The Symposium of Frontiers of Statistics and Data Sciences

June 25-26, 2016, jointly sponsored by

The Hong Kong Polytechnic University and

The Hong Kong Statistical Society

Invited Talk on 25 June 2016, 11:40-12:05, Room Y303

Adequate Sample Size for Using Limiting Distributions

Speaker: Professor Kai Wang NG

Founding Patrick S C Poon Professor in Statistics & Actuarial Science

and retired as Honorary Professor, The University of Hong Kong

(URL: http://www.saasweb.hku.hk/staff/kaing)

Adjunct Professor, University of Alberta, Canada

- **Asymptotic inference means that the inference on $(\theta_1, \cdots, \theta_k)$ based on a CDF$(n, \theta_1, \cdots, \theta_k)$ can be substituted by the inference based on another CDF$(\theta_1, \cdots, \theta_k)$ which is the limit of CDF$(n, \theta_1, \cdots, \theta_k)$ as $n \to \infty$. But these theories so far have not provided any guideline on the size of n such that the aforesaid substitution is good enough.**
- **With high respect and deep trust, and under publication pressure, practitioners happily follow suit as if there is no need to query about n.**
- **Eg. a normal distribution is always assumed for an MLE in such "asymptotic inference" regardless the size of n.**
- **Such practice is pervasive in financial mathematics and econometrics, where the number of model parameters can be very large while the sample size needed for reliable asymptotic inference is never discussed.**
- "*Bright halos often outshine simple truths – anonymous*"
- **Look at two facts that make the size of n crucial.**

(a) Basic Principle for Normality

If X is normal, $g(X)$ cannot be normal for non-linear $g(\cdot)$.

(b) Invariance Principle for Limiting Normality

For θ, its MLE $\hat{\theta}_n \sim N(\theta, \sigma^2(n, \theta))$ *as* $n \to \infty$; and for $\tau = g(\theta)$ where $g(\cdot)$ is any continuous one-one transform, its MLE $\hat{\tau}_n = g(\hat{\theta}_n) \sim N(\ g(\theta), (g'(\theta))^2 \sigma^2(n, g(\theta))\ )$ *as* $n \to \infty$.

Fact 1: Conflict between (a) and (b)

For a finite n at hand, if there is a nonlinear $g(\cdot)$ such that $\hat{\tau} = g(\hat{\theta})$ is very closed to normality then $\hat{\theta}$ is not close to normality. Thus it is *very lucky* that the parametrization at hand is the right one closest to normality.

This explains why Fisher (the founder of MLE) didn't use limiting normality of MLE $\hat{\rho}_n$ and spent a long time until he was satisfied with the reparametrizeation $log((1+\hat{\rho}_n)/(1-\hat{\rho}_n))$.

Fact 2: Non-uniform convergence over parameter values

For the parametrization at hand, the convergence to normality is not uniform over all parameter values.

This can be demonstrated by the ARMA$(1,1)$ model:

$$x_t - \phi x_{t-1} = z_t - \theta z_{t-1}, \quad z_t \sim N(0,1)$$

where $|\phi| < 1, |\theta| < 1,$ and $\phi \neq \theta$ (otherwise not identifiable).

NOTE: The stationary dist. of any segment of ARMA$(1,1)$ is multivariate normal with cov matrix as a known function of (θ, ϕ) and thus can be simulated exactly without the "burn-in" method used in common textbooks.

If n is adequate for normality, the $(1-\alpha)$ confidence intervals of ϕ and θ are respectively

$$\hat{\phi}_n \pm z_{\alpha/2}\sqrt{\frac{(1-\hat{\phi}_n^2)(1-\hat{\phi}_n\hat{\theta}_n)^2}{n(\hat{\phi}_n-\hat{\theta}_n)^2}}, \quad \hat{\theta}_n \pm z_{\alpha/2}\sqrt{\frac{(1-\hat{\theta}_n^2)(1-\hat{\phi}_n\hat{\theta}_n)^2}{n(\hat{\phi}_n-\hat{\theta}_n)^2}}$$

$\therefore$ Check normality by true coverage of C.I.

Based on 10,000 stationary series of $ARMA(1,1)$

Coverage % of C.I. constructed from converged $\hat{\phi}_n$ and $\hat{\theta}_n$

$n/12$	n	**Conv**	ϕ**:95%**	ϕ**:99%**	θ**:95%**	θ**:99%**
		$\phi = 0.5$ **and** $\theta = -0.5$				
20	**240**	**8826**	**94.88**	**98.70**	**93.76**	**98.05**
30	**360**	**9256**	**94.57**	**98.87**	**94.06**	**98.47**
40	**480**	**9482**	**94.83**	**98.78**	**94.60**	**98.65**
		$\phi = -0.5$ **and** $\theta = 0.5$				
20	**240**	**8552**	**94.80**	**98.87**	**94.27**	**98.36**
30	**360**	**9116**	**94.96**	**98.94**	**94.21**	**98.40**
40	**480**	**9428**	**94.94**	**99.13**	**94.65**	**98.62**
		$\phi = 0.6$ **and** $\theta = 0.8$				
120	**1440**	**9323**	**93.67**	**98.05**	**92.99**	**97.09**
240	**2880**	**9774**	**94.14**	**98.52**	**93.71**	**97.88**
480	**5760**	**9949**	**94.86**	**98.79**	**94.19**	**98.52**
		$\phi = 0.8$ **and** $\theta = 0.9$				
480	**5760**	**9556**	**94.32**	**98.37**	**93.56**	**97.89**
960	**11520**	**9877**	**94.30**	**98.71**	**94.05**	**98.30**

Box-Jenkins ARMA(p, q) model with $2 + p + q$ parameters:

$$(x_t - \mu) - \phi_1(x_{t-1} - \mu) - \cdots - \phi_p(x_{t-p} - \mu) = z_t - \theta_1 z_{t-1} - \cdots - \theta_q z_{t-q}$$

Adding the variance structure for z_t, Engle received a Nobel Economic Prize with GARCH(r, s) model:

$$z_t = h^{1/2} e_t, \quad e_t \sim iid\ N(0, 1), \quad h_t = \omega + \sum_{i=1}^{s} \alpha_i z_{t-1}^2 + \sum_{j=1}^{r} \gamma_j h_{t-j}$$

Further variance structures generate EGARCH, IGARCH, HIGARCH ....

Multivariate series models with covariance structures like WAR(r), CAW(p, q), GCAW(p, q, r), etc. even more complex.

Q: Relative to the above series length n for ARMA$(1, 1)$, how many times of n are needed for these complex models?

My Humble Reminder:

“Adequate sample-size is a very important issue

surely in practical work

although not so considered in

teaching/publishing/selling asymptotics”

Thank you for listening!