Foundations/Stats

STATS 743B

Cameron Roopnarine*

23rd February 2023

LECTURE 1
11th January

Order Statistics

Let $X_1, X_2, \dots, X_n$ be a random sample of size n from a population with CDF $F(x)$ and PDF $f(x)$. Let $X_{1:n} \le \cdots \le X_{n:n}$ denote the corresponding order statistics obtained by arranging the X_i's in increasing (non-decreasing) order of magnitude. Then, their distributions, dependence properties, moments, characteristics, etc. can be made use of effectively to develop inferential methods.

Binomial Derivation

The CDF of $X_{r:n}$, for $r = 1, 2, \dots, n$, is

$$\begin{aligned}
F_{r:n}(x) &= \mathbb{P}(X_{r:n} \le x) \\
&= \mathbb{P}(\text{at least } r \text{ of the } X_i\text{'s are} \le x) \\
&= \sum_{i=r}^{n} \mathbb{P}(\text{exactly } i \text{ of } x_i\text{'s are} \le x_i) && \text{because they are mutually exclusive} \\
&= \sum_{i=r}^{n} \binom{n}{i} \big(F(x)\big)^i \big(1 - F(x)\big)^{n-i} \\
&= I_{F(x)}(r, n-r+1),
\end{aligned}$$

where

$$I_p(a, b) = \frac{1}{B(a,b)} \int_0^p t^{a-1}(1-t)^{b-1}\,\mathrm{d}t, \text{ for } 0 < p < 1,$$

is the Incomplete Beta Ratio function.

Pearson's Identity

For $0 < p < 1$,

$$I_p(r, n-r+1) = \sum_{i=r}^{n} \binom{n}{i} p^i (1-p)^{n-i}, \text{ for } r = 1, 2, \dots, n.$$

It connects the survival function of binomial distribution with the cumulative distribution function of beta distribution. (Proof by integration of parts)

*LaTeXer

REMARK 1

The expression of the CDF of $X_{r:n}$ in (1) holds whether the distribution $F(x)$ is continuous or discrete.

If the distribution is continuous, then the PDF of $X_{r:n}$, for $r = 1, \ldots, n$, can be obtained from (1) as

$$\begin{aligned} f_{r:n}(x) &= \frac{\mathrm{d}}{\mathrm{d}x} F_{r:n}(x) \\ &= \frac{\mathrm{d}}{\mathrm{d}x} I_{F(x)}(r, n-r+1) \\ &= \frac{\mathrm{d}}{\mathrm{d}x} \frac{1}{B(r, n-r+1)} \int_0^{F(x)} t^{r-1}(1-t)^{n-r}\,\mathrm{d}t \\ &= \frac{1}{B(r, n-r+1)} \big(F(x)\big)^{r-1} \big(1-F(x)\big)^{n-r} f(x) \\ &= \frac{n!}{(r-1)!(n-r)!} \big(F(x)\big)^{r-1} \big(1-F(x)\big)^{n-r} f(x). \end{aligned}$$

If the population distribution $F(x)$ is discrete, however, the above method can not be used. But, we can find the PDF of $X_{r:n}$ (for $r = 1, \ldots, n$) as

$$\begin{aligned} f_{r:n}(x) &= \mathbb{P}(X_{r:n} = x) \\ &= F_{r:n}(x) - F_{r:n}(x-) \\ &= \frac{1}{B(r, n-r+1)} \int_{F(x-)}^{F(x)} t^{r-1}(1-t)^{n-r}\,\mathrm{d}t. \end{aligned}$$

Derivation from Jacobian

Now, let us focus on the case when the population distribution is continuous. In this case, with $f(x)$ as the PDF, due to independence of X_i's, their joint density is

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i), \quad x_i \in S.$$

Now, let us introduce the transformation

$$X_{1:n} = \min\{X_1, \ldots, X_n\}, \ldots, X_{n:n} = \max\{X_1, \ldots, X_n\}.$$

Then, evidently, it is an $n!$-to-1 transformation, and so the joint density function of $(X_{1:n}, \ldots, X_{n:n})$ is

$$f_{1,\ldots,n:n}(x_1, \ldots, x_n) = n! \prod_{i=1}^{n} f(x_i),\ x_1 < x_2 < \cdots < x_n.$$

From (4), we can obtain, by integrating out $(x_{r+1}, \ldots, x_n)$ and $(x_1, \ldots, x_{r-1})$, the PDF of $X_{r:n}$ $(r = 1, \ldots, n)$ as follows:

$$\int\!\!\int_{x_r < x_{r+1} < \cdots < x_n}\!\!\cdots\!\int f(x_{r+1}) \cdots f(x_n)\,\mathrm{d}x_n\,\mathrm{d}x_{n-1} \cdots \mathrm{d}x_{r+1} = \frac{[1-F(x_r)]^{n-r}}{(n-r)!}$$

and

$$\int\!\!\int_{x_1 < x_2 < \cdots < x_r}\!\!\cdots\!\int f(x_1) \cdots f(x_{r-1})\,\mathrm{d}x_1\,\mathrm{d}x_2 \cdots \mathrm{d}x_{r-1} = \frac{[F(x_r)]^{r-1}}{(r-1)!}$$

so that we obtain the PDF of $X_{r:n}$ as

$$f_{r:n}(x_r) = \frac{n!}{(r-1)!(n-r)!} \big(F(x_r)\big)^{r-1} \big(1-F(x_r)\big)^{n-r} f(x_r).$$

Similarly, from (4), we can obtain, by integrating out $(x_{s+1}, \ldots, x_n)$, $(x_{r+1}, \ldots, x_{s-1})$ and $(x_1, \ldots, x_{r-1})$, the joint PDF of $(X_{r:n}, X_{s:n})$, for $1 \le r < s \le n$, as follows:

$$\int\int \cdots \int_{x_s < x_{s+1} < \cdots < x_n} f(x_{s+1}) \cdots f(x_n) \, \mathrm{d}x_n \, \mathrm{d}x_{n-1} \, \mathrm{d}x_{s+1} = \frac{[1 - F(x_s)]^{n-s}}{(n-s)!},$$

$$\int\int \cdots \int_{x_r < x_{r+1} < \cdots < x_{s-1} < x_s} f(x_{r+1}) \cdots f(x_{s-1}) \, \mathrm{d}x_{r+1} \, \mathrm{d}x_{r+2} \, \mathrm{d}x_{s-1} = \frac{[F(x_s) - F(x_r)]^{s-r-1}}{(s-r-1)!},$$

and

$$\int\int \cdots \int_{x_1 < x_2 < \cdots < x_{r-1} < x_r} f(x_1) \cdots f(x_{r-1}) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \, \mathrm{d}x_{r-1} = \frac{[F(x_r)]^{r-1}}{(r-1)!},$$

so that we can obtain the joint PDF of $(X_{r:n}, X_{s:n})$, for $1 \le r < s \le n$ as

$$f_{r,s:n}(x_r, x_s) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \big(F(x_r)\big)^{r-1} \big(F(x_s) - F(x_r)\big)^{s-r-1} \big(1 - F(x_s)\big)^{n-s} f(x_r) f(x_s), \quad \text{for } x_r < x_s.$$

Multinomial Derivation

For directly deriving the PDF of $X_{r:n}$, let us consider

$$\mathbb{P}(x \le X_{r:n} \le x + \Delta x) = \frac{n!}{(r-1)!1!(n-r)!} \big(F(x)\big)^{r-1} \big(F(x+\Delta x) - F(x)\big)^1 \big(1 - F(x+\Delta x)\big)^{n-r} + \mathcal{O}((\Delta x)^2),$$

where $\mathcal{O}((\Delta x)^2)$ corresponds to more than one x_i in the interval $(x, x+\Delta x)$. Then, we obtain the density of $X_{r:n}$ as follows:

$$\begin{aligned}
f_{r:n}(x) &= \lim_{\Delta x \to 0} \frac{\mathbb{P}(x \le X_{r:n} \le x + \Delta x)}{\Delta x} \\
&= \frac{n!}{(r-1)!1!(n-r)!} \big(F(x)\big)^{r-1} \lim_{\Delta x \to 0} \Big[\big(F(x+\Delta x) - F(x)\big)^1 \\
&\qquad \big(1 - F(x+\Delta x)\big)^{n-r} \Big] + \mathcal{O}((\Delta x)^2) \\
&= \frac{n!}{(r-1)!(n-r)!} \big(F(x)\big)^{r-1} f(x) \big(1 - F(x)\big)^{n-r},
\end{aligned}$$

exactly as before.

Similarly, for deriving the joint PDF of $(X_{r:n}, X_{s:n})$, for $1 \le r < s \le n$, let us consider the multinomial probability

$$\begin{aligned}
&\mathbb{P}(x < X_{r:n} \le x + \Delta x, y < X_{s:n} \le y + \Delta y) \\
&= \frac{n!}{(r-1)!1!(s-r-1)!1!(n-s)!} \big(F(x)\big)^{r-1} \big(F(x+\Delta x) - F(x)\big)^1 \\
&\quad \times \big(F(y) - F(x+\Delta x)\big)^{s-r-1} \big(F(y+\Delta y) - F(y)\big)^1 \big(1 - F(y+\Delta y)\big)^{n-s} \\
&\quad + \mathcal{O}((\Delta x)^2 \Delta y) \to \text{corresponds to more than one } X_i \text{ in } (x, x+\Delta x) \\
&\quad + \mathcal{O}(\Delta x (\Delta y)^2) \to \text{corresponds to more than one } X_i \text{ in } (y, y+\Delta y)
\end{aligned}$$

Then, we obtain the joint density of $(X_{r:n}, X_{s:n})$ as follows:

$$\begin{aligned}
f_{r,s:n}(x,y) &= \lim_{\Delta x \to 0, \Delta y \to 0} \frac{\mathbb{P}(x < X_{r:n} \le x + \Delta x, y < X_{s:n} \le y + \Delta y)}{\Delta x \Delta y} \\
&= \frac{n!}{(r-1)!1!(s-r-1)!1!(n-s)!} \big(F(x)\big)^{r-1} \\
&\times \lim_{\Delta x \to 0} \left[\frac{\big(F(x+\Delta x) - F(x)\big)^1}{\Delta x} \big(F(y) - F(x+\Delta x)\big)^{s-r-1} \right] \\
&\times \lim_{\Delta y \to 0} \left[\frac{\big(F(y+\Delta y) - F(y)\big)^1}{\Delta y} \big(1 - F(y+\Delta y)\big)^{n-s} \right] + \mathcal{O}((\Delta x)^2 \Delta y) + \mathcal{O}(\Delta x (\Delta y)^2) \\
&= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \big(F(x)\big)^{r-1} f(x) \big(F(y) - F(x)\big)^{s-r-1} f(y) \big(1 - F(y)\big)^{n-s}, \ x < y,
\end{aligned}$$

exactly as before.

EXAMPLE 1

Let us consider Uniform$(0,1)$ distribution with

$$f(x) = 1 \text{ for } 0 < x < 1, \qquad F(x) = x \text{ for } 0 < x < 1.$$

Then, from (2), we have the PDF of $X_{r:n}$ (for $1 \le r \le n$) to be

$$f_{r:n}(x) = \frac{1}{B(r, n-r+1)} x^{r-1}(1-x)^{n-r} \text{ for } 0 < x < 1;$$

that is,

$$X_{r:n} \stackrel{\text{d}}{=} \text{Beta}(r, n-r+1).$$

So, we readily have

$$\mathbb{E}[X_{r:n}] = \frac{r}{n+1} = \pi_r, \qquad \text{Var}(X_{r:n}) = \frac{r(n-r+1)}{(n+1)^2(n+2)} = \frac{\pi_r(1-\pi_r)}{n+2}.$$

Similarly, from (6), we have the joint PDF of $(X_{r:n}, X_{s:n})$ for $1 \le r < s \le n$, to be

$$f_{r,s:n}(x,y) = \frac{1}{B(r, s-r, n-s+1)} x^{r-1}(y-s)^{s-r-1}(1-y)^{n-s} \text{ for } 0 < x < y < 1,$$

which implies

$$(X_{r:n}, X_{s:n}) \stackrel{\text{d}}{=} \text{BivBeta}(r, s-r, n-s+1).$$

From this, we can readily find, for $1 \le r < s \le n$,

$$\begin{aligned}
\text{Cov}(X_{r:n}, X_{s:n}) &= \mathbb{E}[X_{r:n} X_{s:n}] - \mathbb{E}[X_{r:n}]\,\mathbb{E}[X_{s:n}] \\
&= \frac{\pi(n-s+1)}{(n+1)^2(n+2)} \\
&= \frac{\pi_r(1-\pi_s)}{n+2};
\end{aligned}$$

observe that they are always positively correlated. Moreover,

$$\rho_{X_{r:n}, X_{s:n}} = \frac{\text{Cov}(X_{r:n}, X_{s:n})}{\sqrt{\text{Var}(X_{r:n})\,\text{Var}(X_{s:n})}} = \sqrt{\frac{\pi_r}{1-\pi_r} \times \frac{1-\pi_s}{\pi_s}},$$

free of n (just a function of proportions $\frac{r}{n+1}$ and $\frac{s}{n+1}$).

Probability Integral Transform

Suppose X is a continuous random variable with CDF $F(x)$ and PDF $f(x)$. Then, the transformed variable $U = F(x)$ is uniformly distributed over $(0,1)$.

Proof: For $u \in (0,1)$, consider

$$\begin{aligned}\mathbb{P}(U \leq u) &= \mathbb{P}(F(x) \leq u) \\ &= \mathbb{P}(X \leq Q(u)) \\ &= F(Q(u)) \\ &= u,\end{aligned}$$

where $Q(u)$ is the quantile function (i.e., it is F^{-1} in the case of absolute continuous function), which means $U = F(x)$ is Uniform$(0,1)$.

REMARK 2

The above result will hold even if the population distribution is not absolutely continuous, but has discontinuities. All we have to do is take Q as the generalized quantile function with right inverse.
Since $F(x)$ is a non-decreasing function, if we have $(X_{1:n}, X_{2:n}, \ldots, X_{n:n})$ as order statistics from a continuous distribution with CDF $F(x)$, then the transformed variables $\big(F(X_{1:n}), F(X_{2:n}), \ldots, F(X_{n:n})\big)$ will be distributed as Uniform$(0,1)$ order statistics, $(U_{1:n}, U_{2:n}, \ldots, U_{n:n})$ no matter what the distribution of $F(\,\cdot\,)$ is!

Probability-Probability Plot

One important application of the previous result is in model validation methods. Because

$$\big(F(X_{1:n}, \ldots, F(X_{n:n}))\big) \text{ and } (U_{1:n}, \ldots, U_{n:n})$$

have identical distributions no matter what the population distribution $F(\,\cdot\,)$ is, we can use it to examine whether an assumed $F(\,\cdot\,)$ is reasonable for the data at hand. This is done by a P-P plot as follows:

- Step 1: From the given data $(x_{1:n}, \ldots, x_{n:n})$, estimate the parameters of the assumed model $F(\,\cdot\,)$;
- Step 2: With the estimated parameter values, say $\hat{\theta}$, find the values of

  $$\big(F(x_{1:n};\hat{\theta}), F(x_{2:n};\hat{\theta}), \ldots, F(x_{n:n};\hat{\theta})\big).$$

  These are the "empirical" (or observed) probabilities;
- Step 3: Plot these against the "theoretical" probabilities

  $$\left(\mathbb{E}[U_{1:n}] = \frac{1}{n+1}, \mathbb{E}[U_{2:n}] = \frac{2}{n+1}, \ldots, \mathbb{E}[U_{n:n}] = \frac{n}{n+1}\right).$$

  A near straight line fit would provide support for the assumed model $F(\,\cdot\,)$.

REMARK 3

One can also indicate variability at each point by estimating $\mathrm{Var}(F(x_{i:n}))$ (using delta method).

Quantile-Quantile Plot

Another important and related application is the Q-Q plot. In it, we invert the distributional relationship to use

$$(X_{1:n}, X_{2:n}, \ldots, X_{n:n}) \text{ and } (F^{-1}(U_{1:n}), F^{-1}(U_{2:n}), \ldots, F^{-1}(U_{n:n}))$$

have identical distribution, where $Q \equiv F^{-1}(\,\cdot\,)$ is the quantile function of the assumed model. Then, the $Q-Q$ plot proceeds as follows:

- Step 1: Determine the order statistics from the given data, $x_{1:n}, x_{2:n}, \ldots, x_{n:n}$, which are the empirical quantiles;
- Plot them against the theoretical quantiles

$$\left(F^{-1}\left(\frac{1}{n+1}\right), F^{-1}\left(\frac{2}{n+1}\right), \ldots, F^{-1}\left(\frac{n}{n+1}\right)\right).$$

Once again, a near straight line fit would provide support for the assumed model $F(\,\cdot\,)$.

REMARK 4

Once again, we can indicate the variability at each point by estimating $\mathrm{Var}(X_{i:n})$.

REMARK 5

In both cases, rather than making a qualitative assessment on "near straight line fit", can can make it quantitatively by using the correlation coefficient or any other "measure of fit" (correlation-type goodness-of-fit test).

REMARK 6

Note that estimation of the model parameters is avoided in Q-Q plot, but is necessary in a P-P plot!

Pivot 1

A pivot is a random variable, which is a function of the data and the parameter of the model, whose distribution is free of the parameter(s).

Pivots are essential quantities for developing inferential methods such as interval estimation, hypothesis tests, etc.

EXAMPLE 2

Let $X_1, \ldots, X_n$ be a random sample from EXP(θ) distribution with PDF

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta},\ x > 0,\ \theta > 0.$$

Then, it is well-known that

$$\sum_{i=1}^{n} X_i \sim \text{GAM}(n, \theta),$$

where n is the shape parameter and θ is the scale parameter of the Gamma distribution. Hence,

$$Y = \frac{\sum_{i=1}^{n} X_i}{\theta}$$

is a pivot (for θ) since its distribution is GAM$(n, 1)$ which is free of the parameter θ.

EXAMPLE 3

Suppose $X_1, \ldots, X_n$ is a random sample from $\mathcal{N}(\mu, \sigma^2)$ distribution. Let $\bar{X}$ and S^2 denote the sample mean and sample variance, respectively. Then, it is known that

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \text{ and } \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

So, if we consider

$$\bar{X} - \mu \sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right),$$

it will not be a pivot if σ^2 is unknown, and it will be a pivot for μ only if σ^2 is known. However, if we consider

$$\frac{\bar{X} - \mu}{S/\sqrt{n}},$$

it will be a pivot (for μ) since its distribution will be Student's t distribution with $n-1$ degrees of freedom, as it is free of both μ and σ^2. Similarly,

$$\frac{(n-1)S^2}{\sigma^2}$$

will be a pivot (for σ^2) as its distribution is central χ^2 distribution with $n-1$ degrees of freedom, and is free of μ and σ^2.

EXAMPLE 4

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from U$(0, \theta)$ distribution. Let $X_{1:n}, \ldots, X_{n:n}$ be the corresponding order statistics. Then,

$$T = \frac{X_{n:n}}{\theta}$$

is a pivot (for θ) since its density function is

$$f_T(t) = nt^{n-1},\ 0 < t < 1,$$

which is free of θ.

REMARK 7

All the pivotal quantities discussed above are all "parametric pivots" as they are pivots for parameters of a specific parametric model assumed.

EXAMPLE 5

Let $X_1, \ldots, X_n$ be a random sample from BERN(π) distribution, where $\pi \in (0,1)$ is the probability of success. Then, it is well-known that

$$Y = \sum_{i=1}^{n} X_i \sim \text{BIN}(n, \pi).$$

As the distribution of Y depends on the parameter π, it is not a pivot. In fact, no exact pivot exists in this case.

However, by the use of Central Limit Theorem, it is known that

$$Z = \frac{Y - n\pi}{\sqrt{n\pi(1-\pi)}} \overset{\text{asymp.}}{\sim} \mathcal{N}(0,1).$$

As the distribution of Z is, asymptotically, standard normal, which is free of π, it can serve as a pivot. But, note that it is only an approximate pivot for π.

Pivot for Population Quantile

Suppose we have a random sample from $\mathcal{N}(\mu, \sigma^2)$ distribution, and that we are interested in inferring about p-th quantile $\xi_p = \mu + \sigma z_p$, where z_p is the p-th quantile of the standard normal distribution. Then, it is evident that, with $\bar{X}$ and S as estimates of μ and σ respectively, then an estimate of ξ_p is

$$\hat{\xi}_p = \hat{\mu} + \hat{\sigma} z_p = \bar{X} + z_p S.$$

Then, the variable

$$\begin{aligned} T &= \frac{\hat{\xi}_p - \xi_p}{S} \\ &= \frac{(\bar{X} + z_p S) - (\mu + z_p \sigma)}{S} \\ &= \frac{(\bar{X} - \mu) + z_p (S - \sigma)}{S} \\ &= \frac{\bar{X} - \mu}{S} + z_p \left(1 - \frac{1}{S/\sigma}\right) \end{aligned}$$

is a pivot as its distribution is free of parameters μ and σ. Hence, this pivot could be used for developing inference for the p-th quantile ξ_p of the normal distribution.

EXERCISE 1

Can you think of a way to find its percentage points?

REMARK 8

Though the above derivation is shown for normal distribution, it can be done similarly for any member of location-scale family of distributions like Logistic, Laplace, Gumbel distributions.

Non-parametric Confidence Interval for Quantile

Now, let us assume we have a random sample $X_1, X_2, \dots, X_n$ from a distribution function $F(x)$ that is continuous. Let ξ_p be the p-th quantile of F. Then, $F(\xi_p) = \mathbb{P}(X \le \xi_p) = p$. We are interested in a confidence interval for ξ_p, but without assuming a specific form of the distribution F, like normal!

Let $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ denote the order statistics obtained from the sample. Let $X_{r:n}$ and $X_{s:n}$ be two selected order statistics, for $1 \le r < s \le n$. Then, we have:

$$\begin{aligned} \mathbb{P}(X_{r:n} \le \xi_p \le X_{s:n}) &= \mathbb{P}(F(X_{r:n}) \le F(\xi_p) \le F(X_{s:n})) \\ &= \mathbb{P}(U_{r:n} \le p \le U_{s:n}) \\ &= \mathbb{P}(p \le U_{s:n}) - \mathbb{P}(p < U_{r:n}) \\ &= 1 - \mathbb{P}(U_{s:n} < p) - \big(1 - \mathbb{P}(U_{r:n} < p)\big) \\ &= \mathbb{P}(U_{r:n} < p) - \mathbb{P}(U_{s:n} < p) \\ &= \sum_{i=r}^{p} \binom{n}{i} p^i (1-p)^{n-i} - \sum_{i=s}^{n} \binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i}, \end{aligned}$$

where $U_{r:n}$ and $U_{s:n}$ are order statistics from Uniform$(0, 1)$ distribution. Thus, $(X_{r:n}, X_{s:n})$ is a non-parametric confidence interval for the p-th population quantile ξ_p, with its coverage probability not depending on F, but only on p and n.

So, for a given confidence level $1-\alpha$, all we need to do is, for a given sample size n and the quantile level p, we need to determine integers r and s such that

$$1 \leq r < s \leq n$$

and

$$\sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i} \approx 1-\alpha.$$

Note that $1-\alpha$ may not be achievable exactly as the binomial distribution is discrete and so has jumps.

REMARK 9

The choice of r and s may not be unique. So, if there is more than one choice of (r, s) satisfying the above conditions, then it would be meaningful to choose that pair (r, s) for which

$$s - r \text{is the smallest}$$

among all these choices satisfying the conditions. This would then correspond to the "narrowest non-parametric confidence interval for population quantile."