DOI：10.13718/j.cnki.xdzk.2016.05.017

聚合风险模型下指数保费的非参数估计[①]

张林娜[1]，　温利民[1,2]，　方　婧[1]

1. 江西师范大学 数学与信息科学学院，南昌 330022；2. 江西财经大学 信息管理学院，南昌 330013

摘要：在聚合风险模型的假设下，研究了聚合风险下指数保费的非参数估计，证明了估计的强相合性和渐近正态性. 最后通过数值模拟的方法验证了估计的收敛速度及渐近正态性.

关　键　词：聚合风险模型；指数保费；非参数估计；相合性；渐近正态性

中图分类号：O211.5　　**文献标志码**：A　　**文章编号**：1673－9868(2016)05－0100－06

在保险精算中，精算师的一个重要任务就是如何为一份保单制定合适的价格. 制定保费的原则既不能太高，否则容易失去市场竞争力，又不能太低，否则会出现偿付不足的情况. 因此，合适的选取保费原理是保险公司的一个重要举措. 指数保费原理具有非负安全负荷性、合理风险附加性、极大损失性、转移不变性、单调性、保留一阶随机控制(FSD)序和停止损失序以及连续性等性质，是保费原理中除期望值原理之外，满足最多优良性质的保费原理[1－2]，是目前保险精算中运用最多的保费原理之一.

一般地，保险公司常常拥有大量相同类型的保单. 为了刻画这些保单在一定时间内的总索赔，我们有两类风险模型可供选择：一类是个体风险模型，另一类是聚合风险模型. 个体风险模型以每张保单为基本对象[3]，考虑某些保单(称为保单组合)在一定时期内可能发生的理赔总量，进而考虑全部保单组合在某段时期的理赔总量. 设有 m 张保单，在一定时期内第 i 张保单发生的总索赔为 Y_i，则这个时期内的总索赔额为：

$$T=\sum_{i=1}^{m}Y_i \tag{1}$$

即为个体风险模型的总索赔额.

注意到，当 m 很大时，有些保单在这段时间内可能取值为 0，而有些保单可能导致多个非零索赔. 若记这个保单组合中取值非零的总索赔次数为 N，这些非零索赔记为 X_i，$i=1,2,\cdots$，则总索赔可表示为

$$S=\sum_{i=1}^{N}X_i \tag{2}$$

式(2) 称为聚合风险模型. 关于聚合风险模型的进一步讨论可参考文献[4－6].

由于聚合风险有较为复杂的结构，总索赔 S 的分布不仅涉及到索赔次数 N 的分布，而且还与一系列索赔额 X_i，$i\geqslant 1$ 有关，因此 S 的分布的精确表达式较难得到，因而聚合索赔 S 的保费亦是未知的. 但保险公司在长期的经营中可能已经对 S 有一定的历史记录(样本)，根据这些信息可以对聚合风险的保费进行估计.

本文将研究聚合风险模型下总索赔的指数保费及其非参数估计问题. 后面的章节安排如下：第 1 节介

---

① 收稿日期：2015－03－02
基金项目：国家自然科学基金项目(71361015)；教育部人文社科基金项目(15YJC910010)；中国博士后科学基金项目(2013M540534)；江西省研究生创新基金项目(YC2014－S162).
作者简介：张林娜(1991－)，女，山西运城人，硕士研究生，主要从事数理统计与保险精算研究.
通信作者：温利民，教授，硕士研究生导师.

绍聚合风险模型和指数保费原理；第2节给出聚合风险模型的非参数估计，并证明估计的强相合性和渐近正态性；第3节利用数值模拟的方法验证非参数估计的收敛速度.

1　聚合风险模型和指数保费原理

考虑某个保险公司的某个保单组合在一定时间(这里指一年)的总索赔额符合(2)式的聚合风险模型. 一般地，在聚合风险模型中常对索赔次数 N 和索赔额过程 $\{X_i, i \geqslant 1\}$ 提出下面的假设：

假设1　索赔次数 N 为取非负整数值随机变量，具有概率分布 $P(N=k)=p_k$，$k=0,1,2,3,\cdots$，记 $M_N(t)=\mathscr{E}[\exp(Nt)]$ 为 N 的矩母函数.

假设2　索赔额 X_j，$j=1,2,\cdots$ 独立同分布，具有概率分布函数 $F_X(x)$、密度函数 $f(x)$、矩母函数 $M_X(t)=\mathscr{E}[\exp(Xt)]$ 以及一阶矩 $\beta_1=\mathscr{E}(X_1)$ 和二阶矩 $\beta_2=\mathscr{E}(X_1{}^2)$ 存在.

假设3　索赔额过程 $\{X_j, j=1, 2, \cdots\}$ 与索赔次数 N 相互独立.

记 χ 为所有满足上面假设的聚合风险的总索赔随机变量的集合，定义聚合风险 S 的保费为 χ 到 $\mathbb{R}$ 的一个映射，称为保费原理.

定义1　设 $S \in \chi$ 是取值非负的聚合风险总索赔，其分布函数为 $F_S(s)$，保费原理就是给风险 S 分配一个实值泛函 $H(\cdot)$，记为 $S \longmapsto H(S)$.

假设一份保单的风险记为 S，则保费就是从 S 到实数集 $\mathbb{R}$ 的一个映射，记为 $H(S)$. 由于风险 S 是由其分布函数来刻画的，并且 S 由分布 $F_S(x)$ 唯一确定，因此，风险 S 的保费也可以记为 $H(F_S(x))$. 在保险精算中，精算师根据不同的保险情况选择不同的保费原理. 比如，期望值保费原理适用于寿险，很少用于财产保险和偶然性保险，方差保费原理经常适用于财产保险，而指数保费原理既用于寿险也用于财产保险. 有关保费原理的讨论可参考文献[7－9].

一种重要的保费计算原理是通过效用函数来描述的. 假设一个保险公司具有最初的资产为 w，其效用函数为 $u(\cdot)$. 在保险公司面临的总索赔为 S 时，一种公平的总保费 $H(S)$ 可以定义为下面方程的解：

$$\mathscr{E}[u(w+H(S)-S)]=u(w) \tag{3}$$

公式(3)意味着公司承保了保单后的期望效用确切地等价于没有承保保单的效用，因此条件(3)表达了保费 $H(X)$ 在效用上是公平的. 如果令 $w=0$ 并规定 $u(0)=0$，则得到零效用保费原理，即保费 $H(S)$ 为下面方程的解：

$$\mathscr{E}[u(H(S)-S)]=0 \tag{4}$$

虽然零效用原理保费在理论上看是非常合理的，但该原理无法在实际中加以运用，因为 $H(S)$ 在方程(4)中一般没有精确的解析式. 但是，如果效用函数取为下面的指数形式：

$$u(x)=1-\exp(-\alpha x) \qquad x>0 \tag{5}$$

这里 $\alpha>0$ 为已知参数，则求解(4)式可得到下面的显示解：

$$P \equiv H(S)=\frac{1}{\alpha}\log[\mathscr{E}[\exp(\alpha S)]] \qquad \alpha>0 \tag{6}$$

定义2　对可保风险 $X \in \chi$，由(6)式定义的保费称为指数保费原理.

注意到(5)式中的常数 α 恰好等于 $-\dfrac{\mathrm{d}}{\mathrm{d}x}\ln\dfrac{\mathrm{d}u}{\mathrm{d}x}$，度量了保险公司对风险的厌恶度. 容易看出指数保费原理(6)是在损失函数 $L(S, P)=(\exp(\alpha P)-\exp(\alpha S))^2$ 下的风险的最优估计[10].

显然，为了得到聚合风险 S 的保费 $H(S)$ 的显示表达式，则首先需要给出 S 的分布 $F_S(s)$. 注意到 S 的分布与索赔额序列 X_i(一般为连续型随机变量)和索赔次数 N(一般为离散型随机变量)都有关. 在聚合风险模型(2)中，总索赔 S 的分布函数为：

$$F_S(x)=\Pr(S \leqslant x)=\sum_{k=0}^{\infty} p_k F^{*k}(x) \tag{7}$$

其中 $F^{*k}(x)$ 表示 $F_X(x)$ 的 k 重卷积.

显然，在一般情况下要得到 S 的分布是非常困难的[4]，解决这个问题的一个可行的办法就是充分利用

样本数据对 $F_S(x)$ 进行估计，从而得到保费 $H(F_S(x))$ 的一个非参数估计.

2　聚合风险模型下指数保费的非参数估计

设保险公司对聚合风险 S 已经有一定年份的索赔记录，设 $S_1, S_2, \cdots, S_n$ 为前 n 年的观测值，其中

$$S_i = \sum_{j=1}^{N_i} X_{ij} \qquad i = 1, 2, \cdots, n$$

而 N_i 为第 i 年的索赔次数，而 X_{ij} 为第 i 年的第 j 次索赔额. 假设 N_i，$i = 1, 2, \cdots, n$ 为相互独立且具有相同分布的取非负整数值的随机变量. 而$\{X_{ij}, i = 1, 2, \cdots, n, j = 1, 2, \cdots\}$相互独立且具有相同的分布，并与索赔次数$\{N_i, i \geqslant 1\}$相互独立. 我们的目标是利用这些样本信息对 S 的指数保费进行估计并作相应的推断.

为了书写方便，记

$$Z = \exp(\alpha S) \qquad \mu_z = \mathscr{E}(Z) \qquad \sigma_z^2 = \mathrm{Var}(Z) \tag{8}$$

由定义 2，指数保费为 $H(S) = \dfrac{1}{\alpha} \log \mathscr{E}(\exp(\alpha S))$，$\alpha > 0$ 为保险公司对风险的厌恶度，则 $H(S)$ 的一个非参数估计为

$$\widehat{H(S)} = \frac{1}{\alpha} \log\left(\frac{1}{n} \sum_{i=1}^{n} \exp\left(\alpha \sum_{j=1}^{N_i} X_{ij}\right)\right) \tag{9}$$

对非参数估计$\widehat{H(S)}$有下面的大样本性质：

命题 1　假设 $\mu_z = \mathscr{E}(\exp(\alpha S))$ 存在，当 $n \to \infty$ 时，$\widehat{H(S)}$ 是指数保费 $H(S)$ 的强相合估计，即

$$\widehat{H(S)} \longrightarrow H(S), \text{ a. s.} \tag{10}$$

证　$\widehat{H(S)}$ 可以表达为

$$\widehat{H(S)} = \frac{1}{\alpha} \log\left(\frac{1}{n} \sum_{i=1}^{n} \exp(\alpha S_i)\right)$$

由于 S_i 对 $i = 1, 2, \cdots, n$ 是独立同分布的，因此由强大数定律，有

$$\frac{1}{n} \sum_{i=1}^{n} \exp(\alpha S_i) \longrightarrow \mu_z, \text{ a. s.}$$

又由于对数函数为连续函数，则

$$\widehat{H(S)} = \frac{1}{\alpha} \log\left(\frac{1}{n} \sum_{i=1}^{n} \exp(\alpha S_i)\right) \longrightarrow \frac{1}{\alpha} \log(\mu_z) = H(S), \text{ a. s.}$$

引理 1[11]　设$\{a_n\}$为一趋于 ∞ 的数列，b 为常数，并且对随机变量序列$\{Z_n\}$有

$$a_n(Z_n - b) \xrightarrow{L} Z \tag{11}$$

又设 $g(\cdot)$ 为可微函数，且 g' 在点 b 处连续，则有

$$a_n[g(Z_n) - g(b)] \xrightarrow{L} g'(b) Z \tag{12}$$

命题 2　当 $n \to \infty$ 时，$\widehat{H(S)}$ 是渐近正态的，即

$$\sqrt{n}\,(\widehat{H(S)} - H(S)) \xrightarrow{L} N(0, \sigma^2) \tag{13}$$

其中渐近方差为

$$\sigma^2 = \frac{1}{\alpha^2 \mu_z^2} \sigma_z^2 \tag{14}$$

证　根据独立同分布的中心极限定理，有

$$\sqrt{n}\,(\overline{Z} - \mu_z) \xrightarrow{L} N(0, \sigma_z^2) \tag{15}$$

令

$$g(x) = \frac{1}{\alpha} \log(x)$$

则

$$g(\mu_z)=\frac{1}{\alpha}\log(\mu_z)$$

那么

$$g'(\mu_z)=\frac{1}{\alpha\mu_z}$$

则由引理1知,

$$\sqrt{n}(\widehat{H(S)}-H(S))=\sqrt{n}(g(\overline{Z})-g(\mu_z))\xrightarrow{L}N(0,(g'(\mu_z))^2\sigma_z^2),\ \text{a. s.} \tag{16}$$

将 $g'(\mu_z)$ 代入,易得

$$\sigma^2=\frac{1}{\alpha^2\mu_z^2}\sigma_z^2 \tag{17}$$

即有

$$\sqrt{n}(\widehat{H(S)}-H(S))\xrightarrow{L}N(0,\sigma^2)$$

3 数值模拟

在本节中,我们将利用数值模拟的方法验证聚合风险下指数保费估计 $\widehat{H(S)}$ 具有大样本性质并检验其收敛速度情况. 假定 X_i,$i=1,2,\cdots$ 独立同分布服从指数分布,密度函数为

$$f(x)=\lambda\exp(-\lambda x)\qquad x>0 \tag{18}$$

因此有

$$M_X(\alpha)=\frac{\lambda}{\lambda-\alpha}$$

$$M_X'(\alpha)=\frac{\mathrm{d}M_X(\alpha)}{\mathrm{d}\alpha}=\frac{\lambda}{(\lambda-\alpha)^2} \tag{19}$$

再假定索赔次数 N 服从 Poisson(θ) 分布,此时由模型假设1-3,可得随机变量 Z 的数学期望为

$$\mu_z=\exp(\theta(M_X(\alpha)-1)) \tag{20}$$

方差为

$$\sigma_z^2=\exp(\theta(M_X(2\alpha)-1))-\exp(2\theta(M_X(\alpha)-1)) \tag{21}$$

在模拟中,假设某保险公司对风险的厌恶度为0.2,即 $\alpha=0.2$,我们取 $\lambda=2$,$\theta=1$,因此指数保费为:

$$H(S)=\frac{1}{\alpha}\log\left[\exp(\theta(M_X(\alpha)-1))\right]=\frac{\theta}{\lambda-\alpha}=0.56 \tag{22}$$

渐近方差为

$$\begin{aligned}\sigma^2=&\frac{1}{\alpha^2\mu_z^2}\left[\exp\left(\theta(M_X(2\alpha)-1)\right)-\exp(2\theta(M_X(\alpha)-1))\right]=\\&\frac{1}{\alpha^2\exp\left(\frac{2\theta\alpha}{\lambda-\alpha}\right)}\left[\exp\left(\frac{2\theta\alpha}{\lambda-2\alpha}\right)-\exp\left(\frac{2\theta\alpha}{\lambda-\alpha}\right)\right]=\\&0.704\ 2\end{aligned} \tag{23}$$

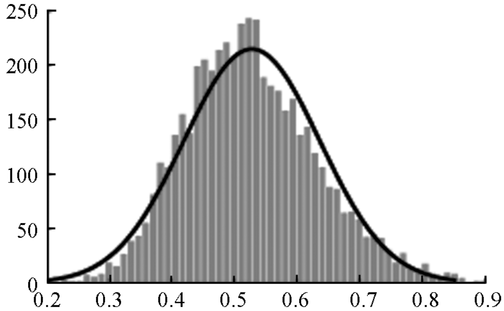
分别取 $n=10,30,50,80,100,200,500,800,1\ 000$ 进行模拟,每次模拟重复5 000次,计算得到估计的平均值及其相应的均方误差如表1.

表1 聚合风险模型下指数保费非参数估计的相合性模拟

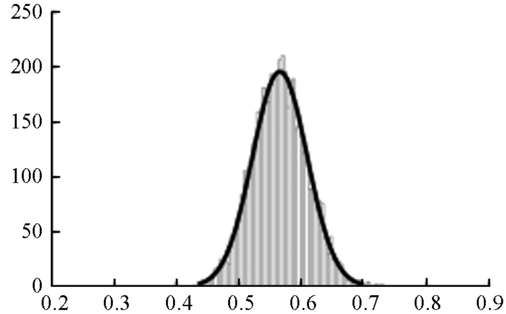
样本容量 n	10	30	50	80	100	200	500	800	1 000
平均值	0.532 0	0.481 5	0.606 4	0.593 5	0.531 4	0.525 4	0.578 7	0.566 2	0.565 4
均方误差	0.033 4	0.016 6	0.010 7	0.006 5	0.004 4	0.002 7	0.001 4	0.000 6	0.000 4

由表1可看出,随着样本容量的增大,平均值与真实值更接近,即偏差越来越小,说明本文提出的非

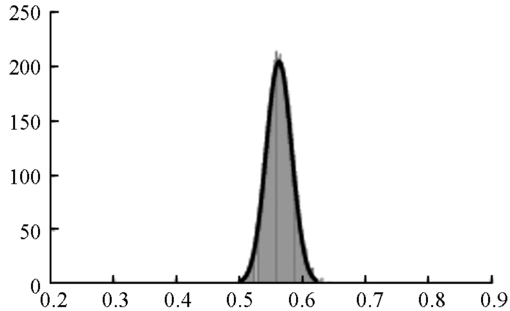
参数估计是渐近无偏的，且均方误差也越来越小. 取不同的 n 分别画出模拟的概率直方图和正态概率图以检验估计的渐近正态性(图 1,2).



(a) n=30

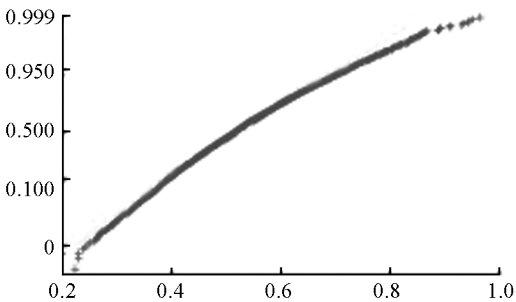


(b) n=200

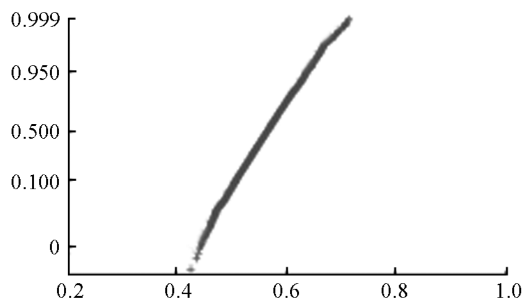


(c) n=1 000

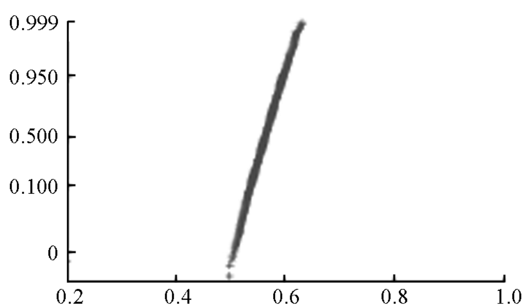
图 1　$\widehat{H(S)}$的频率直方图



(a) n=30



(b) n=200



(c) n=1 000

图 2　$\widehat{\mathrm{H(S)}}$的正态概率图

由图 1,2 可看出随着 n 的增大，$\widehat{H(S)}$ 的频率直方图与正态分布图更接近，且正态概率图中的点更接近一条直线. 说明估计具有很好的渐近正态性.

取定不同的置信水平 α'，在给定不同的 n 下，得到指数保费的 $1-\frac{\alpha'}{2}$ 置信度的置信区间，如表2.

表2 不同置信水平 α' 下 $H(S)$ 的近似置信区间

样本容量 n	30		200		1 000	
置信水平 α'	0.05	0.10	0.05	0.10	0.05	0.10
置信区间	(0.46, 0.50)	(0.45, 0.51)	(0.50, 0.55)	(0.49, 0.56)	(0.54, 0.59)	(0.53, 0.60)

根据表2可以对指数保费 $H(S)$ 进行统计推断，例如区间估计、假设检验等.

参考文献：

[1] GERBER H U. On Additive Premium Calculation Principles [J]. Astin Bulletin, 1974, 8(1): 215-222.

[2] DENUIT M. The Exponential Premium Calculation Principle Revisted [J]. Astin Bulletin, 1999, 29(2): 215-226.

[3] YANG J, ZHOU S, ZHANG Z. The Compound Poisson Random Variable's Approximation to the Individual Risk Model [J]. Insurance: Mathematics and Economics, 2005, 36(1): 57-77.

[4] HERNANDEZ A, FERNANDEZ-SANCHEZ M P, GOMEZ E. On the Independence between Risk Profiles in the Compound Collective Risk Actuarial Model [J]. Mathematics and Computers in Simulation, 2012, 82(8): 1419-1431.

[5] 施久玉. 一类聚合风险模型 [J]. 运筹与管理, 2004, 13(6): 53-60.

[6] 方 婧, 章 溢, 温利民. 聚合风险模型下的信度估计 [J]. 江西师范大学学报(自然科学版), 2012, 36(6): 607-611.

[7] YOUNG V R. Premium Principles [M]. Encyclopedia of Actuarial Science. Hoboken: John Wiley and Sons, 2006.

[8] FURMAN E, ZITIKIS R. Weighted Premium Calculation Principles [J]. Insurance: Mathematics and Economics, 2008, 42(1): 459-465.

[9] BUHLMANN H. An Economic Premium Priciple [J]. Astin Bulletin, 1980, 11(1): 52-60.

[10] WEN L M, WANG W, WANG J L. The Credibility Premium for Exponential Principle [J]. Acta Mathematica Sinica, 2011, 27 (11): 2217-2228.

[11] 茆诗松, 王静龙, 濮晓龙. 高等数理统计 [M]. 2版. 北京: 高等教育出版社, 2006.

The Nonparametric Estimate of Exponential Premium Under Collective Risk Models

ZHANG Lin-na[1], WEN Li-min[1,2], FANG Jing[1]

1. Department of Mathematics and Information Science, Jiangxi Normal University, Nanchang 330022, China;

2. Department of Information and Management, Jiangxi University of Finance and Economics, Nanchang 330013, China

Abstract: The exponential premium principle is one of the most important premium principles and is widely applied in non-life insurance actuarial science. In this paper, the nonparametric estimate of exponential premium is investigated under collective risk models. In addition, the estimator is proved strongly consistent and asymptotically normal. Finally, a numerical simulation method is used to verify the estimated speed of convergence, and the asymptotic normality of the estimator is checked in the simulations.

Key words: collective risk model; exponential premium; nonparametric estimate; consistency; asymptotic normality

责任编辑 张 枸