Consistency of extremum estimators

Binary choice

$y_i^* = x_i'\beta + \varepsilon_i$, $\varepsilon_i | x_i \sim F(\cdot)$

- known, symmetric around zero
- refer to this as the link

Observe $y_i = 1\{y_i^* \geq 0\}$

Then $\Pr[y_i = 1 | x_i] = \Pr[\varepsilon_i > -x_i'\beta | x_i]$

$= 1 - F(-x_i'\beta)$

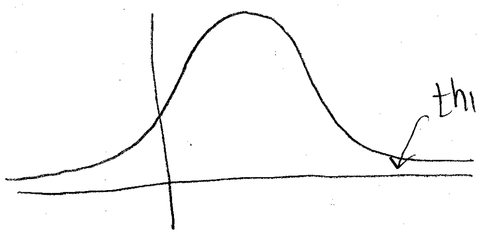
$= F(x_i'\beta)$

Can choose $F(t) = \Lambda(t) = \dfrac{\exp\{t\}}{1+\exp\{t\}}$

- Typically assume that F is known.
- Cannot estimate the scale of β (ie decisions are the same under β and $k\beta$ for $k>0$, and all we observe are the decisions.)

1. Probit: $F(t) = \Phi(t)$ (ie $\varepsilon \sim N(0,1)$)
2. Cauchit: $F(t) = C(t) = \frac{1}{2} + \frac{1}{\pi}\arctan(t)$ (ie $\varepsilon \sim$ Cauchy
3. Gosset: $F(t) = T(t, v)$ (ie $\varepsilon \sim t(\underbrace{v}_{\text{deg of freedom}})$)

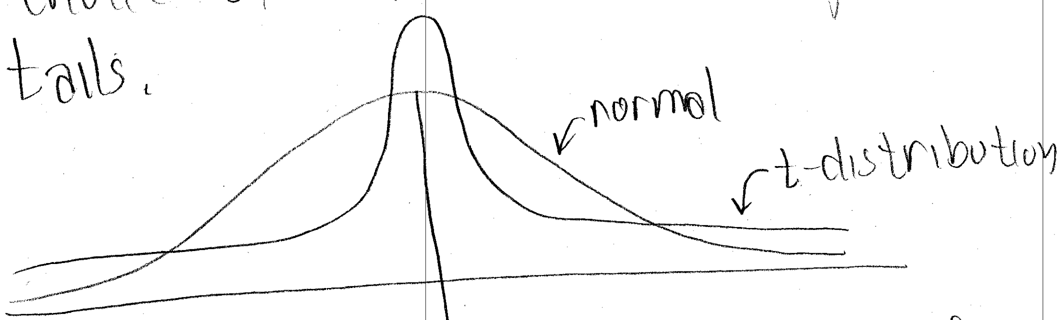
Suppose $\underbrace{\varepsilon_i}_{\substack{\text{logistic} \\ \text{Gumbell}}} = \ln \underbrace{E_i}_{\text{exponential}} = \ln[-\ln \underbrace{U_i}_{\text{uniform}}]$



- not centered around zero
- skewed
- thin tails: exponential tails
- close to normal

Extreme value distributions approximate the distribution of $\underline{Y} = \max\{\underline{X}_1, \ldots, \underline{X}_n\}$, where $\underline{X}_i \overset{iid}{\sim}$ Normal

- Koenker and Yoon discusses the choice of links
  - the choice of links is most important in the tails.



- What is important is that choice of link affects $\widehat{\Pr[y_i=1|x_i]}$ and $\dfrac{\partial \widehat{\Pr[y_i=1|x_i]}}{\partial x}$
  - If you are using $\widehat{\Pr[y_i=1|x_i]}$ as an instrument in another equation, this is important!

<u>Extremum consistency</u>

Let $\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}}\ \hat{Q}(\theta)$

Thm: If 1] $Q(\theta)$ is uniquely minimized at θ_0,

2] Θ is compact

3] $Q(\cdot)$ is continuous

4] $\sup_{\theta \in \Theta} |\hat{Q}(\theta) - Q(\theta)| \xrightarrow{P} 0$, then $\hat{\theta} \xrightarrow{P} \theta_0$

Pf: By uniform convergence, $\hat{Q}(\hat{\theta}) - Q(\hat{\theta}) \xrightarrow{P} 0$ and $\hat{Q}(\theta_0) - Q(\theta_0) \xrightarrow{P} 0$. Since $\hat{\theta} = \arg\min \hat{Q}(\theta)$ and $\theta_0 = \arg\min Q(\theta)$, we have:

$$
\begin{aligned}
Q(\theta_0) \le Q(\hat{\theta}) &= \hat{Q}(\hat{\theta}) + [Q(\hat{\theta}) - \hat{Q}(\hat{\theta})] \\
&\le \hat{Q}(\theta_0) + [Q(\hat{\theta}) - \hat{Q}(\hat{\theta})] \\
&= Q(\theta_0) + [\underbrace{\hat{Q}(\theta_0) - Q(\theta_0)}_{o_p(1)} + \underbrace{Q(\hat{\theta}) - \hat{Q}(\hat{\theta})}_{o_p(1)}] \\
&= Q(\theta_0) + o_p(1)
\end{aligned}
$$

$$\Rightarrow Q(\theta_0) \le Q(\hat{\theta}) \le Q(\theta_0) + o_p(1)$$

$$\Rightarrow Q(\hat{\theta}) \xrightarrow{P} Q(\theta_0).$$

Since Θ is compact and Q is continuous, we have that $\forall$ open subset N of Θ containing θ_0, $\inf_{\theta \notin N} Q(\theta) = Q(\theta^*)$ for some $\theta^* \in \Theta$. By identification, $Q(\theta^*) > Q(\theta_0)$, which gives us

$$\inf_{\theta \notin N} Q(\theta) > Q(\theta_0)$$

Since $Q(\hat{\theta}) \xrightarrow{P} Q(\theta_0)$, $Q(\hat{\theta}) < \inf_{\theta \notin N} Q(\theta)$ with probability approaching one.

Thus, $\hat{\theta} \in N$ $\forall N$, and we have $\hat{\theta} \xrightarrow{P} \theta_0$.

1] ensure that $Q(\theta)$ is minimized at the value of θ that you want.

2] Can generalize this in several directions

(*) Epi-convergence and Stochastic Equi-Semicontinuity is a good reference for micro theory and econometric theory.

3] Don't worry about measurability

We can derive uniform convergence from pointwise convergence from the following Lemma.

Lemma 1: Let Θ be compact, and suppose

i) $\hat{Q}(\theta) \xrightarrow{p} Q(\theta) \quad \forall \theta \in \Theta$ and

ii) (stochastic equicontinuity) $\forall \eta > 0 \; \exists \varepsilon > 0$ and n_ε s.t. $\forall n > n_\varepsilon$,

$$\Pr{}^*\left[\sup_{\|\theta - \theta'\| > \varepsilon} |\hat{Q}(\theta) - \hat{Q}(\theta')| > \eta\right] \leq \eta$$

Then $\hat{\theta} \xrightarrow{p} \theta$.

Lemma 2: Suppose $\hat{Q}(\theta) \xrightarrow{p} Q(\theta) \quad \forall \theta \in \Theta$ and

either

a) Holder: for some $h > 0$,

$|\hat{Q}(\theta) - \hat{Q}(\theta')| \leq B_T \|\theta - \theta'\|^h$, $B_T = O_p(1)$ or

b) $\hat{Q}(\theta)$ is concave or convex over a convex $\Theta \subseteq \mathbb{R}^k$ and $Q(\theta)$ is continuous then uniformly convergence holds.