3.1 Estimation Theory

- Assume that the real $p \times 1$ vector time series $\{Y_t : t = 0, \pm 1, \cdots\}$ is $\mathcal{F}_t-$ measurable, strictly stationary and ergodic, and its conditional density function is given by

$$Y_t|\mathcal{F}_{t-1} \sim f(\theta_0, \tilde{Y}_{t-1}), \quad (1)$$

where $\mathcal{F}_t$ is the σ-field generated by $\{Y_t, Y_{t-1}, \cdots\}$, $\tilde{Y}_t = (Y_t, \cdots, Y_{t-p+1})$ or $\tilde{Y}_t = (Y_t, Y_{t-1}, \cdots)$, and θ_0 is an $m \times 1$ true parameter vector. The structure of the time series $\{Y_t\}$ is characterized by the density f and the parameter θ_0.

Given the random sample $\{Y_1, \cdots, Y_n\}$ and the initial value Y_0, the conditional joint density of $\{Y_1, \cdots, Y_n\}$ is

$$\prod_{t=1}^{n} f(\theta_0, \tilde{Y}_{t-1}).$$

Replacing the true parameter θ_0 by its unknown parameter θ, we have

$$\tilde{L}_n(\theta) \equiv \prod_{t=1}^{n} f(\theta, \tilde{Y}_{t-1}).$$

$\tilde{L}_n(\theta)$ is called the conditional likelihood function in term of θ.

$$L_n(\theta) \equiv \sum_{t=1}^{n} \log f(\theta, \tilde{Y}_{t-1})$$

is called the conditional log-likelihood function in term of θ.

Assume that the parameter space Θ is a compact subset of R^m and θ_0 and θ are in Θ. The maximizer of $\log L_n(\theta)$ on Θ is called the conditional MLE of θ_0, denoted by

$$\widehat{\theta}_n = \text{argmax}_{\Theta} L_n(\theta).$$

Example 3.1. Assume $\{Y_1, \cdots, Y_n\}$ are from the AR(1) model

$$Y_t = \phi Y_{t-1} + \varepsilon_t,$$

with true parameter ϕ_0, where ε_t is i.i.d $N(0, \sigma_0^2)$.

The conditional density of Y_t, given Y_{t-1}, is

$$f(Y_t|Y_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(Y_t - \phi_0 Y_{t-1})^2}{2\sigma_0^2}}.$$

Denote $\theta = (\phi, \sigma^2)'$ and $\theta_0 = (\phi_0, \sigma_0^2)'$. The conditional log-likelihood function in term of θ is

$$L_n(\theta) = -\frac{1}{2\sigma^2} \sum_{t=1}^{n} (Y_t - \phi Y_{t-1})^2 - n \log(\sqrt{2\pi\sigma^2}).$$

Denote

$$S_n(\phi) = \sum_{t=1}^{n} (Y_t - \phi Y_{t-1})^2.$$

The minimizer of $S_n(\phi)$ is called LSE and it is equivalent to the MLE of ϕ_0.

Example 3.2. Assume $\{Y_1, \cdots, Y_n\}$ are from the ARMA(1,1) model

$$Y_t = \phi Y_{t-1} - \psi \varepsilon_{t-1} + \varepsilon_t,$$

with true parameter (ϕ_0, ψ_0), where ε_t is i.i.d $N(0, \sigma_0^2)$.

The conditional density of Y_t, given $(Y_{t-1}, \varepsilon_{t-1})$, is

$$f(Y_t|Y_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(Y_t - \phi_0 Y_{t-1} + \psi_0 \varepsilon_{t-1})^2}{2\sigma_0^2}}.$$

Denote $\theta = (\phi, \psi, \sigma^2)'$ and $\theta_0 = (\phi_0, \psi_0, \sigma_0^2)'$. Given (Y_0, ε_0), the log- conditional joint density function of $\{Y_1, \cdots, Y_n\}$ is

$$\begin{aligned} L_n(\theta_0) &= -n \log(\sqrt{2\pi\sigma_0^2}) \\ &\quad - \frac{1}{2\sigma_0^2} \sum_{t=1}^{n} (Y_t - \phi_0 Y_{t-1} + \psi_0 \varepsilon_{t-1})^2. \end{aligned}$$

The problem is that we do not have ε_{t-1} !!

The first term is

$$Y_1 - \phi_0 Y_0 + \psi_0 \varepsilon_0.$$

Replaced θ_0 by the unknown parameter θ, we have

$$Y_1 - \phi Y_0 + \psi \varepsilon_0.$$

The second term is

$$Y_2 - \phi_0 Y_1 + \psi_0 \varepsilon_1,$$

where

$$\varepsilon_1 = Y_1 - \phi_0 Y_0 + \psi_0 \varepsilon_0.$$

Replaced θ_0 by the unknown parameter θ, the second term is

$$Y_2 - \phi Y_1 + \psi \varepsilon_1,$$

where

$$\varepsilon_1 = Y_1 - \phi Y_0 + \psi \varepsilon_0??$$

Denote

$$\varepsilon_1(\theta) = Y_1 - \phi Y_0 + \psi \varepsilon_0.$$

The second term with unknown parameter θ is

$$Y_2 - \phi Y_1 + \psi \varepsilon_1(\theta).$$

The third term is

$$Y_3 - \phi_0 Y_2 + \psi_0 \varepsilon_2,$$

where

$$\varepsilon_2 = Y_2 - \phi_0 Y_1 + \psi_0 \varepsilon_1.$$

Denote

$$\varepsilon_2(\theta) = Y_2 - \phi Y_1 + \psi \varepsilon_1(\theta).$$

The third term with unknown parameter θ is

$$Y_3 - \phi Y_2 + \psi \varepsilon_2(\theta).$$

Similarly, the t-th term is

$$Y_t - \phi Y_{t-1} + \psi \varepsilon_{t-1}(\theta),$$

where

$$\varepsilon_{t-1}(\theta) = Y_{t-1} - \phi Y_{t-2} + \psi \varepsilon_{t-2}(\theta).$$

Given (Y_0, ε_0), the log- conditional log-likelihood function of $\{Y_1, \cdots, Y_n\}$ is

$$\begin{aligned} L_n(\theta) &= -n \log(\sqrt{2\pi\sigma^2}) \\ &\quad -\frac{1}{2\sigma^2} \sum_{t=1}^{n} [\varepsilon_t(\theta)]^2. \end{aligned}$$

Denote

$$S_n(\theta) = \sum_{t=1}^{n} [\varepsilon_t(\theta)]^2.$$

The minimizer of $S_n(\theta)$ is called LSE and it is equivalent to the MLE of θ_0.

Example 3.3. Assume $\{\varepsilon_1, \cdots, \varepsilon_n\}$ are from the GARCH(1,1) models:

$$\begin{cases} \varepsilon_t = \eta_t \sqrt{h_t}, \\ h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}, \end{cases} \tag{2}$$

with true parameter $\theta_0 = (\omega_0, \alpha_0, \beta_0)'$, where $\eta_t \sim$iid $N(0,1)$.

The conditional density of ε_t, given $(\varepsilon_{t-1}, h_{t-1})$, is

$$f(\varepsilon_t|\varepsilon_{t-1}) = \frac{1}{\sqrt{2\pi h_t}} e^{-\frac{\varepsilon_t^2}{2h_t}}.$$

Given (ε_0, h_0), the log- conditional joint density function of $\{\varepsilon_1, \cdots, \varepsilon_n\}$ is

$$L_n(\theta_0) \ = \ -\frac{1}{2}\sum_{t=1}^{n}[\log h_t + \frac{\varepsilon_{t-1}^2}{h_t}].$$

The problem is that we do not have h_t !!

$$h_1 = \omega_0 + \alpha_0\varepsilon_0^2 + \beta_0 h_0.$$

Denote $\theta = (\omega, \alpha, \beta)'$. Replaced by the unknown parameter θ, we have

$$h_1(\theta) = \omega + \alpha\varepsilon_0^2 + \beta h_0.$$

$$h_2 = \omega_0 + \alpha_0 \varepsilon_1^2 + \beta_0 h_1.$$

Replaced by the unknown parameter θ,

$$h_2(\theta) = \omega + \alpha \varepsilon_1^2 + \beta h_1(\theta).$$

Similarly, the t-th term is

$$h_t(\theta) = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}(\theta).$$

The log- conditional likelihood function of $\{\varepsilon_1, \cdots, \varepsilon_n\}$ is

$$L_n(\theta) \;=\; -\frac{1}{2} \sum_{t=1}^{n} \left[\log h_t(\theta) + \frac{\varepsilon_{t-1}^2}{h_t(\theta)} \right].$$

We assume that the parameter space Θ is a compact subset of R^m, and the true value θ_0 of θ is an interior point in Θ. We use the following OF with the initial value Y_0 to estimate θ_0:

$$L_n(\theta) = \sum_{t=1}^{n} l_t(\theta),$$

where $l_t(\theta) = l(\tilde{Y}_t, \theta)$ is a measurable function with respect to $\tilde{Y}_t$ and is continuous in terms of θ. The estimator of θ_0 the maximizer of $L_n(\theta)$ on Θ, i.e.

$$\widehat{\theta}_n = \mathrm{argmax}_{\Theta} L(\theta).$$

When the dimension of the initial value $\tilde{Y}_0$ is infinite, it need to be replaced by some constant $\widetilde{Y}^*$. To make it simple, we assume that $\tilde{Y}_0$ is available.

We only need the identification condition for the consistency of $\widehat{\theta}_n$ as follows.

Assumption 2.1.

$E \sup_{\theta \in \Theta}[l_t(\theta)] < \infty$, and $E[l_t(\theta)]$ has a unique maximizer at θ_0.

Theorem 2.1 *If Assumptions 2.1 holds, then* $\widehat{\theta}_n \to \theta_0$ *a.s.*.

Proof. By the ergodic theorem,

$$\frac{1}{n}\sum_{t-1}^{n} l_t(\theta) \to E[l_t(\theta)] \; a.s.$$

θ_0 is the unique maximizer of $E[l_t(\theta)]$. $\widehat{\theta}_n$ is maximizer of $\frac{1}{n}\sum_{t-1}^{n} l_t(\theta)$. Intuitively, $\widehat{\theta}_n \to \theta_0$.

For asymptotic normality, we assume that $l_t(\theta)$ has continuously twice differentiable almost surely (a.s.) in terms of θ.

Denote

$$D_t(\theta) = \frac{\partial l_t(\theta)}{\partial \theta} \ \ and \ \ P_t(\theta) = -\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'},$$

$\Sigma = E[P_t(\theta_0)]$ and $\Omega = E[D_t(\theta_0)D_t'(\theta_0)]$. We need one assumption as follows:

Since the true value θ_0 of θ is an interior point in Θ and $\widehat{\theta}_n \to \theta_0$. As n is large, $\widehat{\theta}_n$ is an interior point in Θ. Thus,

$$\frac{1}{n}\sum_{t=1}^{n} D_t(\widehat{\theta}_n) = 0.$$

By Taylor's expansion,

$$\frac{1}{n}\sum_{t=1}^{n} D_t(\widehat{\theta}_n) - \frac{1}{n}\sum_{t=1}^{n} D_t(\theta_0) = \frac{1}{n}\sum_{t=1}^{n} P_t(\widehat{\xi}_n)(\widehat{\theta}_n - \theta_0).$$

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = -\left[\frac{1}{n}\sum_{t=1}^{n} P_t(\widehat{\xi}_n)\right]^{-1} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} D_t(\theta_0).$$

Assumption 2.2. $0 < \Sigma$, $\Omega < \infty$, and the following holds:

(i) $D_t(\theta_0)$ is a martingale difference in terms of $\mathcal{F}_t$;

(ii) $E\sup_{\theta \in V_0(\eta)} \|P_t(\theta)\| < \infty$ for some $\eta > 0$, where $V_0(\eta) = \{\theta : \|\theta - \theta_0\| < \eta\}$.

We now state our second result as follows.

Theorem 2.2 *If $\widehat{\theta}_n \to \theta_0$ a.s. and Assumptions 2.2 holds, then*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \longrightarrow_{\mathcal{L}} N(0, \Sigma^{-1}\Omega\Sigma^{-1}).$$

Σ and Ω are estimated by

$$\widehat{\Sigma}_n = \frac{1}{n}\sum_{t=1}^{n} P_t(\widehat{\theta}_n)$$

$$\widehat{\Omega}_n = \frac{1}{n}\sum_{t=1}^{n} [D_t(\widehat{\theta}_n)D_t(\widehat{\theta}_n)'].$$

$\widehat{\Sigma}^{-1}\widehat{\Omega}_n\widehat{\Sigma}^{-1}$ is the estimator of $\Sigma^{-1}\Omega\Sigma^{-1}$. Denote

$$\widehat{\Sigma}_n^{-1}\widehat{\Omega}_n\widehat{\Sigma}_n^{-1} = (\widehat{\sigma}_{ij})_{m\times m}.$$

Then

$$\sqrt{n}(\widehat{\theta}_{in} - \theta_{i0}) \sim N(0, \widehat{\sigma}_{ii}).$$

t-test for $\theta_{i0} = 0$ v.s. $\theta_{i0} \neq 0$:

$$\frac{\sqrt{n}(\widehat{\theta}_{in} - 0)}{\sqrt{\widehat{\sigma}_{ii}}} \sim N(0, 1).$$

Note that $l_t(\theta) = l(\tilde{Y}_t, \theta)$. The function $l(\cdot, \theta)$ does not depend on t and hence $\{l_t(\theta)\}$ is strictly stationary and ergodic. When $\tilde{Y}_t = (Y_t, Y_{t-1}, \cdots)$ is an infinite dimensional vector, we need the initial value $\tilde{Y}_0$ as in ARMA(1,1) and GARCH(1,1). This involves an initial effect problem.