کاربردی از قضیه حدی مرکزی

محمدحسین پورسعید[۱]

چکیده

در این مقاله، بر پایه تعبیری احتمالی، روش هایی برای تقریب $\sum_{x=a}^{b} g(x)$ و تعداد جوابهای معادله سیاله $x_1 + x_2 + \ldots + x_n = m \;;\; a_i \le x_i \le b_i \;,\; i = 1,2,\ldots n$ ارائه می شود.

واژه های کلیدی: معادله سیاله، تابع مولد، قضیه حدی مرکزی، قضیه تبدیل وارون، روش مونت کارلو.

۱. مقدمه

کاربرد توزیع نرمال، بی تردید بیش از سایر توزیع‌های احتمالی است. علاوه بر مسائل عملی این توزیع، در تقریب تعداد زیادی از توزیعها کارآیی دارد و بخاطر نقش عمده آن در قضیه حدی مرکزی، یکی از مهمترین توزیع های احتمال بشمار می رود.

تحت شرایط بسیار متعدد، وقتی که تعداد متغیرها در حاصل جمع یا میانگین متغیرهای تصادفی زیاد شود تابع توزیع این حاصل جمع یا میانگین به سمت تابع توزیع نرمال میل می کند. صورتی از قضیه حدی مرکزی در زیر آمده است.

اگر $X_1, X_2, X_3, \ldots$ دنباله ای از متغیرهای تصادفی مستقل و همتوزیع با میانگین μ و واریانس متناهی σ^2 باشد و دنباله متغیرهای تصادفی $Z_1, Z_2, Z_3, \ldots$ را بصورت زیر تعریف کنیم.

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \qquad n = 1,2,3,\ldots$$

آنگاه

$$\lim_{n \to \infty} F_{Z_n}(t) = \Phi(t)$$

که در آن $\Phi(0)$ تابع توزیع نرمال استاندارد است[۳].

حال اگر X_i ها همتوزیع نباشند، نمی توان از فرم بالا بهره برد با این حال تحت شرط کراندار بودن و مستقل بودن آن‌ها، توزیع مجانبی $\dfrac{\sum_{i=1}^{n}(X_i - E(X_i))}{\sqrt{\sum_{i=1}^{n} Var(X_i)}}$ نیز بصورت نرمال استاندارد است [۴].

۲. بحث و نتیجه گیری

در قسمت های الف و ب این بخش، روش تقریب تعداد جوابهای مشروط معادله سیاله $x_1 + x_2 + \ldots + x_n = m$ و تعیین بازه اطمینان برای $\sum_{x=a}^{b} g(x)$، به عنوان کاربردهایی از قضیه حدی مرکزی، معرفی می شوند.

الف) معادله سیاله $a_1 x_1 + a_2 x_2 + \ldots + a_n x_n = m$ را در نظر می گیریم. تعداد جوابهای مشروط ۱ معادله را هر گاه $c_i \le x_i \le d_i \;;\; i = 1,2,\ldots,n$ می توان با استفاده از تابع مولد محاسبه کرد که برای n های بزرگ، بکارگیری آن زمانبر است. لذا با ارائه تعبیری احتمالی از مسئله، روشی تقریبی معرفی می شود.

۱- گروه ریاضی دانشگاه لرستان.

به ازای تغییرات $x_i \; ; \; i = ۱,۲,...,n$ مقادیری مختلف را برای $a_۱ x_۱ + a_۲ x_۲ + ... + a_n x_n$ داریم که ممکن است بعضی از آن قادیر برابر m باشند و در واقع تعداد دفعاتی که $\sum_{i=۱}^{n} a_i x_i \; ; \; c_i \le x_i \le d_i$ برابر است با m، تعداد جوابهای مشروط معادله محسوب می شود. بنابر این اگر $x_i : \; i = ۱,۲,...,n$ را متغیر تصادفی گسسته یکنواخت X_i بدانیم، $a_i x_i$ را نیز می توان بصورت متغیری تصادفی با تکیه گاه $\{a_i c_i , a_i(c_i + ۱), a_i(c_i + ۲),...,a_i d_i\}$ در نظر گرفت تا با توجه به عبارت زیر و تقریب $P(\sum_{i=۱}^{n} a_i X_i = m)$، بتوان تعداد جوابهای مشروط معادله را بطور تقریبی محاسبه کرد.

$$P(\sum_{i=۱}^{n} a_i X_i = m) = \frac{N(\sum_{i=۱}^{n} a_i x_i = m, c_i \le x_i \le d_i ; i = ۱,۲,...,n)}{N(\sum_{i=۱}^{n} a_i x_i, c_i \le x_i \le d_i ; i = ۱,۰۲,...,n)}$$

$$\Rightarrow N(\sum_{i=۱}^{n} a_i x_i = m \; , \; c_i \le x_i \le d_i \; ; \; i = ۱,۲,..., n) =$$

$$P(\sum_{i=۱}^{n} a_i X_i = m) \times \prod_{i=۱}^{n} (d_i - c_i + ۱) \qquad (۱)$$

که در آن $N(\circ)$، نماد تعداد است.

با توجه به رابطه (۱)، که $\prod_{i=۱}^{n} (d_i - c_i + ۱)$ نقش ضریب $P(\sum_{i=۱}^{n} a_i X_i = m)$ را دارد، در می یابیم که بزرگی n و عدم دقت تقریب $P(\sum_{i=۱}^{n} a_i X_i = m)$ تاثیری منفی بر دقت تقریب تعداد جوابهای معادله دارد. از طرفی بکارگیری و دقت تقریب قضیه حدی مرکزی، مستلزم بزرگ بودن n است لذا حالتی خاص از معادله را بصورت $x_۱ + x_۲ + ... + x_n = m$ در نظر می گیریم تا به ازای n های نه چندان بزرگ، $P(\sum_{i=۱}^{n} X_i = m)$ و تعداد جوابهای معادله را بتوان با تقریبی مناسب محاسبه کرد.

مثال ۱: تعداد جوابهای معادله سیاله

$$x_۱ + x_۲ + ... + x_۷ = ۳۰$$

هر گاه

$$۱ \le x_۱ \le ۵ \; , ۰ \le x_۲ \le ۸ \; , \; -۲ \le x_۳ \le ۷ \; ,$$
$$۰ \le x_۴ \le ۶ \; , \; ۳ \le x_۵ \le ۹ \; , \; ۳ \le x_۶ \le ۷ \; , \; ۰ \le x_۷ \le ۵$$

را به صورت تقریبی می یابیم.

$$E(X_۱) = \frac{\sum_{x=۱}^{۵} x}{۵} = ۳ \, , \, E(X_۲) = \frac{\sum_{x=۰}^{۸} x}{۹} = ۴ \, , \, E(X_۳) = \frac{\sum_{x=-۲}^{۷} x}{۱۰} = ۲/۵ \, ,$$

$$E(X_۴) = \frac{\sum_{x=۰}^{۶} x}{۷} = ۳ \quad E(X_۵) = \frac{\sum_{x=۳}^{۹} x}{۷} = ۶ \; , \; E(X_۶) = \frac{\sum_{x=۳}^{۷} x}{۵} = ۵ \, ,$$

$$E(X_۷) = \frac{\sum_{x=۰}^{۵} x}{۶} = ۲/۵$$

به همین ترتیب، با محاسبه و جایگذاری $E(X_i^۲); \; i = ۱,...,۶$ داریم:

$$Var(X_۱) = ۲ \, , \, Var(X_۲) = \frac{۲۰}{۳} \, , \, Var(X_۳) = \frac{۳۳}{۴} \, , \, Var(X_۴) = ۴$$

$$Var(X_۵) = ۴ \, , \, Var(X_۶) = ۲ \, , \, Var(X_۷) = \frac{۳۵}{۱۲}$$

$$\Rightarrow \sum_{i=۱}^{۷} E(X_i) = ۲۶ \, , \quad \sum_{i=۱}^{۷} Var(X_i) = \frac{۱۷۹}{۶}$$

$$P(\sum_{i=۱}^{۷} X_i = ۳۰) = P(۲۹۵ < \sum_{i=۱}^{۷} X_i < ۳۰/۵)$$

$$= P(\frac{۲۹/۵ - ۲۶}{\sqrt{\frac{۱۷۹}{۶}}} < \frac{\sum_{i=۱}^{۷} X_i - \sum_{i=۱}^{۷} E(X_i)}{\sqrt{\sum_{i=۱}^{۷} Var(X_i)}} < \frac{۳۰/۵ - ۲۶}{\sqrt{\frac{۱۷۹}{۶}}}) \cong$$

$$P(۰/۶۴۰۷۹۲۱۰۷ < Z < ۰/۸۲۳۸۷۵۵۶۶)$$

برای تقریب بهتر، تاکید می‌شود که در استفاده از جدول نرمال استاندارد باید بطور دقیق عمل کرد.

$$\Phi(۰/۶۴۰۷۹۲۱۰۷) = ۰/۷۳۹۱۶۱۳۹۵ \; ,$$

$$\Phi(۰/۸۲۳۸۷۵۵۶۶) = ۰/۷۹۴۹۸۵۱۵۸$$

$$\Rightarrow P(۰/۶۴۰۷۹۲۱۰۷ < Z < ۰/۸۲۳۸۷۵۵۶۶)$$

$$= ۰/۰۵۵۸۲۳۷۶۳$$

$$\prod_{i=۱}^{۷} (d_i - c_i + ۱) = ۶۶۱۵۰۰$$

$$\Rightarrow (۰/۰۵۵۸۲۳۷۶۳) \times (۶۶۱۵۰۰) \cong ۳۶۹۲۷$$

بنابراین تقریب تعداد جوابها برابر است با ۳۶۹۲۷ که با استفاده از کامپیوتر و تعیین تعداد دقیق جوابها، خطاهای مطلق و نسبی تقریب نیز برابر است با

$$\Delta = ۳۶۹۲۷ - ۳۶۸۹۰ = ۳۷ \; , \; \delta = \frac{۳۷}{۳۶۸۹۰} \approx ۰/۰۰۱$$

ب) یکی از روش های عددی که با استفاده از اعداد تصادفی و شبیه سازی پدیده ها به بررسی مسائل می پردازد، روش مونت کارلو نام دارد. این روش در حل معادلات، محاسبه انتگرال های پیچیده و ... که حل آنها با روش های معمولی بسیار وقت گیر و پر هزینه است، کاربرد دارد.

در مقاله "روش شبیه سازی مونت کارلو و کاربردهای آن در آمار و احتمال" در مورد شبیه سازی متغیرهای تصادفی پیوسته، تولید اعداد تصادفی از هر توزیع دلخواه با استفاده از کامپیوتر و... توضیحاتی ارائه شده است [۲] . با این حال، در این قسمت یک بازه اطمینان برای $\sum_{x=a}^{b} g(x)$ ارائه می شود که با توجه به یکنواخت بودن توزیع مربوطه، استناد به قضیه تبدیل وارون و استفاده از کامپیوتر لزومی نخواهد داشت.

فرض کنید که $\sum_{x=a}^{b} g(x)$ فرمول محاسباتی نداشته و $b-a$ نیز عددی بزرگ باشد. شبیه قسمت الف، توزیع احتمالی متغیر تصادفی یکنواخت گسسته X را بصورت زیر تعریف می کنیم.

$$P(X=x)=\frac{1}{b-a+1} \qquad x=a,a+1,\dots,b$$

بنابراین داریم:

$$E(g(X))=\sum_{x=a}^{b} g(x)P(X=x)=\frac{1}{b-a+1}\sum_{x=a}^{b} g(x)$$

$$\Rightarrow \sum_{x=a}^{b} g(x)=(b-a+1)\times E(g(X))$$

در صورت کراندار بودن $g(x)$، می توان با استفاده از کامپیوتر یا جدول اعداد تصادفی n عدد از $a,a+1,\dots,b$ را تولید کرده، یک بازه اطمینان برای $\sum_{x=a}^{b} g(x)$ ارائه نمود. با این حال، با توجه به سهولت و عمومیت استفاده از ماشین حساب، روشی مناسب معرفی می شود.

همانطور که می دانیم توزیع اعداد تصادفی در ماشین حسابهای معمولی بصورت زیر است.

$$P(Y=y)=\frac{1}{1000} \qquad y=0,0/001,0/002,\dots,0/999$$

با توجه به اینکه X و Y دارای توزیع یکنواخت گسسته هستند، برای تولید اعدادی تصادفی از مجموعه اعداد $\{a,a+1,\dots,b\}$ تبدیلی خطی از Y بصورت $X=\frac{b-a}{0/999}Y+a$ تقریباً مناسب خواهد بود. لازم به ذکر است که بکار گیری روش پیشنهادی در تولید اعداد تصادفی، ممکن است باعث تولید اعداد اعشاری بجای اعدادی صحیح و پوشش ندادن کلیه اعداد a تا b و در نتیجه کاهش دقت در تعیین بازه اطمینان شود.

به هر حال با استفاده از ماشین حساب و تولید اعداد تصادفی $y_1,y_2,\dots,y_n$ ، یک بازه اطمینان $\%100(1-\alpha)$ برای $\sum_{x=a}^{b} g(x)$ بصورت زیر است.

$$\left[(b-a+1)\left(\bar{g}-z_{\frac{\alpha}{2}}\times \frac{S_g}{\sqrt{n}}\right),\ (b-a+1)\left(\bar{g}+z_{\frac{\alpha}{2}}\times \frac{S_g}{\sqrt{n}}\right)\right]$$

که در آن

$$\bar{g}=\frac{\sum_{i=1}^{n} g(x_i)}{n}=\frac{\sum_{i=1}^{n} g\left(\frac{100\cdot(b-a)}{999}y_i+a\right)}{n}$$

$$S_g^2=\frac{\sum_{i=1}^{n} g^2\left(\frac{100\cdot(b-a)}{999}y_i+a\right)-n(\bar{g})^2}{n-1},$$

$$\Phi(z_\alpha)=\alpha$$

مثال ۲: با توجه به روش پیشنهادی و تولید ۳۰ عدد تصادفی(جدول۱)، بازه های اطمینان ۹۵٪ برای

الف: $\sum_{x=1}^{500}\frac{2x^2+3x+4}{2x^3+3x^2+4x+5}=6/34$

ب: $\sum_{x=1}^{500}\frac{x^3+2x^2+3x+4}{2x^3+3x^2+4x+5}=251/71$

پ: $\sum_{x=1}^{500}\frac{x^4+2x^3+3x^2+4x}{2x^3+3x^2+4x+5}=62750/56$

ت: $\sum_{x=1}^{500}\frac{x^5+2x^4+3x^3+4x^2}{2x^3+3x^2+4x+5}=20927250$

به صورت زیر است:

الف: $(3/06,9/45)$

ب: $(250/77,252/44)$

پ: $(44715/52,66084/1)$

ت: $(11981890,22302650)$.

همانطور که ملاحظه می شود، دقت روش پیشنهادی بستگی به فرم تابع $g(x)$ دارد که با تولید بیشتر اعداد تصادفی می توان دقت آنرا نیز افزایش داد.

۲۸ □ سال دهم، شماره دوم ـــــــ