استنباط آماری پیشرفته (۱)

نادر نعمت الهی

گروه آمار

دانشگاه علامه طباطبایی

فهرست مطالب

فصل ۱

کلیات و مفاهیم اولیه

بخش ۱: خلاصه‌ای از نظریه اندازه[۱]

در ابتدا خلاصه‌ای از نظریه اندازه که مورد لزوم مباحث این درس می‌باشد را مرور می‌کنیم.

تعریف ۱-۱ : سیگما میدان $(\sigma - field)$

فرض کنید Ω یک مجموعه باشد و A مجموعه‌ای غیر تهی از زیر مجموعه‌های Ω باشد. A را یک سیگما میدان گویند به شرط آن که شرایط زیر برای آن برقرار باشد

الف) اگر $A \in \mathsf{A}$ آنگاه $A^C \in \mathsf{A}$

ب) اگر $A_1, A_2, \ldots$ مجموعه‌هایی در A باشند آنگاه $\bigcup_{n=1}^{\infty} A_n \in \mathsf{A}$ ■

از تعریف فوق نتیجه می‌شود که اگر A یک سیگما میدان باشد آنگاه $\Omega \in \mathsf{A}$ ، $\varnothing \in \mathsf{A}$ و $\bigcap_{n=1}^{\infty} A_n \in \mathsf{A}$ خواهد بود.

تعریف ۱-۲ : یک اندازه μ یک تابع مجموعه‌ای روی سیگما میدان A از زیر مجموعه‌های Ω است هر گاه در شرایط زیر صدق کند $\mu : A \to [0, \infty)$

الف) اگر $A \in \mathsf{A}$ آنگاه $\mu(A) \geq 0$

ب) اگر $A_1, A_2, \ldots \in \mathsf{A}$ و A_i ها مجزا باشند، یعنی $\forall i \neq j \quad A_i \cap A_j = \phi$ ، آنگاه

$$\mu\left(\bigcup_{n=1}^{\infty} A_i\right) = \sum_{n=1}^{\infty} \mu(A_n)$$ ■

اگر A یک سیگما میدان از زیر مجموعه‌های Ω باشد آنگاه (Ω, A) را یک فضای اندازه‌پذیر[۲] گویند و مجموعه‌ی $A \in \mathsf{A}$ را اندازه‌پذیر گویند. همچنین اگر μ یک اندازه باشد که روی (Ω, A) تعریف شده باشد آنگاه سه تایی $(\Omega, \mathsf{A}, \mu)$ را یک فضای اندازه[۳] می‌نامند. یک اندازه μ سیگما

[۱] Measure Theory
[۲] Measurable Space
[۳] Measure Space

متناهی $(\sigma - finite)$ نامیده می‌شود. هر گاه مجموعه‌های $A_i \in \mathsf{A}$ موجود باشند که $\Omega = \bigcup_i A_i$ و $\mu(A_i) < \infty$. تمام اندازه‌هایی که در این درس بررسی می شوند، سیگما متناهی هستند و به همین دلیل آن ها را تنها اندازه می نامیم.

مثال ۱-۱: اندازه‌شمارنده[۱]. فرض‌کنید $\Omega = \{w_1, w_2, w_3, ...\}$ و تمام زیرمجموعه‌های Ω $\mathsf{A} =$

$$if \;\; A \in \mathsf{A} \qquad \Rightarrow \mu(A) = \begin{cases} A & \text{تعداداعضای مجموعه} \\ \infty & \end{cases} \quad \begin{matrix} \text{اگر } A \text{ متناهی باشد} \\ \text{اگر } A \text{ نامتناهی باشد} \end{matrix}$$

این اندازه را اندازه شمارنده گویند. ■

مثال ۱-۲: اندازه لبگ[۲]. فرض کنید Ω فضا اقلیدسی n بعدی باشد،

$\Omega = \{(w_1, ..., w_n) \,|\, w_i \in \Re, \;\; \forall i\}$ و A کوچکترین سیگما میدانی باشد که شامل تمام مستطیل‌های باز زیر باشد:

$$A = \{(w_1, w_2, ..., w_n) \,|\, a_i < w_i < b_i, \;\; -\infty < a_i < b_i < +\infty\}$$

A شامل تمام زیرمجموعه‌های باز و بسته Ω می‌باشد و به اعضای آن مجموعه‌های برل[۳] گویند. تنها یک اندازه μ می‌توان روی این A معرفی کرد که به مجموعه A فوق اندازه

$$\mu(A) = (b_1 - a_1)(b_2 - a_2)....(b_n - a_n)$$

را نسبت می‌دهد (که حجم مجموعه A است) و به آن اندازه لبگ گویند. ■

توابع اندازه‌پذیر[۴]:

یک تابع حقیقی $f : \Omega \to R$ یک تابع اندازه‌پذیر است اگر برای هر مجموعه برل B داشته باشیم که $f^{-1}(B) \in \mathsf{A}$. برای مثال

$$A \in \mathsf{A} \;\; , \;\; I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

---

[۱] Counting Measure
[۲] Lebesgue Measure
[۳] Borel Sets
[۴] Measurable Functions

I_A یک تابع اندازه‌پذیر است. اگر $A_1,A_2,....,A_K \in \mathsf{A}$ و $A_i \cap A_j = \phi$ آنگاه

$$f_s(x) = a_1 I_{A_1}(x) + a_2 I_{A_2}(x) + ... + a_k I_{A_k}(x)$$

یک تابع اندازه‌پذیر است که به آن تابع ساده گویند. حال اگر یک دنباله از توابع ساده غیرنزولی $f_{s_1}, f_{s_2}, ...$ را در نظر بگیریم و قرار دهیم $f^+(x) = \lim_{n\to\infty} f_{s_n}(x) \in [0,\infty)$ آنگاه f^+ یک تابع اندازه‌پذیر است. همچنین یک تابع دلخواه f را می‌توان بصورت $f = f^+ - f^-$ نوشت که در آن

$$f^+(x) = \max(f(x), 0)$$
$$f^-(x) = -\min(f(x), 0)$$

یک تابع $f: \Omega \to R$ اندازه‌پذیر خواهد بود اگر قسمت های مثبت و منفی آن یعنی f^+ و f^- اندازه‌پذیر باشند و همچنین f^+ و f^- در صورتی اندازه‌پذیر هستند که بصورت حد دنباله‌ای از توابع ساده غیرنزولی باشند.

انتگرال یک تابع نسبت به اندازه μ :

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \Rightarrow \int I_A d\mu = \mu(A)$$

$$f_s(x) = a_1 I_{A_1}(x) + ... + a_k I_{A_k}(x) \Rightarrow \int f_s d\mu = \sum_{i=1}^{k} a_i \mu(A_i)$$

$$f(x) = \lim_{n\to\infty} f_{s_n}(x) \Rightarrow \int f d\mu = \lim_{n\to\infty} \int f_{s_n} d\mu$$

$$f = f^+ - f^- \Rightarrow \int f d\mu = \int f^+ d\mu - \int f^- d\mu$$

مثال ۱-۳: اگر $\Omega = \{x_1, x_2, ...\}$ و μ اندازه شمارنده باشد انگاه برای یک تابع $f: \Omega \to [0,\infty)$ داریم f که

$$f_{s_n}(x) = \begin{cases} f(x_i) & x = x_i \quad i = 1,2,...,n \\ 0 & o.w. \end{cases}$$
$$= f(x_1) I_{\{x_1\}}(x) + f(x_2) I_{\{x_2\}}(x) + ... + f(x_n) I_{\{x_n\}}(x)$$

$$f(x)=\lim_{n\to\infty} f_{s_n}(x)=f(x_1)I_{\{x_1\}}(x)+f(x_2)I_{\{x_2\}}(x)+\dots$$

$$=\begin{cases} f(x_i) & x=x_i \\ 0 & o.w. \end{cases}$$

$$\int f(x)d\mu=\lim_{n\to\infty}\int f_{s_n}(x)d\mu=\lim_{n\to\infty}\sum_{i=1}^{n} f(x_i)=\sum_{i=1}^{+\infty} f(x_i)$$

پس در صورتی f نسبت به μ انتگرال‌پذیر است که $\sum_{i=1}^{+\infty} f(x_i)$ متناهی باشد. ■

واژه تقریبا همه جا[1]

فرض کنید یک گزاره P برای تمام نقاط Ω درست باشد بجز نقاط A که در آن $\mu(A)=0$ (یعنی P برای تمام $\Omega-A$ درست است و $\mu(A)=0$ می باشد) در این صورت گوئیم این گزاره تقریبا همه جا μ درست است. ($a.e.\mu$)

قضیه همگرایی مغلوب[2] :

اگر f_n دنباله‌ای از توابع اندازه‌پذیر باشند و $\lim_{n\to\infty} f_n(x)=f(x)$ $(a.e.\mu)$

و یک تابع انتگرال‌پذیر g چنان موجود باشد که $|f_n(x)|\le g(x)$ $\forall x$

آنگاه f_n و f انتگرال‌پذیر هستند و $\int f d\mu=\lim_{n\to\infty}\int f_n d\mu$ ■

لم فاتو[3]

اگر $\{f_n\}$ دنباله‌ای از توابع نامنفی اندازه‌پذیر باشند، آنگاه

$$\int(\liminf_{n\to\infty} f_n)d\mu\le\liminf_{n\to\infty}\int f_n d\mu$$ ■

[1] almost every where
[2] Dominated Convergence Theorem
[3] Fatou's Lemma

قضیه رادون نیکودیم[1]

فرض کنید f یک تابع نامنفی انتگرال‌پذیر و $A \in \mathsf{A}$ باشد. قرار می‌دهیم:

$$\nu(A) = \int_A f d\mu \qquad (1)$$

در این صورت $\nu(A)$ یک اندازه روی $(\Omega\ ,\ \mathsf{A})$ است و واضح است که

$$\mu(A) = 0 \Rightarrow \nu(A) = 0 \qquad (2)$$

اگر رابطه (۲) برقرار باشد گوییم که ν نسبت به μ مطلقاً پیوسته[2] است. ■

قضیه رادون نیکودیم می گوید که رابطه (۲) شرط لازم و کافی برای آن است که یک تابع f موجود باشد که در رابطه (۱) صدق کند. تابع f را مشتق رادون نیکودیم ν نسبت به μ گویند $f = \frac{\partial \nu}{\partial \mu}$.

این مشتق به صورت $a.e.$ یکتا است یعنی اگر g نیز در رابطه (۱) صدق کند آنگاه:

$$f = g \quad a.e.\mu .$$

قضیه فوبینی[3]

فرض کنید $(\Omega_1,\ \mathsf{A}\ ,\mu)$ و $(\Omega_2,\ \mathsf{B}\ ,\nu)$ دو فضای اندازه و f یک تابع غیرمنفی اندازه‌پذیر روی $\Omega_1 \times \Omega_2$ باشد . در این صورت:

$$\int_{\Omega_1}\left[\int_{\Omega_2} f\ (x,y)d\nu(y)\right]d\mu(x) = \int_{\Omega_2}\left[\int_{\Omega_1} f\ (x,y)d\mu(x)\right]d\mu(y)$$

$$= \int_{\Omega_1\times\Omega_2} f\ \ d(\mu\times\nu)\ \blacksquare$$

اندازه احتمال[4]

فرض کنید Ω فضای نمونه یک آزمایش تصادفی باشد. یک اندازه‌ی P روی فضای اندازه‌پذیر $(\Omega, \mathsf{A}\ )$ که در شرط $P(\Omega) = 1$ صدق کند را اندازه‌ی احتمال گویند و به مقدار $P(A)$ احتمال A گویند و به $(\Omega,\ \mathsf{A}\ ,P)$ فضای احتمال گویند.

---

[1] Radon Nikodym
[2] Absolutely Continuous
[3] Fubini's Theorem
[4] Probability Measure

اگر P نسبت به اندازه μ مطلقاً پیوسته باشد و p مشتق رادون نیکودیم آن باشد آنگاه

$$P(A)=\int_A p d\mu$$

p را تابع چگالی احتمال P نسبت به μ گویند.

در این درس معمولاً Ω را فضای اقلیدسی در نظر می گیریم. اگر اندازه‌ی احتمال P نسبت به اندازه‌ی لبگ مطلقاً پیوسته با چگالی p باشد آنگاه توزیع مربوطه را یک توزیع پیوسته و اگر P نسبت به اندازه شمارنده مطلقاً پیوسته با چگالی p باشد آنگاه توزیع مربوطه را یک توزیع گسسته گویند. در کاربردهای معمولی این درس داریم که :

$$\int f dP=\int fp d\mu=\begin{cases}\int f(x)p(x)dx & (p \text{ پیوسته}) \text{ اندازه لبگ}\\ \sum f(x)p(x) & (p \text{ گسسته}) \text{ اندازه شمارنده}\end{cases}$$

در استنباط آماری معمولاً با یک خانواده از توزیع‌ها به صورت $\mathsf{P}=\{P_\theta:\theta\in\Theta\}$ سر و کار داریم. اگر به ازای هر θ، P_θ نسبت به اندازه μ مطلقاً پیوسته باشد گوییم که خانواده P تحت تسلط μ قرار دارد[۱]. اگر $P_\theta(N)=0$ گوئیم $N=P_\theta-null\ set$ و اگر $\forall\theta\in\Theta$ $P_\theta(N)=0$ گوییم $N=\mathsf{P}$ – null set.

متغیر تصادفی:

در یک فضای احتمال (Ω, A, P) تابع $X:\Omega\to R$ را اگر اندازه‌پذیر باشد یک متغیر تصادفی گویند یعنی اگر برای هر مجموعه بُرل B، $X^{-1}(B)\in\mathsf{A}$ باشد آنگاه Xرا یک متغیر تصادفی گویند

$$A=X^{-1}(B) \quad , \quad P_X(B)=P(X\in B)=P(A)$$

P_X را توزیع متغیر تصادفی X نامیم.

اگر P_X نسبت به اندازه لبگ مطلقاً پیوسته باشد آنگاه X را یک متغیر پیوسته با تابع چگالی p_X گوییم هر گاه $P_X(A)=\int_A p_x d\mu$

و X را گسسته گوئیم هر گاه $P_X(A)=\int_A p_x d\mu \rightsquigarrow$ Counting Measure

[۱] Dominated by μ

در هر صورت p_X را می‌توان چگالی P_X نامید. گرچه در حالت گسسته آن را تابع احتمال نیز گویند.

اگر $A=(-\infty,x]$ آنگاه $F_X(x)=P_X(A),$ F_X را تابع توزیع متغیر تصادفی X نامیم . اگر X پیوسته باشد (P_Xمطلقاً پیوسته است) آنگاه ثابت می‌شود که تابع $F_X(x)$ نیز (با تعریف توابع مطلقاً پیوسته در آنالیز) مطلقاً پیوسته است.

در آنالیز، تابع g را مطلقاً پیوسته نامیم اگر به ازای هر $\varepsilon>0$ یک $\delta>0$ وجود داشته باشد که به ازای هر n،

$$\sum_{i=1}^{n}(b_i-a_i)<\delta \Rightarrow \sum_{i=1}^{n}|f(b_i)-f(a_i)|<\varepsilon$$

بخش ۲: خانواده گروهی از توزیع‌ها[1]

در این درس با دو خانواده از توزیع‌ها سر و کار داریم که یکی خانواده گروهی و دیگری خانواده نمایی از توزیع‌ها می باشد.

تعریف ۲-۱: یک خانواده گروهی از توزیع‌ها به وسیله انجام یک خانواده از تبدیلات روی یک متغیر تصادفی با یک توزیع خاص حاصل می‌شود.

در زیر سه خانواده‌ی مهم از خانواده‌ی گروه‌ها را معرفی می کنیم.

۱- خانواده‌ی مکانی[2]

فرض کنید U یک متغیر تصادفی با یک توزیع احتمال ثابت $f(u)$ باشد. اگر μ مقداری ثابت باشد آنگاه توزیع $X=U+\mu$ عبارت است از $f_X(x)=f(x-\mu)$

توزیع حاصل برای توزیع احتمال ثابت f و $-\infty<\mu<+\infty$ را یک خانواده مکانی از توزیع‌ها گویند.

برای مثال توزیع $N(\mu,1)$ که از توزیع ثابت $N(0,1)$ با جمع با مقدار ثابت μ حاصل می‌شود.

$$Z\sim N(0,1)\Rightarrow X=Z+\mu\sim N(\mu,1),\qquad f_x(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu)^2}=f(x-\mu)$$

[1] Group Families
[2] Location Family

۲- خانواده مقیاسی[۱]

فرض کنید U یک متغیر تصادفی با یک توزیع احتمال ثابت $f(u)$ باشد. اگر σ مقداری ثابت باشد آنگاه توزیع $X = \sigma U$ عبارت است از $f_X(x) = \frac{۱}{\sigma} f(\frac{x}{\sigma})$

توزیع حاصل برای $\sigma > ۰$ را یک خانواده مقیاسی از توزیع‌ها گویند.

برای مثال توزیع $E(\theta)$ که از توزیع ثابت $E(۱)$ با ضرب در مقدار ثابت θ حاصل می‌گردد.

$$U \sim E(۱) \Rightarrow X = \theta U \sim E(\theta), \quad f_X(x) = \frac{۱}{\theta} e^{-\frac{x}{\theta}} = \frac{۱}{\theta} f(\frac{x}{\theta})$$

۳- خانواده‌ی مکانی – مقیاسی[۲]

فرض کنید U یک متغیر تصادفی با یک توزیع احتمال ثابت $f(u)$ باشد. اگر μ و σ مقادیر ثابتی باشند. آنگاه توزیع $X = \mu + \sigma U$ عبارت‌است از $f_X(x) = \frac{۱}{\sigma} f(\frac{x-\mu}{\sigma})$

توزیع حاصل برای $-\infty < \mu < +\infty$ و $\sigma > ۰$ را یک خانواده مکانی – مقیاسی از توزیع‌ها گویند.

برای مثال توزیع $N(\mu, \sigma^۲)$ از توزیع $N(۰,۱)$ به صورت زیر حاصل می‌گردد.

$$U \sim N(۰,۱) \Rightarrow X = \mu + \sigma U \sim N(\mu, \sigma^۲),$$

$$f_X(x) = \frac{۱}{\sigma} \frac{۱}{\sqrt{۲\pi}} e^{-\frac{۱}{۲}(\frac{x-\mu}{\sigma})^۲} = \frac{۱}{\sigma} f(\frac{x-\mu}{\sigma})$$

در زیر تعدادی از خانواده های مکانی مقیاسی آورده شده است.

(i) $E(\alpha, \beta)$ $f_X(x) = \frac{۱}{\beta} e^{-\frac{۱}{\beta}(x-\alpha)} I_{(\alpha+\infty)}(x)$

(ii) $DE(\alpha, \beta)$ $f_X(x) = \frac{۱}{۲\beta} e^{-\frac{۱}{\beta}|x-\alpha|}$

(iii) $C(\alpha, \beta)$ $f_X(x) = \frac{۱}{\pi\{۱ + (\frac{x-\alpha}{\beta})^۲\}}$

---

[۱] scale family

[۲] location – scale family

$$(iv) \qquad U(\theta-\frac{\tau}{۲},\theta+\frac{\tau}{۲})$$

در هر یک از حالات فوق ما یک کلاس از تبدیلات G و یک عمل * داریم که دارای دو خاصیت زیر است.

الف- کلاس G تحت عمل ترکیب بسته است یعنی

$$if \quad g_۱,g_۲ \in G \qquad then \qquad g_۱ * g_۲ \in G$$

ب- کلاس G تحت عمل وارون بسته است یعنی

$$if \quad g \in G \quad then \quad \exists g^{-۱} \in G \qquad ə\, g * g^{-۱} = e \in G$$

که e تبدیل همانی نام دارد.

یک کلاس از تبدیلات که تحت عمل ترکیب و وارون بسته باشد، یک گروه تبدیلات نامیده می‌شود.

حال فرض کنید P یک توزیع خاص و معلوم مربوط به متغیر تصادفی X باشد. در این صورت مجموعه‌ی $G = \{g \mid Y_g = g(X)\}$ را یک خانواده ***گروهی از توزیع‌ها*** گویند.

برای مثال خانواده‌ی مکانی- مقیاسی یک خانواده گروهی از توزیع‌ها می باشد.

$$G = \{g \mid g(U) = \mu + \sigma U, \mu \in \Re, \sigma > ۰\}$$

الف) $X_۱ = g_۱(U) = \mu_۱ + \sigma_۱ U$

$$X_۲ = g_۲(X_۱) = g_۲(g_۱(U)) = \mu_۲ + \sigma_۱(g_۱(U)) = \mu_۲ + \sigma_۱(\mu_۱ + \sigma_۱ U)$$
$$= (\mu_۱\sigma_۱ + \mu_۲) + \sigma_۱\sigma_۲ U \Rightarrow g_۲ * g_۱ \in G$$

$$X = g(U) = \mu + \sigma U \quad \Rightarrow \quad g^{-۱}(U) = \frac{U-\mu}{\sigma}$$

$$Y = g(g^{-۱}(U)) = \mu + \sigma(\frac{U-\mu}{\sigma}) = U \quad \Rightarrow \quad g * g^{-۱} = e \in G$$

یک خانواده از تبدیلات که در شرط $\forall g_۱,g_۲ \in G \;\; g_۱ * g_۲ = g_۲ * g_۱$ صدق کند، یک گروه تبدیلات جابجایی[۱] نامیده می‌شود. برای مثال خانواده‌ی مکانی و خانواده‌ی مقیاسی یک گروه تبدیلات جابجایی هستند، اما خانواده مکانی – مقیاسی چنین نیست.

---

[۱] Commutative

مثال ۲-۱: فرض کنید $\underset{\sim}{U} = (U_1,...,U_n)$ یک بردار تصادفی n بعدی باشد در این صورت می‌توان گروه تبدیلات زیر را روی این بردار تصادفی انجام داد.

$$\underset{\sim}{U} + \underset{\sim}{a} = (U_1 + a, U_2 + a, ..., U_n + a) \qquad a \in R$$

$$b\underset{\sim}{U} = (bU_1, bU_2, ..., bU_n)$$

$$a + b\underset{\sim}{U} = (a + bU_1, a + bU_2, ..., a + bU_n) \qquad a \in R, b > 0 \quad \blacksquare$$

مثال۲- ۲: توزیع چند متغیره نرمال

فرض کنید $\underset{\sim}{U} = (U_1,...,U_p)$ و $U_i \sim N(0,1)$,$i = 1,...,p$, و $U_1,...,U_p$ از یکدیگر مستقل باشند و تبدیل زیر را در نظر بگیرید

$$\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} + B \begin{pmatrix} U_1 \\ U_2 \\ . \\ . \\ U_p \end{pmatrix} \Leftrightarrow \underset{\sim}{X} = \underset{\sim}{a} + B\underset{\sim}{U}$$

که در آن B یک ماتریس وارون پذیر $p \times p$ است. توزیع $\underset{\sim}{X}$ را توزیع نرمال چند متغیره با میانگین و واریانس زیر گویند

$$E(\underset{\sim}{X}) = \underset{\sim}{a} + BE(\underset{\sim}{U}) = \underset{\sim}{a}$$

$$\sum = V(\underset{\sim}{X}) = BV(\underset{\sim}{U})B' = BB'$$

برای بدست آوردن توزیع $\underset{\sim}{X}$ توجه کنید که

$$f(\underset{\sim}{u}) = (2\pi)^{-\frac{p}{2}} e^{-\frac{1}{2}\sum_{i=1}^{p} u_i^2} = (2\pi)^{-\frac{p}{2}} e^{-\frac{1}{2}\underset{\sim}{u}'\underset{\sim}{u}}$$

$$\underset{\sim}{X} = \underset{\sim}{a} + B\underset{\sim}{U} \Rightarrow \underset{\sim}{U} = B^{-1}(\underset{\sim}{X} - \underset{\sim}{a}) \Rightarrow |J| = |B^{-1}|$$

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = |B^{-1}| f_{\underset{\sim}{U}}(B^{-1}(\underset{\sim}{x} - \underset{\sim}{a})) = |B^{-1}| \left\{ e^{-\frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{a})'(B^{-1})'B^{-1}(\underset{\sim}{x} - \underset{\sim}{a})} \right\} (2\pi)^{-p/2}$$

$$= (2\pi)^{-p/2} |\sum|^{-1/2} e^{-\frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{a})'\sum^{-1}(\underset{\sim}{x} - \underset{\sim}{a})} \qquad \blacksquare$$

بخش ۳: خانواده نمایی از توزیع‌ها

تعریف ۳-۱: یک خانواده $\{P_{\underset{\sim}{\theta}} : \underset{\sim}{\theta} \in \Theta\}$ از توزیع‌ها را یک خانواده‌ی نمایی k پارامتری گویند اگر توزیع P_θ برای هر $\underset{\sim}{\theta} \in \Theta$ و هر $x \in S_X^*$ (تکیه‌گاه X) دارای تابع چگالی بفرم زیر نسبت به اندازه‌ی μ باشد

$$p_{\underset{\sim}{\theta}}(x) = \left\{ e^{\sum_{j=1}^{k} c_j(\theta) T_j(x) - B(\theta)} \right\} h(x) \qquad (1)$$

که در آن بایستی

۱- S_X^* بستگی به مقادیر $\underset{\sim}{\theta} = (\theta_1, \theta_2, ..., \theta_k)$ نداشته باشد.

۲- $c_j(\underset{\sim}{\theta})$, $j = 1, ..., k$ توابعی غیرصفر و پیوسته از $\underset{\sim}{\theta} = (\theta_1, ..., \theta_k)$ باشند.

۳- در حالت پیوسته $T_j'(x)$، $j = 1, 2, ..., k$، توابعی پیوسته روی S_X^* و در حالت گسسته $T_j(x)$، $j = 1, 2, ..., k$، توابعی غیر صفر روی S_X^* باشند.

٤- در حالت پیوسته $h(x)$ تابعی پیوسته روی S_X^* باشد.

خانواده $p_{\underset{\sim}{\theta}}$ در (۱) را یک خانواده نمایی پر رتبه[۱] با بُعد k گویند هرگاه شرایط ۱ تا ٤ بالا برقرار باشد و علاوه بر آن

الف- توابع $c_j(\underset{\sim}{\theta})$، $j = 1, 2, ..., k$، به طور تابعی از یکدیگر مستقل باشند.

ب- توابع $T_j(x)$، $j = 1, 2, ..., k$، به طور خطی از یکدیگر مستقل باشند.

ج – فضای پارامتری Θ یک شامل یک مستطیل K بعدی باشد. ■

مثال ۳-۱: خانواده توزیع‌های زیر به یک خانواده نمایی یک پارامتری متعلق می‌باشند.

(a) $B(1, \theta)$ $\theta \in (0, 1)$

(b) $B(n, \theta)$ $\theta \in (0, 1)$ n is $known$

(c) $NB(n, \theta)$ $\theta \in (0, 1)$ n is $known$

(g) $N(0, \sigma^2)$ $\sigma > 0$

(h) $N(\theta, \theta)$ $\theta > 0$

(i) $\Gamma(\alpha, \beta)$ $\beta > 0, \alpha$ $known$

[۱] Full Rank

$(d)\ Ge(\theta) \quad \theta \in (0,1)$ $(j)\ \Gamma(\alpha,\beta) \quad \alpha > 0, \beta \ known$

$(e)\ P(\theta) \quad \theta \in (0,+\infty)$ $(k)\ Be(\alpha,\beta) \quad \alpha > 0, \beta \ known$

$(f)\ N(\mu,1) \quad \mu \in R$ $(l)\ Be(\alpha,\beta) \quad \beta > 0, \alpha \ known$ ■

مثال۳-۲: خانواده توزیع‌های زیر به یک خانواده نمایی دو پارامتری پر رتبه متعلق است.

$(a) \quad N(\mu,\sigma^2) \quad \mu \in R, \sigma > 0$ $(b) \quad \Gamma(\alpha,\beta) \quad \alpha > 0, \beta > 0$

$(c) \quad Be(\alpha,\beta), \alpha > 0, \beta > 0$ ■

مثال۳-۳: خانواده‌ی $N(\theta,\theta^2), \theta > 0$ به یک خانواده نمایی متعلق است اما پررتبه نیست زیرا

$$p_\theta(x) = \frac{1}{\sqrt{2\pi\theta^2}} e^{-\frac{1}{2\theta^2}(x-\theta)^2} = \frac{1}{\sqrt{2\pi\theta^2}} e^{-\frac{1}{2\theta^2}x^2 + \frac{x}{\theta} - \frac{1}{2}} = \left\{ e^{-\frac{1}{2\theta^2}x^2 + \frac{1}{\theta}x - \frac{1}{2}\ln(2\pi\theta^2)} \right\} e^{-\frac{1}{2}}$$

$$c_1(\theta) = -\frac{1}{2\theta^2} \qquad T_1(x) = x^2 \qquad h(x) = e^{-\frac{1}{2}}$$

$$c_2(\theta) = \frac{1}{\theta} \qquad T_2(x) = x \qquad B(\theta) = +\frac{1}{2}\ln(2\pi\theta^2)$$

پارامتر دوبُعدی $(-\frac{1}{2\theta^2}, \frac{1}{\theta})$روی یک منحنی در فضای R^2 قرار می گیرد و به همین دلیل این خانواده را خانواده نمایی منحنی[1] گویند. ■

مثال۳-٤: خانواده توزیع‌های زیر به یک خانواده نمایی متعلق نیستند.

$(a) \quad U(0,\theta), \ \theta > 0$ $(b) \quad E(\alpha,\beta), \ \alpha \in \Re, \ \beta > 0$ ■

خانواده‌ی نمایی:

۱- اگر $X_1, X_2, ..., X_n$ یک نمونه تصادفی (i.i.d) از خانواده نمایی k پارامتری باشند آنگاه توزیع توأم $\underset{\sim}{X} = (X_1, ..., X_n)$ نیز به یک خانواده نمایی k-پارامتری متعلق است، زیرا

[1] Curved Exponential Family

$$p_{\underset{\sim}{\theta}}(x_1,...,x_n)=\prod_{i=1}^{n}p_\theta(x_i)=\prod_{i=1}^{n}e^{\sum_{j=1}^{k}c_j(\underset{\sim}{\theta})T_j(x_i)-B(\underset{\sim}{\theta})}h(x_i)$$

$$=...=e^{\sum_{j=1}^{k}c_j(\underset{\sim}{\theta})T_j^*(\underset{\sim}{x})-nB(\underset{\sim}{\theta})}h^*(\underset{\sim}{x})$$

که در آن $$T_j^*(\underset{\sim}{x})=\sum_{i=1}^{n}T_j(x_i)\quad and\quad h^*(\underset{\sim}{x})=\prod_{i=1}^{n}h(x_i)$$

۲- در یک خانواده نمایی یک پارامتری با تابع چگالی

$$p_\theta(x)=e^{c(\theta)T(x)-B(\theta)}h(x)$$

داریم که $$E_\theta(T(x))=\frac{B'(\theta)}{C'(\theta)}$$

زیرا $$\int e^{c(\theta)T(x)-B(\theta)}h(x)d\mu(x)=1\Rightarrow\int[c'(\theta)T(x)-B'(\theta)]p_\theta(x)d\mu(x)=0$$

$$\Rightarrow c'(\theta)E_\theta(T(x))-B'(\theta)=0\Rightarrow E_\theta(T(x))=\frac{B'(\theta)}{c'(\theta)}$$

فرم متعارف خانواده نمایی[1]

با تغییر پارامتر $\underset{\sim}{\eta}=c_j(\underset{\sim}{\theta})$، $j=1,2,...,k$ در خانواده توزیع‌های نمایی (۱) داریم که

$$p_{\underset{\sim}{\eta}}(x)=\left\{e^{\sum_{j=1}^{k}\eta_jT_j(x)-A(\underset{\sim}{\eta})}\right\}h(x)\qquad(2)$$

که این فرم را فرم متعارف خانواده توزیع‌های نمایی گویند. در این حالت مجموعه‌ی

$$\{\underset{\sim}{\eta}=(\eta_1,...,\eta_k)\mid\int_{-\infty}^{+\infty}h(x)e^{\sum_{j=1}^{k}\eta_jT_j(x)}d\mu(x)<\infty\}$$

را فضای پارامتری طبیعی و $\underset{\sim}{\eta}$ را پارامتر طبیعی گویند. خواص زیر را برای این خانواده می‌توان اثبات کرد.

[1] Canonical Form

۱- (کتاب TSH فصل ۲ ، قضیه ۹) برای هر تابع انتگرال‌پذیر f و هر η در فضای پارامتری طبیعی، انتگرال $\int f(x)e^{\sum_{j=1}^{k}\eta_j T_j(x)}h(x)d\mu(x)$ پیوسته و دارای مشتقات تمام مراتب نسبت به η_j ها است و این مشتقات می توانند به درون انتگرال برده شوند.

۲- (کتاب TSH بخش ۷. ۲ لم ۸) اگر X دارای توزیعی از خانواده نمایی با تابع چگالی بفرم (۲) نسبت به اندازه‌ی μ باشد آنگاه $\underset{\sim}{T}=(T_1,...,T_k)$ دارای توزیعی از خانواده نمایی با چگالی $p_{\underset{\sim}{\eta}}(t_1,...,t_k)=\left\{e^{\sum_{j=1}^{k}\eta_j t_j - A(\underset{\sim}{\eta})}\right\}k(\underset{\sim}{t})$ نسبت به یک اندازه ν می‌باشند (در استنباط آماری ۲ اثبات می‌گردد).

۳- فرض کنید $X_1,X_2,...,X_n$ متغیرهای تصادفی مستقل و هر کدام دارای توزیعی از خانواده نمایی با تابع چگالی زیر باشند

$$p_{\underset{\sim}{\eta}}(x_i)=\left\{e^{\eta T_j(x_i)-A_i(\eta)}\right\}h_i(x_i)$$

در این صورت $\sum_{i=}^{n}T_i(x_i)$ دارای توزیعی از خانواده نمایی یک پارامتری می باشد.

اگر دو خانواده نمایی k-پارامتری (با پارامترهای یکسان) مستقل داشته باشیم آنگاه $(T_1,...,T_k)$ و $(U_1,...,U_k)$ هر کدام دارای توزیعی از خانواده نمایی k پارامتری می‌باشند و طبق قضیه ۲، $(T_1+U_1,...,T_k+U_k)$ نیز دارای توزیعی از خانواده نمایی k- پارامتری است. به وسیله استقرا قضیه ۳ ثابت می‌شود.

٤- اگر X دارای توزیعی از خانواده نمایی k- پارامتری با تابع چگالی بفرم (۲) باشد آنگاه

$$E_\eta(T_j(x))=\frac{\partial}{\partial\eta_j}A(\underset{\sim}{\eta})$$
$$\operatorname{cov}_\eta(T_i(x),T_j(x))=\frac{\partial^2}{\partial\eta_i\partial\eta_j}A(\underset{\sim}{\eta})$$

مثال ۳-۵: فرض کنید $\underset{\sim}{X}=(X_1,...,X_k)$ دارای توزیع چند جمله‌ای با پارامترهای n و $p_1,...,p_k$ باشد. در این صورت:

$$p(\underset{\sim}{x}) = P(X_۱ = x_۱,...,X_k = x_k) = \frac{n!}{x_۱!x_۲!...x_k!} p_۱^{x_۱}...p_k^{x_k}$$

$$= h(\underset{\sim}{x}) e^{x_۱ \ln p_۱ + ... + x_k \ln p_k}$$

$$= h(\underset{\sim}{x}) e^{x_۱ \ln \frac{p_۱}{p_k} + x_۲ \ln \frac{p_۲}{p_k} + ... + x_{k-۱} \ln \frac{p_{k-۱}}{p_k}} \times e^{x_k \ln p_k + \sum_{i=۱}^{k-۱} x_i \ln p_k}$$

$$= h(\underset{\sim}{x}) e^{\sum_{i=۱}^{k-۱} x_i \ln \frac{p_i}{p_k} + n \ln p_k}$$

قرار می‌دهیم $\eta_i = \ln \frac{p_i}{p_k}$ در این صورت

$$A(\underset{\sim}{\eta}) = -n \ln p_k = n \ln(۱ + \sum_{i=۱}^{k-۱} e^{\eta_i})$$

$$(since \quad e^{\eta_i} = \frac{p_i}{p_k} \Rightarrow \sum_{i=۱}^{k-۱} e^{\eta_i} = \frac{۱ - p_k}{p_k} = \frac{۱}{p_k} - ۱ \Rightarrow p_k = (۱ + \sum_{i=۱}^{k-۱} e^{\eta_i})^{-۱})$$

$$\therefore \quad p_{\underset{\sim}{\eta}}(\underset{\sim}{x}) = \left\{ e^{\sum_{i=۱}^{k-۱} \eta_i x_i - A(\underset{\sim}{\eta})} \right\} h(\underset{\sim}{x})$$

که $\underset{\sim}{\eta} = (\eta_۱,...,\eta_{k-۱})$ پارامتر طبیعی است و

$$E(X_i) = \frac{\partial}{\partial \eta_i} A(\underset{\sim}{\eta}) = \frac{n e^{\eta_i}}{۱ + \sum_{i=۱}^{k} e^{\eta_i}} = \frac{n \frac{p_i}{p_k}}{\frac{۱}{p_k}} = np_i$$

$$\text{cov}(X_i, X_j) = \frac{\partial^۲}{\partial \eta_i \partial \eta_j} A(\underset{\sim}{\eta}) = ... = -np_i p_j \qquad \blacksquare$$

تابع مولد گشتاور و تابع مولد انباشته[۱]

اگر $\underset{\sim}{T} = (T_۱,....,T_k)$ آنگاه $\alpha_{r_۱,...,r_k} = E(T_۱^{r_۱}..T_k^{r_k})$ را گشتاور توأم $T_۱,....,T_k$ گویند و تابع مولد گشتاور $\underset{\sim}{T} = (T_۱,....,T_k)$ به صورت زیر تعریف می‌شود

$$M_{\underset{\sim}{T}}(u_۱,...,u_k) = E(e^{u_۱T_۱ + ... + u_kT_k})$$

[۱] Moment Generating fun. and Cumulant Generating fun.

اگر $M_{\underset{\sim}{T}}$ در یک همسایگی حول مبدأ موجود باشد و گشتاورهای $\alpha_{r_1,...,r_s}$ موجود باشند آنگاه بسط تابع $M_{\underset{\sim}{T}}$ به صورت زیر خواهد بود

$$M_{\underset{\sim}{T}}(u_1,...,u_k) = \sum_{r_1}\dots\sum_{r_k}\alpha_{r_1,...,r_k}\frac{u_1^{r_1}..u_k^{r_k}}{r_1!...r_k!}$$

همچنین تابع مولد انباشته $\underset{\sim}{T}$ به صورت زیر تعریف می‌شود

$$K_{\underset{\sim}{T}}(u_1,...,u_k) = \log M_{\underset{\sim}{T}}(u_1,...,u_k) = \sum_{r_1}\dots\sum_{r_k}k_{r_1,...,r_k}\frac{u_1^{r_1}..u_k^{r_k}}{r_1!...r_k!}$$

که در آن $k_{r_1,...,r_k}$ را انباشته توأم $\underset{\sim}{T} = (T_1,....,T_k)$ گویند. اگر تنها یک T داشته باشیم آنگاه

$$M_T(u) = E(e^{uT}) = \sum_r \alpha_r \frac{u^r}{r!}$$

$$K_T(u) = \ln M_T(u) = \sum_r k_r \frac{u^r}{r!}$$

و می‌توان نشان داد که (تمرین)

$$\alpha_1 = k_1 \qquad \alpha_2 = k_2 + k_1^2 \qquad \alpha_3 = k_3 + 3k_1k_2 + k_1^3$$

در خانواده‌ی نمایی از توزیع‌ها توابع $M_{\underset{\sim}{T}}(\underset{\sim}{u})$ و $K_{\underset{\sim}{T}}(\underset{\sim}{u})$ در یک همسایگی نزدیک صفر موجود هستند و:

$$K_{\underset{\sim}{T}}(\underset{\sim}{u}) = A(\underset{\sim}{\eta}+\underset{\sim}{u}) - A(\underset{\sim}{\eta}) \qquad M_{\underset{\sim}{T}}(\underset{\sim}{u}) = \frac{e^{A(\underset{\sim}{\eta}+\underset{\sim}{u})}}{e^{A(\underset{\sim}{\eta})}}$$

(اثبات: تمرین)

حال اگر $X_1, X_2, ..., X_n$ از یکدیگر مستقل باشند و $X = X_1 + X_2 + ... + X_n$ آنگاه

$$M_X(u) = E[e^{(X_1+...+X_n)u}] = \prod_{i=1}^{n} M_{x_i}(u) \qquad k_X(u) = \sum_{i=1}^{n} k_{X_i}(u)$$

مثال۳-۶: فرض کنید $X \sim B(n,p)$ در این صورت

$$p(x) = P(X = x) = \binom{n}{x} p^x q^{n-x} = \binom{n}{x} e^{x\ln\frac{p}{q} + n\ln q}$$

قرار می دهیم $\eta = \ln\frac{p}{q}$ در این صورت

$$A(\eta) = -n\ln q = n\ln(۱+e^{\eta}) \qquad p(x) = \binom{n}{x} e^{x\eta - A(\eta)}$$

$$M_x(u) = \frac{e^{A(\eta+u)}}{e^{A(\eta)}} = \frac{e^{n\ln(۱+e^{\eta+u})}}{e^{n\ln(۱+e^{\eta})}} = \left(\frac{۱+e^{\eta+u}}{۱+e^{\eta}}\right)^n = \left(\frac{۱+\frac{P}{q}e^{u}}{۱+\frac{P}{q}}\right)^n = (q+pe^{u})^n$$

مثال ۳-۷: فرض کنید $X \sim \Gamma(\alpha,\beta)$ در این صورت (α معلوم)

$$p(x) = \frac{۱}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-۱} e^{-\frac{x}{\beta}} = \left[e^{-\frac{۱}{\beta}x - \alpha\ln\beta}\right]\frac{x^{\alpha-۱}}{\Gamma(\alpha)}$$

$$\eta = -\frac{۱}{\beta} \qquad A(\eta) = \alpha\ln\beta = -\alpha\ln\left(\frac{۱}{\beta}\right) = -\alpha\ln(-\eta)$$

$$M_X(u) = \frac{e^{A(\eta+u)}}{e^{A(\eta)}} = \frac{e^{-\alpha\ln(-\eta-u)}}{e^{-\alpha\ln(-\eta)}} = \frac{\left(\frac{۱}{\beta}-u\right)^{-\alpha}}{\left(\frac{۱}{\beta}\right)^{-\alpha}} = \frac{۱}{(۱-\beta u)^{\alpha}}$$

$$k_X(u) = -\alpha\ln(۱-\beta u) \qquad u < \frac{۱}{\beta}$$

بنابراین

$$E(X^{r}) = \frac{\partial^r}{\partial u^r} M_X(u)\big|_{u=۰} = \alpha(\alpha+۱)\ldots(\alpha+r-۱)\beta^r = \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)}\beta^r$$

■

بخش ٤: آماره‌های بسنده[۱]

تعریف٤-۱: فرض $X_{۱},...,X_{n}$ یک نمونه تصادفی از یک خانواده از توزیع‌های $\mathsf{P}=\{P_{\theta}:\theta\in\Theta\}$ باشند (θممکن است یک بردار باشد) آماره $T(\underset{\sim}{X})=T(X_{۱},...,X_{n})$ را یک آماره‌ی بسنده برای θو یا به عبارت دقیق‌تر یک آماره بسنده برای خانواده توزیع‌های P گویند هر گاه توزیع شرطی نمونه تصادفی $X_{۱},...,X_{n}$ به شرط $T(\underset{\sim}{X})=t$ برای هر مقدار t بستگی به θ نداشته باشد.

توجه کنید که آماره‌های بسنده داده‌ها را در خود خلاصه می‌کنند بدون این که اطلاعی در مورد پارامتر مجهول را از دست بدهیم. یعنی داده‌ها را می‌توان در آماره‌ی بسنده T خلاصه کرد و با داشتن مقدار T نیازی به داشتن خود داده‌ها (تا آن جا که مربوط به استنباط درباره پارامترهای مجهول می‌شود) نداریم.

قضیه ٤-۱: اگر $\underset{\sim}{X}|T\overset{D}{=}\underset{\sim}{Y}|T$ آنگاه $\underset{\sim}{X}\overset{D}{=}\underset{\sim}{Y}$

اثبات:

$$\underset{\sim}{X}\text{ چگالی}=p_{\underset{\sim}{X}}(\underset{\sim}{x})=\int p_{\underset{\sim}{X}|T}(\underset{\sim}{x}|t)q(t)d\mu(t)=\int p_{\underset{\sim}{Y}|T}(\underset{\sim}{x}|t)q(t)d\mu(t)$$
$$=p_{\underset{\sim}{Y}}(\underset{\sim}{x})=\underset{\sim}{Y}\text{ چگالی}$$

■

عکس این قضیه برقرار نیست زیرا اگر در پرتاب سه سکه

$X=$ تعداد شیرها $Y=$ تعداد خط ها $T=$ تعداد شیر − تعداد خط

آنگاه $X\overset{D}{=}Y$ اما

$$\begin{cases}P(X=۱|T=۱)=\dfrac{P(X=۱,T=۱)}{P(T=۱)}=\dfrac{۳/۸}{۳/۸}=۱\\ P(X\neq۱|T=۱)=۰\end{cases}$$

$$\begin{cases}P(Y\neq۲|T=۱)=\dfrac{P(Y=۱,T=۱)}{P(T=۱)}=۰\\ P(Y=۲|T=۱)=۱\end{cases}$$

$$\Rightarrow X|T\overset{D}{\neq}Y|T$$

[۱] Sufficient Statistics

حال فرض کنید که مشاهدات $\underset{\sim}{X}$ را کنار گذاشته و مقدار $T=T(\underset{\sim}{X})$ را نگه داریم. اکنون $\underset{\sim}{X}'$ را از توزیع شرطی $\underset{\sim}{X}$ با فرض $T=t$ می‌سازیم (به وسیله شبیه‌سازی) چون $\underset{\sim}{X}|T \overset{D}{=} \underset{\sim}{X}'|T$ پس $\underset{\sim}{X}$ و $\underset{\sim}{X}'$ یک توزیع دارند. در نتیجه با نگه داشتن مقدار T می‌توانیم مشاهداتی همانند (هم‌توزیع) $\underset{\sim}{X}$ بسازیم. این کار را با آماره‌های غیربسنده نمی‌توان انجام داد، چون توزیع شرطی $\underset{\sim}{X}$ با فرض $T=t$ به پارامتر مجهول بستگی دارد و نمی‌توان $\underset{\sim}{X}'$ را بازیافت نمود.

مثال ٤-١: فرض کنید X_1, X_2 یک نمونه تصادفی از توزیع $P(\theta)$ باشد. نشان دهید که $T(\underset{\sim}{X})=X_1+X_2$ یک آماره بسنده برای θ است و از روی آن نمونه‌ی (X_1', X_2') هم توزیع با (X_1, X_2)را بسازید.

$$P(X_1=x_1, X_2=x_2 \mid T(\underset{\sim}{X})=t)=\begin{cases}\dfrac{P(X_1=x_1, X_2, x_2)}{P(T=t)} & x_1+x_2=t \\ 0 & x_1+x_2\neq t\end{cases}$$

$$= \begin{cases}\dfrac{\dfrac{e^{-\theta}\theta^{x_1}}{x_1!}\cdot\dfrac{e^{-\theta}\theta^{x_2}}{x_2!}}{\dfrac{e^{-2\theta}(2\theta)^t}{t!}} & x_1+x_2=t \\ 0 & x_1+x_2\neq t\end{cases} = \begin{cases}\dfrac{t!}{x_1!x_2!2^t} & x_1+x_2=t \\ 0 & x_1+x_2\neq t\end{cases}$$

$$= \begin{cases}\binom{t}{x_1}\left(\frac{1}{2}\right)^{x_1}\left(\frac{1}{2}\right)^{t-x_1} & x_1+x_2=t \\ 0 & x_1+x_2\neq t\end{cases}$$

$$\Rightarrow \underset{\sim}{X}|T=t \sim B(t,\tfrac{1}{2})$$

چون توزیع شرطی به θ بستگی ندارد، پس $T(\underset{\sim}{X})=X_1+X_2$ یک آماره‌ی بسنده برای θ است. حال اگر X_1' و $X_2'=t-X_1'$ را به ترتیب تعداد شیرها و تعداد خط‌ها در t پرتاب یک سکه در نظر بگیریم آن گاه $\underset{\sim}{X}'|t \sim B(t,\frac{1}{2})$ که باتوجه به قضیه ٤-١، $\underset{\sim}{X}' \overset{D}{=} \underset{\sim}{X}$ است. ■

استفاده از تعریف ٤-١ برای بررسی بسندگی یک آماره، نیاز به در دست داشتن آماره بسنده $T(\underset{\sim}{X})$ و همچنین بررسی توزیع شرطی نمونه به شرط T دارد. یک راه ساده برای بدست آوردن یک آماره‌ی

بسنده و بررسی بسندگی آن در قضیه دسته‌بندی نیمن که به قضیه تجزیه به عوامل مشهور است، آمده است.

قضیه ٤-۲: (قضیه تجزیه به عوامل[۱])

فرض کنید $X_1, X_2, \ldots, X_n$ دارای توزیع توأم متعلق به خانواده‌ی $\mathsf{P} = \{P_\theta : \theta \in \Theta\}$ که به وسیله اندازه‌ی μ مغلوب شده است، باشند. شرط لازم و کافی برای آن که آماره $T(\underset{\sim}{X})$ برای خانواده‌ی P بسنده باشد این است که توابع نامنفی g و h وجود داشته باشند بطوری که تابع چگالی p_θ مربوط به P_θ در رابطه زیر صدق کند.

$$p_\theta(\underset{\sim}{x}) = g(\theta, T(\underset{\sim}{x}))h(\underset{\sim}{x}) \qquad (a.e.\mu)$$

اثبات: (کتاب TSH بخش ٦. ۲ قضیه ۸ و نتیجه ۱) ■

مثال ٤-۲: فرض کنید $X_1, \ldots, X_n$ یک نمونه تصادفی از $Be(\theta, 1)$ باشند. یک آماره‌ی بسنده برای θ پیدا کنید.

$$f_\theta(\underset{\sim}{x}) = \prod_{i=1}^{n} \theta x_i^{\theta-1} = \begin{cases} \theta^n (\prod_{i=1}^{n} x_i)^{\theta-1} & = g(\theta, \prod_{i=1}^{n} x_i) \times 1 \\ \theta^n e^{(\theta-1)\sum_{i=1}^{n} \ln x_i} & = g(\theta, \sum_{i=1}^{n} \ln x_i) \times 1 \end{cases}$$

پس $T(X) = \prod_{i=1}^{n} X_i$ و $S(X) = \sum_{i=1}^{n} \ln X_i$ برای θ بسنده می‌باشند.

قضیه ۳: اگر T یک آماره‌ی بسنده برای θ و $T = k(U)$ آنگاه U نیز برای θ بسنده است.

اثبات: T بسنده است، بنابراین

$$p_\theta(\underset{\sim}{x}) = g(\theta, T(\underset{\sim}{x}))h(\underset{\sim}{x}) = g(\theta, k(U(\underset{\sim}{x}))h(\underset{\sim}{x}) = g^*(\theta, U(\underset{\sim}{x}))h(\underset{\sim}{x})$$

پس طبق قضیه تجزیه به عوامل $U(\underset{\sim}{X})$ یک آماره بسنده برای θ است. ■

بنابراین در مثال ٤-۲ $T = \ln(S)$ و $S = e^T$ هر دو برای θ بسنده هستند.

---

[۱] Factorization Theorem.

مثال ٤-٣: فرض کنید $X_1,...,X_n$ یک نمونه تصادفی از $N(\mu,\sigma^2)$ باشند. نشان دهید که $(\bar{X},S^2)$ یک آماره بسنده برای θ است.

مثال ٤-٤: فرض کنید $X_1,...,X_n$ یک نمونه تصادفی از $U(0,\theta)$ باشند. نشان دهید که $X_{(n)}$ یک آماره‌ی بسنده برای θ است.

مثال ٤-٥: فرض کنید $X_1,...,X_n$ یک نمونه تصادفی از یک توزیع P_θ باشند و قرار می‌دهیم $T=(X_{(1)},...,X_{(n)})$ چون کلیه مقادیر $\underset{\sim}{X}$ برابر $n!$ حالت $(X_{(i_1)},X_{(i_2)},...,X_{(i_n)})$ می‌باشد که توزیع شرطی $\underset{\sim}{X}$ به شرط هر کدام از این مقادیر برابر $\frac{1}{n!}$ است و به θ بستگی ندارد پس $T(\underset{\sim}{X})$ یک آماره بسنده برای θ است.

مثال ٤-٦: فرض کنید $X_1,...,X_n$ یک نمونه تصادفی از خانواده ی نمایی k پارامتری باشند در این صورت

$$p_{\underset{\sim}{\theta}}(\underset{\sim}{x})=\left[e^{\sum_{j=1}^{k}C_j(\underset{\sim}{\theta})\sum_{i=1}^{n}T_j(x_i)-nB(\underset{\sim}{\theta})}\right]\prod_{i=1}^{n}h(x_i)=g(\underset{\sim}{\theta},\sum_{i=1}^{n}T_1(x_i),...,\sum_{i=1}^{n}T_k(x_i))$$

پس $\underset{\sim}{T}=(\sum_{i=1}^{n}T_1(X_i),...,\sum_{i=1}^{n}T_k(X_i))$ یک آماره‌ی بسنده برای θ است.

آماره بسنده‌ی مینیمال:

فرض کنید $X_1,X_2,....,X_n$ یک نمونه تصادفی از توزیع $N(\theta,1)$ باشند و قرار دهید

$$T_1(\underset{\sim}{X})=(X_1,X_2,...,X_n)\qquad T_2(\underset{\sim}{X})=(\sum_{i=1}^{m}X_i,\sum_{i=m+1}^{n}X_i)\qquad T_3(\underset{\sim}{X})=\sum_{i=1}^{n}X_i$$

طبق قضیه‌ی تجزیه به عوامل $T_3(\underset{\sim}{X})$ یک آماره‌ی بسنده برای θ است و $T_3=k_1(T_2)$ $T_2=k_2(T_1)$ پس طبق قضیه ٤-٣، T_1 و T_2 نیز برای θ بسنده هستند. همچنین T_3 بیشتر از T_2 و T_2 بیشتر از T_1 داده‌ها را خلاصه می‌کنند.

حال این سوال مطرح است که داده‌ها را تا چه اندازه می‌توان خلاصه کرد. در پاسخ به این سوال بایستی گفت تا جایی می‌توان داده‌ها را خلاصه کرد که اطلاعات داده‌ها در مورد پارامترها از دست نرود. آماره‌ی بسنده‌ای که تا حد ممکن داده‌ها را خلاصه کند آماره‌ی بسنده‌ی مینیمال نامیده می‌شود.

تعریف ٤-۲: آماره‌ی $T(\underset{\sim}{X})$ یک آماره‌ی بسنده‌ی مینیمال است اگر اولاً این آماره بسنده باشد و ثانیاً برای هر آماره‌ی بسنده‌ی دیگر U آماره‌ی T تابعی از آن باشد یعنی $T = K(U)$.

نکات: ۱- آماره‌های بسنده مینیمال یکتا نیستند اما اگر T_1 و T_2 بسنده‌ی مینیمال باشند یک رابطه یک به یک بین آنها برقرار است.

۲- اگر T یک آماره‌ی بسنده‌ی مینیمال باشد و $U = K(T)$ و K یک تابع یک به یک نباشد آن گاه U یک آماره‌ی بسنده نیست.

به وسیله قضایای زیر روش بدست آوردن آماره‌ی بسنده‌ی مینیمال را بررسی می‌کنیم.

قضیه ٤-٤: آماره‌ی U برای θ بسنده است اگر و فقط اگر به ازای هر $\theta_1, \theta_2 \in \Theta$ ، $\frac{p(x;\theta_1)}{p(x:\theta_2)}$ تابعی از U باشد.

اثبات: (کفایت)

$$\frac{p(x;\theta_1)}{p(x;\theta_2)} = \ell(u;\theta_1,\theta_2) \Rightarrow p(x;\theta_1) = \ell(u;\theta_1,\theta_2)p(x;\theta_2)$$

اگر θ_2 را برابر مقدار خاص $\theta_.$ بگیریم آن گاه

$$p(x;\theta_1) = \ell(u;\theta_1,\theta_.)p(x;\theta_.) \quad \forall\theta_1$$
$$= g(U,\theta_1)h(x) \quad \forall\theta_1$$

پس طبق قضیه تجزیه به عوامل U برای θ بسنده است.

(لزوم): اگر U برای θ بسنده باشد آن گاه طبق قضیه تجزیه به عوامل داریم که

$$\left.\begin{array}{l} p(\underset{\sim}{x};\theta_1) = g(\theta_1,u)h(\underset{\sim}{x}) \\ p(\underset{\sim}{x};\theta_2) = g(\theta_2,u)h(\underset{\sim}{x}) \end{array}\right\} \Rightarrow \frac{p(\underset{\sim}{x};\theta_1)}{p(\underset{\sim}{x};\theta_2)} = \frac{g(\theta_1,u)}{g(\theta_2,\underset{\sim}{u})} = \ell(u;\theta_1,\theta_2)$$

■

قضیه ٤-٥: فرض کنید $P = \{p_0, p_1, ..., p_k\}$ خانواده‌ای متناهی متشکل از $k+1$ توزیع باشد که همگی دارای تکیه‌گاه یکسان هستند. در این صورت $T(x) = \left(\frac{P_1(x)}{P_0(x)}, ..., \frac{P_k(x)}{P_0(x)}\right)$ یک آماره بسنده مینیمال برای θ است.

اثبات:

$$p_j(x) = \frac{p_j(x)}{p_0(x)} p_0(x) = g(T(x), j)h(x) \qquad j = 1, ..., k$$

$$p_0(x) = 1 \times p_0(x) = 1 \times h(x)$$

لذا طبق قضیه تجزیه به عوامل $T(\underset{\sim}{X})$ یک آماره بسنده است. حال فرض کنید U یک آماره‌ی بسنده برای P باشد . طبق قضیه ٤-٤ بایستی T تابعی از U باشد پس T یک آماره‌ی بسنده‌ی مینیمال است. ■

قضیه ٤-٦: فرض کنید P خانواده‌ای از توزیع‌ها با تکیه‌گاه یکسان و $P_0 \subset P$ باشد. اگر T برای P_0 بسنده‌ی مینیمال و برای P بسنده باشد، آنگاه برای P بسنده‌ی مینیمال نیز هست.

اثبات: فرض کنید U یک آماره‌ی بسنده برای P است. پس U برای P_0 نیز بسنده است و لذا T تابعی از U است، پس T برای P بسنده‌ی مینیمال است. ■

مثال ٤-٧: فرض کنید $X_1, X_2, ..., X_n$ یک نمونه تصادفی از توزیع $N(\theta, 1)$ باشد. یک آماره‌ی بسنده‌ی مینیمال برای θ پیدا کنید.

$$P_0 = \{N(\theta_0, 1), N(\theta_1, 1)\}$$

$$\frac{p_{\theta_1}(\underset{\sim}{x})}{p_{\theta_0}(\underset{\sim}{x})} = \frac{e^{-\frac{1}{2}\sum_{i=1}^{n}(X_i - \theta_1)^2}}{e^{-\frac{1}{2}\sum_{i=1}^{n}(X_i - \theta_0)^2}} = e^{(\theta_1 - \theta_2)\sum_{i=1}^{n} X_i - \frac{1}{2}(\theta_1^2 - \theta_0^2)} = h(\sum_{i=1}^{n} X_i)$$

که h یک تابع یک به یک است. چون $\sum_{i=1}^{n} X_i$ برای P بسنده و برای P_0 بسنده مینیمال است پس برای خانواده‌ی $N(\theta, 1)$ بسنده‌ی مینیمال است.

مثال ٤-٨: فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از توزیع $U(\theta-\frac{1}{2}, \theta+\frac{1}{2})$ باشد. یک آماره‌ی بسنده‌ی مینیمال برای θ پیدا کنید.

$$T(\underset{\sim}{X}) = (X_{(1)}, X_{(n)})$$

قضیه ٤-٧: فرض کنید X دارای توزیعی از خانواده‌ی نمایی k-پارامتری باشد. در این صورت $T = (T_1, \ldots, T_k)$ یک آماره‌ی بسنده‌ی مینیمال برای این خانواده است به شرط آنکه یکی از شرایط زیر برای این خانواده برقرار باشد:

الف- این خانواده پررتبه باشد.

ب- فضای پارامتری شامل $k+1$ نقطه $\underset{\sim}{\eta}^{(0)}, \underset{\sim}{\eta}^{(1)}, \ldots, \underset{\sim}{\eta}^{(k)}$ باشد که فضای E_k را پدید می‌آورند. به این معنی که آنها به یک زیر فضای محض E_k تعلق ندارند.

اثبات: اگر شرایط (الف) یا (ب) برقرار باشد آنگاه می‌توان $\underset{\sim}{\eta}^{(1)}, \ldots, \underset{\sim}{\eta}^{(k)}$ را در داخل فضای پارامتری چنان یافت که مستقل خطی باشند. حال خانواده توزیع‌های $P_{\cdot} = \{p_{\underset{\sim}{\eta}^{(0)}}, \ldots, p_{\underset{\sim}{\eta}^{(k)}}\}$ را در نظر می‌گیریم که در آن $\underset{\sim}{\eta}^{(0)}$ یک ترکیب خطی غیرمحدب از $\underset{\sim}{\eta}^{(1)}, \ldots, \underset{\sim}{\eta}^{(k)}$ است. بنابراین

$$T(\underset{\sim}{X}) = \left(\frac{p(\underset{\sim}{x}, \underset{\sim}{\eta}^{(1)})}{p(\underset{\sim}{x}, \underset{\sim}{\eta}^{(0)})}, \ldots, \frac{p(\underset{\sim}{x}, \underset{\sim}{\eta}^{(k)})}{p(\underset{\sim}{x}, \underset{\sim}{\eta}^{(0)})}\right) = \left(e^{\sum_{r=1}^{k}(\eta_r^{(1)} - \eta_r^{(0)})t_r}, \ldots, e^{\sum_{r=1}^{k}(\eta_r^{(k)} - \eta_r^{(0)})t_r}\right)$$

و یا معادلاً

$$W(\underset{\sim}{x}) = \left(\sum_{r=1}^{k}(\eta_r^{(1)} - \eta_r^{(0)})t_r, \ldots, \sum_{r=1}^{k}(\eta_r^{(k)} - \eta_r^{(0)})t_r\right)$$

$$= [t_1, \ldots, t_k] \begin{bmatrix} \eta_1^{(1)} - \eta_1^{(0)} & \ldots & \eta_1^{(k)} - \eta_1^{(0)} \\ \vdots & & \\ \eta_k^{(1)} - \eta_k^{(0)} & \ldots & \eta_k^{(k)} - \eta_k^{(0)} \end{bmatrix} = [t_1, \ldots, t_k]B$$

یک آماره‌ی بسنده‌ی مینیمال برای P_0 است. چون $\underset{\sim}{\eta}^{(i)}$ها مستقل خطی هستند پس ستون‌های B مستقل خطی و در نتیجه B وارون‌پذیر است. بنابراین $(T_1, \ldots, T_k)$ برای P_0 بسنده‌ی مینیمال است و چون برای P بسنده است پس برای P بسنده مینیمال نیز هست. ■

مثال ٤-٩: فرض کنید $X_1, \ldots, X_n$ یک نمونه تصادفی از $N(\theta, \theta^2)$ باشند. در این صورت

$$P_\theta(\underset{\sim}{x}) = \{e^{-\frac{1}{2\theta^2}\sum x_i^2 + \frac{1}{\theta}\sum x_i - n\ln(2\pi\theta^2)}\}e^{-\frac{1}{n}}$$

بنابراین آماره $T = (\sum X_i, \sum X_i^2)$ برای θ بسنده است و این آماره برای θ بسنده مینیمال است زیرا در اینجا فضای پارامتری عبارت از $\underset{\sim}{\eta} = (\frac{1}{\theta}, -\frac{1}{2\theta^2})$ است و اگر نقاط

$$\underset{\sim}{\eta}^{(0)} = (1, -\frac{1}{2}) \qquad \underset{\sim}{\eta}^{(1)} = (\frac{1}{2}, -\frac{1}{8}) \qquad \underset{\sim}{\eta}^{(2)} = (\frac{1}{3}, -\frac{1}{18})$$

را در نظر بگیریم آن گاه ماتریس B فوق عبارت است از $\begin{bmatrix} \frac{1}{2}-1 & \frac{1}{3}-1 \\ -\frac{1}{8}+\frac{1}{2} & -\frac{1}{18}+\frac{1}{2} \end{bmatrix}$ که وارون‌پذیر است و طبق قضیه قبل T برای θ بسنده مینیمال است. ■

روش لهمن- شفه برای بدست آوردن آماره بسنده مینیمال

قضیه ٤-٨: خانواده توزیع‌های $\{P_\theta : \theta \in \Theta\}$ با چگالی p_θ را در نظر بگیرید. اگر $T(\underset{\sim}{X})$ آماره‌ای باشد که برای هر دو نقطه $\underset{\sim}{x}$ و $\underset{\sim}{y}$ در تکیه گاه $S^*_{\underset{\sim}{X}}$ داشته باشیم که

$$\frac{p_\theta(\underset{\sim}{x})}{p_\theta(\underset{\sim}{y})} = \text{مستقل از } \theta = K(\underset{\sim}{x}, \underset{\sim}{y}) \Leftrightarrow T(\underset{\sim}{x}) = T(\underset{\sim}{y})$$

آن گاه T برای این خانواده بسنده مینیمال است.

اثبات: ابتدا ثابت می‌کنیم که $T(\underset{\sim}{X})$ یک آماره بسنده است. افرازهای ایجاد شده توسط $T(\underset{\sim}{X})$ را به صورت $A_t = \{\underset{\sim}{x} \mid T(\underset{\sim}{x}) = t\}$ نشان می‌دهیم. فرض کنید $\underset{\sim}{U}_t$ یک نقطه به خصوص در A_t باشد. حال اگر $\underset{\sim}{x} \in A_t$ باشد در این صورت

$$\left.\begin{matrix} \underset{\sim}{x} \in A_t \\ \underset{\sim}{U}_t \in A_t \end{matrix}\right\} \Rightarrow T(\underset{\sim}{x}) = T(\underset{\sim}{u}_t) \Rightarrow p_\theta(\underset{\sim}{x}) = K(\underset{\sim}{x}, \underset{\sim}{U}_t) p_\theta(\underset{\sim}{u}_t)$$

$$\Rightarrow p_\theta(\underset{\sim}{x}) = K(\underset{\sim}{x}, \underset{\sim}{U}_{T(\underset{\sim}{x})}) f_\theta(\underset{\sim}{U}_{T(\underset{\sim}{x})}) = h(\underset{\sim}{x}) g(\theta, T(\underset{\sim}{x}))$$

پس بنا به قضیه تجزیه به عوامل، $T(\underset{\sim}{X})$ یک آماره بسنده است.

حال فرض کنید U یک آماره بسنده دیگر باشد در این صورت

$$U(\underset{\sim}{x})=U(\underset{\sim}{y})\Rightarrow\frac{P_\theta(\underset{\sim}{x})}{P_\theta(\underset{\sim}{y})}=\frac{g(\theta,U(\underset{\sim}{x}))}{g(\theta,U(\underset{\sim}{y}))}=\frac{h(\underset{\sim}{x})}{h(\underset{\sim}{y})}=K(\underset{\sim}{x},\underset{\sim}{y})\quad\theta$$ مستقل از

لذا بنا به فرض $T(\underset{\sim}{x})=T(\underset{\sim}{y})$. یعنی T تابعی از U است پس T بسنده مینیمال است. ■

مثال ٤-١٠: فرض کنید $X_1,X_2,\ldots,X_n$ یک نمونه تصادفی از توزیع $Pa(\alpha,\theta)$ باشند یک آماره بسنده مینیمال برای (α,θ) پیدا کنید.

$$p_{\alpha,\theta}(\underset{\sim}{x})=\prod_{i=1}^{n}\frac{\theta\alpha^\theta}{x_i^{\theta+1}}I_{(\alpha,+\infty)}(x_i)=\frac{\theta^n\alpha^{n\theta}}{(\prod_{i=1}^{n}x_i)^{\theta+1}}I_{(0,x_{(1)})}(\alpha)$$

$$\frac{p_{\alpha,\theta}(\underset{\sim}{x})}{p_{\alpha,\theta}(\underset{\sim}{y})}=\frac{(\prod_{i=1}^{n}y_i)^{\theta+1}}{(\prod_{i=1}^{n}x_i)^{\theta+1}}.\frac{I_{(0,x_{(1)})^{(\alpha)}}}{I_{(0,y_{(1)})^{(\alpha)}}}=1\Leftrightarrow(\prod_{i=1}^{n}x_i,x_{(1)})=(\prod_{i=1}^{n}y_i,y_{(1)})$$

پس $T(\underset{\sim}{X})=(\prod_{i=1}^{n}X_i,X_{(1)})$ یک آماره بسنده مینیمال برای (α,θ) است. ■

توجه: با توجه به رابطه آماره‌های بسنده و آماره‌های بسنده مینیمال، روش تشخیص بسندگی یا عدم بسندگی یک آماره به صورت زیر می باشد. فرض کنید T یک آماره دلخواه باشد. برای تشخیص بسندگی یا عدم بسندگی آماره T، ابتدا توسط روش لهمن – شفه یک آماره بسنده مینیمال S پیدا می‌کنیم اگر S تابعی از آماره T باشد آنگاه T بسنده است و اگر S تابعی از T نباشد. آن گاه T یک آماره بسنده نخواهد بود (اثبات: برهان خلف).

آماره‌های فرعی[1]:

تعریف٤-٣: یک آماره $T(\underset{\sim}{X})$ را فرعی گوئیم هرگاه توزیع آن به پارامتر مجهول بستگی نداشته باشد و یک آماره $T(\underset{\sim}{X})$ را فرعی از مرتبه اول گوئیم هرگاه $E_\theta[T(\underset{\sim}{X})]=c$ و به θ بستگی نداشته باشد.

مثال ٤-١١: اگر $X_1,\ldots,X_n$ یک نمونه تصادفی از یک خانواده مکان $f(x-\theta)$ باشند آنگاه $S(\underset{\sim}{X})=X_{(n)}-X_{(1)}$ یک آماره فرعی است زیرا

[1] Ancillary statistics

$$X_i = Z_i + \theta \qquad i = ۱,...,n$$

Z_i دارای تابع چگالی $f(x)$ است که به θ بستگی ندارد.

$$R = X_{(n)} - X_{(۱)} \Rightarrow F_R(r) = P(R \le r) = P(X_{(n)} - X_{(۱)} \le r)$$
$$= P(\max_{۱\le i\le n} X_i - \min_{۱\le i\le n} X_i \le r) = P(\max_{۱\le i\le n}(Z_i + \theta) - \min_{۱\le i\le n}(Z_i + \theta) \le r)$$
$$= P(Z_{(n)} - Z_{(۱)} \le r)$$

که این احتمال به θ بستگی ندارد پس R یک آماره فرعی است. ■

مثال ٤-۱۲: اگر $X_۱,...,X_n$ یک نمونه تصادفی از یک خانواده مقیاس $\frac{۱}{\sigma} f(\frac{x}{\sigma})$ باشند آنگاه

$$S(\underset{\sim}{X}) = (Y_۱,...,Y_{n-۱}) = (\frac{X_۱}{X_n},...,\frac{X_{n-۱}}{X_n})$$

یک آماره فرعی برای σ است زیرا

$$F_{Y_۱,...,Y_{n-۱}}(y_۱,...,y_{n-۱}) = P(\frac{X_۱}{X_n} \le y_۱,...,\frac{X_{n-۱}}{X_n} \le y_{n-۱})$$
$$= P(\frac{X_۱/\sigma}{X_n/\sigma} \le y_۱,...,\frac{X_{n-۱}/\sigma}{X_n/\sigma} \le y_{n-۱})$$
$$= P(\frac{Z_۱}{Z_n} \le y_۱,...,\frac{Z_{n-۱}}{Z_n} \le y_{n-۱})$$

که در آن $Z_i = \frac{X_i}{\sigma}$ و $f_{Z_i}(z) = \sigma f_X(\sigma z) = f(z)$ بنابراین احتمال فوق به σ بستگی ندارد پس $S(\underset{\sim}{X})$ یک آماره فرعی است. ■

توجه: همانگونه که می‌دانیم آماره‌های بسنده جهت کاهش بُعد داده‌ها مورد استفاده قرار می‌گیرند. آن آماره بسنده‌ای می‌تواند این کاهش بُعد را تا حد ممکن انجام دهد که هیچ تابع غیرثابت آن یک آماره فرعی نباشد یعنی اگر $E[f(T)] = c$ نتیجه دهد که $f(T) = c \quad a.e.$ که با کم کردن c از طرفین، تعریف آماره کامل حاصل می‌شود.

بخش ۵: کامل بودن[۱]

فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از خانواده توزیع‌های $P = \{P_\theta : \theta \in \Theta\}$ باشند اگر $T(\underset{\sim}{X})$ یک آماره باشد آنگاه گویند $P^T = \{P_\theta^T : \theta \in \Theta\}$ خانواده توزیع‌های تولید شده توسط آماره T است.

تعریف ۵-۱: الف- خانواده P_θ^T را کامل گویند اگر برای هر تابع $h(T)$ داشته باشیم که

$$E_\theta[h(T)] = 0 \quad \forall \theta \in \Theta \Rightarrow \quad h_{(T)} = 0 \quad a.e.\ P^T \qquad (1)$$

و آماره $T(\underset{\sim}{X})$ را کامل گویند اگر خانواده توزیع‌های به وجود آمده توسط T کامل باشد (یعنی خانواده‌ای کامل است که در آن تنها برآوردگر نااریب صفر خود برآوردگر صفر باشد).

ب- خانواده P_θ^T را کامل کراندار گویند اگر برای هر تابع حقیقی کراندار $h(T)$ رابطه (۱) برقرار باشد.

توجه: توجه کنید که اگر یک آماره T کامل باشد آن گاه کامل کراندار نیز است اما عکس این مطلب برقرار نیست. و همچنین اگر T یک آماره کامل باشد آنگاه هر تابع یک به یک آن نیز کامل است.

مثال ۱: فرض کنید $X \sim B(\theta, \frac{1}{2})$، نشان دهید که

الف- اگر $\Theta = \{0,1,2,\ldots\}$ آنگاه $P = \{B(\theta, \frac{1}{2}) \mid \theta \in \Theta\}$ کامل است.

ب- اگر $\Theta^* = \{1,2,\ldots\}$ آنگاه $P^* = \{B(\theta, \frac{1}{2}) \mid \theta \in \Theta\}$ کامل نیست.

حل:

$$E_\theta[h(X)] = 0 \Rightarrow \sum_{x=0}^{\theta} h(x)\binom{\theta}{x}\left(\frac{1}{2}\right)^\theta = 0 \Rightarrow \sum_{x=0}^{\theta} h(x)\binom{\theta}{x} = 0 \qquad \forall \theta \in \Theta$$

الف-

$$\theta = 0 \Rightarrow h(0) = 0$$
$$\theta = 1 \Rightarrow h(0) + h(1) = 0 \Rightarrow h(1) = 0$$
$$\theta = k : h(0) = h(1) = \ldots = h(k) = 0 \qquad \text{فرض}$$

[۱] Completeness

$\theta = k+۱ : h(k+۱) = ۰$ حکم

$$h(۰)+h(۱)\binom{k+۱}{۱}+\ldots+h(k)\binom{k+۱}{k}+h(k+۱)=۰ \overset{(۲)}{\Rightarrow} h(k+۱)=۰$$

$$\therefore \quad h(x)=۰ \quad a.e.\ P$$

که نتیجه (۲) از فرض حاصل می‌گردد

ب-

$$\theta = ۱ \Rightarrow h(۰)+h(۱)=۰ \Rightarrow h(۱) = -h(۰) = -a$$

$$\theta = ۲ \Rightarrow h(۰)+۲h(۱)+h(۲)=۰ \Rightarrow h(۲) = -a$$

$$\vdots$$

$$h(x) = \begin{cases} a & x = ۰,۲,\ldots. \\ -a & x = ۱,۳,\ldots. \end{cases} = (۱)^x a, \qquad a = h(۰)$$

پس اگر $a \neq ۰$ آنگاه P یک خانواده کامل نیست. ■

توجه: مثال قبل نشان می‌دهد که کامل بودن به خانواده توزیع ها بستگی دارد و نه به آماره $T(\underset{\sim}{X})$.

مثال ۵-۲: فرض کنید X یک متغیر تصادفی گسسته با تابع احتمال زیر باشد.

$$p_\theta(x) = \begin{cases} \theta & x = ۱ \\ \theta^{x-۲}(۱-\theta)^۲ & x = ۲,۳,\ldots \end{cases} \qquad ۰ < \theta < ۱$$

نشان دهید که X یک آماره کامل نیست اما کامل کراندار است.

حل:

$$۰ = E_\theta[h(X)] = \sum_{x=۱}^{+\infty} h(x)p_\theta(x) = \theta h(۱) + \sum_{x=۲}^{+\infty} h(x)\theta^{x-۲}(۱-\theta)^۲$$

$$\Rightarrow \sum_{x=۲}^{+\infty} h(x)\theta^{x-۲} = -\theta h(۱)(۱-\theta)^{-۲}$$

با توجه به اینکه $\frac{۱}{(۱-\theta)^۲} = \sum_{x=۰}^{\infty}(x+۱)\theta^x$ داریم:

$$\overset{y=x-۲}{\Rightarrow} \sum_{y=۰}^{+\infty} h(y+۲)\theta^y = -h(۱)\sum_{x=۰}^{+\infty}(x+۱)\theta^{x+۱} = -h(۱)\sum_{y=۱}^{+\infty} y\,\theta^y$$

$$\Rightarrow \begin{cases} h(۲)=۰ \\ h(y+۲)=-yh(۱) \end{cases} \quad y=۱,۲,۳,... \quad \Rightarrow h(x)=-(x-۲)h(۱) \qquad x=۱,۲,۳,....$$

بنابراین X کامل نیست زیرا برای $E_\theta[h(X)]=۰$ نتیجه گرفتیم که $h(x)\neq ۰$ (با قرار دادن $h(۱)\neq ۰$). اما X یک آماره کامل کراندار است زیرا اگر $h(x)$ یک تابع کراندار باشد و $h(۱)\neq ۰$ آنگاه $\lim_{x\to\pm\infty} h(x)=\pm\infty$. بنابراین برای اینکه $h(x)$ یک تابع حقیقی کراندار باشد بایستی $h(۱)=۰$ باشد پس $h(x)=۰, \quad \forall x=۱,۲,.....$ بنابراین X یک آماره کامل کراندار است. ■

قضیه ۵-۱: اگر X دارای توزیعی از خانواده نمایی k-پارامتری پررتبه باشد آنگاه آماره $T_{\underset{\sim}{}}=(T_۱,...,T_k)$ برای این خانواده کامل است.

اثبات: (کتاب TSH بخش ۳ .٤ قضیه ۱) ■

با توجه به قضیه فوق می‌توان نتیجه گرفت که در توزیع $N(\mu,\sigma^۲)$ آماره $T(\underset{\sim}{X})=(\bar{X},S^۲)$ و در توزیع $B(n,\theta)$ آماره $T(\underset{\sim}{X})=\bar{X}$ و در توزیع $Be(\theta,۱)$ آماره $T(\underset{\sim}{X})=\sum \ln X_i$ آماره‌های کامل (و بسنده مینیمال) می‌باشند. مثالهای دیگر در غیر از خانواده نمایی در کتاب خوانده شود.

همچنین توجه کنید که اگر خانواده نمایی پررتبه نباشد آنگاه آماره $T_{\underset{\sim}{}}=(T_۱,...,T_k)$ کامل نیست. برای مثال $N(\theta,\theta^۲)$.

مثال ۵-۳: فرض کنید $X_۱,\ldots,X_n$ یک نمونه تصادفی از $U(\theta,\theta+۱)$ باشد. در این صورت به راحتی دیده می‌شود که $U(\underset{\sim}{X})=(X_{(۱)},X_{(n)})$ و یا معادلاً $T(\underset{\sim}{X})=(X_{(n)}-X_{(۱)},\frac{X_{(۱)}+X_{(n)}}{۲})$ یک آماره بسنده مینیمال توأم برای θ است و $S(\underset{\sim}{X})=X_{(n)}-X_{(۱)}$ یک آماره فرعی است و چون S قسمتی از آماره T است پس S و T از یکدیگر مستقل نمی‌باشند.

توجه: توجه کنید که آماره بسنده مینیمال تمام اطلاعات را در مورد پارامتر مجهول θ دارد اما آماره فرعی هیچ اطلاعی در مورد θ ندارد و به نظر می‌رسد که بایستی از یکدیگر مستقل باشند. اما در مثال بالا چنین نبود.

قضیه ۵-۲ (قضیه باسو[۱]): اگر T یک آماره بسنده و کامل کراندار برای خانواده $P = \{p_\theta : \theta \in \Theta\}$ باشد و V یک آماره فرعی برای P باشد آنگاه T و V از یکدیگر مستقل هستند.

اثبات: فرض کنید E یک پیشامد باشد و قرار دهیم $T_E(w) = \begin{cases} ۱ & w \in E \\ ۰ & w \notin E \end{cases}$ در این صورت I_E یک متغیر تصادفی کراندار است و $E(I_E) = P(E)$. حال نشان می‌دهیم که برای هر مجموعه برل A داریم که

$$P(V \in A \mid T = t) = P(V \in A)$$

قرار می دهیم $E = (V \in A)$ در این صورت

$$P(V \in A) = P(E) = E(I_E) \sim> \quad \theta \text{ مستقل از}$$

$$P(V \in A \mid T = t) - P(V \in A) = P(E \mid T = t) - P(E) = E[I_E \mid T = t] - E[I_E]$$
$$= h(t) \quad \theta \text{ مستقل از}$$

$$h(T) = E[I_E \mid T] - E[I_E]$$

$h(T)$ تابعی کراندار است و $E_\theta[h(T)] = E\{E[I_E \mid T = t]\} - E\{E(I_E)\}$

$$= E(I_E) - E(I_E) = ۰ \overset{(۳)}{\Rightarrow} h(T) = ۰ \quad a.e.\ P$$

که در آن (۳) از کامل کراندار بودن T نتیجه می‌گردد. پس

$$P(V \in A \mid T = t) = P(V \in A) \quad a.e.\ P$$

■

قضیه ۵-۳ (قضیه بهادر[۲]): اگر یک آماره بسنده کامل باشد آنگاه آن آماره بسنده مینیمال است.

اثبات: فرض کنید U یک آماره بسنده کامل و T یک آماره بسنده مینیمال باشد، در این صورت $T = K(U)$ و بنابراین

$$E_\theta(U) = E_\theta[E(U \mid T)] = E_\theta[h(T)] = E_\theta[h(K(U))]$$
$$\Rightarrow E_\theta[U - h(k(U))] = ۰ \overset{(۴)}{\Rightarrow} U = h(k(U)) \quad a.e.\ P$$

که در آن (٤) از کامل بودن U نتیجه می‌گردد.

---

[۱] Basu's Theorem
[۲] Bahadour's Theorem

$\Rightarrow U = h(T) \quad a.e.\ P \Rightarrow$ بسنده می‌نیمال است. U ■

توجه: توجه کنید که عکس قضیه بالا درست نیست یعنی اگر یک آماره بسنده مینیمال باشد آنگاه آن آماره ممکن است کامل نباشد. برای مثال توزیع‌های $N(\theta,\theta^2)$ و $U(\theta,\theta+1)$.

سوال: آیا آماره کاملی سراغ دارید که بسنده نباشد.

جواب: $P(U=c)=1$ در این صورت

$$E[h(U)] = E[h(c)] = 0 \Rightarrow P(h(c)=0) = 1$$

پس $U=c$ یک آماره کامل است ولی بسنده نیست. ■

توجه: با توجه به قضیه بهادر، اگر بخواهیم یک آماره کامل در خانواده P پیدا کنیم ابتدا در این خانواده طبق روش لهمن شفه یک آماره بسنده مینیمال پیدا می‌کنیم. اگر این آماره کامل بود که مسئله حل شده است و اگر نبود آنگاه طبق قضیه بهادر می‌توان نتیجه گرفت که در این خانواده آماره بسنده کامل وجود ندارد.

بخش ٦: توابع زیان محدب[1]

فرض کنید در برآورد پارامتر مجهول $\theta \in \Theta$ برآوردگر $\delta \in \mathsf{D}$ (D کلاس کلیه برآوردگرها) را در نظر بگیریم. مهمترین خصوصیت یک برآوردگر خوب آن است که برآوردی را برای θ بدست دهد که تا حد ممکن به مقدار واقعی θ نزدیک باشد. میزان این نزدیکی را با تابع زیان اندازه‌گیری می‌کنیم که با نماد $L(\theta,\delta(\underset{\sim}{X}))$ نمایش داده می‌شود و

$$L:\Theta\times\mathsf{D}\to[0,+\infty)$$
$$L(\theta,\delta)\geq 0 \quad \forall\theta,\delta$$
$$L(\theta,\delta)=0 \quad if \quad \delta=\theta$$

چند تابع زیان مشهور عبارتند از :

تابع زیان درجه دوم یا مربع خطا $(i)\ L(\theta,T)=(\delta-\theta)^2$

تابع زیان قدر مطلق $(ii)\ L(\theta,T)=|\delta-\theta|$

[1] Convex Loss Function

تابع زیان LINEX $(iii)\ L(\theta,T)=b\{e^{a(\delta-\theta)}-a(\delta-\theta)-1\}$

چون تابع زیان بستگی به $\delta(\underset{\sim}{X})$ دارد پس مقدار آن ثابت نیست و یک متغیر تصادفی است. برای مقایسه دو برآوردگر، امید ریاضی تابع زیان این دو برآوردگر را که تابعی از θ می‌باشد با یکدیگر مقایسه می‌کنند که به آن تابع مخاطره[1] گویند. پس

$$R(\theta,\delta)=E_{\theta}[L(\theta,\delta(\underset{\sim}{X}))]$$

در خصوص توابع زیان محدب قضایای مهمی وجود دارد که در این بخش ابتدا توابع محدب را معرفی و سپس این قضایا را بیان می‌کنیم.

تعریف ٦-۱: تابع حقیقی φ را روی فاصله $I=(a,b)$ که $-\infty\le a<b\le\infty$ محدب گوئیم هرگاه برای هر $a<x<y<b$ و هر $0<\delta<1$ داشته باشیم

$$\varphi(\gamma x+(1-\gamma)y)\le\gamma\varphi(x)+(1-\gamma)\varphi(y)$$

و آن را اکیداً محدب گوئیم هرگاه نامساوی اکید باشد.

می‌توان نشان داد که توابع محدب، توابعی پیوسته روی (a,b) هستند.

قضیه ٦-۱: الف – اگر φ روی (a,b) مشتق‌پذیر باشد آنگاه شرط لازم و کافی برای آنکه φ محدب باشد آن است که

$$\varphi'(x)\le\varphi'(y)\qquad\forall\ a<x<y<b$$

و اکیداً محدب است اگر نامساوی اکید باشد.

ب – اگر φ روی (a,b) دوبار مشتق‌پذیر باشد آنگاه شرط لازم و کافی برای آن که φ محدب باشد آن است که

$$\varphi''(x)\ge 0\qquad\forall\ a<x<y<b$$

و اکیداً محدب است اگر نامساوی اکید باشد. ■

[1] Risk Function

قضیه ٦-٢: اگر φ یک تابع محدب روی یک فاصله باز I باشد و X یک متغیر تصادفی با $P(X \in I) = ۱$ و امید ریاضی متناهی باشد آنگاه $\varphi(E(X)) \le E[\varphi(X)]$ و اگر φ اکیداً محدب باشد آنگاه نامساوی اکید است، بجز اینکه X با احتمال یک روی I ثابت باشد.

اثبات: (تمرین) ■

نتیجه: اگر X یک متغیر تصادفی ناثابت مثبت با امید متناهی باشد آنگاه

$$\frac{۱}{E(X)} \le (\frac{۱}{X}), \qquad E(\log X) < \log(E(X))$$

برای توابع زیان محدب قضایای زیر را داریم.

قضیه ٦-٣ (قضیه رائوبلاکول): فرض کنید X یک متغیر تصادفی با توزیعی از خانواده $\mathsf{D} = \{p_\theta : \theta \in \Theta\}$ باشد و T یک آماره بسنده برای D باشد. فرض کنید δ برآوردگری برای تابع پارامتری $g(\theta)$ باشد و تابع زیان $L(\theta,d)$ یک تابع اکیداً محدب بر حسب d باشد. اگر δ دارای امید و مخاطره متناهی $R(\theta,\delta) = E\{L(\theta,\delta(X))\} < \infty$

باشد و قرار دهیم $g(T) = E[\delta(X)|T]$ آنگاه

$$R(\theta,g) < R(\theta,\delta)$$

بجز اینکه با احتمال یک $g(T) = \delta(X)$ باشد.

اثبات: قرار می‌دهیم $\varphi(d) = L(\theta,d)$ و $\delta = \delta(X)$ و فرض کنید $P^{x|t}$ توزیع شرطی X به شرط $T = t$ باشد. در این صورت طبق نامساوی جنسن داریم که

$$L(\theta,E[\delta(X)|T]) < E\{L(\theta,\delta(X))|T\} \Rightarrow L(\theta,g(T)) < E\{L(\theta,\delta(X))|T\}$$

بجز اینکه با احتمال یک $\delta(X) = g(T)$ باشد . حال اگر از طرفین امید ریاضی بگیریم آنگاه

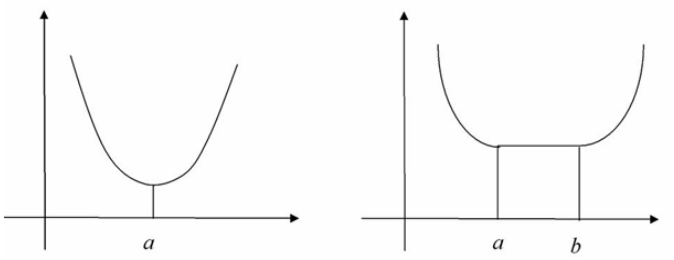
$$E[L(\theta,g(T))] < E\{E[L(\theta,\delta(X))|T]\} = E[L(\theta,\delta(X))] \Rightarrow R(\theta,g) < R(\theta,\delta)$$

بجز اینکه با احتمال یک $\delta(X) = g(T)$ باشد. ■

توجه: قضیه رائوبلاکول می‌گوید که با شرطی کردن یک برآوردگر روی یک آماره بسنده، برآوردگری حاصل می‌گردد که دارای مخاطره کمتری نسبت به برآوردگر اولیه است و این یک توجیه مناسب برای

استفاده از آماره‌های بسنده در برآوردیابی می‌باشد. همچنین توجه کنید که از بسندگی تنها برای این استفاده شد که بگوئیم $g(T)=E[\delta(X)|T]$ یک برآوردگر است و به پارامتر θ بستگی ندارد. اگر در قضیه تابع زیان تنها محدب باشد آنگاه نامساوی به صورت $R(\theta,g)\le R(\theta,\delta)$ تبدیل می‌شود.

لم ٦-١: فرض کنید φ یک تابع محدب روی $(-\infty,+\infty)$ باشد که از پائین کراندار و همچنین غیر یکنوا است. در این صورت φ دارای یک مینیمم مقدار است و مجموعه مقادیر S که φ مقدار مینیمم خود را می‌گیرد یک فاصله بسته است و اگر φ اکیداً محدب باشد این مجموعه تنها یک نقطه است.



اثبات: چون φ محدب و غیر یکنواست پس $\lim_{x\to\pm\infty}\varphi(x)=+\infty$ و چون φ پیوسته است مقدار مینیمم خود را می‌گیرد. چون φ محدب است، مجموعه S یک فاصله است و چون φ پیوسته است، پس این مجموعه بسته است. ■

قضیه ٦-٤: فرض کنید ρ یک تابع محدب باشد که روی $(-\infty,+\infty)$ تعریف شده و X یک متغیر تصادفی باشد که $\varphi(a)=E[\rho(X-a)]$ برای بعضی مقادیر a متناهی باشد. اگر ρ یکنوا نباشد آنگاه $\varphi(a)$ مقدار مینیمم خود را روی یک فاصله بسته می‌گیرد. اگر ρ اکیداً محدب باشد مقدار مینیمم یکتا خواهد بود.

اثبات: چون $a\to\pm\infty \quad as \quad X-a\xrightarrow{P}\pm\infty$ پس $\lim_{t\to\pm\infty}\rho(t)=+\infty$ پس $\lim_{a\to\pm\infty}\varphi(a)=+\infty$ پس φ غیر یکنواست و چون ρ (اکیداً) محدب است پس φ نیز (اکیداً) یکنواست. حال که شرایط لم ٦-١ برای φ برقرار است پس نتیجه از لم ٦-١ حاصل می‌شود. ■

مثال ٦-١: (تابع زیان درجه دوم) : اگر $\rho(t)=t^2$ و $E(X^2)<\infty$، چون ρ تابعی اکیداً محدب است پس $\varphi(a)=E[(X-a)^2]$ مقدار مینیمم خود را در نقطه $a=\mu$ اختیار می‌کند.

مثال ٦-٢: (تابع زیان قدرمطلق خطا) : اگر $\rho(t)=|t|$ و $E(|X|)<\infty$، چون ρ تابعی محدب است پس $\varphi(a)=E[|X-a|]$ مقدار مینیمم خود را در نقاط یک فاصله بسته بدست می‌آورد که همان میانه X است.

نتیجه ٦-١: تحت شرایط قضیه ٦-٤ اگر ρ زوج باشد و X حول μ متقارن باشد آنگاه $\varphi(a)$ مینیمم خود را در $a=\mu$ بدست می آورد.

اثبات: بر طبق قضیه ٦-٤، φ مقدار مینیمم خود را بدست می آورد. اگر $\mu+c$ یک مقدار مینیمم باشد آنگاه $\mu-c$ نیز مینیمم می‌باشد و در نتیجه کلیه نقاط در فاصله $\mu-c$ تا $\mu+c$ مینیمم می‌باشند پس $a=\mu$ نیز مینیمم است. ■

(تعمیم توابع محدب به فضای k بُعدی $\Re^k$ به عهده دانشجو)

(مبحث همگرایی‌ها بخش ۸ به عهده دانشجو)

فصل ۲

ناآریبی

در برآورد نقطه‌ای می‌خواهیم برآوردگری را بدست آوریم که برای تمام مقادیر ممکن پارامتر، تابع مخاطره را مینیمم کند. اما در عمل با توجه به بزرگی کلاس برآوردگرها چنین امکانی وجود ندارد، یعنی در آمار بهترین به معنای مطلق وجود ندارد. بنابراین در آمار همیشه به دنبال بدست آوردن برآوردگرهایی هستیم که دارای خاصیت بهینه باشند. برای بدست آوردن برآوردگر بهینه دو روش زیر بیشتر مورد توجه است :

۱- محدود کردن[۱] کلاس برآوردگرها به یک کلاس کوچکتر

الف- کلاس برآوردگرهای ناارریب ب- کلاس برآوردگرهای پایا

۲- برقراری یک نوع رابطه ترتیب[۲] بین برآوردگرها

الف- برآوردگرهای بیز ب- برآوردگرهای مینیماکس

بخش ۱: برآوردگرهای ناارریب[۳]

تعریف ۱-۱: برآوردگر $T(\underset{\sim}{X})$ را برای $\gamma(\theta)$ ناارریب گویند هرگاه $E_\theta[T(\underset{\sim}{X})]=\gamma(\theta),\ \forall\theta\in\Theta$ و اگر برآوردگر $T(\underset{\sim}{X})$ برای $\gamma(\theta)$ ناارریب نباشد مقدار $b(\theta)=E_\theta[T(\underset{\sim}{X})]-\gamma(\theta)$ را اریبی برآوردگر $T(\underset{\sim}{X})$ در برآورد $\gamma(\theta)$ گویند. ■

توجه کنید که در بعضی موارد ممکن است برآوردگر ناارریب موجود نباشد.

مثال ۱-۱: فرض کنید $X\sim B(n,\theta)$ نشان دهید که $\gamma(\theta)=\frac{۱}{\theta}$ دارای هیچ برآوردگر ناارریبی نیست.

حل: فرض کنید $T(\underset{\sim}{X})$ یک برآوردگر ناارریب $\frac{۱}{\theta}$ باشد، در اینصورت

$$E(T(\underset{\sim}{X}))=\frac{۱}{\theta}\Rightarrow\sum_{x=۰}^{n}T(x)\binom{n}{x}\theta^x(۱-\theta)^{n-x}=\frac{۱}{\theta}$$

$$\Rightarrow T(۰)(۱-\theta)+\sum_{x=۱}^{n}T(x)\binom{n}{x}\theta^x(۱-\theta)^{n-x}=\frac{۱}{\theta}$$

[۱] Restriction
[۲] Ordering
[۳] Unbiased Estimators

حال اگر $\theta \to 0^+$ آنگاه سمت چپ معادله فوق به سمت $T(0)$ و سمت راست بـه سـمت $+\infty$ میـل می‌کند و نشان می‌دهد که $T(\underset{\sim}{X})$ ای وجود ندارد که در رابطه فوق صدق‌کند. ■

تعریف ۱-۲: الف – در خانواده توزیـع‌هـای $P = \{p_\theta : \theta \in \Theta\}$ تـابع حقیقـی $\gamma(\theta)$ را بـرآورد پذیر[1] گویند اگر آماره $h(\underset{\sim}{X})$ وجود داشته باشد بطوری که $E_\theta[h(\underset{\sim}{X})] = \gamma(\theta), \quad \forall \theta \in \Theta$

ب- آماره $\delta(\underset{\sim}{X})$ را برآوردگر ناارایب با واریانس بطور یکنواخت مینـیمم[2] (UMVU) بـرای $\gamma(\theta)$ گویند اگر $E_\theta[h(\underset{\sim}{X})] = \gamma(\theta), \ \forall \theta \in \Theta$ و برای هر برآوردگر ناارایب $h(\underset{\sim}{X})$ پارامتر $\gamma(\theta)$ داشته باشیم که $V_\theta(\delta(\underset{\sim}{X})) \le V_\theta(h(\underset{\sim}{X})) \qquad \forall \theta \in \Theta$

ج- آماره $\delta(\underset{\sim}{X})$ را برآوردگر ناارایب با واریانس بطور موضعی مینـیمم[3] (LMVU) بـرای $\gamma(\theta)$ در $\theta = \theta_.$ گویند اگر این برآوردگر ناارایب بـوده و بـرای هـر برآوردگـر ناارایـب $h(\underset{\sim}{X})$ پـارامتر $\gamma(\theta)$ داشته باشیم که $V_{\theta_.}(\delta(\underset{\sim}{X})) \le V_{\theta_.}(h(\underset{\sim}{X}))$ ■

یک روش برای بدست آوردن برآوردگرهای UMVU یا LMVU یافتن کـلاس کلیـه برآوردگرهـای ناارایب $\gamma(\theta)$ یعنی Δ_γ است و سپس در این کلاس برآوردگرهای UMVU یـا LMVU را بدسـت می‌آوریم. برای این منظور از لم زیر استفاده می‌کنیم.

لم ۱-۱: اگر $\delta_.$ یک برآوردگر ناارایب $\gamma(\theta)$ باشد آنگاه کلیه برآوردگرهـای ناارایـب $\gamma(\theta)$ بـصورت $\delta = \delta_. - U$ مـی‌باشـد کـه در آن U هـر برآوردگـر ناارایـب صـفر اسـت (یعنـی $\forall \theta \in \Theta$ $E_\theta(U) = 0$) و بعلاوه $Var(\delta) = E[(\delta_. - U)^2] - [\gamma(\theta)]^2$

اثبات : واضح است. ■

قرار می‌دهیم

$$\Delta_. = \{U \,|\, E_\theta(U) = 0, \ \forall \theta \in \Theta\} \quad \text{و} \quad \Delta_\gamma = \{\delta \,|\, E_\theta(\delta) = \gamma(\theta), \ \forall \theta \in \Theta\}.$$

[1] Estimable
[2] Uniformly Mminimum Variance Unbiased
[3] Locally Mminimum Variance Unbiased

برای استفاده از لم فوق در حل یک مسئله، ابتدا کلاس کلیه برآوردگرهای ناریب $\gamma(\theta)$ را به دست می‌آوریم و واریانس آنها را در یک نقطه ثابت $\theta_.$ مینیمم می‌کنیم. اگر این مقدار مینیمم به مقدار $\theta_.$ بستگی نداشته باشد (و در نتیجه حل مسئله به مقدار مشخص $\theta_.$ بستگی نداشته باشد) آنگاه برآوردگر حاصل UMVU پارامتر مورد نظراست. در غیر این صورت اگر برآوردگر حاصل به $\theta_.$ بستگی داشته باشد (و در نتیجه کمترین واریانس را در بعضی از نقاط $\theta_.$ بدست آورد) آنگاه این برآوردگر LMVU نامیده می‌شود و در این حالت UMVU پارامتر مورد نظر وجود ندارد.

مثال ۱-۲: فرض کنید X یک متغیر تصادفی گسسته با تابع احتمال زیر باشد

$$p_\theta(x)=\begin{cases}\theta & x=1\\ \theta^{x-2}(1-\theta)^2 & x=2,3,...\end{cases} \qquad 0<\theta<1$$

بررسی کنید که آیا پارامترهای الف) $\gamma_1(\theta)=\theta$ ب) $\gamma_2(\theta)=(1-\theta)^2$ دارای برآوردگر UMVU هستند؟

حل: در یکی از مثال‌های فصل قبل نشان دادیم که

$$E_\theta[U(X)]=0 \Rightarrow U(x)=(x-2)(-h(1))=(x-2)a \qquad x=1,2,....$$

و نتیجه گرفتیم که X یک آماره کامل نیست اما کامل کراندار است. حال قرار می‌دهیم:

$$\delta_1(X)=I_{\{1\}}(X)\Rightarrow E_\theta[\delta_1(X)]=\theta$$

$$\delta_2(X)=I_{\{2\}}(X)\Rightarrow E_\theta[\delta_2(X)]=(1-\theta)^2$$

پس δ_1 و δ_2 به ترتیب برآوردگرهای ناریب برای $\gamma_1(\theta)$ و $\gamma_2(\theta)$ می‌باشند. در هریک از حالات فوق، فرم کلی برآوردگرهای ناریب به فرم زیر می‌باشد:

$$\delta_i^*(X)=\delta_i(X)-U(X) \qquad i=1,2,...$$

و δ_i^* موقعی کمترین واریانس خود را در $\theta_.$ بدست می‌آورد که

$$V_{\theta_.}(\delta_i^*)=E_{\theta_.}[(\delta_i-U)^2]-[\gamma_i(\theta_.)]^2$$

و یا معادلا $E_{\theta_.}[(\delta_i-U)^2]$ مینیمم شود. اما

$$E_{\theta_.}\left[(\delta_i(X)-U(X))^2\right]=\sum_{x=1}^{+\infty}\left(\delta_i(x)-U(x)\right)^2 p_{\theta_.}(x)$$

$$= \sum_{x=۱}^{+\infty} \left(\delta_i(x) - a(x-۲)\right)^۲ p_{\theta_۰}(x)$$

بنابراین مقداری از a را می‌خواهیم پیدا کنیم که مجموع بالا را مینیمم کند.

الف) $i=۱ \Rightarrow g_۱(a) = E_{\theta_۰}[(T_۱(x)-h(x))^۲] = (۱-a)^۲\theta_۰ + \sum_{x=۲}^{+\infty} a^۲(x-۲)^۲\theta_۰^{x-۲}(۱-\theta_۰)^۲$

$$= (۱+a)^۲\theta_۰ + \sum_{y=۰}^{+\infty} a^۲ y^۲ \theta_۰^y (۱-\theta_۰)^۲$$

$$g_۱'(a) = ۲(۱+a)\theta_۰ + ۲a(۱-\theta_۰)^۲ \sum_{y=۰}^{+\infty} y^۲\theta_۰^y = ۲a\left(\theta_۰ + (۱-\theta_۰)^۲ \sum_{y=۰}^{+\infty} y^۲\theta_۰^y\right) + ۲\theta_۰$$

$$g_۱''(a) = cons\,tant > ۰ \Rightarrow g_۱'(a) = ۰ \Rightarrow a_۱^* = \frac{-\theta_۰}{\theta_۰ + (۱-\theta_۰)^۲ \sum_{y=۰}^{+\infty} y^۲\theta_۰^y}$$

بنابراین $\delta_۱^*(X) = \delta_۱(X) - a_۱^*(X-۲)$ به دلیل آنکه $a_۱^*$ به مقدار $\theta_۰$ بستگی دارد یک برآوردگر LMVU برای θ است.

ب-

$$i=۲ \Rightarrow g_۲(\theta) = E_{\theta_۰}\left[\left(\delta_۲(X) - U(X)\right)^۲\right]$$

$$= a^۲\theta_۰ + (۱-\theta_۰)^۲ + \sum_{x=۳}^{+\infty} a^۲(x-۲)^۲\theta_۰^{x-۲}(۱-\theta_۰)^۲$$

$$= a^۲\left[\theta_۰ + \sum_{y=۱}^{+\infty} y^۲\theta_۰^y (۱-\theta_۰)^۲\right] + (۱-\theta_۰)^۲$$

$$g_۲'(a) = ۲a\left[\theta_۰ + (۱-\theta_۰)^۲ \sum_{y=۱}^{+\infty} y^۲\theta_۰^y\right], \qquad g_۲''(a) = cons\,tant > ۰$$

$$\Rightarrow g_۲'(a) = ۰ \Rightarrow a_۲^* = ۰$$

چون $a_۲^*$ بستگی به $\theta_۰$ ندارد بنابراین برآوردگر

$$\delta_۲^*(X) = \delta_۲(X) - a_۲^*(X-۲) = \delta_۲(X)$$

برآوردگری ناریب برای $(۱-\theta)^۲$ است که واریانس را نه تنها برای θ. بلکه برای کلیه $۰\langle\theta\langle ۱$ مینیمم می‌کند پس $T_۲^*(X)$ برآوردگر UMVU برای $(۱-\theta)^۲$ است. ■

با توجه به مثال بالا این سوال مطرح می‌شود که چه برآوردگرهایی برای پارامتر $\gamma(\theta)$ برآوردگرهای UMVU هستند و همچنین چه توابع برآوردپذیر $\gamma(\theta)$ دارای برآوردگرهای UMVU هستند. جواب به این سوال از طریق قضیه زیر داده می‌شود. در این قضیه $\Delta = \left\{\delta \middle| E(\delta^۲)\langle\infty\right\}$ است.

قضیه ۱-۱ (قضیه لهمن– شفه): فرض کنید X یک متغیر تصادفی با توزیعی از خانواده $P = \{P_\theta : \theta \in \Theta\}$ باشد و $\delta \in \Delta$ و Δ. کلاس کلیه برآوردگرهای ناریب صفر باشد. در این صورت شرط لازم و کافی برای آنکه δ یک برآوردگر UMVU برای امید خود $\gamma(\theta) = E_\theta[\delta(X)]$ باشد آن است که :

$$Cov\left(\delta(X), U(X)\right) = ۰ \qquad \forall U \in \Delta., \ \forall \theta \in \Theta$$

توجه: اگر به جای θ مقدار θ. را در قضیه قرار دهیم، آنگاه T یک برآوردگر LMVU خواهد بود.

همچنین چون $E_\theta(U) = ۰$ پس $Cov\left(\delta(X), U(X)\right) = ۰ \Leftrightarrow E[\delta U] = ۰$

اثبات: (شرط کافی): فرض کنید

$$Cov\left(\delta(X), U(X)\right) = ۰ \qquad \forall U \in \Delta., \ \forall \theta \in \Theta$$

و فرض کنید δ^* برآوردگر ناریب دیگری برای $\gamma(\theta)$ باشد. یعنی

$$E_\theta[\delta^*(X)] = E_\theta[\delta(X)] = \gamma(\theta) \Rightarrow E_\theta[\delta^*(X) - \delta(X)] = ۰ \quad \forall \theta \in \Theta$$

بنابراین طبق فرض مساله داریم که

$$Cov\left(\delta(X), \delta^*(X) - \delta(X)\right) = ۰ \quad \forall \theta \in \Theta \Rightarrow Cov\left(\delta(X), \delta^*(X)\right) - V(\delta(X)) = ۰$$

$$\Rightarrow Var_\theta(\delta(X)) = Cov_\theta\left(\delta(X), \delta^*(X)\right) \overset{(۱)}{\leq} \sqrt{Var_\theta(\delta(X)) Var_\theta(\delta^*(X))}$$

$$\Rightarrow \sqrt{V_\theta(\delta(X))} \leq \sqrt{V_\theta(\delta^*(X))} \Rightarrow V_\theta(\delta(X)) \leq V_\theta(\delta^*(X)) \quad \forall \theta \in \Theta$$

که در آن نامساوی (۱) از نامساوی شوارتز بدست می‌آید. پس $\delta(X)$ برآوردگر UMVU پارامتر $\gamma(\theta) = E_\theta(\delta)$ است.

(شـــرط لازم): فـــرض کنیـــد $\delta(X)$ برآوردگـــر UMVU بـــرای $\gamma(\theta)=E_\theta(\delta)$ باشـــد و $E_\theta[U(X)]=0\ \forall\theta\in\Theta$. می‌خواهیم نشان دهیم که $Cov_\theta(\delta(X),U(X))=0$

برای یک مقدار ثابت λ قرار می‌دهیم $\delta^*(X)=\delta(X)+\lambda U(X)$

چون $\delta(X)$ برآوردگر UMVU و $\delta^*(X)$ برآوردگر ناارایب $\gamma(\theta)$ می‌باشند پس

$$V_\theta(\delta(X))\le V_\theta(\delta^*(X))=V_\theta(\delta(X)+\lambda U(X))$$

$$=V_\theta(\delta(X))+\lambda^2V_\theta(U(X))+2\lambda Cov_\theta(\delta(X),U(X))$$

$$\Rightarrow \lambda^2V_\theta(U(X))+2\lambda Cov_\theta(\delta(X),U(X))\ge 0\quad \forall\theta\in\Theta,\ \forall\lambda \qquad (2)$$

معادله $\lambda^2V_\theta(\delta(X))+2\lambda Cov_\theta(\delta(X),U(X))=0$ یک معادله درجه دوم بر حسب λ است که دارای دو ریشه $\lambda_2=\dfrac{-2Cov_\theta(\delta(X),U(X))}{V_\theta(\delta(X))}$ و $\lambda_1=0$ است و در نتیجه بین این دو ریشه علامت معادله منفی خواهد بود. بنابراین اگر نامعادله (۲) بخواهد برای تمام مقادیر λ برقرار باشد بایستی $\lambda_1=\lambda_2=0$ باشد یعنی بایستی $Cov_\theta(\delta(X),U(X))=0$ باشد. ■

مثال ۱-۳: در مثال ۱-۲ کلاس کلیه برآوردگرهای UMVU و همچنین کلاس کلیه پارامترهای $\gamma(\theta)$ برآوردپـذیر دارای برآوردگـر UMVU را بدسـت آوریـد و نـشان دهیـد کـه پـارامتر θ دارای هـیچ برآوردگر UMVU نیست.

حل: کلاس کلیه برآوردگرهای ناارایب صفر عبارت است از

$$U(X)=a(X-2)\quad X=1,2,...,\quad a=-U(1)$$

$\delta(X)$ برآوردگر UMVU برای $\gamma(\theta)=E_\theta(\delta)\Leftrightarrow Cov_\theta(\delta(X),U(X))=0$

$$\Leftrightarrow E_\theta[\delta(X)a(X-2)]=0\Leftrightarrow E_\theta[\delta(X)(X-2)]=0$$

$$\Leftrightarrow \delta(X)(X-2)=\{-\delta(1)(1-2)\}(X-2)=\delta(1)(X-2)$$

$$\Rightarrow \delta(x)=\begin{cases}\delta(1) & x=1,3,4,....\\ b & x=2\end{cases}\Rightarrow \delta(x)=\begin{cases}a & x=1,3,4,....\\ b & x=2\end{cases}$$

که در آن b یک ثابت دلخواه اسـت. بنـابراین فـرم کلـی برآوردگرهـای UMVU بـصورت $\delta(X)$ است.

$$\gamma(\theta)=E_\theta[\delta(X)]=a\theta+b(1-\theta)^2+\sum_{x=3}^{+\infty}a\theta^{x-2}(1-\theta)^2=a\theta+b(1-\theta)^2+a\theta(1-\theta)$$
$$=a(2\theta-\theta^2)+b(1-\theta)^2=-a(1-\theta)^2+a+b(1-\theta)^2$$
$$=a+(b-a)(1-\theta)^2=a+c(1-\theta)^2$$

بنابراین تنها پارامترهای بفرم فوق دارای برآوردگر UMVU هستند، ولی θ دارای برآوردگر UMVU نیست زیرا

$$\theta=a+c(1-\theta)^2\Rightarrow c(1-2\theta+\theta^2)-\theta+a=0\Rightarrow c\theta^2-(2c+1)\theta+a=0$$
$$\overset{\forall\theta\in(0,1)}{\Rightarrow}\begin{cases}a=0\\ c=0\\ 2c+1=0\Rightarrow c=-\frac{1}{2}\end{cases}$$

که به تناقض بر می‌خوریم. ■

نکاتی در مورد قضیه لهمن – شفه:

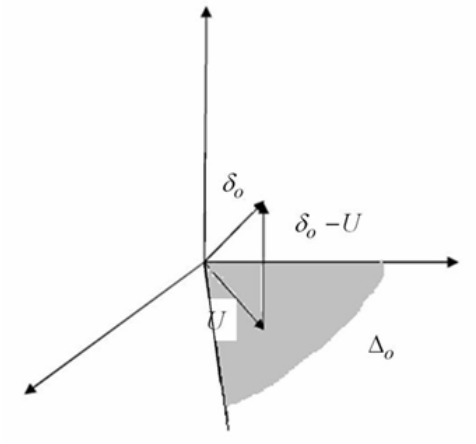
۱) با توجه به مثال قبل، بوسیله این قضیه می‌توان کلیه برآوردگرهای UMVU و همچنین کلیه پارامترهای برآوردپذیر دارای برآوردگر UMVU را بدست آورد و تنها مشکل آن بدست آوردن برآوردگرهای نا اریب صفر است.

۲) تعبیر هندسی: اگر $\Delta=\left\{\delta\middle|E(\delta^2)\langle\infty\right\}$ آنگاه Δ یک فضای برداری است زیرا

$$\forall a,b\in R,\ \delta_1,\delta_2\in\Delta\Rightarrow a\delta_1+b\delta_2\in\Delta$$

و Δ_o نیز یک زیر فضای Δ است. اگر قرار دهیم $\langle\delta_1,\delta_2\rangle=Cov(\delta_1,\delta_2)$ آنگاه یک ضرب داخلی روی Δ خواهیم داشت و با توجه به ضرب داخلی فوق داریم که δ_1,δ_2 فاصله $=d(\delta_1,\delta_2)$

$$=\sqrt{\langle\delta_1-\delta_2,\delta_1-\delta_2\rangle}=\sqrt{Var(\delta_1-\delta_2)}$$



اگر δ برآوردگر UMVU باشد آنگاه به ازای هر $\delta'\in\Delta$ داریم که $\sqrt{V(\delta)}\le\sqrt{V(\delta')}$. حال چون $\delta=\delta_o-U$

می‌خواهیم U را طوری انتخاب کنیم که $V(\delta)=V(\delta_o-U)=d(\delta_o,U)$ مینیمم شود. اما برای مینیمم کردن $d(\delta_o,U)$ بایستی U را تصویر δ_o در زیر فضای Δ_o بگیریم. بنابراین باید δ_o-U بر زیر فضای Δ_o عمود باشد. یعنی

$$Cov(\delta_o-U,U)=0 \Leftrightarrow Cov(\delta,U)=0 \qquad \blacksquare$$

۳) از قضیه لهمن – شفه برای اثبات عدم UMVU بودن یک برآوردگر می‌توان استفاده کرد.

مثال ۱-٤: فرض کنید $X\sim U(\theta,\theta+1)$ می دانیم که $T(X)=X$ یک آماره بسنده مینیمال برای θ است. نشان دهید که $T(X)$ برای θ کامل نیست و $S(X)=X-\frac{1}{2}$ یک برآ وردگر نا اریب θ است که UMVUE آن نیست.

حل:

$$E[h(X)]=0 \Rightarrow \int_{\theta}^{\theta+1} h(x)dx=0 \Rightarrow \frac{d}{d\theta}\int_{\theta}^{\theta+1} h(x)dx=0$$

$$\Rightarrow h(\theta+1)-h(\theta)=0 \Rightarrow h(\theta+1)=h(\theta) \quad \forall \theta\in\Re$$

بنابراین توابع متناوب با دوره تناوب ۱ دارای امید صفر هستند. برای مثال اگر $h(x)=\sin(2\pi x)$ آنگاه $E[\sin(2\pi x)]=0$ اما $P(\sin(2\pi x)=0)=0\neq 1$.

$$E(S(X))=E(X-\frac{1}{2})=\int_{\theta}^{\theta+1}(x-\frac{1}{2})dx=\frac{1}{2}x^2-\frac{1}{2}x\Big|_{\theta}^{\theta+1}=\frac{1}{2}(2\theta)=\theta$$

$$Cov_\theta(S(X),h(X))=E_\theta(S(X)h(X))=E_\theta(X\sin(2\pi X))$$

$$=\int_{\theta}^{\theta+1} x\sin(2\pi x)dx=[-\frac{1}{2\pi}x\cos(2\pi x)+\frac{1}{4x^2}\sin(2\pi x)]_{\theta}^{\theta+1}$$

$$=-\frac{1}{2\pi}(\theta+1)\cos(2\pi\theta)+\frac{1}{4\pi^2}\sin(2\pi\theta)+\frac{1}{2\pi}\theta\cos(2\pi\theta)-\frac{1}{4\pi^2}\sin(2\pi\theta)$$

$$=-\frac{1}{2\pi}\cos(2\pi\theta)\neq 0$$

پس $S(X)=X-\frac{1}{2}$ برآوردگر UMVU پارامتر θ نیست. $\blacksquare$

٤) توسط قضیه لهمن – شفه می‌توان یکتایی برآوردگرهای UMVU را به صورت زیر ثابت کرد.

لـــم ۱-۲: بـــرای هـــر دو متغیـــر تـــصادفی X و Y بـــا گـــشتاورهای دوم متنـــاهی داریـــم کـــه $[Cov(X,Y)]^2 \le V(X)V(Y)$ و تساوی برقرار است اگر و فقـط اگـر ثابـت‌هـای a و b موجـود باشند بطوری که $P(X = a+bY) = 1$.

قضیه ۱-۲: اگر $\delta(X)$ یک برآوردگر UMVU برای پـارامتر $\gamma(\theta)$ باشـد آنگـاه $\delta(X)$ برآوردگـر یکتای[1] UMVU برای $\gamma(\theta)$ است.

اثبات: فرض کنید δ و δ' دو برآوردگر UMVU پارامتر $\gamma(\theta)$ باشند. پس

$$E_\theta(\delta) = E_\theta(\delta') = \gamma(\theta), \qquad V_\theta(\delta) = V_\theta(\delta') \quad \forall \theta \in \Theta \qquad (۳)$$

در نتیجه $E_\theta[\delta - \delta'] = 0$ پس $\delta - \delta'$ برآوردگر نا اریب صفر است. پس طبق قضیه لهمن – شفه

$$o = \operatorname{cov}(\delta, \delta - \delta') = V(\delta) - Cov(\delta - \delta') \Rightarrow Cov(\delta - \delta') = V(\delta)$$

$$\Rightarrow [Cov(\delta, \delta')]^2 = [V(\delta)]^2 \overset{(۳)}{=} V_\theta(\delta) V_\theta(\delta')$$

یعنی در نامساوی کواریانس حالت تساوی برقرار است پس طبق لم ۱-۳ داریم که $P_\theta(\delta = a\delta' + b) = 1$ و در نتیجه طبق (۳) داریم که $P_\theta(\delta = \delta') = 1$ یعنی برآوردگر UMVU یکتا است. ■

توجه: برآوردگر $\delta = c$ برآوردگری UMVU برای امید خود است، زیرا دارای واریانس صـفر اسـت. حال اگر برآوردگرهای ثابت را کنار بگذاریم، برای یک تابع برآوردپذیر $\gamma(\theta)$ سـه حالـت زیـر پـیش می‌آید:

حالت اول: هیچ تابع غیر ثابتی از θ برآوردگر UMVU ندارد.

مثال ۱-۵: فرض کنید متغیر تصادفی X دارای تابع احتمال زیر باشد، نشان دهید کـه هـیچ تـابع غیـر ثابتی از θ برآوردگر UMVU ندارد

x	$\theta-1$	θ	$\theta+1$
$p_\theta(x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

ابتدا کلاس برآوردگرهای ناریب صفر یعنی Δ. را بدست می‌آوریم.

[1] Unique

$$E_\theta[U(X)]=۰\Rightarrow \frac{۱}{۳}[U(\theta-۱)+U(\theta)+U(\theta+۱)]=۰$$

$$\Rightarrow U(\theta-۱)+U(\theta)+U(\theta+۱)=۰$$

$$\left.\begin{array}{l}\theta=۰\Rightarrow U(-۱)+U(۰)+U(۱)=۰\\ \theta=۱\Rightarrow U(۰)+U(۱)+U(۲)=۰\\ \theta=۲\Rightarrow U(۱)+U(۲)+U(۳)=۰\end{array}\right\}\Rightarrow\begin{cases}U(k)=U(k+۳)\\ U(k)+U(k+۱)+U(k+۲)=۰\end{cases}\quad k\in\mathbb{Z}$$

$$\Rightarrow \Delta_{.}=\left\{U\,\middle|\,U(۳k)=a,U(۳k+۱)=b,U(۳k+۲)=-(a+b),\ k\in\mathbb{Z},\ a,b\in\Re\right\}$$

$$\delta \text{ برآوردگر UMVU} \Leftrightarrow \delta U\in\Delta_{.}\Rightarrow\begin{cases}\delta(۳k)U(۳k)=a\\ \delta(۳k+۱)U(۳k+۱)=b\\ \delta(۳k+۲)U(۳k+۲)=-(a+b)\end{cases}$$

$$\begin{array}{l}a\neq ۰,b\neq ۰,\\ a+b\neq .\\ \Rightarrow\end{array}\begin{cases}\delta(۳k)=۱\\ \delta(۳k+۱)=۱\Rightarrow\ \delta(k)=۱\ \forall k\in\mathbb{Z}\\ \delta(۳k+۲)=۱\end{cases}$$

پس δ یک برآوردگر ثابت است که نمی‌توا ند برای $\gamma(\theta)$ نااریب باشد پس UMVU نیز نیست. ■

حالت دوم: تنها بعضی از توابع برآوردپذیر غیر ثابت دارای برآوردگر UMVU هستند. (مثال ۳۱-)

حالت سوم: هر پارامتر برآوردپذیر $\gamma(\theta)$ دارای برآوردگر UMVU است.

یک روش برای به دست آوردن UMVU در این حالت توسط قضیه رائوبلاکول است. طبق قضیه رائوبلاکول اگر T یک آماره بسنده برای θباشد و $\delta(\underset{\sim}{X})$ یک برآوردگر نااریب برای $\gamma(\theta)$ باشد آنگاه $g(T)=E[\delta(\underset{\sim}{X})|T]$ یک آماره است و $E_\theta[g(T)]=E_\theta[\delta(\underset{\sim}{X})]=\gamma(\theta)$ و $V_\theta[g(T)]<V_\theta[\delta(\underset{\sim}{X})]$ و تساوی برقرار است اگر و فقط اگر $P\left(g(T)=\delta(\underset{\sim}{X})\right)=۱$.

بنابراین هر برآوردگر نااریب $\gamma(\theta)$ را می‌توان با امید شرطی نسبت به یک آماره بسنده بهبود بخشید.

حال این سوال مطرح است که آیا به ازای هر برآوردگر نااریب اولیه $\delta(\underset{\sim}{X})$، برآوردگر حاصل از امید شرطی یکتا است؟

قضیه ۱-۳: فرض کنید X دارای توزیعی از خا نواده $P = \{P_\theta : \theta \in \Theta\}$ باشد و T یک آماره بسنده کامل برای P باشد. در این صورت هر تابع برآوردپذیر $\gamma(\theta)$ دارای یک و تنها یک برآوردگر نا اریب است که تابعی از T می باشد.

اثبات: وجود این برآوردگر از قضیه رائوبلاکول بدست می آید و اگر $g_1(T)$ و $g_2(T)$ دو برآوردگر نا اریب $\gamma(\theta)$ باشند آنگاه

با احتمال یک $g_1(T) = g_2(T) \quad \Rightarrow^{completeness} \quad E[g_1(T) - g_2(T)] = 0$ ■

از قضیه رائوبلاکول و قضیه قبل نتیجه زیر را می‌توان بدست آورد.

قضیه ۱-٤: فرض کنید X دارای توزیعی از خانواده $P = \{P_\theta : \theta \in \Theta\}$ باشد و T یک آماره بسنده و کامل برای خانواده P باشد.

الف) برای هر تابع برآوردپذیر $\gamma(\theta)$ یک برآوردگر نااریب وجود دارد که به طور یکنواخت مخاطره را برای هر تابع زیان محدب $L(\theta,d)$ مینیمم می‌کند. این برآوردگر در حالت خاص UMVU است.

ب) برآوردگر UMVU حالت قبل همان برآوردگر نااریب $\gamma(\theta)$است که تابعی از T می‌باشد.این برآوردگر نااریب با مخاطره مینیمم است به شرط آنکه مخاطره آن متناهی و تابع زیان اکیدا محدب باشد. ■

توجه: اگر تابع زیان به غیر از تابع زیان درجه دوم باشد، برآوردگر حاصل را UMRU گویند.

قضایای قبل نه تنها وجود برآوردگر UMVU را تضمین می‌کند بلکه روش‌هایی را برای بدست آوردن برآوردگرهای UMVU ارایه می‌دهد. با داشتن آماره بسنده کامل T در برآورد پارامتر $\gamma(\theta)$ این روش‌ها عبارتند از:

الف) ابتدا یک براوردگر نااریب $\delta(\underset{\sim}{X})$ برای $\gamma(\theta)$ تعیین می‌کنیم در این صورت $g(T) = E[\delta(\underset{\sim}{X})|T]$ برآوردگر UMVU پارامتر $\gamma(\theta)$ است.

مثال ۱-٦ : فرض کنید $X_1, X_2, ..., X_n$ یک نمونه تصادفی از $U(0,\theta)$ باشند و $\gamma(\theta)=\frac{\theta}{2}$. در این صورت $T = X_{(n)}$ یک آماره بسنده و کامل برای θاست و $E[X_1]=\frac{\theta}{2}$، پس $g(T)=E[X_1|T]$ برآوردگر UMVU پارامتر $\gamma(\theta)$ است. اما اگر $X_{(n)}=T=t$ باشد آنگاه $X_1=t$ یا $X_1<t$ خواهد بود. پس اگر $X_{(n)}=t$ باشد آنگاه با احتمال $\frac{1}{n}$، $X_1=t$ و با احتمال $\frac{n-1}{n}$، X_1 دارای توزیع یکنواخت در فاصله $(0,t)$ می‌باشد. بنابراین

$$g(t)=E(X_1|X_{(n)}=t)=t(\frac{1}{n})+\frac{t}{2}(\frac{n-1}{n})=\frac{n+1}{n}.\frac{t}{2}$$

پس $g(T)=\frac{n+1}{n}.\frac{T}{2}$ برآوردگر UMVU پارامتر $\gamma(\theta)=\frac{\theta}{2}$ است. ■

ب) چون برآوردگر UMVU یکتا است، پس برآوردگر UMVU تابع برآوردپذیر $\gamma(\theta)$ با حل معادله زیر بدست می‌آید. $E_\theta[g(T)]=\gamma(\theta) \quad \forall\theta\in\Theta$

مثال ۱-۷: در مثال قبل داریم که $\frac{X(n)}{\theta}\sim Be(n,1)\Rightarrow E(\frac{X_{(n)}}{\theta})=\frac{n}{n+1}$

$$\Rightarrow E(\frac{n+1}{n}.\frac{X(n)}{2})=\frac{\theta}{2}\Rightarrow \frac{n+1}{n}.\frac{X(n)}{2} \text{ is UMVUE of } \frac{\theta}{2}$$ ■

مثال ۱-۸: فرض کنید $X_1, X_2, ..., X_n$ یک نمونه تصادفی از $B(1,\theta)$ باشند. برآوردگر UMVU پارامتر $\gamma(\theta)=\theta(1-\theta)$ را بدست آورید.

حل: می‌دانیم که $T=\sum_{i=1}^{n}X_i\sim B(n,\theta)$ آماره بسنده و کامل برای θ است.

$$E[g(T)]=\theta(1-\theta)\Rightarrow\sum_{t=0}^{n}g(t)\binom{n}{t}\theta^t(1-\theta)^{n-t}=\theta(1-\theta)$$

$$\Rightarrow\sum_{t=0}^{n}g(t)\binom{n}{t}(\frac{\theta}{1-\theta})^t=\theta(1-\theta)^{1-n}$$

قرار می دهیم $\beta=\frac{\theta}{1-\theta}>0$ در این صورت $\theta=\frac{\beta}{1+\beta}$ و $1-\theta=\frac{1}{1+\beta}$ بنابراین

$$\sum_{t=0}^{n} g(t)\binom{n}{t}\beta^t = \frac{\beta}{1+\beta}\left(\frac{1}{1+\beta}\right)^{1-n} = \beta(1+\beta)^{n-2} = \sum_{u=0}^{n-2}\binom{n-2}{u}\beta^{u+1}$$

$$\overset{t=u+1}{=} \sum_{t=1}^{n-1}\binom{n-2}{t-1}\beta^t \quad \forall \beta > 0$$

$$\Rightarrow g(t) = \begin{cases} 0 & t = 0, n \\ \dfrac{\binom{n-2}{t-1}}{\binom{n}{t}} & t \neq 0, n \end{cases} = \begin{cases} 0 & t = 0, n \\ \dfrac{t(n-t)}{n(n-1)} & t \neq 0, n \end{cases} \Rightarrow g(t) = \frac{t(n-t)}{n(n-1)} \quad t = 0, \ldots, n$$

پس $g(T) = \dfrac{T(n-T)}{n(n-1)}$ برآوردگر UMVU پارامتر $\theta(1-\theta)$ است. ■

بخش‌های ۲-۲، ۳-۲ و ٤-۲ به عنوان سمینار برای دانشجویان

بخش ۵: نامساوی اطلاع[1] (نامساوی کرامر - رائو)

فرض کنید δ یک برآوردگر $g(\theta)$ باشد که $E(\delta) = g(\theta)$ و $\psi(x, \theta)$ تابعی از x و θ با گشتاور دوم متناهی باشد، در این صورت طبق نامساوی کواریانس داریم که

$$Var(\delta) \geq \frac{[Cov(\delta, \psi)]^2}{Var(\psi)} \qquad (1)$$

برای یافتن یک کران پایین برای $V(\delta)$ می‌خواهیم سمت راست نامساوی فوق به δ بستگی نداشته باشد و حداقل به $E(\delta) = g(\theta)$ بستگی داشته باشد.

قضیه ۵-۱ (بلیت[2] ۱۹۷٤): شرط لازم و کافی برای آن که $Cov(\delta, \psi)$ بستگی به δ تنها از طریق $g(\theta)$ داشته باشد آن است که $Cov(U, \psi) = 0 \quad \forall U \in \Delta_.$

که در آن $\Delta_. = \{U : E_\theta(U) = 0, E_\theta(U^2) < \infty \quad \forall \theta \in \Theta\}$

---

[1] Information Inequality
[2] Blyth

اثبات: (**کفایت**) فرض کنید δ_1 و δ_2 هر دو برای $g(\theta)$ نا اریب باشند در این صورت $\delta_1 - \delta_2 \in \Delta$.

پس $Co\,v(\delta_1 - \delta_2, \psi) = 0$ و در نتیجه $Cov(\delta_1, \psi) = Cov(\delta_2, \psi)$. یعنی از $E(\delta_1) = E(\delta_2)$ نتیجه گرفتیم که $Cov(\delta_1, \psi) = Cov(\delta_2, \psi)$ پس $Cov(\delta, \psi)$ تنها از طریق $E(\delta)$ به δ وابسته است.

(لـــزوم): فرض کنید $Cov(\delta, \psi)$ تنها از طریق $g(\theta) = E(\delta)$ به δ وابسته باشد. چون $E(\delta + U) = g(\theta)$ پس

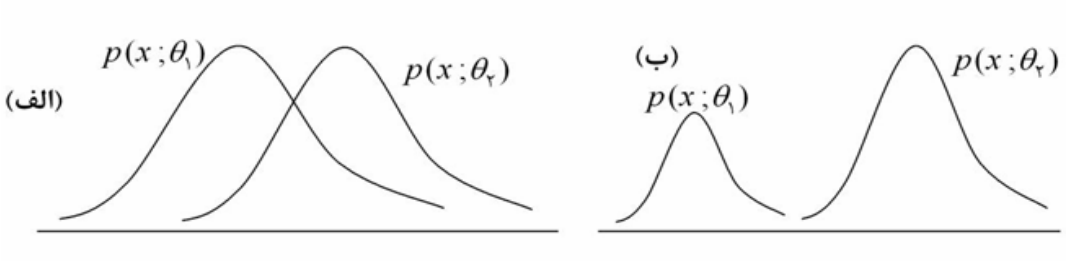
$$Cov(\delta, \psi) = Cov(\delta + U, \psi) \qquad \forall \theta \in \Theta \qquad \forall U \in \Delta.$$

$$\Rightarrow Cov(\delta, \psi) = Cov(\delta, \psi) + Cov(U, \psi)$$

$$\Rightarrow Cov(U, \psi) = 0 \qquad \forall \theta \in \Theta \qquad \forall U \in \Delta.$$

اطلاع فیشر:

نمودارهای زیر را در نظر بگیرید.



سؤال: مشاهده X در کدام حالت اطلاع بیشتری در مورد پارامتر θ به ما می‌دهد؟

جواب: بدن شک در حالت (ب) مشاهده X به ما اطلاع بیشتری می‌دهد.

پس هر چه تابع چگالی X از یک مقدار θبه مقدار دیگر θبیشتر تغییر کند، مشاهده X به ما در مورد مقدار پارامتر θ بیشتر اطلاع می‌دهد. حال توجه کنید که

$$\text{مقدار تغییر نسبی در چگالی در نقطه } X \text{ به ازای تغییر پارامتر به اندازه } \Delta = \left[\frac{P(x;\theta+\Delta) - p(x;\theta)}{\Delta} \cdot \frac{1}{p(x;\theta)}\right]$$

$$\text{میزان تغیر نسبی}^{1} \text{ در چگالی در نقطه } X = \lim_{\Delta \to 0}\left[\frac{P(x;\theta+\Delta)-p(x;\theta)}{\Delta}.\frac{1}{p(x;\theta)}\right]$$

$$= \frac{\frac{\partial}{\partial\theta}p(x;\theta)}{p(x;\theta)} = \frac{\partial}{\partial\theta}\ln p(x;\theta)$$

میانگین توان دوم این تغییر نسبی را میزان اطلاعی که X در مورد پارامتر θدارد می‌نامند و به آن اطلاع فیشر گویند یعنی

$$I_X(\theta) = E_\theta\left[\frac{\partial}{\partial\theta}\ln p(X;\theta)\right]^2 = \int\left(\frac{p'_\theta}{p_\theta}\right)^2 p_\theta d\mu$$

توجه کنید که هر چه این امید ریاضی در یک نقطه θ_0 بیشتر شود آنگاه می‌توان θ_0 را از نقاط همسایه θ بهتر تشخیص داد و بنابراین θ را در نقطه $\theta = \theta_0$ دقیق‌تر برآورد کرد.

حال فرض کنید شرایط زیر برقرار باشند:

الف) Θیک زیر فاصله باز از اعداد حقیقی باشد $\left(\Theta \subseteq R\right)$

ب) توزیع‌های P_θ دارای تکیه‌گاه یکسان باشند و بنابراین $S_X^* = \{x \mid p(x;\theta) > 0\}$ به θ بستگی نداشته باشد.

ج) $\frac{\partial}{\partial\theta}p(x;\theta)$ برای تمام $\theta \in \Theta$ موجود و متناهی باشد.

د) $\frac{\partial}{\partial\theta}\int p(x;\theta)d\mu = \int\frac{\partial}{\partial\theta}p(x;\theta)d\mu$

هـ) $E\left[\frac{\partial}{\partial\theta}\ln p(x;\theta)\right]^2 > 0 \quad \forall\theta\in\Theta$

و) برای هر برآوردگر δ داشته باشیم که: $\frac{\partial}{\partial\theta}E_\theta[\delta] = \frac{\partial}{\partial\theta}\int\delta p_\theta d\mu = \int\delta\frac{\partial}{\partial\theta}p_\theta d\mu$

این شرایط را شرایط نظم[2] گویند.

حال اگر قرار دهیم $\psi(x;\theta) = \frac{\partial}{\partial\theta}\ln p(x;\theta)$ آنگاه برای برآوردگر U که $E_\theta(U) = 0$ داریم که

[1] Relative Rate
[2] Regularity Condition

$$0 = \frac{\partial}{\partial\theta}E_\theta(U) \overset{(ب)}{=} \int U\,\frac{\partial}{\partial\theta}p_\theta d\mu = \int [U\,\frac{\partial}{\partial\theta}\ln p_\theta]p_\theta d\mu = E[U.\frac{\partial}{\partial\theta}\ln p_\theta]$$
$$= Cov(U, \frac{\partial}{\partial\theta}\ln p_\theta)$$

بنابراین شرایط قضیه ۵-۱ برقرار است. همچنین چون

$$E\left[\frac{\partial}{\partial\theta}\ln p_\theta\right] = \int\left(\frac{\partial}{\partial\theta}\ln p_\theta\right)p_\theta d\mu = \int\frac{\partial}{\partial\theta}p_\theta d\mu \overset{(ب)}{=} \frac{\partial}{\partial\theta}\int p_\theta d\mu = \frac{\partial}{\partial\theta}(1) = 0$$

پس $$Cov(\delta,\psi) = E(\delta\psi) = \int\delta\frac{\partial}{\partial\theta}p_\theta d\mu \overset{(ب)}{=} \frac{\partial}{\partial\theta}\int\delta p_\theta d\mu = \frac{\partial}{\partial\theta}E_\theta[\delta] = g'(\theta)$$

بنابراین نامساوی (۱) به صورت زیر تبدیل می‌شود.

$$Var(\delta) \geq \frac{[g'(\theta)]^2}{E\left[\frac{\partial}{\partial\theta}\ln p(x;\theta)\right]^2} = \frac{[g'(\theta)]^2}{I_X(\theta)}$$

این نامساوی را نامساوی اطلاع یا نامساوی کرامر – رائو گویند. به کران پایین بدست آمده برای واریانس، کران پایین کرامر– رائو (CRLB) گویند. توجه کنید که اگر مشتق دوم $\ln p(x;\theta)$ موجود باشد و $\frac{\partial^2}{\partial\theta^2}\int p_\theta d\mu = \int\frac{\partial^2}{\partial\theta^2}p_\theta d\mu$ آنگاه

$$I_X(\theta) = E\left[\frac{\partial}{\partial\theta}\ln p(x;\theta)\right]^2 = -E[\frac{\partial^2}{\partial\theta^2}\ln p(x;\theta)]$$

حال اگر $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از توزیعی با چگالی $p_\theta : \theta \in \Theta$ باشند آنگاه $I(\theta) = E_\theta\left[\frac{\partial}{\partial\theta}\ln p(\underset{\sim}{X},\theta)\right]^2$ را میزان اطلاع فیشر که نمونه تصادفی $X_1, X_2, \ldots, X_n$ در مورد θ دارد، گویند.

$$I(\theta) = E_\theta\left[\frac{\partial}{\partial\theta}\ln p(\underset{\sim}{X};\theta)\right]^2 = V_\theta[\frac{\partial}{\partial\theta}\ln p(\underset{\sim}{X};\theta)] = V_\theta[\frac{\partial}{\partial\theta}\ln\prod_{i=1}^{n}\ln p(X_i;\theta)]$$

$$\overset{(2)}{=} \sum_{i=1}^{n}V_\theta[\frac{\partial}{\partial\theta}\ln p(X_i;\theta)] \overset{(3)}{=} nV_\theta[\frac{\partial}{\partial\theta}\ln p(X_1;\theta)] = nE_\theta[\frac{\partial}{\partial\theta}\ln p(X_1;\theta)]^2 = nI_{X_1}(\theta)$$

که در آن (۲) از استقلال و (۳) از همتوزیع بودن X_iها نتیجه می‌گردد.

مثال ۵-۱: فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از توزیع $N(\theta,1), \theta \in R$ باشند، CRLB را برای واریانس برآوردگر ناارایب $g(\theta)=\theta^2$ بدست آورده و آن را با واریانس برآوردگر UMVU پارامتر θ^2 مقایسه کنید.

$$g(\theta)=\theta^2 \Rightarrow g'(\theta)=2\theta$$

$$\ln f_\theta(x) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2 \Rightarrow \frac{\partial^2}{\partial\theta^2}\ln f_\theta(x) = -1$$

$$\Rightarrow I(\theta) = nI_{X_1}(\theta) = n \Rightarrow CRLB = \frac{(2\theta)^2}{n} = \frac{4\theta^2}{n}$$

در این خانواده $S=\sum_{i=1}^{n}X_i \sim N(n\theta,n)$ آماره بسنده و کامل برای θ است و $T=\frac{S^2-n}{n^2}$ برآوردگر UMVU برای θ^2 است و $V(T)\overset{?}{=}\frac{4\theta^2}{n}+\frac{2}{n^2}>\frac{4\theta^2}{n}=CRLB$ ■

مثال ۵-۲: فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از توزیع نمایی با پارامتر θ باشند. CRLB را برای واریانس برآوردگرهای ناارایب θ و $\frac{1}{\theta}$ بدست آورده و آن را با واریانس برآوردگرهای UMVU مقایسه کنید.

$$\ln f_\theta(x) = -\ln\theta - \frac{x}{\theta} \Rightarrow \frac{\partial^2}{\partial\theta^2}\ln f_\theta(x) = \frac{1}{\theta^2} - \frac{2x}{\theta^3}$$

$$I_{X_1}(\theta) = -E_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f_\theta(X_1)\right] = -\left(\frac{1}{\theta^2}-\frac{2}{\theta^2}\right) = \frac{1}{\theta^2} \Rightarrow I(\theta)=\frac{n}{\theta^2}$$

الف) $g_1(\theta)=\theta \Rightarrow g_1'(\theta)=1 \Rightarrow CRLB = \frac{1}{n/\theta^2} = \frac{\theta^2}{n}$

$$T=\bar{X} \ \ is \ \ UMVUE \ \ of \ \ \theta \Rightarrow V(T)=\frac{\theta^2}{n}=CRLB$$

ب) $g_2(\theta)=\frac{1}{\theta} \Rightarrow g_2'(\theta)=\frac{-1}{\theta^2} \Rightarrow CRLB = \frac{\left(-1/\theta^2\right)^2}{n/\theta^2} = \frac{1}{n\theta^2}$

$$T=\frac{n-1}{\sum X_i} \ \ is \ \ UMVUE \ \ of \ \ \frac{1}{\theta} \Rightarrow V(T)=\frac{1}{(n-2)\theta^2} > \frac{1}{n\theta^2}=CRLB$$

یک سوال که در مورد نامساوی کرامر – رائو مطرح است این است که چه موقع نامساوی بـه مـساوی تبدیل می‌شود و یا به عبارتی برای چـه برآوردگرهـای نـاارایبی واریـانس برآوردگـر بـا CRLB برابـر می‌شود.

قضیه ۵-۲: فرض کنید شرایط نظم برقرار باشد و T برآوردگری باشد کـه $V_\theta(T)<\infty, \quad \forall\theta\in\Theta$.

در این صورت $V_\theta(T)=CRLB$ اگر و فقط اگر تابع چگالی $p_\theta(x)$ متعلق به خـانواده نمـایی یـک پارامتری زیر باشد.

$$p_\theta(x)=e^{c(\theta)T(x)-B(\theta)}h(x)$$

در ایـن حالـت $g(\theta)=E_\theta[T(X)]=\dfrac{B'(\theta)}{c'(\theta)}$ و $I(\theta)=c'(\theta)g'(\theta)$ مـی‌باشـد. (یعنـی تنهـا در خانواده‌های نمایی بفرم فوق و تنها برای توابع پارامتری بفرم فوق واریانس برآوردگر ناارایب $g(\theta)$ بـا CRLB برابر می‌شود.)

اثبات:

$$\rho^{۲}\left(\frac{\partial}{\partial\theta}\ln f_\theta(X),T(X)\right)=۱ \Leftrightarrow \text{تساوی در نامساوی کرامر – رائو برقرار است}$$

$$\Leftrightarrow \frac{\partial}{\partial\theta}\ln f_\theta(X)=a(\theta)T(X)+b(\theta)$$

$$\Leftrightarrow \ln f_\theta(X)=T(X)\int a(\theta)d\theta+\int b(\theta)d\theta$$

$$=T(X)c(\theta)-B(\theta)+R(X)$$

$$\Leftrightarrow f_\theta(X)=e^{c(\theta)T(X)-B(\theta)+R(X)}$$

$$\Leftrightarrow f_\theta(x)=e^{c(\theta)T(X)-B(\theta)}h(X)$$

در این خانواده $g(\theta)=E_\theta\big(T(X)\big)=\dfrac{B'(\theta)}{c'(\theta)}$ و

$$I(\theta)=-E[c''(\theta)T(X)-B''(\theta)]=-\frac{c''(\theta)B'(\theta)-B''(\theta)c'(\theta)}{c'(\theta)}$$

$$= \left[\frac{B'(\theta)}{C'(\theta)}\right]' c'(\theta) = g'(\theta)c'(\theta) \qquad \blacksquare$$

مثال ۵-۳: نشان دهید که تنها توزیع گسسته‌ای که دارای میانگین و واریانس یکسان است و $V_\theta(X)$ در آن با CRLB برابر می‌شود توزیع پواسون است.

$$g(\theta) = E_\theta(X) = V_\theta(X) = \theta$$

$$V_\theta(X) = CRLB \overset{(*)}{\Leftrightarrow} \theta = \frac{(\theta')^2}{I_X(\theta)} = \frac{1}{I_X(\theta)} = \frac{1}{g'(\theta)c'(\theta)} = \frac{1}{c'(\theta)}$$

$$\Rightarrow c'(\theta) = \frac{1}{\theta} \Rightarrow c(\theta) = \ln\theta \overset{(*)}{\Rightarrow} p_\theta(x) = e^{(\ln\theta)x - B(\theta)}h(x)$$

که در آن (*) از قضیه ۵-۲ بدست می‌آید. اما $\theta = g(\theta) = E_\theta(X) = \frac{+B'(\theta)}{1/\theta}$ پس $B'(\theta) = 1$ یا $B(\theta) = \theta$ و در نتیجه $h(x) = x!$ پس $p_\theta(x) = \frac{\theta^x e^{-\theta}}{x!}$ $\blacksquare$

قضیه ۵-۳: اگر $\theta = h(\zeta)$ و h مشتق پذیر باشد آن گاه $I^*(\zeta) = I[h(\zeta)][h'(\zeta)]^2$ اما CRLB برای واریانس برآوردگرهای θ و ζ یکی است. (یعنی CRLB با تبدیلات روی پارامترها تغییر نمی‌کند.)

اثبات:

$$I(\zeta) = E\left[\frac{\partial}{\partial\zeta}\ln f_\zeta(X)\right]^2 = E\left[\frac{\partial}{\partial\zeta}\ln f_\theta(X).\frac{\partial\theta}{\partial\zeta}\right]^2$$

$$= E\left[\frac{\partial}{\partial\theta}\ln f_\theta(X)\right]^2 [\frac{\partial\theta}{\partial\zeta}]^2 = I(\theta)\left[h'(\zeta)\right]^2$$

$$\frac{\partial}{\partial\zeta}E[T] = \frac{\partial}{\partial\theta}E[T]\frac{\partial\theta}{\partial\zeta} = g'(\theta)h'(\zeta)$$

$$\Rightarrow CRLB_\zeta = \frac{[g'(\zeta)]^2}{I(\zeta)} = \frac{[g'(\theta)h'(\zeta)]^2}{I(\theta)[h'(\zeta)]^2} = \frac{[g'(\theta)]^2}{I(\theta)} = CRLB_\theta \qquad \blacksquare$$

مثال ۵-٤: فرض کنید X متعلق به یک خانواده نمایی با تابع چگالی $p_\theta(x) = e^{C(\theta)T(x) - B(\theta)}h(x)$ باشد و فرض کنید $g(\theta) = E_\theta[T(X)]$ در این صورت نشان دهید که $I[g(\theta)] = \frac{1}{Var(T)}$.

حل: $I(\theta)=V_\theta\left[\frac{\partial}{\partial\theta}\ln p_\theta(X)\right]V_\theta\left[c'(\theta)T(X)-B'(\theta)\right]=\left[c'(\theta)\right]^2V_\theta(T)$

$$I(g(\theta))=\frac{I(\theta)}{\left[g'(\theta)\right]^2}=\left[\frac{c'(\theta)}{g'(\theta)}\right]^2V_\theta(T)=\frac{1}{V_\theta(T)}$$

$$\left(Since\ \ V_\theta(T)=\frac{B''(\theta)-c''(\theta)g(\theta)}{[c'(\theta)]^2}=\frac{g'(\theta)}{c'(\theta)}\right)$$

بخش ٦: نامساوی اطلاع (کرامر- رائو) در حالت چند پارامتری

در این قسمت فرض می‌کنیم $\underset{\sim}{\theta}=(\theta_1,\ldots,\theta_k)'$ باشد و شرایط نظم را به صورت زیر در نظر می‌گیریم:

الف- Θ یک گوی باز در فضای R^k باشد، $\Theta\subseteq R^k$.

ب- توزیع‌های $P_{\underset{\sim}{\theta}}$ دارای تکیه گاه یکسان باشند و بنابراین $S_X^*=\{x\mid p_{\underset{\sim}{\theta}}(x)>0\}$ به $\underset{\sim}{\theta}$ بستگی نداشته باشد.

ج- $\frac{\partial}{\partial\theta_i}f_{\underset{\sim}{\theta}}(x)$ برای هر $i=1,2,\ldots,k$ و برای تمام $\underset{\sim}{\theta}\in\Theta$ موجود باشد.

د- $\forall\ \ i=1,2,\ldots,k\quad \frac{\partial}{\partial\theta_i}\int f_{\underset{\sim}{\theta}}(x)d\mu(x)=\int\frac{\partial}{\partial\theta_i}f_{\underset{\sim}{\theta}}(x)d\mu(x)$

هـ- $\frac{\partial}{\partial\theta_i}\ln f_{\underset{\sim}{\theta}}(x),\ \ i=1,2,\ldots,k$ مستقل خطی باشند.

و – برای هر برآوردگر T پارامتر $g(\theta)$ داشته باشیم که

$$\forall\ \ i=1,2,\ldots,k\quad \frac{\partial}{\partial\theta_i}E_\theta[T(\underset{\sim}{X})]=\frac{\partial}{\partial\theta_i}\int T(x)f_{\underset{\sim}{\theta}}(x)d\mu(x)=\int T(x)\frac{\partial}{\partial\theta_i}f_{\underset{\sim}{\theta}}(x)d\mu(x)$$

تعریف٦-١: ماتریس $\underset{k\times k}{I(\underset{\sim}{\theta})}=(I_{ij}(\underset{\sim}{\theta}))$ که در آن

$$I_{ij}(\underset{\sim}{\theta})=Cov\left(\frac{\partial}{\partial\theta_i}\ln f_{\underset{\sim}{\theta}}(X),\frac{\partial}{\partial\theta_j}\ln f_{\underset{\sim}{\theta}}(X)\right)=E[\frac{\partial}{\partial\theta_i}\ln f_{\underset{\sim}{\theta}}(X)\frac{\partial}{\partial\theta_j}\ln f_{\underset{\sim}{\theta}}(X)]$$

را ماتریس اطلاع فیشر $\underset{\sim}{\theta}$ گویند. می‌توان نشان داد که اگر شرط (هـ) برقرار باشد آنگاه مـاتریس $I(\underset{\sim}{\theta})$ معین مثبت است (زیرا ماتریس واریانس کواریانس معین نامنفی اسـت $\left(\cdot \le V(\sum_{i=1}^{n} a_i X_i) = \underset{\sim}{a}'\Sigma\underset{\sim}{a}\right)$ و در صورتی معین مثبت است که متغیرهای تصادفی بردار تـصادفی مربوطـه وابـسته خطـی نباشـند.) همچنین اگر در شرایط (ج) و (د) بتوان مشتق مرتبـه اول را بـا مـشتق مرتبـه دوم عـوض کـرد، آنگـاه می‌توان نشان داد که

$$I_{ij}(\underset{\sim}{\theta}) = -E\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln f_{\underset{\sim}{\theta}}(X)\right]$$

توجه کنید که اگر تغییر پارامتر زیر را داشته باشیم $\theta_i = h_i(\zeta_1,\zeta_2,...,\zeta_k)\ \ i=1,2,...,k$ و قراردهیم $J_{k\times k} = \left[\frac{\partial\theta_i}{\partial\zeta_j}\right]_{ij}$ آنگاه

$$I^*_{ij}(\underset{\sim}{\zeta}) = E\left[\frac{\partial}{\partial\zeta_i}\ln p_{\theta(\zeta)}(X)\frac{\partial}{\partial\zeta_j}\ln p_{\theta(\zeta)}(X)\right]$$
$$= \sum_{r=1}^{k}\sum_{\ell=1}^{k} E\left[\frac{\partial}{\partial\theta_r}\ln p_\theta(X)\frac{\partial}{\partial\theta_\ell}\ln p_\theta(X)\right]\frac{\partial\theta_r}{\partial\zeta_i}.\frac{\partial\theta_\ell}{\partial\zeta_j}$$

$$\Rightarrow I^*(\underset{\sim}{\zeta}) = JI(\underset{\sim}{\theta})J'$$

مثال ٦-١: فرض کنید X دارای چگالی از خانواده نمایی k پارامتری بفرم زیر باشد

$$p_\theta(x) = \left[e^{\sum_{j=1}^{k} c_j(\underset{\sim}{\theta})T_j(x) - B(\underset{\sim}{\theta})}\right]h(x)$$

و قرار دهیم $\tau_i = E(T_i(X))\qquad i=1,2,...,k$ در این صورت نشان دهید که $I(\underset{\sim}{\tau}) = C^{-1}$ که C ماتریس واریانس کواریانس $(T_1,...,T_k)$ است.

حل: اگر تغییر پارامتر $\eta_j = c_j(\underset{\sim}{\theta}),\ \ j=1,...,k$ را بدهیم آنگاه توزیع به صورت

$$p_\theta(x) = \left[e^{\sum_{j=1}^{k}\eta_j T_j(x) - A(\underset{\sim}{\eta})}\right]h(x)$$

در می آید و در این حالت

$$I^{*}_{j,k}(\underset{\sim}{\eta}) = \frac{\partial^{۲}}{\partial\eta_j \partial\eta_k} A(\underset{\sim}{\eta}) \Rightarrow I^{*}(\underset{\sim}{\eta}) = C = \left[Cov(T_j, T_k) \right]_{jk}$$

همچنین $\tau_j = E(T_j(X)) = \frac{\partial}{\partial\eta_j} A(\underset{\sim}{\eta})$ پس طبق مطالب فوق

$$J = \left[\frac{\partial\tau_j}{\partial\eta_i} \right]_{i,j} = \left[\frac{\partial^{۲}}{\partial\eta_i \partial\eta_j} A(\underset{\sim}{\eta}) \right]_{i,j} = \left[Cov(T_i, T_j) \right]_{i,j} = C$$

و بنابراین

$$C = I^{*}(\underset{\sim}{\eta}) = JI(\underset{\sim}{\tau})J' = CI(\underset{\sim}{\tau})C' \;\Rightarrow\; I(\underset{\sim}{\tau}) = C^{-۱} \qquad \blacksquare$$

برای ساختن نامساوی اطلاع ابتدا قضیه زیر را می‌آوریم.

قضیه ۶-۱: برای هر برآوردگر T برای $g(\theta)$ و هر تابع $\psi_i(x, \underset{\sim}{\theta})$ با گشتاور مرتبه دوم متناهی داریم که

$$V_\theta(T) \geq \underset{\sim}{\gamma}' C^{-۱} \underset{\sim}{\gamma} \qquad (۱)$$

که در آن $\underset{\sim}{\gamma}' = (\gamma_۱, \ldots, \gamma_k)$ و $C = (C_{ij})$ و $\gamma_i = Cov(T, \psi_i)$ و $C_{ij} = Cov(\psi_i, \psi_j)$.

طرف راست رابطه (۱) تنها از طریق $g(\theta) = E_\theta(T)$ بستگی به T دارد به شرط آنکه هر کدام از ψ_i ها در شرط $Cov(U, \psi_i) = ۰$ برای هر $U \in \Delta$. صدق کنند.

اثبات: برای هر ثابت $a_۱, a_۲, \ldots, a_k$ داریم که

$$\left[Cov(T, \sum_{i=۱}^{k} a_i \psi_i) \right]^{۲} \leq V(T) V(\sum_{i=۱}^{k} a_i \psi_i)$$

$$\Rightarrow V(T) \geq \frac{\left[Cov(T, \sum_{i=۱}^{k} a_i \psi_i) \right]^{۲}}{V(\sum_{i=۱}^{k} a_i \psi_i)}$$

اما

$$Cov(T, \sum_{i=۱}^{k} a_i \psi_i) = \sum_{i=۱}^{k} a_i Cov(T, \psi_i) = \sum_{i=۱}^{k} a_i \gamma_i = \underset{\sim}{a}' \underset{\sim}{\gamma}$$

$$V\left(\sum_{i=1}^{k} a_i \psi_i\right) = V\left(\underset{\sim}{a}'\underset{\sim}{\psi}\right) = \underset{\sim}{a}' V\left(\underset{\sim}{\psi}\right)\underset{\sim}{a} = \underset{\sim}{a}' C \underset{\sim}{a}$$

با استفاده از نامساوی کشی– شوارتز تعمیم یافته ($\underset{\sim}{a} \neq \underset{\sim}{0}, \forall \underset{\sim}{a}$) $\left[\underset{\sim}{a}'\underset{\sim}{\gamma}\right]^2 \leq (\underset{\sim}{a}' C \underset{\sim}{a})(\underset{\sim}{\gamma}' C^{-1} \underset{\sim}{\gamma})$ که در آن C معین مثبت است) داریم که

$$V(T) \geq \frac{\left[\underset{\sim}{a}'\underset{\sim}{\gamma}\right]^2}{\underset{\sim}{a}' C \underset{\sim}{a}} \quad \forall \underset{\sim}{a} \neq \underset{\sim}{0} \quad \Rightarrow V(T) \geq \max_{\underset{\sim}{a} \neq \underset{\sim}{0}} \frac{\left[\underset{\sim}{a}'\underset{\sim}{\gamma}\right]^2}{\underset{\sim}{a}' C \underset{\sim}{a}} = \underset{\sim}{\gamma}' C^{-1} \underset{\sim}{\gamma}$$

$$\therefore V(T) \geq \underset{\sim}{\gamma}' C^{-1} \underset{\sim}{\gamma} \qquad \blacksquare$$

قضیه ۶-۲: فرض کنید X دارای تابع چگالی $f_\theta(x)$ باشد اگر T برآوردگر برای $g(\theta)$ باشد که $E(T^2) < \infty$ و شرایط نظم برقرار باشد آنگاه

$$V_{\underset{\sim}{\theta}}(T) \geq \underset{\sim}{\alpha}' I^{-1}(\theta) \underset{\sim}{\alpha} \qquad (۲)$$

که در آن $\underset{\sim}{\alpha}' = (\alpha_1, \ldots, \alpha_k)$ و $\alpha_i = \dfrac{\partial}{\partial \theta_i} E_{\underset{\sim}{\theta}}[T(X)]$ و نامساوی فوق را نامساوی اطلاع گویند.

اثبات: در قضیه ۶-۱ قرار می دهیم $\psi_i(x, \underset{\sim}{\theta}) = \dfrac{\partial}{\partial \theta_i} \ln f_\theta(x)$ بنابراین ماتریس C حاصل بنابر شرط نظم (هـ) معین مثبت است و نتیجه به دست می‌آید.

$$\gamma_i = E(T\psi_i) = \int T \frac{\partial}{\partial \theta_i} f_{\underset{\sim}{\theta}}(x) d\mu(x) = \frac{\partial}{\partial \theta_i} E_{\underset{\sim}{\theta}}(T(X)) = \alpha_i \qquad \blacksquare$$

مثال۶-۲: فرض کنید $X_1, X_2, \ldots X_n$ یک نمونه تصادفی از $N(\mu, \sigma^2)$ باشند. CRLB برای واریانس برآوردگر ناریب پارامتر $E(X^2) = \mu^2 + \sigma^2$ را بدست آورید.

حل:

$$\ln f_{\underset{\sim}{\theta}}(\underset{\sim}{X}) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\frac{\partial}{\partial \mu}\ln f_{\underset{\sim}{\theta}}(\underset{\sim}{x}) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) \qquad \frac{\partial^2}{\partial \mu^2}\ln f_{\underset{\sim}{\theta}}(\underset{\sim}{x}) = \frac{n}{\sigma^2}$$

$$\frac{\partial}{\partial \sigma^2}\ln f_{\underset{\sim}{\theta}}(\underset{\sim}{x}) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum(x_i - \mu)^2$$

$$\frac{\partial^2}{\partial (\sigma^2)^2}\ln f_{\underset{\sim}{\theta}}(\underset{\sim}{x}) = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum(x_i - \mu)^2 \qquad \frac{\partial^2}{\partial \mu \partial \sigma^2}\ln f_{\underset{\sim}{\theta}}(\underset{\sim}{x}) = -\frac{1}{\sigma^4}\sum(x_i - \mu)$$

$$\therefore I(\mu,\sigma^{۲}) = \begin{bmatrix} \frac{n}{\sigma^{۲}} & ۰ \\ ۰ & \frac{n}{۲\sigma^{۴}} \end{bmatrix}$$

■ $\underset{\sim}{\alpha}' = (۲\mu, ۱) \quad \Rightarrow CRLB = \underset{\sim}{\alpha}' I^{-۱} \underset{\sim}{\alpha}' = \begin{bmatrix} ۲\mu & ۱ \end{bmatrix} \begin{bmatrix} \frac{\sigma^{۲}}{n} & ۰ \\ ۰ & \frac{۲\sigma^{۴}}{n} \end{bmatrix} \begin{bmatrix} ۲\mu \\ ۱ \end{bmatrix} = \frac{۲\mu^{۲}\sigma^{۲}}{n} + \frac{۲\sigma^{۴}}{n}$

توجه: اگر $I_{ii}(\underset{\sim}{\theta})$ اطلاع فیشر در حالتی باشد که θ_i نامعلوم و مابقی اعضای $\underset{\sim}{\theta}$ بجز θ_i معلوم باشند T آنگاه $I_{ii}^{-۱}(\underset{\sim}{\theta}) \le \left[I^{-۱}(\theta) \right]_{ii}$ و در نتیجه

$$\frac{\left[\frac{\partial g(\underset{\sim}{\theta})}{\partial \theta_i} \right]^{۲}}{I_{ii}(\underset{\sim}{\theta})} \le \left[\frac{\partial g(\underset{\sim}{\theta})}{\partial \theta_i} \right]^{۲} \left[I^{-۱}(\underset{\sim}{\theta}) \right]_{ii}$$

یعنی اینکه تعریف اطلاع فیشر چند پارامتری و نامساوی اطلاع چندپارامتری باعث می‌شود که CRLB (در یک پارامتری) تیزتر[۱] شود. برای رسیدن به این نامساوی توجه کنید که اگر در قضیه ۶-۱ قرار دهیم $\underset{\sim}{a}' = (۰,....,۰,۱,۰,.....,۰)$ آنگاه

$$CRLB = \frac{\left[Cov(T, \psi_i) \right]^{۲}}{V(\psi_i)} = \frac{\underset{\sim}{a}' \underset{\sim}{\alpha}}{\underset{\sim}{\alpha}' I(\theta) \underset{\sim}{\alpha}} \le \underset{\sim}{\alpha}' I^{-۱}(\theta) \underset{\sim}{\alpha} \qquad , \alpha_k = \begin{cases} \frac{\partial g(\underset{\sim}{\theta})}{\partial \theta_i} & k = i \\ ۰ & k \ne i \end{cases}$$

$$\Rightarrow \frac{\left[\frac{\partial g(\underset{\sim}{\theta})}{\partial \theta_i} \right]^{۲}}{I_{ii}(\underset{\sim}{\theta})} \le \left[\frac{\partial g(\underset{\sim}{\theta})}{\partial \theta_i} \right]^{۲} \left[I^{-۱}(\underset{\sim}{\theta}) \right]_{ii} \Rightarrow I_{ii}^{-۱}(\underset{\sim}{\theta}) \le \left[I^{-۱}(\underset{\sim}{\theta}) \right]_{ii}$$

همچنین توجه کنید که اگر پارامترها بر هم متعامد باشند آنگاه $\forall i \ne j \quad I_{ij}(\underset{\sim}{\theta}) = ۰$ و در این صورت نامساوی فوق به تساوی تبدیل می‌شود. مثال ۶-۲ این حالت را نشان می‌دهد.

[۱] Sharp

فصل ۳

پایایی

بخش۱: مفهوم پایائی[1] و کلیات

در فصل قبل کلاس برآوردگرها را محدود به کلاس برآوردگرهای نااریب کردیم و در این کلاس بهترین برآوردگرهای نااریب را پیدا کردیم. در این قسمت می‌خواهیم خود را محدود به کلاس دیگری از برآوردگرها، یعنی برآوردگرهای پایا کنیم و در این کلاس بهترین برآوردگر پایا را پیدا کنیم.

برای پی بردن به مفهوم پایائی به مثال زیر توجه کنید.

مثال۱-۱: فرض کنید متغیر تصادفی X طول عمر یک قطعه برحسب ساعت باشد که دارای توزیع نمایی با تابع چگالی زیر است:

$$f_\theta(x)=\frac{1}{\theta}e^{-x/\theta} \qquad x>0,\ \theta>0$$

می‌خواهیم براساس یک مشاهده X پارامتر θ را براساس تابع زیان $L(\theta,\delta)=(1-\frac{\delta}{\theta})^2$ برآورد کنیم. فرض کنید $\delta(x)$ یک برآورد پارامتر θ براساس X (طول عمر برحسب ساعت) باشد.

حال اگر شخص دیگری طول عمر این قطعه را بر حسب دقیقه اندازه‌گیری کند و در این حالت متغیر تصادفی Y را طول عمر قطعه برحسب دقیقه در نظر بگیرد در این صورت بایستی $Y=60X$ باشد.

اگر قرار دهیم $\eta=60\theta$ در این صورت به راحتی دیده می‌شود که Y دارای تابع چگالی زیر است:

$$f_\eta(y)=\frac{1}{60}f_\theta(\frac{y}{60})=\frac{1}{60\theta}e^{-\frac{y}{60\theta}} \quad\Rightarrow\quad f_\eta(y)=\frac{1}{\eta}e^{-y/\eta} \qquad y>0,\eta>0$$

حال اگر در این اندازه‌گیری جدید برآورد δ^* را برای پارامتر η بدست آوریم انتظار داریم که

$$\delta^*(y)=60\delta(X) \qquad (1)$$

باشد(زیرا منطقی به نظر می‌رسد که برآوردگرهای یکسانی را صرف نظر از نوع اندازه‌گیری بکار ببریم.) در این حالت:

$$L(\eta,\delta^*)=(1-\frac{\delta^*}{\eta})^2=(1-\frac{60\delta}{60\theta})^2=(1-\frac{\delta}{\theta})^2=L(\theta,\delta)$$

بنابراین ساختار کلی مسئله با اندازه‌گیری برحسب دقیقه (یعنی کلاس توابع چگالی، کلاس فضای پارامتری و کلاس توابع زیان) دقیقا با ساختار کلی مسئله با اندازه‌گیری برحسب ساعت یکی می‌شود.

---

[1] Inrariance

در نتیجه منطقی به نظر می‌رسد که در این مسئله برآوردهایی را مورد استفاده قرار دهیم که تحت این دو ساختار یکسان باشند یعنی:

$$\delta(y) = \delta^*(y) \qquad (۲)$$

بنابراین از روابط (۱) و (۲) داریم که:

$$\delta(۶۰x) = ۶۰\delta(x) \qquad (۳)$$

برآوردگر δ که در شرط (۳) صدق می‌کند را برآوردگر هم پایا[۱] می‌نامند و همانطور که دیده می‌شود کلاس برآوردگرهای هم پایا یک قسمت کوچک از کلاس کلیه برآوردگرها است.

مسئله فوق برای حالت کلی‌تر یعنی $Y = cX$, $c > ۰$ نیز برقرار است و در این حالت برآوردگر هم‌پایا به صورت زیر است:

$$\delta(cx) = c\delta(x) \qquad \forall c > ۰ \qquad (٤)$$

چون این رابطه برای هر x برقرار است پس برای $x = \frac{۱}{c}$ نیز برقرار است و در نتیجه از رابطه (٤) داریم که

$$\delta(۱) = \frac{۱}{x}\delta(x) \Rightarrow \delta(x) = \delta(۱)x \Rightarrow \delta(x) = kx$$

بنابراین در این حالت کلاس برآوردگرهای هم‌پایا عبارت است از:

$$\mathbf{D}_E = \{\delta \in \mathbf{D} \mid \delta(x) = kx,\ \forall x,\ \forall k > ۰\}$$

در این فصل می‌خواهیم در بین برآوردگرهای هم‌پایا برای یک پارامتر مجهول آن برآوردگری را پیدا کنیم که دارای کمترین مخاطره[۲] باشد، که به آن برآوردگر هم‌پایا با کمترین مخاطره (MRE[۳]) گویند.

در این مثال داریم که:

$$R(\theta,\delta) = E_\theta\left[(۱ - \frac{\delta(X)}{\theta})^۲\right] = \frac{۱}{\theta^۲}E_\theta\left[(kX - \theta)^۲\right] = \frac{۱}{\theta^۲}\left[V(kX) + \left[E(kX - \theta)\right]^۲\right]$$

$$= \frac{۱}{\theta^۲}\left[k^۲\theta^۲ + (k-۱)^۲\theta^۲\right] = k^۲ + (k-۱)^۲ = ۲k^۲ - ۲k + ۱$$

---

[۱] Equivariance
[۲] Risk
[۳] Minimom Risk Equivariant Estimator

بنابراین تابع مخاطره ثابت است و به θ بستگی ندارد. در نتیجه برآوردگر MRE با مینیمم کردن $R(\theta,\delta)$ نسبت به k بدست می آید:

$$\frac{\partial R(\theta,\delta)}{\partial k} = ۴k - ۲ = ۰ \Rightarrow k = \frac{۱}{۲}, \qquad \frac{\partial^۲ R(\theta,\delta)}{\partial k^۲} = ۴ > ۰$$

بنابراین برآوردگر $\delta^*(X) = \frac{۱}{۲}X$ برآوردگری هم‌پایا با کمترین‌مخاطره است. پس برآوردگر MRE پارامتر θ است. ■

اصول پایائی:

با توجه به مثال بالا می‌توان خود را محدود به کلاس برآوردگرهای پایا کنیم. اما پایائی را می‌توان از دو جنبه در نظر گرفت.

۱-پایائی اندازه‌گیری[۱] (پایائی طبیعی یا منطقی یا تابعی)

این جنبه از پایائی می‌گوید که استنباط نبایستی به نوع اندازه‌گیری داده‌ها بستگی داشته باشد.

(مثلا در مثال ۱-۱ اگر $\delta(X)$ برای برآورد θ بکار برده می‌شود طبیعی به نظر می‌رسد که $\delta^*(Y) = c\delta(X)$ برای برآورد $\eta = c\theta$ بکار برده شود.)

۲- پایائی ساختاری[۲](یا ریاضی)

این جنبه از پایائی می‌گوید که اگر دو مسئله استنباطی دارای یک ساختار از لحاظ فضای پارامتری، تابع چگالی و تابع زیان باشند آنگاه بایستی روش استنباط یکسانی را در دو مسئله مورد استفاده قرار دهیم.

(مثلا در مثال ۱-۱ برآورد $\delta(Y) = \delta(cX)$ را برای برآورد $\eta = c\theta$ بکار ببریم).

با تلفیق این دو جنبه می‌توان برآوردگرهای هم‌پایا را پیدا کرد(مثلا در مثال ۱-۱ $\delta(cX) = c\delta(X)$)

فرمول بندی و تعاریف پایایی

با توجه به مطالب بالا در حالت کلی ما یک نمونه تصادفی $\underset{\sim}{X} = (X_۱, X_۲, ..., X_n)$ با مقادیر $\not{x}$ و تابع چگالی یا احتمال توام $f_\theta(\underset{\sim}{x}) = f_\theta(x_۱, ..., x_n)$ داریم. پس بوسیله یک تبدیل یک به یک و

[۱] Measurement Invariance
[۲] Formal Invariance

پوشای $Y_{\sim} = g(X_{\sim})$ داده‌های $X_{\sim}$ را به داده‌های جدید $Y_{\sim}$ تبدیل می‌کنیم که این تبدیلات یک گروه از تبدیلات G را بوجود می‌آورند.

تعریف ۱-۱: یک گروه از تبدیلات روی $\mathscr{X}$ که آن را با G نمایش می‌دهیم یک مجموعه تبدیلات اندازه‌پذیر یک به یک و پوشا روی $\mathscr{X}$ است که در شرایط زیر صدق می‌کند:

$(i)\ \ if\ \ g_1, g_2 \in \mathsf{G} \Rightarrow g_1 g_2 \in \mathsf{G}$

$(ii)\ \forall g \in \mathsf{G}\ \ \exists g^{-1} \in \mathsf{G}\ \ \ni g^{-1}(g(x)) = x$

$(iii)\ \ \exists e \in \mathsf{G}\ \ \ni e(x) = x$

مثال ۱-۲: در مثال ۱-۱ گروه تبدیلات زیر را داریم $\mathsf{G} = \{g_c : g_c(X) = cX\ , c > 0\}$ که به آن گروه تبدیلات ضربی یا مقیاسی گویند. توابع g_c یک به یک و پوشا هستند.

$$g_{c_2} g_{c_1}(x) = g_{c_2}(g_{c_1}(x)) = g_{c_2}(c_1 x) = c_2 c_1 x = g_{c_2 c_1}(x) \Rightarrow g_{c_2} g_{c_1} = g_{c_2 c_1} \in g$$

$$g_c{}^{-1} = g_{c^{-1}} \in \mathsf{G}\ \ , e = g_1 \in \mathsf{G}$$

تعریف ۱-۲: فرض‌کنید $\mathsf{F} = \{f_\theta(x_{\sim}) : \theta \in \Theta\}$ یک خانواده از توزیع‌های احتمال برای $X_{\sim}$ باشند و G یک گروه از تبدیلات روی x باشد. خانواده F را تحت گروه G پایا[1] گویند اگر برای هر $\theta \in \Theta$ و $g \in \mathsf{G}$ یک $\theta' \in \Theta$ یکتا وجود داشته باشد بطوریکه $Y_{\sim} = g(X_{\sim})$ دارای تابع چگالی $f_{\theta'}(y_{\sim})$ متعلق به خانواده F باشد در این حالت θ' را با $\bar{g}(\theta)$ نمایش می‌دهند. ■

مثلا در مثال ۱-۱ تحت گروه تبدیلات $\mathsf{G} = \{g_c : g_c(X) = cX\ , c > 0\}$ خانواده توزیع نمایی پایا می‌باشد.

$$f_\theta(x) = \frac{1}{\theta} e^{-x/\theta} \quad \begin{matrix} x > 0 \\ \theta > 0 \end{matrix} \quad \xrightarrow[\eta = c\theta]{Y = cX} \quad f_\eta(y) = \frac{1}{2} e^{-y/\eta} \quad \begin{matrix} y > 0 \\ \eta > 0 \end{matrix}$$

اگر قرار دهیم $P_\theta^X(A) = P_\theta(X \in A)$ آنگاه تعریف ۱-۲ می‌گوید که خانواده F تحت تبدیلات G پایا است اگر برای هر $\theta \in \Theta$ و $g \in \mathsf{G}$ یک $\bar{g}(\theta) \in \Theta$ موجود باشد بطوری که $P_\theta^{g(X)} = P_{\bar{g}(\theta)}^X$ و یا:

[1] invariant

$$P_\theta(g(X)\in A)=P_{\bar g(\theta)}(X\in A) \qquad (۵)$$

زیرا

$$P_\theta^{g(X)}(A)=P_\theta(g(X)\in A)=\int_{\{x\,|\,g(x)\in A\}} f_\theta(x)\,d\mu(x)\overset{(*)}{=}\int_{\{x\,|\,g(x)\in A\}} f_{\bar g(\theta)}(g(x))\,d\mu(x)$$
$$\overset{Y=g(X)}{=}\int_{\{y\,|\,y\in A\}} f_{\bar g(\theta)}(y)\,dv(y)=P_{\bar g(\theta)}(X\in A)=P_{\bar g(\theta)}^X(A)$$

که در آن تساوی (*) از پایایی توزیع نتیجه می‌گردد. از رابطه (۵) برای هر تابع انتگرال پذیر h نتیجه می‌شود که

$$E_\theta[h(g(X)]=E_{\bar g(\theta)}[h(X)] \qquad (۶)$$

قضیه۱-۱: اگر توزیع‌های F $f_\theta\in\mathsf{F}$ مجزا باشند و خانواده F تحت تبدیلات G پایا باشد آنگاه گروه تبدیلات $\bar{\mathcal{G}}=\{\bar g: g\in\mathcal{G}\}$ روی Θ تشکیل یک گروه از تبدیلات را می‌دهد.

اثبات: مجزا بودن f_θ بدین معنی است که $\theta_1\neq\theta_2\Rightarrow p_{\theta_1}^X\neq p_{\theta_2}^X$. حال شرایط گروه تبدیل بودن G را بررسی می کنیم:

(i) $\bar g\in\mathsf{G}$ یک به یک است

$$Let\ C=g(A),\ \bar g(\theta_1)=\bar g(\theta_2)\Rightarrow P_{\bar g(\theta_1)}(X\in C)=P_{\bar g(\theta_2)}(X\in C)$$
$$\overset{(۵)}{\Rightarrow}P_{\theta_1}(g(X)\in C)=P_{\theta_2}(g(X)\in C)$$
$$\Rightarrow P_{\theta_1}(X\in g^{-1}(C))=P_{\theta_2}(X\in g^{-1}(C))\Rightarrow P_{\theta_1}(X\in A)=P_{\theta_2}(X\in A)\overset{(*)}{\Rightarrow}\theta_1=\theta_2$$

$(ii)\ \bar g\in\mathsf{G}$ *is onto, i.e.,* $\forall\theta\in\Theta\ \exists\theta^*\in\Theta$ ə $\bar g(\theta^*)=\theta$

$$P_\theta(X\in A)=P_\theta(g^{-1}(g(X))\in A)\overset{(۵)}{=}P_{\overline{g^{-1}}(\theta)}(g(X)\in A)$$
$$=P_{\overline{\bar g(g^{-1}(\theta))}}(X\in A)=P_{\bar g(\theta^*)}(X\in A)$$
$$\overset{(*)}{\Rightarrow}\bar g(\theta^*)=\theta\ where\ \theta^*=\overline{g^{-1}}(\theta)$$

$(iii)\ \bar{g}_1, \bar{g}_2 \in \bar{G} \Rightarrow \bar{g}_2\bar{g}_1 \in \bar{G}$

ابتدا نشان می‌دهیم که $\overline{g_2 g_1} = \bar{g}_2\bar{g}_1$

$$P_{\overline{g_2 g_1}(\theta)}(X \in A) \overset{(1)}{=} P_\theta(g_2 g_1(X) \in A) = P_\theta(g_1(X) \in g_2^{-1}(A)) \overset{(1)}{=} P_{\bar{g}_1(\theta)}(X \in g_2^{-1}(A))$$

$$= P_{\bar{g}_1(\theta)}(g_2(X) \in A)) \overset{(1)}{=} P_{\bar{g}_2\bar{g}_1(\theta)}(X \in A)$$

$$\overset{(*)}{\Rightarrow} \bar{g}_1\bar{g}_2 = \overline{g_2 g_1} \in \bar{G} \qquad (\text{since } g_2 g_1 \in G)$$

$(iv)\ \forall \bar{g} \in \bar{G}\ \ \exists \bar{g}^{-1} \in \bar{G}\ \ \ni \bar{g}\ \ \bar{g}^{-1}(\theta) = \theta$

ابتدا نشان می‌دهیم که $\bar{g}^{-1} = \overline{g^{-1}}$

$$P_\theta(X \in A) = P_\theta(g^{-1}g(X) \in A) \overset{(1)}{=} P_{\overline{g^{-1}(\theta)}}(g(X) \in A) = P_{\bar{g}(\overline{g^{-1}(\theta)})}(X \in A)$$

$$\overset{(*)}{\Rightarrow} \bar{g}(\overline{g^{-1}}(\theta)) = \theta \Rightarrow \overline{g^{-1}} = \bar{g}^{-1}$$

از طرفی

$$g \in G \Rightarrow g^{-1} \in G \Rightarrow \bar{g} \in \bar{G} \ \text{ and } \ \overline{g^{-1}} \in \bar{G}$$

$$\Rightarrow \bar{g}^{-1} = \overline{g^{-1}} \in G \ \text{ and } \ \bar{g}\bar{g}^{-1}(\theta) = \bar{g}\left(\bar{g}^{-1}(\theta)\right) = \overline{gg^{-1}}(\theta) = \bar{e}(\theta) = \theta$$

$(v)\ \exists \bar{e} \in \bar{G}\ \ \ni \bar{e}(\theta) = \theta$

$$P_\theta(X \in A) = P_\theta(e(X) \in A) \overset{(1)}{=} P_{\bar{e}(\theta)}(X \in A) \overset{(*)}{\Rightarrow} \bar{e}(\theta) = \theta,\ \bar{e} \in \bar{G}$$

که در آن کلیه روابط $(*)$ از مجزا بودن بدست آمده‌است. ■

برای مثال، در مثال ۱-۱، $\bar{G} = \{\bar{g}_c \mid \bar{g}_c(\theta) = c\theta, c > 0\}$ گروه تبدیلات بوجود آمده تحت خانواده توزیع های پایای نمایی، روی فضای پارامتری Θ می‌باشد.

تعریف ۱-۳: فرض کنید خانواده $\mathsf{F} = \{f_\theta(\underset{\sim}{x}) : \theta \in \Theta\}$ تحت گروه تبدیلات G پایا باشد. در این صورت تابع زیان $L(\theta, \delta)$ را تحت گروه G پایا گویند اگر برای هر $g \in G$ و هر $\delta \in D$ (کلاس برآوردگرها) یک $\delta^* \in D$ یکتا وجود داشته باشد بطوری که

$$L(\theta, \delta) = L(\bar{g}(\theta), \delta^*) \qquad \forall \theta \in \Theta$$

در این حالت δ^* را بوسیله $\tilde{g}(\delta)$ نشان می‌دهند و مسئله مورد نظر را تحت گروه تبدیلات G پایا گویند. ■

برای مثال، در مثال ۱-۱ تابع زیان $L(\theta,\delta)=(1-\frac{\delta}{\theta})^2$ تحت تبدیلات $G=\{g_c \mid g_c(X)=cX, c>0\}$ پایا بود زیرا

$$L(\bar{g}(\theta),\tilde{g}(\delta)=L(c\theta,c\delta)=(1-\frac{c\delta}{c\theta})^2=(1-\frac{\delta}{\theta})^2=L(\theta,\delta)$$

قضیه۱-۲: اگر خانواده F تحت تبدیلات G پایا باشد، انگاه گروه‌تبدیلات $\tilde{G}=\{\tilde{g}: g\in G\}$ (یعنی گروه تبدیلات بوجود آمده تحت این خانواده روی کلاس برآوردگرها) تشکیل یک گروه از تبدیلات را می‌دهند.

اثبات: تمرین (همانند قضیه۱-۱). ■

برای مثال، در مثال ۱-۱، $\tilde{G}=\{\tilde{g}_c \mid \tilde{g}_c(\delta)=c\delta, c>0\}$ گروه تبدیلات بوجود آمده تحت خانواده توزیع‌های پایای نمایی روی کلاس برآوردگرهای D می باشد.

برآوردگرهای هم‌پایا[1]

در قسمت قبل اصول پایایی را به صورت زیر معرفی کردیم:

۱- **پایائی اندازه‌گیری (طبیعی یا منطقی یا تابعی):** اگر $\delta(x)$ را برای برآورد $h(\theta)$ بکار می‌بریم طبیعی به نظر می‌رسد که $\tilde{g}(\delta(X))$ را برای برآورد $\tilde{g}(h(\theta))$ بکار ببریم.

۲- **پایائی ساختاری(یا ریاضی):** اگر دو مسئله استنباطی دارای یک ساختار از لحاظ فضای پارامتری، تابع چگالی و تابع زیان باشند آنگاه بایستی روش یکسانی را در دو مسئله مورد استفاده قرار دهیم. یعنی $\delta(g(X))$ را برای برآورد $\tilde{g}(h(\theta))$ بکار ببریم.

با توجه به تعاریف بالا و اصول پایائی، برآوردگر هم‌پایا را بصورت زیر تعریف می‌کنیم.

[1] Equivariant Estimators

تعریف۱-٤: فرض کنید مسئله مورد نظر ما تحت گروه تبدیلات G پایا باشد (تعاریف۱-۲و۱-۳) برآوردگر δ را تحت گروه تبدیلات G هم‌پایا گویند هرگاه برای هر $g \in G$ و هر $\underset{\sim}{x} \in \mathcal{X}$ داشته باشیم که $\delta(g(\underset{\sim}{X})) = \tilde{g}(\delta(\underset{\sim}{X}))$ ■

برای مثال، در مثال ۱-۱ شرط فوق معادل است با $\delta(cX) = c\delta(X)$.

با توجه به تعریف هم‌پایایی این سوال مطرح می‌شود که تحت یک تبدیل G کدامیک از توابع $h(\theta)$ دارای برآوردگر هم پایا هستند. اگر در تعریف ۱-٤ G را با $\bar{G}$ و $\mathcal{X}$ را با Θ عوض کنیم آنگاه یک تابع $h(\theta)$ طبق اصول پایایی دارای برآوردگر هم‌پایا است اگر در شرط زیر صدق می‌کند:

$$h(\bar{g}(\theta)) = \tilde{g}(h(\theta))$$

برای مثال، در مثال ۱-۱، $h(\theta) = \theta$ در شرط فوق صدق می کند زیرا :

$$h(\bar{g}(\theta)) = h(c\theta) = c\theta$$
$$\tilde{g}(h(\theta)) = c(h(\theta)) = c\theta$$

اما اگر بخواهیم $h(\theta) = \theta^r$ را برآورد کنیم آنگاه $h(\theta)$ در شرط فوق صدق نمی‌کند زیرا

$$h(\bar{g}(\theta)) = h(c\theta) = c^r\theta^r \qquad \tilde{g}(h(\theta)) = c\theta^r$$

بنابراین، در این مثال اگر بخواهیم $h(\theta) = \theta^r$ را برآورد کنیم بایستی از ابتدا خانواده $\tilde{G}$ را به‌صورتی درنظر بگیریم که $\tilde{g}(h(\theta)) = c^r\theta^r$ باشد یعنی $\tilde{G} = \{\tilde{g}_c \mid \tilde{g}_c(\delta) = c^r\delta\}$.

مثال ۱-۳: فرض کنید $X \sim B(n,p)$ که در آن n معلوم و p پارامتر مجهول است و می‌خواهیم آن را تحت تابع زیان $L(p,\delta) = (\delta - p)^2$ برآورد کنیم. فرض کنید $T(X)$ برآوردگر پارامتر p براساس X باشد.

حال فرض کنید شخصی بخواهد احتمال شکست یعنی $q = 1 - p$ را تحت تابع زیان درجه دوم فوق برآورد نماید. این شخص از تعداد شکست‌ها در n آزمایش یعنی $Y = n - X \sim B(n,q)$ استفاده می‌نماید. بنابراین در این مسئله گروه تبدیلات زیر را داریم:

$$G = \{g_1, g_2\} \qquad g_1(x) = n - x, \quad g_2(x) = x$$

چون براساس این تبدیل $Y = g_1(X) \sim B(n,1-p)$ و $Y = g_2(X) \sim B(n,p)$ پس گروه تبدیلات $\bar{G}$ تولید شده توسط G عبارت است از

$$\bar{G} = \{\bar{g}_1, \bar{g}_2\} \qquad \bar{g}_1(p) = 1-p \quad \bar{g}_2(p) = p$$

حال چون می‌خواهیم $h(p) = p$ یا $h(p) = 1-p$ را برآورد کنیم پس بایستی

$$\tilde{g}(1-p) = \tilde{g}(h(p)) = h(\bar{g}_1(p)) = h(1-p) = p$$
$$\tilde{g}_2(p) = \tilde{g}_2(h(p)) = h(\bar{g}_2(p)) = h(p) = p$$

بنابراین گروه تبدیلات زیر روی D بوجود می‌آید

$$\tilde{G} = \{\tilde{g}_1, \tilde{g}_2\} \qquad \tilde{g}_1(\delta) = 1-\delta, \quad \tilde{g}_2(\delta) = \delta$$

همچنین توجه کنید که

$$L(\bar{g}_1(p), \tilde{g}_1(\delta)) = L(1-p, 1-\delta) = (\delta - p)^2 = L(p,\delta)$$

یعنی تابع زیان درجه دم تحت این تبدیل پایا است.

حال طبق تعریف ۱-٤ برآوردگر δ در این مسئله هم‌پایا است اگر

$$\delta(g(X)) = \tilde{g}(\delta(X)) \Leftrightarrow \delta(n - X) = 1 - \delta(X)$$

از جمله برآوردگرهای پایا در این مسئله $T_1(x) = \frac{x}{n}$ و $T_2(x) = \alpha(\frac{x}{n}) + (1-\alpha)(0.5)$ می‌باشد.

اگر برآوردگرهای پایا بفرم بالا را در نظر بگیریم در این صورت تعیین این برآوردگر در نقاط $T(0), T(1), \ldots T\left(\left[\frac{n}{2}\right]\right)$ کافی می‌باشد زیرا طبق رابطه بالا

$$T(n) = -T(0), \; T(n-1) = -T(1), \ldots.$$

در حالی که اگر یک برآوردگر عادی را بکار ببریم بایستی این برآوردگر را در نقاط $T(0), T(1), \ldots T(n)$ تعیین کنیم و بنابراین برآوردگرهای پایا نیز روشی برای خلاصه کردن داده‌ها ارایه می‌دهند. ■

بهترین برآوردگر هم پایا

حال این سوال مطرح می‌شود که در بین برآوردگرهای هم پایا کدام برآوردگر بهتر است. جواب این سوال این است که برآوردگر هم پایائی که دارای کمترین مخاطره باشد بهترین برآوردگر هم پایا می‌باشد. برای بدست آوردن بهترین برآوردگر هم پایا از تعاریف و قضایای زیر استفاده می‌کنیم.

تعریف ۱-۵: دو نقطه θ_1,θ_2 در Θ را معادل[1] گویند اگر $\bar{g}\in \bar{G}$ موجود باشد که $\theta_2=\bar{g}(\theta_1)$. یک مدار[2] در Θ عبارت از یک کلاس (مجموعه) از نقاط معادل در Θ است. بنابراین مدار $\theta_.$ در Θ عبارت از مجموعه زیر است: $\Theta(\theta_.)=\{\bar{g}(\theta_.)\mid \bar{g}\in \bar{G}\}$. ■

معادل بودن در تعریف ۱-۵ یک رابطه هم‌ارزی را بدست می‌دهد. بنابراین تعریف فوق Θ را به کلاس‌های هم‌ارزی افراز می‌کند. هر یک از این کلاس‌ها را یک مدار (تحت $\bar{G}$) می‌نامیم.

تعریف ۱-۶: یک گروه $\bar{G}$ از تبدیلات روی Θ را انتقالی یا گذرا[3] گویند، هرگاه Θ دارای تنها یک مدار باشد و یا معادلا اگر برای هر دو نقطه $\theta_1,\theta_2\in\Theta$ یک $\bar{g}\in \bar{G}$ موجود باشد بطوریکه $\theta_2=\bar{g}(\theta_1)$. ■

برای مثال، در مثال ۱-۱ $\bar{G}$ تنها دارای یک مدار است $(\theta_2=\frac{\theta_2}{\theta_1}\theta_1=c\theta_1)$ پس یک گروه انتقالی است ولی در مثال ۱-۳ چون $\bar{g}(p)=1-p$ پس مدارها همان زوج‌های مرتب $(p,1-p)$ هستند و در نتیجه $\bar{G}$ انتقالی نیست. در این گروه نقاط $p_1=\frac{1}{3},p_2=\frac{2}{3}$ را می‌توان به هم انتقال داد ولی نقاط $p_1=\frac{1}{3},p_2=\frac{2}{3}$ را نمی‌توان به هم انتقال داد.

قضیه ۱-۳(مهم): تابع مخاطره هر برآوردگر هم‌پایا روی مدارهای Θ مقداری ثابت دارد و یا معادلا

$$R(\theta,\delta)=R(\bar{g}(\theta),\delta)\quad \forall\theta\in\Theta,\bar{g}\in \bar{G}$$

---

[1] Equivalent
[2] Orbit
[3] Transitive

اثبات:

$$R(\theta,\delta)=E_\theta[L(\theta,\delta(\underset{\sim}{X}))]$$

$$=E_\theta[L(\bar{g}(\theta),\tilde{g}(\delta(\underset{\sim}{X})))] \quad \text{(پایائی تابع زیان)}$$

$$=E_\theta[L(\bar{g}(\theta),\delta(g(\underset{\sim}{X})))] \quad \text{(هم‌پایائی برآوردگر } \delta\text{)}$$

$$=E_{\bar{g}(\theta)}[L(\bar{g}(\theta),\delta(\underset{\sim}{X}))] \quad \text{(پایائی توزیع، رابطه (٦))}$$

$$=R(\bar{g}(\theta),\delta) \quad \blacksquare$$

نتیجه ۱-۱: اگر $\bar{G}$ یک گروه انتقالی از تبدیلات روی فضای پارامتری Θ باشد آنگاه تابع مخاطره هر برآوردگر هم پایا ثابت است و به θ بستگی ندارد. ■

توجه: با توجه به نتیجه فوق اگر $\bar{G}$ یک گروه انتقالی باشد آنگاه تابع مخاطره برآوردگرهای هم‌پایا نسبت به θ ثابت است و در نتیجه بهترین برآوردگر هم پایا (MRE) آن برآوردگری است که این تابع مخاطره ثابت را مینیمم می‌کند.

برای مثال، در مثال ۱-۱ دیدیم که تابع مخاطره برآوردگر هم‌پایای $\delta(X)=kX$ برابر $R(\theta,\delta)=۲k^۲-۲k+۱$ بود که به θ بستگی نداشت و مینیمم آن به ازای $k=\frac{۱}{۲}$ بدست آمد، پس برآوردگر MRE عبارت از $\delta^*(X)=\frac{۱}{۲}X$ بود.

تعریف ۱-۷: فرض کنید G گروهی از تبدیلات روی فضای $\mathcal{X}$ باشد. تابع $T(X)$ را روی $\mathcal{X}$ نسبت به G پایا[۱] گویند هرگاه برای هر $x\in\mathcal{X}$ و $g\in G$ داشته باشیم که $T(g(x))=T(x)$. تابع $T(x)$ را پایای ماکزیمال[۲] نسبت به G گویند هرگاه T پایا باشد و در شرط زیر صدق کند

$$T(x_۱)=T(x_۲)\Rightarrow x_۱=g(x_۲) \quad for\ some\ g\in G \qquad \blacksquare$$

مفاهیم فوق برای درک این مطلب که $\mathcal{X}$ می‌تواند به مدارهایی از نقاط که تحت G معادل هستند، مفید می‌باشد. یک تابع پایا تابعی است که روی هر مدار مقداری ثابت دارد. $\left(x'\sim x\Rightarrow T(x')=T(x)\right)$ و یک تابع پایای ماکزیمال تابعی است که روی هر مدار مقداری ثابت

[۱] Invariant
[۲] Maximal Invariant

دارد و مقادیر متفاوتی را به هر یک از مدارهای متفاوت می‌دهد. توجه کنید که اگر G یک گروه انتقالی روی $\mathcal{X}$ باشد آنگاه $\mathcal{X}$ خود یک مدار است و بنابراین تنها تابع پایائی ماکزیمال روی $\mathcal{X}$ تابع ثابت است.

بدیهی است که اگر T یک آماره پایای ماکزیمال باشد، هر تابع یک به یک از آن نیز یک آماره پایای ماکزیمال است. پس آماره پایای ماکزیمال یکتا نیست ولی همه آنها توابعی یک به یک از یکدیگر هستند.

مثال ۱-٤: فرض کنید $\mathcal{X} = \Re^n$ و G گروه تبدیلات مکانی باشد.

$$G = \{g_c : g_c(x_1, x_2, ..., x_n) = (x_1 + c, ..., x_n + c), c \in R\}$$

در این صورت $T(\underset{\sim}{x}) = (x_1 - x_n, ..., x_{n-1} - x_n)$ پایای ماکزیمال است. چون $(x_i + c) - (x_n + c) = x_i - x_n$ پس $T(g_c(\underset{\sim}{x})) = T(\underset{\sim}{x})$ پس T پایا است. حال اگر $T(\underset{\sim}{x}) = T(\underset{\sim}{x}')$ آنگاه $(x_i - x_n) = (x_i' - x_n)$ و در نتیجه با $c = x_n - x_n'$ داریم که $g_c(\underset{\sim}{x}') = \underset{\sim}{x}$ پس T پایای ماکزیمال است. ■

مثال ۱-۵: فرض کنید $\mathcal{X} = \Re^n$ و G گروه تبدیلات مقیاسی باشد.

$$G = \{g_c : g_c(x_1, ..., x_n) = (cx_1, ..., cx_n), c > 0\}$$

اگر $Z = \sum_{i=1}^{n} X_i^2$ آنگاه $T(\underset{\sim}{X}) = \begin{cases} 0 & if \quad Z = 0 \\ (\frac{X_1}{Z}, ... \frac{X_n}{Z}) & if \quad Z \neq 0 \end{cases}$ یک تابع پایای ماکزیمال است.

مثال ۱-٦: فرض کنید $\mathcal{X} = \Re^n$ و G گروه تبدیلات مکانی- مقیاسی باشد

$$G = \{g_{b,c} \mid g_{b,c}(x_1, ..., x_n) = (cx_1 + b, ..., cx_n + b), c > 0, b \in R\}$$

اگر $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ و $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

آنگاه $T(\underset{\sim}{x}) = \begin{cases} 0 & s = 0 \\ (\frac{x_1 - \bar{x}}{s}, ..., \frac{x_n - \bar{x}}{s}) & s \neq 0 \end{cases}$ تابع پایای ماکزیمال است.

توجه: پایایی و پایایی ماکزیمال را روی Θ بطور مشابه می‌توان تعریف کرد.

قضیه ۱-٤: فرض کنید G گروهی از تبدیلات روی فضای $\mathcal{X}$ باشد و $T(X)$ یک پایای ماکزیمال باشد. در این صورت تابع $h(X)$ نسبت به G پایا است اگر و فقط اگر h تابعی از T باشد.

اثبات:

(کفایت): فرض کنید $h(x)=s(T(x))$ در این صورت

$h(g(x))=s(T(g(x))=s(T(x))=h(x)$ پس h پایا است.

(لزوم): فرض کنید h پایا باشد و $T(x_1)=T(x_2)$ در این صورت چون T پایای ماکزیمال است پس $x_1=g(x_2)$ و در نتیجه $h(x_1)=h(g(x_2))=h(x_2)$ یعنی h تابعی از T است. ■

قضیه ۱-۵: اگر h یک آماره پایا باشد آنگاه توزیع آن روی هر مدار Θ مقداری ثابت است.

اثبات: فرض کنید θ_1,θ_2 در یک مدار Θ باشند در این صورت $\bar{g}\in \bar{G}$ وجود دارد که $\theta_2=\bar{g}(\theta_1)$ و بنابراین

$$P_{\theta_2}(h(X)\in A)=P_{\bar{g}(\theta_1)}(h(X)\in A)\overset{(5)}{=}P_{\theta_1}(h(g(x))\in A)\overset{(*)}{=}P_{\theta_1}(h(X)\in A)$$

یعنی توزیع روی هر مدار Θ ثابت است. توجه داشته باشید که تساوی (*) از پایائی h حاصل می‌گردد. ■

نتیجه۱-۲: اگر $\bar{G}$ تقارنی باشد آنگاه Θ تنها یک مدار دارد و در نتیجه توزیع هر آماره پایای h روی آن به θ بستگی ندارد، یعنی h یک آماره فرعی[1] است.

قضیه زیر حالت کلی تر قضیه ۱-۵ را بیان می کند.

قضیه ۱-٦: فرض کنید یک مسئله تحت تبدیلات G پایا باشد و $\upsilon(\theta)$ یک تابع پایای ماکزیمال روی Θ نسبت به $\bar{G}$ باشد. اگر $h(X)$ یک تابع پایا تحت G باشد آنگاه توزیع $h(X)$ تنها به $\upsilon(\theta)$ بستگی دارد.

اثبات: برای هر $A\subset\mathcal{X}$ داریم که

[1] Ancillary

$$P_{\bar{g}(\theta)}(h(X)\in A)\overset{(5)}{=}P_{\theta}(h(g(X))\in A)\overset{(*)}{=}P_{\theta}(h(X)\in A)$$

توجه داشته باشید که تساوی (*) از پایائی h حاصل می‌گردد. یعنی $S(\theta)=P_{\theta}(h(X)\in A)$ یک تابع پایا روی Θ است $(S(\bar{g}(\theta))=S(\theta))$ بنابراین طبق قضیه ۱-٤ $S(\theta)$ تابعی از پایای ماکزیمال $\upsilon(\theta)$ است. چون این مطلب برای هر A برقرار است پس توزیع $h(X)$ تنها به $\upsilon(\theta)$ بستگی دارد. ■

توسط قضیه زیر برآوردگرهای MRE را بدست می‌آوریم.

قضیه ۱-۷: فرض کنید مسئله مورد نظر تحت تبدیلات G پایا باشد(تعاریف۱-۲و ۱-۳) و $\bar{G}$ یک گروه انتقالی روی Θ باشد و $Y=T(X)$ یک تابع پایای ماکزیمال روی $\mathcal{X}$ باشد در این صورت هر برآوردگر هم‌پایای $\delta(X)$ با مخاطره متناهی که امید شرطی $E_{\bar{e}}[L(\bar{e},\delta(X)|Y=y]$ را مینیمم کند برآوردگر MRE پارامتر θ است.

اثبات: برآوردگر MRE از مینیمم کردن $R(\theta,\delta)=E_{\theta}[L(\theta,\delta(X)]$ بدست می آید. اما طبق نتیجه ۱-۱ تابع مخاطره هر برآوردگر هم پایا نسبت به θ ثابت است(چون $\bar{G}$ انتقالی است). بنابراین کافی است این مخاطره را در نقطه $\theta=\bar{e}$ مینیمم کنیم. بنابراین کافی است مخاطره زیر را نسبت به δ مینیمم کنیم

$$R(\bar{e},\delta)=E_{\bar{e}}[L(\bar{e},\delta(\underset{\sim}{X}))]=E_{\bar{e}}\{E_{\bar{e}}[L(\bar{e},\delta(\underset{\sim}{X}))|Y]\}$$
$$=\int E_{\bar{e}}[L(\bar{e},\delta(\underset{\sim}{X}))|Y=y]f_Y(y)d(\mu_y)$$

این انتگرال موقعی مینیمم می شود که $E_{\bar{e}}[L(\bar{e},\delta(\underset{\sim}{X}))|Y=y]$ مینیمم شود و در نتیجه قضیه ثابت است. ■

توجه: اگر در مسئله‌ای آماره بسنده $S(X)$ وجود داشته باشد که پایا باشد یعنی $S(g(X))=\tilde{g}(S(X))$ آنگاه می‌توان برآوردگر هم‌پایای $\delta(X)$ را همین آماره بسنده $S(X)$ در نظر گرفت که با توجه به قضیه باسو و قضیه ۱-۵ این آماره هم‌پایا از تابع پایای ماکزیمال $Y=T(X)$ مستقل است.

بخش۲ : برآوردگر MRE پارامتر مکان[۱]

فرض کنید $\underset{\sim}{X} = (X_1, X_2, ..., X_n)$ دارای توزیع احتمال توام

$$f(\underset{\sim}{x} - \theta) = f(x_1 - \theta, ..., x_n - \theta)$$

باشد که در آن f معلوم و θ پارامتر مکان نامعلوم است. اگر گروه تبدیلات زیر را در نظر بگیریم.

$$G = \{g_a : g_a(x_1, ..., x_n) = (x_1 + a, ..., x_n + a) = \underset{\sim}{x} + a, a \in R\}$$

آنگاه خانواده توزیع‌های فوق نسبت به این تبدیل پایا است، هرگاه گروه تبدیلات $\bar{G}$ بصورت زیر باشد

$$\bar{G} = \{\bar{g}_a : \bar{g}_a(\theta) = \theta + a, a \in R\}$$

که یک گروه انتقالی از تبدیلات است. در برآورد پارامتر $h(\theta) = \theta$ تحت تبدیل G، $h(\theta)$ دارای برآوردگر هم‌پایا است اگر

$$h(\bar{g}(\theta)) = \tilde{g}(h(\theta)) \Leftrightarrow \theta + a = \tilde{g}(\theta)$$

بنابراین گروه تبدیلات $\tilde{G}$ بصورت زیر است $\tilde{G} = \{\tilde{g}_a : \tilde{g}_a(\delta) = \delta + a, a \in R\}$

در نتیجه تحت گروه G تابع زیان $L(\theta, \delta)$ پایا است اگر و فقط اگر

$$L(\theta, \delta) = L(\bar{g}(\theta), \tilde{g}(\delta)) \Leftrightarrow L(\theta, \delta) = L(\theta + a, \delta + a) \overset{?}{\Leftrightarrow} L(\theta, \delta) = \rho(\delta - \theta)$$

(اثبات رابطه سوم:

$$\Rightarrow)\ a = -\theta \Rightarrow L(\theta, \delta) = L(\cdot, \delta - \theta) = \rho(\delta - \theta)$$

$$(\Leftarrow)\ L(\theta + a, \delta + a) = \rho(\delta + a - \theta - a) = \rho(\delta - \theta) = L(\theta, \delta)$$

بنابراین اگر تابع زیان بفرم $L(\theta, \delta) = \rho(\delta - \theta)$ باشد آنگاه مسئله، تحت گروه تبدیلات G پایا است . برای مثال $L(\theta, \delta) = |\delta - \theta|$ یا $L(\theta, \delta) = (\delta - \theta)^2$ از این نوع تابع زیان می باشند. همچنین برآوردگرهای هم‌پایا در این حالت بصورت زیر می‌باشند:

$$\delta(g(\underset{\sim}{X})) = \tilde{g}(\delta(\underset{\sim}{X})) \Leftrightarrow \delta(\underset{\sim}{X} + a) = \delta(\underset{\sim}{X}) + a$$

از جمله برآوردگرهای هم پایا در این حالت $\bar{X}$ و $\frac{X_{(1)} + X_{(n)}}{2}$ می‌باشند.

برای بدست آوردن برآوردگرهای MRE در این حالت از قضیه زیر استفاده می‌کنیم.

---

[۱] MRE Estimator of Location Parameter

قضیه ۲-۱: اگر $\delta_.$ هر برآوردگر هم پایای θ باشد آنگاه شرط لازم و کافی برای آنکه برآوردگر δ برای θ هم پایا باشد آن است که

$$\delta(\underset{\sim}{x}) = \delta_.(\underset{\sim}{x}) + U(\underset{\sim}{x}) \qquad (۱)$$

که در آن $U(\underset{\sim}{x})$ هر تابع پایا نسبت به G است یعنی:

$$U(\underset{\sim}{x} + a) = U(\underset{\sim}{x}) \qquad (۲)$$

اثبات: فرض کنید روابط (۱) و (۲) برقرار باشند در این صورت:

$$\delta(\underset{\sim}{x} + a) = \delta_.(\underset{\sim}{x} + a) + U(\underset{\sim}{x} + a) = \{\delta_.(\underset{\sim}{x}) + a\} + U(\underset{\sim}{x}) = \delta(\underset{\sim}{x} + a)$$

پس $\delta(\underset{\sim}{x})$ یک براوردگر هم پایا است. حال فرض کنید $\delta(\underset{\sim}{x})$ هم پایا باشد و قرار می‌دهیم $U(\underset{\sim}{x}) = \delta(\underset{\sim}{x}) - \delta_.(\underset{\sim}{x})$ در این صورت رابطه (۱) برقرار است و همچنین

$$U(\underset{\sim}{x} + a) = \delta(\underset{\sim}{x} + a) - \delta_.(\underset{\sim}{x} + a) = \delta(\underset{\sim}{x}) + a - \delta_.(\underset{\sim}{x}) - a = \delta(\underset{\sim}{x}) - \delta_.(\underset{\sim}{x}) = U(\underset{\sim}{x})$$

پس رابطه (۲) برقرار است. ■

اگر قرار دهیم $\underset{\sim}{Y} = (Y_۱, \ldots, Y_{n-۱})$ که $Y_i = X_i - X_n \quad , i = ۱, \ldots, n-۱$ و $Y = c$ اگر $n = ۱$ باشد، آنگاه $\underset{\sim}{Y}$ یک تابع پایای ماکزیمال نسبت به G است و طبق قضیه ۱-٤ تابع $U(\underset{\sim}{x})$ پایا است اگر و فقط اگر تابعی از تابع پایای ماکزیمال $\underset{\sim}{Y}$ باشد بنابراین از قضیه ۲-۱ نتیجه می‌شود که

نتیجه ۱: اگر $\delta_.$ هر برآوردگر هم‌پایای θ باشد آنگاه شرط لازم و کافی برای اینکه برآوردگر δ برای θ هم‌پایا باشد این است که تابع v از $n-۱$ متغیر $y_i = x_i - x_n$، $i = ۱, \ldots, n-۱$ وجود داشته‌باشد بطوری که $\delta(\underset{\sim}{x}) = \delta_.(\underset{\sim}{x}) - v(\underset{\sim}{y})$. ■

حال چون گروه $\bar{G}$ انتقالی است و $\underset{\sim}{Y}$ یک تابع پایای ماکزیمال است در این صورت طبق ۱-۷ برآوردگر هم پایای $\delta(\underset{\sim}{X}) = \delta_.(\underset{\sim}{X}) - v(\underset{\sim}{Y})$ که امید شرطی $E_{\theta=.}[\rho(\delta_.(\underset{\sim}{X}) - v(y)) \mid \underset{\sim}{y}]$ را مینیمم کند برآوردگر MRE پارامتر θ است. حال اگر $v^*(\underset{\sim}{y})$ این امید شرطی را مینیمم کند آنگاه $\delta^*(\underset{\sim}{x}) = \delta_.(\underset{\sim}{x}) - v^*(\underset{\sim}{y})$ برآوردگر MRE پارامتر θ است.

نتیجه ۲-۲: در خانواده توزیع‌های مکان و تحت تبدیلات مکان G داریم که

الف- اگر $\rho(\delta-\theta)=(\delta-\theta)^{۲}$ باشد آنگاه $v^{*}(\underset{\sim}{y})=E_{\theta=.}[\delta_{.}(\underset{\sim}{x})\,|\,\underset{\sim}{y}]$ و در نتیجه برآوردگر MRE پارامتر θ عبارت است از

$$\delta^{*}(\underset{\sim}{x})=\delta_{.}(\underset{\sim}{x})-E_{\theta=.}[\delta_{.}(\underset{\sim}{x})\,|\,\underset{\sim}{y}]$$

ب) اگر $\rho(\delta-\theta)=|\delta-\theta|$ باشد آنگاه $v^{*}(\underset{\sim}{y})=med_{\theta=.}[\delta_{.}(\underset{\sim}{x})\,|\,\underset{\sim}{y}]$(میانه توزیع $\delta_{.}(\underset{\sim}{x})\,|\,\underset{\sim}{y}$به ازای $\theta=۰$) و در نتیجه برآوردگر MRE پارامتر θ عبارت است از

$$\delta^{*}(\underset{\sim}{x})=\delta_{.}(\underset{\sim}{x})-med_{\theta=.}[\delta_{.}(\underset{\sim}{x})\,|\,\underset{\sim}{y}]$$

■

مثال ۲-۱: (حالت $n=۱$) اگر X دارای توزیع احتمال $f(x-\theta)$ باشد در این صورت $v(\underset{\sim}{y})$ یک مقدار ثابت است و طبق مطالب فوق $\delta^{*}(x)=x-v^{*}$ برآوردگر MRE پارامتر θ است که v^{*} مقداری است که $E_{.}[\rho(x-v)]$ را مینیمم می‌کند و تحت تابع زیان درجه دوم $\delta^{*}(x)=X-E_{.}(x)$ برآوردگر MRE پارامتر θ است. ■

مثال ۲-۲: فرض کنید $X_{۱},X_{۲},...,X_{n}$ یک نمونه تصادفی از $N(\mu,\sigma^{۲})$ باشند که $\sigma^{۲}$ معلوم است. برآوردگر MRE پارامتر μرا تحت تابع زیان درجه دوم بدست آورید.

حل:
$$f_{X}(x)=\frac{۱}{\sqrt{۲\pi\sigma^{۲}}}e^{-\frac{۱}{۲\sigma^{۲}}(x-\mu)^{۲}}=f(x-\mu)$$

قرار می دهیم $\delta_{.}(\underset{\sim}{X})=\overline{X}$ پس $\delta_{.}(\underset{\sim}{X})$ آماره بسنده کامل و برآوردگر هم‌پایای مکان μ است $(\delta_{.}(\underset{\sim}{X}+a)=\delta_{.}(\underset{\sim}{X})+a)$ و چون تابع پایای ماکزیمال $\underset{\sim}{Y}$ یک آماره فرعی است پس طبق قضیه باسو $\delta_{.}(\underset{\sim}{X})$ و $\underset{\sim}{Y}$ از یکدیگر مستقل هستند.

بنابراین تحت تابع‌زیان درجه‌دوم داریم که

$$\delta^{*}(\underset{\sim}{X})=\delta_{.}(\underset{\sim}{X})-E_{.}[\delta_{.}(\underset{\sim}{X})\,|\,\underset{\sim}{y}]=\overline{X}-E_{.}[\overline{X}]=\overline{X}$$

پس $\overline{X}$ برآوردگر MRE پارامتر μ است. ■

قضیه ۲-۲: فرض کنید F کلاس کلیه توزیع های یک متغیره F با تابع چگالی f و واریانس ثابت متناهی مثلا $\sigma^{۲}=۱$ باشند. همچنین فرض کنید $X_{۱},X_{۲},...,X_{n}$ یک نمونه تصادفی هر یک با

چگالی $f(x_i-\theta)$ با $\theta=E(X_i)$ باشد و $r_n(F)$مخاطره برآوردگر MRE پارامتر θ تحت تابع زیان درجه دوم باشد. در این صورت $r_n(F)$ موقعی مقدار ماکزیمم خود را روی F بدست می‌آورد که F توزیع نرمال باشد.

اثبات: در حالت توزیع نرمال برآوردگر MRE پارامتر θ برابر $\bar{X}$ با مخاطره $r_n(F)=\frac{1}{n}$ است . چون مخاطره $\bar{X}$ در هر توزیعی برابر $\frac{1}{n}$ است بنابراین مخاطره برآوردگر MRE در هر توزیع F بایستی کمتر یا مساوی $\frac{1}{n}$ باشد(MRE مخاطره را مینمم می‌کند) بنابراین توزیع نرمال دارای ماکزیمم مقدار مخاطره برآوردگر MRE در بین کلیه توزیع‌های F است. ■

مثال ۲-۳: فرض کنید $X_1,X_2,...,X_n$ یک نمونه تصادفی از توزیع نمایی دو پارامتری $E(\theta,b)$ باشند که b مقداری معلوم است. برآوردگر MRE پارامتر θ را تحت توابع زیان درجه دوم و قدر مطلق خطا بدست آورید.

حل: $$f_\theta(x)=\frac{1}{b}e^{-\frac{1}{b}(x-\theta)}I_{(\theta,+\infty)}(x)=\frac{1}{b}e^{-\frac{1}{b}(x-\theta)}I_{(0,+\infty)}(x-\theta)=f(x-\theta)$$

قرار می‌دهیم $\delta_0(\underset{\sim}{X})=X_{(1)}$ بنابراین $\delta_0(\underset{\sim}{X})$ آماره بسنده کامل وهمچنین برآوردگر هم‌پایای مکان θ است و طبق قضیه باسو از آماره فرعی $\underset{\sim}{Y}$ (پایای ماکزیمال) مستقل است. بنابراین

الف- تحت تابع زیان درجه دوم

$$v^*(\underset{\sim}{y})=E[X_{(1)}\mid \underset{\sim}{y}]=E_0(X_{(1)})=\int_0^{+\infty}x\left(\frac{n}{b}e^{-\frac{n}{b}x}\right)dx=\frac{b}{n}$$

پس برآوردگر MRE پارامتر θ عبارت است از $\delta^*(\underset{\sim}{X})=X_{(1)}-\frac{b}{n}$ که برآوردگر UMVU نیز هست.

ب- تحت تابع زیان قدر مطلق خطا $v^*(\underset{\sim}{y})=med_0[X_{(1)}\mid \underset{\sim}{y}]=med_0(X_{(1)})$ پس

$$\int_0^{v^*}\frac{n}{b}e^{-\frac{n}{b}x}dx=\frac{1}{2}\Rightarrow -e^{-\frac{n}{b}x}\Big|_0^{v^*}=\frac{1}{2}\Rightarrow 1-e^{-\frac{n}{b}v^*}=\frac{1}{2}\Rightarrow v^*=\frac{b}{n}\ln 2$$

بنابراین برآوردگر MRE پارامتر θ عبارت است از $\delta^*(X) = X_{(1)} - \frac{b}{n}\ln 2$ که این برآوردگر ناااریب نیست. ■

برآوردگر پیتمن[1]

اگر تابع زیان درجه دوم باشد، آنگاه برآوردگر MRE مکان را به فرم راحت‌تری می‌توان بدست آورد.

قضیه ۲-۳: تحت گروه تبدیلات مکانی G و در خانواده توزیع‌های مکان و تحت تابع زیان درجه دوم، برآوردگر MRE مکان عبارت است از:

$$\delta^*(\underset{\sim}{X}) = \delta_0(\underset{\sim}{X}) - E_0[\delta_0(\underset{\sim}{X}) \mid \underset{\sim}{y}] = \frac{\int_{-\infty}^{+\infty} u f(x_1 - u, \ldots, x_n - u)\,du}{\int_{-\infty}^{+\infty} f(x_1 - u, \ldots, x_n - u)\,du}$$

این فرم را بعنوان برآوردگر پیتمن پارامتر θ می‌شناسند.

اثبات: فرض کنید $\delta_0(\underset{\sim}{X}) = X_n$ برآوردگر هم‌پایای θ باشد. بنابراین

$$v^*(\underset{\sim}{y}) = E_0(X_n \mid \underset{\sim}{y})$$

حال تبدیل زیر را در نظر می‌گیریم:

$$\begin{cases} Y_i = X_i - X_n, & i = 1, \ldots, n-1 \\ W = X_n \end{cases}$$

در این صورت $J = 1$ و $f_{\underset{\sim}{Y},W}(y_1, \ldots, y_{n-1}, w) = f(y_1 + w, \ldots, y_{n-1} + w, w)$

و بنابراین

$$f_{X_n \mid \underset{\sim}{Y}}(w \mid \underset{\sim}{y}) = f_{W \mid \underset{\sim}{Y}}(w \mid \underset{\sim}{y}) = \frac{f_{\underset{\sim}{Y},W}(y_1, \ldots, y_{n-1}, w)}{\int f_{\underset{\sim}{Y},W}(y_1, \ldots, y_{n-1}, w)\,dw}$$

$$= \frac{f(y_1 + w, \ldots, y_{n-1} + w, w)\,dw}{\int f(y_1 + w, \ldots, y_{n-1} + w, w)\,dw}$$

$$\Rightarrow v^*(\underset{\sim}{y}) = E_0(X_n \mid \underset{\sim}{Y}) = E_0(W \mid \underset{\sim}{Y}) = \frac{\int w f(y_1 + w, \ldots, y_{n-1} + w, w)\,dw}{\int f(y_1 + w, \ldots, y_{n-1} + w, w)\,dw}$$

[1] Pitman Estimator

$$= \frac{\int w f\ (x_1 - x_n + w\ ,...,x_{n-1} - x_n + w\ ,w\ )dw}{\int f\ (x_1 - x_n + w\ ,...,x_{n-1} - x_n + w\ ,w\ )dw}$$

$$\overset{u = x_n - w}{=} x_n - \frac{\int u f\ (x_1 - u,...,x_{n-1} - u, x_n - u)du}{\int f\ (x_1 - u,...,x_{n-1} - u, x_n - u)du}$$

$$\therefore \delta^*(\underset{\sim}{x}) = x_n - E_.(X_n\ |\underset{\sim}{Y}) = \frac{\int u f\ (x_1 - u,...,x_n - u)du}{\int f\ (x_1 - u,...,x_n - u)du} \qquad \blacksquare$$

مثال۲-٤: فرض کنید $X_1, X_2,..., X_n$ یک نمونه تصادفی از $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$ باشد برآوردگر MRE پارامتر θ را بدست آورید.

حل:

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \begin{cases} \prod_{i=1}^{n} I_{(\theta - \frac{1}{2}, \theta + \frac{1}{2})}(x_i) = \prod_{i=1}^{n} I_{(-\frac{1}{2}, \frac{1}{2})}(x_i - \theta) = f\ (x_1 - \theta,....,x_n - \theta) \\ I_{(x_{(n)} - \frac{1}{2}, x_{(1)} + \frac{1}{2})}(\theta) \end{cases}$$

$$\therefore \delta^*(\underset{\sim}{x}) = \frac{\int u f\ (\underset{\sim}{x} - u)du}{\int f\ (\underset{\sim}{x} - u)du} = \frac{\int u I_{(x_{(n)} - \frac{1}{2}, x_{(1)} + \frac{1}{2})} du}{\int I_{(x_{(n)} - \frac{1}{2}, x_{(1)} + \frac{1}{2})} du} = ... = \frac{x_{(1)} + x_{(n)}}{2}$$

پس $\delta^*(\underset{\sim}{X}) = \frac{X_{(1)} + X_{(n)}}{2}$ برآوردگر MRE برای θ است ولی UMVUE نیست. $\blacksquare$

مثال ۲-۵: فرض کنید $X_1, X_2,..., X_n$ یک نمونه تصادفی از $E(\theta, b)$ باشند که b معلوم است برآوردگر MRE پارامتر θ را بدست آورید.

حل:

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \begin{cases} \prod_{i=1}^{n} \frac{1}{b} e^{-(x_i - \theta)/b} I_{(\cdot, +\infty)}(x_i - \theta) = f\ (x_1 - \theta,....,x_n - \theta) \\ \frac{1}{b^n} e^{(-n\bar{x} + n\theta)/b} I_{(-\infty, x_{(1)})}(\theta) \end{cases}$$

$$\therefore \delta^*(\underset{\sim}{x}) = \frac{\int uf\ (\underset{\sim}{x}-u)du}{\int f\ (\underset{\sim}{x}-u)du} = \frac{\int_{-\infty}^{x_{(1)}} \frac{u}{b^n} e^{(-n\bar{x}+nu)/b} du}{\int_{-\infty}^{x_{(1)}} \frac{1}{b^n} e^{(-n\bar{x}+nu)/b} du} = \frac{\frac{b}{n}(u-\frac{b}{n})e^{\frac{nu}{b}}\Big|_{-\infty}^{x_{(1)}}}{\frac{b}{n}e^{\frac{-nu}{b}}\Big|_{-\infty}^{x_{(1)}}}$$

$$= x_{(1)} - \frac{b}{n}$$

که همان برآوردگر بدست آمده در مثال۲-۳- الف می‌باشد. ■

قضیه ۲-٤: برآوردگر پیتمن پارامتر مکان تابعی از آماره بسنده است.

اثبات: فرض کنید $\underset{\sim}{S} = (S_1,....,S_k)$ یک آماره بسنده توام برای θ باشد. بنابراین طبق قضیه تجزیه به عوامل داریم که $f(x_1-\theta,...,x_n-\theta) = g(\underset{\sim}{S},\theta)h(x_1,...,x_n)$ بنابراین برآوردگر پیتمن پارامتر θ عبارت است از

■ $$\delta^*(\underset{\sim}{x}) = \frac{\int uf\ (\underset{\sim}{x}-u)du}{\int f\ (\underset{\sim}{x}-u)du} = \frac{\int ug(\underset{\sim}{S},u)h(\underset{\sim}{x})du}{\int g(\underset{\sim}{S},u)h(\underset{\sim}{x})du} = \frac{\int ug(\underset{\sim}{S},u)du}{\int g(\underset{\sim}{S},u)du} = k(\underset{\sim}{S})$$

با توجه به مثال‌های بالا دیده می‌شود که برای برآورد پارامتر مکان ارتباطی بین برآوردگرهای MRE و نااریب و UMVUE وجود دارد. این ارتباط در قضیه زیر آورده شده است:

قضیه ۲-۵: فرض کنید تابع زیان درجه دوم باشد در این صورت برای برآورد پارامتر مکان

الف- اگر $\delta(\underset{\sim}{X})$ هر برآورد هم‌پایا با اریب ثابت b باشد آنگاه $\delta(\underset{\sim}{X})-b$ یک برآوردگر هم پایا و نااریب می‌باشد و دارای مخاطره کمتری نسبت به $\delta(\underset{\sim}{X})$ است.

ب- برآوردگر یکتای MRE یک برآوردگر نااریب است.

ج- اگر برآوردگر UMVU موجود باشد و هم‌پایا باشد آنگاه یک برآوردگر MRE است.

اثبات: الف)هم‌پایا بودن و نااریب بودن $\delta(\underset{\sim}{X})-b$ واضح است.

$$R(\theta,\delta-b) = E_\theta[(\delta(\underset{\sim}{X})-b-\theta)^2]$$
$$= E_\theta\left[\left(\delta(\underset{\sim}{X})-\theta\right)^2\right] + b^2 - 2bE_\theta\left[\delta(\underset{\sim}{X})-\theta\right]$$
$$= R(\theta,\delta) - b^2 < R(\theta,\delta)$$

ب- فرض کنید $\delta(\underset{\sim}{X})$ یک برآوردگر یکتای MRE باشد ولی ناریب نباشد و دارای اریب b باشد در این صورت طبق قسمت (الف) $\delta(\underset{\sim}{X})-b$ یک برآوردگر هم‌پایا با مخاطره کمتر نسبت به $\delta(\underset{\sim}{X})$ است که این با یکتا بودن برآوردگر MRE ، $\delta(\underset{\sim}{X})$ متناقض است. پس $\delta(\underset{\sim}{X})$ ناریب می‌باشد.

ج- اگر $\delta(\underset{\sim}{X})$ یک برآوردگر هم‌پایا و UMVU باشد آنگاه این برآوردگر مقدار مخاطره را مینیمم می‌کند. پس یک برآوردگر MRE است. ■

توجه: در مثال‌های فوق دیدیم که اگر تابع زیان درجه دوم نباشد آنگاه ممکن است برآوردگر MRE یک برآوردگر ناریب نباشد. مفهوم معمول ناریبی این است که

$$E_\theta[\delta(\underset{\sim}{X})]=\theta \quad \forall\theta\in\Theta$$

و اگر تابع زیان درجه دوم باشد آنگاه $R(\theta,\delta)=E_\theta[(\delta(\underset{\sim}{X})-\theta)^2]$ موقعی به ازای هر $\theta\in\Theta$ مینیمم مقدار خود را بدست می‌آورد که $\theta=E_\theta[\delta(\underset{\sim}{X})]$ باشد بنابراین

$$E_\theta[(\delta(\underset{\sim}{X})-\theta)^2]\le E_\theta[(\delta(\underset{\sim}{X})-\theta')] \quad \forall\theta'\neq\theta$$

یعنی متوسط فاصله $\delta(\underset{\sim}{X})$ با مقدار واقعی پارامتر، یعنی $E_\theta[(\delta(\underset{\sim}{X})-\theta)^2]$ کمتر است از متوسط فاصله $\delta(\underset{\sim}{X})$ از هر مقدار دیگر پارامتر. این مطلب راه گشای تعریفی برای ناریبی تحت توابع زیان دیگر است. ■

تعریف۲-۱: (**مخاطره ناریب**[1]): فرض کنید در یک مسئله برآوردیابی تابع زیان $L(\theta,\delta(\underset{\sim}{X}))$ باشد. برآوردگر $\delta(\underset{\sim}{X})$ را مخاطره ناریب گویند هرگاه

$$E_\theta[L(\theta,\delta(\underset{\sim}{X}))]\le E_\theta[L(\theta',\delta(\underset{\sim}{X}))] \quad \forall\theta'\neq\theta \qquad ■$$

مثال۲-٦: (ناریبی میانه): اگر تابع زیان قدرمطلق خطا باشد آنگاه

$$g(\theta)=E_\theta[L(\theta,\delta(\underset{\sim}{X}))]=E[|\delta(\underset{\sim}{X})-\theta|]=\int_{-\infty}^{+\infty}|\delta-\theta|f(\delta)d\delta$$

$$=\int_{-\infty}^{\theta}(\theta-\delta)f(\delta)d\delta+\int_{\theta}^{+\infty}(\delta-\theta)f(\delta)d\delta$$

$$g'(\theta)=0 \quad \Rightarrow \quad \int_{-\infty}^{\theta}f(\delta)d\delta=\int_{\theta}^{+\infty}f(\delta)d\delta\Rightarrow\theta=med_\theta(\delta(\underset{\sim}{X})) \qquad ■$$

[1] Risk-unbiased

با توجه به قضیه ۲-۵-ب می‌خواهیم نشان دهیم که برآوردگر MRE یک برآوردگر مخاطره نااریب است. برای این منظور از لم زیر استفاده می‌کنیم.

لم ۲-۱: اگر $\tilde{G}$ یک گروه جابجایی (آبلی) باشد و δ یک برآوردگر هم پایا تحت گروه G باشد آنگاه $\tilde{G}(\delta)$ برای هر $\tilde{g} \in \tilde{G}$ یک برآوردگر هم‌پایا است.

اثبات: برای هر $g_* \in G$ داریم که

$$\tilde{g}\left(\delta(g_*(\underset{\sim}{X}))\right) \overset{(1)}{=} \tilde{g}\left(\tilde{g}_*(\delta(\underset{\sim}{X}))\right) \overset{(2)}{=} \tilde{g}_*\left(\tilde{g}(\delta(\underset{\sim}{X}))\right)$$

که در آن (۱) از هم‌پایا بودن δ و (۲) از آبلی بودن $\tilde{G}$ بدست می‌آید. یعنی $\tilde{g}(\delta)$ یک برآوردگر هم‌پایا است. ■

قضیه ۲-۶: اگر $\bar{G}$ یک گروه انتقالی و $\tilde{G}$ یک گروه جابجایی باشد آنگاه برآوردگر MRE یک برآوردگر مخاطره نااریب است.

اثبات: فرض کنید δ برآوردگر MRE باشد و $\theta', \theta \in \bar{G}$. چون $\bar{G}$ انتقالی است پس $\bar{g} \in \bar{G}$ موجود است بطوری که $\theta = \bar{g}(\theta')$ و بنابراین

$$E_\theta[L\left(\theta', \delta(\underset{\sim}{X})\right)] = E_\theta[L\left(\bar{g}^{-1}(\theta), \delta(\underset{\sim}{X})\right)] \overset{(*)}{=} E[L\left(\theta, \tilde{g}(\delta(\underset{\sim}{X}))\right)] \overset{(**)}{\geq} E_\theta[L(\theta, \delta(\underset{\sim}{X})]$$

که در آن تساوی (*) از پایایی تابع زیان و نامساوی (**) از لم ۲-۱ و MRE بودن δ حاصل می‌گردند. پس δ برآوردگر مخاطره نااریب است. ■

نتیجه ۲-۳: تحت گروه تبدیلات مکان G و تحت تابع زیان درجه دوم، برآوردگر MRE پارامتر θ مخاطره-نااریب(نااریب – میانگین) است. ■

بخش ۳: برآوردگر MRE پارامتر مقیاس[۱]

فرض کنید $\underset{\sim}{X} = (X_1, \ldots, X_n)$ دارای توزیع احتمال توام

$$\frac{1}{\sigma^n} f\left(\frac{\underset{\sim}{x}}{\sigma}\right) = \frac{1}{\sigma^n} f\left(\frac{x_1}{\sigma}, \ldots, \frac{x_n}{\sigma}\right) \qquad \sigma > 0$$

[۱] MRE Estimator of scale-parameter

باشد که در آن f معلوم و σ پارامتر مقیاس نامعلوم است. اگر گروه تبدیلات زیر را در نظر بگیریم

$$G = \{g_c : g_c(x_1, x_2, \ldots, x_n) = (cx_1, \ldots, cx_n) = c\underset{\sim}{x}\ , c > 0\}$$

آنگاه خانواده توزیع‌های فوق نسبت به این تبدیل پایا است هرگاه گروه تبدیلات $\bar{G}$ بصورت زیر باشد

$$\bar{G} = \{\bar{g}_c : \bar{g}_c(\sigma) = c\sigma,\ c > 0\}$$

که یک گروه انتقالی از تبدیلات است. در برآورد پارامتر $h(\sigma) = \sigma^r$ تحت تبدیل G، $h(\sigma)$ دارای برآوردگر هم‌پایا است اگر

$$h(\bar{g}(\sigma)) = \tilde{g}(h(\sigma)) \Leftrightarrow (c\sigma)^r = \tilde{g}(\sigma^r)$$

بنابراین گروه تبدیلات $\tilde{G}$ بصورت زیر است $\tilde{G} = \{\tilde{g}_c : \tilde{g}_c(\delta) = c^r\sigma,\ c > 0\}$

بنابراین تحت گروه G تابع زیان $L(\sigma, \delta)$ پایا است اگر و فقط اگر

$$L(\bar{g}(\sigma), \tilde{g}(\delta)) = L(\sigma, \delta) \Leftrightarrow L(c\sigma, c^r\delta) = L(\sigma, \delta) \overset{(*)}{\Leftrightarrow} L(\sigma, \delta) = \gamma\left(\frac{\delta}{\sigma^r}\right)$$

اثبات(*):

$$\Rightarrow)\ \ c = \frac{1}{\sigma} \Rightarrow L(\sigma, \delta) = L(1, \frac{\delta}{\sigma^r}) = \gamma(\frac{\delta}{\sigma^r})$$

$$\Leftarrow)\ \ L(c\sigma, c^r\delta) = \gamma(\frac{c^r\delta}{(c\sigma)^r}) = \gamma(\frac{\delta}{\sigma^r}) = L(\sigma, \delta)$$

بنابراین اگر تابع زیان بفرم $L(\sigma, \delta) = \gamma(\frac{\delta}{\sigma^r})$ باشد آنگاه مسئله ما تحت گروه تبدیلات G پایا است. برای مثال $L(\sigma, \delta) = (1 - \frac{\delta}{\sigma^r})^2$ و $L(\sigma, \delta) = |1 - \frac{\delta}{\sigma^r}|$ از این نوع تابع زیان هستند.

همچنین برآوردگرهای هم‌پایا در این حالت بصورت زیر می‌باشند

$$\delta\left(g(\underset{\sim}{X})\right) = \tilde{g}\left(\delta(\underset{\sim}{X})\right) \Leftrightarrow \delta(c\underset{\sim}{X}) = c^r\delta(\underset{\sim}{X})$$

از جمله برآوردگرهای هم‌پایای پارامتر $\sigma(r = 1)$ عبارتند از

$$D = \frac{1}{n}\sum_{i=1}^{n}|X_i - \bar{X}| \qquad S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

برای بدست آوردن برآوردگر MRE در این حالت از قضیه زیر استفاده می‌کنیم.

قضیه ۳-۱: اگر $\delta_{\cdot}(\underset{\sim}{X})$ هر برآوردگر هم‌پایای σ^r باشد آنگاه شرط لازم و کافی برای آنکه براوردگر $\delta(\underset{\sim}{X})$ برای σ^r هم‌پایا باشد آن است که

$$\delta(\underset{\sim}{x}) = \frac{\delta_{\cdot}(\underset{\sim}{x})}{U(\underset{\sim}{x})} \qquad (۱)$$

که در آن $U(\underset{\sim}{x})$ هر تابع پایا نسبت به G است یعنی

$$U(c\underset{\sim}{x}) = U(\underset{\sim}{x}) \qquad (۲)$$

اثبات: فرض کنید(۱) و (۲) برقرار باشند در این صورت

$$\delta(c\underset{\sim}{x}) = \frac{\delta_{\cdot}(c\underset{\sim}{x})}{U(c\underset{\sim}{x})} = \frac{c^r \delta_{\cdot}(\underset{\sim}{x})}{U(\underset{\sim}{x})} = c^r \delta(\underset{\sim}{x})$$

پس $\delta(\underset{\sim}{x})$ یک برآوردگر هم‌پایا است. حال فرض کنید $\delta(\underset{\sim}{x})$ برآوردگری هم‌پایا باشد قرار می‌دهیم. $U(\underset{\sim}{x}) = \frac{\delta_{\cdot}(\underset{\sim}{x})}{\delta(\underset{\sim}{x})}$ در این صورت رابطه (۱) برقرار است و همچنین

$$U(c\underset{\sim}{x}) = \frac{\delta_{\cdot}(c\underset{\sim}{x})}{\delta(c\underset{\sim}{x})} = \frac{c^r \delta_{\cdot}(\underset{\sim}{x})}{c^r \delta(\underset{\sim}{x})} = \frac{\delta_{\cdot}(\underset{\sim}{x})}{\delta(\underset{\sim}{x})} = U(\underset{\sim}{x})$$

■

اگر قرار دهیم $\underset{\sim}{Z} = (Z_1,...,Z_n)$ که $Z_i = \frac{X_i}{X_n} i = 1,...,n-1, Z_n = \frac{X_n}{|X_n|}$ (و $Z = c$ اگر $n = 1$ باشد) آنگاه $\underset{\sim}{Z}$ یک تابع پایای ماکزیمال نسبت به G است و طبق قضیه ۱-٤ تابع $U(\underset{\sim}{x})$ پایا است اگر و فقط اگر تابعی از تابع پایای ماکزیمال $\underset{\sim}{Z}$ باشد. بنابراین از قضیه ۳-۱ نتیجه می شود که:

نتیجه۳-۱: اگر $\delta_{\cdot}$ هر برآوردگر هم‌پایای σ^r باشد آنگاه شرط لازم و کافی برای این که برآوردگر δ برای σ^r هم‌پایا باشد این است که تابع w از n متغیر

$$Z_i = \frac{X_i}{X_n} i = 1,...,n-1, \quad Z_n = \frac{X_n}{|X_n|}$$

وجود داشته باشد بطوری که $\delta(\underset{\sim}{x}) = \frac{\delta_{\cdot}(\underset{\sim}{x})}{w(\underset{\sim}{z})}$ ■

حال چون گروه $\bar{G}$ انتقالی است و Z یک تابع پایای ماکزیمال است در این صورت طبق قضیه ۱-۷ برآوردگر هم پایای $\delta(X)=\frac{\delta_{\cdot}(X)}{w(Z)}$ که امید شرطی $E_{\sigma=1}[\gamma\left(\frac{\delta_{\cdot}(X)}{w(z)}\right)|z]$ را مینیمم کند برآوردگر MRE پارامتر σ^r است. حال اگر $w^*(z)$ این امید شرطی را مینیمم کند آنگاه $\delta(X)=\frac{\delta_{\cdot}(X)}{w^*(z)}$ برآوردگر MRE پارامتر σ^r است.

لم ۳-۱: فرض کنید X یک متغیر تصادفی مثبت باشد

الف- اگر $E(X^2)<\infty$ آنگاه $E\left(\frac{X}{c}-1\right)^2$ مقدار مینیمم خود را به ازای $c=\frac{E(X^2)}{E(X)}$ بدست می‌آورد.

ب- اگر $E(X)<\infty$ و X دارای تابع چگالی f باشد آنگاه $E\left(\left|\frac{X}{c}-1\right|\right)$ مقدار مینیمم خود را به ازای مقدار c بدست می‌آورد که در رابطه زیر صدق کند

$$\int_{\cdot}^{c} xf(x)dx=\int_{c}^{+\infty} xf(x)dx$$

(هر مقداری c که در رابطه فوق صدق می کند را میانه – مقیاس[1] توزیع X می‌نامند)

اثبات: تمرین. ■

نتیجه ۳-۲: در خانواده توزیع‌های مقیاس و تحت تبدیلات مقیاس G داریم که

الف- اگر $\gamma\left(\frac{\delta}{\sigma^r}\right)=\left(\frac{\delta}{\sigma^r}-1\right)^2$ (فرم A) آنگاه برآوردگرMRE پارامتر σ^r عبارت‌است از

$$\delta^*(X)=\frac{\delta_{\cdot}(X)E_1[\delta_{\cdot}(X)|z]}{E_1[\delta_{\cdot}^2(X)|z]}$$

ب- اگر $\gamma\left(\frac{\delta}{\sigma^r}\right)=\left|\frac{\delta}{\sigma^r}-1\right|$ (فرم B) آنگاه $\delta^*(X)=\frac{\delta_{\cdot}(X)}{w^*(z)}$ که در آن $w^*(z)$ هر میانه مقیاس توزیع $\delta_{\cdot}(X)$ به شرط Z با $\sigma=1$ می‌باشد.

اثبات: تمرین. ■

[1] Scale-median

مثال ۳-۱: (حالت $n=1$) اگر X داری توزیع احتمال $\frac{1}{\sigma}f\left(\frac{x}{\sigma}\right)$ باشد در این صورت $w^*(\underset{\sim}{z})$ یک مقدار ثابت است و طبق مطالب فوق $\delta^*(\underset{\sim}{X})=\frac{X^r}{w^*}$ برآوردگر MRE پارامتر σ^r است که w^* مقداری است که $E_1\left[\gamma\left(\frac{X^r}{w}\right)\right]$ را مینیمم می‌کند. بنابراین تحت تابع زیان فرم A، $\delta^*(X)=\frac{X^r E_1(X^r)}{E_1(X^{2r})}$ و تحت تابع زیان بفرم B، $\delta^*(X)=\frac{X^r}{w^*}$ که w^* هر میانه مقیاس توزیع X^r می‌باشد، برآوردگرهای MRE پارامتر σ^r هستند. ■

مثال ۳-۲: فرض کنید $X_1,X_2,\ldots,X_n$ یک نمونه تصادفی از $N(0,\sigma^2)$ باشند برآوردگر MRE پارامتر σ^2 را تحت تابع زیان بفرم A بدست آورید.

حل:
$$f_X(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x}{\sigma})^2}=\frac{1}{\sigma}f\left(\frac{x}{\sigma}\right)$$

قرار می‌دهیم $\delta_0(\underset{\sim}{X})=\sum_{i=1}^{n}X_i^2$ در این صورت $\delta_0(\underset{\sim}{X})$ آماره بسنده کامل برای σ^2 و همچنین برآوردگر هم‌پایای σ^2 است (زیرا $\delta_0(c\underset{\sim}{X})=c^2\delta_0(\underset{\sim}{X})$) و طبق قضیه باسو از آماره فرعی $\underset{\sim}{Z}$ (پایای ماکزیمال) مستقل است. بنابراین

$$\delta^*(\underset{\sim}{X})=\frac{\delta_0(\underset{\sim}{X})E_1[\delta_0(\underset{\sim}{X})]}{E_1[\delta_0^2(\underset{\sim}{X})]}=\frac{\sum X_i^2 E(\chi^2_{(n)})}{E_1[(\chi^2_{(n)})^2]}=\frac{n\sum_{i=1}^{n}X_i^2}{2n+n^2}=\frac{1}{n+2}\sum_{i=1}^{n}X_i$$

نشان دهید که این برآوردگر مخاطره ناریب است. ■

مثال ۳-۳: فرض کنید $X_1,X_2,\ldots,X_n$ یک نمونه تصادفی از $U(0,\theta)$ باشند برآوردگر MRE پارامتر θ را تحت توابع زیان A و B بدست آورید.

حل:
$$f_X(x)=\frac{1}{\theta}I_{(0,\theta)}(x)=\frac{1}{\theta}I_{(0,1)}\left(\frac{x}{\theta}\right)=\frac{1}{\theta}f\left(\frac{x}{\theta}\right)$$

قرار می دهیم $\delta_{\circ}(\underset{\sim}{X}) = X_{(n)}$ بنابراین $\delta_{\circ}(\underset{\sim}{X})$ آماره بسنده کامل برای θ است و همچنین برآوردگر هم‌پایای θ است. بنابراین از آماره فرعی (پایای ماکزیمال) $\underset{\sim}{Z}$ مستقل است. درنتیجه

الف- تحت تابع زیان بفرم A با $r = ۱$،

$$\delta^*(\underset{\sim}{X}) = \frac{X_{(n)}E_{۱}(X_{(n)})}{E_{۱}(X^{۲}_{(n)})} = \frac{X_{(n)}\left(\frac{n}{n+۱}\right)}{\frac{n}{n+۲}} = \frac{n+۲}{n+۱}x_{(n)} \qquad (X_{(n)} \sim Be(n,۱))$$

ب- تحت تابع زیان بفرم B با $r = ۱$، $\delta^*(\underset{\sim}{X}) = \frac{X_{(n)}}{w^*}$ که در آن w^* هر میانه مقیاس توزیع $X_{(n)}$ با $\theta = ۱$ است. یعنی

$$\int_{\circ}^{w^*} x\,(nx^{n-۱})dx = \int_{w^*}^{۱} x\,(nx^{n-۱})dx \Rightarrow \ldots \Rightarrow w^* = \sqrt[n+۱]{\frac{۱}{۲}} \Rightarrow \delta^*(\underset{\sim}{X}) = \sqrt[n+۱]{۲}X_{(n)}$$

■

برآوردگر پیتمن:

اگر تابع زیان بفرم A باشد آنگاه برآوردگر MRE پارامتر مقیاس را بفرم راحت‌تری می‌توان بدست آورد.

قضیه ۳-۲: تحت گروه تبدیلات مقیاسی G و در خانواده توزیع‌های مقیاسی و تحت تابع زیان بفرم A برآوردگر MRE پارامتر مقیاس σ^r عبارت است از

$$\delta^*(\underset{\sim}{x}) = \frac{\delta_{\circ}(\underset{\sim}{x})E_{۱}[\delta_{\circ}(\underset{\sim}{x}) \mid \underset{\sim}{z}]}{E_{۱}[\delta^{۲}_{\circ}(\underset{\sim}{x}) \mid \underset{\sim}{z}]} = \frac{\int_{\circ}^{+\infty} \nu^{n+r-۱} f\,(\nu x_{۱},\ldots,\nu x_{n})d\nu}{\int_{\circ}^{+\infty} \nu^{n+۲r-۱} f\,(\nu x_{۱},\ldots,\nu x_{n})d\nu}$$

این فرم را بعنوان برآورد پیتمن پارامتر σ^r می‌شناسند.

اثبات: تمرین (همانند اثبات قضیه۲-۳ می باشد) $\delta_{\circ}(\underset{\sim}{X}) = X_n^r$. ■

مثال۳-٤: فرض‌کنید $X_{۱}, X_{۲}, \ldots, X_n$ یک نمونه تصادفی از توزیع نمایی با پارامتر λ باشند. برآوردگر پیتمن پارامتر λ را بدست آورید.

حل: $r=۱$

$$f_{\underset{\sim}{X}}(\underset{\sim}{x})=\prod_{i=۱}^{n}\frac{۱}{\lambda}e^{-\frac{x_i}{\lambda}}=\frac{۱}{\lambda^n}e^{-\sum_{i=۱}^{n}\frac{x_i}{\lambda}}=\frac{۱}{\lambda^n}f\left(\frac{x_۱}{\lambda},...,\frac{x_n}{\lambda}\right)=\nu^n f\left(\nu x_۱,...,\nu x_n\right)$$

$$\delta^*(\underset{\sim}{x})=\frac{\int_{۰}^{+\infty}\nu^n f\left(\nu\underset{\sim}{x}\right)d\nu}{\int_{۰}^{+\infty}\nu^{n+۱}f\left(\nu\underset{\sim}{x}\right)d\nu}=\frac{\int_{۰}^{+\infty}\nu^n e^{-\nu\sum x_i}d\nu}{\int_{۰}^{+\infty}\nu^{n+۱}e^{-\nu\sum x_i}d\nu}=\frac{\frac{\Gamma(n+۱)}{(\sum x_i)^{n+۱}}}{\frac{\Gamma(n+۲)}{(\sum x_i)^{n+۲}}}=\frac{\sum x_i}{n+۱}$$

پس برآوردگر MRE پارامتر λ برابر $\delta^*(\underset{\sim}{X})=\frac{\sum X_i}{n+۱}$ است(که مخاطره‌نااریب نیز است). ■

مثال ۳-۵: فرض کنید $X_۱,X_۲,...,X_n$ یک نمونه تصادفی از $U(۰,\theta)$ باشند. برآوردگر پیتمن پارامتر θ را بدست آورید.

حل:

$$f_{\underset{\sim}{X}}(\underset{\sim}{x})=\begin{cases}\prod_{i=۱}^{n}\frac{۱}{\theta}I_{(۰,۱)}(\frac{x_i}{\theta})=\frac{۱}{\theta^n}\prod_{i=۱}^{n}I_{(۰,۱)}(\frac{x_i}{\theta})=\frac{۱}{\theta^n}f\left(\frac{x_۱}{\theta},...,\frac{x_n}{\theta}\right)=\nu^n f\left(\nu\underset{\sim}{x}\right)\\ \frac{۱}{\theta^n}I_{(x_{(n)},+\infty)}(\theta)=\nu^n I_{(۰,\frac{۱}{x_{(n)}})}(\nu) \qquad \nu=\frac{۱}{\theta}\end{cases}$$

$$\delta^*(\underset{\sim}{x})=\frac{\int_{۰}^{+\infty}\nu^n f\left(\nu\underset{\sim}{x}\right)d\nu}{\int_{۰}^{+\infty}\nu^{n+۱}f\left(\nu\underset{\sim}{x}\right)d\nu}=\frac{\int_{۰}^{۱/x_{(n)}}\nu^n d\nu}{\int_{۰}^{۱/x_{(n)}}\nu^{n+۱}d\nu}=\frac{\frac{\nu^{n+۱}}{n+۱}\Big|_{۰}^{۱/x_{(n)}}}{\frac{\nu^{n+۲}}{n+۲}\Big|_{۰}^{۱/x_{(n)}}}=\frac{n+۲}{n+۱}x_{(n)}$$

که همان برآوردگر بدست آمده در مثال ۳-۳ است. ■

قضیه ۳-۳: برآوردگر پیتمن پارامتر مقیاس تابعی از آماره بسنده است.

اثبات : تمرین. ■

بخش٤: برآوردگر MRE پارامترهای مکان-مقیاس

فرض کنید $\underset{\sim}{X} = (X_1,\ldots,X_n)$ دارای توزیع احتمال توام $\frac{1}{\sigma^n} f\left(\frac{x_1-\theta}{\sigma},\ldots,\frac{x_n-\theta}{\sigma}\right)$ باشند که در آن f معلوم و θ و σ هر دو نامعلوم هستند. در زیر برآوردگرهای MER برای σ و θ را به طور جداگانه بدست می‌آوریم.

الف- برآوردگرهای MRE پارامتر مقیاس σ:

فرض کنید بخواهیم پارامتر σ^rرا برآورد کنیم . اگر گروه تبدیلات زیر را در نظر بگیریم.

$$\mathsf{G} = \{g_{a,b} \mid g_{a,b}(\underset{\sim}{x}) = a + b\underset{\sim}{x},\ a \in R, b > 0\}$$

آنگاه خانواده توزیع‌های فوق نسبت به این تبدیل پایا است هرگاه گروه تبدیلات $\bar{G}$ به صورت زیر باشد.

$$\bar{G} = \{\bar{g}_{a,b} \mid \bar{g}_{a,b}(\theta,\sigma) = (a + b\theta, b\sigma),\ a \in R, b > 0\}$$

که یک گروه انتقالی از تبدیلات است. در برآورد پارامتر $h(\theta,\sigma) = \sigma^r$ تحت تبدیل G، $h(\sigma)$ دارای برآوردگر هم‌پایا است اگر

$$h(\bar{g}(\theta,\delta)) = \tilde{g}\left(h(\theta,\sigma)\right) \Leftrightarrow h(a + b\theta, b\sigma) = \tilde{g}(\sigma^r) \Leftrightarrow b^r\sigma^r = \tilde{g}(\sigma^r)$$

بنابراین گروه تبدیلات $\tilde{G}$ به صورت زیر است.

$$\tilde{G} = \{\tilde{g}_{a,b} \mid \tilde{g}_{a,b}(\delta) = b^r\delta,\ b > 0, a \in R\}$$

بنابراین تحت گروه G تابع زیان $L(\theta,\sigma,\delta)$پایا است اگر و فقط اگر

$$L(\theta,\sigma,\delta) = L(a + b\theta, b\sigma, b^r\delta) \overset{a=\frac{-\theta}{\sigma}, b=\frac{1}{\sigma}}{\Leftrightarrow} L(\theta,\sigma,\delta) = \gamma\left(\frac{\delta}{\sigma^r}\right)$$

(برای مثال تابع زیان های بفرم A و B).

در این حالت برآوردگرهای هم‌پایا بفرم زیر می‌باشند.

$$\delta\left(g(\underset{\sim}{X})\right) = \tilde{g}\left(\delta(\underset{\sim}{X})\right) \Leftrightarrow \delta(a + b\underset{\sim}{X}) = b^r\left(\delta(\underset{\sim}{X})\right) \qquad (1)$$

برای بدست آوردن برآوردگر MRE در این حالت به صورت زیر عمل می‌کنیم. اگر تنها یک تغییر در مکان داشته باشیم یعنی:

$$g(\underset{\sim}{x}) = \underset{\sim}{x} + a \qquad \tilde{g}(\theta,\sigma) = (\theta + a, \sigma) \qquad (b = ۱)$$

در این صورت رابطه (۱) با $b = ۱$ نتیجه می‌دهد که بایستی $\delta(a + b\underset{\sim}{X}) = \delta(\underset{\sim}{X})$

یعنی δ یک تابع پایا نسبت به G است پس طبق نتیجه ۲-۱، $\underset{\sim}{\delta}$ تابعی از $n-۱$ تفاضل $Y_i = X_i - X_n \quad , i = ۱,...,n-۱$ می‌باشد. اما تابع چگالی توام Y_iها بصورت زیر بدست می‌آید:

$$W_i = X_i - \theta \qquad i = ۱,...,n \Rightarrow f_{\underset{\sim}{W}}(\underset{\sim}{w}) = f_{\underset{\sim}{X}}(w + \theta) = \frac{۱}{\sigma^n} f\left(\frac{\underset{\sim}{w}}{\sigma}\right)$$

$$\begin{cases} Y_i = X_i - X_n = W_i - W_n & i = ۱,...,n-۱ \\ T = W_n \end{cases} \Rightarrow J = ۱$$

$$f_{\underset{\sim}{Y},T}(\underset{\sim}{y},t) = f_{\underset{\sim}{W}}(y_۱ + t,...,y_{n-۱} + t, t) = \frac{۱}{\sigma^n} f\left(\frac{y_۱ + t}{\sigma},...,\frac{y_{n-۱} + t}{\sigma},\frac{t}{\sigma}\right)$$

$$\Rightarrow f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \frac{۱}{\sigma^n}\int_{-\infty}^{+\infty} f\left(\frac{y_۱ + t}{\sigma},...,\frac{y_{n-۱} + t}{\sigma},\frac{t}{\sigma}\right)dt$$

$$\overset{u=\frac{t}{\sigma}}{=} \frac{۱}{\sigma^{n-۱}}\int_{-\infty}^{+\infty} f\left(\frac{y_۱}{\sigma} + u,...,\frac{y_{n-۱}}{\sigma} + u, u\right)du = \frac{۱}{\sigma^{n-۱}} f^*\left(\frac{\underset{\sim}{y}}{\sigma}\right)$$

بنابراین توزیع $\underset{\sim}{Y} = (Y_۱,...,Y_{n-۱})$ به یک خانواده مقیاس متعلق است در نتیجه در این خانواده برای برآورد پارامتر σ^r از برآوردگرهای MRE بدست آمده در بخش قبل استفاده می‌کنیم. بنابراین برآوردگر MRE پارامتر σ^r بفرم $\delta(\underset{\sim}{X}) = \frac{\delta_.(\underset{\sim}{Y})}{w^*(\underset{\sim}{Z})}$ است که $\delta_.(\underset{\sim}{Y})$ هر برآوردگر پایای مقیاس پارامتر σ^r بر اساس $\underset{\sim}{Y}$ است $\left(\delta_.(b\underset{\sim}{Y}) = b^r\delta_.(\underset{\sim}{Y})\right)$ که در آن $Z_{n-۱} = \frac{Y_{n-۱}}{|Y_{n-۱}|}$ و $\underset{\sim}{Z} = (Z_۱,...,Z_{n-۱}),\ Z_i = \frac{Y_i}{Y_{n-۱}},\ i = ۱,۲,...,n-۲$ و $w^*(\underset{\sim}{z})$ هر عددی است که $E_{\sigma=۱}\left\{\gamma\left(\frac{\delta_.(\underset{\sim}{Y})}{w(\underset{\sim}{z})}\right) | \underset{\sim}{z}\right\}$ را مینیمم می‌کند.

نتیجه ٤-۱: اگر تابع زیان بفرم A $\left(L(\theta,\sigma,\delta)=\left(\frac{\delta}{\sigma^r}-۱\right)^۲\right)$ باشد آنگاه براوردگر MRE پارامتر σ^r عبارت است از

$$\delta^*(\underset{\sim}{X})=\frac{\delta_.(\underset{\sim}{Y})E_{\sigma=۱}[\delta_.(\underset{\sim}{Y})\,|\,\underset{\sim}{z}]}{E_{\sigma=۱}[\delta_.^۲(\underset{\sim}{Y})\,|\,\underset{\sim}{z}]}$$

و اگر تابع زیان بفرم B $\left(L(\theta,\sigma,\delta)=\left(\frac{\delta}{\sigma^r}-۱\right)\right)$ باشد آنگاه برآوردگر MRE پارامتر σ^r عبارت است از $\delta^*(\underset{\sim}{X})=\frac{\delta_.(\underset{\sim}{X})}{w^*(\underset{\sim}{z})}$ که در آن $w^*(\underset{\sim}{z})$ هر میانه مقیاس توزیع $\delta_.(\underset{\sim}{Y})$ به شرط $\underset{\sim}{Z}$ با $\sigma=۱$ می‌باشد. ■

مثال ٤-۱: فرض کنید $X_۱,X_۲,\ldots,X_n$ یک نمونه تصادفی از $N(\mu,\sigma^۲)$باشند که در آن μ و $\sigma^۲$ هر دو نامعلوم هستند. برآورد MRE پارامتر $\sigma^۲$ را تحت تابع زیان‌های بفرم A و B بدست آورید.

حل: $T=(\bar{X},\sum_{i=۱}^{n}(X_i-\bar{X})^۲)$ یک آماره بسنده کامل برای μ و $\sigma^۲$ است. قرار می‌دهیم $\delta_.=\sum_{i=۱}^{n}(X_i-\bar{X})^۲$ در این صورت $\delta_.$ یک برآوردگر هم‌پایا برای $\sigma^۲$ است

$\left(\delta_.(a+b\underset{\sim}{X})=b^۲\delta_.(\underset{\sim}{X}),\ \delta_.(\underset{\sim}{Y})=b^۲\delta_.(\underset{\sim}{Y})\right)$ و از آماره پایای ماکزیمال $\underset{\sim}{Z}$(فرعی) مستقل است و

$$\delta_.\sim\chi^۲_{(n-۱)}\quad(\sigma=۱)$$

الف- اگر تابع زیان بفرم A باشد آنگاه

$$\delta^*(\underset{\sim}{X})=\frac{\delta_.E_۱(\delta_.\,|\,\underset{\sim}{z})}{E_۱(\delta_.^۲\,|\,\underset{\sim}{z})}=\frac{\delta_.E_۱(\delta_.)}{E_۱(\delta_.^۲)}=\frac{\delta_.(n-۱)}{۲(n-۱)+(n-۱)^۲}=\frac{\delta_.}{n+۱}=\frac{۱}{n+۱}\sum_{i=۱}^{n}(X_i-\bar{X})^۲$$

ب- اگر تابع زیان بفرم B باشد آنگاه w^* میانه-مقیاس توزیع $\delta_.\,|\,\underset{\sim}{Z}\overset{D}{=}\delta_.$ است.

$$\int_{0}^{w^*} \frac{x}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)} x^{\frac{n-1}{2}-1} e^{-\frac{1}{2}x} dx = \int_{w^*}^{+\infty} \frac{x}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)} x^{\frac{n-1}{2}-1} e^{-\frac{1}{2}x} dx$$

$$\Leftrightarrow \ldots\ldots \Leftrightarrow \int_{0}^{w^*} \frac{1}{2^{\frac{n+1}{2}}\Gamma\left(\frac{n+1}{2}\right)} x^{\frac{n+1}{2}-1} e^{-\frac{1}{2}x} dx = \int_{w^*}^{+\infty} \frac{x}{2^{\frac{n+1}{2}}\Gamma\left(\frac{n+1}{2}\right)} x^{\frac{n+1}{2}-1} e^{-\frac{1}{2}x} dx$$

یعنی w^* میانه توزیع $\chi^2_{(n+1)}$ است و بنابراین $w^* = \chi^2_{0.5}(n+1)$ و در نتیجه

$$\delta^*(\underset{\sim}{X}) = \frac{1}{\chi^2_{0.5}(n+1)} \sum_{i=1}^{n} (X_i - \bar{X})^2$$ ■

مثال ٤-٢: فرض‌کنید $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از توزیع $U(\theta - \frac{1}{2}\sigma, \theta + \frac{1}{2}\sigma)$ باشد. برآوردگر MRE پارامتر σ را تحت توابع زیان بفرم A و B بدست آورید.

حل:

$$f_{\theta,\sigma}(x) = \frac{1}{\sigma} I_{(\theta-\frac{1}{2},\theta+\frac{1}{2})}(x) = \frac{1}{\sigma} I_{(-\frac{1}{2},\frac{1}{2})}\left(\frac{x-\theta}{\sigma}\right)$$

$T = (X_{(1)}, X_{(n)})$ یک آماره بسنده مینیمال برای (θ, σ) است و $\delta_0 = X_{(n)} - X_{(1)}$ برآوردگر هم‌پایای σ است که از آماره پایای ماکزیمال (فرعی) $\underset{\sim}{Z}$ مستقل است و اگر $\sigma = 1$ باشد.

$$\delta_0 = X_{(n)} - X_{(1)} = (X_{(n)} - \theta + \frac{1}{2}) - (X_{(1)} - \theta + \frac{1}{2}) = V_{(n)} - V_{(1)}$$

$$\overset{D}{=} V_{(n-1)} \sim Be(n-1, 2)$$

الف- اگر تابع زیان بفرم A باشد آنگاه

$$\delta^*(\underset{\sim}{X}) = \frac{\delta_0 E_1[\delta_0]}{E_1[\delta_0^2]} = \frac{\delta_0 (\frac{n-1}{n+1})}{\frac{n(n-1)}{(n+1)(n+2)}} = \frac{n+2}{n}\delta_0 = \frac{n+2}{n}(X_{(n)} - X_{(1)})$$

ب- اگر تابع زیان به فرم Bباشد برآوردگر MRE، σ را بدست آورید. ■

توجه: توجه کنید که اگر گروه تبدیلات $\begin{matrix} G & = \{g_b(\underset{\sim}{x}) = b\underset{\sim}{x}, b > 0\} \\ \bar{G} & = \{\bar{g}_b(\sigma) = b\sigma, b > 0\} \end{matrix}$ را به کار می‌بردیم (همانند بخش قبل) آنگاه $\bar{G}$ دیگر یک گروه انتقالی نبود و از روش بکار برده شده در این فصل نمی‌توان

برآوردگر MRE، σ^r را بدست آورد. (اشتاین[۱](۱۹٦٤) نشان داد که $\delta(\underset{\sim}{x})=\psi\left(\frac{\bar{x}}{s}\right)s^{۲}$ که

$\psi\left(\frac{\bar{x}}{s}\right)=\min\left\{\frac{۱}{n+۱},\frac{۱+\frac{n\bar{x}^{۲}}{s}}{n+۲}\right\}$ تحت تبدیل فوق برای $\sigma^{۲}$ در توزیع نرمال دارای مخاطره کمتری نسبت به $\delta_{.}=\frac{۱}{n+۱}\sum_{i=۱}^{n}(X_i-\bar{X})^{۲}$ در مثال٤-۱ است)

ب- برآوردگر MRE پارامتر مکان θ

فرض کنید بخواهیم پارامتر مکان θرا در خانواده توزیع‌های $\frac{۱}{\sigma^n}f\left(\frac{x_۱-\theta}{\sigma},...,\frac{x_n-\theta}{\sigma}\right)$ برآورد کنیم. همان‌گونه که دیدیم این خانواده تحت گروه‌های تبدیلات

$$G = \{g_{a,b} \mid g_{a,b}(\underset{\sim}{x})=a+b\underset{\sim}{x}\}$$
$$\bar{G} = \{\bar{g}_{a,b} \mid \bar{g}_{a,b}(\theta,\sigma)=(a+b\theta,b\sigma)\}$$

پایا است و در برآورد پارامتر θ گروه تبدیلات $\tilde{G}$ به صورت زیر می‌باشد

$$\tilde{G} = \{\tilde{g}_{a,b}(\delta)=a+b\delta\}$$

در این‌صورت مسئله مورد نظر تحت گروه G پایا است اگر و فقط اگر تابع زیان بفرم زیر باشد

$$L(\theta,\sigma,\delta)=L(a+b\theta,\ b\sigma,\ a+b\delta) \overset{a=\frac{-\theta}{\sigma},b=\frac{۱}{\sigma}}{\Leftrightarrow} L(\theta,\sigma,\delta)=\rho\left(\frac{\delta-\theta}{\sigma}\right)$$

در این حالت برآوردگرهای هم‌پایا به فرم زیر می‌باشند

$$\delta\left(g(\underset{\sim}{X})\right)=\tilde{g}\left(\delta(\underset{\sim}{X})\right)\Leftrightarrow\delta(a+b\underset{\sim}{x})=a+b\delta(\underset{\sim}{x}) \qquad (۲)$$

چون $\bar{G}$ انتقالی است پس مخاطره هر برآوردگر هم‌پایا ثابت است. برای بدست آوردن برآوردگر MRE پارامتر θ دو حالت زیر پیش می‌آید.

حالت اول: برای یک مقدار ثابت σ قرار می‌دهیم

$$g_\sigma(x_۱,\ldots,x_n)=\frac{۱}{\sigma^n}f\ (\frac{x_۱}{\sigma},...,\frac{x_n}{\sigma})$$

[۱] Stein

بنابراین خانواده توزیع‌های مکان مقیاس به صورت زیر تبدیل می‌شود.

$$g_\sigma(x_1-\theta,....,x_n-\theta) \qquad (۳)$$

که به یک خانواده مکان متعلق است.

لم ٤-۱: فرض کنید برای خانواده مکان (۳) و تحت تابع زیان بفرم $\rho\left(\frac{\delta-\theta}{\sigma}\right)$ یک برآوردگر MRE δ^* برای θ تحت تبدیلات مکان($G^*=\{g_a(\underset{\sim}{x})=\underset{\sim}{x}+a\}$)موجود باشد و

(i) δ^* به σ بستگی نداشته باشد

(ii) δ^* در رابطه (۲) صدق کند.

آنگاه δ^* مخاطره را در بین تمام برآوردگرهایی که در (۲) صدق می‌کنند، مینیمم می‌کند.

اثبات: فرض کنید δ هر براوردگر هم‌پایای دیگر که در (۲) صدق می‌کند باشد، بنابراین δ تحت G^* نیز هم‌پایا است. همچنین فرض کنید که σ مقداری معلوم باشد. بنابراین طبق فرض قضیه در مورد δ^*، برای این مقدار σ مخاطره δ^* کمتر یا مساوی δ است و چون این موضوع برای هر σ برقرار است پس δ^* برآوردگر MRE در بین تمام برآوردگرهای هم‌پایای (۲) است. ■

مثال ٤-۳: فرض کنید $X_1,X_2,\ldots,X_n$ یک نمونه تصادفی از $N(\mu,\sigma^2)$ باشند که هر دوی μ و σ نامعلوم هستند. طبق مثال۲-۲، $\delta^*=\bar{X}$ برآوردگر MRE پارامتر μ تحت تابع زیان $\rho\left(\frac{\delta-\theta}{\sigma}\right)=\left(\frac{\delta-\theta}{\sigma}\right)^2$ است. چون شرایط (i) و (ii) لم ٤-۱ برقرار است بنابراین $\delta^*=\bar{X}$ برآوردگر MRE پارامتر μ تحت تبدیلات G می‌باشد. ■

مثال ٤-٤: فرض کنید $X_1,X_2,\ldots,X_n$ یک نمونه تصادفی از $U(\theta-\frac{1}{2}\sigma,\theta+\frac{1}{2}\sigma)$ باشند. در این‌صورت طبق مثال۲-٤، $\delta^*=\frac{X_{(1)}+X_{(n)}}{2}$ برآوردگر MRE پارامتر θ تحت تابع زیان $\rho\left(\frac{\delta-\theta}{\sigma}\right)=\left(\frac{\delta-\theta}{\sigma}\right)^2$ است.چون شرایط (i)و (ii) لم ٤-۱ برقرار است، بنابراین δ^* برآوردگر MER پارامتر θ تحت تبدیلات G می‌باشد. ■

توجه: در مثال ۲-۳ دیدیم که در خانواده توزیع‌های $E(\theta,b)$ با b معلوم $\delta^* = X_{(1)} - \frac{b}{n}$ برآوردگر MRE پارامتر θ است. چون شرط (i) لم ٤-۱ برای این برآوردگر صدق نمی‌کند پس از این روش نمی‌توان برآوردگر MRE پارامتر θ را پیدا کرد.

حالت دوم: اگر شرایط حالت اول برقرار نباشد، برای یافتن برآوردگر MRE پارامتر θ همانند بخش‌های قبل به صورت زیر عمل می‌کنیم.

قضیه ٤-۱: فرض کنید $\delta_.$ هر برآوردگر θ باشد که در رابطه زیر صدق کند

$$\delta_.(a+b\underset{\sim}{x}) = a + b\delta_.(\underset{\sim}{x})$$

و δ_1 هر برآوردگر σ باشد که مقادیر مثبت را بگیرد و در رابطه زیر صدق کند

$$\delta_1(a+b\underset{\sim}{x}) = b\delta_1(\underset{\sim}{x})$$

در این صورت δ یک برآوردگر هم‌پایای مکان ($\delta(a+b\underset{\sim}{x}) = a+b\delta(\underset{\sim}{x})$) می‌باشد اگر و فقط اگر

$$\delta(\underset{\sim}{x}) = \delta_.(\underset{\sim}{x}) - w(\underset{\sim}{z})\delta_1(\underset{\sim}{x})$$

که در آن $Y_i = X_i - X_n \quad i = 1,\ldots,n-1$،

$$\underset{\sim}{Z} = (Z_1,\ldots,Z_{n-1}), Z_n = \frac{Y_{n-1}}{|Y_{n-1}|}, Z_i = \frac{Y_i}{Y_{n-1}} \; i = 1,\ldots,n-1$$

اثبات: همانند اثبات قضیه۲-۱ می‌باشد و بعنوان تمرین واگذار می‌گردد. ■

نتیجه ٤-۲: اگر $\delta_.$ و δ_1 برآوردگرهای هم‌پایای θ و σ باشند که در شرایط قضیه ٤-۱ صدق می‌کنند آنگاه برآوردگر MRE پارامتر مکان θ عبارت است از

$$\delta^*(\underset{\sim}{x}) = \delta_.(\underset{\sim}{x}) - w^*(\underset{\sim}{z})\delta_1(\underset{\sim}{x})$$

که در آن برای هر $\underset{\sim}{z}$، $w(\underset{\sim}{z})$ هر عددی است که $E_{.,1}\{\rho[\delta_.(\underset{\sim}{x}) - w^*(\underset{\sim}{z})\delta_1(\underset{\sim}{x})] \mid \underset{\sim}{z}\}$ را مینیمم کند. ■

نتیجه ٤-۳: اگر تابع زیان بفرم $\rho\left(\frac{\delta-\theta}{\sigma}\right)=\left(\frac{\delta-\theta}{\sigma}\right)^{۲}$ (فرم C) باشد آنگاه $w^{*}(\underset{\sim}{z})$ از رابطه زیر بدست می‌آید.

$$w^{*}(\underset{\sim}{z})=\frac{E_{.,۱}[\delta_{.}(\underset{\sim}{X})\delta_{۱}(\underset{\sim}{X})\mid \underset{\sim}{Z}]}{E_{.,۱}[\delta_{۱}^{۲}(\underset{\sim}{X})\mid \underset{\sim}{Z}]}$$

اثبات: تمرین. ■

مثال ٤-۵: فرض کنید $X_{۱},X_{۲},\ldots,X_{n}$ یک نمونه تصادفی از توزیع نمایی $E(\theta,\sigma)$ باشند که در آن θ و σ هر دو نامعلوم هستند. برآوردگرهای MRE پارامترهای θ تحت تابع زیان به فرم C و σ تحت تابع زیان به فرم A را بدست آورید.

حل: به راحتی دیده می‌شود که $T=(T_{۱},T_{۲})=\left(X_{(۱)},\sum_{i=۱}^{n}(X_{i}-X_{(۱)})\right)$ یک آماره بسنده کامل برای (θ,σ) است و $T_{۱}\sim E(\theta,\frac{\sigma}{n})$ و $T_{۲}\sim\frac{\sigma}{۲}\chi^{۲}_{(۲n-۲)}$ و $T_{۱}$ و $T_{۲}$ از یکدیگر مستقل هستند. بنابراین $\frac{۱}{n-۱}\sum_{i=۱}^{n}(X_{i}-X_{(۱)})$ و $X_{(۱)}-\frac{۱}{n(n-۱)}\sum_{i=۱}^{n}(X_{i}-X_{(۱)})$ به ترتیب برآوردگرهای UMVU پارامترهای θ و σ هستند.

الف- در برآورد σ تحت تابع زیان $L(\theta,\sigma,\delta)=(\frac{\delta}{\sigma}-۱)^{۲}$ (فرم A با $r=۱$) $\delta_{۱}=\sum_{i=۱}^{n}(X_{i}-X_{(۱)})$ یک برآوردگر هم‌پایا برای σ است $\left(\delta_{۱}(a+b\underset{\sim}{X})=b\delta_{۱}(\underset{\sim}{X})\right)$ و طبق قضیه باسو از آماره پایای ماکزیمال (فرعی) $\underset{\sim}{Z}$ مستقل است. بنابراین برآوردگر MRE پارامتر σ عبارت است از

$$\delta^{*}(\underset{\sim}{X})=\frac{\delta_{۱}E_{۱}[\delta_{۱}\mid \underset{\sim}{z}]}{E_{۱}[\delta_{۱}^{۲}\mid \underset{\sim}{z}]}=\frac{\delta_{۱}E_{۱}[\delta_{۱}]}{E_{۱}[\delta_{۱}^{۲}]}=\frac{\delta_{۱}(n-۱)}{(n-۱)+(n-۱)^{۲}}=\frac{\delta_{۱}}{n}=\frac{۱}{n}\sum_{i=۱}^{n}(X_{i}-X_{(۱)})$$

(توجه کنید در بخش قبل دیدیم که اگر $\theta=۰$ باشد، آنگاه برآوردگر MRE پارامتر σ عبارت از $\frac{۱}{n+۱}\sum_{i=۱}^{n}X_{i}$ بود).

ب- در برآورد θ تحت تابع زیان به فرم (C)،

$$\delta_{۱}(\underset{\sim}{X}) = \sum_{i=۱}^{n}(X_i - X_{(۱)}) \qquad \text{و} \qquad \delta_{.}(\underset{\sim}{X}) = X_{(۱)}$$

به ترتیب برآوردگرهای هم‌پایا برای θ و σ هستند $\begin{pmatrix} \delta_{.}(a+b\underset{\sim}{x}) = a + b\delta_{.}(\underset{\sim}{x}) \\ \delta_{۱}(a+b\underset{\sim}{x}) = b\delta_{۱}(\underset{\sim}{x}) \end{pmatrix}$ و بنابر قضیه باسو $(\delta_{.}, \delta_{۱})$ و $\underset{\sim}{Z}$ از یکدیگر مستقل هستند. بنابراین برآوردگر MRE پارامتر σ عبارت است از

$$w^{*}(\underset{\sim}{z}) = \frac{E_{.,۱}[\delta_{.}\delta_{۱} \mid \underset{\sim}{z}]}{E_{.,۱}[\delta_{۱}^{۲} \mid \underset{\sim}{z}]} = \frac{E_{.,۱}[\delta_{.}]E_{.,۱}[\delta_{۱}]}{E_{.,۱}[\delta_{۱}^{۲}]} = \frac{\frac{۱}{n}(n-۱)}{(n-۱)+(n-۱)^{۲}} = \frac{\frac{۱}{n}}{n} = \frac{۱}{n^{۲}}$$

$$\Rightarrow \delta^{*}(\underset{\sim}{X}) = \delta_{.}(\underset{\sim}{X}) - w^{*}(\underset{\sim}{z})\delta_{۱}(\underset{\sim}{X}) = X_{(۱)} - \frac{۱}{n^{۲}}\sum_{i=۱}^{n}(X_i - X_{(۱)})$$

(توجه کنید که در بخش‌های قبل دیدیم که برآوردگر MRE پارامتر مکان θ با فرض معلوم بودن σ برابر $X_{(۱)} - \frac{\sigma}{n}$ بود) . ■

توجه: تحت تابع زیان $\left(\frac{\delta-\theta}{\sigma}\right)^{۲}$ هیچ برآوردگر مخاطره ناریب وجود ندارد. توجه کنید که گروه $\tilde{G}$ در این حالت غیرجابجایی است. ■

فصل ۴

تصمیم بیزی

بخش ۱: مقدمه

تمامی روش‌های استنباط آماری که تاکنون بررسی کرده‌ایم یک نوع تصمیم‌گیری می‌باشند و در آنها به کمک نمونه‌های آماری و داده‌ها درباره پارامتر یا توزیع مجهول به تصمیم‌گیری می‌پرداختیم و به آنها آمار کلاسیک گویند. در نظریه تصمیم به کمک نمونه‌های آماری و دیگر جوانب، با بررسی مقدار سود و زیان می‌کوشیم تا به بهترین تصمیم‌گیری دست یابیم.

بنابراین نظریه تصمیم شامل روش‌های دیگری از مسائل استنباط آماری است که روش‌های قبلی را شامل می‌باشد و جنبه‌های جدیدی را نیز در نظر می‌گیرد. در فصل‌های قبل با عناصر یک مسئله تصمیم آشنا شده‌ایم ولی آنها را بطور کلی معرفی نکرده‌ایم . در زیر این عناصر را معرفی می‌کنیم.

داده‌ها: داده‌ها بوسیله یک بردار تصادفی $\underset{\sim}{X}$ با فضای نمونه(مجموعه مقادیر) $\mathcal{X}$ نمایش داده می‌شود.

وضع طبیعت[۱]: پارامتر واقعی و مجهول θ را که می‌خواهیم روی آن استنباط انجام دهیم را وضع طبیعت گویند.

فضای پارامتری[۲]: مجموعه تمام مقادیر ممکن θ را فضای پارامتری گویند و با نماد Θ نمایش می‌دهند.

مدل: مدل عبارت است از مجموعه کلیه توزیع‌های احتمال برای بردار $\underset{\sim}{X}$ که دارای پارامتر مجهول θ است. یعنی مجموعه $\{f(\underset{\sim}{x} \mid \theta) : \theta \in \Theta\}$ که در آن $f(\underset{\sim}{x} \mid \theta)$ تابع چگالی یا تابع احتمال روی $\mathcal{X}$ است.

فضای کارها[۳]: بعد از اینکه $\underset{\sim}{X} = \underset{\sim}{x}$ مشاهده گردید، تصمیمی برای θ اتخاذ می‌گردد. مجموعه کلیه تصمیم‌های ممکن را فضای کارها گویند و با نماد A نمایش می‌دهند و اعضای آن را با $a \in A$ نشان می‌دهند. فضای کارها معین می‌کند که مسئله تصمیم مورد مطالعه، یک مسئله برآورد نقطه‌ای یا فاصله اطمینان و یا آزمون فرض می‌باشد. در مسئله برآورد نقطه‌ای می‌خواهیم نقاطی از فضای کارها را

---

[۱] State of Nature
[۲] Parameter Space
[۳] Action Space

برای θ حدس بزنیم پس $\theta \subset A$ خواهد بود (مثلا" در برآورد پارامتر توزیع دو جمله‌ای $0 < p < 1$ و می‌تواند $A = [0,1]$ باشد) در مسئله آزمون فرض، ما دو کار "قبول H_0" و "رد H_0" را داریم که آن را با a_0 و a_1 نمایش می‌دهیم و بنابراین $A = \{a_0, a_1\}$. در یک مسئله فاصله اطمینان، کارها فواصل یا زیرمجموعه‌هایی از فضای پارامتر هستند و بنابراین $A \subset \Theta$.

تابع زیان: اگر $\theta \in \Theta$ مقدار واقعی وضع طبیعت باشد در این صورت کار $a \in A$ ممکن است درست یا تا اندازه‌ای نادرست و یا کلا نادرست باشد. میزان این نادرستی را با کمیت تابع زیان $L(\theta,a)$ اندازه‌گیری می‌کنیم که مقدار زیان بکار بردن کار a موقعی که θ مقدار وضع طبیعت واقعی باشد را اندازه می‌گیرد. بنابراین $L : \Theta \times A \to [0,+\infty)$ و $L(\theta,a) = 0$ به معنای این است که اگر θ وضع طبیعت واقعی است آنگاه a تصمیم درستی می‌باشد.

تابع تصمیم[1]: تابع تصمیم (یا قاعده تصمیم) قاعده‌ای است که با مشاهده $\underset{\sim}{X} = \underset{\sim}{x}$ معین می‌کند که چه کار $a \in A$ را بایستی بکار ببریم و معمولا آنرا با $\delta(\underset{\sim}{x})$ نمایش می‌دهند. بنابراین

$$\delta(\underset{\sim}{x}) : \mathcal{X} \to A$$

(در مسئله برآورد نقطه‌ای، توابع تصمیم همان برآوردگرها می‌باشند)

فضای توابع تصمیم: مجموعه کلیه توابع تصمیم ممکن را با D نمایش می‌دهیم. در نظریه تصمیم هدف این است که معین کنیم که کدامیک از توابع تصمیم $\delta(\underset{\sim}{x}) \in D$ یک قاعده تصمیم خوب می‌باشد.

تابع مخاطره[2] **(ریسک)**: در مسئله تصمیم روی پارامتر θ براساس تابع تصمیم $\delta(\underset{\sim}{x})$ میزان دقت تابع تصمیم بوسیله تابع مخاطره یعنی $R(\theta,\delta) = E_\theta\left[L\left(\theta,\delta(\underset{\sim}{X})\right)\right]$ اندازه گرفته می‌شود. چون مقدار θ نامعلوم است علاقه‌مند هستیم توابع تصمیمی را بکار ببریم که $R(\theta,\delta)$ را به ازای هر $\theta \in \Theta$ مینیمم کند. اما در حالت کلی چنین کاری ممکن نیست و بنابراین بایستی محدودیت‌هایی روی فضای توابع تصمیم (کلاس برآوردگرها) برای دستیابی به تصمیم بهینه اعمال کنیم. یک روش

[1] Decision Rule
[2] Risk Function

برای پیدا کردن تصمیم بهینه، برقراری یک رابطه ترتیب بین تصمیم‌ها می‌باشد. دو روش اساسی در برقراری رابطه‌ای ترتیبی بین تصمیمات، اصل کمین بیشینه[۱](مینیماکس) و اصل بیز[۲] می‌باشد.
در این فصل و فصل بعد این دو روش را برای بدست آوردن برآورد نقطه‌ای برای پارامتر مجهول θ بررسی می کنیم.

بخش۲- تصمیم بیزی[۳]

در روش‌هائی که قبلا مطالعه کردیم (روش‌های کلاسیک) پارامتر θ را یک مقدار ثابت نامعلوم در نظر گرفتیم و یک نمونه تصادفی $X_1, X_2, \ldots, X_n$ از جمعیت که دارای توزیع $f_\theta(x)$ بود جمع آوری کرده و براساس آن در مورد θ تصمیم‌گیری می‌کردیم. در روش بیزی θرا کمیتی در نظر می‌گیریم که خود یک متغیر تصادفی است و تغییرات آن توسط یک توزیع احتمال (که آنرا توزیع پیشین می‌نامیم) بیان می‌گردد. این توزیع پیشین براساس اعتقادات و تجربیات قبلی آزمایشگر و قبل از مشاهده داده‌ها تعیین می‌گردد. سپس از جمعیت یک نمونه جمع آوری می‌شود و بر اساس آن توزیع پیشین تصحیح می‌گردد. توزیع پیشین تصحیح شده را توزیع پسین می‌نامند و براساس این توزیع پسین تصمیم‌گیری انجام می‌گیرد. برای مثال فرض کنید قطعات تولیدی یک کارخانه دارای توزیع نمایی با میانگین θ ساعت می‌باشند و به علت فرسودگی دستگاه‌ها این میانگین θ نیز در سال تغییر می‌کند و تغییرات آن طبق یک توزیع احتمال $\pi(\theta)$ باشد.

مخاطره و تصمیم بیزی:

فرض کنید $\pi(\theta)$ توزیع احتمال روی فضای پارامتری Θ باشد که به آن توزیع پیشین[۴] θ گویند چون تابع مخاطره $R(\theta,\delta) = E_\theta[L(\theta,\delta(\underset{\sim}{X}))]$ تابعی از θ است و θ متغیر تصادفی است پس $R(\theta,\delta)$ نیز یک متغیر تصادفی است و تابعی از θ است. امید ریاضی (متوسط) $R(\theta,\delta)$ را نسبت به چگالی پیشین θ مخاطره بیز می‌نامند یعنی

[۱] Minimax
[۲] Bayes
[۳] Bayesian Decision
[۴] Prior Distribution

$$r(\pi,\delta)=E[R(\theta,\delta)]=\int_{\Theta}R(\theta,\delta)d\Pi(\theta)=\begin{cases}\int_{\Theta}R(\theta,\delta)\pi(\theta)d\theta & \text{اگر } \pi(\theta) \text{ پیوسته باشد}\\ \sum_{\Theta}R(\theta,\delta)\pi(\theta) & \text{اگر } \pi(\theta) \text{ گسسته باشد}\end{cases}$$

تصمیم بیزی نسبت به توزیع پیشین π آن تصمیم $\delta^{\pi}(\underset{\sim}{x})$ است که مخاطره بیزی را در بین تمام توابع تصمیم مینیمم کند، یعنی تصمیم بیزی آن تابع تصمیمی است که در رابطه زیر صدق کند.

$$r(\pi,\delta^{\pi})=\inf_{\delta\in D} r(\pi,\delta)$$

روش بدست آوردن تصمیم بیزی:

فرض کنید θ دارای توزیع پیشین $\pi(\theta)$ باشد (در اینجا θ هم بعنوان پارامتر و هم به عنوان متغیر تصادفی در نظر گرفته می‌شود) و $\underset{\sim}{X}=(X_1,\ldots,X_n)$ یک نمونه تصادفی جمع‌آوری شده از جمعیتی باشد که توزیع آن به θ بستگی دارد. براساس این نمونه تصادفی در توزیع پیشین $\pi(\theta)$ تجدید نظر می‌کنیم و چگالی پسین θ را به صورت زیر بدست می‌آوریم

$$\pi(\theta\mid\underset{\sim}{x})=\frac{f(\underset{\sim}{x},\theta)}{m(\underset{\sim}{x})}$$

که در آن $f(\underset{\sim}{x},\theta)$ چگالی توام $\underset{\sim}{X}$ و θ و $m(\underset{\sim}{x})$ چگالی کناری $\underset{\sim}{X}$ است بنابراین

$$f(\underset{\sim}{x},\theta)=\pi(\theta)f(\underset{\sim}{x}\mid\theta)\quad,\quad m(\underset{\sim}{x})=\int f(\underset{\sim}{x},\theta)d\Pi(\theta)$$

پس $$\pi(\theta\mid\underset{\sim}{x})=\frac{\pi(\theta)f(\underset{\sim}{x},\theta)}{m(\underset{\sim}{x})}=c(\underset{\sim}{x})\pi(\theta)f(\underset{\sim}{x}\mid\theta)\propto\pi(\theta)f(\underset{\sim}{x}\mid\theta)=\pi(\theta)L(\theta)$$

یعنی برای محاسبه توزیع احتمال پسین θ کافیست حاصلضرب $\pi(\theta)$ در $L(\theta)$ (تابع درست نمایی) را بدست آورده و آنرا به یک توزیع احتمال تبدیل کنیم.

مثال ۲-۱: فرض کنید $X_1,\ldots,X_n$ یک نمونه تصادفی از توزیع $B(1,\theta)$ باشند و θ دارای توزیع پیشین $Be(\alpha,\beta)$ باشد که در آن α و β معلوم هستند. چگالی پسین θرا بدست آورید.

حل:

$$\pi(\theta \mid \underset{\sim}{x}) \propto \pi(\theta)L(\theta) = \left[\frac{1}{Be(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\right]\left[\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\right]$$

$$\propto \theta^{\alpha+\sum x_i-1}(1-\theta)^{n+\beta-\sum x_i-1}$$

$$\Rightarrow \theta \mid \underset{\sim}{x} \sim Be(\alpha+\sum x_i, n+\beta-\sum x_i) \quad \blacksquare$$

برای بدست آوردن تصمیم بیزی از قضیه زیر استفاده می‌کنیم.

قضیه ۲-۱: فرض کنید θ دارای توزیع پیشین $\pi(\theta)$ باشد و به شرط θ، $\underset{\sim}{X}$ دارای توزیعی از خانواده $\{f(\underset{\sim}{x} \mid \theta) \mid \theta \in \Theta\}$ باشد. فرض کنید در مسئله تصمیم با تابع زیان غیرمنفی $L(\theta,\delta)$ شرایط زیر برقرار باشد.

الف- یک تابع تصمیم δ با مخاطره متناهی موجود باشد.

ب- برای هر $\underset{\sim}{x} \in \mathcal{X}$ یک تابع تصمیم $\delta^{\pi}(\underset{\sim}{x})$ موجود باشد که

$$r(\underset{\sim}{x},\delta(\underset{\sim}{x})) = E[L(\theta,\delta(\underset{\sim}{x})) \mid \underset{\sim}{X} = \underset{\sim}{x}] = \int_{\Theta} L(\theta,\delta(\underset{\sim}{x}))d\pi(\theta \mid \underset{\sim}{x})$$

را مینیمم کند (یعنی $r(\underset{\sim}{x},\delta^{\pi}) = \inf_{\delta \in D} r(\underset{\sim}{x},\delta)$).

در این صورت $\delta^{\pi}(\underset{\sim}{x})$ یک تصمیم بیز است ($r(\underset{\sim}{x},\delta)$ را مخاطره پسین گویند).

اثبات: فرض کنید δ هر تصمیمی با مخاطره متناهی باشد. در این صورت چون تابع زیان نامنفی است.

پس $$E\{L(\theta,\delta(\underset{\sim}{x})) \mid \underset{\sim}{X} = \underset{\sim}{x}\} < \infty \qquad a.e.$$

و در نتیجه طبق شرط (ب) داریم که

$$E\{L(\theta,\delta(\underset{\sim}{x})) \mid \underset{\sim}{X} = \underset{\sim}{x}\} \geq E\{L(\theta,\delta^{\pi}(\underset{\sim}{x}) \mid \underset{\sim}{X} = \underset{\sim}{x}\} \qquad a.e.$$

$$\Rightarrow E\left[E\{L(\theta,\delta(\underset{\sim}{x})) \mid \underset{\sim}{X} = \underset{\sim}{x}\}\right] \geq E\left[E\{L(\theta,\delta^{\pi}(\underset{\sim}{x})) \mid \underset{\sim}{X} = \underset{\sim}{x}\}\right]$$

$$\Rightarrow E[L(\theta,\delta(\underset{\sim}{x}))] \geq E[L(\theta,\delta^{\pi}(\underset{\sim}{x}))]$$

که امید ریاضی نسبت به توزیع توام θ و $\underset{\sim}{X}$ است پس $r(\pi,\delta) \geq r(\pi,\delta^{\pi})$ یعنی $\delta^{\pi}(\underset{\sim}{X})$ یک تصمیم بیز است زیرا مخاطره بیز را مینیمم می‌کند. $\blacksquare$

با توجه به قضیه بالا برای بدست آوردن تصمیم بیزی بایستی مخاطره پسین را مینیمم کنیم. در برآورد نقطه‌ای، تصمیم بیزی بدست آمده را برآوردگر بیز[۱] می‌نامیم و در قضیه زیر تحت توابع زیان مختلف این برآوردگر را بدست می‌آوریم.

قضیه ۲-۲: تحت فرضیات قضیه۲-۱ در برآورد نقطه‌ای پارامتر $\gamma(\theta)$داریم که

الف- تحت تابع زیان درجه دوم $L\left(\gamma(\theta),\delta(\underset{\sim}{X})\right)=[\delta(\underset{\sim}{X})-\gamma(\theta)]^{۲}$ برآورد بیز عبارت است از:

$$\delta^{\pi}(\underset{\sim}{x})=E[\gamma(\theta)\,|\,\underset{\sim}{x}]$$

ب- تحت تابع زیان قدر مطلق $L\left(\gamma(\theta),\delta(\underset{\sim}{X})\right)=\left|\delta(\underset{\sim}{X})-\gamma(\theta)\right|$ برآورد بیز، میانه توزیع $\gamma(\theta)\,|\,\underset{\sim}{x}$ است.

ج- تحت تابع زیان درجه دوم وزنی $L\left(\gamma(\theta),\delta(\underset{\sim}{X})\right)=w(\theta)[\delta(\underset{\sim}{X})-\gamma(\theta)]^{۲}$ برآورد بیز عبارت است از:

$$\delta^{\pi}(\underset{\sim}{x})=\frac{E[w(\theta)\gamma(\theta)\,|\,\underset{\sim}{x}]}{E[w(\theta)\,|\,\underset{\sim}{x}]}$$

اثبات (ج):

$$r\left(\underset{\sim}{x},\delta(\underset{\sim}{x})\right)=\int_{\Theta}w(\theta)[\delta(\underset{\sim}{x})-\gamma(\theta)]^{۲}\pi(\theta\,|\,\underset{\sim}{x})d\theta$$

$$=\delta^{۲}(\underset{\sim}{x})E[w(\theta)\,|\,\underset{\sim}{x}]-۲\delta(\underset{\sim}{x})E[w(\theta)\gamma(\theta)\,|\,\underset{\sim}{x}]+E[w(\theta)\gamma^{۲}(\theta)\,|\,\underset{\sim}{x}]$$

■ $$\frac{\partial r(\underset{\sim}{x},\delta(\underset{\sim}{x}))}{\partial\delta(\underset{\sim}{x})}=۰\Rightarrow\delta_{\pi}(\underset{\sim}{x})=.....$$

مثال ۲-۲: در مثال۲-۱ برآوردگر بیز پارامتر θرا تحت تابع زیان درجه دوم بدست آورید.

$$X_{۱},\ldots,X_{n}\sim B(۱,\theta)\qquad\theta\sim Be(\alpha,\beta)\Rightarrow\theta\,|\,\underset{\sim}{x}\sim Be(\alpha+\sum x_{i},\beta+n-\sum x_{i})$$

$$\delta^{\pi}(\underset{\sim}{x})=E[\theta\,|\,\underset{\sim}{x}]=\frac{\alpha+\sum x_{i}}{\alpha+\beta+n}=\left(\frac{\alpha+\beta}{\alpha+\beta+n}\right)\frac{\alpha}{\alpha+\beta}+\left(\frac{n}{\alpha+\beta+n}\right)\bar{x}$$

چون $E(\theta)=\dfrac{\alpha}{\alpha+\beta}$ بنابراین برآوردگر بیز بصورت یک میانگین وزنی از میانگین توزیع پیشین $\dfrac{\alpha}{\alpha+\beta}$(برآورد θ قبل از مشاهده داده‌ها) و میانگین نمونه $\bar{x}$ (برآورد θ بوسیله مشاهدات) می‌باشد.

[۱] Bayes Estimator

توجه کنید که این برآوردگر ناریب نیست و در صورتی که $n \to \infty$ و یا $\alpha = \beta = ۰$ انگاه برآوردگر $\delta^{\pi}(\underset{\sim}{x})$ با برآوردگر UMVU یکی می‌شود.

البته حالت $\alpha = \beta = ۰$ یک توزیع پیشین نمی‌باشد زیرا بایستی $\alpha > ۰, \beta > ۰$ باشد(حالت $\alpha \to ۰$ و با $\beta \to ۰$ را بعدا بحث خواهیم کرد). ■

مثال ۲-۳: فرض کنید $X_۱,\ldots,X_n$ یک نمونه تصادفی n تایی از توزیع $N(\theta,\sigma^۲)$ باشند که در آن $\sigma^۲$ یک مقدار معلوم است. اگر $\theta \sim N(\mu,\tau^۲)$ و تابع زیان بفرم درجه۲ باشد برآوردگر بیز پارامتر θ را بدست آورید.

حل:

$$\pi(\theta \mid \underset{\sim}{x}) \propto \pi(\theta)L(\theta) = \frac{۱}{\sqrt{۲\pi\tau^۲}} e^{-\frac{۱}{۲\tau^۲}(\theta-\mu)^۲} \left(\frac{۱}{\sqrt{۲\pi\sigma^۲}}\right)^n e^{-\frac{۱}{۲\sigma^۲}\sum_{i=۱}^{n}(X_i-\theta)^۲}$$

$$\propto e^{-\frac{\theta^۲}{۲\tau^۲}+\frac{\mu}{\tau^۲}\theta} e^{\frac{\theta}{\sigma^۲}\sum x_i - \frac{n\theta^۲}{۲\sigma^۲}} = e^{-\frac{\theta^۲}{۲}(\frac{۱}{\tau^۲}+\frac{n}{\sigma^۲})+\theta(\frac{\mu}{\tau^۲}+\frac{n\bar{x}}{\sigma^۲})}$$

$$= e^{-\frac{۱}{۲}(\frac{n}{\sigma^۲}+\frac{۱}{\tau^۲})[\theta^۲ - ۲\theta\frac{\frac{\mu}{\tau^۲}+\frac{n\bar{x}}{\sigma^۲}}{\frac{۱}{\tau^۲}+\frac{n}{\sigma^۲}}]} = \exp\left\{-\frac{۱}{۲}(\frac{n}{\sigma^۲}+\frac{۱}{\tau^۲})\left(\theta^۲ - ۲\theta\frac{\mu\left(\frac{\sigma^۲}{n}\right)+\tau^۲\bar{x}}{\tau^۲+\frac{\sigma^۲}{n}}\right)\right\}$$

$$\propto \exp\left\{-\frac{۱}{۲}(\frac{n}{\sigma^۲}+\frac{۱}{\tau^۲})\left(\theta - \frac{\mu\left(\frac{\sigma^۲}{n}\right)+\tau^۲\bar{x}}{\tau^۲+\frac{\sigma^۲}{n}}\right)^۲\right\}$$

$$\therefore \quad \theta \mid \underset{\sim}{x} \sim N\left(\frac{\mu\left(\frac{\sigma^۲}{n}\right)+\tau^۲\bar{x}}{\frac{\sigma^۲}{n}+\tau^۲}, \frac{۱}{\frac{n}{\sigma^۲}+\frac{۱}{\tau^۲}}\right)$$

$$\Rightarrow \delta^{\pi}(\underset{\sim}{x}) = E(\theta \mid \underset{\sim}{x}) = \frac{\frac{\sigma^{۲}}{n}}{\frac{\sigma^{۲}}{n}+\tau^{۲}}\mu + \frac{\tau^{۲}}{\frac{\sigma^{۲}}{n}+\tau^{۲}}\bar{x}$$

یعنی برآوردگر بیز بصورت یک میانگین وزنی از میانگین توزیع پیشین (μ) و میانگین نمونه $\bar{x}$ (برآورد عادی θ) می‌باشد و به سادگی دیده می‌شود که ناااریب نیست. ■

مثال ۲-٤: فرض کنید $X_{۱},\ldots,X_{n}$ یک نمونه تصادفی n تایی از توزیع $U(۰,\theta)$ باشند اگر $\theta \sim U(۰,۱)$ باشد و تابع زیان بصورت $L(\theta,\delta) = \frac{(\delta-\theta)^{۲}}{\theta^{۲}}$ باشد برآوردگر بیز پارامتر θ را بدست آورید.

حل:

$$\pi(\theta \mid \underset{\sim}{x}) \propto \pi(\theta)L(\theta) = I_{(۰,۱)}(\theta)\prod_{i=۱}^{n}\frac{۱}{\theta}I_{(۰,\theta)}(x_i) = I_{(۰,۱)}(\theta)\frac{۱}{\theta^{n}}I_{(x_{(n)},+\infty)}(\theta)$$

$$= \frac{۱}{\theta^{n}}I_{(x_{(n)},۱)}(\theta) \Rightarrow \pi(\theta \mid \underset{\sim}{x}) = c(\underset{\sim}{x})\frac{۱}{\theta^{n}}I_{(x_{(n)},۱)}(\theta)$$

$$\Rightarrow \delta^{\pi}(\underset{\sim}{x}) = \frac{E[\theta w(\theta) \mid \underset{\sim}{x}]}{E[w(\theta) \mid \underset{\sim}{x}]} = \frac{E(\frac{۱}{\theta} \mid \underset{\sim}{x})}{E(\frac{۱}{\theta^{۲}} \mid \underset{\sim}{x})} = \frac{\int_{x_{(n)}}^{۱}\frac{۱}{\theta^{n+۱}}d\theta}{\int_{x_{(n)}}^{۱}\frac{۱}{\theta^{n+۲}}d\theta} = \frac{-\frac{۱}{n\theta^{n}}\Big|_{x_{(n)}}^{۱}}{\frac{-۱}{(n+۱)\theta^{n+۱}}\Big|_{x_{(n)}}^{۱}}$$

$$= \frac{\frac{۱}{n}[-۱+\frac{۱}{x_{(n)}^{n}}]}{\frac{۱}{n+۱}[-۱+\frac{۱}{x_{(n)}^{n+۱}}]} = \frac{n+۱}{n}x_{(n)}\left[\frac{۱-x_{(n)}^{n}}{۱-x_{(n)}^{n+۱}}\right]$$

■

نکاتی در مورد برآوردگر بیز:

نکته۱: در رابطه با برآوردگرهای بیز این سوال مطرح می‌شود که چه خانواده از توزیع‌های پیشین را برای یک مسئله انتخاب کنیم. یکی از روش‌های انتخاب خانواده توزیع‌های پیشین، استفاده از خانواده توزیع‌های مزدوج است.

تعریف۲-۱: فرض کنید F خانواده توزیع‌های $f(\underset{\sim}{x}\mid\theta)$ باشد یعنی $F=\{f(\underset{\sim}{x}\mid\theta)\mid\theta\in\Theta\}$. خانواده Π از توزیعهای پیشین برای θ را یک **خانواده مزدوج** برای F گوئیم هرگاه برای هر $f\in F$ و هر $\pi\in\Pi$ و هر $x\in\mathcal{X}$ توزیع‌های پسین نیزدر خانواده Π قرار گیرند. ■

برای مثال در مثال۲-۲ خانواده $Be(\alpha,\beta)$ یک خانواده مزدوج برای $B(۱,\theta)$ بود و در مثال۲-۳ خانواده $N(\mu,\tau^{۲})$ یک خانواده مزدوج برای $N(\theta,\sigma^{۲})$ بود.

مثال ۲-۵: فرض کنید $X_{۱},X_{۲},\ldots,X_{n}$ یک نمونه تصادفی از توزیع نمایی با تابع چگالی زیر باشند $f(x\mid\theta)=\theta e^{-\theta x}\ ,x>۰$. اگر $\theta\sim\Gamma(\alpha,\beta)$ باشد تحت تابع زیان درجه دوم برآوردگر بیز میانگین توزیع یعنی $\gamma(\theta)=\frac{۱}{\theta}$ را بدست آورید.

حل:

$$\pi(\theta\mid\underset{\sim}{x})\propto\pi(\theta)L(\theta)=\frac{۱}{\beta^{\alpha}\Gamma(\alpha)}\theta^{\alpha-۱}e^{-\theta/\beta}\theta^{n}e^{-n\theta\bar{x}}\propto\theta^{n+\alpha-۱}e^{-(n\bar{x}+\frac{۱}{\beta})\theta}$$

$$\Rightarrow\theta\mid\underset{\sim}{x}\sim\Gamma(n+\alpha,\frac{۱}{n\bar{x}+\beta^{-۱}})$$

$$\delta^{\pi}(\underset{\sim}{x})=E[\gamma(\theta)\mid\underset{\sim}{x}]=E[\frac{۱}{\theta}\mid\underset{\sim}{x}]=\frac{\left(n\bar{x}+\beta^{-۱}\right)^{n+\alpha}}{\Gamma(n+\alpha)}\times\frac{\Gamma(n+\alpha-۱)}{\left(n\bar{x}+\beta^{-۱}\right)^{n+\alpha-۱}}$$

$$=\frac{n\bar{x}+\beta^{-۱}}{n+\alpha-۱}=\left(\frac{\alpha-۱}{n+\alpha-۱}\right)\left(\frac{۱}{\beta(\alpha-۱)}\right)+\left(\frac{n}{n+\alpha-۱}\right)\bar{x}$$

بنابراین برآورد بیز، یک میانگین وزنی از میانگین نمونه $\bar{x}$ (برآوردگر معقول $\frac{۱}{\theta}$) و میانگین $\frac{۱}{\theta}$ در توزیع پیشین $\left(\frac{۱}{\beta(\alpha-۱)}\right)$ می‌باشد. ■

مثال ۲-٦: فرض کنید $X_{۱},\ldots,X_{n}$ یک نمونه تصادفی از توزیع $P(\theta)$ باشند اگر $\theta\sim\Gamma(\alpha,\beta)$ باشد برآورد بیز پارامتر θ تحت تابع زیان درجه دوم را بدست آورید.

$$\pi(\theta \mid \underset{\sim}{x}) \propto \pi(\theta)L(\theta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\theta^{\alpha-1}e^{-\theta/\beta}\frac{e^{-n\theta}\theta^{n\bar{x}}}{\prod_{i=1}^{n} x_i!} \propto \theta^{n\bar{x}+\alpha-1}e^{-\theta(n+\frac{1}{\beta})}$$

$$\Rightarrow \theta \mid \underset{\sim}{x} \sim \Gamma(n\bar{x}+\alpha, \frac{1}{n+\beta^{-1}}) \Rightarrow E(\theta \mid \underset{\sim}{x}) = \frac{n\bar{x}+\alpha}{n+\beta^{-1}}$$

$$\therefore \delta^{\pi}(\underset{\sim}{X}) = \frac{\beta^{-1}}{n+\beta^{-1}}(\alpha\beta) + \frac{n}{n+\beta^{-1}}\bar{X}$$

برآوردگر بیز یک میانگین وزنی از میانگین نمونه $\bar{X}$ (برآوردگر معقول θ) و میانگین توزیع پیشین، $\alpha\beta$است. ■

نکته ۲: در مثال‌های قبل دیدیم که برآوردگرهای بیز ناریب نمی‌باشند . این موضوع در قضیه زیر اثبات می‌شود.

قضیه ۲-۳: فرض کنید θ دارای توزیع پیشین $\pi(\theta)$ باشد و $f(\underset{\sim}{x} \mid \theta)$ توزیع شرطی $\underset{\sim}{X}$ به شرط θ باشد. در برآورد پارامتر $\gamma(\theta)$ تحت تابع زیان درجه دوم، هیچ برآوردگر ناریب $\delta(\underset{\sim}{X})$ نمی‌تواند یک برآوردگر بیز باشد، بجز اینکه $r(\pi,\delta) = E_{\underset{\sim}{X},\theta}\left[\left(\delta(\underset{\sim}{X})-\theta\right)^2\right] = 0$

این امید ریاضی نسبت به چگالی توام $\underset{\sim}{X}$ و θ است.

اثبات: فرض کنید δ یک برآوردگر بیز و ناریب برای $\gamma(\theta)$ باشد. در اینصورت چون δ برآوردگر ناریب $\gamma(\theta)$ است پس $E[\delta(\underset{\sim}{X}) \mid \theta] = \gamma(\theta)$

و چون δ برآوردگر بیز $\gamma(\theta)$ است پس $E[\gamma(\theta) \mid \underset{\sim}{X}] = \delta(\underset{\sim}{X})$

همچنین محاطره بیز عبارت است از

$$r(\Pi,\delta) = E\left[\left(\delta(\underset{\sim}{X})-\gamma(\theta)\right)^2\right] = E[\delta^2(\underset{\sim}{X})] - 2E\left[\left(\gamma(\theta)\delta(\underset{\sim}{X})\right)\right] + E\left[\gamma^2(\theta)\right] \quad (1)$$

از طرفی

$$E[\gamma(\theta)\delta(\underset{\sim}{X})] = E\left[E\left(\gamma(\theta)\delta(\underset{\sim}{X}) \mid \underset{\sim}{X}\right)\right] = E\left[\delta(\underset{\sim}{X})E\left[\gamma(\theta) \mid \underset{\sim}{X}\right]\right] = E[\delta^2(\underset{\sim}{X})] \quad (2)$$

$$E[\gamma(\theta)\delta(\underset{\sim}{X})] = E\left[E\left(\gamma(\theta)\delta(\underset{\sim}{X}) \mid \theta\right)\right] = E\left[\gamma(\theta)E\left[\delta(\underset{\sim}{X}) \mid \theta\right]\right] = E[\gamma^2(\theta)] \quad (3)$$

و در نتیجه از روابط (۱) و (۲) و (۳) نتیجه می‌شود که $r(\pi,\delta)=۰$ است. ■

مثال ۲-۷: در مثال۲-۳ دیدیم که $\delta^{\pi}(\underset{\sim}{X})=\frac{\sigma^{۲}/n}{\sigma^{۲}/n+\tau^{۲}}\mu+\frac{\tau^{۲}}{\sigma^{۲}/n+\tau^{۲}}\bar{X}$ برآوردگر بیز θ در توزیع $N(\theta,\sigma^{۲})$ بود. نشان دهید که $\bar{X}$ برآوردگر UMVU پارامتر θاست نمی‌تواند برآوردگر بیز باشد.

حل: اگر $\bar{X}$ بخواهد برآوردگر نااریب بیز باشد بایستی $E_{\underset{\sim}{X},\theta}\left((\bar{X},\theta)^{۲}\right)=۰$ باشد. اما

$$E_{\underset{\sim}{X},\theta}\left((\bar{X}-\theta)^{۲}\right)=E\left[E(\bar{X}-\theta)^{۲}\mid\theta\right]=E\left[\frac{\sigma^{۲}}{n}\mid\theta\right]=\frac{\sigma^{۲}}{n}\neq۰$$

پس $\bar{X}$ نمی‌تواند برآوردگر بیز باشد.

مثال ۲-۸: در مثال۲-۲ دیدیم که در خانواده توزیع‌های $B(۱,\theta)$ تحت توزیع پیشین $Be(\alpha,\beta)$ و تابع زیان درجه دوم برآوردگر بیز برابر $\delta^{\pi}(\underset{\sim}{X})=\frac{\alpha+n\bar{X}}{\alpha+\beta+n}$ بود. تحت چه توزیع پیشینی $\bar{X}$ یک برآوردگر بیز برای θاست.

حل: $\bar{X}$ در صورتی برآوردگر بیز θ است که

$$۰=E_{\underset{\sim}{X},\theta}\left[(\bar{X}-\theta)^{۲}\right]=E\left[E(\bar{X}-\theta)^{۲}\mid\theta\right]=E\left[Var(\bar{X})\mid\theta\right]=E\left[\frac{\theta(۱-\theta)}{n}\right]$$

$$\Rightarrow E\left[\frac{\theta(۱-\theta)}{n}\right]=۰\overset{۰\le\theta\le۱}{\Rightarrow}P\left(\theta(۱-\theta)=۰\right)=۱\Rightarrow P(\theta=۰\ or\ \theta=۱)=۱$$

$$\Rightarrow P(\theta=۰)+P(\theta=۱)=۱$$

بنابراین در صورتی $\bar{X}$ برآوردگر بیز است که θ دارای توزیع پیشین برنولی باشد و برای چنین توزیع پیشینی داریم که

$$\delta^{\pi}(۱)=۱\qquad\delta^{\pi}(۰)=۰\quad(\text{چون }\theta=۰\text{ یا }۱\text{ است}).\qquad(٤)$$

و هر برآوردگری که در شرط بالا صدق کند براوردگر بیز است. مثلا $\delta^{\pi}(\underset{\sim}{X})=\bar{X}$ در این شرایط صدق می‌کند. البته این چگالی پیشین غیرمعقول است.

نکته ۳: در مثال‌های بالا اکثر برآوردگرهای بیز یکتا بودند و تنها در مثال ۲-۸ با تابع چگالی پیشین غیرمعقول $P(\theta=۰)+P(\theta=۱)=۱$ دیدیم که برآوردگر بیز یکتا نبود. شرط کلی برای یکتا بودن برآوردگر بیز در قضیه زیر آورده شده است.

قضیه ۲-٤: اگر تابع زیان $L(\theta,\delta)$ درجه دوم یا بطور کلی اکیدا محدب باشد و P کلاس کلیه توزیع‌های P_θ باشد، آنگاه برآورد بیز δ^π یکتا ($a.e.\ P$) است هرگاه

الف- مخاطره بیز آن نسبت به π متناهی باشد $r(\pi,\delta^\pi)<\infty$

ب- اگر Q توزیع کناری X باشد یعنی $Q(A)=\int P_\theta(X\in A)d\Pi(\theta)$

آنگاه $$a.e.Q \Rightarrow a.e.\ P$$

(به عبارتی $P(\underset{\sim}{X}\in A)=۰\Rightarrow P(\underset{\sim}{X}\in A\mid\theta)=۰$ و یا $Q(N)=۰\Rightarrow P(N)=۰$)

اثبات: برای تابع زیان درجه دوم از قضیه۲-۲ نتیجه می‌شود که برآوردگر بیز با مخاطره متناهی بایستی در رابطه $\delta^\pi(\underset{\sim}{X})=E\left[\gamma(\theta)\mid\underset{\sim}{X}\right]$ صدق کند بجز روی مجموعه‌ای از نقاط $\underset{\sim}{x}$ مانند N که $Q(N)=۰$ و این بدان معنی است که برآوردگر بیز یکتا است. در حالت تابع زیان اکیدا محدب نیز نتیجه به صورت فوق است زیرا این نوع توابع دارای نقطه مینیمم یکتا می‌باشند. ■

مثال ۲-۹: در مثال۲-۸ دیدیم که تحت چگالی پیشین $P(\theta=۰)+P(\theta=۱)=۱$ برآوردگر بیز پارامتر θ در خانواده $B(۱,\theta)$ در شرط $\delta^\pi(۰)=۰$ و $\delta^\pi(۱)=۱$ صدق می‌کرد و یکتا نبود. علت یکتا نبودن آن برقرارنبودن شرط(ب)قضیه ۲-٤ است زیرا

$$P(X_۱=۱,\ldots,X_n=۱\mid\theta)=\prod_{i=۱}^{n}\theta=\theta^n\neq۰$$

$$P(X_۱=۱,\ldots,X_n=۱)=\sum_{t=۰}^{۱}P(X_۱=۱,\ldots,X_n=۱\mid\theta=t)P(\theta=t)$$

$$=۰\times P(\theta=۰)+۰\times P(\theta=۱)=۰$$

$$\left(f(\underset{\sim}{x}\mid\theta)=\theta^{\sum x_i}(۱-\theta)^{n-\sum x_i}=۰\ \ if\ \ \theta=۱\right)$$

بنابراین $$P(X_۱=۱,\ldots,X_n=۱)=۰\not\Rightarrow P(X_۱=۱,\ldots,X_n=۱\mid\theta)=۰$$ ■

نکته ٤: در بعضی از مسایل ممکن است برای بدست آوردن برآوردگر بیز، تابع چگالی پسین را بطور کامل محاسبه کنیم. به مثال زیر توجه کنید.

مثال۲-۱۰: فرض کنید X دارای توزیع $U(۰,\theta)$ باشد و θ دارای چگالی پیشین بصورت $\pi(\theta)=\begin{cases}\theta e^{-\theta} & \theta>۰\\ ۰ & O.W.\end{cases}$ باشد. تحت هر یک از تابع زیان‌های درجه دوم و قدرمطلق خطا برآوردگرهای بیز را بدست آورید.

حل:

$$\pi(\theta\mid x)=\frac{\pi(\theta)f(x\mid\theta)}{m(x)}=c(x)\pi(\theta)f(x\mid\theta)=c(x)\theta e^{-\theta}\frac{۱}{\theta}I_{(x,+\infty)}(\theta)$$

$$۱=\int\pi(\theta\mid x)d\theta\Rightarrow ۱=\int_x^{+\infty}c(x)e^{-\theta}d\theta=c(x)[-e^{-\theta}]_x^{+\infty}=c(x)e^{-x}$$

$$\Rightarrow c(x)=e^{x}\qquad\therefore\pi(\theta\mid x)=e^{-(\theta-x)}I_{(x,+\infty)}(\theta)$$

الف) $L(\theta,\delta)=(\delta-\theta)^{۲}\Rightarrow\delta^{\pi}(x)=E(\theta\mid x)=x+۱\Rightarrow\delta^{\pi}(X)=X+۱$

ب)

$$L(\theta,\delta)=\mid\delta-\theta\mid\Rightarrow\int_x^m\pi(\theta\mid x)d\theta=\frac{۱}{۲}\Rightarrow\int_x^m e^{-(\theta-x)}d\theta=\frac{۱}{۲}$$

$$\Rightarrow -e^{-(\theta-x)}\mid_x^m=\frac{۱}{۲}\Rightarrow ۱-e^{-(m-x)}=\frac{۱}{۲}\Rightarrow m-x=\log۲\Rightarrow m=x+\log۲$$

$$\Rightarrow\delta^{\pi}(X)=X+\log۲$$

■

نکته ۵: اگر در مسئله‌ای یک آماره بسنده $\underset{\sim}{T}$ موجود باشد آنگاه توزیع پسین به آماره $\underset{\sim}{T}$ وابسته است و در حقیقت $\pi(\theta\mid\underset{\sim}{x})=\pi(\theta\mid\underset{\sim}{t})$. برای اثبات این مطلب توجه کنید که از قضیه تجزیه به عوامل داریم که

$$L(\theta)=f(\underset{\sim}{x}\mid\theta)=g(\underset{\sim}{t}\mid\theta)h(\underset{\sim}{x})$$

پس

$$\pi(\theta\mid\underset{\sim}{x})=\frac{f(\underset{\sim}{x}\mid\theta)\pi(\theta)}{\int f(\underset{\sim}{x}\mid\theta')d\pi(\theta')}=\frac{g(\underset{\sim}{t}\mid\theta)\pi(\theta)}{\int g(\underset{\sim}{t}\mid\theta')d\pi(\theta')}=\pi(\theta\mid\underset{\sim}{t})$$

یعنی توزیع پسین $\pi(\theta\mid\underset{\sim}{t})$ بستگی به $\underset{\sim}{x}$ از طریق $\underset{\sim}{t}$ دارد و بنابراین محاسبه توزیع پسین براساس $\underset{\sim}{x}$ یا $\underset{\sim}{t}$ یکسان است. برای مثال، در مثال۲-۳ داریم که

$$X_1,\ldots,X_n \;\; i.i.d \;\; N(\mu,\sigma^2) \qquad ,T=\bar{X}\sim N(\theta,\frac{\sigma^2}{n}) \qquad ,\theta\sim N(\mu,\tau^2)$$

$$\pi(\theta\,|\,t)\propto f\,(t\,|\,\theta)\pi(\theta)\propto e^{-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2}\,e^{-\frac{1}{2\tau^2}(\theta-\mu)^2}\propto e^{-\frac{\theta^2}{2}(\frac{1}{\tau^2}+\frac{n}{\sigma^2})+\theta(\frac{\mu}{\tau^2}+\frac{n\bar{x}}{\sigma^2})}$$

که همان تابع به دست آمده در مثال۲-۳ است بنابراین

$$\theta\,|T=t\sim N\left(\frac{\mu\left(\frac{\sigma^2}{n}\right)+\tau^2\bar{x}}{\frac{\sigma^2}{n}+\tau^2},\frac{1}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\right) \qquad \blacksquare$$

توزیع‌های پیشین ناآگاهی بخش[1]:

همانگونه که قبلا گفته شد، توزیع پیشین $\pi(\theta)$ براساس اعتقادات و تجربیات قبلی پژوهشگر تشکیل می‌گردد. حال اگر هیچ توزیع پیشینی در دسترس نباشد، می‌توان از توزیع‌های پیشن ناآگاهی بخش استفاده کرد. برای مثال اگر در یک مسئله آزمون فرض ساده در مقابل ساده احتمالات یکسان $\frac{1}{2}$ را به فرض صفر و مقابل نسبت دهیم آنگاه این توزیع پیشین، ناآگاهی بخش است.

یکی از روش‌های بدست آوردن توزیع پیشین ناآگاهی بخش توسط جفریز[2] (۱۹۶۱) ارایه شده است، که در این روش چگالی پیشین را به صورت $\pi(\theta)\propto\sqrt{|I(\theta)|}$ در نظر می‌گیرد که در آن $|I(\theta)|$ دترمینان ماتریس اطلاع فیشر می‌باشد.

مثال ۲-۱۱: اگر $X\sim N(\theta,1)$ آنگاه توزیع پیشین جفریز را برای θ بدست آورید.

$$f\,(x\,|\,\theta)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\theta)^2}\Rightarrow I(\theta)=1\Rightarrow\pi(\theta)=1\;\; -\infty<\theta<+\infty \qquad \blacksquare$$

توجه کنید که توزیع پیشین به دست آمده در مثال ۲-۱۱ یک تابع چگالی نیست. اگر در مسئله‌ای اندازه پیشین در شرط $\int d\Pi(\theta)=\infty$ صدق کند آنگاه پیشین حاصل را ناسَره[3] گویند و توزیع پیشین

[1] Noninformative Priors
[2] Jeffreys
[3] Improper Prior

که در شرط $\int d\Pi(\theta)=۱$ صدق کند را توزیع پیشین سَره[۱] گویند. بنابراین توزیع پیشین بدست آمده در مثال ۲-۱۱ ناسره است.

بایستی توجه کرد که اگر توزیع پیشینی ناسره باشد آنگاه ممکن است مخاطره براساس این توزیع پیشین متناهی باشد و برآوردگر بیز آن قابل محاسبه باشد. برای مثال، در مثال ۲-۱۱ داریم که

$$\pi(\theta\,|\,x)\propto f\ (x\ |\theta)\pi(\theta)=f\ (x\ |\theta)=\frac{۱}{\sqrt{۲\pi}}e^{-\frac{۱}{۲}(x-\theta)^{۲}}$$

پس توزیع پسین سره است و $\theta\,|\,x\sim N\,(x\,,۱)$ در نتیجه $\delta^{\pi}(X\,)=E\,(\theta\,|\,X\,)=X$. برآوردگر بیز حاصل را برآوردگر بیز تعمیم یافته گویند.

تعریف۲-۲: یک برآوردگر $\delta_G^{\pi}(x\,)$ را برآوردگر بیز تعمیم یافته[۲] نسبت به اندازه $\pi(\theta)$ گویند اگر مخاطره پسین $E\left[L\left(\theta,\delta(x)\right)|\,\underset{\sim}{X}=\underset{\sim}{x}\right]$ در $\delta=\delta_G^{\pi}$ برای هر $\underset{\sim}{x}$ مینیمم شود. ■

مثال ۲-۱۲: در مثال ۲-۲ توزیع پیشین جفریز را بدست آورده وسپس براساس توزیع پیشین $\pi(\theta)=\frac{۱}{\theta(۱-\theta)}$، برآوردگر بیز(تعمیم یافته) θ تحت تابع زیان درجه دوم را بدست آورید.

حل:

$$X_{\,i}\sim B\,(۱,\theta)\qquad i=۱,۲,...,n\qquad f_P\,(x\,)=\theta^{x}\,(۱-\theta)^{۱-x}\Rightarrow I\,(\theta)=\frac{n}{\theta(۱-\theta)}$$

توزیع سره است $\pi(\theta)\propto\sqrt{|\,I\,(\theta)\,|}\Rightarrow\Pi(\theta)\propto\frac{۱}{\theta^{۱/۲}(۱-\theta)^{۱/۲}}$

براساس توزیع پیشین ناسره $\pi(\theta)=\frac{۱}{\theta(۱-\theta)}$ داریم

$$\pi(\theta\,|\,\underset{\sim}{x}\,)\propto\theta^{n\bar{x}}\,(۱-\theta)^{n-n\bar{x}}\,\frac{۱}{\theta(۱-\theta)}=\theta^{n\bar{x}-۱}(۱-\theta)^{n-n\bar{x}-۱}\qquad ۰<\theta<۱$$

بنابراین اگر $\bar{x}\neq۰$ یا $\bar{x}\neq۱$ آنگاه توزیع پسین سره است و $\theta\,|\,\underset{\sim}{x}\sim Be\left(n\bar{x},n-n\bar{x}\right)$ و برآوردگر بیز تعمیم یافته θ عبارت است از $\delta_G^{\pi}(\underset{\sim}{x}\,)=E\,(\theta\,|\,\underset{\sim}{x}\,)=\frac{n\bar{x}}{n}=\bar{x}$

---

[۱] Proper
[۲] Generalized Bayes

حال اگر $\bar{x}=0$ باشد آنگاه طبق قضیه ۲-۱ برآوردگر بیز آن مقداری از a است که انتگرال زیر را مینیمم کند $$E[(\theta-a)^2 \mid \underset{\sim}{x}] = c(\underset{\sim}{x})\int_0^1 (\theta-a)^2 \theta^{-1}(1-\theta)^{n-1} d\theta$$

انتگرال فوق در صورتی همگرا است که $a=0$ باشد(بخاطر وجود θ^{-1}) پس در این حالت برآورد بیز تعمیم یافته عبارت است از $\delta_G^{\pi}(\underset{\sim}{x}) = a = 0 = \bar{x}$. به همین ترتیب برای $\bar{x}=1$ می‌توان نشان داد که برآورد بیز تعمیم یافته عبارت است از $\delta_G^{\pi}(\underset{\sim}{x}) = a = 1 = \bar{x}$. بنابراین در هر حالت برآورد بیز تعمیم یافته برابر $\delta_G^{\pi}(\underset{\sim}{x}) = \bar{x}$ است.

همچنین توجه کنید که در مثال۲-۲ تحت توزیع پیشین سره $Be(\alpha,\beta)$ برآورد بیز θ عبارت بود از

$$\delta^{\pi}(\underset{\sim}{x}) = \frac{\alpha + n\bar{x}}{\alpha+\beta+n}$$

حال اگر $\alpha \to 0$ و $\beta \to 0$ آنگاه $\delta^{\pi}(\underset{\sim}{x}) \to \bar{x}$ یعنی $\delta^{\pi}(\underset{\sim}{x})$ به سمت برآورد بیز تعمیم یافته θ تحت توزیع پیشین جفریز میل می‌کند. ■

تعریف۲-۳: یک برآوردگر $\delta_L(\underset{\sim}{x})$ را برآوردگر بیز حدی[1] گویند اگر دنباله‌ای از توزیع‌های پیشین π_m و برآوردگرهای بیز $\delta^{\pi_m}(\underset{\sim}{X})$ موجود باشد بطوری که

$$\delta^{\pi_m}(\underset{\sim}{x}) \to \delta(\underset{\sim}{x}) \ \ (a.e.) \ \ as \ \ \ m \to \infty \qquad \blacksquare$$

بنابراین برآوردگر $\delta_L(\underset{\sim}{X}) = \bar{X}$ در توزیع $B(1,\theta)$ یک برآوردگر بیز حدی برای θ است.

مثال ۲-۱۳: اگر $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از $N(\theta,\sigma^2)$ باشند که σ^2 معلوم است، آنگاه تحت توزیع پیشین ناسَره $\pi(\theta)=1$ (جفریز) برآوردگر بیز تعمیم یافته θ برابر $\delta_G^{\pi}(\underset{\sim}{x}) = \bar{x}$ است و تحت توزیع پیشین سره $N(\mu,\tau^2)$ برآوردگر بیز $\delta^{\pi}(\underset{\sim}{x}) = \dfrac{\mu\left(\frac{\sigma^2}{n}\right) + \tau^2\bar{x}}{\frac{\sigma^2}{n}+\tau^2}$ است و در نتیجه برآوردگر $\delta_L(\underset{\sim}{x}) = \bar{x}$ برآوردگر بیز حدی θ موقعی که $\tau \to \infty$ می‌باشد. ■

[1] Limit of Bayes Estimator

توجه: از دیدگاه بیز، برآوردگرهای بیز حدی مطلوبتر از برآوردگرهای بیز تعمیم یافته هستند. زیرا حد دنباله‌های بیز بایستی به یک برآوردگر بیز (سره) نزدیک باشد، اما برآوردگر بیز تعمیم یافته ممکن است به هیچ براوردگر بیزی (سره) نزدیک نباشد (مسئله ۲-۱۵ را ببینید).

مثال ۲-۱٤: فرض کنید $X \sim N(\mu,\sigma^2)$ که μ و σ^2 نامعلوم هستند. توزیع پیشین جفریز را برای $\underset{\sim}{\theta} = (\mu,\sigma)$ به دست آورید.

$$I_{ij}(\underset{\sim}{\theta}) = -E_{\underset{\sim}{\theta}}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j} logf\ (x \mid \underset{\sim}{\theta})\right]$$

$$logf\ (x \mid \underset{\sim}{\theta}) = -\frac{1}{2}\log(2\pi) - \log\sigma - \frac{(X-\mu)^2}{2\sigma^2}$$

$$I(\underset{\sim}{\theta}) = -E_{\underset{\sim}{\theta}}\begin{bmatrix} -\frac{1}{\sigma^2} & \frac{2(\mu - X)}{\sigma^3} \\ \frac{2(\mu-X)}{\sigma^3} & -\frac{1}{\sigma^2} - \frac{3(X-\mu)^2}{\sigma^4} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix}$$

$$\pi(\underset{\sim}{\theta}) \propto |I(\underset{\sim}{\theta})|^{\frac{1}{2}} = \left|\frac{2}{\sigma^4}\right|^{\frac{1}{2}} = \frac{\sqrt{2}}{\sigma^2} \Rightarrow \pi(\underset{\sim}{\theta}) \propto \frac{1}{\sigma^2}$$

که توزیعی ناسره است. ■

مثال ۲-۱۵: فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از $N(0,\sigma^2)$ باشند . برآوردگر بیز σ^2 را براساس توابع زیان $A)\ L(\sigma^2,\delta) = (\delta - \sigma^2)^2 \quad B)\ L(\sigma^2,\delta) = \frac{(\delta-\sigma^2)^2}{\sigma^4}$ بدست آورید. اگر توزیع پیشین جفریز را بکار ببریم، برآوردگر بیز σ^2را تحت توابع زیان فوق به دست آورید.

حل:

الف-

$$f(\underset{\sim}{x} \mid \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2} = c\tau^{+\frac{n}{2}} e^{-\tau y}$$

که در آن $\tau = \frac{1}{2\sigma^2}$ و $y = \sum_{i=1}^{n} x_i^2$ قرار می‌دهیم $\qquad \frac{1}{2\sigma^2} = \tau \sim \Gamma(g, \frac{1}{\alpha})$

پس $E(\tau)=\frac{g}{\alpha}$، $E(\tau^{۲})=\frac{g(g+۱)}{\alpha^{۲}}$ ، $E(\frac{۱}{\tau})=\frac{\alpha}{g-۱}$ و $E(\frac{۱}{\tau^{۲}})=\frac{\alpha^{۲}}{(g-۱)(g-۲)}$

بنابراین $\pi(\tau\mid \underset{\sim}{x})\propto \tau^{\frac{n}{۲}}e^{-\tau y}\tau^{g-۱}e^{-\alpha\tau}=\tau^{g+\frac{n}{۲}-۱}e^{-(\alpha+y)\tau}$

$$\Rightarrow \tau\mid \underset{\sim}{x}\sim \Gamma(g+\frac{n}{۲},\frac{۱}{\alpha+y})$$

$$A)\Rightarrow \delta^{\pi}(\underset{\sim}{x})=E(\sigma^{۲}\mid \underset{\sim}{x})=E(\frac{۱}{۲\tau}\mid \underset{\sim}{x})=\frac{۱}{۲}\frac{\alpha+y}{g+\frac{n}{۲}-۱}=\frac{\alpha+y}{n+۲g-۲}$$

$$B)\Rightarrow \delta^{\pi}(\underset{\sim}{x})=\frac{E(\sigma^{۲}\frac{۱}{\sigma^{۴}}\mid \underset{\sim}{x})}{E(\frac{۱}{\sigma^{۴}}\mid \underset{\sim}{x})}=\frac{E(۲\tau\mid \underset{\sim}{x})}{E(۴\tau^{۲}\mid \underset{\sim}{x})}=\frac{۱}{۲}\frac{(g+\frac{n}{۲})\Big/(\alpha+y)}{(g+\frac{n}{۲})(g+\frac{n}{۲}+۱)\Big/(\alpha+y)^{۲}}$$

$$=\frac{\alpha+y}{n+۲g-۲}$$

ب- $I(\tau)=-E\left[\frac{\partial^{۲}}{\partial\tau^{۲}}(\frac{n}{۲}\ln\tau-\tau y)\right]=-E\left[-\frac{n}{۲\tau^{۲}}\right]=\frac{n}{۲\tau^{۲}}$

$$\pi(\tau)\propto\sqrt{|I(\tau)|}\propto\frac{۱}{\tau}\Rightarrow \qquad \pi(\tau)=\frac{۱}{\tau}\qquad \tau>۰$$

که یک پیشین ناسره است.

$$\pi(\tau\mid \underset{\sim}{x})\propto \tau^{\frac{n}{۲}-۱}e^{-\tau y}\qquad \Rightarrow \tau\mid \underset{\sim}{x}\sim\Gamma(\frac{n}{۲},\frac{۱}{y})$$

$$A)\Rightarrow \delta_G^{\pi}(\underset{\sim}{x})=E(\sigma^{۲}\mid \underset{\sim}{x})=E(\frac{۱}{۲\tau}\mid \underset{\sim}{x})=\frac{۱}{۲}\frac{y}{\frac{n}{۲}-۱}=\frac{y}{n-۲}$$

$$B)\Rightarrow \delta_G^{\pi}(\underset{\sim}{x})=\frac{E(\sigma^{۲}\frac{۱}{\sigma^{۴}}\mid \underset{\sim}{x})}{E(\frac{۱}{\sigma^{۴}}\mid \underset{\sim}{x})}=\frac{E(۲\tau\mid \underset{\sim}{x})}{E(۴\tau^{۲}\mid \underset{\sim}{x})}=\frac{۱}{۲}\frac{(n/۲)/y}{(\frac{n}{۲})(\frac{n}{۲}+۱)/y^{۲}}=\frac{y}{n+۲}$$

توجه کنید که برآوردگرهای بیز تعمیم یافته در این حالت همان براوردگرهای بیز حدی با $\alpha\to۰$ و $g\to۰$ می باشند. ■

مثال ۲-۱۶: فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از $N(\theta, \sigma^2)$ باشند که در آن θ و σ^2 هر دو نامعلوم هستند. اگر $\tau = \frac{1}{2\sigma^2} \sim \Gamma(\alpha, g)$ و $\pi(\theta) = 1; \theta \in R$، برآوردگر بیز σ^2 را تحت توابع زیان A و B مثال قبل و برآوردگر بیز θرا تحت تابع زیان درجه دوم به دست آورید.

حل:

$$f(\underset{\sim}{x} \mid \theta, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2} \propto \tau^{\frac{n}{2}} e^{-\tau\left[\sum(x_i-\bar{x})^2 + n(\bar{x}-\theta)^2\right]}$$

اگر $Z = \sum_{i=1}^{n}(x_i - \bar{x})^2$ آنگاه $\pi(\theta, \tau) \propto \tau^{g-1} e^{-\alpha\tau}$

$$\therefore \pi(\theta, \tau \mid \underset{\sim}{x}) \propto \tau^{\frac{n}{2}} e^{-\tau(Z + n(\bar{x}-\theta)^2)} \tau^{g-1} e^{-\alpha\tau} = \tau^{\frac{n}{2}+g-1} e^{-\tau\left[\alpha + Z + n(\bar{x}-\theta)^2\right]}$$

$$\Rightarrow \pi(\theta, \tau \mid \underset{\sim}{x}) = C\tau^{\frac{n}{2}+g-1} e^{-\tau\left[\alpha + Z + n(\bar{x}-\theta)^2\right]}$$

$$\pi(\tau \mid \underset{\sim}{x}) = \int_{-\infty}^{+\infty} \pi(\theta, \tau \mid \underset{\sim}{x}) d\theta = C^* \tau^{\frac{n-1}{2}+g-1} e^{-\tau(\alpha+Z)}$$

$$\Rightarrow \tau \mid \underset{\sim}{x} \sim \Gamma\left(\frac{n-1}{2} + g, \frac{1}{\alpha + Z}\right)$$

$$(A) \Rightarrow \delta_G^{\pi}(\underset{\sim}{x}) = E\left[\sigma^2 \mid \underset{\sim}{x}\right] = E\left(\frac{1}{2\tau} \mid \underset{\sim}{x}\right) = \frac{1}{2}\frac{\alpha + Z}{\frac{n-1}{2} + g - 1} = \frac{\alpha + Z}{n + 2g - 3}$$

$$(B) \Rightarrow \delta_G^{\pi}(\underset{\sim}{x}) = \frac{E(2\tau \mid \underset{\sim}{x})}{E(4\tau^2 \mid \underset{\sim}{x})} = \frac{1}{2}\frac{\left(\frac{n-1}{2} + g\right) \Big/ (\alpha + Z)}{\left(\frac{n-1}{2} + g\right)\left(\frac{n-1}{2} + g + 1\right) \Big/ (\alpha + Z)^2} = \frac{\alpha + Z}{n + 2g + 1}$$

$$\pi(\theta \mid \underset{\sim}{x}) = \int_{\cdot}^{+\infty} \pi(\theta, \tau \mid \underset{\sim}{x}) d\tau = C \frac{\Gamma\left(\frac{n}{2} + g\right)}{\left[\alpha + Z + n(\bar{x} - \theta)^2\right]^{\frac{n}{2}+g}}$$

این توزیع حول $\bar{x}$ متقارن است پس $\delta_G^{\pi}(\underset{\sim}{x}) = E(\theta \mid \underset{\sim}{x}) = \bar{x}$ ■

فصل ۵

مینیماکس بودن

و مجاز بودن

بخش ۱:تصمیم مینیماکس[۱]

در فصل قبل تصمیم بیز را از طریق مینیمم کردن مخاطره بیز (که همان میانگین تابع مخاطره نسبت به یک توزیع پیشین θ یعنی $\int_{\Theta} R(\theta,\delta)d\Pi(\theta)$ بود) بدست آوردیم.

مزیتی که روش بیز داشت این بود که اطلاعات تابع مخاطره در یک عدد(مخاطره بیز) خلاصه می‌شد و دو تصمیم بوسیله مقایسه این مخاطره بیزشان با یکدیگر مقایسه می‌شدند. در این فصل روش دیگری برای بدست آوردن یک تصمیم بهینه در نظر می‌گیریم که در این روش آن تصمیمی را انتخاب می‌کنیم که ماکسیمم (سوپریمم) تابع مخاطره نسبت به کلیه مقادیر فضای پارامتری Θ را در بین کلیه توابع تصمیم مینیمم کند.

تعریف ۱-۱: فرض کنید که D مجموعه کلیه توابع تصمیم و $R(\theta,\delta)$ تابع مخاطره باشد. تصمیم $\delta_m \in D$ را تصمیم مینیماکس گویند هرگاه برای هر $\delta \in D$ داشته باشیم که

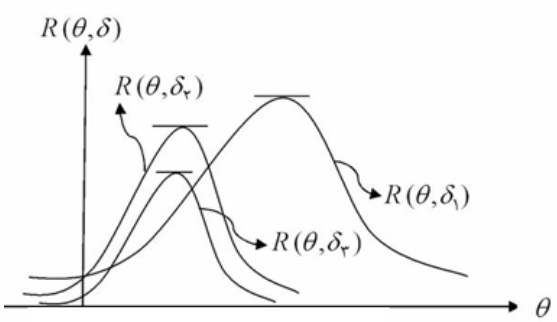
$$\sup_{\theta\in\Theta} R(\theta,\delta_m) \le \sup_{\theta\in\Theta} R(\theta,\delta)$$

و یا به عبارتی $\delta_m \in D$ را تصمیم مینیماکس گویند هرگاه

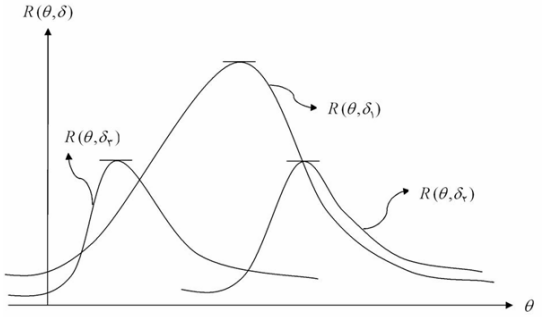
$$\sup_{\theta\in\Theta} R(\theta,\delta_m) = \inf_{\delta\in D} \sup_{\theta\in\Theta} R(\theta,\delta)$$

$$\left(\max_{\theta\in\Theta} R(\theta,\delta_m) = \min_{\delta\in D} \max_{\theta\in\Theta} R(\theta,\delta)\right)$$

به شکل‌های زیر توجه کنید.



δ_3 تصمیم مینیماکس است



δ_2 و δ_3 تصمیم مینیماکس است

[۱] Minimax Decision

با توجه به شکل‌های بالا تصمیم مینیماکس ممکن است یکتا نباشد. تصمیم مینیماکس آن تصمیمی است که در بدترین حالت تابع مخاطره $R(\theta,\delta)$ نسبت به θ، بهترین حالت را انتخاب می‌کند و در حقیقت یک حالت محافظه کارانه را برای انتخاب بهترین تصمیم در پیش می‌گیرد. با این وجود در بیشتر موارد تصمیم‌های معقولانه‌ای را بدست می‌دهد. مثلا در بعضی از مسائل برآورد نقطه‌ای برآورد مینیماکس همان برآورد پایا و UMVU نیز می‌باشد.

تصمیم‌های مینیماکس وابستگی زیادی به تصمیم‌های بیز دارند. یک تصمیم مینیماکس بوسیله مینیمم کردن، ماکسیمم مخاطره می‌خواهد بهترین تصمیم را در بدترین حالت انتخاب کند. بنابراین انتظار می‌رود که تصمیم مینیماکس همان تصمیم بیز با بدترین حالت توزیع پیشین باشد. چنین توزیعی را نامساعدترین توزیع[۱] گویند. به عبارت دقیقتر اینکه اگر π یک توزیع پیشین و δ^{π} تصمیم بیز نسبت به این توزیع پیشین و $r(\pi,\delta^{\pi})$ مخاطره بیز آن باشد، در این صورت توزیع π را نامساعدترین گویند اگر برای هر توزیع پیشین دیگر π' با تصمیم بیز $\delta^{\pi'}$ داشته باشیم که $r(\pi,\delta^{\pi}) \geq r(\pi',\delta^{\pi'})$.

در قضیه زیر شرط مینیماکس بودن یک تصمیم بیز را می‌آوریم.

قضیه ۱-۱: فرض کنید π یک توزیع پیشین روی Θ باشد و $\delta^{\pi}_{\cdot}$ تصمیم بیز نسبت به این توزیع پیشین باشد. اگر تابع مخاطره تصمیم δ^{π} در شرط

$$r(\pi,\delta^{\pi}) = \int R(\theta,\delta^{\pi}) d\Pi(\theta) = \sup_{\theta\in\Theta} R(\theta,\delta^{\pi}) \qquad (۱)$$

صدق کند آنگاه

الف- δ^{π} تصمیم مینیماکس است

ب- اگر δ^{π} تصمیم بیز یکتا نسبت به توزیع پیشین π باشد، آنگاه $\delta^{\pi}_{\cdot}$ تصمیم مینیماکس یکتا است.

ج- توزیع π نامساعدترین توزیع است

اثبات: الف-

$$\sup_{\theta\in\Theta} R(\theta,\delta^{\pi}) \overset{(۱)}{=} \int R(\theta,\delta^{\pi}) d\Pi(\theta) = r(\pi,\delta^{\pi}) \leq r(\pi,\delta) = \int R(\theta,\delta) d\Pi(\theta)$$

[۱] Least Favorable Distribution

$$\leq \int \left[\underset{\theta\in\Theta}{Sup}\, R(\theta,\delta) \right] d\,\Pi(\theta) = \underset{\theta\in\Theta}{Sup}\, R(\theta,\delta)$$

بنابراین δ^{π} تصمیم مینیماکس است

ب- چون δ^{π} تصمیم بیز یکتا است پس $\forall \delta \in D$، $r(\pi,\delta^{\pi}) < r(\pi,\delta)$ و با در نظر گرفتن $<$ بجای $\leq$ در اثبات (الف) نتیجه می‌شود که تصمیم δ^{π} یک تصمیم مینیماکس یکتا است.

ج- فرض کنید π' هر توزیع پیشین دیگر با تصمیم بیز $\delta^{\pi'}$ باشد در این صورت

$$r(\pi',\delta^{\pi'}) = \int R(\theta,\delta^{\pi'}) d\,\Pi'(\theta) \overset{*}{\leq} \int R(\theta,\delta^{\pi}) d\,\Pi'(\theta) \leq Sup R(\theta,\delta^{\pi})$$
$$\overset{(1)}{=} \int R(\theta,\delta^{\pi}) d\,\Pi(\theta) = r(\pi,\delta^{\pi})$$

که در آن * از این واقعیت ناشی می‌گردد که $\delta^{\pi'}$ تصمیم بیز نسبت به π' است. پس توزیع δ^{π} نامساعدترین توزیع است. ■

شرط(۱) بیان می‌کند که میانگین تابع مخاطره $R(\theta,\delta^{\pi})$ با ماکسیمم این تابع برابر باشد یکی از حالت‌هایی که این شرط برقرار می‌شود، این است که تابع مخاطره ثابت باشد، بنابراین:

نتیجه ۱-۱: اگر تصمیم بیز δ^{π} نسبت به توزیع پیشین π دارای مخاطره ثابت باشد، آنگاه δ^{π} تصمیم مینیماکس است.

اثبات:

$$R(\theta,\delta^{\pi}) = C \qquad \forall \theta \in \Theta \Rightarrow \left\{ \begin{array}{l} \underset{\theta\in\Theta}{Sup}\, R(\theta,\delta^{\pi}) = C \\ \int R(\theta,\delta^{\pi}) d\,\Pi(\theta) = C \int d\,\Pi(\theta) = C \end{array} \right\}$$

بنابراین شرط(۱) برقرار است. ■

نتیجه ۱-۲: فرض کنید w_{π} مجموعه نقاطی از فضای پارامتری باشد که تابع مخاطره δ^{π} در روی آن ماکسیمم خود را به دست می‌آورد، یعنی

$$w_{\pi} = \left\{ \theta : R(\theta,\delta^{\pi}) = \underset{\theta'}{Sup}\, R(\theta',\delta^{\pi}) \right\}$$

در این صورت تصمیم بیز δ^{π} یک تصمیم مینیماکس است اگر $\Pi(w_{\pi}) = ۱$.

(به عبارت دیگر یک شرط کافی برای آنکه برآوردگر δ^{π} مینیماکس باشد آن است که یک مجموعه w وجود داشته باشد که $\Pi(w)=1$ و $R(\theta,\delta^{\pi})$ ماکسیمم خود را در تمام نقاط w بدست آورد.)

اثبات: فرض کنید δ هر تصمیم دیگر به غیر از δ^{π} باشد در این صورت

$$\sup_{\Theta} R(\theta,\delta^{\pi}) \overset{(*)}{=} \int_{w_{\pi}} R(\theta,\delta^{\pi})\,d\Pi(\theta) \overset{(**)}{=} \int R(\theta,\delta^{\pi})\,d\Pi(\theta)$$

$$\leq \int R(\theta,\delta)\,d\Pi(\theta) \leq \sup_{\Theta} R(\theta,\delta)$$

که در آن تساوی (*) به این علت است که روی w_{π} مخاطره با سوپریمم برابر می‌شود و تساوی (**) نیز از این مطلب ناشی می‌گردد که $\Pi(w_{\pi}^{c})=0$. پس δ^{π} یک تصمیم مینیماکس است. ■

در حالت برآورد نقطه‌ای تصمیم مینیماکس را برآورد نقطه‌ای مینیماکس گویند.

مثال ۱-۱: فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از توزیع $B(1,\theta)$باشند، تحت تابع زیان درجه دوم برآوردگر مینیماکس θ را بدست آورید.

حل: در مثال ۲-۲ فصل قبل تحت چگالی پیشین $Be(\alpha,\beta)$ برآوردگر یکتای بیز عبارت بود از $\delta^{\pi}(\underset{\sim}{x})=\dfrac{\alpha+\sum x_i}{\alpha+\beta+n}$ که مخاطره آن برابر است با

$$R(\theta,\delta^{\pi}) = E_{\theta}\left[\delta^{\pi}(\underset{\sim}{x})-\theta\right]^2 = V_{\theta}(\delta^{\pi}(\underset{\sim}{X})) + \left\{E[\delta^{\pi}(\underset{\sim}{X})-\theta]\right\}^2$$

$$= V_{\theta}\left(\frac{\alpha+\sum X_i}{\alpha+\beta+n}\right) + \left\{E\left[\frac{\alpha+\sum X_i}{\alpha+\beta+n}-\theta\right]\right\}^2$$

$$= \frac{n\theta(1-\theta)}{(\alpha+\beta+n)^2} + \left(\frac{\alpha+n\theta-\alpha\theta-\beta\theta-n\theta}{\alpha+\beta+n}\right)^2$$

$$= \frac{1}{(\alpha+\beta+n)^2}\left[n\theta(1-\theta)+(\alpha-(\alpha+\beta)\theta)^2\right]$$

$$= \frac{1}{(\alpha+\beta+n)^2}\left[n\theta-n\theta^2+\alpha^2+(\alpha+\beta)^2\theta^2-2\alpha(\alpha+\beta)\theta\right]$$

$$= \frac{1}{(\alpha+\beta+n)^2}\left[\{(\alpha+\beta)^2-n\}\theta^2+\{n-2\alpha(\alpha+\beta)\}\theta+\alpha^2\right]$$

در صورتی $R(\theta,\delta^{\pi})$ به پارامتر θ بستگی ندارد که

$$\begin{matrix} ۲\alpha \\ (\alpha+\beta) \end{matrix}\begin{cases} (\alpha+\beta)^{۲}-n=۰ \\ n-۲\alpha(\alpha+\beta)=۰ \end{cases} \Rightarrow (\alpha+\beta)n-۲\alpha n=۰ \Rightarrow \beta-\alpha=۰ \Rightarrow \alpha=\beta$$

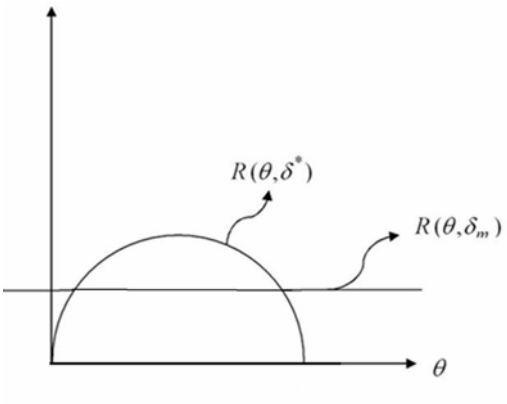
$$\Rightarrow (۲\alpha)^{۲}-n=۰ \Rightarrow \alpha=\frac{\sqrt{n}}{۲}=\beta$$ که با جایگذاری در معادله اول بدست می‌دهد:

بنابراین برآوردگر $\delta_m(X)=\dfrac{\frac{\sqrt{n}}{۲}+\sum X_i}{n+\sqrt{n}}$ برآوردگر یکتای بیز با مخاطره ثابت است و در نتیجه برآوردگر مینیماکس یکتا است.

توجه کنید که برآوردگر UMVU پارامتر θ برابر $\delta^{*}(\underset{\sim}{X})=\bar{X}$ است و به راحتی می‌توان نشان داد که

$$R(\theta,\delta_m)=\frac{۱}{۴(۱+\sqrt{n})^{۲}}, \qquad R(\theta,\delta^{*})=\frac{\theta(۱-\theta)}{n}$$

این دو تابع مخاطره در زیر رسم شده‌اند.



دیده می‌شود که در بیشتر فاصله $\theta\in(۰,۱)$ برآوردگر مینیماکس از برآوردگر UMVU بهتر عمل می‌کند ولی برآوردگر مینیماکس همواره از برآوردگرهای دیگر بهتر نیست. چون

$$\underset{\theta\in(۰,۱)}{Sup}\, R(\theta,\bar{X})=\underset{\theta\in(۰,۱)}{Sup}\,\frac{\theta(۱-\theta)}{n}=\frac{۱}{۴n}>\frac{۱}{۴(۱+\sqrt{n})^{۲}}=\underset{\theta\in(۰,۱)}{Sup}\, R(\theta,\delta_m)$$

بنابراین تحت تابع زیان درجه دوم $\bar{X}$ نمی‌تواند براوردگر مینیماکس باشد. ■

مثال ۱-۲: فرض کنید $X_۱,X_۲,\ldots,X_n$ یک نمونه تصادفی از $B(۱,\theta)$ باشند، تحت تابع زیان $L(\theta,\delta)=\dfrac{(\delta-\theta)^{۲}}{\theta(۱-\theta)}$ و چگالی پیشین $Be(۱,۱)=U(۰,۱)$ برآوردگر بیز پارامتر θ را بدست اورده و نشان دهید که یک برآوردگر مینیماکس است.

حل: ابتدا برآوردگر بیز را بدست می‌آوریم

$$\pi(\theta \mid \underset{\sim}{x}) \propto \pi(\theta) L(\theta) = I_{(۰,۱)}(\theta)\theta^{\sum x_i}(۱-\theta)^{n-\sum x_i}$$

$$\Rightarrow \theta \mid \underset{\sim}{x} \sim Be(n\bar{x}+۱, n-n\bar{x}+۱)$$

$$\delta^{\pi}(\underset{\sim}{x}) = \frac{E\left[\frac{\theta}{\theta(۱-\theta)} \mid \underset{\sim}{x}\right]}{E\left[\frac{۱}{\theta(۱-\theta)} \mid \underset{\sim}{x}\right]} = \frac{\int_{۰}^{۱}\theta^{n\bar{x}}(۱-\theta)^{n-n\bar{x}-۱}d\theta}{\int_{۰}^{۱}\theta^{n\bar{x}-۱}(۱-\theta)^{n-n\bar{x}-۱}d\theta}$$

قرار می دهیم $y = n\bar{x} = \sum_{i=۱}^{n} x_i$ در این صورت اگر $y \neq ۰$ و $y \neq n$ آنگاه

$$\delta^{\pi}(\underset{\sim}{x}) = \frac{Be(n\bar{x}+۱, n-n\bar{x})}{Be(n\bar{x}, n-n\bar{x})} = \frac{\Gamma(n\bar{x}+۱)\Gamma(n-n\bar{x})}{\Gamma(n+۱)} \times \frac{\Gamma(n)}{\Gamma(n\bar{x})\Gamma(n-n\bar{x})} = \frac{n\bar{x}}{n} = \bar{x}$$

اگر $y = ۰$ باشد انگاه طبق قضیه ۲-۱ فصل قبل برآوردگر بیز آن مقداری از a است که انتگرال زیر را مینیمم کند

$$E\left[\frac{(\theta-a)^{۲}}{\theta(۱-\theta)} \mid \underset{\sim}{x}\right] = \int_{۰}^{۱}\frac{(\theta-a)^{۲}}{\theta(۱-\theta)}(۱-\theta)^{n}d\theta = \int_{۰}^{۱}(\theta-a)^{۲}\theta^{-۱}(۱-\theta)^{n-۱}d\theta$$

و این انتگرال بخاطر وجود $\theta^{-۱}$ نامتناهی می‌شود بجز اینکه $a = ۰$ باشد. بنابراین در این حالت برآورد بیز عبارت است از $\delta^{\pi}(\underset{\sim}{x}) = a = ۰ = \frac{۰}{n} = \frac{y}{n} = \bar{x}$. به همین ترتیب برای $y = n$ می‌توان نشان داد که برآورد بیز عبارت است از $\delta^{\pi}(\underset{\sim}{x}) = a = ۱ = \frac{n}{n} = \frac{y}{n} = \bar{x}$ بنابراین در هر حالت $\delta^{\pi}(\underset{\sim}{x}) = \bar{x}$ برآورد بیز است . همچنین

$$R(\theta, \delta^{\pi}) = E\left[\frac{(\bar{X}-\theta)^{۲}}{\theta(۱-\theta)}\right] = \frac{۱}{\theta(۱-\theta)}V_{\theta}(\bar{X}) = \frac{۱}{\theta(۱-\theta)}\frac{\theta(۱-\theta)}{n} = \frac{۱}{n}$$

بنابراین $\delta^{\pi}(\underset{\sim}{X}) = \bar{X}$ یک برآوردگر بیز با مخاطره ثابت است و در نتیجه یک برآوردگر مینیماکس است. ■

لم ۱-۱: فرض کنید δ یک برآوردگر بیز (یا UMVU یا مینیماکس یا مجاز) برای $g(\theta)$ تحت تابع زیان درجه دوم باشد. در این‌صورت $a\delta+b$ یک برآوردگر بیز (یا UMVU یا مینیماکس یا مجاز) برای $ag(\theta)+b$ است

اثبات: چون $$R(ag(\theta)+b, a\delta+b) = a^2 R(g(\theta),\delta)$$

بنابراین نتیجه به راحتی حاصل می‌گردد. ■

مثال ۱-۳: فرض کنید $X \sim B(n,p_1)$، $Y \sim B(n,p_2)$ و X و Y از یکدیگر مستقل باشند. با استفاده از نتیجه ۱-۲ یک برآوردگر مینیماکس برای p_2-p_1 به دست آورید.

حل: با توجه به مثال ۱-۱ به دنبال برآوردگرهایی به فرم $C(Y-X)$ (این برآوردگر تحت یک تبدیل خاص پایا است) هستیم که مینیماکس باشند. تابع مخاطره‌ی این برآوردگرها عبارتند از

$$\begin{aligned} R\left(C(Y-X), p_1, p_2\right) &= E\left[C(Y-X)-(p_2-p_1)\right]^2 \\ &= V\left(C(Y-X)\right)+\left[E\left(C(Y-X)\right)-(p_2-p_1)\right]^2 \\ &= C^2 n\left(p_1(1-p_1)+p_2(1-p_2)\right)+(Cn-1)^2(p_2-p_1)^2 \end{aligned}$$

می‌خواهیم مجموعه‌ای از نقاط w را پیدا کنیم که تابع مخاطره فوق در این نقاط ماکسیمم باشد. بنابراین از تابع فوق نسبت به p_1 و p_2 مشتق جزئی می گیریم.

$$\begin{aligned} &\frac{\partial R}{\partial P_1} = 0 \Rightarrow \left[2(Cn-1)^2 - 2C^2 n\right]p_1 - 2(Cn-1)^2 p_2 = -C^2 n \\ &\frac{\partial R}{\partial P_2} = 0 \Rightarrow -2(Cn-1)^2 p_1 + \left[2(Cn-1)^2 - 2C^2 n\right]p_2 = -C^2 n \end{aligned} \qquad (1)$$

اگر این دستگاه ۲ معادله و ۲ مجهول ترکیب خطی از یکدیگر نباشند آنگاه این دستگاه تنها دارای یک جواب (p_1^o, p_2^o) است و بنابراین π احتمال ۱ را به نقطه (p_1^o, p_2^o) بایستی نسبت دهد و برآوردگر بیز نسبت به این پیشین برآوردگر $\delta(X,Y) = p_2^o - p_1^o$ است که دارای ماکسیمم مخاطره در نقطه (p_1^o, p_2^o) نیست. بنابراین در این حالت نمی‌توان از نتیجه ۱-۲ استفاده کرد.

حال اگر دو معادله‌ی دستگاه فوق یک ترکیب خطی از یکدیگر باشند آنگاه

$$2(Cn-1)^2 - 2C^2 n = 2(Cn-1)^2 \Rightarrow C^2 n = 2(Cn-1)^2 \Rightarrow C = \frac{\sqrt{2n}}{n[\sqrt{2n}\pm 1]}$$

چون $-۱ < p_۲ - p_۱ < ۱$ می‌باشد پس حالت منفی در مخرج که برآوردگر خارج از فضای پارامتری را می‌دهد قابل قبول نیست. با قرار دادن این مقدار C در معادله (۱) به معادله $p_۱ + p_۲ = ۱$ می‌رسیم و بنابراین برآوردگر

$$\delta(X,Y) = \frac{\sqrt{۲n}}{n(\sqrt{۲n}+۱)}(Y - X)$$

روی مجموعه $w = \{(p_۱, p_۲) \mid p_۱ + p_۲ = ۱\}$ دارای مخاطره ماکسیمم است. حال اگر نشان دهیم که $\delta(x,y)$ برآوردگر بیز $p_۲ - p_۱$ نسبت به پیشینی است که احتمال ۱ به مجموعه w نسبت می‌دهد آنگاه طبق نتیجه ۱-۲، $\delta(x,y)$ برآوردگر مینیماکس است. اما $p_۱ + p_۲ = ۱$ نتیجه می دهد که $p_۲ - p_۱ = ۲p_۲ - ۱$ پس طبق لم ۱-۱ کافی است برآوردگر بیز $p_۲$ را نسبت به این پیشین به دست آوریم. حال چون $p_۲ = ۱ - p_۱$ بنابراین برای برآورد $p_۲$ بر اساس n آزمایش با احتمال موفقیت $p_۲$ (بر اساس Y) و n آزمایش با احتمال موفقیت $p_۲ = ۱ - p_۱$ (بر اساس $n - X$) برآوردگر یکتای بیز طبق مثال ۱-۱ عبارت است از:

$$\frac{[Y + (n - X)] + \frac{\sqrt{۲n}}{۲}}{۲n + \sqrt{۲n}}$$

پس برآوردگر یکتای بیز $p_۲ - p_۱ = ۲p_۲ - ۱$ طبق لم ۱-۱ عبارت است از

$$۲\left\{\frac{Y + (n - X) + \frac{\sqrt{۲n}}{۲}}{۲n + \sqrt{۲n}}\right\} - ۱ = \frac{[۲(Y - X) + ۲n + \sqrt{۲n}] - [۲n + \sqrt{۲n}]}{\sqrt{۲n}(\sqrt{۲n}+۱)}$$

$$= \frac{\sqrt{۲n}}{n(\sqrt{۲n}+۱)}(Y - X) = \delta(X,Y)$$

پس $\delta(X,Y)$ برآوردگر مینیماکس یکتای $p_۲ - p_۱$ است. ■

نکات: ۱) توجه کنید که قضیه ۱-۱ می‌گوید که اگر برای یک برآوردگر بیز شرط $\int R(\theta, \delta^{\pi}) d\Pi(\theta) = \underset{\theta \in \Theta}{Sup}\, R(\theta, \delta^{\pi})$ برقرار باشد آنگاه آن برآوردگر بیز، مینیماکس است. یکی از راه‌های برقراری شرط فوق ثابت بودن مخاطره δ^{π} است که در نتیجه ۱-۱ و مثال‌های قبل مورد

بررسی قرار گرفت. اما مثال‌هایی نیز وجود دارد که در آنها مخاطره برآوردگر بیز ثابت نیست اما شرط بالا برقرار می‌باشد و در نتیجه برآودرگر بیز حاصله مینیماکس خواهد بود. به تمرین زیر توجه کنید.

تمرین: فرض کنید $X \sim N(\theta,1)$ که $\theta \in [-m,m]$ و m عددی ثابت و $m > 1 > 0$ اگر

$$\delta^m(x) = mtgh(mx) = m\frac{e^{mx} - e^{-mx}}{e^{mx} + e^{-mx}}$$

الف- نشان دهید که $\delta^m(x)$ برآورد بیز پارامتر θ تحت توزیع پیشین زیر است.

$$.\pi(\theta) = \begin{cases} \frac{1}{2} & \theta = m \\ \frac{1}{2} & \theta = -m \end{cases}$$

ب- مخاطره ی بیز δ^m را محاسبه کنید و نشان دهید که $r(\pi,\delta^m) = R(m,\delta^m)$

ج- نشان دهید که $\underset{-m\le\theta\le m}{Sup} R(\theta,\delta^m) = R(m,\delta^m)$ و بنابراین طبق شرط قضیه ۱-۱ δ^m برآوردگر مینیماکس θ است.

حل تمرین:

$$\pi(\theta \mid \underset{\sim}{x}) = \frac{\pi(\theta)L(\theta)}{m(x)} = \frac{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\theta)^2}\pi(\theta)}{m(\underset{\sim}{x})}$$

(الف

$$m(\underset{\sim}{x}) = \sum_{\theta=-m,m}\pi(\theta)L(\theta) = \frac{1}{\sqrt{2\theta}}\frac{1}{2}\left\{e^{-\frac{1}{2}(x-m)^2} + e^{-\frac{1}{2}(x+m)^2}\right\}$$

$$\therefore \pi(\theta \mid \underset{\sim}{x}) = \begin{cases} \frac{e^{-\frac{1}{2}(x-m)^2}}{e^{-\frac{1}{2}(x-m)^2} + e^{-\frac{1}{2}(x+m)^2}} & \theta = m \\ \frac{e^{-\frac{1}{2}(x+m)^2}}{e^{-\frac{1}{2}(x-m)^2} + e^{-\frac{1}{2}(x+m)^2}} & \theta = -m \end{cases} = \begin{cases} \frac{e^{mx}}{e^{mx} + e^{-mx}} & \theta = m \\ \frac{e^{-mx}}{e^{mx} + e^{-mx}} & \theta = -m \end{cases}$$

$$\delta^m(\underset{\sim}{x}) = E[\theta \mid \underset{\sim}{x}] = m\frac{e^{mx}}{e^{mx} + e^{-mx}} - m\frac{e^{-mx}}{e^{mx} + e^{-mx}} = mtgh(mx)$$

ب) $r(\pi,\delta^m)=\frac{1}{2}\sum_{\theta=-m,m} R(\theta,\delta^m)=\frac{1}{2}R(m,\delta^m)+\frac{1}{2}R(-m,\delta^m)$

$$=\frac{1}{2}E_{\theta=m}\left[(mtgh(mX)-m)^2\right]+\frac{1}{2}E_{\theta=-m}\left[(mtgh(mX)-m)^2\right]$$

$$=\frac{1}{2}E_{\theta=m}\left[\frac{4m^2e^{-2mX}}{(e^{mX}+e^{-mX})^2}\right]+\frac{1}{2}E_{\theta=-m}\left[\frac{4m^2e^{+2mX}}{(e^{mX}+e^{-mX})^2}\right]$$

$$g_n{}^{(x)}=\frac{e^{-2nx}}{(e^{nx}+e^{-nx})^2}, X\sim N(n,1) \qquad ,\gamma=E_{\theta=n}\left[g_n(x)\right]$$

$$\therefore r(\pi,\delta^m)=2m^2\left\{E_{\theta=m}[g_m(X)]+E_{\theta=-m}[g_{-m}(X)]\right\}$$

$$=4m^2E_{\theta=m}[g_m(X)]=4m^2E\left[\frac{e^{-2mX}}{(e^{mX}+e^{-mX})^2}\right]=R(m,\delta^m)$$

ج) $R(\theta,\delta^m)=E_\theta\left[(mtgh(mX)-\theta)^2\right]\le R(m,\delta^m)$

که نامساوی آخر از این واقعیت ناشی می‌گردد که این نامعادله‌ی درجه دوم بر حسب θ در فاصله‌ی $[-m,m]$، مینیمم خود را در $\theta=E\left[mtgh(mX)\right]$ اختیار می‌کند.

۲) از قضیه ۱-۱ موقعی می‌توان استفاده کرد که توزیع پیشین سره باشد. حال اگر توزیع پیشین ناسره باشد به دو صورت زیر می‌توان قضیه ۱-۱ را تعمیم داد.

الف- اگر توزیع پیشین ناسره باشد آنگاه ممکن است توزیع پسین سره باشد و بتوان برآوردگر بیز تعمیم یافته را به دست آورد و امیدوار بود که این برآوردگر مینیماکس باشد (یعنی مینیماکس بودن آن مورد بررسی قرار گیرد).

ب- دنباله‌ای از توزیع‌های پیشین ناسره را یافت که به سمت توزیع پیشین سره مورد نظر میل کنند و از روش بیز حدی که در زیرآورد می‌شود برآوردگر مینیماکس را به دست آورد.

تعریف ۱-۲: دنباله‌ای از توزیع‌های پیشین $\{\pi_n\}$را نامساعدترین گویند اگر برای هر توزیع پیشین π داشته باشیم که

$$r_\pi \le r=\lim_{n\to\infty} r_{\pi_n} \qquad (1)$$

که در آن $r_n=r_{\pi_n}=r(\pi_n,\delta^{\pi_n})=\int R(\theta,\delta^{\pi_n})d\Pi_n(\theta)$ مخاطره‌ی بیز نسبت به π_n است.

قضیه ۱-۲ (روش بیز حدی): فرض کنید $\pi_1,\pi_2,\pi_3,...$ دنباله‌ای از توزیع‌های پیشین باشند و δ^{π_n} تصمیم بیز نسبت به توزیع پیشین π_n با مخاطره بیز $r_n = r(\pi_n,\delta^{\pi_n}) = \int R(\theta,\delta^{\pi_n})d\Pi_n(\theta)$ باشد. اگر $\lim_{n\to\infty} r_n = r$ و δ_m تصمیمی‌باشد که $\sup_{\theta\in\Theta} R(\theta,\delta_m) = r$ آنگاه

الف) δ_m یک تصمیم مینیماکس است.

ب) دنباله‌ی $\{\pi_n\}$ نامساعدترین است.

اثبات: الف) برای هر تصمیم δ داریم که

$$r_n = r(\pi_n,\delta^{\pi_n}) = \int R(\theta,\delta^{\pi_n})d\Pi_n(\theta) \underset{(*)}{\leq} \int R(\theta,\delta)d\Pi_n(\theta) \leq \sup_{\theta\in\Theta} R(\theta,\delta)$$

$$\Rightarrow \lim_{n\to\infty} r_n \leq \sup_{\theta\in\Theta} R(\theta,\delta)$$

که نامساوی (*) از این مطلب بدست می‌آید که δ^{π_n} تصمیم بیز نسبت به π_n است.

بنابراین: $$\sup_{\theta\in\Theta} R(\theta,\delta_m) = r = \lim_{n\to\infty} r_n \leq \sup_{\theta\in\Theta} R(\theta,\delta) \qquad \forall\delta$$

یعنی δ_m یک تصمیم مینیماکس است.

ب) فرض کنید π هر توزیع پیشین باشد.

$$r_n = \int_\Theta R(\theta,\delta^{\pi})d\Pi(\theta) \leq \int_\Theta R(\theta,\delta)d\Pi(\theta) \leq \sup_{\theta\in\Theta} R(\theta,\delta) = r = \lim_{n\to\infty} r_n$$

پس دنباله‌ی $\{\pi_n\}$ نامساعدترین است. ■

نکات: ۱) در قسمت الف قضیه فوق کافی است داشته باشیم که

$$\sup_{\Theta} R(\theta,\delta) = r \leq \lim_{n\to\infty} r_n \qquad (۱)$$

۲) توجه کنید که اگر $\delta^{\pi_n} \to \delta$ آنگاه دلیلی ندارد که $r_{\pi_n} \to r$ موقعی که $n \to \infty$. برای مشاهده یک مثال مشخص به تمرین زیر توجه کنید.

تمرین: فرض کنید $X_1,X_2,....,X_n$ یک نمونه تصادفی از $N(\theta,1)$ باشند و $\theta \sim N(b\sigma^2,\sigma^2)$ که $b \neq 0$. تحت تابع زیان درجه دوم، برآوردگر بیز δ^{π_σ} و مخاطره‌ی بیز آن یعنی $r_{\pi_\sigma} = r_\sigma$ را به دست آورید و نشان دهید که اگر $\sigma \to \infty$ آنگاه $\delta_\sigma \to \delta$ که δ برآوردگری با مخاطره ثابت r است، اما $r_\sigma \not\to r$.

$$\delta_\sigma = \frac{n\overline{x}+b}{n+\frac{۱}{\sigma^۲}} \to \overline{x}+\frac{b}{n} \quad , \quad R(\overline{X}+\frac{b}{n},\theta) = \frac{۱}{n}+\frac{b^۲}{n} = r$$

■ $$r_\sigma = \frac{۱}{n+\frac{۱}{\sigma^۲}} \to \frac{۱}{n} \neq \frac{۱}{n}+\frac{b^۲}{n}$$

۳) قضیه ۱-۲ دارای مطلوبیت کمتری نسبت به قضیه ۱-۱ است زیرا

الف- در قضیه ۱-۲ اگر δ^{π_n} برآوردگر بیز یکتا باشد آنگاه دلیلی ندارد که δ_m برآوردگر مینیماکس یکتا باشد.

ب- برای استفاده از قضیه ۱-۲ ناگزیر هستیم که r و مخاطره ی بیز r_{π_n} را محاسبه کنیم که در بعضی موارد مشکل است. در بعضی موارد می‌توان برای محاسبه r_{π_n} از لم زیر استفاده کرد.

با وجود مشکلات فوق در بسیاری موارد ناگزیر هستیم که از این قضیه یعنی روش بیز حدی استفاده کنیم.

لم۱-۲: اگر δ^π برآوردگر بیز پارامتر $\gamma(\theta)$ نسبت به توزیع پیشین π باشد و اگر $r_\pi = E\left[\left(\delta^\pi(\underset{\sim}{X})-\gamma(\theta)\right)^۲\right]$ مخاطره‌بیز آن باشد، آنگاه $r_\pi = \int Var[\gamma(\theta)\,|\,\underset{\sim}{x}]dP(\underset{\sim}{x})$ و اگر واریانس پسین به $\underset{\sim}{x}$ بستگی نداشته باشد آنگاه $r_\pi = Var[\gamma(\theta)\,|\,\underset{\sim}{x}]$.

اثبات: $$r_\pi = E[(\delta_\pi(\underset{\sim}{x})-\gamma(\theta))^۲] = \int\left\{E\left[\left(\gamma(\theta)-\delta^\pi(\underset{\sim}{x})\right)^۲\,|\,\underset{\sim}{x}\right]\right\}dP(\underset{\sim}{x})$$

■ $$= \int Var(\gamma(\theta)\,|\,\underset{\sim}{x})dP(\underset{\sim}{x})$$

مثال ۱-٤: فرض کنید $X_۱, X_۲, \ldots, X_n$ یک نمونه تصادفی از $N(\theta,\sigma^۲)$ باشد که در آن $\sigma^۲$ معلوم است. تحت تابع زیان درجه دوم نشان دهید که $\overline{X}$ برآوردگر مینیماکس پارامتر θ است.

حل: دنباله توزیع‌های پیشین $N(۰,m^۲),\ m = ۱,۲,۳,\ldots$ را در نظر بگیرید، در این‌صورت

$$\theta\,|\,\underset{\sim}{x} \sim N\left(\frac{m^۲\overline{x}}{m^۲+\frac{\sigma^۲}{n}}, \frac{۱}{\frac{n}{\sigma^۲}+\frac{۱}{m^۲}}\right) \qquad \delta^{\pi_m}(\underset{\sim}{x}) = \frac{m^۲\overline{x}}{m^۲+\frac{\sigma^۲}{n}} \overset{m\to\infty}{\to} \overline{x}$$

مخاطره بیز δ^{π_m} عبارت است از

$$r_m = r(\pi_m, \delta^{\pi_m}) = \int V(\theta \mid \underset{\sim}{x}) dP(\underset{\sim}{x}) = \int \frac{۱}{\frac{n}{\sigma^۲} + \frac{۱}{m^۲}} dP(\underset{\sim}{x}) = \frac{۱}{\frac{n}{\sigma^۲} + \frac{۱}{m^۲}}$$

$$\lim_{m\to\infty} r_m = \lim_{m\to\infty} \frac{۱}{\frac{n}{\sigma^۲} + \frac{۱}{m^۲}} = \frac{\sigma^۲}{n} = R(\theta, \bar{x}) = \sup_{\Theta} R(\theta, \bar{x})$$

بنابراین طبق قضیه ۱-۲، $\bar{X}$ برآوردگر مینیماکس پارامتر θ است. ■

با توجه به مثال قبل برای هر $\sigma^۲$ معلوم داریم که

$$\sup_{\theta} R(\theta, \sigma^۲, \bar{X}) \le \sup_{\theta} R(\theta, \sigma^۲, \delta) \qquad \forall \delta$$

$$\Rightarrow \sup_{\sigma^۲} \sup_{\theta} R(\theta, \sigma^۲, \bar{X}) \le \sup_{\sigma^۲} \sup_{\theta} R(\theta, \sigma^۲, \delta) \qquad \forall \delta$$

حال اگر $\sigma^۲$ نامعلوم باشد از رابطه فوق نتیجه می‌شود که

$$\infty = \sup_{\sigma^۲} \frac{\sigma^۲}{n} = \sup_{\sigma^۲} [\sup_{\theta} R(\theta, \sigma^۲, \bar{X})] \le \sup_{\sigma^۲} \sup_{\theta} R(\theta, \sigma^۲, \delta) \qquad \forall \delta$$

یعنی مخاطره تمام برآوردگرها نامتناهی است بجز اینکه $\sigma^۲$متناهی باشد. بنابراین فرض می‌کنیم $\sigma^۲$ نامعلوم و $\sigma^۲ \le M$ باشد در این صورت $\sup_{\sigma^۲} \sup_{\theta} R(\theta, \sigma^۲, \bar{X}) = \frac{M}{n}$. در این حالت مینیماکس بودن برآوردگر $\bar{X}$ از قضیه زیر نتیجه می‌شود.

قضیه۱-۳: فرض کنید X یک متغیر تصادفی با توزیع F از خانواده $\mathcal{F}_۱$ و $g(F)$ پارامتر مورد نظر باشد و $\mathcal{F}_۰ \subset \mathcal{F}_۱$. اگر δ_o برآوردگر مینیماکس $g(F)$ در $\mathcal{F}_۰$ باشد و

$$\sup_{F\in\mathcal{F}_۰} R(F, \delta_o) = \sup_{F\in\mathcal{F}_۱} R(F, \delta_o)$$

آنگاه δ_o برآوردگر مینیماکس $g(F)$ در $\mathcal{F}_۱$ است.

اثبات: برای هر برآوردگر δ داریم که

$$\sup_{F\in\mathcal{F}_۱} R(F, \delta) \ge \sup_{F\in\mathcal{F}_۰} R(F, \delta) \overset{(*)}{\ge} \sup_{F\in\mathcal{F}_۰} R(F, \delta_o) \overset{(**)}{=} \sup_{F\in\mathcal{F}_۱} R(F, \delta_o)$$

که در آن نامساوی (*) از مینیماکس بودن δ_o در F_0 و تساوی (**) از فرض حاصل می‌گردد. پس δ_o در F_1 مینیماکس است. ■

اگر قرار دهیم

$$F_0 = \{N(\theta, M) : \theta \in R, M < \infty\} \quad و \quad F_1 = \{N(\theta, \sigma^2) : \theta \in R, \sigma^2 \le M < \infty\}$$

آنگاه $F_0 \subset F_1$ و $\sup_{F_0} R(\theta, \sigma^2, \bar{X}) = \frac{M}{n} = \sup_{F_1} R(\theta, \sigma^2, \bar{X})$. بنابراین طبق قضیه ۱-۳ کافی‌است نشان‌دهیم که $\bar{X}$ در F_0 مینیماکس است و در نتیجه $\bar{X}$ در F_1 مینیماکس خواهد بود. اما تحت F_0

$$r_m = \frac{1}{\frac{n}{M} + \frac{1}{m^2}} \xrightarrow{m \to \infty} \frac{M}{n} = R(\theta, \bar{X}) = \sup_{\theta} R(\theta, \bar{X})$$

پس طبق قضیه ۱-۲ $\bar{X}$ در F_0 مینیماکس است. ■

مثال ۱-۵ (ناپارامتری): فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی با توزیع یکسان F که دارای میانگین متناهی θ است، باشد. برآوردگر مینیماکس پارامتر θ تحت تابع زیان درجه دوم را در هر یک از حالات زیر به دست آورید.

الف- واریانس کراندار $V_F(X_i) \le M < \infty$

ب- حدود کراندار $-\infty < a < X_i < b < +\infty$

(توجه کنید اگر ماکسیمم مخاطره هر برآوردگر بی‌نهایت شود آنگاه مسئله مینیماکس منتفی است پس شرایط فوق را برای گریز از این موضوع در مسئله اعمال کرده ایم)

حل : الف- اگر قرار دهیم $F_0 = \left\{N(\theta, \sigma^2) \mid \sigma^2 \le M < \infty\right\}$

$$F_1 = \{F \mid E_F(X_i) = \theta,\ V_F(X_i) \le M < \infty\}$$

در این صورت $F_0 \subset F_1$ و طبق بحث قبل از این مثال $\bar{X}$ در F_0 برآوردگری مینیماکس است و

$$\begin{cases} \underset{F_o}{Sup}\, R(\theta,\sigma^2,\bar{X}) = \dfrac{M}{n} \\ \underset{F_1}{Sup}\, R(\theta,\sigma^2,\bar{X}) = \underset{F_1}{Sup}\; V(\bar{X}) = \underset{F_1}{Sup}\, \dfrac{\sigma^2}{n} = \dfrac{M}{n} \end{cases}$$

پس شرایط قضیه ۱-۳ برقرار است، یعنی $\bar{X}$ در F_1 مینیماکس است.

ب- بدون از دست دادن کلیت مسئله فرض می کنیم $b=1$ و $a=0$، قرار می دهیم

$$F_0 = \{B(1,\theta) \mid 0<\theta<1\} \qquad F_1 = \{F \mid F(1)-F(0)=1\}$$

پس $F_0 \subset F_1$ و در مثال ۱-۱ دیدیم که

$$\delta_o(\underset{\sim}{X}) = \frac{\sqrt{\frac{n}{2}} + \sum_{i=1}^{n} X_i}{n+\sqrt{n}} = \frac{\sqrt{n}}{1+\sqrt{n}}\bar{X} + \frac{1}{2(1+\sqrt{n})}$$

برآوردگر مینیماکس در F_0 می‌باشد. حال طبق قضیه ۱-۳ کافی است نشان دهیم که ماکسیمم مخاطره δ_o در F_0 و F_1 یکی است.

$$F_0 \text{ تحت } : R(\theta,\delta_o) = \frac{1}{4(1+\sqrt{n})^2} \Rightarrow \underset{F_0}{Sup}\, R(\theta,\delta_o) = \frac{1}{4(1+\sqrt{n})^2}$$

$$F_1 \text{ تحت } : R(\theta,\delta_o) = E\left[\frac{\sqrt{n}}{1+\sqrt{n}}\bar{X} + \frac{1}{2(1+\sqrt{n})} - \theta\right]^2$$

$$= \frac{n}{(1+\sqrt{n})^2} V(\bar{X}) + \left(\frac{\sqrt{n}}{1+\sqrt{n}}\theta + \frac{1}{2(1+\sqrt{n})} - \theta\right)^2$$

$$= \frac{1}{(1+\sqrt{n})^2}\left[V_F(X_1) + \left(\frac{1}{2}-\theta\right)^2\right]$$

$$V_F(X_1) = E(X-\theta)^2 = E(X^2) - \theta^2 \le E(X) - \theta^2 = \theta - \theta^2$$

زیرا $0 \le X \le 1 \Rightarrow X^2 \le X$

$$\therefore \quad R(\delta_o,\theta) \le \frac{1}{(1+\sqrt{n})^2}\left[(\theta-\theta^2) + \left(\frac{1}{2}-\theta\right)^2\right] = \frac{1}{4(1+\sqrt{n})^2}$$

$$\Rightarrow \underset{F_1}{Sup}\, R(\theta,\delta_o) = \frac{1}{4(1+\sqrt{n})^2} = \underset{F_0}{Sup}\, R(\theta,\delta_o)$$

پس طبق قضیه ۱-۳، $\delta_o(\underset{\sim}{X})$ در F_1 برای θ مینیماکس است. ■

بخش ۲ : تصمیم‌های مجاز[۱] (پذیرفتنی)

در یک مسئله تصمیم کلاس توابع تصمیم D معمولاً کلاس بزرگی است و انتخاب اینکه کدام تصمیم $\delta \in D$ را مورد استفاده قرار دهیم کار مشکلی می باشد. بنابراین بایستی یک زیر کلاس C از D پیدا کنیم که در این کلاس کلیه تصمیم‌های خوب قرار داشته باشند. حال چون این کلاس C یک کلاس کوچکتر است، پیدا کردن یک تصمیم خوب در آن راحت‌تر می باشد. در این قسمت کلاسی از توابع تصمیم خوب را با استفاده از محک مجاز بودن[۲] تعریف می‌کنیم و نحوه‌ی بدست آوردن این تصمیم‌ها را ارائه می‌دهیم.

مقایسه تصمیم‌ها:

تعریف ۲-۱ : فرض کنید δ_1 و δ_2 دو تصمیم با توابع مخاطره $R(\theta,\delta_1)$ و $R(\theta,\delta_2)$ باشند.

الف- تصمیم δ_1را به خوبی[۳] δ_2گویند اگر: $R(\theta,\delta_1) \le R(\theta,\delta_2) \quad \forall \theta \in \Theta$

ب- تصمیم δ_1 را بهتر از[۴] δ_2 گویند اگر : $R(\theta,\delta_1) \le R(\theta,\delta_2) \quad \forall \theta \in \Theta$

و برای بعضی $\theta \in \Theta$ داشته باشیم که: $R(\theta,\delta_1) < R(\theta,\delta_2).$

ج – تصمیم δ_1 را هم ارز[۵] δ_2 گویند اگر: $R(\theta,\delta_1) = R(\theta,\delta_2) \quad \forall \theta \in \Theta$

د- تصمیم‌های δ_1 و δ_2را غیرقابل مقایسه گویند هرگاه هیچکدام از سه حالت بالا برقرار نباشد.

مثال ۲-۱: فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از $N(\mu,\sigma^2)$ باشد. می دانیم که S^2 برآوردگر UMVU برای پارامتر σ^2 است. نشان دهید که در کلاس برآوردگرهای cS^2 برآوردگری بهتر از S^2 وجود دارد.

حل: در فصل مربوط به برآوردگرهای ناریب نشان دادیم که برآوردگر $\widetilde{S^2} = \frac{n-1}{n+1}S^2 = \frac{1}{n+1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ که یک برآوردگر MRE است دارای مخاطره کمتری

---

[۱] Admissible Decisions
[۲] Admissibility
[۳] As Good As
[۴] Better Than
[۵] Equivalent

نسبت به S^2 است یعنی $\forall \sigma^2 \quad R(\sigma^2,\widetilde{S^2}) < R(\sigma^2,S^2),$ بنابراین برآوردگر $\widetilde{S^2}$ در کلاس cS^2 بهترین برآوردگر است. ■

مثال ۲-۲: فرض کنید X_1,X_2 نمونه ای تصادفی از $N(\mu,\sigma^2)$ باشند برآوردگرهای زیر را برای پارامتر μ در نظر بگیرید.

$$\delta_1 = \frac{1}{2}X_1 + \frac{1}{2}X_2 \ , \ \delta_2 = \frac{1}{4}X_1 + \frac{3}{4}X_2 \ , \ \delta_3 = \frac{3}{4}X_1 + \frac{1}{4}X_2 \ , \ \delta_4 = \frac{1}{2}X_1 + \frac{1}{2}X_2 + 1$$

اگر تابع زیان درجه دوم باشد این ٤ برآوردگر را با یکدیگر مقایسه کنید.

حل:

$$R(\mu,\delta_1) = V(\delta_1) = \frac{1}{2}\sigma^2 \qquad R(\mu,\delta_2) = V(\delta_2) = (\frac{1}{16}+\frac{9}{16})\sigma^2 = \frac{5}{8}\sigma^2$$

$$R(\mu,\delta_3) = V(\delta_3) = (\frac{9}{16}+\frac{1}{16})\sigma^2 = \frac{5}{8}\sigma^2 \qquad R(\mu,\delta_4) = V(\delta_4) = \frac{1}{2}\sigma^2 + 1$$

بنابراین δ_1 بهتر از δ_2 و δ_3 و δ_4 است و δ_2 هم ارز δ_3 است و δ_4 با δ_2 و δ_3 قابل مقایسه نیست زیرا اگر $\sigma^2 \geq 8$ آنگاه $R(\mu,\delta_4) \leq R(\mu,\delta_2)$ و اگر $\sigma^2 < 8$ آنگاه $R(\mu,\delta_4) > R(\mu,\delta_2)$. ■

تعریف ۲-۱ مقایسه‌ای را بین تمام توابع تصمیم بوجود نمی‌آورد و تنها یک ترتیب را روی کلاس توابع تصمیم D تعریف می کند.

تعریف ۲-۲: یک تصمیم δ را مجاز[1] گویند هرگاه تصمیم دیگری مانند $\delta' \in D$ وجود نداشته باشد که از δ بهتر باشد. یک تصمیم δ را غیر مجاز[2] گویند اگر تصمیم دیگری مانند $\delta' \in D$ وجود داشته باشد که از δ بهتر باشد. ■

با توجه به مثال ۲-۱ در خانواده $N(\mu,\sigma^2)$ برآوردگر S^2 برای σ^2 غیر مجاز است زیرا برآوردگر $\widetilde{S^2} = \frac{n-1}{n+1}S^2$ از آن بهتر است. اما خود برآوردگر S^2 نیز مجاز نیست و توسط *Brewster and Zidek* (۱۹۷۴) برآوردگر بهتر از آن پیدا شده است.

[1] admissible
[2] inadmissible

با توجه به تعریف ۲-۳ یک تصمیم مجاز تصمیم خوبی است زیرا هیچ تصمیم دیگری بر آن ارجحیت ندارد و یک تصمیم غیرمجاز تصمیم بدی است زیرا تصمیم دیگری وجود دارد که بر آن ارجحیت دارد. با این وجود توجه کنید که مجاز بودن واقعاً یک خاصیت مثبت نیست و همچنین خاصیت منفی نیز نیست، یعنی یک تصمیم مجاز لزوماً به طور یکنواخت خوب نیست و همچنین به طور یکنواخت نیز بد نمی باشد. اگر بدانیم که یک تصمیم غیر مجاز است، در این صورت بایستی دنبال تصمیمی بگردیم که از آن بهتر باشد. اما اگر بدانیم که یک تصمیم مجاز است دلیل بر آن نیست که حتماً از این تصمیم استفاده کنیم زیرا در خیلی از موارد تصمیم‌های مجاز زیادی وجود دارد (مثال ۲-٤ را ملاحظه کنید) که بعضی معقول و بعضی غیرمعقول هستند و بعضی به سختی قابل محاسبه هستند. به مثال زیر توجه کنید.

مثال ۲-۳: فرض کنید $X \sim B(10,\theta)$ نشان دهید که $\delta(X)=\frac{1}{3}$ برآوردگری مجاز برای θ تحت تابع زیان درجه دوم است و آن را با برآوردگر UMVU پارامتر θ یعنی $\delta^*(X)=\frac{X}{10}$ مقایسه کنید.

حل: تابع مخاطره برآورد δ در $p=\frac{1}{3}$ عبارت است از

$$R(\tfrac{1}{3},\delta)=E[(\delta(x)-\tfrac{1}{3})^2]$$

$$=\sum_{x=0}^{10}(\delta(x)-\tfrac{1}{3})^2 P_{P=\frac{1}{3}}(X=x)=\sum_{x=0}^{10}(\tfrac{1}{3}-\tfrac{1}{3})^2 P_{P=\frac{1}{3}}(X=x)=0$$

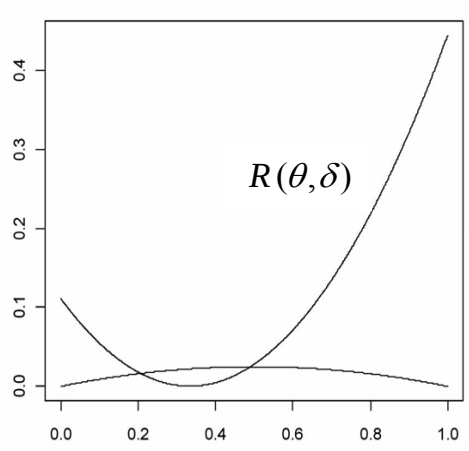
حال فرض کنید $\delta'(x)$ برآوردگر دیگر به خوبی $\delta(x)$ باشد در این صورت

$$R(\tfrac{1}{3},\delta')\le R(\tfrac{1}{3},\delta)=0 \xrightarrow{R(\frac{1}{3},\delta')\ge 0} R(\tfrac{1}{3},\delta')=0 \Rightarrow E[(\delta'(x)-\tfrac{1}{3})^2]=0$$

$$\Rightarrow P\left((\delta'(x)-\tfrac{1}{3})^2=0\right)=1 \Rightarrow P(\delta'(x)=\tfrac{1}{3})=1$$

بنابراین اگر δ' بخواهد به خوبی δ باشد بایستی با احتمال یک با δ برابر باشد. حال اگر δ' بهتر از δ باشد آنگاه به خوبی δ است و در نتیجه باز هم بایستی δ' و δ با احتمال یک با هم برابر باشند و در نتیجه هیچ برآوردگری بهتر از δ وجود ندارد و در نتیجه δ یک برآوردگر مجاز است.

تنها مزیت برآوردگر مجاز $\delta(X)=\frac{۱}{۳}$ این است که در $\theta=\frac{۱}{۳}$ مخاطره آن از هر برآوردگری کمتر است و در مقایسه با برآوردگر UMVUی $\delta^*(X)=\frac{X}{۱۰}$ یک برآوردگر غیرمعقول است. مقایسه مخاطره این دو برآوردگر نشان می دهد که در بیشتر نقاط فاصله $\theta\in(۰,۱)$ برآوردگر UMVU بهتر از $\delta(X)=\frac{۱}{۳}$ عمل می کند و برآوردگر δ تنها در نقطه ی $\theta=\frac{۱}{۳}$ خوب عمل می کند.



$$R(\theta,\delta)=(\frac{۱}{۳}-\theta)^۲$$

$$R(\theta,\delta^*)=\frac{\theta(۱-\theta)}{۱۰}$$

■

با استفاده از تعریف ۲-۱ در زیر کلاسی از توابع تصمیم را معرفی می کنیم که شامل تمام توابع تصمیم خوب است.

تعریف ۲-۳: فرض کنید D کلاسی کلیه توابع تصمیم باشد و C کلاسی از توابع تصمیم و زیر کلاسی از D باشد $C\subset D$.

الف- کلاس C را یک کلاس کامل[۱] گویند اگر برای هر تصمیم $\delta'\notin C$ یک تصمیم $\delta\in C$ وجود داشته که δ بهتر از δ' باشد.

ب- کلاس C را یک کلاس اساسا کامل[۲] گویند اگر برای هر تصمیم $\delta'\notin C$ یک تصمیم $\delta\in C$ وجود داشته باشد که δ به خوبی δ' باشد. ■

قضیه‌های زیر رابطه تصمیم‌های مجاز و کلاس‌های کامل و اساسا کامل را نشان می‌دهد.

قضیه۲-۱: اگر C یک کلاس کامل باشد آنگاه کلاس کلیه تصمیم‌های مجاز A درون C قرار دارد ($A\subset C$).

[۱] complete
[۲] essentially complete

اثبات: فرض کنید δ' یک تصمیم مجاز باشد($\delta' \in A$)ولی $\delta' \notin C$. در این‌صورت طبق تعریف بایستی یک $\delta \in C$ وجود داشته باشد که δ از δ' بهتر باشد و چون δ' مجاز است این امکان ندارد پس بایستی $\delta' \in C$ باشد و در نتیجه $\delta' \in A \Rightarrow \delta' \in C$ پس $A \subset C$.

قضیه۲-۲: فرض کنید C یک کلاس اساسا کامل باشد . اگر $\delta' \notin C$ یک تصمیم مجاز باشد آنگاه یک تصمیم $\delta \in C$ وجود دارد بطوری که هم ارز δ' است.

اثبات: چون C اساسا کامل است و $\delta' \notin C$ پس یک تصمیم $\delta \in C$ وجود دارد بطوری که $R(\theta,\delta) \le R(\theta,\delta'), \quad \forall \theta \in \Theta$ وچون δ' یک تصمیم مجاز است پس نامساوی فوق بطور اکید برای هیچ $\theta \in \Theta$ برقرار نیست پس $R(\theta,\delta) = R(\theta,\delta'), \quad \forall \theta \in \Theta$ یعنی δ و δ' هم ارز هستند. ■

با توجه به تعریف و قضایای بالا بایستی توجه خود را به کلاس توابع تصمیم کامل معطوف کنیم و در این کلاس برآوردگرهای مجاز را پیدا کنیم.

روش‌های بدست آوردن برآوردگرها و تصمیم‌های مجاز

در زیر قضایایی را برای بدست آوردن تصمیم‌های مجاز و غیرمجاز می‌آوریم که عمده انها در مورد برآوردگرهای مجاز است.

قضیه۲-۳: اگر δ^π یک تصمیم بیز یکتا نسبت به چگالی پیشین $\pi(\theta)$ و تابع زیان $L(\theta,\delta)$ باشد آنگاه δ^π یک تصمیم مجاز است.

اثبات: فرض کنید δ^π یک تصمیم مجاز نباشد در این صورت یک تصمیم δ وجود دارد بطوری که :

$$R(\theta,\delta) \le R(\theta,\delta^\pi) \ \forall \theta \in \Theta, \quad R(\theta',\delta) < R(\theta',\delta^\pi) \quad for\ some\ \theta' \in \Theta \qquad (۱)$$

بنابراین

$$r(\pi,\delta) = \int R(\theta,\delta) d\Pi(\theta) \underset{(۱)}{\le} \int R(\theta,\delta^\pi) d\Pi(\theta) = r(\pi,\delta^\pi) \qquad (۲)$$

از طرفی چون δ^π برآوردگر بیز است پس

$$r(\pi,\delta^\pi) \le r(\pi,\delta) \qquad (۳)$$

بنابراین $(۲),(۳) \Rightarrow \quad r(\pi,\delta^{\pi}) = r(\pi,\delta)$

یعنی δ نیز یک تصمیم بیز است که این با یکتا بودن تصمیم بیز متناقض است. پس δ^{π} یک تصمیم مجاز است. ■

قضیه ۲-٤: فرض کنید $\gamma(\theta)$ پارامتر مورد نظر ما باشد و $\gamma(\theta) \in [a,b]$ و تابع زیان $L(\theta,\delta)$ برای $\delta \neq \gamma(\theta)$ مثبت و برای $\delta = \gamma(\theta)$ صفر باشد و تابع زیان $L(\theta,\delta)$ با دور شدن δ از $\gamma(\theta)$ یک تابع صعودی باشد. در این صورت هر برآوردگر δکه مقادیری را در خارج از فاصله‌ی $[a,b]$ با احتمال مثبت بگیرید غیرمجاز است.

اثبات: برآوردگر δ' را به صورت زیر تعریف می کنیم.

$$\delta' = \begin{cases} a & \delta < a \\ \delta & a \le \delta \le b \\ b & \delta > b \end{cases}$$

در این صورت بنابر فرض قضیه اگر $\delta < a$ آنگاه $L(\theta,a) < L(\theta,\delta)$ و اگر $\delta > b$ آنگاه $L(\theta,b) < L(\theta,\delta)$ بنابراین

$$\begin{aligned} R(\theta,\delta') &= E[L(\theta,a)I_{\delta<a}] + E[L(\theta,\delta)I_{a\le\delta\le b}] + E[L(\theta,b)I_{\delta>b}] \\ &\le E[L(\theta,\delta)I_{\delta<a}] + E[L(\theta,\delta)I_{a\le\delta\le b}] + E[L(\theta,\delta)I_{\delta>b}] \\ &= E[L(\theta,\delta)] = R(\theta,\delta) \end{aligned}$$

پس δ برآوردگری غیرمجاز است. ■

قضیه ۲-۵: فرض کنید تابع زیان درجه دوم باشد و X یک متغیر تصادفی با میانگین θ و واریانس $\sigma^{۲}$ باشد در این صورت $aX+b$ یک برآوردگر غیرمجاز برای θ است وقتی که

$$(i)\ a > ۱ \qquad \text{یا} \qquad (ii)\ a < ۰ \qquad \text{یا} \qquad (iii)\ a = ۱,\ b \neq ۰$$

اثبات: $f(a,b) = R(\theta,aX+b) = E[(aX+b-\theta)^{۲}] = Var(aX+b) + [E(aX+b)-\theta]^{۲}$

$$= a^{۲}Var(x) + (a\theta+b-\theta)^{۲} = a^{۲}\sigma^{۲} + [(a-۱)\theta+b]^{۲}$$

$$(i) \quad a > ۱ \quad \Rightarrow \quad f(a,b) \ge a^{۲}\sigma^{۲} > \sigma^{۲} = f(۱,۰) = R(\theta,X)$$

پس $aX+b$ به وسیله X مغلوب[۱] می‌شود (یعنی $(R(\theta,X)<R(\theta,aX+b)$)

$$(ii)\quad a<۰\Rightarrow\ (a-۱)^۲>۱\Rightarrow f(a,b)\geq[(a-۱)\theta+b]^۲=(a-۱)^۲[\theta+\frac{b}{a-۱}]^۲$$
$$>[\theta+\frac{b}{a-۱}]^۲=f(۰,\frac{-b}{a-۱})=R(\theta,\frac{-b}{a-۱})$$

بنابراین $aX+b$ به وسیله برآوردگر ثابت $\delta=\frac{-b}{a-۱}$ مغلوب می‌شود.

$$(iii)\quad a=۱,\ b\neq۰\quad\Rightarrow f(a,b)=f(۱,b)=\sigma^۲+b^۲>\sigma^۲=f(۱,۰)=R(\theta,X)$$

پس $aX+b$ به وسیله X مغلوب می‌شود. ■

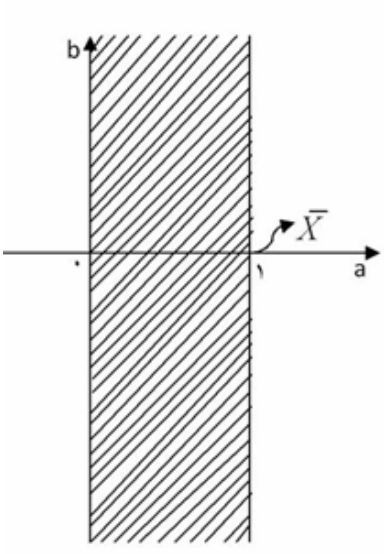
مثال ۲-٤: فرض‌کنید $X_۱,X_۲,...,X_n$ یک نمونه‌تصادفی از $N(\theta,\sigma^۲)$ باشند که در آن $\sigma^۲$ معلوم است. برای چه مقادیر از a و b کلاس برآوردگرهای خطی $a\bar{X}+b$ مجاز می‌باشد.

حل: در فصل قبل دیدیم که برآوردگر بیز یکتای θ تحت توزیع پیشین $N(\mu,\tau^۲)$ عبارت است از:

$$\delta^{\pi}(\underset{\sim}{X})=\frac{n\tau^۲}{\sigma^۲+n\tau^۲}\bar{X}+\frac{\sigma^۲}{\sigma^۲+n\tau^۲}\mu$$

که مخاطره بیز این برآوردگر متناهی است زیرا

$$R(\theta,\delta^{\pi})=(\frac{n\tau^۲}{\sigma^۲+n\tau^۲})^۲\frac{\sigma^۲}{n}+\left[\frac{n\tau^۲\theta+\mu\sigma^۲}{\sigma^۲+n\tau^۲}-\theta\right]^۲\Rightarrow r(\pi,\delta^{\pi})=\int_{\Theta}R(\theta,\delta^{\pi})d\Pi(\theta)<\infty$$



حال چون $۰<\frac{n\tau^۲}{\sigma^۲+n\tau^۲}<۱$ و $-\infty<\frac{\mu\sigma^۲}{\sigma^۲+n\tau^۲}<+\infty$ پس برآوردگر $a\bar{X}+b$ برای $b\in R,۰<a<۱$ یک برآوردگر مجاز است.

چون $E(\bar{X})=\theta$ پس طبق قضیه ۲-۵ برآوردگر $a\bar{X}+b$ برای $a>۱$ (حالت (i)) و $a<۰$ (حالت (ii)) و $a=۱,b\neq۰$ (حالت (iii)) مجاز نیست.

[۱] Dominate

برای حالت $a=۰$ برآوردگر $\delta(\underset{\sim}{X})=b$ مجاز است زیرا این برآوردگر تنها برآوردگری است که در $\theta=b$ دارای مخاطره صفر است، $R(b,\delta)=۰$ (مثال ۲-۳ را ملاحظه کنید).

تنها نقطه‌ای از صفحه مختصات (a,b) که باقی مانده است نقطه $(a,b)=(۰,۱)$ مربوط به برآوردگر $ML, UMVU, MRE$ و می‌نیماکس $\bar{X}$ است. که از دو روش، یکی روش بلیت[۱] (۱۹۵۱) یا روش بیزی حدی[۲] و دیگری روش نامساوی اطلاع مجاز بودن آن را ثابت می کنیم.

الف- روش بیز حدی

قضیه ۲-٦: فرض کنید $\theta\in\Theta\subset R$ و $\delta_.$ یک تابع تصمیم باشد. فرض کنید که $R(\theta,\delta)$ برای هر δ ثابت یک تابع پیوسته از θ باشد. فرض کنید دنباله‌ای از توزیع‌های پیشین π_m با تصمیم‌های بیز متناظر $\delta^m=\delta^{\pi_m}$ و مخاطره‌های بیز متناظر $r_m=r(\pi_m,\delta^m)$ موجود باشند. همچنین فرض کنید r_m^* مخاطره بیز تصمیم $\delta_.$ نسبت به توزیع پیشین π_m باشد. اگر برای هر $\eta>۰$ و $\theta\in\Theta$ داشته باشیم که:

$$\frac{\Pi_m(\theta+\eta)-\Pi_m(\theta-\eta)}{r_m^*-r_m}\to\infty \quad as \quad m\to\infty$$

(که Π_m تابع توزیع پیشین با چگالی پیشین π_m است) آنگاه $\delta_.$ یک تصمیم مجاز است.

اثبات: فرض کنید $\delta_.$ یک تصمیم مجاز نباشد در این صورت یک تصمیم δ^* وجود دارد به طوری که

$$R(\theta,\delta^*)\le R(\theta,\delta_.) \quad \forall\theta\in\Theta$$

$$R(\theta_.,\delta^*)<R(\theta_.,\delta_.) \quad for\ some\ \theta_.\in\Theta$$

قرار می‌دهیم $\varepsilon=R(\theta_.,\delta_.)-R(\theta_.,\delta^*)>۰$

چون تابع مخاطره $R(\theta,\delta)$ برای δ ثابت یک تابع پیوسته از θاست پس

$$\forall\varepsilon>۰\ \exists\eta>۰\ \ni|\theta-\theta_.|\le\eta\Rightarrow\begin{cases}\left|R(\theta,\delta^*)-R(\theta_.,\delta^*)\right|<\varepsilon/۴\\ \left|R(\theta,\delta_.)-R(\theta_.,\delta_.)\right|<\varepsilon/۴\end{cases}$$

[۱] Blyth
[۲] Limiting Bayes Method

$$\Rightarrow R(\theta,\delta_{.})-R(\theta,\delta^{*})>\left(R(\theta_{.},\delta_{.})-\frac{\varepsilon}{۴}\right)-\left(R(\theta_{.},\delta^{*})+\frac{\varepsilon}{۴}\right)$$

$$=R(\theta_{.},\delta_{.})-R(\theta_{.},\delta^{*})-\frac{\varepsilon}{۲}=\varepsilon-\frac{\varepsilon}{۲}=\frac{\varepsilon}{۲}$$

در نتیجه $\forall\theta\in[\theta_{.}-\eta,\theta_{.}+\eta]$ $\quad R(\theta,\delta_{.})-R(\theta,\delta^{*})>\frac{\varepsilon}{۲}$

بنابراین اگر r_m^{**} مخاطره بیز تصمیم δ^{*} باشد یعنی $r_m^{**}=\int R(\theta,\delta^{*})d\Pi_m(\theta)$ در این صورت

$$r_m^{*}-r_m^{**}=\int_{-\infty}^{+\infty}\left[R(\theta,\delta_{.})-R(\theta,\delta^{*})\right]d\Pi_m(\theta)$$

$$\geq\int_{\theta_{.}-\eta}^{\theta_{.}+\eta}\left[R(\theta,\delta_{.})-R(\theta,\delta^{*})\right]d\Pi_m(\theta)$$

$$\geq\frac{\varepsilon}{۲}\int_{\theta_{.}-\eta}^{\theta_{.}+\eta}d\Pi_m(\theta)=\frac{\varepsilon}{۲}[\Pi_m(\theta_{.}+\eta)-\Pi_m(\theta_{.}-\eta)]$$

$$\Rightarrow\frac{r_m^{*}-r_m^{**}}{r_m^{*}-r_m}\geq\frac{\frac{\varepsilon}{۲}\left[\Pi_m(\theta_{.}+\eta)-\Pi_m(\theta_{.}-\eta)\right]}{r_m^{*}-r_m}\to\infty\qquad as\ \ m\to\infty$$

بنابراین $m_{.}$ی وجود دارد که $\frac{r_{m_{.}}^{*}-r_{m_{.}}^{**}}{r_{m_{.}}^{*}-r_{m_{.}}}>۱$ و یا معادلاً $r_{m_{.}}>r_{m_{.}}^{**}$ که این با تصمیم بیز بودن $\delta^{m_{.}}$ نسبت به توزیع پیشین $\pi_{m_{.}}$ متناقض است. بنابراین $\delta_{.}$ یک تصمیم مجاز است. ■

مثال ۲-۵: فرض کنید $X_۱,X_۲,\ldots,X_n$ یک نمونه تصادفی از $N(\theta,\sigma^{۲})$ باشند نشان دهید که $\bar{X}$ یک برآوردگر مجاز برای θ است.

حل: اگر توزیع پیشین $N(۰,m^{۲})$ را برای θ در نظر بگیریم آنگاه

$$\theta\,|\,\underset{\sim}{x}\sim N\left(\frac{nm^{۲}\bar{x}}{\sigma^{۲}+nm^{۲}},\frac{۱}{\frac{n}{\sigma^{۲}}+\frac{۱}{m^{۲}}}\right)\qquad\delta^{m}=\delta^{\pi_m}=\frac{nm^{۲}}{\sigma^{۲}+nm^{۲}}\bar{X}$$

$$r_m=r(\pi_m,\delta^{\pi_m})=E_{\underset{\sim}{X}}[V(\theta\,|\,\underset{\sim}{X})]=\frac{۱}{\frac{n}{\sigma^{۲}}+\frac{۱}{m^{۲}}}$$

همچنین مخاطره بیز برآوردگر $\delta_{\cdot}(\underset{\sim}{X}) = \bar{X}$ عبارت است از

$$r_m^* = r(\pi_m, \delta_{\cdot}) = E[E(\bar{X} - \theta) \mid \theta] = E\left(\frac{\sigma^2}{n}\right) = \frac{\sigma^2}{n}$$

بنابراین

$$\frac{\Pi_m(\theta+\eta) - \Pi_m(\theta-\eta)}{r_m^* - r_m} = \frac{\int_{\theta-\eta}^{\theta+\eta} \frac{1}{\sqrt{2\pi m^2}} e^{-\frac{1}{2m^2}t^2} dt}{\frac{\sigma^2}{n} - \left[\frac{n}{\sigma^2} + \frac{1}{m^2}\right]^{-1}}$$

$$= \frac{\frac{1}{m\sqrt{2\pi}} \int_{\theta-\eta}^{\theta+\eta} e^{-\frac{1}{2m^2}t^2} dt}{\frac{\sigma^4}{n(nm^2+\sigma^2)}} = \frac{n(nm^2+\sigma^2)}{m\sqrt{2\pi}\sigma^4} \int_{\theta-\eta}^{\theta+\eta} e^{-\frac{1}{2m^2}t^2} dt$$

$$\lim_{m\to+\infty} e^{-\frac{1}{2m^2}t^2} = 1 \Rightarrow \lim_{m\to+\infty} \int_{\theta-\eta}^{\theta+\eta} e^{-\frac{1}{2m^2}t^2} dt = \int_{\theta-\eta}^{\theta+\eta} \lim_{m\to+\infty} e^{-\frac{1}{2m^2}t^2} dt = (\theta+\eta) - (\theta-\eta) = 2\eta$$

$$\therefore \quad \frac{\Pi_m(\theta+\eta) - \Pi_m(\theta-\eta)}{r_m^* - r_m} \to +\infty \qquad as \;\; m \to +\infty$$

بنابراین طبق قضیه ۲-۶ (روش بلیت) $\delta_{\cdot}(\underset{\sim}{X}) = \bar{X}$ یک برآوردگر مجاز است. ■

نتیجه: از مثال های ۲-۴ و ۲-۵ نتیجه می‌شود که در توزیع $N(\theta, \sigma^2)$ برآوردگر $a\bar{X} + b$ برای θ مجاز است اگر و فقط اگر

$$(i) \quad 0 \le a < 1 \quad \text{یا} \quad (ii) \quad a = 1, b = 0.$$ ■

ب- روش نامساوی اطلاع[1]:

از این روش می‌توان برای اثبات مجاز بودن یا مینیماکس بودن برآوردگر استفاده کرد (Hodges & Lehmann (۱۹۵۱))

با توجه به اینکه $R(\theta, \delta) = E[(\delta - \theta)^2] = V_\theta(\delta) + b^2(\theta)$

که در آن $b(\theta) = Bias(\delta)$ بنابراین از نامساوی اطلاع با $\gamma(\theta) = \theta$ داریم که

[1] The Information Inequality Method

$$R(\theta,\delta) \geq \frac{[1+b'(\theta)]^2}{nI(\theta)} + b^2(\theta) \qquad (1)$$

که در آن $I(\theta) = \frac{1}{\sigma^2}$ می‌باشد. حال فرض کنید $\bar{X}$ برآوردگری مجاز نباشد در این صورت برآوردگری مانند δ وجود دارد که

$$R(\theta,\delta) \leq R(\theta,\bar{X}) = \frac{\sigma^2}{n} \qquad \forall \theta \in \Theta \qquad (2)$$

و بنابراین

$$\frac{\sigma^2[1+b'(\theta)]^2}{n} + b^2(\theta) \leq R(\theta,\delta) \leq \frac{\sigma^2}{n} \qquad \forall \theta \qquad (3)$$

می‌خواهیم نشان دهیم که $b(\theta) \equiv 0$ یعنی δ یک برآوردگر ناریب است.

$(3) \Rightarrow \quad b^2(\theta) \leq \frac{\sigma^2}{n} \Rightarrow |b(\theta)| \leq \frac{\sigma}{\sqrt{n}} \Rightarrow$ تابع b کراندار است

$$(3) \Rightarrow \frac{\sigma^2[1+b'(\theta)]^2}{n} \leq \frac{\sigma^2}{n} \Rightarrow [1+b'(\theta)]^2 \leq 1 \Rightarrow b'(\theta) \leq 0$$

پس b یک تابع غیرصعودی است.

حال فرض کنید $b(\theta) \not\equiv 0$. در این صورت دو حالت زیر را داریم:

(i) فرض کنید θ_0ی وجود داشته باشد که $b(\theta_0) < 0$ در این صورت

$$b(\theta) < 0 \qquad \forall \theta \geq \theta_0$$

$$(3) \Rightarrow nb^2(\theta) + \sigma^2 b'^2(\theta) + 2\sigma^2 b'(\theta) \leq 0$$

$$\Rightarrow nb^2(\theta) + 2\sigma^2 b'(\theta) \leq 0 \Rightarrow \frac{-b'(\theta)}{b^2(\theta)} \geq \frac{n}{2\sigma^2} \qquad \forall \theta \geq \theta_0$$

$$\Rightarrow \int_{\theta_0}^{\theta} \frac{-b'(t)}{b^2(t)} dt \geq \frac{n}{2\sigma^2} \int_{\theta_0}^{\theta} dt \Rightarrow \frac{1}{b(t)}\Big|_{\theta_0}^{\theta} \geq \frac{n}{2\sigma^2}(\theta - \theta_0)$$

$$\Rightarrow \frac{1}{b(\theta)} - \frac{1}{b(\theta_0)} \geq \frac{n}{2\sigma^2}(\theta - \theta_0)$$

$$\overset{b(\theta)<0}{\Rightarrow} -\frac{1}{b(\theta_0)} > \frac{1}{b(\theta)} - \frac{1}{b(\theta_0)} \geq \frac{n}{2\sigma^2}(\theta - \theta_0)$$

$$\theta \to \infty \Rightarrow -\frac{1}{b(\theta_0)} \to \infty \# \Rightarrow b(\theta_0) = 0$$

(ii) فرض کنید θ_0ی وجود داشته باشد که $b(\theta_0)>0$ در این صورت $\forall \theta \le \theta_0$، $b(\theta)>0$ و همانند قسمت (i) می‌توان به یک تناقض رسید پس $b(\theta_0)=0$ است.

بنابراین $\forall \theta$ $b(\theta)=b'(\theta)=0$ و (۱) نتیجه می‌دهد که $R(\theta,\delta)\ge \frac{\sigma^2}{n}$ و بنابراین $R(\theta,\delta)=\frac{\sigma^2}{n}=R(\theta,\bar{X})$ یعنی $\bar{X}$ یک برآوردگر مجاز و همچنین مینیماکس برای θ است.

(مینیماکس بودن به این دلیل است که اگر برآوردگری مانند δ موجود باشد که $\sup_{\Theta} R(\theta,\delta)\le \sup_{\Theta} R(\theta,\bar{X})=\frac{\sigma^2}{n}$ آنگاه $\forall \theta$, $R(\theta,\delta)\le \frac{\sigma^2}{n}$ و طبق مراحل بالا $R(\theta,\delta)=R(\theta,\bar{X})$ پس $\bar{X}$ برآوردگر مینیماکس است.)

همچنین $\bar{X}$ تنها برآوردگر مینیماکس برای θ است که این موضوع از لم زیر نتیجه می‌شود.

لِم ۲-۱: اگر تابع زیان $L(\theta,\delta)$ اکیداً محدب باشد و δ برآوردگری مجاز برای $\gamma(\theta)$ باشد و δ'هر برآوردگر دیگری باشد که $\forall \theta$ $R(\theta,\delta)=R(\theta,\delta')$ آنگاه $P(\delta'=\delta)=1$.

اثبات: اگر $\delta^*=\frac{\delta+\delta'}{2}$ آنگاه $R(\theta,\delta^*)\overset{(*)}{<}\frac{1}{2}\left[R(\theta,\delta)+R(\theta,\delta')\right]=R(\theta,\delta)$

بجز اینکه با احتمال یک $\delta=\delta'$ باشد.(نامساوی (*) از اکیدا محدب بودن تابع زیان نتیجه می‌گردد.)

بنابراین رابطه فوق متناقض با مجاز بودن δ است بجز اینکه $P(\delta=\delta')=1$ ■

مثال ۲-۶: میانگین نرمال بریده شده[1]

فرض کنید $X_1,X_2,\ldots,X_n$ یک نمونه تصادفی از $N(\theta,\sigma^2)$ باشند که σ^2 معلوم و حالات زیر را برای θ در نظر بگیرید.

الف- اگر $\theta>\theta_0$ باشد آنگاه طبق قضیه ۲-٤، $\bar{X}$ غیرمجاز است و به وسیله برآوردگر $\delta(\underset{\sim}{X})=\begin{cases}\theta_0 & \bar{X}<\theta_0\\ \bar{X} & \bar{X}\ge\theta_0\end{cases}=\max(\theta_0,\bar{X})$ مغلوب می‌شود. اما در زیر به وسیله روش نامساوی اطلاع نشان می دهیم که $\bar{X}$ برآوردگری مینیماکس برای θ است.

[1] Truncated Normal mean

اگر $\bar{X}$ مینیماکس نباشد آنگاه برآوردگری مانند δ و $\varepsilon > ۰$ وجود دارند به طوری که

$$R(\theta,\delta) \le R(\theta,\bar{X}) - \varepsilon = \frac{\sigma^۲}{n} - \varepsilon \qquad \forall \theta > \theta_۰$$

$$\overset{(۳)}{\Rightarrow} \frac{\sigma^۲\left[۱+b'(\theta)\right]^۲}{n} + b^۲(\theta) \le \frac{\sigma^۲}{n} - \varepsilon \qquad \forall \theta > \theta_۰ \qquad (٤)$$

بنابراین

$$(۴) \Rightarrow b^۲(\theta) \le \frac{\sigma^۲}{n} - \varepsilon \Rightarrow \qquad b \text{ کراندار است}$$

$$(۴) \Rightarrow \frac{\sigma^۲\left[۱+b'(\theta)\right]^۲}{n} \le \frac{\sigma^۲}{n} - \varepsilon \Rightarrow \left[۱+b'(\theta)\right]^۲ \le ۱ - \frac{\varepsilon n}{\sigma^۲}$$

$$\Rightarrow ۱ + b'^۲(\theta) + ۲b'(\theta) \le ۱ - \frac{\varepsilon n}{\sigma^۲} \Rightarrow ۲b'(\theta) \le -\frac{\varepsilon n}{\sigma^۲}$$

$$\Rightarrow b'(\theta) \le \frac{-\varepsilon n}{۲\sigma^۲} < ۰ \qquad \forall \theta > \theta_۰$$

کراندار بودن b و اکیداً نزولی بودن b در تناقض می‌باشد، پس $\bar{X}$ مینیماکس است. توجه کنید برآوردگر $\delta(\underset{\sim}{X}) = \max(\theta_۰, \bar{X})$ برآوردگری بهتر از $\bar{X}$ است و می‌توان نشان داد که $\delta(\underset{\sim}{X})$ نیز مینیماکس است(اثبات: تمرین) و بنابراین برآوردگر مینیماکس یکتا نیست.

$$\left(\begin{array}{c} R(\theta,\delta) \le R(\theta,\bar{X}) = \dfrac{\sigma^۲}{n},\ \bar{X} > \theta_۰ \Rightarrow R(\theta,\delta) = R(\theta,\bar{X}) = \dfrac{\sigma^۲}{n} \\ \\ \Rightarrow \underset{\theta>\theta_۰}{Sup}\, R(\theta,\delta) = \dfrac{\sigma^۲}{n} \end{array} \right)$$

ب- اگر $a \le \theta \le b$ باشد آنگاه $\bar{X}$ طبق قضیه ۲-٤ به وسیله برآوردگر

$$\delta^*(\underset{\sim}{X}) = \begin{cases} a & \bar{X} < a \\ \bar{X} & a \le \bar{X} \le b \\ b & \bar{X} > b \end{cases}$$

مغلوب می‌شود پس غیرمجاز است. حال اگر $\bar{X}$ بخواهد برآوردگر مینیماکس باشد آنگاه بایستی مغلوب کننده آن نیز مینیماکس باشد یعنی

$$\underset{a\le\theta\le b}{Sup}\, R(\theta,\delta^*) = \underset{a\le\theta\le b}{Sup}\, R(\theta,\bar{X}) = \frac{\sigma^۲}{n}$$

اما $\forall \theta \in [a,b]$ ، $R(\theta,\delta^*) < R(\theta,\bar{X}) = \frac{\sigma^2}{n}$. به علاوه $R(\theta,\delta^*)$ یک تابع پیوسته از θ است، بنابراین $R(\theta,\delta^*)$ ماکسیمم خود را در یک نقطه $a \le \theta_. \le b$ اختیار می‌کند. در نتیجه $\sup_{a\le\theta\le b} R(\theta,\delta^*) = R(\theta_.,\delta^*) < \frac{\sigma^2}{n}$ که با رابطه قبل در تناقض است پس $\bar{X}$ نمی‌تواند مینیماکس باشد.

ج- اگر $-m \le \theta \le m$ آنگاه در بخش قبل نشان دادیم که برآوردگر $\delta_m(x) = m\frac{e^{+mx}-e^{-mx}}{e^{mx}+e^{-mx}} = mtghmx$ برای θ مینیماکس است. ■

مثال ۲-۷: (**واریانس توزیع نرمال**) فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از $N(0,\sigma^2)$ باشند. اگر $\tau = \frac{1}{2\sigma^2} \sim \Gamma(g, \frac{1}{\alpha})$ آنگاه $\tau \mid \underset{\sim}{x} \sim \Gamma(g+\frac{n}{2}, \frac{1}{\alpha+y})$ که $Y = \sum_{i=1}^{n} X_i^2$. اگر قرار دهیم $r = \frac{n}{2}$ و تابع زیان درجه دوم باشد آنگاه برآوردگر بیز $\frac{1}{\tau} = 2\sigma^2$ عبارت است از

$$\delta^{\pi}(\underset{\sim}{x}) = E(\frac{1}{\tau} \mid \underset{\sim}{x}) = \frac{\alpha+y}{r+g-1}$$

می‌خواهیم برآوردگرهای مجاز و غیرمجاز خطی $\tau = \frac{1}{2\sigma^2}$ به فرم $aY+b$ را تعیین کنیم.

چون $\delta^{\pi}(\underset{\sim}{X}) = \frac{1}{r+g-1}Y + \frac{\alpha}{r+g-1}$ پس طبق قضیه ۲-۳ به نظر می‌رسد که $aY+b$ برای مقادیر $b>0$ و $0<a<\frac{1}{r-1}$، $(g>0)$ مجاز باشد. یکی از حالت‌های این برآوردگرهای مجاز $\frac{1}{r}Y+b$ می‌باشد. از طرفی $E(Y) = n\sigma^2 = \frac{r}{\tau}$ پس $\frac{1}{r}Y$ یک برآوردگر نااریب برای $\frac{1}{\tau}$ است. بنابراین چون $\frac{1}{r}Y+b$ که برآوردگری اریب‌دار است در رابطه

$$R(2\sigma^2, \frac{1}{r}Y+b) = R(2\sigma^2, \frac{1}{r}Y) + b^2 > R(2\sigma^2, \frac{1}{r}Y)$$

صدق می‌کند پس برآوردگری غیرمجاز است. به نظر می‌رسد که اشکالی در این خصوص وجود دارد. قضیه ۲-۳ بیان می‌کرد که برآوردگر بیز یکتا، برآوردگری مجاز است. و دو شرط برای یکتا بودن

برآوردگر بیز وجود داشت که در این جا شرط اول یعنی متناهی بودن مخاطره بیز به ازای هر $b>۰$ و $۰<a<\frac{۱}{r-۱}$ برقرار نیست زیرا

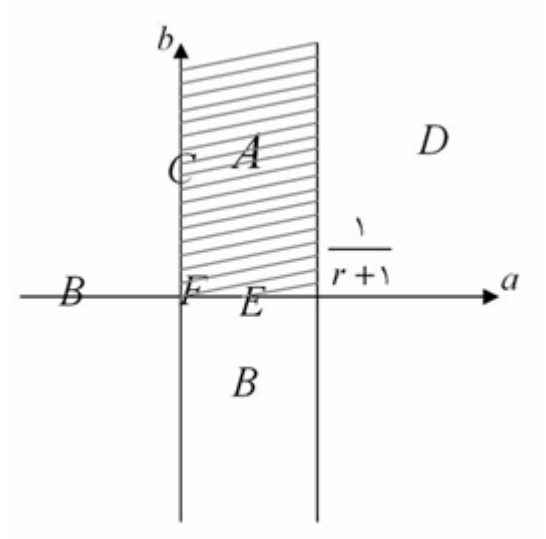
$$R(۲\sigma^۲,\frac{Y+\alpha}{r+g-۱})=E(\frac{Y+\alpha}{r+g-۱}-\frac{۱}{\tau})^۲=V(\frac{Y+\alpha}{r+g-۱})-\left[E(\frac{Y+\alpha}{r+g-۱})-\frac{۱}{\tau}\right]^۲$$

$$=\frac{۱}{(r+g-۱)^۲}\left[\frac{r}{\tau^۲}\right]-\left[\frac{(\frac{r}{\tau})+\alpha}{r+g-۱}-\frac{۱}{\tau}\right]^۲=\frac{۱}{(r+g-۱)^۲}\left\{\frac{r}{\tau^۲}-(\alpha-\frac{g-۱}{\tau})^۲\right\}$$

بنابراین مخاطره بیز متناهی خواهد بود اگر $E(\frac{۱}{\tau^۲})<\infty$ و این امید ریاضی موقعی متناهی است که $g>۲$ باشد $\left(E(\frac{۱}{\tau^۲})=\frac{\alpha^۲}{(g-۱)(g-۲)}\right)$. با اعمال این شرط روی برآوردگر بیز $\delta^{\pi}(\underset{\sim}{X})$ دیده می‌شود که برآوردگر بیز یکتا در شرط زیر صدق می‌کند.

$$b>۰ \text{ و } ۰<a<\frac{۱}{r+۱}$$

و بنابراین برای این مقادیر (a,b) برآوردگر $aY+b$ برای $\frac{۱}{\tau}=۲\sigma^۲$ مجاز است. (A) برای $a<۰$ و همچنین برای $b<۰$ توسط قضیه ۲-٤ برآوردگر $aY+b$ غیرمجاز است. (B) (زیرا مقادیری خارج از دامنه پارامتر $(۰,+\infty)$ را می‌گیرد.) حالت $a=۰$ بر آوردگر $\delta(\underset{\sim}{X})=b,\ b>۰$ مجاز است زیرا این برآوردگر تنها برآوردگری است که در $\frac{۱}{\tau}=b$ دارای مخاطره صفر است $R(b,\delta)=۰$. (مثال ۲-۳ را ملاحظه کنید). (C)



برای حالت $a>\frac{۱}{r+۱}$ (مسئله ۲,۱۲) می‌توان نشان داد که برآوردگر فوق غیرمجاز است. (D) همچنین

$$R(۲\sigma^۲, aY+b) = E[(aY+b-\frac{۱}{\tau})^۲] = a^۲V(Y) + \left[aE(Y)+b-\frac{۱}{\tau}\right]^۲$$

$$= \frac{a^۲r}{\tau^۲} + \left[\frac{ar}{\tau}+b-\frac{۱}{\tau}\right]^۲ = a^۲r(\frac{۱}{\tau^۲}) + \left[(ar-۱)\frac{۱}{\tau}+b\right]^۲$$

پس برای $b=۰$ داریم که $a\neq ۰$ $$R(۲\sigma^۲, aY) = [a^۲r+(ar-۱)^۲]\frac{۱}{\tau^۲} = f(a),$$

$$f'(a) = [۲ar+۲r(ar-۱)]\frac{۱}{\tau^۲} = ۰ \Rightarrow a(r+۱)-۱=۰ \Rightarrow a = \frac{۱}{r+۱}$$

$$f''(a) = (۲r+۲r^۲)\frac{۱}{\tau^۲} > ۰$$

پس $R(\frac{۱}{\tau}, \frac{۱}{r+۱}Y) < R(\frac{۱}{\tau}, aY)$ برای هر $a\neq ۰$ یعنی $aY+b$ برای $b=۰$ و $a\neq \frac{۱}{r+۱}$ و $a\neq ۰$ برآوردگری غیرمجاز است.(E) برای $b=۰$ و $a=۰$ داریم که

$$R(\frac{۱}{\tau}, aY) < R(\frac{۱}{\tau}, ۰) \Leftrightarrow a^۲r+(ar-۱)^۲ < ۱ \Leftrightarrow a^۲r(۱+r)-۲ar < ۰$$

$$\Leftrightarrow a(۱+r)-۲ < ۰ \Leftrightarrow a < \frac{۲}{۱+r}$$

پس برآوردگر $\delta=۰$ به وسیله برآوردگر $\delta^* = aY$ که $a < \frac{۲}{۱+r}$ مغلوب می‌شود پس غیرمجاز است(F). تنها حالت $b\geq ۰$ و $\frac{۱}{r+۱}Y+b$ باقی مانده است که در زیر توسط قضیه کارلین[۱] مجاز بودن آن را ثابت می کنیم. ■

فرض کنید X دارای توزیعی از خانواده نمایی با تابع چگالی

$$p_\theta(x) = \beta(\theta)e^{\theta T(x)} \qquad (۱)$$

نسبت به اندازه μ باشد و Θ فضای پارامتری طبیعی باشد. در این صورت Θ دارای نقاط انتهایی مانند $\underline{\theta}$ و $\overline{\theta}$ است $(-\infty \leq \underline{\theta} \leq \overline{\theta} \leq +\infty)$. برای برآورد $E_\theta(T)$ تحت تابع زیان درجه دوم، برآوردگر $aT+b$ برای $a<۰$ یا $a>۱$ غیر مجاز است و برای $a=۰$ مقداری ثابت دارد پس مجاز

[۱] Karlin (۱۹۵۸)

است. برای اینکه بتوانیم شرط قضیه کارلین را بیان کنیم راحت‌تر است که این برآوردگر را به صورت زیر بنویسیم.

$$\delta_{\lambda,\gamma}(x) = \frac{۱}{۱+\lambda}T + \frac{\gamma\lambda}{۱+\lambda} \qquad (۲)$$

که در آن $۰ \le \lambda < \infty$ معادل $۰ < a \le ۱$ است.

قضیه۲-۷: (قضیه کارلین) تحت فرضیات بالا، شرط کافی برای آنکه برآوردگر (۲) برای $g(\theta) = E_\theta(T)$ تحت تابع زیان درجه دوم مجاز باشد این است که انتگرال تابع $e^{-\gamma\lambda\theta}\left[\beta(\theta)\right]^{-\lambda}$ در $\underline{\theta}$ و $\overline{\theta}$ واگرا باشد. یعنی برای یک $\underline{\theta} < \theta_. < \overline{\theta}$ انتگرال‌های

$$\int_{\theta_.}^{\theta^*} \frac{e^{-\gamma\lambda\theta}}{\left[\beta(\theta)\right]^{-\lambda}} d\theta \qquad and \qquad \int_{\theta^*}^{\theta} \frac{e^{-\gamma\lambda\theta}}{\left[\beta(\theta)\right]^{-\lambda}} d\theta$$

موقعی که به ترتیب $\theta^* \to \overline{\theta}$ و $\theta^* \to \underline{\theta}$ به سمت بی‌نهایت میل کنند.

اثبات: از روش نامساوی اطلاع قضیه را ثابت می‌کنیم. با توجه به اینکه در خانواده نمایی (۱) داریم که

$$g(\theta) = E_\theta(T) = -\frac{\beta'(\theta)}{\beta(\theta)} \qquad , \qquad g'(\theta) = V_\theta(T) = I(\theta)$$

بنابراین طبق نامساوی اطلاع برای هر برآوردگر δ داریم که

$$E_\theta\left[\left(\delta(X) - g(\theta)\right)^۲\right] = V_\theta(\delta(X)) + b^۲(\theta) \ge \frac{[g'(\theta) + b'(\theta)]^۲}{I(\theta)} + b^۲(\theta)$$

$$= \frac{[I(\theta) + b'(\theta)]^۲}{I(\theta)} + b^۲(\theta) \qquad (۳)$$

که در آن $b(\theta) = E_\theta[\delta(X)] - \gamma(\theta)$. چون در خانواده نمایی یک پارامتری هستیم پس برای برآوردگر $\delta = \delta_{\lambda,\gamma}$ نامساوی فوق به تساوی تبدیل می‌شود و اگر $b_{\lambda,\gamma}(\theta)$ اریبی برآوردگر $\delta_{\lambda,\gamma}$ باشد آنگاه

$$\begin{cases} b_{\lambda,\gamma}(\theta) = \dfrac{\lambda}{۱+\lambda}[\gamma - g(\theta)] \\ b'_{\lambda,\gamma}(\theta) = -\dfrac{\lambda}{۱+\lambda} g'(\theta) \end{cases}$$

$$E_\theta\left[\left(b_{\lambda,\gamma}(X)-g(\theta)\right)^2\right]=\frac{[I(\theta)+b'_{\lambda,\gamma}(\theta)]^2}{I(\theta)}+[b_{\lambda,\gamma}(\theta)]^2 \qquad (4)$$

حال اگر $\delta_.$ برآوردگری باشد که

$$E\left[\left(\delta_.(X)-g(\theta)\right)^2\right]\le E_\theta\left[\left(\delta_{\lambda,\gamma}(X)-g(\theta)\right)^2\right] \qquad (5)$$

و $b_.(\theta)$ اریبی آن باشد آن گاه از روابط (۳) و (٤) داریم که

$$\frac{\left[I(\theta)+b'_.(\theta)\right]^2}{I(\theta)}+b_.^2(\theta)\le E\left[\left(\delta_.(X)-g(\theta)\right)^2\right]\le E_\theta\left[\left(\delta_{\lambda,\gamma}(X)-g(\theta)\right)^2\right]$$

$$=\frac{\left[I(\theta)+b'_{\lambda,\gamma}(\theta)\right]^2}{I(\theta)}+[b_{\lambda,\gamma}(\theta)]^2$$

اگر قرار دهیم $h(\theta)=b_.(\theta)-b_{\lambda,\gamma}(\theta)$ از نامساوی فوق داریم که

$$\left\{b_.^2(\theta)-b_{\lambda,\gamma}^2(\theta)\right\}+\frac{\left(b'_.(\theta)-b'_{\lambda,\gamma}(\theta)\right)\left(2I(\theta)+b'_.(\theta)-b'_{\lambda,\gamma}(\theta)\right)}{I(\theta)}\le 0$$

$$\Rightarrow [h(\theta)]^2+2b_{\lambda,\gamma}(\theta)\left[b_.(\theta)-b_{\lambda,\gamma}(\theta)\right]+\frac{h'(\theta)\left[h'(\theta)+2\left(I(\theta)+b'_{\lambda,\gamma}(\theta)\right)\right]}{I(\theta)}\le 0$$

$$\Rightarrow h^2(\theta)-\frac{2\lambda}{1+\lambda}\left(\gamma-g(\theta)\right)h(\theta)+\frac{2}{1+\lambda}h'(\theta)+\frac{[h'(\theta)]^2}{I(\theta)}\le 0$$

$$\Rightarrow h^2(\theta)-\frac{2\lambda}{1+\lambda}\left(\gamma-g(\theta)\right)h(\theta)+\frac{2}{1+\lambda}h'(\theta)\le 0 \qquad (6)$$

قرار می‌دهیم

$$k(\theta)=h(\theta)\beta^{+\lambda}(\theta)e^{\gamma\lambda\theta} \qquad (7)$$

در این صورت

$$k'(\theta)=h'(\theta)\beta^{\lambda}(\theta)e^{\gamma\lambda\theta}+\lambda h(\theta)\beta^{\lambda-1}(\theta)\beta'(\theta)e^{\gamma\lambda\theta}+\gamma\lambda h(\theta)\beta^{\lambda}(\theta)e^{\gamma\lambda-\theta}$$

$$=k(\theta)\left\{\frac{h'(\theta)}{h(\theta)}+\lambda\frac{\beta'(\theta)}{\beta(\theta)}+\gamma\lambda\right\}$$

$$=k(\theta)\left\{\frac{h'(\theta)}{h(\theta)}+\lambda(g(\theta)-\gamma)\right\}=\frac{k(\theta)}{h(\theta)}\{h'(\theta)-\lambda h(\theta)(g(\theta)-\gamma)\}$$

$$(6)\Rightarrow h^2(\theta)+\frac{2}{1+\lambda}\frac{h(\theta)k'(\theta)}{k(\theta)}\le 0\Rightarrow\frac{h(\theta)}{k(\theta)}[h(\theta)k(\theta)+\frac{2}{1+\lambda}k'(\theta)]\le 0$$

$\frac{h(\theta)}{k(\theta)}=\beta^{-\lambda}(\theta)e^{-\gamma\lambda\theta}>0$

$$\Rightarrow \qquad k^{2}(\theta)\beta^{-\lambda}(\theta)e^{-\gamma\lambda\theta}+\frac{2}{1+\lambda}k'(\theta)\le 0 \qquad (8)$$

از این رابطه نتیجه می‌شود که $k'(\theta)<0$ یعنی k یک تابع نزولی از θ است. حال نشان می‌دهیم که $k(\theta)=0 \ \forall\theta$.

فرض کنید که مقداری از θ وجود داشته باشد که $k(\theta)\neq 0$ باشد. یعنی یک مقدار θ_0 وجود دارد که $k(\theta_0)<0 (i)$ یا $k(\theta_0)>0 (ii)$ است.

حالت (i): اگر $k(\theta_0)<0$ آنگاه $k(\theta)<0, \ \forall\theta\ge\theta_0$ حال از رابطه (۸) داریم که

$$\frac{d}{d\theta}[\frac{1}{k(\theta)}]=\frac{-k'(\theta)}{k^{2}(\theta)}\ge\frac{1+\lambda}{2}\beta^{-\lambda}(\theta)e^{-\gamma\lambda\theta}$$

$$\Rightarrow\int_{\theta_0}^{\theta^*}\frac{d}{d\theta}[\frac{1}{k(t)}]dt\ge\frac{1+\lambda}{2}\int_{\theta_0}^{\theta^*}\beta^{-\lambda}(t)e^{-\gamma\lambda t}dt$$

$$\Rightarrow -\frac{1}{k(\theta_0)}>\frac{1}{k(\theta^*)}-\frac{1}{k(\theta_0)}\ge\frac{1+\lambda}{2}\int_{\theta_0}^{\theta^*}\beta^{-1}(t)e^{\gamma\lambda\theta}dt$$

حال اگر $\theta^*\to\infty$ آنگاه سمت راست به ∞ میل می‌کند که با کمتر بودن آن از $-\frac{1}{k(\theta_0)}$ در تناقض است بجز اینکه $k(\theta_0)=0$.

(ii) به طور مشابه می‌توان نشان داد که نمی‌تواند $k(\theta_0)>0$ باشد.

بنابراین برای هر θ، $k(\theta)=0$ و بنابراین برای هر θ، $h(\theta)=0$ خواهد بود، یعنی در نامساوی (۶) حالت تساوی برقرار می‌شود و در نتیجه در نامساوی (۵) نیز حالت تساوی خواهیم داشت و این نشان می‌دهد که برآوردگر $\delta_{\lambda,\gamma}(X)=\frac{T+\gamma\lambda}{1+\lambda}$ برآوردگری مجاز است. ■

مثال۲-۷ (ادامه): با توجه به اینکه در توزیع $N(0,\sigma^{2})$ داریم که

$$f_{\sigma^{2}}(\underset{\sim}{x})=(2\pi\sigma^{2})^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^{2}}\sum x_i^{2}}=\pi^{-\frac{n}{2}}\tau^{r}e^{-\tau y}\qquad y=\sum x_i^{2},\tau=\frac{1}{2\sigma^{2}}$$

پس این توزیع به فرم خانواده نمایی(۱) با مقادیر زیر است

$$T(X)=\frac{Y}{r},\ \theta=-r\tau,\ \beta(\theta)=(\frac{-\theta}{r})^{r},\ \underline{\theta}=-\infty,\ \overline{\theta}=0$$

(توجه کنید که پارامترهای بگونه‌ای انتخاب شده‌اند که $E(T(X)) = \frac{۱}{\tau}$باشد). حال طبق قضیه کارلین برآوردگر $\frac{۱}{۱+\lambda}\frac{Y}{r} + \frac{\gamma\lambda}{۱+\lambda}$ برای $\frac{۱}{\tau}$ مجاز است اگر انتگرال‌های

$$\int_{-\infty}^{-c} e^{-\gamma\lambda\theta}(\frac{-\theta}{r})^{-r\lambda} d\theta = k\int_{c}^{\infty} e^{\gamma\lambda\theta}\theta^{-r\lambda} d\theta \qquad \text{و} \qquad \int_{\cdot}^{c} e^{-\gamma\lambda\theta}\theta^{-r\lambda} d\theta$$

هر دو واگرا باشند. انتگرال اولی موقعی واگرا است که $r\lambda \le ۱$ و $\gamma = ۰$ یا $\gamma\lambda > ۰$ باشد. و در انتگرال دوم عامل $e^{\gamma\lambda\theta}$ نقشی در انتگرال ندارد و موقعی واگرا است که $r\lambda \ge ۱$ باشد. اگر این شرایط را با هم ادغام کنیم آنگاه برآوردگر $\frac{۱}{۱+\lambda}\frac{Y}{r} + \frac{\gamma\lambda}{۱+\lambda}$مجاز است اگر

$$(i) \quad \gamma = ۰,\ \lambda = \frac{۱}{r} \quad \text{یا} \quad (ii) \quad \lambda \ge \frac{۱}{r},\ \gamma > ۰$$

اگر قرار دهیم $a = \frac{۱}{(۱+\lambda)r}$ و $b = \frac{\gamma\lambda}{۱+\lambda}$ آنگاه برآوردگر $aY + b$ برای $\frac{۱}{\tau}$مجاز است اگر

$$(i) \quad a = \frac{۱}{۱+r},\ b = ۰ \quad \text{یا} \quad (ii) \quad ۰ < a \le \frac{۱}{۱+r},\ b > ۰$$

یعنی برآوردگر $\frac{۱}{۱+r}Y + b,\ b \ge ۰$ برای $\frac{۱}{\tau}$ مجاز است. با توجه به نواحی بدست آمده در این مثال می‌توان گفت که برآوردگر $aY + b$ برای $\frac{۱}{\tau} = ۲\sigma^۲$مجاز است اگر و فقط اگر

$$(i) \quad ۰ \le a \le \frac{۱}{r+۱},\ b > ۰ \qquad (ii) \quad a = \frac{۱}{r+۱},\ b = ۰$$

بنابراین برآوردگر $cY + d$ برای $\sigma^۲ = \frac{۱}{۲\tau}$ مجاز است اگر و فقط اگر

$$(i) \quad ۰ \le c \le \frac{۱}{n+۲},\ d > ۰ \qquad (ii) \quad c = \frac{۱}{n+۲},\ d = ۰\ (r = \frac{n}{۲})\ (c = \frac{a}{۲},\ d = \frac{b}{۲})$$

یعنی برآوردگر MRE، $\delta^*(X) = \frac{۱}{n+۲}\sum X_i^۲$ برای $\sigma^۲$ مجاز است. ■

مثال ۲-۸: (واریانس نرمال با میانگین مجهول)

فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از $N(\mu, \sigma^2)$ باشند. در مثال (۷) دیدیم که اگر $\mu = 0$ باشد، آنگاه برآوردگر MRE، $\frac{1}{n+2}\sum_{i=1}^{n} X_i^2$ برای σ^2 مجاز است. حال اگر μ نامعلوم باشد آیا برآوردگر MRE پارامتر σ^2 یعنی $\frac{1}{n+1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ برای σ^2 مجاز است.

حل: اشتاین[1] (۱۹٦٤) نشان داد که برآوردگر $\frac{1}{n+1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ مجاز نیست و بوسیله برآوردگر $\delta^* = \min(\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n+1}, \frac{\sum X_i^2}{n+2})$ مغلوب می‌شود (اثبات تمرین) دلیل توجیهی این برآوردگر این است که فرض کنید بخواهیم آزمون $H_0 : \mu = 0 \; vs \; H_1 : \mu \neq 0$ را آزمون کنیم در این صورت فرض H_0 را قبول می‌کنیم اگر $\frac{|\sqrt{n}\bar{X}|}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}} < C$ و در نتیجه σ^2 را با $\frac{1}{n+2}\sum_{i=1}^{n} X_i^2$ برآورد می‌کنیم و در غیر این صورت فرض H_1 را قبول می‌کنیم و σ^2 را با $\frac{1}{n+1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ برآورد می‌کنیم. اگر قرار دهیم $C = \sqrt{\frac{n-1}{n+1}}$ آنگاه نامساوی فوق به نامساوی $\frac{1}{n+1}\sum_{i=1}^{n}(X_i - \bar{X})^2 < \frac{1}{n+2}\sum_{i=1}^{n} X_i^2$ تبدیل می‌شود و بنابراین با توجه به آزمون فوق برآوردگر δ^* حاصل می‌گردد.

$$n\bar{X}^2 < \frac{1}{n+1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \Rightarrow n\bar{X}^2 < \frac{1}{n+1}\sum_{i=1}^{n} X_i^2 - \frac{n}{n+1}\bar{X}^2$$

$$\Rightarrow \frac{n(n+2)}{n+1}\bar{X}^2 < \frac{1}{n+1}\sum_{i=1}^{n} X_i^2$$

$$\Rightarrow \frac{n}{n+1}\bar{X}^2 < \frac{1}{(n+1)(n+2)}\sum_{i=1}^{n} X_i^2 = (\frac{1}{n+1} - \frac{1}{n+2})\sum X_i^2$$

[1] Stein

$$\Rightarrow \frac{۱}{n+۲}\sum_{i=۱}^{n} X_i^{۲} < \frac{۱}{n+۱}[\sum X_i^{۲} - n\bar{X}^{۲}] = \frac{۱}{n+۱}\sum (X_i - \bar{X})^{۲}$$

برآوردگر δ^* در رده بزرگتری از کلاس‌ها به صورت $\delta(\bar{x},s) = \varphi(\frac{\bar{x}}{s})S^{۲}$ قرار دارد زیرا

$$\delta^* = \min(\frac{S^{۲}}{n+۱}, \frac{\sum X_i^{۲}}{n+۲}) = \min(\frac{۱}{n+۱}, \frac{S^{۲}+n\bar{X}^{۲}}{(n+۲)S^{۲}})S^{۲}$$

$$= \min\left(\frac{۱}{n+۱}, \frac{۱}{n+۲} + \frac{n}{n+۲}(\frac{\bar{X}}{S})^{۲}\right)S^{۲} = \varphi(\frac{\bar{X}}{S})S^{۲}$$

برآوردگرهای به فرم $\delta(\bar{X},S)$ نیز بر برآوردگر $\frac{۱}{n+۱}\sum (X_i - \bar{X})^{۲}$ غلبه دارند(اثبات: تمرین) ■

مثال ۲-۹: (توزیع دو جمله‌ای) فرض کنید $X \sim B(n,p)$ باشد. برای چه مقادیراز a و b برآوردگر $a(\frac{X}{n})+b$ تحت تابع زیان درجه دوم مجاز و برای چه مقادیر غیرمجاز است.

$$f_X(x) = p(X = x) = \binom{n}{x} p^x (۱-p)^{n-x} = \binom{n}{x}(۱-p)^n e^{\frac{x}{n} n \ln \frac{p}{۱-p}}$$

قرار می‌دهیم

$$\theta = n \ln \frac{p}{۱-p},\ T(X) = \frac{X}{n},\ \beta(\theta) = (۱-p)^n = (۱+e^{\frac{\theta}{n}})^{-n}$$

$$g(\theta) = E_\theta(T(X)) = E_\theta(\frac{X}{n}) = p = \frac{e^{\frac{\theta}{n}}}{۱+e^{\frac{\theta}{n}}}$$

$۰ < p < ۱ \Rightarrow \underline{\theta} = -\infty,\ \bar{\theta} = +\infty$

طبق قضیه کارلین برآوردگر $\delta_{\lambda,\gamma}(x) = \frac{X}{n(۱+\lambda)} + \frac{\gamma\lambda}{۱+\lambda}$ برای p مجاز است اگر دو انتگرال زیر واگرا باشند

$$\int_{-\infty}^{-c} e^{-\gamma\lambda\theta}\left(۱+e^{\frac{\theta}{n}}\right)^{\lambda n} d\theta = \int_{c}^{\infty} e^{\gamma\lambda\theta}\left(۱+\frac{۱}{e^{\frac{\theta}{n}}}\right)^{\lambda n} d\theta \qquad (۱)$$

$$\int_c^{\infty} e^{-\gamma\lambda\theta}\left(۱+e^{\frac{\theta}{n}}\right)^{\lambda n} d\theta \qquad (۲)$$

اگر $\lambda < ۰$ باشد، آنگاه $e^{-\gamma\lambda\theta}\left(۱+e^{\frac{\theta}{n}}\right)^{\lambda n} \le e^{-\gamma\lambda\theta}$ و بنابراین هر دو انتگرال نمی‌توانند همزمان واگرا باشند. اگر $\lambda = ۰$ آنگاه هر دو انتگرال واگرا هستند. اگر $\lambda > ۰$ باشد آنگاه

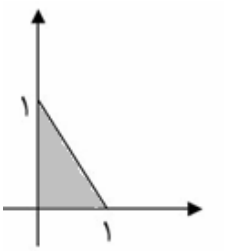
$$۱+e^{\frac{\theta}{n}} > e^{\frac{\theta}{n}} \Rightarrow \left(۱+e^{\frac{\theta}{n}}\right)^{\lambda n} > \left(e^{\frac{\theta}{n}}\right)^{\lambda n}$$

$$\Rightarrow e^{-\gamma\lambda\theta}\left(۱+e^{\frac{\theta}{n}}\right)^{\lambda n} > e^{\lambda\theta}e^{-\gamma\lambda\theta} = e^{(۱-\gamma)\lambda\theta}$$

در این صورت انتگرال دوم به ازای $۱-\gamma \ge ۰$ یا $\gamma \le ۱$ واگراست و در انتگرال اول عامل $\left(۱+e^{-\frac{\theta}{n}}\right)^{\lambda n}$ در همگرایی یا واگرایی تاثیر نمی‌گذارد پس این انتگرال به ازای $\gamma\lambda \ge ۰$ یا $\gamma \ge ۰$ واگرا است.

بنابراین در انتگرال در صورتی همزمان واگرا هستند که $(\lambda > ۰, ۰ \le \gamma \le ۱)$ یا $\lambda = ۰$. با $a = \frac{۱}{۱+\lambda}$ و $b = \frac{\gamma\lambda}{۱+\lambda}$ نتیجه می‌شود که برآوردگر $a(\frac{X}{n})+b$ برای p مجاز است اگر

$(۰ < a < ۱,\ b \ge ۰,\ a+b = \frac{۱+\gamma\lambda}{۱+\lambda} \le ۱)$ یا $a = ۱,\ b = ۰$



که با ادغام این دو حالت داریم که $۰ < a \le ۱,\ b \ge ۰,\ a+b \le ۱$

برای $۰ \le b \le ۱$ و $a = ۰$ برآوردگر $\delta = b$ مجاز است (زیرا این برآوردگر تنها برآوردگری است که در $p = b$ دارای مخاطره صفر است). به راحتی می‌توان نشان داد که برای سایر مقادیر a و b برآوردگر $a(\frac{X}{n})+b$ غیرمجاز است (مسئله۲-٤). بنابراین برآوردگر $a(\frac{X}{n})+b$ برای p مجاز است اگر و فقط اگر (a,b) در مثلث بسته زیر قرار گیرند:

$$\{(a,b) \mid a \ge ۰, b \ge ۰, a+b \le ۱\}$$

■

از قضیه ۲-۷ نتیجه زیر حاصل می‌شود.

نتیجه ۲-۱: اگر فضای پارامتری طبیعی در خانواده نمایی $\beta(\theta)e^{\theta T(x)}$ تمام اعداد حقیقی باشد، یعنی $\underline{\theta}=-\infty$ و $\overline{\theta}=+\infty$ آنگاه T برای برآورد $E_\theta(T)$ تحت تابع زیان درجه دوم مجاز است.

اثبات: با $\lambda=0$ و $\gamma=1$ هر دو انتگرال قضیه کارلین واگرا می‌شوند موقعی که $\theta^*\to\pm\infty$ ■

با توجه به نتیجه فوق در توزیع نرمال $N(\mu,\sigma^2)$ با σ^2 معلوم $f_\mu(\underset{\sim}{x})\propto e^{\frac{n\mu}{\sigma^2}\bar{x}+\frac{n\mu^2}{2\sigma^2}}$

$\bar{X}$ برای μ مجاز است و در توزیع دوجمله‌ای (مثال۲-۹) $\frac{X}{n}$ برای p مجاز است و در توزیع پواسون $f_\theta(\underset{\sim}{x})=e^{-\mu}e^{n\ln\mu\bar{x}}$ که $\theta=n\ln\mu\in(-\infty,+\infty)$ برآوردگر $\bar{X}$ برای μ مجاز است.

رابطه برآوردگرهای مجاز و مینیماکس

قضیه۲-۸: اگر δ یک تصمیم مجاز با مخاطره ثابت باشد انگاه δ یک تصمیم مینیماکس است.

اثبات: فرض کنید δ یک تصمیم مینیماکس نباشد در این صورت تصمیم دیگری مانند δ^* وجود دارد که

$$\sup_{\Theta} R(\theta,\delta^*)<\sup_{\Theta} R(\theta,\delta)=c$$

بنابراین

$$R(\theta,\delta^*)\le \sup R(\theta,\delta^*)<c=R(\theta,\delta)\quad \forall\theta\in\Theta$$

و این با مجاز بودن δ در تناقض است پس δ یک تصمیم مینیماکس است. ■

نتیجه ۲-۲: تحت فرضیات نتیجه ۲-۱ و تحت تابع زیان $\frac{(\delta-g(\theta))^2}{V_\theta(T)}$ برآوردگر T برای $g(\theta)=E_\theta(T)$ برآوردگر یکتای مینیماکس است.

اثبات: طبق نتیجه ۲-۱، T برای $g(\theta)$ تحت تابع زیان $(\delta-g(\theta))^2$ مجاز است پس تحت تابع زیان $\frac{(\delta-g(\theta))^2}{V_\theta(T)}$ نیز مجاز است. حال چون تحت این تابع زیان T دارای مخاطره ثابت است پس طبق قضیه ۲-۸، T برای $g(\theta)$ مینیماکس است. یکتایی این براورد از لم۲-۱ حاصل می‌شود(تابع زیان اکیدا محدب است). ■

قضیه ۲-۹: اگر δ_m یک تصمیم مینیماکس یکتا باشد آنگاه δ_m یک تصمیم مجاز است.

اثبات: اگر δ_m مجاز نباشد آنگاه تصمیم دیگری مانند δ وجود دارد که بر آن غلبه می‌کند و در نتیجه $\sup_{\Theta} R(\theta,\delta) \le \sup_{\Theta} R(\theta,\delta_m)$. یعنی δ نیز یک تصمیم مینیماکس است و این با یکتا بودن برآوردگر مینیماکس δ_m در تناقض است. پس δ_m یک تصمیم مجاز است. ■

مثال۲-۱۰: اگر $X \sim B(n,p)$ آنگاه طبق نتیجه ۲-۲ تحت تابع زیان $\frac{(\delta-p)^2}{pq}$ برآوردگر $\frac{X}{n}$ یک برآوردگر یکتای مینیماکس است پس طبق قضیه ۲-۹ یک برآوردگر مجاز است. ■

مثال۲-۱۱: فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از توزیع $N(\theta,\sigma^2)$ با σ^2 معلوم باشند. تحت تابع زیان درجه دوم

الف- در مثال‌های قبل دیدیم که $\bar{X}$ برآوردگر یکتای مینیماکس برای θ است پس $\bar{X}$ برای θ مجاز است.

ب- در مثال‌های قبل دیدیم که $\bar{X}$ برآوردگر مجاز با مخاطره ثابت $R(\theta,\bar{X}) = \frac{\sigma^2}{n}$ است پس برای θ مینیماکس است. ■

مثال۲-۱۲: (دو توزیع دو جمله‌ای) فرض کنید $X \sim B(m,p)$ و $Y \sim B(n,\pi)$ از یکدیگر مستقل باشند. نشان دهید که برآوردگر $a\frac{X}{m} + b\frac{Y}{n} + c$ تحت تابع زیان درجه دوم و تحت شرایط زیر برای p مجاز است.

$(a=1,\ b=c=0)$ یا $(0 \le a < 1,\ 0 \le c \le 1,\ 0 \le a+c \le 1,\ 0 \le b+c \le 1,\ 0 \le a+b+c \le 1)$

اثبات: فرض کنید برآوردگر دیگری مانند $\delta(X,Y)$ موجود باشد که

$$E\left(a\frac{X}{m} + b\frac{Y}{n} + c - p\right)^2 \ge E\left[(\delta(X,Y) - p)^2\right] \ \forall p$$

$$\Rightarrow \sum_{x=0}^{m}\sum_{k=0}^{n}\left(a\frac{x}{m} + b\frac{k}{n} + c - p\right)^2 P(X=x, Y=k) \quad (1)$$

$$\ge \sum_{x=0}^{m}\sum_{k=0}^{n}(\delta(x,k) - p)^2 P(X=x, Y=k)$$

حال اگر $\pi \to 0$ آنگاه (۱) نتیجه می‌دهد که

$p(X=x)$ برابر $k=0$ که به ازای $p(X=x,Y=k)=\binom{m}{x}p^x q^{m-x}\binom{n}{k}\pi^k(1-\pi)^{n-k}$ است.)

$$\sum_{x=0}^{m}(a\frac{x}{m}+c-p)^2P(X=x)\geq\sum_{x=0}^{m}(\delta(x,0)-p)^2P(X=x)$$

طبق مثال ۲-۹، $a\frac{x}{m}+c$ تحت شرایط این مثال مجاز است پس بایستی $\delta(x,0)=a\frac{x}{m}+c$ برای هر $x=0,1,...,m$. بنابراین عبارت $k=0$ در (۱) حذف می‌شود. عبارات دیگر شامل یک عامل π هستند که می‌تواند حذف شود و بنابراین (۱) نتیجه می‌دهد که

$$\sum_{x=0}^{m}\sum_{k=1}^{n}(a\frac{x}{m}+b\frac{k}{n}+c-p)^2\binom{m}{x}p^x q^{m-x}\binom{n}{k}\pi^{k-1}(1-\pi)^{n-k}\geq \sum_{x=0}^{m}\sum_{k=1}^{n}(\delta(x,k)-p)^2\binom{m}{x}p^x q^{m-x}\binom{n}{k}\pi^{k-1}(1-\pi)^{n-k} \quad (2)$$

دو مرتبه با $\pi\to 0$ رابطه (۲) نتیجه می‌دهد که

$$\sum_{x=0}^{m}(a\frac{x}{m}+\frac{b}{n}+c-p)^2P(X=x)\geq\sum_{x=0}^{m}(\delta(x,1)-p)^2P(X=x)$$

و چون تحت شرایط این مسئله $0\leq a+\frac{b}{n}+c\leq 1$ پس طبق مثال۲-۹ برآوردگر $a\frac{x}{m}+\frac{b}{n}+c$ برای p مجاز است. بنابراین (۲) نتیجه می‌دهد که برای هر $x=0,1,....,m$، $\delta(x,1)=a\frac{x}{m}+\frac{b}{n}+c$. به همین ترتیب اگر با استقراء روی k پیش برویم نتیجه می‌شود که $\forall x=0,...,m \atop k=0,1,....,n$ $\delta(x,k)=a\frac{x}{m}+b\frac{k}{n}+c$. پس برآوردگر $\delta(X,Y)=a\frac{X}{m}+b\frac{Y}{n}+c$ تحت شرایط مسئله برآوردگری مجاز برای p است.

در خارج از ناحیه داده شده در این مثال می‌توان نشان داد که برآوردگر $\delta(X,Y)$ غیرمجاز است(مسئله ۲,۲۱). پس $\delta(X,Y)$ مجاز است اگر و فقط اگر در شرایط این مسئله صدق کند. ■

1- Berger, J. O. (1985). Statistical Decision Theory and Bayesian Analysis, Second Edition, Springer-Verlag, New York.

2- Lehman E. L. and Casella G. (1998). Theory of Point Estimation, Second Edition, Springer, New York.

3- Casella, G. and Berger, R. L. (1990). Statistical Inference, Wodsworth & Brooks, California.

٤- بهبودیان، جواد (١٣٧٤). تصمیم آماری، انتشارات دانشگاه شیراز.