درس ۲۵:

تولید اعداد تصادفی

تهیه شده توسط گروه بهینه‌یاب

www.behinehyab.com

مقدمه

روش‌های متفاوتی برای تولید اعداد تصادفی وجود دارد. تمام این روش ها یک وجه مشترک دارند: اینکه تمام آن‌ها از برنامه های 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.

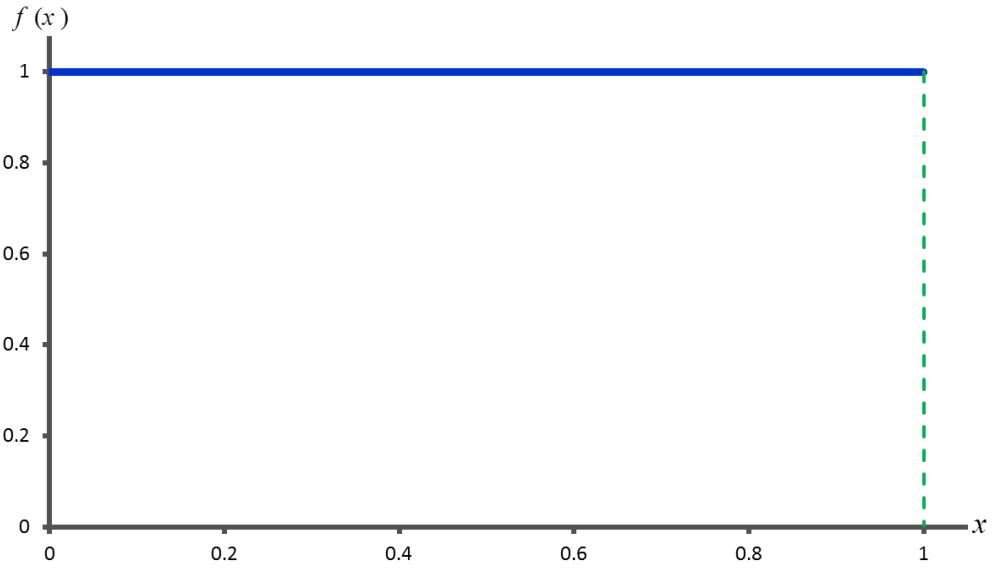
در این درس به ارایه برخی از روش‌های محاسباتی تولید اعداد تصادفی می‌پردازیم، ضوابط ارزیابی، مقایسه و انتخاب را عرضه می‌داریم و سرانجام آزمایش های مربوط به تصادفی بودن اعداد به دست آمده از مولدهای را معرفی می‌کنیم.

خواص اعداد تصادفی

هر دنباله از اعداد تصادفی مانند R_1، R_2 و ... 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.

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

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



شکل۱: تابع چگالی اعداد تصادفی

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

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

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

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

در باره پیامدهای خاصیت **استقلال** و خاصیت توزیع احتمال **یکنواخت** پیوسته برای اعداد تصادفی، می‌توان دو مورد زیر را بیان کرد:

۱- اگر فاصله (۰،۱) به n رده با زیر فاصله **مساوی** تقسیم شود، انتظار می رود که از N مشاهده $\frac{N}{n}$ آن‌ها 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 می برد. 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، مسلما 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 ارایه 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 پیروی** اعداد بدست آمده از یک مولد از تابع چگالی یکنواخت، می‌توان از آزمون های معرفی شده که در این درس ارایه می‌شود 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 می‌کنیم. قبل از تشریح برخی از روش های تولید اعداد تصادفی، بعضی از ملاحظات مهم را فهرست وار ارایه می‌کنیم.

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 الگوریتمی‌پرداخت 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 می‌شود.

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 می‌شود.

روش میان مربعی

این روش با یک عدد اولیه به نام هسته شروع به کار می‌کند. 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 می‌کنند. 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 می‌کند. اگر هسته n رقمی باشد، مربع آن ۱-۲n یا ۲n رقمی خواهد بود و اگر n زوج باشد با حذف $n/2$ ارقام از هر سمت (راست و چپ) مربع آن، می‌توان ارقام میانی را تعیین کرد.

مثال ۱: فرض کنید که به دنباله از اعداد تصادفی چهار رقمی نیاز داریم. هسته را با X_0 نماد گذاری کنید و مقدار ۵۴۹۷ را برای آن در نظر بگیرید. فرض کنید i-امین عددی که مربع می‌شود X_i و i امین عدد تصادفی که تولید می‌شود R_i باشد.

$$X_0 = 5497$$
$$X_0^2 = (5497)^2 = 30\boxed{2170}09 \Rightarrow X_1 = 2170 \Rightarrow R_1 = 0.2170$$
$$X_1^2 = (2170)^2 = 04\boxed{7089}00 \Rightarrow X_2 = 7089 \Rightarrow R_2 = 0.7089$$
$$X_2^2 = (7089)^2 = 50\boxed{2539}21 \Rightarrow X_3 = 2539 \Rightarrow R_3 = 0.2539$$
$$\dots$$

نکته: این روش می‌تواند دچار مشکلاتی شود. اولا نمی‌توان قواعد ساده‌ای 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 می‌رسد. 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:

$$X_0 = 5197$$
$$X_0^2 = (5197)^2 = 27\boxed{0088}09 \Rightarrow X_1 = 0088 \Rightarrow R_1 = 0.0088$$
$$X_1^2 = (0088)^2 = 00\boxed{0077}44 \Rightarrow X_2 = 0077 \Rightarrow R_2 = 0.0077$$
$$X_2^2 = (0077)^2 = 00\boxed{0059}29 \Rightarrow X_3 = 0059 \Rightarrow R_3 = 0.0059$$
$$\dots$$

می‌بینیم که تمام R_i ها با صفر شروع می‌شود.

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=۴ داریم:

$$X_i = 6500$$
$$X_i^2 = (6500)^2 = 42\boxed{2500}00 \Rightarrow X_{i+1} = 2500 \Rightarrow R_{i+1} = 0.2500$$
$$X_{i+1}^2 = (2500)^2 = 06\boxed{2500}00 \Rightarrow X_{i+2} = 2500 \Rightarrow R_{i+2} = 0.2500$$
$$X_{i+2}^2 = (2500)^2 = 06\boxed{2500}00 \Rightarrow X_{i+3} = 2500 \Rightarrow R_{i+3} = 0.2500$$
$$\dots$$

با ثابت ماندن مقدار اعداد تصادفی تولد شده، الگوریتم عملا از هم می‌پاشد.

روش میان ضربی

این روش همانند روش میان مربعی است و با انتخاب دو هسته، X_0 و X_0' که تعداد ارقامشان مساوی است شروع می‌شود. اگر دو هسته n رقمی باشد، آن‌ها را در هم ضرب می‌کنیم و با حذف $n/2$ از ارقام سمت راست و همین تعداد رقم از سمت چپ حاصلضرب، n رقم قرار گرفته در وسط را تعیین و آن را تبدیل به یک عدد اعشاری کوچکتر از یک می‌کنیم. 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_0 و X_1 تکرار می‌کنیم و این روند تکرار می‌شود. برای روشن شدن موضوع مثال زیر را در نظر بگیرید.

مثال ۴: از دو هسته ۷۲۲۹=X_0 و ۲۹۳۸=X_0' 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 عمده ی 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 می شد می‌توان اعداد را تولید کرد. فرض کنید عدد ثابت را با K نماد گذاری می‌کنیم و تعداد ارقام K و X_0 را یکسان در نظر می‌گیریم. تحت این شرایط، تمام محاسبات مانند روش میان ضربی انجام می گیرد.

مثال ۵: با استفاده از دو هسته K=۳۹۸۷ و X_0=۷۲۲۳ و براساس روش مضرب ثابت، دنباله ای 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 کاملا 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 می‌کند که یک دنباله n تایی مانند X_1، X_2 و ... و X_n را دریافت و بقیه دنباله، یعنی X_{n+1}، X_{n+2} و ... را تولید می‌کند. اساس تولید مقدار در این روش استفاده از رابطه

$$X_i \equiv \left(X_{i-1} + X_{i-n}\right) \mod m, \quad i = n+1, n+2, \ldots$$

است. طبق تعریف رابطه $a \equiv (b) \mod m$ 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 می‌دهد.

مثال ۶: اگر مقدار ۴ برای پیمانه یعنی m، در نظر گرفته شود، اعداد ۲و۵و۱۰و۱۴ و ... معادل هم هستند زیرا عددی مانند ۱۰ وجود دارد که باقی مانده تقسیم (۲-۱۰)،(۶-۱۰)، 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-۱ خواهد بود. بنابراین داریم

$$17 \mod 3 = 2$$
$$14 \mod 5 = 4$$
$$37 \mod 2 = 1$$
$$16 \mod 4 = 0$$
$$\ldots$$

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.

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

روش همنهشتی خطی

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

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

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

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

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

اگر مقدار عددی هسته را با X_0 نمادگذاری کنیم، مولد فوق الذکر به صورت رابطه زیر می‌شود.

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

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

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

$$\vdots$$

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

مثال ۸: مقادیر زیر را در نظر بگیرید

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

براساس روش همنهشتی 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، می‌توان از رابطه زیر استفاده کرد.

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

با استفاده از روش همنهشتی خطی داریم:

$$X_0 = 27$$
$$X_1 \equiv [17(27) + 43] \bmod 100 = 502 \bmod 100 = 2 \Rightarrow R_1 = 0.02$$
$$X_2 \equiv [17(2) + 43] \bmod 100 = 77 \bmod 100 = 77 \Rightarrow R_2 = 0.77$$
$$X_3 \equiv [17(77) + 43] \bmod 100 = 1352 \bmod 100 = 52 \Rightarrow R_2 = 0.52$$
$$\vdots$$

نظیر هر مولد دیگر، در مورد مولدهای همنهشتی 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 می‌تواند اعدادی با خواص مورد نظر تولید کند یا نه، باید به آزمون هایی را انجام داد. آزمون های مورد بحث را می‌توان 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 می‌پردازند. با توجه به همیت این دو آزمون به خصوص آزمون مربع کالی، به تفصیل به بررسی آن می‌پردازیم.

توجه: برای مطالعه ادامه آموزش روش‌های ارزیابی استقلال و یکنواختی تابع توزیع اعداد تصادفی و حل تمرین‌های متنوع، جزوه این درس را از طریق وب سایت بهینه یاب به نشانی *www.behinehyab.com* تهیه فرمایید.

برای دریافت بسته‌های آموزشی گروه **بهینه یاب** 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 **بهینه یاب** با ما در تماس باشید.

با تشکر از توجه شما

گروه آموزشی **بهینه یاب**