CHAPTER 4

APPLICATIONS

In this chapter we show that the limit theorem on the distribution of interactive quadruples a two-by-two table, as considered in [1] , is a special case of our Theorem 3.3.2. Furthermore, we shall show that our theorem can be applied to obtain asymptotic distribution of certain non-parametric statistics.

4.1 DISTRIBUTION OF INTERACTIVE QUADRUPLES IN A TWO-BY-TWO TABLES

In [1] , Arom consider independent identically distributed continuous random variables X_i, Y_j, Z_k and W_l, for $i = 1,\dots, P$, $j = 1,\dots, Q$, $k = 1,\dots, R$, $l = 1,\dots, S$ and define I and J to be the numbers of quadruples (X_i, Y_j, Z_k, W_l) such that $X_i > Y_j$, $X_i > Z_k$, $W_l > Y_j$, $W_l > Z_k$ and $X_i < Y_j$, $X_i < Z_k$, $W_l < Y_j$, $W_l < Z_k$ respectively. It was shown that the joint distribution of $\frac{I-E(I)}{\sqrt{Var(I)}}$ and $\frac{J-E(J)}{\sqrt{Var(J)}}$ tends to a bivariate normal distribution as P, Q, R, S tend to infinity. We shall show that our Theorem 3.3.2 is applicable to this situation with a milder assumption.

Let

$$X_1, \ldots, X_P$$

$$Y_1, \ldots, Y_Q$$

$$Z_1, \ldots, Z_K$$

$$W_1, \ldots, W_S$$

be independent random variables such that the random variables in each row are identically distributed. Define

$$f_1(x,y,z,w) = \begin{cases} 1 & \text{when } x > y, \; x > z, \; w > y, \; w > z, \\ 0 & \text{otherwise}, \end{cases}$$

$$f_2(x,y,z,w) = \begin{cases} 1 & \text{when } x < y, \; x < z, \; w < y, \; w < z, \\ 0 & \text{otherwise}. \end{cases}$$

Let

$$U_1 = \sum_{l=1}^{S} \sum_{k=1}^{R} \sum_{j=1}^{Q} \sum_{i=1}^{P} f_1(X_i, Y_j, Z_k, W_l)$$

and

$$U_2 = \sum_{l=1}^{S} \sum_{k=1}^{R} \sum_{j=1}^{Q} \sum_{i=1}^{P} f_2(X_i, Y_j, Z_k, W_l),$$

then $U_1 = I$, $U_2 = J$. By Theorem 3.3.2, we can conclude that $\frac{U_1 - E(U_1)}{\sqrt{Var(U_1)}}$ and $\frac{U_2 - E(U_2)}{\sqrt{Var(U_2)}}$ have a bivariate normal distribution when P, Q, R, S tend to infinity, hence $\frac{I-E(I)}{\sqrt{Var(I)}}$ and $\frac{J-E(J)}{\sqrt{Var(J)}}$ have a bivariate normal distribution when P, Q, R, S tend to infinity.

4.2 A BIVARIATE EXTENSION OF THE U STATISTIC

Let X, Y and Z be three random variables with continuous distribution function F, G and H. Whitney [3], propose a test of the hypothesis

$$H_o : F = G = H ,$$

against the alternatives

$$H_1 : F > G , F > H ,$$

or

$$H_2 : F > G > H ,$$

by using two statistics U, V. To test such a hypothesis with a sample of l X's, m Y's and n Z's from three populations, we arrange the sample values in ascending order and let U count the number of times a Y precedes an X, and V count the number of times a Z precedes an X. The proposed tests use $U \leq K_1$, $V \leq K_2$ and $U \geq K_3$, $V \leq K_4$ as critical regions for testing H_o against H_1 and H_2 respectively. The constants K_i are chosen to give any required significance level. We shall use Theorem 3.3.2 to show that the joint distribution of $\frac{U-E(U)}{\sqrt{Var(U)}}$ and $\frac{V-E(V)}{\sqrt{Var(V)}}$ tends to a bivariate normal distribution when l, m, n tend to infinity.

Let

$$X_1, \ldots, X_l$$
$$Y_1, \ldots, Y_m$$
$$Z_1, \ldots, Z_n ,$$

be independent random variables such that the random variables in each row are identically distributed. Define

$$f_1(x,y,z) = \begin{cases} 1 & \text{when } y < x \,, \\ 0 & \text{otherwise} \,, \end{cases}$$

and

$$f_2(x,y,z) = \begin{cases} 1 & \text{when } z < x \,, \\ 0 & \text{otherwise} \,, \end{cases}$$

put

$$U_1 = \sum_{k=1}^{n} \sum_{j=1}^{m} \sum_{i=1}^{l} f_1(x_i, y_j, z_k)$$

$$U_2 = \sum_{k=1}^{n} \sum_{j=1}^{m} \sum_{i=1}^{l} f_2(x_i, y_j, z_k) \,,$$

then $U = \frac{U_1}{n}$ and $V = \frac{U_2}{m}$.

By Theorem 3.3.2 we can conclude that $\frac{U_1 - E(U_1)}{\sqrt{Var(U_1)}}$ and $\frac{U_2 - E(U_2)}{\sqrt{Var(U_2)}}$ have a bivariate normal distribution as l, m, n tend to infinity, but

$$\frac{U_1 - E(U_1)}{\sqrt{Var(U_1)}} = \frac{nU - E(nU)}{\sqrt{Var(nU)}}$$

$$= \frac{nU - nE(U)}{n\sqrt{Var(U)}}$$

$$= \frac{U - E(U)}{\sqrt{Var(U)}} \;;$$

and

$$\frac{U_2 - E(U_2)}{\sqrt{Var(U_2)}} = \frac{mV - E(mV)}{\sqrt{Var(mV)}}$$

$$= \frac{V - E(V)}{\sqrt{Var(V)}} ,$$

hence the joint distribution of $\frac{U - E(U)}{\sqrt{Var(U)}}$ and $\frac{V - E(V)}{\sqrt{Var(V)}}$ has a bivariate normal distribution as l, m, n tend to infinity.

Observe that Theorem 3.3.2 can be applied to the case of k+1 populationsas well. For each i, $i = 1,\ldots,k$, let X_{ij}, $j = 1,\ldots,n_i$, be sample from the i^{th} population with distribution function F_i. The above test of H_o against H_1 can be generalized to testing

$$H_o : F_0 = F_1 = F_2 = \ldots = F_k ,$$

against the alternative hypothesis

$$H_1 : F_0 > F_1, F_0 > F_2, \ldots, F_0 > F_k .$$

To test the hypothesis with a sample of l_i X_{ij}'s we arrange the sample values in ascending order and let U_i be the number of times an X_i precedes an X_0, for $i = 1,\ldots,k$. We can conclude that the distribution of $\frac{U_i - E(U_i)}{\sqrt{Var(U_i)}}$ for $i = 1,\ldots,k$, tends to a k-variate normal distribution as l_i's tend to infinity.