*The Annals of Mathematical Statistics*
1968, Vol. 39, No. 4, 1350–1353

# ON THE PROPORTION OF OBSERVATIONS ABOVE SAMPLE MEANS IN A BIVARIATE NORMAL DISTRIBUTION[1]

BY CHANDAN K. MUSTAFI

*Rutgers—The State University*

Let $(x_i, y_i)$ $(i = 1, 2, \cdots, n)$ be $n$ independent observations from a bivariate normal distribution with

$$E(x_i) = E(y_i) = 0, \quad E(x_i^2) = E(y_i^2) = 1, \quad E(x_i y_i) = \rho, \quad |\rho| < 1.$$

Let $\bar{x}$ and $\bar{y}$ be the sample means of $x$ and $y$ respectively. Suppose

$$\begin{aligned} P_n &= \text{proportion of } x_1, x_2, \cdots, x_n \text{ above } \bar{x}, \\ Q_n &= \text{proportion of } y_1, y_2, \cdots, y_n \text{ above } \bar{y}. \end{aligned} \tag{1}$$

In this note, we shall derive the limiting bivariate distribution of $(P_n, Q_n)$. This will be a generalization of the result obtained by David [2]. An application of this result will also be pointed out.

THEOREM. *The limiting distribution of* $(n^{\frac{1}{2}}(P_n - \frac{1}{2}), n^{\frac{1}{2}}(Q_n - \frac{1}{2}))$ *is bivariate normal with means zero and dispersion matrix*

$$\begin{pmatrix} \frac{1}{4} - (2\pi)^{-1} & (\arcsin \rho - \rho)(2\pi)^{-1} \\ (\arcsin \rho - \rho)(2\pi)^{-1} & \frac{1}{4} - (2\pi)^{-1} \end{pmatrix}. \tag{2}$$

PROOF. Let $t_1$ and $t_2$ be any two real numbers. Suppose,

$$\begin{aligned} p_1(n) &= \tfrac{1}{2} + n^{-\frac{1}{2}} t_1; \quad q_1(n) = \tfrac{1}{2} - n^{-\frac{1}{2}} t_1; \\ p_2(n) &= \tfrac{1}{2} + n^{-\frac{1}{2}} t_2; \quad q_2(n) = \tfrac{1}{2} - n^{-\frac{1}{2}} t_2. \end{aligned}$$

Let $U_q$ and $\xi_q$ be the sample and the population quantiles of order $q$ for $x$. Let $V_q$ and $\eta_q$ be the corresponding expressions for $y$.

$$\begin{aligned} &P\{n^{\frac{1}{2}}(P_n - \tfrac{1}{2}) \leqq t_1, n^{\frac{1}{2}}(Q_n - \tfrac{1}{2}) \leqq t_2\} \\ &= P\{P_n \leqq p_1(n), Q_n \leqq p_2(n)\} = P\{U_{q_1(n)} \leqq \bar{x}, V_{q_2(n)} \leqq \bar{y}\} \\ &= P\{n^{\frac{1}{2}}(U_{q_1(n)} - \bar{x} - \xi_{q_1(n)}) \leqq -n^{\frac{1}{2}} \xi_{q_1(n)}, \\ &\qquad n^{\frac{1}{2}}(V_{q_2(n)} - \bar{y} - \eta_{q_2(n)}) \leqq -n^{\frac{1}{2}} \eta_{q_2(n)}\}. \end{aligned} \tag{3}$$

Let,

$$a_n = n^{\frac{1}{2}}(U_{q_1(n)} - \bar{x} - \xi_{q_1(n)}), \quad b_n = n^{\frac{1}{2}} \bar{x}, \quad c_n = n^{\frac{1}{2}}(V_{q_2(n)} - \bar{y} - \eta_{q_2(n)}), \quad d_n = n^{\frac{1}{2}} \bar{y}.$$

Then,

$$a_n + b_n = n^{\frac{1}{2}}(U_{q_1(n)} - \xi_{q_1(n)}), \qquad c_n + d_n = n^{\frac{1}{2}}(V_{q_2(n)} - \eta_{q_2(n)}).$$

Received 11 October 1967.

[1] Research partially supported by Grant No. WP00457 of the Department of the Interior.

The vector $\mathbf{Z}_1 = (\bar{x},\bar{y})$ is independent of the vector $\mathbf{Z}_2 = (x_1 - \bar{x}, \cdots, x_n - \bar{x}, y_1 - \bar{y}, \cdots, y_n - \bar{y})$. The vector $(a_n, c_n)$ is a function of $\mathbf{Z}_2$ and the vector $(b_n, d_n)$ is a function of $\mathbf{Z}_1$. Hence, the vector $(a_n, c_n)$ is independent of the vector $(b_n, d_n)$. Let $\alpha_n(s_1, s_2)$, $\beta_n(s_1, s_2)$, $\gamma_n(s_1, s_2)$ be the characteristic functions of $(a_n, c_n)$, $(b_n, d_n)$ and $(a_n + b_n, c_n + d_n)$ respectively. Hence, ([1], P-285)

$$(4) \qquad \gamma_n(s_1, s_2) = \alpha_n(s_1, s_2)\cdot\beta_n(s_1, s_2),$$
$$\lim_{n\to\infty} \gamma_n(s_1, s_2) = [\lim_{n\to\infty} \alpha_n(s_1, s_2)][\lim_{n\to\infty} \beta_n(s_1, s_2)].$$

The asymptotic bivariate distribution of $(a_n + b_n, c_n + d_n)$ can be obtained from a result of Siddiqui ([3], p. 148) if we consider the random variables $(U_{q_1(n)}, V_{q_2(n)})$ instead of $(U_\alpha, V_\beta)$ and specialize to the case of bivariate normal distributions (see also [4]). This leads to the fact that $(a_n + b_n, c_n + d_n)$ has a limiting bivariate normal distribution with means zero and dispersion matrix $\mathbf{\Sigma}$, where

$$(5) \qquad \mathbf{\Sigma} = \begin{pmatrix} \pi/2 & 2\pi F(0, 0) - \pi/2 \\ 2\pi F(0, 0) - \pi/2 & \pi/2 \end{pmatrix}$$

and $F(0, 0) = P\{x_i < 0, y_i < 0\}$. It follows ([1], P-290)

$$(6) \qquad F(0, 0) = \tfrac{1}{4} + \text{arc sin } \rho/2\pi.$$

Hence, from (5) and (6)

$$(7) \qquad \mathbf{\Sigma} = \begin{pmatrix} \pi/2 & \text{arc sin } \rho \\ \text{arc sin } \rho & \pi/2 \end{pmatrix}.$$

Let $\mathbf{S}' = (s_1, s_2)$. Hence,

$$\lim_{n\to\infty} \gamma_n(s_1, s_2) = e^{-\frac{1}{2}\mathbf{S}'\mathbf{\Sigma}\mathbf{S}}.$$

Also,

$$\beta_n(s_1, s_2) = e^{-\frac{1}{2}\mathbf{S}'\mathbf{\Phi}\mathbf{S}},$$

where

$$(8) \qquad \mathbf{\Phi} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

From (4), (7) and (8), we obtain

$$(9) \qquad \lim_{n\to\infty} \alpha_n(s_1, s_2) = e^{-\frac{1}{2}\mathbf{S}'(\mathbf{\Sigma}-\mathbf{\Phi})\mathbf{S}}.$$

We now assert that the matrix $\mathbf{\Sigma} - \mathbf{\Phi}$ is positive definite for each $|\rho| < 1$. Let

$$C(\rho) = \text{arc sin } \rho - \rho.$$

Then, $C'(\rho) = (1 - \rho^2)^{-\frac{1}{2}} - 1 > 0$.

Thus the function $C(\rho)$ is monotone increasing in $\rho$. Also, $C(-1) = -\pi/2 + 1$, $C(1) = \pi/2 - 1$. Therefore,

$$-\pi/2 + 1 < \arcsin \rho - \rho < \pi/2 - 1,$$

and hence

$$(\arcsin \rho - \rho)^2 < (\pi/2 - 1)^2.$$

This shows that the matrix $\mathbf{\Sigma} - \mathbf{\Phi}$ is positive definite.

From (9) it follows that $(a_n, c_n)$ converges in law to $(a, c)$ where $(a, c)$ has a bivariate normal distribution with means 0 and dispersion matrix $\mathbf{\Sigma} - \mathbf{\Phi}$. Now, it can be proved that [2]

$$(10) \qquad \lim_{n\to\infty} (-n^{\frac{1}{2}}\xi_{q_1(n)}) = t_1(2\pi)^{\frac{1}{2}}, \qquad \lim_{n\to\infty} (-n^{\frac{1}{2}}\eta_{q_2(n)}) = t_2(2\pi)^{\frac{1}{2}}.$$

From (3), (10) and the uniformity of convergence it follows that

$$\lim_{n\to\infty} P\{n^{\frac{1}{2}}(P_n - \tfrac{1}{2}) \leqq t_1, n^{\frac{1}{2}}(Q_n - \tfrac{1}{2}) \leqq t_2\}$$
$$= \lim_{n\to\infty} P\{a_n \leqq -n^{\frac{1}{2}}\xi_{q_1(n)}, c_n \leqq -n^{\frac{1}{2}}\eta_{q_2(n)}\} = P\{a/(2\pi)^{\frac{1}{2}} \leqq t_1, c/(2\pi)^{\frac{1}{2}} \leqq t_2\}.$$

Thus, $(n^{\frac{1}{2}}(P_n - \frac{1}{2}), n^{\frac{1}{2}}(Q_n - \frac{1}{2}))$ has a limiting bivariate normal distribution with means zero and dispersion matrix $(2\pi)^{-1}(\mathbf{\Sigma} - \mathbf{\Phi})$ which is the same as (2). This completes the proof of the theorem.

It may be pointed out that the result derived above is independent of a change in location and scale of the variables. Thus, the theorem remains true if the random variables $(x_i, y_i)$ have a bivariate normal distribution with arbitrary means and arbitrary non-singular covariance matrix.

We now use the theorem to prove the following corollary:

COROLLARY. *Let $(x_i, y_i)$ $(i = 1, 2, \cdots, n)$ be independent observations from a bivariate normal distribution with arbitrary means and dispersion matrix. Let $\rho$ be the correlation coefficient between $x_i$ and $y_i$. Let*

$$(11) \qquad \begin{aligned} w_1(n) &= \textit{proportion of observations for which } x_i > \bar{x}, y_i > \bar{y};\\ w_2(n) &= \textit{proportion of observations for which } x_i > \bar{x}, y_i < \bar{y};\\ w_3(n) &= \textit{proportion of observations for which } x_i < \bar{x}, y_i < \bar{y};\\ w_4(n) &= \textit{proportion of observations for which } x_i < \bar{x}, y_i > \bar{y}; \end{aligned}$$

*so that*

$$w_1(n) + w_2(n) + w_3(n) + w_4(n) = 1.$$

*Then,*

(i) *$n^{\frac{1}{2}}(w_1(n) - w_3(n))$ is asymptotically normal with mean 0 and variance*

$$\tfrac{1}{2} + \pi^{-1}(\arcsin \rho - \rho - 1),$$

(ii) *$n^{\frac{1}{2}}(w_2(n) - w_4(n))$ is asymptotically normal with mean 0 and variance*

$$\tfrac{1}{2} - \pi^{-1}(\arcsin \rho - \rho + 1),$$

(iii) *$w_1(n) - w_3(n)$ and $w_2(n) - w_4(n)$ are asymptotically independent.*

PROOF. It follows from (1) and (11)

$$P_n = w_1(n) + w_2(n), \qquad Q_n = w_1(n) + w_4(n).$$

Hence

$$w_1(n) - w_3(n) = P_n + Q_n - 1, \qquad w_2(n) - w_4(n) = P_n - Q_n. \tag{12}$$

The proof of parts (i) and (ii) of the corollary is thus immediate from the theorem. Also, it follows from equations (2) and (12) that the asymptotic covariance between $w_1(n) - w_3(n)$ and $w_2(n) - w_4(n)$ is zero. This completes the proof of part (iii) of the corollary.

## REFERENCES

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

[2] DAVID, H. T. (1962). The sample mean among the moderate order statistics. *Ann. Math. Statist.* **33** 1160–1166.

[3] SIDDIQUI, M. M. (1960). Distribution of Quantiles in Samples from a Bivariate Population. *J. Res. N.B.S.B* **4B** (Math. and Math. Phys.) 145–160.

[4] SRIVASTAVA, O. P., HARKNESS, W. L. and BARTOO, J. B. (1964). Asymptotic distribution of distances between order statistics from bivariate populations. *Ann. Math. Statist.* **35** 748–754.