



MONTE CARLO SANAL DENEY TEKNİĞİNE GİRİŞ

Ders Notu

Mustafa Y. ATA

1. Giriş

Monte Carlo Tahmin Yöntemi

Kumarhaneleri ile ünlü Monoca Prenslığı'nın bir kentinin adı olan "Monte Carlo"(MC), ilk kez 1940'larda A.B.D.'nin Los Alamos eyaletinde nükleer silah dizgeleri üzerine çalışan John von Neumann(1903-1957), Ulam ve Metropolis gibi bilimciler tarafından, bir grup matematiksel yöntemleri nitelendirmek için kullanıldı. Bu yöntemleri aynı amaca yönelik diğer sayısal yöntemlerden ayıran temel özellikleri, söz konusu problemi *sanal bir rulet tekerleği* ile oynanan bir tür şans oyununa dönüştürerek, *problemin yaklaşık çözümlerini* üretmeleridir. Bu yaklaşık çözüme, sanal rulet tekerleğinin ürettiği *rassal sayılardan* problemin *çok büyük sayılarda* özel çözümleri elde edilerek, *istatistik biliminin* temel yasası olan [*Büyük Sayılar Yasası\(BSY\)*](#)'sının mantığı ile ulaşılır.

MC yöntemler, matematiksel olarak modellenen ve doğasında olasılık bulunan gerçek dizgelere ilişkin olaylara olduğu kadar, olasılık içermeyen çok boyutlu belirli tümlevler gibi tümü ile soyut salt matematiksel problemlere de uygulanır. Olasılık içersin veya içermesin, yaklaşık çözümü aranan bu sorunların tümü, "*bir ölçümöte tahmin sorunu*" niteliğindedir.

"MC Tahmin" adıyla bilinen bu yöntemlerin kapsamında ele alınabilen tümü ile soyut salt matematiksel problemlerin sayısal yaklaşık değerlerini bulmak için *yeterli sayıda özel çözüm üretmede* genelde bir sıkıntı yoktur. Doğasında olasılık bulunan gerçek dizgelerin, söz konusu sorunun *yaklaşık çözümü için gerekli sayıda* özel çözümlerini elde etmek ise, söz konusu rassal olayın gerçek deneme maliyetinin zaman ve kaynak açısından çok

düşük olduğu şans oyunları gibi durumlar dışında pek olası değildir.

Gerçek ve Sanal Gerçek

Ancak, bir dizgenin gerçeği yerine, onun *benzeri* olan matematiksel *modeli* üzerinden BSY mantığının işlerlik kazanması için gerekli olan yeterli sayıda söz konusu özel çözümler üretilebilir. Bir dizgenin matematiksel modeli, söz konusu modelin içerdiği tüm varsayımların geçerli olduğu, bir *sanal gerçektir*. Bir dizgenin gerçeğini gözlemek yerine, onun modeline göre sanal gerçeğini üretmeye ise *benzetim* denir. Gerçek bir deneydeki *gözlem değerleri*, somut bir *yığından rasgele örnekleme ile* elde edilen birimlerin söz konusu özelliklerinin *ölçülen* değerleri, sanal bir deneydeki *sanal gözlem değerleri* ise, sanal gerçeğin modelindeki girdi niteliğindeki değişkenlerin [*olasılık dağılımlarından rasgele örnekleme ile hesaplanan*](#) değerlerdir. Bir dizgeye ilişkin sanal gerçeğe göre tanımlanmış sorunların yaklaşık çözümlerini MC yöntemle elde etmede de, tümü ile soyut salt matematiksel sorunların sayısal yaklaşık değerlerini bulmak için *yeterli sayıda özel çözüm üretmede* olduğu gibi bir sıkıntı yoktur. MC yöntemlerin salt matematiksel sorunların sayısal yaklaşık değerlerini bulma amacı ile kullanımı “*MC Hesaplama*”; gerçek dizgeler yerine sanal gerçeklerinin benzetimle üretilerek, dizgeyi tanımlayan ölçümöteletlerin tahmin sorunlarının çözümünde kullanımı ise “*MC Benzetim*” olarak bilinir.

Tümü ile gerçek dizgeler, ancak söz konusu dizge gerçek ortamda gözlenerek modellenenebilir. Modellenmiş dizgelerin alt-dizgelerini oluşturduğu üst-düzey sanal dizgeler tasarlanabilir. Tasarım aşamasındaki bu sanal dizgelerin çıktısına ilişkin öteölçüm tahminlerinin elde edilmesinin söz konusu olduğu her türlü problemin yaklaşık çözümleri, söz konusu *dizgenin canlandırıldığı sanal bir deneyin benzetim verilerinden MC yöntemlerle* elde edilir. Tasarlanan bir dizgenin sanal ortamdan gerçek ortama aktarılması, MC benzetim geçidinden başarı ile geçebilmesine bağlıdır. Sanal atom bombası MC sınamalardan başarı ile geçtikten sonra, 1944’de Los Alamos çöllerinde ilk atom bombası denendi. Savaş sonrasında ise, tüm *sanal nükleer*

savaş benzetimleri böyle bir savaşın galibi olamayacağını gösterdiği için, soğuk savaş döneminde nükleer silahlar, bu teknolojiye sahip olmayanlara “aba altından gösterilen” kağıttan sopalara dönüştü.

Artık günümüzde, kuramsal tasarımı yapılan her dizgenin gerçekten denenmeden önce sıkı bir sanal deney aşamasından başarı ile geçmesi ön koşuldur. Sanal bir deneyle, tasarım aşamasındaki bir dizge canlandırılarak, istenen her koşulda dizgenin işleyişi incelenebilir. MC benzetim sınamalarından başarı ile geçtikten sonra, bir dizgenin gerçek ortamda gözlenecek davranışı ile sanal deney aşamasındaki davranışı arasındaki fark, dizgenin modeli ile dizgenin kendisi arasındaki fark kadardır. Başka bir deyişle, sanal deneyin temelini oluşturan bir dizgeye ilişkin modelin hatası ne kadar az ise, söz konusu dizgeye ilişkin sanal deneyle elde edilen bilgi ile o dizgeye ilişkin olarak gerçekleştirilecek bir deneyden elde edilecek bilgi ile o kadar çok örtüşecektir. Bu açıklamaların ışığında, MC yöntemlerin, nicel temeli olan her türlü kuramsal bilginin sanal verilerle sınanmasında güvenle kullanılabileceği açıktır.

2. Monte Carlo Tahmin

2.1 Monte Carlo Örnek

Yeterince büyük bir n sayıdaki, bir birinden *bağımsız ve aynı* *bağımlı*(b.a.d.) *rassal değişkenler* dizisinin oluşturduğu bir

$$X = \{X_1, X_2, \dots, X_n\} \quad (1)$$

sürecine ilişkin,

$$\{x_1, x_2, \dots, x_n\} \quad (2)$$

gözlem değerleri elde edilebiliyorsa, X 'in *beklenen değeri*

$$\langle X \rangle \equiv \mu_X \equiv \mu \quad (3)$$

için, (2)'deki n çaplı *MC örneğin ortalaması*

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i, \quad (4)$$

tutarlı ve sapmasız bir tahmin edicidir.

2.2 Monte Carlo Yakınsama

MC örnek çapı belli bir $N \ll \infty$ değerine kadar adım adım arttırılarak, belli bir $h > 0$ *tahmin hatası* ile, *Büyük Sayılar Yasası*(BSY)¹ gereği, MC örnek ortalamasının

$$\text{er}_{n \rightarrow N} \bar{x}_n \rightarrow \hat{\mu}_{MC} \quad (5)$$

1

Zayıf BSY : $\text{er}_{n \rightarrow \infty} \text{Ol} \{|\bar{x}_n - \mu| < h\} \rightarrow 1$

Güçlü BSY : $\text{Ol} \left(\text{er}_{n \rightarrow \infty} \bar{x}_n \rightarrow \mu \right) = 1.$

biçiminde yakınsayacağı değere, X 'in beklenen değerinin *MC tahmini* denir.

2.3 Monte Carlo Güven Aralığı

Temel Erim Teoremi(TET)'ne göre, n çaplı Birimleştirilmiş örnek ortalamalarının dağılımı, $n \rightarrow \infty$ için, *Birim Normal Dağılımı* yakınsar:

$$\sqrt{n}(\bar{X}_n - \mu)\sigma^{-1} \sim N(0,1); \quad (6)$$

$$\begin{aligned} \text{er}_{n \rightarrow \infty} \sqrt{n}(\bar{X}_n - \mu)\sigma^{-1} &\rightarrow \Phi(z < z_{\alpha/2}) = \\ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_{\alpha/2}} e^{-\frac{1}{2}z^2} dz &= 1 - \alpha/2. \end{aligned} \quad (7)$$

Buna göre, X sürecinin *değişkesi*,

$$\langle (X - \mu)^2 \rangle \equiv \sigma_X^2 \equiv \sigma^2 \quad (8)$$

olmak üzere, süreç beklenen değeri μ için $\%100(1 - \alpha)$ 'lık

$$Ol \left(\bar{x}_n - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x}_n + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha \quad (9)$$

güven düzeyindeki güven aralığı,

$$\left(\bar{x}_n - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x}_n + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) \quad (10)$$

biçiminde oluşturulabilir ve n çaplı örnek ortalamalarına göre oluşturulacak (10)'daki *ardışık* aralıkların sürecin gerçek beklenen değerini içermesi olasılığı, (9)'a göre $1 - \alpha$ olarak kabul edilir.

2.4 Monte Carlo Tahminlerin Duyarlılığı

Ardışık güven aralıkları genişliği daraldıkça tahminlerin duyarlılığı artar. MC tahminlerin duyarlılığı, (5)'teki MC örnek ortalaması ile sürecin beklenen değeri arasındaki mutlak fark biçiminde tanımlanan MC tahmin hatasının üst sınırı h ile belirlenebilir. h

değerinin ise, (5), (9) ve (10)'a göre, (10)'daki güven aralığının yarı genişliği kadar alınabileceği açıktır:

$$h = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} . \quad (11)$$

Eğer sürecin değişkesi biliniyorsa, süreç beklenen değerinin MC tahminini belli bir h^* duyarlılıkla ve belli bir α anlamlılık düzeyinde elde etmek için gerekli olan MC örnek çapı N , (11)'deki ilişkidten belirlenebilir:

$$N = (z_{\alpha/2} \sigma / h^*)^2 . \quad (12)$$

2.5 Monte Carlo Yakınsama Kıstası

Ancak, süreç beklenen değerini tahmin amacıyla gerçekleştirilecek MC bir sanal deneyde, süreç değişkesinin baştan biliniyor olması pek olası bir durum değildir. Bu nedenle, (8)'deki süreç değişkesinin de, sapmasız ve tutarlı olduğu bilinen

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \quad (13)$$

tahmin edicisi ile elde edilecek MC örnek değişkesinin tahmin değerleri, (10)'da süreç değişkesinin yerine konarak,

$$\left(\bar{x}_n - z_{\alpha/2} \frac{\hat{\sigma}_n}{\sqrt{n}}, \bar{x}_n + z_{\alpha/2} \frac{\hat{\sigma}_n}{\sqrt{n}} \right) \quad (14)$$

rassal genişlikteki ardışık güven aralıkları oluşturulabilir.²

Bir örnek istatistiği olan $\hat{\sigma}_n^2$ de, Güçlü BSY gereği³, örnek çapı yeterince büyük bir $N \ll \infty$ değerine eriştiğinde süreç değişkesini MC tahminini yakınsar:

$$\text{er}_{n \rightarrow N} \hat{\sigma}_n^2 \rightarrow \hat{\sigma}_{MC} . \quad (15)$$

² $\sqrt{n}(\bar{X}_n - \mu)\hat{\sigma}_n^{-1}$ istatistiği, serbestlik derecesi $n-1$ olan bir Student dağılımlıdır. Ancak MC sanal deneylerde örnek çapı çok büyük değerlere ulaştığından, standart normal dağılımlı olduğunu varsaymak uygulamada bir sakınca doğurmaz.

³ Zayıf BSY: $\text{er}_{n \rightarrow \infty} O\ell \{|\hat{\sigma}_n^2 - \sigma^2| < h\} \rightarrow 1$;

Güçlü BSY: $O\ell \left(\text{er}_{n \rightarrow \infty} \hat{\sigma}_n^2 \rightarrow \sigma^2 \right) = 1$.

Dolayısı ile, MC örnek çapı n arttıkça, MC örnek değişkesi süreç değişkesini yakınsayacak ve (14)'deki rassal olarak değişen ardışık güven aralıkları genişliğindeki kararlı daralmanın hızı yavaşlayarak, MC örnek çapı,

$$N = (z_{\alpha/2} \hat{\sigma}_N / h^*)^2 \quad (16)$$

gibi bir değere eriştiğinde, belirlenen bir h^* duyarlılıkla ve belli bir α anlamlılık düzeyinde süreç beklenen değerinin MC tahminine ulaşılmış olacaktır.

MC örnek ortalamasının süreç beklenen değerini, belli bir duyarlılık ve anlamlılıkta yakınsaması için gerekli örnek çapı değerinin (12) ve (16)'da olduğu gibi belirleyen kurallara *MC yakınsama kıstasları* denir. (12)'deki kıstasa göre MC örnek çapı *deterministik*; (16)'daki kıstasa göre ise *stokastik* bir nitelik taşır. Deterministik MC yakınsamada MC örnek çapı sanal deney başlamadan bilinir. Stokastik MC yakınsamada ise, MC örnek çapının değeri, sanal deney sonunda anlaşılır ve deneyin her yinelenmesinde rasgele bir değer alır.

Stokastik MC yakınsama kıstası, MC örnek çapı adım adım arttırılırken, her adımda MC örnek değişkesinin hesaplanmasını gerektirir. Bundan dolayı, MC örnek değişkesinin doğrudan (13)'deki tanıma göre hesaplanması yerine,

$$\begin{aligned} \hat{\sigma}_n^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \\ &\cong \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}_n^2. \end{aligned} \quad (17)$$

algoritmasının kullanılması daha uygundur.

Şekil 1'de $O\ell(Y=1) \equiv O\ell(X < 5)$ ölçümötesinin tahmini amacıyla

$$Z \sim T[1,6], P \sim T[0,1], X = Z^P$$

$$Y = \begin{cases} 1, & X < 5 \\ 0, & X \geq 5 \end{cases}$$

ve MC örnek çapı $n=1(1)500$ olmak üzere gerçekleştirilen bir deney sonucu elde edilen gözlem değerlerinden her adımda ardışık olarak hesaplanan,

MC örnek ortalaması:

$$\bar{y}_n = \frac{1}{n} \sum_{i=1}^n y_i$$

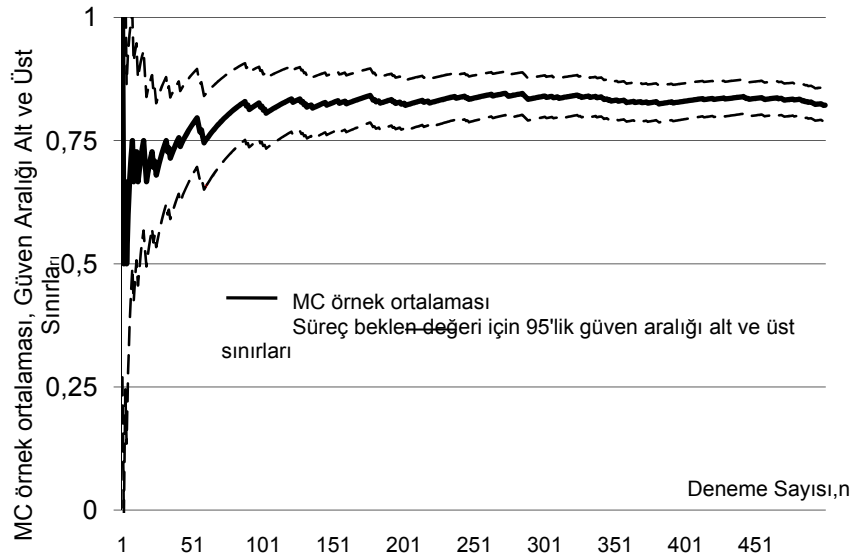
MC örnek değişkesi:

$$\frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}_n^2$$

%95'lik güven aralığı sınırları:

$$\left(\bar{y}_n - z_{\alpha/2} \frac{\hat{\sigma}_n}{\sqrt{n}}, \bar{y}_n + z_{\alpha/2} \frac{\hat{\sigma}_n}{\sqrt{n}} \right)$$

değerlerine göre, rassal güven aralığı genişliğindeki kararlı daralma hızının giderek yavaşladığı MC yakınsama görülmektedir.



Şekil: 1 %95'lik güven aralığında MC yakınsama.

MC tahmin konusunda buraya kadar verilen bilgiler gerçek deneyler için de geçerlidir; yeter ki, istenen sayıda deneysel gözlem elde etme konusunda bir sıkıntı bulunmasın. Şans oyunları dışında, yeterli sayıda gözlem değerine dayalı olarak MC yöntemle ölçümöte tahminine başvurulabilecek deney gerçekleştirmek pek olası değildir. Bundan dolayı

genellikle MC tahmin yöntemi, gözlem üretme maliyetinin yalnızca

- sanal gözlem üreteç tasarımı,
- bilgisayar programlaması ve
- sanal deneyin gerçekleştirilmesi için gerekli bilgisayar kullanım süresine

bağlı olan yaygın olarak *MC benzetim* olarak bilinen *sanal deneylerde* uygulanır.

3. Olasılık Dağılımlarından Rasgele Örnekleme

Bir rassal değişgene ilişkin gerçek gözlemlerden, o rassal değişgenin *deneysel dağılımı* elde edilir ve bu deneysel dağılım, söz konusu rassal değişgenin kuramsal dağılımlardan biri ile temsil edilip edilemeyeceği kararında kanıt olarak kullanılır. Böyle bir istatistiksel karar sonucunda, rassal değişgen için geçerli bir model elde edilmiş olur.

Her hangi bir rassal değişgene ilişkin gerçek gözlem değerleri, bu gözlemlerin ölçüleceği birimler, çerçevesi ve kapsamı çok iyi belirlenmiş bir yığından seçildikleri için, henüz açık biçimi bilinmeyen ama $F(x)$ gibi belli bir dağılımdan rasgele seçilmiş değer niteliğindedirler. *Rasgele örnekleme* adını verdiğimiz bu işlem, X rassal değişgeninin tanım aralığı (a,b) , olasılık yoğunluk fonksiyonu,

$$\int_a^b f(x)dx = 1$$

koşulunu sağlayan $f(x)$ gibi bir fonksiyon ve bu aralığın her hangi bir alt aralığı (c,d) olmak üzere,

$$\{x_i: Ol[x_i \in (c,d)] = \int_c^d f(x)dx, a < x < c\}$$

kümesinin elemanlarını üreten bir yöntemdir.

Deneysel ya da kuramsal dağılımı bilinen bir rassal değişgenle ilgili her türlü bilgi, hatta değerleri o rassal değişgenin dağılımından elde edilebilir. Bir rassal değişgenin dağılımından rasgele örneklem ile elde edilen bu gözlem değerlerine, *sanal gözlem değerleri* diyoruz. Dağılımı belli olan bir rassal

değişgenin gerçek gözlem değerleri ile rassal değişgenin dağılımından rasgele örnekleme yöntemi ile üretilecek sanal gözlem değerleri arasında *istatistiksel* bir fark yoktur. Bu nedenle dağılımı bilinen bir rassal değişgenin, böyle bir bilgiye ulaşılrken kullanılan gözlem değerlerini elde edildikleri biçimde saklamanın bir gereği kalmaz.

Her hangi bir dağılımdan rasgele örnekleme yapılabilmesi, Teorem:1’de ifade edilecek olan, “dağılım fonksiyonunun dağılımı birim tek-düze dağılımdır ve her hangi rassal bir değişgenin değeri, birim tek-düze rassal değişgenler cinsinden ifade edilebilir” gerçeğinden hareket eder. Dolayısı ile, her hangi bir dizgenin temel girdisi, birim tek-düze rassal değişgenlerdir.

Gerçek deneylerde X gibi belli bir değişgene ilişkin gözlem değerleri, tüm gözlem birimlerinin oluşturduğu yığından rasgele seçilen bir gözlem birimi üzerinde ölçülerek elde edilir ve bu değerlerden X ’in olasılık dağılımını belirleyen ölçümötelerin tahmini amaçlanır. Sanal deneylerde ise, olasılık dağılımı bilinen X gibi bir rassal değişgenin değil, X rassal değişgeninin,

$$Y=g(X)$$

gibi bir *işlevinin olasılık dağılımını* belirleyen ölçümötelerinin tahmini amaçlanır ve bu tahminlere X ’in *olasılık dağılımından rasgele örneklemeyle* elde edilen $\{x_{i=1(1)n}\}$ *sanal gözlem değerleri*, $g(.)$ işlevinin *açık biçiminde* yerine konarak elde edilen $\{y_{i=1(1)n}\}$ *sanal gözlem değerleri* ile ulaşılr. Gerçek gözlem değerleri *ölçüm işlemi*; sanal gözlem değerleri ise *hesaplama işlemi* sonucu elde edilen değerlerdir. *Sanal deney çıktısının* matematiksel modeli olan $g(.)$ işlevinin açık biçimini tanımlama, *sanal deney tasarımının çekirdeğini* oluşturur. Sanal deneyin matematiksel modelinin kurulması ve genel tasarımının yapılması konularını daha sonra ele alacağız.

Tasarımı tamamlanan sanal deneyin gerçekleştirilmesi ise özünde, matematiksel modeldeki *olasılık dağılımları bilinen temel girdi* değerlerinin söz konusu olasılık dağılımlarından rasgele örnekleme algoritmalarıyla elde edilip matematiksel modelde yerlerine konarak sanal deney çıktıların hesaplanması

işlemdir. Bu bölümde, olasılık dağılımı verilen X gibi belli bir değişgene ilişkin sanal gözlem değerlerinin, X 'in olasılık dağılımından rasgele örneklemeyle nasıl elde edilebileceği açıklanacaktır. Rassal değişgenin sürekli ya da süreksiz oluşuna, olasılık dağılımının gözlemsel ya da kuramsal oluşuna, kuramsal dağılımın tersinin açık biçiminin bulunup bulunmamasına bağlı olarak uygulanacak rasgele örnekleme yöntemi değişir.

Tüm rasgele örnekleme yöntemleri temelde, *birim tek-düze dağılımlı* bir R rassal değişgenin değerlerini, olasılık dağılımı verilen X gibi bir rassal değişgenin değerlerine dönüştüren ve *üreteç* olarak adlandırılan birer algoritmadır.



Birim tek-düze değişgenleri, $F(x)$ dağılımlı rassal değişgenlere dönüştürerek bu dağılımlardan rasgele örnekleme yapmanın,

- ters dönüşüm,
- kabul-red,

olmak üzere, genelde iki temel yaklaşımı vardır. Sanal gözlem üretme yöntemi olarak bunlardan hangisinin seçileceği, öncelikle rassal değişgenin dağılım ve ters dağılım fonksiyonunun yapısına ve hesaplamaların etkinliğine bağlıdır.

Ters dönüşüm yöntemi olarak bilinen yöntem, rasgele örnekleme tam ve doğrudan kuramsal bir yaklaşım olduğundan, en sık başvurulan yöntemdir. Bu yöntemin doğrudan uygulanması, ancak dağılım fonksiyonunun düzgün ve kesintisiz sürekli artan bir fonksiyon olması durumunda söz konusudur. Bir sistemin girdi ve çıktıların hepsi bu özellikte olmayabilir. Sanal deney modelindeki kimi değişgenlerin dağılımı deneysel, kesikli ve kesin sürekli artmayan nitelikte olabilir. Bu tür dağılımların ters dağılım fonksiyonları

tanımlanabildiği ve bunlardan etkin olarak sanal gözlem üretilebildiği durumlarda da, ters dönüşüm yönteminin belli biçimleri uyarlanabilir.

Öte yandan ters dağılım fonksiyonundan doğrudan sanal gözlem üretmek, kimi rassal değişkenler için etkin bir yol olmayabilir. Böyle durumlarda, rassal değişkenin özelliğine bağlı olarak, yukarıdaki diğer seçeneklerden uygun olanından yararlanılır. Bu temel yaklaşımlar dışında, bazı rassal değişkenler için etkin özel algoritmalar da bulunmaktadır.

Bu bölümde önce, tüm rasgele örnekleme algoritmalarının temel girdisi olan *birim tek-düze dağılımlı* bir rassal değişkene ilişkin sanal gözlem değerlerinin elde edilmesinde yararlanılan klasik üreteçler; birim tek-düze dağılımlı rassal değişkenin sanal gözlem değerleri verildiğinde, uygulamada en çok yararlanılan olasılık dağılımlarından rasgele örnekleme yöntemleri ve bu yöntemleri kullanarak sürekli ve kesikli kuramsal dağılımlar ile deneysel dağılımlardan rasgele örnekleme tanıtılacaktır.

3.1 Birim Tek-Düze Dağılımdan Rasgele Örnekleme

Birim tek-düze bir rassal değişkenin değerlerini elde etmenin bir yolu doğrudan böyle dağılıma sahip olduğu varsayılan gerçek dünyadaki rassal bir değişkeni gözlemektir. Hilesiz bir zar atışının ya da hilesiz bir rulet tekerleği çevrilmesinin tekdüze bir rassal değişkenin değerlerinden biri ile sonuçlanacağı ve bu değerlerin $[0,1)$ aralığına dönüştürülmesi düşünülebilir. Örneğin bu amaçla ağırlık merkezi etrafında tekdüze dönebilen bir tekerleğinin çevresi, 100 eşit dilime ayrılıp her birine 00'dan 99'a kadar bir gösterge numarası verilir ve döndürülerek durdurulursa, sabit yüzey üzerinde işaretlenmiş bir nokta karşısına her seferinde bu yüz mümkün dilimden biri eşit olasılıkla gelecektir. Her denemede sabit nokta karşısında duran dilimin gösterge numarası 100'e oranlanırsa, bu fiziki sürecin çıktısının yüzde birlik duyarlılıkta birim tek-düze bir rassal değişkenin değerleri biçiminde gözlenmiş olacağı açıktır.

Bu duyarlık, ancak tekerlek üzerindeki eşit dilim sayısı arttırılarak sağlanabilir. Sanal bir deneyde en az 10^6 'da birlik bir duyarlık gerektiğinde, bu yöntemin sorunu çözemeyeceği açıktır. Öte yandan, böyle fiziki araçları üretmek mümkün olsa bile, bir sisteme ilişkin sanal gözlem elde etmek için sanal deney laboratuvarında böyle mekanik araçlar kullanmak, ilk bakışta sistemi tümü ile gözlemeğe ya da denemeye göre bir kazanç gibi görülse de, elde edilmesi gereken sanal gözlem sayısının büyüklüğü dikkate alındığında, zaman ve maliyet açısından gerçekleştirilmesinin söz konusu olamayacağı görülür.

Birim tekdüze değişgen değerlerini elde etme amacı ile mekanik araçların yerine elektronik araçlar öngörülebilir. Nitekim,geçen yüzyıl ortalarında devreye giren ilk sayısal bilgisayarlarda, Birim tek-düze rassal bir değişgenin değerlerini elde etme amacı ile uzaydan dünyaya ulaşan kozmik ışın frekanslarını algılayan 'kozmetik ışın sayaçları' kullanılmıştır. Ancak, ister mekanik ister elektronik olsun, birim tek-düze rassal bir değişgen değerinin fiziki araçlarla elde edilmiş olmasının, sanal deney teknikleri açısından önemli bir eksikliği vardır. Sanal deneyin gerçekleştirildiği sırada, tek-düze rassal değişgen değerleri ürettiği varsayılan fiziki süreçte beklenmedik bir gelişme yaşanabilir ve farkına varılamayabilir. Bunu önlemenin yolu, üretilen değerleri bilgisayar belleğinde depolamak ve sanal deneyde kullanmadan önce tek-düze dağılım varsayımını sağlayıp sağlamadıklarına bakmaktır. Böyle bir sınama için çok büyük yer tutacak verilerin bilgisayar belleğinde depolanması ise, ne ekonomik ve ne de sanal deney tekniği açısından hiç etkin olmayan bir yoldur.

Fiziki üreteçlerdeki bu tıkanıklık, tek-düze rassal değişgen üreticinin de sanal bir üreteç olarak tasarlanması ve böylece sanal bir deneyin tümü ile sanal ortamda gerçekleştirilebilmesi ile aşılabılır. Başlangıç değeri verildiğinde, bir özyineleme biçiminde Birim tek-düze dağılıma sahip sanal bir rassal değişgenin ardışık değerlerini hesaplayan bir algoritmadan oluşan bu üreteçlere *sanal rassal sayı üreteçleri* diyoruz.

Bu üreteçler, *tohum* adını verdiğimiz bir ilk başlangıç değerinden başlayarak, görünüşte birim tekdüze dağılıma sahip

rassal bir değişgenin ardışık sanal değerlerini veren özyinelemeye dayalı bir algorithmadan oluştukları için, sonunda ilk başlangıç değeri olarak verilen tohumla ulaşacakları belli bir periyoda sahiptirler. Bu nedenle, tüm deneyler için söz konusu olan *yinelenebilirlik* kıstası, sanal deneylerde, gerçek deneylerde asla söz konusu olamayacak bir kesinlikle sağlanır. Tekdüze dağılımlı rassal bir değişgene ilişkin değerlerin fiziki bir üreteçle elde edilmesi durumunda, böyle bir periyod söz konusu değildir. Belli bir değerden sonra, sanal rassal sayı üreteçlerinin üretilmiş diziyi aynı biçimde yinemesi, her zaman sanal bir deneyde gerekli deneme sayısından daha büyük periyodlu üreteçler kullanıldığında bir sorun oluşturmaz.

3.1.1 Ara-kare yöntemi

İlk önerilen sanal rassal sayı üretici *ara-kare yöntemi* olarak bilinir. J.von Neumann tarafından 1948 de önerilen bu yöntemde, çift haneli bir tohumla başlanarak, her adımda

- bir önceki rassal sayının karesi alınır,
- bulunan sayının hane sayısı tekse, sol başına “0” konarak çift haneye tamamlanır, ve
- üretilcek rassal sayının hane sayısı kadar hane, bu sayının ortasından alınarak bir sonraki rassal sayı olarak belirlenir.

Örneğin, iki haneli rassal bir sayı dizisi üretmek için, tohum **05** olarak alınırsa:

$$1. \text{ rassal sayı: } (05)^2 = 25 \rightarrow \mathbf{25}$$

$$2. \text{ rassal sayı: } (25)^2 = 0625 \rightarrow \mathbf{62}$$

$$3. \text{ rassal sayı: } (62)^2 = 3844 \rightarrow \mathbf{84}$$

$$4. \text{ rassal sayı: } (84)^2 = 7056 \rightarrow \mathbf{05}$$

olarak bulunur ve {25, 62, 84, 05} rassal sayı dizisi elde edilir. Görüldüğü üzere, 4. rassal sayı tohum değeri ile aynı olduğundan bu üreticinin periyodu 4’tür. {25, 62, 84, 05} kümesinin elemanı olmayan ve hane sayısı daha fazla bir tohum seçilerek ara-kare yöntemi ile periyodu daha büyük bir üreteç elde edilebilir. Tohum

değerine çok duyarlı olan bu yöntemle oluşturulan üreteçlerin çoğunlukla kısa periyotlu olmaları ve uzun periyotlu olanların ise yine çoğunlukla rassallık sınamalarından geçemediği görüldüğünden, ciddi sanal deney uygulamalarında artık kullanılmamaktadır. Lehmer tarafından ilk kez 1948 de önerilen *benzerlik yöntemi*, tek-düze dağılımlı rassal sayı üretiminde en yaygın olarak kullanılan yöntemlerin ilki ve temeli olmuştur.

3.1.2 Benzerlik yöntemi

Birim tek-düze dağılımlı rassal sayı üretmede en yaygın olarak kullanılan yöntem, benzerlik yöntemidir. Bu yöntemde de, dizideki sayılar bir öncekinden belli bir formüle göre elde edilir. Özyinelemeye dayalı bu formül için, önce beş tane negatif olmayan sayı belirlenir: (1) hane sayısı, h ; (2) başlangıç değeri yada tohum, r_0 ; (3) çarpan, a ; (4) artış, c ; ve (5) modül, m . Üretilecek Birim tek-düze dağılımlı değişgen değerlerinin istenen duyarlılığına bağlı olarak hane sayısı h , çift bir sayı; modül ise a ve c den büyük olmalıdır. Bu yöntemde, her r_i değeri, bir önceki r_{i-1} değerinden hesaplanan, $a10^h r_{i-1} + c$ değerinin modüle tamsayı bölümünden kalan değerinin 10^h ye oranı olarak belirlenir:

$$r_i = [(a \cdot 10^h \cdot r_{i-1} + c) \bmod(m)] \cdot 10^{-h} \quad (1)$$

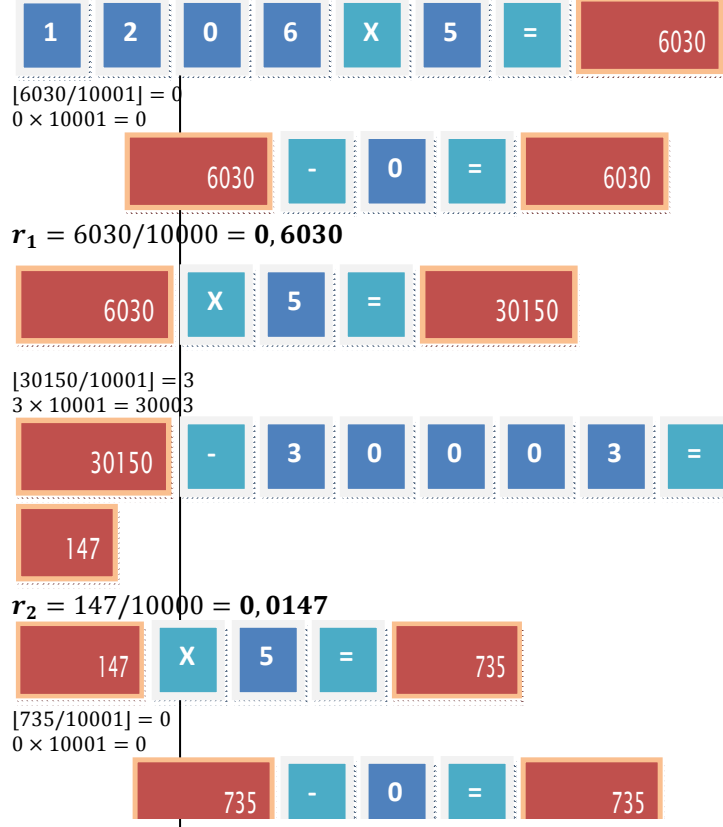
Benzerlik yöntemine ilişkin olarak verilen yukarıdaki algoritma, doğrudan doğruya Birim tekdüze bir rassal değişgenin $[0,1)$ aralığındaki sanal değerlerini üretir. Yalnızca rassal sayı üretilmek istenirse, (1) üreticinin pay ve paydasındaki 10^h çıkarılarak,

$$s_i = (a \cdot s_{i-1} + c) \bmod(m) \quad (2)$$

üretici kullanılabilir. Birim tekdüze rassal değişgen değerlerinin bilgisayarla elde edilmesinde, (1); hesap makinesi yardımı ile elde edilmesinde ise, (2) algoritması daha uygundur. Hesap makinesi ile Birim tekdüze değişgenin sanal değerlerini elde ederken, tamsayı bölme işleminin ve s_i rassal sayılarının 10^h değerine bölünerek $[0,1)$ aralığındaki r_i değerlerine dönüştürülmesinin zihinden yapılması etkinliği artırır.

Örnek: $h=4$; $s_0=0,1206$; $a=5$; $c=0$; ve $m=10^4+1=10001$ olarak seçilir ve siyah-mavi renkli işlemler zihinden yapılırsa, mavi kareler hesap makinesinin tuşlarını, pembe renkli dikdörtgen ekranını göstermek üzere, (2)'deki algoritmadan, $\{r_i; i=1(1)3\}$ dizisi Çizelge:1'de gösterilen adımlar izlenerek bir hesap makinesiyle hızlı bir biçimde hesaplanabilir.

Çizelge:1 Sözde tekdüze dağılımlı rassal bir değişgene ilişkin değerlerin hesap makinesiyle üretilmesi.



Verilen örnekteki gibi artış değeri $c=0$ olan üreteçlere *çarpımsal benzerlik üreteçleri*, $c>0$ olan üreteçlere de *karma benzerlik üreteçleri* denir. Lehmer(1951:141-145)'in önerdiği

$$s_i = (23 \cdot s_{i-1} + c) \bmod (10^8 + 1)$$

çarpımsal benzerlik üreticinin, periyodu 5,882,352 olan bir rassal sayı dizisi ürettiğini belirten Taussky-Todd(1956:17),

$$s_0 = 1; s_i = (5^{17} \cdot s_{i-1} + c) \bmod(2^{42})$$

üretici ile $2^{40} \approx 10^{12}$ periyotlu rassal bir sayı dizisi elde edileceğini göstermişlerdir.

Çarpımsal benzerlik üreticileri, ilk kez Thomson(1958:83-84) tarafından önerilen karma benzerlik üreticilerine göre, daha hızlı ancak daha kısa periyotludurlar.

s_0 , a , c , ve m için en uygun değerlerin seçimi büyük ilgi duyulan bir sorundur. Bir üreticinin periyodu her zaman modül değerinden küçük olduğundan, m için çok büyük değerlerin seçilmesi gerekir. m değerinin aynı zamanda, tamsayı bölme işlemindeki kalanın kolayca bulunmasına ve böylece benzerlik ilişkisinin hızlı hesaplanmasına olanak verecek bir değer olması istenir. Bu sorunu çözmenin kolay bir yolu, m değeri olarak, bir bilgisayarın bir defada işlemleyebildiği ikil hanelerden oluşan bir bilgi biriminin içerdiği ikil hane sayısı olan *bilgisayar kelime uzunluğunu* almaktır. Eğer bilgisayarın kelime uzunluğu k ise, toplama ve çarpma işlemleri k modülüne kendiliğinden indirgenmiş olur. Ancak $m=k$ alınması periyodu küçültebilir. Bu sorun $m=k-1$ alınarak aşılır. Bu nedenle, örneğin kelime uzunluğu 32 ikil hane olan bilgisayarlarda, rassal sayı üreticinin modülü genellikle $m=2^{32}-1$ olarak alınır.

Üreticinin a çarpanının seçimi, tümü ile üreticinin periyodunu arttıracak biçimde yapılır. Periyodun alabileceği en büyük değer olan m 'ye, ancak ve ancak,

- c ve m 'nin tek ortak böleninin 1 olması, başka bir deyişle, c 'nin m 'ye göre asal bir sayı olması;
- $a-1$ 'in, m 'nin her asal böleninin katları olması; ve
- m , 4 'ün bir katı ise, $a-1$ 'in de 4 'ün bir katı olması

durumunda ulaşılabilir. Bu noktalar göz önünde bulundurulduğunda, bilgisayarın kullandığı radiks Z , $m=Z^e$, ve $2 \leq k < e$ olmak üzere; genelde çarpan olarak $a=Z^k+1$ değerinin alınması uygundur. Çarpan değerinin seçimindeki bu basit

yaklaşım karşılık, Knuth(1969:21-24) , örneğin $e=35$ için, $a=2^{23}+2^{14}+2^2+1$ gibi daha karmaşık çarpanlar önermekte ve böyle karmaşık çarpanların kullanılma nedenlerini tartışmaktadır.

Üreteç periyodu m olduğunda, üreteç bir başlangıç değerinden itibaren m tane değer üreteceğinden, başlangıç değeri s_0 tohumunun seçiminde özellikle dikkat edilmesi gereken bir nokta yoktur. Ancak, sanal bir deneyin farklı yinelemelerinde aynı tohumun kullanılması, sanal deneyin önceki deneme sonuçlarının aynı biçimde üretilmesine yol açabilir. Bu nedenle, sanal deneyin farklı yinelemelerinde, bir önceki deneme sonundaki rassal sayı değerinin, bir sonraki denemede tohum olarak alınması yaygın bir uygulamadır. Bununla birlikte, sanal bir deney programının algoritmalarını denetleme amacı ile yapılan denemelerinde ise aynı tohumun kullanılması gerekir. Sanal deney programlamasının bu aşamasında, bir önceki sanal deneyin aynen yinelenmesi, programdaki hataları ayıklamada büyük kolaylık sağlar.

Karma benzerlik üreteçlerinde, artış değeri c 'nin m 'ye göre asal bir sayı olması dışında, genel bir kural yoktur.

Buraya kadar ele alınan benzerlik yöntemlerinin, n tane tohumla başlayıp,

$$r_i = [(10^h \cdot \sum_{j=1}^n r_{i-j} + c_j a_j) \bmod(m)] \cdot 10^{-h} \quad (3)$$

biçiminde genellenebilecek ve *toplamsal benzerlik yöntemleri* olarak adlandırılan çeşitlemeleri de önerilmiştir.

Tüm bu öneriler arasından, bilgisayar teknolojisindeki gelişmeler, Lehmer'in başlangıçta önerdiği biçimdeki çarpımsal benzerlik üreteçlerini ön plana çıkarmıştır. Modül ve çarpan değerleri iyi seçildiğinde, çarpımsal benzerlik üreteçleri ile, rassallık sınamalarından başarı ile geçen ve bir çok sanal deney için yeterli periyoda sahip güvenilir sanal rassal sayı üreteçleri elde edilebilmektedir. Fishman(1996:603-604)'da Fortran dilinde bir uyarlaması da verilen üreteçte, $a=950706376$ ve $m=2^{31}-1$ olmak üzere yaklaşık $p=2.0 \times 10^9$ periyodlu bir sanal rassal sayı dizisi elde edilebilmektedir.

Bu üreteç programlanırken, (1) tanımının doğrudan bir algoritma olarak kullanılması durumunda, programlama diline de bağlı olarak, kimi $a.r_{i-1}$ değerlerinin hesaplanmasındaki tamsayı aritmetik işlem sonuçları, kelime uzunluğu 32-bit olan bilgisayarlarda, $2^{31}-1$ değerini aşacağından *taşma hatası* verir. Bunun için, aritmetik işlemleri parçalayarak taşma hatasını önleyen (1)'dekine denk,

$$\begin{aligned} a_1 &= a \bmod(2^{16}), \quad a_2 = a - a_1; \\ X_i &= a_1 r_{i-1} \bmod(2^{31}), \quad Y_i = a_2 r_{i-1} \bmod(2^{31}); \\ r_i &= \{X_i + Y_i + [a_1 r_{i-1} 2^{-31}] + [a_2 r_{i-1} 2^{-31}]\} \bmod(2^{31} - 1) \end{aligned} \quad (4)$$

algoritması kullanılır. (4) algoritmasının QuickBasic dilinde bir uyarlaması, program ekinde *RAS* adlı bir alt-yordam olarak verilmiştir.

3.1.3 Rassallık Sinamaları

Sanal rassal sayı üretici tarafından üretilen sayı dizisi, tek-düze bir dağılımdan rasgele örneklem ile elde edilecek bir küme ile benzer olmalıdır. Üretilen rassal sayı dizisindeki her hangi bir sayı, dizinin her hangi bir sırasında ve aynı olasılıkla bulunabilmelidir. Rassal sayı üreteçleri, bir önceki rassal sayıdan bir sonrakini elde etmeye dayandığından, bu koşulu sağlamayamazlar. Ancak, üretilen dizi, rassal olduğu bilinen dizilerin özellikleri üzerine kurulu belli istatistiksel sinamalardan geçerse, üreticinin tekdüze rassal sayı ürettiği kabul edilebilir.

Bu bölümde, $[0,1)$ aralığındaki birim tekdüze dağılmış sayı dizilerine uygulanan sinamalardan bir kaçına değineceğiz. Benzer sinamalar, başka dağılımlı diziler için de geliştirilebilir. Bu sinamalar örneklerle daha önce açıklandığından, burada yalnızca rassal sayı dizilerinin sinanmasında nasıl kullanılabileceğine değinilecektir.

Kolmogorov-Simirnov uyum-iyiliği sinaması, dizinin tüm dağılımına uygulanabilir. Bu uygulamaya, *frekans* yada *dağılım-denkliliği sinaması* da denir. Bu sinama, üretilen sayılar kümesinin dağılımını kuramsal birim tek-düze dağılımla karşılaştırır. Birim aralık, bir dizi alt-aralıklara bölünür ve her alt-aralığın üst sınırına

kadar olan sanal rassal sayı dizisinin birikimli görelî sıklığı hesaplanır. Eğer i -inci üst sınır x_i ile ve bu üst sınıra ilişkin birikimli görelî sıklık f_i ile gösterilirse, $F(x)=x$ olduğundan, bu sınıma için tüm gereken, $|f_i - x_i|$ farkının en-büyük değerini bulmaktır.

Dizi sınaması denen bir ikinci sınıma, *ki-kare sınamasını* ardışık sayı ikililerine uygular. Bilgisayar rassal bir sayıyı göstermek üzere sonlu sayıda hane kullanmak zorunda olduğundan, ancak sonlu sayıda farklı sayı üretebilir. Örneğin, rassal sayı sekiz ikilî hane ile gösterilebiliyorsa, ancak $2^8=256$ farklı sayı gösterilebilir. Eğer f tane farklı sayı gösterilebiliyorsa, f^2 tane ardışık sayı ikilisi bulunabilir. Rassal bir dizi için, dizideki her olası sayı ikilisinin kuramsal görelî sıklığı f^{-2} olmalıdır. Olası sayı ikililerinin f^2 tane kombinasyonunun her biri için, bu beklenen sıklıkla bir ki-kare değeri hesaplanabilir ve bu değerler bir sınıma istatistiği elde etmek üzere toplanabilir. Bu yaklaşım, üç, dört, ve daha fazla ardışık sayılara genişletilebilir. Ancak, olası kombinasyon sayısı, özellikle f oldukça büyükse, çok fazla olabilir. Sınıma verileri için n tane bağımsız sayı ikilisi üretmek için, $2n$ tane rassal sayı üretmek gerekir. Üretilen bu diziden birinci ile ikinci, üçüncü ile dördüncü, vb biçimde sayılar eşleştirilir. Geçerli bir ki-kare sınaması için, n en az $5 f^2$ olmalı ve hatta daha fazlası yeğlenmelidir.

Daha genel bir sınıma tasarımında, r tane sayı kombinasyonunun belli kümeleri tanımlanır, her küme için kuramsal sıklık tahmin edilir, ve rassal sayı dizisinin beklenene uyup uymadığını belirlemek üzere, bir ki-kare sınaması kullanılabilir. Bu yaklaşıma dayalı, **poker sınaması** olarak bilinen yaygın bir sınımada $r=5$ alınır ve en olası poker elleri kümeleri oluşturur.

Bir başka sınıma, **permütasyon sınaması** olarak adlandırılır ve yine r tane sayı kümesine dayanır. Tüm dizi, her biri r sayıdan oluşan farklı kümelere ayrılır. r elemanın olası sıralanışlarının her biri için bir sıklık elde edilir ve bu sıklıklar kuramsal beklenenleri ile bir ki-kare sınamasından geçirilirler.

Akış sınaması, rassal sayılar dizisine doğrudan uygulanabilir. Burada ele alınan sayı dizilerinin rassallık sınamaları içinde en ayırd edicisi olup, bu sınamada diğer sınamalara göre daha fazla sayıda üreteç rassallık sınamasından geçememektedir. Poker sınaması bu anlamda ikinci sırada gelmekte, dizi sınaması ve sıklık sınaması ise en az ayırd edici sınama özelliği göstermektedirler. Hemen hemen dikkatlice tasarlanmış her üreteç, dizi ve sıklık sınamalarından geçen diziler üretebilir.

Coveyou ve MacPherson(1969) tarafından önerilen ve Knuth(1969:82-96) tarafından da bir uygulama algoritması verilen **spektral sınama**, ardışık n tane sayının bağımsızlığını ölçmekte ve en ayırd edici sınama olduğu öne sürülmektedir.

Rassal sayı kümelerinin sınanmasında özellikle yararlı oldukları kanıtlanmış diğer iki sınama, U^2 ya da **uzaklık** ve **aralık sınamalarıdır**. Bu sınamalar, daha önce tartışılmadığından, burada ayrıntıları ile tanımlanacaktır.

Uzaklık sınamasında, rassal sayıların ardışık ikilileri, birim kare içindeki noktaların koordinatları olarak kabul edilir. Örneğin, r_1, r_2, r_3 , ve r_4 rassal sayılar dizisini ele alalım. Bu sayılar (x,y) düzleminde, (r_1, r_2) ve (r_3, r_4) noktalarının koordinatları olarak görülebilir. Bu iki nokta arasındaki uzaklığın karesi, $U^2 = (r_1 - r_3)^2 + (r_2 - r_4)^2$ olarak verilebilir. Eğer bu noktalar, birim kare içinde tek-düze dağılımlı iseler, U^2 nin gözlenen değerinin en az x kadar olma olasılığı, Wilson(1891:390).

$$F(x) = \begin{cases} \pi x - \frac{8x^{2/3}}{3} + \frac{x^2}{2}, & x \leq 1; \\ \frac{1+8\sqrt[3]{(x-1)^2}}{3} + (\pi - 2) + 4\sqrt{x-1} - \frac{x^2}{2} - 4x \cdot \text{arcsec}\sqrt{x}, & 1 < x \leq 2 \\ 1, & x > 2 \end{cases} \quad (5)$$

olarak ifade etmiştir. x 'in seçilen bazı değerleri için hesaplanan bu olasılıklar aşağıdaki çizelgedeki gibidir.

Çizelge:2 Birim kare içindeki rasgele iki nokta arasındaki uzaklığın karesinin dağılımı.

x	0.0	0.25	0.5	0.75	1.0	1.5	2.0
$F(x)$	0.0	0.483	0.753	0.905	0.975	0.999	1.0

Jaffray-Gruenberger(1965), bu olasılıkların hesabını daha ayrıntılı tartışmakta, poker ve aralık sınamalarını açıklamaktadırlar. Bu olasılıkları kullanarak, belirlenen aralıklarda U^2 değerleri için kuramsal sıklıklar hesaplanabilir ve bir χ^2 sınaması yapılabilir.

Aralık sınaması, belli bir olayın oluşumu ile aynı olayın bir sonraki oluşumu arasındaki, olayın oluşum sayısı cinsinden ölçülen, *aralığın* uzunluğuna dayanır. Bu yaklaşım rassal sayı dizilerine iki biçimde uygulanabilir.

Birinci yaklaşımda, $0 < a < b < 1$ olmak üzere, birim aralığın (a,b) gibi bir alt aralığı tanımlanır. Aralık uzunluğu, dizinin başından itibaren alt aralık içine düşen sayılar arasındaki dizide bulunan sayıların sayısıdır. Belli bir rassal sayı dizi için bu yaklaşıma göre aralık uzunlukları,

$$\begin{aligned} w_i &= \begin{cases} 0, & (r_i \leq a) \wedge (r_i \geq b) \\ 1, & a < r_i < b \end{cases} \\ U_i &= \begin{cases} U_{i+1}, & (w_i = 0) \wedge \sum_{i=1}^n w_i \geq 1 \\ 0, & w_i = 1 \end{cases} \\ m &= m + 1 \mid (i > 1) \wedge (w_i = 1); U_m = U_{i-1} \blacksquare \end{aligned} \quad (6)$$

algoritması kullanılarak bulunabilir.

Çizelge: 3.a'da, 20 elemanlı bir rassal sayı dizisinin, bu algoritmaya göre bulunan aralık uzunluklarının elde edilişi görülmektedir. Bu örnekte, birim aralığın alt-aralığı olarak seçilen $(0.35,0.5)$ aralığına düşen rassal sayılar, 1., 2., 3., 4., 6., 7., 10., 12., 13., 14., 15., 18.,19.,20.,21., 24., 25., 27., 28., 29., ve 30. sıradaki koyu renkle vurgulanan rassal sayılar olup; bu sayıların belirlediği toplam 20 aralıkta yer alan sayıların sayısı ise, sırası ile 0, 0, 0, 1, 0, 2, 1, 0, 0, 1, 2, 0, 0, 0, 2, 0, 1, 0, 0, ve 0 olarak bulunmuştur. Birim tek düze dağılımlı rassal bir değişgenin (a,b) alt-aralığında olma olasılığı, $b-a$ olduğundan, k uzunluğunda bir aralık gözleme olasılığı, $Ol(k)=(b-a)(1-b+a)^k$ olur. Dolayısı ile, 0,1,2,.....,n uzunluğundaki aralıklar sayılır, ve n 'den daha uzun aralıkların tümü bir araya toplanırsa, tanımlanan olasılıktan kuramsal sıklıklar hesaplanarak, bir χ^2 sınamasına gidilir.

Aralık sınamasının kullanımındaki ikinci bir yaklaşımda ise, rassal sayılar kümesinin tümü, ondalıktan önceki sıfırların göz önüne alınmadığı, 0-9 aralığındaki rakamların oluşturduğu bir haneler dizisi olarak ele alınır. 0-9 aralığındaki rakamlardan biri sinama değeri olarak seçilir, ve bu rakamın yinelenmeleri arasındaki haneler sayılır. Sinama değeri için seçilen rakam h ve rassal sayı dizisinin ondalık hanesi ve ondalıktan önceki sıfırlar atıldıktan sonra birleştirilerek elde edilen rakamlar dizisinin haneleri h_j olmak üzere, bu ikinci yaklaşıma göre verilen bir rassal dizinin aralık uzunlukları,

$$w_i = \begin{cases} 0, & h_j \neq h \\ 1, & h_j = h \end{cases}$$

$$U_j = \begin{cases} U_{j+1}, & (w_j = 0) \wedge \sum_{j=1}^n w_j \geq 1 \\ 0, & w_j = 1 \end{cases}$$

$$m = m + 1 \mid (j > 1) \wedge (w_j = 1); U_m = U_{j-1} \blacksquare \quad (7)$$

algoritması ile bulunabilir. 0-9 aralığında on rakam bulunduğundan, rassal bir dizinin hanelerinde her birinin gözlenme olasılığı 0.1 ve k uzunluğunda bir aralık gözleme olasılığının, $Ol(k)=(0.1)(0.9)^k$ olacağı açıktır. Aralık uzaklıklarına ilişkin gözlenen değerlerle rassal bir dizi için beklenen değerlerin karşılaştırılması yine bir χ^2 sınaması ile yapılır.

Çizelge:3.a ve 3.b'deki aralık uzaklıkları kümesinin her bir elemanının olasılığı Çizelge:4.a ve 4.b. de gösterildiği gibi hesaplanır ve kuramsal sıklıkları en az 5 olacak biçimde aralık uzaklıkları yeniden kümelendikten sonra, rassal sayılar dizisi Çizelge:5.a ve 5.b'de sonuçları özetlenen bir χ^2 sınamasından geçirilebilir. Aralık uzunluğu sınaması için hesaplanan χ^2 istatistikleri, birinci yaklaşım için,

$$\chi^2_{(1,0.05)}=3.84146 < \chi^2=30.2323,$$

ve ikinci yaklaşım için de,

$$\chi^2=4.18294 < \chi^2_{(2,0.05)}=5.99$$

olarak elde edildiğinden, verilen rasgele sayı dizisinin tekdüze dağıldığı, birinci yaklaşıma göre red edilirken, ikinci yaklaşıma göre kabul edilecektir.

Çizelge: 3 Bir rassal sayı dizisinin aralık uzunluklarının bulunuşu:

a) Birinci yaklaşım ($a=0.35, b=0.5$)

b) İkinci yaklaşım ($h=2$).

i	r_i	w_i	U_i	m	U_m
1	0.3931	1	0	-	
2	0.3558	1	0	1	0
3	0.4538	1	0	2	0
4	0.4213	1	0	3	0
5	0.1206	0	1	-	-
6	0.4492	1	0	4	1
7	0.3615	1	0	5	0
8	0.2612	0	1	-	-
9	0.1814	0	2	-	-
10	0.3831	1	0	6	2
11	0.1125	0	1	-	-
12	0.4112	1	0	7	1
13	0.3922	1	0	8	0
14	0.4292	1	0	9	0
15	0.3813	1	0	10	1
16	0.7426	0	1	-	-
17	0.9311	0	2	-	-
18	0.4316	1	0	11	2
19	0.3922	1	0	12	0
20	0.4352	1	0	13	0
21	0.3601	1	0	14	0
22	0.1872	0	1	-	-
23	0.7812	0	2	-	-
24	0.4181	1	0	15	2
25	0.3591	1	0	16	0
26	0.2983	0	1	-	-
27	0.3702	1	0	17	1
28	0.4042	1	0	18	0
29	0.4903	1	0	19	0
30	0.4102	1	0	20	0

i,n	h_i	w_i	U_m
1-4	3-9-3-1	0,0,0,0	-
5-8	3-5-5-8	0,0,0,0	-
9-12	4-5-3-8	0,0,0,0	-
13-16	4-2-1-3	0,1,0,0	-
17-20	1-2-0-6	0,1,0,0	3
21-24	4-4-9-2	0,0,0,1	5
23-28	3-6-1-5	0,0,0,0	-
29-32	2-6-1-2	1,0,0,1	4,2
33-36	1-8-1-4	0,0,0,0	-
37-40	3-8-3-1	0,0,0,0	-
41-44	1-1-2-5	0,0,1,0	10
45-48	4-1-1-2	0,0,0,1	4
49-52	3-9-2-2	0,0,1,1	2,0
53-56	4-2-9-2	0,1,0,1	1,1
57-60	3-8-1-3	0,0,0,0	-
61-64	7-4-2-6	0,0,1,0	6
65-68	9-3-1-1	0,0,0,0	-
69-72	4-3-1-6	0,0,0,0	-
73-76	3-9-2-2	0,0,1,1	11,0
77-80	4-3-5-2	0,0,0,1	3
81-84	3-6-0-1	0,0,0,0	-
85-88	1-8-7-2	0,0,0,1	7
89-92	7-8-1-2	0,0,0,1	3
93-96	4-1-8-1	0,0,0,0	-
97-100	3-5-9-1	0,0,0,0	-
101-104	2-9-8-3	1,0,0,0	8
105-108	3-7-0-2	0,0,0,1	6
109-112	4-0-4-2	0,0,0,1	3
113-116	4-9-0-3	0,0,0,0	-
117-120	4-1-0-2	0,0,0,1	7

Çizelge:4 Aralık Uzunlukları k 'ların Gözlenme Sıklıkları(G_k), Gözlenme Olasılıkları, ve Beklenen Sıklıkları(B_k).

a) Birinci yaklaşıma göre.

k	G_k	$Ol(U=k)$	B_k
0	13	0.15	3.00
1	4	0.1275	2.55
2	3	0.108375	2.17
$k>2$	0	0.614125	12/28
Toplam	20	1.000000	20.00

b) İkinci yaklaşıma göre.

k	G_k	$Ol(U=k)$	B_k
0	2	.100000	2.00
1	2	0.090000	1.80
2	2	.081000	1.62
3	4	0.072900	1.458
4	2	0.065610	1.312
5	1	.059049	1.181
6	2	0.053144	1.063
7	2	.047830	0.957
8	1	0.043047	0.861
9	0	0.038742	0.775
10	1	0.034868	0.697
11	1	0.031381	0.628
$k>11$	0	0.282429	5.648
Toplam	20	1.000000	20.000

Çizelge:5 Bir Sayı Dizinin χ^2 İstatistiği ile Aralık Uzunluğu Rassallık Sınaması.

a) Birinci Yaklaşım.

U	G	B	$(G-B)^2/B$
$0 \leq U \leq 1$	17	5.55	23.622072
$1 < 2 \leq U$	3	14.45	9.0728373
Toplam	20	20.00	$\chi^2=32.694909$

b) İkinci Yaklaşım.

U	G	B	$(G-B)^2/B$
$0 \leq U \leq 3$	10	6.878	1.417110
$4 \leq U \leq 6$	5	3.556	0.586371
$6 < 7 \leq U$	5	9.566	2.179423
Toplam	20	20.000	$\chi^2=4.18294$

3.2 Ters Dönüşüm Yöntemi

Teorem:1 Dağılım İşlevinin Dağılımı: X rassal değişkeninin sürekli artan dağılım işlevi $Y=F(x)$ olsun. $X \rightarrow Y$ dönüşümüne, olasılık tümlev dönüşümü denir. Y 'nin dağılımı birim tekdüze dağılımdır.

İspat: F , olasılık yoğunluk işlevi olduğundan, $-\infty < x < \infty$ için $0 \leq F(x) \leq 1$ olur. Dolayısı ile $Ol(Y < 0) = Ol(Y > 1) = 0$. Eğer F kesin

artan bir işlev ise $0 < y < 1$ aralığındaki her hangi bir y değeri için, $F(x_0) = y$ denkleğini sağlayan tek bir x_0 değeri vardır. Eğer x değerlerinin her hangi bir aralığının tümü için, $F(x) = y$ denkleği sağlanıyorsa, bu aralıktan her hangi bir x değeri $x = x_0$ olarak alınabilir. Eğer G , Y 'nin olasılık yoğunluk işlevi ise, o zaman,

$$G(y) = Ol(Y \leq y) = Ol(X \leq x_0) = F(x_0) = y$$

olur. $G(y) = y$ ve $0 < y < 1$ olduğu için, G birim tekdüze olasılık yoğunluk işlevidir•

Ters dönüşüm yöntemi olarak bilinen yöntem, rasgele örneklemeye tam ve doğrudan kuramsal bir yaklaşım olduğundan, en sık başvuru olan yöntemdir. Bu yöntemin doğrudan uygulanması, ancak dağılım işlevinin düzgün ve sürekli artan bir işlev olması durumunda söz konusudur. Bir dizgenin girdi ve çıktıların hepsi bu özellikte olmayabilir. Sanal deney modelindeki kimi değişkenlerin dağılımı deneysel, kesikli ve kesin sürekli artmayan nitelikte olabilir. Bu tür dağılımların ters dağılım işlevleri tanımlanabildiği ve bunlardan etkin olarak sanal gözlem üretilebildiği durumlarda da, ters dönüşüm yönteminin belli biçimleri uyarlanabilir.

Teorem:2 *Ters Dönüşüm Yöntemi:* Dağılım işlevinin tersi

$$F^{-1}(r) = \text{ebas}\{x \in [a, b] : F(x) \geq r, 0 \leq r \leq 1\}$$

olan X rassal değişkeninin *sürekli ve kesin artan* dağılım işlevi

$$\{F(x), a \leq x \leq b\}$$

olsun. R rassal değişkeninin dağılımı birim tek düze $[0,1]$ ise,

$$X = F^{-1}(R) \quad (8)$$

rassal değişkeninin dağılımı, F' dir.

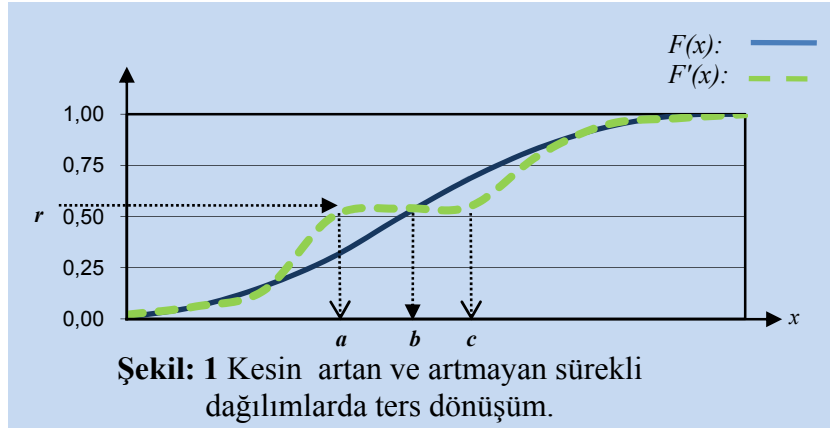
İspat Teorem 1'den

$$Ol(X \leq x) = Ol[F^{-1}(R) \leq x] = Ol[R \leq F(x)] = F(x)$$

olduğunu görmek teoremin ispatı için yeter.

Birim tek düze dağılımlı bir R rassal değişkeninin r_i gibi bir değeri verildiğinde, X rassal değişkeninin kesin sürekli artan $F(x)$

dağılımından (8)'deki ters dönüşümü ile, x_i sanal değerinin elde edilişi Şekil: 1'de gösterildiği gibidir.



$F(x)$ kesin sürekli artan bir fonksiyon olduğu için, şekilde de görüldüğü biçimde, verilen bir r değerine karşılık, bir tek $x=b$ gibi bir değer bulunur.

Teorem:1'in sonucu, doğrudan ters dönüşüm algoritmasının bir tanımıdır.

Algoritma:1 Ters Dönüşüm Algoritması

Amaç: Sürekli ve kesin artan $\{F(x), a \leq x \leq b\}$ dağılımından rasgele örneklem ile X rassal değişkeninin sanal gözlem değerlerini üretmek.

Girdi: $\{F(r) = 1, 0 \leq r \leq 1\}$ dağılımlı R rassal değişkeninin sanal değerleri ve $\{x = F^{-1}(r), 0 \leq r \leq 1\}$ değerlerinin hesaplanabilmesi için gerekli tanımlar.

Çıktı: X

Yöntem: $X \leftarrow F^{-1}(R)$

Ters olasılık dağılımlarından doğrudan ters dönüşüm yöntemiyle rasgele örneklem yapılabilen rassal değişkenler Çizelge:6'da verilmiştir.

Çizelge:6 Olasılık dağılımlarından doğrudan ters dönüşüm yöntemiyle rasgele örnekleme yapılabilen rassal değişkenler.

Değişgen	Dağılım İşlevi, $F(x)$	Ters Dağılım İşlevi, $F^{-1}(x)$
<i>Bernoulli(o)</i>	$F(x) = \begin{cases} 0, & x = 1 \\ 1 - o, & x = 0 \end{cases}$	$X = \begin{cases} 1, & R \leq o \\ 0, & R > o \end{cases}$
<i>Cauchy(θ, λ)</i>	$F(x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \left(x - \frac{\theta}{\lambda} \right), \quad \lambda > 0, -\infty < x < \infty .$	$X = \theta + \lambda. \tan \left[\pi \left(R - \frac{1}{2} \right) \right]$
<i>Erlang(β, γ)</i>	$F(x) = 1 - e^{-\left(\frac{x}{\beta}\right)} \left[\sum_{i=0}^{\gamma-1} \frac{(x/\beta)^i}{i!} \right], \quad \beta > 0, \gamma > 0, \quad 0 \leq x < \infty$	$X = -\beta. \log \left(\prod_{i=1}^{\gamma} R_i \right)$
<i>Geometrik(o)</i>	$F(x) = 1 - (1 - o)^{x+1}, x = 0(1)n$	$X = \left\lceil \frac{\log R}{\log(1 - o)} \right\rceil - 1$
<i>Gumbel(α, β)</i>	$F(x) = e^{-e^{\left(\frac{x-\alpha}{\beta}\right)}}, \alpha: \text{Tepe Değer}, \beta > 0, -\infty < x < \infty .$	$X = \alpha - \beta(-\log R)$
<i>Kesikli tek-düze, ktd(a,b)</i>	$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & x = a(1)b, \\ 1, & x > b \end{cases}$	$X = \lfloor (b - a + 1)R \rfloor + a$

<i>Tek-düze, td(a,b)</i>	$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b, \\ 1, & x \geq b \end{cases}$	$X = (b-a)R + a, \quad R \sim td(0,1).$
<i>Üstel(α)</i>	$F(x) = 1 - e^{-\left(\frac{x}{\alpha}\right)}, \quad 0 \leq x < \infty$	$X = -\alpha \cdot \log R$
<i>Weibull(α, c)</i>	$F(x) = 1 - e^{-\left(\frac{x}{\alpha}\right)^c}, \quad 0 < x < \infty$	$X = \alpha(-\log R)^{\frac{1}{c}}$

Şekil:1’de noktalardan oluşan eğriyle gösterilen $F'(x)$ dağılımı ise kesin sürekli artan bir işlev olmadığından, her r değerinin ters dönüşümü, X rassal değişgeninin sanal değeri olarak tek bir değer değil, seçilen bölgede olduğu gibi, $a < x < c$ gibi bir aralık verebilir. Sürekli ancak kesin artmayan dağılım işlevinin artmayan bölgeleri için söz konusu olan bu durum, kesikli dağılımlarda ve deneysel dağılımlarda geneldir. Bu dağılımlarda ters dönüşüm, Şekil:1’deki gibi, bir aralığı belirler. Bu durumda, sanal değer olarak söz konusu aralığın ilişkin olduğu değişgen değeri, ya da bu aralıktaki değerleri temsil eden bir değer alınır.

Kimi rassal değişgenlerin her hangi bir kuramsal dağılımı bulunmayabilir. Böyle durumlarda söz konusu değişgenin deneysel dağılımından rasgele örneklemeyle sanal gözlem değerlerinin üretilmesinde yine ters dönüşüm yöntemi uygulanabilir. Sürekli bir değişgenin deneysel dağılımı, kuramsal dağılımlarda olduğu gibi matematiksel bir işlev değil, söz konusu rassal değişgen değerlerinin belli sayısal değerler arasında olma olasılıklarını bu değer aralıklarına eşleyen bir çizelge biçimindedir.

Ters dönüşüm yönteminin, deneysel dağılımlara, kesikli dağılımlara ve sürekli dağılımların sürekli artmadığı bölgeleri için genelleştirilerek bir rasgele örnekleme algoritmasının oluşturulmasında, Çizelge:6’daki gibi genelleştirilmiş bir deneysel dağılımın simgeleminden yararlanılabilir.

Çizelge:7 Genelleştirilmiş bir deneysel dağılım.

j	X	$Ol(a_{j-1} \leq X < a_j)$	$Ol(X \leq a_j)$
1	$a_0 \leq X < a_1$	$Ol(X \leq a_1) - Ol(X \leq a_0)$	$Ol(X \leq a_1)$
2	$a_1 \leq X < a_2$	$Ol(X \leq a_2) - Ol(X \leq a_1)$	$Ol(X \leq a_2)$
.	.	.	.
$j+1$	$a_j \leq X < a_{j+1}$	$Ol(X \leq a_{j+1}) - Ol(X \leq a_j)$	$Ol(X \leq a_j)$
.	.	.	.
$k-1$	$a_{k-2} \leq X < a_{k-1}$	$Ol(X \leq a_{k-1}) - Ol(X \leq a_{k-2})$	$Ol(X \leq a_{k-1})$
k	$a_{k-1} \leq X < a_k$	$Ol(X \leq a_k) - Ol(X \leq a_{k-1})$	$Ol(X \leq a_k) = 1$

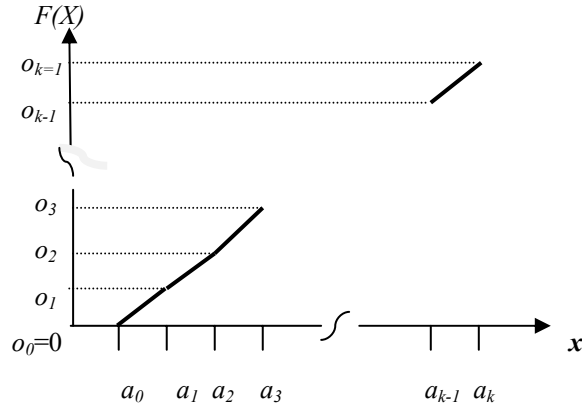
Sürekli bir X değişgenine ilişkin Çizelge:7'deki deneysel dağılımın sınıf aralıkları, j ile gösterilen aralık sıra numaraları ile de temsil edilebileceğinden,

$$O\{X \leq a_j\} = o_j$$

olmak üzere, (9)'daki gibi kesikli bir dağılım biçiminde de ifade edilebilir.

$$F(x) = \begin{cases} (o_1 - o_0) \frac{x - a_0}{a_1 - a_0}, & a_0 \leq x < a_1; \\ \vdots \\ (o_j - o_{j-1}) \frac{x - a_{j-1}}{a_j - a_{j-1}}, & a_{j-1} \leq x < a_j; \\ \vdots \\ (o_k - o_{k-1}) \frac{x - a_{k-1}}{a_k - a_{k-1}}, & a_{k-1} \leq x < a_k. \end{cases} \quad (9)$$

(9)'daki kesikli dağılım işlevinin grafiği Şekil:2'deki gibi, ardışık olarak (a_0, o_0) , (a_1, o_1) , ..., (a_j, o_j) , ..., (a_k, o_k) noktalarını birleştiren kesikli doğru parçalarından oluşur.



Şekil:2 Kesikli dağılım grafiği.

(9)'daki kesikli dağılımın, birim tek-düze dağılımlı R rassal değişgeninin r gibi bir değerine göre ters dönüşümü ise,

$$F^{-1}(r) = \begin{cases} (a_1 - a_0) \frac{r - o_0}{(o_1 - o_0)} + a_0, & o_0 \leq r < o_1; \\ \vdots \\ (a_j - a_{j-1}) \frac{r - o_{j-1}}{(o_j - o_{j-1})} + a_{j-1}, & o_{j-1} \leq r < o_j; \\ \vdots \\ (a_k - a_{k-1}) \frac{r - o_{k-1}}{(o_k - o_{k-1})} + a_{k-1}, & o_{k-1} \leq r < o_k. \end{cases} \quad (10)$$

olarak tanımlanabilir. Kesikli dağılımlar için geçerli olan (10)'nun, deneysel dağılımlar için uyarlaması ise, Çizelge:6'nın 1. ve 2. sütunlarındaki $X \leftrightarrow J$ dönüşümüne göre (11)'deki gibi yapılabilir.

$$J = \begin{cases} 1, & r < o_1; \\ 2, & o_1 \leq r < o_2; \\ \vdots \\ j, & \sum_{i=1}^{j-1} o_i \leq r < \sum_{i=1}^j o_i; \\ \vdots \\ k, & o_{k-1} \leq r < o_k. \end{cases} \quad (11)$$

$X \leftarrow J$ ■

Ters dönüşüm yöntemi, (10)'daki tanıma göre kesikli bir dağılımdan rasgele örnekleme ile X rassal değişgeninin *sanal değerini*; (11)'deki tanıma göre ise, deneysel bir dağılımdan rasgele örnekleme ile X rassal değişgeninin sanal değeri değil, sanal değer in bulunduğu *aralığın sıra numarasını* verir. Aralık sıra numarasının $J \rightarrow X$ dönüşümü ile de sanal değerler elde edilebileceğinden, kesikli ve deneysel dağılımlardan rasgele örnekleme yöntemi olarak genelleştirilmiş ters dönüşümün temel algoritması olarak (11) alınır.

$J \rightarrow X$ dönüşümü; deneysel dağılımlarda, sanal değer in içinde bulunduğu aralığı temsil eden bir değeri; kesikli dağılımlarda ise (11)'deki dönüşüm değerini verecek biçimde tanımlanmak üzere, genelleştirilmiş ters dönüşüm yönteminin bir bilgisayar alt-yordamı biçiminde programlanabilecek algoritması aşağıdaki gibi verilebilir. k 'nın büyük olması durumunda, önce o_j ler büyükten küçüğe doğru sıralanarak Algoritma:2'nin etkinliği artırılabilir.

Algoritma: 2 Kesikli ve Deneysel Dağılımlardan Rasgele Örneklem İçin **Ters Dönüşüm Algoritması**.

Amaç: $\{F_j(x), a_j \leq x \leq b_j, j = 1(1)k\}$ dağılımlı $X \equiv X_j$ rassal değişkeninin sanal gözlem değerlerini üretmek.

Girdi: $\{F(r) = 1, 0 \leq r \leq 1\}$ dağılımlı R rassal değişkeninin sanal değerleri ve $\{F_j^{-1}(r), 0 \leq r \leq 1\}$ değerlerinin hesaplanabilmesi için gerekli tanımlar.

Çıktı: X

Yöntem: $X \leftarrow F_j^{-1}(R)$

$\{j = 1(1)k,$
 $r_i < o_j \text{ ise çık. } \}$
 $x_i \leftarrow j.$

3.3 Red-Kabul Yöntemi

Teorem: 3 $Y=G(x)$ dağılımından y sanal gözlem değerleri elde edilebiliyorsa, s

$$\frac{f(y)}{g(y)} \leq s, \forall y \quad (12)$$

ilişkisini sağlayan her hangi bir sabit olmak üzere; üretilen y değerlerinden,

$$\frac{f(y)}{g(y)} \quad (13)$$

olasılığıyla kabul edilecek x değerleri, $F(x)$ dağılımlı X rassal değişkenine ilişkindir.

İspat: $G(x)$ dağılımından N inci denemede üretilen y değeri kabul edilirse;

$$O\ell(X \leq x) = O\ell(Y_N \leq x) = O\ell\{Y \leq x | R \leq g(y)/sg(y)\}$$

$$\begin{aligned}
&= \frac{O\ell\{Y \leq x, R \leq f(y)/sg(y)\}}{O\ell\{R \leq g(y)/sg(y)\}} \\
&= \frac{\int_{-\infty}^x O\ell\{Y \leq x, R \leq f(y)/sg(y) | Y = y\} g(y) dy}{O\ell\{R \leq f(y)/sg(y)\}} \\
&= \frac{\int_{-\infty}^x [f(y)/sg(y)] g(y) dy}{O\ell\{R \leq f(y)/sg(y)\}} \\
&\underset{x \rightarrow \infty}{\text{er}} \frac{\int_{-\infty}^x [f(y)/sg(y)] g(y) dy}{O\ell\{R \leq f(y)/sg(y)\}} = F(x) \blacksquare
\end{aligned}$$

Teorem:3'e göre, Red-Kabul yöntemiyle rasgele örnekleme yöntemi Algoritma:3'teki adımlar izlenerek uygulanır.

Algoritma: 3 Sürekli dağılımlardan **Red-Kabul** yöntemiyle rassal örnekleme.

Amaç: $F(x)$ dağılımlı rassal değişgeninin sanal gözlem değerlerini üretmek.

Girdi: $\{F(r) = 1, 0 \leq r \leq 1\}$ dağılımlı R rassal değişgeninin sanal değerleri, $\{G^{-1}(r_j)\}$ değerlerinin hesaplanabilmesi için gerekli tanımlar, $s, j=0$.

Çıktı: X

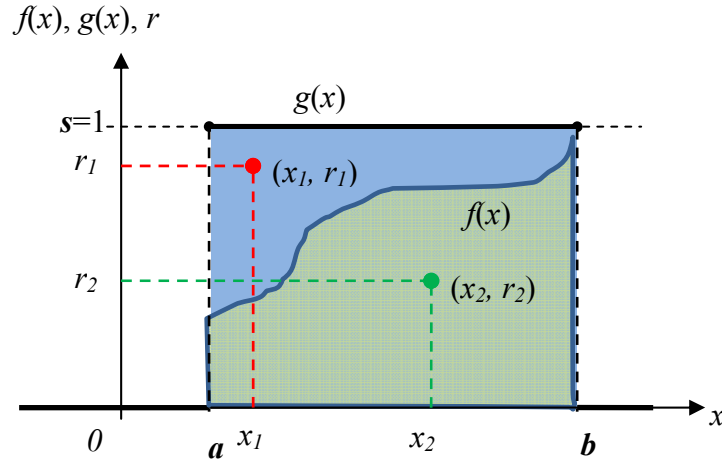
Yöntem: $\{j = j + 1;$

$$y_j = G^{-1}(r_j);$$

$$r_j \leq \frac{f(y_j)}{sg(y_j)} \text{ ise çık}$$

$$X \leftarrow Y \blacksquare$$

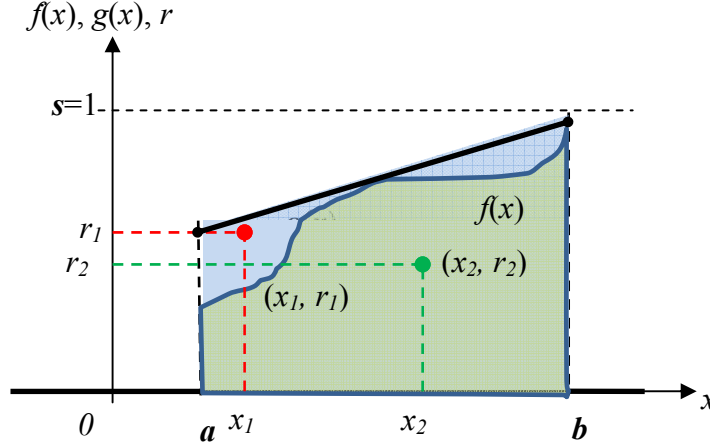
Şekil:3'deki gibi, $a < x < b$ aralığında tanımlı her hangi bir olasılık yoğunluk işlevi $f(x)$ 'den, *red-kabul* yöntemiyle rasgele



Şekil:3 Red-kabul yönteminin temel mantığı.

örnekleme için $g(x) = s = \text{Enb}\{f(x), a < x < b\}$ biçiminde seçilebilir ve $s=1$ olacak biçimde grafik ölçeklendirilebilir. $g(x)$ 'in altında kalan alanın, $\int_a^b g(x)dx = s \cdot (b - a) = 1 \cdot (b - a) = b - a$ olacağı açıktır. Eğer, $g(x) = \frac{1}{b-a}$, $a < x < b$ olarak tanımlanırsa, $\int_a^b g(x)dx = 1$ koşulu sağlanır ve dolayısı ile $X \sim td(a, b)$ olur. 3.5.1'de görüleceği üzere, ters dönüşüm yöntemiyle $X \sim td(a, b)$ rassal değişgeninin sanal gözlem değerlerini, birim tek-düze R tek-düze rassal değişgenin r değerlerinden ters dönüşüm yöntemiyle üretmek son derece kolaydır. Red-kabul yöntemine göre, üretilen $y_i = x_i$ değeri, $y_i < f(y_i)$ koşulunu sağlarsa, $F(x)$ dağılımından rasgele örneklemeyle elde edilen sanal gözlem olarak kabul edilecektir. Şekil:3'de verilen örnekte görsel olarak, mavi ile gösterilen red bölgesine düşen (x_1, r_1) noktasının $y_i < f(y_i)$ koşulunu sağlamadığı; yeşil ile gösterilen kabul bölgesine düşen (x_2, r_2) noktasının ise sağladığı görülmektedir. Buna göre, üretilen x_2 değeri, $F(x)$ dağılımından gelmiş kabul edilecek; x_1 değeri ise red edilecektir.

Red-kabul yönteminin etkinliği, $G(x)$ dağılımından üretilen değerlerin, $F(x)$ dağılımından üretilmiş değerler olarak, (13)'deki kabul edilme olasılığı ile doğru orantılıdır. $f(x)$ 'in zarfı olarak seçilecek $g(x)$ 'e bağlı olarak, Şekil:3'teki zarfın boşa kalan red alanı değişir. Bu nedenle, rasgele örneklemenin kolay yapılabilmesi yanında, zarf işlevi $g(x)$; Şekil:4'teki gibi red bölgesininin olabildiğince küçük olacağı biçimde seçilmelidir.



Şekil:4 Red-kabul yönteminin etkinliği.

3.4 Rassal Değişkenler Arasındaki Matematiksel İlişkilerden Yararlanarak Dolaylı Rasgele Örneklem

Bilinen yöntemlerle doğrudan her hangi bir rassal değişkenin olasılık dağılımından rasgele örnekleme yapılamayan durumlarda, söz konusu rassal değişkenin olasılık dağılımı ile doğrudan rasgele örnekleme yapılabilen rassal değişkenler arasında matematiksel bir bağıntının varlığı araştırılmalıdır. Eğer böyle bir bağıntı varsa, söz konusu rassal değişkenin sanal gözlem değerleri, bu bağıntıdan yararlanılarak dolaylı yoldan elde edilebilir. Rassal değişkenler arasında bilinen matematiksel ilişkiler Çizelge:8'de özetlenmiştir.

Çizelge:8 Rassal değişkenler arasındaki ilişkiler.

<i>X</i>		Matematiksel İlişki ve Koşular	<i>Y</i>
Dağılım	Olasılık Yoğunluk İşlevi, $f(x)$		Dağılım
<i>Birim Normal, N(0,1)</i>	$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty < x < \infty$	$Y = \sum_{i=1}^v X_i^2$	$\chi^2_{(v)}$
<i>Normal(μ, σ^2), N(μ, σ^2)</i>	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < \infty$	$\mu = 0, \sigma^2 = 1$	$N(0,1)$
		$\mu = \sigma^2 = \lambda$	$Poisson(\lambda)$
		$\mu = n.o,$ $\sigma^2 = n.o. (1 - o)$ $n \rightarrow \infty$	$Binom(n, o)$
<i>Üstel(α)</i>	$\alpha e^{-\alpha x}, x > 0$	$r = 1$	$Gama(\alpha, 1)$
		$Y = \sum_{i=1}^{\gamma} X_i, \quad \beta = \alpha.$	$Erlang(\beta, \gamma)$
<i>Gama(α, r)</i>	$\frac{\alpha}{\Gamma(r)} (\alpha x)^{r-1} e^{-\alpha x}, x > 0$	$\alpha = \frac{1}{2}, r = \frac{v}{2}$	$\chi^2_{(v)}$
<i>Binom(n, o)</i>	$\binom{n}{x} o^x (1 - o)^{n-x}, x = 0(1)n$	$n = 1$	<i>Bernoulli(o)</i>

		$\mu = n \cdot o,$ $n \rightarrow \infty$	$Poisson(\lambda)^*$
Negatif Binom (r, o)	$\binom{x+r-1}{r-1} o^{r-1} (1-o)^x, x = 0(1)\infty$	$r = 1$	$Geometrik(o)$
t (r)	$\frac{\Gamma[(r+1/2)]}{\sqrt{\pi r} \Gamma(r/2)} \frac{1}{(1+x^2/r)^{(r+1)/2}}, -\infty < x < \infty$	$Y = \frac{U}{\sqrt{V/r}}; U \sim (N(0,1), V \sim \chi^2_{(r)})$	$t(r)$
F (r_1, r_2)	$\frac{\Gamma\left[\frac{r_1+r_2}{2}\right] (r_1/r_2)^{r_1/2}}{\Gamma(r_1/2) \Gamma(r_2/2)} \frac{1}{(1+r_1 x/r_2)^{(r_1+r_2)/2}}, 0 < x < \infty$	$Y = \frac{U/r_1}{V/r_2}; U \sim \chi^2_{(r_1)}, V \sim \chi^2_{(r_2)}$	F (r_1, r_2)
Cauchy (θ, λ)	$(\pi \lambda)^{-1} [1 + (x - \theta)/\lambda^2]^{-1}, \quad \lambda > 0.$	$Y = \frac{U}{V}; U \sim N(0,1), V \sim N(0,1);$ $\theta = 0, \lambda = 1$	$t(1)^\dagger$
Weibull (c, α, ξ_0)	$c \alpha^{-1} [(x - \xi_0)/\alpha]^{c-1} e^{-\left(\frac{x-\xi_0}{\alpha}\right)^c}, \quad c > 0, \alpha > 0, x > \xi_0.$	$Y = \left[\frac{X - \xi_0}{\alpha}\right]^c, \quad Y > 0$	$\ddot{U}stel(1)^\dagger$
		$Y = -c \cdot \log\left(\frac{X}{\alpha}\right), \xi_0 = 0$	<i>Gumbel:</i> $e^{-y} e^{-e^y}^\dagger$
		$Y = X^c, \theta = \alpha^{-c} \sim Gama(\delta, r)$ ($\theta > 0; \delta > 0, r > 0$)	<i>Pareto</i> †
Erlang (β, γ)	$F1 - e^{-\left(\frac{x}{\beta}\right)} \left[\sum_{i=0}^{\gamma-1} \frac{(x/\beta)^i}{i!} \right], \quad \beta > 0, \gamma > 0, 0 \leq x < \infty$	$Y = X, \quad \gamma = 1.$	$\ddot{U}stel(\beta)$

* Hogg, R. V.-Craig, A. T. (1970:179).

† Johnson, N. L.-Kotz, S. (1970:160; 250; 266; 266).

3.5 Sürekli Rassal Değişgen Dağılımlarından Rasgele Örnekleme

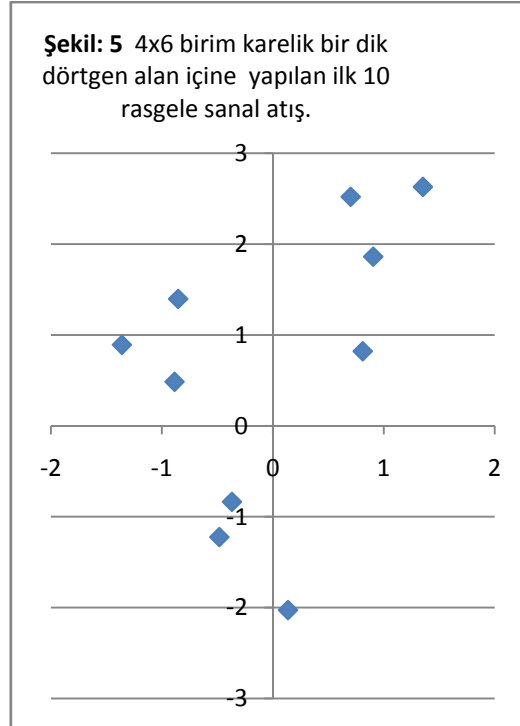
Örnek:1 a, b birim karelik dikdörtgen bir alan içine rasgele atışların koordinatları, koordinat düzleminin başlangıcı söz konusu dikdörtgen alanın merkezi olmak üzere;

$$X \sim td\left(-\frac{a}{2}, \frac{a}{2}\right); Y \sim td\left(-\frac{b}{2}, \frac{b}{2}\right)$$

biçiminde tanımlı *birbirinden bağımsız* iki tek düze rassal değişgendir.

$$a=4, b=6, r_i^* = [10^4 \cdot 23 \cdot r_{i-1}^* \bmod(10^4 + 1)] / 10^4]$$

olmak üzere, Şekil:5'deki böyle bir dik dörtgen alan içine yapılan ilk on rasgele sanal atışa ilişkin koordinatların doğrudan ters dönüşüm yöntemine göre hesaplanması Çizelge:9'da görüldüğü gibidir.



Çizelge:9 4X6 birim karelik bir dik dörtgen alan içine yapılan ilk on rasgele sanal atışın koordinatları.

i	r_i^X	$x_i = 4r_i^X - 2$	r_i^Y	$y_i = 6r_i^Y - 3$
0	0,1234	-	0,4321	-
1	0,8382	1,3528	0,9383	2,6298
2	0,2786	-0,8856	0,5809	0,4854
3	0,4078	-0,3688	0,3607	-0,8358
4	0,3794	-0,4824	0,2961	-1,2234
5	0,7262	0,9048	0,8103	1,8618
6	0,7026	0,8104	0,6369	0,8214
7	0,1598	-1,3608	0,6487	0,8922
8	0,6754	0,7016	0,9201	2,5206
9	0,5342	0,1368	0,1623	-2,0262
10	0,2866	-0,8536	0,7329	1,3974

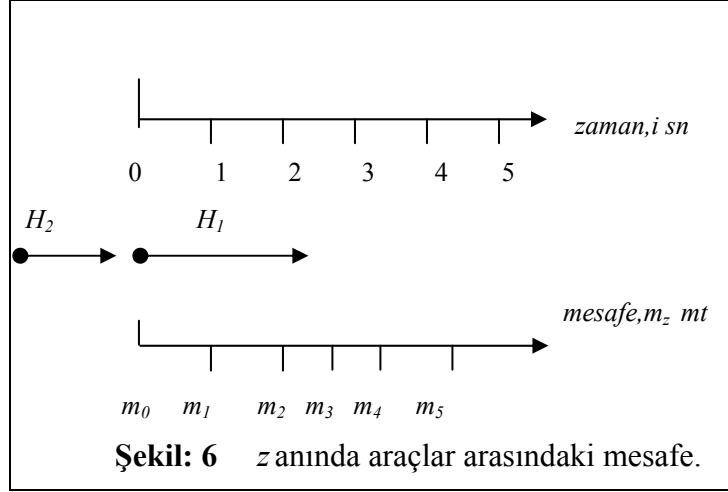
Örnek:2 Bir yarış pistindeki iki araçtan öndekinin hızı $H_1 \sim td(90,120)$ km/saat, arkadakinin hızı ise $H_2 \sim td(80,130)$ km/saat'dır. Yarış başladıktan sonra her iki aracın ortalama hızlarına çıktıkları bir noktada, birinci aracın ikinci araçtan 20 metre olduğu gözlenmiştir. 5 saniye sonra aralarındaki mesafe ne olabilir?

Her iki aracın da ortalama hızlarına çıktıkları başlangıç noktasından sonra, her saniyede bir hız değiştirdikleri varsayımı altında, z. saniye sonunda aralarındaki mesafenin,

$$m_i = 0,27(h_{1i} - h_{2i}), \quad i = 1(1)z$$

$$M_z = m_0 + \sum_{i=1}^z m_i \blacksquare$$

biçiminde ifade edilebileceği, Şekil:6 yardımı ile görülebilir.



Çizelge:10’da verilen birim tek-düze değışgen değerlerine göre, Çizelge:6’da verilen doğrudan ters dönüşüm yöntemiyle $H_1 \sim td(90,120)km/saat$ ve $H_2 \sim td(80,130)km/saat$ dağılımlarından rasgele örnekleme değerleri, m_i ve M_z tanımlarında yerlerine konarak, *bir deneme sonunda*, araçların arasındaki mesafenin 11,778 metre olacağı bulunur.

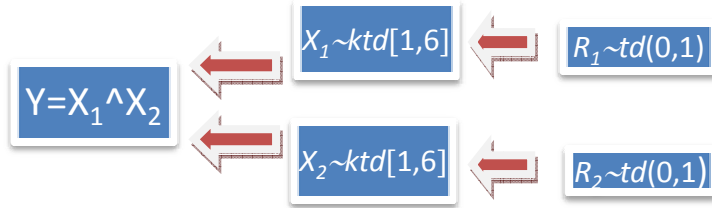
Çizelge:10 $H_1 \sim td(90,120)km/saat$ ve $H_2 \sim td(80,130)km/saat$ hızda giden iki araç arasındaki mesafenin benzetimi.

i, z	$r_i^{H_1}$	h_{1i}	$r_i^{H_{2i}}$	h_{2i}	m_i	M_z
0	0,167239	-	0,023156	-	20	20
1	0,846494	115,395	0,532588	106,629	2,435	22,435
2	0,469343	104,080	0,249512	92,76	3,223	25,658
3	0,794879	113,846	0,738771	116,939	-0,859	24,799
4	0,282199	98,66	0,991717	129,586	-8,644	16,155
5	0,490571	104,717	0,809469	120,473	-4,377	11,778

3.6 Kesikli ve Deneyisel Dağılımlardan Rasgele Örnekleme

Örnek:1 Olgusal dünyada mutlak hilesiz bir yoktur. Ancak, olasılık dağılımı, olan hilesiz bir zarın sanal gözlem değerleri, Çizelge:6'daki tanıma göre kolayca üretilebilir.

Hilesiz iki zarla oynan bir şans oyunu, aşağıdaki gibi tasarlanabilir.



ve olmak üzere, olasılık dağılımları verilen iki rassal değişgen cinsinden biçiminde tanımlanan bir rassal değişgene ilişkin ilk on sanal gözlem değerlerinin elde edilişi Çizelge:11'de verildiği gibidir.

Çizelge:11 Hilesiz iki zarla oynan bir şans oyununda ilk on sanal deneme sonucu.

i	r_{1i}		r_{2i}		
0	0,015432	-	0,278901	-	-
1	0,354936	3	0,414723	3	27
2	0,163528	1	0,538629	4	1
3	0,761144	5	0,388467	3	125
4	0,506312	4	0,934741	6	4096
5	0,645176	4	0,499043	3	64
6	0,839048	6	0,477989	3	216
7	0,298104	2	0,993747	6	64
8	0,856392	6	0,856182	6	46656
9	0,697016	5	0,692186	5	3125
10	0,031368	1	0,920273	6	1

Örnek:2 Hileli bir zarın deneysel dağılımı, Çizelge:12a'daki gibi olsun. Hileli zarın modeli olarak her hangi kuramsal bir dağılım önerilemeyeceği için bu hileli zarın sanal değerleri, verilen deneysel dağılımdan Algoritma:2'de tanımlanan rasgele örnekleme ile elde edilebilir. Bu hileli zarın, sanal ortamda 5 kere atılması sonucu gözlenen sanal değerleri Çizelge:12.b'nin son sütununda görülmektedir.

Çizelge: 12 Hileli Bir Zarın Benzetimi.

*a. Hileli Bir Zarın
Deneysel Dağılımı.*

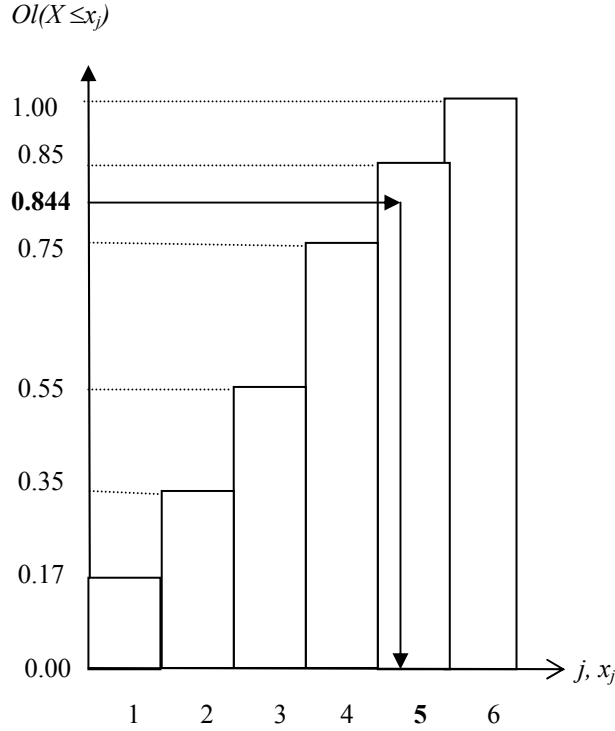
j	a_j	$O(X \leq a_j) = o_j$
1	1	0.17
2	2	0.35
3	3	0.55
4	4	0.75
5	5	0.85
6	6	1.00

*b. Deneysel Dağılımdan Ters Dönüşüm Yöntemi ile Rasgele
Örnekleme.*

i	r_i	$o_{j-1} \leq r_i < o_j$	j	$x_i = a_j$
1	0.844	$0.75 \leq r_i < 0.85$	5	5
2	0.417	$0.35 \leq r_i < 0.55$	3	3
3	0.596	$0.55 \leq r_i < 0.75$	4	4
4	0.709	$0.55 \leq r_i < 0.75$	4	4
5	0.306	$0.17 \leq r_i < 0.35$	2	2

Bu sanal değerlerin elde edilmesinde kullanılan birim tek-düze değişgen değerleri Çizelge:12b'nin ikinci sütununda ondalıktan sonra üç haneye kadar verilmiştir. Çizelge:12b'nin üçüncü sütununda, i . deneme için belirlenen r_i değerinin birikimli olasılık eksenindeki hangi olasılık aralığında olduğu; dördüncü sütunda bu aralığın sıra numarası; ve son sütunda da, birikimli

olasılık eksenindeki aralık sıra numarasına karşı gelen X rassal değişkeninin ters dönüşümle elde edilen sanal değerleri görülmektedir. $r_1=0.844$ değeri, sıra numarası $j=5$ olan $0.75 \leq r_1 < 0.85$ birikimli olasılık değer aralığına düştüğünden, ters dönüşüm yöntemi birinci denemede hileli zarın sanal gözlem değerini $x_1=a_5=5$ olarak verir. Aynı sanal değerın ters dönüşüm yöntemi ile elde edilişi grafiksel olarak Şekil:7’de gösterilmiştir.



Şekil:7 Hileli bir zarın birikimli olasılık diyagramı ve ilk atış sonucunun ters dönüşümle elde edilişi.

Algoritma:2’ye göre, bu örnekteki ilk sanal değerin elde edilişinde izlenen adımlar Çizelge:13’de sıralanmıştır. Döngü içindeki $r_1 < o_j$ koşulu $j=5$ olduğunda sağlandığından, $x_1 = 5$ sanal değeri $x_1=a_5 \leftarrow j=5$ dönüşümü ile elde edilmiştir.

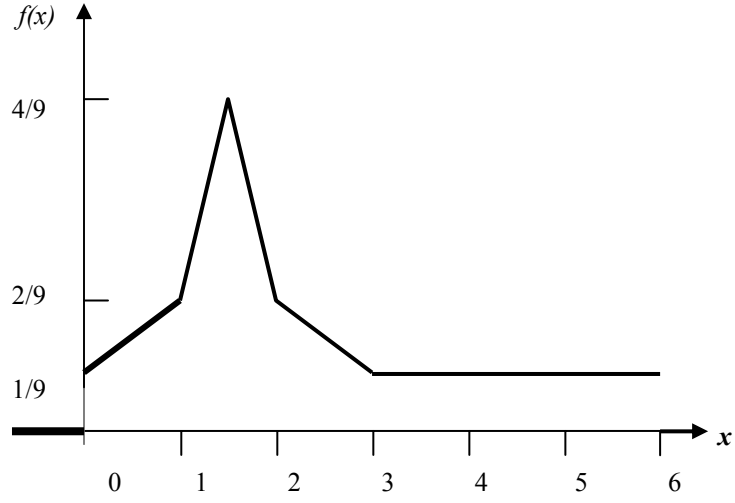
Çizelge:13 Algoritma:2'ye göre hileli zara ilişkin ilk sanal gözlem değeri.

j	$r_j < o_j$	$x_j = a_j$
1	$0.844 > 0.17$	-
2	$0.844 > 0.35$	-
3	$0.844 > 0.55$	-
4	$0.844 > 0.75$	-
5	$0.844 < 0.85$	$x_1 = a_5 = 5$

Örnek: 3 Bir durakta otobüs bekleme süresinin olasılık yoğunluk işlevi,

$$f(x) = \begin{cases} 0, & x < 0 \\ (x+1)/9, & 0 \leq x < 1 \\ (4x-2)/9, & 1 \leq x < 1.5 \\ (10-4x)/9, & 1.5 \leq x < 2 \\ (4-x)/9, & 2 \leq x < 3 \\ 1/9, & 3 \leq x < 6 \\ 0, & 6 \leq x \end{cases}$$

olsun. Verilen olasılık yoğunluk işlevinin grafiği Şekil:8a'daki gibidir.



Şekil:8.a Otobüs bekleme süresinin olasılık yoğunluk işlevi .

Bir yolcunun, sürelerde bekleme olasılıkları ise,

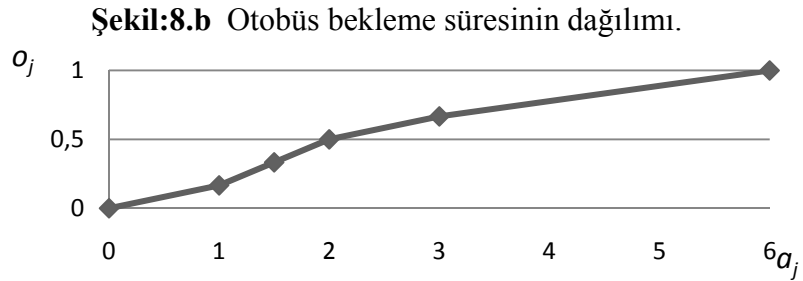
$$Ol(a_j \leq x < a_{j+1}) = \int_{a_j}^{a_{j+1}} f(x)dx$$

tanımına göre hesaplanabilir. Bu bekleme sürelerinin hesaplanan olasılıkları ve birikimli olasılıkları Çizelge: 6'daki düzenlemeye uygun olarak Çizelge:14'de verilmiştir.

Çizelge:14 Bekleme sürelerinin olasılıkları ve birikimli olasılıkları.

j	a_j	$Ol(a_{j-1} \leq X < a_j)$	$Ol(X \leq a_j) = o_j$
0	0	0	0,000
1	1	3/18	0,166
2	1.5	3/18	0,333
3	2	3/18	0,500
4	3	3/18	0,660
5	6	6/18	1,000

Bekleme süresinin dağılımı $F(x)$ 'in grafiğinin Şekil:8.b'de gösterildiği biçimde, (0,0), (1,3/18), (1.5,6/18), (2,9/18), (3,12/18), (6,1) noktalarını birleştiren 5 doğru parçasından oluşacağı görülebilir.



Bu bilgilerin ışığında bekleme süresinin kesikli dağılım işlevi,

$$F(x) = \begin{cases} (3/18)x, & 0 \leq x < 1; \\ (3/18)(2x-1), & 1 \leq x < 1.5; \\ (3/18)(2x-1), & 1.5 \leq x < 2; \\ (3/18)(x+1), & 2 \leq x < 3; \\ (2/18)(x+3), & 3 \leq x < 6. \end{cases}$$

biçiminde ve dağılım fonksiyonunun tersi de,

$$F^{-1}(r) = \begin{cases} 6r, & 0 \leq r < 3/18; \\ 3r+1/2, & 3/18 \leq r < 6/18; \\ 3r+1/2, & 6/18 \leq r < 9/18; \\ 6r-1, & 9/18 \leq r < 12/18; \\ 9r-3, & 12/18 \leq r < 1. \end{cases}$$

olarak yazılabilir.

Elde edilen ters dağılım fonksiyonu, birikimli olasılık eksenindeki beş aralığın her biri için farklı dönüşüm içermektedir. Bu dönüşümlere, ilişkin oldukları aralık sıra numaraları ile erişilebilir. Deneysel dağılımlardan ters dönüşüm tekniği ile rasgele örnekleme yaparken olduğu gibi, Birim tek-düze rassal değişken değeri r_i verilmişken, önce Algoritma:2'den yararlanarak aralık sıra numarası j ; ve sonra da, j . ters dönüşümden sanal gözlem değeri elde edilir. Çizelge:15'de, verilen birim tek-düze rassal değişken değerlerinden, Algoritma:2'de tanımlanan genelleştirilmiş ters dönüşümle bekleme süresinin sanal gözlem değerlerinin nasıl hesaplanacağı gösterilmiştir.

Çizelge:15 Bekleme Süresinin Kesikli Dağılımından Ters Dönüşüm Yöntemi ile Rasgele Örnekleme.

i	r_i	$o_{i-1} \leq r_i < o_i$	j	$x_i = F_j^{-1}(r_i)$
1	0.844	$12/18 \leq r_i < 1$	5	$9(0.844)-3=4.596$
2	0.417	$6/18 \leq r_i < 9/18$	3	$3(0.417)+0.5=1.751$
3	0.596	$9/18 \leq r_i < 12/18$	4	$6(0.596)-1=2.576$