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

دانشمندان علم آمار، این علم را به دو شاخهٔ نظری و تجربی تقسیم می کنند. آمار تجربی بـه روشهایی اطلاق می شود که با استفاده از روشهای شبیه سازی، خواص بـرآوردکننده 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[۳] به دلیل شباهت شبیه سازی آماری به بازیهای شانسی و به دلیل اینکه شهر مونت کارلو مرکز کشور کوچک موناکو، مرکز بازیهای شانسی 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، محاسبهٔ انتگرالهای پیچیده ای که از راههای کلاسیک قابل محاسبه نیستند و مسائل بسیار پیچیدهٔ دیگر، نمونه هایی از کاربرد این روش اند.[۲]

---

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

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

۳ Metropolis

۲. قانون قوی اعداد بزرگ

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

$$P[\lim_{n\to\infty} \overline{X}_n = \mu] = 1 \,.$$

قانون اعداد بزرگ می گوید با احتمال یک، بـرای هـر مقـدار مثبت ε ، مقدار $|\overline{X}_n - \mu|$ فقط بـه تعـداد دفعـات متنـاهی، از ε بزرگتر خواهد شد.

حال اگر جامعهٔ بالا را برنولـی فـرض کنیـم و p احتمـال مـوفقیـت باشد، آنگاه $X_i = 1$ و $X_i = 0$ ، به ترتیب نشان دهندهٔ موفقیت و شکست در آزمایش i ام بوده و داریم:

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

و لذا

$$P[\lim_{n\to\infty} \overline{X}_n = p] = 1 \,.$$

در نتیجه $\sum X_i / n$ برآوردگری برای احتمال موفقیت است.

برای شبیه سازی هر پیشامد، می توان از اعداد تصادفی استفاده کرد. گرچه استفاده از اعداد تصـادفی بـرای شـبیه سـازی متغیرهـای گسسته نیز امکان دارد، ما فقط روش شبیه سازی متغیرهای تصـادفی پیوسته را که ساده تر است به کمک قضیـهٔ زیـر توضیـح مـی دهیـم. برای شبیه سازی متغیرهای تصادفی گسسته و سایر مسائل مربـوط بـه شبیه سازی به [۵] مراجعه کنید.

قضیهٔ تبدیـل وارون: اگر X یـک متغیـر تصـادفی بـا توزیـع یکنواخت بر بازهٔ $(0,1)$ باشد و F تـابع توزیـع تجمعـی دلخـواه از یک متغیـر پیوسـته، و $Y = F^{-1}(X)$ باشـد، آنگاه Y داری تابع توزیع تجمعی F است. یعنی:

$$F_Y(t) = P(Y \le t) = P(F^{-1}(X) \le t)$$
$$= P(X \le F(t)) = F(t)$$

با استفاده از این قضیه می تـوان اعـداد تصـادفی از هـر توزیـع دلخواهی را با استفاده از اعداد تصـادفی یکنواخـت در بـازهٔ $(0,1)$ تولید کرد [۱].

۳. اعداد تصادفی و شبه تصادفی

همان طور که گفته شد، روش مونت کارلو از اعداد تصـادفی برای شبیه سازی پدیده ها استفاده می کند. اعداد تصادفی اغلـب بـر اساس الگوریتمهای خاصی تولید می شوند. البته آنچه که امروزه بـه عنوان اعداد تصادفی مصطلح اسـت، اعـداد شـبه تصـادفی[۴] انـد کـه توسط الگوریتمهای ریاضی خاصی توسط رایانـه تولیـد مـی شـوند. همهٔ الگوریتمهای موجود خوب نیستند. برای مثـال، یـک الگوریتـم ممکن اسـت در بیشـتر مـوارد نوعـی همبسـتگی در دنبالـهٔ اعـدادش مشاهده شود. مثلاً ممکن است هر پنجمین عددش بزرگ باشد. برای تشـخیص کیفیـت تصـادفی بـودن اعـداد تولیـد شـده توسـط یـک الگوریتم، می توان زوجهای تصـــادفی (x,y) از ایـن اعـــداد را در دستگاه محور مختصات دکارتی رسم کـرد. اگر تجمع نقـاط در یک قسمت از صفحهٔ مختصـات بیشـتر باشـد، نشـان دهنـدهٔ کیفیـت پایین ایـن اعـــداد اسـت. اگر اعـــداد تولیـد شـده تقریباً همـان خواص اعداد تصادفی را دارا باشند، در عمل عنوان تصادفی به آنـها اطلاق می شود.

۴. محاسبهٔ 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 پر هزینه است، می توان از روش مونت کارلو برای حل آنها کمک گرفت. در این قسـمت دو مسـألهٔ احتمال را با استفاده از روش مونـت کـارلو حـل کـرده و بـا جـواب تحلیلی آن مقایسه می کنیم.

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

---

۴ Pseudorandom

حل: با در نظر گرفتن شرایط ساختن یک مثلث، جواب تحلیلی این مسأله برابر $p = \frac{1}{4}$ به دست می آید. 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 زایانه 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 می توانند هزاران عدد تصادفی تولید کنند.

مثال ۲ـ اگر نقطهٔ A به طور تصادفی در مربعی به ضلع واحد به طور یکنواخت توزیع شده باشد، مطلوب است احتمال این که فاصلهٔ A تا نزدیکترین ضلع مربع کمتر از فاصلهٔ 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، محاسبهٔ مقدار انتگرال

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

باشد که در آن g(x) تابعی حقیقی در $(-\infty, \infty)$ است. ایدهٔ 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 بازهٔ [a,b] تعریف می کنیم:

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

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

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

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

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

حال یک نمونهٔ تصادفی n تایی به صورت $x_1, x_2, \ldots, x_n$ از چگالی f(x) انتخاب کرده و مقدار $g(x_i)$ ، $i = 1, \ldots, n$ ، را محاسبه می کنیم. پس:

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

و بنابراین:

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

مثال ـ یک مسئلهٔ انتگرالگیری، برای توزیع نرمال مطرح می شود که می خواهیم 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. می دانیم که این انتگرال را نمی توان برحسب ترکیبات متناهی از توابع مقدماتی نوشت[٤]. 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 مقادیرمتناظر آنها که در توزیع نرمال آمده اند، در جدول (۳) مشاهده می کنید.

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

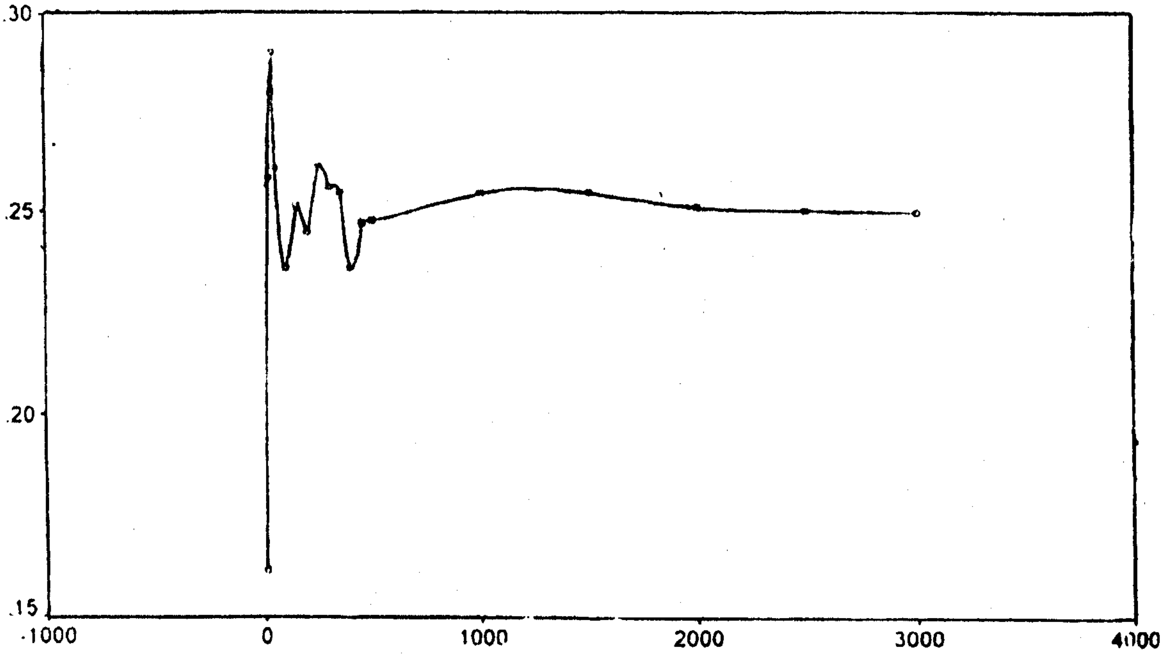
بازهٔ اطمینان ۹۵ درصدی برای p	احتمال p به روش مونت کارلو	تعداد تکرار آزمایش n
(۰/۱۱۴۵ , ۰/۲۰۸۳)	۰/۱۶۱۴	۱۰
(۰/۲۱۸۰ , ۰/۲۲۹۰)	۰/۲۵۸۵	۲۰
(۰/۲۱۳۷ , ۰/۳۴۴۱)	۰/۲۷۷۹	۳۰
(۰/۲۵۳۳ , ۰/۳۲۷۷)	۰/۲۹۰۵	۴۰
(۰/۲۲۵۹ , ۰/۲۹۵۷)	۰/۲۶۰۸	۵۰
(۰/۲۰۵۹ , ۰/۲۶۶۳)	۰/۲۳۶۱	۱۰۰
(۰/۲۲۱۸ , ۰/۲۸۰۸)	۰/۲۵۱۳	۱۵۰
(۰/۲۳۰۸ , ۰/۲۵۹۰)	۰/۲۴۴۹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

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

بازهٔ اطمینان ۹۵ درصدی برای p	احتمال p به روش مونت کارلو	تعداد تکرار آزمایش n
(۰/۴۴۶۰ , ۰/۶۳۴۰)	۰/۵۴۰۰	۱۰
(۰/۳۶۲۰ , ۰/۴۳۸۰)	۰/۴۰۰۰	۲۰
(۰/۳۸۰۶ , ۰/۴۸۶۰)	۰/۴۳۳۳	۳۰
(۰/۴۰۲۴ , ۰/۵۲۷۶)	۰/۴۶۵۰	۴۰
(۰/۴۵۲۲ , ۰/۴۹۹۸)	۰/۴۷۶۰	۵۰
(۰/۳۹۱۲ , ۰/۴۴۴۸)	۰/۴۱۸۰	۱۰۰
(۰/۳۷۱۲ , ۰/۴۱۰۲)	۰/۳۹۰۷	۱۵۰
(۰/۳۸۸۲ , ۰/۴۳۳۸)	۰/۴۱۱۰	۲۰۰
(۰/۴۰۰۷ , ۰/۴۳۴۵)	۰/۴۱۷۶	۲۵۰
(۰/۳۷۷۵ , ۰/۴۲۹۱)	۰/۴۰۳۳	۳۰۰
(۰/۳۸۶۱ , ۰/۴۲۵۳)	۰/۴۰۵۷	۳۵۰
(۰/۳۸۵۵ , ۰/۴۳۶۵)	۰/۴۱۱۰	۴۰۰
(۰/۳۷۳۰ , ۰/۴۵۸۲)	۰/۴۱۵۶	۴۵۰
(۰/۴۱۸۲ , ۰/۴۴۲۶)	۰/۴۳۰۴	۵۰۰
(۰/۳۹۶۹ , ۰/۴۱۸۷)	۰/۴۰۷۸	۱۰۰۰
(۰/۴۰۶۲ , ۰/۴۱۶۰)	۰/۴۱۱۱	۱۵۰۰

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

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



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

مراجع

[۱] بهبودیان، جواد؛ آمار ناپارامتری، 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، ترجمهٔ آذرنوش و نیرومند، (۱۳۷٤)، انتشارات دانشگاه مشهد.