

Simulation of Non Normal Distribution in Simple Steps

By Chandrakant Maheshwari

Introduction

In risk management, assumption of data distribution is important because using that assumption the risk Managers come up with the required risk related numbers (especially Value at Risk and Potential Future Exposure). The accurate calculation of these numbers ensures robust risk management in the organization.

Assumption of normal distribution for a given data series is most favorite among risk managers for the data they have. It is because the simplicity of its implementation as well as the ease of explanation to their reporting managers/peers and finally its theoretic tractability. Unfortunately in reality most of the data series is far from normal even though by theoretical intuition they should follow normal distributions (example: distribution of price returns). Those series demonstrate high amount of skew and kurtosis due to which statistical tests for normality fail. Also there is a limit in availability of data which prevents the risk practitioner to use historical simulation approach to come up with the required risk number. Under such circumstances, the practitioner has to rely on Monte Carlo simulation approach to represent the actual data distribution by a simulated distribution.

This issue is a widely researched in academia and they have also given good practical solutions; but most of those research scripts are quite mathematical in their discussion and pose a challenge for any practitioner who has not received a rigorous training in any quantitative subject.

In this article without going into the mathematical rigor, we present a simple and step by step approach to develop a non-normal distribution using Monte Carlo simulation. By saying simple, we do not mean that we simplify any mathematics around it, but rather refrain from going into the details of their proofs. For any reader who is interested in going into the details can read the papers mentioned in the references in this article. Our purpose is that after reading this article a practitioner with a little quant back ground should be able to understand and then develop his own non normal distribution.

In the next section we present the methodology to generate the non-normal distribution. We generate the non-normal distribution using a non-linear transformation. This involves an optimization exercise, so in the subsequent section, we provide a SAS code where the parameters which will generate required skew and kurtosis values are derived. In the next section we summarize the steps in simple manner.

For users who do not have SAS, we provide tables (Appendix B) for the parameters needed for transformation to get various values of skew and kurtosis.

Methodology

Fleishman (1978) suggested a method for generating non-normal variables with specified values of skew and kurtosis. The idea is simple: Generate a random standard normal deviate X and compute

$$Y = a + bX + cX^2 + dX^3$$

Here Y will be having mean 0 and variance 1 and required skew and kurtosis. However, the difficult part is to determine a , b , c , and d such that Y has the specified values of mean, standard deviation, skew and kurtosis. Given X is standard normal distribution, the expected value of Y will be

$$E(Y) = E(a) + E(bX) + E(cX^2) + E(dX^3)$$

$$E(Y) = a + bE(X) + cE(X^2) + dE(X^3) \text{ and } E(X) = 0, E(X^2) = 1, E(X^3) = 0$$

As discussed $E(Y) = 0$. Hence $a = -c$. The equation now becomes

$$Y = -c + bX + cX^2 + dX^3 \quad (1)$$

For Y to have required skew and kurtosis b, c and d must satisfy the following three equations

$$F_1 = b^2 + 6bd + 2c^2 + 15d^2 - 1 = 0$$

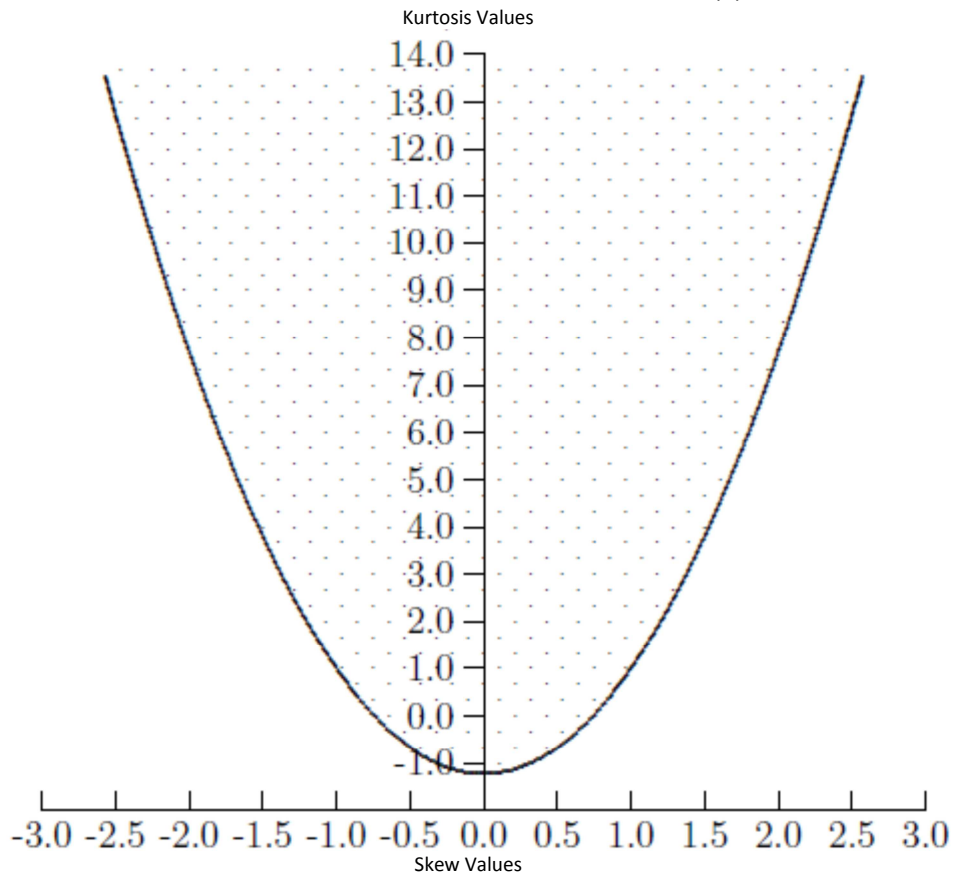
$$F_2 = 2c(b^2 + 24bd + 105d^2 + 2) - \text{RequiredSkewValue} = 0$$

$$F_3 = 24(bd + c^2(1 + b^2 + 28bd) + d^2(12 + 48bd + 141c^2 + 225d^2)) - \text{ReqKurtosis} = 0$$

But before going further, please ensure that the combination of skew and kurtosis falls in the dotted region below.

For other combinations this methodology is not appropriate. The equation below graph follows:

$$\text{Kurtosis} = -1.2264489 + 1.6410373 \times \text{Skew}^2 \quad \text{--- (2)}$$



Summarizing Simulation Steps

Suppose the user wants to generate the non-normal random distribution with mean (μ), standard deviation (σ), skew and kurtosis. He/she should follow the following steps:

- 1) Confirm that the skew and kurtosis combination follows in the graph above or the equation 2.
- 2) Derive the values of parameters b, c , and d using the SAS code above or from table (link in Appendix B).
- 3) Generate X as standard normal distribution.
- 4) Derive Y using equation 1.
- 5) Get the required random number distribution $Z = \mu + \sigma \times Y$

Conclusion

Assumption of normality of data distribution is simple and intuitive for data modeling exercises. But in spite of sound logical assumptions from theoretical intuition (typical example is that returns of an asset are assumed to be normally distributed); real data series fails the statistical tests of normality. This is because of skew-ness and fat tails (kurtosis) in them. Researchers have made a lot of progress to include skew and kurtosis of the data; but most of the papers describing the process are mathematically overwhelming for practitioners and business professionals. We by no means undermine or compromise the importance of mathematical rigor; but believe that as long as right references are known and available to the business professional; he should be able get his assignment done correctly without any need of detailed understanding the rigor. Hence in this paper without including any mathematical rigor we guide the practitioners how they can implement their required non-normal distribution without losing conceptual soundness.

References

1. Fleishman, A. I. (1978). A method for simulating non-normal distributions. Psychometrika, 43(4):521-532.
2. Luo, H (2011). Generation of Non-normal Data – A Study of Fleishman’s Power Method

Appendix A: SAS Code

```
proc nlp tech=trureg outest=res;
min y;
decvar b c d;;
f1 = b*b+6*b*d+2*c*c+15*d*d-1;
f2 = 2*c*(b*b+24*b*d+105*d*d+2) - RequiredSkewValue;
f3 = 24*(b*d + c*c*(1 + b*b + 28*b*d) + d*d*(12 + 48*b*d + 141*c*c + 225*d*d))- RequiredKurtosisValue;
y = (f1 * f1 + f2 * f2+ f3 * f3);
run;
```

Appendix B: Fleishman’s Power Method Coefficient table

Table <http://chandrakant721.wordpress.com/fleishmans-power-method-coefficient-table/>

Chandrakant Maheshwari

Manager – Genpact Smart Decision Services
Financial Services Analytics
Chandrakant.Maheshwari@genpact.com

About the Author

Chandrakant Maheshwari is a manger in the Financial Services Analytics division at Genpact. He has nine years of experience in the area of financial risk management. He is an alumnus of IIT Delhi with a degree in Mathematics and Computing, a 5 yr Integrated M. Tech course.