بنام خدا

استنباط آماری I

یادآوری: توزیع‌های مهم آماری: گسسته و پیوسته

- مطالبی درباره امید ریاضی، واریانس، گشتاورها، چندکها، تابع مولد گشتاور و قضایای حدی

- روش‌های برآورد یابی: روش گشتاورها، روش درستنمایی ماکزیمم، روش بیزی و مینیماکس (کم بیشینه)

- روش بیزی برآورد بیشینه

- تعریف بهترین برآوردگر: برآوردگر یکریخت با کمترین واریانس

- برآورد فاصله‌ای

1 Rohatgi, V. K. & Saleh
An Introduction to Probability Theory and Mathematical Statistics.

2 Casella & Berger
Statistical Inference.

— Lehman I, II

4/ Theory of Point Estimation

5/ Testing Statistical Hypotheses.

6- Bickel & Doksum
Mathematical Statistics

3 Ferguson, T. S. Mathematical Statistics

۷ نمره: آمار ریاضی - پارسیان

آمار ریاضی درباره چگونگی جمع‌آوری و دیگر ارزیابی اطلاعات مربوط به پدیده‌های تصادفی یا عدم قطعیت بحث می‌کند

برای تحلیل اطلاعات جمع‌آوری شده و نتیجه استنباط‌های احتمالی از یک مدل ریاضی داریم

۱- آزمایش تصادفی ۲- فضای نمونه ۳- پیشامد ۴- عملیات روی پیشامدها ۵- فضای پیشامدها

۶- فضای احتمال

- آزمایش تصادفی: عملیاتی که در آن چند نتیجه ممکن وجود دارد که تکرار کننده در تشابه عمل انجام آن یکسان است، ولی نتیجه آزمایش از پیش معلوم نیست نتیجه آزمایش را برآمد می‌نامیم.

- فضای نمونه: مجموعه‌ای که عضوهای آن همگی برآمدهای ممکن یک آزمایش تصادفی هستند

- پیشامد: هر زیرمجموعه A از فضای نمونه S را پیشامد می‌نامیم اگر بتوان برای آن احتمال تعریف کرد.

$A^c \in \mathcal{A} \Leftarrow A \in \mathcal{A}$ اگر -

$\bigcup_{i=1}^{\infty} A_i \in \mathcal{A} \Leftarrow A_1, A_2, \ldots \in \mathcal{A}$ اگر -

از سرانجام متناهی نتیجه می‌شود که

$\bigcap_{i=1}^{n} A_i \in \mathcal{A}$

فضای اندازه $(\Omega, \mathcal{A})$

فضای پیشامدها: مجموعه $\mathcal{M}$ از زیرمجموعه‌های Ω که تشکیل یک σ-جبر را می‌دهد فضای پیشامدها نامیده می‌شود و عضو A را که پیشامد می‌نامند.

فضای احتمال: در آزمایش تصادفی، با هر پیشامد A از فضای نمونه‌ای عدد $P(A)$ را که نمایانگر احتمال رخداد پیشامد A است متناظر می‌کنیم به طوری که در اصول زیر صدق کند:

- برای هر $A \in \mathcal{A}$، $P(A) \geq 0$
- $P(\Omega) = 1$
- اگر $A_1, A_2, \ldots$ دنباله‌ای از پیشامدهای دوبه‌دو ناسازگار از پیشامدها باشد $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

سه‌تایی $(\Omega, \mathcal{A}, P)$ را فضای احتمال یا مدل احتمال می‌نامیم.

$P : \mathcal{A} \to [0,1]$ یک تابع مجموعه‌ای است

متغیر تصادفی - تابع احتمال (چگالی) - تابع توزیع -

فضای اندازه‌پذیر: میدان بورل $\mathcal{B}$ در $\mathbb{R}$ کوچکترین میدان شامل همه بازه‌های $\mathbb{R}$ است شامل همه فاصله‌های باز (a,b) که در $\mathbb{R}$ است.

میدان بورل $\mathcal{B}^k$ در $\mathbb{R}^k$ کوچکترین میدان شامل همه زیرمجموعه‌های $\mathbb{R}^k$ است که شامل همه k-مستطیل‌های باز $\{(x_1, \ldots, x_k) : a_i < x_i < b_i, i = 1, \ldots, k\}$ در $\mathbb{R}^k$ است.

متغیر تصادفی: گیریم X از Ω به $\mathbb{R}$ که تابع اندازه‌پذیر است اگر برای هر $B \in \mathcal{B}$ مجموعه $X^{-1}(B) = \{\omega : X(\omega) \in B\}$ پیشامد باشد یعنی در $\mathcal{A}$ باشد. پیشامد $\{\omega : X(\omega) \in B\}$ را با $(X \in B)$ نشان می‌دهیم و احتمال این پیشامد را با $P(X \in B)$ نشان می‌دهیم.

$\Omega = \{H, T\}$ ، $\mathcal{A}$ σ-جبر مجموعه‌های Ω $\quad X(H) = 1$ ، $X(T) = 0$

$B = (-\infty, x]$ $\quad x \in \mathbb{R}$

$X^{-1}(B) = X^{-1}((-\infty, x]) = \begin{cases} \emptyset & x < 0 \\ \{T\} & 0 \leq x < 1 \\ \{H, T\} & x \geq 1 \end{cases}$

$\Omega = \{TT, HT, TH, TT\}$

تعداد شیر در پرتاب دو سکه $X(\omega) =$

$X(HH) = 2$ ، $X(HT) = X(TH) = 1$ ، $X(TT) = 0$

$X^{-1}(-\infty, x] = \begin{cases} \emptyset & x < 0 \\ \{TT\} & 0 \leq x < 1 \\ \{HT, TH, TT\} & 1 \leq x < 2 \\ \{HH, HT, TH, TT\} & x \geq 2 \end{cases}$

numerical-valued

3

$\{x : A\}$ مقدار تابع A در نقاط x از دامنه تعریف A است. A تابعی است که $\{x : A\}$ را به $A(x)$ نگاشت می‌کند.

① خانواده توزیع چرخشی:

۱- برنولی ۲- دوجمله‌ای ۳- هندسی ۴- دوجمله‌ای منفی ۵- دوجمله‌ای ۶- چندجمله‌ای ۷- فوق هندسی ۸- چندجمله‌ای فوق هندسی

② خانواده توزیع سری توانی: Power Series Dists.

متغیر تصادفی X را دارای توزیع متعلق به خانواده سری توانی گوییم هرگاه احتمال المنوریت در نقطه x به صورت

$$f_\theta(x) = \frac{a(x)\,\theta^x}{c(\theta)} \quad ; \quad x = 0, 1, 2, \dots \qquad \theta > 0$$

که در آن $a(x)$ تابعی نامنفی از x و

$$c(\theta) = \sum_{x=0}^{\infty} a(x)\,\theta^x$$

تابعی از θ است.

این خانواده از یک توزیع یک پارامتری خود را خانواده‌ای که تابعی از آن نامنفی باشد.

خانواده توزیع‌های گسسته است و به‌ازای هر انتخاب از $a(x)$ و $c(\theta)$ یک توزیع گسسته حاصل می‌شود.

توزیع دوجمله‌ای

$$f_p(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, \dots, n$$

$$a(x) = \begin{cases} \binom{n}{x} & x = 0, 1, 2, \dots, n \\ 0 & x = n+1, \dots \end{cases} \qquad , \theta = \frac{p}{1-p}$$

$$c(\theta) = \sum_{x=0}^{\infty} a(x)\,\theta^x = \sum_{x=0}^{n} \binom{n}{x} \theta^x = (1+\theta)^n = \left(1 + \frac{p}{1-p}\right)^n$$

$$= (1-p)^{-n}$$

$$f_\theta(x) = \frac{a(x)\,\theta^x}{c(\theta)} = \begin{cases} \dfrac{\binom{n}{x}\left(\frac{p}{1-p}\right)^x}{(1-p)^{-n}} & x = 0, 1, \dots, n \\ 0 & x = n+1, \dots \end{cases}$$

$$= \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, \dots, n \\ 0 & \text{سایر نقاط} \end{cases} \qquad \sim B(n, p)$$

توزیع دوجمله‌ای منفی

$$f_p(x) = \binom{x+r-1}{x} p^x (1-p)^x \qquad x = 0, 1, 2, \dots$$

$$\text{if } a(x) = \binom{x+r-1}{x} \quad x = 0, 1, 2, \dots \quad , \theta = 1-p$$

$$c(\theta) = \sum_{x=0}^{\infty} a(x)\,\theta^x = \sum_{x=0}^{\infty} \binom{x+r-1}{x} \theta^x \overset{?}{=} (1-\theta)^{-r} = p^{-r}$$

$$f_\theta(x) = \frac{a(x)\theta^x}{c(\theta)} = \frac{\binom{x+r-1}{x}(1-p)^x}{p^{-r}} = \binom{x+r-1}{x} p^r (1-p)^x \quad x = 0, 1, \ldots$$

- توزیع پوآسن Poisson Dist.

$$f_\theta(x) = \frac{e^{-\theta}\theta^x}{x!} \quad x = 0, 1, \ldots \qquad a(x) = \frac{1}{x!} \quad x = 0, 1, \ldots$$

$$c(\theta) = \sum_{x=0}^{\infty} \frac{1}{x!}\theta^x = e^\theta$$

$$f_\theta(x) = e^{-\theta}\theta^x / x! \quad x = 0, 1, \ldots$$

Poisson dist.

- توزیع پوآسن کوتاه شده (قطع شده) از صفر Truncated at zero

فرض کنید $Y \sim P(\theta)$ و متغیر تصادفی X حاصل مشاهده‌ی Y باشد به شرطی که Y مقدار مثبت اختیار کند. به عبارت دیگر X همان Y است به شرطی که Y صفر مشاهده نشود. در این صورت X دارای توزیع پوآسن کوتاه شده (قطع شده) است.

$$f_\theta(x) = P(X = x) = P(Y = x \mid Y > 0) = \frac{P(Y = x, Y > 0)}{P(Y > 0)} = \frac{P(Y = x)}{1 - P(Y = 0)} \quad ; x > 0$$

$$= \frac{e^{-\theta}\theta^x}{x!} \Big/ 1 - e^{-\theta} \quad ; x > 0$$

$$= \frac{\theta^x}{(e^\theta - 1)\, x!} \quad ; x = 1, 2, \ldots$$

(۳) خانواده توزیع‌های تکسونی

تعریف: متغیر تصادفی X را دارای توزیع تکسونی گوییم

$$f_\theta(x) = q^{k(x,\theta)} - q^{k(x+1,\theta)} \quad ; x = 0, 1, \ldots \quad , 0 < q < 1, \; k(0,\theta) = 0, \; k(\infty,\theta) = 0$$

به شرطی که $f_\theta(x)$ در تعریف به ازای حداقل یک x اکیداً مثبت باشد. این خانواده کوچک‌ترین خانواده‌ای از توزیع‌های گسسته است که شامل توزیع دوجمله‌ای نیز باشد.

مثال ۱

توزیع هندسی: $k(x,\theta) = x$ قرار دهید

$$f(x) = q^x - q^{x+1} = q^x(1-q) = pq^x \quad x = 0, 1, \ldots$$

5

تذکر: اگر $f_{\theta=p}(x) = p^x - p^{(x+1)\theta}$ آنگاه x^θ ...

$= (1-p)^{x^\theta} - (1-p)^{(x+1)\theta}$

...

(٤) خانواده تقسیم مکانی - مقیاسی

(Location-Scale family)

تعریف: خانواده تقسیم‌های پیوسته x را متقارن خانواده تقسیم مکانی گوییم هرگاه تابع چگالی احتمال آن به صورت زیر باشد

$$f_\theta(x) = f(x-\theta)$$

مثال $X \sim N(\theta, 1)$ آنگاه

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}$$

تعریف: تقسیم پیوسته نامنفی پیوسته x را متقارن خانواده تقسیم مقیاسی گوییم هرگاه تابع چگالی احتمال آن به صورت زیر باشد

$$f_\theta(x) = \frac{1}{\theta} f(x/\theta)$$

مثال $X \sim N(0, \theta^2)$

$$f_\theta(x) = \frac{1}{\theta\sqrt{2\pi}} e^{-\frac{x^2}{2\theta^2}} = \frac{1}{\theta}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x/\theta)^2}$$

تعریف کلی: اگر تابع چگالی احتمال پیوسته x به صورت زیر باشد آنگاه تقسیم X را متقارن خانواده تقسیم‌های مکانی - مقیاسی گوییم.

$$f_{(\mu,\sigma)}(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

که در آن $\theta = (\mu, \sigma)$، μ پارامتر مکان، σ پارامتر مقیاس است.

تذکر: اگر X_0 دارای تابع توزیع $F_0(x)$، چگالی $f_0(x)$ باشد و θ پارامتر مکانی باشد

بازنویسی $X = \mu + \sigma X_0$ داریم

$F(x) = P(X \le x) = P(\mu + \sigma X_0 \le x)$

$= P(X_0 \le \frac{x-\mu}{\sigma}) = F_0(\frac{x-\mu}{\sigma})$

$f(x) = \frac{dF(x)}{dx} = \frac{1}{\sigma} f_0(\frac{x-\mu}{\sigma})$

تابع توزیع X خانواده مقیاس-مکان تشکیل داد.

μ را پارامتر مکان و σ را پارامتر مقیاس گوییم و قرارداد $\mu = 0$, $\sigma = 1$

توزیع نرمال استاندارد

$N(\mu, \sigma^2)$ توزیع نرمال

$DE(\alpha, \beta)$ توزیع نمایی دوگانه

$\frac{1}{2\beta} e^{\frac{-|x-\alpha|}{\beta}}$ $\quad x \in \mathbb{R}, \alpha \in \mathbb{R}, \beta > 0$

توزیع کوشی، کلموگروف، توزیع لاپلاس

تمرینات ($\beta = \alpha$, $\beta = -\alpha$) نیز جزو خانواده مکان-مقیاس توزیع نرمال هستند

---

② خانواده نمایی: Exponential Dist. family

تعریف: برای بردار تصادفی $\underline{X} = (X_1, X_2, \dots, X_n)$ اگر توابع Q, D روی $\Theta \subseteq \mathbb{R}$, توابع T, S روی $\mathbb{R}^n$ وجود داشته باشند طوری که تابع چگالی احتمال $\underline{X}$ به صورت زیر نوشته شود، گوییم توزیع $\underline{X}$ متعلق خانواده توزیع نمایی است.

$$f_\theta(\underline{x}) = \exp\{Q(\theta) T(\underline{x}) + D(\theta) + S(\underline{x})\}$$

یا به صورت معادل

$$f_\theta(\underline{x}) = \exp\{Q(\theta) T(\underline{x})\} B(\theta) h(\underline{x})$$

$h(\underline{x}) = e^{S(\underline{x})}$, $B(\theta) = e^{D(\theta)}$.

- اگر توزیع X متعلق به خانواده نمایی باشد توزیع توأم نمونه تصادفی $X_1, \dots, X_n$ نیز خانواده نمایی تشکیل می‌دهد.

- اگر توزیع Y تابعی از X باشد یعنی $Y = h(X)$، چنانچه توزیع X خانواده نمایی تشکیل دهد، توزیع Y نیز خانواده نمایی تشکیل می‌دهد. (نکته آزمون فرضیات آماری کنکور)

کنکور TSH

خانواده‌ای نمایی k پارامتری

اگر توزیع $\underline{X}$ بصورت زیر باشد

$$f_{\underline{\theta}}(\underline{x}) = \exp\left\{\sum_{i=1}^{k} Q_i(\underline{\theta}) T_i(\underline{x}) + D(\underline{\theta}) + S(\underline{x})\right\} \quad \underline{\theta} = (\theta_1, \dots, \theta_r) \quad *$$

که در آن $Q_1(\cdot), \dots, Q_k(\cdot)$، D توابعی حقیقی روی Θ که $\Theta \subseteq \mathbb{R}^r$، و $T_1, \dots, T_k$ و S توابعی حقیقی از فضای نمونه‌ای $\underline{X}$ هستند.

آنگاه توزیع $\underline{X}$ متعلق به خانواده‌ای k پارامتری است.

① - اگر $k=1$، خانواده‌ای یک پارامتری حاصل می‌شود.

② - فضای نمونه‌ای $\underline{X}$ و $\underline{\theta}$ هیچ نقشی ندارند. تقلیل پذیر است.

③ - عموماً $r \le k$. اگر $r=k$ خانواده نمایی را رتبه کامل گوییم.

④ - Q_i و T_i توابعی یک به یک نیستند.

⑤ - n مورد از نمونه‌ها نقشی ندارند.

⑥ - کمترین k یک نباشد، به این گونه عددی کوچکتر مثل k که بتوان تابع $\underline{X}$ را بصورت * نشان داد تعریف خانواده نمایی است.

⑦ - با فرض $\eta_i = Q_i(\theta)$، می‌توان $f_{\underline{\theta}}$ در * را بصورت زیر نوشت:

$$f_{\underline{\eta}}(\underline{x}) = \exp\left\{\sum_{i=1}^{k} \eta_i T_i(\underline{x}) + D^*(\underline{\eta}) + S(\underline{x})\right\}$$

که در آن $\underline{\eta} = (\eta_1, \dots, \eta_k)$، η_i ها را پارامتر طبیعی می‌نامیم.

مثال: اگر $\theta \sim U(0,1)$ و توزیع X متعلق به خانواده‌ای یک پارامتری زیر

$$f_\theta(x) = \binom{5}{x} \theta^x (1-\theta)^{5-x} \qquad x = 0, 1, \dots, 5$$

$$= \binom{5}{x} \left(\frac{\theta}{1-\theta}\right)^x (1-\theta)^5 \qquad x = 0, 1, \dots, 5$$

$$= \underbrace{\exp\left\{x \log \frac{\theta}{1-\theta}\right\}}_{T(x)\ \ Q(\theta)} \cdot \underbrace{(1-\theta)^5}_{B(\theta)} \cdot \underbrace{\binom{5}{x}}_{h(x)} \cdot \underbrace{I_{\{0,1,\dots,5\}}(x)}_{S(x)}$$

$n=1$
$k=1$
$r=1$

مثال ۱ – اگر متغیر اول x بصورت زیر باشد

x	۱	۲	۳
$p_\theta(x)$	$\theta/2$	$1/2$	$\frac{1-\theta}{2}$

در نمونه تصادفی x مقدار زیر بدست آمده است.

مطلوب است تابع درستنمایی:

$$T_1(x)=\begin{cases}1 & x=1\\ 0 & x\neq 1\end{cases}$$ تعداد x های برابر ۱

$$f_\theta(x_i)=\begin{cases}\theta/2 & x=1\\ 1/2 & x=2\\ \frac{1-\theta}{2} & x=3\end{cases}=\begin{cases}(\theta/2)^{T_1(x)}\cdot(1/2)^{T_2(x)}\cdot\left(\frac{1-\theta}{2}\right)^{T_3(x)} & x=1,2,3\\ 0 & \text{سایر نقاط}\end{cases}$$

$$=\begin{cases}\left(\frac{\theta}{2}\right)^{T_1(x)}\cdot\left(\frac{1}{2}\right)^{T_2(x)}\cdot\left(\frac{1-\theta}{2}\right)^{1-T_1(x)-T_2(x)} & x=1,2,3\\ 0 & \text{سایر نقاط}\end{cases}$$

$$=\exp\left\{\underbrace{\log\frac{\theta}{1-\theta}}_{Q_1(\theta)}\cdot T_1(x)+\underbrace{\log\left(\frac{1}{1-\theta}\right)}_{Q_2(\theta)}T_2(x)\right\}\cdot\underbrace{\frac{1-\theta}{2}}_{B(\theta)}\cdot\underbrace{I_{\{1,2,3\}}(x)}_{h(x)}$$

رتبه کامل نمایی است.

$n=1,\ k=2,\ r=1$

– رابطه زیر برای محاسبه امید ریاضی بسیار مفید است.

$$E(T(x))=\frac{-D'(\theta)}{Q'(\theta)}$$

اثبات این رابطه با استفاده از $E\left(\frac{\partial}{\partial\theta}\log f_\theta(x)\right)=0$ بدست می‌آید.

$$E(T_i(x))=\frac{-\partial D(\theta)}{\partial\theta_i}\quad\text{و}\quad Q_i(\theta)=\theta_i$$

$$\text{Cov}(T_i(x),T_j(x))=-\frac{\partial^2 D(\theta)}{\partial\theta_i\partial\theta_j}$$

اصول کاهش واریانس:

فرض کنید T یک آماره برای θ باشد و فرض کنید T یک آماره کافی برای θ باشد. آنگاه $E(T|S)$ یک تابع از S است که آماره‌ای نااریب برای θ است و واریانس آن از واریانس T کوچکتر یا مساوی است.

مثال: فرض کنید $X_1,X_2,\dots,X_n$ نمونه‌ای تصادفی از توزیع $N(\mu,\sigma^2)$ باشد. $\bar{X}$ و S^2 آماره‌های کافی برای μ و σ^2 هستند. $\bar{X}$ برآوردگر نااریب μ است و S^2 برآوردگر نااریب σ^2 است. $\bar{X}$ و S^2 مستقل هستند.

$$E(\bar{X}|S^2)=\mu$$

$$E(S^2|\bar{X})=\sigma^2$$

$$\text{Var}(\bar{X})=\frac{\sigma^2}{n}$$

$$\text{Var}(S^2)=\frac{2\sigma^4}{n-1}$$

مثال: فرض کنید $X_1,X_2,\dots,X_n$ نمونه‌ای تصادفی از توزیع $N(\mu,\sigma^2)$ باشد. $\bar{X}$ و S^2 آماره‌های کافی برای μ و σ^2 هستند. $\bar{X}$ برآوردگر نااریب μ است و S^2 برآوردگر نااریب σ^2 است. $\bar{X}$ و S^2 مستقل هستند.

$$E(\bar{X}|S^2)=\mu$$

$$E(S^2|\bar{X})=\sigma^2$$

$$\text{Var}(\bar{X})=\frac{\sigma^2}{n}$$

$$\text{Var}(S^2)=\frac{2\sigma^4}{n-1}$$

a correct definition of the concept involved in this discussion

تعریف دقیقی که مفهوم گفته شده را شامل می‌شود نیازمند اصول ساده‌ای است

تعریف: فرض کنید $\underline{X} = (X_1, \dots, X_n)$ نمونه‌ای تصادفی از جامعه‌ای با توزیع $P_\theta(x)$, $\theta \in \Theta$ باشد. آماره $T = T(\underline{X})$ را کفایت گوییم برای $\theta \in \Theta$ اگر توزیع شرطی $\underline{X}$ به شرط $T = t$ برای هر t به θ بستگی نداشته باشد.

توزیع توأم $(X_1, \dots, X_n)$ تابعی از θ است.

برای تعریف بالا، اگر T برای θ کفایت باشد یعنی با دانستن مقدار T نمونه‌گیری بیشتر (یا اطلاعات بیشتر) درباره θ نمی‌تواند کمکی به برآورد θ کند.

به عبارت دیگر، آماره کفایت تمام اطلاعات نمونه از یک خانواده توزیع را در خود دارد. بنابراین اطلاعاتی که آماره کفایت از نمونه حذف می‌کند، برای برآورد θ بی‌ارزش است.

مثال: فرض کنید $X_1, X_2, \dots, X_n \overset{iid}{\sim} b(1, p)$

در حالتی که یک نمونه تصادفی n تایی از یک توزیع برنولی داریم.

برای این توزیع می‌دانیم که $\sum_{i=1}^n X_i$ یک آماره کفایت برای p است.

$$T(\underline{X}) = \sum_{i=1}^n X_i \sim B(n, p)$$

$$P(X_1 = x_1, \dots, X_n = x_n \mid \sum_{i=1}^n X_i = t) = \frac{P(X_1 = x_1, \dots, X_n = x_n, T = t)}{P(T = t)}$$

$$= \begin{cases} \dfrac{P(X_1 = x_1, \dots, X_n = x_n)}{P(T = t)} & ; \ \sum x_i = t \\ 0 & \text{در غیر این صورت} \end{cases}$$

$$= \begin{cases} \dfrac{p^{\sum x_i}(1-p)^{n - \sum x_i}}{\binom{n}{t} p^t (1-p)^{n-t}} = \dfrac{1}{\binom{n}{t}} & \sum x_i = t \\ 0 & \text{در غیر این صورت} \end{cases}$$

X_1, X_2 ~ $p(\lambda)$ $\Rightarrow X_1+X_2$ is sufficient for λ

$$P(X_1=x_1, X_2=x_2 \mid X_1+X_2=t) \stackrel{?}{=} \binom{t}{x_1}\left(\frac{1}{2}\right)^t \quad ; \quad x_1=0,1,\dots,t,\ x_1+x_2=t$$

مثال: نمونه‌ای از X_1, X_2 که $T=X_1+X_2$ برای λ کفایت نباشد. (غیرتابعی)

$$P(X_1=0, X_2=1 \mid X_1+X_2=1) = \frac{P(X_1=0, X_2=1)}{P(X_1+X_2=1)}$$

$$= \frac{e^{-\lambda}(\lambda e^{-\lambda})}{P(X_1=1, X_2=0)+P(X_1=0, X_2=1)} = \frac{e^{-\lambda}(\lambda e^{-\lambda}) = \lambda e^{-2\lambda}}{\lambda e^{-2\lambda} + \lambda e^{-2\lambda}}$$

$$= \frac{1}{2} .$$

اثبات

فرض کنید $X_1, \dots, X_n$ نمونه‌ای از توزیع با تابع چگالی (یا احتمال) $f_\theta(x)$ باشد. آماره $T(X_1, \dots, X_n)$ برای θ کفایت است اگر و تنها اگر توابعی مانند g و h وجود داشته باشند به طوری که

$$f_\theta(x_1, \dots, x_n) = g_\theta(t, \theta)\, h(x_1, \dots, x_n)$$

که در آن g_θ فقط از طریق $T(x_1, \dots, x_n)$ به $x_1, \dots, x_n$ وابسته است.

فضیه: (فضیه تجزیه، تجزیه‌پذیری، معیار تجزیه) : Factorization criterion theorem

آن را تبدیل کنید و سپس با استفاده از تابع تبدیل ... را بیابید

// $f_{x(n)} = g(\theta, s(x)) \cdot h(x) \Rightarrow g_\theta(\theta, u(x_{(n)})) \cdot h(x) = g^*(\theta, u(x_{(n)})) \cdot h(x)$

اثبات: تابع تبدیل یک‌به‌یک $\psi = k(T)$ و T را در نظر بگیرید

$$f_\theta(x) = g_\theta(t, h(x)) = g_\theta(k^{-1}(\psi), h(x))$$
$$= g^*_\theta(\psi) \cdot h(x)$$

مثال ۱: اگر $X_1, \dots, X_n \overset{iid}{\sim} B(1, \theta) \Rightarrow T = \sum_{i=1}^n X_i$ is sufficient for θ

$T_1 = (X_1, \sum_{i=2}^n X_i)$, $T_2 = (X_1 X_2, X_3, \sum_{i=3}^n X_i)$

مثال ۲: $X_1, X_2, \dots, X_n \overset{iid}{\sim} U(0, \theta)$ که $\theta \in \mathbb{N}$

$X_{(n)}$ آماره‌ای کفایتی برای θ است.

$$P(X = x_i) = \frac{1}{N} \quad x = 1, 2, \dots, N, \quad i = 1, \dots, n$$

$$f_\theta(\underline{x}) = \prod_{i=1}^n f_\theta(x_i) = \frac{1}{N^n} \quad 1 \le x_1, x_2, \dots, x_n \le N$$

$$= \frac{1}{N^n} \Phi(1, \min x_i) \cdot \Phi(\max x_i, N) \qquad \Phi(a,b) = \begin{cases} 1 & b \ge a \\ 0 & b < a \end{cases}$$

$$= \frac{1}{N^n} \Phi(\max x_i, N) \cdot \Phi(1, \min x_i)$$

$$g_N(w, \max x_i) \cdot h(x)$$

$X_{(n)} = \max(X_i)$ is sufficient for N

مثال ۳: $X_1, X_2, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ که μ مجهول

$\underline{T} = (T_1, T_2) = (\sum x_i, \sum x_i^2)$ is sufficient for (μ, σ^2)

یعنی $(\bar{x}, s^2)$ نیز برای (μ, σ^2) کفایتی است.

~~این برای است که $\bar{X}$ برای μ، بهترین تخمین‌گر نااریب است~~

~~$(\sum x_i, \sum x_i^2)$ برای (μ, σ^2) بهترین تخمین‌گر نااریب~~

~~است.~~

~~اگر μ مجهول باشد~~

~~اگر μ معلوم باشد، $\frac{1}{n}\sum (x_i - \mu)^2$ برای σ^2 بهترین تخمین‌گر نااریب است.~~

~~مثال ۸ صفحه ۱۸۵~~

$X_1, \dots, X_n \sim U(-\theta/2, \theta/2)$ $\theta > 0$

$$f_\theta(x) = \begin{cases} 1/\theta & x \in (-\theta/2, \theta/2) \\ 0 & \text{o.w.} \end{cases}$$

$$f_\theta(\underline{x}) = \left(\frac{1}{\theta}\right)^n \quad -\theta/2 < x_1, \dots, x_n < \theta/2$$

$$= (1/\theta)^n \, u(x_{(1)} + \theta/2) \cdot u(\theta/2 - x_{(n)})$$

$$= g(\theta, x_{(1)}, x_{(n)}) \times 1 \qquad T = (X_{(1)}, X_{(n)})$$

بنابراین T کفایتی است.

کفایتی مینیمال

می‌دانیم که توابعی از آماره‌ی کفایتی هم کفایتی هستند. اگر $T(\underline{x})$ کفایتی باشد، آن‌گاه T تابعی از T است. یعنی T اطلاعات θ را خلاصه‌تر می‌کند.

summarization

ـ افراز Π_1 را ظریف‌تر از افراز Π_2 گوییم هرگاه هر مجموعه‌ی افراز Π_1 زیرمجموعه‌ای از یکی از مجموعه‌های افراز Π_2 باشد.

اگر افراز Π_1 ظریف‌تر از Π_2 باشد، آن‌گاه $T_1 = S(T_2)$ برای یک تابع S.

مثال

آماره‌ی T^* را یک آماره‌ی کفایتی مینیمال گوییم هرگاه

۱- T^* کفایتی باشد.

۲- T^* نسبت به هر آماره‌ی کفایتی دیگر ظریف‌تر باشد.

13

قضیه فیشر-نیمن: فرض کنید $U(X)$ و $N(X)$ دو آماره باشند که رابطه‌ی یک‌به‌یک D وجود دارد به‌طوری که برای هر X داریم $N(X) = D(U(X))$. اگر N کفایتی برای θ باشد آنگاه U نیز کفایتی برای θ است.

$$\forall \lambda_1, \lambda_2 \in D : U(\lambda_1) = U(\lambda_2) \Longrightarrow N(\lambda_1) = N(\lambda_2)$$

قضیه: فرض کنید X دارای تابع چگالی یا احتمال $f_\theta(x)$ باشد. اگر برای هر $\theta_0 \in \Theta$ نسبت $\frac{f_\theta(x)}{f_{\theta_0}(x)}$ تابعی از x مانند $T(x)$ باشد آنگاه $T(X)$ یک آماره کفایتی مینیمال است.

تعریف: فرض کنید $\mathcal{P} = \{f_\theta(x) : \theta \in \Theta\}$ خانواده‌ای از توابع چگالی یا احتمال باشد. گوییم $\mathcal{P}$ کامل است هرگاه برای هر تابع g که $E_\theta[g(X)] = 0$ برای هر $\theta \in \Theta$ نتیجه شود $g(x) = 0$ به‌جز روی مجموعه‌ای با احتمال صفر.

قضیه: اگر T کفایتی مینیمال باشد و $\mathcal{P}_T = \{f_{T,\theta}(t) : \theta \in \Theta\}$ کامل باشد، آنگاه T آماره‌ای کامل است.

قضیه: اگر T کفایتی کامل باشد و φ تابعی یک‌به‌یک باشد، آنگاه $\varphi(T)$ نیز کفایتی کامل است.

قضیه: اگر T کفایتی کامل و $S = h(T)$ آماره‌ای ناریب برای θ باشد، آنگاه S یکتا است.

قضیه: اگر T کفایتی کامل و $S = h(T)$ آماره‌ای ناریب برای $\tau(\theta)

اثبات: (شرط لازم) فرض کنید T کفایت باشد. بنابراین بنا به تعریف تابع
$f_\theta(x) = g_\theta(T(x)) \cdot h(x)$
$f_{\theta_0}(x) = g_{\theta_0}(T(x)) \cdot h(x)$

$$\frac{f_\theta(x)}{f_{\theta_0}(x)} = \frac{g_\theta(T(x))}{g_{\theta_0}(T(x))} = g^*_\theta(T(x))$$

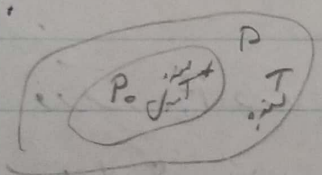
شرط کافی: فرض کنید که نسبت $\frac{f_\theta(x)}{f_{\theta_0}(x)}$ تنها از طریق $T(x)$ به x بستگی داشته باشد. لذا تابعی مانند g وجود دارد که

$$\frac{f_\theta(x)}{f_{\theta_0}(x)} = g_\theta(T(x)) \quad \text{یا} \quad f_\theta(x) = g_\theta(T(x)) \cdot \underbrace{f_{\theta_0}(x)}_{h(x)}$$

پس بنا به تعریف عاملی T کفایت است.

(P₀ ⊂ P)

قضیه: فرض کنید P و P_0 دو خانواده از توزیع‌های احتمال با بردار θ باشند بطوری که برای هر x مجموعه $\{\theta : f_\theta(x) > 0\}$ یکسان باشد. اگر T^* در P_0 کفایت باشد و P کفایت باشد آنگاه T^* در P کفایت است.



اثبات: فرض کنید T هر کفایت باشد و نشان دهید که T برای خانواده توزیع‌های P کفایت است.
چون $P_0 \subset P$ پس T نیز کفایت است و چون T^* تابعی از T است پس T^* نیز کفایت است.

لم: اگر P_0 یک خانواده متناهی از توزیع‌ها باشد که $\Theta = \{\theta_0, \theta_1, \dots, \theta_k\}$ و T کفایت باشد آنگاه

$$T^*(x) = \left(\frac{f_1(x)}{f_0(x)}, \frac{f_2(x)}{f_0(x)}, \dots, \frac{f_k(x)}{f_0(x)}\right)$$

نیز کفایت است.

اثبات: اولاً T^* تابعی از T است زیرا برای $\theta \neq \theta_0$ داریم

$$f_\theta(x) = \underbrace{\frac{f_\theta(x)}{f_{\theta_0}(x)}}_{g_\theta(T^*(x))} \cdot \underbrace{f_{\theta_0}(x)}_{h(x)}$$

پس T^* کفایت است.

حال اگر T یک آماره کفایت مینیمال باشد آنگاه برای هر $\theta \neq \theta_0$، $\theta_0 = 0$

$\frac{f_\theta(x)}{f_0(x)}$ تابعی از T است و چون T^* نسبت به T کفایت دارد پس $\frac{f_\theta(x)}{f_0(x)}$ تابعی از T^* است.

بنابراین به ازای هر $\theta \neq 0$ و $\theta_0 \neq 0$ دلخواه، T^* کفایت T است پس T^* مینیمال است.

T^* تابعی از T است پس $\frac{f_\theta(x)}{f_0(x)}$ تابعی از T^* است.

$$f_\theta(x) = \prod_{i=1}^{n} f_1(x_i)^{I_{(1)}(\theta)} \cdot f_2(x_i)^{I_{(2)}(\theta)} \cdots f_k(x_i)^{I_{(k)}(\theta)} \cdot f_0(x_i)^{1-I_{(1)}(\theta)-I_{(2)}(\theta)-\cdots-I_{(k)}(\theta)} \cdot I_{\{\theta_1,\theta_2,\ldots,\theta_k\}}(\theta)$$

$$= \prod_{i=1}^{n} \left(\frac{f_1(x_i)}{f_0(x_i)}\right)^{I_{(1)}(\theta)} \cdots \left(\frac{f_k(x_i)}{f_0(x_i)}\right)^{I_{(k)}(\theta)} \cdot f_0(x) \cdot I_{\{\theta_1,\theta_2,\ldots,\theta_k\}}(\theta)$$

$$= \left(\frac{f_1(x)}{f_0(x)}\right)^{I_{(1)}(\theta)} \cdots \left(\frac{f_k(x)}{f_0(x)}\right)^{I_{(k)}(\theta)} \cdot I_{\{\theta_1,\theta_2,\ldots,\theta_k\}}(\theta) \cdot f_0(x)$$

تابع کمکی مینیمال کفایت است.

مثالی که ثابت می‌کند کفایت کامل نیست:

$X_1, X_2, \ldots, X_n$ نمونه‌ای از $N(\theta, 1)$

توزیع P_θ خانواده‌ای است که تحت آن $\theta = 0$ یا $\theta = 1$ است.

برای $\frac{f_1(x)}{f_0(x)}$ تابعی از θ است یعنی

$$\frac{f_1(x)}{f_0(x)} = \frac{e^{-\frac{1}{2}\sum(x_i-1)^2}}{e^{-\frac{1}{2}\sum x_i^2}} = e^{\sum x_i - \frac{n}{2}}$$

هر تابعی که هم از $\bar{X}$ و هم از S^2 مستقل باشد تابع ثابت است.

$\bar{X}$ تابعی از θ است و S^2 تابعی از θ نیست پس $\bar{X}$ و S^2 مستقل‌اند.

Theorem 6.2

Casella & Berger (Ancillary Statistics)

تعریف: آماره R را کمکی گوییم که توزیع آن به θ بستگی نداشته باشد.

قضیه: اگر T کفایت کامل باشد و R کمکی باشد آنگاه T و R مستقل‌اند.

برهان: فرض کنید A یک پیشامد دلخواه باشد و $P(R \in A) = c$ (که c به θ بستگی ندارد).

$g(T) = P(R \in A | T)$ را در نظر بگیرید.

$E_\theta[g(T)] = E_\theta[P(R \in A | T)] = P_\theta(R \in A) = c$

$E_\theta[g(T) - c] = 0$ برای هر θ

چون T کفایت کامل است $g(T) - c = 0$ یعنی $g(T) = c$

$P(R \in A | T) = P(R \in A)$ یعنی R و T مستقل‌اند.

مثال: $X_1, X_2, X_3 \overset{iid}{\sim} N(\theta, 1)$ ، $A(\underline{X}) = X_1 + X_2 - 2X_3$

ثابت کنید $A(\underline{X}) \sim N(0, 6)$

مثال: $X_1, \dots, X_n \overset{iid}{\sim} U(\theta, \theta+1)$ ، $A(\underline{X}) = X_{(n)} - X_{(1)} \quad ; \theta \in \mathbb{R}$

ثابت کنید $A(\underline{X})$ ancillary است.

قضیه: if $X_1, \dots, X_n \overset{iid}{\sim} U(0,1) \Rightarrow X_{(i)} \sim Beta(i, n-i+1)$

اثبات: $X_1 - \theta, X_2 - \theta, \dots, X_n - \theta \overset{iid}{\sim} U(0,1) \Rightarrow X_{(i)} - \theta \sim Beta(i, n-i+1)$

$$\Rightarrow \begin{cases} X_{(1)} - \theta \sim Beta(1, n) \\ X_{(n)} - \theta \sim Beta(n, 1) \end{cases} \Rightarrow A(\underline{X}) = X_{(n)} - X_{(1)} \sim Beta(n-1, 2)$$

به طور کلی: $X_{(i)} - X_{(j)} \sim Beta(i-j, n-(i-j)+1)$

حل:

$$f_\theta(x_i) = \begin{cases} 1 & \theta < x_i < \theta + 1 \\ 0 & \text{o.w} \end{cases}$$

$$f_\theta(\underline{x}) = \begin{cases} 1 & x_{(1)} > \theta \ , \ x_{(n)} < \theta + 1 \\ 0 & \text{o.w} \end{cases}$$

$$= U(x_{(1)} - \theta) \cdot U(\theta + 1 - x_{(n)}) \times 1$$

$$= g_\theta(T^*(\underline{x})) \cdot h(\underline{x}) \qquad T^*(\underline{x}) = (x_{(1)}, x_{(n)})$$

پس T^* یک آماره کافی مینیمال است.

$$\frac{f_\theta(\underline{x})}{f_\theta(\underline{y})} = \frac{U(x_{(1)} - \theta) U(\theta + 1 - x_{(n)})}{U(y_{(1)} - \theta) U(\theta + 1 - y_{(n)})} \Rightarrow (x_{(1)}, x_{(n)}) = (y_{(1)}, y_{(n)})$$

پس T^* یک آماره کافی مینیمال است.

ثابت کنید $A(\underline{X})$ ancillary است.

assume $Y_1 = X_{(1)}$ ، $Y_n = X_{(n)}$

چگالی توام Y_1 و Y_n را به دست آورید.

$$f_{1,n}(y_1, y_n) = n(n-1) f(y_1) f(y_n) [F(y_n) - F(y_1)]^{n-2}$$

$$F(x) = \begin{cases} 0 & x < \theta \\ x - \theta & \theta < x < \theta + 1 \\ 1 & x > \theta + 1 \end{cases}$$

$f_{1,n}(y_1, y_n) = n(n-1)(y_n - y_1)^{n-2}$ $\quad \theta < y_1 < y_n < \theta + 1$ برای

حال: $\begin{cases} A = Y_n - Y_1 \\ M = \frac{Y_n + Y_1}{2} \end{cases} \rightarrow \begin{cases} Y_n = \frac{A + 2M}{2} \\ Y_1 = \frac{2M - A}{2} \end{cases}$ $\quad |J| = 1$

می‌دانیم که بین A و M چگونه است.

$$g(a, m) = \begin{cases} n(n-1)a^{n-2} & 0 < a < 1, \ \theta + \frac{a}{2} < m < \theta + 1 - \frac{a}{2} \\ 0 & \text{سایر نقاط} \end{cases}$$

A حاشیه‌ای $= \int_{\theta + \frac{a}{2}}^{\theta + 1 - \frac{a}{2}} n(n-1)a^{n-2} \, dm = n(n-1)a^{n-2}(1-a)$; $0 < a < 1$

که وابسته به θ است. پس A یک آماره زائد است.

- قضیه: اگر $X_1, \dots, X_n$ نمونه‌ای از خانواده مکانی $f_\theta(x) = f(x - \theta)$ باشد آنگاه $A = X_{(n)} - X_{(1)}$ یک آماره زائد از θ است.

و اگر $X_1, \dots, X_n$ نمونه‌ای از خانواده مقیاسی $f_\theta(x) = \frac{1}{\theta} f(\frac{x}{\theta})$ باشد آنگاه $U(x) = \frac{\sum X_i}{X_n}$ آماره‌ای زائد از θ است.

اثبات (در خانواده مکانی): با فرض $Z_i = X_i - \theta$ ابتدا ثابت کنید چگالی Z_i عبارت است از $f_0(z_i)$ که θ ندارد.

A تابع توزیع: $G(a) = P(A \le a) = P(X_{(n)} - X_{(1)} \le a)$

$$= P(X_{(n)} - \theta - (X_{(1)} - \theta) \le a)$$

$$= P(Z_{(n)} - Z_{(1)} \le a)$$

چون اولی ندارد، θ ندارد پس احتمال A نیز θ ندارد، A آماره زائد است.

خانواده مقیاسی: با فرض $Z_i = \frac{X_i}{\theta}$ نشان دهید چگالی Z_i عبارت است از $f_1(z_i)$ و θ ندارد.

اگر فرض $Y_i = \frac{X_i}{X_n}$; $(i = 1, 2, \dots, n-1)$

Y_1 را آماره‌ای غیر کامل بسازید:

$F_{Y_1}(y_1) = P(Y_1 \le y_1) = P(\frac{X_1}{X_n} \le y_1)$ تابع توزیع Y_1:

$= P(\frac{\theta Z_1}{\theta Z_n} \le y_1) = P(\frac{Z_1}{Z_n} \le y_1)$

که مستقل از θ است. پس Y_1 آماره‌ای غیر کامل است.

$\underline{Y} = (Y_1, Y_2, \dots, Y_{n-1})$ آماره‌ای غیر کامل برداری از این مانند

$U(\underline{x}) = \sum_{i=1}^{n-1} Y_i + 1 = \frac{\sum_{i=1}^{n} x_i}{x_n}$

که آماره‌ای غیر کامل است.

مثال: فرض کنید $X_1, X_2, \dots, X_n$ نمونه‌ای تصادفی از خانواده توزیع‌های $U(\theta, \theta+1)$ باشد. نشان دهید $A(\underline{x}) = X_{(n)} - X_{(1)}$ که آماره غیر کامل است.

حل: $X_1, X_2 \sim N(0, \theta^2)$

$V = \frac{X_1}{X_2} \sim C(0,1)$ توزیع کوشی

که مستقل از θ است. پس V آماره‌ای غیر کامل است.

$U = \frac{X_1 + X_2}{X_2}$ که آماره غیر کامل است. $U = \frac{X_1}{X_2} + 1$ پس $\frac{X_1}{X_2}$ نیز آماره غیر کامل است.

آماره‌ی کامل:

گوییم آماره‌ی T برای خانواده توزیع‌های $\{f_\theta : \theta \in \Theta\}$ کامل است اگر برای هر تابع g که $E_\theta(g(T)) = 0$ برای هر $\theta \in \Theta$ نتیجه شود $P_\theta(g(T) = 0) = 1$ برای هر $\theta \in \Theta$.

به عبارت دیگر، خانواده توزیع‌های احتمال $\{P_\theta : \theta \in \Theta\}$ برای متغیر تصادفی T کامل است اگر تنها تابع اندازه‌پذیر g که امید ریاضی صفر دارد، تابع صفر باشد.

$E_\theta(g(X)) = 0 \quad \forall \theta \in \Theta \Rightarrow P_\theta(g(X) = 0) = 1 \quad \forall \theta \in \Theta$

تعریف: گوییم آماره $T(X)$ کامل است اگر تابع اندازه‌پذیر g از T کامل باشد.

T کامل است if $\forall g(T)$: $E_\theta[g(T)]=0 \ \forall \theta\in\Theta \Rightarrow P_\theta(g(T)=0)=1$ $\forall \theta\in\Theta$

حکم قضیه را به زبان ساده‌تر بیان کنیم: آماره T کامل است اگر تنها تابعی از آماره T که امید صفر دارد، تابع صفر باشد. به بیان دیگر تنها تابعی از آماره T که با صفر برابر است، تابع صفر است.

مثال صفحه ۱۱۷ کتاب گفته شده، پس برای تابع g نیز گفته می‌شود.

قضیه: اگر آماره بسنده‌ی کامل باشد آنگاه بسنده مینیمال است.

اثبات

فرض کنید U هر آماره بسنده مینیمال، T آماره‌ی بسنده کامل باشد. به این صورت T بسنده است و U نیز بسنده مینیمال. بنابراین T نتیجه می‌گیریم تابعی مانند h وجود دارد به طوری که

① $U=h(T)$ حال، فرض $\eta(U)=E_\theta(T|U)$

$$E_\theta(\eta(U))=E_\theta(T) \Rightarrow E_\theta[\eta(U)-T]=0 \Rightarrow$$

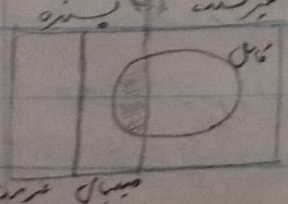
$$E(\eta(h(T))-T)=0$$

و به هم کامل بودن T نتیجه می‌گیریم که

$$\eta(h(T))-T=0 \Rightarrow \eta(h(T))=T$$

$$\eta(U)=T \quad ②$$

از ① و ② نتیجه می‌شود که U و T توابعی یک‌به‌یک از هم هستند و لذا با توجه به فرض مینیمال بودن U، T نیز مینیمال است. (اگر فقط بسنده بود نمی‌توانستیم این را بگوییم)



تذکر: (برعکس آماره کامل می‌توان چند تابع $h(T)$ بسنده‌ی کامل یافت، $h(T)$ را یک‌به‌یک تابع کرانداز نااریب، تابعی گرفت و آن را به این صورت امید صفر کامل را کامل کرانه‌ر گذاشت. کامل کرانداز یکتا بودن خواص مطلوب آن را نتیجه می‌دهد)

این قضیه برای θ تعریف شده، نیاز به تعریف آماره‌ی کامل برای θ است (و لم کفایت به این قضیه کمک می‌کند.)

قضیه باسو: فرض کنید X دارای یکی از اعضای خانواده‌ی $\{f_\theta(x),\theta\in\Theta\}$ است. اگر آماره T بسنده کامل برای θ و R هر آماره‌ی فرعی باشد آنگاه T و R مستقل از همند

$X_1, \dots, X_n \sim U(0, \theta)$; $\theta > 0$

$X_{(1)}, \dots, X_{(n)}$ مشاهدات مرتب شده $\left(\frac{X_1}{X_{n+1}}, \dots, \frac{X_n}{X_{n+1}}\right)$

21 — $\frac{X_{n+1}}{X_{n+1}} = 1$

$X_1, \dots, X_n \sim N(0, \sigma^2)$ نمونه تصادفی از

اثبات: برای اثبات کامل بودن آماره‌ای مثل $T = t$ که R تابعی از T است کافی است نشان دهیم R از T مستقل است. لذا فرض کنید

$$P(R \le r) = F(r) \quad (1)$$

تابع مشخصه زیر را در نظر می‌گیریم:

$$I(x) = \begin{cases} 1 & : R \le r \\ 0 & : \text{در غیر این صورت} \end{cases}$$

$$E_\theta(I(x)) = P(R \le r) \quad \forall \theta \in \Theta$$

$$\eta(t) = E_\theta(I(x) \mid T = t) = P(R \le r \mid T = t) \quad (2)$$

با توجه به تعریف (1) و (2) داریم:

$$E_\theta(\eta(T)) = E_\theta(I(x)) = P(R \le r) \quad \forall \theta \in \Theta$$

$$\Rightarrow E_\theta(\eta(T) - P(R \le r)) = 0 \quad \forall \theta \in \Theta$$

چون T کامل است لذا

$$\eta(t) - P(R \le r) = 0 \quad \text{با احتمال یک}$$

$$\eta(t) = P(R \le r)$$

$$\Rightarrow P_\theta(R \le r \mid T = t) = P_\theta(R \le r)$$

از هر $\theta \in \Theta$، r، t و R، T مستقل از هم هستند.

* قضیه: فرض کنید $\{f_\theta(x) = \exp\{\sum_{i=1}^{k} Q_i(\theta) T_i(x) + D(\theta) + S(x)\} ; x \in A\}$ خانواده‌ای از توابع چگالی باشد. اگر $T = (T_1(\underline{X}), \dots, T_k(\underline{X}))$ یک آماره کافی کامل برای θ باشد آنگاه T یک آماره کامل و کافی مینیمال است.

Lehmann

$X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2) \Rightarrow (\bar{X}, S^2)$ یک آماره کامل و کافی مینیمال است.

$X_1, X_2, \dots, X_n \sim N(\theta, \theta^2) \Rightarrow (\bar{X}, S^2)$ یک آماره کامل و کافی مینیمال نیست

مثال: $X_1, \dots, X_n \sim b(\theta, \theta+1) \Rightarrow T = X_{(n)}$ یک آماره کامل و کافی مینیمال است.

حکم عادی زیاد: عبارت بالا را اثبات نمایید.

راه‌حل دیگر: بیشترین مشاهده را ماکسیمم می‌گیریم $T=X_{(n)}$ آماره‌ی کافی است

$$G(t)=P(X_1\le t)^n=(t/\theta)^n \qquad 0<t<\theta$$

$$g_\theta(t)=nt^{n-1}/\theta^n \qquad 0<t<\theta$$

برای تکمیل $h(t)$ را بررسی می‌کنیم.

$$h^+(t)=\begin{cases} h(t) & h(t)>0 \\ 0 & h(t)\le 0 \end{cases},\quad h^-(t)=\begin{cases} -h(t) & h(t)<0 \\ 0 & h(t)\ge 0 \end{cases}$$

$h(t)=h^+(t)-h^-(t)$ ، $E_\theta(h(T))=0$ ، $\forall\theta>0$

$$\Rightarrow \int_0^\theta h(t)g_\theta(t)dt=\int_0^\theta (h^+(t)-h^-(t))\frac{nt^{n-1}}{\theta^n}dt=0 \quad \forall\theta>0$$

پس اگر A کوچک محدب بزرگ دلخواه باشد، از فرض نتیجه می‌شود که

$$\int_0^\theta h^+(t)t^{n-1}dt=\int_0^\theta h^-(t)t^{n-1}dt \qquad \forall\theta\in A$$

$$\Rightarrow h^+(t)t^{n-1}=h^-(t)t^{n-1} \Rightarrow h^+(t)=h^-(t)$$

بنابراین

$$\Rightarrow h(t)=h^+(t)-h^-(t)=0$$

پس T کامل است، یعنی هیچ تابع غیرثابتی از آن نیز امید صفر ندارد.

مثال: $X_1,X_2,\dots,X_n$ نمونه‌ای از $U(0,\theta)$ است؟ $X_{(n)}$ کامل است.

~~مثال: نمونه‌ای از توزیع نرمال~~

گروه‌تر هستند از ~~توزیع نرمال~~

راه‌حل دیگر:

$$f_\theta(x)=\begin{cases} e^{-(x-\theta)} & x>\theta \\ 0 & x\le\theta \end{cases}$$

متغیر $Y=e^{-X}$ دارای توزیع $U(0,e^{-\theta})$ است

یعنی X_1 و X_2 هر دو با یک پارامتر توزیع یکسان دارند، $Y_{(n)}$ نیز کامل است (چون $Y_{(n)}$ از $X_{(n)}$ تابعی یک‌به‌یک است) و $X_{(n)}$ نیز کامل است.

تذکر: کامل بودن یک آماره دو مفهوم کاملاً متمایز دارد. ممکن است آماره‌ای هم کامل باشد هم نباشد و نه کامل باشد و نه نباشد. کامل بودن یک آماره به توزیع نمونه‌ای بستگی دارد.

مثال: X_1,X_2 نمونه‌ای از $P(\theta)$، $T=X_1$ کامل است یا خیر؟

$T=X_1+2X_2$ کامل است یا خیر؟

$X\sim N(0,\theta)$ کامل است یا خیر؟

$X\sim N(0,\theta^2)$ کامل است یا خیر؟

$X_1, \dots, X_n \sim N(0, \sigma^2)$

$T = (X_{(1)}, X_{(n)})$ کفایت است

اثبات کامل لنت. (خودتان)

۶. فصل نهم: روش‌های برآوردیابی

فرض کنید متغیر تصادفی $\underline{X}$ روی فضای احتمال $(\Omega, \mathcal{A}, P)$ تعریف شده و بردار پارامتر $\underline{\theta}$ است که در فضای پارامتر $\Theta \subseteq \mathbb{R}^k$ قرار دارد. $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ هر یک از پارامترهای $\underline{\theta}$ را تابعی روی Θ در نظر بگیرید. می‌خواهیم $\phi(\underline{\theta})$ را با استفاده از یک نمونه $\underline{X}$ تخمین بزنیم. بهترین منظور تعریف زیر را داریم.

تعریف: اگر $\underline{X} = (X_1, \dots, X_n) \sim F_{\underline{\theta}}(\underline{x})$ (یعنی $X_1, \dots, X_n$ مستقل و هم‌توزیع) باشد، آماره $\delta(\underline{X})$ را برآوردگر (تخمین‌گر) برای $\phi(\underline{\theta})$ گوییم هرگاه $\underline{X}$ در $\mathbb{R}^n$ باشد و

$\phi(\underline{\theta}) : \Theta \longrightarrow \mathbb{R}^r$ (معمولاً $r = 1$)

برای مسئله برآورد نقطه‌ای دو روش کلی زیر معرفی می‌شود.

۱- روش‌هایی با ضمان برآوردگرهای منطقی که امید؟

۲- روش‌هایی که ضمن حفظ برآوردگر خوب از آن، برآوردگر خوب از آن را نیز بدست دهد.

در این فصل، با روش‌های ۱ و بعضی روش‌های ۲ آشنا می‌شویم.

۱- روش گشتاوری: قبل از ارائه این روش برآوردگر، نکته برآوردگری گشتاوری لازم است یادآوری جانشانی‌ها (Substitution Principle) که در آن استفاده می‌شود را بیان کنیم.

اصل جانشانی: Substitution Principle

فرض کنید h تابع مناسبی است که $\phi(\theta) = h(F_\theta)$

گیریم $h(F_n^*)$ برآوردگر منطقی برای $\phi(\theta)$ است که

دیدیم F_n^* برآورد نااریب و سازگار برای F است.

تابع توزیع: F_n^* را به صورت زیر تعریف می‌کنیم.

$$F_n^*(x) = \frac{\text{تعداد } x_i \text{ های کوچکتر یا مساوی } x}{n} = \frac{(\#\, x_i) \le x}{n} = \frac{1}{n}\sum_{i=1}^{n} u(x - x_i)$$

که u تابع پله‌ای واحد است.

$$\Rightarrow Y = nF_n^* \sim B(n, F(x))$$

و هم‌چنین با توجه به مستقل بودن $x_1, x_2, \ldots, x_n$ از یکدیگر و یکسان توزیع بودن آن‌ها، $x_1, x_2, \ldots, x_n$ یک نمونه تصادفی از F است.

$$F_n^* \text{ گام‌ها} = F_n^*(x) - F_n^*(x^-)$$
$$= F_n^*(x) - F_n^*(x-1) = \frac{(\#\, x_i \le x) - (\#\, x_i \le x-1)}{n} = \frac{1}{n}$$

حال برای k ام گشتاورها داریم:

$$\varphi(\theta) = h(F_\theta) = E_\theta x^k = \int x^k \, dF_\theta = \mu_k$$

که $\theta = (\theta_1, \ldots, \theta_r)$ و $\varphi(\theta)$ تابعی از θ است.

$$\widetilde{\varphi}(\theta) = h(F_n^*) = \int x^k \, dF_n^* = \frac{1}{n}\sum_{i=1}^{n} x_i^k = \overline{x^k}$$

توجه: فرض کنید

$$\varphi(\theta) = h(F_\theta) = h(E_\theta x, E_\theta x^2, \ldots, E_\theta x^k)$$

و به طور مشابه برآوردگر

$$\widetilde{\varphi}(\theta) = h(\bar{x}, \overline{x^2}, \ldots, \overline{x^k})$$

در صورتی که $h : \mathbb{R}^k \to \mathbb{R}^k$ تابعی پیوسته باشد، برآوردگر $\varphi(\theta)$ نیز سازگار است.

$$E(XY) \approx \frac{1}{n}\sum_{i=1}^{n} x_i y_i$$

$$\text{Cov}(X,Y) \approx \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}$$

همچنین برآوردگر سازگار برای پارامتر

$\theta = (\mu, \sigma^2)$ ... $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ ...

$E_\theta X = \mu$

$Var(X) = E_\theta X^2 - E_\theta^2 X = \sigma^2$

$\hat{\mu} = \bar{X}$ ... μ ...

$\hat{\sigma}^2 = \bar{X^2} - \bar{X}^2$ ... σ^2 ...

$\hat{\sigma} = \sqrt{\bar{X^2} - \bar{X}^2}$ ... σ ...

$\underline{\theta} = (a, b)$ ... $X_1, X_2, \dots, X_n$ ... $U(a, b)$ ...

$-\infty < a < b < \infty$ ... θ ...

$$f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ \text{...} \end{cases}$$

$$EX = \frac{a+b}{2}, \quad Var(X) = \frac{(b-a)^2}{12} \Longrightarrow$$

$$Var(X) = \frac{(b-a)^2}{12} = \frac{1}{4} \cdot \frac{(b-a)^2}{3}$$

(1) $EX = \frac{a+b}{2}$

(2) $\sqrt{3 Var X} = \frac{b-a}{2}$

(1) + (2) : b

$$\begin{cases} a = EX - \sqrt{3(EX^2 - E^2X)} \\ b = EX + \sqrt{3(EX^2 - E^2X)} \end{cases}$$

$$\hat{a} = \bar{X} - \sqrt{3(\bar{X^2} - \bar{X}^2)} \longrightarrow \begin{cases} \hat{a} = \bar{X} - \sqrt{3} S_b \\ \hat{b} = \bar{X} + \sqrt{3} S_b \end{cases}$$

$$\hat{b} = \bar{X} + \sqrt{3(\bar{X^2} - \bar{X}^2)}$$

$X_1, \dots, X_n \overset{iid}{\sim} B(\theta_1, \theta_2)$ $\theta_1 \in \mathbb{N}$, $0 < \theta_2 < 1$ ...

$EX = \theta_1 \theta_2$ (1)

$Var(X) = \theta_1 \theta_2 (1 - \theta_2) \Rightarrow EX^2 = Var(X) + E^2X$

$= \theta_1 \theta_2 - \theta_1 \theta_2^2 + \theta_1^2 \theta_2^2$ (2)

۲۶

$$\theta_1 = \frac{E(X)}{\theta_2}$$

$$E(X^2) = \theta_2 - \frac{E^2(X)}{\theta_2} + E^2(X)$$

$$\theta_1 = \frac{E^2(X)}{E(X) - (E(X^2) - E^2(X))} \qquad \theta_2 = \frac{E(X) - (E(X^2) - E^2(X))}{E(X)}$$

$$\Rightarrow \begin{cases} \tilde{\theta}_1 = \dfrac{\bar{X}^2}{\bar{X} - S_b^2} \\ \tilde{\theta}_2 = \dfrac{\bar{X} - S_b^2}{\bar{X}} \end{cases}$$

اگر $\bar{X} > S_b^2$ آنگاه برآوردهای گشتاوری مثبت هستند و اگر $\bar{X} < S_b^2$ آنگاه برآوردهای گشتاوری مقدار نامناسب (نامعقول) دارند.

مثال: $X_1, X_2, \ldots, X_n \sim N(\theta, \theta)$

$$E(X) = \theta$$
$$Var(X) = \theta = E(X^2) - E^2(X)$$

$$\Rightarrow \begin{cases} \tilde{\theta} = \bar{X} \\ \tilde{\theta} = S_b^2 \end{cases}$$

برآوردهای گشتاوری برای پارامتر کمیت نامعلوم قرار دارد که تابعی از پارامتر است.

$$\varphi(\theta) = h_0(E_\theta X)$$

برآوردهای ماکزیمم درستنمایی همین کار را انجام می‌دهد.

$$\varphi(\theta) = h_0(E_\theta X, E_\theta X^2)$$

نتیجه: لذا مطابق این قرار داد $\tilde{\theta} = S_b^2$ مقدار نامناسب است.

۱۰ـ۵ـ برآوردگر درستنمایی ماکزیمم:

اصل درستنمایی ماکزیمم این است که برآوردی برای θ انتخاب کنیم که احتمال وقوع (یا چگالی) نمونه را ماکزیمم کند. به عبارت دیگر برآوردی انتخاب کنیم که احتمال وقوع نمونه مشاهده شده را ماکزیمم کند.

ابتدا حالتی را در نظر می‌گیریم که متغیر تصادفی X گسسته باشد.

فرض کنید $(x_1, \dots, x_n)$ مقادیر مشاهده شده نمونه تصادفی $(X_1, \dots, X_n)$ از توزیعی با تابع احتمال $f_\theta(x)$ باشد که $\theta \in \Theta$.

تابع درستنمایی عبارت است از:

$$L(\theta; x_1, \dots, x_n) = f_\theta(x_1, \dots, x_n)$$

که مقدار عددی آن به ازای هر θ برابر احتمال وقوع نمونه مشاهده شده است.

اگر $X_1, \dots, X_n$ مستقل و همتوزیع باشند، تابع درستنمایی عبارت است از:

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

برآوردگر درستنمایی ماکزیمم: $\hat{\theta}(x)$ یا MLE برآوردگری است که

$$L(\hat{\theta}; x_1, \dots, x_n) = \sup_{\theta \in \Theta} L(\theta; x_1, \dots, x_n)$$

حال فرض کنید X پیوسته است و $f_\theta(x)$ تابع چگالی احتمال آن است. تابع درستنمایی ماکزیمم مانند حالت گسسته تعریف می‌شود با این تفاوت که $f_\theta(x)$ تابع چگالی احتمال X است. اگر θ یک پارامتر k بعدی باشد، یعنی $\theta \in \mathbb{R}^k$، برای یافتن برآوردگر درستنمایی ماکزیمم θ معمولاً از دستگاه معادلات زیر استفاده می‌شود

$$\frac{\partial \log L(\theta; x)}{\partial \theta_i} = 0 \quad i = 1, \dots, k$$

و عنوان دستگاه بازار کرم مطلق ما $X_1, \dots, X_n \sim f(x;\theta)$ از θ عنوان MLE.

مثال

① $X_1, \dots, X_n \sim N(\theta, 1)$ $\Rightarrow \hat{\theta} = \bar{X}$

② $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ μ معلوم، σ^2 مجهول

$$\hat{\sigma}^2 = \frac{1}{n}\sum (X_i - \mu)^2$$

③ $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ μ, σ^2 مجهول

$$\hat{\theta} = (\bar{X}, S_b^2)$$

زیرا تابع $L(\mu, \sigma^2)$ در $\mu = \bar{X}$ ماکزیمم است.

زیرا تابع $L(\bar{X}, \sigma^2)$ در $\sigma^2 = S_b^2$ ماکزیمم است.

$\Rightarrow L(\bar{X}, S_b^2) \geq L(\mu, \sigma^2)$ در μ, σ^2

④ $X_1, \dots, X_n \sim U(0, \theta)$ $\Rightarrow \hat{\theta} = X_{(n)}$ $\theta \in (0, 1)$

برای θ آن $\theta = \frac{1}{2}$ و $\theta = \frac{1}{3}$ کدام مقدار بیشتر است؟

⑤ $X \sim B(k, p)$ $\theta = (k, p)$ ، $\Theta = \{(k,p) : k = 2, p = \frac{1}{2}, \frac{1}{3}\}$

برآورد ماکزیمم درستنمایی θ را بیابید و جدول زیر را کامل کنید.

x \ θ	$(2, \frac{1}{2})$	$(2, \frac{1}{3})$	$(3, \frac{1}{2})$	$(3, \frac{1}{3})$	$\hat{\theta}_{MLE}$
0	$\frac{1}{4}$	$\frac{4}{9}$	$\frac{1}{8}$	$\frac{8}{27}$	$(3, \frac{1}{3})$
1	$\frac{1}{2}$	$\frac{4}{9}$	$\frac{3}{8}$	$\frac{12}{27}$	$\frac{1}{2}$
2	$\frac{1}{4}$	$\frac{1}{9}$	$\frac{3}{8}$	$\frac{6}{27}$	$(3, \frac{1}{2})$
3	0	0	$\frac{1}{8}$	$\frac{1}{27}$	$\frac{1}{2}$

$$\hat{\theta}(x) = \begin{cases} (3, \frac{1}{3}) & x = 0 \\ (2, \frac{1}{2}) & x = 1 \\ (3, \frac{1}{2}) & x = 2, 3 \end{cases}$$

۵.

الف) $X_1, X_2, \dots, X_n \overset{i.i.d}{\sim} DU\{1, 2, \dots, \theta\}$ ...

$$L(\theta, \underline{x}) = f_\theta(\underline{x}) = \begin{cases} (\frac{1}{\theta})^n & 1 \le x_{(1)} \le x_{(n)} \le \theta \\ 0 & \text{در غیر این صورت} \end{cases}$$

در مقابل پارامترهای پیوسته، در اینجا $\hat{\theta} < x_{(n)}$ نمی‌تواند برآورد باشد زیرا $L(\hat{\theta}, \underline{x}) = 0$

پس برآورد θ به روش MLE باید $\hat{\theta} \ge x_{(n)}$ باشد. در نتیجه $L(\hat{\theta}, \underline{x}) = (\frac{1}{\hat{\theta}})^n < L(x_{(n)}, \underline{x}) = (\frac{1}{x_{(n)}})^n$

پس $\hat{\theta} = x_{(n)}$ برآورد ماکسیمم درستنمایی θ است.

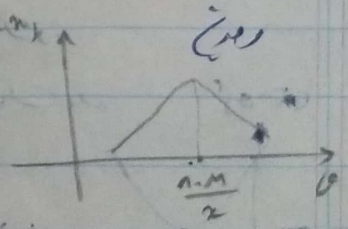
ب) $X \sim HG(\theta, M, n)$ $\hat{\theta} = ?$

$$f_\theta(x) = L(\theta, x) = \frac{\binom{M}{x}\binom{\theta - M}{n - x}}{\binom{\theta}{n}} \quad ; x \in \mathbb{N}, \max(0, n - \theta + M) \le x \le \min(n, M)$$

$$R(\theta) = \frac{f_\theta(x)}{f_{\theta-1}(x)} = \frac{\theta - n}{\theta} \cdot \frac{\theta - M}{\theta - M - n + x}$$

اگر $R(\theta) < 1$ باشد این نشان می‌دهد که $f_\theta(x)$ نسبت به θ نزولی است و اگر $R(\theta) > 1$ باشد نشان می‌دهد که $f_\theta(x)$ نسبت به θ صعودی است.

$$R(\theta) < 1 \iff \theta > n \cdot \frac{M}{x}$$
$$R(\theta) > 1 \iff \theta < n \cdot \frac{M}{x}$$



پس برآورد بیشینه درستنمایی θ عبارت است از:

$$\hat{\theta} \simeq n \cdot \frac{M}{x}$$
$$\hat{\theta} = [n \cdot \frac{M}{x}]$$

نقص برآورد $\theta \ge \max(n, M)$

$\frac{M}{x} \ge 1 \to \theta \ge n \cdot \frac{M}{x} \ge n \Rightarrow \theta \ge \max(n, M)$

$\frac{n}{x} \ge 1 \to \hat{\theta} = M \cdot \frac{n}{x} \ge M$

مثال: اگر $X_1, X_2, \dots, X_n$ نمونه‌ای از $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$ باشد، نشان دهید که تعداد برآوردگرهای بیشینه درستنمایی ماکزیمم برای θ وجود دارد. یکی از آنها $\hat{\theta} = \frac{X_{(1)} + X_{(n)}}{2}$ است.

$$f(\theta, \underline{x}) = \begin{cases} 1 & \theta - \frac{1}{2} < x_{(1)} \le x_{(n)} < \theta + \frac{1}{2} \\ 0 & \text{در غیر این صورت} \end{cases}$$

$$= \begin{cases} 1 & x_{(n)} - \frac{1}{2} < \theta < x_{(1)} + \frac{1}{2} \\ 0 & \text{در غیر این صورت} \end{cases}$$

هر برآوردگری که تساوی زیر را برآورده کند، یک برآوردگر بیشینه درستنمایی ماکزیمم است.

$$x_{(n)} - \frac{1}{2} \quad , \quad x_{(1)} + \frac{1}{2}$$

$$T_1(\underline{x}) = x_{(1)} + \frac{1}{2}$$

$$T_2(\underline{x}) = \frac{x_{(1)} + x_{(n)}}{2}$$

$$T_3(\underline{x}) = x_{(n)} - \frac{1}{2}$$

$$T_\alpha(\underline{x}) = \alpha\left(x_{(n)} - \frac{1}{2}\right) + (1-\alpha)\left(x_{(1)} + \frac{1}{2}\right)$$

که α عددی حقیقی است. برآوردگر $T_\alpha(\underline{x})$ برآوردگر بیشینه درستنمایی ماکزیمم برای θ است. این مثال نشان می‌دهد که MLE می‌تواند یکتا نباشد.

$$\alpha = \frac{1}{2} \longrightarrow T_{\frac{1}{2}}(\underline{x}) = \frac{x_{(1)} + x_{(n)}}{2}$$

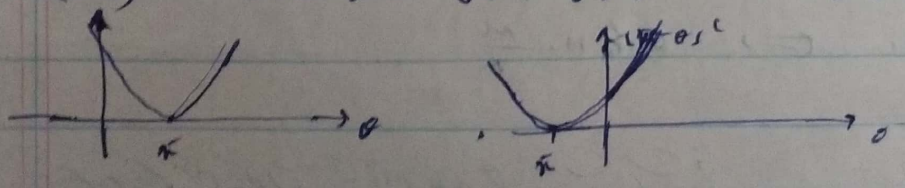
مثال: اگر $X_1, X_2, \dots, X_n$ نمونه‌ای از $N(\theta, 1)$ باشد که $\theta \in \Theta = \mathbb{R}^+$، مطلوب است MLE برای θ.

$$f(\theta, \underline{x}) = (\sqrt{2\pi})^{-n} e^{-\frac{1}{2}\sum(x_i - \theta)^2}$$

$$= (2\pi)^{-n/2} \cdot e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot e^{-\frac{n}{2}(\bar{x} - \theta)^2}$$

ملاحظه می‌کنید که $f(\theta, \underline{x})$ وقتی ماکزیمم می‌شود که $(\bar{x} - \theta)^2$ مینیمم شود.

حالتی که $\bar{x} \in \mathbb{R}^+$ باشد، یعنی $\bar{x} > 0$. در این حالت $\hat{\theta} = \bar{x}$



$$\hat{\theta}(\underline{x}) = \begin{cases} \bar{x} & \bar{x} \ge 0 \\ 0 & \bar{x} < 0 \end{cases}$$

$X_1, X_2, \dots, X_n$ نمونه‌ای از $B(\theta, \frac{1}{2})$; $\theta \in \mathbb{N}$ فرض کنید

$$f_\theta(x) = \binom{\theta}{x} \left(\frac{1}{2}\right)^\theta \quad , \quad x = 0, 1, 2, \dots, \theta \quad , \quad \theta \in \mathbb{N}$$

$$L(\theta, \underline{x}) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \binom{\theta}{x_i} \left(\frac{1}{2}\right)^\theta \quad \text{به شرط } x_{(n)} \le \theta$$

برای محاسبه $\hat{\theta}$ داریم:

$$R(\theta) = \frac{L(\theta, \underline{x})}{L(\theta-1, \underline{x})} = \left(\frac{\theta}{\theta-1}\right)^n \prod_{i=1}^{n} \frac{1}{(\theta - x_i)}$$

همچنین

$$R(\theta+1) = \left(\frac{\theta+1}{\theta}\right)^n \prod_{i=1}^{n} \frac{1}{(\theta+1-x_i)}$$

برآوردگر ماکزیمم درستنمایی $\hat{\theta}$ برآوردگر MLE است اگر و فقط اگر $\hat{\theta}$ کوچکترین عدد صحیح نامنفی باشد که در نامساوی‌های زیر صدق کند:

$$R(\hat{\theta}) \ge 1 \quad , \quad R(\hat{\theta}+1) \le 1$$

به عبارت دیگر

$$\left(\frac{\hat{\theta}}{\hat{\theta}-1}\right)^n \ge \prod_{i=1}^{n} \left(\frac{\hat{\theta}}{\hat{\theta}-x_i}\right) \longrightarrow \left(\frac{1}{2}\right)^n \ge \prod_{i=1}^{n} \left(1 - \frac{x_i}{\hat{\theta}}\right)$$

$$\left(\frac{1}{2}\right)^n \ge \prod_{i=1}^{n} \left(1 - \frac{x_i}{\hat{\theta}}\right) \quad (1)$$

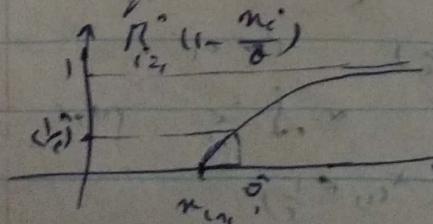
$$\left(\frac{1}{2}\right)^n < \prod_{i=1}^{n} \left(1 - \frac{x_i}{\hat{\theta}+1}\right) \quad (2)$$

برای حل این دستگاه نامساوی‌ها

اگر $\left(\frac{1}{2}\right)^n \ge \prod_{i=1}^{n} \left(1 - \frac{x_i}{\theta}\right)$ برقرار باشد برای θ های بزرگتر از θ نیز برقرار است.

$\hat{\theta} \ge x_{(n)}$

اگر $\hat{\theta}$ کوچکترین عدد صحیح نامنفی باشد که $\hat{\theta} > x_{(n)}$ باشد نتیجه می‌گیریم $\hat{\theta}$ کوچکترین عدد صحیح نامنفی است که $\hat{\theta} \ge x_{(n)}$ باشد.



بنابراین $\hat{\theta}$ برابر است با کوچکترین عدد صحیح نامنفی که در بازه $[x_{(n)}, \infty)$ قرار دارد.

$T=(X_{(1)}, X_{(n)})$ ... MLE ... $\alpha=\frac{1}{2}$ ... $T_{\frac{1}{2}}=\frac{X_{(1)}+X_{(n)}}{2}$ ... MLE ... Invariance property

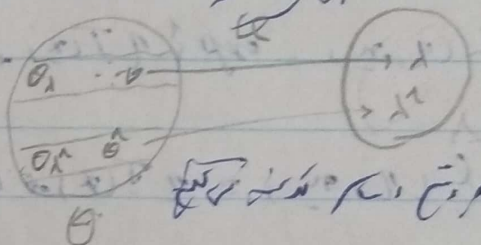
• قضیه (خاصیت پایای MLE): اگر $\hat{\theta}$ برآوردگر بیشینه درستنمایی نارایب θ باشد آنگاه برآوردگر بیشینه درستنمایی نارایب $g(\theta)$ عبارت است از:

اثبات: اگر g را بر Θ نگاشت Λ بنامیم:

$$g:\Theta\longrightarrow\Lambda \quad ; \quad \lambda=g(\theta)$$

به ازای هر $\lambda\in\Lambda$ مجموعه Θ_λ و تعریف زیر را تشکیل می‌دهیم:

$$\Theta_\lambda=\{\theta\in\Theta\ ;\ g(\theta)=\lambda\}$$



واضح است که $\Theta=\bigcup_{\lambda\in\Lambda}\Theta_\lambda$

$$L^*(\lambda)=\sup_{\theta\in\Theta_\lambda} L(\theta,x) \quad \text{تابع درستنمایی القایی}$$

حال اگر $\hat{\theta}$ برآوردگر بیشینه درستنمایی نارایب θ باشد آنگاه

$$L(\hat{\theta})=\sup_{\theta\in\Theta} L(\theta)$$

$\hat{\theta}$ متعلق به یک و فقط یک $\Theta_{\hat{\lambda}}$ خواهد بود مثلاً $\Theta_{\hat{\lambda}}$ و چون $\hat{\theta}\in\Theta$ پس $\hat{\lambda}=g(\hat{\theta})$

حال با بررسی رابطه L^* داریم:

$$L^*(\hat{\lambda})=\sup_{\theta\in\Theta_{\hat{\lambda}}} L(\theta)\geq L(\hat{\theta}) \qquad (1)$$

همچنین داریم:

$$L^*(\hat{\lambda})\leq\sup_{\lambda\in\Lambda} L^*(\lambda)=\sup_{\lambda\in\Lambda}\left(\sup_{\theta\in\Theta_\lambda} L(\theta)\right)=\sup_{\theta\in\Theta} L(\theta)=L(\hat{\theta}) \qquad (2)$$

از (1) و (2) نتیجه می‌گیریم

$$L^*(\hat{\lambda})=L(\hat{\theta})=\sup_{\lambda\in\Lambda} L^*(\lambda)$$

پس $\hat{\lambda}=g(\hat{\theta})$ برآوردگر بیشینه درستنمایی نارایب $g(\theta)$ می‌باشد.

$X_1 \sim X_n$ نمونه $B(1,\theta)$ ، $g(\theta) = V_{ar}(X) = \theta(1-\theta)$

میزان θ برابر

$\hat{g}(\theta) = g(\hat{\theta}) = \bar{X}(1-\bar{X})$

• قضیه: فرض کنید $f_\theta(\underline{x})$ به صورت

$f_\theta(\underline{x}) = \exp\{Q(\theta)T(\underline{x}) + D(\theta) + S(\underline{x})\}$

$\hat{\theta}$ آماره‌ای نااریب برای θ باشد به طوری که $E_\theta(T(\underline{x})) = T(\underline{x})$.

اگر H برد Q یعنی $Q: \Theta \to H$ و $\hat{\theta} \in H$ و H محدب باشد آنگاه $\hat{\theta}$ تنها برآوردگر نااریب یکتا برای $Q(\theta)$ است.

$h(\eta, \underline{x}) = \exp\{\eta T(\underline{x}) + D_0(\eta) + S(\underline{x})\}$

$l(\eta, \underline{x}) = \eta T(\underline{x}) + D_0(\eta) + S(\underline{x})$

$\frac{\partial l(\eta, \underline{x})}{\partial \eta} = T(\underline{x}) + D_0'(\eta) \Rightarrow T(\underline{x}) = -D_0'(\eta)$ ①

$\frac{\partial^2 l(\eta, \underline{x})}{\partial \eta^2} = D_0''(\eta) \leq 0$ ②

T تابع مولد گشتاور $M_{T(\underline{x})}(s) = E(e^{sT}) = \int e^{st} \exp\{\eta \frac{t}{T(\underline{x})} + D_0(\eta) + S(\underline{x})\} d\underline{x}$

$= \exp\{D_0(\eta) - D_0(\eta+s)\} \int \exp\{(\eta+s)t + D_0(\eta+s) + S(\underline{x})\} d\underline{x}$

$= \exp\{D_0(\eta) - D_0(\eta+s)\}$

$M_T'(s) = -D_0'(\eta+s)\exp\{D_0(\eta) - D_0(\eta+s)\}$

$M_T'(s)\Big|_{s=0} = E(T(\underline{x})) \Rightarrow E(T(\underline{x})) = -D_0'(\eta) \overset{①}{=}$

$E_\theta(T(\underline{x})) = T(\underline{x})$

بنابراین با توجه به شرایط قضیه، آماره $T(\underline{x})$ برآوردگر نااریب یکتا برای

$$M_T''(t) = -D_0''(\eta+\delta)\exp\{D_0(\eta)-D_0(\eta+\delta)\} + [D_0'(\eta+\delta)]^2\exp\{D_0(\eta)-D_0(\eta+\delta)\}$$

$$\Rightarrow E(T^2) = M_T''(0) = -D_0''(\eta) + [D_0'(\eta)]^2$$

$$\Rightarrow \mathrm{Var}(T) = -D_0''(\eta) \Rightarrow D_0''(\eta) \le 0$$

لذا $D_0(\eta)$ یک تابع محدب و نزولی F یعنی H نزولی است.

تذکر: قضیه بالا برای خانواده‌های یک پارامتری نیز کامل نیز تحدب پایین و ابتدا است که در اینجا مساوی با بیشتر از یک پارامتر است.

$$E_\theta(T_i(\underline{X})) = T_i(\underline{x}) \qquad i = 1, 2, \dots, k$$

مثال: μ و σ^2 نامعلوم $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$

$$\theta = (\mu, \sigma^2) \qquad \hat{\theta} = ?$$

(با استفاده از تابع درستنمایی)

$$L(\theta, \underline{x}) = \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 - n\ln\sigma\sqrt{2\pi}\right\}$$

$$= \exp\left\{\frac{\mu}{\sigma^2}\sum x_i - \frac{1}{2\sigma^2}\sum x_i^2 - \frac{n\mu^2}{2\sigma^2} - n\ln\sigma\sqrt{2\pi}\right\}$$

$$T_1(\underline{X}) = \sum_{i=1}^{n} X_i \qquad T_2(X) = \sum_{i=1}^{n} X_i^2 \Rightarrow \begin{cases} E_\theta(T_1(\underline{X})) = \sum_{i=1}^{n} x_i \\ E_\theta(T_2(X)) = \sum_{i=1}^{n} x_i^2 \end{cases}$$

$$\Rightarrow n\hat{\mu} = \sum_{i=1}^{n} x_i \longrightarrow \hat{\mu} = \bar{x}$$

$$E(T_2(\underline{x})) = E\sum x_i^2 = \sum E x_i^2 = \sum(\sigma^2+\mu^2) = n(\sigma^2+\mu^2)$$

$$n(\hat{\sigma}^2+\hat{\mu}^2) = \sum_{i=1}^{n} x_i^2 \longrightarrow n\hat{\sigma}^2 = \sum x_i^2 - n\bar{x}^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

مثال: تخمین MLE از θ و θ^2 که

$(X_1, X_2) \sim MB(n, \theta^2, 2\theta(1-\theta), (1-\theta)^2)$

مدل دوجمله‌ای - واریانس.

حل: ثابت نرمالیزه کنندهٔ توزیع توان مشترک خانواده نمایی پارامتری است

$$f_\theta(x_1) = \frac{n!}{x_1! x_2! (n-x_1-x_2)!} \theta^{x_1} (2\theta(1-\theta))^{x_2} (1-\theta)^{2(n-x_1-x_2)}$$

$$= \exp\Big\{ \underbrace{(2x_1+x_2)}_{T_1(x)} \underbrace{\log\frac{\theta}{1-\theta}}_{Q(\theta)} + \underbrace{2n\log(1-\theta)}_{D(\theta)} + x_2 \log 2 + \underbrace{\log\frac{n!}{x_1! x_2! (n-x_1-x_2)!}}_{S(x)} \Big\}$$

لذا آماره کافی مینیمال

$$E_\theta(2X_1 + X_2) = 2n_1 + n_2$$

$$EX_1 = n\theta^2, \quad EX_2 = n\theta(1-\theta)$$

$$E_\theta(2X_1 + X_2) = 2E_\theta(X_1) + EX_2 = 2n\theta^2 + 2n\theta(1-\theta) = 2n_1 + n_2$$

$$\Rightarrow \boxed{\hat{\theta}_2 \; \frac{2X_1 + X_2}{2n}} \qquad \hat{\theta} \in H$$

تابع پارامتری θ

قابل برآورد خطی است.

۳- روش‌های بیز و مینیماکس. Bayes & Minimax Methods

مقدمه: برآوردگیری آماری دو فضای مهم پارامتر (یا حالات طبیعت) و فضای Decision theory

اعمال (مجموعه تصمیم‌ها یا اعمال ممکن) برآوردگر مختلف طبیعت) می‌باشد.

مثال: فرض کنید می‌خواهیم از خانه خارج شویم. حالات طبیعت θ و اعمال ممکن را

در جدول زیر خلاصه می‌کنیم. روش جدول تر از نیاز ما نیست از همین ۳ به یک احتمال
طبیعت و ۳ استراتژی.

احتمال طبیعت / عمل	θ_1	θ_2	میزان زیان
a_1	۴	۴	۴
a_2	۵	۰	۵
a_3	۲	۵	۵

از طرفی میزان احتمال با زیان کمتر و بیشتر

θ	θ_1	θ_2
$\pi(\theta)$	۰٫۸	۰٫۲

logical

و دیگر اطلاعات ندارد هر تصمیم معقول و منطقی است؛

reasonable معقول

که بازی با این سه‌تا دو تا استراتژی نیز معرفی شده است.

۱- استراتژی کمینه‌ی بیشینه (minimax): که در آن عملی که کمترین عمل ماکزیمم شده که
ماکزیمم زیان در حالات مختلف طبیعت را بین همه اعمال کمتر باشد که در مثال ما
عمل a_2 این گونه است.

prior

۲- استراتژی بیز: در این روش که تقریبا احتمال پیشین از حالات مختلف طبیعت
در نظر گرفته می‌شود و با استفاده از آن، میانگین زیان (امید ریاضی) ناشی از انجام و عمل عاملی
می‌شود و عملی را عنوان بهترین عمل انتخاب می‌کند که این میانگین و پیش از همه اعمال دیگر
کمتر باشد.

$R(\theta, a_1) = E_\theta[L(\theta, a_1)] = \pi(\theta_1).L(\theta_1, a_1) + \pi(\theta_2).L(\theta_2, a_1)$ میانگین زیان بیز، اگر عمل a_1 انجام شود

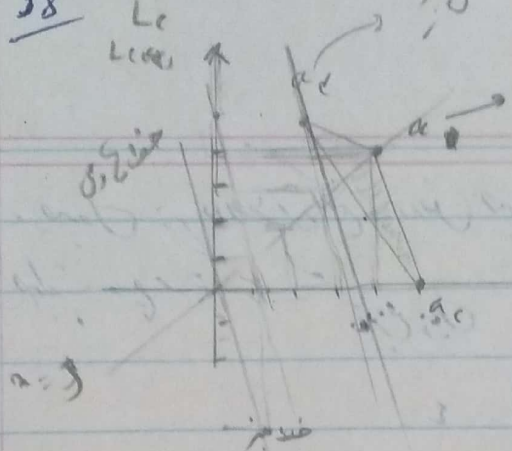
$= ۰٫۸ \times ۴ + ۰٫۲ \times ۴ = ۴$

$R(\theta, a_2) = ۰٫۸ \times ۵ + ۰٫۲ \times ۰ = ۴$ ، $R(\theta, a_3) = ۰٫۸ \times ۲ + ۰٫۲ \times ۵$

$= ۱٫۶ + ۱ = ۲٫۶$

که ملاحظه می‌شود a_3 عمل بیز است. این بیان از راه انتخاب روش کمترین ضرر ملاحظه کرد.

38



اعمال a_1، a_2، a_3 ... اعمال قابل اجرا هستند، مجموعه اعمال ممکن را تشکیل می‌دهند

از بین این اعمال، آن عملی را انتخاب می‌کنیم که بیشترین سود را برای ما داشته باشد

مثال: فرض کنید دو نفر x_1 و x_2 با هم بازی می‌کنند. هر کدام از آنها می‌توانند یکی از دو عمل a_1 یا a_2 را انجام دهند. اگر هر دو عمل یکسان انجام دهند، x_1 برنده است. اگر عمل‌ها متفاوت باشند، x_2 برنده است. برنده یک واحد پول می‌گیرد و بازنده یک واحد پول می‌پردازد. ماتریس بازی را تشکیل دهید.

x_1 \ x_2	a_1	a_2
a_1	1	-1
a_2	-1	1

حال اگر x_1 بداند که x_2 چه عملی را انتخاب می‌کند، بهترین عمل را انتخاب می‌کند.

x_1 \ x_2	a_1	a_2
a_1	1	-1
a_2	-1	1

حال اگر x_2 بداند که x_1 چه عملی را انتخاب می‌کند، بهترین عمل را انتخاب می‌کند.

x_1 \ x_2	a_1	a_2
a_1	1	-1
a_2	-1	1

حال اگر هیچ‌کدام از x_1 و x_2 از انتخاب دیگری خبر نداشته باشند، چه عملی را انتخاب می‌کنند؟

x_1 \ x_2	a_1	a_2
a_1	1	-1
a_2	-1	1

x \ d	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
x_1	a_1	a_2	a_3	a_1	a_1	a_2	a_3	a_4	a_5
x_2	a_1	a_2	a_3	a_2	a_3	a_1	a_2	a_1	a_3

$$R(\theta_1, d_1(x_1)) = E_x[L(\theta_1, d_1(x_1))] =$$

$$= \sum_{j=1}^{2} L(\theta_1, d_1(x_j)) P_{\theta_1}(X = x_j)$$

$$= L(\theta_1, d_1(x_1)) P_{\theta_1}(X = x_1) + L(\theta_1, d_1(x_2)) P_{\theta_1}(X = x_2)$$

$$= L(\theta_1, a_1) P_{\theta_1}(X = x_1) + L(\theta_1, a_1) P_{\theta_1}(X = x_2)$$

$$= 4 \times 0.8 + 4 \times 0.2 = 4$$

θ \ d	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
θ_1	4	5	2	4.2	3.6	3.8	4.4	2.4	2.6
θ_2	4	0	5	4	4.8	3.4	4.6	4.1	5
$B(d)$	4	4	2.6	4.16	3.72	3.56	4.48	2.74	2.18
$M(d)$	4	5	5	4.2	4.8	3.8	4.5	4.1	2.6

$$B(d) = E_\theta[R(\theta, d)] = \sum_{i=1}^{2} R(\theta_i, d) \pi(\theta_i) = R(\theta_1, d) \times 0.8 + R(\theta_2, d) \times 0.2$$

$$M(d) = \max(R(\theta, d))$$

تابع تصمیم:

در نمونه‌گیری $f(x|\theta)$ که $X_1, X_2, \dots, X_n$ مشاهدات $\theta \in \Theta$ (فضای پارامتر)

هر تابع $\delta(x)$ که $x = (x_1, x_2, \dots, x_n)$ را به یک عمل تصمیم $a \in A$ (فضای تصمیم) می‌نگارد را یک تابع تصمیم می‌نامیم.

مثال: δ را یک تابع تصمیم گوییم هرگاه برای هر x، $\delta(x)$ یک تخمین‌گر باشد.

مثلاً $\delta(x) = \bar{x}$ یک تابع تصمیم است.

از انجام هر تصمیم مانند $\delta(x)$ در مقابل هر حالت طبیعت مانند θ مقداری زیان حاصل می‌شود که آن را با $L(\theta, \delta(x))$ نمایش می‌دهیم.

تعریف: L را تابع زیان گوییم هرگاه $L: \Theta \times D \to \mathbb{R}^{\geq 0}$

به‌طوری که $L(\theta, \delta(x))$ معرف میزان زیان حاصل از تصمیم $\delta(x)$ وقتی θ حالت طبیعت باشد.

تعریف: تابع مخاطره (ریسک) را امید

$$R(\theta, \delta) = E_x(L(\theta, \delta(x)))$$

تعریف: (امید ریاضی متغیر x است)

$$R: \Theta \times D \to \mathbb{R}^{\geq 0}$$

در حالت کلی $\Theta = A$ است و تصمیم ممکن است کاهنده باشد لذا تابع زیان باید حاصل خطا را نشان دهد. مثلاً $e = |\delta(x) - \theta|$ اگر θ را تخمین بزنیم.

اگر $A = \{a_1, a_2\}$ که تابع تصمیم باشد از $\mathbb{R}^n$ به C، C را ناحیه بحرانی گوییم هرگاه $x \in C$ عمل a_1 انجام شود و اگر $x \in C^c$ عمل a_2 را انجام دهیم، در این حالت این مسأله یک مسأله آزمون فرض است.

معمولاً توابع تصمیم مورد استفاده و نیز تصمیم بهینه از میان آنها را با استفاده از

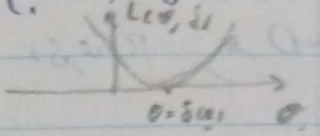
Bayes

Decision function

تابع تصمیم گیری

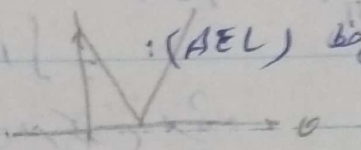
۱- تابع زیان مربع خطا (SEL): square error loss function

$$L_1(\theta, \delta(\underline{x})) = (\theta - \delta)^2$$



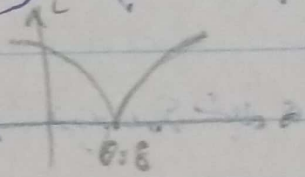
۲- تابع زیان قدرمطلق خطا (AEL): Absolute

$$L_2(\theta, \delta(\underline{x})) = |\theta - \delta|$$



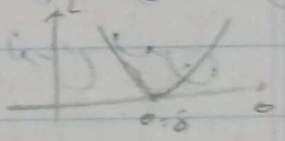
۳- تابع زیان مربع خطای نسبی (RSEL): Relative

$$L_3(\theta, \delta(\underline{x})) = \frac{(\theta - \delta)^2}{|\theta| + 1}$$



۴- تابع زیان نامتقارن خطی: (Varex) نامتقارن

$$L_4(\theta, \delta(\underline{x})) = \begin{cases} (\theta - \delta)^2 & \theta < \delta \\ 10(\theta - \delta)^2 & \theta > \delta \end{cases}$$



۵- تابع زیان خطی نمایی (Linex) نامتقارن

$$L_5(\theta, \delta(\underline{x})) = e^{-c(\theta - \delta)} - c(\theta - \delta) - 1$$

که مقدار c تغییرات تابع زیان نامتقارن مورد نظر را مشخص می‌کند. (c مقداری مثبت است)

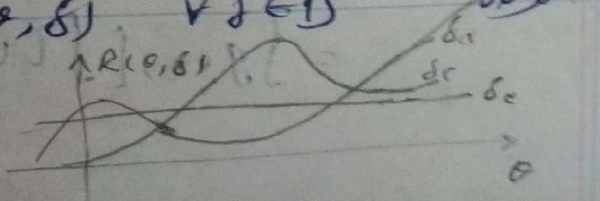
$\theta = \delta \Rightarrow L = 0$ برآورد بدون اریبی

بهترین تابع زیان، تابع زیان مربع خطاست.

ریسک تصمیم عبارت است از $R(\theta, \delta)$ که کمترین δ ی را انتخاب می‌کنیم که کمترین $R(\theta, \delta)$ را به ازای $\theta \in \Theta$ داشته باشد. اما ممکن است که چنین δ ی وجود نداشته باشد و ناچاریم از استراتژی‌های دیگری مثل مینیماکس که مثلاً در آنالیز عددی هم کاربرد دارد استفاده کنیم.

۱- روش مینیماکس (کم‌بیشینه): δ_0 را که تابع تصمیم مینیماکس (کم‌بیشینه) گوییم

$$\sup_{\theta \in \Theta} R(\theta, \delta_0) \leq \sup_{\theta \in \Theta} R(\theta, \delta) \quad \forall \delta \in D$$



که روش بیز، در این روش θ را مانند آن که متغیر تصادفی و تابع گرفته و هیچ آنرا $\pi(\theta)$ نامیده و آن توزیع پیشین (Prior Dist.) گوییم، δ^* را در اینجا برگزیده بیز گوییم اگر $R(\theta,\delta^*)$ کمتر از $R(\theta,\delta)$ باشد برای هر $\delta \in D$ که در آن $R(\theta,\delta)$ ریسک بیز عمل δ را در تابع توزیع پیشین $\pi(\theta)$ نسبت بهبود

$$r(\pi,\delta) = \underset{\pi}{E}(R(\theta,\delta)) = \int R(\theta,\delta)\pi(\theta)\,d\theta$$

توزیع پیشین

توزیع پسین

تابع بیز δ^* را عمل بیز گوییم هرگاه

$$r(\pi,\delta^*) \le r(\pi,\delta) \quad \forall \delta \in D$$

$$r(\pi,\delta^*) = \inf_{\delta} r(\pi,\delta)$$

که در آن $\pi(\theta)$ توزیع پیشین θ، $f(\underline{x}|\theta)$ توزیع نمونه‌ای $\underline{x}$ به شرط θ، توزیع توأم

$f(\underline{x},\theta)$ را توزیع توأم $\underline{x}$ و θ، $g(\underline{x})$ را توزیع کناری $\underline{x}$، $h(\theta|\underline{x})$ را توزیع θ به شرط $\underline{x}$ یا توزیع پسین θ گوییم. Posterior Dist.

یک بیز $r(\pi,\delta) = \int R(\theta,\delta)\pi(\theta)\,d\theta$

$$= \int\int L(\theta,\delta) f(\underline{x}|\theta)\pi(\theta)\,d\underline{x}\,d\theta$$

$$= \int\int L(\theta,\delta) f(\underline{x},\theta)\,d\underline{x}\,d\theta$$

$$= \int\int L(\theta,\delta) h(\theta|\underline{x}) g(\underline{x})\,d\underline{x}\,d\theta$$

• قضیه مهم: اگر تابع زیان $L \in \mathbb{L}$ اکیداً محدب باشد، میانگین توزیع پسین، تابع بیز یکتا است.

اثبات: اگر δ یک تابع تصمیم دلخواه باشد

$$r(\pi,\delta) = \int\int L(\theta,\delta) h(\theta|\underline{x}) g(\underline{x})\,d\underline{x}\,d\theta$$

$$= \int g(\underline{x}) \left[\int L(\theta,\delta) h(\theta|\underline{x})\,d\theta\right] d\underline{x}$$

$K(\delta)$ = عبارت δ

$$K(\delta) = \int L(\theta,\delta)\, h(\theta,\underline{x})\, d\theta = \int (\theta-\delta)^2 h(\theta,\underline{x})\, d\theta$$

برای مینیمم کردن ریسک بیز کافی است $K(\delta)$ را نسبت به δ مینیمم کنیم. از این منظور

$$K'(\delta) = -2\int (\theta-\delta)\, h(\theta,\underline{x})\, d\theta$$

$$K''(\delta) = 2\int h(\theta,\underline{x})\, d\theta = 2 > 0$$

لذا $K(\delta)$ در نقطه $K'(\delta)=0$ مینیمم می‌گردد.

از این معادله می‌توان نتیجه گرفت که

$$K'(\delta)=0 \Rightarrow \int \theta\, h(\theta,\underline{x})\, d\theta = \delta \int h(\theta,\underline{x})\, d\theta = \delta$$

$$\Rightarrow \delta^* = E(\theta|\underline{x})$$

مثال: اگر $x_1, x_2, \ldots, x_n$ نمونه‌ای از توزیع $b(1,\theta)$ باشد و L تابع زیان مربعی باشد، برآوردگر بیز را برای حالتی که θ دارای توزیع پیشین الف) $\pi_1(\theta)=1$ $0<\theta<1$ و ب) $\pi_2(\theta)=\begin{cases}1/2 & \theta=1/2 \\ 1/2 & \theta=0\end{cases}$ بدست آورید.

حل: الف)

$$h(\theta|\underline{x}) = \frac{f(\underline{x}|\theta)\pi(\theta)}{\int f(\underline{x},\theta)\pi(\theta)\, d\theta} \propto f(\underline{x}|\theta)\pi(\theta)$$

$$= \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \qquad 0<\theta<1$$

$$h(\theta|\underline{x}) \propto \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \qquad 0<\theta<1$$

$$\Rightarrow \theta|\underline{x} \sim Beta\left(\sum x_i+1,\ n-\sum x_i+1\right)$$

$$\delta^* = E(\theta|\underline{x}) = \frac{\sum x_i+1}{n+2}$$

$$h(\theta|\underline{x}) = \frac{f(\underline{x}|\theta)\,\pi(\theta)}{\sum_{\theta=0,\frac{1}{2}} f(\underline{x}|\theta)\,\pi(\theta)} = \frac{\frac{1}{2}\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}}{\frac{1}{2}\sum_{\theta=0,\frac{1}{2}} \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}}$$

مثال ب)

- اگر $\sum x_i \neq 0$ باشد، واضح است که نتیجه کسر برابر

$$h(\theta|\underline{x}) = \frac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}}{(\frac{1}{2})^n} = \begin{cases} 1 & \theta = \frac{1}{2} \\ 0 & \theta = 0 \end{cases}$$

بنابراین $\delta^* = E(\theta|\underline{x}) = 1 \times \frac{1}{2} = \frac{1}{2}$ برآوردگر بیز است.

- اگر $\sum x_i = 0$ باشد

$$h(\theta|\underline{x}) = \frac{(1-\theta)^n}{\sum_{\theta=0,\frac{1}{2}} (1-\theta)^n} = \frac{(1-\theta)^n}{1+2^{-n}} = \begin{cases} \frac{2^{-n}}{1+2^{-n}} & \theta = \frac{1}{2} \\ \frac{1}{1+2^{-n}} & \theta = 0 \\ 0 & \text{سایر مقادیر} \end{cases}$$

برای $\sum x_i = 0$ باشد

$$\delta^* = E(\theta|\underline{x}) = \frac{2^{-n}}{(1+2^{-n})} \neq 0, \frac{1}{2} \quad \delta^* \notin \Theta$$

لذا در این حالت برآوردگر بیز وجود ندارد.

مثال ۲- اگر $X_1, X_2, \dots, X_n \overset{i.i.d}{\sim} N(\theta, 1)$ که در آن $\theta \sim N(0,1)$ است،
تخمین بیز تحت تابع زیان مربع خطا چیست؟

$$h(\theta|\underline{x}) \propto f(\underline{x}|\theta) \cdot \pi(\theta)$$

$$h(\theta|\underline{x}) \propto \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum(x_i-\theta)^2} \left(\frac{1}{\sqrt{2\pi}} e^{-\theta^2/2}\right)$$

$$\propto e^{-\frac{\theta^2}{2} + n\theta\bar{x} - \frac{n\theta^2}{2}}$$

$$= \exp\left\{-\frac{n+1}{2}\left(\theta^2 - \frac{2n\theta\bar{x}}{n+1}\right)\right\}$$

$$\propto \exp\left\{-\frac{n+1}{2}\left(\theta - \frac{n\bar{x}}{n+1}\right)^2\right\}$$

$$\Rightarrow \quad \theta|\underline{x} \sim N\left(\frac{n\bar{x}}{n+1}, \frac{1}{n+1}\right)$$

تخمین نقطه‌ای بیز $\delta^* = \frac{n\bar{x}}{n+1} = \frac{\sum x_i}{n+1}$ است.

هر توزیعی که ملاحظه می‌شود، پیش‌توزیع (prior) توزیعی است که پس‌توزیع (posterior) حاصل از آن، از همان خانواده توزیع‌ها باشد. چنین توزیع‌هایی را توزیع مزدوج (conjugate) گوییم.

توزیع نرمال توزیع مزدوج خودش است.

توزیع بتا توزیع مزدوج توزیع ... است.

توزیع گاما ... ...

توزیع ... ...

توزیع ... ...

توزیع ... ...

— تذکر: یکی از اشکالات عمده روش بیزی ضرورت به کار گرفتن توزیع پیشین برای پارامتر مجهول است. حال اگر اطلاعات چندانی در مورد این پارامتر نداشته و نتوانیم توزیع معینی را به عنوان توزیع پیشین در نظر بگیریم. در این حال از چنین توزیع‌هایی استفاده می‌کنیم که اطلاعات زیادی در مورد پارامتر نداشته باشند (non informative) و به آنها توزیع‌های پیشین ناآگاه گفته می‌شود.

مثلاً اگر پارامتر p در یک آزمایش برنولی باشد، توزیع یکنواخت روی بازه $(0,1)$ یک توزیع پیشین ناآگاه است. و یا اگر توزیع نرمال برای پارامتر μ، توزیع نامناسب $\pi(\mu)=1$, $\mu \in \mathbb{R}$ را می‌توان توزیع پیشین ناآگاه در نظر گرفت.

جفری Jeffrey ثابت کرد که اگر $\pi(\theta) \propto \sqrt{I(\theta)}$ که در آن $I(\theta)$ اطلاع فیشر θ است، آنگاه این توزیع پیشین ناآگاه (اطلاع) می‌باشد.

مثال: اگر $\pi(\theta)=1 \quad \forall \theta \in \mathbb{R}$, $X_1, \dots, X_n \sim N(\theta, 1)$ ثابت کنید، تحت SEL، $\bar{X}$ کمترین ریسک بیز برای θ است.

$$h(\theta|\underline{x}) \propto f(\underline{x}|\theta)\pi(\theta)$$
$$\propto \exp\left\{-\frac{n}{2}\theta^2 + \frac{2n\theta\bar{x}}{2}\right\} = \exp\left\{-\frac{n}{2}(\theta^2 - 2\theta\bar{x})\right\}$$
$$\propto \exp\left\{-\frac{n}{2}(\theta-\bar{x})^2\right\}$$

$$\Rightarrow \theta|\underline{x} \sim N(\bar{x}, \tfrac{1}{n}) \Rightarrow \boxed{\delta^* = \bar{x}}$$

مثال: $X_1, \dots, X_n \sim N(\theta, \sigma^2)$ که $\sigma^2 \sim R(\alpha, \beta)$ ثابت SEL برآوردگر بیز σ^2 را بیابید.

حل: قرار دهید $\tau = \frac{1}{\sigma^2}$, $\alpha = 2$, $\beta = \frac{1}{2}$, $s = \sum_{i=1}^{n} x_i^2$

$$h(\tau|\underline{x}) \propto f(\underline{x}|\tau) \cdot \pi(\tau)$$

improper

informative

$$f(\underline{x}|\sigma^2) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum x_i^2\right\} \qquad \tau$$

$$y = \sum x_i^2 \,,\ r = \frac{n}{2} \,,\ \tau = \frac{1}{2\sigma^2}$$

$$\propto \tau^r \cdot e^{-\tau y} \propto f(\underline{x}|\tau)$$

$$\Pi(\tau) \propto \tau^{\alpha-1} \cdot e^{-\beta\tau}$$

$$h(\tau|\underline{x}) \propto (\tau^r e^{-\tau y})(\tau^{\alpha-1} e^{-\beta\tau})$$

$$= \tau^{r+\alpha-1} \cdot e^{-(y+\beta)\tau}$$

$$\Rightarrow \tau|\underline{x} \sim \Gamma\left(r+\alpha, \frac{1}{y+\beta}\right)$$

$$X \sim \Gamma(\alpha,\beta) \Rightarrow \mathbb{E}X^k = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}\beta^k \,,\ \mathbb{E}\left(\frac{1}{X}\right) = \frac{1}{\beta(\alpha-1)}$$

حال می‌خواهیم $\mathbb{E}(\sigma^2|\underline{x})$ که برآوردگر بیز σ^2 است را بیابیم. داریم:

$$\mathbb{E}(\sigma^2|\underline{x}) = \mathbb{E}\left(\frac{1}{2\tau}\middle|\underline{x}\right) = \frac{1}{2}\mathbb{E}\left(\frac{1}{\tau}\middle|\underline{x}\right)$$

$$= \frac{1}{2}\left[\frac{1}{\frac{1}{y+\beta}(r+\alpha-1)}\right] = \frac{\sum x_i^2 + \beta}{n+2\alpha-2}$$

$$\Rightarrow \delta^* = \frac{\sum x_i^2 + \beta}{n+2\alpha-2}$$

در این مثال نشان دادیم برآوردگر بیز یکتا است. اما در حالت کلی برآوردگر مینیماکس الزاما یکتا نیست. گاهی اوقات برآوردگر مینیماکس را برآوردگر بیز نسبت به توزیع پیشین نامناسب می‌یابیم. برآوردگرهای مینیماکس عموما مناسب نیستند.

تعریف. δ_0 را برآوردگر مینیماکس نامیده و θ گوییم اگر

$$\sup_{\theta\in\Theta} R(\theta,\delta_0) = \inf_{\delta\in D} \sup_{\theta\in\Theta} R(\theta,\delta)$$

قضیه: اگر $\delta^*(\underline{x})$ برآوردگری با ریسک ثابت مانند c باشد یعنی $R(\theta,\delta^*)=c$ (یعنی تابع ریسک به ازای تمام θ ها یکسان باشد) آنگاه δ^* برآوردگر مینیماکس است.

اثبات:

δ^* ریسک بیز $= \int R(\theta,\delta^*)\,\pi_0(\theta)\,d\theta$

$= \inf_{\delta\in D} \int R(\theta,\delta)\,\pi_0(\theta)\,d\theta$

$\le \inf_{\delta\in D} \int \sup_{\theta\in\Theta} R(\theta,\delta)\,\pi_0(\theta)\,d\theta = \inf_{\delta\in D}\sup_{\theta\in\Theta} R(\theta,\delta)$ ①

$r(\pi,\delta^*)=E_\theta(R(\theta,\delta^*))$

δ^* ریسک بیز $= \int R(\theta,\delta^*)\,\pi_0(\theta)\,d\theta$

$= c\int \pi_0(\theta)\,d\theta = c$

$= R(\theta,\delta^*)$

$= \sup_{\theta\in\Theta} R(\theta,\delta^*) \ge \inf_{\delta\in D}\sup_{\theta\in\Theta} R(\theta,\delta)$ ②

سمت راست بیز است، sup بزرگتر از هر چیز است

از ① و ② نتیجه می‌شود

δ^* ریسک بیز $= \sup_{\theta\in\Theta} R(\theta,\delta^*) = \inf_{\delta\in D}\sup_{\theta\in\Theta} R(\theta;\delta)$

پس δ^* برآوردگر مینیماکس است.

مثال: اگر $X_1, X_2, \dots, X_n$ از $b(1,\theta)$ و $\theta \sim Beta(\alpha,\beta)$ باشد، آنگاه برآوردگر بیز تحت SEL عبارت است از:

$$h(\theta|\underline{x}) \propto (\theta^{\sum x_i}(1-\theta)^{n-\sum x_i})(\theta^{\alpha-1}(1-\theta)^{\beta-1})$$

$$= \theta^{\sum x_i+\alpha-1}(1-\theta)^{n+\beta-\sum x_i-1}$$

$$\theta|\underline{x} \sim Beta(\alpha+\textstyle\sum x_i,\ n+\beta-\sum x_i).$$

δ^*: برآوردگر بیز $= E(\theta|\underline{x}) = \frac{\alpha + \sum_{i=1}^{n} x_i}{n+\alpha+\beta}$

حال تابع ریسک برآوردگر δ^* یعنی $R(\theta, \delta^*)$ را محاسبه می‌کنیم.

$$R(\theta, \delta^*) = E(\delta^* - \theta)^2$$
$$= E\left[\frac{\alpha + \sum x_i}{n+\alpha+\beta} - \theta\right]^2 = Var\left[\frac{\alpha + \sum x_i}{n+\alpha+\beta} - \theta\right] + \theta^2\left[\frac{\alpha + \sum x_i}{n+\alpha+\beta} - \theta\right]$$

$$= \frac{n\theta(1-\theta)}{(n+\alpha+\beta)^2} + \left(\frac{\alpha + n\theta}{n+\alpha+\beta} - \theta\right)^2$$

$$= \frac{n\theta - n\theta^2 + \alpha^2 + (\alpha+\beta)^2\theta^2 - 2\alpha(\alpha+\beta)\theta}{(n+\alpha+\beta)^2}$$

$$= \frac{[(\alpha+\beta)^2 - n]\theta^2 + [n - 2\alpha(\alpha+\beta)]\theta + \alpha^2}{(n+\alpha+\beta)^2}$$

واضح است که $R(\theta, \delta^*)$ تابعی درجه دوم از θ است که ضرایب θ^2 و θ و عبارت ثابت برابر هستند با

$$\begin{cases} (\alpha+\beta)^2 = n \\ 2\alpha(\alpha+\beta) = n \end{cases} \Rightarrow \begin{cases} \alpha = \sqrt{n}/2 \\ \beta = \sqrt{n}/2 \end{cases}$$

بنابراین اگر توزیع پیشین به صورت $Beta(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2})$ باشد آنگاه $R(\theta, \delta^*)$ تابعی ثابت برابر با

$$R(\theta, \delta^*) = \frac{n/4}{(n+\sqrt{n})^2} = \frac{n}{4(n+\sqrt{n})^2}$$

خواهد بود، یعنی برآوردگر بیز متناظر با این توزیع پیشین برآوردگر مینیمکس است.

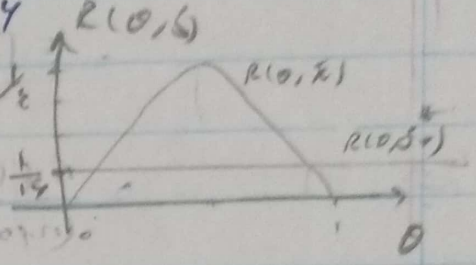
θ برآوردگر مینیمکس $\delta_0 = \frac{\frac{\sqrt{n}}{2} + \sum_{i=1}^{n} x_i}{n+\sqrt{n}}$

مثال: فرض کنید $X_1, \dots, X_n$ نمونه‌ای از توزیع برنولی با پارامتر θ باشد.

$$R(\theta, \bar{X}) = E(\bar{X}-\theta)^2 = Var(\bar{X}) = \frac{\theta(1-\theta)}{n}$$

حالت $n=1$:

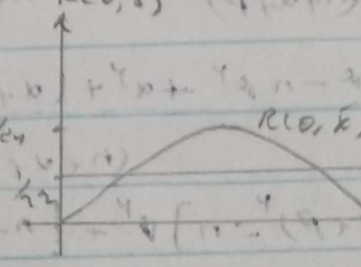
$$R(\theta, X) = \theta(1-\theta) \quad \& \quad R(\theta, \delta_0) = \frac{1}{4}$$



حالت $n=9$:

$$R(\theta, \bar{X}) = \frac{\theta(1-\theta)}{9}$$

$$R(\theta, \delta_0) = \frac{1}{36}$$



ملاحظه می‌شود که به ازای $n=1$، X نسبت به δ_0 برتری ندارد اما به ازای $n=9$ قضیه زیر نشان می‌دهد که δ_0 نسبت به X برتری دارد.

قضیه: فرض کنید $\{\pi_k(\theta)\}$ دنباله‌ای از توزیع‌های پیشین روی Θ و $\{\delta_k^*\}$ دنباله‌ای از تابع تصمیم‌های بیز متناظر با آن‌ها باشد.

$$\lim_{k\to\infty} r(\pi_k, \delta_k^*) = r^*$$

و تابع تصمیم‌گیری مانند δ_0 وجود دارد به طوری که

$$\sup_{\theta\in\Theta} R(\theta, \delta_0) \le r^*$$

آنگاه δ_0 تابع تصمیم مینیماکس است.

اثبات: (برهان خلف) فرض کنید δ_0 تابع تصمیم مینیماکس نباشد. پس تابع تصمیم δ وجود دارد به طوری که

$$\sup_{\theta\in\Theta} R(\theta, \delta) < \sup_{\theta\in\Theta} R(\theta, \delta_0)$$

از طرف دیگر بنا بر فرض ...

$$r(\pi_k, \delta_k^*) = \int R(\theta, \delta_k^*)\, \pi_k(\theta)\, d\theta \le \int R(\theta, \tilde{\delta})\, \pi_k(\theta)\, d\theta$$

(δ_k^* بیز است)

$$\le \int \sup_{\theta\in\Theta} R(\theta, \tilde{\delta})\, \pi_k(\theta)\, d\theta = \sup_{\theta\in\Theta} R(\theta, \tilde{\delta}) < \sup_{\theta\in\Theta} R(\theta, \delta_k)$$

بنابراین ثابت شد برآوردگر بیز منحصر به فرد minimax است.

$$r^* < \sup_{\theta\in\Theta} R(\theta, \delta_k)$$

که این خاصیت در قضیه زیر کاربرد فراوانی دارد.

مثال: فرض کنید $X_1, X_2, \dots, X_n$ نمونه‌ای از $N(\theta, 1)$ باشد و $\theta \sim N(0, k^2)$. با استفاده از قضیه قبل نشان دهید که $\bar{X}$ برآوردگر minimax برای θ است.

$$h(\theta|\underline{x}) \propto e^{-\frac{1}{2}\sum(x_i-\theta)^2} \cdot e^{-\theta^2/2k^2} \propto \exp\left\{\frac{-n\theta^2}{2} + n\theta\bar{x} - \frac{\theta^2}{2k^2}\right\}$$

$$\propto \exp\left\{-\frac{1+nk^2}{2k^2}\left(\theta - \frac{nk^2\bar{x}}{1+nk^2}\right)^2\right\}$$

$$\Rightarrow \theta|\underline{x} \sim N\left(\frac{nk^2\bar{x}}{1+nk^2}, \frac{k^2}{1+nk^2}\right)$$

π_k برآوردگر بیز نسبت به $\delta_k^* = E(\theta|\underline{x}) = \left(\frac{nk^2}{1+nk^2}\right)\bar{x}$

$$R(\theta, \delta_k^*) = E_\theta(\delta_k^* - \theta)^2 = E_x\left(\frac{nk^2}{1+nk^2}\bar{X} - \theta\right)^2$$

$$= Var\left(\frac{nk^2}{1+nk^2}\bar{X}\right) + E^2\left(\frac{nk^2}{1+nk^2}\bar{X} - \theta\right)$$

$$= \left(\frac{nk^2}{1+nk^2}\right)^2 \cdot \frac{1}{n} + \left[\frac{nk^2}{1+nk^2} - 1\right]^2 \theta^2$$

$$= \frac{nk^4}{(1+nk^2)^2} + \frac{1}{(1+nk^2)^2}\theta^2$$

و نهایتاً ریسک بیز δ_k^* عبارت است از:

$$r_k(\pi_k, \delta_k^*) = E_{\pi_k}[R(\theta, \delta_k^*)] =$$

$$= \frac{nk^2}{(1+nk^2)^2} + \frac{1}{(1+nk^2)^2} E_{\pi_k}(\theta^2) =$$

$$= \frac{nk^2}{(1+nk^2)^2} + \frac{k^2}{(1+nk^2)^2} = \frac{k^2(nk^2+1)}{(1+nk^2)^2} = \frac{k^2}{1+nk^2}$$

$$\Rightarrow r^* = \lim_{k\to\infty} r(\pi_k, \delta_k^*) = \frac{1}{n}$$

از طرفی

$$R(\theta, \bar{X}) = E(\bar{X}-\theta)^2 = Var(\bar{X}) = \frac{1}{n}$$

بنابراین

$$\sup_{\theta\in\Theta} R(\theta, \bar{X}) = \frac{1}{n} \leq r^* = \frac{1}{n}$$

و طبق قضیه ۱، $\bar{X}$ مرآد گر کمینه ریسک برای θ است.

تذکر: آنست که مشاهده مرآد گر کمینه ریسک، از دیدگاه بیز با بدترین توزیع پیشین است که توزیع پیشین نظیر ندارد.

تعریف: توزیع پیشین π_0 را بدترین (least favorable) توزیع پیشین گوییم هرگاه برای هر توزیع پیشین دلخواه π

$$r(\pi_0, \delta_0) \geq r(\pi, \delta) \quad \forall \pi$$

که در آن δ_0 مرآد گر بیز نسبت به π_0، و δ مرآد گر بیز نسبت به π است.

unbiasedness, efficiency, consistency
sufficiency
complete

فصل سوم: تعیین کننده‌های برآورد

در این فصل سه معیار برای مقایسه برآوردگرها معرفی می‌کنیم،
نااریبی، کارایی و سازگاری

نااریبی

اگر برای کمیت $g(\theta)$ برآوردی مانند $\delta(\underline{X})$ داشته باشیم:

(الف) نااریبی

δ را برای $g(\theta)$ نااریب گوییم:

$$E(\delta(\underline{X}) - g(\theta)) = 0 \quad \forall \theta \in \Theta$$

یا به عبارت معادل

$$E(\delta(\underline{X})) = g(\theta) \quad \forall \theta \in \Theta$$

برآوردگر $\delta_n(\underline{X})$ را برای $g(\theta)$ مجانباً نااریب گوییم:

asymptotically unbiased

$$\lim_{n\to\infty} E_\theta(\delta_n(\underline{X}) - g(\theta)) = 0$$

یعنی δ_n

بایاس یا اریبی

بایاس برآوردگر $\delta_n(\underline{X})$ برای $g(\theta)$ یعنی $b(\delta_n) = E(\delta_n(\underline{X})) - g(\theta)$

$\delta_n(\underline{X})$ را مجانباً نااریب برای $g(\theta)$ گوییم،

$$\lim_{n\to\infty} b(\delta_n) = 0$$

مثال: اگر $X_1, X_2, \ldots, X_n$ نمونه‌ای تصادفی از جامعه‌ای با میانگین متناهی μ باشد نشان دهید

$\bar{X} = \delta(\underline{X})$ برآوردگری نااریب برای μ است ولی

$\delta_n(\underline{X}) = \frac{n\bar{X}}{n+1}$ برآوردگری نااریب برای μ نیست ولی مجانباً نااریب است زیرا:

$$b(\delta_n) = \frac{n\mu}{n+1} - \mu = \frac{-1}{n+1}\mu \quad \& \quad \lim_{n\to\infty} b(\delta_n) = 0$$

تذکر: گاهی اوقات برای $g(\theta)$ هیچ برآوردگر نااریبی وجود ندارد. در این صورت

می‌گوییم $g(\theta)$، تابع نااریب قابل برآورد (یا تابع قابل برآورد) نیست.

برآورد پذیر

unestimable function

$\theta \in (0,1)$

مثال: اگر $X \sim B(n,\theta)$، در این صورت $g(\theta) = \frac{1}{\theta}$ را برآوردگر نااریب ندارد. زیرا اگر $\delta(x)$ برآوردگر نااریب برای $g(\theta)$ باشد (برای هر θ) آنگاه:

$$E_\theta(\delta(X)) = \sum_{x=0}^{n} \delta(x) \binom{n}{x} \theta^x (1-\theta)^{n-x} = \frac{1}{\theta} \quad \forall \theta \in (0,1)$$

و یا

$$\delta(0)(1-\theta)^n + \sum_{x=1}^{n} \delta(x) \binom{n}{x} \theta^x (1-\theta)^{n-x} = \frac{1}{\theta} \quad \forall \theta \in (0,1)$$

به سادگی می‌توان نشان داد که به ازای هر δ داریم $\theta \to 0$

$$\delta(0) + 0 = \infty$$

حتی اگر $g(\theta) = \theta$ باشد هیچ $\delta(X)$ برآوردگر نااریب نخواهد داشت. این مثال نشان می‌دهد که همیشه $\delta(X)$ برآوردگر نااریبی برای $g(\theta) = \frac{1}{\theta}$ وجود ندارد.

ب) کارایی:

اگر بخواهیم از میان $\delta(X)$ (تمام برآوردگرهای نااریب $g(\theta)$) تعریف کنیم برآوردگر کارا آن است که واریانس کوچکتر یا $g(\theta)$ داشته باشد. این تعریف صحیح نیست زیرا ممکن است (کوچکترین واریانس ممکن از آن برآوردگر باشد) واریانس $g(\theta)$ برای θ های مختلف متفاوت باشد.

۱) $\forall \varepsilon > 0 \quad \forall \theta \in \Theta \quad P_\theta(|\delta(X) - g(\theta)| \geq \varepsilon)$

۲) $\forall r \in \mathbb{N} \quad \forall \theta \in \Theta \quad E_\theta|\delta(X) - g(\theta)|^r$

۳) تابع $L(\theta,\delta)$، $\Pi(\theta)$ معلوم باشد

$$r(\Pi,\delta) = \int_\Theta R(\theta,\delta)\Pi(\theta)\,d\theta$$

که در آن $R(\theta,\delta) = E_\theta(L(\theta,\delta)) = \int L(\theta,\delta) f(x|\theta)\,dx$

۴) تابع $L(\theta,\delta)$ معلوم باشد $\sup_{\theta \in \Theta} R(\theta,\delta)$

تعریف: دنباله برآوردگرهای $\{\delta_n(\underline{x})\}$ را قویاً سازگار گوییم هرگاه

$$P_\theta\left(\lim_{n\to\infty} \delta_n(\underline{x}) = g(\theta)\right) = 1 \qquad \forall \theta \in \Theta$$

به زبان احتمالات (almost surely) $\delta(\underline{x}) \xrightarrow{a.s.} g(\theta)$ نشان می‌دهیم.

تعریف: دنباله برآوردگرهای $\{\delta_n(\underline{x})\}$ را برای $g(\theta)$ در مرتبه r سازگار گوییم هرگاه

$$\lim_{n\to\infty} E_\theta |\delta_n(\underline{x}) - g(\theta)|^r = 0 \qquad r>0$$

به زبان احتمالات $\delta_n(\underline{x}) \xrightarrow{r} g(\theta)$ نشان می‌دهیم.

اگر $r=2$ گوییم δ_n سازگار میانگین مربعات است.

$$\delta_n \xrightarrow{P} g(\theta) \Rightarrow \delta_n \xrightarrow{a.s} g(\theta)$$
$$\Leftrightarrow \delta_n \xrightarrow{P} \delta(\theta)$$

نتیجه:

$$X_n \xrightarrow{a.s} X \Rightarrow X_n \xrightarrow{P} X$$

$$X_n \xrightarrow{a.s} a \Leftrightarrow X_n \xrightarrow{P} a$$

قضیه: اگر δ_n دنباله‌ای از برآوردگرهای سازگار قویاً $g(\theta)$ و c_n دنباله‌ای از اعداد حقیقی باشد، آنگاه

$$\delta_n \xrightarrow{P} g(\theta) \ \&\ c_n \xrightarrow{n\to\infty} 0 \Rightarrow \delta_n + c_n \xrightarrow{P} g(\theta)$$

نتیجه: فرض کنید a, b, c ثابت‌های حقیقی باشند.

$$\left.\begin{array}{l} X_n \xrightarrow{P} a \\ Y_n \xrightarrow{P} b \\ Z_n \xrightarrow{P} c \end{array}\right\} \Rightarrow X_n + Y_n \cdot Z_n \xrightarrow{P} a + bc$$

نتیجه: میانگین نمونه‌ای برآوردگری سازگار قویاً برای میانگین جامعه است.

$$\overline{X^k} \xrightarrow{P} E(X^k)$$

زیرا $X_1^k, \dots, X_n^k$ متغیرهای تصادفی i.i.d هستند

$$\overline{X^k} = \frac{1}{n}\sum X_i^k \Rightarrow E\overline{X^k} = EX^k$$

$$P(|\overline{X^k} - EX^k| \geq \varepsilon) \leq \frac{Var(X^k)}{\varepsilon^2} = \frac{Var X^k}{n\varepsilon^2}$$

$$\lim_{n\to\infty} P(|\overline{X^k} - EX^k| \geq \varepsilon) = 0 \Rightarrow \overline{X^k} \xrightarrow{P} EX^k$$

این قضایا نشان می‌دهد که چگونه می‌توان از همگرایی دنباله‌ای از متغیرهای تصادفی به یک متغیر تصادفی ثابت، همگرایی توابع آن‌ها را نتیجه گرفت.

قضیه: اگر $X_n \xrightarrow{P} x$ و g تابعی پیوسته باشد، آنگاه $g(X_n) \xrightarrow{P} g(x)$.

قضیه: اگر δ_n دنباله‌ای از برآوردگرها برای θ باشد به‌طوری که $\lim_{n\to\infty} MSE(\delta_n) = 0$ (یعنی $Var(\delta_n) \to 0$ و $Bias(\delta_n) \to 0$)، آنگاه δ_n برآوردگری سازگار برای θ است.

$$P(|\delta_n - E(\delta_n)| \geq \varepsilon) \leq \frac{MSE(\delta_n)}{\varepsilon^2}$$

و چون $\lim_{n\to\infty} MSE(\delta_n) = 0$ نتیجه می‌شود که

$$\lim_{n\to\infty} P(|\delta_n - E(\delta_n)| \geq \varepsilon) = 0$$

$$\Rightarrow \delta_n \xrightarrow{P} \theta$$

مثال: اگر $X_1, \dots, X_n$ نمونه‌ای تصادفی از جامعه‌ای با میانگین μ و واریانس σ^2 باشد، آنگاه $\bar{X}$ برآوردگر سازگاری برای μ است.

اثبات: $MSE(\bar{X}) = Var(\bar{X}) = \frac{\sigma^2}{n} \xrightarrow[n\to\infty]{} 0$

پس $\bar{X} \xrightarrow{P} \mu$.

برای S^2 داریم: $Y_i = X_i - \mu \Rightarrow S_Y^2 = S_X^2$

پس کافی است نشان دهیم $S_Y^2 \xrightarrow{P} \sigma^2$

$$S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2\right)$$

$$= \frac{n}{n-1}\bar{Y}^2 - \frac{n}{n-1}\bar{Y}^2$$

$\bar{Y} \xrightarrow{P} 0$ ، $\bar{Y}^2 \xrightarrow{P} \sigma^2$ (با استفاده از قضیه ...)

$$S_Y^2 = \frac{n}{n-1}(\bar{Y}^2 - \bar{Y}^2) \xrightarrow{P} 1 \times (\sigma^2 - 0) = \sigma^2$$

قضیه: اگر $X_n \xrightarrow{P} x \Rightarrow kX_n \xrightarrow{P} kx$

اگر $a_n \xrightarrow{n\to\infty} a$ و $X_n \xrightarrow{P} x \Rightarrow a_n X_n \xrightarrow{P} ax$

$(X_n \xrightarrow{P} x,\ Y_n \xrightarrow{P} y) \Rightarrow X_n \pm Y_n \xrightarrow{P} x \pm y$

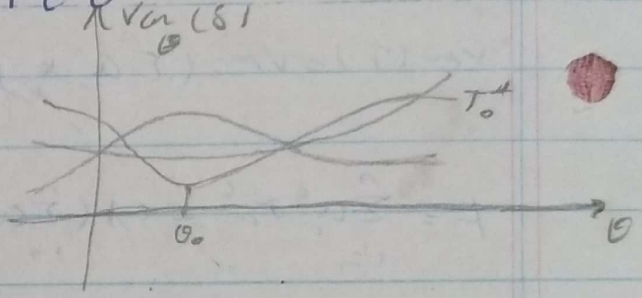
$X_n \cdot Y_n \xrightarrow{P} ?$ $\frac{X_n}{Y_n} \xrightarrow{P} ?$

$\forall \theta \in \Theta$, $\forall T \in U$ ، $\mathrm{Var}_\theta(T^*) \le \mathrm{Var}_\theta(T)$ را (θ) یا کوچکترین واریانس نااریب یکنواخت (UMVUE) یا آماره‌ی نااریب با کمترین واریانس یکنواخت (Uniformly Minimum Variance Unbiased Estimator) می‌نامیم.

T_0^* ، LMVUE را (آماره‌ی نااریب با کمترین واریانس یکنواخت موضعی) در $\theta = \theta_0$ یا (locally Minimum Variance Unbiased Estimator) گوییم اگر:

$\mathrm{Var}_{\theta_0}(T_0^*) \le \mathrm{Var}_{\theta_0}(T)$ $\forall T \in U$

در LMVUE ، T_0^* ، $\theta = \theta_0$ ،



تعریف: اگر $a_1, a_2, \ldots, a_n$ مقادیری ثابت باشند بطوری که $\sum a_i = 1$ ، و $X_1, X_2, \ldots, X_n$ نمونه‌ای تصادفی از توزیعی با پارامتر θ باشد، برآوردگر $T = \sum a_i X_i$ را یک برآوردگر خطی نااریب (LUE) برای $g(\theta)$ گوییم هرگاه $E(T) = g(\theta)$ و کمترین واریانس را در بین تمام برآوردگرهای خطی نااریب داشته باشد. (BLUE یا Best Linear Unbiased Estimator)

بنابراین هر برآوردگر خطی نااریب که بتوان برای $g(\theta)$ یافت که کمترین واریانس را داشته باشد یک BLUE است.

مثال: اگر $X_1, X_2, \ldots, X_n$ نمونه‌ای تصادفی از توزیعی با میانگین μ باشد، نشان دهید $\bar{X}$ یک BLUE برای μ است.

$T(X) = \sum_{i=1}^{n} a_i X_i \xrightarrow{\sum a_i = 1} E(T(X)) = \mu$

$\mathrm{Var}(T(X)) = \sum_{i=1}^{n} a_i^2 \sigma^2$

a_i ها را طوری تعیین می‌کنیم که $\sum a_i^2$ مینیمم شود.

$\sum a_i^2 = \sum (a_i - \frac{1}{n} + \frac{1}{n})^2$

$= \sum_{i=1}^{n} (a_i - \frac{1}{n})^2 + \frac{1}{n} + \frac{2}{n} \sum_{i=1}^{n} (a_i - \frac{1}{n})$ $\longrightarrow a_i = \frac{1}{n}$

پس $\bar{X}$ یک BLUE است.

$$T^* = \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2 \sum_{j=1}^{n} 1/\sigma_j^2}$$

$$T = \sum_{i=1}^{n} a_i x_i \quad ; \quad \sum_{i=1}^{n} a_i = 1$$

$$Var(T) = Var\left(\sum a_i x_i\right) = \sum a_i^2 Var(x_i) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$$

$$L = \sum_{i=1}^{n} a_i^2 \sigma_i^2 - \lambda\left(\sum_{i=1}^{n} a_i - 1\right)$$

$$\frac{\partial L}{\partial a_i} = 2a_i \sigma_i^2 - \lambda = 0 \quad ; \quad i = 1, \dots, n$$

$$\Rightarrow a_i = \frac{\lambda}{\sigma_i^2} \quad ; \quad i = 1, \dots, n \quad \overset{(*)}{\Rightarrow} \quad \sum_{i=1}^{n} a_i = \lambda \sum_{i=1}^{n} \frac{1}{\sigma_i^2}$$

$$\sum_{i=1}^{n} a_i = 1 \Rightarrow \lambda = \frac{1}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$$

از (*) جایگذاری می‌کنیم.

$$a_i = \frac{1}{\sigma_i^2 \sum_{j=1}^{n} 1/\sigma_j^2} \Rightarrow T^* = \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2 \sum_{j=1}^{n} \frac{1}{\sigma_j^2}}$$ است BLUE برای μ

- قضیه (لهمن-شفه): اگر $\mathcal{P}$ یک کلاس پارامتری از توزیع‌های با تابع امتیاز U_θ باشد و T^* یک برآوردگر نااریب از $\tau(\theta)$ باشد که واریانس متناهی دارد و به ازای هر $\theta \in \Theta$ و هر $V \in U_0$ داشته باشیم $E(V \cdot T^*) = 0$ آنگاه T^* یک برآوردگر MVUE برای $\tau(\theta)$ است.

اثبات: (شرح داده نشد)

فرض کنید λ مقدار ثابتی باشد و $V \in U_0$ دلخواه باشد. آنگاه $T^* + \lambda V$ برآوردگر

برآوردگر دیگر

$T = T^* + \lambda V$ برآوردگر نااریب T را در نظر بگیرید. T^* برآوردگر نااریب است که V با آن همبستگی منفی دارد.

$E(T) = E(T^*) + \lambda E(V) = \theta \quad \forall \theta \in \Theta$

$E(V) = 0$ فرض کنید

T^* بهتر از T است.

$Var_\theta(T^*) \le Var_\theta(T) \quad \forall \theta \in \Theta$

$Var_\theta(T^*) \le Var_\theta(T^* + \lambda V) = Var_\theta(T^*) + \lambda^2 Var_\theta(V) + 2\lambda Cov_\theta(T^*, V)$

$\Rightarrow \lambda^2 Var_\theta(V) + 2\lambda Cov_\theta(T^*, V) \ge 0 \quad \forall \theta \in \Theta$

حال فرض کنید V دارای شرایط لازم برای تبدیل به آماره کامل باشد.

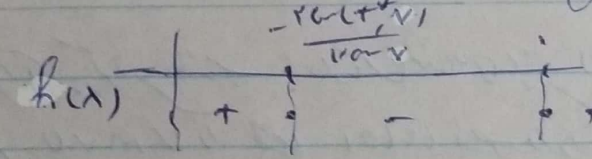
$Cov_\theta(T^*, V) = 0 \quad \forall \theta \in \Theta \Rightarrow E(VT^*) = 0 \quad \forall \theta \in \Theta$

توزیع‌های (توابع) که $Cov(T^*, V)$ آن‌ها صفر است.

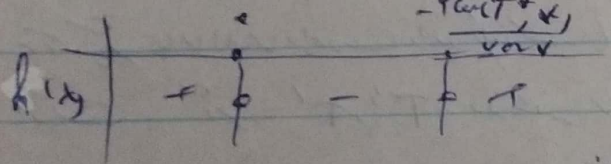
$Cov(T^*, V) > 0$ یا $Cov(T^*, V) < 0$ در این حالت T^* بهتر نیست.

$\lambda' = 0, \quad \lambda'' = \frac{-2Cov(T^*, V)}{Var(V)}$

اگر $Cov(T^*, V) > 0$

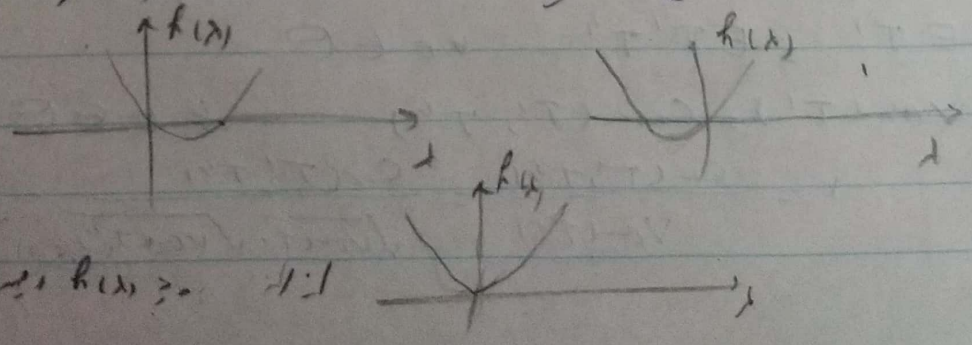


اگر $Cov(T^*, V) < 0$



که با فرض بهتر بودن T^* تناقض دارد. لذا باید

$Cov_\theta(T^*, V) = 0 \quad \forall \theta \in \Theta \Rightarrow E(VT^*) = 0 \quad \forall \theta \in \Theta$



حداقل یکی از این سه حالت باید برقرار باشد که $h(\lambda) \ge 0$

تعریف: فرض کنید T^* برآوردگری باشد که $E(V.T^*)=0 \quad \forall \theta \in \Theta$ برای هر V که $E_\theta V=0 \quad \forall \theta \in \Theta$ و $E_\theta V^2<\infty$. در این صورت اگر T برآوردگر دیگری از $g(\theta)$ باشد و $T \in U_{g(\theta)}$، آنگاه بردار $V=T-T^*$ متعلق به U_0 است و داریم بنابراین

$$E(T^*.V)=E(T^*(T-T^*))=0 \quad \forall \theta \in \Theta$$

$$\Rightarrow E T^{*2}=E(T^*.T) \quad \forall \theta \in \Theta$$

$$E T^{*2}-g^2(\theta)=E(T^*.T)-g(\theta) \quad \forall \theta \in \Theta$$

$$Var(T^*)=Cov(T,T^*), \quad \forall \theta \in \Theta$$

$$=\rho_{T,T^*}\sqrt{Var(T^*)}\sqrt{Var(T)}$$

$$\Rightarrow \sqrt{\frac{Var(T^*)}{Var(T)}}=\rho_{T,T^*} \leq 1 \quad \forall \theta \in \Theta$$

$$Var(T^*) \leq Var(T) \quad \forall \theta \in \Theta$$

پس T^* کمترین واریانس را در $g(\theta)$ دارد.

قضیه ۳۵: اگر U فضای برآوردگرهای نااریب با واریانس متناهی از $g(\theta)$ غیرتهی باشد آنگاه، حداکثر یک برآوردگر $UMVU$ از $g(\theta)$ وجود دارد (یعنی اگر دو برآوردگر $UMVU$ وجود داشته باشند یکتا هستند).

اثبات: اگر T' و T'' دو عضو باشند که هر دو $UMVUE$ هستند آنگاه

$$T'-T'' \in U_0$$

بنابراین $E(T'.(T'-T''))=0 \quad \forall \theta \in \Theta$

$$\Rightarrow E T'^2=E T'T'' \quad \forall \theta \in \Theta$$

$$Var(T')=Cov(T',T'') \quad \forall \theta \in \Theta$$

$$1=\frac{Cov(T',T'')}{Var(T')}=\frac{Cov(T',T'')}{\sqrt{Var(T')}\sqrt{Var(T'')}}=\rho_{T',T''}$$

برای اثبات کنیم، هر برآوردگر خطی اریب بین T' و T'' برابر است با

$\exists a, b$ s.t. $T' = a + bT''$ w.p. 1 $\forall \theta \in \Theta$

$E(T') = a + b\,ET'' \Rightarrow g(\theta) = a + b\,g(\theta) \Rightarrow a = 0, b = 1$

$\Rightarrow T' = T''$ w.p. 1 for all θ.

قضیه ۳: اگر $\{T_i(\underline{x})\}$ برآوردگرهای UMVU برای $\{g_i(\theta)\}$ باشند، آنگاه $T(\underline{x}) = \sum_{i=1}^{k} a_i T_i(\underline{x})$ برآوردگر UMVU برای $g(\theta) = \sum_{i=1}^{k} a_i g_i(\theta)$ است.

اثبات: چون $T_i(\underline{x})$ها UMVU هستند هر یک قضیه ۱

$E_\theta(T_i(\underline{x}) \cdot V(\underline{x})) = 0 \quad \forall \theta \in \Theta, V \in U_0$ برای $i = 1, \dots, k$

$E(T(\underline{x}) \cdot V(\underline{x})) = E\left(\left(\sum_{i=1}^{k} a_i T_i(\underline{x})\right) \cdot V(\underline{x})\right) =$

$\sum_{i=1}^{k} a_i E(T_i(\underline{x}) \cdot V(\underline{x})) = 0 \quad \forall V \in U_0, \forall \theta \in \Theta$

پس بنابر قضیه ۱، $T(\underline{x})$ یک UMVU برای امید ریاضی خودش است

$E_\theta(T(\underline{x})) = E\sum_{i=1}^{k} a_i T_i(\underline{x}) = \sum_{i=1}^{k} a_i ET_i(\underline{x}) = \sum_{i=1}^{k} a_i g_i(\theta) = g(\theta)$

قضیه ۴: اگر $T(\underline{x})$ یک UMVU $g(\theta)$ باشد آنگاه $T^* = T^k(\underline{x})$ نیز برای همین تابع UMVU است.

اثبات: چون $T(\underline{x})$ یک UMVU است، بنابر قضیه ۱

$E(T(\underline{x}) \cdot V) = 0 \quad \forall \theta \in \Theta, \forall V \in U_0$

$U_1 \in U_0$ را در نظر بگیرید، $U_1(\underline{x})$ هم برآوردگر نااریب صفر است پس

$E(T(\underline{x}) \cdot U_1(\underline{x})) = E(T(\underline{x}) \cdot U_1(\underline{x})) = 0 \quad \forall \theta \in \Theta$

یعنی ET^2 برای UMVU است.

64

$$E(T^2 \cdot V) = E(T^2 \cdot \overbrace{T \cdot V}^{V_1}) = E(T^2 \cdot V_1) = E(T \cdot \overbrace{T \cdot V_1}^{V_2}) = E(T \cdot V_2) = 0$$

لطفاً: $U_1(\underline{x}) \in U_0$ داریم $U_1(\underline{x}) = T(\underline{x}) \cdot U_0(\underline{x})$

$$E(T^2 \cdot U(\underline{x})) = E(T(\underline{x}) T U_1(\underline{x})) = E(T(\underline{x}) \cdot T^2 U_0(\underline{x})) = 0$$

پس T^2 یک UMVUE از θ^2 است. با شکل بازنویسی در KGW قضیه برقرار است.

حذف

نکته مهم: در عمل برای پیدا کردن برآوردگر UMVUE یکی از سه حالت زیر پیش می‌آید:

۱- بعضی اوقات از تعداد لطف نابایب ناریب برآوردگر پارامتر θ، برآورد UMVUE بدست می‌آید.

۲- هر خطی از تعداد لطف نابایب ناریب برآوردگر پارامتر θ دارای UMVUE نیست و اصلاً وجود ندارد.

۳- بعضی تعداد لطف نابایب ناریب برآوردگر پارامتر θ دارای تعداد UMVUE بیش از یک است.

مثال بررسی حالت اول: فرض کنید $X_1, \dots, X_n$ نمونه تصادفی از توزیع

$$f_\theta(x) = \begin{cases} \frac{1}{2} & x = \theta-1, \theta, \theta+1 \\ 0 & \text{سایر جاها} \end{cases} \qquad \theta \in \mathbb{Z}$$

ثابت می‌شود $T(\underline{x}) = (X_{(1)}, X_{(n)})$ کلی بارتر تعداد لطف نابایب ناریب برآوردگر پارامتر θ است. (تمرین ۲.۱.۸ کتاب P.E صفحه ۱۲۱)

تابع چگالی $X_{(1)}$ عبارت است از

$$f_\theta(x) = \begin{cases} 0 & x = -1 \\ (1-\theta)^2 \cdot \theta^x & x = 0, 1, \dots \end{cases}$$

ثابت می‌شود $\theta_1(\theta) = (1-\theta)^2$ تعداد نابایب ناریب برآوردگر است، و دارای تابع UMVUE است.

و نیز ثابت می‌شود $\theta_2(\theta) = \theta$ تعداد نابایب ناریب برآوردگر است، ولی دارای تابع UMVUE نیست.

حل: فرض کنید $n=1$ و داریم $T(X) = X$ تعداد نابایب ناریب θ است، پس

$$E X = \sum_{x=-1}^{\infty} x f_\theta(x) = -\theta + \sum_{x=0}^{\infty} x(1-\theta)^2 \theta^x = -\theta + (1-\theta)^2 \cdot \frac{\theta}{(1-\theta)^2} = 0$$

پس نتیجه: اگر $\theta \neq 0$ برآوردگر UMVUE برای θ وجود ندارد.

$$E(T^* \cdot X) = 0 \qquad \forall \theta \in \Theta \qquad (1)$$

از آنجایی که تابع برآوردگر نااریب همراه با حداقل واریانس یکتا است
برآوردگر $U(x)=ax$ است که a مقداری ثابت است. اگر $U(x)$ تابع
برآوردگر نااریب θ باشد:

$$E(U(x))=0 \Rightarrow \sum_{x=-1}^{\infty} U(x) f_\theta(x)=0$$

$$\Rightarrow U(-1)\cdot\theta+(1-\theta)^2\sum_{x=0}^{\infty} U(x)\theta^x=0$$

$$\Rightarrow \sum_{x=0}^{\infty} U(x)\theta^x=\frac{\theta}{(1-\theta)^2}\underbrace{\left(-U(-1)\right)}_{a}$$

$$\Rightarrow \sum_{x=0}^{\infty} U(x)\theta^x=\sum_{x=0}^{\infty} a\cdot x\,\theta^x \Rightarrow U(x)=a\cdot x$$

$$U(x)=a\cdot x \qquad x=-1,0,1,2,\dots \qquad (2)$$

از (1) و (2) نتیجه می‌گیریم $T^*(x)=a\cdot x$; $x=-1,0,1,\dots$

$$\Rightarrow T^*(-1)=a \;,\; T^*(1)=a=2T^*(-1) \;,\; T^*(2)=a=2T^*(-1)$$

$$\Rightarrow T^*(x)=T^*(-1) \quad x=-1,1,2,\dots \;,\; T^*(-1)=b$$

در نتیجه:

$$T^*(x)=\begin{cases} a & x=-1,1,2,\dots \\ b & x=0 \end{cases}$$

$$U(\theta)=E(T^*(x))=\sum_{x=-1}^{\infty} T^*(x) f_\theta(x)=a\theta+b(1-\theta)^2+\sum_{x=1}^{\infty} a(1-\theta)^2\theta^x$$

$$=a\theta+b(1-\theta)^2+a(1-\theta)^2\cdot\frac{\theta}{1-\theta}=2a\theta+b(1-\theta)^2-a\theta^2$$

$$=a(2\theta-\theta^2)+b(1-\theta)^2=a\left(1-(1-\theta)^2\right)+b(1-\theta)^2$$

$$=a+(b-a)(1-\theta)^2$$

حال اگر $a=0$ ، $b=1$ داریم $T^*(x)=\begin{cases}0 & x=-1 \text{ یا } 1 \\ 1 & x=0\end{cases}$

که UMVUE برای $\theta(1-\theta)=\theta(\theta)$ است، چون a و b وجود ندارند که $\theta(\theta)=0$
لذا علی رغم اینکه $\theta(\theta)=0$ تنها تابع نااریب صفر از θ است ولی UMVUE ندارد.

$$T_0(x)=\begin{cases}1 & x=-1 \\ 0 & x=0 \text{ یا } 1\end{cases} \qquad E(T_0(x))=0 \cdot \sum_{x=0}^{\infty} 2^x f_\theta(x)=0$$

و مورد ۴ در قضیه زیر آمده است

قضیه (رائو-بلکول)

قضیه ۵: اگر $\mathcal{X}$ متغیر تصادفی با خانواده توزیع‌های $\{F_\theta(x);\theta\in\Omega\}$ باشد و T آماره کفایت کامل از این خانواده باشد آنگاه $\phi(\theta)$ که تابع نااریب صفر از θ است تنها دارای یک تابع نااریب صفر یکنواخت با کمترین واریانس است که تابعی از T است و این تابع نااریب صفر یکنواخت همان UMVUE است.

اثبات: چون $\phi(\theta)$ تابعی نااریب صفر از θ است پس تابع نااریب صفر دیگری مانند δ_0 از $\phi(\theta)$ وجود دارد. قرار دهید $\delta_*=E(\delta_0|T)$ که در آن T کفایت کامل است.
در این صورت δ_* تابعی نااریب صفر از $\phi(\theta)$ است (زیرا از قضیه $E(\delta_0|T)$ نیز کفایت کامل T و δ_0 تشکیل می‌دهد)

$$E(\delta_*)=E\,E(\delta_0|T)=E(\delta_0)=\phi(\theta)$$

حال از اینکه δ_* قرار دهید δ'_* و δ''_* دو تابع دیگر نااریب از $\phi(\theta)$ باشند که بر حسب T کفایت کامل و این صورت باشند

$$h(T)=\delta'_*(T)-\delta''_*(T)$$

$$E(h(T))=E(\delta'_*(T)-\delta''_*(T))=0 \qquad \forall\theta\in\Omega$$

و چون T کفایت کامل است نتیجه می‌شود که $\delta'_*(T)-\delta''_*(T)=0$ یا احتمال یک

یعنی با احتمال یک $\delta'_*(T)=\delta''_*(T)$.

و در پایان باید گفت که δ_* دارای کمترین واریانس در بین تابع‌های نااریب است

$$\mathrm{Var}_\theta(\delta_0) = \mathrm{Var}_\theta(E(\delta_0|T)) + E_\theta(\mathrm{Var}(\delta_0|T)) \geq \mathrm{Var}_\theta[E(\delta_0|T)]$$

$$\mathrm{Var}_\theta(\delta_0) \geq \mathrm{Var}_\theta(\delta_*) \quad \forall \delta_0 \in U, \forall \theta \in \Theta$$

لذا δ_* ، UMVUE است.

توجه کنید، در صورتی که فضای نمونه‌ای تابع T نقش محوری را بازی کند UMVUE ارائه می‌کند و این بدین معناست که اگر خانواده‌ی کامل نباشد هیچ تابعی از T یک برآوردگر نااریب کامل با واریانس یکنواخت کمینه نداریم. مثال زیر موضوع را روشن می‌کند.

مثال: $X \sim \delta_\theta(x) = \begin{cases} \theta & x=-1 \\ (1-\theta)^2\theta^x & x=0,1,2,\dots \end{cases}$ را در نظر بگیرید خانواده توزیع

$\{\delta_\theta(x); 0<\theta<1\}$ کامل نیست. زیرا با قرار دادن $n=1$ داریم:

$$E(u(x)) = 0 \Rightarrow u(x) = ax \neq 0$$

آقای شمال دارای نام $\theta^2(1-\theta)=\delta(\theta)$ تصور نادرستی از برآوردگر UMVUE دارد

تصور $\delta_*(x) = \begin{cases} 0 & x \neq 0 \\ 1 & x=0 \end{cases}$

مثال: اگر $X_1 \sim b(\theta,1)$ کدام یک از تابعیت‌های زیر دارای UMVUE است

الف) $\delta(\theta)=\theta$ ، ب) $\delta(\theta)=\theta^2$ ج) $\delta(\theta)=P_\theta(X_1=1)$

حل:

چون خانواده‌ی توزیع کامل است، $T=\sum_{i=1}^n X_i$ آماره‌ی کامل کفایت است، بنابراین هر تابع از T که نااریب باشد، UMVUE خواهد بود

$\bar{X} \sim b(\theta, \frac{1}{n})$ می‌دانیم

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = E\left(\frac{T}{n}\right) = \theta$$

پس $\bar{X} = \frac{T}{n}$، UMVUE برای θ است.

ب) اگر $\bar{X}^2$ را برای تخمین θ^2 در نظر بگیریم که تابعی از T است

داریم

$$E(\bar{X}^2) = \mathrm{Var}(\bar{X}) + E^2\bar{X}$$

$$= \frac{1}{n} + \theta^2 \Rightarrow E\left(\bar{X}^2 - \frac{1}{n}\right) = \theta^2$$

پس $\delta_*(T) = \bar{X}^2 - \frac{1}{n} = \left(\frac{T}{n}\right)^2 - \frac{1}{n}$ برای θ^2، UMVUE است.

$\delta^*(T) = E(\delta_0 | T)$

$$\delta_0(x_1) = \begin{cases} 1 & x_1 \le 1 \\ 0 & x_1 > 1 \end{cases}$$

$E(\delta_0(X_1)) = P(X_1 \le 1) = \Phi(0)$

$T = \sum_{i=1}^{n} X_i$

$\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1^2 = 1$, $\sigma_2^2 = n$, $\rho = 1/\sqrt{n}$

$$\rho_{X_1,T} = \frac{Cov(X_1, T)}{\sqrt{Var(X_1)}\sqrt{Var(T)}} = \frac{Cov(X_1, X_1 + X_2 + \dots + X_n)}{\sqrt{Var(X_1)} \cdot \sqrt{Var(T)}} = \frac{1}{\sqrt{n}}$$

$$W = X_1 | T \sim N\left(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho(t - \mu_2), \sigma_1^2(1-\rho^2)\right)$$

$$W = X_1 | T \sim N\left(\frac{t}{n}, 1 - \frac{1}{n}\right) \quad : \quad E W = E(X_1 | T) = \frac{T}{n} = \bar{X}$$

$$\delta^*(t) = E(\delta_0(X_1) | T = t) = \int_{-\infty}^{\infty} \delta_0(x) f_{X_1|T}(x|t)\, dx = P(X_1 \le 1 | T)$$

$$= \int_{-\infty}^{1} f_{X_1|T}(x|t)\, dx = P(W \le 1) = P\left(Z \le \frac{1 - t/n}{\sqrt{1 - 1/n}}\right)$$

$$\Rightarrow \delta^*(T) = P\left(Z \le \frac{1 - T/n}{\sqrt{1 - 1/n}}\right) = \Phi\left(\frac{\sqrt{n}}{\sqrt{n-1}}(1 - \bar{X})\right)$$

مثال: اگر $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$ نمونه تصادفی $(\sum X_i, \sum X_i^2)$ یا $(\bar{X}, S^2)$ آماره کفایتی کامل برای $\theta = (\mu, \sigma^2)$ است.

$\delta_1^*(\underline{X}) = \bar{X}$ is UMVUE for μ

$\delta_2^*(\underline{X}) = S^2$, , , σ^2

$\delta_3^*(\underline{X}) = \sqrt{\frac{n-1}{2}} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})}$, , , σ

مثال: اگر $X_1, X_2, \dots, X_n$ نمونه‌ای از $U\{1, \dots, \theta\}$ باشد که $\theta \in \mathbb{N}$ پارامتر است.

$$f_\theta(x) = \begin{cases} \frac{1}{\theta^n} & 1 \le x_{(1)} \le x_{(n)} \le \theta \\ 0 & \text{o.w} \end{cases}$$

$$f_\theta(x) = \frac{1}{\theta^n} \cdot u(x_{(1)} - 1) \, u(\theta - x_{(n)})$$

پس $x_{(n)}$ آماره کفایت برای θ است.

ثابت می‌کنیم که $x_{(n)}$ کامل نیز می‌باشد. برای این منظور ابتدا

توزیع $Y = x_{(n)}$ را بدست می‌آوریم.

$$G(y) = P(Y \le y) = P(x_{(n)} \le y) = [P(x \le y)]^n = \frac{y^n}{\theta^n} \quad y = 1, \dots, \theta$$

$$g(y) = G(y) - G(y-1) = \frac{y^n}{\theta^n} - \frac{(y-1)^n}{\theta^n} = \frac{y^n - (y-1)^n}{\theta^n} \quad y = 1, \dots, \theta$$

حال فرض کنیم $h(Y)$ تابع ناریبی از $Y = x_{(n)}$ باشد داریم

$$E(h(Y)) = 0 \quad \forall \theta \in \mathbb{N}$$

$$\Rightarrow \sum_{y=1}^{\theta} h(y) \cdot \frac{y^n - (y-1)^n}{\theta^n} = 0 \quad \forall \theta \in \mathbb{N}$$

$$\Rightarrow \sum_{y=1}^{\theta} h(y)\left(y^n - (y-1)^n\right) = 0 \quad \forall \theta \in \mathbb{N}$$

اگر $\theta = 1$ باشد از روی فرض نتیجه می‌شود

$$h(1)(1-0) = 0 \Rightarrow h(1) = 0$$

اگر $\theta = 2$ باشد نتیجه می‌شود.

$$h(1)(1-0) + h(2)(2^n - 1) = 0 \xrightarrow{h(1)=0} h(2) = 0$$

~~$\forall \theta \in \mathbb{N}$~~

به راحتی ثابت می‌شود $h(y) = 0$ ، $y = 1, \dots, \theta$ ، $\forall \theta \in \mathbb{N}$

پس $Y = x_{(n)}$ یک آماره کامل نمونه‌ای است.

از طرفی $EX_1 = \sum_{x=1}^{\theta} \frac{x}{\theta} = \frac{\theta+1}{2}$ پس $\delta_0 = 2X_1 - 1$

از آنجایی که تابع ناریب کارا، حداکثر یک تابع ناریب

۲۰

$$P(X_1 = x,\ X_{(n)} \le y) = P(X_1 = x \mid X_{(n)} \le y)\, P(X_{(n)} \le y)$$

$$= \frac{1}{y} \cdot \frac{y^n}{\theta^n} = \frac{y^{n-1}}{\theta^n}$$

$$P(X_1 = x,\ X_{(n)} = y) = P(X_1 = x,\ X_{(n)} \le y) - P(X_1 = x,\ X_{(n)} \le y-1)$$

$$= \begin{cases} \dfrac{y^{n-1}}{\theta^n} - \dfrac{(y-1)^{n-1}}{\theta^n} & x = 1, 2, \dots, y-1 \\ \dfrac{y^{n-1}}{\theta^n} & x = y \end{cases}$$

$$P(X_1 = x \mid X_{(n)} = y) = \begin{cases} \dfrac{y^{n-1} - (y-1)^{n-1}}{y^n - (y-1)^n} & x = 1, 2, \dots, y-1 \\ \dfrac{y^{n-1}}{y^n - (y-1)^n} & x = y \end{cases}$$

$$E(X_1 \mid y) = \sum_{x=1}^{y} x\, P(X_1 = x \mid X_{(n)} = y)$$

$$= \sum_{x=1}^{y-1} x\, P(X_1 = x \mid X_{(n)} = y) + y\, P(X_1 = y \mid X_{(n)} = y)$$

$$= \frac{y(y-1)}{2} \left[\frac{y^{n-1} - (y-1)^{n-1}}{y^n - (y-1)^n} \right] + \frac{y^n}{y^n - (y-1)^n}$$

$$= \frac{y^{n+1} - (y-1)^{n+1}}{y^n - (y-1)^n}$$

برآورد بیز θ با توجه به آماره کافی $X_{(n)}$

$$\delta_A(X_{(n)}) = \frac{X_{(n)}^{n+1} - (X_{(n)} - 1)^{n+1}}{X_{(n)}^n - (X_{(n)} - 1)^n}$$

مثال: اگر $X \sim B(n, \theta)$ و $\theta(1-\theta)$ تابع برآوردپذیر یکانی باشد

$T = X$ آماره کافی

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

$$\Rightarrow \sum_{t=0}^{n} g(t) \binom{n}{t} \theta^{t-1} (1-\theta)^{n-t-1} = 1 \Rightarrow g(t) \frac{\binom{n}{t}}{\binom{n-2}{t-1}} = 1$$

2)

$$\Rightarrow g(t) = \frac{\binom{n-t}{t-1}}{\binom{n}{t}} = \frac{t(n-t)}{n(n-1)} \Rightarrow g(T) = \frac{T(n-T)}{n(n-1)}$$

θ^2 ... $X_1, \dots, X_n \sim N(0, \theta)$ ... Ω ...

$\frac{X_{(n)}}{\theta} \sim Beta(n, 1)$ ... $X_{(n)}$ ...

$$\Rightarrow E(X_{(n)}) = \frac{n}{n+1}\theta \qquad Var(X_{(n)}) = \frac{n}{(n+2)(n+1)^2}\theta^2$$

$$\Rightarrow E(X_{(n)})^2 = Var + E^2$$

$$= \frac{n}{(n+2)(n+1)^2}\theta^2 + \frac{n^2}{(n+1)^2}\theta^2$$

$$= \frac{n + n^2(n+2)}{(n+2)(n+1)^2}\theta^2 = \frac{n(1+n(n+2))}{(n+2)(n+1)^2}\theta^2$$

$$= \frac{n(1+n^2+2n)}{(n+2)(n+1)^2}\theta^2 = \frac{n}{n+2}\theta^2 \Rightarrow \boxed{g(T) = \frac{n+2}{n}X_{(n)}^2}$$

$X \sim B(n, \theta)$ ... $0 < \theta < 1$ ...

$$h(\theta) = \theta(1-\theta)$$

$$E(\delta(X)) = \theta(1-\theta) \Rightarrow \sum_{x=0}^{n} \delta(x)\binom{n}{x}\theta^x(1-\theta)^{n-x} = \theta(1-\theta)$$

$$\Rightarrow \sum_{x=0}^{n} \delta(x)\binom{n}{x}\left(\frac{\theta}{1-\theta}\right)^x = \theta(1-\theta)^{n-1} \qquad \rho = \frac{\theta}{1-\theta} \Rightarrow \theta = \frac{\rho}{1+\rho}$$

$$1 - \theta = \frac{1}{1+\rho}$$

$$\Rightarrow \sum_{x=0}^{n} \delta(x)\binom{n}{x}\rho^x = \rho(1+\rho)^{n-2} = \rho\sum_{x=0}^{n-2}\binom{n-2}{x}\rho^x$$

$$= \sum_{x=0}^{n-2}\binom{n-2}{x}\rho^{x+1} \qquad x+1 = k$$

$$\Rightarrow \sum_{x=0}^{n}\delta(x)\binom{n}{x}\rho^x = \sum_{k=1}^{n-1}\binom{n-2}{k-1}\rho^k$$

$$\Rightarrow \delta(0) = 0 \quad , \quad \delta(n) = 0 \quad , \quad \delta(x)\binom{n}{x} = \binom{n-2}{x-1} \qquad x = 1, \dots, n-1$$

$$\Rightarrow \delta(x) = \frac{x(n-x)}{n(n-1)} \qquad x = 0, 1, \dots, n$$

$$= \int t(\underline{x}) \frac{\partial_{\theta} f_{\theta}(\underline{x})}{f_{\theta}(\underline{x})} \cdot f_{\theta}(\underline{x}) \, d\underline{x} = \int [t(\underline{x}) \cdot \partial_{\theta} \log f_{\theta}(\underline{x})] f_{\theta}(\underline{x}) \, d\underline{x}$$

$$= E[T(\underline{x}) \cdot \partial_{\theta} \log f_{\theta}(\underline{x})] \quad \textcircled{1}$$

$$= \mathrm{Cov}(T(\underline{x}), \partial_{\theta} \log f_{\theta}(\underline{x}))$$

$$E(\partial_{\theta} \log f_{\theta}(\underline{x})) = \int \partial_{\theta} \log f_{\theta}(\underline{x}) \cdot f_{\theta}(\underline{x}) \, d\underline{x}$$

$$= \int \frac{\partial_{\theta} f_{\theta}(\underline{x})}{f_{\theta}(\underline{x})} \cdot f_{\theta}(\underline{x}) \, d\underline{x} = \int \partial_{\theta} f_{\theta}(\underline{x}) \, d\underline{x}$$

$$= \partial_{\theta} \int f_{\theta}(\underline{x}) \, d\underline{x} = 0 \quad \textcircled{2}$$

$$\psi'(\theta) + b'(\theta) = E(T(\underline{x}) \cdot \partial_{\theta} \log f_{\theta}(\underline{x}))$$

$$\textcircled{2} \Rightarrow = \mathrm{Cov}(T(\underline{x}), \partial_{\theta} \log f_{\theta}(\underline{x}))$$

$$\leq \sqrt{\mathrm{Var}(T(\underline{x}))} \sqrt{\mathrm{Var}(\partial_{\theta} \log f_{\theta}(\underline{x}))}$$

$$= \sqrt{\mathrm{Var}(T(\underline{x}))} \cdot \sqrt{E(\partial_{\theta} \log f_{\theta}(\underline{x}))^2}$$

$$\Rightarrow (\psi'(\theta) + b'(\theta))^2 \leq \mathrm{Var}(T) \cdot E(\partial_{\theta} \log f_{\theta}(\underline{x}))^2$$

$$\Rightarrow \mathrm{Var}(T) \geq \frac{(\psi'(\theta) + b'(\theta))^2}{E(\partial_{\theta} \log f_{\theta}(\underline{x}))^2}$$

نامساوی کرامر رائو

- اگر $\psi(\theta) = \theta$ آنگاه $\mathrm{Var}(T) \geq \frac{(1 + b'(\theta))^2}{I_{\underline{x}}(\theta)}$
- اگر $\psi(\theta) = \theta$ و $b(\theta) = 0$ آنگاه $\mathrm{Var}(T) \geq \frac{1}{I_{\underline{x}}(\theta)}$
- اگر T ناریب باشد آنگاه $\mathrm{Var}(T) \geq \frac{(\psi'(\theta))^2}{I_{\underline{x}}(\theta)}$

آماره T که واریانس آن برابر کران پایین کرامر رائو است را آماره کارا می‌نامیم.

$$f_\theta(\underline{x}) = \prod_{i=1}^{n} f_\theta(x_i)$$

$$\log f_\theta(\underline{x}) = \sum_{i=1}^{n} \log f_\theta(x_i)$$

$$\frac{\partial}{\partial\theta} \log f_\theta(\underline{x}) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f_\theta(x_i)$$

- اگر x_i ها از توزیع یکسان و مستقل (iid) باشند f_θ به شکل زیر است.

$$I_{\underline{x}}(\theta) = E\left[\frac{\partial}{\partial\theta} \log f_\theta(\underline{x})\right]^2 = \mathrm{Var}\left[\frac{\partial}{\partial\theta} \log f_\theta(\underline{x})\right] =$$

$$= n E\left[\frac{\partial}{\partial\theta} \log f_\theta(x_1)\right]^2 = n I_{x_1}(\theta)$$

(*) بنابراین کران کرامر-رائو برابر است با $\mathrm{Var}(T(\underline{x})) \geq \frac{(\psi'(\theta))^2}{n I_{x_1}(\theta)}$

* تنها شرایط منظم می‌توانند نشان داده شوند.

$$E_\theta\left(\frac{\partial}{\partial\theta} \log f_\theta(\underline{x})\right)^2 = -E_\theta\left(\frac{\partial^2}{\partial\theta^2} \log f_\theta(\underline{x})\right)$$

$$0 = E\left(\frac{\partial}{\partial\theta} \log f_\theta(x_1)\right) = \int \frac{\partial}{\partial\theta} \log f_\theta(x) \cdot f_\theta(x)\, dx$$

$$= \int \frac{\partial}{\partial\theta} f_\theta(x)\, dx$$

در اینجا فرض کردیم که θ مستقل است.

$$0 = \int \frac{\partial}{\partial\theta} f_\theta(x)\, dx = \int \frac{\partial}{\partial\theta} f_\theta(x) \cdot \frac{f_\theta(x)}{f_\theta(x)}\, dx$$

$$\Rightarrow E\left[\frac{\frac{\partial}{\partial\theta} f_\theta(x)}{f_\theta(x)}\right] = 0 \quad (*)$$

از اینجا $\frac{\partial}{\partial\theta} \log f_\theta(x) = \frac{\frac{\partial}{\partial\theta} f_\theta(x)}{f_\theta(x)}$

$$\frac{\partial^2}{\partial\theta^2} \log f_\theta(x) = \frac{\frac{\partial^2}{\partial\theta^2} f_\theta(x) \cdot f_\theta(x) - \left(\frac{\partial}{\partial\theta} f_\theta(x)\right)^2}{f_\theta^2(x)}$$

$$= \frac{\frac{\partial^2}{\partial\theta^2} f_\theta(x)}{f_\theta(x)} - \left(\frac{\frac{\partial}{\partial\theta} f_\theta(x)}{f_\theta(x)}\right)^2$$

$$E\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(\underline{x})\right] = E\left[\frac{\frac{\partial^2}{\partial\theta^2} f_\theta(\underline{x})}{f_\theta(\underline{x})}\right] - E\left[\frac{\partial}{\partial\theta}\log f_\theta(\underline{x})\right]^2$$

$$\Rightarrow E\left[\frac{\partial}{\partial\theta}\log f_\theta(\underline{x})\right]^2 = -E\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(\underline{x})\right]$$

قضیه: اگر $T(\underline{x})$ آماره‌ای برای θ باشد، آنگاه
$$I_T(\theta) = I_{\underline{x}}(\theta)$$

اثبات: $f_\theta(\underline{x}) = g_\theta(t)\cdot h(\underline{x})$

$$\frac{\partial}{\partial\theta}\log f_\theta(\underline{x}) = \frac{\partial}{\partial\theta}\log g_\theta(t) + \frac{\partial}{\partial\theta}\log h(\underline{x})$$

$$= \frac{\partial}{\partial\theta}\log g_\theta(t)$$

$$E\left[\frac{\partial}{\partial\theta}\log f_\theta(\underline{x})\right]^2 = E\left[\frac{\partial}{\partial\theta}\log g_\theta(T)\right]^2$$

$$I_{\underline{x}}(\theta) = I_T(\theta)$$

نتیجه: اگر $T(\underline{x})$ آماره‌ای غیر کفایتی برای θ با چگالی $g_\theta(t)$ باشد، آنگاه
$$I_{\underline{x}}(\theta) \geq I_T(\theta)$$

از اثبات قضیه نتیجه می‌شود که
$$E\left[\frac{\partial}{\partial\theta}\log f_\theta(\underline{x}) \mid T=t\right] = \frac{\partial}{\partial\theta}\log g_\theta(t)$$

① $g_\theta(t) = P_\theta(T(\underline{x}) = t) = \sum_{\{\underline{x} \mid T(\underline{x}) = t\}} f_\theta(\underline{x})$

$\Rightarrow \frac{\partial}{\partial \theta} g_\theta(t) = g'_\theta(t) = \sum_{\{\underline{x} \mid T(\underline{x}) = t\}} f'_\theta(\underline{x})$

② $T = t$ داده شده $f_{\underline{x} \mid T=t}(\underline{x}) = \frac{f_{\underline{x},t}(\underline{x},t)}{g_\theta(t)} = \begin{cases} \frac{f_\theta(\underline{x})}{g_\theta(t)} & T(\underline{x}) = t \\ 0 & T(\underline{x}) \neq t \end{cases}$

$E\left[\frac{\partial}{\partial \theta} \log f_\theta(\underline{x}) \mid T = t\right] = \sum_{\{\underline{x} \mid T(\underline{x}) = t\}} \frac{\partial}{\partial \theta} \log f_\theta(\underline{x}) \cdot f_{\underline{x} \mid T=t}(\underline{x})$

$= \sum_{\{\underline{x} \mid T(\underline{x}) = t\}} \frac{\partial}{\partial \theta} \log f_\theta(\underline{x}) \cdot \frac{f_\theta(\underline{x})}{g_\theta(t)} = \frac{\sum_{\{\underline{x}, T(\underline{x}) = t\}} \frac{\partial}{\partial \theta} f_\theta(\underline{x})}{g_\theta(t)}$

$= \frac{\sum_{\{\underline{x} \mid T(\underline{x}) = t\}} f'_\theta(\underline{x})}{g_\theta(t)} = \frac{g'_\theta(t)}{g_\theta(t)} = \frac{\partial}{\partial \theta} \log g_\theta(t)$ ▲

$\geq E\left[\frac{\partial}{\partial \theta} \log f_\theta(\underline{x}) - \frac{\partial}{\partial \theta} \log g_\theta(T)\right]^2 =$: طرفین تساوی

$= I_{\underline{x}}(\theta) + I_T(\theta) - 2E\left[\frac{\partial}{\partial \theta} \log f_\theta(\underline{x}) \cdot \frac{\partial}{\partial \theta} \log g_\theta(T)\right]$ ①

$E\left[\frac{\partial}{\partial \theta} \log f_\theta(\underline{x}) \cdot \frac{\partial}{\partial \theta} \log g_\theta(T)\right] = E\left\{E\left[\frac{\partial}{\partial \theta} \log f_\theta(\underline{x}) \cdot \frac{\partial}{\partial \theta} \log g_\theta(T) \mid T = t\right]\right\}$

$= E\left\{\frac{\partial}{\partial \theta} \log g_\theta(t) \, E\left[\frac{\partial}{\partial \theta} \log f_\theta(\underline{x}) \mid T = t\right]\right\}$

$E\left[\frac{\partial}{\partial \theta} \log f_\theta(\underline{x}) \mid T = t\right]$: داریم $\frac{\partial}{\partial \theta} \log g_\theta(t)$ نسبت به

$= E\left[\frac{\partial}{\partial \theta} \log g_\theta(T)\right]^2 = I_T(\theta)$ ②

از ① و ② :

$\geq I_{\underline{x}}(\theta) - I_T(\theta) \Rightarrow I_{\underline{x}}(\theta) \geq I_T(\theta)$ ▲

78

قضیه: شرط لازم و کافی برای آنکه آماره T از θ کفایت باشد آن است که $\{f_\theta(\underline{x})\}$ خانواده نمایی باشد.

اثبات: نخست فرض کنیم T کفایت باشد. در این صورت

$$f_{T|\theta}(t) = 1$$

و لذا

$$\log f_\theta(\underline{x}) = \log g_\theta(t) + \log h(\underline{x})$$

حال فرض کنیم

$$\exists a, b \text{ s.t. } V = a(\theta)T + b(\theta)$$

آنگاه

$$\log f_\theta(\underline{x}) = a(\theta)T(\underline{x}) + b(\theta)$$

و لذا

$$\log f_\theta(\underline{x}) = T(\underline{x}) \int a(\theta)\,d\theta + \int b(\theta)\,d\theta$$

$$= T(\underline{x})\,Q(\theta) + D(\theta) + S(\underline{x})$$

و در نتیجه

$$f_\theta(\underline{x}) = \exp\{Q(\theta)T(\underline{x}) + D(\theta) + S(\underline{x})\}$$

یعنی توزیع $\underline{X}$ متعلق به یک خانواده نمایی است.

تذکر: یعنی متغیر تصادفی پیوسته X با توزیع $f_\theta(x) = \exp\{Q(\theta)\int a(\theta)\,d\theta + \int b(\theta)\,d\theta + R(x)\}$ دارای تابع چگالی $f_\theta(x) = \exp\{Q(\theta)T(x) + D(\theta) + S(x)\}$ است.

مثال: فرض کنید $X_1, \dots, X_n$ نمونه‌ای تصادفی از توزیع $U(0,\theta)$ باشد. آماره $T = X_{(n)}$ از θ کفایت است.

حل: تابع چگالی توأم $\underline{X}$ عبارت است از

$$f_\theta(\underline{x}) = \frac{1}{\theta^n} I_{(0,\theta)}(x_{(n)})$$

و لذا

$$\log f_\theta(\underline{x}) = -n\log\theta + \log I_{(0,\theta)}(x_{(n)})$$

$$\frac{\partial}{\partial\theta}\log f_\theta(x) = -1 + \frac{x}{\theta} \rightarrow \frac{\partial^2}{\partial\theta^2}\log f_\theta(x) = -\frac{x}{\theta^2}$$

$$I_X(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(X)\right] = -E\left[\frac{-X}{\theta^2}\right] = \frac{EX}{\theta^2} = \frac{\theta}{\theta^2} = \frac{1}{\theta}$$

$$I_{\underline{X}}(\theta) = n\, I_X(\theta) = \frac{n}{\theta}$$

$\bar{X}$ آماره بسنده و کامل برای θ است، چون $E\bar{X} = \theta$ و $T(\underline{X}) = \bar{X}$ UMVUE برای $\tau(\theta) = \theta$ است.

$$Var(T(\underline{X})) = Var(\bar{X}) = \frac{\theta}{n} \quad (1)$$

کران پایین کرامر-رائو برای $T(\underline{X}) = \bar{X}$ برابر است با

$$\frac{(\tau'(\theta) + b'(\theta))^2}{I_{\underline{X}}(\theta)} = \frac{1}{n/\theta} = \frac{\theta}{n} \quad (2)$$

از (1) و (2) نتیجه می‌شود $\bar{X}$ UMVUE است.

ب- با توجه به $Y = \sum_{i=1}^{n} X_i \sim P(n\theta)$ و با فرض Y کفایی بسنده برای θ است. حال فرض کنید

$$T(\underline{X}) = \begin{cases} 1 & Y = 0 \\ 0 & Y \neq 0 \end{cases}$$

$T(\underline{X})$ نیز آماره کفایی بسنده است، چون $E(T(\underline{X})) = e^{-n\theta}$ و $T(\underline{X})$ تابعی از Y است.

$$Var(T(\underline{X})) = ET^2 - E^2T$$
$$= 1\times P(Y=0) - (1\times P(Y=0))^2 = e^{-n\theta} - e^{-2n\theta}$$
$$= e^{-n\theta}(1 - e^{-n\theta}) \quad (3)$$

کران پایین کرامر-رائو برابر است با

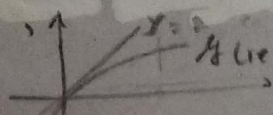
$$C.R.L.B = \frac{(\tau'(\theta) + b'(\theta))^2}{I_{\underline{X}}(\theta)} = \frac{n^2 e^{-2n\theta}}{n/\theta} = n\theta e^{-2n\theta} \quad (4)$$

از آنجا که (3) > (4)

$$e^{-n\theta}(1 - e^{-n\theta}) > n\theta e^{-2n\theta}$$

$$1 - e^{-n\theta} > n\theta e^{-n\theta} \iff 1 > e^{-n\theta}(1 + n\theta)$$

$$\iff e^{n\theta} > 1 + n\theta \xrightarrow{\text{قرار دهید } n\theta = x} e^x > 1 + x \Rightarrow x > 0$$



Efficiency

کارایی برآوردگرهای نااریب از دیدگاه واریانس:

تعریف: اگر δ_1 و δ_2 دو برآوردگر نااریب برای θ باشند کارایی نسبی δ_2 نسبت به δ_1 را با نماد $e(\delta_1|\delta_2)$ تعریف کرده و به صورت زیر تعریف می‌کنیم:

$$e_\theta(\delta_1|\delta_2) = \frac{Var(\delta_2)}{Var(\delta_1)} \begin{cases} >1 & \delta_1 \text{ کاراتر از } \delta_2 \\ =1 & \text{کارایی یکسان} \\ <1 & \delta_2 \text{ کاراتر از } \delta_1 \end{cases}$$

تعریف: برآوردگر نااریب δ^* را از θ کارآترین گوییم.

$$Var_\theta(\delta^*) = \frac{[\psi'(\theta)]^2}{I_X(\theta)} \quad \forall \theta \in \Theta$$

استثنا: توزیع‌هایی که در آن‌ها برآوردگر نااریب با واریانس کمتر از کران پایین فرچت-کرامر-رائو (به شکل منظم) وجود دارد. در این حالت برآوردگر نااریب کارایی یکنواخت (UMVUE) است.

برآوردگر δ^* را کارآترین برآوردگر نااریب $\psi(\theta)$ گوییم هرگاه:

$$\forall \theta \in \Theta, \forall \delta \in U: \quad Var(\delta^*) \le Var(\delta)$$

تعریف: دنباله‌ای از برآوردگرهای $\{\delta_n\}$ را برای $\psi(\theta)$ مجانبی کارآ گوییم هرگاه:

$$\lim_{n\to\infty} e(\delta_n|\delta^*) = 1$$

که در آن δ^* برآوردگر کارا از $\psi(\theta)$ است.

مثال: اگر $X_1, \dots, X_n \sim P(\theta)$ و $\delta_n(\underline{x}) = (1-\frac{1}{n})^{\sum x_i}$ نشان دهید که δ_n برای $\psi(\theta) = e^{-\theta}$ مجانبی کارآ است.

حل: ابتدا نشان می‌دهیم که $E(\delta_n(\underline{X})) = e^{-\theta}$

$$P(X_1 = 0) = e^{-\theta} \qquad g(T) = P(X_1 = 0|T) = \left(1-\frac{1}{n}\right)^T \quad T = \sum_{i=1}^n X_i$$

سوال ۱

$$E(\delta_n(\underline{x})) = \sum_{y_2=0}^{\infty} \left(\frac{n-1}{n}\right)^{y_2} \frac{(n\theta)^{y_2} e^{-n\theta}}{y_2!}$$

$$= \sum_{y_2=0}^{\infty} \frac{((n-1)\theta)^{y_2} e^{-n\theta}}{y_2!} = \sum_{y_2=0}^{\infty} \frac{[(n-1)\theta]^{y_2} e^{-n\theta}}{y_2!} \cdot \frac{e^{-(n-1)\theta}}{e^{-(n-1)\theta}}$$

$$= e^{-\theta} \sum_{y_2=0}^{\infty} \frac{((n-1)\theta)^{y_2} e^{-(n-1)\theta}}{y_2!} = e^{-\theta}$$

این یک برآوردگر نااریب برای θ است.

$$Var(\delta_n(\underline{x})) = E(\delta_n^2(x)) - E^2(\delta_n(x))$$

$$E(\delta^2(x)) = E\left[\left(\frac{n-1}{n}\right)^{y_2}\right]^2 = \sum_{y_2=0}^{\infty} \left(\frac{n-1}{n}\right)^{2y_2} \cdot \frac{e^{-n\theta}(n\theta)^{y_2}}{y_2!}$$

$$= \sum_{y_2=0}^{\infty} \left(\frac{(n-1)^2\theta}{n}\right)^{y_2} \cdot \frac{e^{-n\theta}}{y_2!} = \sum_{y_2=0}^{\infty} \frac{\left[\frac{(n-1)^2\theta}{n}\right]^{y_2}}{y_2!} \cdot \frac{e^{-\frac{(n-1)^2}{n}\theta}}{e^{-\frac{(n-1)^2}{n}\theta}} \cdot e^{-n\theta}$$

$$= e^{-2\theta + \frac{\theta}{n}}$$

$$Var(\delta(\underline{x})) = e^{-2\theta + \theta/n} - e^{-2\theta} = e^{-2\theta}\left(e^{\theta/n} - 1\right)$$

$$e(\delta_n) = \frac{(\psi'(\theta))^2}{I_x(\theta)\, Var(\delta_n)} = \frac{e^{-2\theta}}{n/\theta} \cdot \frac{\theta}{e^{-2\theta}(e^{\theta/n} - 1)} = \frac{\frac{\theta}{n}}{(e^{\theta/n} - 1)} < 1$$

چون

$$\frac{\theta}{n} = x \qquad = \frac{x}{e^x - 1} < 1$$

$$e^x > 1 + x$$
$$e^x - 1 > x$$

پس این برآوردگر برای θ کارایی مجانبی دارد.

$$\lim_{n\to\infty} e(\delta_n) = \lim_{n\to\infty} \frac{\theta/n}{e^{\theta/n} - 1} = \lim_{n\to\infty} \frac{-\theta/n^2}{-\frac{\theta}{n^2} e^{\theta/n}} = 1$$

قضیه حد مرکزی

۹۹