آمار ریاضی ۱
فصل دوم: آماره بسنده کامل

استاد: دکتر اکبر اصغرزاده

خانواده چگالی کامل (خانواده توزیع‌های کامل)

اگر X دارای تابع چگالی احتمال $f(x,\theta)$ که $\theta \in \Theta$، در اینصورت خانواده چگالی‌های (توزیع‌های) X یعنی $\{f(x,\theta) : \theta \in \Theta\}$ را کامل گویند هرگاه برای هرتابع $g(X)$ و هر $\theta \in \Theta$:

$$E_\theta[g(X)] = ۰ \;\Rightarrow\; P_\theta[g(X) = ۰] = ۱.$$

مثال ۱. اگر $X \sim b(۱,\ p)$ که $۰ < p < ۱$. نشان دهید که خانواده‌ی توزیع‌های X یعنی خانواده $\{b(۱,p) : ۰ < p < ۱\}$ یک خانواده کامل است.

حل.

$$f(x,p) = p^x\,(۱-p)^{۱-x},\quad x = ۰, ۱$$
$$f(۰,p) = ۱ - p,\quad f(۱,p) = p.$$

برای هر تابع دلخواه $g(x)$ و هر $۰ < p < ۱$ داریم:

$$E_p[g(X)] = ۰ \Rightarrow \sum_{x=۰}^{۱} g(x)\,f(x,p) = ۰$$

$$\Rightarrow g(۰)(۱-p) + g(۱)\,p = ۰ \;\Rightarrow\; g(۰) + \Big[g(۱) - g(۰)\Big]p = ۰.$$

$$\Rightarrow\ g(0)=0,\quad g(1)-g(0)=0 \Rightarrow\ g(1)=0,$$
$$\Rightarrow\ g(x)=0,\quad \forall\, x=0,1$$

یا

$$P[g(X)=0]=1.$$

لذا خانواده X کامل است.

آماره کامل

یک آماره T را کامل گویند هرگاه خانواده چگالی‌های T یعنی خانواده

$$\{f(t,\theta):\ \ \theta\in\Theta\},$$

کامل باشد. این بدان معنی است که برای هر تابع $g(T)$ و هر $\theta\in\Theta$:

$$E_\theta[g(T)]=0\ \Rightarrow\ P[g(T)=0]=1.$$

مثال ۲. اگر $X_1, \ldots, X_n \overset{iid}{\sim} b(1,\ p)$ که $0 < p < 1$. نشان دهید که آماره $T = \sum_{i=1}^{n} X_i$ یک آماره کامل است.

حل. می‌دانیم که $T = \sum_{i=1}^{n} X_i \sim b(n,p)$.

$$f(t,p) = \binom{n}{t} p^t (1-p)^{n-t}, \quad t = 0, \ldots, n.$$ تابع چگالی T

برای هر تابع $g(T)$ و هر $0 < p < 1$،

$$E_p[g(T)] = 0 \Rightarrow P[g(T) = 0] \overset{?}{=} 1$$

داریم:

$$E_p[g(T)] = \sum_{t=0}^{n} g(t)\, f(t,p) = \sum_{t=0}^{n} g(t) \binom{n}{t} p^t (1-p)^{n-t}$$

$$= (1-p)^n \sum_{t=0}^{n} g(t) \binom{n}{t} \left(\frac{p}{1-p}\right)^t.$$

اگر $E_p[g(T)] = ۰$، نتیجه می‌شود:

$$
\begin{aligned}
&(۱-p)^n \sum_{t=۰}^{n} g(t) \binom{n}{t} \left(\frac{p}{۱-p}\right)^t = ۰ \\
&\Rightarrow \quad \sum_{t=۰}^{n} g(t) \binom{n}{t} \left(\frac{p}{۱-p}\right)^t = ۰ \\
&\Rightarrow \quad \sum_{t=۰}^{n} h(t)\,\gamma^t = ۰, \quad \left(h(t) = g(t)\binom{n}{t},\ \gamma = \frac{p}{۱-p}\right) \\
&\Rightarrow \quad h(۰) + h(۱)\gamma + \cdots + h(n)\,\gamma^n = ۰.
\end{aligned}
$$

چون $\gamma \neq ۰$ پس داریم:

$$
\begin{aligned}
&\Rightarrow \ h(۰) = h(۱) = \cdots = h(n) = ۰ \Rightarrow \ h(t) = ۰, \ \ \forall\,(t = ۰, \ldots, n) \\
&\Rightarrow \ \binom{n}{t} g(t) = ۰ \Rightarrow \ g(t) = ۰, \ \ \forall\, t = ۰, \ldots, n.
\end{aligned}
$$

لذا آماره T کامل است. همچنین خانواده توزیع‌های T یعنی $\{b(n,p) : ۰ < p < ۱\}$ نیز کامل است.

مثال ۳. فرض کنید $X \sim b(۱, p),\ p \in \Theta = \{\frac{۱}{۴},\ \frac{۳}{۴}\}$. نشان دهید که X کامل نیست.

حل. باید نشان دهیم که تابعی از X وجود دارد که امید ریاضی آن برای هر p صفر است ولی خود آن تابع صفر نیست.

$$E_p[g(X)] = ۰,\ \forall p \in \Theta \Rightarrow\ P[g(X) = ۰] \neq ۱.$$

$$E_p[g(X)] = ۰,\ \forall p \in \Theta \Rightarrow \sum_{x=۰}^{۱} g(x)\, f(x, p) = ۰$$

$$\sum_{x=۰}^{۱} g(x)\, p^x\, (۱ - p)^{۱-x} = ۰,\quad \forall p \in \Theta$$

$$\Rightarrow \begin{cases} \sum_{x=۰}^{۱} g(x)\, (\frac{۱}{۴})^x\, (۱ - \frac{۱}{۴})^{۱-x} = ۰, & p = \frac{۱}{۴} \\ \sum_{x=۰}^{۱} g(x)\, (\frac{۳}{۴})^x\, (۱ - \frac{۳}{۴})^{۱-x} = ۰, & p = \frac{۳}{۴} \end{cases}$$

$$\Rightarrow \begin{cases} \frac{۳}{۴} g(۰) + \frac{۱}{۴}\, g(۱) = ۰ \\ \frac{۱}{۴} g(۰) + \frac{۳}{۴}\, g(۱) = ۰ \end{cases}$$

$$\begin{cases} ۳\, g(\circ) + g(۱) = \circ, & (۱) \\ g(\circ) + ۳\, g(۱) = \circ, & (۲) \end{cases}$$

$$\xrightarrow{(۱)-(۲)} \quad ۲\, g(\circ) - ۲\, g(۱) = \circ \Rightarrow \; g(\circ) = g(۱).$$

قرار دهید $g(\circ) = g(۱) = a$. با انتخاب $a \neq \circ$، داریم $g(\circ) \neq \circ$، $g(۱) \neq \circ$. به عبارتی

$$g(x) \neq \circ, \quad \forall x = \circ, ۱.$$

لذا X یا خانواده توزیع‌های تولیدشده توسط X یعنی خانواده $\{b(۱, p) : p \in \{\frac{۱}{۴}, \frac{۳}{۴}\}\}$ کامل نیست.

مثال ۴. اگر $X \sim N(\theta,\ ۱)$، $\theta \in \mathbb{R}$، نشان دهید که X کامل است.

حل. برای هر تابع $g(X)$ و هر $\theta \in \mathbb{R}$:

$$E_\theta[g(X)] = ۰ \ \Rightarrow\ P[g(X) = ۰] \overset{?}{=} ۱$$

$$\begin{aligned}
E_\theta[g(X)] = \int g(x)\, f(x)\, dx &= \int_{-\infty}^{\infty} g(x)\, \frac{۱}{\sqrt{۲\pi}} e^{-\frac{(x-\theta)^۲}{۲}}\, dx \\
&= \frac{۱}{\sqrt{۲\pi}} \int_{-\infty}^{\infty} g(x)\, e^{-\frac{x^۲}{۲} + \theta\, x - \frac{\theta^۲}{۲}}\, dx \\
&= \frac{e^{\frac{\theta^۲}{۲}}}{\sqrt{۲\pi}} \int_{-\infty}^{\infty} g(x)\, e^{-\frac{x^۲}{۲}}\, e^{\theta\, x}\, dx,
\end{aligned}$$

اگر $E_\theta[g(X)] = ۰ \ \Rightarrow\ \int_{-\infty}^{\infty} g(x)\, e^{-\frac{x^۲}{۲}}\, e^{\theta\, x}\, dx = ۰$

$$* \ \Rightarrow\ \int_{-\infty}^{\infty} h(x)\, e^{\theta\, x}\, dx = ۰ \quad (h(x) = g(x)\, e^{-\frac{x^۲}{۲}}),$$

بنا بر خاصیت یکتایی تبدیل لاپلاس داریم:

$$
\begin{aligned}
& h(x) = 0, \quad \forall x \in \mathbb{R}, \\
& \Rightarrow \; g(x) = 0, \quad \forall x \in \mathbb{R},
\end{aligned}
$$

به عبارتی $P[g(X) = 0] = 1$. و لذا X کامل است.

تبدیل لاپلاس:

$$L[f(x)] = \int_{-\infty}^{\infty} e^{-sx} f(x)\, dx.$$

خواص مهم:

۱. $L(0) = 0$

۲. تبدیل لاپلاس منحصربفرد است. یعنی

$$\text{اگر}\; L[f_1(x)] = L[f_2(x)] \;\Rightarrow\; f_1(x) = f_2(x),\; \forall x,$$

$$* : \int_{-\infty}^{\infty} h(x)\, e^{-\theta x}\, dx = 0 \;\Rightarrow\; L[h(x)] = 0 \Rightarrow\; h(x) = 0,\; \forall x.$$

مثال ۵. اگر $X \sim N(0,\theta)$ که $\theta > 0$. نشان دهید X کامل نیست.

حل. داریم $E(X) = 0$ ولی $P(X = 0) = 0 \neq 1$. لذا X کامل نیست.

مثال ۶. اگر $X_1, \dots, X_n \overset{iid}{\sim} U(0,\theta)$، که $\theta > 0$. نشان دهید که $T = X_{(n)}$ یک آماره بسنده کامل است.

حل. داریم:

$$f(x,\theta) = \frac{1}{\theta}, \quad 0 < x < \theta.$$

می‌دانیم که $T = X_{(n)}$ یک آماره بسنده (بسنده مینیمال) برای θ است. برای هر تابع $g(T)$ و هر $\theta > 0$:

$$E_\theta[g(T)] = 0 \;\Rightarrow\; P_\theta[g(T) = 0] \overset{?}{=} 0.$$

داریم:

$$E_\theta[g(T)] = \int g(t)\, f_T(t)\, dt.$$

برای محاسبه تابع چگالی $T = X_{(n)}$ با استفاده از فرمول داریم:

$$f_T(t) = n\, f_{X_1}(t)\,[F_{X_1}(t)]^{n-1} = n\,\frac{1}{\theta}\,(\frac{t}{\theta})^{n-1} = \frac{n\,t^{n-1}}{\theta^n}, \quad 0 < t < \theta.$$

بنابراین اگر $E_\theta[g(T)] = 0$، نتیجه می‌شود:

$$\int_0^\theta g(t)\,f_T(t)\,dt = 0$$

$$\Rightarrow \quad \int_0^\theta g(t)\,\frac{n\,t^{n-1}}{\theta^n}\,dt = 0 \Rightarrow \int_0^\theta g(t)\,t^{n-1}\,dt = 0 \quad *$$

با مشتق‌گیری از طرفین رابطه * نسبت به θ داریم:

$$g(\theta)\,\theta^{n-1} = 0, \quad \Rightarrow\; g(\theta) = 0,\ \forall\theta > 0,$$

لذا $g(t) = 0,\ \forall t > 0$. بنابراین T یک آماره کامل است.

قضیه بهادر

هر آماره بسنده کامل، بسنده مینیمال است.

تذکر. عکس قضیه بهادر درست نیست. بدین معنی که یک آماره بسنده مینیمال لزوماً کامل نیست.

مثال ۷. فرض کنید $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\theta,\ \theta)$. می‌دانیم که آماره $(\sum_{i=1}^n X_i,\ \sum_{i=1}^n X_i^2)$ (یا معادل آن‌ها $((\bar{X},\ \bar{X^2}),\ (\bar{X},\ S^2)$) آماره بسنده مینیمال است ولی این آماره کامل نیست. یعنی می‌توان تابعی از $(\sum_{i=1}^n X_i,\ \sum_{i=1}^n X_i^2)$ پیدا کرد که امید ریاضی آن صفر شود درحالیکه خود آن تابع صفر نیست. داریم:

$$E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n \theta = n\theta,$$

$$E(\sum_{i=1}^n X_i^2) = \sum_{i=1}^n E(X_i^2) = \sum_{i=1}^n \Big(Var(X_i) + E^2(X_i)\Big) = \sum_{i=1}^n (\theta^2 + \theta^2) = 2n\theta^2,$$

$$E(\sum_{i=1}^n X_i)^2 = Var(\sum_{i=1}^n X_i) + E^2(\sum_{i=1}^n X_i) = n\theta^2 + (n\theta)^2 = (n^2 + n)\theta^2.$$

$$\Rightarrow \begin{cases} E\left(\frac{\sum_{i=1}^{n} X_i^2}{2n}\right) = \theta^2 \\ E\left[\frac{(\sum_{i=1}^{n} X_i)^2}{n^2+n}\right] = \theta^2 \end{cases}$$

$$\Rightarrow E\left[\frac{\sum_{i=1}^{n} X_i^2}{2n} - \frac{(\sum_{i=1}^{n} X_i)^2}{n^2+n}\right] = 0.$$

لذا تابعی از $(\sum_{i=1}^{n} X_i, \ \sum_{i=1}^{n} X_i^2)$ پیدا کردیم که امید ریاضی آن صفر است ولی خودش با احتمال یک صفر نمی‌شود. لذا $(\sum_{i=1}^{n} X_i, \ \sum_{i=1}^{n} X_i^2)$ کامل نیست.

آماره بسنده کامل در خانواده نمایی تک پارامتری

$$X_1, \ldots, X_n \overset{iid}{\sim} f(x, \theta), \quad \theta \in \mathbb{R}$$
$$f(x, \theta) = a(\theta)\, b(x)\, e^{c(\theta) d(x)}.$$

در خانواده نمایی تک پارامتری آماره $S = \sum_{i=1}^{n} d(X_i)$ آماره بسنده کامل می‌باشد.

آماره بسنده کامل در خانواده نمایی k پارامتری

$$X_1, \ldots, X_n \overset{iid}{\sim} f(x, \theta), \quad \theta \in \Theta \subseteq \mathbb{R}^k$$
$$f(x, \theta) = a(\underline{\theta})\, b(x)\, e^{\sum_{i=1}^{k} c_i(\theta) d_i(x)}, \quad \underline{\theta} = (\theta_1, \ldots, \theta_k).$$

در خانواده نمایی k پارامتری، آماره بسنده مینیمال به‌صورت زیر خواهد بود.

$$S = \left(\sum_{i=1}^{n} d_1(X_i), \ldots, \sum_{i=1}^{n} d_k(X_i)\right).$$

اگر در خانواده نمایی k پارامتری، بعد آماره با بعد پارامتر یکسان باشد، آنگاه آماره بسنده مینیمال بسنده کامل نیز خواهد بود (در غیر اینصورت آماره بسنده مینیمال بسنده کامل نیست).

مثال ۸. فرض کنید $X_1,\dots,X_n \overset{iid}{\sim} N(\theta,1)$. خانواده فوق به خانواده نمایی تک پارامتری تعلق دارد. لذا آماره بسنده مینیمال و کامل به‌صورت $T=\sum_{i=1}^n d(X_i)$ یعنی $T=\sum_{i=1}^n X_i$ می‌باشد.

مثال ۹. فرض کنید $X_1,\dots,X_n \overset{iid}{\sim} N(\mu,\sigma^2)$. می‌دانیم توزیع نرمال به خانواده نمایی دوپارامتری تعلق دارد. و در آن آماره بسنده مینیمال به‌صورت $S=(\sum_{i=1}^n X_i,\ \sum_{i=1}^n X_i^2)$ می‌باشد. از آنجایی‌که بعد آماره بسنده مینیمال و بعد پارامتر برابر و مقدار ۲ می‌باشد، لذا $(\sum_{i=1}^n X_i,\ \sum_{i=1}^n X_i^2)$ بسنده کامل نیز می‌باشد.

مثال ۱۰. فرض کنید $X_1,\dots,X_n \overset{iid}{\sim} N(\theta,\theta^2)$. در آن آماره بسنده مینیمال به‌صورت $S=(\sum_{i=1}^n X_i,\ \sum_{i=1}^n X_i^2)$ می‌باشد. از آنجایی‌که بعد آماره بسنده و بعد پارامتر یکسان نیستند، یعنی

$$\text{بعد پارامتر} = 1, \qquad \text{بعد آماره بسنده مینیمال} = 2,$$

لذا $S=(\sum_{i=1}^n X_i,\ \sum_{i=1}^n X_i^2)$ کامل نیست.

۱. فرض کنید $X_1, \ldots, X_n \overset{iid}{\sim} Exp(\theta)$. آماره بسنده کامل پارامتر θ را بیابید.

۲. فرض کنید $X_1, \ldots, X_n \overset{iid}{\sim} pois(\lambda)$. آماره بسنده کامل پارامتر λ را بیابید.

۳. فرض کنید $X_1, \ldots, X_n \overset{iid}{\sim} Beta(\theta, 1)$ با تابع چگالی زیر باشد.

$$f(x, \theta) = \theta\, x^{\theta-1}, \quad 0 < x < 1, \quad \theta > 0.$$

آماره بسنده کامل پارامتر θ را بیابید.

۴. فرض کنید X یک متغیر تصادفی با تابع چگالی زیر باشد.

$$f(x|\theta) = \left(\frac{\theta}{2}\right)^{|x|} (1-\theta)^{1-|x|}, \quad x = -1, 0, 1, \quad 0 \leq \theta \leq 1.$$

الف) آیا X یک آماره بسنده کامل است؟ چرا؟

ب) آیا $|X|$ یک آماره بسنده کامل است؟ چرا؟

۵. فرض کنید $X_1, \ldots, X_n \overset{iid}{\sim} N(0,\ \theta)$، $\theta > 0$.

الف) نشان دهید آماره $\bar{X}$ کامل نیست (همچنین آماره $\sum_{i=1}^{n} X_i$).
ب) نشان دهید $\bar{X}^2$ کامل است (همچنین آماره $\sum_{i=1}^{n} X_i^2$).

۶. فرض کنید $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta,\ \theta)$، $\theta > 0$. نشان دهید که $(\bar{X},\ S^2)$ کامل نیست.

۷. فرض کنید $X_1, \ldots, X_n \overset{iid}{\sim} U(\theta,\ \theta + 1)$.

الف) آماره بسنده مینیمال θ را بیابید.
ب) آیا آماره بسنده مینیمال به‌دست آمده کامل است؟ چرا؟