International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Modeling Insurance Claims using a Compound Distribution

Roslina Magdalena Lumbanbatu[1], Fronita Girsang[2], Ita Yapulina Br Surbakti[3]

Department of Mathematics, University of North Sumatera, Indonesia

Abstract: ***A typical model for insurance risk, the so-called collective risk model, treats the aggregate loss as having a compound distribution with two main components: one characterizing the arrival of claims and another describing the severity (or size) of loss resulting from the occurrence of a claim. We review a collection of loss distributions and present methods that can be used to assess the goodness-of-fit of the claim size distribution. The collective risk model is often used in health insurance and in general insurance, whenever the main risk components are the number of insurance claims and the amount of the claims. It can also be used for modeling other non-insurance product risks, such as credit and operational risk.***

Keywords: Insurance risk model; Loss distribution; Claim arrival process

1. Introduction

A loss model or actuarial risk model is a parsimonious mathematical description of the behavior of a collection of risks constituting an insurance portfolio. It is not intended to replace sound actuarial judgment. In fact, according to Willmot (2001), a well formulated model is consistent with and adds to intuition, but cannot and should not replace experience and insight. Moreover, a properly constructed loss model should reflect a balance between simplicity and conformity to the data since overly complex models may be too complicated to be useful.

A typical model for insurance risk, the so-called collective risk model, treats the aggregate loss as having a compound distribution with two main components: one characterizing the frequency (or incidence) of events and another describing the severity (or size or amount) of gain or loss resulting from the occurrence of an event (Kaas et al., 2008; Klugman, Panjer, and Willmot, 2008; Tse, 2009). The stochastic nature of both components is a fundamental assumption of a realistic risk model. In classical form it is defined as follows. If $\{N_t\}_{t>0}$ is a process *counting* claim occurrences and $\{X_k\}, 1<k<\infty$ is an independent sequence of positive independent and identically distributed (i.i.d.) random variables representing claim sizes, then the *risk process* $\{R_t\}_{t>0}$ is given by

$$R_t = u + c(t) - \sum_{i=1}^{N_t} X_i.$$

The non-negative constant u stands for the initial capital of the insurance company and the deterministic or stochastic function of time c(t) for the premium from sold insurance policies. The sum $\{\sum_{i=1}^{N_t} X_i\}$ is the so-called *aggregate claim process*, with the number of claims in the interval (0, t] being modeled by the *counting process* Nt. Recall, that the latter is defined as $N_t = \max\{n : \sum_{i=1}^{n} W_i \le t\}$, where $\{W_i\}_{i=0}^{\infty}$ is a sequence of positive ran- dom variables and $\sum_{i=1}^{u} W_i \equiv 0$. In the insurance risk context Nt is also referred to as the *claim arrival process*.

The collective risk model is often used in health insurance and in general insurance, whenever the main risk components are the number of insurance claims and the amount of the claims. It can also be used for modeling other non- insurance product risks, such as credit and operational risk (Chernobai, Rachev, and Fabozzi, 2007; Panjer, 2006). In the former, for example, the main risk components are the number of credit events (either defaults or downgrades), and the amount lost as a result of the credit event.

A. Claim Arrival Processes

In this section we focus on efficient simulation of the claim arrival process {Nt}. This process can be simulated either via the arrival times {Ti}, i.e. moments when the i-th claim occurs, or the inter-arrival times (or waiting times) $W_i = T_i - T_{i-1}$, i.e. the time periods between successive claims (Burnecki and Weron, 2005). Note that in terms of W_i's the claim arrival process is given by

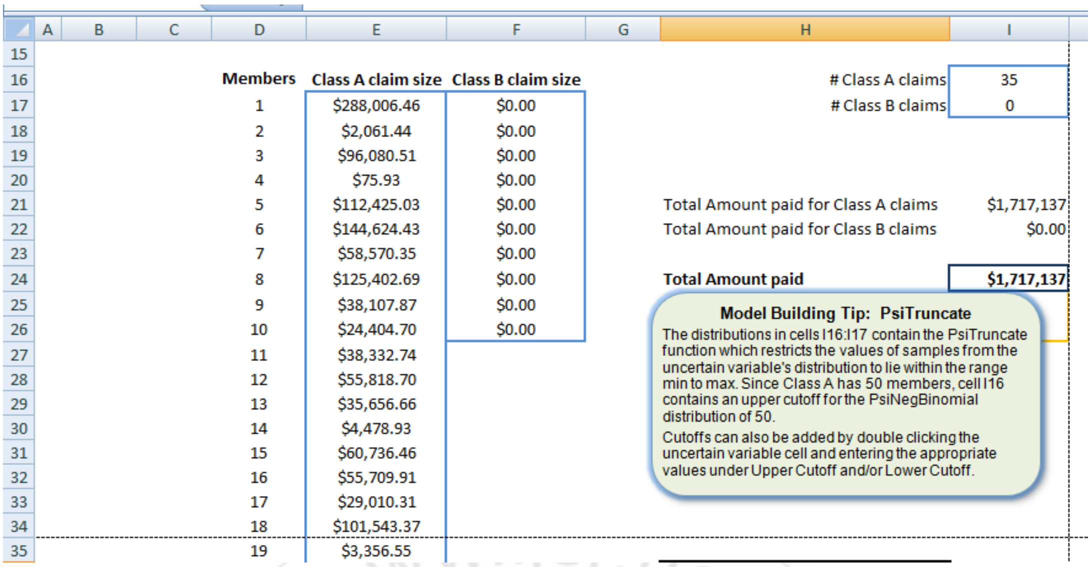
$$N_t = \sum_{n=1}^{\infty} I(T_n \le t).$$

In what follows we discuss four examples of {Nt}, namely Negative Binomial Distribution and Geometric Distribution.

B. Simulation Risk Process

An insurance policy has two basic classes of insured members, class A and class B; there are 50 class A members and 10 class B members. The number of claims from class A members follows a Negative Binomial distribution with parameters (5, 0.25), and the number of claims from class B members follows a geometric distribution with parameter 0.25. The size of a claim from class A members follows an exponential distribution with mean $40,000, and the claim size from class B members is distributed with a Pearson 5 distribution with parameters (10, 2000000).

We would like to characterize the total claims distribution in this case. We model the claim sizes using *PsiExponential* and *PsiPearson5* distributions in cells E17:E66, and F17:F26. The number of claims is also uncertain, and this is modeled using the PsiNegBinomial and Psi*Geometric* distributions in cells I16 and I17. The amount paid in claims depends on the number of claims, and this is computed in cells I21:I22. The overall amount paid is in cell I24, and the Expected total amount paid is computed in cell I25. The maximum probable loss at the 99th percentile is computed in cell I26, using the *PsiPercentile* function.



	A	B	C	D	E	F	G	H	I
15									
16				Members	Class A claim size	Class B claim size		# Class A claims	35
17				1	$288,006.46	$0.00		# Class B claims	0
18				2	$2,061.44	$0.00			
19				3	$96,080.51	$0.00			
20				4	$75.93	$0.00			
21				5	$112,425.03	$0.00		Total Amount paid for Class A claims	$1,717,137
22				6	$144,624.43	$0.00		Total Amount paid for Class B claims	$0.00
23				7	$58,570.35	$0.00			
24				8	$125,402.69	$0.00		Total Amount paid	$1,717,137
25				9	$38,107.87	$0.00			
26				10	$24,404.70	$0.00			
27				11	$38,332.74				
28				12	$55,818.70				
29				13	$35,656.66				
30				14	$4,478.93				
31				15	$60,736.46				
32				16	$55,709.91				
33				17	$29,010.31				
34				18	$101,543.37				
35				19	$3,356.55				

2. Conclusion

The advantage of the mixture model is some possible mathematical tractability. The unobservable variable in the compound distribution model has the advantage of providing a natural interpretation to the resulting model. It also can be interpreted as a way to model the presence of heterogeneity in the insurance risk portfolio.

References

[1] K. Burnecki and R. Weron (2005). Modeling of the Risk Process, *in* P. Cizek, W. H¨ardle, R. Weron (eds.) *Statistical Tools for Finance and Insurance*, Springer-Verlag, Berlin, 319–339.
[2] Chernobai, A.S., Rachev, S.T., and Fabozzi, F.J. (2007). *Operational Risk: A Guide to Basel II Capital Requirements, Models, and Analysis*, Wiley.
[3] Kaas, R., Goovaerts, M., Dhaene, J., and Denuit, M. (2008). *Modern Actuarial Risk Theory: Using R*, Springer.
[4] Klugman, S. A., Panjer, H.H., and Willmot, G.E. (2008). *Loss Models: From Data to Decisions* (3rd ed.), Wiley, New York.
[5] Panjer, H.H. (2006). *Operational Risk : Modeling Analytics*, Wiley.
[6] Tse, Y.-K. (2009). *Nonlife Actuarial Models: Theory, Methods and Evaluation*, Cambridge University Press, Cambridge.
[7] Willmot, G. E. (2001). The nature of modelling insurance losses, *The Munich Re Inaugural Lecture*, December 5, 2001, Toronto.