

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

احمد گائینی^۱ غلامرضا جندقی^۲

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 زمینه‌های کاربرد این روش در آمار و احتمال مورد بررسی قرار می‌گیرد. در هر قسمت، برای نشان دادن چگونگی نتایج آن، مثالهایی مطرح می‌شود.

۱. مقدمه

کشور کوچک موناکو، مرکز بازیهای شانسی بود، به کار گرفته شد. امروزه این روش در بسیاری از علوم مورد استفاده قرار می‌گیرد. کاربرد روش مونت کارلو در مدل سازی مسائل فیزیکی به ما امکان می‌دهد که سیستمهای پیچیده تری را که در عمل بررسی آنها بسیار مشکل است مطالعه کیم. حل معادلاتی که بیانگر کشش و واکنش انتهای یک سیستم اند، محاسبه انتگرالهای پیچیده ای که از راههای کلاسیک قابل محاسبه نیستند و مسائل بسیار پیچیده دیگر، نمونه‌هایی از کاربرد این روش اند.^[۲]

دانشمندان علم آمار، این علم را به دو شاخه نظری و تجربی تقسیم می‌کنند. آمار تجربی به روش‌های اطلاق می‌شود که با استفاده از روش‌های شبیه سازی، خواص بسرآوردهای مطالعه می‌شود. این روشها به شبیه سازی مونت کارلو موسوم شده است. روش‌های مونت کارلو، به روش‌های اطلاق می‌شوند که بر اساس دنباله‌ای از اعداد تصادفی به بررسی مسائل می‌پردازند. نام مونت کارلو اولین بار توسط مت روپولیس^۳ به دلیل شباهت شبیه سازی آماری به بازیهای شانسی و به دلیل اینکه شهر مونت کارلو مرکز

Metropolis^۴

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

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

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

همان طور که گفته شد، روش مونت کارلو از اعداد تصادفی برای شبیه سازی 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 اعداد 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 مسئله 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؟

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 $|\mu - \bar{X}_n|$ فقط به تعداد دفعات متناهی، از ϵ بزرگ‌تر خواهد شد.

حال اگر جامعه بالا را برنولی فرض کنیم و احتمال موفقیت باشد، آنگاه $\bar{X}_n = 1$ و $\mu = 0$ ، به ترتیب نشان دهنده موفقیت و شکست در آزمایش آن بوده و داریم:

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

و لذا

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

در نتیجه $\sum X_i / n$ برآوردگری برای احتمال موفقیت است. برای شبیه سازی هر پیشامد، می‌توان از اعداد تصادفی استفاده کرد. گرچه استفاده از اعداد تصادفی برای شبیه سازی متغیرهای گستته نیز امکان دارد، ما فقط روش شبیه سازی متغیرهای تصادفی پیوسته را که ساده‌تر است به کمک قضیه زیر توضیح می‌دهیم. برای شبیه سازی متغیرهای تصادفی گستته و سایر مسائل مربوط به شبیه سازی 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 $(X)^{-1} = F^{-1}(Y)$ 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 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 $f(x)$ می تواند 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 (x) تحت چگالی (x) $f(x)$ 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 تابی به صورت x_1, x_2, \dots, x_n از چگالی $f(x)$ انتخاب کرده و مقدار $(x_i, g(x_i))$ ، $i = 1, \dots, n$ را محاسبه می کنیم. 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 - یک مسئله انتگرالگیری، برای توزیع نرمال مطروح می شود که می خواهیم احتمالی به صورت زیر را محاسبه کنیم:

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

است. می دانیم که این انتگرال را نمی توان بر حسب ترکیبات متنه ای 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 می کنند. 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 می کنید.

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

تحلیلی این مسئله برابر $\frac{1}{4} p$ به دست می آید. این مسئله از روش

مونت کارلو با استفاده از نرم افزار S-PLUS برنامه نویسی شده و نتایج در جدول (۱) آمده است. [۲].

همان طور که از اعداد درون جدول (۱) مشخص است با افزایش تعداد اعداد تصادفی تولید شده، مقدار احتمال به جواب تحلیلی مسئله نزدیکتر شده است. این نکته در شکل (۱) کاملاً مشخص است. در سالهای اخیر با پیشرفت زیانه ها دیگر نایاب نگران زمان مصرف شده برای محاسبه بود، زیرا با سرعت فوق العاده ای که این زیانه ها دارند، در مدتی بسیار محدود می توانند 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 مسئله برابر $p = \sqrt{2} - 1 = 0.414214$ است. بسا 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، نتایج زیر را در بازه $[a, b]$ تعریف می کنیم:

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

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

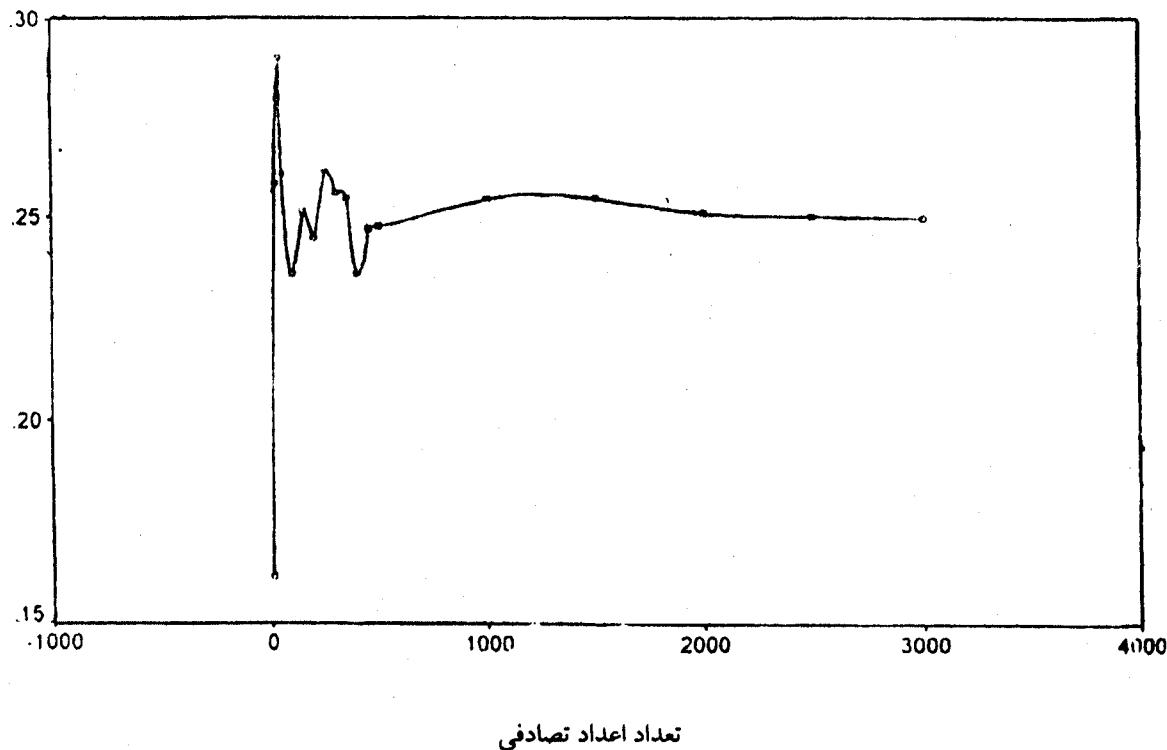
تعداد تکرار آزمایش n	احتمال p به روش مونت کارلو	بازه اطمینان ۹۵ درصدی برای p
۱۰	۰/۱۶۱۴	(۰/۱۱۴۵, ۰/۲۰۸۳)
۲۰	۰/۲۵۸۵	(۰/۲۱۸۰, ۰/۲۲۹۰)
۳۰	۰/۲۷۷۹	(۰/۲۱۳۷, ۰/۳۴۴۱)
۴۰	۰/۲۹۰۰	(۰/۲۰۲۳, ۰/۳۲۷۷)
۵۰	۰/۲۶۰۸	(۰/۲۲۰۹, ۰/۲۹۰۷)
۱۰۰	۰/۲۳۶۱	(۰/۲۰۰۹, ۰/۲۶۶۳)
۱۵۰	۰/۲۰۱۳	(۰/۲۲۱۸, ۰/۲۸۰۸)
۲۰۰	۰/۲۴۴۹	(۰/۲۳۰۸, ۰/۲۰۹۰)
۲۵۰	۰/۲۶۱۳	(۰/۲۴۸۰, ۰/۲۷۸۶)
۳۰۰	۰/۲۰۶۲	(۰/۲۴۳۱, ۰/۲۶۹۳)
۴۰۰	۰/۲۰۴۷	(۰/۲۳۴۸, ۰/۲۷۴۶)
۴۰۰	۰/۲۳۶۱	(۰/۲۱۷۶, ۰/۲۰۳۶)
۴۰۰	۰/۲۴۶۸	(۰/۲۲۸۰, ۰/۲۶۰۱)
۵۰۰	۰/۲۴۷۷	(۰/۲۳۰۳, ۰/۲۶۰۱)
۱۰۰۰	۰/۲۰۳۰	(۰/۲۳۸۷, ۰/۲۷۰۳)
۱۰۰۰	۰/۲۰۳۹	(۰/۲۴۱۲, ۰/۲۶۸۶)
۲۰۰۰	۰/۲۰۱۲	(۰/۲۲۹۳, ۰/۲۶۳۱)
۲۰۰۰	۰/۲۰۰۴	(۰/۲۳۹۶, ۰/۲۶۱۲)
۳۰۰۰	۰/۲۴۹۹	(۰/۲۴۰۰, ۰/۲۰۹۸)

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

n	تعداد تکرار آزمایش	احتمال p به روش مونت کارلو	بازه اطمینان ۹۵ درصدی برای p
۱۰		۰/۰۴۰۰	(۰/۶۳۴۰, ۰/۴۴۶۰)
۲۰		۰/۴۰۰۰	(۰/۳۶۲۰, ۰/۴۳۸۰)
۳۰		۰/۴۳۳۳	(۰/۳۸۰۶, ۰/۴۸۶۰)
۴۰		۰/۴۶۰۰	(۰/۴۰۲۴, ۰/۰۲۷۶)
۵۰		۰/۴۷۶۰	(۰/۴۰۲۲, ۰/۴۹۹۸)
۱۰۰		۰/۴۱۸۰	(۰/۳۹۱۲, ۰/۴۴۴۸)
۱۵۰		۰/۳۹۰۷	(۰/۳۷۱۲, ۰/۴۱۰۲)
۲۰۰		۰/۴۱۱۰	(۰/۳۸۸۲, ۰/۴۲۲۸)
۲۵۰		۰/۴۱۷۶	(۰/۴۰۰۷, ۰/۴۳۴۵)
۳۰۰		۰/۴۰۳۳	(۰/۳۷۷۵, ۰/۴۲۹۱)
۳۵۰		۰/۴۰۵۷	(۰/۳۸۶۱, ۰/۴۲۰۳)
۴۰۰		۰/۴۱۱۰	(۰/۳۸۰۰, ۰/۴۳۶۰)
۴۵۰		۰/۴۱۰۶	(۰/۳۷۳۰, ۰/۴۰۸۲)
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 ≤ z)	P(Z ≤ z) به روش مونت کارلو	بازه اطمینان ۹۵٪ برای (P(Z ≤ z))
-۰/۰	۰/۰۰۰۰	۰/۰۰۰۶	(۰/۴۹۴۱, ۰/۰۰۷۱)
-۰/۱	۰/۰۳۹۸	۰/۰۳۹۹	(۰/۰۳۶۶, ۰/۰۴۳۲)
-۰/۲	۰/۰۷۹۳	۰/۰۸۰۳	(۰/۰۶۸۸, ۰/۰۹۱۸)
-۰/۳	۰/۰۱۷۹	۰/۰۱۸۸	(۰/۰۶۹۶, ۰/۰۲۸۰)
-۰/۴	۰/۰۰۰۴	۰/۰۰۰۸	(۰/۰۴۹۹, ۰/۰۶۳۷)
-۰/۵	۰/۰۹۱۰	۰/۰۹۶۱	(۰/۰۸۶۵, ۰/۰۷۰۵)



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

مراجع

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