آمار و احتمال مهندسی

اساتید: دکتر توسلی‌پور، دکتر وهابی

دانشکده مهندسی برق و کامپیوتر، دانشکدگان فنی، دانشگاه تهران

تمرین ششم ــ نمونه‌گیری، نامساوی‌ها، قانون اعداد بزرگ، قضیه حد مرکزی

طراح: محمدرضا علوی

سوپروایزر: سارا معصومی

تاریخ تحویل: ۱۵ دی ۱۴۰۳

بیشتر بدانیم: نامساوی جنسن

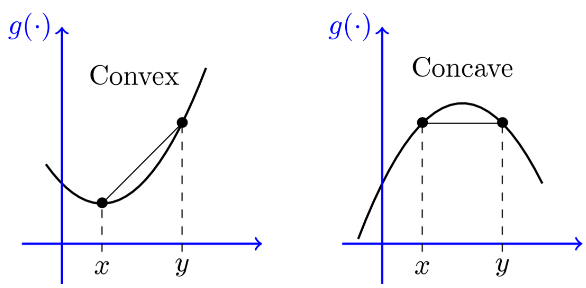
حتماً قبلاً نامساوی زیر را که به نامساوی میانگین حسابی‌ــ‌هندسی[1] معروف است دیده‌اید.

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 \cdot x_2 \cdot \cdots \cdot x_n} \tag{1}$$

این نامساوی برای تمام x_iهای حقیقی و نامنفی برقرار است. آیا می‌توانید با یک روش آمار و احتمالاتی نامساوی بالا را اثبات کنید؟ برای انجام این کار ابتدا باید با توابع محدب[2] و نامساوی جنسن[3] آشنا شویم.

تابع محدب:

تعریف. فرض کنید $f : \mathbb{R} \to \mathbb{R}$ یک تابع دو بار مشتق‌پذیر (تابعی که مشتق دوم دارد) باشد. این تابع محدب است اگر و تنها اگر مشتق دوم آن در تمام نقاط دامنه غیرمنفی باشد. برای مثال تابع x^2 یک تابع محدب است.



شکل ۱: تفاوت تابع محدب و مقعر

نامساوی جنسن:

قضیه. اگر f یک تابع محدب و X یک متغیر تصادفی باشد و همچنین $E[f(X)]$ و $f(E[X])$ وجود داشته باشند خواهیم داشت،

$$E[f(X)] \geq f(E[X])$$

و تساوی تنها در حالتی برقرار است که f یک تابع خطی باشد.

نامساوی جنسن ابزاری قدرتمند است و به طور ویژه در زمینه‌های علم داده، اقتصاد و نظریه اطلاعات استفاده می‌شود.

---

[1] Arithmetic Mean Geometric Mean Inequality

[2] Convex Functions

[3] Jensen's Inequality

حال می‌خواهیم با استفاده از این نامساوی عبارت ۱ را ثابت کنیم. ابتدا از دو طرف عبارت لگاریتم می‌گیریم. پس کافی‌ست نشان دهیم،

$$\ln\left(\frac{x_۱+x_۲+\cdots+x_n}{n}\right) \geq \frac{۱}{n}\left(\ln x_۱+\cdots+\ln x_n\right) \qquad (۲)$$

می‌دانیم تابع $\ln x$ یک تابع مقعر است، در نتیجه تابع $-\ln x$ یک تابع محدب خواهد بود. همچنین فرض کنید X یک متغیر تصادفی است به صورتی که $P(X \in \{x_۱,\ldots,x_n\}) = ۱$ و X هر کدام از مقادیر $x_۱,\ldots,x_n$ را با احتمال یکسان $\frac{۱}{n}$ می‌پذیرد. در این صورت داریم،

$$E[X] = \sum_{i=۱}^{n} \frac{۱}{n} x_i = \frac{x_۱+\cdots+x_n}{n}$$

در نتیجه با استفاده از نامساوی جنسن داریم،

$$\begin{aligned} E[-\ln X] &\geq -\ln E[X] \\ \longrightarrow \ln E[X] &\geq E[\ln X] \\ \longrightarrow \ln\left(\frac{x_۱+x_۲+\cdots+x_n}{n}\right) &\geq \frac{\ln x_۱+\cdots+\ln x_n}{n} \\ \xrightarrow{\text{Exponentiating both sides}} \frac{x_۱+x_۲+\cdots+x_n}{n} &\geq \sqrt[n]{x_۱\cdot x_۲\cdot\cdots\cdot x_n} \end{aligned}$$

و به این شکل نامساوی ۱ ثابت خواهد شد. از نامساوی جنسن برای حل نامساوی‌های دیگر جبری نیز استفاده می‌شود که می‌توانید مثال‌های دیگر آن را در این‌جا مشاهده کنید.

۱. اثر دارو — ۱۵ نمره

یک شرکت داروسازی در حال آزمایش یک داروی جدید است. در این آزمایش زمان تاثیر دارو بعد از مصرف با متغیر تصادفی T بر حسب ساعت نشان داده می‌شود. می‌دانیم میانگین زمان تاثیر دارو پس از مصرف ۴ ساعت است. حال به سوالات زیر پاسخ دهید.

آ. یک کران بالا برای احتمال این که زمان تاثیر دارو حداقل ۱۰ ساعت طول بکشد به دست آورید.

ب. فرض کنید ۵ ساعت از زمان مصرف این دارو توسط یک بیمار گذشته است و دارو هنوز اثر نکرده است؛ اگر تا ۵ ساعت آینده نیز دارو تاثیری روی بیمار نگذارد وضعیت این بیمار وخیم خواهد شد. احتمال این که این اتفاق ناگوار رخ دهد حداکثر چقدر است؟ (می‌دانیم $E(T|T\geq ۵) = ۸$).

ج. فرض کنید T دارای توزیع نمایی است و احتمال دقیق این که زمان تاثیر دارو حداقل ۱۰ ساعت باشد را محاسبه کنید. جواب به دست آمده را با جواب به دست امده در قسمت آ مقایسه کنید، چه نتیجه‌ای می‌گیرید؟

پاسخ :

آ. طبق نامساوی مارکوف:

$$P(T\geq a) \leq \frac{E[T]}{a}$$

در نتیجه،

$$P(T\geq ۱۰) \leq \frac{۴}{۱۰} = ۰/۴$$

ب. اگر $\mathcal{F}$ یک پیشامد باشد، با توجه به نامساوی مارکوف شرطی[۴] خواهیم داشت:

$$P(T \geq a|\mathcal{F}) \leq \frac{E[T|\mathcal{F}]}{a}$$

در نتیجه،

$$P(T \geq ۱۰|T \geq ۵) \leq \frac{E[T|T \geq ۵]}{۱۰} = ۰/۸$$

ج. از آن‌جایی که $E(T) = ۴$، نتیجه می‌گیریم $\lambda = \frac{۱}{۴}$ و خواهیم داشت:

$$P(T \geq t) = e^{-\lambda t} = e^{-\frac{t}{۴}}, \qquad t \geq ۰$$

$$\longrightarrow P(T \geq ۱۰) = e^{-\frac{۱۰}{۴}} \approx ۰/۰۸۲$$

کران بالایی که در قسمت آ برای این احتمال به دست آوردیم ۰/۴ بود و از آن‌جایی که این عدد با احتمال واقعی این پیشامد فاصله قابل توجهی دارد، نتیجه می‌گیریم نامساوی مارکوف در این‌جا تخمین مناسبی را ارائه نمی‌دهد.

۲. سود ماهانه ۱۵ نمره

یک شرکت کوچک درحال بررسی فروش ماهانه خود است. این شرکت می‌داند میانگین فروش ماهانه‌اش ۵۰۰۰۰ دلار و انحراف معیار آن ۸۰۰۰ دلار است. با توجه به این اطلاعات به سوالات زیر پاسخ دهید.

آ. بیشینه احتمال این که در یک ماه فاصله‌ی فروش از میانگین حداقل ۲۴۰۰۰ باشد را به دست آورید.

ب. فرض کنید میانگین هزینه‌های ماهانه این شرکت ۳۰۰۰۰ دلار و انحراف معیار آن ۵۰۰۰ دلار باشد. احتمال این که در یک ماه سود این شرکت کم‌تر از ۲۵۰۰۰ دلار با میانگین سود ماهانه فاصله داشته باشد حداقل چقدر است؟ (فرض کنید در این شرکت مقدار هزینه ماهانه مستقل از مقدار فروش ماهانه است.)

ج. با فرض این که در فصل تعطیلات انحراف معیار فروش ماهانه دو برابر می‌شود، بیشینه احتمال این که فروش در این دوره یا حداقل ۸۰۰۰۰ دلار باشد و یا حداکثر ۲۰۰۰۰ دلار باشد را پیدا کنید.

پاسخ :

آ. فرض می‌کنیم متغیر تصادفی X نشانگر میزان فروش ماهانه این شرکت باشد. با توجه به نامساوی چبیشف[۵] می‌دانیم:

$$P(|X - \mu_X| \geq k) \leq \frac{\sigma_X^۲}{k^۲}$$

پس داریم،

$$P(|X - ۵۰۰۰۰| \geq ۲۴۰۰۰) \leq \frac{۸۰۰۰^۲}{۲۴۰۰۰^۲} = \frac{۱}{۹} \approx ۰/۱۱$$

[۴] Conditional Markov's Inequality
[۵] Chebyshev's Inequality

ب. فرض می‌کنیم متغیر تصادفی Y نشانگر هزینه‌های ماهانه این شرکت باشد. در نتیجه سود ماهانه شرکت را می‌توان با متغیر تصادفی $Z = X - Y$ نشان داد. ابتدا امید و انحراف معیار Z را محاسبه می‌کنیم.

$$\begin{aligned} \mu_Z = E(Z) &= E(X - Y) \\ &= E(X) - E(Y) \\ &= ۵۰۰۰۰ - ۳۰۰۰۰ \\ &= ۲۰۰۰۰ \end{aligned}$$

$$\begin{aligned} \sigma_Z &= \sqrt{\sigma_X^۲ + \sigma_Y^۲} \\ &= \sqrt{۸۰۰۰^۲ + ۵۰۰۰^۲} \\ &\approx ۹۴۳۴ \end{aligned}$$

حال می‌خواهیم حداقل مقدار احتمال $P(|Z - \mu_z| < ۲۵۰۰۰)$ را به دست آوریم. با توجه به نامساوی چبیشف داریم:

$$P(|Z - ۲۰۰۰۰| \geq ۲۵۰۰۰) \leq \frac{۹۴۳۴^۲}{۲۵۰۰۰^۲} \approx ۰/۱۴$$

$$\longrightarrow P(|Z - ۲۰۰۰۰| < ۲۵۰۰۰) = ۱ - P(|Z - ۲۰۰۰۰| \geq ۲۵۰۰۰) > ۱ - ۰/۱۴$$

$$\longrightarrow P(|Z - ۲۰۰۰۰| < ۲۵۰۰۰) > ۰/۸۶$$

پس با احتمال حداقل ۸۶ درصد، سود شرکت در این بازه قرار خواهد داشت.

ج. به‌طور مشابه با توجه به نامساوی چبیشف داریم:

$$\begin{aligned} P(X \geq ۸۰۰۰۰ \text{ or } X \leq ۲۰۰۰۰) &= P(|X - ۵۰۰۰۰| \geq ۳۰۰۰۰) \\ &\leq \frac{۱۶۰۰۰^۲}{۳۰۰۰۰^۲} \approx ۰/۲۸ \end{aligned}$$

## ۳. توابع نااریب	۱۵ نمره

فرض کنید $X_۱, X_۲, ..., X_n$ یک نمونه تصادفی از توزیع $U(۰, \theta)$ باشد.

آ. تابعی از $\bar{X}$ پیدا کنید که برآوردگری نااریب[۶] برای θ باشد.

ب. تابعی از آماره رتبه[۷] $X_{(n)}$ پیدا کنید که برآوردگری نااریب برای θ باشد.

پاسخ :

آ. می‌دانیم امید ریاضی این توزیع یکنواخت برابر است با:

$$E(X) = \frac{\theta + ۰}{۲} = \frac{\theta}{۲}$$

[۶] Unbiased Estimator
[۷] Order Statistic

در نتیجه برای محاسبه میانگین $\bar{X}$ داریم،

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = E(X) = \frac{\theta}{2}$$

در نتیجه $g(\bar{X}) = 2\bar{X}$ یک برآوردگر نااریب برای پارامتر θ است زیرا $E(g(\bar{X})) = \theta$.

ب. از آن‌جایی که توزیع X یکنواخت است داریم،

$$F(x) = \frac{x-0}{\theta} = \frac{x}{\theta}, \qquad 0 \le x \le \theta$$

پس توزیع آماره رتبه آخر به صورت زیر است،

$$\begin{aligned} F_n(x) = P(X_{(n)} \le x) &= P(X_1 \le x, \ldots, X_n \le x) \\ &= F^n(x) \\ &= \frac{x^n}{\theta^n}, \qquad 0 \le x \le \theta \end{aligned}$$

سپس با مشتق‌گیری داریم،

$$f_n(x) = \frac{d}{dx}F_n(x) = \frac{nx^{n-1}}{\theta^n}, \qquad 0 \le x \le \theta$$

حال امیدریاضی این آماره را محاسبه می‌کنیم،

$$\begin{aligned} E(X_{(n)}) = \int_0^\theta x f_n(x)\,dx &= \int_0^\theta x \cdot \frac{nx^{n-1}}{\theta^n}\,dx \\ &= \frac{n}{\theta^n}\int_0^\theta x^n\,dx \\ &= \frac{n}{\theta^n}\left(\frac{x^{n+1}}{n+1}\right)\Bigg|_0^\theta \\ &= \frac{n\theta}{n+1} \end{aligned}$$

در نتیجه $f(X_{(n)}) = \frac{n+1}{n}X_{(n)}$ یک برآوردگر نااریب برای θ است.

۴. شکار پرندگان — ۱۵ نمره

در یک جنگل کوهستانی ۵۵ درصد عقاب‌ها ماده هستند. چند عقاب باید شکار کنید تا با احتمال ۹۰ درصد مطمئن باشید که حداقل نیمی از پرندگان شکار شده ماده هستند؟ هر فرضی را که در راه‌حل خود در نظر گرفته‌اید ذکر کنید.

پاسخ :

می‌دانیم جنسیت یک پرنده یک متغیر تصادفی با توزیع برنولی است. فرض می‌کنیم جنسیت هر پرنده مستقل از پرنده‌های دیگر باشد. پس می‌توان جنسیت پرندگان شکار شده را با $X_1, \ldots, X_n \sim B(1, 0/55)$ نشان داد. به طوری که $X_i = 1$ اگر پرنده iاُم ماده باشد و در

غیر اینصورت $X_i = ۰$.
تعریف می‌کنیم،

$$S_n = \sum_{i=۱}^{n} X_i$$

واضح است که S_n نشانگر تعداد پرنده‌های ماده شکار شده است. حال می‌خواهیم n را طوری تعیین کنیم که داشته باشیم،

$$P\left(S_n \geq \frac{n}{۲}\right) = ۰/۹$$

ابتدا باید توزیع S_n را با استفاده از قضیه حد مرکزی تقریب بزنیم.
می‌دانیم $E(X_i) = p = ۰/۵۵$ و $Var(X_i) = p(۱-p) = ۰/۲۴۷$. در نتیجه با توجه به قضیه حد مرکزی داریم،

$$S_n \sim N(۰/۵۵n, ۰/۲۴۷n)$$
$$\longrightarrow Z = \frac{S_n - ۰/۵۵n}{\sqrt{۰/۲۴۷n}} \sim N(۰, ۱)$$

پس داریم،

$$\begin{aligned} P\left(S_n \geq \frac{n}{۲}\right) &= P\left(\frac{S_n - ۰/۵۵n}{\sqrt{۰/۲۴۷n}} \geq \frac{n/۲ - ۰/۵۵n}{\sqrt{۰/۲۴۷n}}\right) \\ &= P(Z \geq -\frac{۰/۰۵n}{\sqrt{۰/۲۴۷n}}) \\ &= P(Z \leq \frac{۰/۰۵n}{\sqrt{۰/۲۴۷n}}) \\ &= ۰/۹ \end{aligned}$$

با توجه به جدول توزیع نرمال استاندارد داریم، $\frac{۰/۰۵n}{\sqrt{۰/۲۴۷n}} \approx ۱/۲۸۱۶$ و در نتیجه $n \approx ۱۶۳$.

۵. توزیع پارتو　　　۱۰ نمره

فرض کنید $X_۱, ..., X_n$ یک نمونه تصادفی از توزیع پارتو[8] با پارامترهای $\alpha > ۰$ (shape parameter) و $x_m > ۰$ (scale parameter) باشد. تابع چگالی احتمال این توزیع به صورت زیر است،

$$f(x; \alpha, x_m) = \frac{\alpha x_m^\alpha}{x^{\alpha+۱}}, \qquad x \geq x_m$$

آ. توضیح دهید برای چه مقادیری از α می‌توان قضیه حد مرکزی را برای تخمین زدن توزیع میانگین نمونه استفاده کرد؟

ب. برای $\alpha = ۳$ و $x_m = ۱$ توزیع تقریبی $\bar{X}_{۳۰}$ را محاسبه کنید.

پاسخ :

آ. میانگین و واریانس توزیع پارتو به صورت زیر است:

$$\mu = \begin{cases} \frac{\alpha x_m}{\alpha - ۱} & \alpha > ۱, \\ \infty & \alpha \leq ۱ \end{cases}$$
$$\sigma^۲ = \begin{cases} \frac{\alpha x_m^۲}{(\alpha-۱)^۲(\alpha-۲)} & \alpha > ۲, \\ \infty & \alpha \leq ۲ \end{cases}$$

[8] Pareto Distribution

می‌دانیم قضیه حد مرکزی فقط می‌تواند روی توزیع‌هایی اعمال شود که میانگین و واریانس متناهی داشته باشند. پس برای استفاده از این قضیه در این‌جا، باید داشته باشیم $\alpha > ۲$.

ب. ابتدا میانگین و واریانس توزیع را محاسبه می‌کنیم،

$$\mu = \frac{\alpha x_m}{\alpha - ۱} = \frac{۳}{۲} = ۱/۵$$
$$\sigma^۲ = \frac{\alpha x_m^۲}{(\alpha - ۱)^۲(\alpha - ۲)} = \frac{۳}{۴} = ۰/۷۵$$

طبق قضیه حد مرکزی داریم،

$$\bar{X}_{۳۰} \sim N(\mu, \frac{\sigma^۲}{n})$$
$$\longrightarrow \bar{X}_{۳۰} \sim N(۱/۵, ۰/۰۲۵)$$

۶. آماره رتبه عجیب! ۱۵ نمره

متغیر تصادفی X دارای توزیع زیر است:

$$P(X = x) = \frac{(e - ۱)^{-۱}}{x!}, \qquad x = ۱, ۲, ۳, \ldots$$

همچنین $E_۱, E_۲, \ldots$ متغیرهای تصادفی مستقل و هم توزیع از توزیع نمایی با پارامتر $\lambda = ۲$ هستند. توزیع $Z = \min\{E_۱, ..., E_X\}$ را به دست آورید.

پاسخ :

با توجه به قانون احتمال کل[۹] داریم،

$$\begin{aligned}
P(Z > z) &= \sum_{x=۱}^{\infty} P(Z > z|x)P(X = x) \\
&= \sum_{x=۱}^{\infty} P(E_۱ > z, \ldots, E_x > z|x)P(X = x) \\
&= \sum_{x=۱}^{\infty} \prod_{i=۱}^{x} P(E_i > z)P(X = x) \\
&= \sum_{x=۱}^{\infty} P(E_i > z)^x \, P(X = x) \\
&= \sum_{x=۱}^{\infty} (e^{-۲z})^x \, \frac{۱}{(e - ۱)x!} \\
&= \frac{۱}{e - ۱} \sum_{x=۱}^{\infty} \frac{(e^{-۲z})^x}{x!} \\
&= \frac{۱}{e - ۱} \left(e^{e^{-۲z}} - ۱\right), \qquad z \geq ۰
\end{aligned}$$

[۹] Law of total probability

۷. تقریب توزیع میانگین ۱۵ نمره

فرض کنید $\bar{X}_۱$ و $\bar{X}_۲$ میانگین‌های دو نمونه تصادفی مستقل از هم به اندازه n از جامعه‌ای با واریانس $\sigma^۲$ باشند. مقدار n را طوری مشخص کنید که $P\left(|\bar{X}_۱ - \bar{X}_۲| < \frac{\sigma}{۵}\right) \approx ۰/۹۹$ برقرار باشد.

پاسخ :

با استفاده از قضیه حد مرکزی می‌توانیم توزیع $\bar{X}_۱$ و $\bar{X}_۲$ را به صورت زیر تقریب بزنیم،

$$\bar{X}_۱ \sim N(\mu, \frac{\sigma^۲}{n}), \quad \bar{X}_۲ \sim N(\mu, \frac{\sigma^۲}{n})$$

از آن جایی که $\bar{X}_۱$ و $\bar{X}_۲$ مستقل هستند داریم،

$$\bar{X}_۱ - \bar{X}_۲ \sim N(۰, \frac{۲\sigma^۲}{n})$$

می‌خواهیم n را طوری پیدا کنیم که،

$$\begin{aligned} ۰/۹۹ &\approx P\left(|\bar{X}_۱ - \bar{X}_۲| < \frac{\sigma}{۵}\right) \\ &= P\left(\frac{-\sigma/۵}{\sigma/\sqrt{n/۲}} < \frac{\bar{X}_۱ - \bar{X}_۲}{\sigma/\sqrt{n/۲}} < \frac{\sigma/۵}{\sigma/\sqrt{n/۲}}\right) \\ &\approx P\left(-\frac{۱}{۵}\sqrt{\frac{n}{۲}} < Z < \frac{۱}{۵}\sqrt{\frac{n}{۲}}\right) \end{aligned}$$

که در آن $Z \sim N(۰, ۱)$ است. بنابراین باید داشته باشیم،

$$۲P\left(Z \geq \frac{۱}{۵}\sqrt{\frac{n}{۲}}\right) \approx ۰/۰۱ \longrightarrow P(Z \geq \frac{۱}{۵}\sqrt{\frac{n}{۲}}) \approx ۰/۰۰۵$$

با توجه به جدول توزیع نرمال استاندارد $\frac{۱}{۵}\sqrt{\frac{n}{۲}} = ۲/۵۷۶$ و در نتیجه $n \approx ۳۳۲$.