

## On Maximum and Minimum Concomitant Order Statistics with Application

<sup>1</sup>Olosunde, A.A., <sup>2</sup>Alaba, J.G.

<sup>1,2</sup>Department Of Mathematics Obafemi Awolowo University, Ile-Ife. 220005

**Abstract:** This study is aimed at employing the methodology of concomitant order statistics in the analysis of poultry feeds data. The purpose is directed toward finding the distribution of maximum and minimum concomitant order statistics of the cholesterol level given the egg weight. It was estimated on the basis of sample of 96 chickens, which were classified into two groups of 48 chickens each. One group was fed with in-organic copper salt combination while the second group with organic copper salt combination. The estimate of the probability of obtaining both the highest and the lowest are given using numerical approach to solve the integral obtained.

### I. Introduction

Let  $(X_1, Y_2), \dots, (X_n, Y_n)$  be independent and identically distributed random bivariate observations from the distribution  $F(x, y)$ . Also, assume the pairs are ordered by their  $X$  variates and the order statistics of  $X$  are denoted as usual in order statistics by  $X_{i:n}, 1 \leq i \leq n$ . Then, the  $Y$  variate associated with  $X_{i:n}$  is called the concomitant of the  $i$ th order statistics and is usually denoted by  $Y_{[i:n]}$  [David (1973)]. These concomitant order statistics become useful in some problems involving selection. Examples of its importance in Engineering is when the densities of several metal alloy specimens are measured and are selected based on the tensile strength, if these material are subjected to another round of fatigue testing. At each stage the sample is will be censored with respect to the ordered-variate 'time to failure', whilst the associate density is completely known. In another example, if top  $k$  out of  $n$  bulls based on their genetic make-up for breeding, what is average milk yields of their female offspring. In either case,  $X$  represent the initial test which has distribution  $f_{X[i:n]}(X)$ , then the probability density function (pdf) of  $Y_{[i:n]}$  can be written as;

$$f_{Y_{[i:n]}}(y) = \int_{-\infty}^{\infty} f(y|x) f_{i:n}(x) dx \quad (1)$$

where  $f_{i:n}(x)$  is the density function of the  $i$ th order statistics in a sample of  $n$  from the marginal distribution of  $X$ . Similarly, the joint distribution of  $Y_{[i_1:n]}, \dots, Y_{[i_k:n]}, (1 \leq i_1 \dots \leq i_k \leq n)$  can be written as;

$$f_{Y_{[i_1:n]}, \dots, Y_{[i_k:n]}}(y_{i_1}, \dots, y_{i_k}) = \int_{-\infty}^{\infty} \int_{-\infty}^{x_{i_k}} \dots \int_{-\infty}^{x_{i_2}} \dots \prod_{i=1}^k f(y_i | x_i) f_{i_1, \dots, i_k:n}(x) dx_i \quad (2)$$

A substantial number of literature have been written on concomitants of order statistics prominent among them is Bhattacharya (1984) under the name induced order statistics and Nagaraja and David (1994). The most important use of of concomitants arises in selection procedures when  $k (< n)$  samples are chosen on the basis of their  $X$  values. Then the corresponding  $Y$  values represent attributes on an associated characteristics. For example  $X$  might be given such that it represent the score on a screening test and  $Y$  the score on a later test, if the  $k$  best performer a selected from  $X$ . From the data in Olosunde (2013), if the biggest size  $k$  out of  $n$  sample is selected for diet (in our survey, we find that people choice for eggs are primarily based on their sizes), then  $Y_{[n-k+1:n]}, \dots, Y_{[n:n]}$  might represent the average cholesterol level intake; alternatively,  $X$  might be the weight of the eggs and  $Y$  the corresponding cholesterol level which is unknown at time of chosen. Suppose only the highest  $k$  weights are selected for diet, then define  $V_{[k:n]} = \max(Y_{[n-k+1:n]}, \dots, Y_{[n:n]})$ , That is the maximum of the highest cholesterol intake and  $U_{[i:n]} = \min(Y_{[n-i+1:n]}, \dots, Y_{[n:n]})$ , That is the minimum of the lowest cholesterol intake. Then the two statistics is of utmost importance in human health watch. These will serve as the primary aim of this study.

### II. Order Statistics

From the data Olosunde, 2013, we need some preliminary results that we aid in achieving the stated objectives. Given the random variable  $X_{i:n}$  (egg's weight) have been ordered. The probability density function

(pdf) and the cumulative density function (cdf) (expressed in form of negative binomial) of the order statistics (David, 2003) respectively;

$$f_{(n-k,n)}(x) = \frac{n!}{k!(n-k)!} F(x)^{n-k-1} [1-F(x)]^k f(x) \tag{3}$$

and

$$F_{i:n}(x) = P(X_{i:n} \leq x) = \binom{n-1-r}{i-1} \sum_{r=0}^{n-i} [F(x)]^i [1-F(x)]^{n-i-r}, \quad -\infty < x < \infty \tag{4}$$

Where  $f(x)$  is the pdf of the random sample  $X$  having a normal density function.

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x, \mu < \infty, \sigma > 0 \tag{5}$$

Where  $\mu$  and  $\sigma$  are location and scale parameters respectively. The distribution function  $F(x)$  can be expressed as incomplete gamma function

$$F(x) = \frac{\Gamma\left(\frac{1}{2}, \left(\frac{x-\mu}{\sigma}\right)^2\right)}{\sqrt{2}\Gamma\left(\frac{1}{2}\right)} \tag{6}$$

Now from equations 3 and 4 we can derive the pdf and the cdf for the ordered statistics for the random variable  $X$  having normal density function 5 and distribution function 6.

### III. Conditional Distribution of $V_{[k:n]}$ and $U_{[i:n]}$

To develop the distribution of  $V_{[k:n]}$  and  $U_{[i:n]}$ , we need to define some primary distribution necessary for the development. Now without loss of generality let the location parameter  $\mu=0$  and  $\sigma=1$ . Let two random variable  $X$  and  $Y$ , both having a standard normal distribution function have a joint density function  $f(x,y)$  defined by  $f(x|y)/f(y)$ . Define the bivariate normal distribution for random variables  $X$  and  $Y$  to be

$$f(x, y; \rho) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\left\{-\frac{1}{2}[x^2 - 2\rho xy + y^2]\right\}, \quad \rho > 1 \tag{7}$$

(7)

Then the conditional distribution of  $X$  given  $y$  is

$$f(X | y; \rho) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\left\{-\frac{1}{2}[x^2 - 2\rho xy]\right\} \tag{8}$$

(8)

Where in both equations 7 and 8,  $\rho$  is the correlation coefficient. From the information given above then we derived the conditional distribution for  $V_{[k:n]}$  and  $U_{[i:n]}$  using the following propositions.

**Proposition1**

Let  $X_{1:n}, \dots, X_{n:n}$  be order statistics from distribution  $f(x)$  and the  $Y_{1:n}, \dots, Y_{n:n}$  be the associated order statistics. Then the finite-sample c.d.f. of is  $V_{[k:n]}$  given by;

$$P(V_{k:n} \leq y) = \int_{-\infty}^{\infty} \{F(y | x)\}^k f_{X_{n-k:n}}(x) dx$$

**Proof:**

Let

$$P(V_{k:n} \leq y) = F_{k:n}(y) = P(Y_{[n-k+1:n]} \leq y, \dots, Y_{[n:n]} \leq y)$$

$$= \int_{x_0}^{x_k} P(Y_{[n-k+1:n]} \leq y, \dots, Y_{[n:n]} \leq y | X_{[n-k:n]} = x_0, \dots, X_{[n:n]} = x_k) \times f_{X_{[n-k+1:n]}}(x_0) k! \left\{ \prod_{i=1}^k \frac{f_1(x_i)}{1-F_1(x_0)} dx_i \right\} \times f_{X_{n-k:n}}(x_0) dx_0$$

Since the associated statistics are independent. The integrand of the multiple integral in brackets above is a symmetric function of the variables  $x_1, \dots, x_k$ . Thus, it can be expressed as

$$F_{k:n}(y) = \left\{ \int_{x_0} F(y|x) \frac{f(x)}{1-F(x)} dx \right\}^k = \{P(Y \leq y | X > x_0)\}^k = [F(y|x_0)]^k$$

By further simplification, Nagaraja and David (1994). The c.d.f of  $V_{k:n}$  can be expressed as;

$$F_{k:n}(y) = \int_{-\infty}^{\infty} [F(y|x)]^k f_{X_{n-k:n}}(x) dx \tag{9}$$

**Proposition 2**

From the above proposition; we have c.d.f. of  $U_{k:n}$  given as:

$$G_{k:n}(y) = 1 - \int_{-\infty}^{\infty} [1-F(y|x)]^k f_{X_{n-k:n}}(x) dx \tag{10}$$

**Proof:**

From the proof above the result is obvious and follows immediately by noting that  $P(X > x) = 1 - P(X < x)$  and

$$\begin{aligned} F_U(u) &= P(U \leq u) = P[\min(Y_{1:n}, \dots, Y_{n:n}) \leq u] = 1 - P(\min(X_{1:n}, \dots, X_{k:n}, Y_{1:n}, \dots, Y_{k:n}) > u) \\ &= 1 - P(X_{1:n}, \dots, X_{k:n} > u \cap Y_{1:n}, \dots, Y_{k:n} > u) \\ &= 1 - P(X_{i:n} > u)P(Y_{i:n} > u) = 1 - (1 - P(X_{i:n} \leq u))(1 - P(Y_{i:n} \leq u)) \end{aligned}$$

and the result (10) follows.

**IV. Application to Poultry feeds Data**

Next we apply this theory to data in Olosunde, (2013). We want to know the estimated probability for the maximum and the minimum cholesterol level taken by an individual. Using Kolmogorov test statistics, the data is approximately normally distributed with parameters shown in table 4.1. and more so we are considering the bivariate normal distribution data with correlation coefficient 0.87.

**Table 4.1: Data Summary**

				Skewness		Kurtosis	
				STATISTICALS	STD. ERROR	STATISTICS	STD. ERROR
COPPER SALT		Mean	Std. Deviation				
INORGANIC COPPER SALT	WEIGHT	58.3517	3.59707	.108	.343	-1.009	.674
	CHOLESTEROL	195.73	22.13915	-.087	.343	-1.186	.674
ORGANIC COPPER SALT	WEIGHT	59.0998	1.84122	.016	.343	-1.178	.674
	CHOLESTEROL	131.46	37.62636	-.099	.343	-1.129	.674
Total	WEIGHT	58.7257	2.86704	-.124	.246	-.366	.488
	CHOLESTEROL	163.59	44.56972	-.543	.246	-.613	.488

The data summary are provided in table 4.1 above, where the kurtosis and skewness are statistics that characterize the shape and symmetry of the distribution. These statistics are displayed with their standard errors. A measure of the extent to which observations cluster around a central point. For a normal distribution, the value of the kurtosis statistic is zero. Positive kurtosis indicates that the observations cluster more and have longer tails than those in the normal distribution, and negative kurtosis indicates that the observations cluster less and have shorter tails.

In the above we have the following values: -1.009, -1.186, -1.178, -1.129, -.366, -.613 which shows that the weight, cholesterol for the two set of feeds cluster less and have shorter tails meanwhile for the two groups joint together, the multivariate statistics approximates better to multivariate normal (this has been by the help of central limit theorem i.e. larger number of observations). A measure of the asymmetry of a distribution. The normal distribution is symmetric and has a skewness value of 0. A distribution with a significant positive skewness has a long right tail. A distribution with a significant negative skewness has a long left tail. A skewness value more than twice its standard error is taken to indicate a departure from symmetry.

Example (Olosunde, 2013): The eggs selected for consumption has bivariate normal distribution function with parameters as shown in table 4.1. with correlation coefficient,  $\rho=0.87$ . Suppose out of 9 eggs selected, the first k with highest weight were selected, now for cholesterol level test. What is the smallest value

of  $k$  ensuring with probability at least 0.9 that the highest of the  $n$  eggs, as judged by the level of cholesterol is included among the  $k$  selected?.

To solve this problem we need to estimate the lowest  $k$  by using the numerical results when  $X$  and  $Y$  are bivariate normal using numerical computational approach (Simpson rule) in R software by Ihaka and Gentleman to solve (9) and (10).

This yields  $0.56306+0.2207+0.1152+0.0640=0.9304$ .

The required value of  $k$  is 4 in order not to exceed 0.9 maximum.

### References

- [1]. Balakrishnan, N and Kim, J.A. (2005). Point and Interval Estimation for bivariate normal distribution based on progressively Type-II censored data. In *Communication In Statistics Theory and Methods*, Special issue in memory of Milton Sobel.
- [2]. Bhattacharya, P.K. (1974). Convergence of sample paths of normalized sums of induced order statistics. *Annals of Statistics*, 2, 1034-1039.
- [3]. David, H.A. (1973): Concomitants of Order Statistics, *Bulletin of the International Statistical Institute*, 45, 295-300.
- [4]. David, H.A. and Nagarajah, H.N. (1998): Concomitants of Order Statistics, In *Handbook of Statistics 16: order statistics theory and methods*. pp.487-513, North-Holland Academic Publishers, Amsterdam.
- [5]. Ihaka, R., and Gentleman, R. (1996) R. A Language for Data Analysis and Graphics. *Journal Computational and Graphical Statistics* 5, 299-314.
- [6]. Olosunde, A.A. (2013): On Exponential Power Distribution with Poultry feeds data: A case Study. *Journal of Iranian Statistical Society*. to appear in Vol 12, pp. 253-269.
- [7].