经济管理学术文库 • 经济类

非寿险精算中的贝叶斯统计分析

Bayesian Statistical Analysis in Non-life Insurance Actuarial Science

温利民　章 溢／著

经济管理出版社
ECONOMY & MANAGEMENT PUBLISHING HOUSE

经济管理学术文库 • 经济类

非寿险精算中的贝叶斯统计分析

Bayesian Statistical Analysis in Non-life Insurance
Actuarial Science

温利民　章　溢／著

经济管理出版社
ECONOMY & MANAGEMENT PUBLISHING HOUSE

图书在版编目（CIP）数据

非寿险精算中的贝叶斯统计分析/温利民著. —北京：经济管理出版社，2018.8
ISBN 978-7-5096-5872-7

Ⅰ. ①非… Ⅱ. ①温… Ⅲ. ①贝叶斯方法—应用—保险—精算学 Ⅳ. ①F840.4

中国版本图书馆 CIP 数据核字（2018）第 143810 号

组稿编辑：杨国强
责任编辑：杨国强 张瑞军
责任印制：黄章平
责任校对：张晓燕

出版发行：经济管理出版社
（北京市海淀区北蜂窝 8 号中雅大厦 A 座 11 层 100038）
网　　址：www. E-mp. com. cn
电　　话：(010) 51915602
印　　刷：北京玺诚印务有限公司
经　　销：新华书店
开　　本：720mm×1000mm/16
印　　张：12.25
字　　数：213 千字
版　　次：2018 年 9 月第 1 版　2018 年 9 月第 1 次印刷
书　　号：ISBN 978-7-5096-5872-7
定　　价：68.00 元

前 言

本书主要介绍非寿险精算中贝叶斯统计模型与统计推断方法，全书都是基于贝叶斯统计的理论基础之上，因此本书的阅读者应对贝叶斯统计有所了解。

本书的结构大概分为四个部分。第一部分是综述，主要介绍本书的结构、研究的问题以及章节的安排。第二部分介绍了各种保费原理中风险保费的贝叶斯统计推断。第三部分介绍风险度量模型中风险参数及相关风险度量的贝叶斯估计及其统计推断方法。在这一部分中，作者通过结合信度理论与经验贝叶斯方法，研究了帕累托风险模型中的风险参数、在险价值及其推广风险度量以及随机效应分层模型中的风险参数的信度估计问题。第四部分是责任准备金评估模型及其贝叶斯统计推断，研究了聚合数据中随机 B-F 模型索赔均值的信度估计及准备金的贝叶斯估计，并建立了个体数据的广义线性模型，同时研究了准备金的极大似然估计及其统计性质。本书所有内容都是基于非寿险精算的最新文献之上的拓展研究，力图将理论和实际运用相结合，涉及数学方面的内容都有详细的证明，尽量做到通俗易懂。

本书第二作者章溢负责第 5 章、第 7 章、第 8 章和第 10 章的编写工作，其余章节由温利民完成。全书由温利民统稿。

本书可以作为保险精算、金融数学以及概率论与数理统计专业的教学用书或参考用书。该书适用于两类读者：第一类是保险精算、金融、统计等专业的教师和研究人员，以及广大攻读保险、精算方向学位的研究生；第二类是金融、保险、统计、管理的从业人员，特别是正在从事保险精算和金融创新设计的决策人员以及相关的研究工作者。深信本书将对他们的学习、研究和应用有

所禅益。

感谢国家自然科学基金（71761019）和江西省自然科学基金（20171ACB21022）对本书的出版资助！

温利民

江西师范大学

2018 年 9 月

目　录

第一部分　非寿险精算中的贝叶斯统计问题

第二部分　风险保费的贝叶斯统计推断

第三部分 风险度量的贝叶斯统计分析

第四部分 责任准备金的评估及其统计模型

第一部分
非寿险精算中的贝叶斯统计问题

第 1 章　非寿险精算中的贝叶斯统计分析

保险精算是依据经济学的基本原理和知识,利用现代数学方法,对各种保险经济活动未来的财务风险进行分析、估价和管理的一门综合性的应用科学。如研究保险事故的出险规律、保险事故损失额的分布规律、保险人承担风险的平均损失及其分布规律、保险费率和责任准备金、保险公司偿付能力等保险具体问题。对于经营保险业务的各类保险公司,在其经营管理的各个环节上,如计划、统计、展业、财务、投资、研究及培训等方面,都需要精算发挥其特有的作用。保险精算具有计算、分析、预测、服务等多项职能,这些职能可以通过保险精算师的职能具体表现出来。

保险精算根据保险的类别及计算方法的不同又可以分为寿险精算和非寿险精算。寿险精算是以概率论和数理统计为工具研究人寿保险的寿命分布规律、寿险出险规律、寿险产品的定价、责任准备金的计算、保单现金价值的估值等问题的学科。在寿险精算中,利率和死亡率的测算是厘定寿险成本的两个基本问题。由于利率一般由国家控制,所以在相当长的时期里利率并不是保险精算所关注的主要问题,而死亡率的测算即生命表的建立成为寿险精算的核心工作,现在仍然是精算研究的重要课题。根据相当数量的人群在不同年龄的死亡率构成的表格称为生命表。生命表又称“死亡表”(Mortality Table)或寿命表,是对相当数量的人口自出生(或一定年龄)开始,直到这些人口全部去世为止的生存与死亡记录。通常以 10 万(或 100 万)人作为 0 岁的生存人数,然后根据各年中死亡人数、各年末生存人数计算各年龄人口的死亡率、生存率,列成表格,直到此 10 万人全部死亡为止。生命表上所记载的死亡率、生存率是决定人寿保险费的重要依据,是反映一个国家或一个区域人口生存死亡规律的调查统计表。有了生命表,寿险精算的所有问题都能得到解决。

相对而言,非寿险精算问题则比寿险精算复杂得多。由于非寿险保险标的的不同,风险因素复杂多变,导致非寿险精算中损失发生的概率、额度都不

相同。例如财产保险中,财产所在的地域、建筑结构和用途等都将影响损失的发生概率。南方地域比北方地域发生洪灾的可能性大;木质结构的房屋比水泥结构的房屋发生火灾的可能性大;机动车从事营运的比非营运的发生交通事故的可能性大。再如汽车第三者责任保险,司机的性别、年龄、喝酒习惯,汽车的性能、型号以及路段等因素都会影响事故发生的可能性。非寿险精算始终把损失发生的频率、损失发生的规模以及对损失的控制作为它的研究重心。

为了合理地确定非寿险精算中事故发生的频率和额度,一个重要的信息是该保险的历史损失数据。例如汽车保险,在以往的时间内,该汽车向保险公司索赔的次数和索赔额度的数据显然能在一定程度上代表该汽车的未来风险状况。然而,一般来说,单个保单的历史索赔数据是非常稀缺的。在这种小样本的数据中,以往的索赔数据很可能是司机的运气不佳导致的。但是,保险公司同时承保大量的其他保单。因此,一个重要的问题是,是否可以利用其他同类型保单的数据来计算或改进该保单发生风险的概率?解决这类问题的最好方法就是利用贝叶斯统计方法及在此基础上发展起来的信度理论。

在数理统计中,一般假设随机变量 X 的分布为 $F_X(x,\theta)$,其中 θ 是未知参数。贝叶斯统计与频率统计最大的区别是对待 θ 的观点不同。在频率学派中,一般认为 θ 是一个固定的未知参数,从而可以根据样本数据 $x_1,x_2,\cdots,x_n$ 进行估计并做统计推断。而贝叶斯学派认为 θ 是一个未知的随机变量,在以往的经验和数据资料中已经对 θ 有一定的认识,这些认识将形成 θ 的某个概率分布 $\pi(\theta)$。由于这个分布是根据经验或以往历史资料形成并假定出来的,因此被称为先验分布或结构分布。在综合样本数据 $x_1,x_2,\cdots,x_n$ 及先验分布 $\pi(\theta)$ 的基础上对 θ 进行估计并做统计推断。由于风险的非齐次性,导致保单发生损失的概率受到多方面因素的影响,这些因素的综合用随机参数 θ 刻画。根据经验资料和专家观点,将形成 θ 的先验分布 $\pi(\theta)$,而该保单在过去时间内的索赔恰可以看成风险 X 的样本。这些特征与贝叶斯统计的假设相当吻合,因此,非寿险精算中风险概率等相关问题用贝叶斯统计来建立模型是较为妥当的。

§1.1 保费定价及相关统计问题

在非寿险精算中,精算师最重要的职责就是如何为一份保单制定合适的

保费。由于风险是随机变量,而制定的保费必须是一个固定的金额。在厘定保费时,保险公司有两个最为关心的问题:一是如何使征收的保费足够理赔;二是在保费足够理赔的基础上,如何增强保险产品的竞争力。第一个问题要求保费尽量高,以使得总的保费收入减去索赔及相关费用后有剩余,保证保险公司的正常利润。第二个问题要求保费尽量低,以使得保险公司有充足的竞争力,在市场竞争中赢得更多的保单。因此,合理的保费定价显得非常关键。保险公司在厘定费率时,既要考虑到总的保费收入,又要考虑到投保人的预期保费,使得保费在投保人之间公平分摊。

定义 1.1.1　设 X 是取值非负的风险随机变量,其分布函数为 $F_X(x)$,保费原理就是给风险 X 分配一个实值泛函数 $H(\cdot)$,记为 $X \to H(X)$,或 $F_X \to H(F_X)$。

保费原理作为风险 X 的一种度量,在理论上和实际中都要求 $H(X)$ 满足一些条件。

(1)非负安全负荷性(Non-negative loading):$H(X) \geqslant E(X)$。一般来说,收取的保费至少要大于风险 X 的数学期望 $E(X)$,否则根据大数定律,保险公司将处于亏损状态,并最终导致破产的发生。

(2)极大损失(Maximal loss):$H[X] \leqslant \mathrm{esssup}[X]$,其中 $\mathrm{esssup}[X]$ 表示 X 的本性上确界。

(3)转移不变性(Translation invariance):$H[X+C]=H[X]+C$,其中 C 为任意常数。

(4)对独立风险的可加性(Additivity):若 $X,Y \in \mathcal{X}$,且相互独立,则 $H[X+Y]=H[X]+H[Y]$。若两个风险是相互独立的,则合起来承保与单独承保缴纳的保费相同。

(5)累次性(Iterativity):对任意的风险 X,Y,有 $H(X)=H[H(X|Y)]$。即 X 的保费可以分两步进行,首先给定 $Y=y$ 的条件下,计算 X 的条件保费,记为 $g(y)$,即 $g(y)=H(X|Y=y)$,则 $H(X)=H[g(Y)]$。

(6)合理风险附加(No unjustified risk loading):若 $X=C$, a. s. ,其中 C 为大于零的常数,则 $H[C]=C$。如果已经知道(以概率 1)某个风险取值为常数 C,则收取保费 C 是合理的,因为这时没有风险的波动性。

(7)单调性(Monotoni city):若 $X,Y \in \mathcal{X}$,且 $X(\omega) \leqslant Y(\omega)$ 对任意 $\omega \in \Omega$ 成立,则 $H[X] \leqslant H[Y]$。

(8)保留一阶随机控制序(First stochastic order):若 $X,Y \in \mathcal{X}$,且 $S_X(t) \leqslant S_Y(t)$,对任意 $t \geqslant 0$ 成立,则 $H[X] \leqslant H[Y]$。

(9)保留停止损失序(Stop loss order):若 $X, Y\in\mathcal{X}$,且 $E[X-d]_+\leqslant E[Y-d]_+$,对任意 $d\geqslant 0$ 成立,则 $H[X]\leqslant H[Y]$。

(10)连续性(Continuity):$\lim\limits_{a\to 0^+}H[\max(X-a, 0)]=H[X]$ 以及 $\lim\limits_{a\to\infty}H[\min(X, a)]=H[X]$。

在实际运用中,常用的保费原理有:

(1)净保费原理:$H(X)=E(X)$。

(2)期望值原理:$H(X)=(1+\alpha)E(X)$,其中 $\alpha>0$ 为常数。

(3)方差原理:$H(X)=E(X)+\alpha Var(X)$,其中 $\alpha>0$ 为常数。

(4)修改方差原理:$H(X)=E(X)+\alpha\dfrac{Var(X)}{EX}$,其中 $\alpha>0$ 为常数。

(5)标准差原理:$H(X)=E(X)+\alpha\sqrt{Var(X)}$,其中 $\alpha>0$ 为常数。

(6)指数原理:$H(X)=\dfrac{1}{\alpha}\log E[\exp(\alpha X)]$,其中 $\alpha>0$ 为常数。

(7)Esscher 原理:$H(X)=\dfrac{E(Xe^{\alpha X})}{E(e^{\alpha X})}$,其中 $\alpha>0$ 为常数。

(8)Kamp 保费:$H(X)=\dfrac{E[X(1-e^{-\alpha X})]}{E(1-e^{-\alpha X})}$,其中 $\alpha>0$ 为常数。

(9)条件尾期望原理:$H(X)=E(X|X>q)=\dfrac{E[XI(X>q)]}{P(X>q)}$,其中 $q>0$ 为常数。

(10)修正条件尾期望原理:$H(X)=E(X|X>q)+\alpha\dfrac{Var(X|X>q)}{E(X|X>q)}$,其中 $q>0, \alpha>0$ 为常数。

(11)失真保费原理:$H(X)=\int_0^\infty g[S_X(x)]dx$,其中函数 $g(\cdot)$ 为非降失真函数,满足:$g(0)=0, g(1)=1$。

(12)风险调整保费原理:$H(X)=\int_0^\infty (SX(x))^{\frac{1}{p}}dx$,其中 $p>1$ 为常数。

(13)零效用保费原理:保费 H 为方程 $U(0)=E[U(H-X)]$ 的解,其中 $U(x)$ 为效用函数。

(14)分位数保费原理:$H(X)=\min\{p, F_X(p)\geqslant 1-\varepsilon\}$,其中 ε 为给定的小概率。

(15)绝对偏差保费原理:$H(X)=E(X)+\alpha\kappa_x$,其中 $\kappa_x=E|X-F_X^{-1}(\frac{1}{2})|$ 为

X 的绝对中位数离差,$\alpha>0$ 为常数。

显然,并不是每一种保费原理都满足上述十条性质,但满足更多的性质的保费原理在实际中得到更广泛的运用。不同的保险公司可以根据公司自身的情况选取合适的保费原理。然而,保费原理的选取只是保险公司的第一步工作,更重要的是如何利用已有的信息(先验信息和样本信息)对给定的保费进行估计,并制订合适的保费厘定方案。这种结合了先验信息和样本信息得到的保费估计方法,在非寿险精算中称为经验厘定。

经验厘定方法最早起源于 20 世纪初,现已发展成为非寿险精算定价保费的最基本方法。由于大部分经验厘定的保费估计都可以表达为样本信息和先验信息的加权和,并且权重依赖于样本容量的大小,反映了样本的可信度,因此经验厘定方法也称为信度理论。根据保费厘定的方法不同,可以将其分为两个主要分支:第一个分支是建立在频率方法上的有限扰动理论(Li mited Fluctuation);第二个分支是以贝叶斯理论为基础的最精确可信度理论(Greatest Accuracy Credibility)。这两种方法都是希望通过已有的历史数据来合理地制定保费。

有限扰动信度理论是 Mowbray(1914)在对工伤补偿保险的保费厘定时提出的,而 Slellwayer(1925)将有限扰动理论应用于汽车保险,Perryman(1932)对有限扰动理论进行了解释。

最大精确度信度理论也称为欧式经验厘定理论。最早提出最大精确度信度理论方法的是 Whitney(1918),他在研究工伤补偿保险的保费厘定时,假设一个雇主拥有 P 个雇员,总索赔服从二项分布 B(P, M),而索赔概率 M 本身也是正态分布随机变量,事实上这正是贝叶斯框架下的二项—正态模型。根据贝叶斯定理可以估计参数 M,得到下一期的保费估计为信度加权的形式。Keffer(1929)将最大精确度信度原理运用到群体寿险。Bailey(1945)运用最小二乘法对最大精确信度理论建立数学模型,得到正态—正 态、贝塔—二项等分布模型下的信度估计。而真正无分布(任意分布)的最大精确度信度模型是 Bühlmann(1967)提出的。他仍然用最小二乘法,在贝叶斯框架下,将估计限定在样本的线性函数类中,得到的最优保费估计恰好为信度加权形式,并建立了无分布信度理论。由于最大精确度信度理论不依赖于先验分布的具体形式,使得这种方法在非寿险精算中得到广泛的运用,在后来的研究和运用中逐步替代了有限扰动信度理论,因此现代的信度理论或经验厘定一般指最大精确度信度理论。

Bühlmann 信度估计是建立在净保费原理之上的。由于净保费原理的安全

负荷系数为零,研究表明,如果保险公司收取的保费仅仅为净保费,则保险公司必将破产。Bühlmann(1970)首次讨论了方差保费原理的信度估计问题。Gerber(1980)通过将平方损失函数修改为指数加权损失函数的方法建立了Esscher 保费原理下的信度理论。Pan 等(2008)证明了 Gerber(1980)的信度估计不满足相合性,从而提出 Esscher 保费原理中新的信度估计形式。Wen 等(2009)将这一结果推广到广义加权保费原理,Wen 和 Wu(2011)建立了指数保费原理中的信度估计问题。本书的第二部分将围绕保费原理中的信度理论展开,进一步研究零效用保费原理、广义加权保费原理经验厘定问题。

§1.2 风险度量的统计及推断

众所周知,高收益总是伴随高风险,风险的本质就是不确定性。由于投资组合的回报率是不确定的,因此任何投资都有风险。概率论与数理统计就是处理这种不确定性(风险) 的学科。在概率统计中,一般用随机变量描述风险(不确定性)。风险度量是从风险的集合空间到 R 上的一个实函数。常用的风险度量有方差度量、在险价值度量、条件尾期望度量、失真风险度量等。在保险精算中,保单的保费即为这份保单风险的价格,因此保费定价问题也是风险度量的问题之一。在金融风险管理中,对风险的度量和评估是投资者、决策者最为关心的问题,也是保险精算师最为关心的问题之一。

Markowitz(1952)利用方差度量风险,建立了马科维茨(Markowitz) 投资组合理论,并因此获得了 1990 年的诺贝尔经济学奖。但是,研究者发现利用方差度量风险有很大的局限性,即追加或减少一定数量的固定现金于一个资产组合并不影响该资产组合的风险大小,这与人们对风险的直观判断是不吻合的。因此,研究者提出许多其他度量风险的方法,例如半方差、在险价值度量、尾在险价值度量、条件在险价值度量、期望短缺、失真风险度量,等等。Artzner(1999)从公理化的角度探讨一种合理的风险测度方法应该满足什么样的性质这一问题,并首次提出了一致风险度量的概念,将风险度量建立在公理化体系下。自此,几乎所有的风险度量的研究都基于一致性度量之上。

在金融风险管理的运用中,一个现实的问题是,如何根据已有信息对相关的风险度量进行估计? 注意到风险度量是风险 X 的某个实函数,与风险的分布有关,而风险随机变量的分布一般是未知的。通过了解风险的信息(一般指

样本信息)去估计风险的分布,进而估计风险 X 的度量。在经典的统计学中,常利用样本数据来估计风险 X 的大小。假定风险随机变量 X 具有某个概率分布 $F(x, \theta)$,这里 θ 为未知参数,表示 X 的未知风险特征,称为风险参数。设 $X_1, \cdots, X_n$ 为该风险的 n 个观察值。根据贝叶斯决策理论,评估风险 X 的大小不仅利用样本数据信息(观察值),而且利用到参数 θ 的先验分布信息,即:

风险度量的估计 = θ 的先验信息 $\oplus$ X 的样本信息

但是,上述估计的过程中有两个问题:①如何为风险参数选择合适的先验分布;②如何为风险选取合适的样本分布。对第一个问题,贝叶斯理论中有一些选取先验分布的方法,例如共轭先验、无信息先验等。但是,大部分的先验分布选取的方法或者是基于数学上的考虑(例如共轭先验分布),或者只是在某些特定条件下适用(例如无信息先验),至今未找到一种普遍适用的确定先验分布的方法。对第二个问题,目前大多数运用的方法是根据样本来拟合一个理论分布。但是,这样拟合出来的理论分布只是近似分布,而不一定是风险的真实分布。本书的第三部分将研究各种风险模型中风险度量的统计推断问题,分别建立帕累托风险模型、指数—伽马模型以及随机效应风险模型,利用信度理论以及经验贝叶斯理论研究风险参数及风险度量的估计及其统计推断。

§1.3　责任准备金的评估模型

索赔准备金又称赔款准备金,是衡量保险人某一时期内已经发生的保险事故应负的赔款责任及理赔费用的估计金额。其计提的原因在于:在保险公司会计年度内发生的赔案中,总有一部分未能在当年决算结案。根据审慎经营的原则,保险公司对于这些已发生赔案应提取赔款准备金,以正确核算损益,并为预期索赔支付保险金做好资金准备。非寿险责任准备金是保险公司最主要的负债项目。责任准备金评估的充足性和准确性,分别构成了保险公司履行保险赔付责任的能力和经营成本的重要基础,成为保险监管部门的重点监管目标之一。

在众多责任准备金评估模型中,应用最广泛、影响最大的当属链梯法。在早期的保险公司中,精算师将索赔按事故发生的年份进行分类。若以纵轴为事故发生年,横轴为事故发生后的进展年,则索赔数据将形成一个上三角形。