Chapter 1

Quantile Estimation

1.1 Univariate quantile estimation

Given a sample $(y_1, y_2, ..., y_n)$, we are interested in estimating its τ-th quantile, where $\tau \in (0,1)$. To find the quantile from the sample, we can look for q such that

$$\frac{1}{n}\sum \mathbb{I}\left\{y_i \leq q\right\} \approx \tau. \tag{1.1}$$

If we ignore discreteness on the left-hand side, we solve the equation $\frac{1}{n}\sum \mathbb{I}\left\{y_i \leq q\right\} = \tau$. In this chapter, we always work with continuously distributed y. In the population model, q_τ^0 that solves $E\left[y \leq q\right] = \tau$ is the population parameter.

1.1.1 Asymptotic Result

(1.1) characterize the estimation by a method of moment. Now we cast the problem into an m-estimation. Introduce the check function

$$\rho\left(z\right) = z\left(\tau - \mathbb{I}\left(z \leq 0\right)\right).$$

Define $\psi\left(z\right) = \tau - \mathbb{I}\left(z \leq 0\right)$, which be considered as a **subgradient** of $\rho\left(z\right)$. Notice $\rho\left(z\right)$ is continuous but $\psi\left(z\right)$ is discontinuous.

Let

$$S_n\left(q\right) = \frac{1}{n}\sum \rho\left(y_i - q\right) = \frac{1}{n}\sum \left(y_i - q\right)\left(\tau - \mathbb{I}\left\{y_i - q \leq 0\right\}\right).$$

The first-order condition is

$$\frac{\partial}{\partial q}S_n\left(q\right) = \frac{1}{n}\sum\left(\tau - \mathbb{I}\left\{y_i - q \leq 0\right\}\right) \times (-1) = \frac{1}{n}\sum \mathbb{I}\left\{y_i \leq q\right\} - \tau \xrightarrow{p} F\left(y \leq q\right) - \tau.$$

The "second derivative" is

$$\begin{aligned}
\frac{\partial^2}{\partial q^2}S_n\left(q\right) &\triangleq \lim_{\delta \to 0} \frac{\frac{\partial}{\partial q}S_n\left(q+\delta\right) - \frac{\partial}{\partial q}S_n\left(q-\delta\right)}{2\delta} \\
&= \lim_{\delta \to 0} \frac{\frac{1}{n}\sum\left(\mathbb{I}\left\{y_i \leq q+\delta\right\} - \mathbb{I}\left\{y_i \leq q-\delta\right\}\right)}{2\delta} \\
&\xrightarrow{p} \lim_{\delta \to 0} \frac{F\left(y \leq q+\delta\right) - F\left(y \leq q-\delta\right)}{2\delta} \\
&= f_y\left(q\right)
\end{aligned}$$

where the above heuristic calculation implicitly assumes the exchangeability between $\lim_{\delta\to 0}$ and $\xrightarrow{p}$.

If true coefficient q_τ^0 is identified, then by ULLN we have

$$\hat{q} \xrightarrow{p} q_\tau^0.$$

Identification is equivalent to $f_y(q_\tau) > 0$.

Notice $\mathbb{I}\{y_i \leq q\} - \tau$ follows a binary distribution with variance $\tau(1-\tau)$. As a result,

$$\sqrt{n}\left(\hat{q} - q_\tau^0\right) \xrightarrow{d} N\left(0, \frac{\tau(1-\tau)}{f_y^2\left(q_\tau^0\right)}\right).$$

1.2 Quantile Regression

With other regression, use $X_i'\beta$ to mimic θ:

$$S_n(\beta) = \frac{1}{n}\sum \rho_\tau\left(y_i - X_i'\beta\right)$$

FOC

$$\begin{aligned}
\frac{\partial}{\partial\beta} S_n(\beta) &= \frac{1}{n}\sum X_i\left(\mathbb{I}\left\{y_i \leq X_i'\beta\right\} - \tau\right) \\
&\xrightarrow{p} E\left[X_i\psi\left(y_i - X_i'\beta\right)\right] \\
&= E\left[X_i\psi\left(y_i - X_i'\beta_\tau + X_i'\beta_\tau - X_i'\beta\right)\right] \\
&= E\left[X_i\psi\left(e_i + X_i'\left(\beta - \beta_\tau\right)\right)\right] \\
&= E\left[X_i\left(F_{e|X}\left(X_i'\left(\beta - \beta_\tau\right)\right) - \tau\right)\right]
\end{aligned}$$

SOC in the population version:

$$E\left[X_iX_i'f_{e|X}\left(X_i'\left(\beta - \beta_\tau\right)\right)\right]$$

Similarly, by ULLN and LD we have

$$\hat{\beta} \xrightarrow{p} \beta_\tau$$

The identification condition is $Q_\tau = E\left[XX'f_{e|X}(0)\right]$ which is positive definite.
The variance of the score function is

$$\Omega_\tau = E\left[XX'\psi^2(e)\right]$$

By asymptotic normality,

$$\sqrt{n}\left(\hat{\beta} - \beta_\tau\right) \xrightarrow{d} N\left(0, Q_\tau^{-1}\Omega_\tau Q_\tau^{-1}\right)$$

1.2.1 Conditional Quantile

$Q_{y|X}(\tau)$ is a function of X if the conditional quantile $Q_{y|X}(\tau) = X'\beta_\tau$.

$$\tau = F_{y|X}\left(X'\beta\right) = E\left[\mathbb{I}\left\{y \le X'\beta\right\} \mid X\right]$$

$$\begin{aligned}\Omega_\tau &= E\left[XX'E\left[\left(\mathbb{I}\left\{y \le X'\beta_\tau\right\} - \tau\right)^2 \mid X\right]\right] \\ &= \tau\left(1-\tau\right)E\left[XX'\right]\end{aligned}$$

Under this condition,

$$\sqrt{n}\left(\hat{\beta} - \beta_\tau\right) \xrightarrow{d} N\left(0, \tau\left(1-\tau\right)Q_\tau^{-1}E\left[XX'\right]Q_\tau^{-1}\right)$$

If we further assume e is independent of X,
then the Hessian is simplified as $Q_\tau = E\left[XX'\right]f_e\left(0\right)$
thus

$$\sqrt{n}\left(\hat{\beta} - \beta_\tau\right) \xrightarrow{d} N\left(0, \frac{\tau\left(1-\tau\right)}{f_e\left(0\right)}\left(E\left[XX'\right]\right)^{-1}\right)$$

`Zhentao Shi. Feb 21, 2023. Transcribed by Shu Shen.`