به نام خدا

تولید اعداد تصادفی

تعریف

- ارقام تصادفی: به طور یکنواخت روی مجموعه $\{0,1,\ldots,9\}$ توزیع می شود. مانند ارقام تصادفی موجود در جدول پ-۱ کتاب
- اعداد تصادفی: به طور یکنواخت و تصادفی در فاصله $[0,1]$ توزیع می شوند.
- مقادیر تصادفی: از یک توزیع آماری مشخص پیروی می کنند. مثلاً توزیع نرمال با میانگین μ و واریانس σ^2

Table A.1 Random Digits

94737	08225	35614	24826	88319	05595	58701	57365	74759
87259	85982	13296	89326	74863	99986	68558	06391	50248
63856	14016	18527	11634	96908	52146	53496	51730	03500
66612	54714	46783	61934	30258	61674	07471	67566	31635
30712	58582	05704	23172	86689	94834	99057	55832	21012
69607	24145	43886	86477	05317	30445	33456	34029	09603
37792	27282	94107	41967	21425	04743	42822	28111	09757
01488	56680	73847	64930	11108	44834	45390	86043	23973
66248	97697	38244	50918	55441	51217	54786	04940	50807
51453	03462	61157	65366	61130	26204	15016	85665	97714
92168	82530	19271	86999	96499	12765	20926	25282	39119
36463	07331	54590	00546	03337	41583	46439	40173	46455
47097	78780	04210	87084	44484	75377	57753	41415	09890
80400	45972	44111	99708	45935	03694	81421	60170	58457
94554	13863	88239	91624	00022	40471	78462	96265	55360
31567	53597	08490	73544	72573	30961	12282	97033	13676
07821	24759	47266	21747	72496	77755	50391	59554	31177
09056	10709	69314	11449	40531	02917	95878	74587	60906
19922	37025	80731	26179	16039	01518	82697	73227	13160
29923	02570	80164	36108	73689	26342	35712	49137	13482
29602	29464	99219	20308	82109	03898	82072	85199	13103
94135	94661	87724	88187	62191	70607	63099	40494	49069
87926	34092	34334	55064	43152	01610	03126	47312	59578
85039	19212	59160	83537	54414	19856	90527	21756	64783
66070	38480	74636	45095	86576	79337	39578	40851	53503
78166	82521	79261	12570	10930	47564	77869	16480	43972
94672	07912	26153	10531	12715	63142	88937	94466	31388
56406	70023	27734	22254	27685	67518	63966	33203	70803
67726	57805	94264	77009	08682	18784	47554	59869	66320
07516	45979	76735	46509	17696	67177	92600	55572	17245
43070	22671	00152	81326	89428	16368	57659	79424	57604
36917	60370	80812	87225	02850	47118	23790	55043	75117
03919	82922	02312	31106	44335	05573	17470	25900	91080
46724	22558	64303	78804	05762	70650	56117	06707	90035
16108	61281	86823	20286	14025	24909	38391	12183	89393
74541	75808	89669	87680	72758	60851	55292	95663	88326
82919	31285	01850	72550	42986	57518	01159	01786	98145
31388	26809	77258	99360	92362	21979	41319	75739	98082
17190	75522	15687	07161	99745	48767	03121	20046	28013
00466	88068	68631	98745	97810	35886	14497	90230	69264

اعداد تصادفی

- یکی از ویژگیهای هر مدل شبیه سازی تغییرات تصادفی است.

برای مثال در شبیه سازی سیستم های صف، ورودها و مدت زمان خدمت دهی می توانند تصادفی باشند.

- برای نمونه گیری از توزیع های آماری و تولید مقادیر تصادفی نیاز به منبعی از اعدادتصادفی است.
- لذا اعداد تصادفی هم بصورت مستقیم برای نمونه گیری از توزیع یکنواخت در فاصله ۰ و ۱ و هم برای تولید مقادیر تصادفی استفاده می شود.

خواص اعداد تصادفی

هر دنباله از اعداد تصادفی مانند $R_1, R_2, \ldots$ باید دو خاصیت آماری مهم داشته باشد.

- توزیع احتمال یکنواخت در بازه $[0,1]$
- استقلال هر عدد از دیگری

تابع چگالی این گونه متغیر تصادفی بصورت زیر تعریف می شود.

$$f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & otherwise \end{cases}$$

روشهای تولید اعداد تصادفی

- کنار هم قرار دادن ارقام تصادفی به تعداد مناسب و نوشتن ممیز در سمت چپ عدد بدست آمده.
- استفاده از جداول ارقام تصادفی مانند جدول پ-۱
- تولید اعداد تصادفی با استفاده از شیوه های از پیش تعیین شده که در این فصل در مورد تعدادی از آنها صحبت می شود.

اعداد تصادفی

در هنگام تولید اعداد تصادفی ممکن است مسائلی در خصوص نقض خواص استقلال یا توزیع احتمال یکنواخت بروز نماید که در زیر به تعدادی از آنها اشاره می شود.

- اعداد تصادفی تولید شده ممکن است توزیع یکنواخت نداشته باشند.
- میانگین اعداد تصادفی تولید شده ممکن است بیش از حد بزرگ یا بیش از حد کوچک باشد.
- واریانس اعداد تصادفی تولید شده ممکن است تفاوت قابل ملاحظه ای از مقدار متعارف داشته باشند.

اعداد تصادفی

- ممکن است دنباله اعداد تولید شده تغییراتی تناوبی از خود نشان دهد. مثلاً:
  - وجود همبستگی بین اعداد
  - روندهای صعودی یا نزولی بین اعداد مجاور
  - وجود چند عدد بزرگتر از میانگین و به دنبال آن چند عدد کوچکتر از میانگین

برای پی بردن به این مشکلات آزمون هایی جهت بررسی استقلال و یکنواخت بودن داده ها وجود دارند که بعداً در مورد آنها صحبت می شود.

ملاحظات در بکارگیری روشهای تولید اعداد تصادفی

- روش یا الگوریتم تولید اعداد تصادفی باید سریع باشد.

محاسبات برای تولید چند عدد تصادفی چندان وقت گیر نیست اما اگر تعداد زیادی عدد تصادفی نیاز داشته باشیم (معمولاً به این صورت است) روش مورد استفاده باید به گونه ای باشد که تعداد زیادی عدد تصادفی را در زمان کوتاهی تولید نماید.

Basics of Simulation

ملاحظات در بکارگیری روشهای تولید اعداد تصادفی

- روش مورد استفاده نباید به مقدار زیادی حافظه کامپیوتر نیاز داشته باشد.

از آنجا که اجرای شبیه سازی به نگهداری اطلاعات مربوط به رویداد ها، فعالیت ها و ... نیاز دارد و تولید اعداد تصادفی تنها یک بخش از برنامه کامپیوتری می باشد لذا استفاده از روش هایی که حافظه کمتری نیاز دارند ضروری به نظر می رسد.

A.Ghaderi
University of Kurdistan

ملاحظات در بکارگیری روشهای تولید اعداد تصادفی

- طول دنباله اعداد تولید شده باید به اندازه کافی بلند باشد.

منظور از طول دنباله تعداد اعدادی است که بدون تکرار در آن قرار می گیرند.

- اگر طول دنباله اعداد تولید بلند نباشد ممکن است اعداد تصادفی معینی به طور مکرر تولید شوند که در این صورت نشانه از هم پاشیدن الگوریتم است.

ملاحظات در بکارگیری روشهای تولید اعداد تصادفی

- اعداد تصادفی تولید شده به وسیله یک الگوریتم باید در صورت نیاز تکرار پذیر باشند. یعنی با تعریف شرایط اولیه بتوان اعداد تصادفی خاصی را به دفعات تولید نمود.

تولید اعداد تصادفی یکسان برای مقایسه دو سیستم یا بررسی اثر تغییرات در یک سیستم مورد استفاده قرار می گیرد.

- از همه مهمتر اعداد تصادفی تولید شده دارای خواص آماری توزیع یکنواخت و استقلال باشند

روشهای مختلف تولید اعداد تصادفی

Linear Congruential Method(LCM) ۱- مولد همنهشتی خطی

- این روش جزء روشهای اصلی برای تولید اعداد تصادفی می باشد.

- *To produce a sequence of integers, X_1, X_2, … between 0 and m-1 by following a recursive relationship:*

$$X_i = (aX_{i-1} + c) \ mod \ m \quad i = 1,2,\ldots$$

ضریب — مقدار ثابت — پیمانه

- *The selection of the values for a, c, m, and X_0 drastically affects the **statistical properties** and the **cycle length**.*

روشهای مختلف تولید اعداد تصادفی

تعریف همنهشتی:

$$a = b \ mod \ m \quad or \quad a \overset{m}{\equiv} b$$

عبارت فوق یعنی باقیمانده (b-a) بر m مساوی صفر است و یا به عبارت دیگر، باقیمانده تقسیم b بر m یا a بر m با هم برابر است

$$77 = 377 \ mod \ 100 \quad or \quad 77 \overset{100}{\equiv} 377$$

روشهای مختلف تولید اعداد تصادفی

Linear Congruential Method(LCM) ۱- مولد همنهشتی خطی

- *The random integers are being generated [0,m-1], and to convert the integers to random numbers:*

$$R_i = \frac{X_i}{m}, \quad i = 1,2,\ldots$$

- اگر **c>0** باشد مولد را مولد همنهشتی آمیخته،

Mixed Congruential Method

- در صورتیکه **c=0** باشد، مولد بالا را مولد همنهشتی ضربی می نامند.

Multiplicative Congruential Method

روشهای مختلف تولید اعداد تصادفی

فرض کنید داشته باشیم: $X_0 = 27, \quad a = 17, \quad c = 43, \quad m = 100$

$$X_i = (aX_{i-1} + c) \bmod m$$

چون عدد پیمانه (**m**) ۱۰۰ است، لذا مقادیر **X** بدست آمده همواره بین ۰ و ۹۹ می باشند. چرا؟

باقیمانده تقسیم ۵۰۲ بر ۱۰۰

$$X_0 = 27$$

$$X_1 \equiv [17(27) + 43)] \quad \bmod 100 = 502 \quad \bmod 100 = 2 \quad R_2 = \frac{2}{100} = 0.02$$

$$X_2 \equiv [17(2) + 43)] \quad \bmod 100 = 77 \quad \bmod 100 = 77 \quad R_2 = \frac{77}{100} = 0.77$$

$$X_3 \equiv [17(77) + 43)] \quad \bmod 100 = 1352 \quad \bmod 100 = 52 \quad R_2 = \frac{52}{100} = 0.52$$

روشهای مختلف تولید اعداد تصادفی

توجه داشته باشید که اعداد بدست آمده به مجموعه زیر تعلق دارد.

$$I = \left\{ 0, \frac{1}{m}, \frac{2}{m}, \ldots, \frac{m-1}{m} \right\}$$

واضح است که برای داشتن اعداد یکنواخت در بازه ۰ و ۱ عدد **m** باید بزرگ باشد تا هیچ شکاف بزرگی در بازه صفر و یک باقی نماند.

- بهتر است مقدار **a** از رابطه زیر تعیین گردد.

$$a = \sqrt{m}$$

مولد همنهشتی خطی Linear Congruential Method(LCM)

i	X_i $X_0=1$	X_i $X_0=2$	X_i $X_0=3$	X_i $X_0=4$
1	1	2	3	4
2	13	26	39	52
3	41	18	59	36
4	21	42	63	20
5	17	34	51	4
6	29	58	23	
7	57	50	43	
8	37	10	47	
9	33	2	35	
10	45		7	
11	9		27	
12	53		31	
13	49		19	
14	61		55	
15	25		11	
16	5		15	
17	1		3	

- Use $a = 13$, $c = 0$, and $m = 64$
- The period of the generator is very low

طول دوره بسیار کوتاه

Characteristics of a Good Generator

[LCM]

- Maximum Density
  - Such that he values assumed by R_i, $i = 1,2,\ldots$, leave **no large gaps on [0,1]**
  - Problem: Instead of continuous, each R_i is discrete
  - Solution: a **very large integer for modulus m**
    - Approximation appears to be of little consequence
- Maximum Period
  - To achieve maximum density and avoid cycling.
  - Achieve by: **proper choice of a, c, m, and X_0**.
- Most digital computers use a binary representation of numbers
  - Speed and efficiency are aided by a modulus, **m, to be (or close to) a power of 2**., Values such as 2^{48} or 2^{31}-1 are in common use in generators appearing in many simulation languages.

۲- مولد همنهشتی خطی ترکیبی
Combined Linear Congruential Generators

- Reason: **Longer period** generator is needed because of the increasing **complexity of stimulated systems**.
- Approach: Combine two or more multiplicative congruential generators.
- Let $X_{i,1}, X_{i,2}, \ldots, X_{i,k}$, be the ith output from k different multiplicative congruential generators.
  - The jth generator:
    - Has prime modulus m_j and multiplier a_j and period is m_{j-1}
    - Produces integers $X_{i,j}$ is approx ~ Uniform on integers in [1, m-1]
    - $W_{i,j} = X_{i,j} - 1$ is approx ~ Uniform on integers in [1, m-2]

۲- مولد همنهشتی خطی ترکیبی
Combined Linear Congruential Generators

- Suggested form:

$$X_i = \left(\sum_{j=1}^{k} (-1)^{j-1} X_{i,j} \right) mod\ m_1 - 1 \qquad Hence,\ R_i = \begin{cases} \dfrac{X_i}{m_1}, & X_i \succ 0 \\ \dfrac{m_1 - 1}{m_1}, & X_i = 0 \end{cases}$$

The coefficient: Performs the subtraction $X_{i,1-1}$

- The maximum possible period is:

$$P = \frac{(m_1 - 1)(m_2 - 1)...(m_k - 1)}{2^{k-1}}$$

۲- مولد همنهشتی خطی ترکیبی
Combined Linear Congruential Generators

- Example: For 32-bit computers, L'Ecuyer [1988] suggests combining $k = 2$ generators with $m_1 = 2{,}147{,}483{,}563$, $a_1 = 40{,}014$, $m_2 = 2{,}147{,}483{,}399$ and $a_2 = 40{,}692$. The algorithm becomes:

  Step 1: Select seeds
  - $X_{1,0}$ in the range [**1**, (2,147,483,563)] for the 1st generator
  - $X_{2,0}$ in the range [1, (2,147,483,399)] for the 2nd generator.

  Step 2: For each individual generator,
  $X_{1,j+1} = 40{,}014\ X_{1,j}$ mod 2,147,483,563
  $X_{2,j+1} = 40{,}692\ X_{1,j}$ mod 2,147,483,399.

  Step 3: $X_{j+1} = (X_{1,j+1} - X_{2,j+1})$ mod 2,147,483,562.

  Step 4: Return

$$R_{j+1} = \begin{cases} \dfrac{X_{j+1}}{2{,}147{,}483{,}563}, & X_{j+1} > 0 \\ \dfrac{2{,}147{,}483{,}562}{2{,}147{,}483{,}563}, & X_{j+1} = 0 \end{cases}$$

  Step 5: Set $j = j+1$, go back to step 2.

- Combined generator has period: $(m_1 - 1)(m_2 - 1)/2 \sim 2 \times 10^{18}$

آزمون های مربوط به یکنواخت بودن اعداد تصادفی

- همانطور که گفته شد یکی از خواص اعداد تصادفی یکنواخت بودن آن در بازه صفر- یک است. بعد از تولید اعداد به کمک یکی از مولد تولید اعداد تصادفی (عموماً روش همنهشتی خطی) باید از خاصیت یکنواخت بودن آن مطمئن شد. برای آزمون یکنواخت بودن اعداد بدست آمده دو روش مورد بررسی قرار می گیرد.
  - آزمون مربع کای
  - آزمون کالموگروف- اسمیرنف

Tests for Random Numbers

- *Two categories:*
  - *Testing for uniformity:*

    H_0: $R_i \sim U[0,1]$

    H_1: $R_i \not\sim U[0,1]$
    - *Failure to reject the null hypothesis, H_0, means that evidence of non-uniformity has not been detected.*
  - *Testing for independence:*

    H_0: $R_i \sim$ *independently*

    H_1: $R_i \not\sim$ *independently*
    - *Failure to reject the null hypothesis, H_0, means that evidence of dependence has not been detected.*
- *Level of significance α, the probability of rejecting H_0 when it is true:* $\alpha = P(\text{reject } H_0 | H_0 \text{ is true})$

Tests for Random Numbers

- *When to use these tests:*
  - *If a **well-known simulation languages** or **random-number generators** is used, it is probably unnecessary to test*
  - *If the **generator** is **not explicitly** known or documented, e.g., **spreadsheet programs**, symbolic/numerical calculators, tests should be applied to many sample numbers.*

Frequency Tests

- Test of uniformity
- Two different methods:
  - Kolmogorov-Smirnov test
  - Chi-square test

*Both tests are based on the null hypothesis of no significant difference between the **sample distribution**(generated random Numbers) and the **theoretical uniform distribution**.*

Chi-square test

[Frequency Test]

- *Chi-square test uses the sample statistic:*

n is the # of classes

$$X_0^2 = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i}$$

e_i is the expected # in the ith class

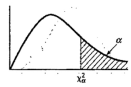
o_i is the observed # in the ith class

  - *Approximately the **chi-square distribution with n-1 degrees** of freedom (where the critical values are tabulated in Table A.6)*
  - *For the uniform distribution, e_i, the expected number in the each class is:*

$$e_i = \frac{N}{n}, \quad \text{where } N \text{ is the total \# of observation}$$

- *Valid only for large samples, e.g. N >= 50*

Table A.6 Percentage Points of The Chi-Square Distribution with ν Degrees of Freedom



ν	$\chi^2_{0.005}$	$\chi^2_{0.01}$	$\chi^2_{0.025}$	$\chi^2_{0.05}$	$\chi^2_{0.10}$
1	7.88	6.63	5.02	3.84	2.71
2	10.60	9.21	7.38	5.99	4.61
3	12.84	11.34	9.35	7.81	6.25
4	14.96	13.28	11.14	9.49	7.78
5	16.7	15.1	12.8	11.1	9.2
6	18.5	16.8	14.4	12.6	10.6
7	20.3	18.5	16.0	14.1	12.0
8	22.0	20.1	17.5	15.5	13.4
9	23.6	21.7	19.0	16.9	14.7
10	25.2	23.2	20.5	18.3	16.0
11	26.8	24.7	21.9	19.7	17.3
12	28.3	26.2	23.3	21.0	18.5
13	29.8	27.7	24.7	22.4	19.8
14	31.3	29.1	26.1	23.7	21.1
15	32.8	30.6	27.5	25.0	22.3
16	34.3	32.0	28.8	26.3	23.5
17	35.7	33.4	30.2	27.6	24.8
18	37.2	34.8	31.5	28.9	26.0
19	38.6	36.2	32.9	30.1	27.2
20	40.0	37.6	34.2	31.4	28.4
21	41.4	38.9	35.5	32.7	29.6
22	42.8	40.3	36.8	33.9	30.8
23	44.2	41.6	38.1	35.2	32.0
24	45.6	43.0	39.4	36.4	33.2
25	49.6	44.3	40.6	37.7	34.4
26	48.3	45.6	41.9	38.9	35.6
27	49.6	47.0	43.2	40.1	36.7
28	51.0	48.3	44.5	41.3	37.9
29	52.3	49.6	45.7	42.6	39.1
30	53.7	50.9	47.0	43.8	40.3
40	66.8	63.7	59.3	55.8	51.8
50	79.5	76.2	71.4	67.5	63.2
60	92.0	88.4	83.3	79.1	74.4
70	104.2	100.4	95.0	90.5	85.5
80	116.3	112.3	106.6	101.9	96.6
90	128.3	124.1	118.1	113.1	107.6
100	140.2	135.8	129.6	124.3	118.5

Source: Robert E. Shannon, *Systems Simulation: The Art and Science*, © 1975, p. 372. Reprinted by permission of Prentice Hall, Upper Saddle River, NJ.

آزمون مربع کای

مثال: فرض کنید یک نمونه ۱۰۰ تایی از اعداد تصادفی تولید شده توسط یک مولد بصورت زیر باشد. به ازای $\alpha = 0.05$ فرض $H_0 : R \sim U[0,1]$ را با استفاده از آزمون مربع کای آزمون کنید.

0.34	0.90	0.25	0.89	0.87	0.44	0.12	0.21	0.46	0.67
0.83	0.76	0.79	0.64	0.70	0.81	0.94	0.74	0.22	0.74
0.96	0.99	0.77	0.67	0.56	0.41	0.52	0.73	0.99	0.02
0.47	0.30	0.17	0.82	0.56	0.05	0.45	0.31	0.78	0.05
0.79	0.71	0.23	0.19	0.82	0.93	0.65	0.37	0.39	0.42
0.99	0.17	0.99	0.46	0.05	0.66	0.10	0.42	0.18	0.49
0.37	0.51	0.54	0.01	0.81	0.28	0.69	0.34	0.75	0.49
0.72	0.43	0.56	0.97	0.30	0.94	0.96	0.58	0.73	0.05
0.06	0.39	0.84	0.24	0.40	0.64	0.40	0.19	0.79	0.62
0.18	0.26	0.97	0.88	0.64	0.47	0.60	0.11	0.29	0.78

آزمون مربع کای

آماره آزمون

$$\sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i} \sim \chi^2_{\alpha,n-1}$$

- تفکیک بازه [0,1] به ۱۰ بازه مساوی به صورت [0,0.1]، [0.1,0.2] و ... . بنابراین، n=10
- شمارش تعداد مشاهدات اتفاق افتاده در هر بازه، یعنی o_i
- محاسبه متوسط تعداد مشاهدات در هر بازه، یعنی e_i=100/10=10
- نکته: پیشنهاد شده است در استفاده از این آزمون، هیچ یک از e_i ها کمتر از ۵ نباشد.

آزمون مربع کای

بازه	o_i	e_i	$o_i - e_i$	$(o_i - e_i)^2$	$(o_i - e_i)^2 / e_i$
۱	۷	۱۰	-۳	۹	۰.۹
۲	۹	۱۰	-۱	۱	۰.۱
۳	۸	۱۰	-۲	۴	۰.۴
۴	۹	۱۰	-۱	۱	۰.۱
۵	۱۴	۱۰	۴	۱۶	۱.۶
۶	۷	۱۰	-۳	۹	۰.۹
۷	۱۰	۱۰	۰	۰	۰.۰
۸	۱۵	۱۰	۵	۲۵	۲.۵
۹	۹	۱۰	-۱	۱	۰.۱
۱۰	۱۲	۱۰	۲	۴	۰.۴
جمع	۱۰۰	۱۰۰	۰		۷.۰

آزمون مربع کای

ناحیه بحرانی → $\chi^2_{\alpha,n-1} = \chi^2_{0.05,9} = 16.9$

از آنجا که عدد بدست آمده (۷) از مقدار ناحیه بحرانی کمتر است لذا فرض صفر را نمی توان رد کرد. یعنی اعداد تصادفی داده شده دارای توزیع یکنواخت در بازه صفر – یک هستند.

Kolmogorov-Smirnov Test [Frequency Test]

- *Compares the continuous cdf, F(x), of the uniform distribution with the empirical cdf, $S_N(x)$, of the N sample observations.*
  - *We know:* $F(x) = x, \quad 0 \le x \le 1$
  - *If the sample from the RN generator is $R_1, R_2, \ldots, R_N$, then the empirical cdf, $S_N(x)$ is:*

$$S_N(x) = \frac{\text{number of } R_1, R_2, \ldots, R_n \text{ which are} \le x}{N}$$

- *Based on the statistic:* $D = \max| F(x) - S_N(x)|$
  - *Sampling distribution of D is known (a function of N, tabulated in Table A.8.)*

Table A.8 Kolmogorov–Smirnov Critical Values

Degrees of Freedom (N)	$D_{0.10}$	$D_{0.05}$	$D_{0.01}$
1	0.950	0.975	0.995
2	0.776	0.842	0.929
3	0.642	0.708	0.828
4	0.564	0.624	0.733
5	0.510	0.565	0.669
6	0.470	0.521	0.618
7	0.438	0.486	0.577
8	0.411	0.457	0.543
9	0.388	0.432	0.514
10	0.368	0.410	0.490
11	0.352	0.391	0.468
12	0.338	0.375	0.450
13	0.325	0.361	0.433
14	0.314	0.349	0.418
15	0.304	0.338	0.404
16	0.295	0.328	0.392
17	0.286	0.318	0.381
18	0.278	0.309	0.371
19	0.272	0.301	0.363
20	0.264	0.294	0.356
25	0.24	0.27	0.32
30	0.22	0.24	0.29
35	0.21	0.23	0.27
Over 35	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

آزمون کالموگروف – اسمیرنف

مراحل آزمون برای بررسی یکنواخت بودن توزیع نمونه داده شده:

- مقادیر را بصورت صعودی مرتب کنید. R_i را iامین عدد کوچک مشاهدات باشد. لذا، $R_{(1)} \le R_{(2)} \le \cdots \le R_{(N)}$
- مقدار i/N را به ازای هر i حساب کنید. (N تعداد اعداد تصادفی مورد آزمون می باشد.)
- مقدار آماره D^+ را بصورت زیر حساب کنید. $D^+ = \max_{1 \le i \le N} \left\{ \frac{i}{N} - R_{(i)} \right\}$

آزمون کالموگروف – اسمیرنف

- مقدار $\frac{i-1}{N}$ را حساب کنید.
- مقدار D^- را بصورت زیر بدست آورید.

$$D^- = \max_{1 \le i \le N} \left\{ R_{(i)} - \frac{i-1}{N} \right\}$$

- آماره آزمون را در انتها بصورت زیر حساب نمایید.

$$D = \max\left\{ D^+, D^- \right\}$$

- با توجه به اندازه نمونه داده شده، یعنی **N**، و سطح آزمون داده شده و جدول، مقدار بحرانی یعنی D_α را تهیه نمایید.
- اگر $D \ge D_\alpha$ ، آنگاه فرض صفر یعنی یکنواخت بودن توزیع مشاهدات رد می شود.

آزمون کالموگروف – اسمیرنف

فرض کنید پنج عدد تصادفی ۰.۴۴ ، ۰.۸۱ ، ۰.۱۴ ، ۰.۰۵ و ۰.۹۳ موجود باشند. به ازای $\alpha = 0.05$ با استفاده از آزمون کالموگروف – اسمیرنف نشان دهید که داده های موجود دارای توزیع یکنواخت هستند یا خیر؟

آزمون کالموگروف – اسمیرنف

i	1	2	3	4	5
$R_{(i)}$	0.05	0.14	0.44	0.81	0.93
$\frac{i}{N}$	0.20	0.40	0.60	0.80	1.00
$\frac{i}{N} - R_{(i)}$	0.15	0.26	0.16	-	0.07
$\frac{i-1}{N}$	0	0.20	0.40	0.60	0.80
$R_{(i)} - \frac{i-1}{N}$	0.05	-	0.04	0.21	0.13

آزمون کالموگروف – اسمیرنف

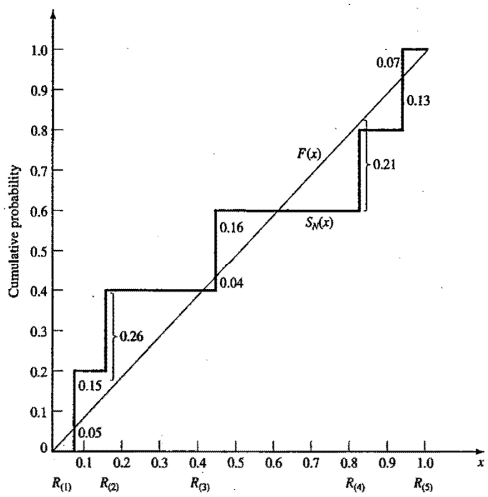


Figure 7.2 Comparison of $F(x)$ and $S_N(x)$.

آزمون کالموگروف – اسمیرنف

- بنابراین مقدار آماره آزمون برابر است با:

$$D = max\{\ D^+, D^-\} = max\{0.26, 0.21\} = 0.26$$

- از طرفی برای $N=5$ و $\alpha = 0.05$ مقدار بحرانی برابر خواهد بود با:

$$D_{0.05} = 0.565$$

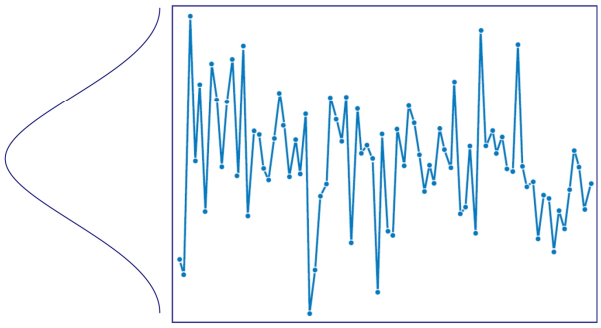
- لذا، نمی توان فرض توزیع احتمال یکنواخت صفر-یک را برای مقادیر داده شده در بالا رد نمود.

آزمون های مربوط به استقلال اعداد تصادفی

- همانطور که پیشتر نیز اشاره شد اعداد تصادفی تولید شده به کمک یک مولد علاوه بر اینکه باید دارای توزیع یکنواخت در بازه صفر – یک باشند، بایستی از یکدیگر مستقل نیز باشند.
- برای پی بردن به خاصیت استقلال اعداد تصادفی تولید شده، می توان آنها را به ترتیبی که تولید شده اند بر روی یک نمودار رسم نمود. چنانچه داده ها روند غیر متعارفی نداشته باشند (روندهای صعودی یا نزولی، روند های سینوسی و ...) می توان پذیرفت که اعداد تولید شده از هم مستقلند.

استقلال اعداد تصادفی

- منظور از استقلال زمانی مشاهدات این است که مشاهدات در طول زمان به یکدیگر وابسته نباشند. به عبارت دیگر



آزمون های مربوط به استقلال اعداد تصادفی

برای اطمینان از اینکه اعداد تولید شده خاصیت مستقل بودن را دارند، آزمونهای مختلفی وجود دارند که در ادامه به دو نمونه از آزمونهای روند و آزمون همبستگی اشاره می شود.

۱- تعداد روندهای صعودی و نزولی

۲- تعداد روندهای بزرگتر و کوچکتر از میانگین.

۳- آزمون همبستگی **Autocorrelation**

Basics of Simulation

آزمون های مربوط به استقلال اعداد تصادفی

تعریف

روند (دنباله): مجموعه ای از مشاهدات یکسان.

طول روند: تعداد مشاهدات موجود در هر روند.

- آزمایش پرتاب یک سکه را در نظر بگیرید. در ۱۰ بار پرتاب یک سکه نتایج زیر ممکن است حاصل شود.

A.Ghaderi
University of Kurdistan

آزمون های مربوط به استقلال اعداد تصادفی

H,T,T,H,H,T,T,T,H,T

تعداد روند ها: ۶

طول روند ها:

روند یک ، ۱مرتبه

روند دو ، ۲ مرتبه

روند سه ،۲ مرتبه

...

آزمون های مربوط به استقلال اعداد تصادفی

اساس آزمون روند را تعداد روند ها یا طول آنها تشکیل می دهد.

تعداد روندها را می توان بر اساس ضابطه صعودی یا نزولی بودن روند تعریف کرد.

در هر روند صعودی، هر عدد از عدد سمت راست خود کوچکتر است.

در هر روند نزولی، هر عدد از عدد سمت راست خود بزرگتر است.

آزمون های مربوط به استقلال اعداد تصادفی

$0.87^{-}\ 0.15^{+}\ 0.23^{+}\ 0.45^{+}\ 0.69^{-}\ 0.32^{-}\ 0.30^{-}\ 0.19^{+}\ 0.24$

$^{-}\ 0.18^{+}\ 0.65^{+}\ 0.82^{+}\ 0.93^{-}\ 0.22^{+}\ 0.81$

$-+++---+-+++-+$

تعداد روندها = ۸

طول روندها: ۱،۳،۳،۱،۱،۳،۱،۱

آزمون های مربوط به استقلال اعداد تصادفی

تعداد روندها ممکن است بیش از حد کم یا زیاد باشد.

در اعداد زیر یک روند آن هم روند صعودی وجود دارد.

۰.۰۲ ۰.۰۸ ۰.۱۸ ۰.۲۳ ۰.۳۶ ۰.۴۲ ۰.۵۵ ۰.۶۳ ۰.۷۲ ۰.۸۹

تولید چنین دنباله ای به وسیله یک مولد چندان محتمل نیست.

آزمون های مربوط به استقلال اعداد تصادفی

دنباله زیر دارای ۹ روند است که ۵ تای آنها صعودی و ۴ تای آنها نزولی است.

۰.۵۷ ۰.۳۶ ۰.۷۹ ۰.۲۸ ۰.۸۴ ۰.۲۶ ۰.۹۶ ۰.۱۵ ۰.۹۳ ۰.۰۸

احتمال وجود این تعداد روند نیز در یک دنباله ده تایی از اعداد تصادفی نیز چندان زیاد نیست.

آنچه بیشتر احتمال دارد اینست که تعداد روندها بین دو حد فوق قرار گیرد.

- در یک دنباله **n** تایی از اعداد تصادفی تعداد بیشترین روند برابر **n-1** و کمترین تعداد روندها برابر یک خواهد بود. چرا؟

آزمون های مربوط به استقلال اعداد تصادفی

۱- تعداد روندهای صعودی و نزولی

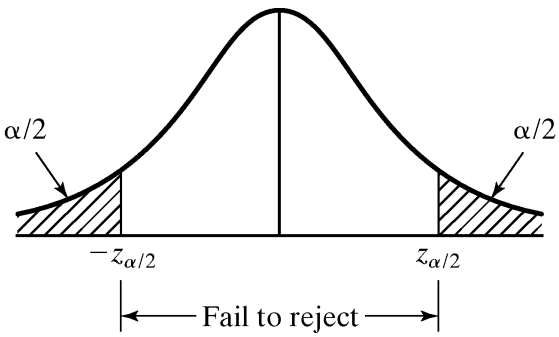
اگر متغیر تصادفی **a** معرف مجموع تعداد روندهای موجود در یک دنباله **n** تایی از اعداد تصادفی باشد، میانگین و واریانس آن را بصورت زیر نشان می دهیم.

$$\mu_a = \frac{(2n-1)}{3} \qquad \sigma_a^2 = \frac{(16n-29)}{90}$$

اگر **n>20** باشد متغیر تصادفی **a** به سمت توزیع نرمال میل می کند.

$$a \sim N[\mu_a, \sigma_a^2] \quad \Rightarrow \quad Z_0 = \frac{a-\mu_a}{\sigma_a} = \frac{a-[(2n-1)/3]}{\sqrt{(16n-29)/90}}$$

Normal Hypothesis Test



اگر $-Z_{\alpha/2} \le Z \le Z_{\alpha/2}$ ⇒ فرض صفر را نمی توان رد کرد.

آزمون های مربوط به استقلال اعداد تصادفی

۱ – تعداد روندهای صعودی و نزولی

به ازای $\alpha = 0.05$ تعیین کنید آیا دنباله متشکل از ۴۰ عدد تصادفی زیر را بر اساس تعداد روندهای صعودی و نزولی می توان دارای خاصیت استقلال دانست یا نه؟

۰٫۲۷	۰٫۳۶	۰٫۶۲	۰٫۵۵	۰٫۹۱	۰٫۷۴	۰٫۹۴	۰٫۸۹	۰٫۶۸	۰٫۴۱
۰٫۲۸	۰٫۱۶	۰٫۳۶	۰٫۰۱	۰٫۰۲	۰٫۵۴	۰٫۰۸	۰٫۷۵	۰٫۷۲	۰٫۱۹
۰٫۵۳	۰٫۸۲	۰٫۳۲	۰٫۲۳	۰٫۴۷	۰٫۱۸	۰٫۶۹	۰٫۹۵	۰٫۰۱	۰٫۱۸
۰٫۲۹	۰٫۶۳	۰٫۵۷	۰٫۱۳	۰٫۴۵	۰٫۸۳	۰٫۰۴	۰٫۷۳	۰٫۴۲	۰٫۳۱

توالی روندهای صعودی و نزولی به شرح زیر است

+ + + − + − + − − − + + − + − − + − + − −

+ − − + − + + − − + + − + − − + + −

تعداد روندها: ۲۶

n= 40

a= 26

آزمون های مربوط به استقلال اعداد تصادفی

۱- تعداد روندهای صعودی و نزولی

$$\mu_a = \frac{2(40)-1}{3} = 26.33 \qquad \sigma_a^2 = \frac{16(40)-29}{90} = 6.79$$

$$Z_0 = \frac{26-26.33}{\sqrt{6.79}} = -0.13$$

$$Z_{0.025} = 1.96$$

فرض استقلال اعداد تصادفی را نمی توان رد کرد.

آزمون های مربوط به استقلال اعداد تصادفی

۲- تعداد روندهای بزرگتر و کوچکتر از میانگین

دنباله زیر را در نظر بگیرید.

۰٫۶۷	۰٫۸۷	۰٫۹۳	۰٫۸۳	۰٫۹۴	۰٫۵۲	۰٫۸۱	۰٫۷۹	۰٫۷۲	۰٫۶۳
۰٫۸۶	۰٫۸۲	۰٫۹۵	۰٫۷۴	۰٫۷۷	۰٫۸۸	۰٫۵۵	۰٫۸۹	۰٫۸۳	۰٫۵۴
۰٫۳۴	۰٫۳۶	۰٫۲۷	۰٫۱۸	۰٫۱۹	۰٫۰۸	۰٫۱۸	۰٫۳۶	۰٫۳۲	۰٫۴۳
۰٫۴۱	۰٫۴۷	۰٫۳۹	۰٫۲۵	۰٫۳۵	۰٫۴۶	۰٫۴۳	۰٫۴۹	۰٫۴۵	۰٫۳۱

توالی روندهای صعودی و نزولی این دنباله به‌شرح زیر است

+ + + − + − + − − − + + − + − − + − + − − + − −

+ − + + − − + + − + − − + +−

آزمون های مربوط به استقلال اعداد تصادفی

۲- تعداد روندهای بزرگتر و کوچکتر از میانگین

این دنباله مانند دنباله مثال قبل است. پس از آزمون قبلی موفق بیرون می آید و استقلال آن پذیرفته می شود.

در این دنباله ۲۰ عدد اول همگی بزرگتر از میانگین 0.495=[(0.99+0)/2] و ۲۰ عدد دیگر کوچکتر از میانگین اند.

- ✓ احتمال رخداد این حالت بسیار ناچیز است.

آزمون های مربوط به استقلال اعداد تصادفی

۲- تعداد روندهای بزرگتر و کوچکتر از میانگین

برای انجام آزمون تعداد روندهای بزرگتر و کوچکتر از میانگین، در دنباله داده شده، هر مشاهده که از میانگین بزرگتر باشد با علامت + و هر مشاهده که از میانگین کوچکتر باشد با علامت − مشخص می شود.

۰٫۴۰ ۰٫۸۴ ۰٫۷۵ ۰٫۱۸ ۰٫۱۳ ۰٫۹۲ ۰٫۵۷ ۰٫۷۷ ۰٫۳۰ ۰٫۷۱
۰٫۴۲ ۰٫۰۵ ۰٫۷۸ ۰٫۷۴ ۰٫۶۸ ۰٫۰۳ ۰٫۱۸ ۰٫۵۱ ۰٫۱۰ ۰٫۳۷

علائم مثبت و منفی برای این مثال، در قالب توالی زیر قابل عرضه است

− + + − − + + + − + − − + + + − − + − −

۱۱ روند

آزمون های مربوط به استقلال اعداد تصادفی

۲- تعداد روندهای بزرگتر و کوچکتر از میانگین

- **n1**: تعداد مشاهدات بزرگتر از میانگین
- **n2**: تعداد مشاهدات کوچکتر از میانگین
- بیشترین تعداد روندها در یک دنباله **n** تایی مساوی **n1+n2**
- کمترین تعداد روندها در یک دنباله **n** تایی مساوی یک

Basics of Simulation

آزمون های مربوط به استقلال اعداد تصادفی

۲- تعداد روندهای بزرگتر و کوچکتر از میانگین

b: تعداد روندهای بزرگتر و کوچکتر از میانگین

$$\mu_b = \frac{2n_1n_2}{n} + \frac{1}{2} \qquad \sigma_b^2 = \frac{2n_1n_2(2n_1n_2 - n)}{n^2(n-1)}$$

$$if \quad n_1 \quad or \quad n_2 > 20 \Rightarrow b \sim N[\mu_b, \sigma_b^2]$$

$$Z_0 = \frac{b - [(2n_1n_2/n) + \frac{1}{2}]}{\left(\frac{2n_1n_2(2n_1n_2 - n)}{n^2(n-1)}\right)^{1/2}}$$

A.Ghaderi
University of Kurdistan

آزمون های مربوط به استقلال اعداد تصادفی

۲- تعداد روندهای بزرگتر و کوچکتر از میانگین

−++++++−−−++−+−−−−−−−++−−−−++−−+−+−−++−

$$b = 17, \quad n_2 = 22, \quad n_1 = 18, \quad n = 40$$

$$\mu_b = \frac{2(18)(22)}{40} + \frac{1}{2} = 20.3$$

$$\sigma_b^2 = \frac{2(18)(22)[2(18)(22) - 40]}{40^2(40-1)} = 9.54$$

$$Z_0 = \frac{17 - 20.3}{\sqrt{9.54}} = -1.07$$

$$Z_{0.025} = 1.96$$

فرض استقلال اعداد تصادفی را نمی توان رد کرد.

آزمون های مربوط به استقلال اعداد تصادفی

۳- آزمون همبستگی، Autocorrelation

- Autocorrelation is concerned with dependence between numbers in a sequence
- Example:

0.12	0.01	0.23	0.28	0.89	0.31	0.64	0.28	0.83	0.93
0.99	0.15	0.33	0.35	0.91	0.41	0.60	0.27	0.75	0.88
0.68	0.49	0.05	0.43	0.95	0.58	0.19	0.36	0.69	0.87

- Numbers at 5-th, 10-th, 15-th, ... are very similar
- Numbers can be
  - Low
  - High
  - Alternating

آزمون های مربوط به استقلال اعداد تصادفی

۳- آزمون همبستگی، Autocorrelation

- Testing the autocorrelation between every m numbers (m is a.k.a. the lag), starting with the i-th number
  - The autocorrelation ρ_{im} between numbers: $R_i, R_{i+m}, R_{i+2m}, R_{i+(M+1)m}$
  - M is the largest integer such that $i+(M+1)m \leq N$
- Hypothesis:

$$H_0: \ \rho_{im} = 0, \quad \text{if numbers are independent}$$
$$H_1: \ \rho_{im} \neq 0, \quad \text{if numbers are dependent}$$

- If the values are uncorrelated:
  - For large values of M, the distribution of the estimator of ρ_{im}, denoted $\hat{\rho}_{im}$ is approximately normal.

Basics of Simulation

آزمون های مربوط به استقلال اعداد تصادفی

۳- آزمون همبستگی، Autocorrelation

- Test statistics is: $Z_0 = \dfrac{\hat{\rho}_{im}}{\hat{\sigma}_{\hat{\rho}_{im}}}$
  - Z_0 is distributed normally with mean $= 0$ and variance $= 1$, and:

$$\hat{\rho}_{im} = \frac{1}{M+1}\left[\sum_{k=0}^{M} R_{i+km} \cdot R_{i+(k+1)m}\right] - 0.25$$

$$\hat{\sigma}_{\rho_{im}} = \frac{\sqrt{13M+7}}{12(M+1)}$$

- If $\rho_{im} > 0$, the subsequence has positive autocorrelation
  - High random numbers tend to be followed by high ones, and vice versa.
- If $\rho_{im} < 0$, the subsequence has negative autocorrelation
  - Low random numbers tend to be followed by high ones, and vice versa.

A.Ghaderi
University of Kurdistan

آزمون های مربوط به استقلال اعداد تصادفی

۳- آزمون همبستگی، Autocorrelation

- Test whether the *3rd, 8th, 13th*, and so on, for the numbers on Slide 60.
  - Hence, $\alpha = 0.05$, $i = 3$, $m = 5$, $N = 30$, and $M = 4$

$$\hat{\rho}_{35} = \frac{1}{4+1}\begin{bmatrix}(0.23)(0.28)+(0.28)(0.33)+(0.33)(0.27)\\+(0.27)(0.05)+(0.05)(0.36)\end{bmatrix} - 0.25$$

$$= -0.1945$$

$$\sigma_{\hat{\rho}_{35}} = \frac{\sqrt{13(4)+7}}{12(4+1)} = 0.128$$

$$Z_0 = -\frac{0.1945}{0.1280} = -1.516$$

  - $z_{0.025} = 1.96$ hence, the hypothesis is not rejected.

خلاصه ای بر مطالب فصل هفتم

- نحوه تولید اعداد تصادفی
- آزمون هایی جهت بررسی یکنواختی اعداد تولید شده
- آزمون هایی جهت بررسی استقلال اعداد تولید شده