آماره‌های کمکی و نقش آنها در استنباط آماری

عبدالرضا سیاره*

چکیده

در این مقاله به بررسی نقش آماره‌های کمکی در حوزه‌هایی از استنباط آماری می‌پردازیم. نخست تعریفهایی از آماره‌های کمکی را ارائه و ملاحظه می‌کنیم که وقتی برآوردگر درستنمایی ماکسیمم یک پارامتر، بسنده نباشد، این برآوردگر به همراه آمارهٔ کمکی مناسب، از لحاظ مقدار اطلاع در مورد پارامتر تحت مطالعه، معادل با تمام نمونه است. سپس آمارهٔ کمکی مناسب برای استنباط را معرفی و روش کاکس در این مورد را تشریح می‌کنیم. به دنبال آن، قضیهٔ باسو و نقش آماره‌های کمکی در این قضیه بررسی شده و با تعریف آمارهٔ مرتبهٔ اول، حالت خاصی از این قضیه و عکس آن بیان شده است.

۱ مقدمه

استنباط آماری به دو روش معمول صورت می‌گیرد. یک روش این است که تمام مقادیر فضای نمونه‌ای را در نظر بگیریم. روش دیگر، فقط به مقدار مشاهده شدهٔ متغیری تصادفی توجه دارد. موضوع استنباط شرطی بر اساس آمارهٔ کمکی برای ایجاد رابطهٔ بین این دو روش پایه‌ریزی شده است. معمولاً کمکی بودن آماره، در مقایسه با سایر خواص آماری یک آماره کمتر مورد توجه قرار می‌گیرد در حالی که کمکی بودن، در بسیاری از موارد می‌تواند به عنوان یک خاصیت خوب در بررسی استقلال آماره‌ها از یکدیگر به کار گرفته شود. در مبحث برآورد نقطه‌ای می‌خواهیم به کمک روشهای متداول، آماره‌ای را به دست بیاوریم که یک برآوردگر برای پارامتر جامعه باشد. یکی از این روشها، روش برآورد درستنمایی ماکسیمم است. این نوع برآوردگر باید دارای خواصی باشد تا آن را برآوردگری «خوب» بنامیم. «خوب» بودن یک برآوردگر مفهومی نسبی است، و در بسیاری از موارد، معیاری برای خوبی یک برآوردگر، بسندگی آن است، اما قضیه‌ای وجود ندارد که بگوید هر برآوردگر درستنمایی ماکسیمم، بسنده نیز هست. بنابراین اگر چنین برآوردگری را به منظور استنباط به کار ببریم، گاهی اوقات مقداری از اطلاع موجود در نمونه را از دست خواهیم داد. لذا اگر روش برآورد درستنمایی ماکسیمم را به کار بگیریم و برآوردگر حاصل بسنده نباشد باید در جستجوی اطلاع از دست رفته باشیم.

اساس پی‌جویی اطلاعات از دست رفته توسط فیشر پی‌ریزی شد. به عقیدهٔ وی در چنین شرایطی، استنباط باید به شرط آمارهٔ کمکی انجام شود تا اطلاعات از دست رفته مجدداً بازیابی شوند. سپس، باسو با مطالعهٔ سیگما میدانهای کمکی و از نقطه نظر افرازی که آمارهٔ کمکی روی فضای نمونه‌ای ایجاد می‌کند، این گونه آماره‌ها را مورد مطالعه قرار داد و نشان داد که اگر چه آماره‌های کمکی به تنهایی آگاهی‌بخش نیستند، اما می‌توان خانواده‌یی از آماره‌های کمکی را یافت که از نظر میزان اطلاع، معادل با تمام نمونهٔ اصلی باشند. باسو این گونه خانواده‌ها را از نظر مقدار اطلاعی که در خود دارند مرتب کرد و استنباط شرطی به شرط آمارهٔ کمکی را مورد بررسی قرار داد. اگر چه باسو نیز به استنباط به شرط آمارهٔ کمکی اعتقاد داشت ولی با ارائهٔ

*) عبدالرضا سیاره، دانشگاه رازی کرمانشاه

مثالی نشان داد، که برای یک مسأله ممکن است چندین آمارهٔ کمکی وجود داشته باشد به طوری که استنباط به شرط هر یک از آنها به نتایج متفاوتی بینجامد.

این مشکل توسط کاکس برطرف شد. وی روشی را برای انتخاب آمارهٔ کمکی از میان سایر آماره‌های کمکی به منظور استنباط شرطی پیشنهاد کرد. در مقالهٔ حاضر سعی شده که این ترتیب تاریخی رعایت شود. در بخش اول مقاله، آمارهٔ کمکی تعریف شده و چند مثال در این مورد ارائه می‌شود. در بخش دوم، نقش آمارهٔ کمکی به همراه برآوردگر درستنمایی ماکسیمم مورد بررسی قرار گرفته است و در بخش سوم به عنوان کاربرد دیگری از کمکی بودن، قضیهٔ باسو را مورد بررسی قرار داده و با تعریف کمکی بودن در حالت خاص، نوعی از عکس قضیهٔ باسو را بیان می‌کنیم.

۲ آماره‌های کمکی

آماره‌های کمکی را اولین بار فیشر در سال ۱۹۲۵ معرفی کرد اما تعریف او دقت چندانی نداشت. تعریف وی به این شرح است. «آمارهٔ کمکی آماره‌ای است که توزیع آن مستقل از پارامتر تحت مطالعه یعنی θ باشد.» در سال ۱۹۵۶ تعریف دقیقتری از آمارهٔ کمکی توسط فیشر به این شرح ارائه شد. «اگر $T(X)$ برآوردگر درستنمایی ماکسیمم θ و $A(X)$ آماره‌ای باشد که توزیع آن فارغ از پارامتر θ بوده و زوج $(T(X), A(X))$ آمارهٔ بسندهٔ مینیمال θ باشد آنگاه $A(X)$ آماره‌ای کمکی برای θ است». همچنین باسو در سال ۱۹۵۵ تعریفی از آمارهٔ کمکی ارائه کرده است. وی می‌نویسد «یک آماره در صورتی می‌تواند ما را دربارهٔ پارامتر مجهول θ مطلع سازد که توزیعش به θ بستگی داشته باشد. آماره‌ای که دارای این ویژگی نباشد، آمارهٔ کمکی نامیده می‌شود.»

باک لند و کندال نیز در سال ۱۹۵۷ تعریفی از این آماره ارائه کرده‌اند: «آمارهٔ کمکی، آماره‌ای است که همراه با $T(X)$، برآوردگر درستنمایی ماکسیمم پارامتر مورد نظر θ، باعث تقلیل هیچ میزانی از اطلاع نشود.» اون (۱۹۴۸) و کاکس و هینکلی (۱۹۷۴) در این مورد می‌نویسند «آمارهٔ $A(X)$ کمکی است هرگاه توزیع آن بستگی به پارامتر موردنظر θ نداشته باشد و علاوه بر این آمارهٔ $T(X)$ی وجود داشته باشد به قسمی که $(T(X), A(X))$ آمارهٔ بسنده مینیمال برای θ باشد.»

با توجه به تعاریف اخیر، آماره‌های کمکی به تنهایی شامل هیچ اطلاعی در مورد پارامتر مجهول یعنی θ نیستند. چنانچه آماره‌ای کمکی نباشد آن را آگاهی بخش می‌نامیم.

مثال ۱: فرض کنید X_1 و X_2 و X_3 یک نمونهٔ تصادفی مستقل سه تایی از توزیع $N(\mu, 1)$ با پارامتر مجهول μ باشد. دو آمارهٔ $T_1 = X_1 - 2X_2 + X_3$ و $T_2 = \bar{X}$ را در نظر می‌گیریم. در این صورت $T_1 \sim N(0, 6)$ و $T_2 \sim N(\mu, \frac{1}{3})$. بنابه تعریف T_1 آماره‌ای کمکی و T_2 آماره‌ای آگاهی‌بخش است، زیرا نسبت به تغییرات μ حساس است.

مثال ۲: فرض کنید X_1 و $X_2, \ldots, X_n$ همتوزیع و مستقل با توزیع یکنواخت روی بازهٔ $(\theta, \theta + 1)$، $\theta \in \mathbb{R}$، باشند. اگر $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ آماره‌های ترتیبی نمونهٔ n تایی باشد و تعریف کنیم $Y = X_{(n)} - X_{(1)}$، در این صورت:

$$f(x) = 1 \qquad \theta < x < \theta + 1$$
$$= 0 \qquad \text{سایر جاها}$$

$$g(X_{(1)}, X_{(n)}) =$$
$$\begin{cases} n(n-1)(X_{(n)} - X_{(1)})^{n-2} & \theta < X_{(1)} < X_{(2)} < \theta + 1 \\ 0 & \text{سایر جاها} \end{cases}$$

با تبدیل $Y = X_{(n)} - X_{(1)}$ و $U = \frac{X_{(n)} + X_{(1)}}{2}$ خواهیم داشت

$$h(y; \theta) = n(n-1)y^{n-2}(1-y) \qquad 0 < y < 1$$
$$= 0 \qquad \text{سایر جاها}$$

پس $Y \sim \mathrm{Bet}(n-1, 2)$. بنابراین توزیع Y به پارامتر مجهول θ بستگی ندارد و $Y = X_{(n)} - X_{(1)}$ یک آمارهٔ کمکی است.

مثال ۳: فرض کنید $X_1, X_2, \ldots, X_n$ نمونه‌ای تصادفی از مدل مکانی $P_L(x - \theta)$ باشد. اگر A را به صورت

$$A = (X_{(n)} - X_{(1)}, X_{(n)} - X_{(2)}, \ldots, X_{(n)} - X_{(n-1)})$$

تعریف کنیم، که در آن $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$، آنگاه می‌توان نشان داد که توزیع A به θ بستگی ندارد. بنابراین A آماره‌ای کمکی برای θ است.

مثال ۴: فرض کنید $X_1, X_2, \ldots, X_n$ نمونه‌ای تصادفی به اندازه n از مدل مقیاسی $P(x; \theta) = \frac{1}{\theta} P_s(\frac{x}{\theta})$ برای $x > 0$ و $\theta > 0$ باشد. اگر A را به صورت

$$A = \left(\frac{X_{(n)}}{X_{(1)}}, \frac{X_{(n)}}{X_{(2)}}, \ldots, \frac{X_{(n)}}{X_{(n-1)}}\right)$$

تعریف کنیم که در آن $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$، به سادگی معلوم می‌شود که توزیع A به θ بستگی ندارد. بنابراین A آماره‌ای کمکی است.

۳ نقش آمارهٔ کمکی در کنار برآوردگر درستنمایی ماکسیمم

با توجه به تعریف فیشر (۱۹۵۶) از آمارهٔ کمکی در حالت یک پارامتری، اگر زوج (T, A) برای θ بسندهٔ مینیمال باشد، آمارهٔ کمکی A باید حاوی مقداری اطلاع در مورد θ باشد. در این صورت A را مکمل T نامیده و اطلاع موجود در A را اطلاع کمکی می‌نامیم و به دنبال معیاری برای اندازه‌گیری آن هستیم.

به عقیدهٔ فیشر یک معیار منطقی برای اندازه‌گیری اطلاع کمکی عبارت است از $\lambda(\theta) = I(\theta) - J(\theta)$ که در آن $I(\theta)$ و $J(\theta)$ به ترتیب میزان اطلاع موجود در تمام نمونهٔ X و میزان اطلاع موجود در آمارهٔ T است که به صورت زیر تعریف می‌شوند:

$$\begin{aligned} I(\theta) &= E\left(\frac{\partial}{\partial\theta}\log f(X|\theta)|\theta\right)^{2} \\ &= -E\left(\frac{\partial^{2}}{\partial\theta^{2}}\log f(X|\theta)|\theta\right). \\ J(\theta) &= E\left(\frac{\partial}{\partial\theta}\log g(T|\theta|\theta\right)^{2} \\ &= -E\left(\frac{\partial^{2}}{\partial\theta^{2}}\log g(T|\theta)|\theta\right). \end{aligned}$$

بنابراین $\lambda(\theta)$ در واقع میزان اطلاع از دست رفته در استفاده از آمارهٔ T به عنوان برآورد کنندهٔ θ است. در اینجا قضیهٔ اصلی این بخش را بدون اثبات بیان می‌کنیم.

قضیهٔ ۱: تحت شرایط نظم، موارد زیر برقرار است.

الف) برای تمام θها، $\lambda(\theta) > 0$.

ب) $\lambda(\theta) = 0$ اگر و تنها اگر T آماره‌ای بسنده برای θ باشد.

ج) اگر آمارهٔ T همراه با آمارهٔ A برای θ بسنده باشد، آنگاه اطلاع حاصل از زوج (T, A) برابر $I(\theta)$ است.

د) اگر A آماره‌ای کمکی و زوج (T, A) آماره‌ای بسنده برای θ باشد، آنگاه $I(\theta) = E(J(\theta)|A)$.

بنابراین اگر T برآوردگر درستنمایی ماکسیمم θ، بسنده نباشد، استفاده از آن موجب از دست رفتن اطلاعاتی به اندازهٔ $\lambda(\theta)$ می‌شود. اما اگر آمارهٔ کمکی Aیی وجود داشته باشد به طوری که زوج (T, A) برای θ بسنده باشد، بنابر قسمت (د) قضیهٔ اخیر، کافی است توزیع شرطی T به شرط آمارهٔ کمکی A را در نظر بگیریم تا هیچ اطلاعی از دست نرود.

مثال معروف زیر را که به مسألهٔ نیل معروف است در نظر بگیرید.

مثال ۵: فرض کنید $(X_1, Y_1), (X_2, Y_2), \ldots, (X_k, Y_k)$ نمونه‌ای تصادفی از جامعه‌ای با چگالی توأم زیر باشد:

$$f(x,y) = \begin{cases} e^{-(\theta x + \frac{y}{\theta})} & x > 0, y > 0, \theta > 0 \\ 0 & \text{سایر جاها} \end{cases}$$

به آسانی می‌توان نشان داد که $T = \sqrt{\sum_1^n Y_i / \sum_1^n X_i}$ برآوردگر درستنمایی ماکسیمم θ بوده امّا برای θ بسنده نیست. از طرفی زوج (T, A) که در آن $A = \sqrt{(\sum_1^n X_i)(\sum_1^n Y_i)}$ برای θ بسنده است. با انجام محاسبهٔ مختصری می‌توان نشان داد که A آماره‌ای کمکی است، زیرا تابع چگالی احتمال T و A عبارت است از

$$f(t,a) = \begin{cases} \frac{2}{(\Gamma(n))^{2}} a^{2n-1} \frac{1}{t} e^{-a(\frac{\theta}{t}+\frac{t}{\theta})} & t > 0 a > 0 \theta > 0 \\ 0 & \text{سایر جاها} \end{cases}$$

تابع چگالی A به صورت زیر محاسبه می‌شود و به θ بستگی ندارد.

$$\begin{aligned} f(a) &= \int_0^\infty f(t,a;\theta)dt \\ &= \frac{2}{(\Gamma(n))^{2}} a^{2n-1} \int_0^\infty \frac{1}{t} e^{-a(\frac{\theta}{t}+\frac{t}{\theta})} dt. \end{aligned}$$

برای محاسبهٔ انتگرال اخیر به طریق زیر عمل می‌کنیم

$$\begin{aligned} (\frac{\theta}{t} + \frac{t}{\theta})^n &= \sum_{k=0}^{n} \binom{n}{k} (\frac{t}{\theta})^{-n+2k} \\ e^{-a(\frac{\theta}{t}+\frac{t}{\theta})} &= \sum_{n=0}^{\infty} \frac{(-1)^n a^n}{n!} \sum_{k=0}^{n} \binom{n}{k} (\frac{t}{\theta})^{-n+2k}. \end{aligned}$$

بنابراین

$$\begin{aligned} f(a) = \frac{2}{(\Gamma(n))^{2}} a^{2n-1} \int_0^\infty \frac{1}{t} \sum_{n=0}^{\infty} \\ \frac{(-1)^n a^n}{n!} \sum_{k=0}^{n} \binom{n}{k} (\frac{t}{\theta})^{-n+2k} dt. \end{aligned}$$

از همگرایی سریها و انتگرال استفاده کرده با محاسباتی مختصر خواهیم داشت:

$$\begin{aligned} f(a) = \frac{2}{(\Gamma(n))^{2}} a^{2n-1} 2 \sum_{n=0}^{\infty} \\ \frac{(-1)^n a^n}{(n!)^2} = \frac{4}{(\Gamma(n))^{2}} a^{2n-1} k_0(2a), \end{aligned}$$

که در آن $k_0(2a)$ تابع بسل $k_0(x)$ یعنی تابع $k_0(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(n!)^2 2^{2n}}$

در $x = 2a$ است. بنابراین توزیع A مستقل از θ بوده و A آماره‌ای کمکی برای θ است.

حال اطلاع موجود در نمونه و اطلاع موجود در آمارهٔ T را محاسبه می‌کنیم. با اندکی محاسبه معلوم می‌شود که $I(\theta) = \frac{2n}{\theta^2}$.

برای محاسبهٔ $J(\theta)$، لازم است که چگالی T را محاسبه کنیم:

$$
\begin{aligned}
f(t) &= \int_0^\infty f(t, a; \theta) da \\
&= \frac{2\Gamma(2n)}{(\Gamma(n))^2} \frac{1}{t} \left(\frac{\theta}{t} + \frac{t}{\theta}\right)^{-2n} \quad t > 0, \theta > 0
\end{aligned}
$$

مقدار اطلاع موجود در T را با محاسبهٔ $E(\frac{\partial \log f(T)}{\partial \theta})^2 = J(\theta)$ به دست می‌آوریم:

$$
\begin{aligned}
\left(\frac{\partial \log f(t)}{\partial \theta}\right)^2 &= \frac{4n^2}{\theta^2} . \frac{(\frac{t}{\theta} - \frac{\theta}{t})^2}{(\frac{t}{\theta} + \frac{\theta}{t})^2} \\
& \frac{4n^2}{\theta^2} . \frac{(\frac{t}{\theta} + \frac{\theta}{t})^2 - 4}{(\frac{t}{\theta} + \frac{\theta}{t})^2} \\
&= \frac{4n^2}{\theta^2} \left(1 - 4(\frac{t}{\theta} + \frac{\theta}{t})^{-2}\right)
\end{aligned}
$$

از طرفی،

$$
\begin{aligned}
E(\frac{T}{\theta} + \frac{\theta}{T})^{-2} &= \int_0^\infty \frac{2\Gamma(2n)}{(\Gamma(n))^2} (\frac{\theta}{t} + \frac{t}{\theta})^{-2(n+1)} \frac{1}{t} dt \\
&= \frac{2(\Gamma(n+1))^2 \Gamma(2n)}{2(\Gamma(n))^2 \Gamma(2n+2)} = \frac{n}{2(2n+1)}.
\end{aligned}
$$

بنابراین

$$
J(\theta) = \frac{4n^2}{\theta^2} (1 - 4 \frac{n}{2(2n+1)}) = \frac{2n}{\theta^2} . \frac{2n}{2n+1}
$$

و

$$
\lambda(\theta) = I(\theta) - J(\theta) = \frac{2n}{\theta^2} (\frac{1}{2n+1}).
$$

پس مقدار اطلاع گمشده کمتر از نصف مقدار اطلاع حاصل از یک جفت مشاهده است.

چگالی شرطی T به شرط $A = a$ عبارت است از

$$
f(t|a) = \frac{f(t,a)}{f(a)} = \frac{1}{2k_0(2a)} e^{-a(\frac{t}{\theta} + \frac{\theta}{t})} . \frac{1}{t},
$$

از طرفی،

$$
\begin{aligned}
J(\theta|a) &= -E\left(\frac{\partial^2}{\partial \theta^2} \log f(T|a, \theta) | \theta, a\right) \\
&= -E(\frac{\partial^2}{\partial \theta^2}(-\log T - \log 2k_0(2a) - a(\frac{T}{\theta} + \frac{\theta}{T}))) \\
&= \frac{2n}{\theta^2} . \frac{k_1(2a)}{k_0(2a)},
\end{aligned}
$$

$$
\begin{aligned}
k_1(\omega) &= \frac{\omega}{2} - \frac{\omega^3}{2^3 2!} + \frac{\omega^5}{2^5 2! 3!} - \cdots \\
&+ (-1)^k \frac{\omega^{2k+1}}{2^{2k+1} k! (k+1)!} + \cdots .
\end{aligned}
$$

بنابراین، داریم

$$
E(J(\theta|a)) = \frac{2n}{\theta^2} = I(\theta).
$$

پس استنباط بر اساس $f(t|a)$ منطقی‌تر به نظر می‌رسد، زیرا اطلاعی که بر اساس توزیع شرطی T به شرط آمارهٔ کمکی A محاسبه می‌شود تمام اطلاعات موجود در نمونه را به ما می‌دهد، و هیچ اطلاعی از دست نمی‌رود. این مطلب نشان می‌دهد که مقداری اطلاع در A موجود است.

جالب توجه است که اگر $f(t|a)$ را برای برآورد θ به کار بگیریم باز هم برآوردگر درستنمایی ماکسیمم θ مساوی T خواهد بود، با این تفاوت که زمانی که آمارهٔ کمکی را نیز در استنباط دخالت می‌دهیم هیچ اطلاعی از دست نمی‌رود. بنابراین اطلاعات از دست رفته از طریق آمارهٔ کمکی بازیافت شده و استنباط کامل می‌شود.

آمارهٔ کمکی برای پارامتر جامعه معمولاً یکتا نیست. بنابراین لازم است آمارهٔ کمکی را که می‌خواهیم در استنباط شرطی به کار ببریم، دقیقاً مشخص نماییم. بدیهی است برای دستیابی به چنین آماره‌ای، باید قادر به مرتب کردن آماره‌های کمکی براساس میزان اطلاع موجود در آنها باشیم. آماره‌های کمکی براساس میزان اطلاع به ترتیب صعودی عبارت‌اند از آمارهٔ کمکی مینیمال، آمارهٔ کمکی لامینال و آمارهٔ کمکی ماکسیمال. در قسمت بعد آمارهٔ کمکی ماکسیمال را تعریف کرده، نشان می‌دهیم در مسألهٔ خاص این آماره یکتا نیست.

۱.۳ سیگما میدانهای کمکی

مشکل اساسی که در استنباط شرطی به شرط آمارهٔ کمکی وجود دارد، یکتا نبودن آمارهٔ کمکی است. بنابراین ممکن است آماره‌های کمکی مختلف در میزان اطلاع کمکی متفاوت باشند. سؤالی که در اینجا مطرح می‌شود این است که کدام آمارهٔ کمکی را در استنباط شرطی به کار ببریم. برای این منظور لازم است ابتدا سیگما میدانهای کمکی را بررسی کنیم.

فضای احتمال $(\Omega, \mathcal{B}, P)$ را در نظر بگیرید. آمارهٔ T یک نگاشت از Ω به T تعریف می‌کند. هر نقطهٔ $t \in T$ زیرمجموعه‌ای از Ω را به صورت $E_t = \{x; T(x) = t\}$ تعریف کرده و خانوادهٔ $\{E_t\}$ یک افراز از Ω را به وجود می‌آورد. برعکس هر افراز از Ω توسط آماره‌ای مثل T تولید می‌شود.

مثال ۶: فرض کنید $\Omega = \{x_1, x_2, x_3\}$ و توزیع احتمال X به شکل زیر باشد:

$$
\begin{array}{c|ccc}
x & x_1 & x_2 & x_3 \\
\hline
\rho_X(x) & \frac{1}{4}-\theta & \frac{1}{2} & \frac{1}{4}+\theta
\end{array}
$$
$$-\frac{1}{4} < \theta < \frac{1}{4}$$

آماره‌های T_1 و T_2 را به صورت زیر تعریف می‌کنیم:

$$T_1 = \begin{cases} 1 & x = x_2 \\ 0 & x = x_1, x_3 \end{cases} \qquad T_2 = \begin{cases} 1 & x = x_1, x_2 \\ 0 & x = x_3 \end{cases}$$

مجموعه‌های افرازکننده Ω توسط T_1 عبارت‌اند از

$$E_0 = \{x_1, x_3\}, \qquad E_1 = \{x_2\}$$

و مجموعه‌های افرازکننده Ω توسط T_2 عبارت‌اند از

$$E'_0 = \{x_3\}, \qquad E'_1 = \{x_1, x_2\}$$

E_0 و E_1 و E'_0 و E'_1 هرکدام یک پیشامد هستند. احتمال وقوع E_0 و E_1 به θ بستگی ندارد در حالی که احتمال وقوع E'_0 و E'_1 به θ وابسته است. بدین جهت پیشامدهای E_0 و E_1 را پیشامدهای کمکی می‌نامند. در یک مسألهٔ مفروض، خانوادهٔ همهٔ پیشامدهای کمکی را با $\mathscr{A}$ نشان می‌دهیم. اگر $\mathscr{A}$ یک سیگما میدان نیز باشد، آن را سیگما میدان کمکی می‌نامیم. در این صورت آمارهٔ T کمکی است اگر سیگما میدان تولید شده توسط آن کمکی باشد.

مثال ۷: در مثال ۶، $\mathscr{A} = \{\{x_2\}, \{x_1, x_3\}, \phi, \Omega\}$ یک سیگما میدان کمکی است. پس T_1 آماره‌ای کمکی است.

سیگما میدان کمکی در یک مسأله یکتا نیست، زیرا اگر $\mathscr{A}$ داده شده باشد و $\mathscr{F}$ یک سیگما میدان باشد، به طوری که $\mathscr{F} \subseteq \mathscr{A}$، $\mathscr{F}$ نیز کمکی است. لذا سیگما میدانهای کمکی را می‌توان مرتب کرد. براین اساس سیگما میدان کمکی ماکسیمال را تعریف می‌کنیم:

سیگما میدان کمکی $\mathscr{M}$ را ماکسیمال گوییم اگر هیچ سیگما میدان کمکی دیگری مانند $\mathscr{M}^*$ موجود نباشد به طوری‌که $\mathscr{M} \subset \mathscr{M}^*$. همچنین می‌توان ثابت کرد، برای هر سیگما میدان کمکی $\mathscr{F}_0$، یک سیگما میدان کمکی ماکسیمال $\mathscr{M}$ وجود دارد به طوری که $\mathscr{F}_0 \subset \mathscr{M}$. سیگما میدان کمکی ماکسیمال آگاهی بخش‌تر از هر سیگما میدان کمکی دیگری است.

۲.۳ یکتا نبودن سیگما میدان کمکی ماکسیمال

اگر A و B دو پیشامد کمکی وابسته به Ω باشند و $A \cap B$ نیز کمکی باشد A و B را همدیس می‌گوییم و می‌نویسیم $A \sim B$. خانوادهٔ کلیهٔ مجموعه‌های کمکی B را که برای هر $A \in \mathscr{A}$ با A همدیس هستند Γ_0 می‌نامیم. دو عضو بدیهی Γ_0 عبارت‌اند از ϕ و Ω. زیرا این دو مجموعه با هر مجموعهٔ کمکی همدیس هستند. Γ_0 یک سیگما میدان است و اگر $\{\mathscr{M}_i\}_{i\in I}$ خانواده‌ای از همهٔ سیگما میدانهای کمکی ماکسیمال باشد. آنگاه $\bigcup_i \mathscr{M}_i = \mathscr{A}$. همچنین $\mathscr{A}$ یک سیگما میدان است اگر و تنها اگر $\mathscr{A} = \Gamma_0$. بدیهی است که اگر $\{\mathscr{M}_i\}$ شامل تنها یک سیگما میدان $\mathscr{M}_0$ باشد آنگاه $\Gamma_0 = \mathscr{M}_0 = \mathscr{A}$، یعنی تمام مجموعه‌های کمکی همدیس هستند. لذا اگر مجموعه‌های کمکی همدیس نباشند خانوادهٔ $\{\mathscr{M}_i\}$ حداقل دارای دو عضو است. بنابراین سیگما میدان کمکی ماکسیمال در حالت کلی یکتا نیست، در نتیجه آماره‌های کمکی وابسته نیز یکتا نخواهند بود.

مثال ۸: فرض کنید $\Omega = \{x_1, x_2, x_3, x_4, x_5\}$ و X دارای توزیع احتمال زیر باشد.

$$
\begin{array}{c|ccccc}
x & x_1 & x_2 & x_3 & x_4 & x_5 \\
\hline
P_X(x) & \frac{1}{2} & \theta & \theta & \frac{1}{4}-\theta & \frac{1}{4}-\theta
\end{array}
$$
$$0 < \theta < \frac{1}{4}$$

در این مثال:

$$\begin{aligned}\mathscr{A} = \{&\{x_1\}, \{x_2, x_4\}, \{x_2, x_5\}, \{x_3, x_4\}, \{x_3, x_5\},\\ &\{x_1, x_2, x_4\}, \{x_1, x_2, x_5\}, \{x_1, x_3, x_4\},\\ &\{x_1, x_3, x_5\}, \{x_2, x_3, x_4, x_5\}, \Omega, \phi\}\end{aligned}$$

$$\Gamma_0 = \{\{x_1\}, \{x_2, x_3, x_4, x_5\}, \Omega, \phi\}$$

و $\mathscr{A} \neq \Gamma_0$. بنابراین $\mathscr{A}$ یک سیگما میدان نیست. در این مثال می‌توان دو سیگما میدان کمکی ماکسیمال $\mathscr{M}_1$ و $\mathscr{M}_2$ را پیدا کرد. اگر $\mathscr{M}_1$ میدان تولید شده توسط $\{x_2, x_4\}$ و $\{x_1\}$ و $\mathscr{M}_2$ میدان تولید شده توسط $\{x_2, x_5\}$ و $\{x_1\}$ باشد، آنگاه

$$\begin{aligned}\mathscr{M}_1 = \{&\{x_1\}, \{x_2, x_4\}, \{x_2, x_3, x_4, x_5\}, \{x_1, x_3, x_5\},\\ &\{x_1, x_2, x_4\}, \{x_3, x_5\}, \Omega, \phi\}\end{aligned}$$

$$\begin{aligned}\mathscr{M}_2 = \{&\{x_1\}, \{x_2, x_5\}, \{x_1, x_2, x_5\}, \{x_2, x_3, x_4, x_5\},\\ &\{x_1, x_3, x_4\}, \{x_3, x_4\}, \Omega, \phi\}.\end{aligned}$$

اگر $\mathscr{F}$ یک سیگما میدان دلخواه وابسته به Ω باشد در این صورت $\mathscr{F} \subset \mathscr{M}_1$ یا $\mathscr{F} \subset \mathscr{M}_2$. $\mathscr{M}_1$ و $\mathscr{M}_2$ همدیس نیستند، زیرا اگر

پیشامدهای کمکی $\{x_{۲}, x_{۵}\}$ و $\{x_{۱}, x_{۳}, x_{۵}\}$ را در نظر بگیریم، داریم $\{x_{۲}, x_{۵}\} \cap \{x_{۱}, x_{۳}, x_{۵}\} = \{x_{۵}\}$ و $P_\theta(\{x_{۵}\}) = \frac{۱}{۴} - \theta$، لذا $\{x_{۵}\}$ کمکی نیست. سیگما میدانهای کمکی $\mathscr{M}_{۱}$ و $\mathscr{M}_{۲}$ ماکسیمال هستند، لذا آمارهٔ کمکی ماکسیمال یکتا نخواهد بود.

پس اگر قرار است استنباط شرطی، براساس آمارهٔ کمکی باشد، بهتر است آمارهٔ کمکی ماکسیمال به کار برده شود. زیرا از هر آماره کمکی دیگری آگاهی‌بخش‌تر است. ولی این گونه آماره‌ها همواره یکتا نیستند. از طرفی استفاده از آماره‌های کمکی ماکسیمال مختلف، منجر به نتایج متفاوتی می‌شود. به مثال بعد توجه کنید.

مثال ۹: تابع احتمال زیر را در نظر بگیرید:

x	۱	۲	۳	۴	۵	۶
$P_X(x)$	$\frac{۱-\theta}{۱۲}$	$\frac{۲-\theta}{۱۲}$	$\frac{۳-\theta}{۱۲}$	$\frac{۱+\theta}{۱۲}$	$\frac{۲+\theta}{۱۲}$	$\frac{۳+\theta}{۱۲}$

$$-۱ < \theta < ۱$$

برآورد درستنمایی ماکسیمم θ عبارت است از

$$T(X) = \begin{cases} ۱ & x = ۴, ۵, ۶ \\ ۰ & x = ۱, ۲, ۳ \end{cases}$$

$T(X)$ برای θ بسنده نیست. به عنوان مثال $T = ۰$ را در نظر بگیرید، با توجه به اینکه

$$P(X = ۱ | T = ۰) = \begin{cases} \frac{P(X=۱)}{P(T=۰)} & T = ۰ \\ ۰ & T \neq ۰ \end{cases}$$

$$= \begin{cases} \frac{۱-\theta}{۶-۲\theta} & T = ۰ \\ ۰ & T \neq ۰ \end{cases}$$

بستگی به θ دارد. لذا $T(X)$ برای θ بسنده نیست.

در اینجا شش آمارهٔ کمکی غیر معادل که مکمل T هستند وجود دارد. این شش آماره در جدول زیر آمده است.

x	۱	۲	۳	۴	۵	۶
$Y_{۱}(X)$	۰	۱	۲	۰	۱	۲
$Y_{۲}(X)$	۰	۱	۲	۰	۲	۱
$Y_{۳}(X)$	۰	۱	۲	۱	۰	۲
$Y_{۴}(X)$	۰	۱	۲	۲	۰	۱
$Y_{۵}(X)$	۰	۱	۲	۱	۲	۰
$Y_{۶}(X)$	۰	۱	۲	۲	۱	۰

به عنوان مثال افرازی که $Y_{۱}$ روی فضای نمونه‌ای ایجاد می‌کند عبارت است از:

$$\{(۱, ۴), (۲, ۵), (۳, ۶)\}.$$

و افراز حاصل از $Y_{۲}$ روی فضای نمونه‌ای عبارت است از:

$$\{(۱, ۴), (۲, ۶), (۳, ۵)\}.$$

$Y_{۱}, Y_{۲}, \ldots, Y_{۶}$ آماره‌های کمکی ماکسیمال هستند. زیرا هیچ سیگما میدان کمکی وجود ندارد که سیگما میدانهای حاصل از $Y_{۱}, \ldots, Y_{۶}$ زیرمجموعهٔ آن باشند. $T(X)$ مقادیر خود را از مجموعهٔ $\{۰, ۱\}$ و مثلاً $Y_{۱}(X)$ مقادیر خود را از مجموعهٔ $\{۰, ۱, ۲\}$ اختیار می‌کنند، لذا زوج $(T, Y_{۱})$ مقادیر $(۰, ۰), (۰, ۱), (۰, ۲), (۱, ۰), (۱, ۱), (۱, ۲)$ را می‌پذیرد. با محاسبهٔ $P((T, Y_{۱}) = (t, y_{۱}))$ برای مقادیر مختلف خواهیم داشت:

$$P\{(T, Y_{۱}) = (۰, ۰)\} = P(X = ۱)$$
$$P\{(T, Y_{۱}) = (۰, ۱)\} = P(X = ۲)$$
$$P\{(T, Y_{۱}) = (۰, ۲)\} = P(X = ۳)$$
$$P\{(T, Y_{۱}) = (۱, ۰)\} = P(X = ۴)$$
$$P\{(T, Y_{۱}) = (۱, ۱)\} = P(X = ۵)$$
$$P\{(T, Y_{۱}) = (۱, ۲)\} = P(X = ۶)$$

بنابراین زوج $(T, Y_{۱})$ معادل X و برای θ بسنده است. دوتاییهای (T, Y_i)، $i = ۲, \ldots, ۶$ نیز برای θ بسنده هستند. پس شش آمارهٔ کمکی ماکسیمال وجود دارد که می‌توان هر یک از آنها را برای استنباط شرطی به کار برد. این موضوع را برای یک مقدار مشخص X بررسی می‌کنیم.

فرض کنید $X = ۵$ مشاهده شده باشد. در این صورت $T = ۱$، $(Y_{۱}, Y_{۲}, Y_{۳}, Y_{۴}, Y_{۵}, Y_{۶}) = (۱, ۲, ۰, ۰, ۲, ۱)$، و

$$P(T = ۰ | Y_{۱} = ۱) = \frac{۲-\theta}{۴} \Longrightarrow P(T = ۱ | Y_{۱} = ۱) = \frac{۲+\theta}{۴}$$
$$P(T = ۰ | Y_{۲} = ۲) = \frac{۳-\theta}{۵} \Longrightarrow P(T = ۱ | Y_{۲} = ۲) = \frac{۲+\theta}{۵}$$
$$P(T = ۰ | Y_{۳} = ۰) = \frac{۱-\theta}{۳} \Longrightarrow P(T = ۱ | Y_{۳} = ۰) = \frac{۲+\theta}{۳}$$
$$P(T = ۰ | Y_{۴} = ۰) = \frac{۱-\theta}{۳} \Longrightarrow P(T = ۱ | Y_{۴} = ۰) = \frac{۲+\theta}{۳}$$
$$P(T = ۰ | Y_{۵} = ۲) = \frac{۳-\theta}{۵} \Longrightarrow P(T = ۱ | Y_{۵} = ۲) = \frac{۲+\theta}{۵}$$
$$P(T = ۰ | Y_{۶} = ۱) = \frac{۲-\theta}{۴} \Longrightarrow P(T = ۱ | Y_{۶} = ۱) = \frac{۲+\theta}{۴}$$

با توجه به محاسبات بالا ملاحظه می‌شود که استفاده از آماره‌های کمکی ماکسیمال مختلف، نتایج متفاوتی را به وجود می‌آورد و قادر به استنباط یکسان در مورد θ نیستیم.

در مثال اخیر دیدیم به دلیل تعدد آماره‌های کمکی ماکسیمال موفق به انجام استنباط به شرط آمارهٔ کمکی نشدیم. در قسمت بعد نشان می‌دهیم آمارهٔ کمکی می‌تواند به عنوان ملاکی برای دقت آزمایش در نظر گرفته شود. سپس از این خاصیت استفاده کرده ملاک کاکس برای انتخاب آمارهٔ کمکی ماکسیمال را معرفی می‌کنیم.

۳.۳ آمارهٔ کمکی به عنوان ملاکی برای دقت آزمایش

فرض کنید A آماره‌ای کمکی برای پارامتر θ باشد. از قبل می‌دانیم که اندازهٔ اطلاع در نمونهٔ X در مورد پارامتر θ به وسیلهٔ $I_X(\theta)$ اندازه‌گیری می‌شود. از طرفی

$$\begin{aligned}I_X(\theta) &= -\int_{\mathbf{R}} \frac{\partial^۲}{\partial\theta^۲}\log f(x;\theta) f(x;\theta)dx\\ &= -\int_{\mathbf{R}}[\frac{\partial^۲}{\partial\theta^۲}\log f(x;\theta)][\int_{\mathbf{R}} f(x,a;\theta)da]dx\\ &= \int_{\mathbf{R}}[-\int_{\mathbf{R}}\frac{\partial^۲}{\partial\theta^۲}\log f(x;\theta) f(x|a;\theta)dx]f(a)da\\ &= \int_{\mathbf{R}}(I_{X|A=a})f(a)da = E(I_{X|A}).\end{aligned}$$

بنابراین $I_X(\theta) = E(I_{X|A})$. پس اطلاع حاصل دربارهٔ θ از روی xهایی به دست می‌آید که برای آنها آمارهٔ کمکی A، مقدار معلوم a را انتخاب کرده باشد. همچنین نتیجه می‌گیریم که، آمارهٔ کمکی، فضای نمونه‌ای را براساس اطلاع موجود دربارهٔ پارامتر θ افراز می‌کند. به عبارتی می‌دانیم $I_X(\theta)$ دقت آزمایش است که معادل $E(I_{X|A})$ شده است. پس هر چه اطلاع حاصل از مقدار مشهود آمارهٔ کمکی بزرگتر باشد، دقت آزمایش بیشتر می‌شود.

مثال ۱۰: فرض کنید متغیر تصادفی X دارای توزیع احتمال زیر باشد:

x	۱	۲	۳	۴
$P_X(x)$	$\frac{۱}{۴}$	$\frac{۱-\theta}{۴}$	$\frac{۱-\theta}{۲}$	$\frac{۳\theta}{۴}$

$۰ < \theta < ۱$

آمارهٔ A را به شکل زیر تعریف می‌کنیم:

$$A = \begin{cases} ۱ & x = ۲,۳,۴\\ ۰ & x = ۱\end{cases}$$

A آماره‌ای کمکی است. اطلاع موجود در X به شکل زیر محاسبه می‌شود

$$\begin{aligned}I_X(\theta) &= -[۰ + \frac{۱-\theta}{۴}\cdot\frac{-۱}{(۱-\theta)^۲} + \frac{۱-\theta}{۲}\cdot\frac{-۱۰}{(۱-\theta)^۲} + \frac{۳\theta}{۴}\cdot\frac{-۱}{\theta^۲}]\\ &= \frac{۳}{۴\theta(۱-\theta)}.\end{aligned}$$

از طرفی

$$P(X|A=۰) = \begin{cases} ۱ & x = ۱\\ ۰ & x = ۲,۳,۴\end{cases}$$

$$P(X|A=۱) = \begin{cases} ۰ & x = ۱\\ \frac{۱-\theta}{۳} & x = ۲\\ \frac{۲(۱-\theta)}{۳} & x = ۳\\ \theta & x = ۴\end{cases}$$

$$\begin{aligned}I_{X|A=۰} &= -E[\frac{\partial^۲}{\partial\theta^۲}\log P(X;\theta)|A=۰] = ۰\\ I_{X|A=۱} &= -E[\frac{\partial^۲}{\partial\theta^۲}\log P(X;\theta)|A=۱] = \frac{۱}{\theta(۱-\theta)}\\ E(I_{X|A}) &= \sum_a I_{X|A=a}P(a) = \frac{۳}{۴\theta(۱-\theta)} = I_X(\theta).\end{aligned}$$

همان طور که مشاهده می‌شود آمارهٔ کمکی A فضای نمونه‌ای را به دو مجموعهٔ $\{۱\}$ و $\{۲,۳,۴\}$ افراز می‌کند. در اینجا $I_{X|A=۰} = ۰$. زیرا با مشاهدهٔ $A = ۰$ که معادل $x = ۱$ است، اطلاعی در مورد θ حاصل نمی‌شود. پس وقتی که $A = ۰$ آزمایش فاقد دقت است. از طرفی $I_{X|A=۱} = \frac{۱}{\theta(۱-\theta)}$، حاوی اطلاع در مورد θ است. پس دقت آزمایش با مشاهدهٔ $A = ۱$ زیاد می‌شود.

۴.۳ روش کاکس برای انتخاب آمارهٔ کمکی ماکسیمال

آنچه به نام روش کاکس در تعیین آمارهٔ کمکی ماکسیمال شناخته شده است، بر اساس دقت آزمایش بنا گردیده است. به عقیدهٔ وی آمارهٔ کمکی ماکسیمال باید برای استنباط شرطی انتخاب شود که به بهترین صورت ممکن فضای نمونه‌ای را براساس اطلاع موجود دربارهٔ پارامتر θ افراز می‌کند. به عبارتی اگر $A_۱, A_۲, \ldots, A_k$ آماره‌های کمکی ماکسیمال در یک مسألهٔ استنباطی باشند، آمارهٔ کمکی A_j از میان آنها انتخاب می‌شود، هرگاه:

$$\mathrm{Var}(I_{X|A_j}) \geq \mathrm{Var}(I_{X|A_i})\ i = ۱,۲,\ldots,k.$$

می‌دانیم هر آمارهٔ کمکی، فضای نمونه‌ای را افراز می‌کند. از طرفی $I_{X|A_i=a_i}$ اطلاع موجود در زیرمجموعه‌ای از فضای نمونه‌ای را که برای آن $A_i = a_i$، محاسبه می‌کند. لذا ایدهٔ کاکس، آن است که آمارهٔ کمکی ماکسیمالی

انتخاب شود که در زیرمجموعه‌هایی از فضای نمونه‌ای که با شرط کردن روی A_iهای متفاوت حاصل می‌شود، افراز به دست آمده از $A_j = a_j$ از نظر اطلاع موجود درباره پارامتر θ بهترین افراز باشد.

مثال ۱۱. توزیع چندجمله‌ای با چهار سلول را در نظر بگیرید. احتمالها به ترتیب عبارت‌اند از

$$P_1(\theta) = \frac{1}{6}(1-\theta), \quad P_2(\theta) = \frac{1}{6}(1+\theta)$$
$$P_3(\theta) = \frac{2-\theta}{6}, \quad P_4(\theta) = \frac{2+\theta}{6}.$$

که در آن θ پارامتری ناشناخته است و $|\theta| < 1$. بنابراین

$$L(\theta) = \frac{n!}{n_1!n_2!n_3!n_4!}\left(\frac{1-\theta}{6}\right)^{n_1}\left(\frac{1+\theta}{6}\right)^{n_2}\left(\frac{2-\theta}{6}\right)^{n_3}\left(\frac{2+\theta}{6}\right)^{n_4}$$

اگر N_i برای $i = 1, 2, 3, 4$ متغیر تصادفی تعداد مشاهدات در هر سلول باشد، داریم:

$$N_1 \sim b(n, \frac{1-\theta}{6}), \quad N_2 \sim b(n, \frac{1+\theta}{6})$$
$$N_3 \sim b(n, \frac{2-\theta}{6}), \quad N_4 \sim b(n, \frac{2+\theta}{6}).$$

اکنون با محاسبات مختصری می‌توان نشان داد اطلاع موجود در $\mathbf{N}$ در مورد θ عبارت است از $I(\theta) = \frac{n(2-\theta^2)}{(1-\theta^2)(4-\theta^2)}$. اگر قرار دهیم $A_1 = N_1 + N_2$ و $A_2 = N_1 + N_4$ آنگاه $A_1 \sim b(n, \frac{1}{3})$، $A_2 \sim b(n, \frac{1}{2})$ بنابراین A_1 و A_2 آماره‌های کمکی ماکسیمال هستند، زیرا هر آمارهٔ کمکی دیگری، تابعی از آماره‌های کمکی A_1 یا A_2 است. حال $I_{\mathbf{N}|A_1=a_1}$ و $I_{\mathbf{N}|A_2=a_2}$ را محاسبه می‌کنیم:

$$I_{\mathbf{N}|A_1=a_1} = -E\left(\frac{\partial^2}{\partial\theta^2}\log L(\theta)|A_1 = a_1\right).$$

برای محاسبهٔ $I_{\mathbf{N}|A_1=a_1}$ به توابع احتمال $N_i|A_1 = a_1$، $i = 1, 2, 3, 4$ نیاز داریم. توزیع N_2 و N_1 سه‌جمله‌ای $(n, P_1, P_2, P_3 + P_4)$ است. با محاسبهٔ مختصری می‌توان نشان داد که

$$P(N_1 = n_1|A_1 = a_1) = \binom{a_1}{n_1}\left(\frac{1-\theta}{2}\right)^{n_1}\left(\frac{1+\theta}{2}\right)^{a_1-n_1}.$$

بنابراین $N_1|A_1 = a_1 \sim b(a_1, \frac{1-\theta}{2})$ و به طور مشابه داریم:

$$N_2|A_1 = a_1 \sim b(a_1, \frac{1+\theta}{2})$$
$$N_3|A_1 = a_1 \sim b(n-a_1, \frac{2+\theta}{4})$$
$$N_4|A_1 = a_1 \sim b(n-a_1, \frac{2+\theta}{4}).$$

در نتیجه

$$\begin{aligned}
I_{\mathbf{N}|A_1=a_1} &= \frac{E(N_1|A_1 = a_1)}{(1-\theta)^2} + \frac{E(N_2|A_1 = a_1)}{(1+\theta)^2} \\
&+ \frac{E(N_3|A_1 = a_1)}{(2-\theta)^2} + \frac{E(N_4|A_1 = a_1)}{(2+\theta)^2} \\
&= \frac{a_1}{(1-\theta)^2} + \frac{n-a_1}{(4-\theta^2)} \\
&= \frac{3a_1 + n(1-\theta^2)}{(1-\theta^2)(4-\theta^2)}
\end{aligned}$$

به همین ترتیب می‌توان نشان داد که

$$I_{\mathbf{N}|A_2=a_2} = \frac{2a_2\theta + n(1-\theta)(2+\theta)}{(1-\theta^2)(4-\theta^2)}$$

با انجام محاسبات ساده‌ای خواهیم داشت:

$$\mathrm{Var}(I_{\mathbf{N}|A_1=a_1}) = \frac{2n}{(1-\theta^2)^2(4-\theta^2)^2}$$
$$\mathrm{Var}(I_{\mathbf{N}|A_2=a_2}) = \frac{n\theta^2}{(1-\theta^2)^2(4-\theta^2)^2}$$

و برای هر $\theta \in (-1, 1)$ داریم:

$$\mathrm{Var}(I_{\mathbf{N}|A_1=a_1}) > \mathrm{Var}(I_{\mathbf{N}|A_2=a_2})$$

به عقیدهٔ کاکس باید آمارهٔ کمکی A_1 را به A_2 ترجیح دهیم و در استنباط شرطی از آمارهٔ کمکی A_1 استفاده کنیم.

۴ آماره‌های کمکی و قضیه باسو

می‌دانیم آمارهٔ بسندهٔ مینیمال، لازم نیست کامل باشد. ولی در صورتی که آمارهٔ بسندهٔ کامل وجود داشته باشد، مینیمال نیز هست. بنابراین آمارهٔ بسندهٔ کامل در فشردگی داده‌ها مؤثر است. این موضوع در قضیهٔ باسو نشان داده شده است.

قضیهٔ ۲ (قضیهٔ باسو): فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونهٔ تصادفی از چگالی $f(x;\theta)$، $\theta \in \Theta$ و T یک آمارهٔ بسندهٔ کامل کراندار باشد. در این صورت T از هر آمارهٔ کمکی A مستقل است. البته این قضیه برای آماره‌های بسندهٔ کامل نیز برقرار است.

یک نکتهٔ اساسی در این قضیه این است که عکس قضیه، در حالت کلی برقرار نیست. اما با تعریف خاصی از آمارهٔ کمکی و یا آمارهٔ کامل، می‌توان نوعی از عکس قضیهٔ باسو را بیان کرد.

در مثال زیر نشان داده می‌شود که عکس قضیهٔ باسو برقرار نیست. در واقع در این مثال خواهیم دید اگر T یک آمارهٔ بسنده، و A یک آمارهٔ کمکی مستقل از T باشد، کامل کراندار بودن T را نمی‌توان نتیجه گرفت.

مثال ۱۲: فرض کنید X متغیری تصادفی با توزیع زیر باشد:

x	-5	-4	-3	-2	-1	1	2	3	4	5
$P_X(x)$	$\alpha' p^2 q$	$\alpha' p q^2$	$\frac{1}{2}p^3$	$\frac{1}{2}q^3$	$\gamma' pq$	γpq	$\frac{1}{2}q^3$	$\frac{1}{2}p^3$	αpq^2	$\alpha p^2 q$

α و α' و γ و γ' مقادیر ثابتی هستند که در شرط $\alpha + \gamma = \alpha' + \gamma' = \frac{3}{2}$ صدق می‌کنند. اگر $0 < p < 1$ و $q = 1 - p$ باشد $P_X(x)$ یک تابع احتمال خواهد بود. آمارهٔ $T = |X|$ آمارهٔ بسندهٔ مینیمال برای پارامتر توزیع یعنی p است. اگر $S = \{x; x > 0\}$ و A را به شکل زیر تعریف کنیم:

$$A = I_{(s)}(X) = \begin{cases} 1 & X \in S \\ 0 & X \in S^c \end{cases}$$

در این صورت

$$P(A = 1) = P(X > 0) = \sum_{i=1}^{5} P(X = i) = \frac{1}{2}$$

$$\Longrightarrow P(A = 0) = P(X \leq 0) = \frac{1}{2}.$$

بنابراین توزیع A به P بستگی ندارد. یعنی A آماره‌ای کمکی است.

با محاسبهٔ کوواریانس T و A خواهیم داشت:

$$\operatorname{Cov}(T, A) = 5\alpha p^2 q + 4\alpha pq^2 + \gamma pq$$
$$- \frac{1}{2}(\alpha + \alpha')(5p^2 q + 4pq^2) - \frac{1}{2}(\gamma + \gamma')pq.$$

در عبارت اخیر اگر $\alpha \neq \alpha'$، آنگاه $\operatorname{Cov}(T, A) \neq 0$. در نتیجه آمارهٔ بسندهٔ T از آمارهٔ کمکی A مستقل نیست. اما اگر $\alpha = \alpha'$، در این صورت، $\operatorname{Cov}(T, A) = 0$ و آماره‌های T و A ناهمبسته می‌شوند. توزیع توأم دو آمارهٔ T و A عبارت است از

$a \backslash t$	1	2	3	4	5	$p_A(a)$
0	$\gamma' pq$	$\frac{1}{2}q^3$	$\frac{1}{2}p^3$	$\alpha' pq^2$	$\alpha' p^2 q$	$\frac{1}{2}$
1	γpq	$\frac{1}{2}q^3$	$\frac{1}{2}p^3$	αpq^2	$\alpha p^2 q$	$\frac{1}{2}$
$p_T(t)$	$(\gamma + \gamma')pq$	q^3	p^3	$(\alpha + \alpha')pq^2$	$(\alpha + \alpha')p^2 q$	

اگر فرض کنیم $\alpha = \alpha'$ و شرط $\alpha + \gamma = \alpha' + \gamma' = \frac{3}{2}$ را در نظر بگیریم، نتیجه خواهیم گرفت که $\gamma = \gamma'$. بر این اساس دو آمارهٔ T و A از یکدیگر مستقل خواهند بود، زیرا حاصلضرب احتمالهای حاشیه‌ای مساوی احتمالهای توأم T و A است. بنابراین آمارهٔ بسندهٔ T از آمارهٔ کمکی A تحت شرایطی مستقل است، اما T آمارهٔ کامل کراندار نیست. زیرا اگر قرار دهیم $E(f(T)) = 0$ خواهیم داشت:

$$[-f(2) + f(3) + (\alpha + \alpha')f(4) - (\alpha + \alpha')f(5)]p^3 +$$
$$[-(\gamma + \gamma')f(1) + 3f(2) - (\alpha + \alpha')(2f(4) - f(5))]p^2$$
$$+ (\gamma + \gamma')f(1) - 3f(2) + (\alpha + \alpha')f(4)]p + f(2) = 0.$$

بنابراین

$$\begin{cases} f(2) = 0 \\ (\gamma + \gamma')f(1) - 3f(2) + (\alpha + \alpha')f(4) = 0 \\ -(\gamma + \gamma')f(1) + 3f(2) - (\alpha + \alpha')(2f(4) - f(5)) = 0 \\ -f(2) + f(3) + (\alpha + \alpha')(f(4) - (\alpha + \alpha')f(5)) = 0 \end{cases}$$

با فرض $a = f(1)$ خواهیم داشت:

$$f(1) = a,\ f(2) = f(3) = 0,\ f(4) = f(5) = -\frac{\gamma + \gamma'}{\alpha + \alpha'}a.$$

چون a دلخواه است پس برای هر t، $f(t) \neq 0$. وقتی یک مدل کامل نیست، آمارهٔ بسندهٔ مینیمال شامل آمارهٔ کمکی است. از طرفی، کامل بودن یک آمارهٔ بسندهٔ مینیمال، موجب جدایی کامل قسمت حاوی اطلاع نمونه از آمارهٔ کمکی می‌شود. با این مقدمه چند مثال را بررسی کرده و سپس به دسته‌بندی توزیعها خواهیم پرداخت.

مثال ۱۳: فرض کنید $X_1, X_2, \ldots, X_n$ نمونه‌ای تصادفی از توزیع $N(\theta, 1)$ باشد. می‌دانیم آمارهٔ بسندهٔ مینیمال برای θ عبارت است از $\bar{X}$ و این آماره کامل است.

مثال ۱۴: فرض کنید $X_1, X_2, \ldots, X_n$ نمونه‌ای تصادفی از توزیع نمایی زیر باشد:

$$f(x; \theta) = e^{-(x-\theta)} \qquad x > \theta$$
$$= 0 \qquad \text{سایر جاها}$$

آمارهٔ $X_{(1)} = \min\{X_1, X_2, \ldots, X_n\}$ برای θ، بسنده و کامل است.

مثال ۱۵: فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونهٔ تصادفی از توزیع $u(\theta - \frac{1}{2}, \theta + \frac{1}{2})$ باشد. آمارهٔ بسندهٔ مینیمال برای θ آمارهٔ $(X_{(1)}, X_{(n)})$ است. به سادگی تحقیق می‌شود که این آماره کامل نیست.

مثال ۱۶: فرض کنید $X_1, X_2, \ldots, X_n$ نمونه‌ای تصادفی از توزیع کوشی $C(\theta, 1)$ باشد. در این صورت $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ آمارهٔ بسندهٔ مینیمال برای θ است اما این آماره، کامل نیست، زیرا

$$f_\theta(x) = \frac{1}{\pi(1 + (x - \theta)^2)},\ x \in \mathbb{R}, \theta \in \mathbb{R}$$

و اگر $g(\cdot)$ تابع چگالی آماره‌های ترتیبی $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ باشد، آنگاه

$$g(x_{(1)},x_{(2)},\ldots,x_{(n)}) = n!\prod_{i=1}^{n}\frac{1}{\pi(1+(x_{(i)}-\theta)^2)}$$
$$-\infty < x_{(1)} < x_{(2)} < \cdots < x_{(n)} < \infty.$$

تابع g تابعی زوج برحسب $x_{(i)} - \theta$ است. هر تابع فردی مثل h از $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ که حاصلضرب آن با g انتگرالپذیر لبگ باشد ایجاب می‌کند که برای هر $\theta \in \Theta$،

$$\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} h(x_{(1)},x_{(2)},\ldots,x_{(n)})$$
$$g(x_{(1)},x_{(2)},\ldots,x_{(n)})dx_{(1)}\ldots dx_{(n)} = 0$$

بی‌آنکه لزوماً $h = 0$. بنابراین آمارهٔ بسندهٔ مینیمال در این مورد کامل نیست.

مثال ۱۷: فرض کنید $X_1, X_2, \ldots, X_n$ نمونه‌ای تصادفی از توزیع لجستیک $L(\theta, 1)$ با تابع چگالی زیر باشد:

$$f_\theta(x) = \frac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}\ x\in\mathbb{R},\theta\in\mathbb{R},$$

در این صورت $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ آمارهٔ بسندهٔ مینیمال برای θ است، اما این آماره کامل نیست.

توزیعهایی که در مثالهای ۱۱ تا ۱۵ مورد بررسی قرارگرفتند، همگی از خانوادهٔ توزیعهای مکانی $f(x-\theta)$ هستند. این توزیعها را می‌توان در دو گروه دسته‌بندی کرد. گروه اول را شامل توزیعهایی در نظر می‌گیریم که برای آنها آمارهٔ بسندهٔ مینیمال، یک بعدی است. در این گروه توزیعهای نرمال و نمایی قرار می‌گیرند. گروه دوم را شامل توزیع یکنواخت و توزیعهایی که برای آنها، آمارهٔ بسندهٔ مینیمال، آماره‌های ترتیبی حاصل از n مشاهده است در نظر می‌گیریم، پس در این گروه توزیعهای کوشی و لجستیک نیز قرار می‌گیرند.

در توزیعهای گروه اول آمارهٔ بسندهٔ مینیمال، یک بعدی است، یعنی تمام اطلاعات در مورد θ، از طریق آماره‌ای یک بعدی استخراج می‌شود، حال آنکه در توزیعهای گروه دوم، باید به آماره‌ای با بعد بیش از یک مراجعه کنیم. نکتهٔ قابل توجه در اینجا این است که، خانوادهٔ توزیعهای آمارهٔ بسندهٔ مینیمال در توزیعهای نرمال و نمایی، کامل نیز هستند. اما در سه توزیع دیگر، خانوادهٔ توزیعهای آمارهٔ بسندهٔ مینیمال، کامل نیست. همچنین در مثال ۳ دیدیم که تفاضلهای $Y_i = X_{(n)} - X_{(i)}$، $i = 1, 2, \ldots, n-1$ در خانوادهٔ مکانی تشکیل یک آمارهٔ کمکی $n-1$ بعدی را می‌دهد و هر Y_i یک آمارهٔ کمکی است.

در توزیعهای یکنواخت، کوشی و لجستیک آمارهٔ کمکی $Y_i = X_{(n)} - X_{(i)}$ تابعی از آمارهٔ بسندهٔ مینیمال است. بنابراین در این سه مورد، بسندگی موفقیتی درکنار گذاشتن آمارهٔ کمکی نداشته است. به این معنی که آمارهٔ بسندهٔ مینیمال، مستقل از آمارهٔ کمکی نیست.

در توزیعهای نرمال و نمایی آمارهٔ بسندهٔ مینیمال، کامل نیز هست و از آمارهٔ کمکی $Y_i = X_{(n)} - X_{(i)}$ مستقل است. اما در سه توزیع دیگر آمارهٔ بسندهٔ مینیمال، کامل نبوده و از آمارهٔ کمکی نیز مستقل نیستند. بنابراین می‌توانیم مدلهای کامل را به عنوان مدلهایی که در آنها آمارهٔ بسندهٔ مینیمال، به طور کامل، درکنار گذاشتن آماره‌های کمکی موفق بوده‌اند، در نظر بگیریم.

قضیهٔ باسو نیز بر همین موضوع تأکید دارد. مثالهای اخیر و قضیهٔ باسو نشان می‌دهند که استقلال آمارهٔ کمکی از آمارهٔ بسندهٔ مینیمال به دلیل ویژگیهای این دو نوع آماره نیست، بلکه معلول خاصیت آماری دیگری است، که آن خاصیت، کامل بودن آمارهٔ بسندهٔ مینیمال می‌باشد.

همان طورکه اشاره شد، در حالت کلی، عکس قضیهٔ باسو برقرار نیست. لیمن در مقالهٔ خود در سال ۱۹۸۱ با تعریف آمارهٔ کمکی مرتبهٔ اول، آماره‌های Fـ کامل و F_1ـ کامل، قضایایی در راستای قضیهٔ باسو و عکس آن ارائه کرده است.

در ادامهٔ مقاله، آمارهٔ کمکی مرتبهٔ اول را تعریف کرده و بر اساس این تعریف، حالت خاصی از قضیهٔ باسو و عکس آن را بیان و ثابت می‌کنیم.

تعریف ۱: آمارهٔ $V = V(X)$ را کمکی مرتبهٔ اول گوییم هرگاه $E_\theta(V)$ بستگی به پارامتر مجهول θ نداشته باشد. بنابراین آمارهٔ کمکی مرتبهٔ اول، برآوردگر نااریب صفر به اضافهٔ یک مقدار ثابت است، یعنی برای هر $\theta \in \Theta$، $E_\theta(V) = C$ که در آن C از θ مستقل است.

مثال ۱۸: فرض کنید X دارای تابع احتمال $P_\theta(x)$ باشد:

$$P_\theta(x) = (\frac{\theta}{2})^{|x|}(1-\theta)^{1-|x|} \qquad x=-1,0,1, \ 0<\theta<1$$

در این صورت داریم:

$$E(X) = (-1)(\frac{\theta}{2}) + 0 + (1)(\frac{\theta}{2})$$

بنابراین X آمارهٔ کمکی مرتبهٔ اول است.

لم ۱: اگر T آمارهٔ بسنده و کامل کراندار باشد در این صورت آمارهٔ کراندار $S = S(X)$ یک برآوردگر نااریب صفر است اگر و تنها اگر $\eta(t) = E(S|t) = 0 \ [a.e.p^T]$.

اکنون می‌توانیم حالت خاصی از قضیهٔ باسو را با تکیه بر آمارهٔ کمکی مرتبهٔ اول اثبات کنیم.

قضیه شرط لازم و کافی برای اینکه T، کامل کراندار باشد آن است که هر آمارهٔ کراندار کمکی مرتبهٔ اول با هر تابع حقیقی کراندار از T ناهمبسته باشد.

برهان (شرط لازم): فرض کنید آمارهٔ بسندهٔ T، کامل کراندار باشد. اگر V یک آمارهٔ کمکی مرتبهٔ اول باشد بی‌آنکه از کلیت کاسته شود، فرض کنید $E(V) = ۰$. زیرا اگر $E(V) = C$، قرار می‌دهیم $V' = V - C$ و V' را آمارهٔ کمکی مرتبهٔ اول در نظر می‌گیریم و خواهیم داشت $E(V') = ۰$. باید نشان دهیم برای هر تابع حقیقی و کراندار $f(t)$، آمارهٔ V و $f(T)$ ناهمبسته‌اند:

$$\begin{aligned} \mathrm{Cov}_\theta(f(T), V) &= E_\theta(f(T)V) - E_\theta(f(T))E_\theta(V) \\ &= E(f(T)V) \\ &= E_\theta[E(f(T)V|T)] \\ &= E_\theta[f(T)E(V|T)] \end{aligned}$$

اما بنابر لم ۱ داریم $E(V|T) = ۰$. بنابراین $\mathrm{Cov}_\theta(f(T), V) = ۰$.

(شرط کافی): فرض کنید آمارهٔ بسندهٔ T، کامل کراندار نباشد. پس تابع کرانداری مثل f وجود دارد به طوری که $E_\theta(f(T)) = ۰$، اما $f(T) \neq ۰$. آمارهٔ $V = V(X)$ آمارهٔ کمکی مرتبهٔ اول است. بنابراین $E(V) = ۰$. پس برای یک مقدار مشخص $\theta_\circ$، با احتمال مثبت قرار می‌دهیم $V(X) = f(T)$.

$$\mathrm{Cov}_{\theta_\circ}(V, f(T)) = E_{\theta_\circ}(Vf(T)) = E_{\theta_\circ}(f^۲(T)) > ۰$$

بنابراین V و $f(T)$ همبسته‌اند. پس اگر $f(T)$ و V ناهمبسته باشند، آمارهٔ T کامل کراندار است و کفایت شرط نیز ثابت می‌شود.

قضیهٔ ۳ برای آمارهٔ بسندهٔ کامل نیز برقرار است به شرط آنکه واریانس آمارهٔ T را محدود فرض کنیم. باید توجه داشت که در قضیهٔ ۳ فقط نوعی از قضیهٔ باسو و عکس آن اثبات شده است. زیرا در این قضیه شرط استقلال آمارهٔ بسندهٔ T از آمارهٔ کمکی V، به شرط ناهمبسته بودن این دو آماره تبدیل شده است. این قضیه را قضیهٔ باسوی مرتبهٔ اول می‌نامیم.

مثال ۱۹: تابع احتمال مثال ۱۸ را در نظر بگیرید:

$$P_\theta(x) = \left(\frac{\theta}{۲}\right)^{|x|}(۱-\theta)^{۱-|x|} \qquad x = -۱, ۰, ۱ \quad ۰ < \theta < ۱$$

$$= ۰ \qquad \text{سایر جاها}$$

دیدیم که $V = X$ آمارهٔ کمکی مرتبهٔ اول است، آمارهٔ $T = |X|$ برای θ بسنده است. زیرا توزیع شرطی X، با فرض مقدار T، به پارامتر θ بستگی ندارد. این موضوع به سادگی از توزیع توأم زیر قابل بررسی است. برای این منظور، احتمالات شرطی را محاسبه می‌کنیم.

$t \backslash x$	-۱	۰	۱	$P_T(t)$
۰	۰	$۱-\theta$	۰	$۱-\theta$
۱	$\frac{\theta}{۲}$	۰	$\frac{\theta}{۲}$	θ

$$P(X=۰|t=۰) = ۱$$

$$P(X=۱|t=۰) = P(X=-۱|t=۰) = P(X=۰|t=۱) = ۰$$

$$P(X=۱|t=۱) = P(X=-۱|t=۱) = \frac{۱}{۲}.$$

تابعی کراندار از T مثل $g(T)$ را در نظر می‌گیریم،

$$\begin{aligned} \mathrm{Cov}(X, g(T)) &= E(Xg(T)) \\ &= (-۱)g(۱)\frac{\theta}{۲} + (۱)g(۱)\frac{\theta}{۲} = ۰, \end{aligned}$$

و بنا بر قضیهٔ باسوی مرتبهٔ اول، آمارهٔ بسندهٔ $T = |X|$ کامل کراندار است

مراجع

[1] BASU, D. (1955), "On statistics independent of a complete sufficient statistic," *Sankhya*, **5**, P. 377-380.

[2] BASU, D. (1958), "On statistics independent of a complete sufficient statistic," *Sankhya*, **20**, P. 223-226.

[3] BASU, D. (195), "The family of ancillary statistics" *Sankhya*, (A) **21**, P. 247-256.

[4] BASU, D. (1964), "Recovery of ancillary information" *Sankhya* (A) **26**, P. 3-16.

[5] BUEHLER, R. J. (1982) Some ancillary statistics and their properties. *Journal of the American statistical Association.* **77**, P. 581-594.

[6] CASELLA, G. and BERGER, R. L, (1990) Statistical inference. Thomson information pub.

[7] Cox, D. R. (1971), "The choice between alternative ancillary statistics," *Journal of the Royal Statistical Society*, **Ser. B, 33**, P. 251-255.

[8] COX, D. R, and HINKLEY, D. V. (1974) Theoretical statistics; London; Chapman and Hall.

[9] FISHER, R. A. (1959) "Statistical methods and scientific inference", Hafner publishing company New York.

[10] GODAMBE, V. P., (1984) "On ancillarity and Fisher information in the presence of a nuisance parameter. *Biometrika*, **71**. 3. P. 626-629.

[11] LANDERS, D., and ROGGE; L. (1972), "Minimal sufficient σ-fields and minimal sufficient statistics." *Annals of Mathematical Statistics*, 43, p. 2045-2049.

[12] LEHMMAN. N., E. L., and SCHEFFE, H. (1950-1956) "Completencess, Similar Regions, and Unbiased Estimation" *Sankhya*, **10**, P. 305-340; **15**, P. 219-236 and **17**, P. 250.

[13] LEHMMAN, E. L. (1981) "An Interpretation of Completeness and Basu's Theorem ," *J. A. S. A.*, **76**, 374, P. 335-340.

[14] LEHMMAN, E. L. (1983) "Theorey of Point Estimation; "John Wiley & Sons.

[15] LINDGREN, B. W., (1976) "Statistical theory", Third Edition. Memillan.

[۱۶] پارسیان، احمد. بسندگی و بسندگی مینیمال، سال دهم، شمارهٔ دوم، ۱۳۷۰ فرهنگ و اندیشهٔ ریاضی.

[۱۷] سیاره، عبدالرضا. آمارهٔ کمکی همراه برآوردکنندهٔ درستنمایی ماکسیمم و آمارهٔ بسنده، ۱۳۷۳، رسالهٔ کارشناسی‌ارشد، دانشگاه تربیت معلم.