فصل ۱

آزمون‌های فرض (آزمون فرض‌ها)

۱.۱ مقدمه

در فصل هفتم روشی از استنباط را مطالعه کردیم که برآورد نقطه‌ای نامیده می‌شود. در این فصل به روش دیگر می‌پردازیم که آزمون فرض‌ها ناام دارد. همانند فصل پیش، این فصل نیز به دو بخش تقسیم می‌شود که نیاز به یافتن آن و ارزش دهی به آزمون‌های فرض‌ها را در برمی‌گیرد. با تعریف فرض آماری شروع می‌کنیم.

تعریف ۱.۱.۱. فرض آماری عبارتی دربازه‌ی پارامتر جامعه است.

تعریف فرض آماری نسبتاً جامع است، اما نکته‌ی مهم آن است که فرض آماری گزاره‌ای را در مورد جامعه می‌سازد. هدف آزمون فرض تصمیم‌گیری براساس نمونه‌ای از جامعه است که کدام یک از دو فرض مکمل درست است.

تعریف ۲.۱.۱. دو فرض مکمل در یک مساله‌ی آزمون فرض، فرض صفر و فرض مقابل خوانده می‌شوند که آن‌ها به ترتیب با H_0 و H_1 نمایش می‌دهیم.

اگر θ نمایش پارامتر جامعه باشد، ساختار کلی فرض صفر و فرض مقابل به صورت

$$H_0 : \theta \in \Theta_0 \quad \nu_s \quad H_1 : \theta \in \Theta_0^c$$

است که در آن Θ_0 زیرمجموعه‌ای از فضای پاراکتر و Θ_0^c متمم (مکمل) آن است. برای مثال، اگر θ ممیانگین تغییرات در فشارخون یک بیمار پس از مصرف دارو باشد، ممکن است علاقه‌متد به آزمون $H_0 : \theta = 0$ در مقابل $H_1 : \theta \neq 0$ باشیم. فرض صفر بیان می‌کند که به طور میانگین داروی مورد نظر تأثیری بر فشار خون ندارد، و فرض مقابل بیان می‌کند که تاثیری وجود دارد. این وضعیت رایج، که در آن H_0 بیانگر آن است که درمان موثر نیست به واژه‌ی فرض صفر [2]

[2] در یک مساله‌ی آزمون فرض، آزمایش کننده پس از مشاهده‌ی نمونه باید تصمیم بگیرد که H_0 را به عنوان یک گزاره‌ی درست بپذیرد یا آن را به عنوان یک گزاره نادرست رد کند و H_1 را بپذیرد.

منجر می‌شود. به عنوان مثال دیگر، یک مصرف کننده ممکن است علاقه‌مند باشد نسبت به تعداد محصولات نامنطبق که توسط یک کارخانه تولید می‌شود. اگر نسبت تعداد محصولات نامنطبق باشد، مصرف کننده ممکن است علاقه‌مند به آزمون $H_0 : \theta \geq \theta_0$ در مقابل $H_1 : \theta < \theta_0$ باشد. مقدار θ_0، نسبت ماکزیمم پذیرفته شده برای محصولات نامنطبق می‌باشد و H_0 بیان می‌کند که نسبت محصولات نامنطبق بسیار بالاست. مسایلی که در آن‌ها موضوع مورد اهمیت آزمون کیفیت یک محصول است، مسایل نمونه‌گیری برای پذیرش عنوان شده‌اند.

تعریف ۳.۱.۱. یک شیوه‌ی آزمون فرض یا آزمون فرض، قاعده‌ای است که مشخص می‌کند:

i) به ازای کدام مقادیر نمونه تصیمیم گرفته شود که فرض H_0 درست است.

ii) به ازای کدام مقادیر نمونه H_0 رد می‌شود و H_1 به عنوان گزاره‌ی درست پذیرفته می‌شود.

زیر مجموعه‌ای از فضای نمونه که بر آن H_0 رد می‌شود ناحیه‌ی بحرانی و متمم (مکمل) ناحیه‌ی بحرانی، ناحیه پذیرش نامیده می‌شود.

با یک نگاشت فلسفی، بعضی افراد نگران تمایز بین "رد کردن H_0" و "پذیرش H_1" هستند. در واژه‌ی اول، چیزی در مورد این که آزمایش کننده آن را بپذیرد وجود ندارد و فقط بیان می‌شود که H_0 باید رد شود. به طور مشابه، چنین تمایزی بین "پذیرش H_0" و "عدم رد H_0" نیز وجود دارد. عبارت اول مستلزم آن است که آزمایش کننده که تائید کند گزاره‌ای که در مشخص شده است، در حالی که عبارت دوم نشان می‌دهد آزمایش کننده واقعاً باور به H_0 ندارد اما شواهدی مبنی بر رد آن را نیز در اختیار ندارد. معمولاً ما نگران چنین موضوعی نخواهیم بود و نگرشی به مساله‌ی آزمون فرض را معرفی خواهیم کرد که در آن مساله‌ای با دو رویکرد در تصمیم‌گیری وجود دارد، رویکردهای تائید H_0 یا تائید H_1.

به طور معمول، یک آزمون فرض برحسب یک آماره‌ی آزمون $W(X) = W(X_1, ..., X_n)$ که تابعی از نمونه است مشخص می‌شود. برای مثال، شیوه‌ی آزمون ممکن است به این صورت باشد که H_0 رد می‌شود اگر $\bar{X}$، میانگین نمونه‌ای، بیشتر از ۳ باشد. در این حالت $W(X) = \bar{X}$ آماره‌ی آزمون است و ناحیه‌ی رد، مجموعه‌ی $\{(x_1, ..., x_n) : \bar{x} > 3\}$ است. در بخش ۲.۸، روش‌های انتخاب آماره‌ی آزمون و ناحیه‌های رد بررسی شده‌اند. معیارهایی برای ارزش‌گذاری آزمون‌ها در بخش ۳.۸ معرفی شده‌اند. همانند برآوردگرهای نقطه‌ای، روش‌های یافتن آزمون‌ها ضمانتی را تامین نمی‌کنند؛ آزمون‌ها قبل از آن‌که ارزشمند تلقی شوند تا به کار بیایند باید ارزیابی شوند.

۲.۱ روش‌های یافتن آزمون‌ها

به جزئیات چهار روش یافتن شیوه‌های آزمون که در موقعیت‌های مختلف مفید هستند و نسبت به جنبه‌های گوناگون مساله ؟؟؟ دارند پرداخته‌ایم. با یک روش بسیار کلی شروع می‌کنیم که همیشه کاربردی است و نیز در بعضی حالت‌ها ویژگی‌های بهینه دارد.

۱.۲.۱ آزمون‌های نسبت درستنمایی

شیوه‌ی آزمون فرض نسبت درستنمایی مربوط به برآوردگری ماکسیمم درستنمایی است که در بخش ۲.۲.۷ بحث شده است، و به همان اندازه کاربرد دارد. یادآوری می‌کنیم اگر $X_۱, ..., X_n$ یک نمونه‌ی تصادفی از جامعه‌ای با تابع چگالی احتمال یا تابع جرم احتمال $f(x \mid \theta)$ (θ ممکن است بردار باشد) باشد، تابع درستنمایی به صورت زیر تعریف می‌شود:

$$(L\theta \mid x_۱, ..., x_n) = L(\theta \mid x) = f(x \mid \theta) = \prod_{i=۱}^{n} f(x_i \mid \theta).$$

فرض کنیم Θ تمام فضای پارامتر باشد. آزمون نسبت درستنمایی به صورت زیر تعریف می‌شود:

تعریف ۱.۲.۱. آماره‌ی آزمون نسبت درستنمایی برای آزمون $H_۰ : \theta \in \Theta_۰$ در مقابل $H_۱ : \theta \in \Theta_۰^c$ به صورت زیر است:

$$\lambda(x) = \frac{\sup_{\Theta_۰} L(\theta \mid x)}{\sup_{\Theta} L(\theta \mid x)}$$

یک آزمون نسبت درستنمایی *(LRT)* شیوه آزمونی است که ناحیه رد آن، $\{x : \lambda(x) \leq c\}$ باشد که در آن $۰ < c \leq ۱$.

مفهوم LRT بیش از همه ممکن است در موقعیتی درک شود که $f(x \mid \theta)$ تابع جرم احتمال یک متغیر تصادفی گسسته است. در این حالت، صورت کسر $\lambda(x)$ ماکسیمم احتمال نمونه‌ی مشاهده شده است که روی پارامترهای مربوط به فرض صفر مجاسبه می‌شود. (تمرین ۴.۸ را ببینید.) مخرج کسر $\lambda(x)$ ماکسیمم احتمال نمونه‌ی مشاهده شده روی تمام فضای پارامتر است. نسبت این دو ماکسیمم کوچک خواهد بود اگر به ازای مقدارهای پارامتر در فرض مقابل، نمونه‌ی مشاهده شده شانس بیشتری برای وقوع نسبت به همه‌ی مقدارهای پارامتر در فرض صفر داشته باشند. در چنین موقعیتی معیار LRT بیانگر آن است که $H_۰$ باید رد شود و $H_۱$ به عنوان گزاره‌ی درست تایید شود. روش‌های انتخاب عدد در بخش ۳.۸ بحث خواهد شد.

اگر به مساله‌ی ماکسیمم سازی روی فضای پارامتر (ماکسیمم سازی بدون محدودیت) و روی زیرمجموعه‌ای از آن (ماکسیمم سازی با محدودیت) تمرکز کنیم، تناظر بین LRT ها و MLE ها روشن‌تر می‌گردد. فرض کنید $\hat{\theta}$، برآوردگر ماکسیمم در تنهایی θ، وجود داشته باشد، آن‌گاه $\hat{\theta}$ با ماکسیمم سازی بدون محدودیت $L(\theta \mid x)$ حاصل می‌شود. همچنین می‌توانیم برآوردگر در تنهایی ماکسیمم θ را روی فضای پارامتر Θ_0 با ماکسیمم سازی با محدودیت به دست آورده، آن را با $\hat{\theta}_0$ نمایش دهیم. یعنی $\hat{\theta} = \hat{\theta}_0(x)$ مقداری از $\theta \in \Theta_0$ است که $L(\theta \mid x)$ را ماکسیمم می‌کند. در این صورت، آماره‌ی LRT به صورت زیر است:

$$\lambda(x) = \frac{L(\hat{\theta}_0 \mid x)}{\hat{\theta} \mid x}$$

مثال ۲.۲.۱. (*LRT* نرمال) فرض کنیم $X_1, ..., X_n$ یک نمونه‌ی تصادفی از جامعه‌ی $N(\theta, 1)$ باشد. آزمون $H_0 : \theta = \theta_0$ در مقابل $H_1 : \theta \neq \theta_0$ را در نظر بگیرید. در این‌جا θ_0 یک عدد ثابت است که قبل از انجام آزمایش توسط محقق مشخص می‌شود. با توجه به آن که فقط یک مقدار از θ توسط H_0 مشخص شده است، صورت کسر $\lambda(x)$، $L(\theta_0 \mid x)$ خواهد بود. در مثال ۵.۲.۷، *MLE* (بدون محدودیت) پارامتر θ، میانگین نمونه‌ای $\bar{X}$ حاصل شده است. بنابراین صورت کسر $\lambda(x)$ برابر $L(\bar{x} \mid x)$ خواهد بود. بنابراین آماره‌ی *LRT* به صورت زیر است:

$$\lambda(x) = \frac{(2\pi)^{-\frac{n}{2}} \exp[-\sum_{i=1}^{n}(x_i - \theta_0)^2/2}{(2\pi)^{-\frac{n}{2}} \exp[-\sum_{i=1}^{n}(x_i - \bar{x})^2/2} = \exp[(-\sum_{i=1}^{n}(x_i - \theta_0)^2 + \sum_{i=1}^{n}(x_i - \bar{x})^2)/2]. \tag{1.1}$$

عبارت $\lambda(x)$ با توجه به عبارت زیر ساده می‌شود:

(UIT) استفاده کنیم. یک دلیل آن است که *(UIT)* برای هر $\theta \in \Theta_0$ دارای احتمال خطای نوع اول کوچک‌تر است. به علاوه، اگر H_0 رد شود، ممکن است بخواهیم که آزمون‌های انفرادی $H_{0\gamma}$ را بررسی کنیم که چرا تاکنون در مورد استنباط‌های بر پایه آزمون‌های فردی بحث نکرده‌ایم. احتمال‌های خطا برای چنین استنباط‌هایی باید بررسی شود قبل از آن که یک رویه برای استنباطی نهایی شود. اما امکان به دست آوردن اطلاعات اضافی با بررسی $H_{0\gamma}$های انفرادی، به جای بررسی *(LRT)* کلی آشکار است.

حال اندازه‌ی *(IUT)* ها را بررسی می‌کنیم. یک کران ساده برای اندازه‌ی یک *(IUT)* به اندازه‌های آزمون‌های انفرادی مرتبط است که در تعریف *(IUT)* به کار گرفته شده‌اند. به یاد بیاورید که در چنین موقعیتی فرض صفر به صورت اجتماع بیان می‌شود، یعنی آزمون زیر را انجام می‌دهیم، $H_0 : \theta \in \Theta_0$ در مقابل $H_1 : \theta \in \Theta_0^c$ که در آن $\Theta_0 = \bigcup_{\gamma \in \Gamma} \Theta_\gamma$

یک *(IUT)* دارای ناحیه‌ی بحرانی به شکل $R = \bigcap_{\gamma\in\Gamma} R_\gamma$ است که R_γ ناحیه‌ی رد آزمون $H_{\circ\gamma} : \theta \in \Theta_\gamma$ می‌باشد.

قضیه ۳.۲.۱. فرض کنیم α_γ اندازه‌ی آزمون $H_{\circ\gamma}$ با ناحیه‌ی بحرانی R_γ باشد. در این صورت *(IUT)* با ناحیه‌ی رد $R = \bigcap_{\gamma\in\Gamma} R_\gamma$ یک آزمون در سطح $\alpha = \sup_{\gamma\in\Gamma} \alpha_\gamma$ است.

برهان. فرض کنیم $\theta \in \Theta_\circ$. در این صورت برای γای، $\theta \in \Theta_\gamma$ و $P_\theta(X \in R) \leq P_\theta(X \in R_\gamma) \leq \alpha_\gamma \leq \alpha$ چون $\theta \in \Theta_\circ$ دلخواه است، (IUT) آزمونی در سطح α است. $\square$

معمولا، ناحیه‌های بحرانی انفرادی R_γ به گونه‌ای انتخاب می‌شوند که برای هر γ، $\alpha_\gamma = \alpha$. در چنین حالتی، قضیه‌ی ۸.۳.۲۳ بیان می‌کند که (IUT) حاصل در سطح α است.

قضیه‌ی ۸.۳.۲۸ که کران بالایی برای اندازه‌ی یک (IUT) را فراهم می‌کند، گاهی مفیدتر از قضیه‌ی ۸.۳.۲۱ است که کران بالایی برای اندازه‌ی یک (UIT) را در اختیار قرار می‌دهد. قضیه‌ی ۸.۳.۲۱ فقط برای (UIT) هایی به کار می‌رود که از آزمون‌های نسبت درستنمایی ساخته می‌شوند. در مقابل، قضیه‌ی ۸.۳.۲۳ برای هر (IUT) به کار می‌رود.

کران قضیه‌ی ۸.۳.۲۱ اندازه‌ی (LRT) است که در یک مساله پیچیده ممکن است محاسبه‌ی آن مشکل باشد. در قضیه‌ی ۸.۳.۲۳، (LRT) نیازی نیست در به دست آوردن کران بالا به کار رود. هر آزمون $H_{\circ\gamma}$ با اندازه‌ی شناخته شده α_γ می‌تواند به کار گرفته شود و سپس کران بالای اندازه‌ی (IUT) برحسب اندازه‌های مشخص و معلوم α_γ برای $\gamma \in \Gamma$ حاصل خواهد شد.

(IUT) در قضیه‌ی ۸.۳.۲۳ یک آزمون در سطح α است. اما اندازه‌ی (IUT) ممکن است خیلی کمتر از α باشد. (IUT) ممکن است بسیار محتاط عمل کند. قضیه‌ی زیر شرایطی را در اختیار می‌گذارد که تحت آن‌ها اندازه‌ی (IUT) دقیقاً α است و (IUT) چندان محافظه کار نیست.

قضیه ۴.۲.۱. آزمون $H_\circ : \theta \in \bigcup_{j=1}^{k} \Theta_j$ را در نظر بگیرید که در آن k عدد صحیح مثبت متناهی است. برای هر $j = 1, \ldots, k$، فرض کنیم R_j ناحیه‌ی بحرانی آزمون در سطح α برای $H_{\circ j}$ باشد. فرض کنید برای iای از $1, \ldots, k$، یک دنباله از نقطه‌های پارامتری وجود داشته باشد که $\theta_l \in \Theta_i, l = 1, 2, \ldots$ به گونه‌ای که،

i) $\lim_{l\to\infty} P_{\theta_l}(X \in R_i) = \alpha$،

ii) برای هر $j = 1, \ldots, k$، $j \neq i$، $\lim_{l\to\infty} P_{\theta_l}(X \in R_j) = 1$

در این صورت، *(IUT)* به ناحیه‌ی رد $R = \bigcap_{j=1}^{k} R_j$ آزمون با اندازه‌ی α است.

برهان. از قضیه‌ی ۲۳.۳.۸، R یک آزمون در سطح α است، یعنی،

$$\sup_{\theta\in\Theta_\circ} P_\theta(X\in R)\le \alpha \tag{۲.۱}$$

اما، چون همه‌ی نقطاط پارامتری θ_l در $\Theta_\circ \subset \Theta_i.$ $\theta_l \in$ صدق می‌کنند، خواهیم داشت:

$$\begin{aligned}
\sup_{\theta\in\Theta_\circ} P_\theta(X\in R) &\ge \lim_{l\to\infty} P_{\theta_l}(X\in R)\\
&= \lim_{l\to\infty} P_{\theta_l}(X\in \bigcap_{j=۱}^{k} R_j)\\
&\ge \lim_{l\to\infty} \sum_{j=۱}^{k} P_{\theta_l}(X\in R_j) - (k-۱) \qquad \text{نامساوی بن فروض}\\
&= (k-۱)+\alpha-(k-۱) \qquad (\text{از } (i), (ii))\\
&= \alpha
\end{aligned}$$

این و (۷.۳.۸) مستلزم آن است که آزمون دارای اندازه‌ی دقیق α است. □