Module-5

Asymptotically Efficient Estimators

Consistent Estimators

Let $X_1, \ldots, X_n$ be a ~~random~~ sample with $\underline{X} = (X_1, \ldots, X_n)$ having d.f. $F_{\underline{\theta}}(\cdot)$ where $\underline{\theta} \in \Theta \subseteq \mathbb{R}^k$ and for each fixed $\underline{\theta} \in \Theta$ the functional form of $F_{\underline{\theta}}$ is known. Let $f_{\underline{\theta}}(\cdot)$ be the pmf/pdf corresponding to $F_{\underline{\theta}}$, $\underline{\theta} \in \Theta$. Let $g: \Theta \to \mathbb{R}^m$ be a given estimand. Based on the information contained in ~~random~~ sample $\underline{X}$ it is desired to estimate $g(\underline{\theta})$. Suppose one decides to use the estimator $\delta_n \equiv \delta_n(\underline{X})$, $n = 1, 2, \ldots$. As the sample sizes increases we have more and more information about $\underline{\theta}$ (and hence about estimand $g(\underline{\theta})$). For $n \to \infty$ the minimal one would expect is that δ_n converges to $g(\underline{\theta}_0)$ $= (g_1(\underline{\theta}_0), \ldots, g_m(\underline{\theta}_0))$ in some stochastic sense. Thus one may be interested in stochastic convergence behavior of ~~the estimand $g(\underline{\theta}) = (g_1(\underline{\theta}), \ldots, g_m(\underline{\theta}))$~~ of the sequence of estimators $\{\delta_n\}_{n \geq 1}$, as $n \to \infty$; here $\delta_n = (\delta_{1,n}, \ldots, \delta_{m,n})$.

For $\underline{x} = (x_1, \ldots, x_m) \in \mathbb{R}^m$ and $\underline{y} = (y_1, \ldots, y_m) \in \mathbb{R}^m$ let

$$\|\underline{x} - \underline{y}\| = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2}$$

denote the Euclidean distance between points $\underline{x}$ and $\underline{y}$.

Definition The sequence $\{\delta_n\}_{n \geq 1}$ of estimators is said to be consistent for $g(\underline{\theta})$ if $\delta_n \xrightarrow{P} g(\underline{\theta})$, as $n \to \infty$, or equivalently, if for every $\varepsilon > 0$

$$\lim_{n \to \infty} P_{\theta}(\|\delta_n(\underline{X}) - g(\underline{\theta})\| \geq \varepsilon) = 0, \quad \forall\, \underline{\theta} \in \Theta.$$

Remark: (i) $\{\delta_n\}_{n \geq 1}$ is consistent for $g(\underline{\theta}) = (g_1(\underline{\theta}), \ldots, g_m(\underline{\theta}))$ iff $\{\delta_{i,n}\}_{n \geq 1}$ is consistent for $g_i(\underline{\theta})$ for every $i = 1, \ldots, m$.

(ii) $\{\delta_n\}_{n \geq 1}$ is consistent for $g(\underline{\theta})$ and $h: \mathbb{R}^m \to \mathbb{R}^k$ is a continuous function then $\{h(\delta_n)\}_{n \geq 1}$ is consistent for $h(g(\underline{\theta}))$

(iii) If

$$\lim_{n\to\infty} \sum_{i=1}^{m} E_{\underline{\theta}}\left((\delta_{i,n} - g_i(\underline{\theta}))^2 \right) = 0, \quad \forall\, \underline{\theta} \in \Theta,$$

then $\{\delta_n\}_{n\geq 1}$ is consistent for $g(\underline{\theta})$. In particular if $\forall\, \underline{\theta} \in \Theta$,

$\lim_{n\to\infty} E_\theta(\delta_{i,n}(\underline{X})) = g_i(\underline{\theta})$, $\forall\, i=1,\dots,m$ and $\lim_{n\to\infty} Var_\theta(\delta_{i,n}(\underline{X})) = 0$, $\forall\, i=1,\dots,m$,

then $\{\delta_n\}_{n\geq 1}$ is consistent for $g(\underline{\theta})$.

(IV) If $\{\delta_n\}_{n\geq 1}$ is consistent for $g(\underline{\theta})$ and $\{a_n\}_{n\geq 1}$ is a sequence of real numbers such that $a_n \to 1$ as $n\to\infty$. Then $\{a_n \delta_n\}_{n\geq 1}$ is also consistent for $g(\underline{\theta})$.

Some results from asymptotic statistics.

(1) Let $X_1, \dots, X_n$ be a random sample from a distribution. If $E_\theta(X_1) = \mu \in (-\infty, \infty)$, $\forall \theta$, then

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mu. \quad \text{(Strong Law of Large Numbers)}$$

In addition if $Var_\theta(X_1) = \sigma^2 \in (0, \infty)$, $\forall \theta$, then

$$\sqrt{n}\,(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2). \quad \text{(Central Limit Theorem)}$$

(2) Let $\underline{Y}_n = (Y_{1,n}, \dots, Y_{m,n})$, $n=1,2,\dots$, be a sequence of i.i.d. r.v.s with common mean $E_{\underline{\theta}}(\underline{Y}_1) = \mu \in \mathbb{R}^m$ and common covariance matrix Σ (a positive definite matrix). Then

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} \underline{Y}_i \xrightarrow{a.s.} \underline{\mu}, \quad \text{as } n\to\infty$$

In fact

$$\sqrt{n}\,(\bar{Y}_n - \underline{\mu}) \xrightarrow{d} N_m(\underline{0}, \Sigma), \quad \text{as } n\to\infty$$

(3) Let $\underline{Z}_n = (Z_{1,n}, \dots, Z_{m,n})$, $n=1,2,\dots$, be a sequence of r.v.s such that for some $\mu \in \mathbb{R}^m$ $\sqrt{n}\,(\underline{Z}_n - \mu) \xrightarrow{d} \underline{Z}$, as $n\to\infty$. Let $\underline{h}: \mathbb{R}^m \to \mathbb{R}^k$ be a given function with continuous first order partial derivatives in a neighborhood of μ. Define $\nabla h(\underline{\mu}) = \left(\left(\frac{\partial h_i(\underline{\mu})}{\partial \mu_j}\right)\right)_{m\times k}$.

Then

$$\sqrt{n}\,(h(\underline{Z}_n) - h(\underline{\mu})) \xrightarrow{d} (\nabla h(\underline{\mu}))' \underline{Z}.$$

(4) Let $\{Z_n\}_{n\geq 1}$ be a sequence of r.v.s such that, for some $\mu \in \mathbb{R}$, $\sqrt{n}(Z_n-\mu) \xrightarrow{d} Z$, as $n\to\infty$.

(a) Suppose that $h: \mathbb{R}\to\mathbb{R}$ is a differentiable function with $h'(\mu)\neq 0$. Let $C_n = 1+\frac{a}{n}+o(\frac{1}{n^2})$, $n=1,2,3,\ldots$ for some real constant a. Then

$$\sqrt{n}\left(h(C_n Z_n)-h(\mu)\right) \xrightarrow{d} h'(\mu)\, Z.$$

(b) If $h'(\mu)=0$ and $h''(\mu)\neq 0$ then

$$\sqrt{n}\left(h(C_n Z_n)-h(\mu)\right) \xrightarrow{d} \frac{h''(\mu)}{2} Z^2.$$

Assignment Problems

(1) Let $X_1, X_2, \ldots$ be a sequence of i.i.d. r.v.s with $E_{\theta}(X_1)=\mu$ $\forall\, \theta\in\Theta$ and $Var_{\theta}(X_1)=\sigma^2$, $\forall\, \theta\in\Theta$. Show that $\bar{X}_n$ is a consistent estimator of μ. and $\hat{S}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$ is a consistent estimator of σ^2. In addition show that

$$\sqrt{n}(\bar{X}_n-\mu) \xrightarrow{d} N(0,\sigma^2)$$

and

$$\sqrt{n}(S^2-\sigma^2) \xrightarrow{d} N\left(0, (D-1)\sigma^4\right),$$

where D is the kurtosis of the distribution. Also find the asymptotic distribution of $\sqrt{n}\begin{pmatrix}\bar{X}-\mu\\ S^2-\sigma^2\end{pmatrix}$.

(2) Let $\begin{pmatrix}X_i\\ Y_i\end{pmatrix}$, $i=1,2,\ldots$ be a sequence of i.i.d. r.v.s with $E(X_1)=\mu_1$, $E(Y_1)=\mu_2$, $Var(X_1)=\sigma_1^2$, $Var(Y_1)=\sigma_2^2$, $Corr(X_1,Y_1)=\rho$, $E(X_1^4)<\infty$, and $Var(X_1^4)<\infty$.

Let $S_1^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$, $S_2^2=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})^2$, $S_{12}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})$ and $r=\frac{S_{12}}{S_1S_2}$.

Show that

(a) $$\sqrt{n}\left(\begin{pmatrix}S_1^2\\ S_2^2\\ S_{12}\end{pmatrix}-\begin{pmatrix}\sigma_1^2\\ \sigma_2^2\\ \sigma_{12}\end{pmatrix}\right)\to N_3(0,\Sigma),$$

where, for $T_1 = X_1 - \mu_1$ and $U_1 = Y_1 - \mu_2$

$$\Sigma = \begin{pmatrix} Cov(T_1^2, T_1^2) & Cov(T_1^2, U_1^2) & Cov(T_1^2, T_1U_1) \\ Cov(T_1^2, U_1^2) & Cov(U_1^2, U_1^2) & Cov(U_1^2, T_1U_1) \\ Cov(T_1U_1, T_1^2) & Cov(T_1U_1, U_1^2) & Cov(T_1U_1, T_1U_1) \end{pmatrix}$$

(b) Show that

$$\sqrt{n}\,(r-\rho) \xrightarrow{d} N(0, (1-\rho^2)^2), \text{ as } n\to\infty$$

$$\sqrt{n}\left(\frac{1}{2}\ln\frac{1+r}{1-r} - \frac{1}{2}\ln\frac{1+\rho}{1-\rho}\right) \xrightarrow{d} N(0,1) \text{ as } n\to\infty,$$

(Variance Stabilizing Constant)

(c) r is consistent estimator of ρ

(3) Let $X_1, X_2, \ldots$ be a sequence of i.i.d. r.v.s with finite mean μ. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, $n = 1, 2, \ldots$. Define

$$T_n = \begin{cases} 0 & n = 1, 2, \ldots, 10^{10} \\ \bar{X}_n & n \geq 10^{10}+1 \end{cases}, \quad n = 1, 2, \ldots$$

Show that $\{T_n\}_{n\geq 1}$ is consistent for μ. (Note: Use this example to infer that consistency is a minimal requirement in asymptotic statistics)

(4) Let $X_1, X_2, \ldots$ be a sequence of i.i.d. $U(0, \theta)$, where $\theta \in \Theta = (0, \infty)$. Let $X_{(n)} = \max\{X_1, \ldots, X_n\}$, $n = 1, 2, \ldots$ Show that $n(\theta - X_{(n)}) \xrightarrow{d} Z \sim Exp(1)$, as $n\to\infty$. Hence conclude that $X_{(n)}$ is a consistent estimator of θ.

(5) Let $X_1, X_2, \ldots$ be a sequence of i.i.d. r.v.s with common p.m.f./p.d.f. $f_\theta(x) = e^{\theta T(x) - \psi(\theta)}$, $x \in \chi \subseteq \mathbb{R}$, $\theta \in \Theta \subseteq \mathbb{R}$, where Θ is an open set. Show that $\frac{1}{n}\sum_{i=1}^{n} T(X_i)$ is a consistent estimator of $\psi'(\theta)$.

$\boxed{4/5}$

Also show that

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} T(X_i) - \psi'(\theta)\right) \xrightarrow{d} N(0, \psi''(\theta))$$

$\boxed{///}$

Example (Consistent Estimators May not exist) Let $X_1, \ldots, X_n$ be iid $N(\theta+\mu, \sigma^2)$ where $\mu, \theta \in \Theta = \mathbb{R}$ and $\sigma > 0$ is known.

Then $\bar{X}_n$ is consistent for $\theta+\mu$. However consistent estimators of θ (or μ) do not exist as if $\delta_n(\underline{X})$ is a consistent estimator for θ. Then, by symmetry δ_n is also consistent estimator of μ. This implies that $\theta = \mu$, which is not true ($X_n \xrightarrow{P} c$, $X_n \xrightarrow{d} d \Rightarrow c = d$)

Definition Let X be distributed according to d.f. $F_{\theta,\mu}$. If there exist pairs (θ_1, μ_1) and (θ_2, μ_2) with $\theta_1 \neq \theta_2$ for which $P_{\theta_1,\mu_1} = P_{\theta_2,\mu_2}$, the parameter θ is said to be unidentifiable.

Notes: Unidentifiable parameters can not be estimated consistently.

Consistency only reveals that, for large n, $\|\delta_n - g(\theta)\|$ is likely to be small and it does not tell us about the order of error (e.g. $\frac{1}{n}$, $\frac{1}{\sqrt{n}}$, $\frac{1}{\ln n}$, etc.). For an estimator δ_n let

$$R_{\delta_n}(\underline{\theta}) = E_\theta(\|\delta_n - g(\underline{\theta})\|^2), \quad \underline{\theta} \in \Theta$$

denotes its mean squared error. For any consistent estimator δ_n, generally

$$\lim_{n\to\infty} R_{\delta_n}(\underline{\theta}) = 0, \quad \forall\, \underline{\theta} \in \Theta;$$

which only tells us that, for large n the risk in estimating $g(\underline{\theta})$ by $\delta_n(\underline{\theta})$ is very small. However for comparison of two consistent estimators one would be interested in knowing at what rates the risks of two estimators are going to zero as $n\to\infty$.

For large sample sizes, performance evaluation of consistent estimators can be done using one of the following two approaches:

(I) The limiting risk approach (or the limiting variance approach)

(II) The asymptotic distribution approach. (or the asymptotic variance approach)

The Limiting Risk Approach

For an estimator $\delta_n(\underline{X})$, let

$$\equiv \quad R_{\delta_n}(\underline{\theta}) = E_{\underline{\theta}}(\|\delta_n(\underline{X}) - g(\underline{\theta})\|^2), \quad \underline{\theta} \in \Theta, \quad n = 1, 2, \ldots$$

For any consistent estimator δ_n, generally

$$\lim_{n\to\infty} R_{\delta_n}(\theta) = 0, \quad \forall\, \theta \in \Theta$$

Suppose that $\{\delta_n\}_{n\geq 1}$ is consistent for $g(\underline{\theta})$, and

$$\lim_{n\to\infty} R_{\delta_n}(\theta) = 0, \quad \forall\, \theta \in \Theta \quad \ldots\ldots (A)$$

Let $\{k_n\}_{n\geq 1}$ be a sequence of real numbers. Define

$$R^*_{\delta_n}(\underline{\theta}) = k_n R_{\delta_n}(\underline{\theta}), \quad \underline{\theta} \in \Theta, \quad n = 1, 2, \ldots$$

Consider the following two extreme situations.

Case I: $\{k_n\}_{n\geq 1}$ is bounded.

In this case, in view of (I),

$$\lim_{n\to\infty} R^*_{\delta_n}(\underline{\theta}) = 0, \quad \forall\, \underline{\theta} \in \Theta$$

Case II. $k_n \to \infty$ sufficiently fast

In this case we may have

$$\lim_{n\to\infty} R^*_{\delta_n}(\underline{\theta}) = \infty, \quad \forall\, \underline{\theta} \in \Theta.$$

The above two cases, being extreme, suggest that there might exist an intermediate sequence $\{k_n\}_{n\geq 1}$ with $k_n \to \infty$ (as $n \to \infty$) and

$$0 < \lim_{n\to\infty} R^*_{\delta_n}(\underline{\theta}) < \infty, \qquad \forall\, \underline{\theta} \in \Theta \quad \cdots \text{(B)}$$

Commonly there will exist a sequence $\{k_n\}_{n\geq 1}$ such that (B) holds. We shall then say that the risk $R_n(\theta)$ tends to zero at rate $1/k_n$. Note that the error of the rate is not uniquely determined. If $1/k_n$ is a possible rate then so is $1/k'_n$ for any sequence $\{k'_n\}$ such that $k'_n/k_n \to c \in (0,\infty)$.

Definition A sequence of estimators $\{\delta_n\}_{n\geq 1}$ of $g(\underline{\theta})$ is said to be unbiased in limit if

$$\lim_{n\to\infty} E_{\underline{\theta}}(\delta_n(\underline{X})) = g(\underline{\theta}), \qquad \forall\, \underline{\theta} \in \Theta.$$

Any consistent estimator is generally unbiased in limit and satisfies

$$\lim_{n\to\infty} R_{\delta_n}(\underline{\theta}) = 0, \qquad \forall\, \underline{\theta} \in \Theta.$$

In that case

$$\lim_{n\to\infty} R_{\delta_n}(\underline{\theta}) = \lim_{n\to\infty} V_{\delta_n}(\underline{\theta}) = \lim_{n\to\infty} \sum_{i=1}^{m} Var_{\theta}(\delta_{in}) = 0,$$

and the limiting risk approach is the same as limiting variance approach.

Remark: (B) $\Rightarrow$ $\lim_{n\to\infty} k_{\delta_n}(\theta) > 0 \;\; \forall\, \theta \in \Theta \Rightarrow \{\delta_n\}_{n\geq 1}$ is consistent for $g(\theta)$.

Definition Let $\{\delta_n\}_{n\geq 1}$ be a sequence of estimators. Suppose that $\{k_n\}_{n\geq 1}$ is a sequence of real numbers such that $k_n \to \infty$, as $n \to \infty$, and

$$\lim_{n\to\infty} R^*_{\delta_n}(\theta) = \lim_{n\to\infty} (k_n R_{\delta_n}(\theta)) = \gamma^2 \in (0,\infty).$$

Then γ^2 is called the limiting risk or limit of the risks.

Note: (i) If $\{\delta_n\}_{n\geq 1}$ is consistent, generally limiting risk (or limit of risks) is the limiting variance (or limit of the variances)

(ii) If $\{\delta_n\}_{n\geq 1}$ is a sequence of estimators with error rate $\frac{1}{k_n}$ and k'_n tends to ∞ more slowly (or faster) than k_n (i.e. $\frac{k'_n}{k_n} \to 0$ (or ∞)), then $\lim_{n\to\infty} k'_n R_{\delta_n}(\theta) = 0$ (or ∞).

Definition Suppose that $\{\delta_n^{(1)}\}$ and $\{\delta_n^{(2)}\}$ are two sequences of estimators of $g(\theta)$ such that for some $\alpha > 0$ and any sequence $n' \equiv n'(n)$ ($n' \to \infty$, as $n \to \infty$)

$$\lim_{n\to\infty} n^{\alpha} R_{\delta_n^{(1)}}(\theta) = \lim_{n\to\infty} n^{\alpha} R_{\delta_{n'}^{(2)}}(\theta) = \tau^2 \in (0,\infty), \ \forall \theta \in \Theta.$$

$\tau^2 \equiv \tau^2(\theta)$

Then the limiting risk efficiency (LRE) of $\{\delta_n^{(1)}\}_{n\geq 1}$ relative to $\{\delta_n^{(2)}\}_{n\geq 1}$ is defined by

$$l_{\delta^{(1)},\delta^{(2)}}(\theta) = \lim_{n\to\infty} \left(\frac{n'(n)}{n} \right), \quad \theta \in \Theta,$$

provided the limit exists and is independent of the particular sequence $\{n'(n)\}_{n\geq 1}$ chosen.

Example Let $X_1, \dots, X_n$ be i.i.d. $N(\theta, 1)$, where $\theta \in \Theta = \mathbb{R}$ is unknown. Let $[n]$ denote largest integer not exceeding n. Define $\delta_n^{(1)} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\delta_n^{(2)} = \frac{1}{[\frac{n}{2}]}\sum_{i=1}^{[n/2]} X_i$. Then

$$R_{\delta_n^{(1)}}(\theta) = \frac{1}{n} \quad \text{and} \quad R_{\delta_n^{(2)}}(\theta) = \frac{1}{[\frac{n}{2}]}, \ \forall \theta \in \Theta.$$

Then

$$\lim_{n\to\infty} n R_{\delta_n^{(1)}}(\theta) = \lim_{n\to\infty} n R_{\delta_{2n}^{(2)}}(\theta) = 1, \quad \forall \theta \in \Theta.$$

and therefore

$$l_{\delta^{(1)},\delta^{(2)}}(\theta) = \lim_{n\to\infty} \left[\frac{n'(n)}{n} \right] = \lim_{n\to\infty} \left(\frac{2n}{2} \right) = 2,$$

with $n'(n) = 2n$, $n = 1, 2, \dots$. This suggests that to obtain the same limiting risk $\delta_n^{(1)}$ requires half as many observations required by $\delta_n^{(2)}$.

Note: If the normalizing factor is not of the form n^{α}, $\alpha > 0$, then the ratio of sample sizes should not be used to measure LRE.

Example Suppose that

$$\lim_{n\to\infty} \ln n\, R_{\delta_n}(\theta) = \gamma^2 \equiv \gamma^2(\theta) \in (0, \infty).$$

Then, for any positive integer m,

$$\lim_{n\to\infty} \ln n\, R_{\delta_{mn}}(\theta) = \gamma^2$$

and the ratio

$$\lim_{n\to\infty} \frac{mn}{n} = m$$

Can not be used to measure LRE. of δ_n against itself.

Theorem Suppose that, for some $\alpha > 0$,

$$\lim_{n\to\infty} n^\alpha R_{\delta_n^{(i)}}(\theta) = \tau_i^2 \in (0, \infty), \quad i = 1, 2.$$

Then the LRE of $\{\delta_n^{(1)}\}_{n\geq 1}$ relative to $\{\delta_n^{(2)}\}_{n\geq 1}$ is

$$l_{\delta^{(1)}, \delta^{(2)}}(\theta) = \left(\frac{\gamma_2^2}{\gamma_1^2}\right)^{\frac{1}{\alpha}}.$$

Proof. Let $n' \equiv n'(n)$ be such that for $\lim_{n\to\infty} n^\alpha R_{\delta_n^{(1)}}(\theta) = \lim_{n\to\infty} n^\alpha R_{\delta_{n'}^{(2)}}(\theta) = \gamma^2 \in (0, \infty)$. (one such seq. $n' \equiv n$ e.g.)

Then

$$n^\alpha R_{\delta_{n'}^{(2)}}(\theta) = \left(\frac{n}{n'}\right)^\alpha (n')^\alpha R(\theta, \delta_{2n'}^{(2)})$$

$$\lim_{n\to\infty} [n^\alpha R_{\delta_{n'}^{(2)}}(\theta)] = \lim_{n\to\infty} \left[\left(\frac{n}{n'}\right)^\alpha \gamma_2^2\right]$$

$$\Rightarrow \lim_{n\to\infty} [n^\alpha R_{\delta_n^{(1)}}(\theta)] = \lim_{n\to\infty} \left(\frac{n}{n'}\right)^\alpha \gamma_2^2$$

$$\Rightarrow \quad \gamma_1^2 = \lim_{n\to\infty} \left(\frac{n}{n'}\right)^\alpha \gamma_2^2$$

$$\Rightarrow \quad l_{\delta^{(1)}, \delta^{(2)}}(\theta) = \lim_{n\to\infty} \left(\frac{n'(n)}{n}\right) = \left(\frac{\gamma_2^2}{\gamma_1^2}\right)^{1/\alpha}.$$

Example Let $X \sim Bin(n, \theta)$, $\theta \in \Theta = (0, 1)$. Consider the estimand $g(\theta) = \theta$. Then

$$\delta_n^{(1)} = \frac{X}{n} \text{ is the UMVUE}$$

$$\delta_n^{(2)} = \frac{X + \sqrt{n}/2}{n + \sqrt{n}} \text{ is the minimax estimator}$$

$$R_{\delta_n^{(1)}}(\theta) = E_\theta\left(\left(\frac{X}{n} - \theta\right)^2\right) = \frac{\theta(1-\theta)}{n}, \quad \theta \in (0,1)$$

$$R_{\delta_n^{(2)}}(\theta) = E_\theta\left(\left(\frac{X + \sqrt{n}/2}{n+\sqrt{n}} - \theta\right)^2\right) = \frac{n}{4(n+\sqrt{n})^2} = \frac{1}{4(\sqrt{n}+1)^2}$$

$$\lim_{n\to\infty} n R_{\delta_n^{(1)}}(\theta) = \theta(1-\theta), \quad \theta \in \Theta$$

$$\lim_{n\to\infty} n R_{\delta_n^{(2)}}(\theta) = \lim_{n\to\infty} \frac{n}{4(\sqrt{n}+1)^2} = \frac{1}{4}.$$

$$l_{\delta_n^{(2)}, \delta_n^{(1)}}(\theta) = 4\theta(1-\theta) \le 1, \quad \forall\, \theta \in \Theta$$

$$l_{\delta_n^{(2)}, \delta_n^{(1)}}(1/2) = 1.$$

Theorem Let $X_1, X_2, \ldots, X_n$ be iid rvs.

(a) Suppose that X_1 has finite first four moments with $E_\theta(X_1) = \mu$, $Var_\theta(X_1) = \sigma^2$, $E_\theta((X_1-\mu)^3) = \mu_3$ and $E((X_1-\mu)^4) = \mu_4$. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Then, for each θ,

(i) $E_\theta(\bar{X}) = \mu$

(ii) $E_\theta((\bar{X}-\mu)^2) = Var_\theta(\bar{X}) = \frac{\sigma^2}{n}$

(iii) $E_\theta((\bar{X}-\mu)^3) = \frac{\mu_3}{n^2} = O\left(\frac{1}{n^2}\right)$

(iv) $E_\theta((\bar{X}-\mu)^4) = \frac{\mu_4}{n^3} + \frac{3(n-1)}{n^3}\sigma^4 = O\left(\frac{1}{n^2}\right)$

(b) Suppose that X_1 has finite first $2k$ moments (where k is a positive integer, $k \ge 2$). Then

(i) $E_\theta((\bar{X}-\mu)^{2k-1}) = O\left(\frac{1}{n^k}\right)$

(ii) $E_\theta((\bar{X}-\mu)^{2k}) = O\left(\frac{1}{n^k}\right)$.

Proof. (a) Proofs of (i) and (ii) are obvious.

$$E_\theta((\bar{X}-\mu)^3) = \frac{1}{n^3} E_\theta\left(\left(\sum_{i=1}^n (X_i-\mu)\right)^3\right)$$

$$= \frac{1}{n^3}\Big[E_\theta\Big(\sum_{i=1}^n (X_i-\mu)^3 + 3\sum_{i=1}^n \sum_{\substack{j=1\\ i\neq j}}^n (X_i-\mu)^2(X_j-\mu)$$

$$+ \sum_{i=1}^n \sum_{j=1}^n \sum_{\substack{k=1\\ i\neq j\neq k}}^n (X_i-\mu)(X_j-\mu)(X_k-\mu)\Big)\Big]$$

$$= \frac{n\mu_3}{n^3} = \frac{\mu_3}{n^2} = O\left(\frac{1}{n^2}\right)$$

$$E_\theta((\bar{X}-\mu)^4) = \frac{1}{n^4} E_\theta\left(\left(\sum_{i=1}^n (X_i-\mu)\right)^4\right)$$

$$= \frac{1}{n^4} E_\theta\Big[\sum_{i=1}^n (X_i-\mu)^4 + 4\sum_{i=1}^n \sum_{\substack{j=1\\ i\neq j}}^n (X_i-\mu)^3(X_j-\mu)$$

$$+ 6\sum_{i=1}^n \sum_{\substack{j=1\\ i\neq j}}^n (X_i-\mu)^2(X_j-\mu)^2 + 0\Big]$$

$$= \frac{\mu_4}{n^3} + \frac{6n(n-1)\sigma^4}{n^4} = O\left(\frac{1}{n^2}\right)$$

(b)
$$E((\bar{X}-\mu)^{2k}) = \frac{1}{n^{2k}} E\left(\left(\sum_{i=1}^n (X_i-\mu)\right)^{2k}\right)$$

$$= \frac{1}{n^{2k}} E\Big[\sum_{i=1}^n (X_i-\mu)^{2k} + 2k\sum_{i=1}^n \sum_{\substack{j=1\\ i\neq j}}^n (X_i-\mu)^{2k-1}(X_j-\mu)$$

$$+ \binom{2k}{2}\sum_{i=1}^n \sum_{\substack{j=1\\ i\neq j}}^n (X_i-\mu)^{2k-2}(X_j-\mu)^2$$

$$+ \binom{2k}{2}\binom{2k-2}{1}\binom{2k-3}{1}\sum_{i=1}^n \sum_{j=1}^n \sum_{\substack{k=1\\ i\neq j\neq k}}^n (X_i-\mu)^{2k-2}(X_j-\mu)(X_k-\mu)$$

$$+ \cdots + \binom{2k}{2}\binom{2k-2}{2}\cdots\binom{2}{2}\sum_{i_1=1}^n \sum_{i_2=1}^n \cdots \sum_{\substack{i_k=1\\ i_1\neq i_2\neq\cdots\neq i_k}}^n (X_{i_1}-\mu)^2(X_{i_2}-\mu)^2\cdots(X_{i_k}-\mu)^2\Big]$$

$$= \frac{n\mu_{2k}}{n^{2k}} + \frac{2k(2k-1)n(n-1)}{n^{2k}}\mu_{2k-2}\,\mu_2 + \cdots + \frac{\lfloor 2k}{2^k}\,\frac{n(n-1)\cdots(n-k+1)}{n^{2k}}\sigma^{2k}$$

$$= O\left(\frac{1}{n^k}\right)$$

$\boxed{\text{///}}$

(c) $E\left((\bar{X}-\mu)^{2k-1}\right) = \frac{1}{n^{2k-1}} E\left[\left(\sum_{i=1}^{n}(X_i-\mu)\right)^{2k-1}\right]$

$= \frac{1}{n^{2k-1}}\Big[\sum_{i=1}^{n}(X_i-\mu)^{2k-1} + \binom{2k-1}{2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}(X_i-\mu)^{2k-2}(X_j-\mu) + \binom{2k-1}{2}\sum\sum_{i\neq j}(X_i-\mu)^{2k-\ldots}(X_j-\mu)\ldots$

$+\cdots + \binom{2k-1}{2}\binom{2k-3}{2}\cdots\binom{5}{2}\sum_{i_1=1}\cdots\sum_{i_{k-1}}^{n}(X_{i_1}-\mu)^2\cdots(X_{i_{k-2}}-\mu)^2(X_{i_{k-1}}-\mu)^3$

$= \frac{n}{n^{2k-1}}\mu_{2k-1} + \binom{2k-1}{2}\frac{n(n-1)}{n^{2k-1}}\mu_{2k-3}\mu^2 + \cdots$

$+ \frac{\lfloor 2k-1}{2^{k-2}}\frac{n(n-1)\cdots(n-k+2)}{n^{2k-1}}\sigma^{2(k-2)}\mu_3$

$= O\left(\frac{1}{n^k}\right).$

Theorem Let $X_1,\ldots,X_n$ be i.i.d. with $E(X_1)=\mu$, $Var(X_1)=\sigma^2$ and, for some $k\geq 3$ a function h has k derivatives, the kth derivatives of h and h^2 are bounded and the first k moments of X_1 exist.

Then

$$E(h(\bar{X})) = h(\mu) + \frac{\sigma^2}{2n}h''(\mu) + R_n$$

and

$$Var(h(\bar{X})) = \frac{\sigma^2}{n}[h'(\mu)]^2 + R_n,$$

where $R_n = O\left(\frac{1}{n^2}\right)$ (i.e., $\exists\, n_0$ and $c<\infty$ such that $R_n \equiv R_n(\mu) < \frac{c}{n^2}$, $\forall\, n>n_0$, $\forall\, \mu$).

Proof. Suppose that $|h^{(k)}(x)| \leq M$, $\forall\, x\in I$, where $P(X_1\in I)=1$.

Then

$$h(\bar{X}) = h(\mu) + h'(\mu)(\bar{X}-\mu) + h''(\mu)\frac{(\bar{X}-\mu)^2}{2} + \cdots + h^{(k-1)}(\mu)\frac{(\bar{X}-\mu)^{k-1}}{\lfloor k-1} + h^{(k)}(\xi_n(\mu,\bar{x}))\frac{(\bar{X}-\mu)^k}{\lfloor k},$$

where $\xi(\mu,\bar{x})$ depends on μ and $\bar{X}$ lies between μ and $\bar{X}$.

Let

$$R_n(\bar{X},\mu) \equiv \frac{(\bar{X}-\mu)^k}{\lfloor k} h^{(k)}(\xi_n(\mu,\bar{x}))$$

Then

$$|R_n(\bar{X},\mu)| \le \frac{M}{\underline{|2}} (\bar{X}-\mu)^2 \text{ and by last theorem,}$$

$$E(h(\bar{X})) = h(\mu) + \frac{\sigma^2}{2n} h''(\mu) + O\left(\frac{1}{n^2}\right) \quad \ldots \quad (I)$$

Now consider $\psi(x) = h^2(x)$, $x \in I$, so that, for $x \in I$,

$$\psi'(x) = 2h(x)\, h'(x)$$

$$\psi''(x) = 2\left[h(x)\, h''(x) + (h'(x))^2\right]$$

Applying (I) to $\psi(\cdot)$ we get

$$E(h^2(\bar{X})) = h^2(\mu) + \frac{\sigma^2}{2n} \times 2\left[h(\mu)\, h''(\mu) + (h'(\mu))^2\right] + O\left(\frac{1}{n^2}\right)$$

$$= h^2(\mu) + \frac{\sigma^2}{n}\left[h(\mu)\, h''(\mu) + (h'(\mu))^2\right] + O\left(\frac{1}{n^2}\right)$$

$$(E(h(\bar{X})))^2 = h^2(\mu) + \frac{\sigma^2}{n} h(\mu)\, h''(\mu) + O\left(\frac{1}{n^2}\right)$$

$$\Rightarrow \operatorname{Var}(h(\bar{X})) = E(h^2(\bar{X})) - (E(h(\bar{X})))^2$$

$$= \frac{\sigma^2}{n} (h'(\mu))^2 + O\left(\frac{1}{n^2}\right).$$

Theorem: Suppose that assumptions of last theorem hold and let $\{c_n\}_{n\ge 1}$ be a sequence of real numbers such that

$$c_n = 1 + \frac{a}{n} + O\left(\frac{1}{n^2}\right),$$

for some real constant a. Let $\delta_n(\bar{X}) = h(c_n \bar{X})$, for some function $h(\cdot)$, $n \ge 1$. Then

$$E(\delta_n(\bar{X})) = h(\mu) + \frac{a\mu}{n} h'(\mu) + \frac{\sigma^2}{2n} h''(\mu) + O\left(\frac{1}{n^2}\right).$$

$$\operatorname{Var}(\delta_n(\bar{X})) = \frac{\sigma^2}{n} (h'(\mu))^2 + O\left(\frac{1}{n^2}\right).$$

Proof. Let $\psi(x) = h(tx)$, $t>0$, $x \in I$, Then

$$E(\psi(\bar{X})) = \psi(\mu) + \frac{\sigma^2}{2n} \psi''(\mu) + O\left(\frac{1}{n^2}\right) \quad \ldots \quad (I)$$

$$\operatorname{Var}(\psi(\bar{X})) = \frac{\sigma^2}{n} [\psi'(\mu)]^2 + O\left(\frac{1}{n^2}\right). \quad \ldots \quad (II)$$

$$\psi'(x) = t\, h'(tx), \quad x \in I$$

$$\psi''(x) = t^2\, h''(tx), \quad x \in I$$

Thus (F) yields

$$E[h(t\bar{X})] = h(t\mu) + \frac{\sigma^2}{2n} t^2 h''(t\mu) + O\left(\frac{1}{n^2}\right), \quad t>0.$$

$$E[h(c_n\bar{X})] = h(c_n\mu) + \frac{\sigma^2}{2n} c_n^2 h''(c_n\mu) + O\left(\frac{1}{n^2}\right)$$

For a function $k(\cdot)$, define

$$\alpha(t) = k(t\mu) \quad t>0.$$

Then, provided k has derivatives upto order m,

$$\alpha(t) = \alpha(1) + (t-1)\alpha'(1) + \frac{(t-1)^2}{\lfloor 2}\alpha''(1) + \cdots + \frac{(t-1)^{m-1}}{\lfloor m-1}\alpha^{(m-1)}(1) + \frac{t^m}{\lfloor m}\alpha^{(m)}(\xi),$$

for some ξ between 1 and t.

$$k(t\mu) = k(\mu) + (t-1)\mu k'(\mu) + \frac{(t-1)^2}{\lfloor 2}\mu^2 k''(\mu) + \cdots + \frac{(t-1)^{m-1}}{\lfloor m-1}\mu^{m-1}k^{(m-1)}(\mu) + \frac{(t-1)^m}{\lfloor m}\mu^m k^{(m)}(\xi) \quad \text{....} \quad (III)$$

$$h(c_n\mu) = h(\mu) + (c_n-1)\mu h'(\mu) + \frac{(c_n-1)^2}{\lfloor 2}\mu^2 h''(\mu) + \cdots + \frac{(c_n-1)^{m-1}}{\lfloor m-1}\mu^{m-1}h^{(m-1)}(\mu) + \frac{(c_n-1)^m}{\lfloor m}\mu^m h^{(m)}(\xi_1),$$

(taking $t=c_n$ and $k=h$ in (III))

for some ξ_1 between 1 and c_n

$$\boxed{h(c_n\mu) = h(\mu) + \frac{a\mu}{n}h'(\mu) + O\left(\frac{1}{n^2}\right)}$$

$$h''(c_n\mu) = h''(\mu) + (c_n-1)\mu h^{(3)}(\mu) + \cdots + \frac{(c_n-1)^{k-3}}{\lfloor k-3}\mu^{k-3}h^{(k-1)}(\mu) + \frac{(c_n-1)^{k-2}}{\lfloor k-2}\mu^{k-2}h^{(k)}(\xi_2),$$

for some ξ_2 between 1 and c_n (taking $t=c_n$ and $k=h'$ in (III))

$$\boxed{h''(c_n\mu) = h''(\mu) + \frac{a\mu}{n}h^{(3)}(\mu) + O\left(\frac{1}{n^2}\right)}$$

$$E(h(c_n\bar{X})) = h(\mu) + \frac{a\mu}{n}h'(\mu) + \frac{\sigma^2}{2n}c_n^2\left\{h''(\mu) + \frac{a\mu}{n}h^{(3)}(\mu) + O\left(\frac{1}{n^2}\right)\right\}$$

$$= h(\mu) + \frac{a\mu}{n}h'(\mu) + \frac{\sigma^2}{2n}\left\{1 + O\left(\frac{1}{n}\right)\right\}\left\{h''(\mu) + O\left(\frac{1}{n}\right)\right\}$$

$$= h(\mu) + \frac{a\mu}{n}h'(\mu) + \frac{\sigma^2}{2n}h''(\mu) + O\left(\frac{1}{n^2}\right)$$

$$\boxed{E[\delta_n(\bar{X})] = E[h(c_n\bar{X})] = h(\mu) + \frac{a\mu}{n}h'(\mu) + \frac{\sigma^2}{2n}h''(\mu) + O\left(\frac{1}{n^2}\right)}$$

From (II)

$$Var(h(t\bar{X})) = \frac{\sigma^2}{n}[t\,h'(t\mu)]^2 + O\left(\frac{1}{n^2}\right)$$

$$Var(h(c_n\bar{X})) = \frac{\sigma^2}{n}c_n^2\left(h'(c_n\mu)\right)^2 + O\left(\frac{1}{n^2}\right)$$

$$= \frac{\sigma^2}{n}\left\{1+O\left(\frac{1}{n}\right)\right\}\left[h'(c_n\mu)\right]^2 + O\left(\frac{1}{n^2}\right)$$

Using (III) with $k \equiv h'$ and $t = c_n$, we get

$$h'(c_n\mu) = h'(\mu) + (c_n-1)\mu h''(\mu) + \frac{(c_n-1)^2}{\underline{|2}}\mu^2 h^{(3)}(\mu) + \cdots + \frac{(c_n-1)^{k-2}}{\underline{|k-2}}\mu^{k-2}h^{(k-1)}(\mu) + \frac{(c_n-1)^{k-1}}{\underline{|k-1}}\mu^{k-1}h^{(k)}(\xi_3),$$

for some ξ_3 between 1 and c_n.

$$h'(c_n\mu) = h'(\mu) + O\left(\frac{1}{n}\right)$$

$$Var(c_n\bar{X}) = \left\{\frac{\sigma^2}{n} + O\left(\frac{1}{n^2}\right)\right\}\left\{h'(\mu) + O\left(\frac{1}{n}\right)\right\}^2 + O\left(\frac{1}{n^2}\right)$$

$$= \frac{\sigma^2}{n}\left(h'(\mu)\right)^2 + O\left(\frac{1}{n^2}\right).$$

Example Let $X_1, \ldots, X_n$ be iid $N(\theta, 1)$ where $\theta \in \Theta = \mathbb{R}$. Consider estimation of $g(\theta) = \Phi(u_0-\theta)$ for a given real constant u_0 and the MLE $\delta_n^{(2)}(\underline{X}) = \Phi(u_0-\bar{X})$. Show that for the UMVUE $\delta_n^{(1)}(\underline{X}) = \Phi\left(\sqrt{\frac{n}{n-1}}(u_0-\bar{X})\right)$,

$$E_\theta(\delta_n^{(1)}(\underline{X})) = \Phi(u_0-\theta) - \frac{u_0(\ldots)}{2n}\phi(u_0-\theta) + O\left(\frac{1}{n^2}\right)$$

$$E_\theta(\delta_n^{(2)}(\underline{X})) = \Phi(u_0-\theta) - \frac{1}{2n}\phi(u_0-\theta)(u_0-\theta) + O\left(\frac{1}{n^2}\right)$$

$$Var_\theta(\delta_n^{(1)}(\underline{X})) = \frac{1}{n}\phi^2(u_0-\theta) + O\left(\frac{1}{n^2}\right).$$

$$Var_\theta(\delta_n^{(2)}(\underline{X})) = \frac{\phi^2(u_0-\theta)}{n} + O\left(\frac{1}{n^2}\right).$$

__Solution__ We have

$$\delta_n(\underline{X}) = h(c_n\bar{X}),$$

where $c_n = \sqrt{\frac{n}{n-1}} = \left(1-\frac{1}{n}\right)^{-\frac{1}{2}} = 1 + \frac{1}{2n} + O\left(\frac{1}{n^2}\right)$, and

$$h(x) = \Phi(u_0-x), \quad x \in \mathbb{R}$$

$$h'(x) = -\phi(u_0-x), \quad h''(x) = \phi'(u_0-x) = -(u_0-x)\phi(u_0-x)$$

$$h^{(3)}(x) = \phi(u_0-x) + (u_0-x)\phi'(u_0-x) = \phi(u_0-x) - (u_0-x)^2\phi(u_0-x)$$

$$h^{(4)}(x) = -\phi'(u_0-x) + 2(u_0-x)\phi(u_0-x) - (u_0-x)^2\phi'(u_0-x)$$

$$= 3(u_0-x)\,\phi(u_0-x) - (u_0-x)^3\phi(u_0-x)$$

$$= \left(3(u_0-x) - (u_0-x)^3\right)\phi(u_0-x) \to 0, \quad \text{as } x\to\pm\infty$$

$\Rightarrow h^{(4)}(x)$ is bounded.

$$E_\theta(\delta_n^{(1)}(X)) = E(h(C_n\bar{X}))$$

$$= h(\theta) + \frac{a\theta}{n}h'(\theta) + \frac{1}{2n}h''(\theta) + O\left(\frac{1}{n^2}\right)$$

$$= \Phi(u_0-\theta) - \frac{\theta}{2n}\phi(u_0-\theta) - \frac{1}{2n}(u_0-\theta)\phi(u_0-\theta) + O\left(\frac{1}{n^2}\right)$$

$$= \Phi(u_0-\mu) - \frac{u_0}{2n}\phi(u_0-\theta) + O\left(\frac{1}{n^2}\right)$$

$$Var_\theta(\delta_n^{(1)}(X)) = Var(h(C_n\bar{X}))$$

$$= \frac{1}{n}(h'(\theta))^2 + O\left(\frac{1}{n^2}\right)$$

$$= \frac{\phi^2(u_0-\theta)}{n} + O\left(\frac{1}{n^2}\right)$$

Assignment Problems

(1) Let $X_1 \ldots X_n$ be i.i.d $Exp(\theta)$ where $\theta\in\Theta=(0,\infty)$, here $\theta = E_\theta(X_1)$. Consider estimators $\delta_n^{(1)} = e^{-\bar{X}}$ and $\delta_n^{(2)} = \frac{1}{n}\sum_{i=1}^n I(X_i\geq 1)$ for estimating $g(\theta) = e^{-\theta}$. Are $\delta_n^{(1)}$ and $\delta_n^{(2)}$ consistent estimators? Find their limiting risks and find $l_{\delta^{(1)},\delta^{(2)}}(\theta)$, $\theta\in\Theta$.

(2) Let $X_1, X_2, \ldots, X_n$ be i.i.d. $N(\mu,\sigma^2)$ where $\theta=(\mu,\sigma)\in\Theta=\mathbb{R}\times(0,\infty)$. Consider the UMVUE $\delta_n^{(1)} = \bar{X}^2 - \frac{S^2}{n(n-1)}$ and the MLE $\delta_n^{(2)} = \bar{X}^2$, where $S^2 = \sum_{i=1}^n (X_i-\bar{X})^2$ for estimating $g(\theta)=\mu^2$. Are $\delta_n^{(1)}$ and $\delta_n^{(2)}$ consistent estimators? Find their limiting risks and find $l_{\delta^{(1)},\delta^{(2)}}(\theta)$, $\theta\in\Theta$.

(3) Let $X_1, \ldots, X_n$ be i.i.d. $U(0,\theta)$; where $\theta\in\Theta=(0,\infty)$. For estimating $g(\theta)=\theta^4$. Consider estimators $\delta_n^{(1)} = \frac{n+5}{n+4}X_{(n)}^4$ and $\delta_n^{(2)} = 16\bar{X}^4$. Are $\delta_n^{(1)}$ and $\delta_n^{(2)}$ consistent estimators. Find their limiting risks and find $l_{\delta^{(1)},\delta^{(2)}}(\theta)$, $\theta\in\Theta$.

(4) Let $X_1 \ldots X_n$ be iid $Bin(1,\theta)$, $\theta\in\Theta$. For estimating $g(\theta) = Var_\theta(X_1)$, let $\delta_n^{(1)}$ be the UMVUE and let $\delta_n^{(2)}$ be the MLE. Find the LRE $l_{\delta_n^{(1)},\delta_n^{(2)}}(\theta)$ and ARE $e_{\delta_n^{(1)},\delta_n^{(2)}}(\theta)$, if they exist.

Remark: (a) Assumptions of last theorem are satisfied when $h(\cdot)$ is a polynomial.

(b) A drawback of limiting risk approach as a large sample measure is that in certain situations the moments (and hence risk) may not exist. However it may happen that the region where risk does not exist shrinks to empty set as $n \to \infty$. In such situations the asymptotic distribution approach may be more realistic.

The Asymptotic Distribution Approach

As discussed before, consistency only tells us that for large n, the error $\|\delta_n - g(\theta)\|$ is likely to be small and it does not tell us whether the order of the error is $\frac{1}{n}, \frac{1}{\sqrt{n}}, \frac{1}{\log n}$ etc.

To get an idea consider a consistent estimator $\{\delta_n\}_{n\geq 1}$ and a sequence $\{k_n\}_{n\geq 1}$ of positive real constants. For any fixed $\varepsilon > 0$, consider

$$P_n(\varepsilon) = P_\theta\left(\|\delta_n(\underline{X}) - g(\theta)\| < \frac{\varepsilon}{k_n}\right), \quad n = 1, 2, \ldots$$

Consider the following two possibilities.

__Case I__ $\{k_n\}_{n\geq 1}$ is bounded.

In this case

$$\lim_{n\to\infty} P_n(\varepsilon) = 1, \quad \forall \varepsilon > 0$$

__Case II.__ $k_n \to \infty$, sufficiently fast

In this case we may have

$$\lim_{n\to\infty} P_n(\varepsilon) = 0, \quad \forall \varepsilon > 0$$

This suggests that, for a given $\varepsilon > 0$, there might exist an intermediate sequence $\{k_n\}_{n\geq 1}$ with $k_n \to \infty$ and

$$0 < \lim_{n\to\infty} P_n(\varepsilon) < 1.$$

Commonly there will exist a sequence $\{k_n\}_{n\geq 1}$ diverging to ∞ and a limiting continuous d.f. H s.t.

$$\lim_{n\to\infty} P\left(k_n(\delta_n - g(\theta)) \leq \underline{\varepsilon}\right) = H(\underline{\varepsilon}) \quad \ldots\ldots (*)$$

We shall then say that the error $\|\delta_n - g(\theta)\|$ tends to zero with rate $\frac{1}{k_n}$.

Remark:

(a) $(*) \Rightarrow k_n(\delta_n - g(\theta)) \xrightarrow{d} Y$, where Y has d.f. $H(\cdot)$

$\Rightarrow \delta_n - g(\theta) \xrightarrow{p} 0$

$\Rightarrow \{\delta_n\}_{n\geq 1}$ is consistent for $g(\theta)$.

(b) The error rate is not uniquely determined. If $\frac{1}{k_n}$ is a possible rate, so is $\frac{1}{m_n}$ for any sequence $\{m_n\}_{n\geq 1}$ for which $\frac{m_n}{k_n} \to c \neq 0$, as $n \to \infty$ ($m_n(\delta_n - g(\theta)) = \frac{m_n}{k_n} k_n(\delta_n - g(\theta)) \to cY$)

(c) If $\{m_n\}_{n\geq 1}$ diverges to ∞ more slowly (or faster) than $\{k_n\}$ i.e. $\lim_{n\to\infty} \frac{m_n}{k_n} = 0$ $(=\infty)$, then $m_n(\delta_n - g(\theta)) \xrightarrow{p} 0$ (or ∞).

Lemma: $Y_n \xrightarrow{d} Y \Leftrightarrow E(\psi(Y_n)) \to E(\psi(Y))$, for every bounded continuous real-valued function $\psi(\cdot)$.

Remark: If $\{\delta_n\}_{n\geq 1}$ is consistent for $g(\theta)$, generally one has, for some sequence $\{k_n\}_{n\geq 1}$ of real constants,

$$k_n(\delta_n - g(\theta)) \xrightarrow{d} N(0, \tau^2(\theta)).$$

Then

$$P(k_n|\delta_n - g(\theta)| < \varepsilon) \approx 2\,\Phi\left(\frac{\varepsilon}{\tau(\theta)}\right) - 1, \text{ for large } n.$$

Thus the large sample behavior of the estimators $\{\delta_n\}_{n\geq 1}$ can be studied in terms of asymptotic variance $\tau^2(\theta)$.

Definition For an estimator $\{\delta_n\}_{n\geq 1}$ suppose that $k_n(\delta_n - g(\theta)) \xrightarrow{d} N(a(\theta), \tau^2(\theta))$ as $n \to \infty$, where $\{k_n\}_{n\geq 1}$ is some sequence of real numbers.

(a) $\tau^2(\theta)$ is called the asymptotic variance of $\{\delta_n\}_{n\geq 1}$;

(b) $a(\theta)$ is called the asymptotic bias of $\{\delta_n\}_{n\geq 1}$

(c) If $a(\theta) = 0$ $\forall \theta \in \Theta$ then $\{\delta_n\}_{n\geq 1}$ is called asymptotically unbiased.

(d) If $\{\delta_n\}_{n\geq 1}$ is asymptotically unbiased (i.e. $a(\theta) = 0$ $\forall \theta$) then $g(\theta)$ is called the asymptotic mean of $\{\delta_n\}_{n\geq 1}$

(Limiting variance and Asymptotic Variance may not be the same). To see this, let $Z \sim N(0,1)$ and $Y_n \sim N(0, \gamma_n^2)$, $n=1,2,\ldots$ where $\gamma_n \to \infty$, as $n \to \infty$. Define

$$Z_n = \begin{cases} Z & \text{w.p. } \pi_n \\ Y_n & \text{w.p. } 1-\pi_n \end{cases}, \quad n=1,2,\ldots,$$

where $\pi_n \to 1$. Then

$$\begin{aligned} F_{Z_n}(z) &= P(Z \le z)\pi_n + P(Y_n \le z)(1-\pi_n) \\ &= \Phi(z)\pi_n + \Phi\left(\frac{z}{\gamma_n}\right)(1-\pi_n) \\ &\to \Phi(z) \end{aligned}$$

Asymptotic Variance $= 1$

$$E(Z_n) = E(Z)\pi_n + E(Y_n)(1-\pi_n) = 0, \quad n=1,2,\ldots$$

$$\begin{aligned} Var(Z_n) &= E(Z_n^2) \\ &= E(Z^2)\pi_n + E(Y_n^2)(1-\pi_n) \\ &= \pi_n + \gamma_n^2(1-\pi_n). \end{aligned}$$

Choose γ_n and π_n s.t. $\gamma_n \to \infty$, $\pi_n \to 1$ but $\gamma_n^2(1-\pi_n) \not\to 0$ (e.g. $\gamma_n = \sqrt{n}$, $\pi_n = 1 - \frac{1}{n}$, $n \ge 1$)

<u>Remark</u> Suppose that $Y_n = k_n(\delta_n - E(\delta_n)) \xrightarrow{d} Y$, where $E(Y) = 0$. Then, provided the limit exists,

$$\lim_{n\to\infty} E(Y_n^2) \geq E(Y^2)$$

i.e.

limit of variance $\geq$ asymptotic variance.

The equality is attained when $\delta_n \equiv h(\bar{X})$, where $\bar{X}$ is the sample mean based on i.i.d. observations, and, for some $k \geq 3$, first k derivatives of h exist and $h^{(k)}$ is bounded.

Most estimators of interest are consistent and suitably normalized have asymptotic normal distribution with ~~the~~ the ~~...~~ asymptotic mean same as the estimand $g(\theta)$ and asymptotic variance $v(\theta)$.

Definition Let $\{\delta_n\}_{n\geq 1}$ be a sequence of consistent estimators for the estimand $g(\theta)$. Suppose that there exists a sequence $\{k_n\}_{n\geq 1}$ of real numbers such that

$$k_n(\delta_n - g(\theta)) \xrightarrow{d} N(0, v(\theta)) \quad \theta \in \Theta$$

where $v(\theta) > 0$ $\forall \theta \in \Theta$. Then $\{\delta_n\}_{n\geq 1}$ is called Consistent and Asymptotically Normal (CAN) estimator of $g(\theta)$

Clearly, the performance of a CAN estimator $\{\delta_n\}_{n\geq 1}$ can be evaluated through its asymptotic variance $v(\theta)$. Among CAN estimators the one for which the error $\|\delta_n - g(\theta)\|$ converges to zero at the faster rate is preferred and among CAN estimators having the same rate of error convergence the one with smaller asymptotic variance is preferred. It would be of interest to find, among CAN estimators having given rate of error, the one with the smallest asymptotic variance $v(\theta)$ $\forall \theta \in \Theta$.

Rao-Cramer Bound (Information Inequality) Let $X_1, X_2, \ldots, X_n$ be i.i.d. with p.d.f./p.m.f. $f_\theta(x)$, where $\theta \in \Theta \subseteq \mathbb{R}$. Suppose that

(a) Θ is an open interval;

(b) ~~distribution~~ support of X_1 does not depend on θ;

(c) for every $\theta \in \Theta$ $f_\theta(\cdot)$ is twice differentiable w.r.t. θ with second derivative continuous in θ;

(d) the integral $\int f_\theta(x)\,dx$ can be twice differentiated ~~twice~~ under the integral sign, so that, for all $\theta \in \Theta$

$$E_\theta\left(\frac{\partial}{\partial\theta} \ln f_\theta(X_1)\right) = 0$$

and $$E_\theta\left(-\frac{\partial^2}{\partial\theta^2} \ln f_\theta(X_1)\right) = E_\theta\left(\left(\frac{\partial}{\partial\theta} \ln f_\theta(X_1)\right)^2\right) = I(\theta)$$

(e) For any $\theta_0 \in \Theta$ $\exists$ $c > 0$ and a function M (both depending on θ_0) such that

$$\left| \frac{\partial^2}{\partial \theta^2} \ln f_\theta(x) \right| \le M(x), \quad \forall\ x \in A, \quad \theta_0 - c < \theta < \theta_0 + c$$

and

$$E_{\theta_0}(M(X)) < \infty,$$

where $A = \{x \in \mathbb{R} : f_\theta(x) > 0\}$.

Let $\{\delta_n\}_{n\ge1}$ be any estimator such that

$$\sqrt{n}\,(\delta_n - g(\theta)) \xrightarrow{d} N(0, v(\theta)), \quad v(\theta) > 0, \ \forall\ \theta \in \Theta$$

Then

$$v(\theta) \geq \frac{[g'(\theta)]^2}{I(\theta)}, \quad \dots \quad (I)$$

$\forall\ \theta \in \Theta$, except on a set of Lebesgue measure 0.

Remark: (a) Under the conditions (a)–(d) of above theorem, Rao-Cramer bound (information inequality) for $\{\delta_n\}_{n\ge1}$ is

$$Var_\theta(\delta_n) \geq \frac{[\frac{d}{d\theta} E_\theta(\delta_n)]^2}{n\, I(\theta)} . \quad \dots \quad (II)$$

Thus if $E_\theta(\delta_n) = g(\theta)$, $\forall\ \theta \in \Theta$. Then

$$Var_\theta(\delta_n) \geq \frac{(g'(\theta))^2}{n I(\theta)}.$$

Now if

$$\sqrt{n}\,(\delta_n - g(\theta)) \xrightarrow{d} N(0, v(\theta))$$

then

$$\liminf\,[Var_\theta(\sqrt{n}(\delta_n - g(\theta)))] \geq v(\theta)$$

$\Rightarrow$ (I) is a consequence of (II) provided

$$\liminf\,[Var_\theta(\sqrt{n}(\delta_n - g(\theta))] = v(\theta).$$

Definition A sequence $\{\delta_n\}_{n\ge1}$ of estimators is said to be asymptotically efficient for estimating $g(\theta)$ if the following conditions are satisfied

(a) $\sqrt{n}\,(\delta_n - g(\theta)) \xrightarrow{d} N(0, v(\theta))$, as $n \to \infty$

(b) $$v(\theta) = \frac{(g'(\theta))^2}{I(\theta)}, \quad \forall\ \theta \in \Theta$$

Remark: There exist estimators $\{\delta_n\}_{n\geq 1}$ for which

$$\sqrt{n}\,(\delta_n - g(\theta)) \xrightarrow{d} N(0, v(\theta))$$

but

$$v(\theta) \geq \frac{[g'(\theta)]^2}{I(\theta)}$$

getting violated at some θ's (such θ's have the Lebesgue measure zero), called points of superefficiency.

Example Let $X_1, \ldots, X_n$ be iid $N(\theta, 1)$, $\theta \in \Theta = \mathbb{R}$. Then $\bar{X}$ is the MLE for $g(\theta)=\theta$, which is consistent and

$$\sqrt{n}\,(\bar{X} - \theta) \xrightarrow{d} N(0, 1)$$

Here $I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2} \ln f_\theta(X_1)\right]$,

$$\left[\frac{g'(\theta)}{I(\theta)}\right]^2 = 1 = v(\theta)$$

$\Rightarrow$ $\bar{X}$ is asymptotically efficient

Now, for $0 \leq a < 1$ consider

$$\delta_n = \begin{cases} \bar{X}, & \text{if } |\bar{X}| \geq \frac{1}{n^{1/4}} \\ a\bar{X}, & \text{if } |\bar{X}| < \frac{1}{n^{1/4}} \end{cases}$$

$$P_\theta(\sqrt{n}(\delta_n - \theta) \leq x) = P_\theta\left(\sqrt{n}(a\bar{X} - \theta) \leq x,\ |\bar{X}| < \frac{1}{n^{1/4}}\right) \quad \cdots (A)$$
$$+ P_\theta\left(\sqrt{n}(\bar{X} - \theta) \leq x,\ |\bar{X}| \geq \frac{1}{n^{1/4}}\right)$$

Case I $\theta \neq 0$

$$P\left(|\bar{X}| < \frac{1}{n^{1/4}}\right) = P\left(-\frac{1}{n^{1/4}} < \bar{X} < \frac{1}{n^{1/4}}\right)$$
$$= \Phi\left(\sqrt{n}\left(\frac{1}{n^{1/4}} - \theta\right)\right) - \Phi\left(\sqrt{n}\left(-\frac{1}{n^{1/4}} - \theta\right)\right)$$
$$\to 0, \quad \text{as } n \to \infty \quad \text{(Consider } \theta<0 \text{ and } \theta>0 \text{ separately)}$$

$\Rightarrow$ first term in (A) goes to zero as $n \to \infty$. Also

$$\lim_{n\to\infty} P_\theta(\sqrt{n}(\bar{X} - \theta) \leq x) = \lim_{n\to\infty} P_\theta\left(\sqrt{n}(\bar{X} - \theta) \leq \min\left\{x, \sqrt{n}\left(-\frac{1}{n^{1/4}} - \theta\right)\right\}\right)$$
$$+ P\left(\sqrt{n}\left(\frac{1}{n^{1/4}} - \theta\right) \leq \sqrt{n}(\bar{X} - \theta) \leq x\right)$$

$= \lim_{n\to\infty} P_\theta(\sqrt{n}(\bar{X}-\theta) \le x)$ (again considering cases of $\theta<0$ and $\theta>0$ separately)

$= \Phi(x)$

Case II $\theta=0$.

$P(|\bar{X}| \ge \frac{1}{n^{1/4}}) = 1 - P(|\bar{X}| < \frac{1}{n^{1/4}})$

$= 1 - [\Phi(n^{1/4}) - \Phi(-n^{1/4})] \to 0$

Thus

$$\lim_{n\to\infty} P(\sqrt{n}(\delta_n-\theta) \le x) = \lim_{n\to\infty} P_\theta(\sqrt{n}(a\bar{X}-\theta) \le x,\ |\bar{X}| < \frac{1}{n^{1/4}})$$

$$= \lim_{n\to\infty} P_\theta(\sqrt{n}\bar{X} \le \frac{x}{a},\ -\frac{\sqrt{n}}{n^{1/4}} < \sqrt{n}\bar{X} < \frac{\sqrt{n}}{n^{1/4}})$$

$$= \lim_{n\to\infty} P_\theta(\sqrt{n}\bar{X} \le \frac{x}{a})$$

$$= \Phi(\frac{x}{a})$$

Thus

$$\sqrt{n}(\delta_n-\theta) \xrightarrow{d} N(0, v(\theta)),$$

where

$$v(\theta) = \begin{cases} 1, & \text{if } \theta\neq 0 \\ a^2, & \text{if } \theta=0 \end{cases}$$

Clearly, for $0<a<1$, $a^2<1$. Thus

$$v(\theta) \ge \frac{[g'(\theta)]^2}{I(\theta)} = 1$$

is violated at $\theta=0$. ~~That~~ Note that the Lebesgue measure of the set $\{0\}$ is 0.

Definition (Asymptotic Relative Efficiency). Suppose that $\{\delta_n^{(1)}\}_{n\ge1}$ and $\{\delta_n^{(2)}\}_{n\ge1}$ be two sequences of estimators of $g(\theta)$ such that for some $\alpha>0$ and for any sequence $n' \equiv n'(n)$ ($n'\to\infty$ as $n\to\infty$)

$$n^\alpha(\delta_n^{(1)} - g(\theta)) \to N(0, \tau_1^2(\theta))$$

$$\text{and } n'^\alpha(\delta_{n'}^{(2)} - g(\theta)) \to N(0, \tau_2^2(\theta)).$$

Then the asymptotic relative efficiency (ARE) of

$S_n^{(1)}$ relative to $S_n^{(2)}$ is defined by

$$e_{S^{(1)}, S^{(2)}}(\theta) \equiv e_{S^{(1)}, S^{(2)}} = \lim_{n\to\infty} \frac{n'(n)}{n}, \quad \theta \in \Theta$$

provided the limit exists and is independent of the subsequence n'.

<u>Remark</u> $e_{S^{(1)}, S^{(2)}} = \frac{1}{2} \Rightarrow$ half as many observations are therefore required with $S_n^{(2)}$ as with $S_n^{(1)}$. Thus $S_n^{(2)}$ is twice as efficient as $S_n^{(1)}$.

<u>Theorem</u> Suppose that, for some $\alpha > 0$, $n^\alpha (S_n^{(i)} - g(\theta)) \xrightarrow{d} N(0, \tau_i^2(\theta))$, $i = 1, 2$, $\tau_i^2(\theta) > 0$, $i = 1, 2$, $\theta \in \Theta$. Then

$$e_{S^{(1)}, S^{(2)}}(\theta) = \left(\frac{\tau_2^2(\theta)}{\tau_1^2(\theta)} \right)^{\frac{1}{\alpha}}, \quad \theta \in \Theta$$

<u>Proof</u> Suppose that

$$n^\alpha (S_n^{(1)} - g(\theta)) \xrightarrow{d} \tau_1^2(\theta)$$

$$n^\alpha (S_{n'}^{(2)} - g(\theta)) \xrightarrow{d} \tau_1^2(\theta).$$

Then

$$n^\alpha (S_{n'}^{(2)} - g(\theta)) = \frac{n^\alpha}{(n')^\alpha} \underbrace{(n')^\alpha (S_{n'}^{(2)} - g(\theta))}_{\downarrow d \ \tau_2^2(\theta)}$$

$$\downarrow d$$

$$\tau_1^2(\theta) \leftarrow n^\alpha (S_n^{(1)} - g(\theta))$$

$$\Rightarrow \lim_{n\to\infty} \left(\frac{n'(n)}{n} \right)^\alpha = \frac{\tau_2^2(\theta)}{\tau_1^2(\theta)}.$$

$$\Rightarrow e_{S^{(1)}, S^{(2)}}(\theta) = \left(\frac{\tau_2^2(\theta)}{\tau_1^2(\theta)} \right)^{\frac{1}{\alpha}}.$$

<u>Remark</u> (a) One can generalize the definitions of the ARE and LRE to cover cases in which asymptotic dist. is not normal (but normalizing factor is of order n^α).

(b) If the normalizing factor is not of the form n^α, $\alpha > 0$, then the ratio of sample sizes can not be used to measure ARE.

Example If $\lim_n (\delta_n - g(\theta)) \xrightarrow{d} N(0, v(\theta))$ then for any positive integer m

$$\lim_n (\delta_{mn} - g(\theta)) \xrightarrow{d} N(0, v(\theta))$$

$\Rightarrow$ ARE of δ relative to δ is

$$\lim_{n\to\infty} \frac{mn}{n} = m$$

and m is arbitrary.

Example Let $X_1 \ldots X_n$ be iid $N(\theta, 1)$, $\theta \in \Theta = \mathbb{R}$. For estimating $g(\theta) = P_\theta(X_1 \le u_0) = \Phi(u_0 - \theta)$, the UMVUE is

$$\delta_n^{(1)} = \Phi\left(\sqrt{\tfrac{n}{n-1}}\,(u_0 - \bar{X})\right)$$

but

$$\delta_n^{(2)} = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le u_0)$$

Let $h(x) = \Phi(u_0 - x)$, $x \in \mathbb{R}$. We have

$$\sqrt{n}(\bar{X} - \theta) \xrightarrow{d} Z \sim N(0, 1)$$

$h'(x) = -\phi(u_0 - x)$, $x \in \mathbb{R}$. Thus

$$\sqrt{n}\left(h\left(\sqrt{\tfrac{n}{n-1}}\bar{X}\right) - h(\theta)\right) \xrightarrow{d} h'(\theta) Z$$

$$\sqrt{n}(\delta_n^{(1)} - g(\theta)) \xrightarrow{d} N(0, \phi^2(u_0 - \theta))$$

Also

$$\sqrt{n}(\delta_n^{(2)} - g(\theta)) \xrightarrow{d} N(0, g(\theta)(1 - g(\theta)))$$

$$e_{\delta^{(2)}, \delta^{(1)}}(\theta) = \frac{\phi^2(u_0 - \theta)}{g(\theta)(1 - g(\theta))} = \frac{\phi^2(u_0 - \theta)}{\Phi(u_0 - \theta)(1 - \Phi(u_0 - \theta))}$$

$$e_{\delta^{(2)}, \delta^{(1)}}(u_0) = \frac{2}{\pi} \approx 0.637, \quad e_{\delta^{(2)}, \delta^{(1)}}(-\infty) = 0$$

Consistency of Method of Moments Estimators

Let $X_1, X_2, \ldots, X_n$ be a random sample where X_1 has a d.f. $F \in \mathcal{P} = \{F_{\underline{\theta}} : \underline{\theta} = (\theta_1, \ldots, \theta_k) \in \Theta\}$, $\Theta \subseteq \mathbb{R}^k$. Suppose that, for each $\underline{\theta} \in \Theta$ the functional form of $F_{\underline{\theta}}$ is known but $\underline{\theta} \in \Theta$ is unknown. Further suppose that, for every $r = 1, 2, \ldots, k$,

$$m_r(\underline{\theta}) = E_{\underline{\theta}}(X_i^r)$$

exists and is finite.

Define sample moments

$$A_r(\underline{X}) = \frac{1}{n} \sum_{i=1}^{n} X_i^r, \quad r = 1, 2, \ldots, k.$$

Clearly $(A_1(\underline{X}), \ldots, A_k(\underline{X}))$ is a consistent estimator of $\underline{\theta} = (\theta_1, \ldots, \theta_k)$.

In method of moments estimation procedure, one equates $A_r(\underline{X})$, $r = 1, \ldots, k$, sample moments with population moments $m_r(\underline{\theta})$ and solves simultaneous equation to get estimator, i.e., one solves simultaneous equations

$$m_r(\underline{\theta}) = A_r(\underline{X}) \quad r = 1, 2, \ldots, k$$

to get a solution $\hat{\underline{\theta}} \equiv \hat{\underline{\theta}}(\underline{X}) = (\hat{\theta}_1(\underline{X}), \ldots, \hat{\theta}_k(\underline{X}))$. The estimator $\hat{\underline{\theta}}(\underline{X})$ is called the Method of Moments Estimator (MME) of $\underline{\theta}$.

If $h : \Theta \to \Lambda$ is a mapping from Θ onto $\Lambda \subseteq \mathbb{R}^m$ and $\hat{\underline{\theta}}(\underline{X})$ is a MME of $\underline{\theta}$, then $h(\hat{\underline{\theta}}(\underline{X}))$ is called a MME of $h(\underline{\theta})$. In particular, if $\hat{\underline{\theta}}(\underline{X}) = (\hat{\theta}_1(\underline{X}), \ldots, \hat{\theta}_k(\underline{X}))$ is a MME of $\underline{\theta} = (\theta_1, \ldots, \theta_k)$ then $\hat{\theta}_i(\underline{X})$ is called a MME of θ_i, $i = 1, \ldots, k$.

Remark: (a) The method of moments estimation procedure is not applicable when one or more $m_r(\underline{\theta})$ $(r = 1, \ldots, k)$ do not exist (Example: Cauchy Distribution).

(b) The MME may not exist when simultaneous equations

$$m_r(\underline{\theta}) = A_r(\underline{X}), \quad r = 1, 2, \ldots, k$$

do not possess a solution.

(c) In situations where one or more $m_r(\underline{\theta})$ $(r = 1, \ldots, k)$ do not depend on $\underline{\theta}$ the MME may not exist. In those situations, one may ignore ~~An~~ equating such population moments with sample moments and add additional equations by considering population additional moments and equating them with corresponding sample moments. We call such an estimator as modified MME.

(d) MME may not be unique as the underlying equations

$$m_r(\underline{\theta}) = A_r(\underline{X}), \quad r = 1, 2, \ldots, k$$

may have more than one solution.

(d) If $m_1(\underline{\theta}), \ldots, m_k(\underline{\theta})$ are continuous functions of $\underline{\theta}$ and the Jacobian

$$\left| \frac{\partial(m_1(\underline{\theta}), \ldots, m_k(\underline{\theta}))}{\partial\theta_1 \cdots \partial\theta_k} \right| \neq 0,$$

then the MME $\hat{\underline{\theta}}(\underline{X}) = (\hat{\theta}_1(\underline{X}), \ldots, \hat{\theta}_k(\underline{X}))$ is consistent for $\underline{\theta}$. In particular $\hat{\theta}_i(\underline{X})$ is consistent for θ_i, $i = 1, 2, \ldots, k$. In fact if $h : \Theta \to \mathbb{R}$ is a continuous function of $\underline{\theta}$ then the MME $h(\hat{\underline{\theta}}(\underline{X}))$ is a consistent estimator of $h(\underline{\theta})$.

(e) We have

$$\sqrt{n}\,\big((A_1(\underline{X}), \ldots, A_k(\underline{X})) - (m_1(\underline{\theta}), \ldots, m_k(\underline{\theta}))\big) \xrightarrow{d} N_k(0, \Sigma^*),$$

for some $\Sigma^* > 0$. Under the assumptions in (d)

$$\sqrt{n}\,(\hat{\underline{\theta}}(\underline{X}) - \underline{\theta}) \xrightarrow{d} N_k(0, \Sigma),$$

for some $\Sigma > 0$.

Example Let $X_1, X_2, \ldots, X_n$ be i.i.d. with common p.d.f.

$$f_{\underline{\theta}}(x) = \frac{1}{\Gamma\alpha\, \mu^{\alpha}} e^{-\frac{x}{\mu}} x^{\alpha-1}, \quad x>0, \quad \underline{\theta} = (\mu, \alpha) \in (0, \infty)^2.$$

Then

$$m_1(\underline{\theta}) = E_{\underline{\theta}}(X_1) = \alpha\mu$$

$$m_2(\underline{\theta}) = E_{\underline{\theta}}(X_i^2) = \alpha(\alpha+1)\mu^2$$

Clearly $m_1(\underline{\theta})$ and $m_2(\underline{\theta})$ are continuous functions of $\underline{\theta}$

and

$$\left| \frac{\partial(m_1(\underline{\theta}), m_2(\underline{\theta}))}{\partial(\alpha, \mu)} \right| = \begin{vmatrix} \mu & \alpha \\ (2\alpha+1)\mu^2 & 2\alpha(\alpha+1)\mu \end{vmatrix} = \alpha\mu^2 > 0$$

Thus the MME $\hat{\underline{\theta}}(\underline{x}) = (\hat{\alpha}(\underline{x}), \hat{\mu}(\underline{x}))$ is consistent

$\hat{\alpha}\ \hat{\mu} =$

$$m_1(\hat{\alpha}, \hat{\mu}) = A_1(\underline{x}) = \overline{X} \qquad \cdots \quad (\mathrm{I})$$

$$m_2(\hat{\alpha}, \hat{\mu}) = \hat{\alpha}(\hat{\alpha}+1)\hat{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \qquad \cdots \quad (\mathrm{II})$$

Solving (I) and (II), we get

$$\overline{X}^2 + \overline{X}\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \qquad \&$$

$$\Rightarrow \quad \hat{\mu} = \frac{S^2}{\overline{X}}, \quad \text{where } S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

and $\hat{\alpha} = \dfrac{\overline{X}^2}{S^2}$

Clearly, as $n \to \infty$,

$$\hat{\mu} = \frac{S^2}{\overline{X}} \xrightarrow{p} \frac{m_2(\underline{\theta}) - (m_1(\underline{\theta}))^2}{m_1(\underline{\theta})} = \frac{\alpha(\alpha+1)\mu^2 - \alpha^2\mu^2}{\alpha\mu} = \mu$$

$$\hat{\alpha} = \frac{\overline{X}^2}{S^2} \xrightarrow{p} \frac{m_1^2(\underline{\theta})}{m_2^2(\underline{\theta}) - m_1^2(\underline{\theta})} = \frac{\alpha^2\mu^2}{\alpha(\alpha+1)\mu^2 - \alpha^2\mu^2} = \alpha.$$

Example (MME may not be a function of minimal sufficient statistic and may be inadmissible)

Let $X_1, X_2, \ldots, X_n$ be iid $U(0, \theta)$, $\theta \in \Theta = (0, a)$. We have

$$m_1(\theta) = \frac{\theta}{2}.$$

The MME $\hat{\theta}(\underline{X})$ is given by

$$\frac{\hat{\theta}(\underline{X})}{2} = \bar{X}$$

$$\Rightarrow \hat{\theta}(\underline{X}) = 2\bar{X},$$

which is not a function of minimal sufficient statistic $T(\underline{X}) = X_{(n)}$. Under any loss function $L(\theta, a)$ that is strictly convex in $a \in \mathcal{A} = (0, a)$, for every $\theta \in \Theta$, the estimator

$$\delta_0(\underline{X}) = E_\theta(\hat{\theta}(\underline{X}) \mid X_{(n)})$$

dominates MME $\hat{\theta}(\underline{X}) = 2\bar{X}$.

$$\delta_0(\underline{X}) = E_\theta(2\bar{X} \mid X_{(n)})$$

$$= 2 E_\theta(X_1 \mid X_{(n)})$$

$$= 2 E_\theta\left(\frac{X_1}{X_{(n)}} \cdot X_{(n)} \mid X_{(n)}\right) \quad \text{(Basu's Thm)}$$

$$= 2 X_{(n)} E_\theta\left(\frac{X_1}{X_{(n)}} \mid X_{(n)}\right) = 2 X_{(n)} E_\theta\left(\frac{X_1}{X_{(n)}}\right)$$

$$E_\theta(X_1) = E_\theta\left(\frac{X_1}{X_{(n)}} \cdot X_{(n)}\right)$$

$$= E_\theta\left(\frac{X_1}{X_{(n)}}\right) E_\theta(X_{(n)}) \quad \text{(Basu's Thm)}$$

$$E_\theta\left(\frac{X_1}{X_{(n)}}\right) = \frac{E_\theta(X_1)}{E_\theta(X_{(n)})} = \frac{\theta/2}{\frac{n}{n+1}\theta} = \frac{n+1}{2n}$$

$$\delta_0(\underline{X}) = \frac{n+1}{n} X_{(n)}$$

$$\hat{\theta}(\underline{X}) = 2\bar{X} \xrightarrow{p} \theta.$$

$$\sqrt{n}(\hat{\theta}(\underline{X}) - \theta) = \sqrt{n}(2\bar{X} - \theta)$$

$$= 2\sqrt{n}\left(\bar{X} - \frac{\theta}{2}\right)$$

$E_\theta(X_1) = \frac{\theta}{2}$

$E_\theta(X_1^2) = \frac{\theta^2}{3}$, $\quad V_\theta(X_1) = \frac{\theta^2}{3} - \left(\frac{\theta}{2}\right)^2 = \frac{\theta^2}{12}$

$$\sqrt{n}\left(\bar{X} - \frac{\theta}{2}\right) \xrightarrow{d} Z \sim N\left(0, \frac{\theta^2}{12}\right)$$

$$\Rightarrow \sqrt{n}\left(\hat{\theta}(\underline{X}) - \theta\right) = 2\sqrt{n}\left(\bar{X} - \frac{\theta}{2}\right) \xrightarrow{d} 2Z \sim N\left(0, \frac{\theta^2}{3}\right)$$

$E_\theta(X_{(n)}) = \frac{n}{n+1}\theta$, $\quad E_\theta(X_{(n)}^2) = \frac{n}{n+2}\theta^2$

$$V_\theta(X_{(n)}) = \left(\frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2\right)\theta^2 = \frac{n\theta^2}{(n+1)(n+2)} \to 0$$

$$E_\theta(X_{(n)}) \to \theta$$

$$\Rightarrow \quad X_{(n)} \xrightarrow{P} \theta$$

$$\Rightarrow \quad \delta_0(\underline{X}) \xrightarrow{P} \theta.$$

Example Let $X_1, X_2, \ldots X_n$ be iid $N(0, \sigma^2)$ where $\sigma \in (0, \infty) = \Theta$ is unknown. Here

$$m_1(\sigma) = E_\sigma(X_1) = 0$$

and

$$0 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

does not have any solution in σ. Thus MME does not exist.

Here we may take

$$m_2(\sigma) = E_\sigma(X_i^2) = \sigma^2$$

and take the modified MME $\hat{\sigma}^2(\underline{X})$ as

$$\hat{\sigma}^2(\underline{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i^2,$$

which is a function of minimal sufficient statistic.

$$T(\underline{X}) = \sum_{i=1}^{n} X_i^2.$$

$E(X_i^2) = \sigma^2$, $\quad V(X_i^2) = E(X_i^4) - (E(X_i^2))^2 = 2\sigma^4$

$$\sqrt{n}\left(\hat{\sigma}^2(\underline{X}) - \sigma^2\right) \xrightarrow{d} N(0, 2\sigma^4)$$

$$\hat{\sigma}^2(\underline{X}) \xrightarrow{P} \sigma^2.$$

Assignment Problems

(1) Let $X_1, X_2, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$, where $\underline{\theta} = (\mu, \sigma) \in \mathbb{R} \times (0, \infty)$ is unknown. Find the MLE $\hat{\underline{\theta}}(\underline{X}) = (\hat{\mu}(\underline{X}), \hat{\sigma}^2(\underline{X}))$ of $\underline{\theta} = (\mu, \sigma)$. Show that $\hat{\mu}(\underline{X})$ and $\hat{\sigma}^2(\underline{X})$ are CAN estimators of μ and σ.

(2) Let $X_1, X_2, \ldots, X_n$ be i.i.d with common pdf

$$f_\theta(x) = \frac{1}{2\theta} e^{-\frac{|x|}{\theta}}, \quad -\infty < x < \infty$$

where $\theta \in \Theta = (0, \infty)$ is unknown. Show that MME does not exist. Find modified MME of θ and show that it is CAN.

(3) Repeat problem 2 with

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

where $\theta \in \Theta = (0, \infty)$ is unknown. Also find the MME of $g(\theta) = \frac{1}{\theta^2}$ and find its asymptotic distribution. Is it CAN?

(4) Let $X_1, X_2, \ldots, X_n$ be a random sample from a population having pdf

$$g_{\underline{\theta}}(x) = \frac{1}{\sigma} e^{-\frac{1}{\sigma}(x-\mu)}, \quad x > \mu,$$

where $\underline{\theta} = (\mu, \sigma) \in \Theta = \mathbb{R} \times (0, \infty)$ is unknown. Find the MME of $\underline{\theta}$ and $g(\underline{\theta}) = \mu + \sigma$. Are they CAN?

(5) Let $X_1, X_2, \ldots, X_n$ be a random sample from $B(\theta_1, \theta_2)$, where $\underline{\theta} = (\theta_1, \theta_2) \in \Theta = (0, \infty)^2$ is unknown. Find MMEs of θ_1 and θ_2 and show that they are CAN.

(6) Repeat Problem 5 with $B(\theta_1, \theta_2)$ replaced by $U(\theta_1, \theta_2)$, where $\underline{\theta} = (\theta_1, \theta_2) \in \Theta = \{(x, y) \in \mathbb{R}^2 : -\infty < x \le y < \infty\}$

(7) (MME may not be unique). Let $X_1, X_2, \ldots, X_n$ be a random sample from the population with pdf

$$f_\theta(x) = \frac{1}{\theta^2} e^{-\frac{x}{\theta^2}}, \quad x > 0,$$

where $\theta \in \Theta = \mathbb{R}$ is unknown. Show that MME is not unique.

(8) (MME may be absurd). Let $X_1, X_2, \ldots, X_n$ be a random sample from $Bin(n, p)$ where $\underline{\theta} = (n, p) \in \{1, 2, \ldots\} \times (0, 1) = \Theta$ is unknown. Find the MME $\hat{n}(\underline{X})$ of n and show that it may be absurd.

(9) Let $X_1, X_2, \ldots, X_n$ be i.i.d. with

$$P_\theta(X_1 = 1) = \frac{2(1-\theta)}{2-\theta} = 1 - P_\theta(X_1 = 2),$$

where $\theta \in \Theta = (0, 1)$ is unknown. Find the MME of θ and show that it is CAN.

Maximum Likelihood Estimators and their Consistency.

Let $\underline{X_1}, \ldots, \underline{X_n}$ be a random sample from a population having d.f. $F_{\underline{\theta}}$, where $\underline{\theta} \in \Theta \subseteq \mathbb{R}^k$ is unknown. Let $\underline{\theta}_0 \in \Theta$ be the true value of $\underline{\theta} \in \Theta$. Let $f_{\underline{\theta}}$ denote the pmf/pdf associated with $F_{\underline{\theta}}$, $\underline{\theta} \in \Theta$. The joint p.m.f./p.d.f. of $\underline{X} = (\underline{X_1}, \ldots, \underline{X_n})$ is

$$g_\theta(\underline{x}) = \prod_{i=1}^{n} f_{\underline{\theta}}(\underline{x_i}), \quad \underline{x} = (x_1, \ldots, x_n) \in \mathbb{R}^{mn}$$

Definition For a observed value $\underline{x}$ of the sample $\underline{X}$, the function

$$L_{\underline{x}}(\underline{\theta}) = \prod_{i=1}^{n} f_{\underline{\theta}}(\underline{x_i}),$$

as a function of $\underline{\theta} \in \Theta$, is called the likelihood function of the observed sample $\underline{x}$.

For given observed sample $\underline{x}$, $L_{\underline{x}}(\underline{\theta})$ is the likelihood of the observed sample $\underline{x}$ coming from $F_{\underline{\theta}}$, $\theta \in \Theta$. If the true value $\underline{\theta}_0$ of $\underline{\theta}$ is unknown to us then a natural estimate of $\underline{\theta}_0$ is that value $\hat{\underline{\theta}}(\underline{x})$ of $\underline{\theta}$ for which $F_{\underline{\theta}}$ is most likely to have produced the observed sample $\underline{x}$, i.e. $\hat{\underline{\theta}}(\underline{x})$ is that value of $\underline{\theta}$ which maximizes the chances of getting the observed sample $\underline{x}$.

Definition For a observed sample $\underline{x}$, let $\hat{\underline{\theta}} = \hat{\underline{\theta}}(\underline{x})$ be such that

$$L_{\underline{x}}(\hat{\underline{\theta}}) = \sup_{\theta \in \Theta} L_{\underline{x}}(\underline{\theta}).$$

Then $\hat{\underline{\theta}}(\underline{x})$ is called the maximum likelihood estimate (MLE) of θ and $\hat{\underline{\theta}}(\underline{X})$ (a random variable) is called maximum likelihood estimator (MLE) of $\underline{\theta}$.

For observed sample $\underline{x}$, the MLE maximizes the likelihood function $L_{\underline{x}}(\underline{\theta})$, $\theta \in \Theta$. Since the maximization of $L_{\underline{x}}(\underline{\theta})$ is equivalent to maximization of $l_{\underline{x}}(\underline{\theta}) = \ln L_{\underline{x}}(\underline{\theta})$, one may obtain MLE by maximizing

$$l_{\underline{x}}(\underline{\theta}) = \ln L_{\underline{x}}(\underline{\theta}) = \sum_{i=1}^{n} \ln f_{\underline{\theta}}(x_i), \quad \underline{\theta} \in \Theta$$

The function $l_{\underline{x}}(\underline{\theta})$ is called the log-likelihood function. Usually it is much easier to work with the log-likelihood function $l_{\underline{x}}(\underline{\theta})$.

When, for given observed sample $\underline{x}$, $l_{\underline{x}}(\underline{\theta})$ is a differentiable function of $\underline{\theta}$, MLE, if it exists, is a critical point i.e. a solution of the simultaneous equations

$$\frac{\partial}{\partial \theta_i} l_{\underline{x}}(\underline{\theta}) = 0, \quad i = 1, 2, \ldots, k.$$

<u>Remark:</u> (a) The MLE may not be a critical point and it may be on boundary of Θ. Let $X_1, X_2, \ldots, X_n$ be i.i.d. $N(\theta, 1)$, where $\theta \in \Theta = [0, \infty)$ is unknown. Then

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

$$\frac{d}{d\theta} l_{\underline{x}}(\theta) = \sum_{i=1}^{n} (x_i - \theta) = n(\bar{x} - \theta)$$

Thus $l_{\underline{x}}(\theta)$ is $\uparrow$ in $(-\infty, \bar{x})$ and $\downarrow$ in $(\bar{x}, \infty)$. Since $\theta \in \Theta = [0, \infty)$, the MLE is

$$\hat{\theta}(\underline{x}) = \begin{cases} \bar{x}, & \text{if } \bar{x} > 0 \\ 0, & \text{if } \bar{x} \leq 0 \end{cases}$$

Here for $\bar{x} \leq 0$, $\hat{\theta}(\underline{x})$ belongs to the boundary of $\Theta = [0, \infty)$

(b) The MLE is a function of minimal sufficient statistic (using factorization theorem) but it itself may not be a sufficient statistic. Let $X_1, X_2, \ldots, X_n$ be i.i.d. $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$ where $\theta \in \Theta = \mathbb{R}$. Then, for given sample $\underline{x}$,

$$L_{\underline{x}}(\theta) = \begin{cases} 1, & x_{(n)} - \frac{1}{2} \leq \theta \leq x_{(1)} + \frac{1}{2} \\ 0, & \text{otherwise.} \end{cases}$$

is constant on $[x_{(n)} - \frac{1}{2}, x_{(1)} + \frac{1}{2}]$.
Thus $\hat{\theta}_1(\underline{x}) = \frac{x_{(1)} + x_{(n)}}{2}$ $\left(x_{(n)} - \frac{1}{2} \leq \frac{x_{(1)} + x_{(n)}}{2} \leq x_{(1)} + \frac{1}{2}\right)$ is the MLE. In general, for any $\alpha \in [0, 1]$ $\hat{\theta}_2(\underline{x}) = \alpha(x_{(n)} - \frac{1}{2}) + (1-\alpha)(x_{(1)} + \frac{1}{2})$ is the MLE. Clearly $\hat{\theta}_1(\underline{X})$ is a function of minimal sufficient statistic $T(\underline{X}) = (X_{(1)}, X_{(n)})$ but $\hat{\theta}_1(\underline{X})$ is not minimal sufficient.

(c) The example considered in (b) above suggests that the MLE may not be unique.

(d) Even when the log-likelihood function $l_{\underline{x}}(\theta)$ is smooth and the MLE is in the interior of Θ, the simultaneous equations

$$\frac{\partial}{\partial \theta_i} l_{\underline{x}}(\underline{\theta}) = 0, \quad i = 1, \ldots, k$$

may have multiple roots (points of local maxima/minima) and various roots have to be checked for global maxima.

(d) For observed sample $\underline{x}$ let $l_{\underline{x}}(\underline{\theta})$ be a differentiable function of $\underline{\theta}$ on $\Theta_{\underline{x}} = \{\underline{\theta} \in \Theta : L_{\underline{x}}(\underline{\theta}) > 0\}$. Define

$$H_{\underline{x}}(\underline{\theta}) = \left(\left(\frac{\partial^2 l_{\underline{x}}(\underline{\theta})}{\partial\theta_i \partial\theta_j} \right)\right), \quad \underline{\theta} \in \Theta$$

Let $H_{\underline{x}}^{(k)}(\underline{\theta})$ be the matrix obtained by deleting last $(n-k)$ rows and last $(n-k)$ columns of $H_{\underline{x}}(\underline{\theta})$, $k = 1, 2, \ldots, n$

Let $\hat{\underline{\theta}}(\underline{x}) = (\hat{\theta}_1(\underline{x}), \ldots, \hat{\theta}_k(\underline{x}))$ be a critical point of $l_{\underline{x}}(\underline{\theta})$,

i.e.

$$\frac{\partial}{\partial\theta_i} l_{\underline{x}}(\underline{\theta}) \Big|_{\underline{\theta} = \hat{\underline{\theta}}(\underline{x})} = 0, \quad i = 1, 2, \ldots, k.$$

Then

(a) $\hat{\underline{\theta}}(\underline{x})$ is a local minimum if $H_{\underline{x}}(\hat{\underline{\theta}}(\underline{x}))$ is positive definite, i.e. all eigen values of $H_{\underline{x}}(\hat{\underline{\theta}}(\underline{x}))$ are positive or equivalently
$\det(H_{\underline{x}}^{(k)}(\hat{\underline{\theta}})) > 0, \quad k = 1, 2, \ldots, n,$

(b) $\hat{\underline{\theta}}(\underline{x})$ is a local maximum if $H_{\underline{x}}(\hat{\underline{\theta}}(\underline{x}))$ is negative definite, i.e. eigen values of $H_{\underline{x}}(\hat{\underline{\theta}}(\underline{x}))$ are negative or equivalently
$(-1)^k \det(H_{\underline{x}}^{(k)}(\hat{\underline{\theta}}(\underline{x}))) > 0 \quad k = 1, 2, \ldots, n$

(c) $\hat{\underline{\theta}}(\underline{x})$ is a saddle point if $H_{\underline{x}}(\hat{\underline{\theta}}(\underline{x}))$ has both positive and negative eigen values but $|H_{\underline{x}}^{(k)}(\hat{\underline{\theta}}(\underline{x}))| \neq 0$, $k = 1, \ldots, n$

(d) the test is inconclusive if $|H_{\underline{x}}(\hat{\underline{\theta}}(\underline{x}))| = 0$.

For a given function $g : \Theta \to \mathbb{R}^p$, let the estimand be the function $\eta = g(\underline{\theta})$, $\underline{\theta} \in \Theta$. For observed sample $\underline{x}$

$$L^*_{\underline{x}}(\eta) = \sup_{\{\underline{\theta} \in \Theta : g(\underline{\theta}) = \eta\}} L_{\underline{x}}(\underline{\theta}), \quad \eta \in g(\Theta) = \{g(\underline{\theta}) : \underline{\theta} \in \Theta\} = \Lambda,$$

is called the likelihood function induced by $\eta = g(\underline{\theta})$.

Definition For given sample observation $\underline{x}$, let $\hat{\eta} \equiv \hat{\eta}(\underline{x})$ be such that

$$L_{\underline{x}}^{*}(\hat{\eta}(\underline{x})) = \sup_{\eta \in \Lambda} L_{\underline{x}}(\eta).$$

Then $\hat{\eta}(\underline{x})$ is called the maximum likelihood estimate of η and $\hat{\eta}(\underline{X})$ is called the maximum likelihood estimator η.

Theorem If, for given sample observation $\underline{x}$, $\hat{\underline{\theta}}(\underline{x})$ is MLE of $\underline{\theta}$ then $g(\hat{\underline{\theta}}(\underline{x}))$ is MLE of $\eta = g(\underline{\theta})$.

Proof. Let $\eta_0 = g(\hat{\theta})$ be the true value of η. For given sample observation $\underline{x}$

$$L_{\underline{x}}^{*}(\hat{\eta}(\underline{x})) = \sup_{\eta \in \Lambda} L_{\underline{x}}^{*}(\eta)$$

$$= \sup_{\eta \in \Lambda} \sup_{\{\underline{\theta} \in \Theta : g(\underline{\theta}) = \eta\}} L_{\underline{x}}(\underline{\theta})$$

$$\geq \sup_{\{\underline{\theta} \in \Theta : g(\underline{\theta}) = \eta_0\}} L_{\underline{x}}(\underline{\theta})$$

$$= L_{\underline{x}}^{*}(\eta_0).$$

Consider the following assumptions:

A1: For $\underline{\theta}_1, \underline{\theta}_2 \in \Theta$, $F_{\underline{\theta}_1} \equiv F_{\underline{\theta}_2}$ if and only if $\underline{\theta}_1 = \underline{\theta}_2$ (i.e. the family $\mathcal{P} = \{F_{\underline{\theta}} : \underline{\theta} \in \Theta\}$ is identifiable;

A2: The d.f.s $F_{\underline{\theta}}$, $\underline{\theta} \in \Theta$ have the same support (distributions with supports depending on parameter $\underline{\theta} \in \Theta$ are out of consideration).

A3: Θ is an open set.

The following theorem provides a theoretical justification for maximizing the likelihood function (equivalently the log-likelihood function) to obtain a reasonable estimate of $\underline{\theta}$.

Theorem (a) Suppose that assumptions A_1 and A_2 hold. Then, for any $\underline{\theta} \neq \underline{\theta}_0$ ($\underline{\theta}, \theta_0 \in \Theta$)

$$E_{\theta_0}(l_{\underline{x}}(\theta_0)) > E_{\theta_0}(l_{\underline{x}}(\theta))$$

(b) For any $\underline{\theta} \neq \underline{\theta}_0$,

$$\lim_{n\to\infty} P_{\theta_0}(l_{\underline{x}}(\underline{\theta}_0) > l_{\underline{x}}(\theta)) = 1.$$

Proof. (a) Consider, for $\underline{\theta} \neq \theta_0$,

$$l_{\underline{x}}(\theta) - l_{\underline{x}}(\underline{\theta}_0) = \ln \frac{L_{\underline{x}}(\theta)}{L_{\underline{x}}(\theta_0)}$$

By the Jensen inequality

$$E_{\theta_0}(l_{\underline{x}}(\theta) - l_{\underline{x}}(\theta_0)) = E_{\theta_0}\left(\ln \frac{L_{\underline{x}}(\theta)}{L_{\underline{x}}(\theta_0)}\right)$$

$$\leq \ln E_{\theta_0}\left(\frac{L_{\underline{x}}(\theta)}{L_{\underline{x}}(\theta_0)}\right)$$

$$= \ln \int \frac{g_{\underline{\theta}}(\underline{x})}{g_{\theta_0}(\underline{x})} g_{\theta_0}(\underline{x})\, d\mu$$

$$= \ln \int g_{\underline{\theta}}(\underline{x})\, d\mu = 0.$$

and the equality is attained if and only if

$$P_{\theta_0}(L_{\underline{x}}(\underline{\theta}) = L_{\underline{x}}(\theta_0)) = 1,$$

which is not the case by assumption (A1).

(b) For $\underline{\theta} \neq \theta_0$,

$$T(\underline{x}) = l_{\underline{x}}(\theta) - l_{\underline{x}}(\theta_0)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \ln \frac{f_\theta(x_i)}{f_{\theta_0}(x_i)}$$

$$\xrightarrow{a.s.} E_{\theta_0}\left(\ln \frac{f_\theta(x_1)}{f_{\theta_0}(x_1)}\right) < \ln E_{\theta_0}\left(\frac{f_\theta(x_1)}{f_{\theta_0}(x_1)}\right) = 0$$

Thus

$$\lim_{n\to\infty} P_{\theta_0}(l_{\underline{x}}(\theta) < l_{\underline{x}}(\theta_0)) = P_{\theta_0}(T(\underline{x}) < 1) = 1.$$

Remark: The above theorem suggests that, under conditions A1 and A2, the MLE $\hat{\theta}(\underline{x})$ of θ is close to the true value of $\underline{\theta}_0$ and hence is a reasonable estimator.

For the case $\theta \in \Theta \subseteq \mathbb{R}$ we have the following two theorems.

Theorem Suppose that assumptions A_1 A_2 and A_3 hold and that, for almost all x, $f_\theta(x)$ is a differentiable function of $\theta \in \Theta \subseteq \mathbb{R}$. Then there exist a sequence $\{\hat{\theta}_n^*(\underline{x})\}_{n\geq 1}$ such that

(a) $P_{\theta_0}(\hat{\theta}_n^*(\underline{x})$ is a root of the equation $l_{\underline{x}}'(\theta)=0) \to 1$, as $n\to\infty$

(b) $P_{\theta_0}(\hat{\theta}_n^*(\underline{x})$ is a local maxima of $l_{\underline{x}}(\underline{\theta})) \to 1$, as $n\to\infty$

(c) $\hat{\theta}_n^*(\underline{x}) \xrightarrow{P} \theta_0$.

Proof. Consider arbitrary $\varepsilon>0$. Since Θ is open and $\theta_0\in\Theta$, there exists $m_0\in\mathbb{N}$ s.t. $\frac{1}{m_0}<\varepsilon$ and $(\theta_0-\frac{1}{m_0}, \theta_0+\frac{1}{m_0})\subseteq\Theta$. Then $(\theta_0-\frac{1}{m}, \theta_0+\frac{1}{m})\subseteq\Theta$, $\forall\, m\geq m_0$. For $m\geq m_0$ and $n\in\mathbb{N}$, define

$$R_{m,n}=\{\underline{x}\in\mathbb{R}^n : l_{\underline{x}}(\theta_0)>l_{\underline{x}}(\theta_0-\tfrac{1}{m}) \text{ and } l_{\underline{x}}(\theta_0)>l_{\underline{x}}(\theta_0+\tfrac{1}{m})\} \quad \cdots \text{(I)}$$

Using last lemma, we have, for any $m\geq m_0$,

$$\lim_{n\to\infty} P_{\theta_0}(l_{\underline{x}}(\theta_0)>l_{\underline{x}}(\theta_0-\tfrac{1}{m}))=1 \text{ and } \lim_{n\to\infty} P_{\theta_0}(l_{\underline{x}}(\theta_0)>l_{\underline{x}}(\theta_0+\tfrac{1}{m}))=1$$

$$\Rightarrow P_{\theta_0}(R_{m,n})\geq P_{\theta_0}(l_{\underline{x}}(\theta_0)>l_{\underline{x}}(\theta_0-\tfrac{1}{m}))+P_{\theta_0}(l_{\underline{x}}(\theta_0)>l_{\underline{x}}(\theta_0+\tfrac{1}{m}))-1 \quad (P(A\cap B)\geq P(A)+P(B)-1)$$

$\cdots$ (II)

$$\Rightarrow \lim_{n\to\infty} P_{\theta_0}(R_{m,n})=1$$

Fix $\underline{x}\in R_{m,n}$. Since $l_{\underline{x}}(\underline{\theta})$ is continuous on closed and bounded interval $[\theta_0-\frac{1}{m}, \theta_0+\frac{1}{m}]$, by virtue of (I), $l_{\underline{x}}(\underline{\theta})$ attains local maximum inside $[\theta_0-\frac{1}{m}, \theta_0+\frac{1}{m}]$, i.e. $\exists\ \hat{\theta}_{m,n}(\underline{x})\in(\theta_0-\frac{1}{m}, \theta_0+\frac{1}{m})$ such that

$$l_{\underline{x}}'(\hat{\theta}_{m,n}(\underline{x}))=0$$

Since $\frac{1}{m}<\varepsilon$, $\forall\, m\geq m_0$, we have

$$P_{\theta_0}(|\hat{\theta}_{m,n}(\underline{x})-\theta_0|<\varepsilon)\geq P_{\theta_0}(|\hat{\theta}_{m,n}(\underline{x})-\theta_0|<\tfrac{1}{m})$$
$$\geq P_{\theta_0}(\underline{X}\in R_{m,n})$$

Using (II) we get

$$\lim_{n\to\infty} P_{\theta_0}\left(|\hat{\theta}_{m,n}(\underline{x}) - \theta_0| < \varepsilon\right) = 1, \quad \forall\, m \geq m_0$$

However $\hat{\theta}_{m,n}$ depends on m (and hence $\varepsilon > 0$).

Fix $\underline{x} \in R_{m,n}$ and define

$$M_n(\underline{x}) = \inf\{ |\hat{\theta}_n(\underline{x}) - \theta_0| : l'_{\underline{x}}(\hat{\theta}_n(\underline{x})) = 0 \}$$

Clearly $M_n(\underline{x})$ exists and $\exists$ a sequence $\{\tilde{\hat{\theta}}_{n,k}(\underline{x})\}_{k\geq 1}$ such that

$l'_{\underline{x}}(\tilde{\hat{\theta}}_{n,k}(\underline{x})) = 0$ and

$$M_n(\underline{x}) = \lim_{k\to\infty} |\tilde{\hat{\theta}}_{n,k}(\underline{x}) - \theta_0|$$

Let

$$\hat{\theta}^*_n(\underline{x}) = \lim_{k\to\infty} \tilde{\hat{\theta}}_{n,k}(\underline{x})$$

Then $\lim_{n\to\infty} P_{\theta_0}(\underline{x} \in R_{m,n}) = 1$ and thus

$$P_{\theta_0}\left(|\hat{\theta}^*_n(\underline{x}) - \theta_0| < \varepsilon\right) \geq P_{\theta_0}\left(\lim_{k\to\infty} |\tilde{\theta}_{n,k}(\underline{x}) - \theta_0| < \varepsilon\right)$$

$$= P_{\theta_0}\left(M_n(\underline{x}) < \varepsilon\right)$$

$$\geq P_{\theta_0}\left(|\hat{\theta}_{m,n}(\underline{x}) - \theta_0| < \varepsilon\right)$$

$\Rightarrow$

$$\Rightarrow \lim_{n\to\infty} P_{\theta_0}\left(|\hat{\theta}^*_n(\underline{x}) - \theta_0| < \varepsilon\right) = 1$$

Also, for any $\underline{x} \in R_{m,n}$

$$l'_{\underline{x}}(\tilde{\theta}_{n,k}(\underline{x})) = 0, \quad \forall\, k = 1, 2, \ldots$$

$$\Rightarrow \lim_{k\to\infty} l'_{\underline{x}}(\tilde{\hat{\theta}}_{n,k}(\underline{x})) = 0$$

$$\Rightarrow l'_{\underline{x}}\left(\lim_{k\to\infty} \tilde{\hat{\theta}}_{n,k}(\underline{x})\right) = 0$$

$$\Rightarrow l'_{\underline{x}}(\hat{\theta}^*_n(\underline{x})) = 0.$$

[derivative can not have discontinuity of first kind i.e. if $\lim_{x\to p+} h'(x)$ and $\lim_{x\to p-} h'(x)$ exist then h' is continuous at p]

Remark: (a) In case of multiple roots of the equation $l'_{\underline{x}}(\theta) = 0$ the above theorem does not say which sequence of roots of the equation $l'_{\underline{x}}(\underline{\theta}) = 0$ should be chosen to ensure consistency.

(b) The above theorem does not guarantee that for any given n, however large, the likelihood function has any local maxima at all.

(c) If for every n and $\underline{x}$, the likelihood equation $l'_{\underline{x}}(\theta)=0$ has a unique root then the above theorem guarantees that it will be consistent and with probability tending to 1 will maximize the likelihood (i.e. will be MLE) (See the Corollary stated below).

(d) Under certain conditions a result similar to the above result also holds when $\theta \in \Theta \subseteq \mathbb{R}^k$ $(k \geq 2)$

<u>Corollary</u> Under the assumptions of last theorem suppose that the equation $l'_{\underline{x}}(\theta)=0$ has unique root $\hat{\theta}^*_n(\underline{x})$ for each n and $\underline{x}$. Then $\{\hat{\theta}^*_n(\underline{x})\}_{n\geq 1}$ is consistent for θ. Moreover with probability tending to 1, $\{\hat{\theta}^*_n\}$ maximizes the likelihood (i.e. $\hat{\theta}^*_n$ is the MLE) and

<u>Theorem</u> Suppose that assumptions $A_1 - A_3$ hold. In addition suppose that

(a) for every $x \in \mathcal{X}^* = \{x \in \mathbb{R} : f_\theta(x) > 0\}$, $f_\theta(x)$ is ~~twice~~ thrice differentiable w.r.t. θ and the third derivative is continuous in $\theta \in \Theta$;

(b) $\int \frac{\partial^3}{\partial\theta^3} f_\theta(x)\,dx = \frac{\partial^2}{\partial\theta^3} \int f_\theta(x)\,dx = 0, \ \forall\, \theta \in \Theta$;

(c) $I(\theta) = E_\theta\left(\left(\frac{\partial}{\partial\theta} \ln f_\theta(X_1)\right)^2\right) \in (0, \infty) \quad \forall\, \theta \in \Theta$

(d) for any $\theta^* \in \Theta$, $\exists$ $c > 0$ and a function $M(x)$ (c and $M(x)$ may depend on θ^*) such that

$$\left|\frac{\partial^3}{\partial\theta^3} \ln f_\theta(x)\right| \leq M(x) \quad \forall\ x \in \mathcal{X}^*,\ \theta \in (\theta^*-c, \theta^*+c)$$

and

$$E_{\theta_0}(M(X_1)) < \infty.$$

Let $\{\hat{\theta}^*_n(\underline{X})\}_{n\geq 1}$ be any consistent sequence of roots of $l'_{\underline{x}}(\theta)=0$. Then

$$\sqrt{n}\,(\hat{\theta}^*_n(\underline{X}) - \theta_0) \xrightarrow{d} N\left(0, \frac{1}{I(\theta_0)}\right),$$

i.e. $\hat{\theta}_n^*$ is CAN. and asymptotically efficient.

<u>Proof.</u> For observed sample point $\underline{x}$, using Taylor's series expansion, we have

$$0 = l'_{\underline{x}}(\hat{\theta}_n^*(\underline{x})) = l'_{\underline{x}}(\theta_0) + (\hat{\theta}_n^*(\underline{x})-\theta_0)\, l''_{\underline{x}}(\theta_0) + \frac{(\hat{\theta}_n^*(\underline{x})-\theta_0)^2}{2}\, l'''_{\underline{x}}(\hat{\theta}_n^{**})$$

where $\hat{\theta}_n^{**}$ lies between θ_0 and $\hat{\theta}_n^*(\underline{x})$

$$\sqrt{n}\,(\hat{\theta}_n^*(\underline{x})-\theta_0) = \frac{-\sqrt{n}\, l'_{\underline{x}}(\theta_0)}{l''_{\underline{x}}(\theta_0) + \frac{(\hat{\theta}_n^*-\theta_0)}{2}\, l'''_{\underline{x}}(\hat{\theta}_n^{**})} = \frac{-\, l'_{\underline{x}}(\theta_0)/\sqrt{n}}{\frac{1}{n} l''_{\underline{x}}(\theta_0) + \frac{(\hat{\theta}_n^*-\theta_0)}{2n}\, l'''_{\underline{x}}(\hat{\theta}_n^{**})}$$

Since $E_{\theta_0}\left(\frac{\frac{\partial}{\partial\theta_0} f_{\theta_0}(x_1)}{f_{\theta_0}(x_1)}\right) = 0$, we have

$$\frac{1}{\sqrt{n}}\, l'_{\underline{x}}(\theta_0) = \frac{1}{\sqrt{n}}\, \frac{\partial}{\partial\theta_0}\left(\sum_{i=1}^n \ln f_{\theta_0}(x_i)\right)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{f'_{\theta_0}(x_i)}{f_{\theta_0}(x_i)}$$

$$= \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^n \frac{f'_{\theta_0}(x_i)}{f_{\theta_0}(x_i)} - E_{\theta_0}\left(\frac{f'_{\theta_0}(x_1)}{f_{\theta_0}(x_1)}\right)\right]$$

$$\xrightarrow{d} N(0,\ I(\theta_0)),$$

where $I(\theta_0) = Var_{\theta_0}\left(\frac{f'_{\theta_0}(x_1)}{f_{\theta_0}(x_1)}\right) = Var_{\theta_0}\left(\frac{\partial}{\partial\theta_0}\ln f_{\theta_0}(x_1)\right)$

$$\frac{1}{n}\, l''_{\underline{x}}(\theta_0) = \frac{1}{n}\sum_{i=1}^n \frac{f_{\theta_0}(x_i)\, f''_{\theta_0}(x_i) - (f'_{\theta_0}(x_i))^2}{f^2_{\theta_0}(x_i)}$$

$$= \frac{1}{n}\sum_{i=1}^n \frac{f''_{\theta_0}(x_i)}{f_{\theta_0}(x_i)} - \frac{1}{n}\sum_{i=1}^n\left(\frac{f'_{\theta_0}(x_i)}{f_{\theta_0}(x_i)}\right)^2$$

$$\xrightarrow{as} E_{\theta_0}\left(\frac{f''_{\theta_0}(x_1)}{f_{\theta_0}(x_1)}\right) - E_{\theta_0}\left(\left(\frac{f'_{\theta_0}(x_1)}{f_{\theta_0}(x_1)}\right)^2\right)$$

$$= -E_{\theta_0}\left(\left[\frac{\partial}{\partial\theta_0}\ln f_{\theta_0}(x_1)\right]^2\right) = -I(\theta_0)$$

$$\left|\frac{1}{n} l_{\underline{x}}'''(\hat{\theta}_n^{**})\right| \leq \frac{1}{n}\left|\frac{\partial^3}{\partial\theta^3}\left(\sum_{i=1}^{n} \ln f_\theta(x_i)\right)\right|_{\theta=\hat{\theta}_n^{**}}$$

$$\leq \frac{\sum_{i=1}^{n} M(x_i)}{n}$$

Also, $\hat{\theta}_n^{*} \xrightarrow{P} \theta_0 \Rightarrow \hat{\theta}_n^{**} \xrightarrow{a.s.} \theta_0$, and

$$\frac{1}{n}\sum_{i=1}^{n} M(x_i) = E_{\theta_0}(M(x_1))$$

Thus

$$\frac{(\hat{\theta}_n^{*}-\theta_0)}{2n} l_{\underline{x}}'''(\hat{\theta}_n^{**}) \xrightarrow{a.s.} 0. \quad \text{Consequently}$$

$$\sqrt{n}\,(\hat{\theta}_n^{*}(\underline{x})-\theta_0) \xrightarrow{d} N\left(0, \frac{1}{I(\theta_0)}\right)$$

<u>Corollary</u>: Under the assumptions of last theorem suppose that the $l_{\underline{x}}'(\theta)=0$ has unique root for every $\underline{x}$ and n. Then the MLE is CAN and asymptotically efficient.

<u>Example</u> (One parameter exponential family). Let

$$f_\theta(x) = e^{\theta T(x) - \psi(\theta)} h(x) \quad \dots \quad (I)$$

w.r.t. σ-finite measure μ, where $\theta \in \Theta = \{\theta \in \mathbb{R} : \int e^{\theta T(x)} h(x)\, d\mu(x) < \infty\}$

Then it can be seen that assumptions of last theorem are satisfied with the role of Θ taken by Θ° (the interior of Θ). Moreover,

(a) for any $\theta \in \Theta^\circ$ all moments of $T(x_1)$ exist and $\psi(\theta)$ is infinitely differentiable at any $\theta \in \Theta^\circ$.

(b) $E_\theta(T(x_1)) = \psi'(\theta)$, $Var_\theta(T(x_1)) = \psi''(\theta)$;

(c) $I(\theta) < \infty$, $\forall\, \theta \in \Theta^\circ$ and $I(\theta) = \psi''(\theta)$, $\theta \in \Theta^\circ$

Theorem For the one parameter exponential family (I), let $\theta_0 \in \Theta^\circ$. Assume that $\psi''(\theta) > 0 \;\; \forall \theta \in \Theta^\circ$. Then the equation $l'_{\underline{x}}(\theta) = 0$ has a unique solution $\hat{\theta}^*_n(\underline{x})$ for every n and $\underline{x}$ such that

(a) $\lim_{n \to \infty} P_\theta($ the $\hat{\theta}^*_n(\underline{x})$ is MLE of $\theta) = 1, \;\; \forall \theta \in \Theta^\circ$;

(b) $\sqrt{n}\,(\hat{\theta}^*_n(\underline{x}) - \theta_0) \xrightarrow{d} N\left(0, \frac{1}{\psi''(\theta_0)}\right)$.

Proof. We have, for $\theta \in \Theta^\circ$

$$L_{\underline{x}}(\theta) = \left(e^{\theta \sum_{i=1}^n T(x_i) - n\psi(\theta)}\right) \prod_{i=1}^n h(x_i)$$

$$l_{\underline{x}}(\theta) = \theta \sum_{i=1}^n T(x_i) - n\psi(\theta) + \sum_{i=1}^n \ln h(x_i)$$

$$\frac{\partial}{\partial \theta} l_{\underline{x}}(\theta) = 0 \iff \psi'(\theta) = \frac{1}{n}\sum_{i=1}^n T(x_i)$$

$$\frac{1}{n}\sum_{i=1}^n T(X_i) \xrightarrow{a.s.} E_\theta(T(X_1)) = \psi'(\theta)$$

Also, since $\psi''(\theta) > 0, \; \forall \theta \in \Theta^\circ$, $\psi'(\theta)$ is a strictly increasing function of $\theta \in \Theta^\circ$. Thus, for large n $(n \to \infty)$, w.p. 1 there exists a unique root of the equation

$$l'_{\underline{x}}(\theta) = 0, \quad \forall \underline{x} \text{ and } n$$

Now the result follows from the last Corollary

Asymptotic Confidence Intervals

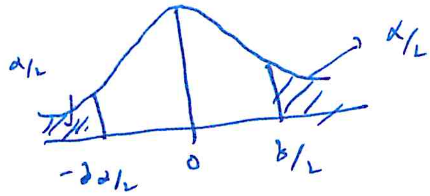
(I) Using MLE

Under the assumptions of last but one theorem suppose that $\hat{\theta}_n(\underline{x})$ is the unique solution of $l'_{\underline{x}}(\theta) = 0$, for every n and $\underline{x}$ and is MLE of $\theta \in \Theta \subseteq \mathbb{R}$. Then

$$\sqrt{n}\,(\hat{\theta}^*_n(\underline{x}) - \theta_0) \to N\left(0, \frac{1}{I(\theta_0)}\right)$$

For a given $\alpha \in (0, 1)$.

$$\lim_{n\to\infty} P_{\theta_0}\left(-z_{\alpha/2} \le \frac{\sqrt{n}\,(\hat{\theta}_n^*(\underline{x})-\theta_0)}{\sqrt{I(\theta_0)}} \le z_{\alpha/2}\right) = 1-\alpha$$



In this case

$$-z_{\alpha/2} \le \frac{\sqrt{n}\,(\hat{\theta}_n^*(\underline{x})-\theta_0)}{\sqrt{I(\theta_0)}} \le z_{\alpha/2}$$

Can be solved to get a $100(1-\alpha)\%$ C.I for θ_0.

By SLLN

$$\frac{1}{n}\sum_{i=1}^{n} -\frac{\partial^2}{\partial\theta_0^2}\ln f_{\theta_0}(x_i) \xrightarrow{as} I(\theta_0) \quad \ldots\ldots \quad (\text{II})$$

If $I(\theta)$ is a continuous function of θ then

$$I(\hat{\theta}_n^*) \xrightarrow{p} I(\theta_0)$$

$$\Rightarrow \frac{\sqrt{n}\,(\hat{\theta}_n^*(\underline{x})-\theta_0)}{\sqrt{I(\hat{\theta}_n^*(\underline{x}))}} \xrightarrow{d} N(0,1)$$

Thus

$$\left[\theta_0 - z_{\alpha/2}\frac{\sqrt{I(\hat{\theta}_n^*(\underline{x}))}}{\sqrt{n}},\ \theta_0 + z_{\alpha/2}\frac{\sqrt{I(\hat{\theta}_n^*(\underline{x}))}}{\sqrt{n}}\right]$$

is also taken as $100(1-\alpha)\%$ C.I. for θ_0.

By (II), if $I(\theta)$ is a smooth fn of θ then

$$\hat{I}(\underline{x}) = \left(\frac{1}{n}\sum_{i=1}^{n} -\frac{\partial^2}{\partial\theta^2}\ln f_\theta(x_i)\right)_{\theta=\hat{\theta}_n^*(\underline{x})}$$

Can also be used as an estimate of $I(\theta_0)$ to get C.I. for θ_0.

Using Score Function

Let us call the function

$$S(\underline{X};\theta) = \frac{\partial}{\partial\theta}\ln g_\theta(\underline{x}) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f_\theta(x_i) = \sum_{i=1}^{n} S_i$$

the score function

$E_{\theta_0}(S_1)=0$, $Var_{\theta_0}(S_1) = I(\theta_0)$. By CLT

$$\frac{\sqrt{n}\,(\bar{S}-0)}{\sqrt{I(\theta_0)}} \xrightarrow{d} N(0,1)$$

$$P_{\theta_0}\left(-z_{\alpha/2} \le \frac{\sqrt{n}\,\bar{S}}{\sqrt{I(\theta_0)}} \le z_{\alpha/2}\right) = 1-\alpha$$

In the above one may replace $I(\theta_0)$ by $I(\hat{\theta}_n^*)$ or $\hat{I}(\underline{x})$ provided $I(\theta)$ is a smooth function.

Example Let $X_1, X_2, \ldots, X_n$ be iid $N(0,\theta)$ where $\theta\in\Theta=(0,\infty)$

Then $\hat{\theta}_n^*(\underline{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ is the unique MLE with

$$I(\theta) = -E_\theta\left(\frac{\partial^2}{\partial\theta^2}\ln f_\theta(x_1)\right)$$

$$= \frac{1}{2\theta^2}$$

$$P_{\theta_0}\left(-1.96 \le \frac{\sqrt{n}\,(\hat{\theta}_n^*(\underline{X})-\theta_0)}{\frac{1}{\sqrt{2}\,\theta_0}} \le 1.96\right) \simeq 0.95$$

$$\Rightarrow \left[\frac{\frac{1}{n}\sum_{i=1}^{n}x_i^2}{1+1.96\sqrt{\frac{2}{n}}},\ \frac{\frac{1}{n}\sum_{i=1}^{n}x_i^2}{1-1.96\sqrt{\frac{2}{n}}}\right]$$

is a 95% C.I for θ_0

Assignment: Let $X_1, X_2, \ldots, X_n$ be i.i.d. Poisson (θ), where $\theta \in \Theta = (0, \infty)$. Find different asymptotic 95% CIs based on MLE (provided it exists) and the score function