Available online at www.sciencedirect.com
SCIENCE DIRECT®

Insurance: Mathematics and Economics 36 (2005) 57–77

INSURANCE MATHEMATICS & ECONOMICS

www.elsevier.com/locate/econbase

The compound Poisson random variable's approximation to the individual risk model

Jingping Yang*, Shulin Zhou, Zhenyong Zhang

LMAM, School of Mathematical Sciences, Peking University, Beijing 100871, China

Received December 2003; received in revised form October 2004; accepted 14 October 2004

Abstract

In this paper we study approximating the total loss associated with the individual insurance risk model by a compound Poisson random variable. By minimizing the expectation of the absolute deviation of the compound Poisson random variable from the true total loss, we investigate not only the optimal compound Poisson random variable but also the numerical calculation of the approximation error. We also discuss the influence of the Poisson parameter on the approximation error.

JEL classification: C63; O29

2000 Subject and Insurance Branch Codes: IM20

Keywords: Compound Poisson approximation; Individual risk model; Wasserstein distance; Comonotonic

1. Introduction

In actuarial science, the approximation of the individual risk model by a compound Poisson model plays an important role. This is mainly due to the compound Poisson model's advantages in recursive calculation (Panjer, 1981), combination and decomposition (Panjer and Willmot, 1992, Chapter 6 or Kaas et al., 2001, Chapter 3). In the development on this topic, there are a lot of papers concerned with the approximation. Bühlmann et al. (1977) illustrated why a cautious insurer should prefer the compound Poisson model to the individual risk model in the sense of stop-loss order. Gerber (1979, Chapter 4) gave a description of the choice of the Poisson parameter and introduced two cases which are often used in the later discussions. Gerber (1984), Hipp (1985, 1986), Michel (1987), De Pril and Dhaene (1992), Sundt (1993) and Dhaene and Sundt (1997) investigated the error bounds for approximation in

* Corresponding author.
E-mail addresses: yangjp@math.pku.edu.cn (J. Yang), szhou@math.pku.edu.cn (S. Zhou).

0167-6687/$ – see front matter
doi:10.1016/j.insmatheco.2004.10.003

terms of distribution or stop-loss premium. Kaas et al. (1988b) discussed the approximation of the aggregate claims and the stop-loss premiums by approximating the aggregate claims by the sum of a compound Poisson random variable (r.v.) and another r.v. determined by stop-loss order. Kuon et al. (1993) studied the approximation quality when the portfolio keeps growing.

The former papers are mainly focused on approximation in the aggregate claims distribution and related functions, such as stop-loss premiums. However, in practice one initially concerns about the individual risk model. Thus one natural question arises: given the observation data from the individual risk model, how should one determine the r.v.'s in the corresponding compound Poisson model? From the practical viewpoint, the r.v.'s in the approximation model have to be determined from the observation data.

The aim of this paper is to develop a method for carrying out such an approximation. By minimizing the expectation of the absolute deviation of compound Poisson r.v.'s from the total loss associated with the individual risk model, we present an optimal approximation model. We also give a numerical method to evaluate the approximation error. Finally we discuss the influence of the Poisson parameter on the approximation error. Throughout this paper, it is always assumed that the individual risks are independent. We first consider the case that the individual risks are homogenous, then apply the homogenous results to approximate the heterogenous risk model.

Consider a portfolio containing n homogenous insurance risks. Let X_i denote the loss associated with the ith risk, $i = 1, 2, \ldots, n$. Assume that $X_1, X_2, \ldots, X_n$ are independent identically distributed with common distribution F, where the variance $\mathrm{Var}(X_1)$ is finite and $0 < F(0) < 1$. The total loss for the portfolio $\{X_1, X_2, \ldots, X_n\}$ can be written as

$$S_n = \sum_{i=1}^{n} X_i. \tag{1.1}$$

The number of claims for the portfolio is denoted as N_n, i.e.,

$$N_n = \#\{i : X_i > 0, i \le n\}.$$

For the technical reason, an auxiliary sequence of risks $\{X_{n+1}, X_{n+2}, \ldots, \}$ is introduced, where $X_1, X_2, \ldots$ are independent identically distributed. Denote

$$M_1 = \inf\{i \ge 1 : X_i > 0\},$$
$$M_n = \inf\{i > M_{n-1} : X_i > 0\}, \quad n \ge 2.$$

It is known that $P(M_i < \infty) = 1$. M_i is the index of the risk that a claim occurs and M_i is strictly increasing with respect to i. By using the auxiliary risks $\{X_{n+1}, X_{n+2}, \ldots\}$, we can define a claim sequence

$$Y_i = X_{M_i}, \quad i \ge 1,$$

which is a subsequence of $\{X_i, i \ge 1\}$ and is well-defined.

It is proved that $Y_i, i \ge 1$ are independent with common distribution function

$$F_Y(x) = \frac{F(x) - F(0)}{1 - F(0)}, \quad x \ge 0.$$

It can also be shown that N_n is Binomial(n, q) with parameter $q = 1 - F(0)$,

$$P(N_n = m) = C_n^m q^m (1-q)^{n-m}, \quad m = 0, 1, \ldots, n,$$

and N_n is independent of the sequence $Y_i, i \geq 1$. Here C_n^m denotes $\binom{n}{m}$. Based on the above method, the link between the homogenous individual risk model and the compound binomial model has been established. We refer to Li and Yang (2001) for the above theoretical results.

By the definition of $\{Y_i\}$, we know that S_n in (1.1) can be expressed as

$$S_n = \sum_{i=1}^{N_n} Y_i. \tag{1.2}$$

Each term Y_i of S_n in (1.2) corresponds to an actual claim while there are many terms X_i in (1.1) which may equal zero. Among these terms in (1.2), the individual claims Y_i, $i \geq 1$ are independent and identically distributed, N_n is the number of claims in the portfolio $\{X_1, X_2, \ldots, X_n\}$, and the independence between N_n and Y_i is fulfilled. Summing over the first N_n claims of the sequence $Y_i, i \geq 1$, we thus obtain the total loss S_n associated with the portfolio $\{X_1, X_2, \ldots, X_n\}$.

Now we consider the approximation model. It is natural to assume that the approximation model has the same claim sequence as the individual risk model. The number of claims $N(\theta)$, as a Poisson r.v., should be estimated from the observation data. Furthermore, the independence between $N(\theta)$ and $Y_i, i \geq 1$ is assumed. An approximation to S_n is defined by

$$S^* = \sum_{i=1}^{N(\theta)} Y_i. \tag{1.3}$$

In the above approximation, the main problem is to determine the Poisson r.v. $N(\theta)$.

Let $F_{\text{poi}(\theta)}$ denote Poisson distribution with mean θ, and $F_{\text{bin}(n,q)}$ denote binomial distribution with parameters (n, q), where $q = 1 - F(0)$. The family of all Poisson r.v.'s, which have common mean θ and are independent of the claim sequence $Y_i, i \geq 1$, is denoted as $R(F_{\text{poi}(\theta)})$. As mentioned above, the optimal Poisson r.v. $N_n^0(\theta) \in R(F_{\text{poi}(\theta)})$ is determined by the following minimizing principle:

$$H_n(\theta) =: E\left|S_n - \sum_{i=1}^{N_n^0(\theta)} Y_i\right| = \inf_{N(\theta)\in R(F_{\text{poi}(\theta)})} E|S_n - S^*|, \tag{1.4}$$

where S_n and S^* are given, respectively, in (1.2) and (1.3). Here $H_n(\theta)$ represents the approximation error.

We organize our paper in the following frame. In Sections 2–6 we investigate the homogenous individual risk models. We first concentrate on the case that the Poisson parameter θ is fixed, then study the influence of the Poisson parameter on the approximation error. In Section 2 we show the existence of the optimal Poisson r.v. $N_n^0(\theta)$, and give an explicit expression for $N_n^0(\theta)$. In Section 3 we prove some optimal properties of $N_n^0(\theta)$. In Section 4 we present a numerical method to evaluate the approximation error. In Section 5 we apply our approximation argument to the functions of the total loss. In Section 6 we demonstrate that $H_n(\theta)$ has a unique minimum point, and make a comparison between the minimum Poisson parameter and those parameters normally used in literatures. In Section 7 we apply our main results to the heterogenous individual risk model. In Appendix A we give the detailed proof of the main theorems in Section 6.

2. The existence of the optimal Poisson r.v.

Given the risk sequence $\{X_1, X_2, \ldots\}$, a r.v. U can be defined as below. Construct r.v.'s $U_m, m = 0, 1, \ldots, n$ such that $U_m, m = 0, 1, \ldots, n$ and $X_i, i \geq 1$ are independent. It is assumed that U_m is uniformly distributed over

$(F_{\mathrm{bin}(n,q)}(m-1)), F_{\mathrm{bin}(n,q)}(m)]$ with probability density function

$$\frac{1}{F_{\mathrm{bin}(n,q)}(m) - F_{\mathrm{bin}(n,q)}(m-1)}.$$

Then define

$$U = \sum_{m=0}^{n} U_m I_{\{N_n = m\}}.$$

Under the above construction, U_m, $m = 0, 1, 2, \ldots, n$ are independent of the number of claims N_n and the claim sequence Y_i, $i = 1, 2, \ldots$. Thus by the independence between N_n and Y_i, U is also independent of Y_i. Moreover,

$$P(U \le x | N_n = m) = \begin{cases} 1, & x \ge F_{\mathrm{bin}(n,q)}(m), \\ \dfrac{x - F_{\mathrm{bin}(n,q)}(m-1)}{F_{\mathrm{bin}(n,q)}(m) - F_{\mathrm{bin}(n,q)}(m-1)}, & F_{\mathrm{bin}(n,q)}(m-1) \le x < F_{\mathrm{bin}(n,q)}(m), \\ 0, & x < F_{\mathrm{bin}(n,q)}(m-1). \end{cases} \tag{2.1}$$

The inverse function $F^{-1}_{\mathrm{bin}(n,q)}(y)$ of $F_{\mathrm{bin}(n,q)}(y)$ is defined as

$$F^{-1}_{\mathrm{bin}(n,q)}(y) = \inf\{x : F_{\mathrm{bin}(n,q)}(x) \ge y\}, \quad y \in [0, 1]$$

and the inverse function $F^{-1}_{\mathrm{poi}(\theta)}(y)$ of $F_{\mathrm{poi}(\theta)}(y)$ is defined as

$$F^{-1}_{\mathrm{poi}(\theta)}(y) = \inf\{x : F_{\mathrm{poi}(\theta)}(x) \ge y\}.$$

Lemma 1. *The r.v. U is uniformly distributed over* $[0, 1]$, *and*

$$N_n = F^{-1}_{\mathrm{bin}(n,q)}(U).$$

Proof. For fixed $s \in (0, 1)$, denote $M = \sup\{k : F_{\mathrm{bin}(n,q)}(k) \le s, k = 0, 1, \ldots\}$. It follows from (2.1) that:

$$P(U \le s | N_n = M + 1) = \frac{s - F_{\mathrm{bin}(n,q)}(M)}{F_{\mathrm{bin}(n,q)}(M+1) - F_{\mathrm{bin}(n,q)}(M)}$$

and

$$P(U \le s | N_n = m) = 0, \quad m > M + 1; \qquad P(U \le s | N_n = m) = 1, \quad m \le M.$$

Thus

$$\begin{aligned} P(U \le s) &= \sum_{m=0}^{n} P(U \le s | N_n = m) P(N_n = m) \\ &= \sum_{m=M+1}^{n} P(U \le s | N_n = m) P(N_n = m) + \sum_{m=0}^{M} P(U \le s | N_n = m) P(N_n = m) \end{aligned}$$

$$= P(U \le s|N_n = M+1)P(N_n = M+1) + \sum_{m=0}^{M} P(N_n = m)$$

$$= \frac{s - F_{\mathrm{bin}(n,q)}(M)}{F_{\mathrm{bin}(n,q)}(M+1) - F_{\mathrm{bin}(n,q)}(M)} \times \{F_{\mathrm{bin}(n,q)}(M+1) - F_{\mathrm{bin}(n,q)}(M)\} + F_{\mathrm{bin}(n,q)}(M) = s.$$

That is, the r.v. U is uniformly distributed over [0, 1].

Next we prove that $N_n = F^{-1}_{\mathrm{bin}(n,q)}(U)$ holds. Fix non-negative integer m, $0 \le m \le n$. Suppose that $N_n = m$. Thus $F_{\mathrm{bin}(n,q)}(m-1) < U \le F_{\mathrm{bin}(n,q)}(m)$ follows. Equivalently, $F^{-1}_{\mathrm{bin}(n,q)}(U) = m$. Hence $N_n = F^{-1}_{\mathrm{bin}(n,q)}(U) = m$. The lemma is proved. □

In the following theorem we prove the existence of the optimal Poisson r.v. $N_n^0(\theta)$, and then give a mathematical expression for the corresponding approximation error.

Theorem 1. *The Poisson r.v. $N_n^0(\theta)$ defined by $N_n^0(\theta) = F^{-1}_{\mathrm{poi}(\theta)}(U)$ satisfies (1.4), i.e., $F^{-1}_{\mathrm{poi}(\theta)}(U)$ is a possible choice for the optimal Poisson r.v. with mean θ. Moreover, the approximation error $H_n(\theta)$ satisfies*

$$H_n(\theta) = E(Y_1)E|F^{-1}_{\mathrm{bin}(n,q)}(U) - F^{-1}_{\mathrm{poi}(\theta)}(U)| = E(Y_1)\int_0^{\infty} |F_{\mathrm{bin}(n,q)}(x) - F_{\mathrm{poi}(\theta)}(x)|\,\mathrm{d}x. \tag{2.2}$$

Proof. Let $R(F_{\mathrm{bin}(n,q)}, F_{\mathrm{poi}(\theta)})$ denote the family containing all random vectors (W, V), where W has distribution $F_{\mathrm{bin}(n,q)}$ and V has distribution $F_{\mathrm{poi}(\theta)}$, and (W, V) is independent of the claim sequence Y_i, $i \ge 1$. Using the independence relation between Y_i, $i \ge 1$ and $R(F_{\mathrm{bin}(n,q)}, F_{\mathrm{poi}(\theta)})$, we find that

$$\inf_{(W,V)\in R(F_{\mathrm{bin}(n,q)},F_{\mathrm{poi}(\theta)})} E\left|\sum_{i=1}^{W} Y_i - \sum_{i=1}^{V} Y_i\right| = \inf_{(W,V)\in R(F_{\mathrm{bin}(n,q)},F_{\mathrm{poi}(\theta)})} E\left|I_{\{V<W\}}\sum_{i=V+1}^{W} Y_i + I_{\{V>W\}}\sum_{i=W+1}^{V} Y_i\right|$$
$$= E(Y_1)\inf_{(W,V)\in R(F_{\mathrm{bin}(n,q)},F_{\mathrm{poi}(\theta)})} E|W-V|. \tag{2.3}$$

From the result of L_1-Wassertein distance (Barrio et al., 1999, or Shorack and Wellner, 1986), it holds that

$$\inf_{(W,V)\in R(F_{\mathrm{bin}(n,q)},F_{\mathrm{poi}(\theta)})} E|W-V| = E|F^{-1}_{\mathrm{bin}(n,q)}(U) - F^{-1}_{\mathrm{poi}(\theta)}(U)| = \int_0^{\infty} |F_{\mathrm{bin}(n,q)}(x) - F_{\mathrm{poi}(\theta)}(x)|\,\mathrm{d}x. \tag{2.4}$$

It is obvious that

$$H_n(\theta) = \inf_{N\in R(F_{\mathrm{poi}(\theta)})} E\left|\sum_{i=1}^{N_n} Y_i - \sum_{i=1}^{N} Y_i\right| = \inf_{(W,V)\in R(F_{\mathrm{bin}(n,q)},F_{\mathrm{poi}(\theta)})} E\left|\sum_{i=1}^{W} Y_i - \sum_{i=1}^{V} Y_i\right|. \tag{2.5}$$

Combining Eqs. (2.3)–(2.5), we obtain (2.2) and complete the proof. □

In this paper we choose the optimal Poisson r.v. as

$$N_n^0(\theta) = F^{-1}_{\mathrm{poi}(\theta)}(U). \tag{2.6}$$

In modern risk theory, comonotonicity is an important concept. The two risks P and Q are comonotonic if there exist two non-decreasing real-valued functions u, v and a risk Z such that

$$P = u(Z), \qquad Q = v(Z)$$

(see Wang et al., 1997). By Lemma 1 and (2.6), $N_n^0(\theta)$ and N_n are both non-decreasing functions of the r.v. U. Thus they are comonotonic. It means that with the varying of U, $N_n^0(\theta)$ increases when N_n increases, and $N_n^0(\theta)$ decreases when N_n decreases.

3. Some results on $N_n^0(\theta)$

The joint distribution $F_{N_n,N(\theta)}$ of $(N_n, N(\theta))$ satisfies

$$\max\{F_{\text{bin}(n,q)}(x) + F_{\text{poi}(\theta)}(y) - 1, 0\} \le F_{N_n,N(\theta)}(x, y) \le \min\{F_{\text{bin}(n,q)}(x), F_{\text{poi}(\theta)}(y)\}. \tag{3.1}$$

See Joe (1997) for its related discussion. Note that the distribution function of $(N_n, N_n^0(\theta))$ equals the upper bound $\min\{F_{\text{bin}(n,q)}(x), F_{\text{poi}(\theta)}(y)\}$.

3.1. Some optimal properties

We will show some optimal properties of $N_n^0(\theta)$ in the following theorem.

Theorem 2. *The optimal r.v.* $N_n^0(\theta)$ *in* (2.6) *is optimal in the following senses*:

$$E(N_n - N_n^0(\theta))^2 = \inf_{N(\theta)\in R(F_{\text{poi}(\theta)})} E(N_n - N(\theta))^2 \tag{3.2}$$

and

$$\text{Var}(N_n - N_n^0(\theta)) = \inf_{N(\theta)\in R(F_{\text{poi}(\theta)})} \text{Var}(N_n - N(\theta)). \tag{3.3}$$

Proof. Recalling inequality (3.1) we have that for $x, y \in [0, \infty)$,

$$P(N_n > x, N(\theta) > y) \le \min\{P(N_n > x), P(N(\theta) > y)\}.$$

Applying Hoeffding's identity (Joe, 1997, p. 23), we obtain

$$\begin{aligned} E\{N_n \times N(\theta)\} &= \int_0^\infty \int_0^\infty P(N_n > x, N(\theta) > y)\,\mathrm{d}x\,\mathrm{d}y \le \int_0^\infty \int_0^\infty \min\{P(N_n > x), P(N(\theta) > y)\}\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_0^\infty \int_0^\infty P(N_n > x, N_n^0(\theta) > y)\,\mathrm{d}x\,\mathrm{d}y = E\{N_n N_n^0(\theta)\}. \end{aligned}$$

Thus we have

$$\begin{aligned} E(N_n - N(\theta))^2 &= EN_n^2 + E(N(\theta))^2 - 2E\{N_n \times N(\theta)\} \ge EN_n^2 + E(N(\theta))^2 - 2E\{N_n \times N_n^0(\theta)\} \\ &= E(N_n - N_n^0(\theta))^2, \end{aligned}$$

which implies (3.2).

By using (3.2) and

$$\text{Var}(N_n - N(\theta)) = E(N_n - N(\theta))^2 - (EN_n - EN(\theta))^2 = E(N_n - N(\theta))^2 - (nq - \theta)^2,$$

we obtain (3.3). Thus we complete the proof of the theorem. □

The above theorem asserts the optimality of $N_n^0(\theta)$ in the senses of (3.2) and (3.3).

3.2. The joint probability of $(N_n, N_n^0(\theta))$

In this section we give several formulas to evaluate the joint distribution of $(N_n, N_n^0(\theta))$.

Given $N_n = m, m = 0, 1, \ldots, n$, the conditional probability

$$\begin{aligned} P(N_n^0(\theta) = k | N_n = m) &= P(F_{\mathrm{poi}(\theta)}^{-1}(U) = k | U \in (F_{\mathrm{bin}(n,q)}(m-1), F_{\mathrm{bin}(n,q)}(m)]) \\ &= P(U \in (F_{\mathrm{poi}(\theta)}(k-1), F_{\mathrm{poi}(\theta)}(k)) | U \in (F_{\mathrm{bin}(n,q)}(m-1), F_{\mathrm{bin}(n,q)}(m)]) \\ &= \frac{\min\{F_{\mathrm{poi}(\theta)}(k), F_{\mathrm{bin}(n,q)}(m)\} - \max\{F_{\mathrm{poi}(\theta)}(k-1), F_{\mathrm{bin}(n,q)}(m-1)\}}{F_{\mathrm{bin}(n,q)}(m) - F_{\mathrm{bin}(n,q)}(m-1)} \vee 0. \end{aligned}$$

Then the joint probability

$$\begin{aligned} P(N_n = m, N_n^0(\theta) = k) &= P(N_n^0(\theta) = k | N_n = m) P(N_n = m) \\ &= (\min\{F_{\mathrm{poi}(\theta)}(k), F_{\mathrm{bin}(n,q)}(m)\} - \max\{F_{\mathrm{poi}(\theta)}(k-1), F_{\mathrm{bin}(n,q)}(m-1)\}) \vee 0. \end{aligned} \quad (3.4)$$

Furthermore, the probability $P(N_n^0(\theta) = N_n)$ can be evaluated by

$$\begin{aligned} P(N_n^0(\theta) = N_n) &= \sum_{i=1}^{n} P(N_n^0(\theta) = N_n = i) \\ &= \sum_{i=0}^{n} (\min\{F_{\mathrm{poi}(\theta)}(i), F_{\mathrm{bin}(n,q)}(i)\} - \max\{F_{\mathrm{poi}(\theta)}(i-1), F_{\mathrm{bin}(n,q)}(i-1)\}) \vee 0. \end{aligned}$$

Table 1 gives some values of the joint probabilities of $(N_n, N_n^0(\theta))$ in the case $n = 1000, \theta = 1, p = 0.001, EY_1 = 1$, calculated by (3.4). Table 2 gives some results in the case $n = 10, \theta = 0.01, p = 0.001, EY_1 = 1$. The values in the two tables show the comonotonic trend between N_n and $N_n^0(\theta)$.

Table 1
The probability $P(N_n = i, N_n^0(\theta) = j)$ when $n = 1000, \theta = 1, p = 0.001$ and $EY_1 = 1$

	$j = 0$	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$
$i = 0$	0.3677	0	0	0	0	0	0
$i = 1$	0.0002	0.3679	3.0687E−08	0	0	0	0
$i = 2$	0	0	0.1840	9.2054E−05	0	0	0
$i = 3$	0	0	0	0.0612	6.1323E−05	0	0
$i = 4$	0	0	0	0	0.0153	2.2969E−05	0
$i = 5$	0	0	0	0	0	0.0030	6.1147E−06
$i = 6$	0	0	0	0	0	0	0.0005
$i = 7$	0	0	0	0	0	0	0

Table 2
The probability $P(N_n = i, N_n^0(\theta) = j)$ when $n = 10, \theta = 0.01, p = 0.001$ and $EY_1 = 1$

	$j = 0$	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$
$i = 0$	0.9900	0	0	0	0	0	0
$i = 1$	4.9535E−06	0.0099	4.9072E−06	0	0	0	0
$i = 2$	0	0	4.4595E−05	4.6050E−08	0	0	0
$i = 3$	0	0	0	1.1896E−07	2.0435E−10	0	0
$i = 4$	0	0	0	0	2.0817E−10	5.7532E−13	0
$i = 5$	0	0	0	0	0	2.498E−13	8.8819E−16
$i = 6$	0	0	0	0	0	0	2.2205E−16

4. Evaluating the approximation error

$H_n(\theta)$ measures the expectation of the absolute approximation error. In this section we will show how to compute the error.

At the beginning we will list a useful lemma.

Lemma 2. *If* $\theta \geq -n\log(1-q)$, *then* $F_{\mathrm{bin}(n,q)}(t) \geq F_{\mathrm{poi}(\theta)}(t)$ *for all t. Conversely, if* $F_{\mathrm{bin}(n,q)}(0) \geq F_{\mathrm{poi}(\theta)}(0)$, *then* $\theta \geq -n\log(1-q)$.

Lemma 2 can be obtained directly by Theorem 1 of De Pril and Dhaene (1992).

Theorem 3.

(a) *If* $\theta \geq -n\log(1-q)$, *then*

$$H_n(\theta) = (\theta - nq)EY_1. \tag{4.1}$$

(b) *If* $0 \leq \theta < -n\log(1-q)$, *then*

$$H_n(\theta) = \left\{2\sum_{i=0}^{n}\sum_{j=0}^{i}(i-j)P(N_n = i, N_n^0(\theta) = j) + \theta - nq\right\}EY_1. \tag{4.2}$$

Proof.

(a) In the case $\theta \geq -n\log(1-q)$, by Lemma 2 we have

$$P(N_n > t) \leq P(N_n^0(\theta) > t), \quad t \in [0, \infty].$$

Then using (2.2), we obtain

$$H_n(\theta) = EY_1 \int_0^\infty |P(N_n \leq t) - P(N_n^0(\theta) \leq t)|\,\mathrm{d}t = EY_1 \int_0^\infty \{P(N_n^0(\theta) > t) - P(N_n > t)\}\,\mathrm{d}t$$
$$= EY_1(EN_n^0(\theta) - EN_n) = EY_1(\theta - nq).$$

(b) In the case $0 \leq \theta < -n\log(1-q)$, by Theorem 1 we have

$$H_n(\theta) = EY_1 \times E|N_n - N_n^0(\theta)| = EY_1 \times \{2E(N_n - N_n^0(\theta))_+ - E(N_n - N_n^0(\theta)\}$$
$$= EY_1 \times \{2E(N_n - N_n^0(\theta))_+ + \theta - nq\}$$
$$= \left\{2\sum_{i=0}^{n}\sum_{j=0}^{i}(i-j)P(N_n = i, N_n^0(\theta) = j) + \theta - nq\right\}EY_1,$$

where the probability function of $(N_n, N_n^0(\theta))$ can be calculated by (3.4). □

The Poisson parameter θ is often chosen as $\theta = nq$ or $\theta = -n\log(1-q)$ (Gerber, 1979, Chapter 4). Tables 3–5 provide some numerical values of $H_n(nq)$ and $H_n(-n\log(1-q))$ when $EY_1 = 1$. By comparing $H_n(nq)$ with $H_n(-n\log(1-q))$, one can analyze the accuracy of the approximation. For instance, in the case $nq = 0.05$, $H_{10}(nq) > H_{10}(-n\log(1-q))$, $H_{100}(nq) < H_{100}(-n\log(1-q))$ and $H_{1000}(nq) < H_{1000}(-n\log(1-q))$. Generally speaking, one cannot assert that the parameter nq is optimal than the parameter $-n\log(1-q)$.

Table 3
The expected errors when $n = 10$

q	nq	$n \log(1-q)$	$H_n(nq)$	$H_n(-n\log(1-q))$
0.001	0.01	0.010005	0.000010	0.000005
0.005	0.05	0.050125	0.000239	0.000125
0.01	0.1	0.100503	0.000911	0.000503
0.05	0.5	0.512933	0.015587	0.012933
0.1	1	1.053605	0.038402	0.053605
0.5	5	6.931472	0.524205	1.931472

Table 4
The expected errors when $n = 100$

q	nq	$n \log(1-q)$	$H_n(nq)$	$H_n(-n\log(1-q))$
0.0001	0.01	0.010001	9.90116E−07	$5.00033E-07$
0.001	0.1	0.100050	0.000091	0.000050
0.005	0.5	0.501254	0.001520	0.001254
0.01	1	1.005034	0.003694	0.005034
0.05	5	5.129329	0.044504	0.129329
0.1	10	10.536051	0.128624	0.536051
0.5	50	69.314718	1.653039	19.314718

Table 5
The expected errors when $n = 1000$

q	nq	$n \log(1-q)$	$H_n(nq)$	$H_n(-n\log(1-q))$
0.0001	0.1	0.100005	0.000009	0.000005
0.001	1	1.000500	0.000368	0.000500
0.005	0.5	5.012542	0.004393	0.012542
0.01	10	10.050336	0.012545	0.050336
0.05	50	51.293294	0.142642	1.293294

5. Approximation to the total loss and related functions

Given the values $\{N_n, Y_1, Y_2, \ldots, Y_{N_n}\}$ determined from the portfolio $\{X_1, X_2, \ldots, X_n\}$, the optimal Poisson r.v. in (2.6) can be determined as follows. First simulate the r.v. U by (2.1), then calculate the Poisson r.v. $N_n^0(\theta) = F_{\text{poi}(\theta)}^{-1}(U)$. In the case $N_n \geq N_n^0(\theta)$, the values of the claim sequence $Y_1, Y_2, \ldots, Y_{N_n^0(\theta)}$ can be obtained from the set $\{Y_1, Y_2, \ldots, Y_{N_n}\}$. Otherwise, simulation is needed to obtain the r.v.'s $\{Y_{N_n+1}, \ldots, Y_{N_n^0(\theta)}\}$.

Now we consider the total amount of the type $\sum_{i=1}^n g(X_i)$, where g is a non-negative measurable function and $g(0) = 0$. An approximation for $\sum_{i=1}^n g(X_i)$ is

$$\sum_{i=1}^{N_n^0(\theta)} g(Y_i).$$

The corresponding approximation error equals

$$h_n(\theta, g) =: \sum_{i=1}^{n} g(X_i) - \sum_{i=1}^{N_n^0(\theta)} g(Y_i).$$

Note that when $g(x) = x$,

$$S_n = \sum_{i=1}^{N_n^0(\theta)} Y_i + h_n(\theta, g).$$

Theorem 4. *For the non-negative function g with* $g(0) = 0$,

$$Eh_n(\theta, g) = (nq - \theta)Eg(Y_1) \tag{5.1}$$

and

$$E|h_n(\theta, g)| = \frac{Eg(Y_1)}{EY_1} H_n(\theta). \tag{5.2}$$

Proof. It is easy to prove (5.1). Here we only prove (5.2) in detail.

Since $g(0) = 0$, we have

$$\sum_{i=1}^{n} g(X_i) = \sum_{i=1}^{N_n} g(Y_i).$$

Thus

$$E|h_n(\theta, g)| = E\left|\sum_{i=1}^{N_n} g(Y_i) - \sum_{i=1}^{N_n^0(\theta)} g(Y_i)\right| = \frac{Eg(Y_1)}{EY_1} H_n(\theta).$$

Therefore, the theorem is proved. □

We would like to mention one interesting fact when $\theta = -n\log(1 - q)$. In this case, by Lemma 2 we have

$$F^{-1}_{\mathrm{poi}(\theta)}(s) \geq F^{-1}_{\mathrm{bin}(n,q)}(s), \quad s \in (0, 1)$$

Then

$$N_n = F^{-1}_{\mathrm{bin}(n,q)}(U) \leq F^{-1}_{\mathrm{poi}(\theta)}(U) = N_n^0(\theta).$$

Hence

$$\sum_{i=1}^{n} g(X_i) = \sum_{i=1}^{N_n} g(Y_i) \leq \sum_{i=1}^{N_n^0(\theta)} g(Y_i).$$

Thus the approximated total amount is always greater than the actual aggregate claims.

For the portfolio $\{X_1, X_2, \ldots, X_n\}$, we consider the excess-of-loss reinsurance treaty with retention M. The direct insurer pays the amount

$$S_{\mathrm{D},n}(M) = \sum_{i=1}^{n} X_i \wedge M$$

and the reinsurer pays the amount

$$S_{\mathrm{R},n}(M) = \sum_{i=1}^{n}(X_i - M)_+,$$

where $X_i \wedge M = \min\{X_i, M\}$ and $(X_i - M)_+ = \max\{X_i - M, 0\}$. Use the two compound Poisson r.v.'s

$$S_{\mathrm{D},n}^{AP}(M) = \sum_{i=1}^{N_n^0(\theta)} Y_i \wedge M \quad \text{and} \quad S_{\mathrm{R},n}^{AP}(M) = \sum_{i=1}^{N_n^0(\theta)} (Y_i - M)_+$$

to approximate $S_{\mathrm{D},n}(M)$ and $S_{\mathrm{R},n}(M)$, respectively. By Theorem 4, the expected approximation errors are

$$E|S_{\mathrm{R},n}(M) - S_{\mathrm{R},n}^{AP}(M)| = \frac{E(Y_1 \wedge M)}{EY_1} H_n(\theta), \qquad E|S_{\mathrm{D},n}(M) - S_{\mathrm{D},n}^{AP}(M)| = \frac{E(Y_1 - M)_+}{EY_1} H_n(\theta),$$

respectively. Note that the sum of the expected approximation errors equals $H_n(\theta)$, i.e.,

$$E|S_{\mathrm{R},n}(M) - S_{\mathrm{R},n}^{AP}(M)| + E|S_{\mathrm{D},n}(M) - S_{\mathrm{D},n}^{AP}(M)| = H_n(\theta).$$

6. The uniqueness of the Poisson parameter to minimizing $H_n(\theta)$

In this section we investigate whether there exists a unique θ_n^0, such that

$$H_n(\theta_n^0) = \min_{\theta \geq 0} H_n(\theta). \tag{6.1}$$

The answer to this problem is affirmative.

For $k = 0, 1, 2, \dots, n-1$, denote $\theta^{(k)}$ to be the solution of equation

$$F_{\mathrm{bin}(n,q)}(k) = F_{\mathrm{poi}(\theta^{(k)})}(k).$$

Then we have the following result.

Theorem 5. *It holds that*

$$\theta^{(n-1)} > \theta^{(n-2)} > \cdots > \theta^{(1)} > \theta^{(0)}.$$

Denote

$$\mathrm{sign}(F_{\mathrm{bin}(n,q)}(x) - F_{\mathrm{poi}(\theta)}(x)) = \begin{cases} 1 & \text{if } F_{\mathrm{bin}(n,q)}(x) > F_{\mathrm{poi}(\theta)}(x), \\ 0 & \text{if } F_{\mathrm{bin}(n,q)}(x) = F_{\mathrm{poi}(\theta)}(x), \\ -1 & \text{if } F_{\mathrm{bin}(n,q)}(x) < F_{\mathrm{poi}(\theta)}(x). \end{cases}$$

In the following we give a theorem on the uniqueness of θ_n^0 and present an approach to solve θ_n^0. Here for simplicity we assume $EY_1 = 1$.

Theorem 6. *There exists a unique θ_n^0 with $0 \leq \theta_n^0 \leq -n\log(1-q)$, such that (6.1) holds. Further, for $\theta \in (\theta^{(k+1)}, \theta^{(k)})$, $k = 0, 1, 2, \ldots, n-1$, the derivative function of H_n with respect to θ satisfies*

$$H_n'(\theta) = \sum_{k=0}^{n-1} (\text{sign}(F_{\text{bin}(n,q)}(k) - F_{\text{poi}(\theta)}(k)) - 1)\frac{\mathrm{e}^{-\theta}\,\theta^k}{k!} + 1. \tag{6.2}$$

$H_n'(\theta)$ is strictly increasing for $\theta \neq \theta^{(k)}$, $k = 0, 1, 2, \ldots, n-1$ and

$$H_n'(\theta_n^0-) \leq 0, \qquad H_n'(\theta_n^0+) \geq 0, \tag{6.3}$$

where $H_n'(\theta_n^0-)$ and $H_n'(\theta_n^0+)$ denote, respectively, the left and right limits of H_n' at point θ_n^0.

We will prove Theorems 5 and 6 in Appendix A.

In the case $n = 1$, (6.2) can be used to find θ_1^0. For $\theta < -\log(1-q)$, $F_{\text{bin}(1,q)}(0) < F_{\text{poi}(\theta)}(0)$. Then by (6.2) we have

$$H_1'(\theta) = -2\,\mathrm{e}^{-\theta} + 1.$$

Solving $H_1'(\theta) = 0$, we have $\theta = \log 2$. Moreover, $H_1'(\theta) < 0$ for $\theta < \log 2$. Then we can conclude that in the case $q > 0.5$, $H_n(\theta)$ achieves its minimum $\log 2 + q - 1$ at point $\theta_1^0 = \log 2$; when $q \leq 0.5$, $H_n(\theta)$ achieves its minimum $-q - \log(1-q)$ at point $\theta_1^0 = -\log(1-q)$.

In general, θ_n^0 can be solved numerically. By Theorem 6, θ_n^0 is the solution of $H_n'(\theta) = 0$ if the solution exits. Otherwise, it satisfies $H_n'(\theta-) \geq 0$ and $H_n'(\theta+) \leq 0$. Thus by Theorem 6, we can focus on the strictly increasing function $h(\theta)$ defined as

$$h(\theta) = \sum_{k=0}^{n-1} (\text{sign}(F_{\text{bin}(n,q)}(k) - F_{\text{poi}(\theta)}(k)) - 1)\frac{\mathrm{e}^{-\theta}\,\theta^k}{k!} + 1, \quad \theta \in [0, \infty)$$

and use the Newtonian method to locate θ_n^0.

Table 6 lists some numerical results. It can be seen that when n is large, the difference between θ_n^0 and nq is very small. Thus nq is a good approximation for θ_n^0 when nq is large while $-n\log(1-q)$ is a good approximation when nq is small.

7. Approximation to the heterogenous individual risk model

In practice, the individual risks X_i of a portfolio are often independent, but not identically distributed. In this section, we will apply the above results to the heterogenous individual risk model.

We can divide the heterogenous portfolio into several independent homogenous portfolios, then approximate every homogenous portfolio separately by our method. Assume that the portfolio can be divided into m homogenous portfolios and the aggregate claims from the ith portfolio can be represented by $IS(i)$ for $i = 1, 2, \ldots, m$. For the ith homogenous portfolio,

$$IS(i) = \sum_{j=1}^{N^{(i)}} Y_{i,j},$$

Table 6
The optimal θ_n^0 and the corresponding error $H_n(\theta_n^0)$

n	q	nq	$-n\log(1-q)$	θ_n^0	$H_n(nq)$	$H_n(-n\log(1-q))$	$H_n(\theta_n^0)$
2	0.001	0.002	0.002001	0.002001	1.997E−06	1.001E − 06	1.001E−06
2	0.01	0.02	0.020101	0.020101	0.000197	0.000101	0.000101
2	0.10	0.2	0.210721	0.210721	0.017462	0.010721	0.010721
2	0.5	1	1.386294	0.961278	0.235797	0.386294	0.226086
10	0.001	0.01	0.010005	0.010005	0.00001	0.000005	0.000005
10	0.01	0.1	0.100503	0.100503	0.000911	0.000503	0.000503
10	0.10	1	1.053605	0.99907	0.038402	0.053605	0.038161
10	0.5	5	6.931472	4.95961	0.524205	1.931472	0.519686
100	0.001	0.1	0.100050	0.100050	0.000091	0.000050	0.000050
100	0.01	1	1.005034	1.000	0.003694	0.005034	0.003694
100	0.1	10	10.536051	9.9991	0.128624	0.536051	0.128548
100	0.5	50	69.314718	49.959	1.653039	19.314718	1.651654
1000	0.001	1	1.0005	1.000	0.000368	0.000500	0.000368
1000	0.01	10	10.050336	10.00	0.012545	0.050336	0.012545

where $N^{(i)}$ is the number of claims and the claim sequences $Y_{i,j}$, $j = 1, 2, \ldots$, are determined as in Section 1, $i = 1, 2, \ldots, m$. Then the aggregate claims for the heterogenous portfolio is

$$IS = \sum_{i=1}^{m} IS(i) = \sum_{i=1}^{m}\sum_{j=1}^{N^{(i)}} Y_{i,j}.$$

Note that the following properties hold:

1. $N^{(i)}$, $i = 1, 2, \ldots, m$, are independent.
2. For a given i, $Y_{i,j}$, $j = 1, 2, \ldots$ are independent identically distributed, here the common distribution is denoted as F_i. Further, $Y_{i,j}$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots,$ are independent.
3. Claim sequences are independent of the numbers of claims.

Let $CS(i, \theta_i)$ be compound Poisson with Poisson r.v. $N(i, \theta_i)$ and claim sequence $Y_{i,j}$, $j = 1, 2, \ldots, N(i, \theta_i)$ and $Y_{i,j}$, $j = 1, 2, \ldots$ are independent. Here θ_i is the Poisson parameter. Then

$$CS(i, \theta_i) = \sum_{j=1}^{N(i,\theta_i)} Y_{i,j}.$$

The ith group's optimal Poisson r.v. determined by (2.6) is denoted as $N^0(i, \theta_i)$, the optimal Poisson parameter satisfying (6.1) is denoted as $\theta^0(i)$. The independence between the Poisson r.v.'s $N(i, \theta_i)$, $i = 1, 2, \ldots, m$ is assumed.

Denote

$$CS^*(i, \theta_i) = \sum_{j=1}^{N^0(i,\theta_i)} Y_{i,j},$$

$$CS(\theta_1, \theta_2, \ldots, \theta_m) = \sum_{j=1}^{m} CS(i, \theta_i)$$

and

$$CS^{*}(\theta_1, \theta_2, \ldots, \theta_m) = \sum_{j=1}^{m} CS^{*}(i, \theta_i).$$

$CS(\theta_1, \theta_2, \ldots, \theta_m)$ and $CS^{*}(\theta_1, \theta_2, \ldots, \theta_m)$ are compound Poisson r.v.'s, each with Poisson parameter $\sum_{i=1}^{m} \theta_i$ and claim distribution

$$\sum_{j=1}^{m} \frac{\theta_j}{\sum_{i=1}^{m} \theta_i} F_j.$$

Among all $CS(\theta_1, \theta_2, \ldots, \theta_m)$'s, we will prove that $CS^{*}(\theta_1, \theta_2, \ldots, \theta_m)$ is optimal in the following sense.

Theorem 7. *It holds that*

$$E(IS - CS^{*}(\theta_1, \theta_2, \ldots, \theta_m))^2 = \inf_{N(i,\theta_i), i \le m} E(IS - CS(\theta_1, \theta_2, \ldots, \theta_m))^2 \tag{7.1}$$

and

$$E|IS - CS^{*}(\theta_1, \theta_2, \ldots, \theta_m)| \le \sum_{i=1}^{m} E|IS(i) - CS^{*}(i, \theta_i)| \le \sum_{i=1}^{m} E|IS(i) - CS(i, \theta_i)|. \tag{7.2}$$

Proof. According to the above notations, we have

$$\begin{aligned} E(IS - CS(\theta_1, \theta_2, \ldots, \theta_m))^2 &= \mathrm{Var}(IS - CS(\theta_1, \theta_2, \ldots, \theta_m)) + \{E(IS - CS(\theta_1, \theta_2, \ldots, \theta_m))\}^2 \\ &= \sum_{i=1}^{m} \mathrm{Var}(IS(i) - CS(i, \theta_i)) + \left\{ \sum_{i=1}^{m} EY_{i,1} \times E(N^{(i)} - \theta_i) \right\}^2 \\ &= \sum_{i=1}^{m} E|N^{(i)} - N(i, \theta_i)|\, \mathrm{Var}(Y_{i,1}) + \sum_{i=1}^{m} \mathrm{Var}(N^{(i)} - N(i, \theta_i)) E(Y_{i,1}^2) \\ &\quad + \left\{ \sum_{i=1}^{m} E(Y_{i,1}) E(N^{(i)} - \theta_i) \right\}^2 . \end{aligned}$$

Thus by Theorems 1 and 2, it holds that

$$\begin{aligned} &E(IS - CS^{*}(\theta_1, \theta_2, \ldots, \theta_m))^2 \\ &= \sum_{i=1}^{m} E|N^{(i)} - N^{0}(i, \theta_i)|\, \mathrm{Var}(Y_{i,1}) + \sum_{i=1}^{m} \mathrm{Var}(N^{(i)} - N^{0}(i, \theta_i)) E(Y_{i,1}^2) + \left\{ \sum_{i=1}^{m} E(Y_{i,1}) E(N^{(i)} - \theta_i) \right\}^2 \\ &\le \sum_{i=1}^{m} E|N^{(i)} - N(i, \theta_i)|\, \mathrm{Var}(Y_{i,1}) + \sum_{i=1}^{m} \mathrm{Var}(N^{(i)} - N(i, \theta_i)) E(Y_{i,1}^2) + \left\{ \sum_{i=1}^{m} E(Y_{i,1}) E(N^{(i)} - \theta_i) \right\}^2 \\ &= E(IS - CS(\theta_1, \theta_2, \ldots, \theta_m))^2, \end{aligned}$$

thus (7.1) holds. It is obvious that (7.2) holds. Now we complete the proof. □

When every homogenous portfolio's size is large enough, it is advised to choose $\theta_i = EN^{(i)}$, $i = 1, 2, \dots, m$. For this choice of the Poisson parameter, the approximation $CS^*(\theta_1, \theta_2, \dots, \theta_m)$ produces good approximation effects. It can be seen from the following two aspects:

(1) First, from Table 6 we can find that

$$\theta^0(i) \sim EN^{(i)}.$$

Thus $EN^{(i)}$ is a good approximation for the optimal Poisson parameter $\theta^0(i)$.

(2) Second, (7.2) shows that the approximation error is bounded by the sum of the approximation errors for each homogenous portfolio. Thus the corresponding error is small when every homogenous portfolio's size is large enough.

8. Conclusions

In this paper, we presented a new method to approximate the individual risk model by a compound Poisson r.v. We investigated the determination of the r.v.'s in the compound Poisson model and the calculation of the approximation error. During our discussion, we first focused on the homogenous individual risk models, then applied the results to the heterogenous individual risk models. Numerical results showed that our approximation model provides a good approximation.

Acknowledgements

The authors would like to thank the referees for their helpful suggestions. The first author was supported by NSFC grants 19831020, 10471008 and the second author was supported by NSFC grant 60372024.

Appendix A

In what follows, we will prove Theorems 5 and 6. First we introduce several lemmas.

Lemma A.1. *There exists a non-negative integer $C \leq n-1$ such that*

$$\frac{P(N_n = k+1)}{P(N_n = k)} \begin{cases} > \dfrac{P(N(\theta) = k+1)}{P(N(\theta) = k)}, & k < C, \\ < \dfrac{P(N(\theta) = k+1)}{P(N(\theta) = k)}, & n > k > C, \\ \geq \dfrac{P(N(\theta) = k+1)}{P(N(\theta) = k)}, & k = C. \end{cases}$$

Proof. It is easy to verify that for $k + 1 \leq n$,

$$\frac{P(N_n = k+1)}{P(N_n = k)} = \frac{C_n^{k+1} q^{k+1} (1-q)^{n-k-1}}{C_n^k q^k (1-q)^{n-k}} = \frac{n-k}{k+1} \frac{q}{1-q} \tag{A.1}$$

and

$$\frac{P(N(\theta)=k+1)}{P(N(\theta)=k)}=\frac{\theta}{k+1}. \tag{A.2}$$

Comparing (A.1) with (A.2), we observe that

$$\frac{P(N_n=k+1)}{P(N_n=k)}\geq\frac{P(N(\theta)=k+1)}{P(N(\theta)=k)}$$

if and only if $(n-k)\{q/(1-q)\}\geq\theta$, i.e.,

$$k\leq n-\frac{\theta(1-q)}{q}.$$

Denote

$$C=\min\left\{n-1,\left[n-\frac{\theta(1-q)}{q}\right]\right\}.$$

Here $[x]$ denotes the integral part of x. Thus if $C<n-1$, for the integer $n>k>C$ it holds that

$$\frac{P(N_n=k+1)}{P(N_n=k)}<\frac{P(N(\theta)=k+1)}{P(N(\theta)=k)}.$$

And if $k<C$,

$$\frac{P(N_n=k+1)}{P(N_n=k)}>\frac{P(N(\theta)=k+1)}{P(N(\theta)=k)}$$

holds. In the case $k=C$, it holds that

$$\frac{P(N_n=k+1)}{P(N_n=k)}\geq\frac{P(N(\theta)=k+1)}{P(N(\theta)=k)}$$

Hence the lemma is proved. □

It is easy to verify the following lemma. We omit its proof here.

Lemma A.2. *Suppose* $a_n>0$, $b_n>0$ *and* $a_{n+1}/a_n\geq b_{n+1}/b_n$, $n\geq 0$. *Then for* $n\geq 1$,

$$\frac{a_n+a_{n-1}+\cdots+a_0}{a_{n-1}+\cdots+a_0}\geq\frac{b_n+b_{n-1}+\cdots+b_0}{b_{n-1}+\cdots+b_0}.$$

Lemma A.3. *Suppose that for non-negative integer* M, $F_{\text{bin}(n,q)}(M)\geq F_{\text{poi}(\theta)}(M)$. *Then*

$$F_{\text{bin}(n,q)}(M+1)>F_{\text{poi}(\theta)}(M+1). \tag{A.3}$$

Proof. It is trivial for the case $M\geq n$. Now we consider the case $M<n$. Assume $\theta<-n\log(1-q)$. Otherwise, by Lemma 4.1 we see that $F_{\text{bin}(n,q)}(k)\geq F_{\text{poi}(\theta)}(k)$, $k=0,1,2,\ldots$, thus (A.3) holds. We will divide our proof into two cases $M\leq C$ and $M>C$, where C is the integer defined in Lemma A.1.

Case A.1 $(M \le C)$.

Since $\theta < -n\log(1-q)$, it leads to $P(N_n = 0) < P(N(\theta) = 0)$. Then according to the definition of C and $F_{\text{bin}(n,q)}(M) \ge F_{\text{poi}(\theta)}(M)$, by Lemmas A.1 and A.2 we conclude that

$$P(N_n = M) > P(N(\theta) = M).$$

Also by Lemma A.1 we have

$$\frac{P(N_n = M+1)}{P(N_n = M)} \ge \frac{P(N(\theta) = M+1)}{P(N(\theta) = M)}.$$

Then it follows that:

$$P(N_n = M+1) > P(N(\theta) = M+1).$$

Thus

$$F_{\text{bin}(n,q)}(M+1) = F_{\text{bin}(n,q)}(M) + P(N_n = M+1) > F_{\text{poi}(\theta)}(M) + P(N(\theta) = M+1) = F_{\text{poi}(\theta)}(M+1),$$

which implies (A.3).

Case A.2 $(M > C)$.

If $P(N_n = M+1) \ge P(N(\theta) = M+1)$, then it is obvious that (A.3) holds. Next we consider the case $P(N_n = M+1) < P(N(\theta) = M+1)$.

Since $P(N_n = M+1) < P(N(\theta) = M+1)$ and $M > C$, by Lemma A.1 we have

$$P(N_n = k) < P(N(\theta) = k), \quad k \ge M+1.$$

Then

$$\begin{aligned} P(N_n \le M+1) &= 1 - \sum_{k=M+2}^{n} P(N_n = k) > 1 - \sum_{k=M+2}^{n} P(N(\theta) = k) \\ &> 1 - \sum_{k=M+2}^{\infty} P(N(\theta) = k) = P(N(\theta) \le M+1). \end{aligned}$$

Thus the lemma is proved. □

Proof of Theorem 5. Fix $k < n$. From the definition $F_{\text{bin}(n,q)}(k) = F_{\text{poi}(\theta^{(k)})}(k)$ and Lemma A.3, we have

$$F_{\text{bin}(n,q)}(k+1) > F_{\text{poi}(\theta^{(k)})}(k+1).$$

Since for fixed k the function $F_{\text{poi}(\theta)}(k+1)$ is strictly decreasing about θ, we have

$$\theta^{(k+1)} < \theta^{(k)}.$$

Now we complete the proof of the theorem.

Lemma A.4. *For $\theta \neq \theta^{(i)}$, $i = 0, 1, \ldots, n-1$, the derivative $H_n'(\theta)$ of $H_n(\theta)$ exists and*

$$H_n'(\theta) = \sum_{k=0}^{n-1} (\mathrm{sign}(F_{\mathrm{bin}(n,q)}(k) - F_{\mathrm{poi}(\theta)}(k)) - 1) \frac{\mathrm{e}^{-\theta}\,\theta^k}{k!} + 1. \tag{A.4}$$

Proof. Fix $\theta_0 > 0$ and $\theta_0 \neq \theta^{(i)}$, $i = 0, 1, \ldots, n-1$. Denote

$$g(\theta, x) = F_{\mathrm{bin}(n,q)}(x) - F_{\mathrm{poi}(\theta)}(x).$$

First consider the case $\theta > \theta_0$. From (2.2) we have

$$\begin{aligned}
\frac{H_n(\theta) - H_n(\theta_0)}{\theta - \theta_0} &= \frac{1}{\theta - \theta_0} \int_0^\infty (|g(\theta, x)| - |g(\theta_0, x)|)\, \mathrm{d}x \\
&= \frac{1}{\theta - \theta_0} \int_0^\infty (g(\theta, x) I_{\{g(\theta,x)>0\}} - g(\theta_0, x) I_{\{g(\theta_0,x)>0\}})\, \mathrm{d}x \\
&\quad + \frac{1}{\theta - \theta_0} \int_0^\infty (-g(\theta, x) I_{\{g(\theta,x)<0\}} + g(\theta_0, x) I_{\{g(\theta_0,x)<0\}})\, \mathrm{d}x =: A + B.
\end{aligned} \tag{A.5}$$

By using the fact that $g(\theta, x) > g(\theta_0, x)$ for $\theta > \theta_0$ and $g(\theta_0, x) \neq 0$, we have

$$\begin{aligned}
A &= \frac{1}{\theta - \theta_0} \int_0^\infty (g(\theta, x) I_{\{g(\theta,x)>0\}} - g(\theta_0, x) I_{\{g(\theta,x)>0\}})\, \mathrm{d}x \\
&\quad + \frac{1}{\theta - \theta_0} \int_0^\infty (g(\theta_0, x)(I_{\{g(\theta,x)>0\}} - I_{\{g(\theta_0,x)>0\}})\, \mathrm{d}x \\
&= \frac{1}{\theta - \theta_0} \int_0^\infty (F_{\mathrm{poi}(\theta_0)}(x) - F_{\mathrm{poi}(\theta)}(x)) I_{\{g(\theta,x)>0\}}\, \mathrm{d}x \\
&\quad + \frac{1}{\theta - \theta_0} \int_0^\infty (F_{\mathrm{bin}(n,q)}(x) - F_{\mathrm{poi}(\theta_0)}(x)) I_{\{g(\theta,x)>0,\, g(\theta_0,x)<0\}}\, \mathrm{d}x.
\end{aligned} \tag{A.6}$$

For the case $g(\theta, x) > 0$ and $g(\theta_0, x) < 0$, it holds that $F_{\mathrm{bin}(n,q)}(x) > F_{\mathrm{poi}(\theta)}(x)$ and $F_{\mathrm{bin}(n,q)}(x) < F_{\mathrm{poi}(\theta_0)}(x)$. Thus in this case it follows that:

$$\begin{aligned}
0 &\leq \int_0^\infty \left| \frac{F_{\mathrm{bin}(n,q)}(x) - F_{\mathrm{poi}(\theta_0)}(x)}{\theta - \theta_0} \right| I_{\{g(\theta,x)>0,\, g(\theta_0,x)<0\}} |\, \mathrm{d}x \\
&\leq \int_0^\infty \frac{F_{\mathrm{poi}(\theta_0)}(x) - F_{\mathrm{poi}(\theta)}(x)}{\theta - \theta_0} I_{\{g(\theta,x)>0,\, g(\theta_0,x)<0\}}\, \mathrm{d}x.
\end{aligned} \tag{A.7}$$

For $x \in [0, \infty)$,

$$\frac{\partial}{\partial\theta} F_{\mathrm{poi}(\theta)}(x) = \frac{\mathrm{e}^{-\theta}\,\theta^{[x]}}{[x]!} \leq \frac{(2\theta_0)^{[x]}}{[x]!}, \quad \theta < 2\theta_0,$$

where $[x]$ is the integral part of x. Then by the dominated convergence theorem and $g(\theta, x) \to g(\theta_0, x)$ we have

$$\frac{1}{\theta - \theta_0} \int_0^\infty (F_{\mathrm{poi}(\theta_0)}(x) - F_{\mathrm{poi}(\theta)}(x)) I_{\{g(\theta,x)>0\}}\, \mathrm{d}x \to \int_0^\infty \frac{\mathrm{e}^{-\theta_0}\,\theta_0^{[x]}}{[x]!} I_{\{g(\theta_0,x)>0\}}\, \mathrm{d}x, \tag{A.8}$$

and by (A.7) we obtain

$$\left|\int_0^\infty \frac{F_{\mathrm{bin}(n,q)}(x) - F_{\mathrm{poi}(\theta)}(x)}{\theta-\theta_0} I_{\{g(\theta,x)>0,\, g(\theta_0,x)<0\}}\,\mathrm{d}x\right| \le \int_0^\infty \frac{F_{\mathrm{poi}(\theta_0)}(x) - F_{\mathrm{poi}(\theta)}(x)}{\theta-\theta_0} I_{\{g(\theta,x)>0,\, g(\theta_0,x)<0\}}\,\mathrm{d}x \to 0. \tag{A.9}$$

Combining (A.8) and (A.9), we conclude that

$$\lim_{\theta\searrow\theta_0} A = \int_0^\infty \frac{\mathrm{e}^{-\theta_0}\,\theta_0^{[x]}}{[x]!} I_{\{g(\theta_0,x)>0\}}\,\mathrm{d}x.$$

Similarly, we have

$$\lim_{\theta\searrow\theta_0} B = -\int_0^\infty \frac{\mathrm{e}^{-\theta_0}\,\theta_0^{[x]}}{[x]!} I_{\{g(\theta_0,x)<0\}}\,\mathrm{d}x.$$

Summarizing the above two equations and (A.5), we deduce that

$$\lim_{\theta\searrow\theta_0} \frac{H_n(\theta)-H_n(\theta_0)}{\theta-\theta_0} = \int_0^\infty \frac{\mathrm{e}^{-\theta_0}\,\theta_0^{[x]}}{[x]!} \mathrm{sign}(F_{\mathrm{bin}(n,q)}(x) - F_{\mathrm{poi}(\theta_0)}(x))\,\mathrm{d}x. \tag{A.10}$$

Similarly,

$$\lim_{\theta\nearrow\theta_0} \frac{H_n(\theta)-H_n(\theta_0)}{\theta-\theta_0} = \int_0^\infty \frac{\mathrm{e}^{-\theta_0}\,\theta_0^{[x]}}{[x]!} \mathrm{sign}(F_{\mathrm{bin}(n,q)}(x) - F_{\mathrm{poi}(\theta_0)}(x))\,\mathrm{d}x \tag{A.11}$$

holds. Thus combining (A.10) and (A.11), we obtain

$$\begin{aligned} H_n'(\theta_0) &= \int_0^\infty \frac{\mathrm{e}^{-\theta}\,\theta_0^{[x]}}{[x]!} \mathrm{sign}(F_{\mathrm{bin}(n,q)}(x) - F_{\mathrm{poi}(\theta_0)}(x))\,\mathrm{d}x \\ &= \sum_{k=0}^{n-1} \mathrm{sign}(F_{\mathrm{bin}(n,q)}(k) - F_{\mathrm{poi}(\theta_0)}(k)) \frac{\mathrm{e}^{-\theta_0}\,\theta_0^k}{k!} + \sum_{k=n}^{\infty} \frac{\mathrm{e}^{-\theta_0}\,\theta_0^k}{k!} \\ &= \sum_{k=0}^{n-1} (\mathrm{sign}(F_{\mathrm{bin}(n,q)}(k) - F_{\mathrm{poi}(\theta_0)}(k)) - 1) \frac{\mathrm{e}^{-\theta_0}\,\theta_0^k}{k!} + 1. \quad \square \end{aligned}$$

Proof of Theorem 6. Fix $\theta \in (\theta^{(i+1)}, \theta^{(i)})$. Since

$$F_{\mathrm{bin}(n,q)}(i) = F_{\mathrm{poi}(\theta^{(i)})}(i), \qquad F_{\mathrm{bin}(n,q)}(i+1) = F_{\mathrm{poi}(\theta^{(i+1)})}(i+1),$$

we have

$$F_{\mathrm{bin}(n,q)}(i) < F_{\mathrm{poi}(\theta)}(i), \qquad F_{\mathrm{bin}(n,q)}(i+1) > F_{\mathrm{poi}(\theta)}(i+1).$$

Thus by Lemma A.3 we assert that

$$F_{\mathrm{bin}(n,q)}(k) < F_{\mathrm{poi}(\theta)}(k), \quad k \le i$$

and

$$F_{\mathrm{bin}(n,q)}(k) > F_{\mathrm{poi}(\theta)}(k), \quad k \geq i+1.$$

Then by Lemma A.4 we have

$$H'(\theta) = -2\sum_{k=0}^{i} \frac{\mathrm{e}^{-\theta}\,\theta^k}{k!} + 1,$$

which is strictly increasing with respect to $\theta \in (\theta^{(i+1)}, \theta^{(i)})$.

From the above equation it is easy to verify

$$H'(\theta^{(i)}-) < H'(\theta^{(i)}+), \quad i = 0, 1, 2, \ldots, n-1.$$

Thus we conclude that $H'(\theta)$ is strictly increasing with respect to $\theta \neq \theta^{(i)}$.

Since $H_n(\theta)$ is continuous, from Theorem 3 we assert that the minimum point θ_n^0 exists and $\theta_n^0 \leq n\log(1-q)$. By the monotonicity of $H_n'(\theta)$ with respect to θ, we assert that $H_n'(\theta_n^0+) \geq 0$ and $H_n'(\theta_n^0-) \leq 0$. Since $H_n'(\theta)$ is strictly increasing, thus θ_n^0 is unique. Hence the uniqueness is proved.

As (6.2) has been proved in Lemma A.4, we now complete the proof of Theorem 6. □

References

Barrio, E.D., Gine, E., Matran, C., 1999. Central limit theorems for the Wasserstein distance between the empirical and the true distributions. The Annals of Probability 27 (2), 1009–1071.

Bühlmann, H., Gagliardi, B., Gerber, H.U., Straub, E., 1977. Some inequalities for stop-loss premiums. Astin Bulletin 9, 75–83.

Dhaene, J., Sundt, B., 1997. On error bounds for approximations to aggregate claims distributions. Astin Bulletin 27 (2), 243–262.

De Pril, N., Dhaene, J., 1992. Error bounds for the compound Poisson approximations of the individual risk model. Astin Bulletin 22 (2), 135–148.

Gerber, H.U., 1979. An Introduction to Mathematical Risk Theory. S.S. Hueber Foundation Monograph No. 8. Richard D. Irwin, Homewood, IL.

Gerber, H., 1984. Error bounds for the compound Poisson approximation. Insurance: Mathematics and Economics 3, 191–194.

Hipp, C., 1985. Approximation of aggregate claims distributions by compound Poisson approximation. Insurance: Mathematics and Economics 4, 227–232.

Hipp, C., 1986. Improved approximations for the aggregate claims distribution in the individual model. Astin Bulletin 16 (2), 89–100.

Joe, H., 1997. Multivariate Models and Dependence Concepts. Chapman & Hall, New York, pp. 35–52.

Kaas, R., Van Heerwaarden, Goovaerts, M.J., 1988b. Between individual and collective model for the total claims. Astin Bulletin 18 (2), 169–174.

Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M., 2001. Modern Actuarial Risk Theory. Kluwer Academic Publishers, London.

Kuon, S., Radtke, M., Reich, A., 1993. An approximate way to switch from the individual risk model to the collective one. Astin Bulletin 23 (1), 23–32.

Li, X.D., Yang, J.P., 2001. Compound Poisson approximation to individual risk models. Acta Scientiarum Naturalium, Universitatis Pekinensis 37 (5), 609–617 (in Chinese).

Michel, R., 1987. An improved error bound for the compound Poisson approximation of a family of a nearly homogeneous portfolio. Astin Bulletin 17 (2), 165–169.

Panjer, H.H., 1981. Recursive evaluation of a family of compound distributions. Astin Bulletin 12 (1), 22–26.

Panjer, H.H., Willmot, G.E., 1992. Insurance Risk Models. Society of Actuaries, Schaumburg.

Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Applications to Statistics. Wiley, New York, p. 64.

Sundt, B., 1993. Improved error bounds for Bertram's method. Astin Bulletin 23 (2), 301–303.

Wang, S., Young, V.R., Panjer, H.H., 1997. Axiomatic characterization of insurance prices. Insurance: Mathematics and Economics 21, 173–183.

Further reading

Hesselager, O., 1995. Order relations for some distributions. Insurance: Mathematics and Economics 16, 129–134.
Kaas, R., Van Heerwaarden, Goovaerts, M.J., 1988a. On stop-loss premiums for the individual risk model. Astin Bulletin 18, 91–98.
Sundt, B., Dhaene, J., 1991. On approximating claims distributions and stop-loss premiums by truncation. Insurance: Mathematics and Economics 10, 133–136.
Sundt, B., Dhaene, J., 1996. On bounds for the difference between the stop-loss transforms of the two compound distributions. Astin Bulletin 26 (2), 225–231.