Insurance: Mathematics and Economics 60 (2015) 1–10

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Bayesian nonparametric predictive modeling of group health claims

Gilbert W. Fellingham [a,*], Athanasios Kottas [b], Brian M. Hartman [c]

[a] *Brigham Young University, USA*

[b] *University of California, Santa Cruz, USA*

[c] *University of Connecticut, USA*

HIGHLIGHTS

- Models employed for healthcare data are parametric and relatively inflexible.
- A more flexible, nonparametric model outperforms the current methods.
- Improved predictions for 84% of renewals and 88% of new policies.
- Effective models are important as healthcare costs rise around the world.

ARTICLE INFO

Article history:
Received May 2013
Received in revised form October 2014
Accepted 19 October 2014
Available online 3 November 2014

Keywords:
Dirichlet process prior
Multimodal prior
Forecasting
New business
Risk management

ABSTRACT

Models commonly employed to fit current claims data and predict future claims are often parametric and relatively inflexible. An incorrect model assumption can cause model misspecification which leads to reduced profits at best and dangerous, unanticipated risk exposure at worst. Even mixture models may not be sufficiently flexible to properly fit the data. Using a Bayesian nonparametric model instead can dramatically improve claim predictions and consequently risk management decisions in group health practices. The improvement is significant in both simulated and real data from a major health insurer's medium-sized groups. The nonparametric method outperforms a similar Bayesian parametric model, especially when predicting future claims for new business (entire groups not in the previous year's data). In our analysis, the nonparametric model outperforms the parametric model in predicting costs of both renewal and new business. This is particularly important as healthcare costs rise around the world.

1. Introduction

As George Box famously said, "essentially, all models are wrong, but some are useful" (Box and Draper, 1987). This is especially true when the process being modeled is either not well understood or the necessary data are unavailable. Both are concerns in health insurance. Our knowledge of the human body and understanding of what makes it sick are limited, but the main difficulty is lack of available data; limited by both technology/cost (e.g. DNA sequences and complete blood panels) and privacy (e.g. patient records especially of prospective policyholders). This is even more prevalent in group health where data on the individual policyholders can be sparse. Bayesian nonparametric (BNP) models are a flexible option to describe both current and prospective healthcare claims. As will be shown, in modeling group health claims BNP models are superior to traditional Bayesian parametric models. Both model types could be used in premium calculations for small groups or prospective blocks of business, and to calculate experience-based refunds. Precise estimation is especially important now as healthcare costs continue to consume an increasing share of personal wealth around the world. The importance of proper prediction is exemplified and described in both Klinker (2010) and Harville (2014).

One of the principles of Bayesian methods very familiar to actuaries is improvement in the process of estimating, say, the pure premium for a block of business by "borrowing strength" from related experience through credibility. For example, if the size of a block is small enough, the exposure in previous years may be limited. In this case, estimates of future costs may be based more heavily on other, related experience in an effort to mitigate the effects of small sample random variation. We refer to Klugman (1992) for a thorough review of credibility, especially from a Bayesian perspective.

* Correspondence to: 223H TMCB, Provo, UT 84602, USA. Tel.: +1 801 422 2806; fax: +1 801 422 0635.

E-mail address: gwf@byu.edu (G.W. Fellingham).

http://dx.doi.org/10.1016/j.insmatheco.2014.10.011
0167-6687/

Hierarchical Bayesian models offer an extremely useful paradigm for prediction in this setting. However, in somewhat simplistic terms, successful Bayesian model specification hinges on selecting scientifically appropriate prior distributions. When there is an unanticipated structure in the function defining the prior, posterior distributions (and prediction) will, by definition, be flawed.

This leads us to consider a Bayesian nonparametric model formulation. Bayesian nonparametric methods build from prior models that have large support over the space of distributions (or other functions) of interest. An increased probability of obtaining more precise prediction comes with the increased flexibility of BNP methods. We refer to Dey et al. (1998), Walker et al. (1999), Müller and Quintana (2004), Hanson et al. (2005), and Müller and Mitra (2013) for general reviews on the theory, methods, and applications of Bayesian nonparametrics. We also refer to Zehnwirth (1979) for an early application of BNP methods in credibility. In this paper, we will demonstrate why BNP methods are useful when building statistical models, especially when prediction is the primary inferential objective.

A brief outline of the paper follows. First, we specify the mathematical structure of the models in the full parametric and nonparametric settings. The parametric model is described first since the nonparametric setting parallels and extends the parametric setting. We provide more detail for the nonparametric setting since it is less familiar. Additionally, we provide the algorithms necessary to implement the nonparametric model in the Appendix. We next present a small simulation study to demonstrate the performance of the two models in situations where the structure used to generate the data is known. Finally, we present results from analyses of claims data from 1994 and compare the two formulations by evaluating their performance in predicting costs in 1995.

2. The models

2.1. The hierarchical parametric Bayes model

We present the traditional parametric Bayesian model first since the nonparametric formulation is based on the parametric version. To develop the parametric model, we need to characterize the likelihood and the prior distributions of the parameters associated with the likelihood. There are two things to consider when thinking about the form of the likelihood: the probability a claim will be made and the amount of the claim, given a claim is made. The probability a claim is made differs from group to group and in our data is around 0.70. Thus, about 30% of the data are zeros, meaning no claim was filed for those particular policies. We chose to deal with this by using a likelihood with a point mass at zero with probability π_i for group i. The parameter π_i depends on the group membership.

The cost of a claim given that a claim is paid is positively skewed. We choose a gamma density for this portion of the likelihood with parameters γ and θ. In a previous analysis of this data, Fellingham et al. (2005, p. 11) indicated that "the gamma likelihood for the severity data is not rich enough to capture the extreme variability present in this type of data". However, we will show that with the added richness furnished by the nonparametric model, the gamma likelihood is sufficiently flexible to model the data.

Let $f(y;\gamma,\theta)$ denote the density at y of the gamma distribution with shape parameter γ and scale parameter θ. Hence,

$$f(y;\gamma,\theta)=\frac{1}{\theta^{\gamma}\Gamma(\gamma)}y^{\gamma-1}\exp\left(\frac{-y}{\theta}\right). \tag{1}$$

The likelihood follows using a compound distribution argument:

$$\prod_{i=1}^{N_g}\prod_{\ell=1}^{L_i}\left[\pi_i I(y_{i\ell}=0)+(1-\pi_i)f(y_{i\ell};\gamma_i,\theta_i)I(y_{i\ell}>0)\right], \tag{2}$$

where i indexes the group number, N_g is the number of groups, ℓ indexes the observation within a specific group, L_i is the number of observations within group i, π_i is the proportion of zero claims for group i, θ_i and γ_i are the parameters for group i, $y_{i\ell}$ is the cost per day of exposure for each policyholder, and I denotes the indicator function. Thus, we have a point mass probability for $y_{i\ell}=0$ and a gamma likelihood for $y_{i\ell}>0$.

As discussed in the opening section, the choice of prior distributions is critical. One of the strengths of the full Bayesian approach is the ability it gives the analyst to incorporate information from other sources. Because we had some previous experience with the data that might have unduly influenced our choices of prior distributions, we chose to use priors that were only moderately informative. These priors were based on information available for other policy types. We did not use any of the current data to make decisions about prior distributions. Also, we performed a number of sensitivity analyses in both the parametric and the nonparametric settings and found that the results were not sensitive to prior or hyperprior specification in either case.

For the first stage of our hierarchical prior specification, we need to choose random-effects distributions for the parameters π_i and (γ_i,θ_i). We assume independent components conditionally on hyperparameters. In particular,

$$\begin{aligned}&\pi_i\mid\mu_\pi\overset{\text{ind.}}{\sim}\text{Beta}(\mu_\pi,\sigma_\pi^2),\quad i=1,\ldots,N_g,\\&\gamma_i\mid\beta\overset{\text{ind.}}{\sim}\text{Gamma}(b,\beta),\quad i=1,\ldots,N_g,\\&\theta_i\mid\delta\overset{\text{ind.}}{\sim}\text{Gamma}(d,\delta),\quad i=1,\ldots,N_g.\end{aligned} \tag{3}$$

Here, to facilitate prior specification, we work with the Beta distribution parameterized in terms of its mean μ_π and variance σ_π^2, that is, with density given by

$$\frac{1}{\text{Be}(c_1,c_2)}\pi^{c_1-1}(1-\pi)^{c_2-1},\quad \pi\in(0,1), \tag{4}$$

where $c_1=\sigma_\pi^{-2}(\mu_\pi^2-\mu_\pi^3-\mu_\pi\sigma_\pi^2)$, $c_2=\sigma_\pi^{-2}(\mu_\pi-2\mu_\pi^2+3\mu_\pi^3-\sigma_\pi^2+\mu_\pi\sigma_\pi^2)$, and $\text{Be}(\cdot,\cdot)$ denotes the Beta function, $\text{Be}(r,t)=\int_0^1 u^{r-1}(1-u)^{t-1}\text{d}u$, $r>0$, $t>0$ (Forbes et al., 2011). We choose specific values for the hyperparameters σ_π^2, b, and d and assign reasonably non-informative priors to μ_π, β and δ. We note that sensitivity analyses showed that the values chosen for the hyperparameters had virtually no impact on the outcome. For the prior distributions, we take a uniform prior on $(0,1)$ for μ_π and inverse gamma priors for β and δ with shape parameter equal to 2 (implying infinite prior variance) and scale parameters A_β and A_δ, respectively. Hence, the prior density for β is given by $A_\beta^2\beta^{-3}\exp(-A_\beta/\beta)$ (with an analogous expression for the prior of δ). Further details on the choice of the values for σ_π^2, b, d, A_β, and A_δ in the analysis of the simulated and real data are provided in Sections 3 and 5, respectively.

The posterior for the full parameter vector

$$(\{(\pi_i,\gamma_i,\theta_i):i=1,\ldots,N_g\},\mu_\pi,\beta,\delta)$$

is then proportional to

$$\begin{aligned}&\left[\prod_{i=1}^{N_g}\frac{\beta^{-b}}{\Gamma(b)}\gamma_i^{b-1}\exp\left(\frac{-\gamma_i}{\beta}\right)\frac{\delta^{-d}}{\Gamma(d)}\theta_i^{d-1}\right.\\&\quad\left.\times\ \exp\left(\frac{-\theta_i}{\delta}\right)\frac{1}{\text{Be}(c_1,c_2)}\pi_i^{c_1-1}(1-\pi_i)^{c_2-1}\right]\\&\quad\times\left[\prod_{i=1}^{N_g}\prod_{\ell=1}^{L_i}\{\pi_i I(y_{i\ell}=0)+(1-\pi_i)f(y_{i\ell};\gamma_i,\theta_i)\right.\\&\quad\left.\times\ I(y_{i\ell}>0)\}\right]p(\mu_\pi)p(\beta)p(\delta),\end{aligned} \tag{5}$$