بخش دوم: آمار

دیدیم که تئوری احتمال یک تئوری کاملاً ریاضی بود که بر اساس یک اصول اولیه و استنتاجات ریاضی از یک مدل مفروض (احتمال یک سری وقایع)، احتمال وقایع دیگری را به دست می‌آوردیم. مدلها فرض شده‌اند و کاری به اینکه با واقعیت تطابق داشته باشند یا نه، نداریم. آمار در مورد کاربرد تئوری احتمال در مسائل واقعی بحث می‌کند و کمک می‌کند که به وسیلهٔ نتایج مشاهدات واقعی (نمونه‌هایی از جامعه) بتوانیم استنتاجاتی (که به احتمال نزدیک یک درست هستند) انجام دهیم. البته این آمار استنتاجی است که بحث ما در مورد آن خواهد بود. شاخهٔ دیگر آمار، آمار توصیفی است که در مورد نحوهٔ دسته‌بندی و ارائهٔ اطلاعات آماری بحث می‌کند (رجوع کنید به کتاب Kreyszig، صفحات ۸۳۸ تا ۸۴۹).

مبحث عمدهٔ آمار استنتاجی، تخمین و آزمون فرضیه است که هر دو کاربرد زیادی در مهندسی برق دارند. مانند سیستم هدایت و کنترل موشک، آشکارسازی هدف در رادار، تعقیب اهداف راداری و ...

- آمار
  - توصیفی
  - استنتاجی
    - تخمین
    - آزمون فرضیه

فصل ۸:

تولید اعداد تصادفی و شبیه‌سازی کامپیوتری

۱. تعریف دنبالۀ اعداد تصادفی

۲. تولید دنبالۀ اعداد تصادفی با توزیع یکنواخت

۳. تولید دنبالۀ اعداد تصادفی با توزیع دلخواه

۴. شبیه‌سازی مونت‌کارلو

تعریف دنبالهٔ اعداد تصادفی:

به جای اینکه سیستم فیزیکی را که دارای ورودی‌های تصادفی یا رفتار تصادفی است عیناً مشاهده کنیم و اثرات پارامترهای مختلف را ببینیم یا با تحمل هزینهٔ فراوان بسازیم (و بعد بفهمیم آن طور که مطلوب است کار نمی‌کندا)، می‌توانیم از شبیه‌سازی کامپیوتری استفاده کنیم (مانند مراحل گیرندهٔ مخابراتی، آشکارساز رادار و ...). امروزه شبیه‌سازی در سراسر دنیا و در رشته‌های مختلف مهندسی کاربرد فراوان دارد. مثلاً طرح اتومبیل ابتدا در کامپیوتر، تعیین محل قراردادن دکلهای مخابرات سیار با شبیه‌سازی امواج و ... .

خصوصاً در مواردی که محاسبهٔ مستقیم مشکل است یا پارامترهای متعددی در مسأله دخیل هستند یا به عنوان مؤیدی برای محاسبات.

به منظور شبیه‌سازی احتیاج به مولد اعداد تصادفی داریم. همچنین برای نمونه‌برداری تصادفی از یک جامعه از اعداد تصادفی استفاده می‌شود. بعضاً برای حل مسائل عددی مشکل از اعداد تصادفی (شبیه‌سازی مونت‌کارلو) استفاده می‌شود. تولید اعداد تصادفی کاربرد بسیار مهمی در ایمن ساختن مخابرات (رمزنگاری) دارد.

دنبالهٔ اعداد تصادفی:

دنباله‌ای از نمونه‌های متغیرهای تصادفی i.i.d. را دنبالهٔ اعداد تصادفی گویند: $x_1, x_2, x_3, \ldots, x_n, \ldots$ (هر دنبالهٔ نمونهٔ فرآیند $\mathbf{x}_n(\omega)$ مشروط بر i.i.d. بودن $\mathbf{x}_i$ها).

طبیعتاً تولید اعداد تصادفی فقط توسط یک پدیدهٔ فیزیکی تصادفی امکان‌پذیر است. آنچه به وسیلهٔ کامپیوتر می‌توان تولید کرد، دنبالهٔ اعداد شبه تصادفی (Pseudo Random) است، یعنی توسط الگوریتمی به طور Deterministic تولید می‌شوند، ولی خواص دنبالهٔ اعداد تصادفی را دارا هستند و لذا تصادفی به نظر می‌رسند. آزمونهایی وجود دارد که تصادفی بودن دنباله را بررسی می‌کنند و در فصل ۸ کتاب آمده‌اند.

برای اینکه دنبالهٔ اعداد شبه تصادفی (از این به بعد تسامحاً می‌گوییم تصادفی) با توزیع f_x داده شده را تولید کنیم، ابتدا تولید دنبالهٔ اعداد تصادفی با توزیع یکنواخت را بررسی می‌کنیم.

تولید دنبالۀ اعداد تصادفی با توزیع یکنواخت:

تولید دنبالۀ اعداد تصادفی با توزیع یکنواخت روشهای گوناگونی دارد و در این زمینه بسیار کار شده است. یک روش که در عین ساده بودن، دنبالۀ حاصل از آن از نظر تصادفی بودن خیلی مطلوب است، الگوریتم Lehmer می‌باشد.

یادآوری: گوییم a همنهشت (Conquence) b است در هنگ n، هرگاه $a-b=kn$ باشد. در این صورت می‌نویسیم:

$$a = b \, mod \; n$$

مثلاً داریم:

$$17 = 1 \, mod \; 4$$
$$19 = 3 \, mod \; 16$$

الگوریتم Lehmer:

$$z_n = az_{n-1} \, mod \; m$$

یعنی:

$$z_n = a^n z_0 \, mod \; m$$

فرض می‌کنیم m یک عدد اول بسیار بزرگ بوده و z_0 عددی دلخواه و کوچکتر از m باشد. عدد $a < m$ طوری انتخاب می‌شود که: $\forall n < m-1 : a^n \neq 1 \, mod \; m$ (در این صورت: $a^{m-1} = 1 \, mod \; m$ خواهد بود). به این ترتیب پریود تکرار دنباله ماکزیمم می‌شود $(m-1)$ (و کلیهٔ اعداد 1 تا $m-1$ یک و فقط یک بار ظاهر می‌شوند).

مثلاً می‌توان $m = 2^{31} - 1$ و $a = 2^7 - 1$ را در نظر گرفت.

اگر بگیریم: $u_i = \frac{z_i}{m}$، با توجه به بزرگی m، u_i دنباله‌ای پیوسته با توزیع یکنواخت $u(0,1)$ به نظر می‌رسد. یعنی می‌توان فرض کزد که: $\mathbf{u} \sim u(0,1)$.

نرم‌افزار MATLAB دنبالهٔ تصادفی یکنواخت برای شما تولید می‌کند.

تولید دنبالهٔ اعداد تصادفی با توزیع دلخواه:

۱. روش تبدیل معکوس

۲. روش رد کردن

۳. روش‌های خاص

۱. روش تبدیل معکوس (Inverse Transform Method):

در فصل ۴ و تمرینهای آن دیدیم که اگر متغیر تصادفی $\mathbf{x}$ دارای توزیع انباشتهٔ F باشد، در این صورت $\mathbf{u} = \mathrm{F}(\mathbf{x})$ دارای توزیع یکنواخت $\mathrm{u}(0,1)$ خواهد بود و به عکس اگر $\mathbf{u} \sim \mathrm{u}(0,1)$ باشد، در این صورت $\mathbf{x} = \mathrm{F}^{-1}(\mathbf{u})$ دارای تابع توزیع انباشتهٔ F خواهد بود.

و نیز اگر $\mathbf{x}$ دارای توزیع F باشد و بخواهیم از روی آن متغیر تصادفی $\mathbf{y}$ با تابع توزیع داده شدهٔ G را بسازیم، کافی است بگیریم:

$$\mathbf{y} = \underbrace{\mathrm{G}^{-1}(\overbrace{\mathrm{F}(\mathbf{x})}^{\text{یکنواخت}})}_{\text{دارای توزیع انباشتهٔ } \mathrm{G}}$$

مثال ۱: تولید متغیر تصادفی نمایی (از روی یکنواخت):

$$F(x) = 1 - e^{-\lambda x}$$

$$\mathbf{x} = F^{-1}(\mathbf{u})$$

$$\mathbf{u} = F(\mathbf{x}) = 1 - e^{-\lambda \mathbf{x}} \Rightarrow -\lambda \mathbf{x} = \ln(1 - \mathbf{u}) \Rightarrow \mathbf{x} = -\frac{1}{\lambda}\ln(1 - \mathbf{u})$$

$$\mathbf{u} \sim u(0,1) \Rightarrow 1 - \mathbf{u} \sim u(0,1)$$

پس:

$$x_i = -\frac{1}{\lambda}\ln u_i$$

مثال ۲: تولید متغیر تصادفی رایلی:

$$f(y)=\frac{y}{\sigma^2}e^{-\frac{y^2}{2\sigma^2}} \quad , \quad F(y)=1-e^{-\frac{y^2}{2\sigma^2}}$$

$$\mathbf{u}=F(\mathbf{y})=1-e^{-\frac{\mathbf{y}^2}{2\sigma^2}} \Rightarrow \mathbf{y}^2=-2\sigma^2\ln(1-\mathbf{u})$$

پس:

$$y_i=\sqrt{-2\sigma^2\ln u_i}$$

ضمناً اگر x_i دارای توزیع نمایی با پارامتر λ باشد، $y_i=\sqrt{x_i}$ دارای توزیع رایلی با پارامتر $\sigma^2=\frac{1}{2\lambda}$ خواهد بود.

۲. روش رد کردن (Rejection Method):

اکنون می‌خواهیم دنبالهٔ تصادفی y_i با تابع چگالی g (توزیع انباشتهٔ G) را تولید کنیم.

فرض: دنبالهٔ x_i با تابع چگالی f (توزیع انباشتهٔ F) قابل تولید است و :

$$\forall x: \frac{g(x)}{f(x)} \le \alpha \quad , \quad f(x)=0 \Rightarrow g(x)=0$$

(چون **x** در منطقهٔ $f(x)=0$ ظاهر نمی‌شود، **y** نیز در آن منطقه ظاهر نخواهد شد.)

روش: ابتدا داریم: $i=j=1$؛

۱. x_i را (با توزیع f) تولید می‌کنیم:

$$f_{\mathbf{x}}(x)=f(x)$$

۲. u_i را با توزیع یکنواخت ($u(0,1)$) تولید می‌کنیم (**u** مستقل از **x** است):

$$f_{\mathbf{u}}(u)=\begin{cases}1 & 0<u<1\\ 0 & \text{otherwise}\end{cases}$$

۳.اگر $u_i \le \frac{g(x_i)}{\alpha f(x_i)}$ باشد، $y_j=x_i$ قرار می‌دهیم ($i=i+1$ و $j=j+1$، بازگشت به قدم اول). در غیر این صورت این x_i را رد می‌کنیم ($i=i+1$، بازگشت به قدم اول).

اثبات: $f_{\mathbf{y}}(y)\overset{?}{=}g(y) \Leftrightarrow F_{\mathbf{y}}(y)\overset{?}{=}G(y)$؛

داریم:

$$F_{\mathbf{y}}(y)=P\{\mathbf{y}\le y\}=P\{\mathbf{x}\le y\mid \mathbf{u}\le \frac{g(\mathbf{x})}{\alpha f(\mathbf{x})}\}=\frac{P\{\mathbf{x}\le y,\mathbf{u}\le \frac{g(\mathbf{x})}{\alpha f(\mathbf{x})}\}}{P\{\mathbf{u}\le \frac{g(\mathbf{x})}{\alpha f(\mathbf{x})}\}}$$

$$\to f_{\mathbf{xu}}(x,u)=f_{\mathbf{x}}(x)f_{\mathbf{u}}(u)=\begin{cases} f(x) & 0<u<1 \\ 0 & \text{otherwise}\end{cases}$$

$$\Rightarrow P\{\mathbf{x}\le y,\mathbf{u}\le \frac{g(\mathbf{x})}{\alpha f(\mathbf{x})}\}=\int_{-\infty}^{y}\int_{0}^{\frac{g(x)}{\alpha f(x)}}\underbrace{f_{\mathbf{xu}}(x,u)}_{\boxed{0<u<1 \text{ برای } f(x)}}du\,dx$$

با توجه به اینکه $\frac{g(x)}{\alpha f(x)}\le 1$ است:

$$=\int_{-\infty}^{y}f(x)\frac{g(x)}{\alpha f(x)}dx=\frac{G(y)}{\alpha}$$

لذا:

$$P\{\mathbf{u}\le \frac{g(\mathbf{x})}{\alpha f(\mathbf{x})}\}=P\{\mathbf{x}\le +\infty,\mathbf{u}\le \frac{g(\mathbf{x})}{\alpha f(\mathbf{x})}\}=\frac{G(+\infty)}{\alpha}=\frac{1}{\alpha}\Rightarrow F_{\mathbf{y}}(y)=\frac{\frac{G(y)}{\alpha}}{\frac{1}{\alpha}}=G(y)$$

تعداد تکرار حلقه برای یک بار تولید اعداد تصادفی y_j توزیع هندسی با پارامتر $p=\frac{1}{\alpha}$ است، لذا:

$$\frac{q}{p}=\frac{1-\frac{1}{\alpha}}{\frac{1}{\alpha}}=\alpha-1=$$ متوسط تعداد شکستها (برای یک بار پیروزی)

$\Rightarrow \alpha =$ متوسط تعداد کل (برای یک بار پیروزی)

مثال: تولید دنبالهٔ تصادفی نرمال (از روی نمایی) به روش رد کردن و مخلوط کردن:

$$f(x)=e^{-x} : x>0 \quad \rightarrow \quad h(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

ولی:

$$f(x)=0 \not\Rightarrow h(x)=0$$

پس ابتدا دنباله با توزیع g (نرمال یکطرفه) را تولید می‌کنیم:

$$g(x)=\frac{2}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} : x>0$$

$$\frac{g(x)}{f(x)}=\frac{\frac{2}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}}{e^{-x}}=\frac{2}{\sqrt{2\pi}}e^{-\frac{x^2-2x}{2}}=\sqrt{\frac{2}{\pi}}e^{\frac{1}{2}}e^{-\frac{x^2-2x+1}{2}}=\sqrt{\frac{2e}{\pi}}e^{-\frac{(x-1)^2}{2}}$$

$$\Rightarrow\frac{g(x)}{f(x)}\le\sqrt{\frac{2e}{\pi}}=\alpha=1.32\simeq\frac{4}{3}$$

یعنی تقریباً در هر ۴ بار، ۳ بار موفقیت.

پس ابتدا x_i نمایی تولید می‌کنیم (که چگونگی آن را قبلاً دیدیم):

$$\mathbf{x}\sim\exp(1)$$

سپس u_i یکنواخت تولید می‌کنیم ($\mathbf{u}$ مستقل از $\mathbf{x}$ است):

$$\mathbf{u}\sim u(0,1)$$

اگر $u_i\le e^{-\frac{(x_i-1)^2}{2}}$ باشد، $y_j=x_i$ قرار می‌دهیم. در غیر این صورت x_i را رد می‌کنیم.

حال برای اینکه از روی $\mathbf{y}$ که نرمال یکطرفه است، متغیر تصادفی $\mathbf{z}$ نرمال (استاندارد) بسازیم، توجه می‌کنیم که:

$$h(z)=\frac{1}{2}g(z)+\frac{1}{2}g(-z)$$

پس v_i یکنواخت تولید می‌کنیم ($\mathbf{v}$ مستقل از $\mathbf{x}$ و $\mathbf{u}$ است):

$$\mathbf{v}\sim u(0,1)$$

اگر $0<v_i<\frac{1}{2}$ بود، $z_i=y_i$ قرار داده و اگر $\frac{1}{2}<v_i<1$ بود، $z_i=-y_i$ قرار می‌دهیم.

(با توجه به اینکه: $f(x)=\sum_k f(x|A_k)P(A_k)$)

نرم‌افزار MATLAB می‌تواند دنبالهٔ تصادفی نرمال نیز تولید کند.

۳. روش‌های خاص:

متغیرهای تصادفی مختلف را با توجه به خواص ویژهٔ هر یک به روشهایی خاص می‌توان تولید کرد.

مثال ۱: تولید دنبالهٔ نرمال: با توجه به قضیهٔ حد مرکزی، مجموع تعداد زیادی متغیر تصادفی یکنواخت، نرمال می‌شود. پس:

$$z_1 = u_1 + \cdots + u_n \quad , \quad z_2 = u_{n+1} + \cdots + u_{2n} \quad , \quad \cdots$$

یعنی:

$$z_i = \sum_{j=1}^{n} u_{(n-1)i+j}$$

مقدار ۱۰ تا ۱۵ برای n کافی است.

مثال ۲: روش دیگری برای تولید دنبالهٔ نرمال؛ تولید دو دنبالهٔ نرمال استاندارد و مستقل از هم:

یادآوری: اگر متغیرهای تصادفی $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند، داریم:

$$\begin{cases} \mathbf{x} \sim N(0,1) \\ \mathbf{y} \sim N(0,1) \end{cases} \Leftrightarrow \begin{cases} \mathbf{R} = \sqrt{\mathbf{x}^2 + \mathbf{y}^2} \sim \text{Rayleigh}(1) \\ \Phi = \tan^{-1} \dfrac{\mathbf{y}}{\mathbf{x}} \sim u(0,2\pi) \end{cases}$$

اگر u_i و v_i دنباله‌های تصادفی یکنواخت و مستقل باشند، آنگاه $r_i = \sqrt{-2\ln u_i}$ دارای توزیع رایلی و $2\pi v_i$ دارای توزیع یکنواخت $(0,2\pi)$ است. پس دنباله‌های زیر:

$$x_i = \sqrt{-2\ln u_i}\cos(2\pi v_i)$$
$$y_i = \sqrt{-2\ln u_i}\sin(2\pi v_i)$$

نرمال و مستقل هستند.

این هم توسط نرم‌افزار MATLAB مستقیماً قابل تولید است.

مثال ۳: تولید دنبالهٔ تصادفی ارلانگ:

می‌دانیم که مجموع n تا متغیر تصادفی نمایی مستقل با پارامتر λ دارای توزیع ارلانگ با پارامتر n و λ است. پس x_i نمایی با پارامتر λ تولید می‌کنیم و سپس قرار می‌دهیم:

$$y_i = \sum_{j=1}^{n} x_{(n-1)i+j}$$

$$\mathbf{x} \sim \exp(\lambda) \Rightarrow \mathbf{y} \sim \mathrm{Erlang}(n,\lambda)$$

شبیه‌سازی مونت‌کارلو (Monte Carlo Simulation):

به دست آوردن یک کمیت غیرتصادفی توسط روش آماری (با تولید دنبالهٔ تصادفی) را شبیه‌سازی مونت‌کارلو گویند.

مثال ۱: فرض کنید می‌خواهیم انتگرال زیر را حساب کنیم:

$$I = \int_0^1 g(u)du$$

(اگر انتگرال در محدودهٔ دیگری را بخواهیم با تغییر متغیر می‌توانیم به این محدوده ببریم.)

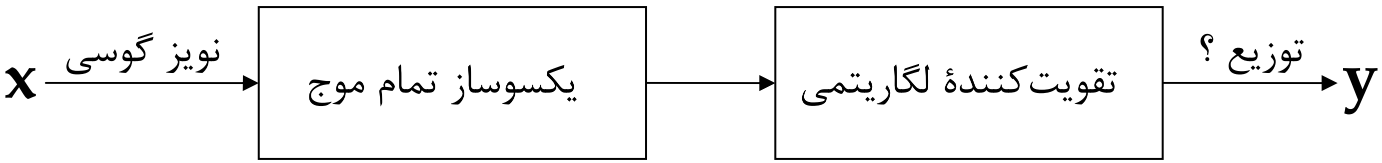
اگر متغیر تصادفی $\mathbf{u}$ را در نظر بگیریم که: $\mathbf{u} \sim u(0,1)$ و فرض کنیم: $\mathbf{x} = g(\mathbf{u})$ باشد، آنگاه:

$$\eta_{\mathbf{x}} = E(g(\mathbf{u})) = \int_{-\infty}^{+\infty} g(u)f(u)du = \int_0^1 g(u)du = I$$

در شبیه‌سازی مونت‌کارلو، دنبالهٔ تصادفی یکنواخت u_i را تولید کرده و سپس $x_i = g(u_i)$ را تولید می‌کنیم. آنگاه با توجه به اینکه: $\eta_{\mathbf{x}} \simeq \frac{1}{n}\sum_{i=1}^{n} x_i$، داریم:

$$I \simeq \frac{1}{n}\sum_{i=1}^{n} g(u_i)$$

مثال ۲: می‌خواهیم $F(y)=P\{\mathbf{y}\le y\}$ (کمیتی غیرتصادفی) را تخمین بزنیم. مثلاً:



اصولاً روش تحلیلی می‌تواند دشوار یا بعضاً بدون جواب بسته باشد.

دنبالهٔ y_i را تولید می‌کنیم (با تولید x_i و طی مراحل سیستم). اگر تعداد y_iهایی که کوچکتر از y هستند را n_y بنامیم و تعداد کل y_iهای تولید شده برابر n باشد، داریم:

$$F(y)\simeq \frac{n_y}{n}$$