

روش شبیه سازی مونت کارلو و کاربردهای آن در آمار و احتمال

غلامرضا جندقی^۱ احمد گائینی

چکیده

یکی از روشهای عددی که امروزه در بسیاری از علوم مورد توجه قرار گرفته، روش شبیه سازی مونت کارلوست. در این مقاله، ابتدا روش شبیه سازی مونت کارلو معرفی شده و سپس زمینه های کاربرد این روش در آمار و احتمال مورد بررسی قرار می گیرد. در هر قسمت، برای نشان دادن چگونگی نتایج آن، مثالهایی مطرح می شود.

۱. مقدمه

کشور کوچک موناکو، مرکز بازیهای شانس بود، به کار گرفته شد. امروزه این روش در بسیاری از علوم مورد استفاده قرار می گیرد. کاربرد روش مونت کارلو در مدل سازی مسائل فیزیکی به ما امکان می دهد که سیستمهای پیچیده تری را که در عمل بررسی آنها بسیار مشکل است مطالعه کنیم. حل معادلاتی که بیانگر کنش و واکنش اتمهای یک سیستم اند، محاسبه انتگرالهای پیچیده ای که از راههای کلاسیک قابل محاسبه نیستند و مسائل بسیار پیچیده دیگر، نمونه هایی از کاربرد این روش اند.[۲]

دانشمندان علم آمار، این علم را به دو شاخه نظری و تجربی تقسیم می کنند. آمار تجربی به روشهایی اطلاق می شود که با استفاده از روشهای شبیه سازی، خواص برآوردکننده ها مطالعه می شود. این روشها به شبیه سازی مونت کارلو موسوم شده است. روشهای مونت کارلو، به روشهایی اطلاق می شوند که بر اساس دنباله ای از اعداد تصادفی به بررسی مسائل می پردازند. نام مونت کارلو اولین بار توسط متروپولیس^۲ به دلیل شباهت شبیه سازی آماری به بازیهای شانس و به دلیل اینکه شهر مونت کارلو مرکز

^۲ Metropolis

^۱ گروه ریاضی و آمار، دانشگاه امام حسین (ع)

^۲ گروه ریاضی و آمار، دانشگاه امام حسین (ع)

۲. قانون قوی اعداد بزرگ

اگر $X_1, X_2, \dots, X_n$ دنباله‌ای از متغیرهای تصادفی مستقل و هم‌توزیع با میانگین متناهی μ باشد، آنگاه

$$P[\lim_{n \rightarrow \infty} \bar{X}_n = \mu] = 1.$$

قانون اعداد بزرگ می‌گوید با احتمال یک، برای هر مقدار مثبت ε ، مقدار $|\bar{X}_n - \mu|$ فقط به تعداد دفعات متناهی، از ε بزرگتر خواهد شد.

حال اگر جامعه بالا را برنولی فرض کنیم و p احتمال موفقیت باشد، آنگاه $X_i = 1$ و $X_i = 0$ ، به ترتیب نشان دهنده موفقیت و شکست در آزمایش i ام بوده و داریم:

$$E(X_i) = \mu = p, \quad \bar{X}_n = \frac{\sum X_i}{n}$$

ولذا

$$P[\lim_{n \rightarrow \infty} \bar{X}_n = p] = 1.$$

در نتیجه $\sum X_i / n$ برآورد گری برای احتمال موفقیت است. برای شبیه‌سازی هر پیشامد، می‌توان از اعداد تصادفی استفاده کرد. گرچه استفاده از اعداد تصادفی برای شبیه‌سازی متغیرهای گسسته نیز امکان دارد، ما فقط روش شبیه‌سازی متغیرهای تصادفی پیوسته را که ساده‌تر است به کمک قضیه زیر توضیح می‌دهیم. برای شبیه‌سازی متغیرهای تصادفی گسسته و سایر مسائل مربوط به شبیه‌سازی به [۵] مراجعه کنید.

قضیه تبدیل وارون: اگر X یک متغیر تصادفی با توزیع یکنواخت بر بازه $(0, 1)$ باشد و F تابع توزیع تجمعی دلخواه از یک متغیر پیوسته، و $Y = F^{-1}(X)$ باشد، آنگاه Y داری تابع توزیع تجمعی F است. یعنی:

$$\begin{aligned} F_Y(t) &= P(Y \leq t) = P(F^{-1}(X) \leq t) \\ &= P(X \leq F(t)) = F(t) \end{aligned}$$

با استفاده از این قضیه می‌توان اعداد تصادفی از هر توزیع دلخواهی را با استفاده از اعداد تصادفی یکنواخت در بازه $(0, 1)$ تولید کرد [۱].

۳. اعداد تصادفی و شبه تصادفی

همان‌طور که گفته شد، روش مونت کارلو از اعداد تصادفی برای شبیه‌سازی پدیده‌ها استفاده می‌کند. اعداد تصادفی اغلب بر اساس الگوریتمهای خاصی تولید می‌شوند. البته آنچه که امروزه به عنوان اعداد تصادفی مصطلح است، اعداد شبه تصادفی^۴ اند که توسط الگوریتمهای ریاضی خاصی توسط رایانه تولید می‌شوند. همه الگوریتمهای موجود خوب نیستند. برای مثال، یک الگوریتم ممکن است در بیشتر موارد نوعی همبستگی در دنباله اعدادش مشاهده شود. مثلاً ممکن است هر پنجمین عددش بزرگ باشد. برای تشخیص کیفیت تصادفی بودن اعداد تولید شده توسط یک الگوریتم، می‌توان زوجهای تصادفی (X, Y) از این اعداد را در دستگاه محور مختصات دکارتی رسم کرد. اگر تجمع نقاط در یک قسمت از صفحه مختصات بیشتر باشد، نشان دهنده کیفیت پایین این اعداد است. اگر اعداد تولید شده تقریباً همان خواص اعداد تصادفی را دارا باشند، در عمل عنوان تصادفی به آنها اطلاق می‌شود.

۴. محاسبه احتمال به کمک روش مونت کارلو

از آنجا که برخی از مسائل احتمال، بسیار پیچیده و حل آنها با روشهای معمولی بسیار وقت گیر و پرهزینه است، می‌توان از روش مونت کارلو برای حل آنها کمک گرفت. در این قسمت دو مسأله احتمال را با استفاده از روش مونت کارلو حل کرده و با جواب تحلیلی آن مقایسه می‌کنیم.

مثال ۱- اگر پاره خطی به طول واحد به تصادف به سه قسمت تقسیم شود، احتمال این که بتوان با سه قسمت مذکور، یک مثلث ساخت چقدر است؟

^۴ Pseudorandom

با قرار دادن $f(x)$ در معادله (۱) خواهیم داشت:

$$I = (b - a) \int_a^b g(x) f(x) dx \quad (2)$$

توجه کنید که $f(x)$ می‌تواند به عنوان تابع چگالی احتمال یکنواخت در بازه $[a, b]$ در نظر گرفته شود. با در نظر گرفتن $f(x)$ به عنوان یک چگالی احتمال، معادله (۲) همان امید ریاضی $g(x)$ تحت چگالی $f(x)$ خواهد بود. یعنی:

$$I = (b - a) \int_a^b g(x) f(x) dx = (b - a) \bar{g} \quad (3)$$

حال یک نمونه تصادفی n تایی به صورت $X_1, X_2, \dots, X_n$ از چگالی $f(x)$ انتخاب کرده و مقدار $g(x_i)$ ، $i = 1, \dots, n$ ، را محاسبه می‌کنیم. پس:

$$\bar{g} = \frac{1}{n} \sum_{i=1}^n g(x_i)$$

و بنابراین:

$$I = (b - a) \bar{g} = (b - a) \left[\frac{1}{n} \sum_{i=1}^n g(x_i) \right]$$

مثال ۱ - یک مسئله انتگرالگیری، برای توزیع نرمال مطرح می‌شود که می‌خواهیم احتمالی به صورت زیر را محاسبه کنیم:

$$P(Z < z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

است. می‌دانیم که این انتگرال را نمی‌توان برحسب ترکیبات متناهی از توابع مقدماتی نوشت [۴]. بنابراین راهی جز تقریب این انتگرال نخواهیم داشت. این انتگرال را معمولاً با استفاده از روشهای عددی مثل روش سیمپسون برآورد می‌کنند. در اینجا ما با استفاده از روش مونت کارلو مقدار تقریبی این انتگرال را محاسبه کرده ایم. نتایج محاسباتی را، همراه با مقادیر متناظر آنها که در توزیع نرمال آمده اند، در جدول (۳) مشاهده می‌کنید.

حل: با در نظر گرفتن شرایط ساختن یک مثلث، جواب

تحلیلی این مسأله برابر $p = \frac{1}{4}$ به دست می‌آید. این مسأله از روش مونت کارلو با استفاده از نرم افزار S-PLUS برنامه نویسی شده و نتایج در جدول (۱) آمده است. [۳]

همان طور که از اعداد درون جدول (۱) مشخص است با افزایش تعداد اعداد تصادفی تولید شده، مقدار احتمال به جواب تحلیلی مسأله نزدیکتر شده است. این نکته در شکل (۱) کاملاً مشخص است. در سالهای اخیر با پیشرفت رایانه ها دیگر نباید نگران زمان مصرف شده برای محاسبه بود، زیرا با سرعت فوق العاده ای که این رایانه ها دارند، در مدتی بسیار محدود می‌توانند هزاران عدد تصادفی تولید کنند.

مثال ۲ - اگر نقطه A به طور تصادفی در مربعی به ضلع واحد

به طور یکنواخت توزیع شده باشد، مطلوب است احتمال این که فاصله A تا نزدیکترین ضلع مربع کمتر از فاصله آن تا نزدیکترین قطر مربع باشد.

حل: جواب تحلیلی این مسأله برابر

$p = \sqrt{2} - 1 = 0.414214$ است. با استفاده از روش مونت کارلو مقدار احتمال به ازای تعدادی مختلف از اعداد تصادفی در جدول (۲) آمده است.

۵. انتگرالگیری به روش مونت کارلو

فرض کنید هدف، محاسبه مقدار انتگرال

$$I = \int_a^b g(x) dx \quad (1)$$

باشد که در آن $g(x)$ تابعی حقیقی در $(-\infty, \infty)$ است. ایده اصلی این است که باید تعریف انتگرال را طوری بیان کرد که از طریق روش مونت کارلو قابل محاسبه باشد. برای این کار، تابع زیر را در بازه $[a, b]$ تعریف می‌کنیم:

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{سایر جاها} \end{cases}$$

جدول ۱: مقدار احتمال محاسبه شده و بازه اطمینان ۹۵ درصدی برای آن (مربوط به مثال ۱)

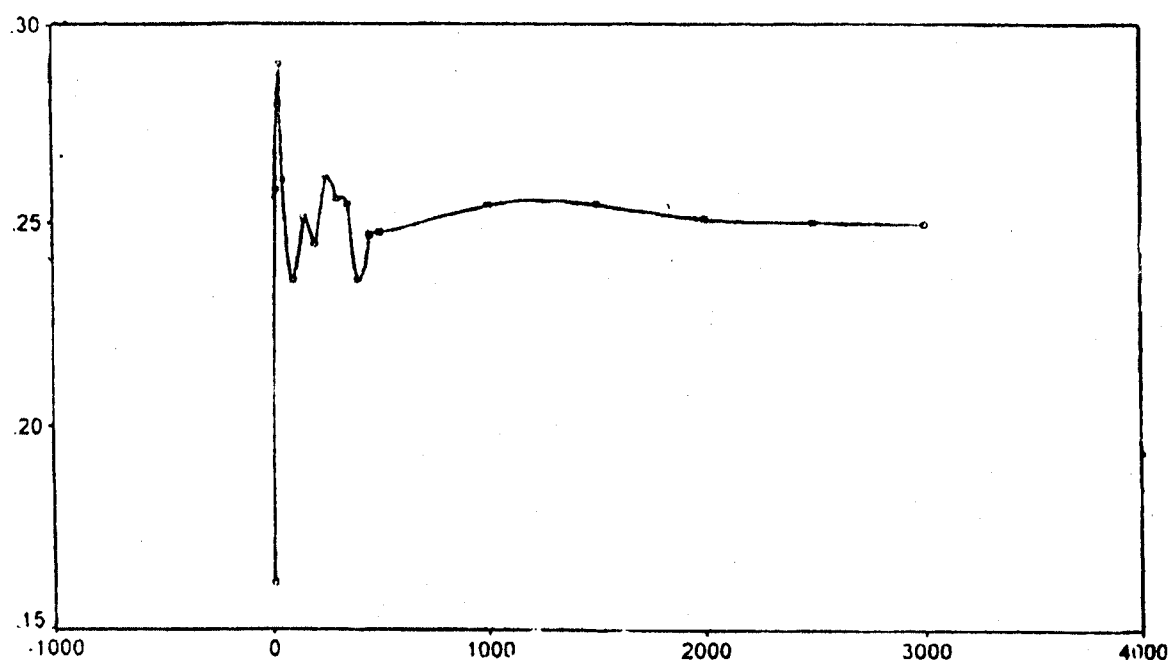
تعداد تکرار آزمایش n	احتمال p به روش مونت کارلو	بازه اطمینان ۹۵ درصدی برای p
۱۰	۰/۱۶۱۴	(۰/۱۱۴۵, ۰/۲۰۸۳)
۲۰	۰/۲۵۸۵	(۰/۲۱۸۰, ۰/۲۲۹۰)
۳۰	۰/۲۷۷۹	(۰/۲۱۳۷, ۰/۳۴۴۱)
۴۰	۰/۲۹۰۵	(۰/۲۵۳۳, ۰/۳۲۷۷)
۵۰	۰/۲۶۰۸	(۰/۲۲۵۹, ۰/۲۹۵۷)
۱۰۰	۰/۲۳۶۱	(۰/۲۰۵۹, ۰/۲۶۶۳)
۱۵۰	۰/۲۵۱۳	(۰/۲۲۱۸, ۰/۲۸۰۸)
۲۰۰	۰/۲۴۴۹	(۰/۲۳۰۸, ۰/۲۵۹۰)
۲۵۰	۰/۲۶۱۳	(۰/۲۴۸۰, ۰/۲۷۴۶)
۳۰۰	۰/۲۵۶۲	(۰/۲۴۳۱, ۰/۲۶۹۳)
۳۵۰	۰/۲۵۴۷	(۰/۲۳۴۸, ۰/۲۷۴۶)
۴۰۰	۰/۲۳۶۱	(۰/۲۱۷۶, ۰/۲۵۴۶)
۴۵۰	۰/۲۴۶۸	(۰/۲۲۸۵, ۰/۲۶۵۱)
۵۰۰	۰/۲۴۷۷	(۰/۲۳۰۳, ۰/۲۶۵۱)
۱۰۰۰	۰/۲۵۴۵	(۰/۲۳۸۷, ۰/۲۷۰۳)
۱۵۰۰	۰/۲۵۴۹	(۰/۲۴۱۲, ۰/۲۶۸۶)
۲۰۰۰	۰/۲۵۱۲	(۰/۲۳۹۳, ۰/۲۶۳۱)
۲۵۰۰	۰/۲۵۰۴	(۰/۲۳۹۶, ۰/۲۶۱۲)
۳۰۰۰	۰/۲۴۹۹	(۰/۲۴۰۰, ۰/۲۵۹۸)

جدول ۲: مقدار احتمال محاسبه شده و بازه اطمینان ۹۵ درصدی برای آن (مربوط به مثال ۲)

تعداد تکرار آزمایش n	احتمال p به روش مونت کارلو	بازه اطمینان ۹۵ درصدی برای p
۱۰	۰/۵۴۰۰	(۰/۴۴۶۰ , ۰/۶۳۴۰)
۲۰	۰/۴۰۰۰	(۰/۳۱۲۰ , ۰/۴۳۸۰)
۳۰	۰/۴۲۳۳	(۰/۳۸۰۶ , ۰/۴۸۶۰)
۴۰	۰/۴۶۵۰	(۰/۴۰۲۴ , ۰/۵۲۷۶)
۵۰	۰/۴۷۶۰	(۰/۴۵۲۲ , ۰/۴۹۹۸)
۱۰۰	۰/۴۱۸۰	(۰/۳۹۱۲ , ۰/۴۴۴۸)
۱۵۰	۰/۳۹۰۷	(۰/۳۷۱۲ , ۰/۴۱۰۲)
۲۰۰	۰/۴۱۱۰	(۰/۳۸۸۲ , ۰/۴۳۳۸)
۲۵۰	۰/۴۱۷۶	(۰/۴۰۰۷ , ۰/۴۳۴۵)
۳۰۰	۰/۴۰۳۳	(۰/۳۷۷۵ , ۰/۴۲۹۱)
۳۵۰	۰/۴۰۵۷	(۰/۳۸۶۱ , ۰/۴۲۵۳)
۴۰۰	۰/۴۱۱۰	(۰/۳۸۵۵ , ۰/۴۳۶۵)
۴۵۰	۰/۴۱۵۶	(۰/۳۷۳۰ , ۰/۴۵۸۲)
۵۰۰	۰/۴۳۰۴	(۰/۴۱۸۲ , ۰/۴۴۲۶)
۱۰۰۰	۰/۴۰۷۸	(۰/۳۹۶۹ , ۰/۴۱۸۷)
۱۵۰۰	۰/۴۱۱۱	(۰/۴۰۶۲ , ۰/۴۱۶۰)

جدول ۳: مقادیر احتمال، همراه با مقادیر متناظر آنها که در جداول آماری موجودند

Z	$P(Z \leq z)$	به روش مونت کارلو $P(Z \leq z)$	بازه اطمینان ۹۵٪ برای $P(Z \leq z)$
۰/۰	۰/۵۰۰۰	۰/۵۰۰۶	(۰/۴۹۴۱ , ۰/۵۰۷۱)
۰/۱	۰/۵۳۹۸	۰/۵۳۹۹	(۰/۵۳۶۶ , ۰/۵۴۳۲)
۰/۲	۰/۵۷۹۳	۰/۵۸۰۳	(۰/۵۶۸۸ , ۰/۵۹۱۸)
۰/۳	۰/۶۱۷۹	۰/۶۱۸۸	(۰/۶۰۹۶ , ۰/۶۲۸۰)
۰/۴	۰/۶۵۵۴	۰/۶۵۶۸	(۰/۶۴۹۹ , ۰/۶۶۳۷)
۰/۵	۰/۶۹۱۵	۰/۶۹۶۱	(۰/۶۸۶۵ , ۰/۷۰۵۷)



تعداد اعداد تصادفی

شکل ۱: نمودار مقدار احتمال محاسبه شده بر حسب تعداد اعداد تصادفی تولید شده (مربوط به مثال ۱)

مراجع

- [۱] بهبودیان، جواد؛ آمار نابارامتری، انتشارات دانشگاه شیراز (۱۳۷۱)، ص ۲۰.
- [2] Hammersley, J. M. & Handscomb, D. C. , *Spoltiswood*, Ballantyn Co. Ltd., London, 1964, pp 5-25.
- [3] Wilks, A. K., *The New S Language*, Wadsworth & Brooks Cole Advanced Books & Software, Pacific Grove, California, 1988.
- [4] Thomas, G. B., Finny & R. L., *Calculus and Analytic Geometry*, 7-th ed., Addison _ Weslev. 1988, pp 390-400.
- [۵] شبیه سازی، تألیف اس.ام. راس، ترجمه آذرنوش و نیرومند، (۱۳۷۴)، انتشارات دانشگاه مشهد.