Available online at www.sciencedirect.com

SCIENCE DIRECT®

Insurance: Mathematics and Economics 36 (2005) 119–135

INSURANCE
MATHEMATICS & ECONOMICS

www.elsevier.com/locate/ime

Hierarchical Bayesian collective risk model: an application to health insurance

Helio S. Migon*, Fernando A.S. Moura[1]

Departamento de Métodos Estatística, Instituto de Matemática, Universidade Federal do Rio de Janeiro (UFRJ), CP 68507, CEP 21945-970 Rio de Janeiro, Brazil

Received December 2003 ; received in revised form November 2004; accepted 29 November 2004

Abstract

This paper deals with the main statistical steps involved in building an insurance plan, with special emphasis on an application to health insurance. The pure premium is predicted based on the available past information concerning the number and the amount of losses, and also the population exposed to risk. Both the size and the number of losses are treated in a stochastic manner. The claims are assumed to follow a Poisson process and the claim sizes are independent and identically distributed non-negative random variables. The model proposed is a generalization of the collective risk model, usually applied in practice. The evolution of the population at risk is also stochastically described via a nonlinear hierarchical growth model. Furthermore, a theoretical decision framework is adopted for evaluating the premium. Model selection and premium calculation are obtained from the predictive distribution, incorporating all the uncertainties involved.

JEL classification: C11; C15; IM11; IM30

Keywords: Collective risk model; Health insurance; Aggregate claim amount; Predictive distribution; Monte Carlo Markov chain

1. Introduction

Insurance can be viewed as a strategy for spreading out the losses caused by unexpected events. Generally, the insurance industry is faced with three different operational problems: evaluation of the premium, simulation studies of the future behavior of the risk portfolio and the use of principles of statistical control to reduce unnecessary expenditures. Often the premium risk is calculated based on past experience. As soon as an insurance company is set

* Corresponding author. Tel.: +55 21 2562 8290.
[1] Tel.: +55 21 2562 7513.
E-mail addresses: migon@im.ufrj.br (H.S. Migon), fmoura@im.ufrj.br (F.A.S. Moura).

0167-6687/$ – see front matter
doi:10.1016/j.insmatheco.2004.11.006

up and a portfolio of risk formed, some simulations should be carried out in order to access the future behavior of the plan designed. These simulations must consider the evolution of the portfolio in time and many other uncertainties presented in the model. There is plenty of room to incorporate expert opinions on a subjective basis as well as to mix information from the whole industry using credibility theory. This is quite a popular attitude in actuarial practice, as mentioned in Hogg and Klugman (1984). Therefore, before data are observed, some prior might be assigned to the parameters involved in the premium calculation and then the predictive distribution can be evaluated. After new data are available, the plan must be reassessed and monitored in order to avoid unexpected behaviors (Rosenberg, 2001).

Clearly, uncertainty is a fundamental characteristic of the insurance business, although classical actuarial mathematics was mainly based on deterministic approaches. Even when randomness is taken into account, many actuarial models do not contemplate all the uncertainties involved. For example, the aggregate claim probability distribution function is often used in place of the predictive distribution via numerical approximations of the convolution techniques (Pai, 1997). It is worth pointing out that the predictive distribution can be easily obtained through the Bayesian paradigm.

In this paper, adherence to the Bayesian perspective is adopted, taking into account all sources of uncertainties. The practical and theoretical advantages of this point of view are well known. The reader is referred to Migon and Gamerman (1999) for a fuller account of the Bayesian paradigm and a critical comparison with the frequentist approach. The size and the number of claims are treated in a stochastic manner. The claims are assumed to follow a Poisson process $\{N(t), t \geq 0\}$ with rate $\lambda > 0$ and the claim sizes are independent and identically distributed non-negative random variables, with mean μ and variance σ^2.

Many innovative aspects of actuarial practice are addressed in this work. The model proposed is heavily based on the principles of credibility theory, applied to the collective risk model. It is a generalization of the collective risk model which is usually applied in practice under a frequentist approach. The aggregate claim size distribution introduced in this paper generalizes the exponential case and the insured population under risk is stochastically modeled via a nonlinear hierarchical growth model. Furthermore, a theoretical decision framework for the premium evaluation, which includes a hierarchical modeling of the claim frequency and value, is fully introduced.

An application to the health insurance industry is presented. The main categories of medical events considered are: doctor's office visits, diagnostic exams, physical therapy and in-patient care. The premium for a particular contract must take into account the age of the insured and the various event categories. Therefore, actuaries are concerned with the probabilistic description of all the uncertainties involved in the loss process.

Although the uncertainty in the parameters describing the claim frequency and value distribution could be related to some explanatory variables, for example, some patient medical record, this kind of information is not available in our study and often very difficult to obtain. Therefore, we will only assume that they are stochastically related. In the absence of an individual explanatory variable, the main advantage of hierarchical Bayesian modeling is to borrow strength from the age class, improving the estimation of the model parameters. Nevertheless, if some auxiliary information were available, some improvements on model predictions could be achieved.

The hierarchical Bayesian collective risk model is implemented in WinBugs (Spiegelhalter et al., 2000), showing how easy it is to deal with it in practice.

It is worth pointing out that the Bayesian literature applied to actuarial modeling is growing very fast. Some examples include Makov et al. (1996), DuMouchel (1983), Pai (1997), Herzod (1994), Klugman (1992), Haberman and Renshaw (1996) and many others. A brief survey of Bayesian solutions to some actuarial problems and a description of the current state of research in the area can be found in Makov (2001) and references therein, where the main areas of application and research, involving the Bayesian paradigm, include: experience rating, compound claim modeling and graduation.

In this paper, the insured population and the aggregate claim process are modeled on a parametric basis and computationally intensive methods are used to make all the inferences (see Usabel, 1998). The ergodic average is used to approximate any function of interest of the predictive distribution. The attractiveness of sampling-based