STAT 620: Asymptotic Statistics **Spring 2022**

Lecture: Apr 19

Lecturer: Xianyang Zhang

1 Multivariate delta method

If the vector-valued function $f : \mathbb{R}^k \to \mathbb{R}^m$ is differentiable at θ and

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

then

$$\sqrt{n}(f(T_n) - f(\theta)) \to^d D_\theta T,$$

where $D_\theta \in \mathbb{R}^{m\times k}$ such that

$$f(\theta + h) - f(\theta) = D_\theta h + o(||h||).$$

2 Functional delta method

$\Phi(P)$ is the parameter of interest where Φ is a map from a probability measure P to $\mathbb{R}^k$ for some $k > 0$. For simplicity, we shall consider the case $k = 1$.

2.1 Examples

- Mean: $\Phi(P) = E_P[X]$
- kth central moment: $\Phi(P) = E_P[(X - E_P X)^k]$
- Quantile: $\Phi(P) = F^{-1}(q)$, where F is the cdf associated with P.

Given $X_1, \dots, X_n \sim^{i.i.d} P$, and a parameter of interest $\Phi(P)$, we estimate $\Phi(P)$ by $\Phi(P_n)$, where $P_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$.

2.2 Gateaux derivative

Our goal is to get the asymptotic distribution of $\Phi(P_n)$. Define the kth derivative $\Phi_P^{(k)}(H)$ of the map $t \mapsto \Phi(P + tH)$ at $t = 0$, where H is a perturbation direction. When $k = 1$,

$$\Phi_P^{(1)}(H) = \frac{\partial}{\partial t}\Phi(P + tH)\bigg|_{t=0}$$

is Gateaux derivative and we assume that it exists.

2.3 Taylor type expansion

If the derivatives exist, we have a Taylor type expansion

$$\Phi(P+tH)-\Phi(P)=t\Phi'_P(H)+\frac{1}{2}t^2\Phi_P^{(2)}(H)+\cdots+\frac{t^m}{m!}\Phi_P^{(m)}(H)+o(t^m)$$

with some regularity conditions in the hindsight. Setting $t=1/\sqrt{n},\ H=\sqrt{n}(P_n-P)=G_n$, we have $P+tH=P_n$. Thus the Taylor expansion transforms to

$$\Phi(P_n)-\Phi(P)=\frac{1}{\sqrt{n}}\Phi'_P(G_n)+\frac{1}{2n}\Phi_P^{(2)}(G_n)+\cdots+\frac{1}{m!}\frac{1}{n^{m/2}}\Phi_P^{(m)}(G_n)+o_p(n^{-m/2}).$$

This expansion is the so-called Von-Mises expansion. Assume Φ'_P is a linear map, that is, for any H_1,H_2, we have

$$\Phi'_P(H_1+H_2)=\Phi'_P(H_1)+\Phi'_P(H_2).$$

Setting $m=1$ in the Von-Mises expansion, we have

$$\begin{aligned}\Phi(P_n)-\Phi(P)&=\frac{1}{\sqrt{n}}\Phi'_P(G_n)+o_p(n^{-1/2})\\ \text{(by linearity)}\ &=\frac{1}{n}\sum_{i=1}^n\Phi'_P(\delta_{X_i}-P)+o_p(n^{-1/2}).\end{aligned}$$

So,

$$\sqrt{n}(\Phi(P_n)-\Phi(P))=\frac{1}{\sqrt{n}}\sum_{i=1}^n\Phi'_P(\delta_{X_i}-P)+o_p(1).$$

Now, we can see that there is hope for applying Central Limit Theorem.

2.4 Influence function

The function $x\mapsto\Phi'_P(\delta_x-P)$ is called influence function. Observe

$$\begin{aligned}\Phi'_P(\delta_x-P)=&\ \frac{d}{dt}\Phi(P+t(\delta_x-P))\Big|_{t=0}\\ =&\ \frac{d}{dt}\Phi((1-t)P+t\delta_x)\Big|_{t=0},\end{aligned}$$

where $(1-t)P+t\delta_x$ characterizes a contaminated model. This expression suggests that the influence function measures some sensitivity of a contaminated system. This model is also used in Robust M-estimation.

3 Asymptotic distribution for sample quantile

We now apply the above tool for getting the asymptotic distribution for sample quantile. Write $\Phi(F)$ as the parameter of interest, this is valid and same as $\Phi(P)$ since F uniquely characterizes P. Define $\Phi(F)=F^{-1}(q)$ that is the qth quantile. We need to calculate

$$\Phi'_F(S_x-F)=\ \frac{d}{dt}\Phi(F_t)\Big|_{t=0}$$

where $F_t = (1-t)F + tS_x$ (we drop the dependence on x for notational purpose) and S_x is the cdf of δ_x. Assume qth quantile is unique for F_t for every t, that is, $F_t(\Phi(F_t)) = q$. Taking derivative on both sides with respect to t we get,

$$\begin{aligned}
&\frac{d}{dt}F_t(\Phi(F_t))\bigg|_{t=0} = 0, \\
&\frac{d}{dt}(1-t)F(\Phi(F_t)) + tS_x(\Phi(F_t))\bigg|_{t=0} = 0, \\
&-F(\Phi(F)) + f(\Phi(F))\Phi'_F(S_x - F) + S_x(\Phi(F)) = 0.
\end{aligned}$$

Thus, we get

$$\begin{aligned}
\Phi'_F(S_x - F) &= \frac{F(\Phi(F)) - S_x(\Phi(F))}{f(\Phi(F))} \\
&= \frac{q - S_x(F^{-1}(q))}{f(F^{-1}(q))} \\
&= \begin{cases} \frac{q-1}{f(F^{-1}(q))} & F^{-1}(q) \geq x, \\ \frac{q}{f(F^{-1}(q))} & F^{-1}(q) < x. \end{cases}
\end{aligned}$$

Now suppose $X \sim F$. We get

$$q = F(F^{-1}(q)) = P(X \leq F^{-1}(q)).$$

Therefore, we have

$$E\Phi'_F(S_X - F) = 0$$

and

$$\text{var}(\Phi'_F(S_X - F)) = E(\Phi'_F(S_X - F)^2) = \frac{(q-1)^2 q + q^2(1-q)}{f^2(F^{-1}(q))} = \frac{q(1-q)}{f^2(F^{-1}(q))}.$$

From the Central Limit theorem, we have

$$\sqrt{n}(\Phi(P_n) - \Phi(P)) = \sqrt{n}(F_n^{-1}(q) - F^{-1}(q)) \to^d N\left(0, \frac{q(1-q)}{f^2(F^{-1}(q))}\right).$$

4 Bootstrap inference

4.1 Motivation

1. Approximate complicated/nonstandard limiting distributions
2. Theoretically bootstrap may provide a more accurate approximation.

Let $X_1, \ldots, X_n \sim^{i.i.d} P$. Consider the parameter θ and an estimator $\hat{\theta}_n$. We can construct confidence interval as

$$P\left(\hat{\theta}_n - K_{\frac{\alpha}{2}}\hat{\sigma} \leq \theta \leq \hat{\theta}_n - K_{1-\frac{\alpha}{2}}\hat{\sigma}\right) - 1 - \alpha,$$

where $\hat{\sigma}$ is a variance estimator for $\hat{\theta}_n$. Informally, bootstrap provides a way to compute $K_{\frac{\alpha}{2}}$ and $K_{1-\frac{\alpha}{2}}$ systematically by resampling.

4.2 Asymptotic approximation

Suppose $X_1, \ldots, X_n \sim^{i.i.d} P \mid \theta$. Estimate P by some P_n. With $\hat{\theta}_n \equiv \hat{\theta}_n(X_1, \ldots, X_n)$ and $\hat{\sigma}_n \equiv \hat{\sigma}_n(X_1, \ldots, X_n)$, we have an asymptotic characterization of $\frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n}$. Thus $K_{\frac{\alpha}{2}}$ and $K_{1-\frac{\alpha}{2}}$ can be estimated based on the asymptotic distribution of $\frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n}$.

4.3 Bootstrap approximation

With the sample size n, suppose we already have the estimators P_n and $\hat{\theta}_n$. Now let $X_1^*, \ldots, X_n^* \sim^{i.i.d} P_n \mid \hat{\theta}_n$. Based on the boostrap samples, we have the estimators $\hat{\theta}_n^* \equiv \hat{\theta}_n^*(X_1^*, \ldots, X_n^*)$ and $\hat{\sigma}_n^* \equiv \hat{\sigma}_n^*(X_1^*, \ldots, X_n^*)$. Thus in theory, we aim to show, in almost sure sense, that

$$\frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}_n^*} \rightarrow \text{ the limiting distribution of } \frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n}$$

as $n \to \infty$. Let

$$K_\alpha^* = \arg\min_x \left\{ P_n \left(\frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}_n^*} \leq x \,\middle|\, X_1, \ldots, X_n \right) \geq 1 - \alpha \right\}.$$

We hope that if P_n is a good approximator to P, then

$$P\left(\frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n} \leq K_\alpha^* \right) \approx 1 - \alpha.$$

Thus, we have a way from bootstrap to approximate $K_{\alpha/2}$ by $K_{\alpha/2}^*$.