Asymptotic normality of numbers of observations near order statistics from stationary processes

Krzysztof Jasiński

Faculty of Mathematics and Computer Science,
Nicolaus Copernicus University, Toruń

December 1, 2016

Notation:

- $\mathbb{X} = (X_n, n \geq 1)$ – sequence of random variables (rv's) with cdf F.
- $X_{1:n} \leq \ldots \leq X_{n:n}$ – order statistics based on $(X_1, \ldots, X_n)$.
- $\text{supp}(F)$ – support of cdf F.

We set
$\gamma_0 := \inf \, supp(F) = \inf\{x \in \mathbb{R} : \, F(x) > 0\}$,

$\gamma_1 := \sup \, supp(F) = \sup\{x \in \mathbb{R} : \, F(x) < 1\}$.

By γ_λ we denote the unique λth quantile of F where $\lambda \in (0, 1)$.

Notation:

- $\mathbb{X} = (X_n, n \geq 1)$ – sequence of random variables (rv's) with cdf F.
- $X_{1:n} \leq \ldots \leq X_{n:n}$ – order statistics based on $(X_1, \ldots, X_n)$.
- $\text{supp}(F)$ – support of cdf F.

We set
$\gamma_0 := \inf \textit{supp}(F) = \inf\{x \in \mathbb{R} : F(x) > 0\}$,

$\gamma_1 := \sup \textit{supp}(F) = \sup\{x \in \mathbb{R} : F(x) < 1\}$.

By γ_λ we denote the unique λth quantile of F where $\lambda \in (0, 1)$.

Notation:

- $\mathbb{X} = (X_n, n \geq 1)$ – sequence of random variables (rv's) with cdf F.
- $X_{1:n} \leq \ldots \leq X_{n:n}$ – order statistics based on $(X_1, \ldots, X_n)$.
- $\text{supp}(F)$ – support of cdf F.

We set
$\gamma_0 := \inf \mathit{supp}(F) = \inf\{x \in \mathbb{R} : F(x) > 0\}$,

$\gamma_1 := \sup \mathit{supp}(F) = \sup\{x \in \mathbb{R} : F(x) < 1\}$.

By γ_λ we denote the unique λth quantile of F where $\lambda \in (0, 1)$.

- For $a > 0$ and $1 \leq k \leq n$ we define the following counting rv's

$$K_{k:n}(-a, 0) := \#\{j \in \{1, \ldots, n\} : X_j \in (X_{k:n} - a, X_{k:n})\}$$

and

$$K_{k:n}(0, a) := \#\{j \in \{1, \ldots, n\} : X_j \in (X_{k:n}, X_{k:n} + a)\}.$$

Interpretation

These two rv's provide information on how many observations fall into the open left and right a-vincity of the order statistic $X_{k:n}$.

Their properties have been studied since 1997 - Pakes and Steutel discussed the behavior of $K_{n:n}(-a, 0)$.

- For $a > 0$ and $1 \leq k \leq n$ we define the following counting rv's

$$K_{k:n}(-a, 0) := \#\{j \in \{1, \ldots, n\} : X_j \in (X_{k:n} - a, X_{k:n})\}$$

and

$$K_{k:n}(0, a) := \#\{j \in \{1, \ldots, n\} : X_j \in (X_{k:n}, X_{k:n} + a)\}.$$

Interpretation

These two rv's provide information on how many observations fall into the open left and right a-vincity of the order statistic $X_{k:n}$.

Their properties have been studied since 1997 - Pakes and Steutel discussed the behavior of $K_{n:n}(-a, 0)$.

- For $a > 0$ and $1 \leq k \leq n$ we define the following counting rv's

$$K_{k:n}(-a, 0) := \#\{j \in \{1, \ldots, n\} : X_j \in (X_{k:n} - a, X_{k:n})\}$$

and

$$K_{k:n}(0, a) := \#\{j \in \{1, \ldots, n\} : X_j \in (X_{k:n}, X_{k:n} + a)\}.$$

Interpretation

These two rv's provide information on how many observations fall into the open left and right a-vincity of the order statistic $X_{k:n}$.

Their properties have been studied since 1997 - Pakes and Steutel discussed the behavior of $K_{n:n}(-a, 0)$.

We consider the extended version of these rv's

$$K_{k_n:n}(A) := \#\{j \in \{1, \ldots, n\} : \quad X_{k_n} - X_j \in A\}, \quad 1 \leq k_n \leq n,$$

where A is a Borel subset of real numbers and $k_n/n \to \lambda \in [0, 1]$.

In the literature three cases are discussed:

1. the case of central order statistics, when $\lambda \in (0, 1)$,
2. the case of extreme order statistics, when k_n or $n - k_n$ is fixed,
3. the case of intermediate order statistics, when $\lambda \in \{0, 1\}$ and both k_n and $n - k_n \to \infty$.

We consider the extended version of these rv's

$$K_{k_n:n}(A) := \#\{j \in \{1, \ldots, n\} : \quad X_{k_n} - X_j \in A\}, \quad 1 \leq k_n \leq n,$$

where A is a Borel subset of real numbers and $k_n/n \to \lambda \in [0, 1]$.

In the literature three cases are discussed:

1. the case of central order statistics, when $\lambda \in (0, 1)$,
2. the case of extreme order statistics, when k_n or $n - k_n$ is fixed,
3. the case of intermediate order statistics, when $\lambda \in \{0, 1\}$ and both k_n and $n - k_n \to \infty$.

Our purpose is to...

- . . . study limiting behavior of suitably centered and normed versions of rv's

$$K_{k_n:n}(A) := \#\{j \in \{1, \ldots, n\} : \quad X_{k_n} - X_j \in A\},$$

under some conditions.

Recent results of this problem

With the assumption that $X_1, X_2, \ldots$ are i.i.d. rv's ...

- ... and imposing some restrictions of the cdf F, Iliopoulos, Dembińska and Balakrishnan (2012) and Dembińska (2012) studied the case of central order statistics.

All these restrictions require that F is **continuous**.

- Dembińska (2014) studied the case of extreme order statistics when F is continuous with left bounded support.

  Moreover, there were obtained analogous result for **discontinuous** F, devoted to three cases: extreme, intermediate and central.

Discontinuity of F here means that the corresponding λth quantiles (including γ_0, γ_1) are not accumulation points of its support.

We extend the discontinuous results to strictly stationary and ergodic observations satisfying an extra condition which guarantees some local independence.

- Dembińska (2014) studied the case of extreme order statistics when F is continuous with left bounded support.

  Moreover, there were obtained analogous result for **discontinuous** F, devoted to three cases: extreme, intermediate and central.

Discontinuity of F here means that the corresponding λth quantiles (including γ_0, γ_1) are not accumulation points of its support.

We extend the discontinuous results to strictly stationary and ergodic observations satisfying an extra condition which guarantees some local independence.

Lemma 1

Let $\mathbb{X} = (X_n, n \geq 1)$ be a strictly stationary and ergodic sequence of rv's with an **arbitrary** cdf F and $1 \leq k_n \leq n$ such that $k_n/n \to \lambda \in [0, 1]$. Then

(a) $X_{k_n:n} \xrightarrow{\text{a.s.}} \gamma_0$ as $n \to \infty$ provided that $\lambda = 0$ and $\gamma_0 > -\infty$,

(b) $X_{k_n:n} \xrightarrow{\text{a.s.}} \gamma_1$ as $n \to \infty$ provided that $\lambda = 1$ and $\gamma_1 < \infty$,

(c) $X_{k_n:n} \xrightarrow{\text{a.s.}} \gamma$ as $n \to \infty$ if $\lambda \in (0, 1)$.

Main result - discontinuous F

Let $\mathbb{X} = (X_n, n \geq 1)$ strictly stationary sequence with **discontinuous** F. Let $\sum_{n=1}^{\infty} \alpha_{\mathbb{X}}(n) < \infty$ and $1 \leq k_n \leq n$, $k_n/n \to \lambda \in [0,1]$. Moreover, assume that

- if $\lambda = 0$, then $\gamma_0 > -\infty$;
- if $\lambda = 1$, then $\gamma_1 < \infty$;
- if $\lambda \in (0,1)$, then there exists a unique λth quantile γ_λ.

Then, for any $A \in \mathcal{B}(\mathbb{R})$, as $n \to \infty$,

$$\sqrt{n}\left(\frac{K_{k_n:n}(A)}{n} - \mathrm{P}(X_1 \in \gamma_\lambda - A)\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma^2_{\mathbb{X}}),$$

provided that γ_λ is not an accumulation point of support of F and

$$\sigma^2_{\mathbb{X}} = p(1-p) + 2\sum_{j=2}^{\infty}\left(\mathrm{P}(X_1 \in \gamma_\lambda - A, X_j \in \gamma_\lambda - A) - p^2\right) \neq 0,$$

where $p = \mathrm{P}(X_1 \in \gamma_\lambda - A)$.

Sketch of proof

- $\sum_{n=1}^{\infty} \alpha_{\mathbb{X}}(n) < \infty$ implies that $\alpha_{\mathbb{X}}(n) \to 0$. So $\mathbb{X}$ is an α-mixing process.

Lemma 2

Under the assumptions of Theorem, the random sequence $\mathbb{Y} = (Y_n, n \geq 1)$, where $Y_n = I(X_n \in \gamma_\lambda - A) - \mathrm{P}(X_1 \in \gamma_\lambda - A)$, $n \geq 1$, satisfies the following conditions:

(a) $\mathrm{E}(Y_1) = 0, \quad \mathrm{P}(|Y_1| \leq C) = 1$ for some constant C,

(b) $\mathbb{Y}$ is an α-mixing strictly stationary process,

(c) $\sum_{n=1}^{\infty} \alpha_{\mathbb{Y}}(n) < \infty$, where $\alpha_{\mathbb{Y}}(n)$, $n \geq 1$, are the α-mixing coefficients of the process $\mathbb{Y}$.

Sketch of proof

- Combining Lemma 2, Theorem 10.3 of Bradley (2007) and Theorem 2 of Rio (1995) yield $\sigma_{\mathbb{X}}^2 > 0$.
- Let

$$U_n := \sqrt{n}\left(\frac{K_{k_n:n}(A)}{n} - \mathrm{P}(X_1 \in \gamma_\lambda - A)\right), \quad W_n := \sum_{i=1}^{n} \frac{Y_i}{\sqrt{n}}.$$

- Using central limit theorem (Theorem 10.3 of Bradley (2007)) we get $W_n \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma_{\mathbb{X}}^2)$.
- $\mathbb{X} = (X_n, n \geq 1)$ is strictly stationary and α-mixing, so is ergodic too.
- By Lemma 1, discontinuity of F implies

$$\mathrm{P}(X_{k_n:n} = \gamma_\lambda \text{ for all large } n\ ) = 1.$$

Sketch of proof

- Combining Lemma 2, Theorem 10.3 of Bradley (2007) and Theorem 2 of Rio (1995) yield $\sigma^2_{\mathbb{X}} > 0$.
- Let

$$U_n := \sqrt{n}\left(\frac{K_{k_n:n}(A)}{n} - \mathrm{P}(X_1 \in \gamma_\lambda - A)\right), \quad W_n := \sum_{i=1}^{n} \frac{Y_i}{\sqrt{n}}.$$

- Using central limit theorem (Theorem 10.3 of Bradley (2007)) we get $W_n \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma^2_{\mathbb{X}})$.
- $\mathbb{X} = (X_n, n \geq 1)$ is strictly stationary and α-mixing, so is ergodic too.
- By Lemma 1, discontinuity of F implies

$$\mathrm{P}(X_{k_n:n} = \gamma_\lambda \text{ for all large } n\ ) = 1.$$

Sketch of proof

- Combining Lemma 2, Theorem 10.3 of Bradley (2007) and Theorem 2 of Rio (1995) yield $\sigma^2_{\mathbb{X}} > 0$.
- Let

$$U_n := \sqrt{n}\left(\frac{K_{k_n:n}(A)}{n} - \mathrm{P}(X_1 \in \gamma_\lambda - A)\right), \quad W_n := \sum_{i=1}^{n} \frac{Y_i}{\sqrt{n}}.$$

- Using central limit theorem (Theorem 10.3 of Bradley (2007)) we get $W_n \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma^2_{\mathbb{X}})$.
- $\mathbb{X} = (X_n, n \geq 1)$ is strictly stationary and α-mixing, so is ergodic too.
- By Lemma 1, discontinuity of F implies

$$\mathrm{P}(X_{k_n:n} = \gamma_\lambda \text{ for all large } n\ ) = 1.$$

Sketch of proof

- Combining Lemma 2, Theorem 10.3 of Bradley (2007) and Theorem 2 of Rio (1995) yield $\sigma_{\mathbb{X}}^2 > 0$.
- Let

$$U_n := \sqrt{n}\left(\frac{K_{k_n:n}(A)}{n} - \mathrm{P}(X_1 \in \gamma_\lambda - A)\right), \quad W_n := \sum_{i=1}^{n} \frac{Y_i}{\sqrt{n}}.$$

- Using central limit theorem (Theorem 10.3 of Bradley (2007)) we get $W_n \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma_{\mathbb{X}}^2)$.
- $\mathbb{X} = (X_n, n \geq 1)$ is strictly stationary and α-mixing, so is ergodic too.
- By Lemma 1, discontinuity of F implies

$$\mathrm{P}(X_{k_n:n} = \gamma_\lambda \text{ for all large } n\ ) = 1.$$

Sketch of proof

- Combining Lemma 2, Theorem 10.3 of Bradley (2007) and Theorem 2 of Rio (1995) yield $\sigma^2_{\mathbb{X}} > 0$.
- Let

$$U_n := \sqrt{n}\left(\frac{K_{k_n:n}(A)}{n} - \mathrm{P}(X_1 \in \gamma_\lambda - A)\right), \quad W_n := \sum_{i=1}^{n} \frac{Y_i}{\sqrt{n}}.$$

- Using central limit theorem (Theorem 10.3 of Bradley (2007)) we get $W_n \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma^2_{\mathbb{X}})$.
- $\mathbb{X} = (X_n, n \geq 1)$ is strictly stationary and α-mixing, so is ergodic too.
- By Lemma 1, discontinuity of F implies

$$\mathrm{P}(X_{k_n:n} = \gamma_\lambda \text{ for all large } n\ ) = 1.$$

Sketch of proof

- Hence

$$U_n - W_n = \frac{\sum_{i=1}^{n} [I(X_i \in X_{k_n:n} - A) - I(X_i \in \gamma_\lambda - A)]}{\sqrt{n}} \xrightarrow{\text{a.s.}} 0.$$

- By Slutsky's lemma,

$$U_n = (U_n - W_n) + W_n \xrightarrow{\text{d}} \mathcal{N}(0, \sigma^2_{\mathbb{X}}),$$

which completes the proof.

Sketch of proof

- Hence

$$U_n - W_n = \frac{\sum_{i=1}^{n} [I(X_i \in X_{k_n:n} - A) - I(X_i \in \gamma_\lambda - A)]}{\sqrt{n}} \xrightarrow{\text{a.s.}} 0.$$

- By Slutsky's lemma,

$$U_n = (U_n - W_n) + W_n \xrightarrow{\text{d}} \mathcal{N}(0, \sigma^2_{\mathbb{X}}),$$

which completes the proof.

Bibliografia

1. Bradley, R.C., (2007). Introduction to Strong Mixing Conditions, Volume 1, Kendrick Press, Heber City (Utah).
2. Dembińska, A., (2012). Asymptotic properties of numbers of observations in random regions determined by central order statistics. *J. Statist. Plann. Inf.* **142**, 516–528.
3. Dembińska, A., (2014). Asymptotic normality of numbers of observations in random regions determined by order statistics. *Statistics: A Journal of Theoretical and Applied Statistics* **48**, 508–523.
4. Iliopoulos, G., Dembińska, A. & Balakrishnan, A., (2012). Asymptotic properties of numbers of observations near sample quantiles. *Statistics* **46**, 85–97.
5. Jasiński K. (2016). Asymptotic normality of numbers of observations near order statistics from stationary processes. *Stat. Probab. Lett.* **119**, 259–263.

Bibliografia

6. Pakes A., Steutel F.W. (1997). On the number of records near the maximum. *Austral. J. Statist.* **39**, 179–193.
7. Rio, E., (1995). The functional law of the iterated logarithm for stationary strongly mixing sequences. *Ann. Probab.* **23**, 1188–1203.