n}{v} + n\ln\lambda - \sum_{i=1}^{n}\ln(x_i)$$
$$\frac{dL}{d\lambda} = \frac{nv}{\lambda} \tag{4.7}$$

Bunlar sıfıra eşitlenirse

$$\hat{v} = \frac{n}{\sum_{i=1}^{n}\ln\frac{x_i}{\lambda}} \tag{4.8}$$

ifadeleri elde edilir. Birinci ifade v parametresinin tahmininde kullanılabilir. Fakat $n > 0$ ve $v > 0$ olduğundan ikinci ifadede bir çelişki meydana gelir. λ 'nın tahmin edicisi olarak;

$$\hat{\lambda} = \min(x_1, x_2, \ldots, x_n) \tag{4.9}$$

ifadesi kullanılmaktadır [18].

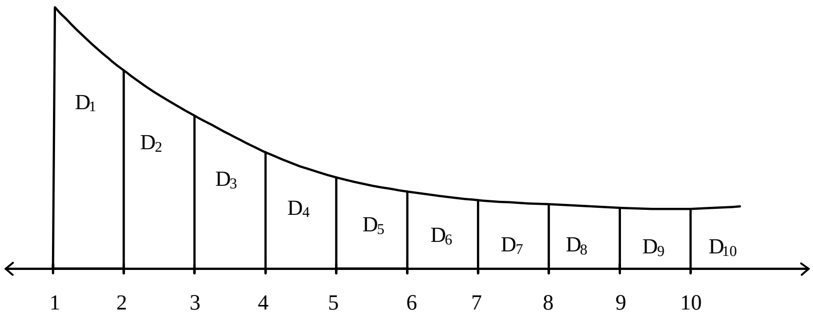
Pareto dağılımında $v = 2$, $\lambda = 1$ alınırsa,

$$f(x) = \frac{2}{x^3} \tag{4.10}$$

olasılık yoğunluk fonksiyonu elde edilir. Bu fonksiyondan faydalanarak $n = 10$ büyüklüğünde bir örneklem oluşturulacaktır. Bunun için $\lambda = 1$ olduğundan alt sınır 1 olarak alınabilir. Burada örneklemin seçileceği aralığın üst sınırının belirlenmesi gerekir. Bunun için aşağıdaki integrali ele alalım.

$$\int_{1}^{Üst} f(x)\,dx = 0{,}99$$
$$\Rightarrow -\frac{1}{(Üst)^2}\Bigg|_{1}^{Üst} = 0{,}99$$
$$\Rightarrow -\frac{1}{(Üst)^2} + 1 = 0{,}99$$
$$\Rightarrow -\frac{1}{(Üst)^2} = -0{,}01$$
$$\Rightarrow -\frac{1}{(Üst)^2} = -\frac{1}{100} \tag{4.11}$$
$$\Rightarrow Üst = 10$$

Alt sınır 1 ve üst sınır 10 olarak elde edildi. Aralık genişliği $h = 1$ olarak alınırsa örneklemin seçileceği aralık 9 eşit parçaya bölünür. Bu aralıkların yoğunlukları $D_i \ (i=1,2,\ldots,9)$ yoğunlukları aşağıda hesaplanmıştır.



Şekil 4.1. $v=2,\ \lambda=1$ ve h = 1 için Pareto dağılımında örneklemlerin seçileceği aralıklar

$1-inci$ aralık için

$$D_1=\int_1^2 f(x)\,dx=-\frac{1}{(x)^2}\Bigg|_1^2=-\frac{1}{(2)^2}+1\Rightarrow\frac{3}{4}\quad\Rightarrow D_1=0{,}75 \tag{4.12}$$

$2-inci$ aralık için

$$D_2=\int_2^3 f(x)\,dx=-\frac{1}{(x)^2}\Bigg|_2^3=-\frac{1}{(3)^2}+\frac{1}{4}\quad\Rightarrow D_2=0{,}14 \tag{4.13}$$

$3-üncü$ aralık için

$$D_3=\int_3^4 f(x)\,dx=-\frac{1}{(x)^2}\Bigg|_3^4=-\frac{1}{16}+\frac{1}{9}\quad\Rightarrow D_3=0{,}049 \tag{4.14}$$

$4-üncü$ aralık için

$$D_4=\int_4^5 f(x)\,dx=-\frac{1}{(x)^2}\Bigg|_4^5=-\frac{1}{25}+\frac{1}{16}\quad\Rightarrow D_4=0{,}0225 \tag{4.15}$$

$5-inci$ aralık için

$$D_5=\int_5^6 f(x)\,dx=-\frac{1}{(x)^2}\bigg|_5^6=-\frac{1}{36}+\frac{1}{25}\qquad\Rightarrow D_5=0{,}0122 \tag{4.16}$$

$6-ıncı$ aralık için

$$D_6=\int_6^7 f(x)\,dx=-\frac{1}{(x)^2}\bigg|_6^7=-\frac{1}{49}+\frac{1}{36}\qquad\Rightarrow D_6=0{,}0074 \tag{4.17}$$

$7-inci$ aralık için

$$D_7=\int_7^8 f(x)\,dx=-\frac{1}{(x)^2}\bigg|_7^8=-\frac{1}{64}+\frac{1}{49}\qquad\Rightarrow D_7=0{,}0048 \tag{4.18}$$

$8-inci$ aralık için

$$D_8=\int_8^9 f(x)\,dx=-\frac{1}{(x)^2}\bigg|_8^9=-\frac{1}{81}+\frac{1}{64}\qquad\Rightarrow D_8=0{,}0033 \tag{4.19}$$

$9-uncu$ aralık için

$$D_9=\int_9^{10} f(x)\,dx=-\frac{1}{(x)^2}\bigg|_9^{10}=-\frac{1}{100}+\frac{1}{81}\qquad\Rightarrow D_9=0{,}0011 \tag{4.20}$$

$10-uncu$ aralık için

$$D_{10}=\int_{10}^{\infty} f(x)\,dx=-\frac{1}{(x)^2}\bigg|_{10}^{\infty}=-\frac{1}{\infty^2}+\frac{1}{100}\qquad\Rightarrow D_{10}=0{,}01 \tag{4.21}$$

Her aralıktan $\|n\,D_i\|$tane tesadüfi örneklem seçilmelidir. Dolayısıyla $1-inci$ aralıktan 7, $2-inci$ aralıktan 1 tane tesadüfi örneklem noktası seçilir. $\|n\,D_j\|<1 \quad (j=3,4,\ldots,9)$olduğundan diğer 7 aralıktan bu bakış açısıyla eleman seçilememektedir. Dolayısıyla $(3,9)$aralığında tesadüfi 2 tane örneklem noktası seçilir.

Aralıklardan seçilen tesadüfi örneklem noktaları sırasıyla $1-inci$ aralıktan $1{,}25;\ 1{,}64;\ 1{,}83;\ 1{,}09;\ 1{,}03;\ 1{,}55;\ 1{,}23$ olmak üzere 7 tane, $2-inci$ aralıktan 2,47 olup 1 tane, geriye kalan kısımdan ise tesadüfi olarak $3{,}34$ ve $4{,}56$ olmak üzere 2 tane örneklem seçilir.

Pareto dağılımının dağılım fonksiyonu ve ters dönüşüm fonksiyonu aşağıdaki gibidir.

$$F(x) = 1 - \left(\frac{\lambda}{x}\right)^{v}$$

$$F^{-1}(u) = \frac{\lambda}{\sqrt[v]{1-u}} \tag{4.22}$$

Burada son ifade de Pareto dağılımının parametreleri $v = 2, \lambda = 1$ alınırsa,

$$F^{-1}(u) = \frac{1}{\sqrt[2]{1-u}} \tag{4.23}$$

ters dağılım fonksiyonu elde edilir. (0,1) aralığında tesadüfi olarak seçilen $0{,}23; 0{,}8; 0{,}2; 0{,}85; 0{,}57; 0{,}33; 0{,}47; 0{,}63; 0{,}92; 0{,}77$ noktaları fonksiyonda yerine yazılırsa aşağıdaki gibi bir örneklem oluşturulacaktır.

$$F^{-1}(u) = \frac{1}{\sqrt[2]{1-u}} = F^{-1}(0{,}23) = \frac{1}{\sqrt[2]{(1-0{,}23)}} = 1{,}1396$$

$$F^{-1}(u) = \frac{1}{\sqrt[2]{1-u}} = F^{-1}(0{,}8) = \frac{1}{\sqrt[2]{(1-0{,}8)}} = 2{,}2360$$

$$F^{-1}(u) = \frac{1}{\sqrt[2]{1-u}} = F^{-1}(0{,}2) = \frac{1}{\sqrt[2]{(1-0{,}2)}} = 1{,}1180$$

$$F^{-1}(u) = \frac{1}{\sqrt[2]{1-u}} = F^{-1}(0{,}85) = \frac{1}{\sqrt[2]{(1-0{,}85)}} = 2{,}5819$$

$$F^{-1}(u) = \frac{1}{\sqrt[2]{1-u}} = F^{-1}(0{,}57) = \frac{1}{\sqrt[2]{(1-0{,}57)}} = 1{,}5249$$

$$F^{-1}(u) = \frac{1}{\sqrt[2]{1-u}} = F^{-1}(0{,}33) = \frac{1}{\sqrt[2]{(1-0{,}33)}} = 1{,}2216$$

$$F^{-1}(u) = \frac{1}{\sqrt[2]{1-u}} = F^{-1}(0,47) = \frac{1}{\sqrt[2]{(1-0,47)}} = 1,3736$$

$$F^{-1}(u) = \frac{1}{\sqrt[2]{1-u}} = F^{-1}(0,63) = \frac{1}{\sqrt[2]{(1-0,63)}} = 1,6439$$

$$F^{-1}(u) = \frac{1}{\sqrt[2]{1-u}} = F^{-1}(0,92) = \frac{1}{\sqrt[2]{(1-0,92)}} = 3,5355$$

$$F^{-1}(u) = \frac{1}{\sqrt[2]{1-u}} = F^{-1}(0,77) = \frac{1}{\sqrt[2]{(1-0,77)}} = 2,0851$$

(4.24)

Materyal ve yöntem bölümünde bahsedilen metotla elde edilen $1,25; 1,64; 1,83; 1,09; 1,03; 1,55; 1,23; 247; 3,34; 4,56$ örneklem noktalarına maksimum olabilirlik yöntemi kullanılarak parametre tahmini yapılırsa $\hat{\lambda} = 1,03$ ve $\hat{v} = 1.846693$ değeri bulunur. Ters dönüşüm yöntemiyle elde edilen $1,1396; 2,2360; 1,1180; 2,5819; 1,5249; 1,2216;$ $1,3736; 1,6439; 3,5355; 2,0851$ örneklem noktalarına maksimum olabilirlik yöntemi kullanılarak parametre tahmini yapılırsa $\hat{\lambda} = 1,1180$ ve $\hat{v} = 2,346298$ değerleri bulunur. Bu tahmini değerler örneklem elde etmek için kullanılan $\lambda = 1,\ v = 2$ parametreleri ile karşılaştırılırsa yeni yöntemle elde edilen tahminlerin gerçek değerlere yakın olduğu görülebilir.

5. SONUÇLAR VE ÖNERİLER

Bu bölümde ilk olarak ters dönüşüm yöntemi ile materyal ve yöntem bölümünde bahsedilen metot aynı anda ele alınarak örneklem üretme yöntemi geliştirildi. Bu yeni yöntem eşit aralıklara bölünmüş ters dönüşüm yöntemi olarak adlandırıldı. Daha sonra ters dönüşüm ve eşit aralıklara bölünmüş ters dönüşüm yöntemleriyle örneklemler üretilip; bu örneklem noktaları kullanılarak maksimum olabilirlik yöntemi ile parametre tahminleri yapıldı. Son olarak da tahmini parametre değerleri gerçek parametrelerle karşılaştırılıp hangi örneklem yöntemiyle daha iyi değerler elde edileceği karşılaştırıldı.

Ters dönüşüm yönteminde ilk olarak $(0,1)$ aralığında tesadüfü noktalar seçilir. Bu noktalar ters dağılım fonksiyonunda yazılarak örneklem elde edilir. Eşit aralıklara bölünmüş ters dönüşüm yönteminde ise $(0,1)$ aralığı h uzunluğuna sahip eşit d tane aralığa bölünür ve her bir aralıktan tesadüfü olarak $\|n\,h\|$ tane nokta seçilir. Bu noktalar ters dağılım fonksiyonunda yazılarak örneklem elde edilir. Şimdi bu iki örneklem üretme metodunu karşılaştırmak için; R-programı kullanılarak elde edilen Çizelge 5.1. ve Çizelge 5.2.’ deki değerler incelenecektir. Çizelge 5.1. ve Çizelge 5.2.’ deki değerleri elde etmek için kullanılan kodlar ekler bölümünde verilmiştir (Ek 1-3).

Çizelge 5.1.’ de ters dönüşüm ve eşit aralıklara bölünmüş ters dönüşüm yöntemleriyle örneklemler üretilip; bu örneklem noktaları kullanılarak maksimum olabilirlik yöntemi ile parametre tahminleri yapılmıştır ve eşit aralıklara bölünmüş ters dönüşüm yöntemleriyle üretilen örneklemlerle yapılan parametre tahminlerinin örneklem üretmek için kullanılan parametrelere daha yakın değerler verdiği gözlenmektedir.

Çizelge 5.1. Maksimum olabilirlik yöntemi ile parametre tahminine örneklem üretme yönteminin etkisi

Örneklem üretmek için kullanılan parametreler		Örneklem Büyüklüğü	Örneklem üretmek için kullanılan yöntem			
			Ters dönüşüm yöntemi		Eşit aralıklara bölünmüş ters dönüşüm yöntemi(h = 0,1)	
λ	v	n	$\hat{\lambda}$	$\hat{v}$	$\hat{\lambda}$	$\hat{v}$
1	2	10	1,020823	2,931854	1,032145	2,205037
		100	1,008811	1,640817	1,006355	2,026853
		1000	1,000201	2,059077	1,000673	2,029222
		10000	1,000015	2,011633	1,000066	2,001859
		100000	1,000000	1,998244	1,000001	1,999799
	5	10	1,008685	7,989972	1,003217	5,408494
		100	1,001036	5,952521	1,004010	5,167252
		1000	1,000190	5,243540	1,000089	4,967841
		10000	1,000028	4,929686	1,000032	5,024877
		100000	1,000010	4,989179	1,000006	4,999461
2	2	10	2,073863	1,207943	2,070858	2,425619
		100	2,007351	1,877883	2,011478	2,071316
		1000	2,000371	1,968991	2,000742	2,022839
		10000	2,000096	1,980925	2,000155	2,007881
		100000	2,000020	1,993970	2,000004	1,999913
	5	10	2,018641	6,135338	2,010037	5,476851
		100	2,002471	5,351646	2,002954	5,207154
		1000	2,000388	4,774587	2,000295	4,967445
		10000	2,000004	4,989885	2,000002	4,997070
		100000	2,000003	5,012237	2,000008	5,000337
3	2	10	3,048477	1,632519	3,086142	2,265919
		100	3,013201	2,357715	3,012928	1,934013
		1000	3,000990	1,988550	3,000074	1,990013
		10000	3,000351	1,955700	3,000568	2,001215
		100000	3,000021	1,994644	3,000009	1,999402
	5	10	3,057857	8,823205	3,011835	4,875996
		100	3,004627	6,283527	3,012197	5,159949
		1000	3,000342	4,891518	3,000518	4,989606
		10000	3,000047	5,045223	3,000011	5,002419
		100000	3,000002	5,013560	3,000001	4,999161

Çizelge 5.2.' de eşit aralıklara bölünmüş ters dönüşüm yönteminde aralık genişliği farklı değerler alınıp örneklemler üretilmiştir; bu örneklem noktaları kullanılarak maksimum olabilirlik yöntemi ile parametre tahminleri yapılmıştır ve aralık genişliği azaldıkça üretilen örneklemlerle yapılan parametre tahminlerinin, örneklem üretmek için kullanılan parametrelere daha yakın değerler verdiği gözlenmektedir.

Çizelge 5.2. Maksimum olabilirlik yöntemi ile parametre tahminine aralık boyunun etkisi

Örneklem üretmek için kullanılan parametreler		Örneklem Büyüklüğü	Örneklem üretmek için kullanılan yöntem: Eşit aralıklara bölünmüş ters dönüşüm yöntemi					
λ	v	n	h = 0,1		h = 0,01		h = 0,001	
			$\hat{\lambda}$	$\hat{v}$	$\hat{\lambda}$	$\hat{v}$	$\hat{\lambda}$	$\hat{v}$
1	2	1000	1,000313	2,038003	1,000656	2,005471	1,000047	2,000790
		10000	1,000028	2,006018	1,000031	2,000131	1,000057	1,999929
		100000	1,000003	1,997159	1,000007	2,000532	1,000000	1,999963
	5	1000	1,000105	4,935495	1,000051	5,009958	1,000055	4,999189
		10000	1,000081	5,018903	1,000007	4,998758	1,000017	4,999685
		100000	1,000002	5,005502	1,000004	4,999464	1,000002	5,000110
2	2	1000	2,001018	1,986934	2,000963	1,998916	2,000825	2,001696
		10000	2,000161	2,005036	2,000016	2,001525	2,000131	2,000407
		100000	2,000007	2,002185	2,000001	1,999648	2,000007	1,999943
	5	1000	2,000254	4,969656	2,001002	5,021775	2,000385	4,995362
		10000	2,000042	5,016794	2,000024	4,999341	2,000012	5,000384
		100000	2,000001	5,002851	2,000001	5,000282	2,000005	5,000233
3	2	1000	3,000210	2,013149	3,001072	1,994974	3,000449	2,001767
		10000	3,000100	1,995142	3,000585	2,001222	3,000188	1,999559
		100000	3,000003	2,002666	3,000011	1,999513	3,000006	1,999931
	5	1000	3,000452	4,975532	3,000039	5,010835	3,000578	5,003822
		10000	3,000117	5,006558	3,000008	4,999479	3,000069	5,000498
		100000	3,000004	4,998889	3,000020	5,000441	3,000004	4,999950

KAYNAKLAR

[1] Turanlı M, Güriş S, 2015. Temel İstatistik. Der Yayınları, İstanbul.

[2] Aytaç M, 2004. Matematiksel İstatistik. Ezgi Kitabevi, Bursa.

[3] Akdeniz F, 2015. Olasılık ve İstatistik. Akademisyen Kitabevi, Ankara.

[4] Çıngı H, 1990. Örnekleme Kuramı. Hacettepe Üniversitesi Fen Fakültesi Basımevi, Ankara.

[5] Alzaatreh A, Famoye F, Lee C, 2012. Gamma-Pareto Distribution and Its Applications, 11 (1): 78-94.

[6] Balcı M, 2008. Matematik Analiz 1. Balcı Yayınları, Ankara.

[7] Güriş S, Çağlayan E, 2013. Ekonometri Temel Kavramlar. Der Kitabevi, İstanbul.

[8] Özdemir AF, Yıldıztepe E, Binar M, 2010. Akademik Bilişim 2010, 10-12 Şubat 2010, Muğla Üniversitesi.

[9] Baydoğan MG, Orbay B, Çetin U, 2014. XIX. Türkiye’de İnternet Konferansı 2014, 27 Kasım 2014,Yaşar Üniversitesi, İzmir.

[10] Esin A, 1975. İstatistik ve Tatbiki Matematik. Ankara İktisadi ve Ticari İlimler Akademisi Basımevi, Ankara.

[11] Öztürk F, Özbek L, 2004. Matematiksel Modelleme ve Simülasyon. Gazi Kitabevi, Ankara.

[12] Bratley P, Fox BL, Schrage LE, 1987. A Guide to Simulation. Springer-Verlang, New York.

[13] Abramowitz M, Stegun IA, 1964. Handbook of Mathematical Function with Formulas, Graphs and Mthematical Tables. Tenth Printing, Washington, D.C.

[14] Kennedy WJ, Gentle JE, 1980. Statistical Computing. Marcel Decker. New York.

[15] Tadimakalla PR, 1980. A Look at the Burr and Related Distributions, International Statistical Review 48:337-344.

[16] Berman M, 1970. Gaussion Processes with Stationary Increments: Local Times and Sample Function Properties, The Annals of Mathematical Statistics 1970, 41(4):1261-1272.

[17] Tadimakalla PR, 1980. Simple Rejection Methods for Sampling from the Normal Distribution, Proceeding of the 10th Annual Conference of AIDS.

[18] Prakash G, 2009. Some Estimators for the Pareto Distribuiton, Journal of Scientific Research,1(2):236-247.

EKLER

Ek -1

Pareto dağılımından ters dönüşüm yöntemiyle veri üretmek için R programlama kodları:

```
lmd <- 3                                   # lamda skaler parametre

v <- 5                                     # şekil parametresi

n <-100000                                 # örneklem büyüklüğü

tsd =r unif(n, 0, 1)                       #(0,1) aralığında n tane tesadüfi sayı üretir

tersdonusum=( (lmd) / ( (1-tsd)^(1/v) ) )  # v ve lamda parametreli Pareto dağılımı için
                                           n tane tesadüfi sayı üretir
```

Ek -2

Pareto dağılımından eşit aralıklara bölünmüş ters dönüşüm yöntemiyle veri üretmek için R programlama kodları:

```
h <- 0.001                                  # aralığın boyu

lmd <- 3                                    # lamda skaler parametre

v <- 5                                      # şekil parametresi

n<-100000                                   # örneklem büyüklüğü

I=0                                         # örneklemin alt sınır

S=1                                         # örneklemin üst sınırı

A=as.integer((S-I)/h)                       # aralık sayısı

U=A+1                                       # aralık adeti

Sha=(A*h)+I                                 # üst sınırdan önceki değer

sındeg=seq(from = I, to =Sha, length.out =U)  # aralıkların sınır değerleri

tn=(as.integer(n*h))                        # aralıktan seçilen sayı adeti

orn <- double()

for(i in 1:A)
{
atn=tn
k= sındeg[i]
m= sındeg[i]+h
if (m > 1) {break}
orn <- c(orn, runif(atn, k, m))
}
```

```
ke= length(orn)                                   # toplamda kaçta değer atanmış
kalan = n-k                                       # n örneklemi tamamlamak için kaçta tane seçim yapılmalı
kalana=m                                          # kalanlar için alt sınır
kalanu=1                                          # kalanlar için üst sınır
orn <- c(orn, runif(kalan, kalana, kalanu))
esitaralıklıtersdonusum=((lmd)/((1-orn)^(1/v)))
```

Ek -3

Pareto dağılımından üretilen örneklem için; maksimum olabilirlik yöntemi ile parametre tahmini için R programlama kodları:

```
pareto.MLE <-  function(X)
{
n  <-  length (X)
m <- min (X)
a <- n / sum ( log(X)-log(m) )
return( c(m,a) )
}

pareto.MLE(esitaralıklıtersdonusum)
```

ÖZGEÇMİŞ

1989 yılında Elazığ'da doğdu. İlköğretimi Elazığ Dumlupınar İlköğretim Okulu'nda ve liseyi ise Mehmet Akif Ersoy Lisesi'nde tamamladı. 2010 yılında kazandığı Fırat Üniversitesi Fen Fakültesi İstatistik Bölümü'nden 2014 yılında mezun oldu. Daha sonra 2016 yılının şubat ayında Bitlis Eren Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı'nda Uygulamalı İstatistik programında yüksek lisansa başladı ve lisansüstü eğitimine devam etmektedir.

Seval ŞAHİN