556: MATHEMATICAL STATISTICS I
THE JOINT DISTRIBUTION OF THE SAMPLE QUANTILES

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

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

then the random variables

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

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_js be defined as above. Then the joint density for the Y_js is given by

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

Now

$$\left.\begin{array}{rcl} S_1 & = & Y_1 \\ S_2 & = & Y_1 + Y_2 \\ S_3 & = & Y_1 + Y_2 + Y_3 \\ \vdots & & \vdots \\ S_n & = & \sum\limits_{j=1}^{n} Y_j \\ S_{n+1} & = & \sum\limits_{j=1}^{n+1} Y_j \end{array}\right\} \Longleftrightarrow \left\{\begin{array}{rcl} Y_1 & = & S_1 \\ Y_2 & = & S_2 - S_1 \\ Y_3 & = & S_3 - S_2 \\ \vdots & & \vdots \\ Y_n & = & S_n - S_{n-1} \\ Y_{n+1} & = & S_{n+1} - S_n \end{array}\right.$$

and so the Jacobian of the transformation from $(Y_1, \ldots, Y_{n+1}) \longrightarrow (S_1, \ldots, S_{n+1})$ is 1, and hence the joint density for $(S_1, \ldots, S_{n+1})$ is given by

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

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

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

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

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

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

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

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

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

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

RESULT 3: Let the S_k be defined as in Result 1. Then, if $k_1(n)$ is a sequence of integers such that

$$k_{1n}\longrightarrow\infty \qquad \text{while} \qquad \frac{k_{1n}}{n}\longrightarrow p_1$$

for some p_1 with $0<p_1<1$, it follows that

$$\sqrt{n+1}\left(\frac{S_{k_{1n}}}{n+1}-\frac{k_{1n}}{n+1}\right)\xrightarrow{d} N\left(0,p_1\right) \text{ as } n\longrightarrow\infty$$

Proof: We have

$$\sqrt{n+1}\left(\frac{S_{k_{1n}}}{n+1}-\frac{k_{1n}}{n+1}\right)=\sqrt{\frac{k_{1n}}{n+1}}\times\sqrt{k_{1n}}\left(\frac{S_{k_{1n}}}{k_{1n}}-1\right)\xrightarrow{d}\sqrt{p_1}\times N\left(0,1\right)\equiv N\left(0,p_1\right)$$

as $n\longrightarrow\infty$ and $k_{1n}\longrightarrow\infty$.

Corollary: Using the same approach, if

$$\frac{k_{1n}}{n}\longrightarrow p_1 \qquad \text{and} \qquad \frac{k_{2n}}{n}\longrightarrow p_2$$

for $0<p_1<p_2<1$, then if $D_n=\sum\limits_{j=k_{1n}+1}^{k_{2n}}Y_j$,

$$\begin{aligned}\sqrt{n+1}\left(\frac{(S_{k_{2n}}-S_{k_{1n}})}{n+1}-\frac{k_{2n}-k_{1n}}{n+1}\right) &= \sqrt{\frac{k_{2n}-k_{1n}}{n+1}}\sqrt{k_{2n}-k_{1n}}\left(\frac{D_n}{k_{2n}-k_{1n}}-1\right)\\ &\xrightarrow{d} \sqrt{p_2-p_1}\times N\left(0,1\right)\equiv N\left(0,p_2-p_1\right).\end{aligned}$$

Similarly

$$\sqrt{n+1}\left(\frac{1}{n+1}\left(S_{n+1}-S_{k_{2n}}\right)-\frac{n+1-k_{2n}}{n+1}\right)\xrightarrow{d} N\left(0,1-p_2\right)$$

where the limiting variables in the three cases are independent, as

$$\begin{aligned}S_{k_{1n}} &= \sum_{j=1}^{k_{1n}}Y_j\\ (S_{k_{2n}}-S_{k_{1n}}) &= \sum_{j=k_{1n}+1}^{k_{2n}}Y_j\\ (S_{n+1}-S_{k_{2n}}) &= \sum_{j=k_{2n}+1}^{n+1}Y_j\end{aligned}$$

are independent.

RESULT 4: Let

$$Z_1 = \frac{S_{k_{1n}}}{n+1} \qquad Z_2 = \frac{(S_{k_{2n}} - S_{k_{1n}})}{n+1} \qquad Z_3 = \frac{(S_{n+1} - S_{k_{2n}})}{n+1}$$

and suppose that

$$\sqrt{n}\left(\frac{k_{1n}}{n} - p_1\right) \longrightarrow 0 \qquad \text{and} \qquad \sqrt{n}\left(\frac{k_{2n}}{n} - p_2\right) \longrightarrow 0$$

as $n \longrightarrow \infty$. Then

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

as $n \longrightarrow \infty$, where $\Sigma = \text{diag}(p_1, p_2 - p_1, 1 - p_2)$.

Proof: We have, as $n \longrightarrow \infty$,

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

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

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, the diagonal nature of Σ) follows from the independence of $S_{k_{1n}}, (S_{k_{2n}} - S_{k_{1n}})$, and $(S_{n+1} - S_{k_{2n}})$.

RESULT 5: If $U_{(1)}, \ldots, U_{(n)}$ are the order statistics from a random sample of size n from a $Uniform\,(0,1)$ distribution, and if $n \longrightarrow \infty$, $k_{1n} \longrightarrow \infty$ and $k_{2n} \longrightarrow \infty$ in such a way that

$$\sqrt{n}\left(\frac{k_{1n}}{n} - p_1\right) \longrightarrow 0 \qquad \text{and} \qquad \sqrt{n}\left(\frac{k_{2n}}{n} - p_2\right) \longrightarrow 0$$

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

$$\sqrt{n}\left(\left(\begin{array}{c} U_{(k_{1n})} \\ U_{(k_{2n})} \end{array}\right) - \left(\begin{array}{c} p_1 \\ p_2 \end{array}\right)\right) \stackrel{d}{\longrightarrow} N\left(0, \left[\begin{array}{cc} p_1(1-p_1) & p_1(1-p_2) \\ p_1(1-p_2) & p_2(1-p_2) \end{array}\right]\right).$$

Proof: Define

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

$$\therefore \ g\left(\frac{S_{k_{1n}}}{n+1}, \frac{S_{k_{2n}} - S_{k_{1n}}}{n+1}, \frac{S_{n+1} - S_{k_{2n}}}{n+1}\right) = \frac{1}{S_{n+1}}\left[\begin{array}{c} S_{k_{1n}} \\ S_{k_{2n}} \end{array}\right]$$

which has the same distribution as $\left(U_{(k_{1n})}, U_{(k_{2n})}\right)^{\mathsf{T}}$, by Result 1. By the Delta Method

$$\sqrt{n}\left(\left(\begin{array}{c} U_{(k_{1n})} \\ U_{(k_{2n})} \end{array}\right) - \left(\begin{array}{c} p_1 \\ p_2 \end{array}\right)\right) \stackrel{d}{\longrightarrow} N\left(0, \dot{g}(\mu)\, \Sigma \dot{g}(\mu)^{\mathsf{T}}\right)$$

where Σ 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

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

RESULT 6: If $X_{(1)}, \ldots, X_{(n)}$ 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 $k_{1n} = \lceil np_1 \rceil$ and $k_{2n} = \lceil np_2 \rceil$

$$\sqrt{n}\left(\left(\begin{array}{c} X_{(k_{1n})} \\ X_{(k_{2n})} \end{array}\right) - \left(\begin{array}{c} x_{p_1} \\ x_{p_2} \end{array}\right)\right) \xrightarrow{d} N\left(0, \left[\begin{array}{cc} \dfrac{p_1(1-p_1)}{\{f_X(x_{p_1})\}^2} & \dfrac{p_1(1-p_2)}{f_X(x_{p_1}) f_X(x_{p_2})} \\ \\ \dfrac{p_1(1-p_2)}{f_X(x_{p_1}) f_X(x_{p_2})} & \dfrac{p_2(1-p_2)}{\{f_X(x_{p_2})\}^2} \end{array}\right]\right)$$

Proof: We use the Delta Method 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 $y_2 = p_2$.

By properties of the multivariate normal distribution, we have that the marginal distribution of $X_{(k_{1n})}$ can be approximated for large n by using the relationship

$$\sqrt{n}(X_{(k_{1n})} - x_{p_1}) \xrightarrow{d} N\left(0, \frac{p_1(1-p_1)}{\{f_X(x_{p_1})\}^2}\right)$$

For example, if $p_1 = 1/2$, x_{p_1} is the **median** $x_{F_X}(0.5)$ of the distribution, and $X_{(k_{1n})}$ is the **sample median** $\widetilde{X}_n(0.5)$, and we have that

$$\sqrt{n}(\widetilde{X}_n(0.5) - x_{F_X}(0.5)) \xrightarrow{d} N\left(0, \frac{1}{4\{f_X(x(0.5))\}^2}\right)$$

If F_X is the $N(\mu, \sigma^2)$ distribution, then $x_{F_X}(0.5) = \mu$ and

$$f_X(x(0.5)) = f_X(\mu) = \left(\frac{1}{2\pi\sigma^2}\right)^{1/2}$$

so this result says that

$$\sqrt{n}(\widetilde{X}_n(0.5) - \mu) \xrightarrow{d} N\left(0, \frac{\pi\sigma^2}{2}\right) \stackrel{.}{\sim} N\left(0, 1.57\sigma^2\right)$$

which contrasts with the exact result for the sample mean

$$\sqrt{n}(\overline{X}_n - \mu) \sim N\left(0, \sigma^2\right).$$