دونکته از این معنی:

# قضیه بهادر

سید محمود طاهری [1]

چکیده

در این مقاله، نخست مفهوم قضیه بهادر تشریح می‌شود. سپس روشهای متفاوت اثبات آن ارایه و مقایسه می‌شوند. با چند مثال، توضیح می‌دهیم که عکس قضیه همواره برقرار نیست.

## مقدمه

از زیباییهای ریاضیات، روشهای متفاوت اثبات یک قضیه است. هر روش، پنجره‌ای به معنی و مفهوم پوشیده در قضیه می‌گشاید و نشان می‌دهد که آن قضیه چگونه با مباحث دیگر پیوند می‌خورد. بررسی روشهای گوناگون اثبات یک قضیه، از دیدگاه آموزشی نیز حائز اهمیت است. به این‌ترتیب، عملاً نشان داده می‌شود که چگونه می‌توان از ابزارهای مختلف ریاضی استفاده کرده و آنها را برای رسیدن به یک هدف هماهنگ نمود.

در این مقاله به یکی از قضایای آمار ریاضی که نخستین بار توسط بهادر، آماردان هندی، ارایه و اثبات شد می‌پردازیم. در بخش اول معنا و مفهوم پنهان در قضیه تشریح می‌شود. در بخش دوم، سه اثبات برای آن بیان می‌کنیم. در بخش پایانی با چند مثال، نشان می‌دهیم که عکس قضیه همواره برقرار نیست.

فرض بر این است که خواننده با مفاهیمی مانند آماره، آماره فرعی، بسندگی، بسندگی مینیمال و کامل بودن آشنا می‌باشد.

## ۱. نکته اول: مفهوم قضیه بهادر

نخست توضیحاتی درباره آماره بسنده، آماره بسنده مینیمال و آماره فرعی می‌دهیم. می‌دانیم که ویژگی یک آماره بسنده، این است که اطلاعاتی را که داده‌ها درباره پارامتر مجهول بیان می‌دارند، فشرده و درخود متمرکز می‌کند. آماره بسنده‌ای که بیشترین فشردگی و کاهش در داده‌ها را انجام دهد، بسنده مینیمال نام دارد. به بیان ریاضی، آماره بسنده‌ای که تابعی از هر آماره بسنده دیگر باشد، آماره بسنده مینیمال است.

حال سؤال این است که: از دید شهودی چه ویژگی باعث

[1] سید محمود طاهری، دانشکده علوم ریاضی، دانشگاه صنعتی اصفهان (Taheri@cc.iut.ac.ir)

می‌شود و تضمین می‌کند که یک آماره، بیشترین فشردگی در داده‌ها را متحمّل (متقبّل ؟!) شود؟ جواب به این سؤال پیوند نزدیک با مفهوم آماره فرعی (کمکی) دارد. ابتدا اشاره می‌کنیم که یک آماره، در صورتی می‌تواند به ما درباره پارامتر مجهول آگاهی ببخشد که توزیع آماره، به آن پارامتر بستگی داشته باشد. اگر آماره‌ای اینگونه نباشد، آماره فرعی نامیده می‌شود[۶]. به همین جهت آماره‌ای را که فرعی نباشد، یعنی توزیع آن به پارامتر مجهول بستگی داشته‌باشد، آگاهی‌بخش می‌نامند. بدیهی است که اگر $T$ یک آماره فرعی و $g(T)$ تابعی از آن باشد، آنگاه $E_\theta[g(T)]$ مقداری ثابت است و به $\theta$ بستگی ندارد. آماره‌ای که ویژگی اخیر را (برای هر تابع دلخواه $g(t)$) داشته‌باشد، آماره فرعی مرتبه اول نامیده می‌شود.

اکنون فرض کنید $T$ یک آماره بسنده باشد. اگر بتوان تابعی از $T$ ساخت که آماره فرعی باشد، این بدین معنی است که: «در $T$ آماره فرعی وجود دارد» و به عبارت دیگر: «هنوز در $T$ چیزهایی هست که اطلاعی درباره $\theta$ به ما نمی‌دهد» و لذا می‌توان نتیجه گرفت که: «$T$ هنوز کاملاً فشرده نشده است». با این توضیحات، آشکار می‌شود که یک آماره بسنده $T$ در بیان و فشردگی داده‌ها کاملاً تواناست، اگر هیچ تابعی (غیر از تابع ثابت) از $T$ یافت نشود که آماره فرعی، و در واقع حتی فرعی مرتبه اول، باشد. یعنی اگر $E_\theta[g(T)] = \circ \quad \forall\theta$ آنگاه با احتمال یک $g(T) = C$، که با کم کردن ثابت $C$ از طرفین به شرط زیر می‌رسیم:

$$E_\theta[g(T)] = \circ \ \forall\theta \Rightarrow P_\theta[g(t) = \circ] = ۱ \ \forall\theta$$

و این چیزی نیست جز تعریف کامل بودن یک آماره[۱۰]. پس کامل بودن یک آماره بسنده، به طور ذاتی و اساسی منجر به بیشترین فشردگی ممکن در داده‌ها می‌شود و لذا تعجب ندارد اگر ادعا شود که:

قضیه بهادر: هر آماره بسنده کامل، بسنده مینیمال است.

## ۲. نکته دوم: اثبات‌های گوناگون برای قضیه بهادر

در ادامه فرض می‌کنیم که $X_۱, \ldots, X_n$، نمونه‌ای تصادفی از جامعه‌ای با تابع چگالی احتمال $f(x;\theta)$ باشد.

**روش اول:**[۱۰] فرض کنید $T$ آماره بسنده کامل و $S$ آماره بسنده مینیمال باشد. ثابت می‌کنیم که $T$ و $S$ معادل هم و در واقع یک آماره‌اند. بدین معنی که $T$ تابعی از $S$ است، و $S$ تابعی از $T$.

بدیهی است که $S$ تابعی از $T$ باشد، زیرا آماره بسنده مینیمال تابعی از هر آماره بسنده است. کافی است ثابت کنیم $T$ تابعی از $S$ است.

تابع $f(T)$ را به صورت زیر در نظر بگیرید:

$$f(T) = T - E(T|S) = T - g(S)$$

اما $E[f(T)] = \circ$، پس بنابه کامل بودن $T$ داریم: $P_\theta[f(T) = \circ] = ۱ \ \forall\theta$ و یعنی $P_\theta[T = g(S)] = ۱ \ \forall\theta$.

**روش دوم:** فرض کنید $T$ آماره بسنده کامل باشد، ثابت می‌کنیم $T$ تابعی از هر آماره بسنده دلخواه $S$ است و لذا مینیمال است. دو تابع $h(S)$ و $L(T)$ را به صورت زیر تعریف می‌کنیم:

$$h(S) = E(T|S)$$

$$L(T) = E[h(S)|T]$$

داریم

$$E[L(T)] = E[h(S)] = E[T]$$

و بنابه قضیه رائو–بلکول

$$Var[L(T)] \leq Var[h(S)] \leq Var(T) \qquad (۱)$$

از طرف دیگر $E[L(T) - T] = \circ$ و چون $T$ کامل است پس با احتمال یک: $L(T) = T$، و لذا $Var[L(T)] = Var(T)$.

پس نابرابریها در (۱) در واقع برابری هستند یعنی $T$ و $h(S)$ و $L(T)$ سه برآوردگر نااریب $E_\theta(T)$ می‌باشند که واریانس‌های یکسان دارند؛ به ویژه $Var[h(S)] = Var(T)$. اما مجدداً بنابه قضیه رائو–بلکول، برابری واریانسهای دو برآوردگر $T$ و $h(S)$، هنگامی است که این دو برآوردگر با احتمال یک برابر باشند، یعنی $T = h(S)$. یعنی آماره بسنده کامل $T$ تابعی از آماره بسنده دلخواه $S$ است.

یادآور می‌شویم که اثبات نابرابریهای (۱) و این موضوع را که برابری واریانسهای $T$ و $h(S)$ نتیجه می‌دهد که $T = h(S)$، می‌توان مستقیماً برپایه نامساوی گشتاوری یعنی آنچه اساس قضیه رائو–بلکول است، بدست آورد[۲].

**روش سوم:**[۷] ثابت می‌کنیم هر آماره بسنده که مینیمال نباشد، کامل نخواهد بود.

برای این، فرض کنید $T$ یک آماره بسنده باشد ولی مینیمال نباشد. تابعی مانند $h(T) \neq 0$ می‌سازیم که امید ریاضی آن صفر است.

**ساختن تابع $h(T)$ با شرط فوق:** چون آماره $T$ مینیمال نیست پس آماره بسنده $T^*$ و تابعی مانند $g$ وجود دارد که $T^*(X) = g[T(X)]$ به قسمی که برای دست کم یک $a$ و دو مقدار $c_1$ و $c_2$ داشته‌باشیم:

$$g(c_1) = g(c_2) = a$$

$$P_\theta[T^*(X) = a] > 0 \quad (\text{برای دست کم یک } \theta)$$

$$P[T(X) = c_i | T^*(X) = a] > 0 \qquad i = 1, 2$$

اکنون تابع $h$ را اینگونه تعریف می‌کنیم:

$$h(c) = \begin{cases} \frac{(-1)^i}{P_\theta[T(X)=c_i|T^*(X)=a]} & c = c_i, i = 1,2 \text{ اگر} \\ 0 & \text{سایر نقاط} \end{cases}$$

داریم:

$$\begin{aligned} E[h(T(X))] &= \sum_{i=1}^{2} \frac{(-1)^i}{P_\theta[T(X) = c_i|T^*(X) = a]} \\ &\times P[T(X) = c_i] \\ &= (-1+1) \times P[T^*(X) = a] \\ &= 0 \end{aligned}$$

(دقت کنید که پیشامد $T(X) = c_i$ زیر مجموعه پیشامد $T^*(X) = a$ می‌باشد). پس برای هر $\theta$، $E_\theta h(T) = 0$. اما طبق فرض برای دست‌کم یک $\theta$، $P_\theta[T^*(X) = a] > 0$ و لذا مثلاً $P_\theta[T(X) = c_1] > 0$، که در این حالت $h(T(X)) = h(c_1) \neq 0$ و در نتیجه $T$ کامل نیست.

**مقایسه سه روش اثبات:** روش اول، ساده‌تر از سایر روشهاست و از این نظر بهترین است. در روش دوم، از شیوه زیبایی استفاده شده است که همانا بکارگیری موقت (کاتالیزور وار) تابع $L(T)$ است؛ به منظور سود جستن از کامل بودن $T$ و استفاده از قضیه رائو–بلکول. البته طولانی بودن و اتکاء به قضیه رائو–بلکول می‌تواند دو نقطه ضعف این روش تلقی شود. روش سوم کمی پیچیده است و ازنظر سادگی و روانی به روش دوم و به ویژه به روش اول نمی‌رسد. اما دو نکته است که مزیت روش سوم می‌باشد:

**الف)** در روش سوم فرض نمی‌شود که آماره بسنده مینیمال وجود دارد. اما در روش اول باید وجود آنرا فرض کنیم. گرچه تحت شرایطی ضعیف آماره بسنده مینیمال همیشه وجود دارد [۴–۱۲]، ولی به هرحال این یک پیش‌فرض است که در روش سوم احتیاجی به آن نیست.

**ب)** روش سوم تقریباً یک روش ساختاری است. به این بیان که در آن، مستقیماً از تعریف کامل بودن استفاده شده و تابعی که در شرط دلخواه صدق می‌کند ساخته‌می‌شود. این، به ویژه از دید ریاضیدان شهودگرا، یک مزیت است. (البته از دید چنین ریاضیدانی پای استدلال سوم چوبین است: شهودگرایان، و به تعبیر دیگر ساختارگرایان، به معادل بودن گزارهٔ $\sim A \Rightarrow \sim B$ با گزارهٔ $B \Rightarrow A$ اعتقاد ندارند!)

## ۳. چند بحث تکمیلی

**۱.۳** مینیمال بودن، به خودی خود تضمین نمی‌کند که یک آماره شامل اطلاعات فرعی و غیر لازم درباره $\theta$ نباشد (کامل

بودن چنین است). مثالهای فراوانی وجود دارند که می‌توانیم از یک آماره بسنده مینیمال، تابعی بسازیم که آماره فرعی باشد، یعنی توزیع آن به $\theta$ بستگی نداشته‌باشد. این مثالها نشان می‌دهند که:

لزوماً هر آماره بسنده مینیمال، کامل نیست و لذا عکس قضیه بهادر همواره برقرار نیست. نتیجه آنکه، بسنده کامل بودن یک شرط قوی‌تر از بسنده مینیمال بودن است و البته در استنباط آماری یک مفهوم بسیار کارساز می‌باشد [۱۰و۱۱].

برای توضیح مطلب، در ادامه چند مثال ارایه شده است که در آنها آماره بسنده مینیمال، کامل نیست.

**مثال ۱.** سکه نااریبی را $\theta$بار می‌ریزیم و $\theta \in \{1, 2, \ldots\}$. به آسانی می‌توان تحقیق کرد که $X$ :تعداد شیرها، یک آماره بسنده مینیمال برای $\theta$ است. اما تابع زیر که امید ریاضی آن همواره صفر است، نشان می‌دهد که $X$ کامل نیست.

$$T(X) = \begin{cases} 1 & \text{اگر } X \text{ زوج باشد} \\ 0 & \text{اگر } X \text{ فرد باشد} \end{cases}$$

**مثال ۲.** فرض کنید $X_1, \ldots, X_n$، نمونه‌ای تصادفی از توزیع لجستیک زیر باشد

$$f(x;\theta) = \frac{e^{-(x-\theta)}}{[1 + e^{-(x-\theta)}]^2} \ , \quad \theta \in R$$

در اینجا $T(X) = (X_{(1)}, \ldots, X_{(n)})$ آماره بسنده مینیمال است، در حالیکه مثلاً $S(T) = X_{(n)} - X_{(1)}$ یک آماره فرعی است.

**مثال ۳.** فرض کنید $X_1, \ldots, X_n$، نمونه‌ای تصادفی از توزیع یکنواخت $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$ باشد. در اینجا $T(X) = (X_{(1)}, X_{(n)})$ آماره بسنده مینیمال است اما کامل نیست، زیرا مثلاً $E_\theta[X_{(n)} - X_{(1)} - \frac{n-1}{n+1}] = 0 \ \forall\theta$. حتی در حالت $n = 1$ نیز $X$ کامل نیست. جالب آنکه در این حالت $X$، برآوردگر نااریب با کمترین واریانس (UMVUE) برای $\theta$ نیست و اصولاً در این حالت UMVUE برای $\theta$ وجود ندارد. برای جزئیات به [۷] صفحه ۱۶۶ مراجعه کنید.

**۲.۳** بحثی که در بخش اول ارایه کردیم به فهم شهودی قضیهٔ باسو نیز کمک می‌کند. چنانچه آماره‌ای بسنده کامل باشد، بدین معنی است که علاوه بر داشتن همه اطلاعات لازم درباره $\theta$، هیچ اطلاعی بیشتر از آنچه درباره $\theta$ لازم است ندارد. پس چنین آماره‌ای ارتباطی با یک آماره فرعی (که شامل هیچ اطلاعی درباره $\theta$ نیست) نمی‌تواند داشته باشد. از این رو طبیعی به نظر می‌رسد که هر آماره فرعی مستقل از یک آماره بسنده کامل باشد.

درباره آماره کامل و قضیهٔ باسو و مباحث مربوط به آنها، مقاله [۹] حاوی نکات ارزنده‌ای است. همچنین در مقاله جالب و روان [۳] مباحث مختلف مربوط به آماره فرعی بررسی شده‌است که برای خواننده علاقمند، می‌تواند مفید باشد.

# مراجع

[۱] بهبودیان، جواد؛ آمار ریاضی، انتشارات امیرکبیر، ۱۳۷۰

[۲] پارسیان، احمد؛ مبانی آمار ریاضی، انتشارات دانشگاه صنعتی اصفهان، ۱۳۷۸

[۳] سیّاره، عبدالرضا؛ آماره‌های کمکی و نقش آنها در استنباط آماری، اندیشه آماری، سال سوم، شماره اول، ۱۳۷۷

[4] Bahadur, R. R. (1954) *Sufficiency and Statistical Decision Functions*, Ann. Math. Stat. 25: 423-462.

[5] Bahadur, R. R. (1957) *On Unbiased Estimates of Uniformly Minimum Variance*, Sankhya 18: 211-224.

[6] Basu, D. (1955) *On Statistics Independent of a Complete Sufficient Statistic*, Sankhya, 5: 377-380.

[7] Kiefer, J. C. (1987) *Introduction to Statistical Inference*, Springer.

[8] Lehman, E. L. and Scheffe, H. (1950) *Completeness, Similar regions, and Unbiased Estimation*, Sankhya 10: 305-340.

[9] Lehman, E. L. (1981) *An Interpretation of Completeness and Basu's Theorem*, JASA, 76: 335-340.

[10] Lehmann, E. L. (1983) *Theory of Point Estimation*, John Wiley.

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

[12] Pitcher, T. S. (1957) *Sets of Measures not Admitting Necessary and Sufficient Statistics or Subfields*, Ann. Math. Stat. 26: 267-268.

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