Consistency and asymptotic normality

Example: (Poisson)

$$X_i \overset{i.i.d}{\sim} \text{Pois}(\vartheta) \qquad g(\vartheta) = \vartheta$$

$$T_n = \frac{1}{n} \sum X_i \quad (\text{MLE estimator})$$

$$Var_\vartheta(T_n) = \frac{\vartheta}{n}$$

As $n \longrightarrow \infty$

1. LLN: $T_n \xrightarrow{P} \vartheta$
2. CLT: $\sqrt{n}\,(T_n - \vartheta) \xrightarrow{d} \mathcal{N}_0(0, Var(X_i))$
   $= \mathcal{N}_0(0, \vartheta)$
3. sampling distribution of T_n is approximately $\mathcal{N}_0\left(\vartheta, \frac{\vartheta}{n}\right)$

(normal approximation)

Definition: Convergence in distribution

Let $\{T_n\}_{n=1}^{\infty}$ and T^* be r.v. we say that ($T_n \xrightarrow[n\to\infty]{d} T^*$) T_n converges in distribution if for all continuous bounded functions f,

$$\lim_{n\to\infty} E[f(T_n)] = E[f(T^*)]$$

Note: Convergence in probability implies convergence in distribution

Theorem (Slutsky) Let $(\{T_n, A_n\}, T^*)$ be a collection of $\mathbb{R}^p$-r.v. and $a \in \mathbb{R}^p$ be a vector of constants. Assume that $T_n \xrightarrow[n\to\infty]{d} T^*$ and $A_n \xrightarrow[n\to\infty]{P} a$. Then

$$A_n^T T_n \xrightarrow[n\to\infty]{d} a^T \cdot T^*.$$

Recall: MLE estimator T_n of $g(\theta)$

$$T_n = \arg\max_{\theta} \log P_\theta(x).$$

we showed in the last lecture that the MLE estimator is an M-estimator and consistent $T_n \xrightarrow[n\to\infty]{P} T^*$

Goal: Show that the MLE estimator is asymptotically normal and there is a general formula for its assymptotic variance.

Theorem: (Asymptotic normality of MLE estimators)

Let $X_i \overset{i.i.d}{\sim} P_{\theta_0}(x)$ for $\theta_0 \in \Theta$ (true parameter) and let $P_\theta(x)$ be the likelihood of our statistical model, and let T_n be the MLE estimator on $X_1, \dots, X_n$. Suppose the following assumptions hold:

(A1) parameter space Θ : compact
$\theta_0 \in \Theta^\circ$ i.e. (in the interior of Θ not in the boundary.

(A2) The log-likelihood $\ell(\theta)$ is differentiable in θ.

(A3) T_n has unique value of $\theta \in \Theta$ that solves the equation $0 = \ell'(\theta)$.
(indentifiability.)

(A4) Uniform integrability for the

score function.

$\forall \varepsilon > 0, \exists k > 0$ s.t. $\sup_{\vartheta} E_{\vartheta}\left[|S(x)| \mathbb{1}_{\{S(x) > k\}}\right] < \varepsilon$

(A5) The map $\vartheta \mapsto \log p_{\vartheta}(x)$ continuous

$\exists$ integrable function $M(x)$ s.t.

$|\log p_{\vartheta}(x)| \le M(x) \quad \forall \vartheta \in \Theta$.

$E_{\vartheta_0}[M(x)] < \infty$.

Then T_n is asymptotically normal

with $\sqrt{n}\,(T_n - \vartheta_0) \xrightarrow[n\to\infty]{d} \mathcal{N}_0(0, I^{-1}(\vartheta_0))$

Where

$I(\vartheta) = Var_{\vartheta}(S_{\vartheta}(x)) = -E_{\vartheta}[S'_{\vartheta}(x)]$

$S_{\vartheta}(x) = \frac{\partial}{\partial \vartheta} \log p_{\vartheta}(x) \qquad S'_{\vartheta}(x) = \frac{\partial^2}{\partial \vartheta^2} \log p_{\vartheta}(x)$

Note: • T_n is asymptotically unbiased

The bias of T_n is less than order $1/\sqrt{n}$. Otherwise $\sqrt{n}(T_n - \vartheta_0)$ should not converge to a distribution with zero mean.

• The variance of T_n is approximately $1/nI(\vartheta_0)$.

In particular the standard error is of order $1/\sqrt{n}$, and the variance is the main contributig factor to the mean square error of T_n.

- if ϑ_0 is the true parameter the sampling distribution of T_n is approximately $\mathcal{N}_0\left(\vartheta_0, \frac{1}{n I(\vartheta_0)}\right)$

<u>Example</u>: (Poisson)

$$\log p_\vartheta(x) = \log \frac{\vartheta^x e^{-\vartheta}}{x!} = x \log \vartheta - \vartheta - \log x!$$

So the score function and its derivative:

$$s_\vartheta(x) = \frac{\partial}{\partial \vartheta} \log p_\vartheta(x) = \frac{x}{\vartheta} - 1, \quad s'_\vartheta(x) = \frac{\partial^2}{\partial \vartheta^2} \log p_\vartheta(x)$$

$$= -\frac{x}{\vartheta^2}.$$

Fisher information matrix:

$$I(\vartheta) = -E_\vartheta[s'_\vartheta(x)] = 1/\vartheta$$

So $\sqrt{n}\,(T_n - \vartheta) \xrightarrow[n\to\infty]{d} \mathcal{N}_0(0, \vartheta)$

Lemma: (properties of the score function)

For $\vartheta \in \Theta$: $E_\vartheta(S_\vartheta(x)) = 0$

$Var_\vartheta(S_\vartheta(x)) = -E(S'_\vartheta(x))$

Proof: By chain rule of differentiation

$$S_\vartheta(x)\, p_\vartheta(x) = \left(\frac{\partial}{\partial\vartheta} \log p_\vartheta(x)\right) p_\vartheta(x)$$

$$= \frac{\frac{\partial}{\partial\vartheta} p_\vartheta(x)}{p_\vartheta(x)} \cdot p_\vartheta(x)$$

$$= \frac{\partial}{\partial\vartheta} p_\vartheta(x) \quad . (*)$$

Since $\int p_\vartheta(x)\, dx = 1$

$$E_\vartheta(S_\vartheta(x)) = \int S_\vartheta(x)\, p_\vartheta(x)\, dx = \int \frac{\partial}{\partial\vartheta} p_\vartheta(x)\, dx$$

$$= \frac{\partial}{\partial\vartheta} \int p_\vartheta(x)\, dx = 0.$$

we differentiate this identity with respect to ϑ:

$$0 = \frac{\partial}{\partial\theta} E_\theta(s_\theta(x))$$

$$= \frac{\partial}{\partial\theta} \int s_\theta(x) p_\theta(x)\, dx$$

$$= \int s'_\theta(x) p_\theta(x) + s_\theta(x)\left(\frac{\partial}{\partial\theta} p_\theta(x)\right) dx$$

$$\overset{(*)}{=} \int s'_\theta(x) p_\theta(x) + s^2_\theta(x)\, p_\theta(x)\, dx$$

$$= E_\theta(s'_\theta(x)) + E_\theta(s^2_\theta(x))$$

$$= E_\theta(s'_\theta(x)) + Var_\theta(s_\theta(x)) \quad \blacksquare$$

Sketch proof: Since T_n maximizes $\ell(\theta)$, we must have $\ell'(T_n) = 0$. Consistency of T_n ensures that $T_n \xrightarrow[n\to\infty]{P} \theta_0$. This allows us to apply a first order Taylor expansion to the equation $0 = \ell'(T_n)$ around $\theta = \theta_0$.

$$0 \approx \ell'(\theta_0) + (T_n - \theta_0)\ell''(\theta_0) \text{ (linearization)}$$

So $\sqrt{n}(T_n - \vartheta_0) \approx -\sqrt{n}\,\dfrac{l'(\vartheta_0)}{l''(\vartheta_0)}$

$$= -\frac{\frac{1}{\sqrt{n}}\, l'(\vartheta_0)}{\frac{1}{n}\, l''(\vartheta_0)} .$$

For the denominator, by the LLN

$$\frac{1}{n} l''(\vartheta_0) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2}{\partial \vartheta^2}\left(\log p_\vartheta(x)\right)\Big|_{\vartheta=\vartheta_0}$$

$$= \frac{1}{n}\sum_{i=1}^{n} s'_{\vartheta_0}(x_i) \xrightarrow[n\to\infty]{P} E_{\vartheta_0}\left(s'_{\vartheta_0}(x)\right)$$

$$= -I(\vartheta_0)$$

For the nominator recall by Lemma (properties of the score)

$s_\vartheta(x)$ has zero mean and variance $I(\vartheta_0)$ when $X_i \sim p_{\vartheta_0}(x)$. Then by CLT

$$\frac{1}{\sqrt{n}} l'(\vartheta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\partial}{\partial \vartheta}\left(\log p_\vartheta(x)\right)\Big|_{\vartheta=\vartheta_0}$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1} s_{\vartheta_0}(x_i) \xrightarrow[n\to\infty]{d} \mathcal{N}_0(0, I(\vartheta_0))$$

By continuous mapping Theorem

slutsky's lemma

$$\sqrt{n}(T_n - \vartheta_0) \rightarrow \frac{1}{I(\vartheta_0)} N(0, I)$$

<u>Note</u>: Continuous mapping Theorem

Let $x_n \xrightarrow{P} x$ g continuous

$$\Rightarrow \quad g(x_n) \rightarrow g(x)$$

$$\frac{1}{n}\sum x_i \xrightarrow{P} \mu \quad g(x) = x^2$$

The $\left(\frac{1}{n}\sum x_i\right)^2 \xrightarrow{P} \mu^2$