MANA130083.01 Nonparametric *Spring 2025*

Lecture 1

Instructor: Prof. Bowen Gang *Scribes: Yize Wang, Jingyi Zhou*

1.1 Order Statistics

Take a random sample, $X_1, X_2, ..., X_n$. Then order them from smallest to largest, $X_{(1)} \leqslant X_{(2)} \leqslant X_{(3)} \leqslant ... \leqslant X_{(n)}$, $X_{(r)}$ is called the r-th order statistic.

- minimum: $X_{(1)}$
- maximum: $X_{(n)}$
- median: $X_{(\frac{n+1}{2})}$, where n odd; $\frac{1}{2}[X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}]$, where n even
- range: $X_{(n)} - X_{(1)}$

1.2 Quantile Function

Let F_X be the CDF of X, then $Q_X(p) = F_X^{-1}(p) = inf[x : F_X(x) \geqslant p]$, where $0 \leqslant p \leqslant 1$. Properties of quantile functions:

1 $\mathbb{E}[X] = \int_0^1 Q_X(p)dp$ (expectation)

Proof.

$$\begin{aligned}\int_0^1 Q_X(p)\,dp &= \int_0^1 F_X^{-1}(p)\,dp \\ &\overset{\text{Let } F_X^{-1}(p)=t}{\underset{p=F_X(t)}{=\!=}} \int_{t_0}^{t_1} tf(t)\,dt \quad \text{where } t_0 = F_X^{-1}(0), t_1 = F_X^{-1}(1), \\ &= \mathbb{E}[X]\end{aligned}$$

□

2 $\mathbb{E}[X^2] = \int_0^1 Q_X^2(p)dp$ (second moment)
The proof is similar to property 1.

1.3 Empirical CDF

CDF is unknown most of the time. We can try to estimate CDF. Because we have $F_X(x) = Pr(X \leqslant x)$, we can simply count how many observations are less or equal to x in our sample. Say our sample is $X_1, X_2, ..., X_n$, then

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leqslant x)$$

Theorem. $n\hat{F}_n(x) \sim Bin\left(n, F_X(x)\right)$

Proof.

$$n\hat{F}_n(x) = \sum_{i=1}^{n} I(X_i \leqslant x)$$

Note, The probability of $I(X_i \leqslant x)$ taking 1 is $P(X_i \leqslant x) = F_X(x)$. □

Corollary. *The following properties are immediate from binomial distribution.*

1 *(unbiased)* $\mathbb{E}[\hat{F}_n(x)] = F_X(x)$

2 $Var[\hat{F}_n(x)] = \frac{1}{n} F(x)\left(1 - F(x)\right)$

3 *(By Chebyshev)* $\hat{F}_n(x) \xrightarrow{p} F_X(x)$

In terms of the convergence property of empirical CDF, something stronger can be said about it.

Theorem. *(Glivenko-Cantelli)*

$$\sup_x |\hat{F}_n(x) - F(x)| \xrightarrow{a.s.} 0$$

where almost sure convergence means

$$\Pr\left(\lim_{n\to\infty} \sup_x |\hat{F}_n(x) - F(x)| = 0\right) = 1$$

Example. *Two examples: One interval example ($S \sim \mathcal{U}(0,1)$, $X_1(s) = s + \mathbb{I}_{(0,1)}(s)$, $X_2(s) = s + \mathbb{I}_{(0,1/2)}(s)$ ...); Another indicator function example (where $\lim_{n\to\infty} f_n(x) = 0$ but $\sup_x |f_n(x)| = 1$)*

This theorem is stronger because it is uniform convergence, which is about the 'worst case'. It is stronger than 'point-wise' convergence. Next we will discuss about the convergence rate.

Theorem. *(Dvoretzky-Kiefer-Wolfowitz Inequality)*

$$\Pr\left(\sup_x |\hat{F}_n(x) - F(x)| > \epsilon\right) \leqslant 2e^{-2n\epsilon}$$

So the empirical CDF is a good approximation of the actual CDF given that the above probability decreases exponentially.

$$\Pr\left\{\frac{\sqrt{n}\left[\hat{F}_n(x) - F_X(x)\right]}{\sqrt{F_X(x)\left(1 - F_X(x)\right)}} \leqslant t\right\} = \Phi(t)$$

where $\Phi(\cdot)$ is the CDF of $\mathcal{N}(0,1)$.

1.4 Empirical Quantile Function

$$\hat{Q}_n(u) = \begin{cases} X_{(1)}\ if\ 0 < u \leqslant \frac{1}{n} \\ X_{(2)}\ if\ \frac{1}{n} < u \leqslant \frac{2}{n} \\ \vdots \\ X_{(n)}\ if\ \frac{n-1}{n} < u \leqslant 1 \end{cases}$$

It is equivalent to say

$$\hat{Q}_n(u) = inf\left[x : \hat{F}_n(x) \geqslant u\right]$$

1.5 Properties and Applications of Order Statistics

Theorem. *(Property of order statistics)*

$$\begin{aligned}\Pr\left(X_{(r)} \leqslant t\right) &= \sum_{i=r}^{n} Pr\left[n\hat{F}_n(t) = i\right] \\ &= \sum_{i=r}^{n} \binom{n}{i} [F_X(t)]^i [1 - F_X(t)]^{n-i}\end{aligned}$$

Theorem. *(Another property of order statistics)*

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!}[F_X(x)]^{r-1}[1 - F_X(x)]^{n-r} f_X(x)$$

A special case: if $X \sim U[0, 1]$, $F_X(t) = t$ where $0 < t < 1$, then

$$F_{X_{(r)}}(t) = \sum_{i=r}^{n} \binom{n}{i} t^i (1-t)^{n-i}$$

$$f_{X_{(r)}}(t) = \frac{n!}{(r-1)!(n-r)!} t^{r-1}(1-t)^{n-r}, 0 < t < 1$$

which implies that $X_{(r)} \sim Beta(r, n - r + 1)$.

Note that here both pdf and CDF are not distribution free.

Corollary. *CDF is the integral of PDF:*

$$\sum_{i=r}^{n} \binom{n}{i} t^i (1-t)^{n-i} = \frac{1}{B(r, n-r+1)} \int_0^t x^{r-1}(1-x)^{n-r} dx$$

Why do we care about the uniform distribution? Think about the following probability integral transform.

$$U_{(r)} = F_X(X_{(r)})$$

where $U_{(r)}$ is the r-th order statistics from $U[0, 1]$. This is based on the following fact: if $X \sim F_X$ and $Y = F_X(X)$, then $Y \sim U[0, 1]$. To prove this fact, think about the CDF of Y:

$$\begin{aligned}F_Y(u) &= Pr\left(Y \leqslant u\right) \\ &= Pr\left(F_X(X) \leqslant u\right) \\ &= Pr\left(F^{-1}(F_X(X)) \leqslant F^{-1}(u)\right) \\ &= Pr\left(X \leqslant Q(u)\right) \\ &= u\end{aligned}$$

This fact can be used to generate random number (a random number generator, though actually it is pseudo random). If $U \sim U[0,1]$ and $X = F_X^{-1}(U)$, then $X \sim F_X$. Note that the computer can generate uniformly distributed random numbers by the following steps

$$\begin{aligned} a_0 &= initial\ number \\ a_{n+1} &\equiv ba_n + c \mod m \\ Then\ \frac{a_0}{m}&, \frac{a_1}{m}, \frac{a_2}{m}, \ldots \sim U[0,1] \end{aligned}$$

Here is also the first random number generator proposed by Von Neumann:

$$m_{n+1} = m_n^2 \mod 10000$$

It is not truly random for sure though it passes all tests of randomness. The concerns are reasonable but this is exactly what the first-generation computer uses.

Suppose we want to generate random numbers following exponential distributions with parameter 2: $X \sim Exp(2)$. It has CDF: $F_X(x) = 1 - e^{-\frac{x}{2}}$. We can first generate $U \sim U[0,1]$ and apply the inverse of the CDF: $X = -2ln(1-U)$. Then $X \sim F_X$.

1.6 Further Discussion of Order Statistics Density

1.6.1 Joint Distribution

The joint density is given by

$$f_{X_{(1)},\ldots,X_{(n)}}(y_1, \ldots, y_n) = n! \prod_{i=1}^{n} f_X(y_i)$$

Note that n samples have $n!$ permutations and that is why there is a factorial above.

1.6.2 Marginal Density

We have already known that the PDF of k-th order statistics is given by

$$\begin{aligned} f_{X_{(r)}}(x) &= \binom{n}{k-1}\binom{n-k+1}{1}[F(x)]^{k-1} f(x)[1-F(x)]^{n-k} \\ &= \frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}[1-F(x)]^{n-k} f(x) \end{aligned}$$

Here is the strict proof. To begin with, we define some intervals: $I_1 = (-\infty, x]$, $I_2 = (x, x+h]$, $I_3 = (x+h, \infty)$. Correspondingly, $P_1 = Pr(x \in I_1) = F_X(x)$, $P_2 = Pr(x \in I_2) = F_X(x+h) - F_X(x)$, $P_3 = Pr(x \in I_3) = 1 - F_X(x+h)$. Then

$$\begin{aligned} f_{X_{(r)}}(x) &= \lim_{h\to 0} \frac{F_{X_{(r)}}(x+h) - F_{X_{(r)}}(x)}{h} \\ &= \lim_{h\to 0} \binom{n}{r-1, 1, n-r} P_1^{r-1} P_2 P_3^{n-r} \\ &= \frac{n!}{(r-1)!1(n-r)!}\left[F_X(x)\right]^{r-1} \lim_{h\to 0}\left\{\frac{F_{X_{(r)}}(x+h) - F_{X_{(r)}}(x)}{h}\left[1 - F_X(x+h)\right]^{n-r}\right\} \\ &= \frac{n!}{(r-1)!(n-r)!}\left[F_X(x)\right]^{r-1} f_X(x)\left[1 - F_X(x)\right]^{n-r} \end{aligned}$$

For the joint distribution of two order statistics, we have also known that

$$f_{X(r),X(s)}(x,y)$$
$$= \begin{cases} \frac{n!}{(r-1)!(s-r-1)!(n-s)!}\left[F_X(x)\right]^{r-1}\left[F_X(y)-F_X(x)\right]^{s-r-1}\left[1-F_X(y)\right]^{n-s}f_X(x)f_X(y), x<y \\ 0, else. \end{cases}$$

where $r < s$. Here is the strict proof. Similarly, we define $I_1 = (-\infty, x]$, $I_2 = (x, x+h]$, $I_3 = (x+h, y]$, $I_4 = (y, y+t]$, $I_5 = (y+t, \infty)$. Correspondingly, $P_1 = F_X(x)$, $P_2 = F_X(x+h) - F_X(x)$, $P_3 = F_X(y) - F_X(x+h)$, $P_4 = F_X(y+t) - F_X(y)$, $P_5 = 1 - F_X(y+t)$. Then

$$\begin{aligned} f_{X_{(r)},X_{(s)}}(x,y) &= \lim_{\substack{h\to 0\\ t\to 0}} \frac{F_{X_{(r)},X_{(s)}}(x+h,y+t) - F_{X_{(r)},X_{(s)}}(x,y+t) - F_{X_{(r)},X_{(s)}}(x+h,y) + F_{X_{(r)},X_{(s)}}(x,y)}{ht} \\ &= \lim_{\substack{h\to 0\\ t\to 0}} \frac{Pr\left(x \leqslant X_{(r)} \leqslant x+h, y \leqslant X_{(s)} \leqslant y+t\right)}{ht} \\ &= \lim_{\substack{h\to 0\\ t\to 0}} \binom{n}{r-1,1,s-r-1,1,n-s} P_1^{r-1}P_2P_3^{s-r-1}P_4P_5^{n-s} \end{aligned}$$

Plug in P_1, P_2, P_3, P_4 and P_5, we get

$$f_{X_{(r)},X_{(s)}}(x,y)$$
$$\begin{aligned} &= \binom{n}{r-1,1,s-r-1,1,n-s}\left[F_X(x)\right]^{r-1}\lim_{\substack{h\to 0\\ t\to 0}}\left\{\frac{F_X(x+h)-F_X(x)}{h}\left[F_X(y)-F_X(x+h)\right]^{s-r-1}\right\} \\ &\cdot \lim_{\substack{h\to 0\\ t\to 0}}\left\{\frac{F_X(y+t)-F_X(t)}{t}\left[1-F_X(y+t)\right]^{n-s}\right\} \\ &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!}\left[F_X(x)\right]^{r-1}\left[F_X(y)-F_X(x)\right]^{s-r-1}\left[1-F_X(y)\right]^{n-s}f_X(x)f_X(y) \end{aligned}$$

A special case: if $X \sim U[0,1]$, then

$$f_{X_{(r)},X_{(s)}}(x,y) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!}x^{r-1}(y-x)^{s-r-1}(1-y)^{n-s}$$

The above result has several applications.

1 Distribution of median
Suppose we have samples $X_1, X_2, ..., X_n$, where n is even, the median is given by $U = \frac{1}{2}\left[X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}\right]$. By method of Jacobian, we can calculate the distribution of median.

2 Distribution of range
Similarly we can calculate the distribution of $X_{(n)} - X_{(1)}$.

1.6.3 Moments of Order Statistics

Now we know the distribution of order statistics, so we can calculate the moments.

$$\begin{aligned} \mathbb{E}[X_{(r)}^k] &= \frac{n!}{(r-1)!(n-r)!}\int_{-\infty}^{\infty} y^k\left[F_X(y)\right]^{r-1}\left[1-F_X(y)\right]^{n-r}f_X(y)dy \\ &= \frac{n!}{(r-1)!(n-r)!}\int_0^1 \left[Q_X(u)\right]^k u^{r-1}\left[1-u\right]^{n-r}du \ \ (\text{let } y = Q_X(u)) \\ &= \mathbb{E}[Q_X(u)]^k \end{aligned}$$

where the expectation is taken with respect to u. Note that $u^{r-1}(1-u)^{n-r}$ is the kernel of beta distribution. U follows beta distribution. A special case: $X \sim U[0,1]$, $Q_X(u) = u$, then

$$\mathbb{E}[X_{(r)}^k] = \frac{n!(r+k-1)!}{(n+k)!(r-1)!}$$

which also implies that $\mathbb{E}[X_{(r)}] = \frac{r}{n+1}$. We then look at the variance: $Var[X_{(r)}] = \mathbb{E}[X_{(r)}^2] - \left[\mathbb{E}[X_{(r)}]\right]^2 = \frac{r(n-r+1)}{(n+1)^2(n+2)}$ For the covariance and correlation between $X_{(r)}$ and $X_{(s)}$, we can predict that it is positive and it decreases when sample size n increases. The covariance is calculated as following.

$$\mathbb{E}[X_{(r)}X_{(s)}] =$$

$$\frac{n!}{(r-1)!(s-r-1)!(n-s)!}\int_{-\infty}^{\infty}\int_{-\infty}^{y} xy\left[F_X(x)\right]^{r-1}\left[F_X(y)-F_X(x)\right]^{s-r-1}\left[1-F_X(y)\right]^{n-s} f_X(x)f_X(y)dxdy$$

If we let $F_X(x) = u$, $F_X(y) = v$, then $du = f_X(x)dx$, $dv = f_X(y)dy$. And the above integral becomes

$$\mathbb{E}[X_{(r)}X_{(s)}] = \frac{n!}{(r-1)!(s-r-1)!(n-s)!}\int_0^1\int_0^{Q_X(v)} Q_X(u)Q_X(v)u^{r-1}(v-u)^{s-r-1}(1-v)^{n-s}dudv$$

where in the integral x corresponds to $X_{(r)}$ and y corresponds to $X_{(s)}$. Because we also have $X \sim U[0,1]$, $Q_X(u) = u$, $Q_X(v) = v$, then

$$\mathbb{E}[X_{(r)}X_{(s)}] = \frac{n!}{(r-1)!(s-r-1)!(n-s)!}\int_0^1\int_0^{v} u^r v(v-u)^{s-r-1}(1-v)^{n-s}dudv$$

We can then evaluate the integral by the following variable transformation. If we let $z = \frac{u}{v}$, then the integral becomes a kernel corresponding to beta distribution:

$$\mathbb{E}[X_{(r)}X_{(s)}] = \frac{r(s+1)}{(n+1)(n+2)}$$

So

$$\begin{aligned}
Cov\left[X_{(r)}, X_{(s)}\right] &= \mathbb{E}\left[X_{(r)}X_{(s)}\right] - \mathbb{E}\left[X_{(r)}\right]\mathbb{E}\left[X_{(s)}\right] \\
&= \frac{r(s+1)}{(n+1)(n+2)} - \frac{rs}{(n+1)^2} \\
&= \frac{r(n-s+1)}{(n+1)^2(n+2)},\ for\ r < s \\
Corr\left[X_{(r)}, X_{(s)}\right] &= \sqrt{\frac{r(n-s+1)}{s(n-r+1)}},\ for\ r < s
\end{aligned}$$

A special case: if $r = 1$ and $s = n$ then

$$Corr\left[X_{(r)}, X_{(s)}\right] = \frac{1}{n}$$

1.6.4 Approximation of Moments of Order Statistics

To introduce the topic, consider the following problem: If we know $\mathbb{E}[X] = \mu$ and we want to estimate $\mathbb{E}[X^k]$, is it reasonable to use μ^k? The answer is yes: it is not a nonsense estimation because it is actually a 1-st order approximation.

We can implement Taylor theorem. Let z be a random variable and g a smooth function, which means that its derivatives of all orders exist. Then

$$g(z) = g(\mu) + \sum_{i=1}^{\infty} \frac{(z-\mu)^i}{i!} g^{(i)}(\mu)$$

where $\mu = \mathbb{E}[z]$. Take expectations on both sides,

$$\mathbb{E}\left[g(z)\right] = g(\mu) + \frac{\sigma^2}{2!} g^{(2)}(\mu) + \sum_{i=3}^{\infty} \frac{\mathbb{E}\left[(z-\mu)^i\right]}{i!} g^{(i)}(\mu)$$

which gives the 2nd order approximation of $g(z)$. Note that the result is consistent with Jensen's inequality which is about convex functions. To approximate $Var\left(g(z)\right)$, we can first calculate

$$g(z) - \mathbb{E}\left[g(z)\right]$$
$$= (z-\mu)g^{(1)}(\mu) + \frac{g^{(2)}(\mu)}{2!}\left[(z-\mu)^2 - Var[z]\right] + \sum_{i=3}^{\infty} \frac{g^{(i)}(\mu)}{i!}\left[(z-\mu)^i - \mathbb{E}[(z-\mu)^i]\right]$$

Similarly,

$$\left[g(z) - \mathbb{E}[g(z)]\right]^2$$
$$= (z-\mu)^2\left[g^{(i)}(\mu)\right]^2 + \frac{1}{4}\left[g^{(2)}(\mu)\right]^2\left[Var^2[z] - 2Var[z](z-\mu)^2\right] - g^{(1)}(\mu)g^{(2)}(\mu)Var[z](z-\mu) + h(z)$$

where $h(z)$ is high-order terms. Take expectations on both sides, we get

$$Var\left[g(z)\right] = \underbrace{\underbrace{\sigma^2[g^{(1)}(\mu)]^2}_{first\ order} - \frac{1}{4}[g^{(2)}(\mu)]^2\sigma^4}_{second\ order\ approximation} + \mathbb{E}[h(z)]$$

There is also an application of this approximation: if $X \sim N(\mu, \sigma^2)$, then $g(X) \sim N\left(g(\mu), \sigma^2[g^{(1)}(\mu)]^2\right)$ approximately (also called delta method). So how does it relate to order statistics? Let $z = U_{(r)}$, $X_{(r)} = F_X^{-1}\left(U_{(r)}\right)$, $g(\cdot) = Q_X(\cdot)$, and also $\mu = \mathbb{E}[z] = \frac{r}{n+1}$, $\sigma^2 = Var[z] = \frac{r(n-r+1)}{(n+1)^2(n+2)}$, then the 1st order approximation is given by

$$\mathbb{E}\left[X_{(r)}\right] = F_X^{-1}\left(\frac{r}{n+1}\right), Var\left[X_{(r)}\right] = \frac{r(n-r+1)}{(n+1)^2(n+2)}\left\{f_X\left[F_X^{-1}(\frac{r}{n+1})\right]\right\}^{-2}$$

The next question is about the asymptotic distribution of $X_{(r)}$. Let $\frac{r}{n} \to p \in (0,1)$, then we have the following theorem.

Theorem. *(root n consistent)*

$$\left[\frac{n}{p(1-p)}\right]^{\frac{1}{2}} f_X(\mu)\left[X_{(r)} - \mu\right] \xrightarrow{D} N(0,1)$$

where $\mu = F_X^{-1}(p)$.

Proof. (sketch) First we can consider the simplest case $X \sim U[0,1]$. There is no loss of generality because we can transform uniform distribution to any distribution as long as we know the CDF. We have known that

$$f_{U_{(r)}}(u) = \frac{n!}{(r-1)!(n-r)!} u^{r-1}(1-u)^{n-r}$$

where $0 < u < 1$. If we standardize $U_{(r)}$: let $Z_{(r)} = \frac{U_{(r)} - \mu}{\sigma}$, then

$$f_{Z_{(r)}} = n\binom{n-1}{r-1}\sigma\mu^{r-1}(1-\mu)^{n-r}e^{v}$$

where $v = (r-1)ln\left(1+\frac{\sigma z}{\mu}\right) + (n-r)ln\left(1-\frac{\sigma z}{1-\mu}\right)$. Because we have $ln\,(1+x) = \sum_{i=1}^{\infty}(-1)^{i-1}\frac{x^i}{i}$, we can let $C_1 = \frac{\sigma}{\mu}$, $C_2 = \frac{\sigma}{1-\mu}$ and then get

$$v = (r-1)\left(C_1 z - C_1^2\frac{z^2}{2} + C_1^3\frac{z^3}{3} - ...\right) - (n-r)\left(C_2 z + C_2^2\frac{z^2}{2} + C_2^3\frac{z^3}{3} - ...\right)$$

We then observe that $C_1 \to \left(\frac{1-p}{pn}\right)^{\frac{1}{2}}$, $C_2 \to \left(\frac{p}{(1-p)n}\right)^{\frac{1}{2}}$ and the coefficients for z, $\frac{z^2}{2}$ and $\frac{z^3}{3}$ go to 0, 1, 0 respectively. So we have $\lim\limits_{n\to\infty} v = -\frac{z^2}{2}$. Substitute back, also by Stirling's formula($k! \approx \sqrt{2\pi}e^{-k}k^{k+\frac{1}{2}}$), we get

$$n\binom{n-1}{r-1}\sigma\mu^{r-1}(1-\mu)^{n-r} \to \frac{1}{\sqrt{2\pi}}$$

and

$$\lim_{n\to\infty} f_{Z_{(r)}} = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$$

$\square$

Here the proof is for $U(0,1)$. For a general distribution, we can immediately have the following transformation and get similar results.

$$\begin{aligned} X_{(r)} &= F_X^{-1}\left(U_{(r)}\right) \\ \mathbb{E}[X_{(r)}] &\to F_X^{-1}(p) \\ \mathrm{Var}[X_{(r)}] &\approx \frac{p(1-p)}{n}[f_X(\mu)]^{-2} \end{aligned}$$