STAT 620: Asymptotic Statistics **Spring 2022**

Lecture: Feb 15

Lecturer: Xianyang Zhang

1 Exponential Family

Consider the exponential family

$$f(x|\beta) = h(x)\exp\left[\eta(\beta)^\top T(x) - \gamma(\beta)\right],$$

where $T : \chi \to \mathbb{R}^d$. Set $\theta = \eta(\beta)$, we can rewrite the distribution as

$$f(x|\theta) = h(x)\exp\left[\theta^\top T(x) - A(\theta)\right],$$

with $A(\theta) = \gamma(\eta^{-1}(\theta))$. As $\int f(x|\theta)d\mu = 1$, we have

$$A(\theta) = \log\left(\int h(x)\exp\left\{\theta^\top T(x)\right\}dx\right).$$

From this we can calculate the likelihood and further get the expectation of $T(x)$ as a function of $A(\theta)$. To see this, note that

$$l(\theta) = \log f(x|\theta) = \log\left[h(x)\right] + \theta^\top T(x) - A(\theta)$$

which implies that

$$\nabla l(\theta) = T(x) - \nabla A(\theta), \quad \nabla^2 l(\theta) = -\nabla^2 A(\theta).$$

Thus we get

$$\mathcal{I}(\theta) = -\mathbb{E}\left[\nabla^2 l(\theta)\right] = \nabla^2 A(\theta).$$

On the other hand, we observe from the definition of $A(\theta)$ that

$$\nabla A(\theta) = \frac{\int T(x)h(x)\exp\left[\theta^\top T(x)\right]dx}{\exp\left[A(\theta)\right]} = \int T(x)h(x)\exp\left[\theta^\top T(x) - A(\theta)\right]dx = \mathbb{E}_\theta\left[T(X)\right],$$

and

$$\nabla^2 A(\theta) = \nabla\mathbb{E}_\theta\left[T(X)\right] = \int T(x)\nabla f(x|\theta)dx.$$

Noting that $\nabla f(x|\theta) = (T(x) - \nabla A(\theta))f(x|\theta)$, we have

$$\begin{aligned}
\nabla^2 A(\theta) &= \int T(x)(T(x) - \nabla A(\theta))^\top f(x|\theta)dx \\
&= \mathbb{E}_\theta\left[T(X)T(X)^\top\right] - \mathbb{E}_\theta\left[T(X)\right]\nabla A(\theta)^\top \\
&= \mathbb{E}_\theta\left[T(X)T(X)^\top\right] - \mathbb{E}_\theta\left[T(X)\right]\mathbb{E}_\theta\left[T(X)\right]^\top \\
&= \mathrm{cov}_\theta\left[T(X)\right].
\end{aligned}$$

2 Maximum Likelihood Estimator for Exponential Family

Suppose $X_1, X_2, \ldots X_n \overset{i.i.d}{\sim} f(x|\theta)$. We aim to find

$$\max_{\theta\in\Theta}\sum_{i=1}^n\left[\theta^\top T(X_i) - A(\theta)\right].$$

Let $g(\theta) = \sum_{i=1}^{n} \left[\theta^\top T(X_i) - A(\theta)\right]$ and note that

$$g'(\theta) = \sum_{i=1}^{n} \left[T(X_i) - \nabla A(\theta)\right].$$

Thus the MLE satisfies that

$$\frac{1}{n}\sum_{i=1}^{n} T(X_i) = \nabla A(\theta) = \mathbb{E}_\theta\left[T(X)\right].$$

Denote the MLE by $\hat{\theta}_n$. We know that $\hat{\theta}_n$ is also a moment estimator that solves the moment equations defined based on T. Moreover, we have

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

which follows from the asymptotic normality for moment estimator.

2.1 Example

Consider the Poisson distribution $f(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}$. Rewriting this in terms of the exponential distribution, we get

$$f(x|\lambda) = \exp\left[x\log\lambda - \lambda - \log x!\right].$$

It implies that in this case $\theta = \log\lambda$, $T(x) = x$ and $A(\theta) = e^\theta$. From this, we see that

$$\nabla A(\theta) = \nabla^2 A(\theta) = e^\theta \text{ and } \hat{\theta}_n = \log(\bar{X}).$$

Thus we get

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} N\left(0, e^{-\theta}\right).$$

3 Asymptotic Relative Efficiency

3.1 Definition

Let us assume that $\hat{\theta}_n$ and T_n are the estimators of the parameter $\theta \in \mathbb{R}$. We aslo assume that

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} N\left(0, \sigma^2(\theta)\right).$$

Let $m(n) \to \infty$ such that

$$\sqrt{n}\left(T_{m(n)} - \theta\right) \xrightarrow{d} N\left(0, \sigma^2(\theta)\right).$$

The ARE of $\hat{\theta}_n$ with respect to T_m is defined as the limit $\lim_{n\to\infty}\frac{m(n)}{n}$.

3.2 Sample Size

Suppose $\frac{m(n)}{n} \to c$ as $n \to \infty$ for some constant c. If $c \geq 1$ then we need cn samples corresponding to n samples to get an estimate of same quality as $\hat{\theta}_n$.

3.3 Confidence Interval

Next we ask the question that what is the asymptotic distribution of $\sqrt{n}(T_n - \theta)$. Let us assume that $\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0, \tau^2(\theta))$. Note that

$$\sqrt{n}(T_{m(n)} - \theta) = \sqrt{\frac{n}{m(n)}}\sqrt{m(n)}(T_{m(n)} - \theta).$$

By comparing the variance, we get that

$$\lim_n \frac{n}{m(n)}\tau^2(\theta) = \sigma^2(\theta) \implies \frac{\sigma^2(\theta)}{\tau^2(\theta)} \approx \frac{n}{m(n)} \implies \tau^2(\theta) \approx c\sigma^2(\theta).$$

Next we construct a $(1-\alpha)100\%$ asymptotic confidence Interval $\mathcal{I}_\alpha$ for θ such that

$$P\left[\theta \in \mathcal{I}_\alpha\right] \to (1-\alpha) \in (0,1).$$

We consider two intervals:

$$\mathcal{I}_{\hat{\theta}_n,\alpha} : \left(\hat{\theta}_n - z_{\alpha/2}\sqrt{\frac{\sigma^2(\theta)}{n}}, \hat{\theta}_n + z_{\alpha/2}\sqrt{\frac{\sigma^2(\theta)}{n}}\right),$$

and

$$\mathcal{I}_{T_n,\alpha} : \left(T_n - z_{\alpha/2}\sqrt{\frac{c\sigma^2(\theta)}{n}}, T_n + z_{\alpha/2}\sqrt{\frac{c\sigma^2(\theta)}{n}}\right).$$

From this, we can calculate the ratio of the lengths of two confidence intervals

$$\frac{\mathcal{I}_{T_n,\alpha}}{\mathcal{I}_{\hat{\theta}_n,\alpha}} \approx \sqrt{c}.$$

4 Super Efficiency

Let $\hat{\theta}_n$ and T_n be two estimators of the parameter θ. Suppose

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} N\left(0, I_\theta^{-1}\right),$$

and

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0, \tau^2(\theta)),$$

where $\tau^2(\theta) \le I_\theta^{-1}$ and there is some point θ such that $\tau^2(\theta) < I_\theta^{-1}$. In this case T_n is said to be a supper efficient estimator for θ.

4.1 Hodges' Estimator

Suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\theta, 1)$ and $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Let

$$T_n = \begin{cases} \hat{\theta}_n & \text{if } |\hat{\theta}_n| \ge n^{-1/4}, \\ 0 & \text{otherwise .} \end{cases}$$

If $\theta = 0$, then

$$\begin{aligned} P_0\left(\sqrt{n}T_n = 0\right) &= P\left(|\hat{\theta}_n| < n^{-1/4}\right) \\ &= P\left(-n^{-1/4} < \hat{\theta}_n < n^{-1/4}\right) \\ &= P\left(-n^{1/4} < \sqrt{n}\hat{\theta}_n < n^{1/4}\right) \to 1 \end{aligned}$$

as $n \to \infty$. Therefore

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} 0.$$

When $\theta \neq 0$,

$$\sqrt{n}(T_n - \theta) = \sqrt{n}(\hat{\theta}_n - \theta)\mathbf{1}\{|\hat{\theta}_n| \geq n^{-1/4}\} + \sqrt{n}(0 - \theta)\mathbf{1}\{|\hat{\theta}_n| < n^{-1/4}\}.$$

We claim that $\mathbf{1}\{|\hat{\theta}_n| \geq n^{-1/4}\} \xrightarrow{a.s.} 1$. To see this, note that

$$\begin{aligned} P(|\hat{\theta}_n| > n^{-1/4}) &= P(\hat{\theta}_n > n^{-1/4}) + P(\hat{\theta}_n < n^{-1/4}) \\ &= P(\sqrt{n}(\hat{\theta}_n - \theta) > \sqrt{n}(n^{-1/4} - \theta)) + P(\sqrt{n}(\hat{\theta}_n - \theta) < \sqrt{n}(-n^{-1/4} - \theta)) \to 1 \end{aligned}$$

as $n \to \infty$. This is a famous counterexample of an estimator which is "superefficient".