# Notes on order statistics

Rui Wang

Tuesday 20$^{\text{th}}$ November, 2018

## 1 Introduction

There are several books on order statistics. This notes only consists very basic facts.

## 2 Basic properties

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample where $X_i \in \mathbb{R}$. The order statistics is defined as

$$T(\mathbf{X}) = (X_{(1)}, \ldots, X_{(n)}),$$

where $X_{(1)} \leq \cdots \leq X_{(n)}$ are the ordered $X_i$'s. It can be seen that $T(\mathbf{X})$ is a measurable map from $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ to $\mathbb{T}, \mathcal{B}(\mathbb{T})$, where $\mathbb{T}$ is the set of all ordered $n$-tuples.

If $X_1, \ldots, X_n$ are iid random variables with **continuous distribution function**, it can be proved that with probability 1, $X_{(1)} < \cdots < X_{(n)}$.

The following statistics are equivalent:

- order statistics
- Empirical measure
- $U(x) = (\sum x_i, \sum x_i^2, \ldots, \sum x_i^n)$
- $V(x) = (\sum_i x_i, \sum_{i<j} x_i x_j, \sum_{i<j<k} x_i x_j x_k, \ldots, x_1 x_2 \cdots x_n)$

See TSH Example 2.4.1.

A useful property is:

**Proposition 1.** *A statistic $\phi$ is a function of order statistic iff it is a symmetric function (coordinates can be exchanged).*

## 3 completeness

**Definition 1.** A statistic $T$ is said to be **complete** if $E_\theta[f(T)] = 0$ for all $\theta \in \Omega$ implies $f(t) = 0(a.e.\mathcal{P})$.

- if $\mathcal{P}_0$, $\mathcal{P}_1$ are two families of distributions such that $\mathcal{P}_0 \subset \mathcal{P}_1$ and **every null set of** $\mathcal{P}_0$ **is also a null set of** $\mathcal{P}_1$, then a sufficient statistic $T$ that is complete for $\mathcal{P}_0$ is also complete for $\mathcal{P}_1$.
- Let $\mathcal{P}_0$ be the class of binomial distributions $b(p,n)$, $0 < p < 1$, n=fixed, and let $\mathcal{P}_1 = \mathcal{P}_0 \cup \{Q\}$ where $Q$ is the Poisson distribution with expectation 1. Then $\mathcal{P}_0$ is complete but $\mathcal{P}_1$ is not.

See TPE Problem 6.32.

### 3.1 completeness of order statistic

**Theorem 1** (TSH Example 4.3.4)**.** *Let $X_1, \dots, X_N$ be i.i.d. with cdf $F \in \mathcal{F}$, where $\mathcal{F}$ is the family of all* ***absolutely continuous distributions****. Then the set of order statistics $T(X) = (X_{(1)}, \dots, X_{(N)})$ was complete.*

*Proof.* Denote $T(X) = (X_{(1)}, \dots, X_{(N)})$ and $T'(X) = (\sum X_i, \sum X_i^2, \dots, \sum X_i^N)$. They are equivalent. So it's enough to proof the completeness of $T'(X)$. Consider the family of densities $\mathcal{F}_0 \subset \mathcal{F}$ as

$$f(X) = C(\theta_1, \dots, \theta_N) \exp(-x^{2N} + \theta_1 x + \dots + \theta_N x^N)$$

It's well defined for all $\theta$'s. The density of a sample of size $N$ is

$$C^N \exp(-\sum x_i^{2N} + \theta_1 \sum x_i + \dots + \theta_N \sum x_i^N)$$

which constitutes an exponential family. So $T'(X)$ is complete for $\mathcal{F}_0$. Becuase the null set of $\mathcal{F}_0$ is the null set of Lebesgue measure, hence the null set of $\mathcal{F}$. $\square$

**Theorem 2** (TSH Problem 4.13)**.** *The order statistic is complete for the family of all continuous distributions.*

*Proof.* Suppose $\phi$ is an **integrable symmetric function**, that is

$$\int \phi(x_1, \dots, x_n) dF(x_1) \dots dF(x_n) = 0$$

. Replace $F$ by $\alpha_1 F_1 + \dots + \alpha_n F_n$, where $0 \le \alpha_i \le 1$, $\sum \alpha_i = 1$. In fact, it's not necessary to impose $\sum \alpha_i = 1$ (just multiply a constant to the equation). It can be deduced that

$$\int \phi(x_1, \dots, x_n) dF_1(x_1) \dots dF_n(x_n) = 0$$

for all continuous $F_i$. This last equation remains valid if the $F_i$ are replaced by $I_{a_i}(x)F(x) + (1 - I_{a_i}(x))F(a_i)$, where $I_{a_i}(x) = 1$ if $\leq a_i$ and =1 otherwise. This implies that $\phi = 0$ except on a set which has measure 0 under $F \times F \ldots \times F$ for all continuous $F$. □

### 3.2 $c_j$-order statistics

**Definition 2.** The $c_j$-order statistics of a sample of vectors are the vectors arranged in increasing order according to their $j$th components. (See TPE Problem 3.4.17)

The following property still holds:

A statistic $\phi$ is a function of $c_j$-order statistics iff it is a symmetric function (exchangable for observations).

**Theorem 3.** *The $c_j$-order statistic is complete for the family of all continuous distributions.*

The proof is essetially the same as previous theorem once replacing $I_a(x)$ by $I_{a_1,\ldots,a_p}(x_1,\ldots,x_p)$.

**Theorem 4** (adapted from TSH Example 4.3.4)**.** *suppose $X_1,\ldots,X_n$ are i.i.d., $T$ is order statistic. Then for a statistic $\delta$, we have*

$$E[\delta|T] = \frac{1}{n!}\sum \delta(X_{i_1},\ldots,X_{i_n})$$

*, that is, the symmetrization of $\delta$.*

*Proof.* Use the definition of conditional expectation. □

# Appendices

## Appendix A haha1

## Appendix B haha2

## Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant Nos. xxxxx, xxxx.