

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

فَسُبْحَانَ اللَّهِ الَّذِي بِيَدِهِ مَلَكُوتُ كُلِّ شَيْءٍ وَإِلَيْهِ تُرْجَعُونَ

فصل سوم

مولدهای اعداد تصادفی یکنواخت

مولدهای اعداد تصادفی

مولدهای تصادفی برای تولید اعداد تصادفی به کار میروند. از اعداد تصادفی تولید شده برای سیمولۀ پارامترهای تصادفی استفاده میشود. اعداد تصادفی تولید شده میتوانند از توزیع خاصی پیروی کنند یا اینکه دارای استقلال و یکنواختی باشند. به طور کلی میتوان مولدهای تصادفی را به دو دسته تقسیم کرد:

۱) مولدهای اعداد تصادفی غیریکنواخت

۲) مولدهای اعداد تصادفی یکنواخت: اعداد دارای خاصیت استقلال و یکنواختی هستند. هر عدد تصادفی نمونه مستقلی از یک توزیع آماری یکنواخت با پارامترهای ۰ و ۱ است. به عنوان استقلال یکنواختی و خواص توزیع آماری پیوسته میتوان چنین گفت:

الف) اگر بازۀ (۰ و ۱) به n قسمت مساوی تقسیم شود، انتظار میرود که از N مشاهده، $\frac{N}{n}$ در هر بخش قرار گیرد (یکنواختی)
ب) احتمال حصول یک مشاهده در هر بخش مستقل از سایر مشاهدات است (استقلال)

۱- مولدهای همنهشتی

دو عدد a و c را همنهشت مدول m گوئیم، اگر باقیمانده تقسیم آنها بر m برابر باشد و میگوئیم a همنهشت است با c به پیمانه m .

$$a \equiv c \pmod{m}$$

$$18 \equiv 8 \pmod{5}$$

(مثال) ۱۸ و ۸ به پیمانه ۵ هم نهشت هستند.

فرمول کلی زیر اساس مولدهای همنهشتی را تشکیل میدهد. در این فرمول اگر $c > 0$ باشد روش هم نهشتی مرکب و اگر $c = 0$ باشد، همنهشتی ضربی خواهد بود.

$$X_{i+1} = (ax_i + c) \bmod m$$

أ. مولد همنهشتی مرکب

$$X_{i+1} = (ax_i + c) \bmod m \quad i = 1, 2, \dots, c > 0$$

در این فرمول برای تولید رشته اعداد تصادفی به X_0 نیاز داریم و تقسیم $(ax_i + c)$ بر m آنقدر ادامه می یابد تا باقیمانده آن از m کمتر شود. پس عدد بعدی X_{i+1} برابر مقدار این باقیمانده میگردد لذا پروسه با X_0 شروع میشود که به آن مقدار اولیه میگویند. از نظر جبری مولد رابطه عبارت است از $X_{i+1} = ax_i + c - mk_{i+1}$ که k_i معرف بزرگترین عدد صحیح موجود در $\frac{ax_i + c}{m}$ می باشد.

مثال) با روش هم نهشتی مرکب و پارامترهای $m = 7, x_0 = 3, c = 3, a = 5$ مولدی تصادفی تولید کنید.

$$X_{i+1} = (ax_i + c) \bmod m$$

$$X_{i+1} = (5x_i + 3) \bmod 7$$

i	0	1	2	3	4	5	6	7
X_i	3	4	2	6	5	0	3	4
R_i		4	2	6	5	0	3	4

4

به فاصله تکرار اعداد تولیدی دوره مولد (r) میگویند که در این مثال $r=6$ است. اگر پارامترهای a و c تغییر کنند، دوره مولد هم تغییر میکند.

مثال) با روش هم نهشتی مرکب و پارامترهای $m = 7, x_0 = 3, c = 10, a = 18$ مولد اعداد تصادفی را ایجاد کنید.

$$X_{i+1} = (ax_i + c) \bmod m \implies X_{i+1} = (18x_i + 10) \bmod 7$$

i	0	1	2	3	4	5	6	7	8
X_i	3	1	0	3	1	0	3	1	0
R_i		1	0	3	1	0	3	1	0

در این مثال $r=3$ است یعنی دوره مولد کوتاهی دارد پس این مولد خوبی به حساب نمی آید.

مثال با روش هم نهستی مرکب و پارامترهای $m = 10, x_0 = 3, c = 7, a = 11$ مولد اعداد تصادفی را ایجاد کنید.

$$X_{i+1} = (ax_i + c) \bmod m \implies X_{i+1} = (11x_i + 7) \bmod 10$$

i	0	1	2	3	4	5	6	7	8	9	10	11
X_i	3	0	7	4	1	8	5	2	9	6	3	0
R_i		0	7	4	1	8	5	2	9	6	3	0

دوره تناوبش زیاد است پس مولد خوبی به حساب می آید.

عدد m محدوده اعداد تولیدی را مشخص میکند یعنی اعداد تولیدی میتوانند از 0 تا $m-1$ باشند، پس برای آنکه دامنه اعداد تولیدی زیاد باشد لازم است m را بزرگ انتخاب کنیم. اگر بخواهیم ماکزیمم دوره مولد را داشته باشیم یعنی $r=m$ بشود، باید اولاً a و m نسبت به هم اول باشند، ثانیاً اگر k یک عامل m است، باقیمانده تقسیم a بر k یک بشود، ثالثاً اگر α طول کلمه حافظه باشد، رابطه $(ax_i + c) < 2^\alpha$ برقرار باشد.

ادعا شده است که ضریب همبستگی زوجهای صحیح متوالی x_i و x_{i+1} در محدوده $\left[\frac{1}{a} - \left(\frac{6c}{am} \right) \left(1 - \frac{c}{m} \right) \right] \pm \frac{a}{m}$ دارد و انتخاب a به صورت $a = \sqrt{m}$ کوچکترین مقدار ضریب همبستگی پیوسته مرتبه یک بین x_i و x_{i+1} خواهد بود.

ب. مولد همنهشتی ضربی

اگر در فرمول کلی همنهشتی یعنی $X_{i+1} = (ax_i + c) \bmod m$ قرار دهیم $c = 0$ آنگاه فرمول حاصل مربوط به همنهشتی ضربی خواهد بود یعنی:

$$X_{i+1} = (ax_i) \bmod m$$

مثال با استفاده از روش همنهشتی ضربی با پارامترهای $m = 15, x_0 = 3, a = 7$ مولد اعداد تصادفی را به دست آورید.

$$X_{i+1} = (7x_i) \bmod 15$$

i	0	1	2	3	4	5
X_i	3	6	12	9	3	6
R_i		6	12	9	3	6

چون $m = 15$ پس تمام اعداد تصادفی تولید شده بین ۰ و ۱۴ خواهند بود.

روش همنهشتی ضربی از روشهای خوب و مناسب است و به آزمونهای یکنواختی و استقلال پاسخ مثبت داده است.

سوال چگونه میتوان مقادیر بهینه تر برای پارامترهای a, m و x_0 اختیار کرد؟

میتوان مولدهای دیگری هم تولید کرد که عملکرد مولد به نوع کامپیوتر انتخابی بستگی داشته باشد. به صورت زیر:

پارامتر x_0 باید فرد باشد.

$$m = 2^r \quad (r > 2)$$

$a = k \cdot 8 \pm 3$ که k یک عدد صحیح مثبت است.

دوره توالی تصادفی های تولیدی حداکثر و معادل 2^{r-2} است. برای تبدیل به یک مولد خوب تصادفی، بهتر است پارامترها را به شکل زیر انتخاب کنیم:

(۱) پارامتر m یکی بیشتر از بزرگترین عدد صحیح داخل یک کلمه کامپیوتر (word) انتخاب گردد. مثلاً اگر ۳۲ بیتی است و با ارزش ترین بیت، بیت علامت است پس بزرگترین عدد میشود:

$$2^{31} - 1 = 2147483647$$

پس مقدار m میشود: $m = 2147483648$ یعنی m میشود:

$$m = 2^{\text{wordlength}-1}$$

7

اگر m از word بزرگتر باشد، سرریز یا overflow رخ میدهد و فقط باقیمانده نامشخص داریم که در بعضی از زبانها مثل فرترن هیچ عملی صورت نمیگیرد.

(۲) پارامتر x_0 نسبت به m باید اول باشد چون m توانی از ۲ است، عدد صحیح مثبت فرد برای پارامتر x_0 در نظر گرفته میشود. چون اگر پارامتر x_0 زوج باشد تمام x_i های تولیدی زوج خواهند بود.

(۳) انتخاب صحیح a که باید نسبت به m اول باشد و a عدد فرد است. بهترین انتخاب برای a :

$$a = k.8 \pm 3$$

که در محدوده $2^{b/2}$ قرار بگیرد که در آن b تعداد بیت های کلمه ماشین است. مثلاً اگر کامپیوتر ۳۲ بیتی است طول کلمه $a = 2^{16} + 3 = 65539$ انتخاب میشود.

این روش از بهترین روشها برای تولید اعداد تصادفی است و زبانهای SIMSCRIPT و GPSS از آن استفاده میکنند.

ج. مولد همنهشتی تجمعی (۱)

در این روش با استفاده از دو عدد قبلی، عدد مورد نظر تولید میشود یعنی با شروع از X_0 و X_1 که معادله آن این چنین است:

$$X_{i+1} \equiv (X_i + X_{i-1}) \bmod m \quad i = 1, 2, \dots$$

روش مناسبی نیست چون سرعتش پایین است. به آزمونهای آماری پاسخ مثبتی نداده است و رفتار شناخته شده ای ندارد.

مثال) با روش همنهشتی تجمعی (۱) و پارامترهای $X_0 = 3$ و $X_1 = 7$ اعداد تصادفی تولید کنید. ($m=?$)

i	1	2	3	4
X_{i-1}	3	7	1	7
X_i	7	1	7	7
X_{i+1}	1	7	7	
R_i	1	7	7	

مثال در روش همنهشتی تجمعی (۱) با انتخاب خاصی به دنباله های شناخته شده ای دست می یابیم مثلاً اگر پارامترهای $X_0 = 0$ و $X_1 = 1$ باشند، آنگاه دنباله فیبوناتچی تولید میشود.

i	1	2	3	4
X_{i-1}	0	1	1	2
X_i	1	1	2	3
X_{i+1}	1	2	3	5
R_i	1	2	3	5

د. مولد همنهشتی تجمعی (۲)

با استفاده از یک سلسله n عددی و حرکت از طرفین، اعداد تصادفی تولید میشوند. رفتار شناخته شده ای ندارد و برای شروع باید رشته $x_1, x_2, \dots, x_n$ را داشته باشیم و در واقع با استفاده از جویباری از اعداد به تولید تصادفی میرسیم.

$$X_i \equiv (X_{i-1} + X_{n-i}) \bmod m, \quad i = 1, 2, \dots, n, \quad x_1, x_2, \dots, x_n$$

چون عمل ضرب نداریم بسیار سریعتر عمل میکند ولی در کل رفتارش به رشته ابتدایی وابسته است.

مثال) با روش همنهشتی تجمعی (۲) و $m=100$ و رشته ۵۷، ۳۴، ۸۹، ۸۲، ۱۶ اعداد تصادفی تولید کنید.

$$X_6 = (X_1 + X_5) \bmod 100 = 73 \bmod 100 = 73$$

$$X_7 = (X_6 + X_2) \bmod 100 = 107 \bmod 100 = 7$$

$$X_8 = (X_7 + X_3) \bmod 100 = 96 \bmod 100 = 96$$

۲- مولد میان مربعی

این روش توسط وان نیومن و متراپلیس ارائه شد. روش کار به این صورت است که ابتدا یک عدد اولیه به نام هسته (X_0) در نظر میگیریم. این عدد محور عملیات قرار میگیرد به این ترتیب که این عدد را به توان ۲ میرسانیم، بعد ارقام میانی آن را انتخاب میکنیم. مثلاً اگر هسته n رقم باشد، مربع هسته $2n$ رقم میشود، حالا برای انتخاب ارقام میانی، $\frac{n}{2}$ رقم از سمت راست و $\frac{n}{2}$ رقم از چپ را حذف میکنیم و n رقم وسط که باقی مانده است را انتخاب میکنیم و به صورت ممیزدار کوچکتر از یک به عنوان عدد تصادفی آن مرحله انتخاب میشود. آن n رقم انتخابی هم میشود هسته مرحله بعد. این روال را به دفعات تکرار میکنیم. البته انتخاب هسته اولیه مهم است چون با انتخاب نامناسب X_0 ممکن است صفرهای تولیدی انتشار یابند و عدد حاصل را به صفر میل دهند.

روش میان مربعی در کل روش کندی است. از آزمونهای موفق آماری به حساب نمی آید چون برای عملکرد مطلوب الگوریتم جهت تولید X_0 نمیتوان قانونی پیدا کرد. در ضمن اشکالاتی دارد که ممکن است در خروجی تولیدی R_i باعث بروز مشکل شود.

مثال گستره ۸ رقمی داریم پس ۴ رقم وسط را باید انتخاب کنیم. شروع از $X_0=5497$ به عنوان هسته

$$X_0 = 5497$$

$$X_0^2 = 30217009 \implies X_1 = 2170 \implies R_1 = 0.2170$$

$$X_1^2 = 04708900 \implies X_2 = 7089 \implies R_2 = 0.7089$$

$$X_2^2 = 50211396 \implies X_3 = 2113 \implies R_3 = 0.2113$$

$$X_3^2 = 04464769 \implies X_4 = 4647 \implies R_4 = 0.4647$$

...

مثال تصادفی تولیدی با روش میان مربعی با عدد محوری ۵۱۹۷ را انجام دهید. (انتشار صفرها)

$$X_0 = 5197$$

$$X_0^2 = 27008809 \implies X_1 = 0088 \implies R_1 = 0.0088$$

$$X_1^2 = 00007744 \implies X_2 = 0077 \implies R_2 = 0.0077$$

$$X_2^2 = 00005929 \implies X_3 = 0059 \implies R_3 = 0.0059$$

$$X_3^2 = 00003481 \implies X_4 = 0034 \implies R_4 = 0.0034$$

نکته: در روش میان مربعی، با انتخاب نامناسب X_0 یا تولید میانی نامناسب X_i هیچ تولیدی ممکن است نداشته باشیم. مثلاً اگر عدد محوری تولیدی صفر باشد، عملیات تولید اعداد تصادفی متوقف می‌گردد.

نکته: تولید اعداد میانی ممکن است در یک حلقه تکرار قرار بگیرد و مولد یک عدد تصادفی تکراری تولید کند. مانند مثال زیر که با رسیدن به عدد محوری ۶۵۰۰ تولید عدد تصادفی تکرار میشود و باید متوقف گردد.

$$X_0^2 = 42250000 \implies X_1 = 2500 \implies R_1 = 0.2500$$

$$X_1^2 = 06250000 \implies X_2 = 2500 \implies R_2 = 0.2500$$

$$X_2^2 = 06250000 \implies X_3 = 2500 \implies R_3 = 0.2500$$

۳- روش میان ضربی

در این روش با دو عدد با تعداد ارقام مساوی ($x_0, \dot{x}_0$) کار را شروع میکنیم. این دو عدد را در هم ضرب میکنیم بعد ارقام میانی حاصل ضرب را انتخاب میکنیم (مشابه روش میان مربعی). این عدد به صورت ممیزدار کوچکتر از ۱ به عنوان عدد تصادفی این مرحله انتخاب میشود. در کل روش کندی است و به آزمونهای آماری پاسخ مثبتی ندارد و ناهنجاری های روش میان مربعی را دارد منتها برای دو عدد محوری به جای یک عدد محوری.

مثال) با روش میان ضربی اعداد تصادفی را با هسته های $X_0 = 7229$ و $\dot{X}_0 = 2938$ ایجاد نمایید.

$$U_1 = X_0 \times \dot{X}_0 = (7229)(2938) = 21238802 \implies X_1 = 2388 \implies R_1 = 0.2388$$

$$U_2 = X_0 \times X_1 = (7229)(2388) = 17262852 \implies X_2 = 2628 \implies R_2 = 0.2628$$

$$U_3 = X_1 \times X_2 = (2388)(2628) = 06275664 \implies X_3 = 2756 \implies R_2 = 0.2756$$

۴- روش ثابت میان ضربی (مضرب ثابت)

این روش تفاوت کمی با روش میان ضربی دارد و عدد تصادفی از حاصل ضرب یک عدد در یک مقدار ثابت و سپس انتخاب ارقام میانی به دست می آید. یعنی یکی از اعداد محوری، ثابت فرض میشود پس میشود شروع از x_0 و یک ثابت k با تعداد ارقام یکسان.

$$X_{i+1} = K \cdot X_i$$

مثال) از روش مضرب ثابت با $x_0 = 7223$ و $k = 3987$ اعداد تصادفی را تولید کنید.

$$X_1 = k \times x_0 = 3987 \times 7223 = 28798101 \quad , \quad x_1 = 7981 \quad , \quad R_1 = 0.7981$$

$$X_2 = k \times x_1 = 3987 \times 7981 = 31820247 \quad , \quad x_2 = 8202 \quad , \quad R_2 = 0.8202$$

این روش، روش کندی است و به آزمونهای آماری پاسخ مثبت نداده است. همان ایرادهای روش میان ضربی را دارد و عملکردش تا حدی به ثابت k بستگی دارد.

به طور خلاصه داریم:

مولد اعداد تصادفی یکنواخت	آیا به آزمونهای آماری پاسخ مثبتی داده است؟
همنهشتی مرکب	
همنهشتی ضربی	بله
همنهشتی تجمعی (۱)	خیر
همنهشتی تجمعی (۲)	خیر
میان مربعی	خیر
میان ضربی	خیر
مضرب ثابت (ثابت میان ضربی)	خیر

۵- آزمون اعداد تصادفی

سلسله اعداد تصادفی تولیدی ۲ ویژگی دارد: یکنواختی و استقلال که این دو ویژگی باید آزمایش شده و صحتشان مشخص گردد. یعنی اینکه بررسی شود که اعداد به صورت یکنواخت توزیع شده اند و تکرار اعداد از استقلال برخوردار است و اینکه پیوستگی خاصی وجود نداشته باشد. هر دوی این ویژگی ها باید وجود داشته باشند. برای بررسی این دو ویژگی آزمونهای مختلفی وجود دارد که در ادامه با آنها آشنا میشویم.

ا. آزمون فراوانی (Frequency)

آزمون یکنواختی سلسله اعداد تصادفی را انجام میدهد. یعنی تکرار آنها بیش از آنچه انتظار می رود رخ ندهد. از روش های مناسب آن آزمون کای دو (خی دو) و آزمون کولموگروف اسمیرنوف می باشد. روش کای دو مناسب نمونه های بیش از ۵۰ و روش کولموگروف اسمیرنوف مناسب نمونه های کمتر از ۵۰ می باشد و جداول آماری خاص خود را دارد.

ب. آزمون امتداد (RUN)

آزمون استقلال سلسله اعداد تصادفی را انجام میدهد. امتداد دنباله به طرف بالا (افزایش) یا پایین (کاهش) به وسیله مقایسه مقادیر واقعی با مقادیر انتظاری است (حرکت زیر یا بالای میانه). از جدول آماری کای دو استفاده میکند.

ج. آزمون همبستگی (Autocorrelation)

آزمون استقلال سلسله اعداد تصادفی را انجام میدهد. همبستگی بین اعداد تولیدی با نمونه ها مقایسه میشود و انتظار همبستگی صفر دارد. از جدول آماری کای دو استفاده میکند.

د. آزمون فاصله - شکاف (Gap)

آزمون استقلال سلسله اعداد تصادفی را انجام میدهد. این امر را با شمارش تعداد ارقامی که بین تکرارهای یک رقم خاص ظاهر میشود، صورت میدهد. از جدول آماری کولموگروف اسمیرنوف برای مقایسه با تعداد فواصل منتظره استفاده میکند.

ه. آزمون پوکر (Poker)

آزمون استقلال سلسله اعداد تصادفی را انجام میدهد. اعداد گروه بندی شده را با هم مانند دسته پوکر بحث میکند و سپس دسته های یافت شده با دسته های انتظاری مقایسه می گردند. از جدول آماری کای دو استفاده میکند.

به طور خلاصه در جدول زیر داریم:

آزمون	آزمونی که روی سلسله اعداد تصادفی انجام می دهد	جدول آماری ای که استفاده می کند
فراوانی	یکنواختی توزیع	کای دو (χ^2) ، کولموگروف اسمیرنوف (D)
امتداد	استقلال	کای دو
همبستگی	استقلال	کای دو
فاصله-شکاف	استقلال	کولموگروف اسمیرنوف
پوکر	استقلال	کای دو

۶- یکنواختی و استقلال

یکنواختی و استقلال را میتوان به صورت زیر تعریف کرد:

الف) یکنواختی و عدم یکنواختی:

یکنواختی $H_0: R_i \sim u[0, 1]$

عدم یکنواختی $H_1: R_i \not\sim u[0, 1]$

ب) استقلال و عدم استقلال:

استقلال ، استقلال $H_0: R_i \sim$

وابستگی ، استقلال $H_1: R_i \not\sim$

با استفاده از استقلال یکنواختی یک ضریب α (احتمال) تعریف میکنیم که معمولاً مقادیر 0.01 یا 0.05 را به α نسبت

میدهم.
$$\alpha = P\left(\frac{H_0}{H_1 \text{ صحت}}\right)$$

این ضریب در واقع احتمال رد شدن آزمونی را روی یک سری اعداد مشخص میکند. مثلاً اگر ۱۰۰ سری آزمون انجام شود و $\alpha = 0.01$ باشد یعنی احتمال می‌رود که این سری اعداد از ۱ آزمون رد شود، پس به عبارت دیگر میتوان گفت با ضریب اطمینان $P = 1 - \alpha = 0.99$ به نتیجه عمل مطمئن بود. البته حتماً خروجی باید بررسی و تأیید شود یعنی اگر یک مولد تصادفی به تمامی آزمونها پاسخ مثبت داد باز دلیل تصادفی بودنش نیست چون شاید از الگوی خاصی پیروی نمیکرده که قبلاً بروز نکرده است.

۲۰ آ. آزمون فراوانی

جهت آزمون یکنواختی دو روش متفاوت وجود دارد که درجه توافق بین تفکیک نمونه ها و انتظار ما را برآورده میکند.

• روش کولموگروف-اسمیرنوف

این روش برای آزمون یکنواختی تصادفی های تولیدی استفاده میشود و مناسب نمونه های کم است ($N \leq 50$).

$$F(x) = x \quad 0 \leq x \leq 1$$

$$R_1, R_2, \dots, R_N$$

چنانچه تعداد نمونه های تصادفی تولیدی N باشد، داریم:

$$S_N(x) = \frac{\text{تعداد اعداد } R_1, R_2, \dots, R_N \text{ که کوچکتر یا مساوی } x \text{ هستند}}{N \text{ (کل نمونه)}} \quad S_N(x) \text{ عبارت است از:}$$

هرچه N بزرگتر شود، $S_N(x)$ تقریب بهتری از $F(x)$ را میدهد. این آزمون بر اساس بیشترین مقدار D استوار است. D فاصله بین $S_N(x)$ و $F(x)$ است و تابعی از N است.

$$D = \max |F(x) - S_N(x)|$$

روش:

گام ۱) مرتب کردن اعداد تصادفی تولیدی (نمونه ها) از کوچکتر به بزرگتر (فرض میکنیم $R(1)$ کوچکترین مقدار را دارد).

گام ۲) محاسبه D^+ و D^- به شرح زیر:

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

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

گام ۳) محاسبه ماکزیمم D^+ و D^- :

$$D = \max(D^+, D^-)$$

گام ۴) تعیین مقدار بحرانی $D\alpha$ از جدول آماری آزمون در سطح α و نمونه های N .

گام ۵) اگر مقدار آماری D (نمونه ها) از مقدار بحرانی $D\alpha$ بزرگتر باشد، یکنواختی نمونه ها تایید نمیشود: ($D \leq D\alpha$)

صحت بررسی: اگر $D \leq D_\alpha$ باشد یعنی اختلافی بین توزیع صحیح نمونه ها $\{R_1, R_2, \dots, R_N\}$ و توزیع یکنواخت وجود ندارد و لذا یکنواخت هستند.

مثال) فرض کنید ۵ عدد تصادفی زیر تولید شده اند، با استفاده از مراحل تست کولموگروف اسمیرنوف با سطح $\alpha = 0.05$ در رابطه با نمونه ها بررسی یکنواختی آن را آزمون کنید.

گام ۱ و ۲)

$R_{(i)}$	0.05	0.14	0.44	0.81	0.93
$\frac{i}{N}$	0.2	0.4	0.6	0.8	1
$\frac{i}{N} - R_{(i)}$	0.15	0.26	0.16	-	0.07
$R_{(i)} - \frac{i-1}{N}$	0.05	-	0.04	0.21	0.13

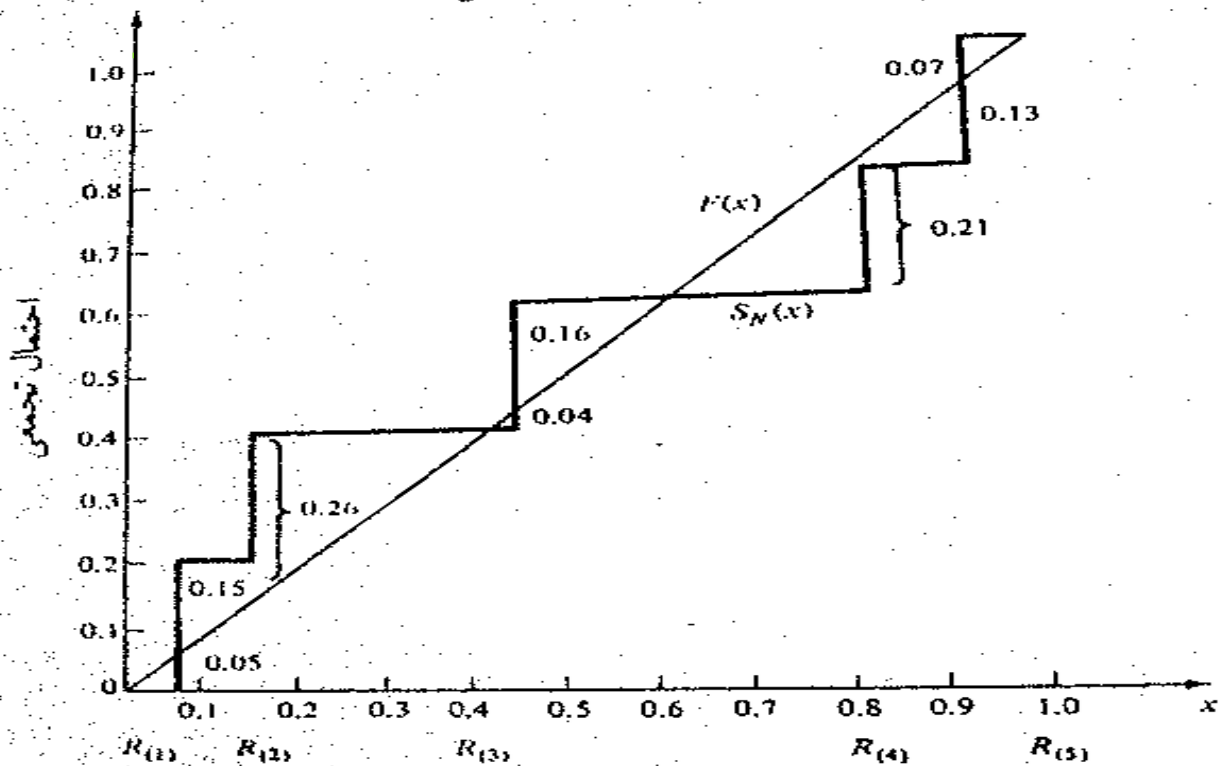
گام ۳) $D = 0.26$

گام ۴) یافتن $D_{0.05}$ از جدول آماری آزمون $D_\alpha = 0.565$

گام ۵) چون $D \leq D_\alpha$ ($0.26 \leq 0.565$) است پس نمونه ها دارای توزیع یکنواخت هستند.

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	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

سحل ۱-۲) معادیر بحرانی دولمودرف اسمیرنوف



شکل ۴-۲ مقایسه $S_N(x)$ و $F(x)$

• آزمون کای دو (خی دو)

این آزمون برای تعیین یکنواختی اعداد و ارتباط با مشاهدات و انتظار بسیار خوب است. برای نمونه های بیشتر از ۵۰ عدد استفاده میشود. اساس این روش بر تقسیم بندی دسته های مشاهدات است. فراوانی اعداد تصادفی تولیدی در هر دسته را با فراوانی انتظار مقایسه و نزدیکی آنها را می سنجد. دسته ها هیچ گونه روی هم افتادگی نباید داشته باشند و تعداد دسته ها باید

$$chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

۳ یا بیشتر باشد. بعد کای دو را به صورت زیر می یابیم:

کای دو مجموع اختلاف مشاهدات و رخ داده ها در دسته هاست. هرچه مشاهدات و انتظارات از یکدیگر فاصله بگیرند، مقدار $(O_i - E_i)^2$ بیشتر میشود و لذا chi^2 افزایش می یابد و اگر این دو مقدار یکسان باشند، chi^2 صفر میشود. روال کار از این قرار است:

- نمونه ها به n دسته تقسیم میگردند که باید $n \geq 3$ باشد.
- O_i تعداد مشاهدات در i امین دسته است.
- E_i تعداد انتظار مشاهده در i امین دسته است.
- $E_i = \frac{N}{n}$ که در آن N تعداد کل نمونه های مشاهده شده است (انتظار مشاهده یکسان)
- نیاز به جدول کای دو است که مقدار بحرانی را از آن بیابیم تا با chi^2 مقایسه شود.
- اگر chi^2 مشاهده شده، از مقدار بحرانی از جدول کوچکتر باشد یکنواختی نمونه ها صحیح است.

▪ یافتن مقدار بحرانی از جدول بر اساس درجه آزادی ($V=n-1$) و پارامتر α میتوان گفت که توزیع نمونه های chi^2 تقریباً توزیع کای دو با $(n-1)$ درجه آزادی است. اگر $chi^2 \leq chi^2_{v,p}$ باشد، آزمون یکنواختی تایید میشود.

مثال) استفاده از آزمون chi^2 با $\alpha = 0.05$ برای آزمون توزیع یکنواخت جهت ۵۰۰۰ نمونه را انجام بدهید.

▪ تقسیم بندی اعداد به دسته های ۱۰۰ تایی ($n=10$)

▪ کل دسته های موجود $n=10$ می باشد.

▪ کل مشاهدات $N=5000$

▪ انتظار مشاهده در هر دسته $E_i = \frac{5000}{10} = 500$

▪ مقدار بحرانی $chi^2_{v,p}$

i	محدوده	O_i	E_i	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1	0-99	468	500	1024	2.048
2	100-199	519	500	361	0.722
3	200-299	480	500	400	0.8
4	300-399	495	500	25	0.05
5	400-499	508	500	64	0.128
6	500-599	426	500	5476	10.952
7	600-699	497	500	9	0.018
8	700-799	515	500	225	0.45
9	800-899	463	500	1369	2.738
10	900-999	529	500	841	1.682

$\chi^2 = 19.588$ ، درجه آزادی $\nu=9$ ، یافتن $p=1-\alpha = 0.95$ ، یافتن $\chi^2_{9,0.95} = 16.9$ از روی جدول مربوطه

چون $19.588 \geq 16.9$ است پس نمونه ها دارای توزیع یکنواختی نیستند و تایید نمیشوند.

v	$X_{0.005}^2$	$X_{0.01}^2$	$X_{0.025}^2$	$X_{0.05}^2$	$X_{0.10}^2$
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

ب. آزمون امتداد (RUN Test)

این آزمون برای بررسی استقلال نمونه ها است و اساس آن بر امتداد اعداد است. افزایش یا کاهش مشاهده ها در سری تولیدی مورد بررسی قرار میگیرد و به عنوان توالی وقایع اتفاقی آزمون را سپری میکند. مقداری را از آزمون به دست می آوریم و با مقدار بحرانی به دست آمده از جدول کای دو مقایسه میکنیم.

رشته اعداد را مینویسیم. اگر از عدد کوچک به بزرگ برویم + یا H و اگر از عدد بزرگ به عدد کوچک برویم - یا T میگذاریم. طول رشته حاصل را میشماریم. حالا تعداد این افزایش کاهش ها را میشماریم که میشود تعداد امتداد یا همان RUN . اگر تعداد اعداد تولیدی N باشد، حداکثر و حداقل امتدادها برابر خواهد بود با:

$$1 \leq RUNS \leq N - 1$$

به مثال های صفحات ۱۰۸ و ۱۰۹ کتاب مراجعه کنید.

اما روش کار این آزمون به ترتیب زیر است:

تعداد کل امتدادها: $a = RUNS$

میانگین: $\mu_a = \frac{2N-1}{3}$

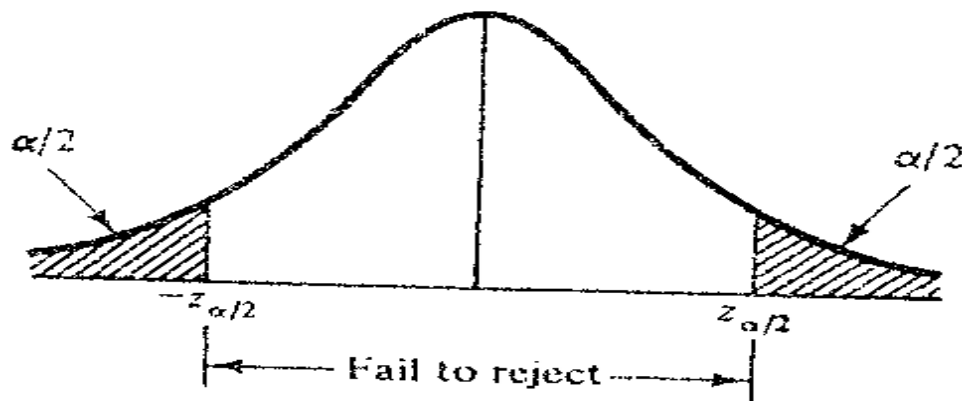
واریانس: $\sigma_a^2 = \frac{16N-29}{90}$

برای $N > 20$ توزیع a با توزیع نرمال $N(\mu_a, \sigma_a^2)$ تقریب زده میشود. این تقریب برای آزمون استقلال اعداد تولیدی است. پس پارامتر Z_0 به شکل زیر و به عنوان پارامتر اصلی تعریف میگردد.

$$Z_0 = \frac{a - \mu_a}{\sigma_a} \quad \text{پارامتر اصلی:}$$

$$Z_0 = \frac{a - [\frac{2N-1}{3}]}{\sqrt{(16N-29)/90}} \quad \text{با جایگذاری}$$

با یافتن $Z_{\alpha/2}$ از جدول کای دو، اگر $-Z_{\alpha/2} \leq Z_0 \leq Z_{\alpha/2}$ باشد، آزمون استقلال موفق بوده است.



مثال) به ازای $\alpha = 0.05$ تعیین کنید که آیا توالی ۴۰ عدد تصادفی زیر دارای خاصیت استقلال است؟

ج. آزمون همبستگی

موضوع تحقیق در صورت علاقه در غیر اینصورت در فصل های آتی تدریس خواهد شد.

د. آزمون فاصله

موضوع تحقیق در صورت علاقه

ه. آزمون پوکر

جهت بررسی استقلال به کار میرود و اساس آن بر تکرار ارقام است که در یک سری اعداد تکرار میشوند. گاهی اوقات این تکرارها بی قاعده اند (مثال ص ۱۱۲ کتاب). در هر موقعیت یک جفت ارقام مثل هم، در اعداد تولیدی ظاهر میشوند. در اعداد سه رقمی فقط سه امکان وجود دارد:

۱- ارقام میتوانند متفاوت باشند.

۲- ارقام می توانند یکسان باشند.

۳- یک جفت مانند هم باشند.

احتمال هر کدام می شود:

$$P(\text{سه رقم سوم با دو رقم اول و دوم متفاوت باشد}) \times P(\text{دومین رقم متفاوت است از اولین رقم}) = P(\text{سه رقم مختلف}) \\ = (0.9)(0.8) = 0.72$$

$$P(\text{سومین رقم مثل اولین رقم است}) \times P(\text{دومین رقم مثل اولین رقم است}) = P(\text{سه رقم یکسان}) \\ = (0.1)(0.1) = 0.01$$

$$P(\text{یک جفت یکسان}) = 1 - 0.72 - 0.01 = 0.27 \implies P(\text{یک جفت یکسان}) = \binom{3}{2} (0.1)(0.9) = 0.27$$

نمونه سوال امتحانی

نمونه سوالها را در فایل جداگانه ای در اسرع وقت برایتان ان شالله ارسال خواهم کرد.