Journal of the Korean
Statistical Society
Vol. 26, No. 3, 1997

Asymptotic Properties of Upper Spacings

Seokhoon Yun[1]

Abstract

It is well known that the spacings, the differences of two successive order statistics, in a random sample of size n from a distribution function F are independent and exponentially distributed if F is itself the exponential distribution. In this paper we obtain an asymptotically similar result on a fixed number of upper spacings as $n \to \infty$ for a general F under the assumption that F is in the domain of attraction of some extreme value distribution. For a heavy or short tailed F, appropriate log transformations of the sample should be proceeded to get the result. As a by-product, we also get that each upper spacing diverges in probability to ∞ and converges in probability to 0 as $n \to \infty$ for a heavy and short tailed F, respectively, which is fully expected.

Key Words : Spacings; Domain of attraction; Extreme value distribution.

[1]Department of Applied Statistics, University of Suwon, Suwon, Kyonggi-do, 445-743, Korea.

1. INTRODUCTION

Let $X_1, ..., X_n$ be a random sample from a distribution function F, and let $X_{1,n} \geq \cdots \geq X_{n,n}$ be their order statistics. The spacings are then defined by

$$S_{i,n} := X_{i,n} - X_{i+1,n}, \; i = 1, ..., n,$$

with the convention that $X_{n+1,n} := 0$.

It is well known that if F is the standard exponential distribution (Exp(1)), the spacings are independent and $S_{i,n}$ is also exponentially distributed with mean $1/i$ (Exp(i)) for each $i = 1, ..., n$ (see Theorem 1.6.1 of Reiss (1989)). This tells us that the maximum can be written as a weighted sum of i.i.d. standard exponential random variables:

$$X_{1,n} = \sum_{i=1}^{n} i^{-1} V_i, \text{ where } V_i = iS_{i,n} \sim \text{Exp}(1). \tag{1.1}$$

This property however does not hold for a general F. In fact, if $F(x) = 1 - e^{-a(x-b)}$, $x \geq b$ ($a > 0$, $b \in \Re$), then the spacings are independent and these distribution functions are the only continuous distribution functions so that the spacings are independent (see Theorem 1.6.3 of Galambos (1987)).

In this paper we show that, for a general F, an asymptotically similar result can be obtained on a fixed number of upper spacings $S_{1,n}, ..., S_{r,n}$, say, as $n \to \infty$ by imposing the domain of attraction assumption on F. Specifically, assume that F is in the domain of attraction of some extreme value distribution G, that is, there exist some constants $a_n > 0$ and $b_n \in \Re$ such that

$$F^n(a_n x + b_n) \to G(x) \text{ as } n \to \infty \tag{1.2}$$

for all x in the support of G. Since this condition is satisfied by most of absolutely continuous distributions F, it may be considered as a general regularity condition rather than a particular strict condition.

Weissman (1978) showed that, under condition (1.2), the random vector

$$(a_n^{-1}(X_{1,n} - b_n), a_n^{-1}(X_{2,n} - b_n), ..., a_n^{-1}(X_{r+1,n} - b_n))$$

converges in distribution to some random vector as $n \to \infty$, and further obtained the explicit form of the density of the joint limiting distribution in the case of $G(x) = \exp(-e^{-x})$, $x \in \Re$, the standard Gumbel distribution. We here derive the explicit form of the density of the joint limiting distribution for a general G using the von Mises representation for the extreme value distributions. This density and the von Mises representation are then particularly helpful for describing various asymptotic properties of the upper spacings including an asymptotically similar result as in (1.1). In particular, we show that each upper spacing diverges in probability to ∞ and converges in probability to 0 as $n \to \infty$ for a heavy and short tailed F, respectively. This information is important to data analysts for judging whether the sample data came from a heavy or short tailed distribution.

2. ASYMPTOTIC PROPERTIES OF UPPER SPACINGS

According to Leadbetter, Lindgren and Rootzén (1983), assumption (1.2) is equivalent to the condition that there exists a $\xi \in \Re$ such that

$$\lim_{u \uparrow x_F} \frac{1 - F(u + g(u)x)}{1 - F(u)} = (1 + \xi x)^{-1/\xi}, \ 1 + \xi x > 0, \tag{2.1}$$

where $x_F := \sup\{x : F(x) < 1\}$, and

$$\begin{array}{ll} x_F = \infty \text{ and } g(u) = \xi u & \text{if } \ \xi > 0; \\ g(u) \text{ is some strictly positive function} & \text{if } \ \xi = 0; \\ x_F < \infty \text{ and } g(u) = -\xi(x_F - u) & \text{if } \ \xi < 0. \end{array}$$

In this case, if we take $b_n = \inf\{x : F(x) \geq 1 - 1/n\}$ and $a_n = g(b_n)$, then we have

$$F^n(a_n x + b_n) \to \Omega_\xi(x) := \exp\{-(1+\xi x)^{-1/\xi}\}, \ 1+\xi x > 0, \ \text{as } n \to \infty. \tag{2.2}$$

The case $\xi = 0$ is interpreted as the limit $\xi \to 0$, i.e., $\Omega_0(x) = \exp(-e^{-x})$, $x \in \Re$. The Ω_ξ defined in (2.2) using the shape parameter ξ is called the von Mises representation for the extreme value distributions. If $\xi > 0$ in (2.2), then the corresponding F is usually referred to as a heavy tailed distribution; if $\xi < 0$, then the corresponding F as a short tailed distribution. It is noteworthy that, when $\xi = 0$, the function g is unique up to asymptotic equivalence, i.e., if there is another $\tilde{g}$ satisfying (2.1), then $\tilde{g}(u) \sim g(u)$ as $u \uparrow x_F$. The G appearing in (1.2) can be generally obtained from Ω_ξ by taking a linear transformation with positive slope of the argument x, i.e., $G(x) = \Omega_\xi(ax + b)$ for some $a > 0$ and $b \in \Re$. In the following, we therefore assume (2.2) instead of (1.2), which is much more efficient to explain the asymptotic behavior of the upper spacings.

The following theorem gives the explicit form of the density of the joint limiting distribution of the $r+1$ largest normalized maxima of a random sample $X_1, ..., X_n$ from F. The issue here is the explicit form of the density rather than the convergence itself.

Theorem 1. Assume that (2.2) holds for some $\xi \in \Re$. Then, for each fixed $r = 1, 2, ...,$

$$\left(\frac{X_{1,n}-b_n}{a_n}, \frac{X_{2,n}-b_n}{a_n}, ..., \frac{X_{r+1,n}-b_n}{a_n}\right) \xrightarrow{d} (Y_1, Y_2, ..., Y_{r+1}) \sim H_M \text{ as } n \to \infty,$$

where H_M is an absolutely continuous, $(r+1)$-dim. distribution function with density

$$h_M(y_1, ..., y_{r+1}) = \frac{\Omega'_\xi(y_1)}{\Omega_\xi(y_1)} \cdots \frac{\Omega'_\xi(y_r)}{\Omega_\xi(y_r)} \Omega'_\xi(y_{r+1}), \; y_1 > \cdots > y_{r+1}, \; 1+\xi y_i > 0.$$

Proof. The proof follows along the similar lines as Leadbetter et. al (1983) who gave the explicit form of the joint limiting distribution function for $r = 1$. Specifically, for $y_1 > \cdots > y_{r+1}$ with $1 + \xi y_i > 0$, let $u_{i,n} = a_n y_i + b_n$ and let $N_{i,n} = \sum_{k=1}^n I(X_k > u_{i,n})$, $i = 1, ..., r+1$, where I denotes the indicator function. The $N_{i,n}$ is the number of exceedances of $u_{i,n}$ by $X_1, ..., X_n$. If we put $p_{i,n} = 1 - F(u_{i,n})$, then (2.2) is equivalent to

$$np_{i,n} = n(1 - F(a_n y_i + b_n)) \to -\log \Omega_\xi(y_i) \text{ as } n \to \infty. \tag{2.3}$$

Therefore, we have

$$\begin{aligned}
&P\left(\frac{X_{1,n}-b_n}{a_n} \le y_1, ..., \frac{X_{r+1,n}-b_n}{a_n} \le y_{r+1}\right) \\
&= P(X_{1,n} \le u_{1,n}, ..., X_{r+1,n} \le u_{r+1,n}) \\
&= P(N_{1,n} = 0, N_{2,n} \le 1, ..., N_{r+1,n} \le r) \\
&= \sum_{k_2=0}^{1} \sum_{k_3=0}^{2-k_2} \cdots \sum_{k_{r+1}=0}^{r-k_2-\cdots-k_r} P(N_{1,n} = 0, N_{2,n} = k_2, ..., N_{r+1,n} = k_2 + \cdots + k_{r+1}) \\
&= \sum_{k_2} \cdots \sum_{k_{r+1}} \binom{n}{k_2} (p_{2,n} - p_{1,n})^{k_2} \cdots \binom{n-k_2-\cdots-k_r}{k_{r+1}} (p_{r+1,n} - p_{r,n})^{k_{r+1}} \\
&\qquad \cdot (1 - p_{r+1,n})^{n-k_2-\cdots-k_{r+1}} \\
&\to \sum_{k_2} \cdots \sum_{k_{r+1}} \frac{[\log \Omega_\xi(y_1) - \log \Omega_\xi(y_2)]^{k_2}}{k_2!} \cdots \frac{[\log \Omega_\xi(y_r) - \log \Omega_\xi(y_{r+1})]^{k_{r+1}}}{k_{r+1}!} \Omega_\xi(y_{r+1}) \\
&\equiv H_M(y_1, ..., y_{r+1}) \text{ as } n \to \infty,
\end{aligned}$$

since, for instance, by (2.3),

$$\begin{aligned}\binom{n}{k_2}(p_{2,n}-p_{1,n})^{k_2} &= \frac{n(n-1)\cdots(n-k_2+1)(p_{2,n}-p_{1,n})^{k_2}}{k_2!}\\ &\to \frac{[\log\Omega_\xi(y_1)-\log\Omega_\xi(y_2)]^{k_2}}{k_2!} \text{ as } n\to\infty.\end{aligned}$$

Finally, one can show that H_M has the density h_M above. □

The following theorem gives the joint limiting distribution of the r upper rescaled spacings in terms of its density.

Theorem 2. Assume that (2.2) holds for some $\xi \in \Re$. Then, for each fixed $r = 1, 2, \ldots,$

$$\left(\frac{S_{1,n}}{a_n},\ldots,\frac{S_{r,n}}{a_n},\frac{X_{r+1,n}-b_n}{a_n}\right) \xrightarrow{d} (Z_1,\ldots,Z_r,Z_{r+1}) \sim H_S \text{ as } n\to\infty,$$

where H_S is absolutely continuous with density

$$\begin{aligned}h_S(z_1,\ldots,z_{r+1}) &= \frac{\Omega'_\xi(z_1+\cdots+z_{r+1})}{\Omega_\xi(z_1+\cdots+z_{r+1})}\cdots\frac{\Omega'_\xi(z_r+z_{r+1})}{\Omega_\xi(z_r+z_{r+1})}\Omega'_\xi(z_{r+1}),\\ &\quad z_1,\ldots,z_r>0,\ 1+\xi z_{r+1}>0.\end{aligned}$$

Proof. The assertion follows from Theorem 1 by taking the transformation

$$\begin{aligned}Z_i &= Y_i - Y_{i+1},\ i=1,\ldots,r,\\ Z_{r+1} &= Y_{r+1}.\ \square\end{aligned}$$

We now state the asymptotic behavior of the upper spacings, which is described differently according to the degree of the tail heaviness of F.

Corollary 1. Assume that (2.2) holds for some $\xi \in \Re$.
If $\xi = 0$, then $a_n^{-1}S_{1,n},\ldots,a_n^{-1}S_{r,n},a_n^{-1}(X_{r+1,n}-b_n)$ are asymptotically independent, each $a_n^{-1}S_{i,n}$ is asymptotically Exp(i)-distributed, and $\exp\{-a_n^{-1}(X_{r+1,n}-b_n)\}$ asymptotically $\Gamma(r+1,1)$-distributed as $n\to\infty$.
If $\xi > 0$, then each $S_{i,n} \xrightarrow{p} \infty$ as $n\to\infty$.
If $\xi < 0$, then each $S_{i,n} \xrightarrow{p} 0$ as $n\to\infty$.

Proof. If $\xi = 0$, then $\Omega_0(x) = \exp(-e^{-x})$ and $\Omega'_0(x) = \Omega_0(x)e^{-x}$, and therefore, from Theorem 2, we have

$$\begin{aligned}h_S(z_1,\ldots,z_{r+1}) &= \left\{\prod_{i=1}^{r} ie^{-iz_i}\right\}\left\{\frac{1}{r!}\exp(-(r+1)z_{r+1}-e^{-z_{r+1}})\right\},\\ &\quad z_1,\ldots,z_r>0,\ z_{r+1}\in\Re. \qquad (2.4)\end{aligned}$$

This density shows everything required.
If $\xi > 0$, then $a_n = g(b_n) = \xi b_n \to \xi x_F = \infty$ as $n \to \infty$ and so the assertion follows.
If $\xi < 0$, then $a_n = g(b_n) = -\xi(x_F - b_n) \to 0$ as $n \to \infty$ and so the assertion follows. □

The following corollary shows that an asymptotically similar result as in (1.1) can be obtained if $\xi = 0$.

Corollary 2. Assume that (2.2) holds with $\xi = 0$. Then, for each fixed $r = 1, 2, \ldots$, the following (i) and (ii) hold.

(i) $a_n^{-1}(X_{1,n} - b_n) \xrightarrow{d} \sum_{i=1}^{r} i^{-1}V_i - \log W_{r+1}$ as $n \to \infty$, where $V_1, \ldots, V_r, W_{r+1}$ are independent random variables such that $V_i \sim \mathrm{Exp}(1)$ and $W_{r+1} \sim \Gamma(r+1, 1)$.

(ii) $a_n^{-1} \sum_{i=1}^{r}(X_{i,n} - X_{r+1,n}) \xrightarrow{d} W_r \sim \Gamma(r, 1)$ as $n \to \infty$.

Proof. (i) From Theorem 2, we have

$$a_n^{-1}(X_{1,n} - b_n) = \sum_{i=1}^{r} a_n^{-1} S_{i,n} + a_n^{-1}(X_{r+1,n} - b_n) \xrightarrow{d} \sum_{i=1}^{r} Z_i + Z_{r+1} \text{ as } n \to \infty.$$

Here, $Z_1, \ldots, Z_r, Z_{r+1}$ are independent with $Z_i \sim \mathrm{Exp}(i)$ and $e^{-Z_{r+1}} \sim \Gamma(r+1, 1)$ by (2.4). Taking $V_i = iZ_i$ and $W_{r+1} = e^{-Z_{r+1}}$ completes the proof.
(ii) From Theorem 2,

$$a_n^{-1} \sum_{i=1}^{r}(X_{i,n} - X_{r+1,n}) = \sum_{i=1}^{r} i a_n^{-1} S_{i,n} \xrightarrow{d} \sum_{i=1}^{r} iZ_i \text{ as } n \to \infty.$$

By (2.4), $Z_1, 2Z_2, \ldots, rZ_r$ are i.i.d. Exp(1) random variables, which completes the proof. □

Remark 1. Let F be the standard exponential distribution. Then (2.2) holds with $\xi = 0$, and moreover $a_n = 1$, $b_n = \log n$, and so comparing (1.1) with Corollary 2 (i), we obtain an interesting result as follows. If $V_1, V_2, \ldots$ are i.i.d. Exp(1) random variables, then, for each fixed $r = 1, 2, \ldots$,

$$\sum_{i=r}^{n} i^{-1} V_i - \log n \xrightarrow{d} -\log W_r \text{ as } n \to \infty,$$

where $W_r \sim \Gamma(r,1)$. This suggests one way of computing the mean and variance of $-\log W_r$:

$$E(-\log W_r) = \lim_{n\to\infty}\left\{\sum_{i=r}^{n} i^{-1}E(V_i) - \log n\right\} = \lim_{n\to\infty}\left\{\sum_{i=1}^{n} i^{-1} - \log n\right\} - \sum_{i=1}^{r-1} i^{-1}$$
$$= \gamma - \sum_{i=1}^{r-1} i^{-1},$$
$$\mathrm{Var}(-\log W_r) = \lim_{n\to\infty}\sum_{i=r}^{n} i^{-2}\mathrm{Var}(V_i) = \sum_{i=1}^{\infty} i^{-2} - \sum_{i=1}^{r-1} i^{-2} = \pi^2/6 - \sum_{i=1}^{r-1} i^{-2},$$

where $\gamma = 0.57722\cdots$ is the Euler's constant. In particular, $-\log W_1$ has the standard Gumbel distribution $\Omega_0(x) = \exp(-e^{-x})$, $x \in \Re$, whose mean and variance are γ and $\pi^2/6$, respectively (For a direct computation, see Johnson and Kotz (1970)).

If $\xi \neq 0$, then $a_n^{-1}S_{1,n}, \dots, a_n^{-1}S_{r,n}$ are neither asymptotically independent nor asymptotically exponentially distributed. In this case, appropriate log transformations of the sample should be proceded to obtain asymptotically similar results as in the first part of Corollary 1. Assuming that (2.2) holds for some $\xi \neq 0$, we define, for $\xi > 0$,

$$S_{i,n}^{(+)} := \log X_{i,n} - \log X_{i+1,n},\ i = 1, \dots, n-1,$$

and, for $\xi < 0$,

$$S_{i,n}^{(-)} := \log(x_F - X_{i,n}) - \log(x_F - X_{i+1,n}),\ i = 1, \dots, n-1.$$

Theorem 3. Assume that (2.2) holds for some $\xi \neq 0$. For each fixed $r = 1, 2, \dots$, let $Z_1, \dots, Z_r, Z_{r+1}$ be independent random variables such that $Z_i \sim \mathrm{Exp}(i)$, $i = 1, \dots, r$, and $e^{-Z_{r+1}} \sim \Gamma(r+1, 1)$.
If $\xi > 0$, then

$$\left(\frac{S_{1,n}^{(+)}}{\xi}, \dots, \frac{S_{r,n}^{(+)}}{\xi}, \frac{\log X_{r+1,n} - \log b_n}{\xi}\right) \xrightarrow{d} (Z_1, \dots, Z_r, Z_{r+1}) \text{ as } n \to \infty.$$

If $\xi < 0$, then

$$\left(\frac{S_{1,n}^{(-)}}{\xi}, \dots, \frac{S_{r,n}^{(-)}}{\xi}, \frac{\log(x_F - X_{r+1,n}) - \log(x_F - b_n)}{\xi}\right) \xrightarrow{d} (Z_1, \dots, Z_r, Z_{r+1}) \text{ as } n \to \infty.$$

Proof. If $\xi > 0$, then $a_n = g(b_n) = \xi b_n$ and so applying the transformation

$$Z_i = \xi^{-1}[\log(1+\xi Y_i) - \log(1+\xi Y_{i+1})],\ i = 1, ..., r,$$
$$Z_{r+1} = \xi^{-1}\log(1+\xi Y_{r+1})$$

to Theorem 1 gives the result. If $\xi < 0$, then $a_n = g(b_n) = -\xi(x_F - b_n)$ and so applying the same transformation as above gives the result. □

The following corollary shows that, when $\xi \neq 0$, asymptotically similar results as in (1.1) can be obtained for the appropriately log transformed maxima. The proof is immediate from Theorem 3 and so omitted.

Corollary 3. Assume that (2.2) holds for some $\xi \neq 0$. For each fixed $r = 1, 2, ...$, let $V_1, ..., V_r, W_{r+1}$ be independent random variables such that $V_i \sim \text{Exp}(1)$ and $W_{r+1} \sim \Gamma(r+1, 1)$.
If $\xi > 0$, then $\xi^{-1}(\log X_{1,n} - \log b_n) \xrightarrow{d} \sum_{i=1}^{r} i^{-1}V_i - \log W_{r+1}$ as $n \to \infty$.
If $\xi < 0$, then $\xi^{-1}[\log(x_F - X_{1,n}) - \log(x_F - b_n)] \xrightarrow{d} \sum_{i=1}^{r} i^{-1}V_i - \log W_{r+1}$ as $n \to \infty$.

Table1. Some examples of distributions

Distribution name	$F(x)$	ξ	$g(u)$
exponential	$1-e^{-x},\ x>0$	0	1
extreme value (Type I)	$\exp(-e^{-x}),\ x \in \Re$	0	1
gamma	$\Gamma(\alpha+1,1)\ (\alpha>0)$	0	1
normal	N(0,1)	0	$1/u$
Rayleigh	$1-\exp(-x^2/2),\ x>0$	0	$1/u$
Weibull	$1-\exp(-x^\alpha),\ x>0\ (\alpha>0)$	0	$\alpha^{-1}u^{1-\alpha}$
lognormal	LN(0,1)	0	$u/(1+\log u)$
logistic	$1/(1+e^{-x}),\ x \in \Re$	0	$1+e^{-u}$
-	$1-e^{1/x},\ x<0$	0	u^2
Pareto	$1-x^{-\alpha},\ x>1\ (\alpha>0)$	$1/\alpha$	u/α
extreme value (Type II)	$\exp(-x^{-\alpha}),\ x>0\ (\alpha>0)$	$1/\alpha$	u/α
t	$t(m)\ (m=1,2,...)$	$1/m$	u/m
F	$F(m,m)\ (m=1,2,...)$	$2/m$	$2u/m$
-	$1-1/[2+x(\log x-1)],\ x>1$	1	u
uniform	$x,\ 0<x<1$	-1	$1-u$
extreme value (Type III)	$\exp[-(-x)^\alpha],\ x<0\ (\alpha>0)$	$-1/\alpha$	$-u/\alpha$
truncated exponential	$(1-e^{-x})/(1-e^{-x_0}),\ 0<x<x_0$	-1	x_0-u
-	$1-(1-x)^\alpha,\ 0<x<1\ (\alpha>0)$	$-1/\alpha$	$(1-u)/\alpha$

Remark 2. If $\xi > 0$, then, since $iS_{i,n}^{(+)}$, $i = 1, ..., r$ are asymptotically i.i.d. with $\mathrm{Exp}(1/\xi)$, an obvious estimator for ξ is given by

$$\hat{\xi}_n = \frac{1}{r}\sum_{i=1}^{r} iS_{i,n}^{(+)} = \frac{1}{r}\sum_{i=1}^{r} \log X_{i,n} - \log X_{r+1,n},$$

which is known as the Hill estimator. For asymptotic properties of $\hat{\xi}_n$, see Davis and Resnick (1984) and Haeusler and Teugels (1985).

In Table 1, we provide some examples of well-known distributions with various values of ξ and $g(u)$.

REFERENCES

(1) Davis, R. A. and Resnick, S. I. (1984). Tail estimates motivated by extreme value theory. *The Annals of Statistics*, **12**, 1467–1487.

(2) Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics* (2nd edn.). Krieger, Florida.

(3) Haeusler, E. and Teugels, J. L. (1985). On asymptotic normality of Hill's estimator for the exponent of regular variation. *The Annals of Statistics*, **13**, 743–756.

(4) Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions-1.* Wiley, New York.

(5) Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer, New York.

(6) Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics.* Springer, New York.

(7) Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, **73**, 812–815.