1 Asymptotic Normality

1.1 Delta Method

Asymptotic Normality of MLE

Let $\psi(u,v)$ be a function and let $v = f_\psi(F)$ be the unique solution of

$$\phi_F(v) = \int_{\mathbb{R}} \psi(u,v)dF(u) = 0. \tag{1}$$

This equation is closely related with the optimization problem

$$\int_{\mathbb{R}} \widetilde{\psi}(u,v)dF_\theta(u) \to \min, \;\; \psi(u,v) = \frac{\partial}{\partial v}\widetilde{\psi}(u,v). \tag{2}$$

More accurately, every solution of (2) is a solution of (1). So, if the problem (2) has an unique solution, then it's $f(F)$ and it's natural to estimate it by $f(\widehat{F}_n)$.

Example 1. Consider a parametric model: $X_i \sim F_\theta$, where F_θ is a continuous distribution with density function $p(x;\theta)$, $\theta \in \Theta$, where Θ is open subset of $\mathbb{R}$.
Consider the problem (2) for $\widetilde{\psi}(u,v) = \ln p(u;v)$. Then we need to find a minimal value of

$$\int_{\mathbb{R}} \ln p(u;v)p(u;\theta)du - \int_{\mathbb{R}} \ln p(u;\theta)p(u;\theta)du = \int_{\mathbb{R}} \ln \frac{p(u;v)}{p(u;\theta)}p(u;\theta)du \geq \ln\left(\int_{\mathbb{R}} \frac{p(u;v)}{p(u;\theta)}p(u;\theta)du\right) = 0.$$

Therefore, θ is the unique solution of our problem. So, it's natural to estimate θ by $f(\widehat{F}_n)$

$$\widehat{\theta} = f(\widehat{F}_n) : \; \frac{1}{n}\sum_{i=1}^{n} \ln p(X_i;\widehat{\theta}) = 0.$$

It's a maximal likelihood estimator for θ.

To prove asymptotically normality of $\widehat{\theta}$ we need to find $L_F(D; f_\psi)$. Consider $F + \varepsilon D$ and let

$$L_F(D; f_\psi) = \lim_{\varepsilon \to 0} \frac{f_\psi(F + \varepsilon D) - f_\psi(F)}{\varepsilon}.$$

By definition,

$$f_\psi(F + \varepsilon D) = \left\{v : \int_{\mathbb{R}} \psi(u,v)dF(u) + \varepsilon \int_{\mathbb{R}} \psi(u,v)dD = 0\right\},$$

so, $f_\psi = \phi_F^{-1}\left(-\varepsilon \int_{\mathbb{R}} \psi(u,v)dD(u)\right)$, where $v = f_\psi(F + \varepsilon D)$. If ϕ_F is the differentiable with respect to v, then

$$\phi_F^{-1}\left(-\varepsilon \int_{\mathbb{R}} \psi(u,v)dD(u)\right) = \phi_F^{-1}(0) - \varepsilon \phi_F^{-1}(0) \int_{\mathbb{R}} \psi(u, f_\psi(F))dD(u) + o(\varepsilon), \;\; \varepsilon \to 0.$$

Therefore,

$$L_F(D; f_\psi) = -\phi_F^{-1}(0) \int_{\mathbb{R}} \psi(u,v)dD(u).$$

Since

$$\phi_F^{-1}(t) = \frac{1}{\phi_F'(\phi_F^{-1}(t))} = \left.\frac{1}{\int_{\mathbb{R}} \frac{\partial}{\partial v}\psi(u,v)dF(u)}\right|_{\phi_F^{-1}(t)}.$$

Suppose that $\psi(u,v)$ is differentiable with respect to v and

$$\int_{\mathbb{R}} \frac{\partial}{\partial v}\psi(u,v)dF(u) \neq 0.$$

Then ϕ_F^{-1} is differentiable and

$$I_F(x) = -\frac{\psi(x, f_\psi(F))}{\int_{\mathbb{R}} \frac{\partial}{\partial v}\psi(u,v)dF(u)\Big|_{v=f(F)}}.$$

Formally, we need to assume some regularity conditions to show that f_ψ is Hadamard differentiable. Particularly, for MLE

$$L_F(x) = -\frac{\frac{\partial}{\partial\theta}\ln p(x;\theta)}{\int_{\mathbb{R}} \frac{\partial^2}{\partial\theta^2}\ln p(x;\theta)p(x;\theta)dx}$$

and

$$\sigma^2(F) = \frac{\int_{\mathbb{R}} \left(\frac{\partial}{\partial\theta}\ln p(x;\theta)\right)^2 p(x;\theta)dx}{\left(\int_{\mathbb{R}} \frac{\partial^2}{\partial\theta^2}\ln p(x;\theta)p(x;\theta)dx\right)^2}.$$

Since,

$$\int_{\mathbb{R}} \frac{\partial^2}{\partial\theta^2}\ln p(x;\theta)p(x;\theta)dx = \int_{\mathbb{R}} \frac{p(x;\theta)\frac{\partial^2}{\partial\theta^2}p(x;\theta) - \left(\frac{\partial}{\partial\theta}p(x;\theta)\right)^2}{p(x;\theta)^2}p(x;\theta)dx = \frac{\partial^2}{\partial\theta^2}\int_{\mathbb{R}} p(x;\theta)dx - \int_{\mathbb{R}}\left(\frac{\partial}{\partial\theta}\ln p(x;\theta)\right)^2 p(x;\theta)dx = -\int_{\mathbb{R}}\left(\frac{\partial}{\partial\theta}\ln p(x;\theta)\right)^2 p(x;\theta)dx$$

and

$$\sigma^2(F) = \frac{1}{\int_{\mathbb{R}}\left(\frac{\partial}{\partial\theta}\ln p(x;\theta)\right)^2 p(x;\theta)dx} = \frac{1}{I(\theta)},$$

where $I(\theta)$ is a Fisher's information. Therefore, MLE is an asymptotically normal estimator.

Confidence Intervals

Using asymptotically normal estimators, we can construct asymptotical confidence intervals. Really, if

$$\sqrt{n}\frac{f(\widehat{F}_n) - f(F)}{\sigma(F)} \xrightarrow{d} Z \sim \mathcal{N}(0,1),$$

then

$$P_F\left(f(F) \in \left(f(\widehat{F}_n) - \frac{z_{1-\alpha/2}\sigma(F)}{\sqrt{n}}, f(\widehat{F}_n) + \frac{z_{1-\alpha/2}\sigma(F)}{\sqrt{n}}\right)\right) \to 1-\alpha.$$

However, we don't know $\sigma(F)$. If σ is a weakly continuous functonal of F, we can use $\sigma(\widehat{F}_n)$ instead of $\sigma(F)$, since

$$\frac{\sigma(F)}{\sigma(\widehat{F}_n)} \xrightarrow{P} 1.$$

The confidence interval

$$\left(f(F) \in \left(f(\widehat{F}_n) - \frac{z_{1-\alpha/2}\sigma(\widehat{F})}{\sqrt{n}}, f(\widehat{F}_n) + \frac{z_{1-\alpha/2}\sigma(\widehat{F})}{\sqrt{n}}\right)\right)$$

is an asymptotical confidence interval of the confidence level $1-\alpha$. It is called *delta-method* interval. Obviously,

$$\overline{L_F^2(X)} = \frac{1}{n}\sum_{i=1}^{n} L_F^2(X_i) \xrightarrow{P} \sigma^2(F).$$

However, we don't know $L_F(x)$. Therefore, we need to estimate $L_F(x)$. Let

$$\widehat{L}(x) = \widehat{L}(x;f) = \lim_{\varepsilon\to 0}\frac{f((1-\varepsilon)\widehat{F}_n + \varepsilon\delta_x) - f(\widehat{F}_n)}{\varepsilon}$$

be a Gateuax derivative of f at the point $\widehat{F}_n$. If $L_F(x)$ is a continuous functional of F under the uniform norm, then

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{L}^2(X_i)$$

is a consistent estimator for $\sigma^2(F)$. Therefore,

$$P_F\left(f(F) \in \left(f(\widehat{F}_n) - \frac{z_{1-\alpha/2}\widehat{\sigma}}{\sqrt{n}}, f(\widehat{F}_n) + \frac{z_{1-\alpha/2}\widehat{\sigma}}{\sqrt{n}}\right)\right) \to 1-\alpha.$$

This interval is called the *infinitesmall jackknife* interval.

Example 2. Let $f(F) = \int_{\mathbb{R}} a(u)dF(u)$. Then

$$L_F(x) = \lim_{\varepsilon\to 0}\frac{f((1-\varepsilon)F + \varepsilon\delta_x) - f(F)}{\varepsilon} = a(x) - \int_{\mathbb{R}} a(u)dF(u).$$

Therefore, L_F is a continuous functional of F and

$$\overline{\widehat{L}^2(X)} = \frac{1}{n}\sum_{i=1}^{n}\left(a(X_i) - \overline{a(X)}\right)^2 = \overline{a(X)^2} - \overline{a(X)}^2$$

is a consistent estimator of $\sigma^2(F)$. So,

$$\left(\overline{a(X)} - \frac{z_{1-\alpha/2}\sqrt{\overline{a(X)^2} - \overline{a(X)}^2}}{\sqrt{n}}, \overline{a(X)} + \frac{z_{1-\alpha/2}\sqrt{\overline{a(X)^2} - \overline{a(X)}^2}}{\sqrt{n}}\right)$$

is an asymptotical $1-\alpha$ confidence interval for $\mathbf{E}a(X)$.

Example 3. Consider $f(F) = F^{-1}(1/2)$. Then

$$\sqrt{n}\frac{f(\widehat{F}_n) - f(F)}{\frac{1}{2p(x_{1/2})}} \xrightarrow{d} Z \sim \mathcal{N}(0,1).$$

Therefore, we need to estimate $p(x_{1/2})$. We can't estimate it by $\overline{\widehat{L}(X)}$ since f is not Hadamard differentiable at $\widehat{F}$.

The Jackknife Method

Let's consider another related estimator for $\sigma^2(F)$. Let

$$\widehat{F}_{n-1,i}(x) = \widehat{F}_{n-i}(x; x_1, ...x_{i-1}, x_{i+1}, ..., x_n) = \frac{1}{n-1}\sum_{j\neq i} I_{x_j \le x}.$$

Then

$$L_{\widehat{F}_n}(x_i) = \lim_{\varepsilon\to 0}\frac{f\left((1-\varepsilon)\widehat{F}_n + \varepsilon\delta_{x_i}\right) - f(\widehat{F}_n)}{\varepsilon} \approx \frac{f\left(\left(1+\frac{1}{n-1}\right)\widehat{F}_n - \frac{1}{n-1}\delta_{x_i}\right) - f(\widehat{F}_n)}{-\frac{1}{n-1}} = (n-1)(f(\widehat{F}_n) - f(\widehat{F}_{n-1,i})).$$

Therefore, it's natural to estimate $\sigma^2(F) = \int_{\mathbb{R}} L^2(x)dF(x) = D_F L^2(X)$ by

$$\frac{1}{n-1}\sum_{i=1}^{n}\left(L_{\widehat{F}_n}(x_i) - \overline{L_{\widehat{F}_n}(x)}\right)^2 \approx (n-1)\sum_{i=1}^{n}\left(f(\widehat{F}_{n-1,i} - \overline{f(\widehat{F}_{n-1,\cdot})}\right)^2 =: nS^2_{jack}.$$

So, it's natural to use an interval

$$f(F) \in \left(f(\widehat{F}_n) - z_{1-\alpha/2} S_{jack}, f(\widehat{F}_n) + z_{1-\alpha/2} S_{jack} \right).$$

This interval is called the *jackknife* interval. It's often used in practice for estimating variance $D_F \widehat{\theta}(X_1, ..., X_n)$ of an estimator $\widehat{\theta}$. The jackknife estimator is

$$S^2_{jack} = \frac{n-1}{n} \sum_{i=1}^{n} (\widehat{\theta}(-i) - \overline{\widehat{\theta}})^2, \ \widehat{\theta}(-i) := \widehat{\theta}(X_1, ..., X_{i-1}, X_{i+1}, ..., X_n), \ \overline{\widehat{\theta}} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\theta}(-i).$$

Example 4. Let $\widehat{\theta}(X_1, ..., X_n) = \overline{X}$. Then

$$\widehat{\theta}(-i) = \frac{1}{n-1} (\overline{X} n - X_i), \ \overline{\widehat{\theta}} = \overline{X}.$$

So

$$S^2_{jack} = \frac{S_0^2}{n}.$$

It's a natural estimator for the $\mathbf{D}\overline{X}$.

Example 5. Consider a sample $X_1, ..., X_{2n+1}$ such that $X_{(n-1)} = X_{(n)} = X_{(n+1)}$ and let $\widehat{\theta}(X_1, ..., X_n) = MED$. Then $\widehat{\theta}(-i) = \widehat{\theta}(X_1, ..., X_n) = MED$ and $S^2_{jack} = 0$. In this situation the jackknife estimator is nonapplicable since $L_F(x; MED)$ is not continuous as functional of F.