

Shannon Information for Concomitants of Generalized Order Statistics in Farlie-Gumbel-Morgenstern (FGM) Family

¹S. TAHMASEBI AND ²J. BEHBOODIAN

¹Department of Statistics, Shiraz University, Shiraz 71454, Iran

²Department of Mathematics, Shiraz Islamic Azad University, Shiraz, Iran

¹stahmasby@yahoo.com, ²Behboodian@stat.susc.ac.ir

Abstract. In the present paper Shannon's entropy for concomitants of generalized order statistics in FGM family is obtained. Application of this result is given for order statistics, record values, k -record values, and progressive type II censored order statistics. Also, we show that the Kullback-Leibler distance among the concomitants of generalized order statistics is distribution-free.

2010 Mathematics Subject Classification: Primary 62B10; Secondary 62G30

Keywords and phrases: Concomitants, generalized order statistics, Kullback-Leibler distance, Shannon's entropy.

1. Introduction

The concept of generalized order statistics was introduced by Kamps [6] as a unified approach to a variety of models of ordered random variables. The random variables $X(1, n, m, k)$, $X(2, n, m, k), \dots, X(n, n, m, k)$ are called generalized order statistics based on the absolutely continuous distribution function F with density function f , if their joint density function is given by

$$f_{1,2,\dots,n}(x_1, x_2, \dots, x_n) = k \left(\prod_{j=1}^{n-1} \gamma_j \right) \left(\prod_{i=1}^{n-1} (1 - F(x_i))^m f(x_i) \right) (1 - F(x_n)) f(x_n),$$

on the cone $F^{-1}(0) < x_1 \leq \dots \leq x_n < F^{-1}(1)$ of \mathbb{R}^n , with parameters $n \in \mathbb{N}$, $k > 0$, $m \in \mathbb{R}$ such that $\gamma_r = k + (n-r)(m+1) > 0$ for all $1 \leq r \leq n$. Let (X_i, Y_i) , $i = 1, 2, \dots, n$ be a random sample of size n from a continuous bivariate distribution. If the pairs are ordered by their X values, then the Y values associated with the r th-order statistic $X_{(r)}$ of X will be denoted by $Y_{[r]}$, $1 \leq r \leq n$, and be called the concomitant of the r th-order statistic. The concomitants are of interest in selection and prediction problems. An excellent review on concomitants of order statistics is given by David and Nagaraja [3]. The FGM family discussed in Johnson

Communicated by Anton Abdulbasah Kamil.

Received: July 28, 2010; Revised: November 22, 2010.

and Kotz [5] provides a flexible family that can be used in such contexts, which is specified by the distribution function(df)

$$(1.1) \quad F_{X,Y}(x,y) = F_X(x)F_Y(y)[1 + \alpha(1 - F_X(x))(1 - F_Y(y))],$$

where $-1 \leq \alpha \leq 1$, and $f_X(x)$, $f_Y(y)$, and $F_X(x)$, $F_Y(y)$ are marginal pdf and cdf of X and Y , respectively. Concomitants can also be defined in the case of generalized order statistics (see Kamps [6], Bairamov and Eryilmaz [1]). For the FGM family with df given by (1.1), the density function of the concomitant of r th-generalized order statistic $Y_{[r,n,m,k]}$, $1 \leq r \leq n$, is given by Beg and Ahsanullah [2], as follows:

$$(1.2) \quad g_{[r,n,m,k]}(y) = f_Y(y) [1 + C^*(r,n,m,k)\alpha(2F_Y(y) - 1)],$$

where

$$C^*(r,n,m,k) = 1 - \frac{2 \prod_{j=1}^r \gamma_j}{(\gamma_1 + 1)(\gamma_2 + 1) \dots (\gamma_r + 1)}.$$

The Shannon entropy for a continuous random variable X with probability density function $f_X(x)$ is defined as

$$(1.3) \quad H(X) = - \int_{-\infty}^{+\infty} f_X(x) \log f_X(x) dx = - \int_0^1 \log f_X(F_X^{-1}(u)) du.$$

The rest of this paper is organized as follows. In Section 2, we derive an analytical expression of entropy for concomitants of generalized order statistics in FGM family. Application of this result is given for order statistics, record values, k -record values, and progressive type II censored order statistics. In Section 3, we show that the Kullback-Leibler distance between $Y_{[r,n,m,k]}$ and $Y_{[s,n,m,k]}$ in this family is distribution-free and is only a function of the sample size n , the indices r, s, m , and the association parameter α .

2. Entropy for concomitants of generalized order statistics

Theorem 2.1. *If $Y_{[r,n,m,k]}$ is the concomitant of r th-generalized order statistics from (1.1), then the Shannon entropy of $Y_{[r,n,m,k]}$ for $1 \leq r \leq n$, $\alpha \neq 0$ is given by*

$$(2.1) \quad H(Y_{[r,n,m,k]}) = W(r, \alpha, n, m, k) + H(Y)(1 - \alpha C^*(r, n, m, k)) - 2\alpha C^*(r, n, m, k)\phi_f(u),$$

where

$$(2.2) \quad W(r, \alpha, n, m, k) = \frac{1}{4\alpha C^*(r, n, m, k)} \left\{ (1 - C^*(r, n, m, k)\alpha)^2 \log(1 - C^*(r, n, m, k)\alpha) - (1 + C^*(r, n, m, k)\alpha)^2 \log(1 + C^*(r, n, m, k)\alpha) \right\} + \frac{1}{2},$$

and $\phi_f(u) = \int_0^1 u \log f_Y(F_Y^{-1}(u)) du$.

Proof. By (1.2) and (1.3), we have

$$\begin{aligned} H(Y_{[r,n,m,k]}) &= -E_{g_{[r,n,m,k]}(y)}[\log f_Y(y)] - E_{g_{[r,n,m,k]}(y)}[\log(1 + C^*(r, n, m, k)\alpha(2F_Y(y) - 1))] \\ &= H(Y)(1 - \alpha C^*(r, n, m, k)\alpha) - 2\alpha C^*(r, n, m, k) \int_0^1 u \log f_Y(F_Y^{-1}(u)) du \end{aligned}$$

$$(2.3) \quad -E_{g_{[r,n,m,k]}(y)} [\log(1 + C^*(r, n, m, k)\alpha(2F_Y(y) - 1))].$$

Now, we need to find $E_{g_{[r,n,m,k]}(y)} [\log(1 + C^*(r, n, m, k)\alpha(2F_Y(y) - 1))]$. First, we derive an expression for $E_{g_{[r,n,m,k]}(y)} [(1 + C^*(r, n, m, k)\alpha(2F_Y(y) - 1))^t]$. So, we have:

$$(2.4) \quad \begin{aligned} u(t) &= E_{g_{[r,n,m,k]}(y)} [(1 + C^*(r, n, m, k)\alpha(2F_Y(y) - 1))^t] \\ &= \int_{-\infty}^{+\infty} f_Y(y) [1 + C^*(r, n, m, k)\alpha(2F_Y(y) - 1)]^{t+1} dy \end{aligned}$$

By change of variable $[(1 + C^*(r, n, m, k)\alpha(2F_Y(y) - 1))] = z$, we obtain

$$(2.5) \quad u(t) = \frac{1}{2\alpha C^*(r, n, m, k)} \left[\frac{(1 + C^*(r, n, m, k)\alpha)^{t+2} - (1 - C^*(r, n, m, k)\alpha)^{t+2}}{t + 2} \right].$$

$$(2.6) \quad \begin{aligned} \frac{-\partial u(t)}{\partial t} \Big|_{t=0} &= -E_{g_{[r,n,m,k]}(y)} [\log(1 + C^*(r, n, m, k)\alpha(2F_Y(y) - 1))] = W(r, \alpha, n, m, k) \\ &= \frac{1}{4\alpha C^*(r, n, m, k)} \left\{ (1 - C^*(r, n, m, k)\alpha)^2 \log(1 - C^*(r, n, m, k)\alpha) \right. \\ &\quad \left. - (1 + C^*(r, n, m, k)\alpha)^2 \log(1 + C^*(r, n, m, k)\alpha) \right\} + \frac{1}{2}. \end{aligned}$$

By substituting (2.6) in (2.3) the result follows. ■

As an application of the representation (2.1) consider the following special cases.

Case 1: If we put $m = 0$ and $k = 1$, then Shannon's entropy for the concomitant of r th-order statistic is given by

$$(2.7) \quad H(Y_{[r]}) = I_{\alpha,n}(r) + H(Y) \left(1 + \left(\frac{n-2r+1}{n+1} \right) \alpha \right) + 2\alpha \left(\frac{n-2r+1}{n+1} \right) \phi_f(u),$$

where $I_{\alpha,n}(r) = W(r, \alpha, n, 0, 1)$. Note that some of the interesting results for $H(Y_{[r]})$ were presented by Tahmasebi and Behboodian [7].

We consider the concomitants of order statistics whenever $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ are independent but otherwise arbitrarily distributed. Now, let us consider the FGM family with df

$$(2.8) \quad F_{X_i, Y_i}(x, y) = F_{X_i}(x)F_{Y_i}(y)[1 + \alpha_i(1 - F_{X_i}(x))(1 - F_{Y_i}(y))], -1 \leq \alpha_i \leq 1,$$

where $F_{X_i}(x) = F_X(x)$ and $F_{Y_i}(y) = F_Y(y)$. Then in the particular cases, the pdf's of $Y_{[1]}$ and $Y_{[n]}$ are given by Eryilmaz [4] as follows:

$$(2.9) \quad f_{[1]}(y) = f_Y(y) \left[1 + \frac{n-1}{(n+1)n} \sum_{j=1}^n \alpha_j(1 - 2F_Y(y)) \right],$$

$$(2.10) \quad f_{[n]}(y) = f_Y(y) \left[1 - \frac{n-1}{(n+1)n} \sum_{j=1}^n \alpha_j(1 - 2F_Y(y)) \right].$$

Now, in the following theorem Shannon's entropy for concomitants of extremes of order statistics is represented.

Theorem 2.2. :Let (X_i, Y_i) , $i = 1, 2, 3, \dots, n$ be independent random vectors from (2.8). If $Y_{[1]}$ and $Y_{[n]}$ are concomitants of extremes of order statistics, then

$$(2.11) \quad H(Y_{[1]}) = Z_{\alpha_j}(n) + H(Y) \left(1 + \frac{n-1}{(n+1)n} \sum_{j=1}^n \alpha_j \right) + 2 \frac{n-1}{(n+1)n} \sum_{j=1}^n \alpha_j \phi_f(u),$$

$$(2.12) \quad H(Y_{[n]}) = Z_{\alpha_j}(n) + H(Y) \left(1 - \frac{n-1}{(n+1)n} \sum_{j=1}^n \alpha_j \right) - 2 \frac{n-1}{(n+1)n} \sum_{j=1}^n \alpha_j \phi_f(u),$$

where

$$(2.13) \quad Z_{\alpha_j}(n) = \frac{n(n+1)}{4 \sum_{j=1}^n \alpha_j (n-1)} \left\{ \left(1 - \left(\frac{n-1}{n(n+1)} \right) \sum_{j=1}^n \alpha_j \right)^2 \log \left(1 - \left(\frac{n-1}{n(n+1)} \right) \sum_{j=1}^n \alpha_j \right) - \left(1 + \left(\frac{n-1}{n(n+1)} \right) \sum_{j=1}^n \alpha_j \right)^2 \log \left(1 + \left(\frac{n-1}{n(n+1)} \right) \sum_{j=1}^n \alpha_j \right) \right\} + \frac{1}{2},$$

and $\phi_f(u)$ is defined in Theorem 2.1.

Proof. The proof is similar to the proof of the Theorem 2.1. ■

Case 2: If we put $m = -1$ and $k = 1$, then Shannon’s entropy for the concomitant of r th-record value is as follows:

$$(2.14) \quad H(R_{[r]}) = C_{\alpha}(r) + H(Y)(1 + \alpha(2^{1-r} - 1)) + 2\alpha(2^{1-r} - 1)\phi_f(u),$$

where $C_{\alpha}(r) = W(r, \alpha, n, -1, 1)$. Also, in this case for $k > 1$, we get the similarly result for k records.

Remark 2.1. From Eq. (2.7) if $r = n = 2^b - 1$, then $H(Y_{[2^b-1]}) = H(R_{[b]})$.

Now, let $X_{1:n}^{\tilde{R}} \leq X_{2:n}^{\tilde{R}} \leq \dots \leq X_{n:n}^{\tilde{R}}$ be the progressive type II censored order statistics which can be viewed as a special case of generalized order statistics, then we denote $Y_{[r:n]}^{\tilde{R}}$ as the r th progressive type II censored concomitant, $1 \leq r \leq n$, based on an absolutely continuous distribution F . The pdf of $Y_{[r:n]}^{\tilde{R}}$ is given by Bairamov and Eryilmaz [1] as follows:

$$(2.15) \quad f_{Y_{[r:n]}^{\tilde{R}}}(y) = c_{r-1} \sum_{i=1}^r \frac{a_{i,r}}{\gamma_i} f_{Y_{[1:\gamma_i]}}(y),$$

where $c_{r-1} = \prod_{j=1}^r \gamma_j$ and $a_{i,r} = \prod_{j=1}^r (\frac{1}{\gamma_j} - \gamma_i)$, $j \neq i$, $1 \leq i \leq r \leq n$, $n \geq 2$, and $f_{Y_{[1:\gamma_i]}}(y)$ is the pdf of the first concomitant of ordinary order statistics from a sample of size γ_i . In the following example, we present a analytical expression of entropy for $Y_{[r:n]}^{\tilde{R}}$ in FGM family with uniform marginals.

Example 2.1. Let (X, Y) be a random variable from (1.2) with joint distribution function

$$F_{X,Y}(x, y) = xy \{ 1 + \alpha(1-x)(1-y) \}, \quad 0 \leq x, y \leq 1, \quad -1 \leq \alpha \leq 1.$$

Then, by using (2.15) the density function of $Y_{[r:n]}^{\bar{R}}$ is

$$f_{Y_{[r:n]}^{\bar{R}}}(y) = c_{r-1} \sum_{i=1}^r \frac{a_{i,r}}{\gamma_i} \left[1 + \alpha(1-2y) \left(\frac{\gamma_i - 1}{\gamma_i + 1} \right) \right]$$

Now, by using (2.1), we get

$$(2.16) \quad H(Y_{[r:n]}^{\bar{R}}) = -\log c_{r-1} + c_{r-1} V_{\gamma_i, a_{i,r}}(\alpha),$$

where

$$(2.17) \quad V_{\gamma_i, a_{i,r}}(\alpha) = \frac{1}{8\alpha \sum_{i=1}^r \frac{a_{i,r}}{\gamma_i} \left(\frac{\gamma_i - 1}{\gamma_i + 1} \right)} \left\{ \left(\sum_{i=1}^r \frac{a_{i,r}}{\gamma_i} \left(1 - \alpha \left(\frac{\gamma_i - 1}{\gamma_i + 1} \right) \right) \right)^2 \times \right. \\ \left[2 \log \left(\sum_{i=1}^r \frac{a_{i,r}}{\gamma_i} \left(1 - \alpha \left(\frac{\gamma_i - 1}{\gamma_i + 1} \right) \right) \right) - 1 \right] \\ - \left(\sum_{i=1}^r \frac{a_{i,r}}{\gamma_i} \left(1 + \alpha \left(\frac{\gamma_i - 1}{\gamma_i + 1} \right) \right) \right)^2 \times \\ \left. \left[2 \log \left(\sum_{i=1}^r \frac{a_{i,r}}{\gamma_i} \left(1 + \alpha \left(\frac{\gamma_i - 1}{\gamma_i + 1} \right) \right) \right) - 1 \right] \right\}.$$

we can easily show that $H(Y_{[r:n]}^{\bar{R}})$ has the following properties:

- (i) $H(Y_{[r:n]}^{\bar{R}}) = -\log c_{r-1} + c_{r-1} V_{\gamma_i, a_{i,r}}(-\alpha) < 0$.
- (ii) $H(Y_{[r:n]}^{\bar{R}})$ is increasing (decreasing) in α for $-1 \leq \alpha < 0$ ($0 < \alpha \leq 1$).

3. Kullback-Leibler distance

The Kullback-Leibler distance for two continuous random variables Z_1 and Z_2 with pdf's f_1 and f_2 , respectively, is given by

$$(3.1) \quad K(Z_1, Z_2) = \int_{-\infty}^{+\infty} f_1(z) \log \left(\frac{f_1(z)}{f_2(z)} \right) dz = E_1 \left(\log \frac{f_1(z)}{f_2(z)} \right),$$

where E_1 denotes the expectation with respect to f_1 . $K(Z_1, Z_2) \geq 0$, where equality holds if and only if $f_1(z) = f_2(z)$ almost everywhere. In the following theorem, we show that the Kullback-Leibler distance between concomitants of r th and s th generalized order statistics in FGM family is distribution-free and is only a function of the sample size n , the indices r, s, m , and the association parameter α .

Theorem 3.1. Let $Y_{[r,n,m,k]}$ and $Y_{[s,n,m,k]}$ be the concomitants of r th- and s th- generalized order statistics in FGM family. Then the Kullback-Leibler distance between $Y_{[r,n,m,k]}$ and $Y_{[s,n,m,k]}$ is

$$(3.2) \quad K(Y_{[r]}, Y_{[k]}) = -W(r, \alpha, n, m, k) + U(r, s, \alpha, n, m, k) + \frac{C^*(r, n, m, k)}{C^*(s, n, m, k)} W(s, \alpha, n, m, k),$$

where

$$U(r, s, \alpha, n, m, k) = \frac{C^*(s, n, m, k) - C^*(r, n, m, k)}{2\alpha(C^*(s, n, m, k))^2} \times$$

$$\begin{aligned} & \{ (1 - C^*(s, n, m, k)\alpha)^2 \log(1 - C^*(s, n, m, k)\alpha) \\ & - (1 + C^*(s, n, m, k)\alpha)^2 \log(1 + C^*(s, n, m, k)\alpha) \} + 2 - \frac{2C^*(r, n, m, k)}{C^*(s, n, m, k)}, \end{aligned}$$

and $W(r, \alpha, n, m, k)$ is defined by (2.2).

Proof. By using (2.1), we have

$$(3.3) \quad H(Y_{[r,n,m,k]}) = W(r, \alpha, n, m, k) - E_{g_{[r,n,m,k]}(y)}[\log f_Y(y)],$$

From (3.1) and (3.3), we get

$$(3.4) \quad K(Y_{[r,n,m,k]}, Y_{[s,n,m,k]}) = -W(r, \alpha, n, m, k) - E_{g_{[r,n,m,k]}}[\log(1 + C^*(s, n, m, k)\alpha(2F_Y(y) - 1))].$$

We see that to determine an expression for the Kullback-Leibler distance between $Y_{[r,n,m,k]}$ and $Y_{[s,n,m,k]}$, we need to find $E_{g_{[r,n,m,k]}}[\log(1 + C^*(s, n, m, k)\alpha(2F_Y(y) - 1))]$. Derivation of this expectation is based on the following strategy: first we write

$$\begin{aligned} G(t) &= E_{g_{[r,n,m,k]}}[(1 + C^*(s, n, m, k)\alpha(2F_Y(y) - 1))^t] \\ &= \int_{-\infty}^{+\infty} f_Y(y) (1 + C^*(r, n, m, k)\alpha(2F_Y(y) - 1)) (1 + C^*(s, n, m, k)\alpha(2F_Y(y) - 1))^t dy \\ &= \frac{C^*(s, n, m, k) - C^*(r, n, m, k)}{2\alpha(C^*(s, n, m, k))^2} \left[\frac{(1 + C^*(s, n, m, k)\alpha)^{t+1} - (1 - C^*(s, n, m, k)\alpha)^{t+1}}{t+1} \right] \\ &\quad + \frac{C^*(r, n, m, k)}{2\alpha(C^*(s, n, m, k))^2} \left[\frac{(1 + C^*(s, n, m, k)\alpha)^{t+2} - (1 - C^*(s, n, m, k)\alpha)^{t+2}}{t+2} \right]. \end{aligned}$$

So, we have

$$\begin{aligned} (3.5) \quad -G'(0) &= -E_{g_{[r,n,m,k]}}[\log(1 + C^*(s, n, m, k)\alpha(2F_Y(y) - 1))] \\ &= \frac{C^*(s, n, m, k) - C^*(r, n, m, k)}{2\alpha(C^*(s, n, m, k))^2} \times \\ &\quad \{ (1 - C^*(s, n, m, k)\alpha)^2 [\log(1 - C^*(s, n, m, k)\alpha) - 1] \\ &\quad - (1 + C^*(s, n, m, k)\alpha)^2 [\log(1 + C^*(s, n, m, k)\alpha) - 1] \} \\ &\quad + \frac{C^*(r, n, m, k)}{8\alpha(C^*(s, n, m, k))^2} \\ &\quad \times \{ (1 - C^*(s, n, m, k)\alpha)^2 [2\log(1 - C^*(s, n, m, k)\alpha) - 1] \\ &\quad - (1 + C^*(s, n, m, k)\alpha)^2 [2\log(1 + C^*(s, n, m, k)\alpha) - 1] \} \\ &= U(r, s, \alpha, n, m, k) + \frac{C^*(r, n, m, k)}{C^*(s, n, m, k)} W(s, \alpha, n, m, k). \end{aligned}$$

If we substitute (3.5) in (3.4) the result follows. ■

Acknowledgement. We are grateful to the editor and referees for their helpful comments and suggestions which improved the presentation of the paper.

References

- [1] I. Bairamov and S. Eryilmaz, Spacings, exceedances and concomitants in progressive type II censoring scheme, *J. Statist. Plann. Inference* **136** (2006), no. 3, 527–536.
- [2] M. I. Beg and M. Ahsanullah, Concomitants of generalized order statistics from Farlie-Gumbel-Morgenstern distributions, *Stat. Methodol.* **5** (2008), no. 1, 1–20.
- [3] H. A. David and H. N. Nagaraja, Concomitants of order statistics, in *Order Statistics: Theory & Methods*, 487–513, 1999, Handbook of Statist., 16 North-Holland, Amsterdam.
- [4] S. Eryilmaz, Concomitants in a sequence of independent nonidentically distributed random vectors, *Comm. Statist. Theory Methods* **34** (2005), no. 9-10, 1925–1933.
- [5] N. L. Johnson and S. Kotz, On some generalized Farlie-Gumbel-Morgenstern distributions, *Comm. Statist.* **4** (1975), 415–427.
- [6] U. Kamps, *A Concept of Generalized Order Statistics*, Teubner Skripten zur Mathematischen Stochastik. Teubner, Stuttgart, 1995.
- [7] S. Tahmasebi and J. Behboodan, Information properties for concomitants of order statistics in Farlie-Gumbel-Morgenstern (FGM) family, *Communication in Statistics -Theory and Methods* **41** (2012), 1954–1968.

