Induced Order Statistics

Po-Ning Chen, Professor

Institute of Communications Engineering

National Chiao Tung University

Hsin Chu, Taiwan 30010, R.O.C.

Induced order statistics

Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be i.i.d. random vector.

Let $D(1), \ldots, D(n)$ be anti-ranks based on $\{Y_i\}$.

In other words, by sorting the 2-dimensional random vector $\{(X_i, Y_i)\}$ according to ascending $\{Y_i\}$, we obtain:

$$(X_{D(i)}, Y_{(i)}).$$

Then $X_{D(1)}, X_{D(2)}, \ldots, X_{D(n)}$ are named the *induced order statistics* or *concomitants of order statistics*.

Example $\phi_1, \phi_2, \ldots, \phi_n$ are received log-likelihood ratios.
Form a 2-dimensional vector sequence as $(\phi_1, |\phi_1|), (\phi_2, |\phi_2|), \ldots, (\phi_n, |\phi_n|)$.
Then we can sort $\{\phi_i\}$ according to $\{|\phi_i|\}$, when doing the decoding/demodulating process.

Absolute order statistics

Definition $X_{[1]}, X_{[2]}, \ldots, X_{[n]}$ are *absolute order statistics* if $X_1, \ldots, X_n$ are sorted according to their absolute values.

Hence,

$$|X_{[1]}| \leq |X_{[2]}| \leq \cdots \leq |X_{[n]}|.$$

How to determine the distributions of $\{X_{[i]}\}$?

Proposition Any *symmetric* random variable X satisfies that

$$\Pr\left[X \leq x\right] = \Pr\left[G|X| \leq x\right],$$

where $G \perp\!\!\!\perp X$ and $\Pr[G = -1] = \Pr[G = 1] = 1/2$.

Here, *symmetric* means that $\Pr[X \leq -|x|] = 1 - \Pr[X \leq |x|]$.

Absolute order statistics

Proposition (Egorov and Nevzorov 1975) For a sequence of **symmetric** random variables $X_1, \ldots, X_k$, define $Y_k = |X_k|$. Then

$$(X_1, \ldots, X_n) \text{ has the same distribution as } (B_1 Y_1, \ldots, B_n Y_n),$$

where $B_1, \ldots, B_n$ are i.i.d. with equal marginal probability over $\{-1, +1\}$, and is independent of $Y_1, \ldots, Y_n$.

- The above proposition can be generalized to *quasi-symmetric* random variable.

Definition (quasi-symmetric) A random variable X is *quasi-symmetric* if

$$p \Pr[X < -|x|] = (1 - p) \Pr[X > |x|],$$

for some $0 \le p \le 1$.

- $p = 1/2$ reduces *quasi-symmetric* to *symmetric*.
- $p = 1$ reduces *quasi-symmetric* to *nonnegative*.
- $p = 0$ reduces *quasi-symmetric* to *nonpositive*.

Absolute order statistics

Proposition For a sequence of **quasi-symmetric** random variables $X_1, \ldots, X_k$ with parameter p, define $Y_k = |X_k|$. Then

$$(X_1, \ldots, X_n) \text{ has the same distribution as } (B_1 Y_1, \ldots, B_n Y_n),$$

where $B_1, \ldots, B_n$ are i.i.d. with

$$\Pr[B_i = +1] = p \quad \text{and} \quad \Pr[B_i = -1] = 1 - p,$$

and is independent of $Y_1, \ldots, Y_n$.

- In light of the above proposition, the distribution of $X_{[1]}, \ldots, X_{[n]}$ can be derived if the parent distribution is quasi-symmetric.

Absolute order statistics

For $x > 0$,

$$\begin{aligned}
\Pr[X_{[k]} < -x] &= \Pr[B_k Y_{(k)} < -x] \\
&= \Pr[B_k = +1] \Pr[Y_{(k)} < -x] + \Pr[B_k = -1] \Pr[-Y_{(k)} < -x] \\
&= p \Pr\left[Y_{(k)} < -x\right] + (1-p) \Pr\left[-Y_{(k)} < -x\right] \\
&= p \Pr\left[Y_{(k)} < -x\right] + (1-p) \Pr\left[V_{(n-k+1)} < -x\right],
\end{aligned}$$

and

$$\begin{aligned}
\Pr[X_{[k]} < x] &= \Pr[B_k Y_{(k)} < x] \\
&= \Pr[B_k = +1] \Pr[Y_{(k)} < x] + \Pr[B_k = -1] \Pr[-Y_{(k)} < x] \\
&= p \Pr\left[Y_{(k)} < x\right] + (1-p) \Pr[-Y_{(k)} < x] \\
&= p \Pr\left[Y_{(k)} < x\right] + (1-p) \Pr[V_{(n-k+1)} < x],
\end{aligned}$$

where $V_k = -|X_k|$.

For $x \in \Re$,

$$\Pr[X_{[k]} < x] = p \Pr\left[Y_{(k)} < x\right] + (1-p) \Pr[V_{(n-k+1)} < x].$$

Absolute order statistics

Theorem (Egorov and Nevzorov 1976) Let $X_1, \ldots, X_n$ be i.i.d. with marginal cdf $F(\cdot)$. Assume $F(\cdot)$ has inverse function, and has 1st-order-differentiable density $f(\cdot)$ satisfying

$$\sup_{\{x \in \Re \,:\, f(x) > 0\}} |f'(x)| \leq M.$$

Then

$$\sup_{x \in \Re} \left| \Pr\left[\frac{f(q_k)(X_{(k)} - q_k)}{\beta_2} < x \right] - \Phi(x) \right| \leq C \left(\frac{1}{\sqrt{k}} + \frac{1}{\sqrt{n-k+1}} + \frac{M\beta_2}{f^2(q_k)} \right),$$

where C is an absolute constant, $q_k = F^{-1}\left(\frac{k}{n+1}\right)$,

$$\beta_2 = \frac{\sqrt{k(n-k+1)}}{(n+1)\sqrt{n+2}}.$$

For $x \in \Re$,

$$\Pr[X_{[k]} < x] = p \Pr\left[Y_{(k)} < x\right] + (1-p) \Pr[V_{(n-k+1)} < x].$$

We can apply Egorov-Nevzorov theorem to $Y_{(k)}$ and $V_{(n-k+1)}$ to give an estimate of $\Pr[X_{[k]} < x]$.

Absolute order statistics

Let $G(x)$ and $g(x)$ be the cdf and pdf of $|X|$, respectively.

Recall that $F(x)$ and $f(x)$ be the cdf and pdf of X, respectively.

Then for $x > 0$,

$$\begin{aligned}
G(x) &= \Pr[|X| < x] \\
&= \Pr[-x < X < x] \\
&= \Pr[X < x] - \Pr[X < -x] \\
&= \Pr[X < x] - \frac{(1-p)}{p}\Pr[X > x] \quad \left(= \left(1 - \frac{p}{1-p}\Pr[X < -x]\right) - \Pr[X < -x]\right) \\
&= \frac{1}{p}\Pr[X < x] - \frac{(1-p)}{p} \quad \left(= 1 - \frac{1}{1-p}\Pr[X < -x]\right) \\
&= \frac{F(x) - (1-p)}{p}, \quad \left(= 1 - \frac{1}{1-p}F(-x)\right)
\end{aligned}$$

which implies that

$$g(x) = \frac{1}{p}f(x) = \frac{1}{1-p}f(-x).$$

Absolute order statistics

For $x < 0$, the cdf $\bar{G}(x)$ and pdf $\bar{g}(x)$ of $-|X|$ are respectively equal to:

$$\bar{G}(x) = \frac{1 - F(-x)}{p} = \frac{1}{1-p}F(x) \quad \text{and} \quad \bar{g}(x) = \frac{1}{p}f(-x) = \frac{1}{1-p}f(x).$$

Observe that

$$q_k(G) = G^{-1}\left(\frac{k}{n+1}\right) = F^{-1}\left(1 - \frac{n-k+1}{n+1}p\right)$$

and

$$q_{n-k+1}(\bar{G}) = \bar{G}^{-1}\left(\frac{n-k+1}{n+1}\right) = -F^{-1}\left(1 - \frac{n-k+1}{n+1}p\right).$$

For convenience, let $h = F^{-1}\left(1 - \frac{n-k+1}{n+1}p\right)$ and $\bar{h} = -h$.

Absolute order statistics

By rewriting the theorem as ($z = q_k + \beta_2 x / f(q_k)$):

$$\sup_{z\in\Re}\left|\Pr\left[X_{(k)} < z\right] - \Phi\left(\frac{f(q_k)(z-q_k)}{\beta_2}\right)\right| \le C\left(\frac{1}{\sqrt{k}} + \frac{1}{\sqrt{n-k+1}} + \frac{M\beta_2}{f^2(q_k)}\right),$$

and replacing $X_{(k)}$ in the theorem by $Y_{(k)}$ and $V_{(n-k+1)}$, we obtain:

$$\begin{aligned}
&\sup_{z\in\Re}\left|\Pr\left[X_{[k]} < z\right] - p\Phi\left(\frac{g(h)(z-h)}{\beta_2}\right) - (1-p)\Phi\left(\frac{g(h)(z+h)}{\beta_2}\right)\right| \\
= \ &\sup_{z\in\Re}\left|p\Pr[Y_{(k)} < z] - p\Phi\left(\frac{g(h)(z-h)}{\beta_2}\right) + (1-p)\Pr[V_{n-k+1} < z] - (1-p)\Phi\left(\frac{\bar{g}(\bar{h})(z-\bar{h})}{\beta_2}\right)\right| \\
\le \ & pC\left(\frac{1}{\sqrt{k}} + \frac{1}{\sqrt{n-k+1}} + \frac{M\beta_2}{g^2(h)}\right) + (1-p)C\left(\frac{1}{\sqrt{n-k+1}} + \frac{1}{\sqrt{k}} + \frac{\bar{M}\beta_2}{\bar{g}^2(\bar{h})}\right),
\end{aligned}$$

where

$$M = \sup_{\{x\in\Re\,:\,g(x)>0\}} |g'(x)| = \frac{1}{p}\sup_{\{x\ge 0\,:\,f(x)>0\}} |f'(x)|$$

and

$$\bar{M} = \sup_{\{x\in\Re\,:\,\bar{g}(x)>0\}} |\bar{g}'(x)| = \sup_{\{z\le 0\,:\,f(-z)>0\}} \left|-\frac{1}{p}f'(-z)\right| = \frac{1}{p}\sup_{\{x\ge 0\,:\,f(x)>0\}} |f'(x)| = M.$$

Absolute order statistics

Consequently,

$$\sup_{z\in\Re}\left|\Pr\left[X_{[k]}<z\right]-p\Phi\left(\frac{g(h)(z-h)}{\beta_2}\right)-(1-p)\Phi\left(\frac{g(h)(z+h)}{\beta_2}\right)\right|$$
$$\leq\ C\left(\frac{1}{\sqrt{k}}+\frac{1}{\sqrt{n-k+1}}+\frac{M\beta_2}{g^2(h)}\right).$$

The above inequality is useful when taking $k=\alpha n$, where

$$\frac{\beta_2\sqrt{n}}{\sqrt{\alpha(1-\alpha)}}=\frac{n\sqrt{(1-\alpha)n+1}}{(n+1)\sqrt{(1-\alpha)(n+2)}}=1+o(1)\text{ as }n\to\infty$$

and

$$h=h_{p,\alpha}+o(1)=F^{-1}\big(1-(1-\alpha)p\big)+o(1)\text{ as }n\to\infty$$

the inequality becomes:

$$\sup_{z\in\Re}\left|\Pr\left[X_{[\alpha n]}<z\right]-p\Phi\left(\frac{f(h_{p,\alpha})(z-h_{p,\alpha})}{p\sqrt{\alpha(1-\alpha)}}\sqrt{n}\right)-(1-p)\Phi\left(\frac{f(h_{p,\alpha})(z+h_{p,\alpha})}{p\sqrt{\alpha(1-\alpha)}}\sqrt{n}\right)\right|\leq C'\frac{1}{\sqrt{n}}.$$

Absolute order statistics

We conclude that as $n \to \infty$,

$$\Pr\left[X_{[\alpha n]} < z\right] = \begin{cases} 0, & \text{if } z < -h_{p,\alpha}; \\ 1-p, & \text{if } -h_{p,\alpha} \le z < h_{p,\alpha}; \\ 1, & \text{if } z \ge h_{p,\alpha} \end{cases}$$

In other words, $X_{[\alpha n]}$ converges in distribution to a random variable that takes values $-h_{p,\alpha}$ and $h_{p,\alpha}$ with probabilities $(1-p)$ and p.

- The above result was derived for quasi-symmetric parent distributions.
- In 1982, Egorov and Nevzorov further generalized their result to an i.i.d. parent sequence $X_1, \ldots, X_n$ with

$$G(x) = \Pr[|X| \le x]$$

whose inverse function exists.

Absolute order statistics

Theorem Suppose that $k/n \to \alpha$ for $0 < \alpha < 1$ as $n \to \infty$, and let $g_\alpha = G^{-1}(\alpha)$. Then for 1st-order-differentiable parent density $f(\cdot)$,

$$\Pr\left[X_{[k]} < x\right] \stackrel{n\to\infty}{\longrightarrow} \begin{cases} 0, & \text{if } x \le g_\alpha; \\ \dfrac{\lambda_1}{\lambda_1+\lambda_2}, & \text{if } -g_\alpha \le x < g_\alpha \\ 1, & \text{if } x \ge g_\alpha, \end{cases}$$

where $\lambda_1 = f(g_\alpha)$ and $\lambda_2 = f(-g_\alpha)$, provided that $f'(\pm g_\alpha) < \infty$.

Sum approximation

Let $S_{[k]} = \sum_{j=1}^{n-k} a_j X_{[j]}$, where $X_{[j]}$ is the order statistics over a population of size n, whose parent distribution is **symmetric**.

Theorem (Egorov and Nevzorov 1975) Fix a sequence of i.i.d. symmetric variables $X_1, \ldots, X_n, \ldots$. Let $G(x)$ be the cdf of $|X_1|$. Define $t_n = G^{-1}((n - k_n)/n)$, and let $X^{(t_n)} = X I_{[|x| \le t_n]}$. Put

$$s^2_{[k_n]} = \sum_{j=1}^{n-k_n} a_j^2 \mathrm{Var}\left[X^{(t_n)}_{[j;n-k_n]} \right],$$

where $\{X^{(t_n)}_{[j;n-k_n]}\}_{j=1}^{n-k}$ are order statistics for $X_1^{(t_n)}, X_2^{(t_n)}, \ldots, X_{n-k_n}^{(t_n)}$.
Then if

$$0 < \liminf_{n\to\infty} \frac{k_n}{n} \le \limsup_{n\to\infty} \frac{k_n}{n} < 1,$$

$$\sup_{x\in\Re} \left| \Pr\left[\frac{S_{[k_n]}}{s_{[k_n]}} < x \right] - \Phi(x) \right| = O\left(\sqrt{\frac{\log(n)}{n}} \right).$$