

درس ۲۵:

تولید اعداد تصادفی

تهیه شده توسط گروه بهینه یاب



www.behinehyab.com

مقدمه

روش‌های متفاوتی برای تولید اعداد تصادفی وجود دارد. تمام این روش‌ها یک وجه مشترک دارند: اینکه تمام آن‌ها از برنامه‌های کامپیوتری مولد دنباله اعداد استفاده می‌کنند به طوری که از لحاظ نظری اعداد مزبور در فاصله صفر تا یک **توزیع یکنواخت** دارند و از لحاظ آماری نیز هر عدد از سایر اعداد موجود در دنباله **مستقل** است.

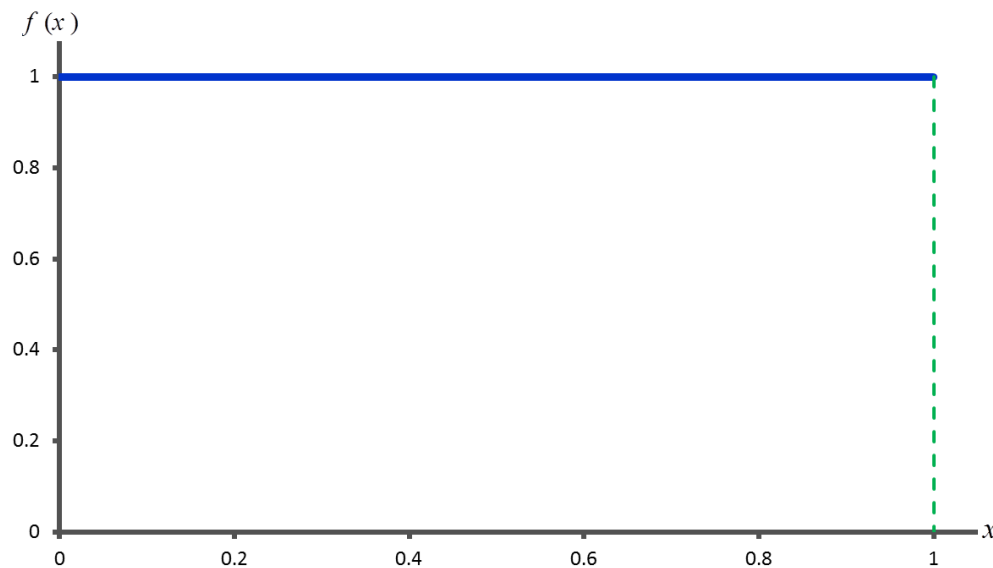
در این درس به آرایه برخی از روش‌های محاسباتی تولید اعداد تصادفی می‌پردازیم، ضوابط ارزیابی، مقایسه و انتخاب را عرضه می‌داریم و سرانجام آزمایش‌های مربوط به تصادفی بودن اعداد به دست آمده از مولدهای را معرفی می‌کنیم.

خواص اعداد تصادفی

هر دنباله از اعداد تصادفی مانند $R_1, R_2, \dots$ باید دو خاصیت آماری مهم داشته باشد. این دو خاصیت عبارت از توزیع احتمال **یکنواخت** و **استقلال** است. هر عدد تصادفی مانند R_i نمونه مستقلی از توزیع احتمال یکنواخت و پیوسته با پارامترهای **صفر و یک** محسوب می‌شود. تابع چگالی این گونه متغیر تصادفی به صورت زیر تعریف می‌شود.

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & x < 0, x > 1 \end{cases}$$

نمایش این تابع چگالی در **شکل ۱** نشان داده شده است.



شکل ۱: تابع چگالی اعداد تصادفی

تابع تجمعی، امید ریاضی، و واریانس R_i به ترتیب به صورت زیر محاسبه می شود.

$$P\{R \leq r\} = \int_0^r dx = r, \quad 1 > r > 0$$

$$E(R) = \int_0^1 x dx = \frac{1}{2} x^2 \Big|_0^1 = \frac{1}{2}$$

$$Var(R) = \int_0^1 x^2 dx - [E(R)]^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

در باره پیامدهای خاصیت **استقلال** و خاصیت توزیع احتمال **یکنواخت** پیوسته برای اعداد تصادفی،

می توان دو مورد زیر را بیان کرد:

۱- اگر فاصله $(0, 1)$ به n رده با زیر فاصله **مساوی** تقسیم شود، انتظار می رود که از N مشاهده $\frac{N}{n}$

آن ها در هر رده قرار گیرد.

۲- احتمال حصول یک مشاهده در یک رده، **مستقل** از محل قرار گرفتن سایر مشاهده ها است.

تولید اعداد شبه تصادفی

می‌بینیم که در عنوان این بخش از کلمه شبه به معنی **غیرواقعی** استفاده شده است. در واقع، منظور این نیست که در این بخش به تولید تصادفی غیرواقعی می‌پردازیم. مقصود از قید کلمه شبه تاکید بر این مطلب است که استفاده از یک روش قطعی و مشخص برای تولید اعداد تصادفی، امکان بالقوه تصادفی بودن واقعی را از بین می‌برد. در صورتی که روش مورد استفاده، قطعی و مشخص باشد، مجموعه اعداد تصادفی تولید شده توسط آن **تکرار پذیر** خواهد بود. پس، اعداد تولید شده واقعاً تصادفی **نیست**. اما باید توجه داشت هدف هر روش این است که به نحوی اعداد تصادفی در محدود صفر و یک تولید کند تا اعداد تولید شده در بالاترین حد ممکن دو خاصیت **استقلال** و **توزیع یکنواخت** را داشته باشد.

به هنگام تولید اعداد شبه تصادفی، مسلماً بروز مسائل یا ارتکاب خطاهایی قابل تصور است. این نوع مسائل یا خطاها به خواص **استقلال** و توزیع احتمال **یکنواخت** مربوط می‌شود. مثال هایی در این زمینه در زیر ارائه شده است.

- ۱- اعداد تصادفی ممکن است توزیع احتمال یکنواخت نداشته باشند.
- ۲- اعداد تصادفی ممکن است جدا از هم (نه پیوسته) باشد
- ۳- میانگین اعداد تولید شده ممکن است بیش از حد بزرگ یا بیش از حد متعارف کوچک باشد.
- ۴- واریانس اعداد تصادفی تولید شده ممکن است تفاوت قابل توجهی از مقدار متعارف داشته باشد.
- ۵- ممکن است دنباله اعداد تولید شده تغییراتی تناوبی از خود نشان دهد. بعضی از مثال ها عبارتند از:

(الف) وجود همبستگی بین اعداد

(ب) وجود رابطه مقداری بین اعداد مجاور به این ترتیب که اعداد مجاور روندی صعودی یا نزولی از خود نشان دهد.

ج) وجود چند عدد بزرگتر از میانگین و به دنبال آن وجود چند عدد کوچکتر از میانگین

به منظور کشف **عدم استقلال** و یا **عدم پیرو** اعداد بدست آمده از یک مولد از تابع چگالی یکنواخت، می توان از آزمون های معرفی شده که در این درس ارایه می شود استفاده کرد. اگر متعاقب انجام آزمون های مربوط مولدی غیرقابل قبول تشخیص داده شد باید به جستجوی مولد دیگری پرداخت.

تعداد روش های تولید اعداد (شبه) تصادفی فراوان است. از این بعد کلمه شبه را حذف می کنیم. قبل از تشریح برخی از روش های تولید اعداد تصادفی، بعضی از ملاحظات مهم را فهرست وارد ارایه می کنیم.

۱- روش یا الگوریتم تولید اعداد تصادفی باید **سریع** باشد. محاسبات شبیه سازی شاید آنقدر پرهزینه نباشد ولی جمع محاسبات لازم در یک بررسی شبیه سازی ممکن است قابل توجه باشد. در این صورت باید به جستجوی الگوریتمی پرداخت که هزینه تولید در مورد آن بسیار کم باشد.

۲- روش نباید نیاز به **مقدار زیادی حافظه** کامپیوتری داشته باشد. اشغال حافظه ممکن است بسیار پر هزینه باشد و بر انجام بخش های دیگر بررسی شبیه سازی از لحاظ نیاز به حافظه تاثیر منفی بگذارد.

۳- طول دنباله اعداد تولید شده باید به اندازه کافی بلند باشد. منظور از طول دنباله، تعداد اعدادی است که بدون تکرار دنباله در آن قرار می گیرد. حالت خاصی از تکرار دنباله ناظر به از **هم پاشیدن الگوریتم** است. اگر عدد تصادفی معینی به طور مکرر تولید شود می گویند الگوریتم از هم پاشیده است. این امر در یکی از مثال های بخش بعد توضیح داده شده است.

۴- صرفنظر از وضعیت سیستم در دست شبیه سازی، باید بتوان با **تعیین نقطه شروع**، الگوریتم مولد را در تولید مجموعه مشخصی از اعداد تصادفی به کار گرفت. رعایت این امر در بررسی های شبیه سازی لازم است و از آن برای تسهیل مقایسه دو سیستم استفاده می شود.

۵- اعداد تصادفی تولید شده باید تا حدود زیادی از خواص آماری توزیع **یکنواخت و استقلال**

برخوردار باشد.

روش های تولید اعداد تصادفی

روش های رایج تولید اعداد تصادفی در این بخش بررسی می شود.

روش میان مربعی

این روش با یک عدد اولیه به نام هسته شروع به کار می کند. روال کار چنین است که هسته را **مربع** می کند و ارقام میانی آن را تعیین و پس از نوشتن صفر و ممیز در سمت چپ ارقام میانی، اولین عدد تصادفی را تولید می کنند. به منظور تولید عدد تصادفی دوم، باید ارقام میانی در مرحله قبل را مربع، ارقام میانی حاصل را تعیین و به عدد اعشاری تبدیل کرد و همین روند ادامه پیدا می کند. اگر هسته n رقمی باشد، مربع آن $2n-1$ یا $2n$ رقمی خواهد بود و اگر n زوج باشد با حذف $n/2$ ارقام از هر سمت (راست و چپ) مربع آن، می توان ارقام میانی را تعیین کرد.

مثال ۱: فرض کنید که به دنباله از اعداد تصادفی چهار رقمی نیاز داریم. هسته را با X نماد گذاری کنید و مقدار ۵۴۹۷ را برای آن در نظر بگیرید. فرض کنید i -امین عددی که مربع می شود X_i و i امین عدد تصادفی که تولید می شود R_i باشد.

$$X_0 = 5497$$

$$X_0^2 = (5497)^2 = 30\boxed{2170}09 \Rightarrow X_1 = 2170 \Rightarrow R_1 = 0.2170$$

$$X_1^2 = (2170)^2 = 04\boxed{7089}00 \Rightarrow X_2 = 7089 \Rightarrow R_2 = 0.7089$$

$$X_2^2 = (7089)^2 = 50\boxed{2539}21 \Rightarrow X_3 = 2539 \Rightarrow R_3 = 0.2539$$

...

نکته: این روش می تواند دچار مشکلاتی شود. اولاً نمی توان قواعد ساده ای برای تعیین مقدار هسته ارایه کرد که عملکرد مطلوب الگوریتم را تضمین کند. همچنین، با ظهور رقم صفر در سمت چپ ارقام میانی، دنباله اعداد تصادفی تولید شده به سرعت به انتها می رسد. در روشن شدن این موضوع به مثال زیر توجه کنید.

مثال ۲: فرض کنید که هسته **مثال ۱** به ۵۱۹۷ تغییر کند. در این صورت:

$$X_0 = 5197$$

$$X_0^2 = (5197)^2 = 27\boxed{0088}09 \Rightarrow X_1 = 0088 \Rightarrow R_1 = 0.0088$$

$$X_1^2 = (0088)^2 = 00\boxed{0077}44 \Rightarrow X_2 = 0077 \Rightarrow R_2 = 0.0077$$

$$X_2^2 = (0077)^2 = 00\boxed{0059}29 \Rightarrow X_3 = 0059 \Rightarrow R_3 = 0.0059$$

...

می بینیم که تمام R_i ها با صفر شروع می شود.

روش فوق عیب دیگری هم دارد که از هم پاشیدگی شدن آن است. از هم پاشیده شدن الگوریتم به خاطر حصول مقداری تکراری یا مقدار صفر برای ارقام میانی است.

مثال ۳: تصور کنید که به هنگام استفاده از روش میان مربعی، مقدار ۶۵۰۰ برای یکی از X ها تولید شده است. در این صورت، به ازای $n=4$ داریم:

$$X_i = 6500$$

$$X_i^2 = (6500)^2 = 42\boxed{2500}00 \Rightarrow X_{i+1} = 2500 \Rightarrow R_{i+1} = 0.2500$$

$$X_{i+1}^2 = (2500)^2 = 06\boxed{2500}00 \Rightarrow X_{i+2} = 2500 \Rightarrow R_{i+2} = 0.2500$$

$$X_{i+2}^2 = (2500)^2 = 06\boxed{2500}00 \Rightarrow X_{i+3} = 2500 \Rightarrow R_{i+3} = 0.2500$$

...

با ثابت ماندن مقدار اعداد تصادفی تولد شده، الگوریتم عملاً از هم می پاشد.

روش میان ضربی

این روش همانند روش میان مربعی است و با انتخاب دو هسته، X و X' که تعداد ارقامشان مساوی است شروع می شود. اگر دو هسته n رقمی باشد، آن ها را در هم ضرب می کنیم و با حذف $n/2$ از ارقام سمت راست و همین تعداد رقم از سمت چپ حاصل ضرب، n رقم قرار گرفته در وسط را تعیین و آن را تبدیل به یک عدد اعشاری کوچکتر از یک می کنیم. اگر n رقم قرار گرفته در وسط حاصل ضرب با X_1 نماد گذاری شود، به منظور تولید عدد تصادفی دوم، محاسبات فوق را با X و X_1 تکرار می کنیم و این روند تکرار می شود. برای روشن شدن موضوع مثال زیر را در نظر بگیرید.

مثال ۴: از دو هسته $X_0 = 7229$ و $X'_0 = 2938$ استفاده و براساس روش میان ضربی، اعداد تصادفی چهار

رقمی تولید کنید.

$$U_1 = X'_0 X_0 = (2938)(7229) = 21238802 \Rightarrow X_1 = \sqrt{21238802} = 2388$$

$$\Rightarrow R_1 = 0.2388$$

$$U_2 = X_0 X_1 = (7229)(2388) = 17262852 \Rightarrow X_2 = \sqrt{17262852} = 2628$$

$$\Rightarrow R_2 = 0.2628$$

$$U_3 = X_1 X_2 = (2388)(2628) = 6275664 \Rightarrow X_3 = \sqrt{6275664} = 2756$$

$$\Rightarrow R_3 = 0.2756$$

نکته: این روش شباهت بسیاری به روش میان مربعی دارد. طول دنباله برای این روش بلندتر و

یکنواختی توزیع اعداد تولید شده بیشتر است. به هر صورت، این روش نیز عیب عمده‌ای از هم پاشیده شدن را دارد.

روش مضرب ثابت

این روش تفاوت ناچیزی با روش میان ضربی دارد. در واقع، با ثابت نگه داشتن یکی از دو عددی که در روش قبل در یکدیگر ضرب می‌شد می‌توان اعداد را تولید کرد. فرض کنید عدد ثابت را با K نماد گذاری می‌کنیم و تعداد ارقام K و X را یکسان در نظر می‌گیریم. تحت این شرایط، تمام محاسبات مانند روش میان ضربی انجام می‌گیرد.

مثال ۵: با استفاده از دو هسته $K = 3987$ و $X_0 = 7223$ و براساس روش مضرب ثابت، دنباله‌ای از اعداد

تصادفی چهار رقمی تولید کنید.

$$V_1 = KX_0 = (3987)(7223) = 28798101 \Rightarrow X_1 = 7981 \Rightarrow R_1 = 0.7981$$

$$V_2 = KX_1 = (3987)(7981) = 31820247 \Rightarrow X_2 = 8202 \Rightarrow R_2 = 0.8202$$

این روش همان معایب روش میان ضربی را دارد. عملکرد مناسب این روش تا حد زیادی به انتخاب

مقدار ثابت K بستگی دارد.

روش همنهشتی جمعی

اساس نظری این روش از روش های پیش گفته متفاوت است و عملکرد آن کاملاً شناخته شده نیست. خواهیم دید که این روش به سبب بی نیازی به عمل ضرب، سرعت قابل توجهی دارد. این روش چنین شروع به کار می کند که یک دنباله n تایی مانند $X_1, X_2, \dots$ و X_n را دریافت و بقیه دنباله، یعنی $X_{n+1}, X_{n+2}, \dots$ را تولید می کند. اساس تولید مقدار در این روش استفاده از رابطه

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

است. طبق تعریف رابطه $a \equiv (b) \pmod{m}$ معادل این است که بگوییم باقیمانده تقسیم $(b-a)$ بر m مساوی صفر است. مثال بعد چگونگی انجام محاسبات را نشان داده و مثال پس از آن سرعت زاید الوصف روش همنهشتی جمعی در تولید اعداد صحیح تصادفی نشان می دهد.

مثال ۶: اگر مقدار ۴ برای پیمانه یعنی m ، در نظر گرفته شود، اعداد ۲ و ۵ و ۱۰ و ۱۴ و ... معادل هم هستند زیرا عددی مانند ۱۰ وجود دارد که باقی مانده تقسیم $(10-2), (10-6), \dots$ و به عدد ۴ مساوی صفر است. در حالت کلی، اگر پیمانه با m مشخص شود، نتیجه نهایی، یک عدد صحیح بین صفر و $m-1$ خواهد بود. بنابراین داریم

$$17 \pmod{3} = 2$$

$$14 \pmod{5} = 4$$

$$37 \pmod{2} = 1$$

$$16 \pmod{4} = 0$$

...

مثال ۷: فرض کنید n مساوی ۵ و دنباله اعداد صحیح تصادفی X_1, X_2, X_3, X_4, X_5 به ترتیب مساوی ۵۷، ۳۴، ۸۹، ۹۲، ۱۶ باشد. برای m مقدار ۱۰۰ را در نظر بگیرید. از روش همنهشتی جمعی استفاده کنید و این دنباله را ادامه دهید.

$$\begin{aligned}
X_6 &\equiv (X_5 + X_1) \pmod{100} = 73 \pmod{100} = 73 \\
X_7 &\equiv (X_6 + X_2) \pmod{100} = 107 \pmod{100} = 7 \\
X_8 &\equiv (X_7 + X_3) \pmod{100} = 96 \pmod{100} = 96 \\
X_9 &\equiv (X_8 + X_4) \pmod{100} = 188 \pmod{100} = 88 \\
X_{10} &\equiv (X_9 + X_5) \pmod{100} = 104 \pmod{100} = 4 \\
X_{11} &\equiv (X_{10} + X_6) \pmod{100} = 77 \pmod{100} = 77 \\
&\vdots
\end{aligned}$$

روش همنهشتی خطی

تصور کنید که دنباله اعداد صحیح $\{X_i\}$ طبق رابطه تکرار پذیر است.

$$X_i \equiv (aX_{i-1} + c) \pmod{m}, \quad i = 1, 2, \dots$$

در رابطه فوق، ضریب a ، مقدار ثابت c و پیمانه m همگی اعداد صحیح است. علامت همنهشتی، یعنی

$\equiv$ چنین می‌رساند که از نظر جبری هر X_i از طریق رابطه زیر تعریف می‌شود.

$$X_i = aX_{i-1} + c - mK_i$$

K_i معرف بزرگترین عدد صحیح موجود در $\frac{aX_{i-1} + c}{m}$ است. اگر $c > 0$ باشد، مولد زیر را یک مولد

همنهشتی آمیخته می‌نامند.

$$X_i \equiv (aX_{i-1} + c) \pmod{m}, \quad i = 1, 2, \dots$$

نکته: در صورتی که $c = 0$ باشد، مولد مزبور را **همنهشتی ضربی** می‌خوانند.

اگر مقدار عددی هسته را با X نمادگذاری کنیم، مولد فوق الذکر به صورت رابطه زیر می‌شود.

$$X_1 \equiv (aX_0 + c) \bmod m$$

$$X_2 \equiv (aX_1 + c) \bmod m = \left[a^2 X_0 + (a+1)c \right] \bmod m = \left[a^2 X_0 + \left(\frac{a^2 - 1}{a - 1} \right) c \right] \bmod m$$

$$X_3 \equiv (aX_2 + c) \bmod m = \left[a^3 X_0 + \left(\frac{a^3 - 1}{a - 1} \right) c \right] \bmod m$$

⋮

$$X_i \equiv (aX_{i-1} + c) \bmod m = \left[a^i X_0 + \left(\frac{a^i - 1}{a - 1} \right) c \right] \bmod m$$

مثال ۸: مقادیر زیر را در نظر بگیرید

$$X_0 = 27, a = 17, c = 43, m = 100$$

براساس روش همنهستی خطی، دنباله ای از اعداد تصادفی تولید کنید. چون مقدار پیمانه مساوی ۱۰۰ است، تمام اعداد صحیح تصادفی به دست آمده از این روش از **صفر** تا **۹۹** به طور یکنواخت پراکنده خواهد بود. برای تولید اعداد تصادفی در محدود صفر و یک، می توان از رابطه زیر استفاده کرد.

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

با استفاده از روش همنهستی خطی داریم:

$$X_0 = 27$$

$$X_1 \equiv [17(27) + 43] \bmod 100 = 502 \bmod 100 = 2 \Rightarrow R_1 = 0.02$$

$$X_2 \equiv [17(2) + 43] \bmod 100 = 77 \bmod 100 = 77 \Rightarrow R_2 = 0.77$$

$$X_3 \equiv [17(77) + 43] \bmod 100 = 1352 \bmod 100 = 52 \Rightarrow R_3 = 0.52$$

⋮

نظیر هر مولد دیگر، در مورد مولدهای همنهستی خطی نیز باید به آزمایش این مطلب دست زد که اعداد تصادفی تولید شده در این روش تا چه میزان از دو شرط **استقلال** و توزیع آماری **یکنواخت** برخوردار است.

آزمون های اعتبار مولدهای تصادفی

به منظور تعیین این مطلب که آیا مولد در دست بررسی می تواند اعدادی با خواص مورد نظر تولید کند یا نه، باید به آزمون هایی را انجام داد. آزمون های مورد بحث را می توان برحسب دو خاصیت فوق در **دو** گروه رده بندی کرد. در ادامه این آزمون ها معرفی می شوند.

آزمون فراوانی

از آزمون های اساسی که به منظور تشخیص اعتبار هر مولد تازه از لحاظ همگونی توزیع آماری اعداد تولید شده با **توزیع احتمال یکنواخت** انجام می گیرید، **آزمون فراوانی** است. دو روش مختلف برای انجام این آزمون وجود دارد: یکی مبتنی بر کاربرد آماره تقریبی **مربع کای** و دیگری مبتنی بر استفاده از آماره **کالموگروف-اسمیرنف** است. این دو آزمون به بررسی میزان همگونی توزیع احتمال یک نمونه از اعداد تصادفی تولید شده با توزیع احتمال یکنواخت می پردازند. با توجه به اهمیت این دو آزمون به خصوص آزمون مربع کالی، به تفصیل به بررسی آن می پردازیم.

آزمون مربع کای

در این بخش به بررسی کاربرد آزمون مربع کای می پردازیم. این آزمون از آماره زیر استفاده می کند.

$$\chi_0^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

که در آن:

O_i : تعداد مشاهدات در i امین دسته بندی

E_i : امید تعداد در i -امین دسته بندی

n : تعداد دسته بندی ها است.

برای محاسبه E_i از رابطه زیر استفاده می‌کنیم که در آن فرض می‌شود که طول دسته بندی ها یکسان است.

$$E_i = \frac{N}{n}$$

که در رابطه فوق، تعداد کل مشاهدات برابر N است.

نکته: می‌توان نشان داد که تابع توزیع نمونه χ_0' به طور تقریبی دارای تابع توزیع مربع کای با $n-1$ درجه آزادی است.

مثال ۹: یک نمونه ($n=100$) از اعداد تصادفی تولید شده توسط یک مولد در زیر دیده می‌شود.

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

با استفاده از آزمون مربع کای با مقدار $\alpha=0.05$ کنترل کنید که آیا اعداد تصادفی دارای توزیع یکنواخت است. محاسبات مورد نیاز برای این آزمون در جدول زیر آمده است.

شماره بازه	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1	8	10	-2	4	0.4
2	8	10	-2	4	0.4
3	10	10	0	0	0
4	9	10	-1	1	0.1
5	12	10	2	2	0.2
6	8	10	-2	2	0.2
7	10	10	0	0	0
8	14	10	4	16	1.6
9	10	10	0	0	0
10	11	10	1	1	0.1
جمع	100	100	0	-	3.4

در این آزمون تعداد بازه ها برابر ۱۰ و طول بازه ها یکسان در نظر گرفته شد. این بازه ها به صورت

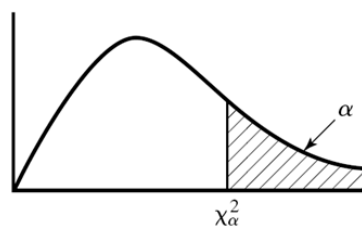
$[0, 0.1), [0.1, 0.2), \dots, [0.9, 1)$ است. مقدار χ'_0 برابر ۳.۴ است. در صورتیکه این مقدار را با مقدار بحرانی

$\chi'_{0.05,9} = 16.9$ است. برای تعیین مقادیر بحرانی می توانیم از جدول ذیل استفاده کنیم. به دلیل این که

مقدار $\chi'_0 = 3.4$ از $\chi'_{0.05,9} = 16.9$ کمتر است لذا فرضیه یکنواخت بودن توزیع متغیرهای تصادفی را

نمی توان با احتمال $1 - \alpha$ رد کرد.

نکته: توصیه می شود که n و N به نحوی انتخاب شود که $E_i \geq 5$ رابطه برقرار باشد.



توزیع مربع کای با
درجه آزادی ν

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

آزمون کالموگروف-اسمیرنف

این آزمون برای مقایسه تابع توزیع تجمعی پیوسته $F(x)$ با تابع توزیع تجربی $S_N(x)$ از نمونه با N داده است. از تعریف داریم:

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

در صورتیکه از متغیر تصادفی X به تعداد N مشاهده به صورت $X_1, X_2, \dots, X_N$ بدست آید، تابع تجمعی تجربی x به صورت زیر تعریف می شود.

$$S_N(x) = \frac{\text{number of } x_1, x_2, \dots, x_N \text{ which are } \leq x}{N}$$

تعریف: تصور کنید که $X_1, X_2, \dots, X_N$ یک نمونه از توزیع احتمال $F(x)$ را تشکیل می دهد. در این صورت آماره D به صورت زیر محاسبه می شود.

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

تابع توزیع D دانسته و از جدول زیر برای مقدار N محاسبه می شود. برای آزمون این که تابع توزیع اعداد تصادفی به صورت یکنواخت است باید مراحل زیر را انجام داد:

گام ۱: داده ها را از کوچکترین تا بزرگترین مرتب کنید. عدد $R_{(i)}$ را i -امین کوچکترین مشاهده در نظر بگیرید که در آن

$$R_{(1)} \leq R_{(2)} \leq \dots \leq R_{(N)}$$

گام ۲: مقدارهای زیر را محاسبه کنید.

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

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

گام ۳: مقدار D را به صورت زیر محاسبه کنید.

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

گام ۴: با استفاده از جدول مقدار بحرانی **کالموگروف-اسمیرنف**، مقدار D_α برای سطح اطمینان α

و تعداد نمونه N را بدست آورید.

گام ۵: اگر $D > D_\alpha$ باشد آنگاه فرض تابع توزیع یکنواخت رد می‌شود. اگر $D > D_\alpha$ باشد آنگاه

می‌توان به این نتیجه رسید که تفاوت معناداری بین $R_{(1)} \leq R_{(2)} \leq \dots \leq R_{(N)}$ و تابع توزیع یکنواخت با احتمال $1 - \alpha$ وجود ندارد.

مقادیر بحرانی آزمون کالموگروف-اسمیرنف

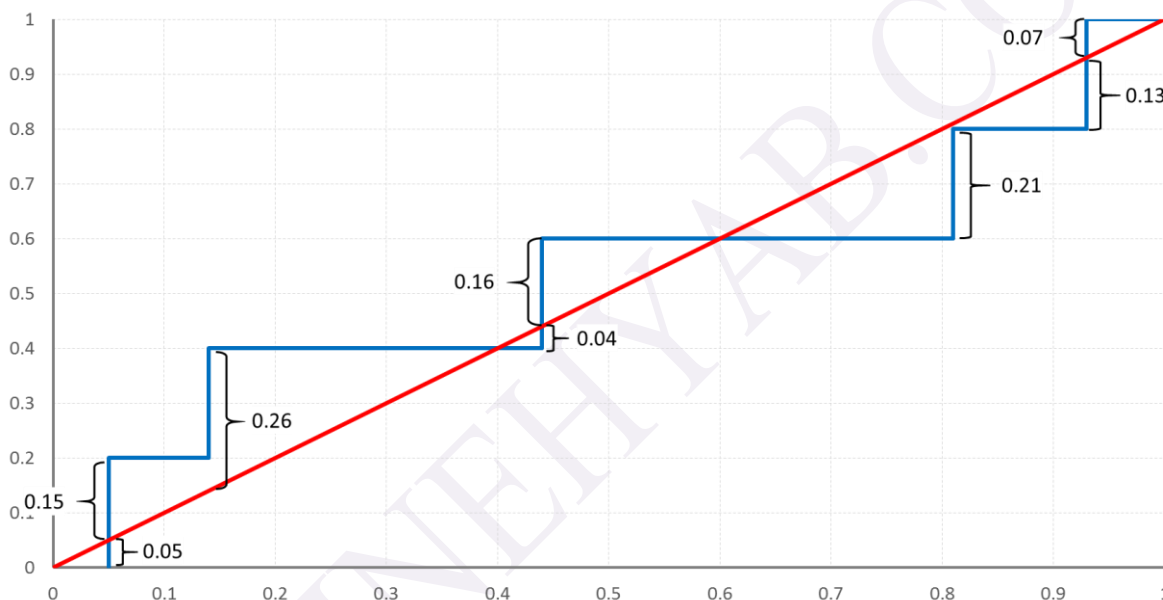
Degrees of Freedom (N)	$D_{0.10}$	$D_{0.05}$	$D_{0.01}$
1	0.950	0.975	0.995
2	0.776	0.842	0.929
3	0.642	0.708	0.828
4	0.564	0.624	0.733
5	0.510	0.565	0.669
6	0.470	0.521	0.618
7	0.438	0.486	0.577
8	0.411	0.457	0.543
9	0.388	0.432	0.514
10	0.368	0.410	0.490
11	0.352	0.391	0.468
12	0.338	0.375	0.450
13	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	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

مثال ۱۰: فرض کنید ۵ عدد تصادفی ۰.۴۴، ۰.۸۱، ۰.۱۴، ۰.۰۵، و ۰.۹۳ را به ازای $\alpha = 0.05$ مورد

آزمایش قرار می‌دهیم تا معلوم شود آیا توزیع احتمال یکنواخت دارد یا نه. در اولین گام باید مقادیر فوق را به ترتیب مقدار نوشت. در این صورت، R_i ($i = 1, 2, \dots, 5$) است. محاسبات لازم در مورد این مساله در جدول زیر نشان داده شده است.

$R_{(i)}$	0.05	0.14	0.44	0.81	0.93
i/N	0.20	0.40	0.60	0.80	1.00
$i/N - R_{(i)}$	0.15	0.26	0.16	-	0.07
$R_{(i)} - (i-1)/N$	0.05	-	0.04	0.21	0.13

باید توجه داشت که تابع تجمعی متغیر تصادفی یکنواخت صفر و یک به صورت i/N نوشته می‌شود. به علاوه مقادیر بدست آمده برای D^+ و D^- ، به ترتیب برابر با $D^+ = \frac{i}{N} - R_{(i)} = 0.26$ و $D^- = R_{(i)} - \left(\frac{i-1}{N}\right) = 0.21$ می‌شود. در نتیجه $D = \max\{0.26, 0.21\} = 0.26$. براساس جدول مقادیر بحرانی، به ازای $\alpha = 0.05$ و $N = 5$ خواهیم داشت $D_\alpha = 0.565$. در نتیجه نمی‌توان فرض توزیع احتمال یکنواخت صفر و یک را برای مقادیر داده شده در بالا را رد کرد. نمایش D^+ و D^- در شکل ۲ آورده شده است.



شکل ۲: نمایش D^+ و D^-

آزمون شکاف

از آزمون شکاف می‌توان به منظور بررسی شکاف‌ها یا فواصل موجود بین تکرارهای متوالی هر یک از ارقام پایه عددی استفاده کرد. به طور کلی، رابطه زیر محاسبه احتمالات نظیر شکاف‌هایی به طول x را تشکیل می‌دهد.

$$P \{ \text{وجود دقیقاً } x \text{ رقم غیر } t \text{ بین دو تکرار متوالی } t \} = 0.9^x 0.1, x = 0, 1, 2, \dots$$

در این آزمون از آماره مربع کای استفاده می‌شود.

مثال ۱۱: دنباله زیر را در نظر بگیرید.

4	1	3	5	1	7	2	8	2	0	7	9	1	3	5	2	7	9
4	1	6	3	3	9	6	3	4	8	2	3	1	9	4	4	6	8
4	1	3	8	9	5	5	7	3	9	5	9	8	5	3	2	2	3
7	4	7	0	3	6	3	5	9	9	5	5	5	0	4	6	8	0
4	7	0	3	3	0	9	5	7	9	5	1	6	6	3	8	8	8
9	2	9	1	8	5	4	4	5	0	2	3	9	7	1	2	0	3
6	3																

به ازای $\alpha = 0.05$ ، آزمون شکاف را برای دنباله ۱۱۰ رقمی فوق را انجام دهیم.

برای رقم‌های صفر تا نه، شکاف‌ها در شکل نشان داده شده است.

شکاف بین رقم صفر (۰)

4	1	3	5	1	7	2	8	2	0	7	9	1	3	5	2	7	9
4	1	6	3	3	9	6	3	4	8	2	3	1	9	4	4	6	8
4	1	3	8	9	5	5	7	3	9	5	9	8	5	3	2	2	3
7	4	7	0	3	6	3	5	9	9	5	5	5	0	4	6	8	0
4	7	0	3	3	0	9	5	7	9	5	1	6	6	3	8	8	8
9	2	9	1	8	5	4	4	5	0	2	3	9	7	1	2	0	3
6	3																

شکاف بین رقم یک (۱)

4	1	3	5	1	7	2	8	2	0	7	9	1	3	5	2	7	9
4	1	6	3	3	9	6	3	4	8	2	3	1	9	4	4	6	8
4	1	3	8	9	5	5	7	3	9	5	9	8	5	3	2	2	3
7	4	7	0	3	6	3	5	9	9	5	5	5	0	4	6	8	0
4	7	0	3	3	0	9	5	7	9	5	1	6	6	3	8	8	8
9	2	9	1	8	5	4	4	5	0	2	3	9	7	1	2	0	3
6	3																

شکاف بین رقم دو (۲)

4	1	3	5	1	7	2	8	2	0	7	9	1	3	5	2	7	9
4	1	6	3	3	9	6	3	4	8	2	3	1	9	4	4	6	8
4	1	3	8	9	5	5	7	3	9	5	9	8	5	3	2	2	3
7	4	7	0	3	6	3	5	9	9	5	5	5	0	4	6	8	0
4	7	0	3	3	0	9	5	7	9	5	1	6	6	3	8	8	8
9	2	9	1	8	5	4	4	5	0	2	3	9	7	1	2	0	3
6	3																

شکاف بین رقم سه (۳)

4	1	3	5	1	7	2	8	2	0	7	9	1	3	5	2	7	9
4	1	6	3	3	9	6	3	4	8	2	3	1	9	4	4	6	8
4	1	3	8	9	5	5	7	3	9	5	9	8	5	3	2	2	3
7	4	7	0	3	6	3	5	9	9	5	5	5	0	4	6	8	0
4	7	0	3	3	0	9	5	7	9	5	1	6	6	3	8	8	8
9	2	9	1	8	5	4	4	5	0	2	3	9	7	1	2	0	3
6	3																

شکاف بین رقم چهار (۴)

4	1	3	5	1	7	2	8	2	0	7	9	1	3	5	2	7	9
4	1	6	3	3	9	6	3	4	8	2	3	1	9	4	4	6	8
4	1	3	8	9	5	5	7	3	9	5	9	8	5	3	2	2	3
7	4	7	0	3	6	3	5	9	9	5	5	5	0	4	6	8	0
4	7	0	3	3	0	9	5	7	9	5	1	6	6	3	8	8	8
9	2	9	1	8	5	4	4	5	0	2	3	9	7	1	2	0	3
6	3																

شکاف بین رقم پنج (۵)

4	1	3	5	1	7	2	8	2	0	7	9	1	3	5	2	7	9
4	1	6	3	3	9	6	3	4	8	2	3	1	9	4	4	6	8
4	1	3	8	9	5	5	7	3	9	5	9	8	5	3	2	2	3
7	4	7	0	3	6	3	5	9	9	5	5	5	0	4	6	8	0
4	7	0	3	3	0	9	5	7	9	5	1	6	6	3	8	8	8
9	2	9	1	8	5	4	4	5	0	2	3	9	7	1	2	0	3
6	3																

شکاف بین رقم شش (۶)

4	1	3	5	1	7	2	8	2	0	7	9	1	3	5	2	7	9
4	1	6	3	3	9	6	3	4	8	2	3	1	9	4	4	6	8
4	1	3	8	9	5	5	7	3	9	5	9	8	5	3	2	2	3
7	4	7	0	3	6	3	5	9	9	5	5	5	0	4	6	8	0
4	7	0	3	3	0	9	5	7	9	5	1	6	6	3	8	8	8
9	2	9	1	8	5	4	4	5	0	2	3	9	7	1	2	0	3
6	3																

شکاف بین رقم هفت (۷)

4	1	3	5	1	7	2	8	2	0	7	9	1	3	5	2	7	9
4	1	6	3	3	9	6	3	4	8	2	3	1	9	4	4	6	8
4	1	3	8	9	5	5	7	3	9	5	9	8	5	3	2	2	3
7	4	7	0	3	6	3	5	9	9	5	5	5	0	4	6	8	0
4	7	0	3	3	0	9	5	7	9	5	1	6	6	3	8	8	8
9	2	9	1	8	5	4	4	5	0	2	3	9	7	1	2	0	3
6	3																

شکاف بین رقم هشت (۸)

4	1	3	5	1	7	2	8	2	0	7	9	1	3	5	2	7	9
4	1	6	3	3	9	6	3	4	8	2	3	1	9	4	4	6	8
4	1	3	8	9	5	5	7	3	9	5	9	8	5	3	2	2	3
7	4	7	0	3	6	3	5	9	9	5	5	5	0	4	6	8	0
4	7	0	3	3	0	9	5	7	9	5	1	6	6	3	8	8	8
9	2	9	1	8	5	4	4	5	0	2	3	9	7	1	2	0	3
6	3																

شکاف بین رقم نه (۹)

4	1	3	5	1	7	2	8	2	0	7	9	1	3	5	2	7	9
4	1	6	3	3	9	6	3	4	8	2	3	1	9	4	4	6	8
4	1	3	8	9	5	5	7	3	9	5	9	8	5	3	2	2	3
7	4	7	0	3	6	3	5	9	9	5	5	5	0	4	6	8	0
4	7	0	3	3	0	9	5	7	9	5	1	6	6	3	8	8	8
9	2	9	1	8	5	4	4	5	0	2	3	9	7	1	2	0	3
6	3																

رقم تکراری	تعداد شکاف ها	طول شکاف																
		47	9	3	2	2	21	6										
0	7	47	9	3	2	2	21	6										
1	8	2	7	6	10	6	45	9	10									
2	8	1	6	12	22	0	38	8	4									
3	17	10	7	0	2	3	8	5	5	2	4	1	14	0	9	14	5	2
4	10	17	7	5	0	2	18	12	3	23	0							
5	13	10	26	0	3	2	11	2	0	0	12	2	12	2				
6	7	3	9	24	9	14	0	22										
7	8	4	5	26	10	1	16	6	22									
8	9	19	7	3	8	21	16	0	0	4								
9	13	5	5	7	8	4	1	14	0	14	2	8	1	9				

محاسبات لازم به طور استفاده از آماره مربع کای در جدول زیر نشان داده شده است.

طول شکاف (x)	احتمال $p(x)$	فراوانی انتظاری $E_i(x) = (110-10)p(x)$	فراوانی مشاهده شده (O_i)	$\frac{(O_i - E_i)^2}{E_i}$
۰	۰.۱۰	۱۰	۱۲	۰.۴
۱	۰.۰۹	۹	۶	۱.۰
۲	۰.۰۸۱	۸.۱	۱۱	۱.۰۴
۳	۰.۰۷۲۹	۷.۲۹	۶	۰.۲۳
۴	۰.۰۶۵۶۱	۶.۵۶۱	۵	۰.۳۷
۵	۰.۰۵۹۰۵	۵.۹۰۵	۸	۰.۷۴
۶	۰.۰۵۳۱۱	۵.۳۱۱	۴	۰.۳۲
۷-۸	۰.۰۹۰۸۸	۹.۰۸۸	۱۰	۰.۰۹
۹-۱۱	۰.۱۰۵۰۵	۱۰.۵۰۵	۱۲	۰.۲۱
۱۲-۱۵	۰.۰۹۷۱	۹.۷۱۰	۹	۰.۰۵
۱۶-۱۹	۰.۰۶۳۷	۶.۳۷۰	۵	۰.۲۹
۲۰-۲۷	۰.۰۷۱۳	۷.۱۳۰	۹	۰.۴۹
۲۸-۴۷	۰.۰۵۰۳	۵.۰۳۰	۳	۰.۸۲
جمع	۱.۰۰۰۰۰	۱۰۰.۰۰۰	۱۰۰	۶.۰۵

چون به ازای مقدار ۰.۰۵ برای سطح معنادار بودن و ۱۲ درجه آزادی، مقدار مندرج در جدول مربع کای

از ۶.۰۵ بزرگتر است ($6.05 < 21.0 = \chi^2_{12-1;0.05}$). لذا دنباله در دست بررسی از آزمون شکاف موفق بیرون

می آید.

آزمون روند

تعداد روندهای صعودی و نزولی

فرض کنید که ۴۰ عدد تصادفی زیر در یک دنباله به کارگیری یک الگوریتم تولید شده است:

0.08	0.09	0.23	0.29	0.42	0.55	0.58	0.72	0.89	0.91
0.11	0.16	0.18	0.31	0.41	0.53	0.71	0.73	0.74	0.84
0.02	0.09	0.30	0.32	0.45	0.47	0.69	0.74	0.91	0.95
0.12	0.13	0.29	0.36	0.38	0.54	0.68	0.86	0.88	0.91

در صورت اعمال آزمون های **مربع کای** و **کالموگروف-اسمیرنف**، فرض همگونی اعداد تصادفی فوق با تابع چگالی یکنواخت پذیرفته می شود. اما با یک نگاه به اعداد هر سطر می توان دید که ده عدد مندرج در هر سطر روندی صعودی دارد. در صورتی که ۴۰ عدد فوق در قالب دنباله زیر تولید می شد، تردیدی کمتر در مورد استقلال آن ها وجود می داشت.

0.41	0.68	0.89	0.84	0.74	0.91	0.55	0.71	0.36	0.30
0.09	0.72	0.86	0.08	0.54	0.02	0.11	0.29	0.16	0.18
0.88	0.91	0.95	0.69	0.09	0.38	0.23	0.32	0.91	0.53
0.31	0.42	0.73	0.12	0.74	0.45	0.13	0.47	0.58	0.29

هر آزمون روند، براساس ترتیب قرار گرفتن اعداد تصادفی در یک دنباله، به بررسی فرض استقلال می پردازد.

اساس آزمون روند را تعداد روندها، یا طول آن ها تشکیل می دهد. تعداد روند ها را می توان براساس ضابطه صعودی یا نزولی بودن روند تعریف کرد. در هر روند صعودی، هر عدد از عدد سمت راست خود

کوچکتر است و در یک روند نزولی، هر عدد از عدد سمت راست خود بزرگتر است. برای روشن شدن موضوع مثال زیر را در نظر بگیرید.

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					

همان طور که دیده می‌شود، بین هر دو عدد متوالی، یک علامت مثبت (+) و منفی (-) نوشته شده است. مثلاً، بین دو عدد نخست (یعنی ۰.۸۷ و ۰.۱۵)، علامت منفی قرار داده شده است. از این علامت چنین بر می‌آید که عدد دوم کوچکتر از عدد اول است. چون دنباله فوق شامل ۱۵ عدد تصادفی است تعداد علائم + و - برابر ۱۴ است. دنباله ۱۴ علامت به شرح زیر است:

-	+	+	+	-	-	-	+	-	+	+	+	-	+
---	---	---	---	---	---	---	---	---	---	---	---	---	---

تعداد روند مساوی ۸ است. طول هشت روند مزبور، به ترتیب، مساوی با ۱، ۱، ۳، ۱، ۱، ۳، ۳، ۱ است.

-	+	+	+	-	-	-	+	-	+	+	+	-	+
1	2			3			4	5	6		7	8	

تعداد روندهای صعودی و نزولی یکسان و مساوی چهار است. تعداد روندها ممکن است بیش از حد کم یا زیاد باشد. مثلاً در دنباله زیر تنها یک روند و آن هم روند صعودی وجود دارد.

0.02	+	0.08	+	0.18	+	0.23	+	0.36	+	0.42	+	0.55	+	0.63	+	0.72	+	0.89
------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------

تولید یک چنین دنباله ای با استفاده از یک مولد معتبر چندان محتمل نیست. به عنوان مثال دیگری، دنباله ده عدد تصادفی زیر را در نظر بگیرید که دارای ۹ روند است که ۵ تای آن صعودی و ۴ تای دیگر نزولی است.

0.08	+	0.93	-	0.15	+	0.96	-	0.26	+	0.84	-	0.28	+	0.79	-	0.36	+	0.57
------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------	---	------

احتمال وجود این تعداد روند در یک دنباله ده تایی از اعداد تصادفی نیز چندان زیاد نیست. آنچه بیشتر احتمال دارد، این است که تعداد روندهای بین دو حد فوق قرار گیرد. اگر یک دنباله از n عدد تصادفی تشکیل شود، **بیشترین** تعداد روندهای مساوی $n-1$ و **کمترین** تعداد روندها مساوی یک خواهد بود.

اگر متغیر تصادفی a معرف مجموع تعداد روندهای موجود در یک دنباله n تایی از اعداد تصادفی باشد، میانگین و واریانس آن به شرح زیر تعریف می‌شود.

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

به علاوه نشان داده می‌شود که به ازای $n > 20$ ، متغیر تصادفی a به سمت تابع **توزیع نرمال** با میانگین μ_a و واریانس σ_a^2 میل می‌کند. از تقریب اخیر می‌توان برای آزمایش استقلال اعداد تصادفی هر دنباله استفاده کرد. برای این منظور از آماره زیر استفاده می‌شود.

$$Z_0 = \frac{a - \mu_a}{\sigma_a} = \frac{a - \left[\frac{(2n-1)}{3} \right]}{\sqrt{\frac{(16n-29)}{90}}}$$

اگر سطح معنا دار بودن را a نمادگذاری کنیم، عبارت $-Z_{\frac{\alpha}{2}} \leq Z_0 \leq Z_{\frac{\alpha}{2}}$ ناظر بر استقلال اعداد تصادفی است.

مثال ۱۲: به ازای $\alpha = 0.05$ تعیین کنید آیا دنباله متشکل از ۴۰ عدد تصادفی زیر را بر اساس تعداد روندهای صعودی و نزولی می‌توان دارای خاصیت استقلال دانست یا نه؟

0.41	0.68	0.89	0.94	0.74	0.91	0.55	0.62	0.36	0.27
0.19	0.72	0.75	0.08	0.54	0.02	0.01	0.36	0.16	0.28
0.18	0.01	0.95	0.69	0.18	0.47	0.23	0.32	0.82	0.53
0.31	0.42	0.73	0.04	0.83	0.45	0.13	0.57	0.63	0.29

توالی روندهای صعودی و نزولی به شرح زیر است.

+	+	+	-	+	-	+	-	-	-	+	+	-	+	-	-	+	-	+	-	-
+	-	-	+	-	+	+	-	-	+	+	-	+	-	-	+	+	-			

تعداد روندها در دنباله فوق بالغ بر ۲۶ است. براساس مقادیر $n = 40$ و $a = 26$ ، مقادیر نظیر μ_a ، σ_a^2

و Z_0 عبارت است از:

$$\mu_a = \frac{2(40) - 1}{3} = 26.33$$

$$\sigma_a^2 = \frac{16(40) - 29}{90} = 6.79$$

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

چون مقدار مندرج در جدول نرمال برای $Z_{0.025}$ ، مساوی ۱.۹۶ است، فرض استقلال اعداد تصادفی را

نمی‌توان رد کرد.

تعداد روندهای بزرگتر و کوچکتر از میانگین

در زمینه آزمایش استقلال یک دنباله از اعداد تصادفی، آزمودن تعداد روندهای صعودی و نزولی کافی

نیست. دنباله صفحه بعد را در نظر بگیرید.

0.63	0.72	0.79	0.81	0.52	0.94	0.83	0.93	0.87	0.67
0.54	0.83	0.89	0.55	0.88	0.77	0.74	0.95	0.82	0.86
0.43	0.32	0.36	0.18	0.08	0.19	0.18	0.27	0.36	0.34
0.31	0.45	0.49	0.43	0.46	0.35	0.25	0.39	0.47	0.41

+	+	+	-	+	-	+	-	-	-	+	+	-	+	-	-	+	-	+	-	-
+	-	-	+	-	+	+	-	-	+	+	-	+	-	-	+	+	-			

این دنباله، درست مانند دنباله مثال قبل است. بنابراین دنباله فوق از آزمون تعداد روندهای صعودی و نزولی موفق بیرون می‌آید و استقلال آن پذیرفته می‌شود. اما قابل تفکر است که ۲۰ عدد اول همگی بزرگتر از میانگین $\left[\frac{(0.99+0.00)}{2} = 0.495 \right]$ و ۲۰ عدد دیگر کوچکتر از میانگین است. احتمال رخ دادن چنین اتفاقی بسیار ناچیز است. تا تغییر دادن تعریف روند، می‌توان از شیوه تجربه و تحلیل عرضه شده بالا به منظور بررسی وضعیتی نظیر آنچه در این جا مطرح است استفاده کرد. بنابراین، روند را به عنوان تداوم مقادیر اعداد تصادفی یک دنباله در سطحی بالاتر از میانگین یا در سطحی پایین تر از میانگین تعریف می‌کنیم. هر مشاهده که از میانگین کوچکتر باشد با علامت «-» مشخص می‌شود و هر مشاهده که از میانگین بزرگتر باشد با علامت «+» مشخص می‌شود. مثلاً، دنباله ۲۰ تایی زیر را در نظر بگیرید

0.40	0.84	0.75	0.18	0.13	0.92	0.57	0.77	0.30	0.71
0.42	0.05	0.78	0.74	0.68	0.03	0.18	0.51	0.10	0.37

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

-	+	+	-	-	+	+	+	-	+	-	-	+	+	+	-	-	+	-	-
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

-	+	+	-	-	+	+	+	-	+	-	-	+	+	+	-	-	+	-	-
1	2	3	4	5	6	7	8	9	10	11									

در این مورد جمعا یازده روند وجود دارد. طول اولین روند مساوی یک است. طول روند بعدی مساوی دو است و برای مابقی به همین صورت ادامه می‌دهیم. فرض کنید تعداد مشاهدات بزرگتر و کوچکتر از میانگین، به ترتیب با n_1 و n_2 نمادگذاری می‌شود. در این صورت بیشترین تعداد روندهای در یک دنباله n تایی مساوی $n = n_1 + n_2$ و کمترین تعداد آن‌ها مساوی یک خواهد بود. اگر متغیر تصادفی b معرف تعداد روندهای بزرگتر و کوچکتر از میانگین باشد، در می‌یابیم که میانگین (μ_a) ، و واریانس (σ_a^2) ، به شرح زیر تعریف می‌شود.

$$\mu_b = \frac{2n_1n_2}{n} + \frac{1}{2}$$

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

همچنین نشان داده می‌شود که اگر n_1 و یا n_2 از ۲۰ بزرگتر باشد، توزیع احتمال b به سمت تابع توزیع نرمال با میانگین μ_b و واریانس σ_b^2 میل می‌کند. بنابراین، می‌توان آماره نرمال استاندارد زیر را به منظور بررسی استقلال اعداد موجود در هر دنباله تعریف کرد:

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

اگر سطح معنا دار بودن را α نمادگذاری کنیم، عبارت $-Z_{\frac{\alpha}{2}} \leq Z_0 \leq Z_{\frac{\alpha}{2}}$ ناظر بر استقلال توالی اعداد مورد نظر است.

مثال ۱۳: تعیین کنید آیا تعداد روندهای بزرگتر یا کوچکتر از میانگین برای دنباله مثال ۱۲ به طور

غیرعادی زیاد است یا نه. توالی علائم مثبت و منفی به شرح زیر است:

-	+	+	+	+	+	+	+	-	-	-	+	+	-	+	-	-	-	-	-	-	+	+
-	-	-	-	+	+	-	-	+	-	+	-	-	+	+	-							

به ازای مقادیر $n = 40, n_1 = 18, n_2 = 22, b = 17$ مقادیر μ_b, σ_b^2, Z_0 به شرح زیر محاسبه می‌شود:

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

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

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

تذکر می‌دهیم که تنها به دلیل صدق رابطه $n_2 > 20$ می‌توان از تقریب نرمال استفاده کرد. چون مقدار مندرج در جدول نرمال برای $Z_{0.025}$ مساوی ۱.۹۶ است، نمی‌توان فرض استقلال اعداد دنباله فوق را رد کرد و به عبارتی می‌توان فرض استقلال را پذیرفت.

آزمون روند براساس طول روند

از مشخصه های مهم یک روند، طول است. مثلاً دنباله زیر را در نظر بگیرید:

0.16	0.27	0.58	0.63	0.45	0.21	0.72	0.87	0.27	0.15	0.92	0.85
------	------	------	------	------	------	------	------	------	------	------	------

در این دنباله، دو عدد کوچکتر از میانگین و دو عدد بزرگتر از میانگین است و این روند تا به آخر ادامه دارد. فرض کنید دنباله فوق با روند مورد بحث باز هم ادامه یابد. با اعمال آزمون تعداد روندهای بزرگتر و کوچکتر از میانگین، دنباله فوق دارای خاصیت استقلال تشخیص داده خواهد شد. اما، باید انتظار داشت که طول روند ها تنها معادل ۲ نباشد و مقادیر دیگری را نیز بپذیرد.

فرض کنید تعداد روندهایی با طول i در یک دنباله n تایی از اعداد تصادفی، Y_i باشد، در می‌یابیم که در یک دنباله مستقل و براساس ضابطه روندهای صعودی و نزولی، امید ریاضی Y_i به شرح زیر تعریف می‌شود:

$$E(Y_i) = \frac{2[n(i^2 + 3i + 1) - (i^3 + 3i^2 - i - 4)]}{(i + 3)!}, \quad i \leq n - 2$$

$$E(Y_i) = \frac{2}{n!}, \quad i = n - 1$$

اگر ضابطه روندهای بزرگتر و کوچکتر از میانگین ملاک عمل قرار گیرد، امید ریاضی Y_i به طور تقریبی از رابطه

$$E(Y_i) = \frac{nw_i}{E(I)} \quad n > 20$$

به دست می‌آید که در آن w_i معرف احتمال تقریبی مساوی طول روند با i و $E(I)$ نیز امید ریاضی تقریبی طول هر روند است. w_i و $E(I)$ به شرح زیر تعریف می‌شود.

$$w_i = \left(\frac{n_1}{n}\right)^i \left(\frac{n_2}{n}\right) + \left(\frac{n_1}{n}\right) \left(\frac{n_2}{n}\right)^i, \quad n > 20$$

$$E(I) = \frac{n_1}{n_2} + \frac{n_2}{n_1}, \quad n > 20$$

امید ریاضی تقریبی مجموع تعداد تمام روندها (با هر طول) در هر دنباله n تایی، $E(A)$ ، از رابطه

$$E(A) = \frac{n}{E(I)} \quad n > 20$$

بدست می‌آید. آزمون مناسب در این زمینه، آزمون مربع کای است. اگر تعداد روندهای مشاهده شده با طول i با O_i نمادگذاری شود، آماره آزمون عبارت خواهد بود از:

$$\chi_0^2 = \sum_{i=1}^L \frac{[O_i - E(Y_i)]^2}{E(Y_i)}$$

که در آن L براساس ضابطه روندهای صعودی و نزولی، به صورت $L = n - 1$ و بر اساس ضابطه روندهای بزرگتر و کوچکتر از میانگین، به صورت $L = n$ تعریف می‌شود. اگر فرض استقلال صحیح باشد، χ_0^2 تقریباً طبق **توزیع مربع کای** با $L - 1$ درجه آزادی تعریف می‌شود.

مثال ۱۴: آیا می‌توان دنباله اعداد تصادفی زیر را بر اساس طول روندهای صعودی و نزولی و به ازای

$a = 0.05$ مستقل دانست؟

0.30	0.48	0.36	0.01	0.54	0.34	0.96	0.06	0.61	0.85
0.48	0.86	0.14	0.86	0.89	0.37	0.49	0.60	0.04	0.83
0.42	0.83	0.37	0.21	0.90	0.89	0.91	0.79	0.57	0.99
0.95	0.27	0.41	0.81	0.96	0.31	0.09	0.06	0.23	0.77
0.73	0.47	0.13	0.55	0.11	0.75	0.36	0.25	0.23	0.72
0.60	0.84	0.70	0.30	0.26	0.38	0.05	0.19	0.73	0.44

دنباله علایم + و - برای این مثال به شرح زیر است:

+	-	-	+	-	+	-	+	+	-	+	-	+	+	-	+	+	-	+	-	-	+	-	+	-	-	+	-	-
+	+	+	-	-	-	+	+	-	-	-	+	-	+	-	-	-	+	-	+	-	-	-	+	-	+	+	-	

طول روندها نیز به شرح زیر است:

1	2	1	1	1	1	2	1	1	1	2	1	2	1	1
1	1	2	1	1	1	2	1	2	3	3	2	3	1	1
1	2	1	1	1	2	1	1	2	1					

فراوانی مشاهده شده برای روندهایی به طول ۱، ۲، و ۳ در جدول زیر درج شده است:

طول روند (i)	1	2	3
فراوانی مشاهده شده (O _i)	26	9	5

در زیر، نحوه محاسبه امید ریاضی تعداد روندها با طول مختلف نشان داده شده است.

$$E(Y_1) = \frac{2}{4!} [60(1+3+1) - (1+3-1-4)] = 25.08$$

$$E(Y_2) = \frac{2}{5!} [60(4+6+1) - (8+12-2-4)] = 10.77$$

$$E(Y_3) = \frac{2}{6!} [60(9+9+1) - (27+27-3-4)] = 3.04$$

میانگین مجموع تعداد روندهای صنعتی و نزولی عبارت است:

$$\mu_a = \frac{[2(60)-1]}{3} = 39.67$$

مجموع $E(Y_i)$ برای $i = 1, 2, 3$ برابر ۳۸.۸۹ است. بنابراین، امید ریاضی تعداد روندها که طولشان از

۳ بیشتر است، از تفاضل $\mu_a - \sum_{i=1}^3 Y_i = 0.78$ بدست می‌آید. قبلاً توضیح داده شد که ضابطه واحدی برای

کمترین مقدار فراوانی انتظاری وجود ندارد. اگر در این مورد، کمترین مقدار فراوانی مساوی ۵ در نظر گرفته

شود، باید رده ای را که فراوانی انتظاری آن کمتر از ۵ است در یکی از رده های مجاور ادغام کرد. بدیهی

است که فراوانی های مشاهده شده نظیر نیز باید در هم ادغام و یک واحد از L کم شود. با رعایت این

ملاحظات، جدول زیر به دست می‌آید.

طول روند (i)	فراوانی مشاهده شده (O _i)	فراوانی انتظاری (E(Y) _i)	$\frac{[O_i - E(Y_i)]^2}{E(Y_i)}$
1	26	25.08	0.03
2	9	10.77	0.02
≥ 3	5	3.82	
	40	39.67	0.05

مقدار استخراج شده از جدول مربع کای برای مساوی ۳.۸۴ است. نظر به این که

$\chi^2_{0.05;1} = 3.84 < \chi^2_0 = 0.05$ است، براساس آزمون فوق نمی‌توان استقلال را مردود دانست.

آزمون همبستگی

آزمون های همبستگی ناظر به بررسی استقلال اعداد هر دنباله است. مثلاً، دنباله زیر را در نظر بگیرید:

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

در نگاه اول، به نظر می‌رسد که اعداد فوق تصادفی است و احتمالا از تمام آزمون های ارایه شده در این فصل موفق بیرون خواهد آمد. ولی با دقت بیشتر معلوم می‌شود که اعداد قرار گرفته در موقعیت هایی که مضربی از ۵ است، از نظر مقدار بسیار بزرگ است. بدیهی است که نمونه ای به کوچکی ۳۰ عدد تصادفی برای رد کردن مولد کافی نیست ولی نکته این است که اعداد موجود در دنباله فوق ممکن است به هم مرتبط باشد.

در این آزمون، به بررسی همبستگی موجود میان اعداد i ام به بعد در یک دنباله است که هر دو تای آنها فاصله ای معادل k از یکدیگر داشته باشد. به این ترتیب، همبستگی میان $R_i, R_{i+k}, R_{i+2k}, \dots, R_{i+(K+1)k}$ یا ρ_{ik} مورد توجه است. اگر تعداد اعداد موجود در دنباله مساوی با n باشد، K بزرگترین عدد صحیح مثبتی است که در رابطه $i + (K+1)k \leq n$ صدق کند. در این آزمون، تنها یک زیر مجموعه $(K+2)$ تایی از اعداد مورد بررسی قرار می‌گیرد.

چون همبستگی غیرصفر مترادف با عدم وجود استقلال است آزمون فرض مناسب به صورت

$$\begin{cases} H_0 : \rho_{ik} = 0 \\ H_1 : \rho_{ik} \neq 0 \end{cases}$$

نوشته می‌شود. اگر $R_i, R_{i+k}, R_{i+2k}, \dots, R_{i+(K+1)k}$ ناهمبسته باشد، به ازای مقادیر بزرگ K توزیع احتمال برآورد کننده پارامتر ρ_{ik} ، یعنی $\hat{\rho}_{ik}$ ، تقریبا توزیع احتمال نرمال خواهد داشت. در این صورت، آماره آزمون به شکل

$$Z_0 = \frac{\hat{\rho}_{ik}}{\sigma_{\hat{\rho}_{ik}}}$$

نوشته می‌شود. اگر فرض استقلال صحیح و مقدار K بزرگ باشد آماره توزیع احتمال نرمال استاندارد خواهد داشت. فرمول محاسبه $\hat{\rho}_{ik}$ و $\sigma_{\hat{\rho}_{ik}}$ به شرح زیر است:

$$\hat{\rho}_{ik} = \frac{1}{K+1} \left[\sum_{m=0}^K R_{i+mk} \times R_{i+(m+1)k} \right] - \frac{1}{4}$$

$$\sigma_{\hat{\rho}_{ik}} = \frac{\sqrt{13K+7}}{12(K+1)}$$

اگر سطح معنا دار بودن را α نمادگذاری کنیم، عبارت $-Z_{\frac{\alpha}{2}} \leq Z_0 \leq Z_{\frac{\alpha}{2}}$ ناظر بر استقلال توالی اعداد

مورد نظر است.

مثال ۱۴: تحقیق کنید آیا اعداد سوم، هشتم، سیزدهم، هجدهم، و ... موجود در دنباله زیر همبستگی

دارد یا نه. سطح معناداری را ۰.۰۵ در نظر بگیرید.

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

چون اولین عدد زیر دنباله، سومین عدد است، لذا $i = 3$ است. به علاوه، $k = 5, K = 4, n = 30$ است.

مقدار K به این صورت محاسبه می شود که بزرگترین عدد صحیحی است که در مورد آن رابطه

$$3 + (K+1) \leq n$$

برقرار است. در این صورت داریم:

$$\hat{\rho}_{35} = \frac{1}{4+1} [(0.23)(0.28) + (0.28)(0.33) + (0.33)(0.27) + (0.27)(0.05) + (0.05)(0.36)] - \frac{1}{4}$$

$$= -0.1945$$

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

$$Z_0 = \frac{-0.1945}{0.128} = -1.516 > -Z_{0.025} = -1.96$$

بنابر محاسبات بالا، نمی توان فرض استقلال را مردود دانست.

تمرین‌ها

تمرین ۱: براساس روش همنهستی خطی، سه عدد صحیح تصادفی با دو رقم اعشار تولید کنید. فرض

کنید $X_0 = 27, a = 8, c = 47, m = 100$.

حل:

$$X_0 = 27, a = 8, c = 47, m = 100$$

$$X_1 = (8 \times 27 + 47) \bmod 100 = 63 \rightarrow R_1 = \frac{63}{100} = 0.63$$

$$X_2 = (8 \times 63 + 47) \bmod 100 = 51 \rightarrow R_2 = \frac{51}{100} = 0.51$$

$$X_3 = (8 \times 51 + 47) \bmod 100 = 55 \rightarrow R_3 = \frac{55}{100} = 0.55$$

تمرین ۲: با استفاده از مقادیر $X_0 = 117, a = 43, m = 1000$ ، چهار عدد تصادفی سه رقمی براساس

روش همنهستی ضربی تولید کنید.

حل:

$$X_0 = 117, a = 43, m = 1000$$

$$X_1 = (43 \times 117) \bmod 1000 = 31$$

$$X_2 = (43 \times 31) \bmod 1000 = 333$$

$$X_3 = (43 \times 333) \bmod 1000 = 319$$

$$X_4 = (43 \times 319) \bmod 1000 = 717$$

تمرین ۳: دنباله اعداد تصادفی 0.68, 0.11, 0.98, 0.73, 0.54 تولید شده است. به ازای مقدار ۰.۰۵

برای سطح معنادار بودن، براساس آزمون کالموگروف-اسمیرنوف فرض همگونی توزیع احتمال اعداد فوق با

تابع چگالی یکنواخت صفر و یک را انجام دهید.

حل:

$R_{(i)}$	0.11	0.54	0.68	0.73	0.98
$\frac{i}{N}$	0.20	0.40	0.60	0.80	1.0
$\frac{i}{N} - R_{(i)}$	0.09	-	-	0.07	0.02
$R_{(i)} - \frac{(i-1)}{N}$	0.11	0.34	0.28	0.13	0.18

$$D^+ = \max_{1 \leq i \leq N} \left(\frac{i}{N} - R_{(i)} \right) = 0.09$$

$$D^- = \max_{1 \leq i \leq N} \left(R_{(i)} - \frac{i-1}{N} \right) = 0.34$$

$$D = \max(D^+, D^-) = 0.34$$

مقدار $D_{0.05} = 0.565$ است که از روی جدول بدست می‌آید. با توجه به این که

$0.34 = D < D_{0.05} = 0.565$ است لذا فرضیه توزیع یکنواخت را نمی‌توان با احتمال ۹۵ درصد رد کرد.

تمرین ۴: یکصد عدد دورقمی موجود در دنباله **مثال ۹** را برعکس بنویسید. در این دنباله اولین عدد ۰.۳۴ است که در این صورت به ۰.۴۳ تغییر می‌یابد. به ازای مقدار ۰.۰۵ برای سطح معنادار بودن، براساس آزمون مربع کای، فرض همگونی توزیع احتمال اعداد دنباله تغییر یافته با تابع چگالی یکنواخت صفر و یک را انجام دهید.

حل:

ده بازه که به صورت از $10i - 9$ تا $10i$ که در آن $i = 1, 2, \dots, 10$ است در نظر بگیرید. با شمارش اعداد (بدون در نظر گرفتن اعشار) که در این ده بازه قرار می‌گیرند و همچنین مقدار $E_i = 10$ ، می‌توان جدول زیر را تولید کرد:

شماره بازه	ابتدای بازه	انتهای بازه	O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
1	01	10	9	10	0.1
2	11	20	9	10	0.1
3	21	30	9	10	0.1
4	31	40	6	10	1.6
5	41	50	17	10	4.9
6	51	60	5	10	2.5
7	61	70	10	10	0.0
8	71	80	12	10	0.4
9	81	90	7	10	0.9
10	91	100	16	10	3.6
جمع			100		14.2

از جدول فوق نتیجه می‌شود که $\chi_o^2 = 14.2$. از جدول تابع توزیع مربع کای داریم که $\chi_{0.05}^2 = 16.9$. لذا $14.2 = \chi_o^2 < \chi_{0.05}^2 = 16.9$ داریم و می‌توان این نتیجه را گرفت که نمی‌توان فرضیه یکنواخت بودن تابع توزیع این ۱۰۰ عدد را رد کرد.

تمرین ۵: پنجاه عدد دو رقمی نخست از **مثال ۹** را در نظر بگیرید. به ازای $a = 0.05$ و براساس تعداد روندهای صعودی و نزولی آیا می‌توان استقلال این اعداد را مردود دانست؟

حل:

این دنباله دارای ۲۷ روند یا ران یا Run به صورت زیر دارد:

+	-	+	-	-	-	+	+	+	+	-	+	-	+	+	+	-	-	+	+
1	2	3	4	5	6	7	8	9	10	11									
+	-	-	-	-	+	+	+	-	+	-	+	-	-	+	-	+	-	+	-
			12	13	14	15	16	17	18	19	20	21	22	23	24				
-	-	+	+	-	-	+	+												
			25	26	27														

با توجه به $N = 50, a = 27$ داریم

$$\mu_a = \frac{(2N - 1)}{3} = 33$$

$$\sigma_a^2 = \frac{(16N - 29)}{90} = 8.5667$$

$$Z_0 = \frac{(a - \mu_a)}{\sigma_a} = -2.05 < -1.96 = -Z_{0.025}$$

لذا فرض استقلال اعداد را می‌توان رد کرد.

برای دریافت بسته‌های آموزشی گروه **بهینه‌یاب** به وب سایت ما به نشانی

www.behinehyab.com مراجعه کنید.

در صورت هر گونه سوال از طریق ایمیل به نشانی **behinehyab@gmail.com** و یا

بخش "تماس با ما" وب سایت گروه **بهینه‌یاب** با ما در تماس باشید.

با تشکر از توجه شما

گروه آموزشی **بهینه‌یاب**