یک نکته ازاین معنی: کامل بودن

سید محمود طاهری*

چکیده

مفهوم کامل بودن از دیدگاه شهودی و ساختاری بررسی می‌شود. با چند مثال تشریح می‌کنیم که چگونه یک خانواده از توزیعها کامل می‌شود و یا کامل بودن را از دست می‌دهد. در پایان دو مقایسه بین ویژگی‌های کامل بودن، بسندگی و نااریبی انجام می‌شود.

۱ مقدمه

آیا تاکنون به این پرسش اندیشیده‌اید که رابطهٔ بین مفهوم آماری کامل بودن با مفهوم کامل و تمام که در زبان روزمرّه به کار می‌بریم چیست؟ مقالهٔ حاضر تلاشی است در پاسخ به این پرسش.

شایان یادآوری است که مباحث مربوط به کامل بودن بسیار است[۲]. امّا آنچه در اغلب کتابهای آماری پوشیده مانده است، مفهوم شهودی آن می‌باشد. در بخش بعد توضیح می‌دهیم که کامل بودن از دید شهودی به چه معناست. روش ما این است که یک خانواده از توابع را می‌سازیم و به مرور آن را کامل می‌کنیم. همچنین با مثالی دیگر توضیح می‌دهیم که چگونه خانواده‌ای که کامل است، این ویژگی را از دست می‌دهد. در بخش سوم دو مقایسه بین ویژگی‌های کامل بودن، بسندگی و نااریبی انجام می‌دهیم.

۲ کامل بودن

تعریف ۱. فرض کنید $P=\{f(x;\theta),\theta\in\Theta\}$ خانواده‌ای از توابع چگالی (یا جرم) احتمال باشد. گوییم این خانواده کامل است اگر رابطهٔ

$$E_\theta g(X)=0,\qquad \forall\,\theta\in\Theta \tag{۱}$$

نتیجه دهد که

$$P_\theta[g(X)=0]=1,\qquad \forall\,\theta\in\Theta. \tag{۲}$$

به عبارت ساده، خانواده P کامل است اگر جز برآوردگر بدیهی $g(X)=0$ برآوردگر نااریب دیگری برای صفر وجود نداشته باشد.

ملاحظه می‌کنید که کامل بودن در واقع یک ویژگی برای خانواده‌ای از توزیع‌هاست. گاهی این ویژگی به یک آماره نسبت داده می‌شود. این هنگامی است که خانواده توزیع‌های آماره کامل باشد. یعنی:

تعریف ۲. آمارهٔ $T(X)$ را کامل گوییم اگر خانواده توزیع‌های آماره T کامل باشد.

به بیان ساده آمارهٔ T کامل است اگر هر تابع غیر ثابتی از T در نظر بگیریم، امید ریاضی آن به θ بستگی داشته باشد، و این (با کمی مسامحه) یعنی این طور نباشد که بتوان از T، θ را حذف کرد.

دو تعریف فوق، تعریف‌های متداول کامل بودن است. برای درک شهودی مفهوم کامل بودن به توضیح زیر و دو مثال بعد آن توجه کنید.

*عضو هیئت علمی دانشگاه صنعتی اصفهان

چنانچه فضای پارامتر، Θ، کوچک باشد غالباً توابع بسیاری وجود دارد که در شرط (۱) صدق کنند. با اضافه شدن یک مقدار جدید مثلاً θ^* به Θ، بسا که برای تعدادی از توابع قبلی $E_{\theta^*}g(X) \neq 0$ و لذا مجموعهٔ کوچکتری از توابع در شرط (۱) صدق کنند. پس با بزرگتر شدن فضای پارامتر، شرط روی $g(x)$ قویتر می‌گردد و ردهٔ توابع $g(x)$ که در (۱) صدق می‌کنند کوچکتر می‌شود. حال اگر Θ آن قدر بزرگ شود که دیگر هیچ تابعی (جز تابع بدیهی $g(x) = 0$) در شرط (۱) صدق نکند، گوییم خانواده P کامل شده است. از اینجا به بعد هر چه Θ را بزرگتر کنیم کامل بودن P همچنان برقرار می‌ماند. (در واقع غیر از $g(x) = 0$ تابعی باقی نمانده است که حذف شود!)

مثال ۱. می‌خواهیم بررسی کنیم که آیا خانواده توزیع‌های دوجمله‌ای $P = \{b(3,\theta), \theta \in \Theta = \{\frac{1}{4}, \frac{3}{4}\}\}$ کامل است یا خیر. فرض کنید $X \sim b(3,\theta)$. چون X تنها مقادیر ۰و۱و۲و۳ را اختیار می‌کند، پس شرط (۱) به صورت زیر در می‌آید.

$$\sum_x g(x)P_\theta(X=x) = 0, \qquad \forall\, \theta \in \Theta$$

$$g(0)P_\theta(X=0) + g(1)P_\theta(X=1) + g(2)P_\theta(X=2) + g(3)P_\theta(X=3) = 0, \qquad \forall \theta \in \Theta$$

چون فضای پارامتر دو مقدار دارد، لذا شرط بالا معادل این است که

$$\begin{cases} \frac{27}{64}g(0) + \frac{27}{64}g(1) + \frac{9}{64}g(2) + \frac{1}{64}g(3) = 0 & \text{برای } \theta = \frac{1}{4} \\ \frac{27}{64}g(0) + \frac{27}{64}g(1) + \frac{9}{64}g(2) + \frac{1}{64}g(3) = 0 & \text{برای } \theta = \frac{1}{4} \end{cases}$$

که یک دستگاه دو معادله و چهار مجهولی است که بی‌نهایت جواب دارد. یعنی بی‌نهایت تابع غیر متحد با صفر می‌توان تعریف کرد که در شرط (۱) صدق کنند. مثلاً تابع زیر

$$g(x) = \begin{cases} 1 & x = 0, 1 \\ -\frac{181}{27} & x = 2 \\ \frac{19}{3} & x = 3 \end{cases}$$

در نتیجه خانواده P کامل نیست.

اکنون فضای پارامتر را بزرگتر می‌کنیم (تا بتوانیم شرط (۱) را قویتر کنیم). مثلاً می‌گیریم $\Theta = \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$. در این حالت، مشابه بحث بالا، شرط (۱) پس از ساده‌سازی به شکل زیر درمی‌آید.

$$\begin{cases} 27g(0) + 27g(1) + 9g(2) + g(3) = 0 & \text{برای } \theta = \frac{1}{4} \\ g(0) + 3g(1) + 3g(2) + g(3) = 0 & \text{برای } \theta = \frac{1}{2} \\ g(0) + 9g(1) + 27g(2) + 27g(3) = 0 & \text{برای } \theta = \frac{3}{4} \end{cases}$$

که یک دستگاه سه معادله و چهار مجهولی است که باز هم بی‌نهایت جواب دارد. اما دقت کنید که با بزرگتر شدن Θ، ردهٔ توابع $g(x)$ که در (۱) صدق می‌کنند کوچکتر شده است. مثلاً تابع $g(x)$ بالا در معادلهٔ مربوط به مقدار پارامتر $\theta = \frac{1}{2}$ و در نتیجه در شرط (۱) صدق نمی‌کند. اما به هر حال خانواده فوق هنوز کامل نشده است و می‌توان $g(x)$های بسیاری یافت که در هر سه معادلهٔ بالا و یعنی در شرط (۱) صدق کنند.

باز هم فضای پارامتر را با افزودن یک عضو دیگر به آن، بزرگتر می‌کنیم. می‌گیریم $\Theta = \{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{3}{4}\}$. با افزودن یک عضو دیگر در این حالت شرط (۱) پس از ساده سازی به شکل زیر درمی‌آید.

$$\begin{cases} 27g(0) + 27g(1) + 9g(2) + g(3) = 0 & \text{برای } \theta = \frac{1}{4} \\ 8g(0) + 12g(1) + 6g(2) + g(3) = 0 & \text{برای } \theta = \frac{1}{3} \\ g(0) + 3g(1) + 3g(2) + g(3) = 0 & \text{برای } \theta = \frac{1}{2} \\ g(0) + g(1) + 27g(2) + 27g(3) = 0 & \text{برای } \theta = \frac{3}{4} \end{cases}$$

دستگاه چهار معادله و چهار مجهولی بالا هیچ جواب غیر بدیهی ندارد و تنها جواب آن عبارت است از $g(x) = 0, x = 0, 1, 2, 3$. پس طبق تعریف، در این حالت شرط (۱)، شرط (۲) را نتیجه می‌دهد و لذا خانواده توزیع‌های $\{b(3,\theta), \theta \in \{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{3}{4}\}\}$ کامل است. واضح است که هر چه به فضای پارامتر بالا بیفزاییم، کامل بودن خانواده حفظ می‌شود.

در ادامه مثالی می‌آوریم تا روشن شود در حالتی که حتی فضای پارامتر شامل بی‌نهایت مقدار است، حذف یک مقدار پارامتر هم ممکن است باعث شود کامل بودن از دست برود.

مثال ۲ [۴]. فرض کنید $P = \{P_N, N \geq 1\}$ که در آن

$$P_N(X=k) = \begin{cases} \frac{1}{N} & k = 1, 2, \ldots, N \\ 0 & \text{سایر نقاط} \end{cases}$$

خانواده P کامل است، زیرا رابطهٔ

$$E_N g(X) = \frac{1}{N}\sum_{k=0}^{N} g(k) = 0, \qquad \forall N \geq 1$$

فقط وقتی برقرار است که $g(k) = 0, k = 0, 1, 2, 3$.

اکنون مقدار دلخواهی مانند n_0 را از فضای پارامتر حذف می‌کنیم. در این حالت خانواده $P_0 = \{P_N, N \geq 1, N \neq n_0\}$

کامل نیست. کافی است یک تابع غیر متحد با صفر داشته باشیم که $E_N g(X) = 0، \quad \forall N \in P_0$. تابع $g(k)$ را به صورت زیر در نظر بگیرید.

$$g(k) = \begin{cases} 0 & k = 1, 2, \ldots, n_0 - 1, n_0 + 2, \ldots \\ c & k = n_0 \\ -c & k = n_0 + 1 \end{cases}$$

که در آن c یک عدد ثابت مخالف صفر باشد. به آسانی دیده می‌شود که

$$E_N g(X) = 0, \quad \forall N \geq 1, N \neq n_0,$$

یعنی $g(k)$ یک برآوردگر نااریب صفر برای خانواده P_0 است و لذا P_0 کامل نیست.

۳ دو مقایسه بین کامل بودن، بسندگی و نااریبی

الف) از تعریف کامل بودن و دو مثال بالا روشن است که چنانچه خانواده P کامل باشد با بزرگ شدن فضای پارامتر، کامل بودن خانواده حفظ می‌شود؛ ولی با کوچک شدن آن ممکن است این خاصیت از دست برود (برای جزئیات، [۱] را ببینید). از این نظر کامل بودن مخالف نااریبی و مخالف بسندگی عمل می‌کند. توضیح آنکه اگر آماره T برای تمام $\theta \in \Theta$ نااریب باشد با محدود کردن Θ، نااریبی T حفظ می‌شود در حالی که با گسترش Θ ممکن است نااریبی T از دست برود. همچنین اگر آماره T برای خانواده توزیع‌های P بسنده باشد آنگاه برای هر زیر مجموعه از P نیز بسنده است اما ممکن است با بزرگتر کردن خانواده P، آماره T دیگر بسنده نباشد.

ب) ویژگی کامل بودن از لحاظ پایائی نیز جالب توجه است. اگر آماره T کامل باشد و $S = h(T)$، آنگاه S نیز کامل است. زیرا فرض کنید $E_\theta g(S) = 0، \quad \forall \theta \in \Theta$، این یعنی $E_\theta g[h(T)] = 0، \quad \forall \theta \in \Theta$. اکنون از کامل بودن T نتیجه می‌شود که $P_\theta[g[h(T)] = 0] = 1، \quad \forall \theta \in \Theta$ و یا $P_\theta[g(S) = 0] = 1 \quad ،\forall \theta \in \Theta$. دربارهٔ بسندگی، ویژگی پایائی در حالت خاص یک به یک بودن تابع h برقرار است. بدین معنی که اگر T آماره بسنده‌ای برای θ باشد، آنگاه هر تابع یک به یک از T نیز برای θ بسنده است. بدیهی است که دربارهٔ نااریبی، این گونه پایائی وجود ندارد. البته این نکته هست که اگر T برآوردگر نااریب θ باشد و $g(T)$ تابع خطی از T باشد، آنگاه $g(T)$ برآوردگر نااریب $g(\theta)$ است.

مراجع

[1] D. A. Fraser (1965), *Nonparametric Methods in Statistics*, John Wiley.

[2] E. L. Lehmann (1991), *Testing Statistical Hypotheses*, John Wiley.

[3] E. L. Lehmann and Casella (1997), *Theory of Point Estimation*, Springer.

[4] V. K. Rohatgi (1976), *An Introduction to Probability Theory and Mathematical Statistics*, John Wiley.

[5] J. Shao (1999), *Mathematical Statistics*, Springer.