IN THE NAME OF GOD

On Some Bayesian Statistical Models in Actuarial Science With Emphasis on Claim Count

By Maryam Abolghasem Joshaghani

Supervisor: **Dr. Farzad Eskandari**

Advisor: **Dr. Mohammad Reza Salehi Rad**

Allameh Tabatabai'e University

E.C.O College of Insurance

Farvardin 1387

TO

MY PARENTS

ACKNOWLEDGEMENT

MY THANKS GO FIRSTLY TO DR. ESKANDARI MY SUPERVISOR AT ECONOMIC FACULTY OF ALLAMEH TABATABAI'E FOR HIS ASSISSTANCE AND ENCOURAGEMENT WITH THIS THESIS PROJECT.

I AM ALSO GRATEFUL FOR THE ADVICE OFFERED BY DR. SALEHI RAD AT E.C.O COLLEGE OF INSURANCE.

MY EXTREME THANKSGIVINGS GO TOWARD DR. GANJ ALI AND DR. NEMATOLLAHI FOR THEIR HELPS AND JUDGEMENTS.

Content

Abstract

The aim of this thesis is to describe some aspect of insurance issues from the Bayesian point of view. We use advanced computational techniques to estimate posterior models among different distributions for claim counts and also various models for both claim amounts and counts. We construct a flexible kind of Markov Chain Monte Carlo methods, Gibbs Sampler, and implement it in various illustrated examples.

Key Words: Bayesian Theory, Bayesian Approach, MCMC (Markov Chain Monte Carlo) Methods, Noninformative Prior Distribution. Gibbs Sampler, WinBUGS.

Preface

Bayesian inference is a useful and important device because of combining the prior distributions of the parameters and the likelihood functions of the data in evaluating the posterior distributions of parameters. Then mostly, it will give us more precise and comprehensive results. In chapter 1, principles of the Bayesian inference will be introduced. Nevertheless, there are many obstacles in getting posterior distributions in statistical problems, especially when noninfomative prior distributions for parameters are used. So the use of MCMC methods particularly, Gibbs sampling can be beneficial. This subject will be investigated in chapter 2.

In chapter 3, four models for claim amounts and counts in insurance will be explained. Applying these models and other similar models in the field of insurance is a case that has been paid attention in recent years. Implementing these models has been done by WinBUGS software. It is a kind of software that have many applications in the Bayesian inference and can be downloaded freely on the world wide web.

In chapter 4, all the explained models in chapter 3 will get implemented in WinBUGS and some statistics and plots will be presented. Also three statistical distributions; Negative Binomial, Generalized Poisson and Zero Inflated Poisson, in the begging of the chapter are implemented in WinBUGS. All the data are collected from two private and governmental insurance companies (requested not to reveal their names). Rumina (2006) worked on Models 2 and 4 to be described in chapters 3 and 4, theoretically and practically, respectively. She considered missing data and tried to estimate claim amounts and counts for them. The procedure in the presented thesis is different. Here, the parameters of the models and their history plots are estimated and shown, respectively, also residuals for Models 1 and 3 are found. Programming codes especially for models 2 and 4, prior distributions, the proofs for finding the posterior distributions of the models' parameters are different. We wish, the presented thesis would be useful and beneficial.

Chapter 1: The Bayesian Methods

Chapter 1: The Bayesian Methods

1.1 Introduction

The representation of uncertainty about parameters or hypotheses as probabilities is central to the Bayesian inference. Under this framework, we can calculate probability that a parameter lies in a given interval or the probability of a hypothesis about a parameter or set of parameters. Let $P(H_0)$ denote our prior beliefs about the truth of a hypothesis, H_0, for example that the excess relative risk of thrombosis for women taking the pill exceeds 2, with H_1 being that the relative risk was under 2. Suppose our actual data at hand show a relative risk of 3.6. Then the probability or the likelihood function of the data x, given our prior belief is the conditional probability $P(x|H)$, with H denoting H_0 for simplicity. Bayes theorem expresses the updated probability statement about H as $P(H|x) = \frac{P(x|H)P(H)}{P(x)}$ where $P(x)$ is the probability of the data averaged over all possible hypotheses.

Here $P(x)$ would be the probability or the marginal distributions of the data over H_0 and H_1, namely $P(x|H_0)P(H_0) + P(x|H_1)P(H_1)$follows from the rule for the joint probability of x and H, $P(x,H) = P(x|H)P(H) = P(H|x)P(x)$ and the probability $P(H|x)$ denotes our updated or posterior probability beliefs about H_0 given the data. In a sense it pools the prior beliefs with the evidence at hand. We view the model or parameters H as random, and since the divisor $P(x)$ is independent of H, We can re-express the result as $P(H|x) \propto P(x|H)P(H)$.[1]

1.2 The Bayesian Paradigm

In this part we explain Bayesian paradigm more mathematically. In the Bayesian paradigm, one of the interest is a quantity θ where its value is unknown. What is known is a probability distribution $\pi(\theta)$ that expresses our current relative opinion as to the likelihood that various possible values of θ are the true value.

The second item is the probability distribution $f(x|\theta)$. It describes the relative likelihood of various values of X being obtained when the experiment is conducted, given that θ happens to be the true parameter value. This is called the model distribution and is the one element that is common to both Bayesian and classical analysis.

The next step is to use Bayes theorem to compute, $\pi^*(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}$ as a posterior distribution of θ. We represent our opinion after running the experiment and getting new results. Now, the final step is to use the posterior density to draw some conclusions which are available for the problem at hand. We will consider 3 items;

1- The first is a point estimate of one of the parameters in the vector θ. We have $\Theta = (\theta_1, \dots, \theta_k)'$ and the posterior distribution of θ_i is

$$\pi^*(\theta_i|x) = \int \pi^*(\theta_i, \dots, \theta_k|x)\, d\theta_1, \dots, d\theta_{i-1} d\theta_{i+1} d\theta_k$$

2- Secondly, the Bayesian Central Limit Theorem indicates that under suitable conditions the posterior density can be approximated by the normal distribution and so the confidence interval is approximately

$$E(\theta_i|x) \pm Z_{1-\frac{\alpha}{2}} \sqrt{var(\theta_i|x)}$$

The approximation improves as the number of observations increases.

3- Finally, of greater interest is the value of a future observation. Suppose the density of this observation is $g(y|\theta)$. It is not necessary that this density matches to the model produced the observation x, but it must depend on the same parameter θ. The predictive density is then $f^*(y|x) = \int g(y|\theta)\pi^*(\theta|x)d\theta$,and it represents all of our knowledge about a future observation.[6]

1.3 The Prior Distribution and the Likelihood Function

In words the full result is that;

$$Posterior\ distribution = \frac{likelihood\ Function * prior\ distribution}{\sum(likelihood * prior)}$$

where the denominator is a fixed normalizing factor which ensures that the posterior probabilities sum to 1. So,

$$Posterior\ Distribution \propto Likelihood\ Function \times Prior\ Distribution$$

This expression simply states the common-sense principle, the updated knowledge combines prior knowledge with the data at hand. The relative influence of the prior and the data at hand on the updated beliefs depends on how much weight we give to the prior (how "informative" we make it) and the strength of the data. For example a large data sample would tend to have a predominant influence on our updated beliefs unless our prior was extremely specific.

If our sample example was small and combined with a prior which was informative the prior distribution would have a relatively greater influence on the updated belief.

1.4 The Sampling Perspective: Estimation

The modern approach to Bayesian estimation has become closely linked to sampling-based simulation methods. Traditional classical estimation methods are involved to find a single optimum estimate, such as the maximum likelihood estimate. This is equivalent to find the mode of likelihood.

The sampling perspective used in Bayesian estimation instead focuses on estimating the entire density or distribution of the parameter. This density estimation is based on a long run of samples from the posterior density. The samples are of parameters themselves, or of function of parameters. Sampling is continued until the stationary distribution is equivalent to the posterior $\pi^*(\theta|x)$.

If there are n observation and p parameters, then the required number of iterations T from a single sampling chain to reach stationary is typically considerably larger than the size of the sample of observations. It will tend to increase with both p and n, and also the complexity of the model.

A variety of Markov chain Monte Carlo (MCMC) methods have been proposed to sample from posterior densities. There are essentially ways of extending the range of single parameter sampling methods to multivariate situations, where each parameters or subset of parameters in the posterior density may have different densities.

One form of MCMC methods, called Gibbs sampling, samples in turn from each parameter θ in the posterior density, while regarding all other parameters as fixed. This sampling method is the main basis of the BUGS program. In the next chapters, we will expand this subject completely.[1]

1.5 Prediction from Sampling

In classical statistics the prediction of out-of-sample data z often concludes values calculating moments or probabilities from the assumed likelihood for y evaluated at the selected point estimate θ_m, namely $p(y|\theta_m)$.

In the Bayesian method, the information about θ is not constrained in a single point estimate but in the posterior density $p(\theta|y)$and so prediction is corresponded based on averaging $p(y|\theta)$over this posterior density. If the sampling approach is used then the information about θ is contained in a long-run of sampled values from this posterior density. So the prediction of out-of-sample data z given the observed data y is, for θ discrete, the sum $p(z|y) = \sum_{\theta} p(z|\theta)p(\theta|y)$ and is an integral over the product $p(z|\theta)p(\theta|y)$ when θ is continuous.

1.6 Categorical Data Models: Binomial and Poisson Distributions

With categorical rather than continuous data, the major base-line distributions are binomial and Poisson. Categorical data occur when data is only available as recorded in discrete categories. Often originally continuous data may be converted to a discrete categorization.

1.6.1 Binomial Outcomes

With binomial data there is a single parameter of interest, the probability of certain outcome π. The data mechanism distinguishes two possible outcomes as "Success" and "failure", and with probabilities π and $1-\pi$, respectively. The two outcomes of a binomial trial are mutually exclusive and exhaustive. Thus the binomial describes the distribution of x success out of n trials. The binomial density is proportional to the product of probability π over the x success, and of $1-\pi$ over the $n-x$ failure. Thus, we can write;

$p(x|\pi) \propto \pi^x(1-\pi)^{n-x}$ *and* $x \sim bin(n, x)$

The parameter of interest is the probability π, with the x success and $n-x$ failures being the data. One way to represent the size of π is, assigning probabilities to a small number of possible alternative values. But π can have an infinity of values between 0 and 1, and so its prior may also be represented by a continuous density.

For reasons of conjugacy, a convenient prior density for the binomial probability is the beta density with parameters a and b (both positive), denoted $beta(a, b)$ such that; $p(\pi) \propto \pi^{a-1}(1-\pi)^{b-1}$. The posterior density of π is then also a beta with parameters $a + x$ and $b + n - x$ specifically: $p(\pi|x,n) \propto \pi^{a+x-1}(1-\pi)^{b+n-x-1}$. So the parameters of the beta prior density state a successes and b failures. In a beta density with parameters α and β the mean is the ratio $\pi = \frac{\alpha}{\alpha+\beta}$ of total "successes" to total "events". The variance is $var(\pi) = \frac{\pi(1-\pi)}{(\alpha+\beta+1)}$. Accordingly, $\alpha + \beta + 1 = a + b + n + 1$ and so $\alpha + \beta + 1$ is sometimes called the extended sample size. For α and β both larger than 10 the beta density is approximated by a normal curve. For relatively small α and β the approximation is better for π closer to 0.5.

1.6.2 Poisson Distribution for Event Counts

The typical application of the binomial distribution is to take a sample of a given size and count the number of sample members characterized by a certain attribute or not. There are circumstances, when the number of times an event occurs can be counted without there being any notion of counting when the event did not occur, there are also many instances when there is a converse event but if the event is rare then there may be a choice between a binomial or Poisson model:

The greater the rarity of the event, the more appropriate the Poisson becomes. Often the number of events can be seen against an exposure a certain extent (e.g. a population, a geographic area, etc.). For example, other things equal, we would expect the number of new cases of a rare infection to be greater for a large population.

We would have observed x events for a mean λ which is the product of an underlying rate μ and an exposure E, such that $\lambda = \mu E$. Usually E is assumed known. (i.e. a fixed constant). Then we conclude that $x \sim Poisson(\lambda)$. So $x \sim Poisson(\mu E)$, where $\lambda = \mu E$, with μ to be estimated.

Hence the likelihood of x events can be seen to be proportional to $e^{-\lambda}\lambda^{x}$. At a known exposure level, with $\lambda = \mu E$, then this is in turn proportional to $e^{-\mu E}\mu^{x}$, since $E^{x}$is a constant. If we observe events $x_1, x_2, x_3, \dots, x_k$ for a sample of size k then the likelihood over the sample will be proportional to $e^{-k\lambda}\lambda^{T}$, where $T = \sum_{i=1}^{k} x_i = k\bar{x}$.

If we observe events $x_1, x_2, x_3, \dots, x_k$ corresponding to fixed exposures $E_1, E_2, \dots, E_k$ and we denote $\varepsilon = \sum_i E_i$ as the total exposure in the sample, then the likelihood is proportional to $e^{-\varepsilon\mu}\mu^{T}$.

If we adopt a gamma $G(a,b)$ for λ, such that $p(\lambda) \propto \lambda^{a-1}e^{-b\lambda}$ then the posterior density for λ will be of a gamma form $G(a+T, b+\varepsilon)$. If we assume a $G(a,b)$ prior for μ, then the posterior density for μ will be of the form $G(a+T, b+\varepsilon)$.

Totally, if one wishes to estimate a probability of survival, classical statisticians would consider it legitimate to use prior experience and information to form the opinion that for example a binomial or Poisson model is appropriate for the experiment but that it is illegitimate to use prior experience and information to form a prior opinion about the probability of survival. This is the dictum regardless of the size of the experiment relative to the previous experience and information. Bayesian statistics goes two steps beyond this classical dictum in that it holds that it is not only legitimate to form an opinion based upon the prior information and experience, but mandatory to do so. The Bayesian statistician views the experimental data as evidence to be assimilated into the experience and knowledge of the experimenter.[1]

1.7 Analysis Through the Bayesian Approach

Suppose that an insurance company had decided to issue a new policy but hesitates to put it into issue to the market because sales volume ($\tilde{\mu}$) per sales policyholders may not be sufficient to cover the cost. Management therefore decides to carry out some research by taking a sample of potential policyholders. The final decision whether or not to go into issue will turn primarily on the evidence obtained from this sample, but will also incorporate any additional prior evidence of sales volume that was obtained on the basis of marketing judgment.

Suppose that the executive making the decision puts the latter information in the from of a probability distribution of guesses, say Normal with mean μ_0 and standard deviation σ_0. The sample is now taken and it is assumed that it is wished to revise the distribution of $\tilde{\mu}$ to take into account of the new evidence from the sample, which may be that the sample mean $\bar{x}$ is assumed Normal with mean $\tilde{\mu}$ and standard deviation σ_1. Then, using Beyes's theorem, the posterior distribution of $\tilde{\mu}$ will be:

$$\theta = \frac{exp\left(-\frac{1}{2}\left\{\left(\frac{\bar{x}-\tilde{\mu}}{\sigma_1}\right)^2 + \left(\frac{\tilde{\mu}-\mu_0}{\sigma_0}\right)\right\}\right)}{2\pi\sigma_0\sigma_1}$$

$$p(\tilde{\mu}|\bar{x}) = \frac{1}{2\pi\sigma_0\sigma_1} \frac{exp\left(-\frac{1}{2}\left\{\left(\frac{\bar{x}-\tilde{\mu}}{\sigma_1}\right)^2+\left(\frac{\tilde{\mu}-\mu_0}{\sigma_0}\right)\right\}\right)}{\int \theta d\bar{x}}$$

And hence $p(\tilde{\mu}|\bar{x})$ is Normal with $mean = \frac{\mu_0\sigma_1^2+\bar{x}\sigma_0^2}{\sigma_1^2+\sigma_0^2} = \left(\frac{\mu_0/\sigma_0^2+\bar{x}/\sigma_1^2}{1/\sigma_0^2+1/\sigma_1^2}\right)$

and $standard\ deviation = \sqrt{\frac{\sigma_0^2\sigma_1^2}{\sigma_0^2+\sigma_1^2}}$

Notice that the posterior mean is a weighted average of the prior mean and the sample mean, the weights being the reciprocals of the variances of the two distributions.

Now, suppose that; $\mu_0 = 18,\ \sigma_0^2 = 3.0$ and that a random sample of 100 from an estimated population of 20000 gives $\bar{x} = 15, s = 12$ and $\sigma_1 = \frac{12}{\sqrt{100}} = 1.2$

Thus the posterior distribution of $\tilde{\mu}$ is Normal with parameters

$$mean = \frac{(18/_{3.0})^2+(15/_{1.2})^2}{(1/_{3.0})^2+(1/_{1.2})^2} = 15.4.$$

$$and\ standard\ deviation = \sqrt{\frac{(3.0)^2(1.2)^2}{(3.0)^2+(1.2)^2}} = 1.113$$

the striking thing about this example is the fact that the evidence supplied by the sample has virtually overwhelmed the executive's original beliefs about the sales of the new product.

1.8 The Effect of the Prior Distribution

Now, it is worth considering whether the shaped assumed for the prior distribution is a critical factor. So, two extremely contrasting prior distributions are considered, namely:

a) Normal, mean 1 and standard derivation 1,