به نام زندگی

شرط لازم و کافی برای تساوی در نامساوی اطلاعات

- شرط کافی : در خانواده توزیع نمایی $P_\theta(y) = c(\theta)e^{g(\theta)T(y)}\,h(y)$ با شرایط گفته شده،

اگر $\hat{\theta}(y) = T(y)$ در نظر بگیریم به کران نامساوی اطلاعات (کمترین مقدار واریانس تخمین) می‌رسیم.

در ادامه می‌خواهیم شرط لازم برای رسیدن به کران نامساوی اطلاعات را نیز به دست بیاوریم.

بنابراین فرض می کنیم که کران نامساوی اطلاعات دستیابی برقرار است و نتایج حاصل را بررسی می کنیم.

به یاد داریم که نامساوی اطلاعات به صورت زیر است،

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

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

$$\Rightarrow \left[\frac{\partial}{\partial\theta} E_\theta\{\hat{\theta}(y)\}\right]^2 \leqslant E_\theta\Big\{\underbrace{\left(\hat{\theta}(y) - E_\theta\{\hat{\theta}(y)\}\right)^2}_{f(x)}\Big\} E_\theta\Big\{\underbrace{\left(\frac{\partial}{\partial\theta}\log p_\theta(y)\right)^2}_{g(x)}\Big\}$$

تساوی در نامساوی شوارتز

$\Longrightarrow$

$g(x) = k\, f(x)$

تساوی در کران نامساوی اطلاعات

$$\frac{\partial}{\partial\theta}\log P_\theta(y) = k(\theta)\left(\hat{\theta}(y) - E_\theta\{\hat{\theta}(y)\}\right) \quad (*)$$

برای به دست آوردن شرط لازم برای تساوی، باید رابطه ی $(*)$ را بررسی کنیم و ببینیم که لزوم $P_\theta(y)$ تحت این شرط به چه شکلی خواهد بود همچنین $\hat{\theta}(y)$ به چه شکلی خواهد بود. بنابراین رابطه $(*)$ را به صورت زیر ساده می کنیم.

$$\frac{\partial}{\partial\theta}\log P_\theta(y) = k(\theta)\left(\hat{\theta}(y) - \underbrace{E_\theta\{\hat{\theta}(y)\}}_{f(\theta)}\right) \qquad (*)$$

برای انجام عملیات فرض می کنیم

$$\Lambda = (a, b)$$

همچنین برای سادگی در نوشتار، قرار می دهیم

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

از طرفین رابطه‌ی $(*)$ انتگرال‌گیری می کنیم.

$$\xrightarrow{\int_a^\theta (*)\, d\theta} \quad \int_a^\theta \frac{\partial}{\partial\theta_1}\log P_{\theta_1}(y)\, d\theta_1 = \int_a^\theta k(\theta_1)\left(\hat{\theta}(y) - f(\theta_1)\right) d\theta_1$$

$$\Rightarrow \log p_{\theta_1}(y) \Big|_{\theta_1=a}^{\theta_1=\theta} = \int_a^{\theta} k(\theta_1)\left(\hat{\theta}(y) - f(\theta_1)\right) d\theta_1$$

$$\Rightarrow \log p_{\theta}(y) - \log p_a(y) = \int_a^{\theta} k(\theta_1)\left(\hat{\theta}(y) - f(\theta_1)\right) d\theta_1$$

$$\Rightarrow \log p_{\theta}(y) = \log p_a(y) + \int_a^{\theta} k(\theta_1)\left(\hat{\theta}(y) - f(\theta_1)\right) d\theta_1 \qquad (**)$$

طرفین عبارت e

$$\Longrightarrow \quad e^{\log P_\theta(y)} = P_\theta(y)$$

$$P_\theta(y) = \underbrace{e^{\log P_a(y)}}_{h(y)} \exp\left(\int_a^\theta k(\theta_1)\left(\hat{\theta}(y) - f(\theta_1)\right) d\theta_1\right)$$

تابعیت فقط از y

تابعیت فقط از θ

$$\hat{\theta}(y) \int_a^\theta k(\theta_1)\, d\theta_1$$

$$\Longrightarrow P_\theta(y) = h(y) \exp\left(\int_a^\theta k(\theta_1)\hat{\theta}(y)\, d\theta_1 - \underbrace{\int_a^\theta k(\theta_1) f(\theta_1)\, d\theta_1}_{\text{تابعیت فقط از } \theta}\right)$$

تابعی از y

$e^{\text{تابعی از } \theta \times \text{تابعی از } y}$

تابعی از θ

بنابراین $P_\theta(y)$ را می‌توان به فرم $P_\theta(y) = C(\theta)\, e^{g(\theta)T(y)}\, h(y)$ نوشت.

$$h(y) = e^{\log P_a(y)}$$

$$c(\theta) = \exp\left(-\int_a^{\theta} k(\theta_1) f(\theta_1)\, d\theta_1\right)$$

$$g(\theta) = \int_a^{\theta} k(\theta_1)\, d\theta_1$$

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

که در آن

$\Rightarrow$ یعنی $P_\theta(y)$ فرم نمایی داشته باشد، $\hat{\theta}(y)=T(y)$

$\Leftarrow$ شرط لازم برای برقراری تساوی در کران نامساوی اطلاعات (رسیدن به تخمین گر با کمترین واریانس) این است که دسته $\{P_\theta(y), \theta\in\Lambda\}$ با شرایط گفته شده، را داشته باشیم $P_\theta(y)=c(\theta)e^{g(\theta)T(y)}h(y)$

و $\hat{\theta}(y) = T(y)$ در نظر گرفته شود.

بنابراین شرط لازم و کافی برای رسیدن به کران نامساوی اطلاعات (تخمین گر با کمترین واریانس) در وی خانواده توزیع های $\{P_\theta(y), \theta \in \Lambda\}$ با شرایط گفته شده، این است که داشته باشیم $P_\theta(y) = c(\theta)\, e^{g(\theta)T(y)}\, h(y)$ (یعنی $P_\theta(y)$ فرم نمایی داشته باشد) و در نظر بگیریم $\hat{\theta}(y) = T(y)$

در ادامه می خواهیم شرط برقراری تساوی در نامساوی اطلاعات را بررسی کنیم و به نتایجی در مورد $k(\theta)$ برسیم.

دیدیم که شرط تساوی به صورت زیر است

$$\frac{\partial}{\partial\theta}\log P_\theta(y) = k(\theta)\left(\hat{\theta}(y) - f(\theta)\right) \qquad (*)$$

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

می‌خواهیم بدانیم که در حالت تساوی $k(\theta)$ چه رابطه‌ای با پارامترهای نامساوی اطلاعات دارد.

طرفین رابطه $(*)$ را $\Longrightarrow$ به توان دو می‌رسانیم

$$\left(\frac{\partial}{\partial\theta}\log P_\theta(y)\right)^2 = k^2(\theta)\left(\hat{\theta}(y) - f(\theta)\right)^2 \qquad (1)$$

(با توجه به نامساوی اطلاعات)

$$\xrightarrow[\text{می گیریم } E_\theta(.)]{\text{از طرفین رابطه ①}} \quad \underbrace{E_\theta\left\{\left(\frac{\partial}{\partial\theta}\log P_\theta(y)\right)^2\right\}}_{I_\theta} = k^2(\theta)\,\underbrace{E_\theta\left\{\left(\hat{\theta}(y) - f(\theta)\right)^2\right\}}_{Var_\theta(\hat{\theta}(y))}$$

$$\Rightarrow \quad I_\theta = k^2(\theta)\, Var_\theta(\hat{\theta}(y)) \qquad ②$$

از طرف دیگر می دانیم که نامساوی اطلاعات نیز در حالت تساوی برقرار است یعنی

$$Var_\theta(\hat{\theta}(y)) = \frac{[f'(\theta)]^2}{I_\theta} \qquad ③$$

③ , ② $\Longrightarrow$

$$k(\theta) = \frac{I_\theta}{f'(\theta)}$$

باضرب کردن $Var_\theta(\hat{\theta}(y))$

این رابطه برای به دست آوردن فرم $P_\theta(y)$ در حالت تساوی، کمک کننده است.

* تخمین بیشترین شباهت

Maximum Likelihood (ML)

همانطور که در بحث MVUE بیان کردیم، برای به دست آوردن تخمین MVUE نیاز به محاسبه یک آماره کافی کامل داریم که لازم می‌دارد مجموعه $\{P_\theta ; \theta \in \Lambda\}$ یک مجموعه خانواده توزیع‌های کامل باشد و یک آماره کافی روی آن به دست آوریم.

در بسیاری از مسائل شرط بالا برقرار نیست و ما به دنبال راه‌حلی برای تخمین هستیم که بتوانیم در حالت کلی برای همه مجموعه‌های $\{P_\theta \,;\, \theta \in \Lambda\}$ و بدون داشتن اطلاعات اولیه در مورد پارامتر θ (یعنی $w(\theta)$ را در اختیار نداریم) پارامتر θ را تخمین بزنیم. بنابراین به دنبال راه‌حلی هستیم که بدون هیچ شرطی روی $\{P_\theta \,;\, \theta \in \Lambda\}$ بتوانیم تخمین θ را به دست بیاوریم. یکی از راه‌حلهای متداول برای این منظور، روش تخمین ML است.

از بحث تخمین MAP می‌دانیم که

$$
\begin{aligned}
\hat{\theta}_{MAP}(y) &= \underset{\theta}{\mathrm{Arg\,Max}}\left(w(\theta|y)\right) \\
&= \underset{\theta}{\mathrm{Arg\,Max}}\left(\frac{P(y|\theta)\,P(\theta)}{P(y)}\right) \\
&= \underset{\theta}{\mathrm{Arg\,Max}}\left(\overbrace{P(y|\theta)}^{P_\theta(y)}\ \overbrace{P(\theta)}^{w(\theta)}\right) \\
&= \underset{\theta}{\mathrm{Arg\,Max}}\left\{P_\theta(y)\,w(\theta)\right\}
\end{aligned}
$$

در مسأله‌ای که می‌خواهیم تخمین θ را بدون داشتن اطلاعات اولیه از θ به دست بیاوریم یا
به عبارت دیگر اطلاعات خاصی در مورد θ در اختیار نداریم که بتوانیم $w(\theta)$ را تعیین کنیم.
فرض می‌کنیم که توزیع θ روی مجموعه Λ به صورت یکنواخت است. یعنی $w(\theta)=cte$
(از نظر تئوری اطلاعات در حالت گسسته، توزیع یکنواخت ماکزیمم آنتروپی را دارد و در این حالت
داریم برای بدترین حالت ممکن، تخمین را به دست می‌آوریم) در این حالت تخمین MAP
به صورت زیر خواهد بود.

$$\hat{\theta}_{MAP}(y) = \underset{\theta}{\mathrm{Arg\,Max}} \{ P_\theta(y) \} \equiv \hat{\theta}_{ML}(y)$$

تابع $P(y|\theta) \equiv P_\theta(y)$ ، هر تابع یکنوا (mono-tone) از آن را یک تابع شباهت (likelihood function) می‌گوییم. بنابراین این تخمین‌گر، تخمین‌گری است که تابع شباهت را ماکزیمم می‌کند و به آن تخمین‌گر بیشترین شباهت می‌گوییم.

* توجه به این نکته لازم است که نداشتن اطلاعات اولیه (در اختیار نداشتن $w(\theta)$) معادل با یکنواخت بودن توزیع θ نیست. علاوه بر این اگر Λ یک مجموعه نامتناهی باشد، فرض یکنواخت بودن توزیع θ، فرض درستی نیست. اما این دیدگاه برای رسیدن به جواب تخمین ML یک دیدگاه ساده است که ما را با قضیه‌های ریاضی مرتبط

بدون ML به تخمین برای رسیدن به تخمین ریاضی های قضیه و روابط) نمی کند دگرگیری ML تخمین.

فرض یکنواخت بودن توزیع θ، در مراجع وجود دارد)

در هر صورت، چنانکه دیدیم

$$\hat{\theta}_{ML}(y) = \operatorname{Arg}\max_{\theta} \{ P_\theta(y) \}$$

به دست آوردن تخمین ML با یک روش سرراست و ساده امکان پذیر است و همچنین تخمین ML ویژگی های خاصی دارد که در ادامه به آن می پردازیم. اما پیاده سازی عملی تخمین گر ML پیچیده است (پیچیدگی از مرتبه نمایی نسبت به ابعاد مسئله است)

برای به دست آوردن تخمین ML می‌توانیم به صورت زیر عمل کنیم.

$$\hat{\theta}_{ML}(\underline{y}) = \operatorname{Arg}\max_{\theta} \{ P_\theta(\underline{y}) \}$$

$$= \operatorname{Arg}\max_{\theta} \{ \log P_\theta(\underline{y}) \}$$

log یک تابع اکیداً صعودی است

$$\xrightarrow[P_\theta(\underline{y})]{\text{با فرض پیوستگی}} \quad \frac{\partial}{\partial\theta} \log P_\theta(\underline{y}) \Big|_{\theta = \hat{\theta}_{ML}} = 0$$

معادله شباهت

Likelihood equation

در ادامه می‌خواهیم ویژگی‌های این معادله را بررسی کنیم. به طور خاص به دنبال رابطه‌ای بین تخمین ML و MVUE هستیم.

فرض می‌کنیم $\hat{\theta}(y)$ یک تخمین بدون بایاس از θ است که به کران کرامر-رائو می‌رسد.

$$Var_{\theta}\left(\hat{\theta}(y)\right) = \frac{1}{I_{\theta}}$$

به عبارت دیگر $\hat{\theta}(y)$ یک MVUE است.

از طرف دیگر دیدیم که برای برقراری تساوی در کران نامساوی فوق شرایط زیر باید برقرار باشد.

$$\log P_\theta(y) = \log h(y) + \int_a^\theta k(\theta_1)\left(\hat\theta(y) - f(\theta_1)\right) d\theta_1 \qquad (I)$$

(با توجه به رابطه (8a))

همچنین می دانیم که در حالت تساوی داریم

$$k(\theta) = \frac{I_\theta}{f'(\theta)}$$

با توجه به اینکه $\hat\theta(y)$ بدون بایاس است، داریم.

$$f(\theta) = E_\theta\{\hat\theta(y)\} = \theta \implies f'(\theta) = 1 \implies k(\theta) = I(\theta)$$

$$\xRightarrow{(I)} \log P_\theta(y) = \log h(y) + \int_a^\theta I_{\theta_1}\left(\hat\theta(y) - \theta_1\right) d\theta_1$$

از طرف دیگر، معادله شباهت به صورت زیر خواهد شد

$$\frac{\partial}{\partial \theta} \log P_\theta(y) \Big|_{\theta = \hat{\theta}_{ML}} = 0$$

$$\frac{\partial}{\partial \theta} \log P_\theta(y) \Big|_{\theta = \hat{\theta}_{ML}} = Z_\theta \left(\hat{\theta}(y) - \theta \right) \Big|_{\theta = \hat{\theta}_{ML}} = 0$$

$$\Longrightarrow \hat{\theta}(y) = \hat{\theta}_{ML}$$

یعنی $\hat{\theta}(y)$ با فرض های گفته شده، همان تخمین ML است.

به عبارت دیگر هر تخمین بدون بایاس $\hat{\theta}(y)$ که به کران کرامر-رائو برسد یک تخمین‌گر ML (بیشترین شباهت) است.

یعنی $\hat{\theta}(y)$ یک MVUE است

تخمین‌گری که معادله شباهت را برقرار می‌کند

اگر $\hat{\theta}$ یک MVUE باشد $\Longrightarrow$ $\hat{\theta}$ یک تخمین ML است

$\not\Longleftarrow$

ولی عکس این مطلب لزوماً درست نیست یعنی هر جواب معادله شباهت (که تخمین‌گر ML است) لزوماً MVUE نیست

مثال : تخمین ML برای پارامتر یک توزیع نمایی

فرض کنیم مشاهدات y_1 ، y_2 ، ... ، y_n را در اختیار داریم و می‌دانیم این مشاهدات

متغیرهای تصادفی iid با توزیع نمایی با پارامتر θ هستند. $\Lambda = (0, \infty)$

$\Gamma = \mathbb{R}^n$

$$P_\theta(y_k) = \begin{cases} \theta e^{-\theta y_k} & y_k \geq 0 \\ 0 & oth. \end{cases}$$

(y_k ها توزیع نمایی با پارامتر θ دارند)

تخمین ML از θ را می‌خواهیم از روی مشاهدات y_1 ، y_2 ، ... ، y_n به دست بیاوریم

$\hat{\theta}_{ML}(\underline{y})$

می دانیم که

$$\hat{\theta}_{ML}(\underline{y}) = \operatorname{Arg\,Max}_{\theta} \{ P_\theta(\underline{y}) \}$$

بنابراین در اولین مرحله $P_\theta(\underline{y})$ را به دست می آوریم.

$$P_\theta(\underline{y}) = \prod_{k=1}^{n} P_\theta(y_k) = \prod_{k=1}^{n} \theta e^{-\theta y_k} \qquad \{y_k\}_{k=1}^{n} \geq 0$$

$\{Y_k\}_{k=1}^{n}$ متغیرهای iid

$$\bar{y} = \frac{\sum_{k=1}^{n} y_i}{n}$$

$$\Longrightarrow P_\theta(\underline{y}) = \theta^{n} e^{-\theta \sum_{k=1}^{n} y_k} = \theta^{n} e^{-n\theta\bar{y}} \qquad \{y_k\}_{k=1}^{n} \geq 0$$

$$\hat{\theta}_{ML}(\underline{y}) = \underset{\theta}{\mathrm{Arg\,Max}} \left\{ P_\theta(\underline{y}) \right\}$$

$$= \underset{\theta}{\mathrm{Arg\,Max}} \left\{ \theta^n e^{-n\theta\bar{y}} \right\}$$

$$\overset{\log(.)}{=} \mathrm{Arg\,Max} \left\{ \underbrace{n\log\theta - n\theta\bar{y}}_{\log P_\theta(\underline{y})} \right\}$$

$$\frac{\partial}{\partial\theta} \left\{ \underbrace{n\log\theta - n\theta\bar{y}}_{\log P_\theta(\underline{y})} \right\} \Bigg|_{\theta = \hat{\theta}_{ML}} = 0$$

معادله شباهت

حل معادله شباهت $\Longrightarrow$

$$\frac{\partial}{\partial\theta}\left(n\log\theta - n\theta\bar{y}\right)\Big|_{\theta=\hat{\theta}_{ML}} = 0$$

$$\frac{n}{\theta} - n\bar{y}\Big|_{\theta=\hat{\theta}_{ML}} = 0$$

$$\Longrightarrow \quad \hat{\theta}_{ML}(\underline{y}) = \frac{1}{\bar{y}} = \frac{n}{\sum_{k=1}^{n} y_k} \quad , \quad \bar{y} = \frac{1}{n}\sum_{k=1}^{n} y_k$$

با توجه به اینکه مشتق دوم $\frac{\partial^2}{\partial\theta^2}\log P_\theta(y) = -\frac{n}{\theta^2}$ به ازای $\hat{\theta}_{ML}(y)$ که در بالا

به دست آمده است، منفی است در نتیجه جواب به دست آمده عبارت $P_\theta(\underline{y})$ را ماکزیمم می کند.

$$\hat{\theta}_{ML}(\underline{y}) = \frac{1}{\bar{y}} \quad , \quad \bar{y} = \frac{\sum_{k=1}^{n} y_k}{n}$$

نکته اول: در مورد یک متغیر تصادفی نمایی y_k با پارامتر θ می دانیم که

$$E\{y_k\} = \frac{1}{\theta} \quad , \quad k=1,2,\dots,n$$

از طرف دیگر به دست آوردیم

$$\hat{\theta}_{ML}(\underline{y}) = \frac{1}{\bar{y}}$$

($\bar{y}$ میانگین عددی متغیرهای تصادفی $y_1, \ldots, y_n$ است)

* از قانون اعداد بزرگ، می دانیم که اگر n به اندازه ی کافی بزرگ باشد، میانگین ها به سمت میانگین های آماری (یعنی امید ریاضی) همگرا می شوند. یعنی

$$\{\bar{y}\} \xrightarrow[n\to\infty]{P} Ey = \frac{1}{\theta}$$

(همگرایی در احتمال)

یا به عبارت دیگر

$$\left\{\frac{1}{\bar{y}}\right\} \xrightarrow[n\to\infty]{P} \theta$$

(همگرایی در احتمال)

بنابراین می‌توان گفت که تخمین‌گر $\hat{\theta}_{ML}(\underline{y})$، ML برای n های به اندازه کافی بزرگ، با مفهوم احتمال به سمت مقدار واقعی پارامتر θ همگرا می‌شود.

این خاصیت برای تمامی تخمین‌گرهای ML برقرار است و به آن خاصیت سازگاری Consistency گفته می‌شود.

تخمین ML در حد وقتی $n \to \infty$، به سمت مقدار واقعی پارامتر مورد تخمین همگرا می‌شود.

Consistency of ML Estimation.