I

Order Statistics

Def: The order statistics of a random sample $X_1, \cdots, X_n$ are the sample values placed in ascending order, denoted by

$$X_{(1)}, \cdots, X_{(n)}$$

Note: Order statistics should be translated as "顺序统计量".

Statistics related to order statistics:

1° Sample range $\quad R = X_{(n)} - X_{(1)}$

2° Sample median
$$M = \begin{cases} X_{(\frac{n+1}{2})}, & n \text{ odd} \\ \frac{1}{2}\left(X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}\right), & n \text{ even} \end{cases}$$

3° Interquartile range $\quad IQR = X_{(n+1-\{\frac{n}{4}\})} - X_{(\{\frac{n}{4}\})}$

where $\{b\}$ is b round to nearest integrer.

Lemma: $\boxed{X_1, \cdots, X_n \text{ is a random sample from a continuous}}$ distribution $f(x)$, ~~pdf $f_X(x)$~~ $(X_1, \cdots, X_n) \sim f(\underline{x})$

then the joint density function of $\boxed{\text{order statistics}}$ $X_{(1)}, \cdots, X_{(n)}$ is

$$f_{X_{(1)}, \cdots, X_{(n)}}(\underline{x}) = \sum_{\pi \in S_n} f(\underline{x}_{\pi^{-1}}) \, I(x_1 < \cdots < x_n)$$

where π is a permutation of subscripts, $I(\cdot)$ is the indicator function.

proof: For continuous distribution,

$$P(X_{(1)} < X_{(2)} < \cdots < X_{(n)}) = 1$$

$\therefore$ We can regard the sample as distinct.

Let $I_1, \cdots, I_n$ be ordered, nonoverlapping intervals,

$\cancel{X_{(1)} \in I_1, \cdots, X_{(n)} \in I_n}$

Then $P(x_{(1)} \in I_1, \cdots, x_{(n)} \in I_n)$

$$= P\Big(\bigcup_{\pi \in S_n} \{ X_{\pi(1)} \in I_1, \cdots, X_{\pi(n)} \in I_n \}\Big)$$

$$= \sum_{\pi \in S_n} P(X_{\pi(1)} \in I_1, \cdots, X_{\pi(n)} \in I_n)$$

$$= \sum_{\pi \in S_n} \int_{I_1} \cdots \int_{I_n} f_\pi(\underline{x})\, d\underline{x}$$

$\cancel{P(X_{\pi(1)} = x_1, \cdots, X_{\pi(n)} = x_n) = P}$

$\because$ $P(X_{\pi(1)} \le x_1, \cdots, X_{\pi(n)} \le x_n) = P(X_1 \le x_{\pi^{-1}(1)}, \cdots, X_n \le x_{\pi^{-1}(n)})$

Define $F_\pi(\underline{x}) = F(x_{\pi^{-1}(1)}, \cdots, x_{\pi^{-1}(n)})$, and

$\underline{X}_\pi = (X_{\pi(1)}, \cdots, X_{\pi(n)})$

then $\underline{X}_\pi \sim F_\pi$

If density exists, then $f_\pi(\underline{x}) = f(\underline{x}_{\pi^{-1}})$

$\therefore$ $LHS = \sum_{\pi \in S_n} \int_{I_1} \cdots \int_{I_n} f(\underline{x}_{\pi^{-1}})\, d\underline{x}$

$$= \int_{I_1} \cdots \int_{I_n} \sum_{\pi \in S_n} f(\underline{x}_{\pi^{-1}})\, d\underline{x}$$

$\therefore$ The joint density function of order statistics is

$\cancel{p(x_{(1)}, \cdots, x_{(n)}) = \sum_{\pi \in S_n} p(\underline{x}_{\pi^{-1}})\, I(x_1 < \cdots}$

$$f_{X_{(1)}, \cdots, X_{(n)}}(\underline{x}) = \sum_{\pi \in S_n} f(\underline{x}_{\pi^{-1}})\, I(x_1 < \cdots < x_n)$$ $\square$.

Note: $\{\underline{x} : x_1 < \cdots < x_n\}$ is a hyper-tetrahedron on which the order statistics are defined.

~~Special cases of pdf of order statistics.~~

$\cancel{1)\ \underline{X} \text{ is exchangeable}: n!\, f(\underline{x})\, I(x_1 < \cdots < x_n)}$

Thm: $\boxed{X_1, \cdots, X_n \text{ Random Sample} \quad n! \prod_{i=1}^n f(x_i)\, I(x_1 < \cdots < x_n)}$

Note: ~~3)~~ $X_1, \cdots, X_n$ ~~iid~~ Random Sample from Uniform$(0,1)$: $n!\, I(0 < x_1 < \cdots < x_n < 1)$

Thm: $X_1, \cdots, X_n$ is a random sample from a continuous distribution with cdf $F(x)$, and pdf $f(x)$. Then the pdf of $X_{(r)}$ is

$$f_{X_{(r)}}(x) = \binom{n}{r-1 \;\; 1 \;\; n-r} F(x)^{r-1} f(x) \left(1-F(x)\right)^{n-r}$$

The cdf of $X_{(r)}$ is

$$F_{X_{(r)}}(x) = \sum_{k=r}^{n} \binom{n}{k} F(x)^{k} \left(1-F(x)\right)^{n-k}$$

proof:
$$
\begin{aligned}
f_{X_{(r)}}(x) &= \int\!\!\cdots\!\!\int_{\mathbb{R}^{n-1}} n! \prod_{i=1}^{r-1} f(x_i) \cdot f(x) \cdot \prod_{j=r+1}^{n} f(x_j) \\
&\qquad \mathbb{1}(x_1 < \cdots < x_{r-1} < x < x_{r+1} < \cdots < x_n)\, dx_1 \cdots dx_{r-1}\, dx_{r+1} \cdots dx_n \\
&= n!\, f(x) \cdot \left[\int\!\!\cdots\!\!\int_{\mathbb{R}^{r-1}} \prod_{i=1}^{r-1} f(x_i)\, \mathbb{1}(x_1 < \cdots < x_{r-1} < x)\, dx_1 \cdots dx_{r-1} \right] \\
&\qquad \cdot \left[\int\!\!\cdots\!\!\int_{\mathbb{R}^{n-r}} \prod_{j=r+1}^{n} f(x_j)\, \mathbb{1}(x < x_{r+1} < \cdots < x_n)\, dx_{r+1} \cdots dx_n \right] \\
&= n!\, f(x) \cdot \left[\int_{-\infty}^{x} dx_{r-1} \int_{-\infty}^{x_{r-1}} dx_{r-2} \cdots \int_{-\infty}^{x_2} dx_1 \cdot \prod_{i=1}^{r-1} f(x_i) \right] \\
&\qquad \cdot \left[\int_{x}^{+\infty} dx_{r+1} \int_{x_{r+1}}^{+\infty} dx_{r+2} \cdots \int_{x_{n-1}}^{+\infty} dx_n \cdot \prod_{j=r+1}^{n} f(x_j) \right] \\
&= n!\, f(x)\, \frac{F(x)^{r-1}}{(r-1)!}\, \frac{[1-F(x)]^{n-r}}{(n-r)!} \\
&= \binom{n}{r-1 \;\; 1 \;\; n-r} F(x)^{r-1} f(x) \left(1-F(x)\right)^{n-r}
\end{aligned}
$$

where $\binom{n}{r-1 \;\; 1 \;\; n-r} \equiv \dfrac{n!}{(r-1)! \cdot 1! \cdot (n-r)!}$

$$
\begin{aligned}
F_{X_{(r)}}(x) &= P(X_{(r)} \le x) \\
&= P\left(\bigcup_{j=r}^{n} \left\{ |\{i : X_i \le x\}| = j \right\} \right) \\
&= \sum_{j=r}^{n} P\left(|\{i : X_i \le x\}| = j \right)
\end{aligned}
$$

$$= \sum_{j=r}^{n} \binom{n}{j} F(x)^j \left(1 - F(x)\right)^{n-j} \qquad \square.$$

Thm: $X_1, \cdots, X_n$ is a random sample from a continuous distribution with cdf $F(x)$ and pdf $f(x)$. Then the joint pdf of $X_{(i)}$ and $X_{(j)}$, $(1 \le i < j \le n)$ is

$$f_{X_{(i)}, X_{(j)}}(u, v) = \binom{n}{i-1,\, 1,\, j-i-1,\, 1,\, n-j} F(u)^{i-1} f(u) \left(F(v) - F(u)\right)^{j-i-1} f(v) \cdot \left(1 - F(v)\right)^{n-j} \mathbb{1}(u < v)$$

Thm: ~~The pdf of uniform order statistic $U_{(r)}$ is~~

Uniform order statistic

$$U_{(r)} \sim B(r, n-r+1) = \frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)} u^{r-1} (1-u)^{n-r}$$

proof:

$$f_{U_{(r)}}(u) = \binom{n}{r-1,\, 1,\, n-r} F_U(u)^{r-1} f_U(u) \left(1 - F_U(u)\right)^{n-r}$$

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

$$= \frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)} u^{r-1} (1-u)^{n-r} \qquad \square.$$