J.Math. Res.& Exposition, 1988, No. 4

A Note on "Non-independence and Uncorrelatedness"*

Y. Wu (吴月华)
(Nanjing Institute of Technology)

Z. Liu (刘志军)
(Inst. of Systems Science, Acad. Sinca, China)

An elementary fact in the theory of probability is that there exists such random variables X, Y which are both non independent and uncorrelated. A well known example is that (X, Y) has a density function

$$f(x, y) = \frac{1-a}{2\pi} e^{-1/2(x^2+y^2)} + \frac{a}{\pi} 1_{(x^2+y^2 \le 1)}$$

with $|a|$ sufficient small. In this example, the marginal distribution of X and Y are both normal. A question naturally presents itsselfthat given two known distribution function F and G; doth having non zero and finite variance, can we co construct a bivariate distribution $H(x, y)$ such that

1° $H(x, \infty) = F(x)$, $H(\infty, y) = G(y)$

2° If (X, Y) possesses distribution function H, then X, Y are non independent and uncorrelated.

Chen and He [1] answered this question by proving the following:

Theorem. The necessary and sufficient condition for such a bivariate distribution to exist is that at most one of F and G can be two point distribution.

Chen and He proved their theorem by a very complicated construction process. In this note, we give a simple expression for H. By doing so, we not only greatly simplify their proof, but also extend their result to n random variables. Specifically speaking, we shall prove:

Theorem Suppose that $F_1, \cdots, F_n$ are n one dimensional distributions having non zero and finite variances. The necessary and sufficient condition for the existence of a n dimensionsl distribution $H(x_1, \cdots, x_n)$ sutisfying

1° $H(\infty, \cdots, \infty, x_i, \infty, \cdots, \infty) = F_i(x_i)$, $i = 1, \cdots, n$;

2° if $(x_1, \cdots, x_n)$ possesses distribution funetion H, then for any $i \neq j$, X_i and X_j are both non independent and uncorrelated.

is that at most one of $F_1, \cdots, F_n$ can be a two point distribution.

Proof The necessity of this condition is obvious and has been pointed out in [1]. For sufficiency, without lossing generality we assume that the

* *Received Feb. 7, 1987.*

support of F_i has at least three points, for $i=1,\cdots,n-1$ (in any case, this can be achieved by rearranging the order of $X_1,\cdots,X_n$). Define n numbers $a_1,\cdots,a_n$:

$$a_i=\int_{-\infty}^{\infty}F_i^2(x)[1-F_i(x)]dx\Big/\int_{-\infty}^{\infty}F_i(x)[1-F_i(x)]dx. \qquad (1)$$

By the assumptions imposed on F_i, a_i is well-defined. Now we put

$$H(x_1,\cdots,x_n)=\prod_{j=1}^{n}F_j(x_j)\{1+a\sum_{1\le i<j\le n}[1-F_i(x_i)][1-F_j(x_j)][a_i-F_i(x_i)]\} \qquad (2)$$

where $a\neq0$ is a constant, $|a|$ sufficiently small.

First we show that the function H defined by (2) is a n dimensional distribution function. Since it is trivial to see that H is left continuous in each argument, and

$$\lim_{x_i\to-\infty}H=0, \qquad \lim_{\substack{x_i\to\infty\\ i=1,\cdots,n}}H=1.$$

One only needs to show that for any $x_i<y_i$, $i=1,\cdots,n$, it is true that

$$\Delta_1\Delta_2\cdots\Delta_nH(x_1,\cdots,x_n)\geq0,$$

Where Δ_j is the difference-operator on x_j (from x_j to y_j). Write

$$\Delta=\prod_{j=1}^{n}[F_i(y_i)-F_i(x_i)].$$

It is easy to verify that

$$\Delta_1\cdots\Delta_nH=\Delta\{1+a\sum_{1\le i<j\le n}[a_i-(a_i+1)[F_i(x_i)+F_i(y_i)]+F_i^2(x_i)+$$
$$+F_i(x_i)F_i(y_i)+F_i^2(y_i)][1-F_j(x_j)-F_j(y_j)]\}. \qquad (3)$$

Since $0<a_i\leq1$, we have

$$\Big|\sum_{1\le i<j\le n}[a_i-(a_i+1)(F_i(x_i)+F_i(y_i))+F_i^2(x_i)+F_i(x_i)F_i(y_i)+F_i^2(y_i)]\cdot$$
$$\cdot[1-F_j(x_j)-F_j(y_j)]\Big|\leq4n^2 .$$

Therefore, if we choose $a\in(0,1/4n^2)$, then, noticing that $\Delta>0$, from (3) we have $\Delta_1\cdots\Delta_nH\geq0$. This completes the proof that H is a n-dimensional distribution.

Now we prove that for any $i<j$, X_i,X_j are dependent (remember that $(X_1,\cdots,X_n)$ has H as its distribution). For this purpose note that the marginal distribution of (X_i,X_j) is

$$G_{ij}(x_i,x_j)=F_i(x_i)F_j(x_j)[1+a(1-F_i(x_i))(1-F_j(x_j))(a_i-F_i(x_i))] \qquad (4)$$

Since the support of F_i has at least three points. In this case we have

$$F_i(x_i)(1-F_i(x_i))(a_i-F_i(x_i))\not\equiv0 \quad \text{on } x_i\in(-\infty,\infty).$$

Also, since F_j is non-degenerated, we hove

$$F_j(x_j)[1-F_j(x_j)]\not\equiv0 \text{ on } x_j\in(-\infty,\infty).$$

This shows that $G_{ij}(x_i,x_j)\not\equiv F_i(x_i)F_j(x_j)$, on $(x_i,x_j)\in R^2$, Proving that X_i,

X_j are non independent.

Finally, we prove that X_i, X_j are uncorrelated. In doing this job we make uss of the well known fact that

$$\mathrm{Cov}(X_i, X_j) = \iint_{-\infty}^{\infty} (G_{ij}(x_i, x_j) - F_i(x_i)F_j(x_j)]dx_i dx_j$$

By (4), we have

$$\mathrm{Cov}(X_i, X_j) = a\int_{-\infty}^{\infty} F_i(x_i)[1 - F_i(x_i)][a_i - F_i(x_i)]dx_i \cdot$$

$$\int_{-\infty}^{\infty} F_j(x_j)[1 - F_j(x_j)]dx_j$$

which is indeed zero, by the definition of a_i (1). This completes the proof of the Theorem.

Chen and He also mentioned an possible extension to the general case: Given two distribution funetions F and G with m and n dimensions respectively, one is asked to construct a $(m+n)$ dimensional random vector $(X_1, \cdots, X_m, Y_1, \cdots, Y_n)$, such that

1° the marginal distribution of $(X_1, \cdots, X_m)$ is F, while that of $(Y_1, \cdots, Y_n)$ is G;

2° for any X_i and Y_j, X_i, Y_i are both non independent and uncorrelated.

We have not succeeded yet in employing the method of this note to deal with this question, though it seems probable that their general problem might be solved by a suitable modification of our method.

Acknowledgement. The authors give sincere thanks to Dr.s Zhidong Bai, Lincheng Zhao and Baiqi Miao, for thir guidance and useful suggstions.

References

[1] X. Chen & S. He, Non-independence and uncorrelatedness, To appear in 《Applied Probability and Statistics》 (in Chinese).