SOME RESULTS ON THE ORDER STATISTICS OF THE MULTIVARIATE NORMAL AND PARETO TYPE 1 POPULATIONS

By K. V. Mardia

University of Rajasthan, Jaipur

0. Preliminaries. Let X_i and R_i be the minimum observation and the range of the ith variate, $i = 1, \cdots, k$, in a random sample of size n from a k-variate continuous population having the probability density function (p.d.f.) $f(x_1, \cdots, x_k)$. We require the following results which can be easily derived:

1^0. The probability function (p.f.) for variables to exceed $X_1, \cdots, X_k$, is

$$\int_{X_1}^{\infty} \cdots \int_{X_k}^{\infty} h(u_1, \cdots, u_k)\, du_1 \cdots du_k = \left\{\int_{X_1}^{\infty} \cdots \int_{X_k}^{\infty} f(x_1, \cdots, x_k)\, dx_1 \cdots dx_k\right\}^n,$$

where $h(X_1, \cdots, X_k)$ is the p.d.f. of $(X_1, \cdots, X_k)$.

2^0. The p.d.f. of $(R_1, \cdots, R_k)$, for $n = 2$, is

$$\sum \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1 + a_1, \cdots, x_k + a_k) f(x_1 + b_1, \cdots, x_k + b_k)\, dx_1 \cdots dx_k,$$

where a_i, b_i take the values R_i or 0 such that $a_i + b_i = R_i$, $i = 1, \cdots, k$, and $\sum$ denotes the summation over all such possible combinations of a_i, b_i, $i = 1, \cdots, k$.

3^0. The p.d.f. of $(R_1, \cdots, R_k)$, for $n = 3$, is

$$\int_0^{R_1} \cdots \int_0^{R_k} \sum{}' \left[\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left\{\prod_{j=1}^{3} f(x_1 + a_{1j}, \cdots, x_k + a_{kj})\right\} \cdot dx_1 \cdots dx_k\right] dS_1 \cdots dS_k,$$

where a_{ij}, $j = 1, 2, 3$, take the values R_i or S_i or 0 such that $a_{i1} + a_{i2} + a_{i3} = R_i + S_i$, $i = 1, \cdots, k$, and $\sum'$ denotes the summation over all such possible combinations of a_{ij}, $i = 1, \cdots, k; j = 1, 2, 3$.

1. Introduction and summary. The p.d.f. of the minimum observation X_1 in a random sample of size n from a univariate Pareto population with the p.d.f.

$$\begin{aligned} f(x_1; a, p) &= pa^p / x_1^{p+1}, & x_1 > a > 0, \\ &= 0, & x_1 \leqq a, p > 0, \end{aligned}$$

is again $f(X_1; a, np)$. It has been shown here that the distribution of $(X_1, \cdots, X_k)$ from a k-variate Pareto type 1 population (Mardia [7]) with the p.d.f.

Received 30 April and 8 May 1962; revised 12 February 1964.

$$f_1(x_1, \cdots, x_k; p) = (p+k-1)^{(k)} \Big/ \left(\prod_{i=1}^{k} a_i\right)\left\{\left(\sum_{i=1}^{k} a_i^{-1}x_i\right) - k + 1\right\}^{p+k} \tag{1.1}$$
$$x_i > a_i > 0, i = 1, \cdots, k; p > 0,$$

is again Pareto type 1 with the p.d.f. $f_1(X_1, \cdots, X_k; np)$ This remarkable property of preserving the form of multivariate population in the joint distribution of statistics is well known for the distribution of sample means from a multivariate normal population.

The distributions of ranges for the random samples of size 2 and 3, drawn from univariate normal population has been derived by Mackay [4] and Mackay and Pearson [5]. These have been given here for the random samples drawn from the multivariate normal and Pareto type 1 populations.

2. Distribution of minimums from Pareto type 1 population. Suppose $(x_1, \cdots, x_k)$ has the p.d.f. given by (1.1). We have

$$\int_{x_1}^{\infty} \cdots \int_{x_k}^{\infty} f_1(z_1, \cdots, z_k; p)\, dz_1 \cdots dz_k = \left\{\left(\sum_{i=1}^{k} a_i^{-1}x_i\right) - k + 1\right\}^{-p}. \tag{2.1}$$

On using (2.1) in the result 1^0, we get the p.f. for variables to exceed $X_1, \cdots, X_k$, as

$$\int_{X_1}^{\infty} \cdots \int_{X_k}^{\infty} h_1(u_1, \cdots, u_k)\, du_1 \cdots du_k = \left\{\left(\sum_{i=1}^{k} a_i^{-1}X_i\right) - k + 1\right\}^{-np}, \tag{2.2}$$

where $h_1(X_1, \cdots, X_k)$ is the p.d.f. of $(X_1, \cdots, X_k)$ in this case. On comparing the right hand sides of (2.1) and (2.2), we find that

$$\begin{aligned} \int_{X_1}^{\infty} \cdots \int_{X_k}^{\infty} & h_1(u_1, \cdots, u_k)\, du_1 \cdots du_k \\ &= \int_{X_1}^{\infty} \cdots \int_{X_k}^{\infty} f_1(z_1, \cdots, z_k; np)\, dz_1 \cdots dz_k. \end{aligned} \tag{2.3}$$

So that the p.f. for variables to exceed $X_1, \cdots, X_k$, is again Pareto type 1 with the index parameter np instead of p and thus we have

THEOREM 2.1. *If the parent population is Pareto type 1 with the* p.d.f. $f_1(x_1, \cdots, x_k; p)$ *then the* p.d.f. *of* $(X_1, \cdots, X_k)$ *is* $f_1(X_1, \cdots, X_k; np)$.

3. Distribution of ranges from the normal population. Let $\phi(x_1, \cdots, x_k)$ be the p.d.f. of a multivariate normal population with the mean vector $\mu = (\mu_1, \cdots, \mu_k)'$ and covariance matrix $\Sigma = (\sigma_{ij})$. We can prove

LEMMA 3.1.

$$\begin{aligned} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} & \left\{\prod_{j=1}^{m} \phi(x_1 + a_{1j}, \cdots, x_k + a_{kj})\right\} dx_1 \cdots dx_k \\ &= \{(2\pi)^{k(m-1)} m^k |\Sigma|^{(m-1)}\}^{-\frac{1}{2}} \left[\exp\left\{-\frac{1}{2}\sum_{j=1}^{m} (\mathbf{a}_j - \bar{\mathbf{a}})'\Sigma^{-1}(\mathbf{a}_j - \bar{\mathbf{a}})\right\}\right], \end{aligned} \tag{3.1}$$

where $\mathbf{a}_j' = (a_{1j}, \cdots, a_{kj}), j = 1, \cdots, m; \bar{\mathbf{a}} = m^{-1}\sum_{j=1}^{m} \mathbf{a}_j$.

THEOREM 3.1. *The* p.d.f. *of the ranges* $R_1, \cdots, R_k$, *for* $n = 2$, *in the case of multivariate normal population, is the* p.d.f. *derived on transforming* $U_1, \cdots, U_k$ *to* $R_1, \cdots, R_k$, *by the transformation* $R_1 = |U_1|, \cdots, R_k = |U_k|$, *where* $U_1, \cdots, U_k$, *are* $N(0, 2\Sigma)$ *and* 0 *stands for the column vector of* k *zeros.*

PROOF. On using Lemma 3.1, for $m = 2$, in the result 2^0, we get the p.d.f. of $R_1, \cdots, R_k$, as

$$\{(4\pi)^k|\Sigma|\}^{-\frac{1}{2}} \sum{}^* \exp\left\{-\frac{1}{4}\sum_{i,j}^{k} c_i c_j R_i R_j \sigma^{ij}\right\}, \tag{3.2}$$

where $\Sigma^{-1} = (\sigma^{ij})$, c_i takes values $+1$ or -1, and $\sum^*$ denotes the summation over all possible combinations of c_i, $i = 1, \cdots, k$.

On utilizing Lemma 3.1 for $m = 3$, in the result 3^0, we obtain

THEOREM 3.2. *If the parent population is* $N(\mu, \Sigma)$, *the* p.d.f. *of the ranges* $R_1, \cdots, R_k$, *for* $n = 3$, *in the notation of the result* 3^0, *is*

$$\begin{aligned}&\{3^{\frac{1}{2}k}(2\pi)^k|\Sigma|\}^{-1}\int_0^{R_1}\cdots\int_0^{R_k}\\ &\qquad\cdot\sum{}'\left\{\exp -\frac{1}{2}\sum_{j=1}^{3}(\mathbf{a}_j - \bar{\mathbf{a}})'\Sigma^{-1}(\mathbf{a}_j - \bar{\mathbf{a}})\right\} dS_1 \cdots dS_k,\end{aligned} \tag{3.3}$$

where $\mathbf{a}_j' = (a_{1j}, \cdots, a_{kj})$; $j = 1, 2, 3$; $\bar{\mathbf{a}} = \frac{1}{3}(\mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_3)$.

If we put $k = 1$ in the above theorems, we obtain the results discovered by Mackay [4] and Mackay and Pearson [5].

4. Distribution of ranges from Pareto type 1 population.

LEMMA 4.1. *If* $V(x)$ *is an algebraic function of* x *then*

$$\begin{aligned}&\int_{a_1}^{\infty}\cdots\int_{a_k}^{\infty} V\left\{\left(\sum_{i=1}^{k} a_i^{-1}x_i\right) - k + 1\right\} dx_1 \cdots dx_k\\ &\qquad = \left[\prod_{i=1}^{k} a_i \Big/ (k-1)!\right]\int_1^{\infty}(x-1)^{k-1}V(x)\,dx.\end{aligned} \tag{4.1}$$

PROOF. On using the transformations $u_i = a_i^{-1}x_i - 1$, $i = 1, \cdots, k$, in the integral in the left hand side of (4.1), the integral becomes

$$\left(\prod_{i=1}^{k} a_i\right)\int_0^{\infty}\cdots\int_0^{\infty} V\left\{\left(\sum_{i=1}^{k} u_i\right) + 1\right\} du_1 \cdots du_k.$$

Now, the application of Liouville's integral (see, Edward [2], 160–161), gives the desired result.

We obtain the following theorem on utilizing the above lemma in the results 2^0 and 3^0.

THEOREM 4.1. *If the parent population is Pareto type 1 with the* p.d.f. $f_1(x_1, \cdots, x_k; p)$ *then the* p.d.f. *of the ranges* $R_1, \cdots, R_k$, *for* $n = 2$, *is*

$$\frac{\{(p+k-1)^{(k)}\}^2}{(k-1)!\left(\prod_{i=1}^{k} a_i\right)} \sum{}^{**} \int_1^\infty \cdot \frac{(u-1)^{k-1}\,du}{\left[\left\{u+\sum_{i=1}^{k}(d_i R_i/a_i)\right\}\left\{u+\sum_{i=1}^{k}(1-d_i)R_i/a_i\right\}\right]^{p+k}}, \tag{4.2}$$

where d_i takes values 0 *or* 1, *and* $\sum^{**}$ *denotes the summation taken over all possible combinations of* d_i, $i = 1, \cdots, k$; *and the* p.d.f. *of* $(R_1, \cdots, R_k)$ *for* $n = 3$, *in the notation of the result* 3^0, *is*

$$\frac{\{(p+k-1)^{(k)}\}^3}{(k-1)!\left(\prod_{i=1}^{k} a_i\right)^2} \int_0^{R_1} \cdots \int_0^{R_k} \cdot \left\{\sum{}' \int_1^\infty \frac{(u-1)^{k-1}\,du}{\prod_{j=1}^{3}\left[u+\sum_{i=1}^{k}(a_i a_{ij})\right]^{p+k}}\right\} dS_1 \cdots dS_k. \tag{4.3}$$

When $k = 2$ and p is an integer, the integral appearing in (4.1) can be evaluated by the partial fraction method. Partial fractions of such integrands are dealt with in Mardia [6].

5. Acknowledgments. The author is very grateful to Professor A. P. Dempster for many valuable suggestions. The author also wishes to thank the referees for their helpful remarks, and Dr. G. C. Patni for his encouragement.

REFERENCES

[1] Cramér, Harald (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[2] Edward, J. (1922). *A Treatise on Integral Calculus*, **2**. Macmillan, London.

[3] Gumbel, E. J. (1958). *Statistics of Extremes*. Columbia Univ. Press.

[4] Mackay, A. T. (1935). The distribution of the difference between the extreme observation and the sample mean in sample of n from a normal universe. *Biometrika* **27** 466–471.

[5] Mackay, A. T. and Pearson, E. S. (1933). A note on the distribution of range in samples of n. *Biometrika* **25** 415–420.

[6] Mardia, K. V. (1961). An important integral and application in partial fraction. *Math. Student* **29** 15–20.

[7] Mardia, K. V. (1962). Multivariate Pareto distributions. *Ann. Math. Statist.* **32** 1008–1015.

[8] Mardia, K. V. Exact distributions of extremes, ranges and mid-ranges in sample from any multivariate population (in press).