# M3/M4S3 STATISTICAL THEORY II
## THE JOINT DISTRIBUTION OF THE SAMPLE QUANTILES

**RESULT 1:** If $Y_1, Y_2, ..., Y_{n+1} \sim Exponential\,(1)$ are independent random variables, and $S_1, S_2, ..., S_{n+1}$ are defined by

$$S_k = \sum_{j=1}^{k} Y_j \qquad k = 1, 2, ..., n+1$$

then the random variables

$$\left[\frac{S_1}{S_{n+1}}, \frac{S_2}{S_{n+1}}, ..., \frac{S_n}{S_{n+1}}\right]$$

given that $S_{n+1} = s$, say, have the same distribution as the order statistics from a random sample of size $n$ from the Uniform distribution on $(0,1)$.

**Proof:** Let the $Y_j$s be defined as above. Then the joint density for the $Y_j$s is given by

$$\exp\left\{-\sum_{j=1}^{n+1} y_j\right\} \qquad y_1, y_2, ..., y_{n+1} > 0.$$

Now

$$\left.\begin{array}{ll} S_1 & = Y_1 \\ S_2 & = Y_1 + Y_2 \\ S_3 & = Y_1 + Y_2 + Y_3 \\ & \\ S_n & = \sum_{j=1}^{n} Y_j \\ S_{n+1} & = \sum_{j=1}^{n+1} Y_j \end{array}\right\} \Leftrightarrow \left\{\begin{array}{ll} Y_1 & = S_1 \\ Y_2 & = S_2 - S_1 \\ Y_3 & = S_3 - S_2 \\ & \\ Y_n & = S_n - S_{n-1} \\ Y_{n+1} & = S_{n+1} - S_n \end{array}\right.$$

and so the Jacobian of the transformation from $(Y_1, ..., Y_{n+1}) \to (S_1, ..., S_{n+1})$ is

$$\begin{vmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & & 1 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{vmatrix} = 1$$

and hence the joint density for $(S_1, ..., S_{n+1})$ is given by

$$\exp\{-s_{n+1}\} \qquad 0 < s_1 < s_2 < ... < s_{n+1}.$$

The marginal distribution for $S_{n+1}$ is $Gamma\,(n+1, 1)$ and thus the conditional distribution of $(S_1, ..., S_n)$ given $S_{n+1} = s$ is

$$\frac{\exp\{-s\}}{\dfrac{1}{\Gamma(n+1)} s^n \exp\{-s\}} = \frac{n!}{s^n} \qquad 0 < s_1 < s_2 < ... < s.$$

Finally, conditional on $S_{n+1} = s$, define the joint transformation

$$V_j = \frac{S_j}{s} \Leftrightarrow S_j = sV_j \qquad j = 1, 2, ..., n$$

which has Jacobian $s^n$. Then, conditional on $S_{n+1} = s$, $(V_1, ..., V_n)$ have joint pdf equal to $n!$ for $0 < v_1 < v_2 < ... < v_n < 1$. Finally, if $U_1, ..., U_n$ are independent random variables each having a Uniform distribution on $(0,1)$, then $(U_1, ..., U_n)$ have joint pdf equal to 1 on the unit hypercube in $n$ dimensions, and thus the corresponding order statistics $U_{(1)}, ..., U_{(n)}$ also have joint pdf equal to

$$n! \qquad 0 < u_1 < u_2 < ... < u_n < 1.$$

**RESULT 2:** Let the $S_k$ be defined as in Result 1. Then

$$\sqrt{k}\left(\frac{1}{k}S_k - 1\right) \xrightarrow{\mathfrak{L}} N(0,1) \text{ as } k \to \infty$$

**Proof:** We have that $S_k$ is the sum of $k$ independent and identically distributed $Exponential(1)$ random variables, $Y_1, ..., Y_k$, so that $E[Y_j] = Var[Y_j] = 1$. Thus the Central Limit Theorem applies, and the result follows.

**RESULT 3:** Let the $S_k$ be defined as in Result 1. Then, if

$$\frac{k_1}{n} \to p_1$$

for some $p_1$ with $0 < p_1 < 1$,

$$\sqrt{n+1}\left(\frac{1}{n+1}S_{k_1} - \frac{k_1}{n+1}\right) \xrightarrow{\mathfrak{L}} N(0,p_1) \text{ as } n \to \infty$$

**Proof:** We have

$$\sqrt{n+1}\left(\frac{1}{n+1}S_{k_1} - \frac{k_1}{n+1}\right) = \sqrt{\frac{k_1}{n+1}}\sqrt{k_1}\left(\frac{1}{k_1}S_{k_1} - 1\right) \xrightarrow{\mathfrak{L}} \sqrt{p_1}N(0,1) \equiv N(0,p_1)$$

as $n \to \infty$ (so that by assumption $k_1 \to \infty$ also).
**Corollary:** Using the same approach, if

$$\frac{k_1}{n} \to p_1 \qquad \text{and} \qquad \frac{k_2}{n} \to p_2$$

for $0 < p_1 < p_2 < 1$, then

$$\sqrt{n+1}\left(\frac{1}{n+1}(S_{k_2} - S_{k_1}) - \frac{k_2 - k_1}{n+1}\right) = \sqrt{\frac{k_2 - k_1}{n+1}}\sqrt{k_2 - k_1}\left(\frac{1}{k_2 - k_1}\sum_{j=k_1+1}^{k_2} Y_j - 1\right)$$

and the right-hand side converges in law to $\sqrt{p_2 - p_1}N(0,1) \equiv N(0, p_2 - p_1)$. Similarly

$$\sqrt{n+1}\left(\frac{1}{n+1}(S_{n+1} - S_{k_2}) - \frac{n+1-k_2}{n+1}\right) \xrightarrow{\mathfrak{L}} N(0, 1 - p_2)$$

where the limiting variables in the three cases are independent, as $S_{k_1}, (S_{k_2} - S_{k_1})$, and $(S_{n+1} - S_{k_2})$ are independent.

**RESULT 4:** Let

$$\begin{aligned} Z_1 &= \frac{1}{n+1}S_{k_1} \\ Z_2 &= \frac{1}{n+1}(S_{k_2} - S_{k_1}) \\ Z_3 &= \frac{1}{n+1}(S_{n+1} - S_{k_2}) \end{aligned}$$

and suppose that

$$\sqrt{n}\left(\frac{k_1}{n}-p_1\right)\to 0 \text{ and } \sqrt{n}\left(\frac{k_2}{n}-p_2\right)\to 0$$

as $n\to\infty$. Then

$$\sqrt{n+1}\left(\left(\begin{array}{c} Z_1 \\ Z_2 \\ Z_3 \end{array}\right)-\left(\begin{array}{c} p_1 \\ p_2-p_1 \\ 1-p_2 \end{array}\right)\right)\xrightarrow{\mathfrak{L}} N(0,\Sigma)$$

as $n\to\infty$, where $\Sigma = diag\,(p_1, p_2-p_1, 1-p_2)$.

**Proof:** We have

$$\sqrt{n+1}\left(\frac{1}{n+1}S_{k_1}-p_1\right)-\sqrt{n+1}\left(\frac{1}{n+1}S_{k_1}-\frac{k_1}{n+1}\right)=\sqrt{n+1}\left(\frac{k_1}{n+1}-p_1\right)\to 0$$

as $n\to\infty$ by assumption, so

$$\sqrt{n+1}\left(\frac{1}{n+1}S_{k_1}-p_1\right) \text{ and } \sqrt{n+1}\left(\frac{1}{n+1}S_{k_1}-\frac{k_1}{n+1}\right)$$

have the same asymptotic distribution, and thus the result follows from Result 3. The proof is similar for the other two terms. Independence (that is, ths diagonal nature of $\Sigma$) follows from the independence of $S_{k_1}, (S_{k_2}-S_{k_1})$, and $(S_{n+1}-S_{k_2})$.

**RESULT 5:** If $U_{(1)},...,U_{(n)}$ are the order statistics from a random sample of size $n$ from a $Uniform\,(0,1)$ distribution, and if $n\to\infty$, $k_1\to\infty$ and $k_2\to\infty$ in such a way that

$$\sqrt{n}\left(\frac{k_1}{n}-p_1\right)\to 0 \text{ and } \sqrt{n}\left(\frac{k_2}{n}-p_2\right)\to 0$$

for $0<p_1<p_2<1$, then

$$\sqrt{n}\left(\left(\begin{array}{c} U_{(k_1)} \\ U_{(k_2)} \end{array}\right)-\left(\begin{array}{c} p_1 \\ p_2 \end{array}\right)\right)\xrightarrow{\mathfrak{L}} N\left(0,\left[\begin{array}{cc} p_1(1-p_1) & p_1(1-p_2) \\ p_1(1-p_2) & p_2(1-p_2) \end{array}\right]\right).$$

**Proof:** Define

$$g(x_1,x_2,x_3)=\frac{1}{x_1+x_2+x_3}\left[\begin{array}{c} x_1 \\ x_1+x_2 \end{array}\right]$$

which yields first derivative

$$\dot{g}(x_1,x_2,x_3)=\frac{1}{(x_1+x_2+x_3)^2}\left[\begin{array}{ccc} x_2+x_3 & -x_1 & -x_1 \\ x_3 & x_3 & -(x_1+x_2) \end{array}\right].$$

Now

$$g\left(\frac{S_{k_1}}{n+1},\frac{S_{k_2}-S_{k_1}}{n+1},\frac{S_{n+1}-S_{k_2}}{n+1}\right)=\frac{1}{S_{n+1}}\left[\begin{array}{c} S_{k_1} \\ S_{k_2} \end{array}\right]$$

which has the same distribution as $\left(U_{(k_1)},U_{(k_2)}\right)^T$, by Result 1. By Cramer's Theorem

$$\sqrt{n}\left(\left(\begin{array}{c} U_{(k_1)} \\ U_{(k_2)} \end{array}\right)-\left(\begin{array}{c} p_1 \\ p_2 \end{array}\right)\right)\xrightarrow{\mathfrak{L}} N\left(0,\dot{g}(\mu)\,\Sigma\dot{g}(\mu)^T\right)$$

where $\Sigma$ is as defined in the Result 4, where here $\mu=(p_1,p_2-p_1,1-p_2)^T$. It can be easily verified that

$$\dot{g}(\mu)\,\Sigma\dot{g}(\mu)^T=\left[\begin{array}{cc} p_1(1-p_1) & p_1(1-p_2) \\ p_1(1-p_2) & p_2(1-p_2) \end{array}\right]$$

and thus the result follows.

**RESULT 6:** If $X_{(1)}, ..., X_{(n)}$ are the order statistics from a random sample of size $n$ from a distribution with continuous distribution function $F_X$ and density $f_X$ which is continuous and non-zero in a neighbourhood of quantiles $x_{p_1}$ and $x_{p_2}$ corresponding to probabilities $p_1 < p_2$, then if $k_1 = \lceil np_1 \rceil$ and $k_2 = \lceil np_2 \rceil$

$$\sqrt{n}\left(\left(\begin{array}{c} X_{(k_1)} \\ X_{(k_2)} \end{array}\right) - \left(\begin{array}{c} x_{p_1} \\ x_{p_2} \end{array}\right)\right) \stackrel{\mathfrak{L}}{\to} N\left(0, \left[\begin{array}{cc} \dfrac{p_1(1-p_1)}{\{f_X(x_{p_1})\}^2} & \dfrac{p_1(1-p_2)}{f_X(x_{p_1}) f_X(x_{p_2})} \\ \dfrac{p_1(1-p_2)}{f_X(x_{p_1}) f_X(x_{p_2})} & \dfrac{p_2(1-p_2)}{\{f_X(x_{p_2})\}^2} \end{array}\right]\right)$$

**Proof:** We use the **Delta Method** (Cramer's Theorem) on the result from Result 5, with the transformation

$$g(y_1, y_2) = \left[\begin{array}{c} F_X^{-1}(y_1) \\ F_X^{-1}(y_2) \end{array}\right]$$

so that

$$\dot{g}(y_1, y_2) = \left[\begin{array}{cc} \dfrac{1}{f_X\left(F_X^{-1}(y_1)\right)} & 0 \\ 0 & \dfrac{1}{f_X\left(F_X^{-1}(y_2)\right)} \end{array}\right]$$

with $y_1 = p_1$ and $y_2 = p_2$.