Journal of İstanbul Kültür University
2006/3 pp. 129-133

LOJİSTİK HARİTA İLE RASGELE SAYI ÜRETİLMESİ VE İSTATİSTİKİ YÖNTEMLERLE SINANMASI

Fatih ÖZKAYNAK[1] ve A. Bedri ÖZER[2]

ÖZET

Rasgelelik günlük yaşamda insanlığı kuşatmıştır. Günümüzde rasgele sayılara şifreleme, güvenli haberleşme vb. birçok alanda ihtiyaç duyulmaktadır. Bunun için sahte rasgele sayıları üreten algoritmalar kullanılmaktadır. Ancak adından da anlaşılacağı gibi bu sayılar kestirilebilirler, çok güvenli değillerdir ve kaçınılmaz olarak rakamların kendilerini tekrar etmeleri önemli bir sorun oluşturmaktadır. Bu çalışmada kestirilebilirliğin ve tekrarlama frekansının azaltılması için rasgele sayı üretiminde kaotik bir sistem olan lojistik harita kullanılmıştır. Kullanılan yöntemin geçerliliği Ki-Kare ve Kolmogorov-Simirnov testleri ile sınanmıştır. Yapılan testler sonucu sahte rasgele sayıları üreten diğer yöntemlere göre daha iyi sonuçlar gözlenmiş ve önerilen yöntemin geçerliliği bu testlerle gösterilmiştir.

1. GİRİŞ

Bilgisayar derleyicilerinde ve sistemlerde rasgele sayılara ihtiyaç duyulmaktadır. Bu sayıları elde etmek için rasgele sayı üreteçleri kullanılmaktadır. Kullanılan bu üreteçlerin aslında sahte rasgele (pseudo random) sayı ürettikleri bilinmektedir [1]. Üreteçlerde kullanılan yöntemler bilindiğinden bu sayıların kestirilebilirliği mümkündür. Güvenlikle ilgili sistemlerde ve şans oyunlarında bu istenmeyen bir durumdur. Sayıların kestirilememesi için farklı arayışlar devam etmektedir. Sahte rasgele sayılarla ilgili bir diğer problemde tekrarlama frekansının yüksekliğidir. Aynı sayılara birden fazla defa rastlanmaması güvercin deliği teoremine göre mümkün değildir. Örneğin 0 ile 1000 arası rasgele sayı üreten bir üreteçten 100'den fazla sayı üretilirse aynı sayıların tekrar edilmesi kaçınılmazdır. Bu yüzden rasgele sayı üretiminde önemli olan mümkün olduğunca sayıların az tekrarlanmasıdır yani tekrarlama frekansının azaltılmasıdır.

Gelişigüzel ve tahmin edilemez davranış gösteren sistemlere kaotik sistemler denir. Kaosu tanımlayan yegane özellik, başlangıç şartlarına bağımlılıktır. Kaotik sistem, birbirine çok yakın iki noktadan bırakıldığında sistemin izleyeceği yörüngeler tamamen birbiriyle ilişkisizdir, ve bu iki yörünge gittikçe birbirinden uzaklaşır [2,3]. Kaosun en belirtici özelliği olan başlangıç şartlarına bağımlılığı, ilk olarak Poincare ortaya atmıştır. Henri Poincare bu olguyu söyle açıklamıştır: '...başlangıç şartlarındaki çok küçük farklar, sonuç olgusunda çok büyük farklar meydana getirmektedir. Başlangıçtaki küçük hata, daha sonra devasa bir hatayı meydana getirmektedir. Kestirim imkansız hale gelmekte ve rastlantısal bir olguya sahip olmaktayız'.

Poincare'nin de değindiği gibi kaotik sistemlerde başlangıç şartlarına bağlı olarak kestirim imkansız hale gelmektedir. Bu özellik, kaotik sistemlerin rasgele sayı üreteci olarak kullanılabileceği fikrini doğurmaktadır. Kaotik bir sisteme farklı başlangıç şartları verildiğinde kestirim imkansız olacağı aşikardır. Böylece sahte rasgele sayı üreten üreteçlerin dezavantajlarından biri olan kestirilebilirlik yok edilmiştir.

[1] Fırat Üniversitesi, Bilgisayar Mühendisliği Bölümü, (0424) 237000 (5026), (0424) 2415526, ozkaynak_fatih@hotmail.com
[2] Fırat Üniversitesi, Bilgisayar Mühendisliği Bölümü, (0424) 237000 (5026), (0424) 2415526, bedriozer@firat.edu.tr

Önerilen kaotik sayı üretecinin tekrarlama frekansı da önemlidir. Tekrarlama frekansının test edilmesi için istatistiki yöntemler mevcuttur. Bu çalışmada kaotik rasgele sayı üreteci olarak basitliğinden dolayı lojistik harita kullanılmıştır. Fakat diğer kaotik sistemlerle de rasgele sayı üretimi gerçekleştirilebilir. Yöntemin sınanması için ise Ki-Kare testi kullanılmıştır. Yine bu çalışmada test sonuçları sahte rasgele sayı üreten üreteçlerle kıyaslanmıştır.

2. RASGELE SAYI ÜRETİM YÖNTEMLERİ

Rasgele sayı üretmek için birçok çalışma yapılmıştır [4]. Bunlardan en çok bilinenleri Orta Kare ve Lineer Congruental algoritmalarıdır. Orta kare algoritması 0000'dan 9999'a kadar rasgele sayılar üretmek için herhangi bir dört basamaklı sayı seçip, bu sayının karesi alınır. Bu sayının ortasındaki dört basamağı alınarak ilk basamak tek basamaklı rasgele bir sayı, ilk iki basamak iki basamaklı rasgele bir sayıdır. Bu şeklinde devam ederek yeni sayılar elde edilir [4]. Orta kare algoritması önceleri kabul edilebilir idi. Ancak, rakamların kaçınılmaz olarak kendilerini tekrar etmeleri önemli bir sorun oluşturur.

Lineer Congruental algoritması Lehmer tarafından tekrarlama frekansını azaltmak için geliştirilmiştir [5]. Bu metot, aynı zamanda rasgele sayı üretmek için kullanılan yöntemlerin en iyi bilinenidir. Çekirdek (başlangıçtaki x_n değeri) keyfi bir sayı içermek üzere

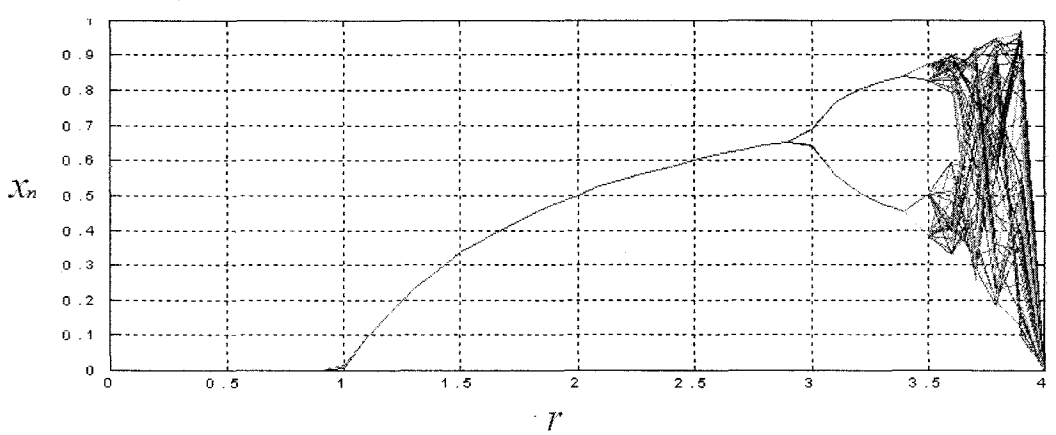
$$x_{n+1} = ax_n + c(\mathrm{mod}\, m)$$

denklem rasgele bir sayı elde etmek için bir önceki rasgele sayı a gibi bir sabit ile çarpılıp, c gibi bir sayı eklenip m gibi diğer bir sabit ile bölümünden kalan sayı hesaplanır. Bu sayının 0 ile m-1 arasında olduğu açıktır. Bu yöntem her ne kadar basit görünüyorsa da iyi sonuçlar elde etmek için a, c ve m parametrelerin seçimine dikkat edilmelidir.

Lojistik harita kaosun gösterilmesi için ayrık zamanlı bir dinamik sistemdir. Sınırlı bir çevredeki popülasyonu modellemek için geliştirmiştir [3]. Lojistik harita bir boyutludur ve doğrusal değildir. Model:

$$x_{n+1} = rx_n(1 - x_n)$$

şeklindedir. Bu model, r parametresinin alacağı değere göre kaotik yada kaotik olmayan davranış gösterecektir. Sistemin hangi doğrusal olmayan parametre değerleri için, kaosa girip girmediğini anlamak için çatallaşma diyagramı kullanılmaktadır. Lojistik harita için çatallaşma diyagramı aşağıda verilmiştir.



r parametresinin 1 ile 3 arasındaki değerlerinde sistem kaotik değildir ve tek bir denge noktası vardır, sistem bu noktaya yaklaşarak orada sabit kalır. r parametresinin 3 ile 3.5 arasındaki değerleri için iki denge noktası vardır ve herhangi birine yaklaşarak orada sabit kalır. Fakat 3.5 değerinden sonra sistem kaosa girmiştir ve davranışı kestirilemez [3].

3. RASGELELİĞİN TEST EDİLMESİ

Rasgele sayı üreteçlerinin, ürettiği rasgele sayı dizilerinin rasgelelik tanımına uygun olup olmadığının test edilmesi gereklidir. Ancak bu testlerin hiçbiri tamamen üretilen dizilerin rasgeleliğe uygunluğu ispat etmez. Verilen testleri geçen bir dizi rasgele olmayan bir özellik taşıyabilir. Bu nedenle yapılan testler sadece belirli bir anlamlılık düzeyinde önerilen yöntemin red edilip edilmeyeceği sonucuna varır.

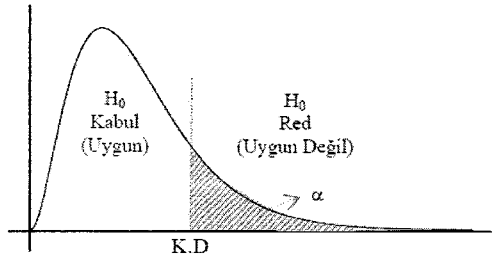
Yığın üzerinde alınan örneklerden yığınla ilgili kararlar verilmesi beklenir. Bu kararlara istatistiksel kararlar denir. Yeni bir ilacın hastalığın tedavisinde etkili olup olmadığı, bir madeni paranın hileli olup olmadığı veya üretilen sayıların rasgele olup olmadığı örnek verilerine dayanarak karar verilebilir. Kararlara varmak için verilen yığınla ilgili varsayımlar yapılması yararlı olmaktadır. Bu şekilde doğru olan yada olmayan varsayımlarda "istatistiksel hipotezler" olarak adlandırılır. Hipotez testleri parametrik ve parametrik olmayan testler olarak sınıflanmaktadır. Bu çalışmada rasgeleliği test etmek için parametrik bir test olan Ki-Kare, parametrik olmayan test içinse Kolmogorov-Simirnov testi kullanılmıştır.

4. Kİ KARE VE KOLMOGOROV-SİMİRNOV TESTİ

Ki-Kare testinde; gözlenen frekansların (o_i), belli bir hipoteze göre elde edilen beklenen frekanslardan (e_i) önemli ölçüde fark edip etmediği araştırılır.

$$\chi^2 = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \ldots + \frac{(o_k - e_k)^2}{e_k} = \sum_{j=1}^{k} \frac{(o_j - e_j)^2}{e_j}$$

$\chi^2 = 0$ ise gözlenen ve beklenen frekanslar tam uyumludur; $\chi^2 > 0$ ise frekanslar tam uyumlu değildir. χ^2 arttıkça, gözlenen ve beklenen frekanslar arasındaki farklılık artar. Kritik değer (K.D), α önem seviyesi ve s.d = m - 1 - r serbestlik derecesine göre hazırlanmış χ^2 kritik değerler tablosundan belirlenir. Burada m tahmin edilen parametre sayısıdır. Hesaplanan χ^2 değeri kritik değerler tablosundaki değerden daha küçükse hipotez kabul edilir veya en azından red edilmez.



Ki-Kare testinden daha güvenilir sonuçlar almak için bazı noktalara dikkat edilmelidir. Kategori sayısı ikiden fazla ise (r>2) her bir beklenen frekans beş veya daha büyük olmalıdır. Küçük beklenen frekanslarla çalışılması hesaplanan χ^2 ye büyük katkı yapacağından hipotezin red edilme olasılığını arttırır.

χ^2 testinin uygulanabilmesi için beklenen frekansların 5'den büyük olması istenir. Kolmogorov-Simirnov testi böyle bir şarta dayanmadığı için kolayca uygulanabilmektedir. Test gözlenen ve beklenen değerlerin kümülatif nispi frekansları arasındaki mutlak fark ile

belirlenmektedir. Hesaplanan maksimum mutlak fark kritik değerden büyük ise hipotez red edilir.

5. ÖNERİLEN YÖNTEMİN TESTİ

Başlangıçta çekirdek değeri 0.3 alınıp r parametresine 3.5 ile 4 arasında farklı değerler verilerek 0 ile 1 aralığında rasgele sayılar üretilebilir. Bu çalışmada r=3.8 değeri için 0 ile 1 arasında üretilen 1000 sayının son iki basamağı alınarak 0-99 arasında rasgele sayılar üretilmiştir. Test sonucu:

$$\chi^2 = \frac{(8-10)^2}{10} + \frac{(13-10)^2}{10} + \frac{(10-10)^2}{10} + \ldots + \frac{(8-10)^2}{10} + \frac{(10-10)^2}{10} = 98.4$$

100 kategori olduğundan serbestlik derecesi s.d = 100-1 = 99 dur. Bu serbestlik derecesinde çeşitli güven değerleri için P değerleri aşağıda verilmiştir.

Kİ-KARE TESTİ				
Güven Değerleri	0.05	0.025	0.1	0.005
P	124.342	129.561	135.807	140.169

Yapılan çalışmada hesaplanan değer 98.4 dür. Bu değer tüm güven değerlerindeki ki-kare değerlerinden daha küçük çıkmıştır.

Kolmogorov-Simirnov testi için ise gözlenen ve beklenen frekanslar arasındaki farkların mutlak değerleri ve kritik değerler aşağıdaki tablolarda gösterilmiştir.

Kategori	o_i	Nispi o_i	F_0	e_i	Nispi e_i	F_e	$\lvert F_0-F_e \rvert$
0	8	8/1000=0.008	0.008	10	10/1000=0.01	0.01	0.002
1	13	13/1000=0.013	0.021	10	10/1000=0.01	0.02	0.007
2	10	10/1000=0.01	0.031	10	10/1000=0.01	0.03	0.001
.	.	..	..	.	.	.	.
99	10	10/1000=0.01	1.000	10	10/1000=0.01	1.00	0.00
Toplam	1000	1		1000			

Test sonucu hesaplanan D değeri 0.027'dir. Bu değer aşağıdaki kritik değerlerden daha küçük olduğu için tüm güven aralıklarında hipotez uygundur.

Kolmogorov-Smirnov Tablosu					
Örnek Boyutu	D Değerleri				
	0.20	0.10	0.05	0.02	0.01
100	0.10563	0.12067	0.13403	0.14987	0.16081

6. SONUÇ

Bu çalışmada rasgele sayı üretimi için kaotik bir sistem olan lojistik harita kullanılmıştır. Kaotik sistemler kestirilemediğinden dolayı bu görev için uygundurlar. Yapılan çalışmada üretilen sayılar, istatistiksel bir yöntemler olan Ki-Kare ve Kolmogorov-Smirnov testlerine tabii tutulmuş ve tatmin edici sonuçlar elde edilmiştir

KAYNAKLAR

[1] P. L'Ecuyer, Random number generation, In Handbook of Simulation Principles, Methodology, Advances, Applications and Practice, 1998

[2] S. H. Strogatz, Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering, Cambridge Pres, 1994

[3] N. Baykal, T. Beyan, Bulanık Mantık Uzman Sistemler ve Denetleyiciler, Bıçaklar kitabevi, 2004

[4] Robinson, Stephanie, Dessart, Donald, Random-number generators: a mysterious use of algorithms, 1998 p. 243-50

[5] M. Law, W. Kelton, Simulation Modeling and Analysis, 2000