APPLICATIONES MATHEMATICAE
38,4 (2011), pp. 477–483

Marek Kałuszka and Andrzej Okolewski (Łódź)

A NOTE ON ORDER STATISTICS FROM SYMMETRICALLY DISTRIBUTED SAMPLES

Abstract. We present a first moment distribution-free bound on expected values of L-statistics as well as properties of some numerical characteristics of order statistics, in the case when the observations are possibly dependent symmetrically distributed about the common mean. An actuarial interpretation of the presented bound is indicated.

1. Introduction. Let $X_1, \ldots, X_n$ be real-valued random variables defined on a common probability space $(\Omega, \mathcal{F}, P)$ with finite means μ. Denote by $X_{1:n} \le \cdots \le X_{n:n}$ the order statistics based on the sample $X_1, \ldots, X_n$. Let λ_k be real numbers. The corresponding L-statistic is defined by $\sum_{k=1}^n \lambda_k X_{k:n}$. A review of the developments dealing with L-statistics is presented e.g. in Serfling (1980, Chapter 8). A comprehensive survey of the current knowledge about bounds for expectations of L-statistics has been given by Rychlik (1998, 2001).

In financial context L-statistics accommodate numerous indices of economic inequality as well as risk measures of actuarial science. In particular, they constitute a natural class of estimators for (closely related to each other) spectral and distorted probability measures of risk (see Dowd et al. 2008; Wang, 1996). Some properties of empirical spectral risk measures based on independent observations are discussed e.g. by Acerbi (2002) and Greselin et al. (2009). Since the assumption of mutual independence of risks is often violated in actuarial and financial practice, the study of the impact of dependence among risks has become a major topic in these sciences nowadays (cf. Denuit et al., 2001). For example, Darkiewicz et al. (2005) showed that there is no strict relation between concave distortion

2010 *Mathematics Subject Classification*: Primary 62G30; Secondary 62E99.
Key words and phrases: dependent observations, order statistics, L-statistics, spectral risk measures, Spearman ρ, Kendall τ.

DOI: 10.4064/am38-4-6

risk measures and Pearson's r, Spearman's ρ and Kendall's τ dependency measures.

In this paper we study properties of L-statistics in the case when the underlying observations are symmetrically distributed about μ, i.e.

$$(X_1 - \mu, \dots, X_n - \mu) \stackrel{d}{=} (\mu - X_1, \dots, \mu - X_n), \tag{1}$$

where $\stackrel{d}{=}$ means equality in distribution. Several examples of dependent random variables $Y_1, \dots, Y_n$ which are symmetrically distributed about zero are given below. Of course, the corresponding random variables $Y_1+\mu, \dots, Y_n+\mu$ are symmetrically distributed about μ. Unless otherwise stated we assume that $i = 1, \dots, n$.

Mixing. Let $(Y_1, \dots, Y_n)$ have the distribution function of the form

$$\mathbf{P}(Y_1 \le t_1, \dots, Y_n \le t_n) = \int_{\Theta} \mathbf{P}(Y_1 \le t_1, \dots, Y_n \le t_n \mid \Theta = \theta)\, dG(\theta),$$

where G is the distribution function of Θ and $Y_1, \dots, Y_n$ are conditionally independent given $\Theta = \theta$ with each conditional distribution $Y_i|\Theta = \theta$ being symmetric about zero. For example, one can consider $Y_i = V_i Z_i$, where $\Theta = (V_1, \dots, V_n)$ is an arbitrary random vector and $(Z_1, \dots, Z_n) \stackrel{d}{=} (-Z_1, \dots, -Z_n)$ is a random vector independent of Θ.

Markov dependence. Let $(\varepsilon_i)_{i=1}^n$ be a sequence of random variables independent of a random variable Y_1 symmetrically distributed about zero. Let $Y_i = f_i(Y_{i-1}, \varepsilon_i)$, $i = 2, \dots, n$, where $f_i : \mathbb{R}^2 \to \mathbb{R}$ are Borel functions such that $f_i(-x, y) = -f_i(x, y)$.

Markov dependence of order 2. For $(V_0, V_1, \dots, V_n)$ having the same distribution as $(-V_0, -V_1, \dots, -V_n)$, define $Y_i = g_i(V_{i-1}, V_i)$, where $g_i : \mathbb{R}^2 \to \mathbb{R}$ are such that $g_i(-x, -y) = -g_i(x, y)$. The following examples may be of interest: $g_i(x, y) = -xy$ and $g_i(x, y) = x + y$.

Generalized AR(1). Assume $\varepsilon_1, \dots, \varepsilon_n$ is a sequence of i.i.d. random variables such that $\varepsilon_i \stackrel{d}{=} -\varepsilon_i$ and Y_1 is a random variable symmetrically distributed about zero and independent of $(\varepsilon_i)_{i=1}^n$. Define $Y_i = f_i(Y_{i-1}, \varepsilon_i)$, $i = 2, \dots, n$, with $f_i : \mathbb{R}^2 \to \mathbb{R}$ satisfying $f_i(-x, -y) = -f_i(x, y)$. For $f_i(x, y) = a_i x + y$, $a_i \in \mathbb{R}$, we get the autoregressive process of order one.

Generalized ARCH(1). Suppose Y_1 and $(\varepsilon_i)_{i=1}^n$ are as in the previous example. Set $Y_i = h_i(Y_{i-1})\varepsilon_i$, where $h_i(-x) = h_i(x)$, $x \in \mathbb{R}$, $i = 2, \dots, n$. If $h_i(y) = (a_i y^2 + b_i)^{1/2}$, $a_i, b_i > 0$, we get the autoregressive conditional heteroskedasticity model of order one. The last two models can be extended to ARMA(p,q) and GARCH(p,q), respectively.

In Section 2 we show that expected values of some L-statistics are greater than or equal to the mean μ. The result is an extension of Rychlik's (2009, Proposition 1) bound for a single order statistic from the i.i.d. sample to the case of symmetrically distributed and thus possibly dependent and possibly nonidentically distributed observations. Furthermore, we present some properties of characteristics of order statistics from a sample satisfying (1) such as the skewness coefficient, the Pearson correlation coefficient, the Spearman ρ and the Kendall τ.

2. Results. We assume that the integrals appearing in this section exist and are finite. Moreover, we denote by $\lfloor x \rfloor$ the smallest integer greater than or equal to x.

PROPOSITION 2.1. *Let $X_1, \ldots, X_n$ be symmetrically distributed about μ. If $\sum_{j=1}^{k} (\lambda_{n-j+1} - \lambda_j) \geq 0$ for $1 \leq k \leq \lfloor n/2 \rfloor$, then*

$$\mathbf{E} \sum_{k=1}^{n} \lambda_k X_{k:n} \geq \mu \sum_{k=1}^{n} \lambda_k. \tag{2}$$

Equality occurs in (2) *if for $1 \leq k \leq \lfloor n/2 \rfloor$ either $\mathbf{P}(X_{n-k+1:n} = X_{n-k:n}) = 1$ or $\sum_{j=1}^{k} (\lambda_{n-j+1} - \lambda_j) = 0$.*

Proof. Put $Y_i = X_i - \mu$, $i = 1, \ldots, n$, and observe that

$$\mathbf{P}(Y_{k:n} \leq t) = \mathbf{P}\Big(\sum_{i=1}^{n} \mathbf{I}(Y_i \leq t) \geq k\Big) = \mathbf{P}\Big(\sum_{i=1}^{n} \mathbf{I}(-Y_i \leq t) \geq k\Big), \quad t \in \mathbb{R},$$

where $\mathbf{I}(A) = 1$ if A is true and $\mathbf{I}(A) = 0$ if A is not true. Denote $Z_i = -Y_i$, $i = 1, \ldots, n$. Then $Z_{k:n} = -Y_{n-k+1:n}$ for any $k = 1, \ldots, n$ and

$$\mathbf{P}(Y_{k:n} \leq t) = \mathbf{P}(Z_{k:n} \leq t) = \mathbf{P}(-Y_{n-k+1:n} \leq t)$$

for arbitrary t and k. Hence

$$-\mathbf{E}\, Y_{k:n} = \mathbf{E}\, Y_{n-k+1:n} \tag{3}$$

for every k. Of course, $\mathbf{E}\, Y_{k:n} \leq \mathbf{E}\, Y_{n-k+1:n}$ for $1 \leq k < (n+1)/2$, and consequently

$$\mathbf{E}\, Y_{n-k+1:n} \geq 0 \tag{4}$$

for such k's. If n is an odd number, then

$$-\mathbf{E}\, Y_{(n+1)/2:n} = \mathbf{E}\, Y_{(n+1)/2:n}, \tag{5}$$

which implies that $\mathbf{E}\, Y_{(n+1)/2:n} = 0$. Set $\Lambda = \sum_{k=1}^{n} \lambda_k$ and $\lambda'_k = \lambda_k - (\Lambda - 1)/n$ for $k = 1, \ldots, n$. From (3)–(5) and Abel's identity it follows that

$$\begin{aligned}
\sum_{k=1}^{n} \lambda_k \mathbf{E}\, Y_{k:n} &= \sum_{k=1}^{n} \lambda'_k \mathbf{E}\, Y_{k:n} + \frac{\Lambda - 1}{n} \sum_{k=1}^{n} \mathbf{E}\, Y_k \\
&= \sum_{k=1}^{\lfloor n/2 \rfloor} (\lambda'_{n-k+1} - \lambda'_k) \mathbf{E}\, Y_{n-k+1:n} \\
&= \mathbf{E}\, Y_{n-\lfloor n/2 \rfloor+1:n} \sum_{k=1}^{\lfloor n/2 \rfloor} (\lambda_{n-k+1} - \lambda_k) \\
&\quad + \sum_{k=1}^{\lfloor n/2 \rfloor - 1} \mathbf{E}\,(Y_{n-k+1:n} - Y_{n-k:n}) \sum_{j=1}^{k} (\lambda_{n-j+1} - \lambda_j) \geq 0. \quad \blacksquare
\end{aligned}$$

REMARK 2.2. (i) The bound (2) shows that any empirical spectral and Wang's premium principle based on symmetrically distributed and thus possibly dependent risks has the desired property of nonnegative risk loading (cf. Young, 2004).

(ii) If $X_1, \ldots, X_n$ are arbitrary integrable random variables defined on a common probability space, and $\sum_{j=1}^{n} \lambda_j = 1$ and $\sum_{j=1}^{k} \lambda_j \leq k/n$ for $k = 1, \ldots, n-1$, which is a stronger assumption than that of Proposition 2.1 (cf. Rychlik, 1998, Section 4.1), then by Abel's identity,

$$\begin{aligned}
\sum_{k=1}^{n} \lambda_k X_{k:n} &= X_{n:n} \sum_{k=1}^{n} \lambda_k + \sum_{k=1}^{n-1} (X_{k:n} - X_{k+1:n}) \sum_{j=1}^{k} \lambda_j \\
&\geq X_{n:n} + \sum_{k=1}^{n-1} (X_{k:n} - X_{k+1:n}) \sum_{j=1}^{k} \frac{1}{n} = \frac{1}{n} \sum_{k=1}^{n} X_{k:n} = \frac{1}{n} \sum_{k=1}^{n} X_k,
\end{aligned}$$

and consequently $\mathbf{E} \sum_{k=1}^{n} \lambda_k X_{k:n} \geq (1/n) \sum_{k=1}^{n} \mathbf{E}\, X_k$. Equality is attained if for $1 \leq k \leq n$ either $\mathbf{P}(X_{k:n} = X_{k+1:n}) = 1$ or $\sum_{j=1}^{k} \lambda_j = k/n$. It is worth noting that for the case of identically distributed observations with common symmetric distribution, the result follows from Rychlik's (1998, eq. (53)) bound.

(iii) Under the assumptions of Proposition 2.1 with $\sum_{j=1}^{k} (\lambda_{n-j+1} - \lambda_j) \geq 0$ replaced by $\sum_{j=1}^{k} (\lambda_{n-j+1} - \lambda_j) \leq 0$, the upper bound $\mathbf{E} \sum_{k=1}^{n} \lambda_k X_{k:n} \leq \mu$ is satisfied.

(iv) Proposition 2.1 remains valid with the conditions $\sum_{j=1}^{k} (\lambda_{n-j+1} - \lambda_j) \geq 0$ and (1) replaced by $\sum_{j=1}^{k} (\lambda_{n-j+1} - a\lambda_j) \geq 0$ and $(X_1 - \mu, \ldots, X_n - \mu) \stackrel{d}{=} (a(\mu - X_1), \ldots, a(\mu - X_n))$, where a is a positive real number.

(v) Some relations between numerical characteristics of order statistics from symmetrically distributed observations follow directly from the property $X_{k:n} - \mu \stackrel{d}{=} \mu - X_{n-k+1:n}$. For example, $X_{k:n}$ and $X_{n-k+1:n}$ have the same variance and kurtosis while their skewness coefficients are the opposite numbers.

The next results will provide some relations between characteristics of pairs of order statistics. Define the sample median as $X_{(n+1)/2:n}$ if n is odd and as the arithmetic mean of $X_{n/2:n}$ and $X_{n/2+1:n}$ if n is even.

PROPOSITION 2.3. *Let the assumptions of Proposition* 2.1 *be satisfied. Then for arbitrary* $k, l = 1, \ldots, n$,

$$\mathbf{corr}(X_{k:n}, X_{l:n}) = \mathbf{corr}(X_{n-k+1:n}, X_{n-l+1:n}). \tag{6}$$

Moreover, the sample median and the sample quasi-ranges $X_{n-k+1:n} - X_{k:n}$, $k = 1, 2, \ldots, \lfloor n/2 \rfloor$, *are uncorrelated.*

Proof. Writing $Y_i = X_i - \mu$ and $Z_i = -Y_i$, $i = 1, \ldots, n$, we get

$$\begin{aligned} \mathbf{P}(Y_{k:n} \le t, Y_{l:n} \le s) &= \mathbf{P}\Big(\sum_{i=1}^{n} \mathbf{I}(Y_i \le t) \ge k, \sum_{i=1}^{n} \mathbf{I}(Y_i \le s) \ge l\Big) \\ &= \mathbf{P}\Big(\sum_{i=1}^{n} \mathbf{I}(-Y_i \le t) \ge k, \sum_{i=1}^{n} \mathbf{I}(-Y_i \le s) \ge l\Big) \\ &= \mathbf{P}(Z_{k:n} \le t, Z_{l:n} \le s) \\ &= \mathbf{P}(-Y_{n-k+1:n} \le t, -Y_{n-l+1:n} \le s). \end{aligned} \tag{7}$$

Hence, for any k and l,

$$\begin{aligned} \mathbf{cov}(Y_{k:n}, Y_{l:n}) &= \mathbf{cov}(-Y_{n-k+1:n}, -Y_{n-l+1:n}) \\ &= \mathbf{cov}(Y_{n-k+1:n}, Y_{n-l+1:n}), \end{aligned} \tag{8}$$

which implies (6). If n is odd, then applying (8) with $k = (n+1)/2$ yields

$$\mathbf{cov}(Y_{(n+1)/2:n}, Y_{n-l+1:n} - Y_{l:n}) = 0. \tag{9}$$

If n is even, then putting $k = n/2 + 1$ and $k = n/2$ in (8) gives

$$\mathbf{cov}(Y_{n/2:n}, Y_{n-l+1:n}) = \mathbf{cov}(Y_{(n+2)/2:n}, Y_{l:n}) \tag{10}$$

and

$$\mathbf{cov}(Y_{n/2:n}, Y_{l:n}) = \mathbf{cov}(Y_{(n+2)/2:n}, Y_{n-l+1:n}). \tag{11}$$

Combining (10) with (11) we get

$$\mathbf{cov}\big(\tfrac{1}{2}(Y_{n/2:n} + Y_{(n+2)/2:n}), Y_{n-l+1:n} - Y_{l:n}\big) = 0,$$

which together with (9) leads to the second statement. ∎

Similar relations can also be established for the Kendall τ and the Spearman ρ. Let us recall the definitions of these coefficients. The *Kendall coefficient* of random variables X, Y is defined by

$$\tau(X, Y) = \mathbf{E}\,\mathrm{sgn}((X - X')(Y - Y')),$$

where (X', Y') is an independent copy of (X, Y) and $\operatorname{sgn}(x) = \mathbf{I}(x > 0) - \mathbf{I}(x < 0)$, $x \in \mathbb{R}$. The *Spearman coefficient* of random variables X, Y with distribution functions F, G, respectively, is defined as

$$\rho(X, Y) = \mathbf{cov}(F(X), G(Y)).$$

PROPOSITION 2.4. *Let $k, l = 1, \ldots, n$. Under the assumptions of Proposition* 2.1,

(i) $\tau(X_{k:n}, X_{l:n}) = \tau(X_{n-k+1:n}, X_{n-l+1:n})$,
(ii) *if $(X_1, \ldots, X_n)$ has a continuous distribution function, then*

$$\rho(X_{k:n}, X_{l:n}) = \rho(X_{n-k+1:n}, X_{n-l+1:n}).$$

Proof. From (7) we see that $(X_{k:n} - \mu, X_{l:n} - \mu) \overset{d}{=} (\mu - X_{n-k+1:n}, \mu - X_{n-l+1:n})$. Set $Y_i = X_i - \mu$. By the definition of Kendall's coefficient, $\tau(X_{k:n}, X_{l:n}) = \tau(Y_{k:n}, Y_{l:n})$. Let $(Y_1', \ldots, Y_n')$ be an independent copy of $(Y_1, \ldots, Y_n)$. Since $(Y_{k:n}, Y_{l:n}) \overset{d}{=} (-Y_{n-k+1:n}, -Y_{n-l+1:n})$ and $(Y_{k:n}^{'}, Y_{l:n}^{'}) \overset{d}{=} (-Y_{n-k+1:n}^{'}, -Y_{n-l+1:n}^{'})$, we conclude that

$$(Y_{k:n}, Y_{l:n}, Y_{k:n}^{'}, Y_{l:n}^{'}) \overset{d}{=} (-Y_{n-k+1:n}, -Y_{n-l+1:n}, -Y_{n-k+1:n}^{'}, -Y_{n-l+1:n}^{'}).$$

Therefore

$$\begin{aligned}\tau(Y_{k:n}, Y_{l:n}) &= \mathbf{E}\operatorname{sgn}[(Y_{k:n} - Y_{k:n}^{'})(Y_{l:n} - Y_{l:n}^{'})] \\ &= \mathbf{E}\operatorname{sgn}[(Y_{n-k+1:n} - Y_{n-k+1:n}^{'})(Y_{n-l+1:n} - Y_{n-l+1:n}^{'})] \\ &= \tau(Y_{n-k+1:n}, Y_{n-l+1:n}),\end{aligned}$$

which is equivalent to (i). Denote by F_k the distribution function of $X_{k:n}$. For $k, l = 1, \ldots, n$ we have

$$\begin{aligned}\mathbf{cov}(F_k(X_{k:n} - \mu + \mu), F_l(X_{l:n} - \mu + \mu)) \\ = \mathbf{cov}(F_k(\mu - X_{n-k+1:n} + \mu), F_l(\mu - X_{n-l+1:n} + \mu)).\end{aligned} \tag{12}$$

For $x \in \mathbb{R}$,

$$\begin{aligned}F_k(x) &= \mathbf{P}(X_{k:n} \le x) = \mathbf{P}(Y_{k:n} \le x - \mu) = \mathbf{P}(-Y_{n-k+1:n} \le x - \mu) \\ &= 1 - \mathbf{P}(Y_{n-k+1:n} < \mu - x) = 1 - F_{n-k+1}(2\mu - x),\end{aligned}$$

and so $F_k(2\mu - x) = 1 - F_{n-k+1}(x)$. From (12) we obtain

$$\begin{aligned}\mathbf{cov}(F_k(X_{k:n}), F_l(X_{l:n})) &= \mathbf{cov}(1 - F_{n-k+1}(X_{n-k+1:n}), 1 - F_{n-l+1}(X_{n-l+1:n})) \\ &= \mathbf{cov}(F_{n-k+1}(X_{n-k+1:n}), F_{n-l+1}(X_{n-l+1:n})),\end{aligned}$$

which gives (ii). ∎

Acknowledgements. We wish to thank the referee for showing that using Abel's identity significantly relaxes the assumption of Proposition 2.1 on the coefficients λ_k, as well as for other valuable comments which led to improvements in the paper. The research was supported by the Polish Ministry of Science and Higher Education Grant no. N N111 431337.

References

C. Acerbi (2002), *Spectral measures of risk: A coherent representation of subjective risk aversion*, J. Banking Finance 26, 1505–1518.

G. Darkiewicz, J. Dhaene and M. Goovaerts (2005), *Risk measures and dependencies of risks*, Braz. J. Probab. Statist. 19, 155–178.

M. Denuit, J. Dhaene and C. Ribas (2001), *Does positive dependence between individual risks increase stop-loss premiums?*, Insurance Math. Econom. 28, 305–308.

K. Dowd, J. Cotter and G. Sorwar (2008), *Spectral risk measures: properties and limitations*, J. Financial Services Res. 34, 61–75.

F. Greselin, M. L. Puri and R. Zitikis (2009), *L-functions, processes, and statistics in measuring economic inequality and actuarial risk*, Statist. Interface 2, 227–245.

T. Rychlik (1998), *Bounds for expectations of L-estimates*, in: N. Balakrishnan and C. R. Rao (eds.), Order Statistics: Theory and Methods, Handbook of Statist. 16, North-Holland, Amsterdam, 105–145.

T. Rychlik (2001), *Projecting Statistical Functionals*, Springer, New York.

T. Rychlik (2009), *Tight evaluations for expectations of small order statistics from symmetric and symmetric unimodal populations*, Statist. Probab. Lett. 79, 1488–1493.

R. J. Serfling (1980), *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

S. Wang (1996), *Premium calculation by transforming the layer premium density*, Astin Bull. 16, 71–92.

V. R. Young (2004), *Premium principles*, in: J. L. Teugels and B. Sundt (eds.), Encyclopedia of Actuarial Science, Vol. 3, Wiley, New York.

Marek Kałuszka, Andrzej Okolewski
Institute of Mathematics
Technical University of Łódź
Wólczańska 215
90-005 Łódź, Poland
E-mail: kaluszka@p.lodz.pl
oko@p.lodz.pl

Received on 14.1.2011;
revised version on 13.4.2011 (2073)