به نام زندگی

در بحث MVUE ، دیدیم که برای یافتن تخمین‌گر MVUE نیاز به داشتن یک آماره

کافی کامل داریم یا به بیان دیگر مجموعه خانواده توزیع‌های $\{P_\theta \, ; \, \theta \in \Lambda\}$ یک

مجموعه خانواده توزیع‌های کامل است. بسیاری از توابع احتمالی که با آنها سروکار داریم

کامل هستند اما علاقه‌مندیم روش‌های تخمینی را بررسی کنیم که با وجود اطلاعات

اولیه محدود از پارامتر مورد تخمین، بتوانند جواب تخمین مناسبی را روی هر تابع احتمالی

به دست بدهند.

قبل از اینکه به این بحث بپردازیم، می خواهیم که با استفاده از بحث های قبل، یک کران برای واریانس تخمین گرهای مختلف (چه بدون بایاس، چه دارای بایاس با استفاده از هر روش تخمین) به دست بیاوریم. به این کران، کران نامساوی اطلاعات می گوییم

قضیه نامساوی اطلاعات

فرض کنیم $\hat{\theta}(X)$ یک تخمین دلخواه از θ روی مجموعه $\{P_\theta ; \theta \in \Lambda\}$ باشد و شرایط زیر برقرار باشند.

۱- مجموعه‌ی Λ یک مجموعه (یا بازه) باز باشد.

۲- در مجموعه $\{P_\theta \,;\, \theta \in \Lambda\}$ همه‌ی اعضا ($P_\theta(y)$) دارای مجموعه پشتیبان S (support set) یکسان باشند.

یادآوری: مجموعه پشتیبان یا support set برای یک تابع چگالی احتمال $f(y)$ به صورت زیر تعریف می‌شود

$$S = \{ y \mid f(y) > 0 \}$$

۳- عبارت $\frac{\partial}{\partial \theta} P_\theta(y)$ وجود داشته باشد و محدود باشد.

$$\forall \theta \in \Lambda \quad , \quad \forall y \in S$$

۴- عبارت $\frac{\partial}{\partial\theta}\int h(y)\,p_\theta(y)\,dy$ وجود داشته باشد و دانسته باشیم

$$\frac{\partial}{\partial\theta}\int h(y)\,p_\theta(y)\,dy = \int h(y)\,\frac{\partial}{\partial\theta}p_\theta(y)\,dy \qquad \forall\theta\in\Lambda$$

* برای $\begin{cases} h(y)=1 \\ h(y)=\hat{\theta}(y) \end{cases}$

در این صورت داریم،

$$Var_\theta(\hat{\theta}(y)) \geq \frac{\left[\frac{\partial}{\partial\theta} E_\theta\{\hat{\theta}(y)\}\right]^2}{I_\theta}$$

نامساوی اطلاعات

که در آن

$$I_\theta \triangleq E_\theta\left\{\left(\frac{\partial}{\partial\theta}\log p_\theta(y)\right)^2\right\}$$

I_θ را تابع اطلاعات فیشر گوییم.

Fisher's Information Function

اثبات : به عنوان تمرین برای مطالعه (احتمالا)

$$\mathrm{Var}_\theta\left(\hat{\theta}(y)\right) I_\theta \geq \left[\frac{\partial}{\partial\theta} E_\theta\left\{\hat{\theta}(y)\right\}\right]^2$$

$$E_\theta\left\{\Big(\underbrace{\hat{\theta}(y) - E_\theta\{\hat{\theta}(y)\}}\Big)^2\right\} E_\theta\left\{\Big(\underbrace{\frac{\partial}{\partial\theta}\log p_\theta(y)}\Big)^2\right\} \geq \left[\frac{\partial}{\partial\theta} E_\theta\left\{\hat{\theta}(y)\right\}\right]^2$$

یادآوری نامساوی شوارتز :

$$\left|\int f(x)\, g^*(x)\, dx\right|^2 \leq \int |f(x)|^2 dx \int |g(x)|^2 dx$$

$$\left|\langle f(x), g(x)\rangle\right|^2 \leq \langle f(x), f(x)\rangle \langle g(x), g(x)\rangle$$

$\exists c ;\ f(x) = c\, g(x)$ تساوی زمانی برقرار است که دو تابع متناسب باشند

$$|\underbrace{E\{xy^*\}}_{\text{ضرب داخلی x, y}}|^2 \le \underbrace{E|x|^2}_{\text{ضرب داخلی x در خودش}} \; \underbrace{E|y|^2}_{\text{ضرب داخلی y در خودش}}$$

$$E^2 x \le E x^2$$

فرم ساده نامساوی شوارتز
برای حالت حقیقی

$$|E x|^2 \le E|x^2|$$

برای حالت مختلط

$$E_\theta\{\dots\} \equiv \underbrace{E_{y|\theta}\{\dots\}}_{\text{با فرتاسیدن فرآیند}}$$

$$Var_\theta(\dots) \equiv \underbrace{Var_{y|\theta}(\dots)}_{\text{فرتاسیدن فرآیند}}$$

در قضیه نامساوی اطلاعات اگر شرط پنجم به صورت زیر برقرار باشد

۵- عبارت $\frac{\partial^2}{\partial\theta^2} P_\theta(y)$ وجود داشته باشد $\forall\theta\in\Lambda,\ \forall y\in S$

و داشته باشیم

$$\frac{\partial^2}{\partial\theta^2}\int P_\theta(y)\,dy = \int \frac{\partial^2}{\partial\theta^2} P_\theta(y)\,dy$$

در این صورت I_θ تابع اطلاعات فیشر به صورت زیر نیز قابل محاسبه است.

$$I_\theta = -E_\theta\left\{\frac{\partial^2}{\partial\theta^2}\log P_\theta(Y)\right\}$$

(اثبات برای مطالعه)

• نامساوی اطلاعات برای یک حالت خاص (حالتی که تخمین گر $\hat{\theta}(y)$ بدون بایاس باشد)

اگر تخمین گر $\hat{\theta}(y)$ بدون بایاس باشد یعنی

$$E_\theta\{\hat{\theta}(y)\} = \theta$$

در این صورت نامساوی اطلاعات به شکل ساده ی زیر قابل بیان خواهد بود،

$$\underbrace{\mathrm{Var}_\theta(\hat{\theta}(y)) \geq \frac{1}{I_\theta}}_{\text{کران کرامر- رائو}}$$

Cramer-Rao Lower Bound (CRLB)

زیرا $\frac{\partial}{\partial\theta} E_\theta\{\hat{\theta}(y)\} = 1$

* ناتساوی اطلاعات برای خانواده توزیع‌های نمایی

در این بخش می‌خواهیم نامساوی اطلاعات را برای خانواده توزیع‌های نمایی بررسی کنیم. هدف این است که به یک شرط لازم و کافی برای رسیدن به کران نامساوی اطلاعات برسیم. به عبارت دیگر می‌خواهیم یک شرط لازم و کافی برای اینکه نامساوی اطلاعات در حالت تساوی برقرار باشد، پیدا کنیم.

فرض کنیم

$$P_\theta(y) = c(\theta)\exp\big(g(\theta)T(y)\big)\,h(y)$$

$$\forall \theta \in \Lambda \,, \quad \forall y \in S$$

که در آن توابع $c(\cdot)$ ، $T(\cdot)$ ، $h(\cdot)$ و $g(\cdot)$ توابع حقیقی هستند و

$g(\theta)$ یک تابع مشتق پذیر نسبت به θ با مشتق $g'(\theta)$ باشد و

$$E_\theta\{|T(y)|\} < \infty \qquad (**)$$

و

$$\frac{\partial}{\partial\theta}\int e^{g(\theta)T(y)}\,h(y)\,dy = \int \frac{\partial}{\partial\theta} e^{g(\theta)T(y)}\,h(y)\,dy$$

در این صورت شرایط 1 تا 4 در قضیه اساسی اطلاعات برقرار است.

با توجه به اینکه می دانیم $\int_{y\in\Gamma} P_\theta(y)\,dy = 1$ همواره برقرار است، خواهیم داشت

$$\int_{\Gamma} c(\theta)\, e^{g(\theta)T(y)}\, h(y)\, dy = 1$$

$$\Longrightarrow \quad c(\theta) = \left[\int_{\Gamma} e^{g(\theta)T(y)}\, h(y)\, dy \right]^{-1} \qquad (*)$$

برای بررسی کران نامساوی اطلاعات روی خانواده توزیع های نمایی ، ابتدا I_θ تابع اطلاعات فیشر را محاسبه می کنیم.

$$I_\theta \triangleq E_\theta\left\{\left(\frac{\partial}{\partial\theta}\log p_\theta(y)\right)^2\right\} = -E_\theta\left\{\frac{\partial^2}{\partial\theta^2}\log p_\theta(y)\right\}$$

اگر شرط (۵) برقرار باشد

می دانیم که

$$p_\theta(y) = c(\theta)\, e^{g(\theta)T(y)}\, h(y)$$

$$\Longrightarrow \log P_\theta(y) = \log c(\theta) + g(\theta)T(y) + \log h(y)$$

$$\xRightarrow[(*)]{c(\theta)} \log P_\theta(y) = g(\theta)T(y) + \log h(y) - \log \underbrace{\int e^{g(\theta)T(y)} h(y)\,dy}_{u(\theta) = \frac{1}{c(\theta)}}$$

$$\overbrace{\phantom{\int e^{g(\theta)T(y)} h(y)\,dy}}^{u'(\theta)}$$

$$\xRightarrow{\frac{\partial}{\partial\theta}} \frac{\partial}{\partial\theta} \log P_\theta(y) = g'(\theta)T(y) - c(\theta) \underbrace{\frac{\partial}{\partial\theta} \int e^{g(\theta)T(y)} h(y)\,dy}_{\int \frac{\partial}{\partial\theta} e^{g(\theta)T(y)} h(y)\,dy}$$

$$\frac{\partial}{\partial\theta} \log u(\theta) = \frac{u'(\theta)}{u(\theta)}$$

$$\frac{1}{u(\theta)} = c(\theta)$$

$$\Longrightarrow \frac{\partial}{\partial\theta} \log P_\theta(y) = g'(\theta)T(y) - c(\theta)\int g'(\theta)T(y)\,e^{g(\theta)T(y)}h(y)\,dy$$

$$\Longrightarrow \frac{\partial}{\partial\theta} \log P_\theta(y) = g'(\theta)T(y) - g'(\theta)\underbrace{\int T(y)\,\underbrace{c(\theta)\,e^{g(\theta)T(y)}h(y)}_{P_\theta(y)}\,dy}_{E_\theta\{T(y)\}}$$

$$\Longrightarrow \frac{\partial}{\partial\theta} \log P_\theta(y) = g'(\theta)\left[T(y) - E_\theta\{T(y)\}\right]$$

می دانیم

$$I_\theta = E_\theta \left\{ \left(\frac{\partial}{\partial \theta} \log p_\theta(y) \right)^2 \right\}$$

$$\Rightarrow \quad I_\theta = \underbrace{E_\theta}_{y|\theta} \left\{ (g'(\theta))^2 \left[T(y) - E_\theta\{T(y)\} \right]^2 \right\}$$

$$\Rightarrow \quad I_\theta = (g'(\theta))^2 \, E_\theta \left\{ \left[T(y) - E_\theta\{T(y)\} \right]^2 \right\}$$

$$\Rightarrow \quad I_\theta = (g'(\theta))^2 \, Var_\theta(T(y))$$

بنابراین نامساوی اطلاعات برای هر تخمین‌گر دلخواه $\hat{\theta}(y)$ روی مجموعه خانواده

توزیع‌های نمایی $\{P_\theta \,;\, \theta \in \Lambda\}$ که در آن $P_\theta(y) = c(\theta)\, e^{g(\theta)T(y)}\, h(y)$

به صورت زیر خواهد بود (شرایط گفته شده نیز برقرار هستند)

$$\mathrm{Var}_\theta\left(\hat{\theta}(y)\right) \geqslant \frac{\left[\frac{\partial}{\partial\theta} E_\theta\{\hat{\theta}(y)\}\right]^2}{\left(g'(\theta)\right)^2 \mathrm{Var}_\theta\left(T(y)\right)}$$

به دنبال پیدا کردن یک شرط لازم و کافی برای برقراری تساوی در نامساوی اطلاعات هستیم.

اگر تخمین گر $\hat{\theta}(y)$ را برابر $T(y)$ درنظر بگیریم، داریم.

$$\hat{\theta}(y) = T(y)$$

و برای محاسبه‌ی کران نامساوی اطلاعات، داریم

$$\frac{\partial}{\partial \theta} E_\theta \{ \hat{\theta}(y) \} = \frac{\partial}{\partial \theta} E_\theta \{ T(y) \}$$

$$= \frac{\partial}{\partial \theta} \underbrace{\int T(y) P_\theta(y) \, dy}$$

$$E_\theta\{T(y)\} = \int T(y)\, P_\theta(y)\, dy$$

$$= \int T(y)\, c(\theta)\, e^{g(\theta)T(y)}\, h(y)\, dy$$

$$\overset{(*)}{\underset{\uparrow\, c(\theta)}{=}} \frac{\int T(y)\, e^{g(\theta)T(y)}\, h(y)\, dy}{\int e^{g(\theta)T(y)}\, h(y)\, dy} = E_\theta\{T(y)\}$$

$$\frac{\partial}{\partial\theta} E_\theta\{T(y)\}$$

$$\frac{\partial}{\partial\theta}\frac{f(\theta)}{u(\theta)} = \frac{f'(\theta)\,u(\theta) - u'(\theta)\,f(\theta)}{[u(\theta)]^2}$$

$$u(\theta) = \frac{1}{c(\theta)}$$

$$\Longrightarrow \frac{\partial}{\partial \theta} E_\theta \{ \hat{\theta}(y) \} = \frac{\partial}{\partial \theta} E_\theta \{ T(y) \}$$

$$\frac{\partial}{\partial \theta} E_\theta \{ T(y) \} = [c(\theta)]^2 \Big[\frac{1}{c(\theta)} \int g'(\theta)\, T^2(y)\, e^{g(\theta) T(y)}\, h(y)\, dy$$

$$- f(\theta) \int g'(\theta)\, T(y)\, e^{g(\theta) T(y)}\, h(y)\, dy \Big]$$

$$\Longrightarrow \frac{\partial}{\partial \theta} E_\theta \{ T(y) \} = g'(\theta) \int T^2(y)\, \underbrace{c(\theta)\, e^{g(\theta) T(y)}\, h(y)}_{p_\theta(y)}\, dy$$

$$- \Big(\int T(y)\, \underbrace{c(\theta)\, e^{g(\theta) T(y)}\, h(y)\, dy}_{p_\theta(y)} \Big)^2$$

$$\Longrightarrow \frac{\partial}{\partial\theta} E_\theta\{T(y)\} = g'(\theta)\underbrace{\left[E_\theta\{T^2(y)\} - E_\theta^2\{T(y)\}\right]}_{Var_\theta(T(y))}$$

$$\Longrightarrow \frac{\partial}{\partial\theta} E_\theta\{\hat{\theta}(y)\} = \frac{\partial}{\partial\theta} E_\theta\{T(y)\} = g'(\theta)\, Var_\theta(T(y))$$

به این ترتیب در نامساوی اطلاعات برای خانواده توزیع‌های نمایی و به ازای تخمین‌گر $\hat{\theta}(y)=T(y)$ داریم.

$$\frac{\left[\frac{\partial}{\partial\theta} E_\theta\{T(y)\}\right]^2}{(g'(\theta))^2 \, Var_\theta(T(y))} = \frac{\cancel{(g'(\theta))^2}\,\left(Var_\theta(T(y))\right)^2}{\cancel{(g'(\theta))^2}\, Var_\theta(T(y))} = Var_\theta(T(y))$$

$$= Var_\theta(\hat{\theta}(y))$$

بنابراین با شرایط گفته شده، نامساوی اطلاعات در حالت تساوی برقرار می‌شود.

$\Longleftarrow$ در خانواده توزیع‌های نمایی به فرم $P_\theta(y) = c(\theta)\, e^{g(\theta)T(y)}\, h(y)$ در صورت برقراری شرایط (**)، اگر تخمین‌گر $\hat{\theta}(y) = T(y)$ را در نظر بگیریم، به کران نامساوی اطلاعات می‌رسیم. یعنی به کمترین مقدار واریانس تخمین می‌رسیم. (شرط کافی رسیدن به کران نامساوی اطلاعات)

به عبارت دیگر تاکنون توانستیم شرط کافی برای رسیدن به کران نامساوی اطلاعات را به
دست بیاوریم.

(توجه به این نکته لازم است که $T(Y)$ ای که به شرط کافی به دست آوردیم، اگر بدون
بایاس باشد، همان MVUE است زیرا به کمترین واریانس تخمین رسیده است)

برای کامل شدن مسئله، باید شرط لازم برای رسیدن به کران نامساوی اطلاعات
را نیز به دست بیاوریم. یعنی کران نامساوی اطلاعات را در حالت تساوی بنویسیم و
تساوی را تحلیل کنیم و ببینیم تحت چه شرایطی، تساوی برقرار است. این شرایط همان

شرط لازم برای رسیدن به کران نامساوی اطلاعات است.