Aymptotic (Large Sample) Theory

A random sequence A_n is:

$$\begin{aligned}
&(a)\ \ o_p(1) \text{ if } A_n \xrightarrow{p} 0 &(1)\\
&(b)\ \ o_p(B_n) \text{ if } A_n/B_n \xrightarrow{p} 0 &(2)\\
&(c)\ \ O_p(1) \text{ if } \forall\epsilon>0, \exists M: \lim_{n\to\infty}\mathbb{P}(|A_n|>M)<\epsilon &(3)\\
&(d)\ \ O_p(B_n) \text{ if } A_n/B_n = O_p(1) &(4)
\end{aligned}$$

If $Y_n \rightsquigarrow Y \implies Y_n = O_p(1)$
If $\sqrt{n}(Y_n - c) \rightsquigarrow Y \implies Y_n = O_p(1/\sqrt{n})$

Distances Between Distributions

For distributions P and Q with pdfs p and q:

$$\begin{aligned}
&(a)\ \ V(P,Q) = \sup_A |P(A) - Q(A)| \quad \textbf{total variation} \text{ distance} &(5)\\
&(b)\ \ K(P,Q) = \int p\log(p/q) \quad \textbf{Kullback-Leibler} \text{ divergence} &(6)\\
&(c)\ \ d_2(P,Q) = \int (p-q)^2 \quad \mathbf{L_2} \text{ distance} &(7)
\end{aligned}$$

A model is **identifiable** if: $\theta_1 \neq \theta_2 \implies K(\theta_1, \theta_2) > 0$.

Consistency

$\hat{\theta}_n = T(X^n)$ is **consistent** for θ if $\hat{\theta}_n \xrightarrow{p} \theta$ (ie if $\hat{\theta}_n - \theta = o_p(1)$).
To show consistency, can show: $\text{Bias}^2(\hat{\theta}_n) + \text{Var}(\hat{\theta}_n) \to 0$.
The MLE is consistent under regularity conditions.
MLE not consistent when number of params (or support?) grows.

Score and Fisher Information

The **score function** is $S(\theta) = \frac{\partial}{\partial\theta} l(\theta) = \frac{\partial}{\partial\theta}\sum_{i=1}^n \log p(x_i|\theta)$.
The **Fisher information** is defined as

$$I_n(\theta) = \mathbb{E}_\theta\left[S(\theta)^2\right] = \text{Var}_\theta\left[S(\theta)\right] = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2} l(\theta)\right] \qquad (8)$$

and $I_n(\theta) = -n\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log p(X_1;\theta)\right] = nI_1(\theta)$.
The **observed information** $\hat{I}_n(\theta) = -\sum_i \frac{\partial^2}{\partial\theta^2}\log p(X_i;\theta)$.
Vector case: $S(\theta) = \left[\frac{\partial l(\theta)}{\partial\theta_i}\right]_{i=1,\ldots,K}$ $\quad I_{ij} = -\mathbb{E}_\theta\left[\frac{\partial^2 l(\theta)}{\partial\theta_i\partial\theta_j}\right]_{i,j=1,\ldots,K}$

Efficiency and Robustness

For an estimator $\hat{\theta}_n(X^n)$ of θ, where $X^n \overset{\text{iid}}{\sim} p(x|\theta)$:
If $\sqrt{n}(\hat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, v^2)$, then v^2 is the **asymptotic-Var**$(\hat{\theta}_n)$.
E.g. for $\hat{\theta}_n = \overline{X}_n$: $\quad v^2 = \sigma^2 = \text{Var}(X_i) = \lim_{n\to\infty} n\text{Var}(\overline{X}_n)$.
In general, asymptotic-Var$(\hat{\theta}_n)$ $v^2 \neq \lim_{n\to\infty} n\text{Var}(\hat{\theta}_n)$.
We will use approx: $\text{Var}(\hat{\theta}_n) \approx v^2/n$.
For param $\tau(\theta)$, $v(\theta) = \frac{|\tau'(\theta)|^2}{I_1(\theta)}$ is the **Cramer-Rao lower bound**.
since, for most estimators $\hat{\theta}_n$, the asymptotic-Var$(\hat{\theta}_n) \geq v(\theta)$.
If $\sqrt{n}(\hat{\theta}_n - \tau(\theta)) \rightsquigarrow \mathcal{N}(0, v(\theta))$ (ie if $v^2 = v(\theta)$) $\implies \hat{\theta}_n$ **efficient**.
(usually) $\sqrt{n}(\tau(\hat{\theta}_{\text{mle}}) - \tau(\theta)) \rightsquigarrow \mathcal{N}(0, v(\theta)) \implies$ MLE efficient.
The **standard error** of **efficient** $\hat{\theta}_n$ is $se = \sqrt{\text{Var}(\hat{\theta}_n)} \approx \sqrt{\frac{1}{I_n(\theta)}}$.
The **estimated standard error** of **efficient** $\hat{\theta}_n$ is $\hat{se} \approx \sqrt{\frac{1}{I_n(\hat{\theta}_n)}}$.
For efficient $\hat{\theta}_n$, $\hat{\tau} = \tau(\hat{\theta}_n)$, $se \approx \sqrt{\frac{|\tau'(\theta)|^2}{I_n(\theta)}}$, and $\hat{se} \approx \sqrt{\frac{|\tau'(\hat{\theta}_n)|^2}{I_n(\hat{\theta}_n)}}$.
In general, **asymptotic normality** is when:
$\frac{\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n)}{\sqrt{\text{Var}(\hat{\theta}_n)}} \rightsquigarrow \mathcal{N}(0,1) \implies \hat{\theta}_n \rightsquigarrow \mathcal{N}(\mathbb{E}(\hat{\theta}_n), \text{Var}(\hat{\theta}_n))$.
If $\sqrt{n}(W_n - \tau(\theta)) \rightsquigarrow \mathcal{N}(0, \sigma_W^2)$ and $\sqrt{n}(V_n - \tau(\theta)) \rightsquigarrow \mathcal{N}(0, \sigma_V^2)$
$\implies$ **asymptotic relative efficiency** $\text{ARE}(V_n, W_n) = \sigma_W^2/\sigma_V^2$.
Often there is a tradeoff between efficiency and robustness. (?)

Hypothesis Testing

Null hypothesis $H_0: \theta \in \Theta_0$, **alternative** $H_1: \theta \in \Theta_1$.
Type I error: If H_0 true but we reject H_0.
To construct a test:

1. Choose a test statistic $W = W(X_1, \ldots, X_n)$
2. Choose a rejection region R (9)
3. If $W \in R$, reject H_0 otherwise retain H_0

For rejection region R, the **power function** $\beta(\theta) = \mathbb{P}_\theta(X^n \in R)$.
Want **level-**α test ($\sup_{\theta\in\Theta_0}\beta(\theta) \leq \alpha$) that maximizes $\beta(\theta_1)$.
A level-α test with power fn β is **uniformly most powerful** if:
$\beta(\theta) \geq \beta'(\theta)\ \forall\theta \in \Theta_1\ \forall\beta' \neq \beta$.

Neyman-Pearson Test

For simple $H_0: \theta = \theta_0$ and $H_1: \theta = \theta_1$, reject H_0 if $\frac{L(\theta_1)}{L(\theta_0)} > k$.
where k chosen s.t. $\mathbb{P}(\frac{L(\theta_1)}{L(\theta_0)} > k) = \alpha$.

Wald Test

For $H_0: \theta = \theta_0$ and $H_1: \theta \neq \theta_0$, reject H_0 if $\left|\frac{\hat{\theta}_n - \theta_0}{se}\right| > z_{\alpha/2}$.
where $z_{\alpha/2}$ is the inverse standard-normal CDF of $1 - \frac{\alpha}{2}$.
and $\hat{\theta}_n$ is an unbiased estimator for θ.
and $se = \sqrt{\text{Var}(\hat{\theta}_n)}$. Can also use $\hat{se} =_{\text{eg.}} \sqrt{S_n^2/n}$.
and if $\hat{\theta}_n$ efficient, can approx: $se \approx \sqrt{\frac{1}{I_n(\theta)}}$ or $\hat{se} \approx \sqrt{\frac{1}{I_n(\hat{\theta}_n)}}$.

Likelihood Ratio Test

For $H_0: \theta \in \Theta_0$ and $H_1: \theta \notin \Theta_0$, reject H_0 if $\lambda(x^n) = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} \leq c$.
where $L(\hat{\theta}_0) = \sup_{\theta\in\Theta_0} L(\theta)$ and $L(\hat{\theta}) = \sup_{\theta\in\Theta} L(\theta)$.
and c chosen s.t. $\mathbb{P}(\lambda(x^n) \leq c) = \alpha$.
Thm: under $H_0: \theta = \theta_0 \implies W_n = -2\log\lambda(X^n) \rightsquigarrow \chi_1^2$
$\implies \lim_{n\to\infty}\mathbb{P}_{\theta_0}(W_n > \chi_{1,\alpha}^2) = \alpha$.
Also: for $\theta = (\theta_1, \ldots, \theta_k)$, if H_0 fixes some of the parameters
$\implies -2\log\lambda(X^n) \rightsquigarrow \chi_\nu^2$, where $\nu = \dim(\Theta) - \dim(\Theta_0)$.

P-Values

The **p-value** $p(x^n)$ is the smallest α-level s.t. we reject H_0.
Thm: For a test of the form: reject H_0 when $W(x^n) > c$,
$\implies p(x^n) = \sup_{\theta\in\Theta_0}\mathbb{P}_\theta(W(X^n) \geq W(x^n)) = \sup_{\theta\in\Theta_0}[1 - F(W(x^n)|\theta)]$.
Thm: Under $H_0: \theta = \theta_0$, $p(x^n) \sim \text{Unif}(0,1)$.

Permutation Test