ON GENERATION OF RANDOM CORRELATION MATRICES

Hsiang-Chuan Liu

Abstract

We provide a geometric method to generate a pxp Correlation matrix ρ with (i, j) th entry:

$\rho_{ij} = \text{Cos}\theta_i \text{Cos}\theta_j \text{Cos}(\phi_i - \phi_j) + \text{Sin}\theta_i \text{Sin}\theta_j$

where θ_i and θ_j are independent variates from $U(0,\pi)$

ϕ_i and ϕ_j are independent variates from $U(0,2\pi)$

$i, j = 1, 2, \cdots\cdots, p$

θ and ϕ are statistically independent

Introduction

The P-variate normal distribution with mean vector u and variance-covariance matrix Σ (non-singular)is defined by the pdf:

$$f(x) = (2\pi)^{\frac{-p}{2}} \Sigma^{\frac{-1}{2}} \exp[-\frac{1}{2}(\underline{x} - \underline{u})\Sigma^{-1}(\underline{x} - \underline{u}) \quad \cdots\cdots(1)$$

we denote this distribution :

$$N_p(\underline{u}, \Sigma) \quad \cdots\cdots(2)$$

In simulation, an obvious method to generate a P vector $\underline{X}$ with this distribution is to generate $Z_1, Z_2, \cdots\cdots, Z_p$ as independent $N_1(0, 1)$variates forming the P vector Z having $n_p(Q, I_p)$ then form $\underline{X}$ as

$$\underline{X} = A\underline{Z} + \underline{u} \quad \cdots\cdots(3)$$

where A is Sed and

$$A = V/^{\frac{1}{2}} \rho^{\frac{1}{2}} \quad \cdots\cdots(4)$$

$$\Sigma = V/^{\frac{1}{2}} \rho^{\frac{1}{2}} \quad \cdots\cdots(5)$$

$$V/^{\frac{1}{2}} = \text{diag}(\sqrt{\sigma_{11}} \cdot \sqrt{\sigma_{22}} \cdots \sqrt{\sigma_{pp}}) \quad \cdots\cdots(6)$$

is the standard deviation matrix, ρ is the correlation matrix.

If we want first to choose some different u's and Σ's.

It is easy to select u and V/. but somewhat difficult to select ρ , since any correlation matrix must be non-negative definite.

Here we'll propose a geometric method.

Geometric Meaning of Correlation Coefficient

In sample case :

Let $[(X_1 , Y_1) , (X_2 , Y_2) , \cdots\cdots , (X_p , Y_p)]$

be a random sample with size p , denote

$$\underline{X} = [x_1, x_2 , \cdots\cdots , x_p]$$
$$= [X_1 - \overline{X} , X_2 - \overline{X} , \cdots , X_p - \overline{X}]$$
$$\underline{Y} = (y_1 , y_2 , \cdots y_p) = (Y_1 - \overline{Y} , Y_2 - \overline{Y} , \cdots\cdots Y_p - \overline{Y})$$

where $\begin{cases} \overline{X} = \frac{1}{p}\sum_{i=1}^{p} X_1 \\ \overline{Y} = \frac{1}{p}\sum_{i=1}^{p} Y_1 \end{cases}$

then

$$\underline{X} \cdot \underline{Y} = \sum_{i=1}^{p} x_i \, y_i = \sum_{i=1}^{p} [X_i - \overline{X}][Y_i - \overline{Y}]$$
$$= [p - 1]S_{xy} \cdots\cdots (8)$$
$$X = \sqrt{x \cdot x} = [\sum_{i=1}^{p} x_i^2]^{\frac{1}{2}}$$
$$= [\sum_{i=1}^{p} [X_i - \overline{X}]^2]^{\frac{1}{2}}$$
$$= \sqrt{[p - 1]}\, S_x \cdots\cdots (9)$$
$$Y = [\sum_{i=1}^{p} (Yi - \overline{Y})^2]^{\frac{1}{2}}$$
$$= \sqrt{[p - 1]}\, S_y \cdots\cdots (10)$$

hence

$$\gamma_{xy} = \frac{\sum_{i=1}^{p} [X_i - \overline{X}][Y_i - \overline{Y}]}{\sqrt{[\sum_{i=1}^{p} [X_i - \overline{X}]][\sum_{i=1}^{p} [Y_i - \overline{Y}]]}} = \frac{S_{xy}}{S_x S_y}$$

$$= \frac{\underline{x} \cdot \underline{y}}{\| \underline{x} \| \cdot \| \underline{y} \|} = \mathrm{Cos}\theta_{xy} \cdots\cdots (11)$$

In population case :

[Definition 1]

Let $g = g(X, Y)$, $h = h(X, Y)$ be two functions of random vector (X, Y)

(i) The inner porduct of two random functions g and h is defined as

$$g \cdot h = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)h(x, y)f(x, y)dxdy \quad \cdots\cdots(12)$$

(ii) The norm of random function g is defined as

$$\| g \| = \sqrt{g \cdot g} \quad \cdots\cdots(13)$$

here suppose

$$\begin{cases} g(x, y) = x - u_x \\ h(x, y) = y - u_y \end{cases} \quad \cdots\cdots(14)$$

where $\begin{cases} u_x = \int_{-\infty}^{\infty} xf(x)dx \\ u_y = \int_{-\infty}^{\infty} yf(y)dy \end{cases}$

$f(x)$ and $f(y)$ are marginal pdf's of X and Y respectively
then

$$g \cdot h = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - u_x)(y - u_y)f(x, y)dxdy$$

$$= \sigma_{xy} \quad \cdots\cdots(15)$$

where σ_{xy} is the covariance of X and Y

$$\| g \| = \sqrt{g \cdot g} = \left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - u_x)^2 f(x, y)dxdy\right]^{\frac{1}{2}}$$

$$= \left[\int_{-\infty}^{\infty}(x - u_x)^2 f(x)dx\right]^{\frac{1}{2}}$$

$$= \sigma_x \quad \cdots\cdots(16)$$

similarly we have

$$\| h \| = \sqrt{h \cdot h} = \sigma_y \quad \cdots\cdots(17)$$

where σ_x and σ_y are standard deviationl of x and y respectively

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{g \cdot h}{\| g \| \cdot \| h \|} = \mathrm{Cos}\theta_{gh} \quad \cdots\cdots(18)$$

Therefore, we know that any correlation coefficient of two random variables may be viewed as the cosine of the angle between two corresopnding vectors, conversely, any cosine of the angle between two vectors can be viewed as a correlation coefficient of two corresponding random vectors.

Generating Random Correlation Matrices

In view of the above discussion, we can generate any pxp correlation matrix by considering the cosine of the angles of the sampling vectors from the unit sphere :

$$S = \{(X_1 , X_2 , X_3) \mid X_1^2 + X_2^2 + X_3^2 = 1 , \quad X_1 , X_2 , X_3 \in R\} \cdots\cdots (19)$$

Let $\underline{X}_i = (X_{1i} , X_{2i} , X_{3i})$ be any vector belonging to S

$i = 1 , 2 , \cdots\cdots , P$

write $\underline{X}_1$ in spherical coordinate :

$$\exists ! \theta_i \in (0 , \pi) , \quad \exists ! \phi_i \in (0 , 2\pi)$$

such that

$$\begin{cases} X_{1i} = \mathrm{Cos}\theta_i \ \mathrm{Cos}\phi_i \\ X_{2i} = \mathrm{Cos}\theta_i \ \mathrm{Sin}\phi_i \\ X_{3i} = \mathrm{Sin}\theta_i \end{cases} \cdots\cdots (20)$$

then we have

$$\rho_{ij} = \frac{\underline{X}_i \cdot \underline{X}_j}{|\underline{X}_i| \cdot |\underline{X}_j|} = \underline{X}_i \cdot \underline{X}_j$$

$$= \mathrm{Cos}\theta_i \ \mathrm{Cos}\phi_i \ \mathrm{Cos}\theta_j \ \mathrm{Cos}\phi_j + \mathrm{Cos}\theta_i \ \mathrm{Sin}\phi_i \ \mathrm{Cos}\theta_j \ \mathrm{Sin}\phi_j + \mathrm{Sin}\theta_i \ \mathrm{Sin}\theta_j$$

$$= \mathrm{Cos}\theta_i \ \mathrm{Cos}\theta_j \ \mathrm{Cos}(\phi_i - \phi_j) + \mathrm{Sin}\theta_i \ \mathrm{Sin}\theta_j \cdots\cdots (21)$$

Therefore the desired radom correlation matrix is

$$\rho = [\rho_{ij}]_{p \times p} \cdots\cdots (22)$$

Conclusion

In consideration of the geometric meaning of correlation coefficients, we obtain a generating method to build a pxp random correlation matrix ρ with(i.j)th entry :

$$\rho_{ij} = \mathrm{Cos}\theta_i \ \mathrm{Cos}\theta_j \ \mathrm{Cos}(\phi_i - \phi_j) + \mathrm{Sin}\theta_i \ \mathrm{Sin}\theta_j \cdots\cdots (23)$$

where $\begin{cases} \theta_i \text{ and } \theta_j \text{ are independent variates} \\ \text{from U } (o,\pi) \\ \phi_i \text{ and } \phi_j \text{ are independent variates} \\ \text{from U } (o,2\pi) \\ i,j = 1,2, \cdots\cdots ,p \\ \theta \text{ and } \phi \text{ are statistically independent} \end{cases}$

References

Andersom, T.W.(1984) An Introduction to multivariate Statistical Analysis, 2nd ed. John Weley & Sons.

Chalmers, C.P.(1975) Generation of Correlation Matrices with a Given Eigen-Structure, JSCS, 133-139.

Kennedy, Jr. W.J. & Gentle, J.E.(1980) Statistical Computing, translated by Taipei Hua Tai Book Store(1982).

Lang, S. (1985) Linear Algebra 2nd ed. Addison Wesley Pubishing Co.

Maindonald (1984) Statistical Computation translated by Taipei Hua Tai Book Store.