Akademeia (2014) 4(1): ea0123

Applied Science
ARTICLE

Compound Poisson approximation in an individual risk model and approximation bounds for a real portfolio in Iran

Maryam Teimourian Sefideh Khan

Department of Mathematics and Statistics, Roudehen Branch, Islamic Azad University, Roudehen, Iran

The distribution function of the sum of random variables plays an important role in the management of portfolios and consequently in actuarial science. However it is generally problematic and so some approximations have been introduced. Though all approximation s show error, one of the best distribution functions for approximation is the Compound Poisson distribution function. In order to improve the distribution approximation of the sum of the random variables, finding the upper bound is of crucial importance. My focus is mostly on upper bounds of approximation that have no restriction on distribution of random variables, and the results are in a better order than bounds previously reported. We review these bounds and compare them. In the end, we applied them to a real portfolio.

KEYWORDS: Compound Poisson approximation, individual risk model, magic factor, total variation distance

Two models exist for aggregate risk in a certain portfolio: Individual risk models and collective risk models. Our focus will be on the individual risk model and as in non-life insurance mathematics, the "time" aspect will be ignored. In an individual risk model, each portfolio contract is important. In addition each of these contracts has its own allocated possibility. The total claim in this model is equal to the sum of the claims of contracts in the portfolio:

$$S_n = X_1 + X_2 + \dots + X_n \qquad (1)$$

The random variable X_i is the claim of i-th contract with cumulative distribution function (CDF) F_i. The individual risk model is used in portfolios with the fixed number of contracts. We want to find the model of, for instance, the aggregate risk in group insurances whose number of personnel is fixed and specific. The aim is computing the distribution function in individual risk model. In the first method the distribution function of random variables S_n is used to compute:

$$P(S_n \le s) = P(X_1 + X_2 + \dots + X_n \le s) \qquad (2)$$

and if the random variables, $X_i, i \in \{1,2,\dots,n\}$, are independent then **equation 2** is equal to the convolution of $F_{X_i}(s)$:

$$P(S_n \le s) = F_{X_1}(s) * F_{X_2}(s) * \dots * F_{X_n}(s) \qquad (3)$$

Even when the random of variable X_i has the Bernoulli distribution, the computation of the distribution function (df) of S_n is not feasible when "n" is large. Another method is using moment

Correspondence: Teimourian.m@riau.ac.ir
Received: 19 November 2013; **Accepted:** 2 February 2014

generating functions. The moment generating function of S_n by assuming independent of the random variables X_i is equal to:

$$M_{S_n}(t) = E(e^{S_n t}) = M_{X_1}(t) * M_{X_2}(t) * \dots * M_{X_n}(t) \quad (4)$$

Only in very specific cases may M_{S_n} be identified by the moment generating function of some known distribution, and the problem thus solved. Therefore, finding the distribution function of S_n in individual risk model, particular for a big portfolio is impossible so that approximation is used to find the total claim model. For approximation Poisson distribution and compound Poisson distribution are often used. For more information refer to [1].

Obviously, each approximation has its errors. Our aim is to minimize this error. One of the methods to compute approximation error is the total variation distance measure which is used to find an upper bound for the approximation. The definition of the total variation distance for two random variables, X and Y, is as follows:

Let X and Y be random variables in the probability space (Ω, F, P) and let B denote the σ -algebra over R. The total variation distance between these random variables is equal to:

$$d_{TV}(X,Y) = \sup_{A \in B} |P(X \in A) - P(Y \in A)| \quad (5)$$

If X and Y have CDF F_X and Q_Y respectively, then the total variation distance between them is denoted by $d_{TV}(F_X, Q_Y)$. The total variation distance has many properties which we mention, some in the following:

- $d_{TV} \le 1$
- $d_{TV}(X,Y) \le p(X \ne Y)$
- $d_{TV}(X,Y) = \frac{1}{2}\sum_k |p(X=k) - p(Y=k)|$
- If $\{X_1, \dots, X_n\}$ and $\{Y_1, \dots, Y_n\}$ are the sequences of discrete and independent random variables then

  $$d_{TV}(\sum_i X_i, \sum_i Y_i) \le \sum_i d_{TV}(X_i, Y_i)$$
- If F^{*m} and Q^m are m-fold convolutions of F and Q respectively, then

  $$d_{TV}(F^{*m}, Q^{*m}) \le m d_{TV}(Q, F)$$
- Let $\{a_k\}, \{b_k\}$ be numeric sequences in $[0,1]$ whenever $\sum_k a_k = \sum_k b_k$ and if ξ_k is the Dirac measure in k-th point then

  $$d_{TV}(\sum_k a_k F^{*k}, \sum_k b_k F^{*k}) \le d_{TV}(\sum_k a_k \xi_k, \sum_k b_k \xi_k)$$

In this paper, we review the literature available on upper bounds for the approximation of $\ell(S_n)$ and compare them. In section 2 we review some bounds which have been suggested in the literature and compare them. In section 3, we present a real portfolio in one Iranian insurance company and compute the bounds for it.

Literature review: Basic inequalities

At first, the Poisson approximation can be used for approximating the distribution function of total claim in a portfolio. Kintchine and Deoblin [2, 3] proved that

$$d_{TV}(\ell(S_n), Poisson) \le \sum p_i^2 =: \lambda_2 \quad (6)$$

where the claims are Bernoulli random variables with success probability p_i and $\ell(S_n)$ is the df of the total claim, S_n.

If portfolio is big or n is large, then **equation 6** may exceed one. LeCam [4] improved this upper bound for approximation when the claims are Bernoulli random variables with success probability p_i with the restriction of the $p_0 = max_{i \in \{1,\dots,n\}} \le \frac{1}{4}$, as

$$d_{TV}(\ell(S_n), Poisson) \le \frac{8\lambda_2}{\lambda} \quad (7)$$

He also proved [5] that

$$C_1 \min\{\lambda_2, 1\} \le d_{TV}(\ell(S_n), Poisson) \le C_2 \min\{\lambda_2, 1\} \quad (8)$$

where C_1 and C_2 denote positive absolute constants. Subsequently many approximation bounds for Poisson approximation with total claim were presented. Among them, Stein [6] used a different method for approximation. This method is known as the “Stein method”. It gives the inequalities

$$\frac{\lambda_2}{32}\min\{\frac{1}{\lambda}, 1\} \le d_{TV}(\ell(S_n), Poisson) \le \lambda_2 \min\{\frac{1}{\lambda}, 1\} \quad (9)$$

It can be seen that $\lambda_2 \min(\frac{1}{\lambda}, 1) \ge \left(\frac{\lambda_2}{\lambda}\right)^2$, therefore d_{TV} is small if $\frac{\lambda_2}{\lambda}$ is small and the upper bound of **equation 9** is better than that of **equation 6** by the order of λ_2. The factor $\frac{1}{\lambda}$ is named the magic factor because it is an important factor in improving the upper bound approximation of d_{TV}.

Roos [7] used Kerstain's method and found an upper bound approximation with a better order for the total claim when the claims are multivariate $X_i = (X_{i1}, \dots, X_{il})$. In his method

$$P(X_i = e_r) = p_{i,r}$$

$$P(X_i = 0) = 1 - \sum_{r=1}^{k} p_{i,r}$$

$$i \in \{1,2,\dots,n\}, r \in \{1,\dots,k\}$$

where the vector e_r has the value one in the r-th component and zero in other components, and $\lambda(r) = \sum_{i=1}^{n} p_{i,r}$ is the mean of r-th component of S_n . The aim is to find the approximation error for approximate $\ell(S_n)$ and multivariate Poisson distribution, $P(\lambda)$ with independent component, and mean vector $\lambda = (\lambda_{(1)}, \dots, \lambda_{(k)})$. The probability distribution function for multivariate Poisson distribution is

$$P(\lambda)\{m\} = \prod_{r=1}^{k} \frac{e^{-\lambda(r)} \lambda_{(r)}^{m_r}}{m_r!}, m = (m_1, \dots, m_k)$$

Define

$$g(x) = 2\sum_{s=2}^{\infty} x^{s-2} \frac{s-1}{s!}$$

$$\alpha = \sum_{i=1}^{n} g(2\sum_{r=1}^{k} p_{i,r}) \min\{2^{-3/2} \sum_{r=1}^{k} \frac{p_{i,r}^2}{\lambda(r)}, (\sum_{r=1}^{k} p_{i,r})^2\}$$

$$\beta = \sum_{i=1}^{n} \min\{\sum_{r=1}^{k} \frac{p_{i,r}^2}{\lambda(r)}, (\sum_{r=1}^{k} p_{i,r})^2\}$$

then:

$$d_{TV}(\ell(S_n), P(\lambda)) \leq 8.8\beta \qquad (10)$$

and if $\alpha < \frac{1}{2e}$ then

$$d_{TV}\big(\ell(S_n), P(\lambda)\big) \leq \frac{\alpha}{1-2\alpha e} \qquad (11)$$

Upper bounds under more assumptions

In all the upper bounds computed so far, the Poisson distribution is used to approximate $\ell(S_n)$, but the compound Poisson approximation is an important distribution that we need to find. In many situations, the number of claims is not fixed and the claim for every contract has a specific distribution and distributions in all of the claims in a portfolio are not identical. Suppose, for instance, a portfolio contains "n" contracts with claims $X_1, \dots, X_n$. We denoted the probability of the event that i-th claim is nonzero by $p_i = P(X \neq 0)$, and by Q_i the conditional distribution of X_i given that it occurs. The compound Poisson approximation with

$$\lambda = \sum_{i=1}^{n} p_i$$

and probability distribution function

$$CPo(\lambda, Q) = \sum_{k=0}^{\infty} e^{-k} \frac{\lambda^k}{k!} Q^{*k}$$

when

$$Q = \frac{1}{\lambda} \sum_{i=1}^{n} p_i Q_i$$

Q^{*k} is a k-fold convolution of Q with itself. In order to obtain upper bound with a better order than λ_2, Roos [8] used suitable assumptions on $Q_1, \dots, Q_n$. Suppose that

$$Q_i, i \in \{1, \dots, n\}$$

can be decomposed in the form

$$Q_i = \sum_{i=1}^{\infty} p_{i,r} U_r$$

Suppose $p_{i,r} \in [0,1]$ with $\sum_{r=1}^{\infty} p_{i,r} = 1$ and a sequence of probability measures $U_1, U_2, \dots$ on (R^l, B^l) which are not allowed to depend on i. Let

$$q_r = \frac{1}{\lambda} \sum_{i=1}^{n} p_i p_{i,r}$$

$$v_i = \sum_{r=1}^{\infty} \frac{p_{i,r}^2}{q_r}; i \in \{1,2,\dots,n\}$$

and also

$$\alpha_1(x) = \sum_{i=1}^{n} g(2p_i) p_i^2 \min\{\frac{x v_i}{\lambda}, 1\}$$

$$\beta_1 = \sum_{i=1}^{n} p_i^2 \min\{\frac{v_i}{\lambda}, 1\}$$

He proved that

$$d_{TV}(\ell(S_n), CPo(\lambda, Q)) \leq 8.8\beta_1 \quad (12)$$

and if $\alpha_1\left(2^{-\frac{3}{2}}\right) < \frac{1}{2^e}$ then

$$d_{TV}(\ell(S_n), CPo(\lambda, Q)) \leq \frac{\alpha_1\left(2^{-\frac{3}{2}}\right)}{1-2e\alpha_1(2^{-\frac{3}{2}})} \quad (13)$$

These upper bounds are not specifically used for the sum of Bernoulli random variables but for all distributions of the claims. If the distribution of random variables is discrete then decomposition assumption for the random variables is satisfied. But, in applications, the distribution of the claims is often continuous and decomposition condition may not be satisfied.

In continuous case we suppose that i-th claim is an L-dimension random vector. Therefore, for every contract in a portfolio an L-dimension claim $X_i = \left(X_{i_1}, \dots, X_{i_n}\right), i \in \{1, \dots, n\}$ is made to show the claims that are independent but not identical. If we define

$$Q = \frac{1}{\lambda}\sum_{i=1}^{n} Q_i p_i$$

then Q_i is absolutely continuous with respect to Q. thus using Random-Nikodym Theorem, Q_i has a density with respect to Q. We denote this density function by f_i. Roos [9] proved that if

$$\tilde{\beta} = \sum_{i=1}^{n} p_i^2 \min\{\frac{1}{\lambda}\int f_i^2 dQ, 1\}$$

and

$$\tilde{\alpha}(x) = \sum_{i=1}^{n} g(2p_i)p_i^2 \min\{\frac{x}{\lambda}\int f_i^2 dQ, 1\}, x \in [0, \infty)$$

then we will have

$$d_{TV}(\ell(S_n), CPo(\lambda, Q)) \leq 8.8\tilde{\beta} \quad (14)$$

and if $\tilde{\alpha}\left(2^{-\frac{3}{2}}\right) < \frac{1}{2^e}$ then

$$d_{TV}(\ell(S_n), CPo(\lambda, Q)) \leq \frac{\tilde{\alpha}\left(2^{-\frac{3}{2}}\right)}{1-2e\tilde{\alpha}\left(2^{-\frac{3}{2}}\right)} \quad (15)$$

It is obvious that if $l = 1$ then this model is the same as one dimensional individual risk model. Suppose the existence a constant C as

$$\min_i\{\int f_i^2 dQ\} < C$$

If

$$\tilde{\beta} = \sum_i p_i^2 \min\{\frac{1}{\lambda}\int f_i^2 dQ, 1\} \leq C\frac{\lambda_2}{\lambda},$$

then the upper bound in **equation 14** is equal to $\frac{\lambda_2}{\lambda}$ but if $\tilde{\beta} \leq \lambda_2$, then this will be of order λ_2. Roof [9] explains “In the literature the additional factor $\frac{1}{\lambda}$ is sometimes called a magic factor, since, on the one hand, it is highly desirable, but on the other hand, the proof of its existence turns out to be difficult.”

Then these bound has the same order as that of Kintchine and Deoblin but their proof was for the Bernoulli case. One can simply prove that the above-mentioned upper bound includes previous bounds that specify the Bernoulli claims. If, for instance,

$$\nu_i =: \frac{\sum_{r=1}^{\infty}}{q_r}$$

then

$$\sum_i p_i^2 \int f_i^2 dQ \leq \sum_i p_i^2 \nu_i$$

then **equations 14 and 15** have a better order than **equations 12 and 13**. On the other hand if $Q_1 \cong \cdots \cong Q_2$ then $\int f_i^2 dQ \cong 1$ and $\tilde{\beta} = \lambda_2 min\{\frac{1}{\lambda}, 1\}$. Then the upper bounds of **equation 14** have the same order as the upper bound as in **equation 9** which is proved when the claims are the Dirac measure.

A case study

In this section, we compute the bounds above for a portfolio of property insurance containing auto insurance and cargo insurance and fire insurance of Asia Insurance Company. For auto insurance, we only focus on compulsory third party insurance and body auto insurance. Also, we mean every kind of vehicle, bus, truck, train, tractor and the like. In body auto insurance all kinds of damage to the insured vehicle, such as accident, fire or theft, are considered.

Finding the claims model

To find the claims models of auto insurance in this portfolio, we select a sample size 121 of auto insurance contracts with damages in the previous year and find the proper distribution. These claims have lognormal distribution with density

$$h_1(x) = \frac{1}{1.4\sqrt{2\pi x}}\exp(\frac{-(\log(x) - \log(4664727))^2}{2 * (1.4)^2})$$

To find the model for the claims of passenger accident insurance in this portfolio, we select a sample size 292 of passenger accident insurance contracts with damages in the previous year and find the proper distribution. These claims have a lognormal distribution with density

$$h_2(x) = \frac{1}{0.98\sqrt{2\pi x}}\exp(\frac{-(\log(x) - \log(5512008))^2}{2 * (1.34)^2})$$

To find the claims model of cargo insurance in this portfolio, we select a sample size 109 of cargo insurance contracts with damages in the previous year and find the proper distribution. These claims have lognormal distribution with density

$$h_3(x) = \frac{1}{1.76\sqrt{2\pi x}}\exp(\frac{-(\log(x) - \log(9479379))^2}{2 * (1.76)^2})$$

To find the claims model of fire insurance in this portfolio, we select a sample size 60 of fire insurance contracts with damages in the previous year and finding the proper distribution. These claims have lognormal distribution with density

$$h_4(x) = \frac{1}{2.03\sqrt{2\pi x}}\exp(\frac{-(\log(x) - \log(9505642))^2}{2 * (2.03)^2})$$

The aim of this section is to compute upper bound for this approximation and to show how to minimize them. To compute upper bound approximation of the distribution of claims of this portfolio, we must know the probability of damage occurrence in every contract. We used empirical method to estimate the probability of damage occurrence.

Using the previous data recorded in Asia Insurance Company in the last 9 years, we estimated the probability of occurrence of claims of body auto insurance as $p_1 = 0.3$, the probability of occurrence of claims compulsory third party insurance as $p_2 = 0.71$, the probability of occurrence of claims fire insurance as $p_3 = 0.026$, the probability of occurrence of claims of cargo insurance as $p_4 = 0.019$. We consider 3 cases, because of possible underestimation and overestimation.

First case

In this case, we compute upper bound with probabilities that are estimated by empirical method. Thus we have $p_1 = 0.3, p_2 = 0.71, p_3 = 0.026$, $p_4 = 0.019$, $\lambda = 10.75, \lambda_2 = 2.42$. These results are summarized in **table 1**.

Second case

In this case, we compute upper bound with probabilities that are estimated by empirical method but we consider the possibility of underestimation. Thus we have $p_1 = 0.35, p_2 = 0.08, p_3 = 0.03$, $p_4 = 0.02$, $\lambda = 12.4, \lambda_2 = 3.287$. These results are summarized in **table 2**.

Third case

In this case, we compute upper bound with probabilities that are estimated by empirical method while considering overestimation. Thus we have $p_1 = 0.25, p_2 = 0.065, p_3 = 0.02$, $p_4 = 0.01$, $\lambda = 8.95, \lambda_2 = 1.70175$. These reultss are summarized in **table 3**. The value of the upper bound for this portfolio is reported in **table 4**.

As shown in Table 4, we can conclude that

- as mentioned before, the Kintchine and Deoblin upper bound is useless for a big portfolio. There are 100 contracts in our portfolio and column 6 is more than 1.
- The bound derived from **equation 15** is better than of **equation 14**, because it has a fixed constant. On the other hand, **equation 14** is more appropriate on the grounds of the absence of a singularity.
- The bound of **equation 14** has order $\frac{\lambda_2}{\lambda}$. Thus if this ratio is small, then our bounds can have a better order. In this study this ratio in the third case is smaller than other cases and the upper bounds in this situation are better than others.

Conclusion

Insurance companies need the distribution function of total claim of every portfolio to estimate the total claim of that portfolio to compute items such as the premium of each contract. Thus finding df of sum of random variables in is very important in actuarial science. If the claim of i-th contract is denoted by X_i which is a random variable, then attempt is made to find the distribution of

$$S_n = \sum_{i=1}^{n} X_i$$

However finding the distribution function for big portfolios is impossible. Thus, it is approximated by an appropriate distribution. Poisson and compound

Poisson distributions provide suitable approximations for this purpose. We review the upper bounds of approximation which were found by compound Poisson distribution. These bounds resulted in a better order. These bounds are not possible to obtain without any restriction on the claim distributions. At the end, we compute these upper bounds for a real portfolio of Asia Insurance Company and compared them. In an actual case, the upper bounds with magic factor are better bounds for compound Poisson approximation in individual risk model.

Table 1. Results summary for the first case

Distribution	Probability	Claims
Lognormal(log(4664727),1.4)	$p_1 = 0.3$	$X_{\{1\}}, \dots, X_{\{25\}}$
Lognormal(log(5512008),1.34)	$p_2 = 0.71$	$X_{\{26\}}, \dots, X_{\{55\}}$
Lognormal(log(9505642),2.03)	$p_3 = 0.026$	$X_{\{56\}}, \dots, X_{\{75\}}$
Lognormal(log(9479379),1.76)	$p_4 = 0.019$	$X_{\{76\}}, \dots, X_{\{100\}}$

Table 2. Results summary for the second case

Distribution	Probability	Claims
Lognormal(log(4664727),1.4)	$p_1 = 0.35$	$X_{\{1\}}, \dots, X_{\{25\}}$
Lognormal(log(5512008),1.34)	$p_2 = 0.08$	$X_{\{26\}}, \dots, X_{\{55\}}$
Lognormal(log(9505642),2.03)	$p_3 = 0.03$	$X_{\{56\}}, \dots, X_{\{75\}}$
Lognormal(log(9479379),1.76)	$p_4 = 0.02$	$X_{\{76\}}, \dots, X_{\{100\}}$

Table 3. Results summary for the third case

Distribution	Probability	Claims
Lognormal(log(4664727),1.4)	$p_1 = 0.25$	$X_{\{1\}}, \dots, X_{\{25\}}$
Lognormal(log(5512008),1.34)	$p_2 = 0.065$	$X_{\{26\}}, \dots, X_{\{55\}}$
Lognormal(log(9505642),2.03)	$p_3 = 0.02$	$X_{\{56\}}, \dots, X_{\{75\}}$
Lognormal(log(9479379),1.76)	$p_4 = 0.01$	$X_{\{76\}}, \dots, X_{\{100\}}$

Table 4. General results summary

Case	$\frac{\lambda_2}{\lambda}$	Bound 14	Bound 15	Bound 16
1	0.2256	0.0721	0.0355	2.4237
2	0.2650	0.0899	0.0453	3.2850
3	0.1901	0.0071	0.00396	1.69

Works Cited

1. Kass, R., Govaerts, M., Dhaene, J., Denuit, M. (2001). *Modern actuarial risk theory.* Kluwer Academic Publishers.
2. Doeblin, W., Yu, X., Kucuk, I. (1939). *Sur les sommes d'un grand nombre de variables aléatoires indépendantes.* Bulletin des Sciences Mathématiques, Deuxième Série, (**63**): 23–32.
3. Khintchine, Y. (1933). *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung.* Springer, Berlin.
4. Le Cam, L. (1960). *An approximation theorem for the Poisson binomial distribution.* Pacific Journal of Mathematics, (**10**): 1181–1197.
5. Le Cam, L. (1965). *On the distribution of sums of independent random variables.* In: Neyman, J., Le Cam, L. (eds.), Bernoulli, Bayes, Laplace (Proceedings of the Bernoulli-Bayes-Laplace Seminar, Statistical Laboratory, University of California, Berkeley, 1963), Springer, New York: 179-202.
6. Stein, C. (1970). *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables.* Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, **2.**
7. Roos, B. (1999). *On the rate multivariate Poisson convergence.* Journal of Multivariate Analysis (**69**): 120-134.
8. Roos, B. (2003). *Kerstan's method for compound Poisson approximation.* Annals of Probability, (**31**): 1754–1771.
9. Roos, B. (2007). *On variational bounds in the compound Poisson approximation of the individual risk model.* Mathematics and Economics, (**40**): 403-414.