A course in large sample theory Winter 2022

Asymptotic theory of sample quantiles and extreme statistics

Advisor: Peirong Xu *Speaker: Bowen Zhou*

1 Asymptotic distribution of sample quantiles

Let $X_1, \cdots, X_n$ be i.i.d. sample from distribution F on the real line and assume F is continuous so that all observations are distinct with probability 1. Rearranging the observations in creasing order, $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$. These variables are called order statistics. We first derive the distribution of order statistics. We assume X has density f in the following discussion.

Lemma 1. *Let $X_1, \cdots, X_n$ be i.i.d. sample from distribution F, and they have density f. Then,*

(a) The density of the order statistic $X_{(m)}$ is

$$f_m(x) = m \left(\begin{array}{c} n \\ m \end{array} \right) (F(x))^{m-1}(1 - F(x))^{n-m} f(x).$$

(b) The joint density of $(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ is

$$g(y_1, y_2, \cdots, y_n) = \begin{cases} n! f(y_1) f(y_2) \cdots f(y_n), & y_1 < y_2 < \cdots < y_n, \\ 0, & \text{otherwise.} \end{cases}$$

For X, the transformation $U = F(X)$ follows a uniform distribution, $\text{Unif}(0, 1)$. So we can study the asymptotic properties of uniform distribution and then derive the general result using the inverse transformation $g(u) = F^{-1}(u)$ in Cramer's theorem.

Theorem 2 (Cramer)**.** *Let $\boldsymbol{g}$ be a function such that $\dot{\boldsymbol{g}}$ is continuous in a neighborhood of $\boldsymbol{\mu}$. $\boldsymbol{X}_n$ is a sequence of random vectors such that $\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu}) \xrightarrow{d} \boldsymbol{X}$ then*

$$\sqrt{n}(\boldsymbol{g}(\boldsymbol{X}_n) - \boldsymbol{g}(\boldsymbol{\mu})) \xrightarrow{d} \dot{\boldsymbol{g}}(\boldsymbol{\mu}) \boldsymbol{X}.$$

Particularly, if $\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu}) \xrightarrow{d} \mathsf{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$, then

$$\sqrt{n}(\boldsymbol{g}(\boldsymbol{X}_n) - \boldsymbol{g}(\boldsymbol{\mu})) \xrightarrow{d} \mathsf{N}(\boldsymbol{0}, \dot{\boldsymbol{g}}(\boldsymbol{\mu}) \boldsymbol{\Sigma} \dot{\boldsymbol{g}}(\boldsymbol{\mu})^\top).$$

Now we suppose $U_1, \cdots, U_n$ be i.i.d. sample of $\text{Unif}(0, 1)$, and $U_{(1)}, \cdots, U_{(n)}$ are order statistics. Then by lemma 1, the density of $U_{(m)}$ is

$$f_m(u) = m \left(\begin{array}{c} n \\ m \end{array} \right) u^{m-1}(1 - u)^{n-m},$$

and joint density of $(U_{(1)}, \cdots, U_{(n)})$ is

$$g(y_1, y_2, \cdots, y_n) = \begin{cases} n!, & y_1 < y_2 < \cdots < y_n, \\ 0, & \text{otherwise.} \end{cases}$$

Then we derive the mean and variance of $U_{(m)}$. The beta function and gamma function may be useful there.

Lemma 3. *The mean and variance of the order statistic $U_{(m)}$ is*

$$\mathbb{E}(U_{(m)}) = \frac{m}{n+1}, \quad \mathrm{Var}(U_{(m)}) = \frac{1}{n+2}\left(\frac{m}{n+1}\right)\left(1-\frac{m}{n+1}\right).$$

In the following discussion, we assume $m/n \to p$ when $k, n \to \infty$ and $0 < p < 1$. In this case,

$$\mathbb{E}(U_{(m)}) \to p, \quad \mathrm{Var}(\sqrt{n}U_{(m)}) \to p(1-p),$$

so we may expect that

$$\sqrt{n}(U_{(m)} - p) \to \mathsf{N}(0, p(1-p)).$$

If we consider two order statistics, $U_{(m_1)} < U_{(m_2)}$, we first calculate the covariance.

Lemma 4. *For $U_{(m_1)} < U_{(m_2)}$, we have*

$$\mathrm{Cov}(U_{(m_1)}, U_{(m_2)}) = \frac{m_1(n+1-m_2)}{(n+1)^2(n+2)}.$$

In the case that $m_1/n \to p_1$ and $m_2/n \to p_2$, then

$$\sqrt{n}\mathrm{Cov}(U_{(m_1)}, U_{(m_2)}) \to p_1(1-p_2),$$

so we may expect that

$$\sqrt{n}\begin{pmatrix} U_{(m_1)} - p_1 \\ U_{(m_2)} - p_2 \end{pmatrix} \xrightarrow{d} \mathsf{N}\left(\mathbf{0}, \begin{pmatrix} p_1(1-p_1) & p_1(1-p_2) \\ p_1(1-p_2) & p_2(1-p_2) \end{pmatrix}\right).$$

To prove (1) and (1), we show first that the joint distribution of $(U_{(1)}, \cdots, U_{(n)})$ has the well-known representation as the distribution of ratios of waiting times.

Lemma 5. *Let $Y_1, \cdots, Y_n$ be i.i.d. sample of $Y \sim \exp(1)$, and let $S_j = \sum_{i=1}^j Y_i$ for $j = 1, \cdots, n+1$. Then the distribution of*

$$\left(\frac{S_1}{S_{n+1}}, \cdots, \frac{S_n}{S_{n+1}}\right)$$

is the same as the order statistics of a sample of size n from Unif$(0,1)$.

Proof. First the joint density of $(Y_1, \cdots, Y_{n+1})$ is

$$f_Y(y_1, \cdots, y_{n+1}) = \exp\left\{-\sum_{i=1}^{n+1} y_i\right\} I(y_i > 0, \forall i),$$

Taking trasformation

$$\begin{cases} S_1 = Y_1, \\ S_2 = Y_1 + Y_2, \\ \quad\vdots \\ S_{n+1} = Y_1 + \cdots + Y_{n+1}, \end{cases}$$

the joint density of $(S_1, \cdots, S_{n+1})$ is

$$f_S(s_1, \cdots, s_{n+1}) = \exp\{-s_{n+1}\} I(0 < s_1 < \cdots < s_{n+1}).$$

Then taking transformation

$$\begin{cases} Z_1 = \frac{S_1}{S_{n+1}}, \\ Z_2 = \frac{S_2}{S_{n+1}}, \\ \quad \vdots \\ Z_n = \frac{S_n}{S_{n+1}}, \\ S_{n+1} = S_{n+1}, \end{cases}$$

the joint density of $(Z_1, \cdots, Z_n, S_{n+1})$ is

$$\begin{aligned} f_{(Z,S_{n+1})}(z_1, \cdots, z_n, s_{n+1}) &= \exp\{-s_{n+1}\} s_{n+1}^n I(0 < z_1 < \cdots < z_n < 1, s_{n+1} > 0) \\ &= \exp\{-s_{n+1}\} s_{n+1}^n I(s_{n+1} > 0) I(0 < z_1 < \cdots < z_n < 1), \end{aligned}$$

we can see that $(Z_1, \cdots, Z_n)$ is independent of S_{n+1}. According to the properties of exponential distribution, $S_{n+1} \sim \Gamma(n+1, 1)$ and the density of S_{n+1} is

$$f_{n+1}(x) = \frac{1}{n!} x^n e^{-x} I(x > 0),$$

so the joint density of $(Z_1, \cdots, Z_n)$ is

$$f_Z(z_1, \cdots, z_n) = n! I(0 < z_1 < \cdots < z_n < 1),$$

which is same as the density of $(U_{(1)}, \cdots, U_{(n)})$. □

We are now ready to derive the asymptotic normality of the order statistics. The proof includes three steps:

(1) Asymptotic normality of $(S_m, S_{n+1} - S_m)^\top$;

(2) Let $g(x_1, x_2) = x_1/(x_1 + x_2)$ so $U_{(m)} = g(S_m, S_{n+1} - S_m)$;

(3) Application of Cramer theorem 2.

Theorem 6. *For $U_{(m)}$, the m-th order statistic of* Unif$(0,1)$*, suppose*

$$\sqrt{n}\left(\frac{m}{n} - p\right) \to 0,$$

then we have

$$\sqrt{n}(U_{(m)} - p) \to \mathsf{N}(0, p(1-p)).$$

Proof. Note that CLT implies

$$\begin{aligned} &\sqrt{m}\left(\frac{S_m}{m} - 1\right) \xrightarrow{d} \mathsf{N}(0,1), \\ &\sqrt{n+1-m}\left(\frac{S_{n+1} - S_m}{n+1-m} - 1\right) \xrightarrow{d} \mathsf{N}(0,1). \end{aligned}$$

Or equivalently,

$$\begin{aligned} &\frac{1}{\sqrt{m}}(S_m - m) \xrightarrow{d} \mathsf{N}(0,1), \\ &\frac{1}{\sqrt{n+1-m}}(S_{n+1} - S_m - (n+1-m)) \xrightarrow{d} \mathsf{N}(0,1). \end{aligned}$$

Under the assumption $m/n \to p$, we have

$$\sqrt{n}\Big(\frac{S_m}{n} - \frac{m}{n}\Big) = \sqrt{\frac{m}{n}}\frac{1}{\sqrt{m}}(S_m - m) \to \sqrt{p}\mathsf{N}(0,1) = \mathsf{N}(0,p),$$

similarly,

$$\sqrt{n}\Big(\frac{S_{n+1} - S_m}{n} - \frac{n+1-m}{n}\Big) \to \mathsf{N}(0, 1-p).$$

With the independence of S_m and $S_{n+1} - S_m$, we have

$$\sqrt{n}\Bigg[\begin{pmatrix} \frac{S_m}{n} \\ \frac{S_{n+1}-S_m}{n} \end{pmatrix} - \begin{pmatrix} \frac{m}{n} \\ \frac{n+1-m}{n} \end{pmatrix}\Bigg] \to \mathsf{N}\Bigg(\mathbf{0}, \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix}\Bigg).$$

Next, notice that

$$\begin{aligned}&\sqrt{n}\Bigg[\begin{pmatrix} \frac{S_m}{n} \\ \frac{S_{n+1}-S_m}{n} \end{pmatrix} - \begin{pmatrix} p \\ 1-p \end{pmatrix}\Bigg] \\ =&\sqrt{n}\Bigg[\begin{pmatrix} \frac{S_m}{n} \\ \frac{S_{n+1}-S_m}{n} \end{pmatrix} - \begin{pmatrix} \frac{m}{n} \\ \frac{n+1-m}{n} \end{pmatrix}\Bigg] + \sqrt{n}\Bigg[\begin{pmatrix} p \\ 1-p \end{pmatrix} - \begin{pmatrix} \frac{m}{n} \\ \frac{n+1-m}{n} \end{pmatrix}\Bigg],\end{aligned}$$

by assumption,

$$\sqrt{n}\Bigg[\begin{pmatrix} p \\ 1-p \end{pmatrix} - \begin{pmatrix} \frac{m}{n} \\ \frac{n+1-m}{n} \end{pmatrix}\Bigg] \to \mathbf{0},$$

so

$$\sqrt{n}\Bigg[\begin{pmatrix} \frac{S_m}{n} \\ \frac{S_{n+1}-S_m}{n} \end{pmatrix} - \begin{pmatrix} p \\ 1-p \end{pmatrix}\Bigg] \to \mathsf{N}\Bigg(\mathbf{0}, \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix}\Bigg).$$

Let $g(x_1, x_2) = x_1/(x_1 + x_2)$ then $U_{(m)} = g(S_m, S_{n+1} - S_m)$ and $g'(p, 1-p) = (1-p, -p)^\top$. With the application of Cramer theorem 2, we have

$$\begin{aligned}\sqrt{n}(U_{(m)} - p) &\to \mathsf{N}\Bigg(0, (1-p, -p)\begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix}\begin{pmatrix} 1-p \\ -p \end{pmatrix}\Bigg) \\ &= \mathsf{N}(0, p(1-p)).\end{aligned}$$

□

The asymptotic normality of two order statistics is similar with the single case.

Theorem 7. *For $U_{(m_1)} < U_{(m_2)}$, the m_1-th and m_2-th order statistics of* Unif(0,1)*, suppose*

$$\sqrt{n}\Big(\frac{m_1}{n} - p_1\Big) \to 0, \quad \sqrt{n}\Big(\frac{m_2}{n} - p_2\Big) \to 0,$$

then

$$\sqrt{n}\begin{pmatrix} U_{(m_1)} - p_1 \\ U_{(m_2)} - p_2 \end{pmatrix} \to \mathsf{N}\Bigg(\mathbf{0}, \begin{pmatrix} p_1(1-p_1) & p_1(1-p_2) \\ p_1(1-p_2) & p_2(1-p_2) \end{pmatrix}\Bigg).$$

Proof. By lemma 5,

$$U_{(m_1)} = \frac{S_{m_1}}{S_{n+1}}, \quad U_{(m_2)} = \frac{S_{m_2}}{S_{n+1}}.$$

Now we consider r.v. $S_{m_1}, S_{m_2} - S_{m_1}$ and $S_{n+1} - S_{m_2}$, by CLT, we have

$$\begin{aligned}
&\frac{1}{\sqrt{m_1}}(S_{m_1} - m_1) \to \mathsf{N}(0,1),\\
&\frac{1}{\sqrt{m_2 - m_1}}(S_{m_2} - S_{m_1} - (m_2 - m_1)) \to \mathsf{N}(0,1),\\
&\frac{1}{\sqrt{n+1-m_2}}(S_{n+1} - S_{m_2} - (n+1-m_2)) \to \mathsf{N}(0,1).
\end{aligned}$$

By assumption $m_1/n \to p_1$ and $m_2/n \to p_2$,

$$\begin{aligned}
&\sqrt{n}\left(\frac{S_{m_1}}{n} - \frac{m_1}{n}\right) \to \mathsf{N}(0,p_1),\\
&\sqrt{n}\left(\frac{S_{m_2} - S_{m_1}}{n} - \frac{m_2 - m_1}{n}\right) \to \mathsf{N}(0,p_2 - p_1),\\
&\sqrt{n}\left(\frac{S_{n+1} - S_{m_2}}{n} - \frac{n+1-m_2}{n}\right) \to \mathsf{N}(0,1 - p_2).
\end{aligned}$$

Then with the independence of $S_{m_1}, S_{m_2} - S_{m_1}$ and $S_{n+1} - S_{m_2}$,

$$\sqrt{n}\left[\begin{pmatrix} S_{m_1}/n \\ (S_{m_2} - S_{m_1})/n \\ (S_{n+1} - S_{m_2})/n \end{pmatrix} - \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}\right] \to \mathsf{N}\left(\mathbf{0}, \begin{pmatrix} p_1 & 0 & 0 \\ 0 & p_2 - p_1 & 0 \\ 0 & 0 & 1 - p_2 \end{pmatrix}\right),$$

let

$$\boldsymbol{g}(x_1, x_2, x_3) = \begin{pmatrix} x_1/(x_1 + x_2 + x_3) \\ (x_1 + x_2)/(x_1 + x_2 + x_3), \end{pmatrix}$$

then

$$\dot{\boldsymbol{g}}(p_1, p_2 - p_1, 1 - p_2) = \begin{pmatrix} 1 - p_1 & -p_1 & -p_1 \\ 1 - p_2 & 1 - p_2 & -p_2 \end{pmatrix}.$$

Checking that

$$\dot{\boldsymbol{g}}(p_1, p_2 - p_1, 1 - p_2)\begin{pmatrix} p_1 & 0 & 0 \\ 0 & p_2 - p_1 & 0 \\ 0 & 0 & 1 - p_2 \end{pmatrix}\dot{\boldsymbol{g}}(p_1, p_2 - p_1, 1 - p_2)^\top = \begin{pmatrix} p_1(1 - p_1) & p_1(1 - p_2) \\ p_1(1 - p_2) & p_2(1 - p_2) \end{pmatrix},$$

the proof is completed. $\square$

Generally, we may consider multivariate case.

Theorem 8. *For $U_{(m_1)} < U_{(m_2)} < \cdots < U_{(m_k)}$, and suppose $\sqrt{n}(m_j/n - p_j) \to 0, 0 < p_1 < \cdots < p_k < 1$, then*

$$\sqrt{n}\begin{pmatrix} U_{(m_1)} - p_1 \\ U_{(m_2)} - p_2 \\ \vdots \\ U_{(m_k)} - p_k \end{pmatrix} \to \mathsf{N}\left(\mathbf{0}, \begin{pmatrix} p_1(1-p_1) & p_1(1-p_2) & \cdots & p_1(1-p_k) \\ p_1(1-p_2) & p_2(1-p_2) & \cdots & p_2(1-p_k) \\ \vdots & \vdots & \vdots & \vdots \\ p_1(1-p_k) & p_2(1-p_k) & \cdots & p_k(1-p_k) \end{pmatrix}\right).$$

Now consider $m_j = \lceil np_j \rceil$, then the condition $\sqrt{n}(m_j/n - p_j) \to 0$ is easily to be verified, so for sample quantiles, we obtain

$$\sqrt{n}\begin{pmatrix} U_{(\lceil np_1 \rceil)} - p_1 \\ U_{(\lceil np_2 \rceil)} - p_2 \\ \vdots \\ U_{(\lceil np_k \rceil)} - p_k \end{pmatrix} \to \mathsf{N}\left(\mathbf{0}, \begin{pmatrix} p_1(1-p_1) & p_1(1-p_2) & \cdots & p_1(1-p_k) \\ p_1(1-p_2) & p_2(1-p_2) & \cdots & p_2(1-p_k) \\ \vdots & \vdots & \vdots & \vdots \\ p_1(1-p_k) & p_2(1-p_k) & \cdots & p_k(1-p_k) \end{pmatrix}\right).$$

Now consider the general distribution F. Applying the transformation $g(y) = F^{-1}(y)$ to $U_{(\lceil np_j \rceil)} - p_j$, we obtain

Corollary 9. *$X_{(1)} < \cdots < X_{(n)}$ are order statistics of sample of size n from a distribution F with density f, then*

$$\sqrt{n}\begin{pmatrix} X_{(\lceil np_1 \rceil)} - x_{p_1} \\ X_{(\lceil np_2 \rceil)} - x_{p_2} \\ \vdots \\ X_{(\lceil np_k \rceil)} - x_{p_k} \end{pmatrix} \to \mathsf{N}\left(\mathbf{0}, \begin{pmatrix} \frac{p_1(1-p_1)}{f^2(x_{p_1})} & \frac{p_1(1-p_2)}{f(x_{p_1})f(x_{p_2})} & \cdots & \frac{p_1(1-p_k)}{f(x_{p_1})f(x_{p_k})} \\ \frac{p_1(1-p_2)}{f(x_{p_1})f(x_{p_2})} & \frac{p_2(1-p_2)}{f^2(x_{p_2})} & \cdots & \frac{p_2(1-p_k)}{f(x_{p_2})f(x_{p_k})} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{p_1(1-p_k)}{f(x_{p_1})f(x_{p_k})} & \frac{p_2(1-p_k)}{f(x_{p_2})f(x_{p_k})} & \cdots & \frac{p_k(1-p_k)}{f^2(x_{p_k})} \end{pmatrix}\right),$$

where x_{p_j} is p_j-quantile of F.

Example 1. *Consider $X_1, \cdots, X_n$ be a sample of size n from $\mathsf{N}(\mu, \sigma^2)$. Let $\bar{X}_n$ be the sample mean and $m_n = X_{(\lceil 0.5n \rceil)}$ be sample median. By CLT and corollary 9 we have*

$$\begin{aligned} \sqrt{n}(\bar{X}_n - \mu) &\to \mathsf{N}(0, \sigma^2), \\ \sqrt{n}(m_n - \mu) &\to \mathsf{N}(0, \pi\sigma^2/2). \end{aligned}$$

The sample mean is more efficient to estimate μ.

Example 2. *The Cauchy distribution $\mathcal{C}(\mu, \sigma)$ has density*

$$f(x) = \frac{1}{\pi\sigma}\frac{1}{1 + [(x-\mu)/\sigma]^2}.$$

The population median, first quartile and third quartile are μ, $\mu - \sigma$ and $\mu + \sigma$. From corollary 9,

$$\begin{aligned} &\sqrt{n}(m_n - \mu) \to \mathsf{N}\left(0, \frac{\pi^2\sigma^2}{4}\right), \\ &\sqrt{n}\begin{bmatrix} X_{(n/4)} - (\mu - \sigma) \\ X_{(3n/4)} - (\mu + \sigma) \end{bmatrix} \to \mathsf{N}\left(\mathbf{0}, \pi^2\sigma^2\begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}\right). \end{aligned}$$

Hence

$$\sqrt{n}\left(\frac{X_{(3n/4)} - X_{(n/4)}}{2} - \sigma\right) \to \mathsf{N})(0, \pi^2\sigma^2/4).$$

2 Asymptotic theory of extreme order statistics

Asymptotic normality does not hold for $p = 0$ and 1. For example, consider $U_{(n)}$, the n-th order statistic of Unif$(0, 1)$, then

$$\mathbb{E}U_{(n)} = \frac{n}{n+1}, \mathrm{Var}(U_{(n)}) = \frac{n}{(n+1)^2(n+2)},$$

thus $\sqrt{\mathrm{Var}(U_{(n)})} = O(n) \gg O(\sqrt{n})$ and it's meaningless to analyze $\sqrt{n}(U_{(n)} - 1)$. However,

$$\begin{aligned} &U_{(n)} \in [0, 1] \;\Rightarrow\; \frac{U_{(n)} - \mathbb{E}U_{(n)}}{\sqrt{\mathrm{Var}(U_{(n)})}} \in \left[-\sqrt{n(n+2)}, \sqrt{\frac{n+2}{n}}\right] \to (-\infty, 1], \\ &\mathbb{P}\left\{\frac{U_{(n)} - \mathbb{E}U_{(n)}}{\sqrt{\mathrm{Var}(U_{(n)})}} \le x\right\} = \int_0^{x\sqrt{\mathrm{Var}(U_{(n)})} + \mathbb{E}U_{(n)}} nx^{n-1}\mathrm{d}x = \left(\frac{x}{n+1}\sqrt{\frac{n}{n+2}} + \frac{n}{n+1}\right)^n \to e^{x-1}. \end{aligned}$$

We can see that the limit distribution of $U_{(n)}$ is an anlogue of exponential distribution. Moreover, it is easy to show

$$n(U_{(n)} - 1) \to -\mathrm{Exp}(1),$$

because

$$\mathbb{P}\Big\{n(U_{(n)} - 1) \le x\Big\} = \mathbb{P}\Big\{U_{(n)} \le \frac{x}{n} + 1\Big\} = \Big(\frac{x}{n} + 1\Big)^n \to \exp(x).$$

Now we consider the general case. In the following discussion, we will always denote M_n as the maximum of n observations. The problem is to determine if there exist a_n and $b_n > 0$ such that $(M_n - a_n)/b_n$ converges to a degenerate distribution G, or equivalently,

$$\mathbb{P}\bigg(\frac{M_n - a_n}{b_n} \le x\bigg) = \mathbb{P}(M_n \le a_n + b_n x) = F(a_n + b_n x)^n \to G(x).$$

We will show that there are three different classes of the limiting G's: Weibull class, Frechet class and Gumbel class.

Definition 10. *A function $c : [0, \infty) \to \mathbb{R}$ is slowly varying if for every $x > 0$,*

$$\frac{c(tx)}{c(t)} \to 1, \quad \text{as } t \to \infty.$$

Any function $c(x)$ converging to a positive finite number as $x \to \infty$ is slowly varying. Now we introduce the three classes of limiting distribution of extreme order statistic M_n.

Theorem 11. *Let $F(x)$ denote the distribution function of a random variable X, and let x_0 denote the upper boundary, possibly $+\infty$, of the distribution of $X : x_0 = \sup\{x : F(x) < 1\}$.*

(a) If $x_0 = +\infty$, and $1 - F(x) = x^{-\gamma}c(x)$ for some $\gamma > 0$ and some slowly varying $c(x)$, then

$$F(b_n x)^n \to G_{1,\gamma}(x) = \begin{cases} \exp\{-x^{-\gamma}\}, & x > 0, \\ 0, & x \le 0, \end{cases}$$

where b_n is such that $1 - F(b_n) = 1/n$.

(b) If $x_0 < \infty$, and $1 - F(x) = (x_0 - x)^\gamma c(1/(x_0 - x))$ for some $\gamma > 0$ and some slowly varying $c(x)$, then

$$F(x_0 + b_n x)^n \to G_{2,\gamma}(x) = \begin{cases} \exp\{-(-x)^\gamma\}, & x < 0, \\ 0, & x \ge 0, \end{cases}$$

where b_n is such that $1 - F(x_0 - b_n) = 1/n$.

(c) If there exists a function $R(t)$ such that for $\forall x$,

$$\frac{1 - F(t + xR(t))}{1 - F(t)} \to e^{-x}$$

as $t \to x_0$, then

$$F(a_n + b_n x)^n \to G_3(x) = \exp\{-e^{-x}\},$$

where $1 - F(a_n) = 1/n$ and $b_n = R(a_n)$.

Remark. $G_{1,\gamma}$, $G_{2,\gamma}$ and G_3 is called Weibull distribution , Frechet distribution and Gumbel distribution, respectively.

Remark. Theorem 11 tells us that

- If conditions of (a) holds, then

$$\frac{M_n}{b_n} \to G_{1,\gamma},$$

where $1 - F(b_n) = 1/n$;

- If conditions of (b) holds, then

$$\frac{M_n - x_0}{b_n} \to G_{2,\gamma},$$

where $1 - F(x_0 - b_n) = 1/n$;

- If conditions of (c) holds, then

$$\frac{M_n - a_n}{b_n} \to G_3,$$

where $1 - F(a_n) = 1/n$ and $b_n = R(a_n)$.

Proof of Theorem 11

(a) Note that $b_n \to \infty$ and

$$\frac{1}{n} = 1 - F(b_n) = b_n^{-\gamma} c(b_n),$$

so for $x > 0$,

$$\begin{aligned}
F^n(b_n x) &= \big(1 - (b_n x)^{-\gamma} c(b_n x)\big)^n \\
&\to \exp\Big\{-\lim_{n\to\infty} n(b_n x)^{-\gamma} c(b_n x)\Big\} \\
&= \exp\Big\{-x^{-\gamma} \lim_{n\to\infty} n b_n^{-\gamma} c(b_n x)\Big\} \\
&= \exp\left\{-x^{-\gamma} \lim_{n\to\infty} \frac{b_n^{-\gamma} c(b_n x)}{b_n^{-\gamma} c(b_n)}\right\} \\
&= \exp\big\{-x^{-\gamma}\big\};
\end{aligned}$$

for $x \le 0$, $F^n(b_n x) \to 0$.

(b) Note that $b_n > 0$, $b_n \to 0$ and

$$\frac{1}{n} = 1 - F(x_0 - b_n) = b_n^{\gamma} c\left(\frac{1}{b_n}\right),$$

hence for $x < 0$,

$$\begin{aligned}
F^n(x_0 + b_n) &= \left(1 - (-b_n x)^{\gamma} c\left(-\frac{1}{b_n x}\right)\right)^n \\
&\to \exp\left\{-\lim_{n\to\infty} n(-b_n x)^{\gamma} c\left(-\frac{1}{b_n x}\right)\right\} \\
&= \exp\left\{-(-x)^{\gamma} \lim_{n\to\infty} n b_n^{\gamma} c\left(-\frac{1}{b_n x}\right)\right\} \\
&= \exp\left\{-(-x)^{\gamma} \lim_{n\to\infty} \frac{b_n^{\gamma} c(-1/(b_n x))}{b_n^{\gamma} c(1/b_n)}\right\} \\
&= \exp\{-(-x)^{\gamma}\},
\end{aligned}$$

for $x \ge 0$, $F^n(x_0 + b_n) = 1$.

(c)

$$
\begin{aligned}
F^n(a_n+b_nx) &= [1-(1-F(a_n+b_nx))]^n \\
&\to \exp\Big\{-\lim_{n\to\infty} n(1-F(a_n+b_nx))\Big\} \\
&= \exp\Big\{-\lim_{n\to\infty} n(1-F(a_n))\frac{1-F(a_n+R_n(a_n)x)}{1-F(a_n)}\Big\} \\
&= \exp\Big\{-\lim_{n\to\infty} \frac{1-F(a_n+R_n(a_n)x)}{1-F(a_n)}\Big\} \\
&= \exp\{e^{-x}\}.
\end{aligned}
$$

□

Remark.

- If $c(x)\to a$ when $x\to+\infty$ for some constant a, and let b'_n satisfies $1/n = b_n'^{-\gamma}a$, then the result of (a) also holds when we replace b_n with b'_n. Specifically,

$$F^n(b'_nx)\to G_{1,\gamma}(x).$$

- If $c(x)\to a$ when $x\to+\infty$ for some constant a, and let b'_n satisfies $1/n = b_n'^{\gamma}a$, then the result of (b) also holds when we replace b_n with b'_n. Specifically,

$$F(x_0+b_nx)^n\to G_{2,\gamma}(x).$$

- In case (c), note that

$$\frac{1-F(t+xR(t))}{1-F(t)} = \mathbb{P}(X>t+xR(t)|X>t),$$

so that the condition that this converge to $\exp(-x)$ means that there is a change of scale, $R(t)$, so that the conditional distribution is approximately exponential with parameter 1.

Remark. The converse of theorem 11 is true.

Remark. The three families of distributions may be related to the exponential distribution as follows. If $Y\sim\mathrm{Exp}(1)$, then

$$Y^{-1/\gamma}\sim G_{1,\gamma},\quad -Y^{1/\gamma}\sim G_{2,\gamma},\quad -\log(Y)\sim G_3.$$

Example 3 (t_ν-distribution)**.** *Density of t_ν-distribution is*

$$f(x)=\frac{c}{(\nu+x^2)^{(\nu+1)/2}}.$$

Observe that

$$\lim_{x\to+\infty}\frac{f(x)}{g(x)}=1,$$

where $g(x) = c/(x^{\nu+1})$. We denote this property as $f(x) \sim g(x)$, whcih is called asymptotic equivalence. Thus,

$$\begin{aligned} 1 - F(x) &= \int_x^\infty f(y)dy \\ &= \int_x^\infty g(y)dy \frac{\int_x^\infty f(y)dy}{\int_x^\infty g(y)dy} \\ &= \frac{c}{\nu} x^{-\nu} \frac{\int_x^\infty f(y)dy}{\int_x^\infty g(y)dy} \\ &\triangleq x^{-\nu} c(x). \end{aligned}$$

It is easy to show $c(x)$ is slowly varying. Hence case (a) holds with $\gamma = \nu$, and

$$\frac{1}{n} = 1 - F(b_n) = b_n^{-\nu} c(x) \sim b_n^{-\nu} \frac{c}{\nu} \Rightarrow b_n = \left(\frac{cn}{\nu}\right)^{1/\nu}.$$

Then,

$$\frac{M_n \nu^{1/\nu}}{c^{1/\nu} n^{1/\nu}} \xrightarrow{d} G_{1,\gamma}.$$

For Cauchy distribution, $\nu = 1$ and $c = 1/\pi$ so that

$$\frac{\pi}{n} M_n \xrightarrow{d} G_{1,1}.$$

Example 4 (Beta distribution)**.** *The* Beta *distribution have density*

$$f(x) = cx^{\alpha-1}(1-x)^{\beta-1} I(0 < x < 1),$$

where $c = \Gamma(\alpha+\beta)/(\Gamma(\alpha)\Gamma(\beta))$. So $x_0 = 1$, and

$$\lim_{x \to 1} \frac{f(x)}{c(1-x)^{\beta-1}} = 1,$$

so similar with Example 3, one can show that

$$1 - F(x) = (1-x)^\beta c(x),$$

where $c(x)$ is such that $\lim_{x\to\infty} c(x) = c/\beta$. Hence case (b) holds with $\gamma = \beta$,

$$\frac{1}{n} = 1 - F(1 - b_n) \sim \frac{c}{\beta} b_n^\beta \Rightarrow b_n = \left(\frac{\beta}{cn}\right)^{1/\beta}.$$

and

$$\left(\frac{cn}{\beta}\right)^{1/\beta} (M_n - 1) \to G_{2,\beta}.$$

For the Unif$(0,1)$*,*

$$n(M_n - 1) \to G_{2,1} = -\mathrm{Exp}(1).$$

Example 5 (Exponential-distribution)**.** *For exponential distribution* Exp(1)*, we have*

$$\mathbb{P}(X > t + x | X > t) = \exp(-x)$$

for all $t > 0, x > 0$*. So we have* $R(t) = 1$ *for all* t *in case (c) and hence* $b_n = 1$ *for all* n*. Now we solve for* a_n*:*

$$1 - F(a_n) = \exp(-a_n) = \frac{1}{n} \Rightarrow a_n = \log(n),$$

that is

$$M_n - \log(n) \to G_3.$$

Example 6 (Weibull-distribution)**.** *If* $X \sim G_{1,1}$*, then the distribution function of* $-X$ *is*

$$F(x) = \begin{cases} 1 - \exp(1/x), & x < 0, \\ 1, & x \geq 0. \end{cases}$$

Then

$$\frac{1 - F(t + xR(t))}{1 - F(t)} = \exp\left\{\frac{1}{t + xR(t)} - \frac{1}{t}\right\} = \exp\left\{-\frac{xR(t)}{t(xR(t) + t)}\right\}.$$

We want to choose $R(t)$ *so that* $R(t)/(t(xR(t) + t)) \to 1$ *as* $t \nearrow x_0 = 0$*.* $R(t) = t^2$ *works.*

$$\frac{1}{n} = 1 - F(a_n) = \exp\left(\frac{1}{a_n}\right) \Rightarrow a_n = -\frac{1}{\log n}$$

and

$$b_n = R(a_n) = a_n^2 = \frac{1}{(\log n)^2}.$$

Hence

$$(\log n)^2(M_n + 1/\log n) \to G_3.$$

Example 7 (Gumbel-distribution)**.** *If* $X \sim G_3$*, then*

$$\mathbb{P}(M_n - \log n \leq x) = \mathbb{P}(M_n \leq x + \log n) = \exp(-e^{-x-\log n})^n = \exp(-e^{-x}),$$

that is, the distribution of $M_n - \log n$ *is exact* G_3*.*

Example 8 (Normal-distribution)**.** *The standard normal distribution* $\mathsf{N}(0,1)$ *falls in case (c). The distribution function is*

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{u^2}{2}\right\} du.$$

First we claim that

$$1 - \Phi(x) \sim \frac{1}{\sqrt{2\pi}} \frac{1}{x} \exp\left(-\frac{x^2}{2}\right) \text{ as } x \to \infty.$$

It can be easily proved by L'Hospital's rule. This claim implies that

$$\begin{aligned} \frac{1 - \Phi(t + xR(t))}{1 - \Phi(t)} &\sim \frac{\exp\{-(t + xR(t))^2/2\}}{t + xR(t)} \cdot \frac{t}{\exp\{-t^2/2\}} \\ &= \frac{t}{(t + xR(t))} \exp\{-txR(t) - x^2R(t)^2/2\}. \end{aligned}$$

This converges to e^{-x} if we let $R(t) = 1/t$. Thus we have case (c) with $b_n = 1/a_n$ and $1 - \Phi(a_n) = 1/n$, and conclude that

$$a_n(M_n - a_n) \to G_3.$$

In fact, we can find an asymptotic expression for a_n:

$$a_n \sim \sqrt{2\log n} - (\log\log n + \log 4\pi)/2\sqrt{2\log n},$$

and $b_n = 1/a_n \sim 1/\sqrt{2\log n}$, so

$$\sqrt{2\log n}M_n - w\log n + \frac{1}{2}\log\log n + \frac{1}{2}\log 4\pi \to G_3.$$

The details can be found in page 99 of [1].

The following are some simple simulations to verify the asymptotic distribution in Example 3-8. For each distribution, we generate a sample of size $n = 100$, and then plot the empirical distribution of the maximum order statistic over 1000 trials.

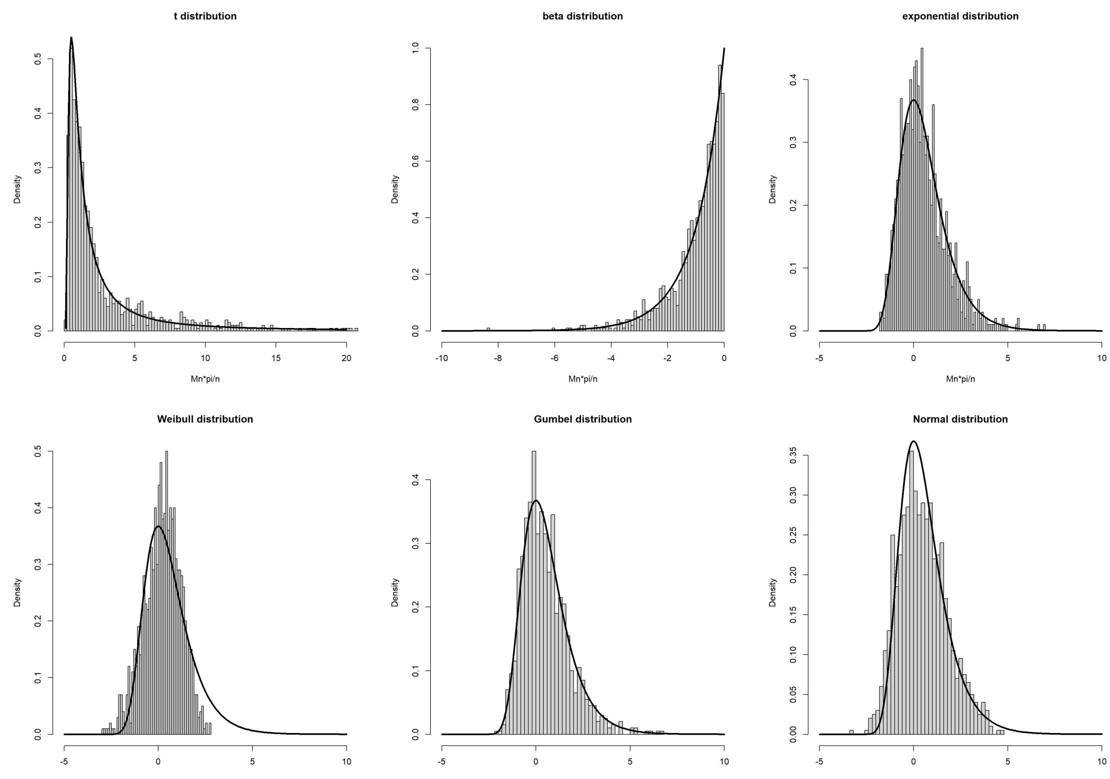

Figure 1: Empirical vs Asymptotic distribution of maximum order statistics

References

[1] T. S. Ferguson, *A course in large sample theory.* Routledge, 2017.