Application of Convolution in Individual Risk Model with non-i.i.d. Data: A Case Study

By:

Deep Mukherjee

Abstract:

Purpose of this paper is to explain how the concept of convolution of continuous distributions could be applied to insurance problems. Building a problem in a very general way, the paper dealt with a complex scenario: where the continuous distributions to be summed are not identical, and no closed form of convoluted distribution exists in literature. Some solutions to the problem are mentioned, and Fenton-Wilkinson method in particular is explained in great detail.

Keywords: Convolution, Individual Risk Model, Fenton-Wilkinson Method, Simulation.

1. Introduction:

One of the most important concepts in Mathematical Statistics is that of a convolution. To put in a simple manner: if a new random variable U = X + Y, where X and Y are random variables with PDFs $f_X(x)$ and $g_Y(y)$ respectively, then PDF of the sum (ie. Variable U) is called the convolution of X and Y. As we are interested in continuous random variables, mathematically speaking, a convolution is defined as:

$$\mathbf{C(u) = f(x) \otimes g(x) = \int_{space} f(x)\, g(u-x)\, dx}$$

$$\mathbf{= g(x) \otimes f(x) = \int_{space} g(x)\, f(u-x)\, dx}$$

The convolution integral is continuously evaluated at each shift u by multiplication and integration of *f (x)* times *g (u-x)* for all values of u running from -infinity to + infinity. Note: *f(u)* and *g(u)* are the same as *f(x)* and *g(x);* the variable symbol makes no difference. *g(-x)* is *g(x)* that is flipped or reversed in variable and *g(u-x)* is the function *g(-x)* shifted along the X axis by an amount u. Another point to note is: it doesn't matter which function one takes first, *i.e.* the convolution operation is commutative. The following pictorial illustration [Fig.1] might be helpful to understand how convolution actually works.

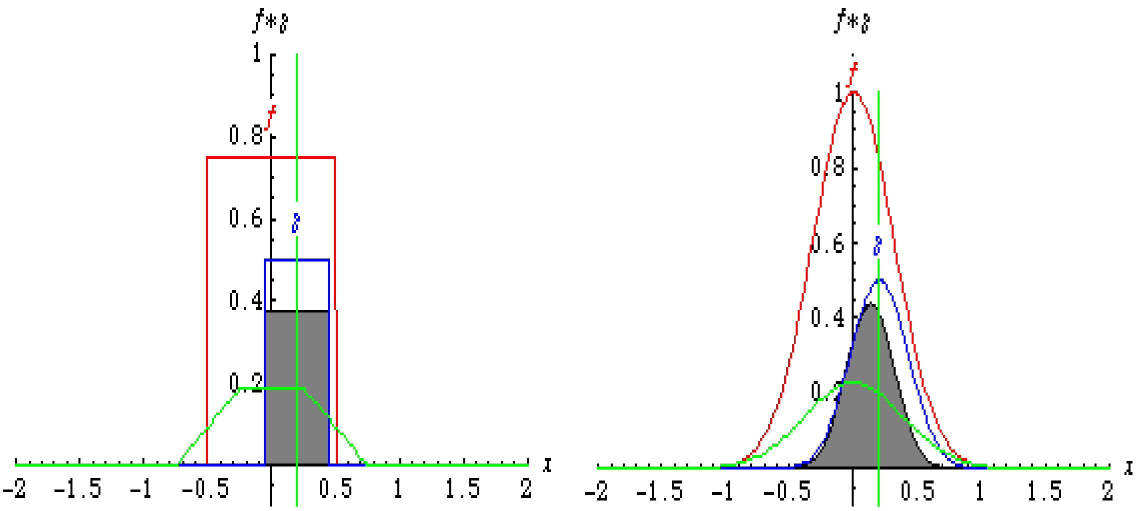

Figure 1: Convolution of two Uniform distributions (left) and two Normal distributions (right)

In each plot, the green curve shows the convolution of the blue and red curves as a function of *u*, the position indicated by the vertical green line. The gray region indicates the product *f (x) g (u-x)* as a function of *u*, so its area as a function of *u* is precisely the convolution.

2. How to convolute?

Some simple convolution results are readily available for use. A few are listed below (for continuous distributions):

(1) If two independent Normal random variables $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.
(2) If two independent Uniform random variables $X_1 \sim U[0,1]$ and $X_2 \sim U[0,1]$ then their $X_1 + X_2$ has the so-called triangular distribution on [0,2].
(3) If two independent Gamma variables $X_1 \sim \Gamma(\alpha, \beta_1)$ and $X_2 \sim \Gamma(\alpha, \beta_2)$ then $X_1 + X_2 \sim \Gamma(\alpha, \beta_1 + \beta_2)$

But for more complex cases, one can find three alternative methods available in statistics literature: *(1) Analytic method* [based on transformation like Fast Fourier Transformation (FFT) or Laplace Transformation], *(2) Numeric method* [based on Monte Carlo Simulation] and (3) *Method of Moments* [will be described in detail, as we proceed]. We are not in the position to comment on the relative robustness of these alternatives, but we can certainly point out some aspects from the point of applicability. It is not easy to use transformations always and one will land up with some very complex form, which may be beyond interpretation. MATLAB software is the most efficient in doing FFT. Concept wise Simulation is simpler and can be done in many mediums, but one cannot get any parametric form as the end result. Method of moments is simple to understand and can be worked out even in MS Excel/Access.

3. Application of convolution in insurance problems:

When pricing insurance contracts, the initial approach is to estimate the mean and variance of insured (aggregate) loss from that contract. Consider the popular Individual Risk Model. In this framework we assume that the portfolio (/policy) consists of *n* insurance policies (/exposures) and X_k denotes the claim made in respect of the policy (/exposure) *k*. Then aggregate amount of claim (from that portfolio/ policy) is $S = X_1 + X_2 + \ldots\ldots + X_n$, where X_k is the loss on insured unit *k* and *n* is the number of risk units insured (known and fixed at the beginning of the period). The X_k's are usually postulated to be independent random variables (*but not necessarily identically distributed*). The next steps are to specify distributions for frequency and severity and eventually compute E (S) and V (S). Then use the deviation from the mean or average value to measure the risk taken and to calculate the loading or extra part of the premium. This suggests that we could use the variance as a measure of risk.

However, the standard deviation (or, variance) has few disadvantages:

- It treats the negative and the positive deviations from the mean in the same way. (As for an insurance company, it is only concerned with positive deviations from mean as in these cases actual claim amount that it has to pay is more than the expectation and might not be covered by the premium earned.) Markowitz, who developed CAPM, realized this and he said that other measures can be used, e.g. the semi-variance.
- In insurance the use of the standard deviation is doubtful when we have to deal with probability distributions with fat tails. In such cases the standard deviation might not even exist.

Some brought forward the concept of "Coherent Measures of Risk" to provide a better measurement of risk. The main point is that computation of these measures (eg.: Value at Risk, Expected Tail Loss) depend on the underlying parameters and distributional assumptions about S. So for greater understanding of the risk coming from a collection of exposures, distribution (either parametric or non-parametric) of the aggregate claim size is very much required. Derivation of the same requires convoluting distributions.

4. An Example of how to convolute in complex scenarios:

What follows next is an example showing how to do convolution, when no already established convolution result exists. Lognormal distribution is picked for that purpose, as it is one of the most useful loss distributions in practice. In property insurance, it is quite expected that loss distribution will not remain the same over the entire range of sum insured. Similarly, in the health insurance loss distribution may vary over different age groups, sex and many other factors. Even if the functional form remains same, parameters may change. So, this adds more to the first problem (not having any closed form of convoluted distribution). Data is not i.i.d – losses though independent, are not coming from identical distributions. The example is built in a general set-up. Applying CART or CHAID algorithm on the past loss record, one can find out the groups or classes (of sum insured/age/sex or some other relevant factor) such that within each class loss is homogeneous but between the groups it is heterogeneous.

Let us consider N exposures in a policy. Assume, we already know K mutually exclusive and collectively exhaustive bands and found out individual claim sizes in each band follows lognormal distribution with different parameters (from study on past losses). Assume, we know the binomial and multinomial distributions of number of loss for these K bands. To start with, one can group N exposures in K bands. N_k denotes the number of exposures in k-th band. So, $N = N_1 + N_2 + \ldots + N_K$. Let n_k denotes the number of losses from k-th band. Before we proceed further, let us make some restrictive assumptions, to avoid complexity:

- An exposure will suffer from a loss maximum once in an underwriting year.
- Once hit by a loss, the exposure will not be a part of the 'active' exposure set for the rest of the underwriting year.

According to the above assumptions, n, the number of losses from the policy can take any value from 0 to N. n_k can take any value from 0 to N_k. Note, $n = n_1 + \ldots + n_K$. So, four losses (say, $n = 4$) can occur in many ways – different combinations of $n_1, n_2, \ldots, n_K$'s. In each case, the variables or distributions to be summed will change. To take care of all possible sums, one has to list down all possible combinations of $n_1, n_2, \ldots, n_K$'s, keeping the assumptions in mind. To know in how many ways four losses (for example) can occur, one has to look at the coefficient of x^4 in the expression below:

$$\prod_{k=1}^{K} (1 + x + x^2 + \ldots + x^{N_k})$$

Let α_n denotes the coefficient of x^n. Then the total number of possibilities $\alpha = \sum_{n=0}^{N} \alpha_n$.

There is another case: '0' aggregate loss, when no exposure is hit by loss. For greater understanding of the above logic, please refer the table below:

Table 1: The below table shows a glimpse of different possibilities of loss occurring

No. of losses	Group 1	Group 2	……	Group K-1	Group K	
n = 1	1	0	……	0	0	
	0	1	……	0	0	
	……		……		……	
	0	0	……	0	1	(Total:K rows)
n = 2	2	0	……	0	0	
	1	0	……	0	1	
	……		……		……	
	0	0	……	0	2	(Total:alpha_2 rows)
……	……	……	……	……	……	
n = N	N_1	N_2	……	N_K-1	N_K	1 row

For $n=1$ case, it easy to logically find out the possible number of cases (α_1) = K. Each case or possibility (i.e. each row in the above table) can occur with certain probability. So, the next task is to calculate those probabilities. Assume the number of losses from the policy, $n \sim$ Bin (N, p) and we know 'p'. Then, one can easily compute binomial $P(n = n_0)$'s. There are several combinations in which n_0 losses can occur and these probabilities can be computed using Multinomial distribution. So, the final probability formula for α_{n0} possible combinations is:

$$\text{Binomial } P(n = n_0) \times \text{Multinomial } P\left(n_1 = n_{1,0}, n_2 = n_{2,0}, \ldots\ldots, n_K = n_{K,0} \;\middle|\; \sum_{k=1}^{K} n_{k,0} = n_0\right)$$

In the most complex scenario (refer the last row of Table1), to get the convoluted distribution of aggregate claim, one has to add X_1 for N_1 times, X_2 for N_2 times and so on. Assume, $X_i \sim \text{LogN}(\mu_i, \sigma^2_i)$ with pdfs, $p_{X\,i}(x)$, $\forall$ i = 1,K and all are independent. Closed-form expressions for the pdf or cdf of the lognormal sum (ΣX) are not available. However, the lognormal sum can be well approximated by different methods like *Fenton-Wilkinson* (F-W) method or *Schwartz-Yeh* method. The first one uses the method of moments (the 3rd method of convolution, mentioned before) with some modification. It also made a crucial assumption: Sum of lognormal is lognormal, though it cannot be proved mathematically. Fenton and Wilkinson suggested a transformation of the original lognormal data produces good result in approximating the convoluted lognormal distribution. First, develop a new normal variable Y as $2.306 \times \log_{10}X$. Find the parameters of normal distribution -- μ_Y and σ^2_Y. Note, they are suggesting a transformation in place of taking natural logarithm to convert lognormal data to normal. Then take exponential of that new variable Y to produce 'transformed' lognormal variable. The 'original' and 'transformed' lognormal data will not vary much (We have checked this on some simulated lognormal data). Now, calculate the mean and variance of the 'transformed' lognormal from μ_Y and σ^2_Y. From now on, we will call our 'transformed' lognormal as X. F-W method computes lognormal parameters - μ_S and σ^2_S by exactly matching the first and second central moments (mean and variance, respectively) of S, aggregate claim size and Σ X. Mathematically speaking, to sum K distributions one has to solve the following equations to solve parameters - μ_S and σ^2_S:

$$\int_0^\infty x p_X(x)dx = \sum_{i=1}^{K} \int_0^\infty x p_{X_i}(x)dx, \qquad \text{(1a)}$$

$$\int_0^\infty (x-\mu_X)^2 p_X(x)dx = \sum_{i=1}^{K} \int_0^\infty (x-\mu_{X_i})^2 p_{X_i}(x)dx, \quad \text{(1b)}$$

Mean (/variance) of ΣX can be written as sum of mean (/variance) because X_i's are independent. For α number of convolution possibilities, parameters - μ_S and σ^2_S need to be solved using eqs (1a-b). The 'final' convoluted distribution is the weighted combination of these α+1-convoluted distributions (remember '0' aggregate loss) of S = Σ X, weights being the binomial×multinomial probabilities (for each of α possibilities) derived before. For S=0, in place of convoluted lognormal cdf we used a probability mass P(S=0) =1 multiplied by weight: binomial probability P(n = 0).

We applied the theory explained above on an imaginary data set. We also cross-validated the F-W result. For that, we used the numeric method of convolution – Monte Carlo Simulation. We calculated cumulative probabilities of α+1 possible cases and accordingly assigned ranges in [0,1] to all α+1 cases. We drew random numbers from U[0,1] several times to choose cases from α+1 possibilities. For α cases (where $n \geq 1$), random number is drawn from 'original' lognormal distributions and then their sum is taken. In the $(\alpha+1)^{st}$ case ('0' aggregate loss) we took zero values. In this way we generated several values of S. Hence one can derive the histogram of S and empirical cdf from that. As you increase the number of simulation, one can get smoother empirical cdf. If the parametric lognormal cdf, provided by F-W method, is close enough to the empirical cdf of S, then one can say that F-W approximation is doing good. We have done the same and below graph suggests that F-W approximation is indeed very good.

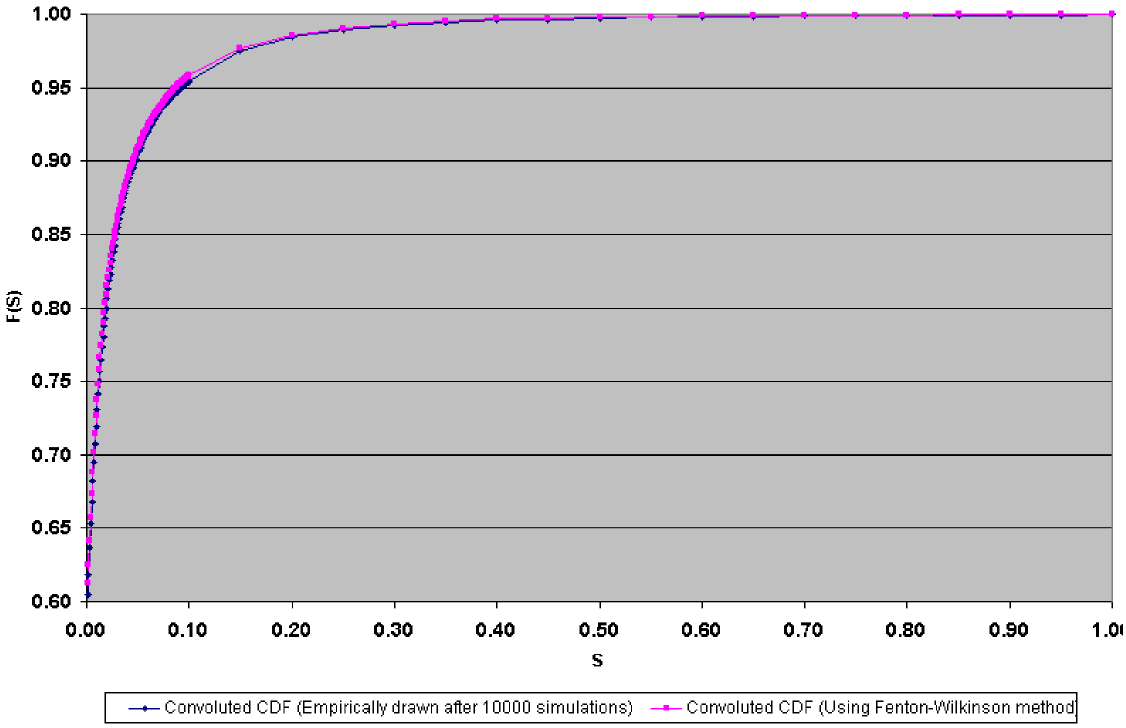

Figure 2: Comparison of two methods of lognormal convolution – F-W and Simulation

5. Conclusion:

Convolution is necessary for better understanding of the aggregate loss distribution. Though any closed form of convolution does not exist for lognormal -- one of the most popular loss distributions in insurance, Fenton-Wilkinson approximation is there for lognormal sum. But F-W method has certain limitations as well. The approximation breaks down if variance of the distribution in use is higher than certain limit. Researchers have also noticed that F-W method fails to give good result for both the tails. Generally, it does not give good result for the lower tail. But still, its simplicity makes it a popular method of lognormal convolution in the field of engineering.

***

References:

1) C.H. von Lanzenauer & W.N. Lundberg, "The n-fold Convolution of a mixed density and mass function" (available in www.casact.org)

2) L.F. Fenton, "The Sum of Lognormal Probability Distributions in scatter transmission systems", IRE Trans. Comm. Vol. 8, pp. 57-67, March 1960

3) J Wu, N. Mehta & J Zhang, "A flexible Lognormal Sum Approximation Method", Mitsubishi Electric Research Laboratories, December 2005 (available in internet)

Acknowledgement:

I would like to thank Sandeep Basak and Biresh Giri, AASI for their valuable assistance. Sandeep was instrumental in the development of this paper. Biresh's thorough editorial review contributed several improvements to the paper's content and style.