STAT 620: Asymptotic Statistics **Spring 2022**

Lecture: Jan 27

Lecturer: Xianyang Zhang

1 Variance-Stabilizing transformations

1.1 A Motivating Example

Consider the observations $X_1, X_2, \ldots X_n \overset{\text{iid}}{\sim} N(0, \sigma^2)$. Let $\overline{X^2} = \frac{1}{n}\sum_{i=1}^n X_i^2$ and notice that

$$\begin{aligned}\operatorname{var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i^2 - \sigma^2)\right) &= \frac{1}{n}\sum_{i=1}^n \operatorname{var}(X_i^2) \\ &= \operatorname{var}(X_1^2) \\ &= E(X_1^4) - E(X_1^2)^2 \\ &= 2\sigma^4.\end{aligned}$$

By the Central Limit Theorem, we have

$$\sqrt{n}(\overline{X^2} - \sigma^2) \xrightarrow{d} N(0, 2\sigma^4).$$

Consider some function g that is differentiable at σ^2. By the delta method,

$$\sqrt{n}(g(\overline{X^2}) - g(\sigma^2)) \xrightarrow{d} g'(\sigma^2)N(0, 2\sigma^4) = N(0, 2\sigma^4(g'(\sigma^2))^2).$$

The goal is to find a g so that the variance is independent of σ. In particular, we need $2\sigma^4(g'(\sigma^2))^2 = c$, that is

$$g'(\sigma^2) = \sqrt{\frac{c}{2}}\frac{1}{\sigma^2},$$

where c is some constant. Thus we have

$$g'(t) = \sqrt{\frac{c}{2}}\frac{1}{t} \implies g(t) = a + b\log t.$$

A particular choice is $g(t) = \log t$ in which case

$$\sqrt{n}(\log(\overline{X^2}) - \log(\sigma^2)) \xrightarrow{d} N(0, 2).$$

From this, we can construct a $1-\alpha$ confidence interval for the variance σ^2 based on observations $X_1, X_2, \ldots X_n$:

$$\left[e^{\log(\overline{X^2}) - z_{\alpha/2}\sqrt{\frac{2}{n}}}, e^{\log(\overline{X^2}) + z_{\alpha/2}\sqrt{\frac{2}{n}}}\right],$$

where $z_{\alpha/2}$ is the $\alpha/2$ upper percentile of standard normal distribution.

1.2 Variance stabilizing transformation

Consider a sequence $\{\hat{\theta}_n\}$ of estimators of $\theta \in \mathbb{R}$ such that

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

A function g is called variance stabilizing transformation if

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \xrightarrow{d} N(0, 1).$$

By the delta method,

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

Essential condition: $g'(\theta) = \frac{1}{\sigma(\theta)}$. If $g'(\theta) = 0$, we use the second order delta method.

2 Second-Order Delta Method

Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be twice differentiable at θ and $r_n(T_n - \theta) \xrightarrow{d} T$. Then, if $\nabla\phi(\theta) = 0$, we have

$$r_n^2(\phi(T_n) - \phi(\theta)) \xrightarrow{d} \frac{1}{2}T^\top \nabla^2\phi(\theta)T.$$

2.1 Proof

By Taylor's expansion

$$\phi(t) = \phi(\theta) + \nabla\phi(\theta)^\top(t - \theta) + \frac{1}{2}(t - \theta)^\top\nabla^2\phi(\theta)(t - \theta) + R(t - \theta)$$

where $R(h) = o\left(||h||^2\right)$. Under our assumption, $\nabla\phi(\theta) = 0$ and thus

$$\phi(t) = \phi(\theta) + \frac{1}{2}(t - \theta)^\top\nabla^2\phi(\theta)(t - \theta) + R(t - \theta).$$

Replacing t by T_n, and multipling by r_n^2 on both sides, we obtain

$$r_n^2(\phi(T_n) - \phi(\theta)) = r_n^2\frac{1}{2}(T_n - \theta)^\top\nabla^2\phi(\theta)(T_n - \theta) + r_n^2 R(T_n - \theta).$$

Notice that

$$r_n^2 R(T_n - \theta) = r_n^2 o_P(||T_n - \theta||^2) = o_P(r_n^2||T_n - \theta||^2) = o_p(1).$$

Thus,

$$r_n^2(\phi(T_n) - \phi(\theta)) = \frac{1}{2}r_n(T_n - \theta)^\top\nabla^2\phi(\theta)r_n(T_n - \theta) + o_p(1) \xrightarrow{d} \frac{1}{2}T^\top\nabla^2\phi(\theta)T$$

by the continuous mapping theorem.

2.2 Kullback-Leibler divergence

Recall that

$$D_{KL}(P, Q) = \int p(x) \log \frac{p(x)}{q(x)} dx,$$

where p and q are densities for P, Q respectively. Let P_t be the distribution of Bernoulli(t) for $t \in (0,1)$. We are interested in deriving the limiting distribution of $D_{KL}(P_t, P_\theta)$. Note that in the discrete case,

$$D_{KL}(P_t, P_\theta) = \sum_{X=0,1} p_t(X) \log \frac{p_t(X)}{p_\theta(X)} = (1-t)\log\left(\frac{1-t}{1-\theta}\right) + t\log\left(\frac{t}{\theta}\right).$$

Let $\phi(t) = D_{KL}(P_t, P_\theta)$. Note that $\phi(\theta) = 0$ and

$$\phi'(t) = \log(\frac{t}{1-t}) - \log(\frac{\theta}{1-\theta}).$$

As $\phi'(\theta) = 0$, we shall employ the second order delta method.

$$\phi''(t) = \frac{1}{t} + \frac{1}{1-t} = \frac{1}{t(1-t)}.$$

Let $\hat{\theta}_n = n^{-1}\sum_{i=1}^n X_i$, where $X_1, \ldots, X_n \sim^{i.i.d}$ Bernoulli(θ). By the central limit theorem, we have

$$\sqrt{n}(\hat{\theta}_n - \theta) \to^d T,$$

where $T \sim N(0, \theta(1-\theta))$. Applying the second order delta method,

$$\begin{aligned} nD_{KL}(P_{\hat{\theta}_n}, P_\theta) &= n(\phi(\hat{\theta}_n) - \phi(\theta)) \\ &\xrightarrow{d} \frac{1}{2}\frac{T^2}{\theta(1-\theta)} \\ &= \frac{1}{2}\chi_1^2. \end{aligned}$$

3 Moment Estimator

Let $X_1, \ldots, X_n \overset{i.i.d}{\sim} P_{\theta_0}$, where $\theta_0 \in \Theta \subset R^k$. Method of moments consists of estimating θ by the solution of a system of equations. Assume exactly identified system, i.e., k equations and k parameters:

$$\frac{1}{n}\sum_{i=1}^n f_j(X_i) = E_\theta f_j(X), \quad j = 1, 2, \ldots, k,$$

where $X \sim P_\theta$.

Notation:

$$\begin{aligned} P_n(x) &= \frac{1}{n}\sum_{i=1}^n 1\{X_i \le x\}, \\ P_n f &= \frac{1}{n}\sum_{i=1}^n f(X_i), \\ P_\theta f &= E_\theta[f(X)], \end{aligned}$$

where $X \sim P_\theta$ and P_n is the empirical distribution function. With the above notation, we can write the system of equations as follows. Let

$$P_n f = P_\theta f = e(\theta),$$

where $f = (f_1, \ldots, f_k)^\top$. If e is invertible, the estimator is given by $\hat{\theta}_n = e^{-1}(P_n f)$, where $e^{-1} : R^k \to R^k$. The true parameter $\theta_0 = e^{-1}(P_{\theta_0} f)$. By the law of large numbers,

$$P_n f = \frac{1}{n}\sum_{i=1}^n f(X_i) \to^p E_{\theta_0} f(X) = P_{\theta_0} f.$$

If e^{-1} is continuous, then by continuous mapping theorem, $\hat{\theta}_n \to^p \theta_0$. By the central limit theorem, we have

$$\sqrt{n}(P_n f - P_{\theta_0} f) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f(X_i) - Ef(X_i)) \to^d N(0, P_{\theta_0} f f^\top - P_{\theta_0} f P_{\theta_0} f^\top).$$

By the delta method,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n}(e^{-1}(P_n f) - e^{-1}(P_{\theta_0} f)) \to^d N(0, (De(\theta_0))^{-1}(P_{\theta_0} f f^\top - P_{\theta_0} f P_{\theta_0} f^\top)((De(\theta_0))^{-1})^\top),$$

where we have used the fact that $De^{-1}(P_{\theta_0} f) = (De(\theta_0))^{-1}$.