Spring 2007 John Rust
Economics 623 Tydings 4115A

Empirical Process Proof of the Asymptotic Distribution of Sample Quantiles

Definition: *Given $\theta \in (0,1)$, the θ^{th}* **quantile** *of a random variable $\tilde{X}$ with CDF F is defined by:*

$$\mu_\theta \equiv F^{-1}(\theta) = \inf\{x \,|\, F(x) \geq \theta\}.$$

Note that $\mu_{.5}$ is the *median,* $\mu_{.25}$ is the 25^{th} percentile, etc. Further if we define the 0^{th} quantile as $\mu_0 = \lim_{\theta\to 0}\mu_\theta$ and define μ_1 similarly, it is easy to see that these are the lower and upper points in the support of $\tilde{X}$ (i.e. the minimum and maximum possible values of $\tilde{X}$ which might be $-\infty$ and $+\infty$ if $\tilde{X}$ has unbounded support). Note also that if F is strictly increasing in a neighborhood of μ_θ, then $\mu_\theta = F^{-1}(\theta)$ is the usual inverse of the CDF F. If F happens to have "flat" sections, say an interval of points x satisfying $F(x) = \theta$, then μ_θ is the smallest x in this interval. The following lemma, a slightly modified version of a lemma from R. J. Serfling, (1980) *Approximation Theorems of Mathematical Statistics* Wiley, New York, provides some basic properties of the *quantile function* $F^{-1}(\theta)$:

Lemma 1: *Let F be a CDF. The quantile function $F^{-1}(\theta)$, $\theta \in (0,1)$ is non-decreasing and left continuous, and satisfies:*

1. $F^{-1}(F(x)) \leq x, \quad -\infty < x < \infty$
2. $F(F^{-1}(\theta)) \geq \theta, \quad 0 < \theta < 1$
3. If F is strictly increasing in a neighborhood of $\mu_\theta = F^{-1}(\theta)$ we have: $F(F^{-1}(\theta)) = \theta$ and $F^{-1}(F(\mu_\theta)) = \mu_\theta$.
4. $F(x) \geq \theta$ *if and only if* $x \geq F^{-1}(\theta)$.

Definition: *Let $(\tilde{X}_1,\ldots,\tilde{X}_N)$ be a random sample of size N from a CDF F. Then the* **sample quantile** *$\hat{\mu}_\theta$, $\theta \in (0,1)$ is defined by:*

$$\hat{\mu}_\theta = F_N^{-1}(\theta),$$

where F_N is the empirical CDF defined by:

$$F_N(x) = \frac{1}{N}\sum_{i=1}^N I\{\tilde{X}_i \leq x\}.$$

Thus $\hat{\mu}_{.5} = F_N^{-1}(.5)$ is the *sample median* $F_N^{-1}(.5) = \text{med}(\tilde{X}_1,\ldots,\tilde{X}_N)$, and $\hat{\mu}_0 = F_N^{-1}(0)$ is the *sample minimum,* $F_N^{-1}(0) = \min(\tilde{X}_1,\ldots,\tilde{X}_N)$, and $\hat{\mu}_1 = F_N^{-1}(1)$ is the *sample maximum,* $F_N^{-1}(1) = \max(\tilde{X}_1,\ldots,\tilde{X}_N)$. Since empirical CDF's have jumps of size $1/N$ (unless more than one of the $\{\tilde{X}_i\}$'s take the same value), then we can bound the maximum difference between θ and $F(F^{-1}(\theta))$ in Lemma 1-2 as follows:

Lemma 2: *Let $(\tilde{X}_1, \ldots, \tilde{X}_N)$ be a random sample from a CDF F and suppose that in this sample each $\tilde{X}_i$ happens to be distinct, so that by reindexing we have $\tilde{X}_1 < \tilde{X}_2 < \cdots < \tilde{X}_{N-1} < \tilde{X}_N$. Then for all $\theta \in (0,1)$ we have:*

$$|F_N(F_N^{-1}(\theta)) - \theta| \le \frac{1}{N}.$$

The following example might help to illustrate the meaning of Lemma 2. Suppose that $N = 3$ and $\theta = .5$. Then the empirical median of the three points $\{\tilde{X}_1, \tilde{X}_2, \tilde{X}_3\}$ is just $\tilde{X}_3 = \hat{\mu}_{.5} = F_N^{-1}(.5)$. The reason is that $\tilde{X}_3$ is the smallest value of x satisfying

$$F_N(x) \ge .5 \tag{1}$$

However it is clear that

$$F_N(\hat{\mu}_{.5}) = F_N(F_N^{-1}(.5)) = F_N(\tilde{X}_3) = \frac{2}{3} \neq .5 \tag{2}$$

but it still is the case that

$$|F_N(F_N^{-1}(.5)) - .5| = |\frac{2}{3} - .5| \le \frac{1}{N} = \frac{1}{3},$$

consistent with Lemma 2.

The following theorem shows that the asymptotic distribution of the sample quantiles $\hat{\mu}_\theta$ for $\theta \in (0,1)$ are normally distributed. It is important to note that we exclude the two cases $\theta = 0$ and $\theta = 1$ in this theorem since the asymptotic distribution of these *extreme value statistics* is very different and generally non-normal.

Theorem: *Let $(\tilde{X}_1, \ldots, \tilde{X}_N)$ be IID draws from a CDF F with continuous density f. Then if $f(\mu_\theta) > 0$, we have:*

$$\sqrt{N}\,(\hat{\mu}_\theta - \mu_\theta) = \sqrt{N}\left(F_N^{-1}(\theta) - F^{-1}(\theta)\right) \Longrightarrow N(0, \sigma^2),$$

where:

$$\sigma^2 = \frac{\theta(1-\theta)}{f(\mu_\theta)^2}.$$

Proof: The Central Limit Theorem for *IID* random variables implies that for any x in the support of F we have:

$$\sqrt{N}\,(F_N(x) - F(x)) \Longrightarrow N(0, \gamma^2),$$

where $\gamma^2 = F(x)[1 - F(x)]$. Letting $x = \mu_\theta = F^{-1}(\theta)$ and using result 3 of Lemma 1 we have:

$$\sqrt{N}\,(F_N(\mu_\theta) - F(\mu_\theta)) = \sqrt{N}\left(F_N(F^{-1}(\theta)) - F(F^{-1}(\theta))\right) \Longrightarrow N(0, \theta(1-\theta)).$$

Furthermore, the property of *stochastic equicontinuity* from the theory of *empirical processes* (see D. Andrews, (1996) *Handbook of Econometrics* (vol. 4) for an accessible introduction and definition of stochastic equicontinuity), we have that the result given above is unaffected if we replace μ_θ by a consistent estimate $\hat{\mu}_\theta$:

$$\sqrt{N}\,(F_N(\hat{\mu}_\theta) - F(\hat{\mu}_\theta)) = \sqrt{N}\left(F_N(F_N^{-1}(\theta)) - F(F_N^{-1}(\theta))\right) \Longrightarrow N(0, \theta(1-\theta)).$$

Now note that part 2. of Lemma 1 implies that

$$\sqrt{N}\left(F_N(F_N^{-1}(\theta)) - F(F_N^{-1}(\theta))\right) \geq \sqrt{N}\left(\theta - F(F_N^{-1}(\theta))\right).$$

However since the true CDF F has a density, the probability of observing duplicate $\{\tilde{X}_i\}$'s is zero, so Lemma 2 implies that with probability 1 we have:

$$\sqrt{N}\left(F_N(F_N^{-1}(\theta)) - F(F_N^{-1}(\theta))\right) = \sqrt{N}\left(\theta - F(F_N^{-1}(\theta))\right) + O_p(1/\sqrt{N}),$$

which implies that:

$$\sqrt{N}\left(\theta - F(F_N^{-1}(\theta))\right) \Longrightarrow N(0, \theta(1-\theta)).$$

Now we apply the Delta theorem, i.e. we do a Taylor series expansion of $F(F_N^{-1}(\theta))$ about the limiting point $\mu_\theta = F^{-1}(\theta)$ to get:

$$F(F_N^{-1}(\theta)) = F(F^{-1}(\theta)) + f(\tilde{\mu}_\theta)\left[F_N^{-1}(\theta) - F^{-1}(\theta)\right],$$

where $\tilde{\mu}_\theta$ is a point on the line segment between $\hat{\mu}_\theta$ and μ_θ. Rewriting this, assuming that $f(\tilde{\mu}_\theta) > 0$, we have

$$\sqrt{N}\left[F_N^{-1}(\theta) - F^{-1}(\theta)\right] = \frac{-1}{f(\tilde{\mu}_\theta)}\sqrt{N}\left(\theta - F(F_N^{-1}(\theta))\right), \tag{3}$$

where we used result 3 of Lemma 1 to write $F(F^{-1}(\theta)) = \theta$ since F is assumed to be strictly increasing with positive density $f(\mu_\theta)$ at $\mu_\theta = F^{-1}(\theta)$. Using the result above and result 3 of Lemma 1 we have:

$$\sqrt{N}\left(F_N^{-1}(\theta) - F^{-1}(\theta)\right) = \sqrt{N}\left(\frac{\theta - F(F_N^{-1}(\theta))}{-f(\tilde{\mu}_\theta)}\right) \Longrightarrow N(0, \sigma^2),$$

where we have used Slutsky's Theorem and the fact that $\tilde{\mu}_\theta \to \mu_\theta$ since $\hat{\mu}_\theta \to \mu_\theta$ with probability 1.