ECO College of Insurance
Allameh Tabatabai University

A Bayesian Approach to Bonus Malus System

An Application in Automobile insurance

By:

Fatemeh Atatalab

Supervisor:

Dr. Amir Teimour Payandeh

Advisor:

Dr. Ghadir Mahdavi

(Date: February – 2012)

Acknowledgements

I would like to thank all those who gave me valuable help to complete my thesis. I would like to thank my supervisor Dr. Amir Teimour Payandeh; for his constant support during this research. Without his support this thesis couldn’t be accomplished. I am thankful to my adviser Dr. Ghadir mahdavi who help me to go ahead with my thesis. I am also grateful to Dr. Reza Ofoghi, the examiner of my thesis, for his helpful comments.

Of course, I am grateful to my parents for their patience and love; my kind brother and dear sister who supported me during my graduation. Without their support this work would never have come into existence.

Tehran, Iran   Fatemeh

Atatalab

February, 2012

Table of content

Abstract

With the deregulation of BMS, it is important to obtain rules in order to transfer policyholders from one BMS to another. This thesis proposes a solution to this problem. It will be shown in this study that an insured who is transferred from a BMS to another one in which level will stay in new system.

We contributed to the work of Pitrebois et al. (2006) by applying to Iranian Auto Data. The aim of this thesis is to show how to develop rules allowing to transfer a policyholder from a given BMS to another one.

We follow the Norberg method to calculate the relativity premium and use it to obtain measure of discrepancy between two systems. The losses considered in this study are the standard quadratic loss, exponential loss and LINEX loss functions. By a numerical example we show that using exponential loss function and LINEX loss function leads to new levels which are lower than the levels created by using quadratic loss functions. Also this thesis uses credibility factor based on Payandeh (2010) and obtains it for Poisson-Gamma and Zero-inflated Poisson gamma distribution under several loss functions.

We observe that the different metrics we have chosen do not provide extremely different results.

Using LINEX loss function, the levels in which policyholder is situated are reduced. As much as a is away from zero, the policyholder with LINEX loss function is transferred to lower levels. This means that they will pay lower premiums than when the loss function is exponential or quadratic. When a LINEX loss is used, the size of maluses is reduced.

Keywords: Credibility factor, LINEX loss function, Balance-quadratic loss function, Zero inflated Poisson gamma distribution, transition rules

Chapter 1

Preliminaries

In this Chapter, we collect some essential elements for other chapters of this thesis. A considerable part of the chapter collected from Bailey (1945, 1950), Bühlman (1967), Varian (1975), Lemaire (1979), Willmot (1987), Tremblay(1992), Holtan (1994), Zellner (1994), Lemaire (1995), Coene (1996), Ferreira (1997), Denuit & Dhaene (2000), Frangos & Vrontos (2000), Press (2003), Morillo & Bermúdez (2003), Pitrebois et al. (2005), Payandeh (2010).

1.1 Bonus-Malus system

BMS is a merit-rating technique third party liability automobile insurance used in most of Europe, Asia and some Latin American and African countries. Depending on the risks allocated to them, the policyholders are divided into a finite number of bonus-malus classes. Based on their claims histories, their classes are modified at each renewal. When a rating system is used, the amount of premium paid by the insured depends on the rating factors of the current period but also on claim history. In practice, a BMS consists of a finite number of levels, each with its own relative premium. New policyholders enter to a specified level. After each year, the policy goes up or down according to transition rules and to the number of claims filed. The premium charged to a policyholder is obtained by applying the relative premium associated to his current level in the system.

BMS allows to match individual premium to risk and to increase incentives for road safety, by taking into consideration the past record. They can be justified by asymmetrical information between the insurance company and the policyholders. Indeed, they encourage policyholders to drive carefully (i.e., they counteract moral hazard) and respond to adverse selection in automobile insurance. Adverse selection occurs whenever the policyholders have a better knowledge of their claim behavior than the insurer does, and take advantage of this additional information about their driving patterns, known to them but unknown to the insurer. Experience rating is a response to adverse selection, by penalizing the more numerous claims of those with more dangerous driving patterns. BMS is in general independent of the claim amount; a crucial issue for the insured is therefore to make a decision whether it is beneficial or not to report small claims (in order to avoid an increase in premium). Small size claims are likely to be defrayed by the policyholders themselves, and not to be reported to the company.

This phenomenon, known as the hunger for bonus, limits claim handling costs since small claims are not reported to the insurer (decreasing the administrative burden). BMS can be developed using the theory of Markov Chains. This has already been done by Loimaranta (1972), Norberg (1976), Borgan, Hoem & Norberg (1981), Gilde & Sundt (1989). They assumed that a BMS can be dealt under the framework of homogeneous Markov Chains.

Lemaire (1979, 1985, and 1995) developed the model to obtain a financially balanced BMS using the expected-values premium calculation principle and the Negative Binomial as the claim frequency distribution. Panjer (1987) proposed the Generalized Poisson-Pascal distribution (in fact, the Hofmann distribution), which includes three parameters, for the modeling of the number of automobile claims. Ruohonen (1987) considered a model for the claim number process. This model is a mixed Poisson process with a three-parameter Gamma distribution as the structure function and is compared with the two-parameter Gamma model giving the Negative Binomial distribution.

Willmot (1987) compared the Poisson-Inverse Gaussian distribution to the Negative Binomial one and concluded that the fits are superior with the Poisson-Inverse Gaussian in all the six cases studied by Gossiaux & Lemaire (1981), see also Besson & Partrat (1990). Tremblay (1992) used the Poisson-Inverse Gaussian distribution. He discussed an optimal BMS considering the quadratic loss function and also using zero-utility premium calculation principle. Tremblay (1992) obtained the design of an optimal BMS using the quadratic error loss function, the Poisson-Inverse Gaussian as the claim frequency distribution and the zero-utility premium calculation.

Holtan (1994) introduced an alternative approach to BMS that eliminates these disadvantages. He suggested the use of very high deductibles that may be borrowed by the insured to the insurance company. Although technically acceptable, this approach obviously causes considerable practical problems. While Holtan (1994) assumed a high deductible which is constant for all insureds, and thus independent of the level they occupied in the BMS at the claim occurrence time. Lemaire (1995) used the quadratic loss function and negative binomial as acclaim frequency distribution for determining the optimal BMS. Coene (1996) introduced a financially balanced BMS and he used a quadratic loss function of the differences between the premiums for an optimal BMS with an infinite number of classes, weighted by the stationary probability of being in a specific class and by imposing variance constrain on the system. . Coene & Doray (1996) considered the design of an optimal BMS by minimizing a quadratic function of the difference between the premium for an optimal BMS with an infinite number of classes.

Denuit (1997) demonstrated that the Poisson-Goncharov distribution introduced by Lefèvre & Picard (1996) provides an appropriate probability model to describe the annual number of claims incurred by an insured motorist. Meng & Whitmore (1999) presented a model which employed the Negative Binomial distribution for individual-level claims and Pareto distribution as the distribution for claim propensities within the portfolio. Walhin & Paris (1999) developed an optimal BMS using a finite Poisson mixture as a claim frequency distribution and also using the Negative Binomial and the Poisson-Inverse Gaussian.

More recently, Frangos & Vrontos (2001) obtained an optimal BMS based both on the number of accidents of each policyholder and on the size of loss for each accident that occurred.

Centeno et al. (2001) use non-homogeneous Markov Chains to model such system. They used open models and explain the open model as close to reality. They said that the closed model overvaluates the probabilities of the extreme classes. When the linear scale is used, we obtain a smaller premium in all the classes for the open model with expectation of class 1 and 2, see Ceneto, et al (1999) for more detail. Pitrebois et al. (2005) developed the model assuming varying deductibles which changes from each level to the next. In this model, the premium correction in the malus zone could be replaced by deductibles because of motivating the insured to be careful about risky situation and preventing them leaving the insurance company after claim. Combining BMS with varying deductibles presents a number of advantages: **1-** According to signal theory, insureds choosing varying deductible should be good drivers. **2-** Even if the insured leaves the company after a claim, he has to pay for the deductible.

Pitrebois et al. (2006) have argued that it is nonsense to oblige insurance companies to use the same bonus-malus scale whenever they used different a priori segments. They try to show how to transfer policyholders from a BMS to the other one? For this aim, they used distance formula and earn minimum result for each system.

1.2 Bayesian model

Bayesian ideas and techniques were introduced into actuarial science in a big way in the late 1960s when the papers of Bühlman (1967, 1969) and Bühlman & Straub (1970) laid down the foundation to the empirical Bayes credibility approach, which is still being used extensively in the insurance industry. Because risk classification in insurance involves unobserved risk characteristics, Bayesian modeling offers an intellectually acceptable approach. Indeed, these characteristics are usually modeled by the introduction of a random effect in the classification process. Consequently, a posteriori analysis following claims experience is an interesting task because a Bayes revision of the heterogeneity component allows estimating more precisely these unobserved characteristics. At each insured period, the random effects can be updated for past claim experience, revealing some individual information. Bayesian analysis is used to design an optimal BMS with infinite number of classes at first. Based on the distribution of the number of claim in the portfolio, the posterior claim frequencies are obtained. These claim frequencies are then used to calculate the posterior premiums.

Bayes' theorem is simply a restatement of the conditional probability. Suppose that $A_1, \dots, A_k$ is any set of mutually exclusive and exhaustive events, and that events B and A_j are of special interest. Bayes' theorem for events provides **a** way to find the conditional probability of A, given B in terms of the conditional probability of B given A_j **.** For this reason, Bayes' theorem is sometimes called a theorem about "inverse probability." Bayes' theorem for events is given by:

$$P\{A_j|B\} = \frac{P\{B|A_j\}P\{A_j\}}{\sum_{i=1}^{k} P\{B|A_i\}P\{A_i\}}.$$

For $P\{B\} \neq 0$. $P\{A_j\}$ is your personal prior probability of event A_j. It is your degree of belief about event A_j prior to your having any information about event B that may bear on A_j. $P\{A_j|B\}$ denotes your posterior probability of event A_j in that it is your degree of belief about event A_j posterior to you having the information about B. This equation is a general form of Bayes' theorem for events. Suppose $\boldsymbol{C}$ denote any event, and $\bar{C}$ **.** Denote the complementary event to $\boldsymbol{C}$ ***(C*** and $\bar{C}$ are mutually exclusive and exhaustive). Then, for any other event B, $P\{B\} \neq 0$, Bayes' theorem for events becomes:

$$P\{C|B\} = \frac{P\{B|C\}P\{C\}}{P\{B|C\}P\{C\} + P\{B|\bar{C}\}P\{\bar{C}\}}.$$

This is the Bayes' Theorem for Complementary Events.

Prior and Posterior distribution

The prior distribution is a key part of Bayesian inference (see Bayesian methods and modeling) and represents the information about an uncertain parameter ɵ that is combined with the probability distribution of new data yield the posterior distribution, which is turn is used for future inferences and decisions involving ɵ.

The key issues in setting up a prior distribution are:

1- What information is going into the prior distribution?

2- The properties of the resulting posterior distribution.

According to Bayesian rule, we can express posterior probability of certain event H given some data with the formula

$$P(H|data) = \frac{P(data|H)P(H)}{P(data)}.$$

The probability of H given the data is called the posterior probability of H.

The posterior distribution summarizes the current state of knowledge about all the uncertain quantities (including unobservable parameters and also missing, latent, and unobserved potential data) in a Bayesian methods and modeling).

1.3 Trajectory

New drivers start in level l_0 of the scale. Experienced drivers arriving in the portfolio are not necessarily placed in level l_0 , but in a level corresponding to their claim history or to the level occupied in the BMS used by a competitor. As referred to Denuit et al (2007) the trajectory of the insured in the BMS is modeled by a sequence $\{L_1, L_2, \dots\}$ of random variable valued in{0, 1, …..,s}, such that L_t is the level occupied during the (t+1)-th year, i.e. during the time interval (t, t+1). Since movements in the scale occur once a year, the insured occupies level L_t from time t until time t+1. Once the number N_t of claims reported by the insured during (t−1, t) is known, this information is used to evaluate again the position of the driver in the system. The $L_t\,s$ obviously depend on the past numbers of claims $N_1, N_2, \dots, N_t$ reported by the insured.

$$L_t = \max\{\min\{L_{t-1}(1 - I_t) + N_t \times penalty\ induced\ by\ each\ claim\ , s\}, 0\}$$

Where

$$I_t = \begin{cases} 1, & if \quad N_t \geq 1 \\ 0, & otherwise. \end{cases}$$

indicates whether at least one claim has been reported in year t. This representation of the $L_t\,s$ clearly shows that the future trajectory of the insured in the scale is independent of the levels occupied in the past, provided that the present level is given.

1.4 Transition rules

The probability of moving from one level to another depends on the number of claims reported during the current year. Therefore, we can introduce more formally the transition rules which impose the transfer from one level to another level once the number of claims is known. If k Claims are reported,

$$t_{ij} = \begin{cases} 1, & if\ the\ policy\ gets\ transferred\ from\ level\ i\ \ to\ j\ , \\ 0, & otherwise. \end{cases}$$

The $t_{ij}(k)s$ are put in matrix from $T(k)$ i.e.

$$T(k) = \begin{pmatrix} t_{00}(k) & \cdots & t_{0s}(k) \\ \vdots & \ddots & \vdots \\ t_{s0}(k) & \cdots & t_{ss}(k) \end{pmatrix}$$

Then, $T(k)$ is a 0-1 matrix having in each row exactly one 1. (See Denuit et al., 2007).

1.5 Transition probability

It is assumed that *N1*, *N2*,… are independent and for example poison distributed with parameter ν. The trajectory will be denoted as $\{L_1(\nu), L_2(\nu), \dots\}$ to emphasize the dependence upon the annual expected claim frequency. Let *Pℓ1ℓ2*(υ) be the probability of moving from level $l_1$to level l_2 for a policyholder with annual mean claim frequency ν, that is,

$$P_{l_1 l_2} = P[L_{k+1}(\nu) = l_2 | l_k(\nu) = l_1$$

With $l_1, l_2 \in \{0,1,2, \dots, s\}$. clearly, the $p_{l_1 l_2}(\nu)$ satisfy

$$P_{l_1 l_2} \geq 0 \quad for\ all\ \ l_1\ and\ \ l_2\ \ , and \quad \sum_{l_2=0}^{s} P_{l_1 l_2}(\nu) = 1.$$

Moreover, the transition probabilities can be expressed using the $t_{ij}(.)s$ introduced above. To see this, it suffices to write

$$P_{l_1 l_2}(\nu) = \sum_{n=0}^{+\infty} P[L_{k+1}(\nu) = l_2 | N_{k+1} = n, L_k(\nu) = l_1] P[N_{k+1} = n | L_k(\nu) = l_1]$$

$$= \sum_{n=0}^{\infty} \frac{\nu^n}{n!} \exp(-\nu)\, t_{l_1 l_2}(n).$$

Note that, we have used the fact that N_{k+1} and $L_k(\nu)$ are independent (since $L_k(\nu)$ depends on $N_1, \dots, N_k$, so that

$$P[N_{k+1} = n | L_k(\nu) = l_1] = P[N_{k+1} = n] = \frac{\nu^n}{n!} \exp(-\nu).$$

The transition probabilities allow the actuary to compute the probability of any trajectory in the scale. Specifically, since the probability that a certain policyholder with expected annual claim frequency ν is in level $l_1, \dots, l_n$ at time 1,…,n is simply the probability of going from l_0 to l_n via the intermediate levels $l_1, \dots, l_{n-1}$, we have

$$P[L_n(\nu) = l_1, \dots, L_n(\nu) = l_n | L_0(\nu) = l_0] = P_{l_0 l_1}(\nu) \dots P_{l_{n-1} l_n}(\nu).$$

Furthermore, it is enough to know the current position in the scale to determine the probability of being transferred to any other level in the bonus-malus scale. Formally,

$$P[L_n(\nu) = l_n | L_{n-1}(\nu) = l_{n-1}, \dots, L_0(\nu) = l_0] = P_{l_{n-1} l_n}(\nu),$$

Whenever $P[L_{n-1}(\nu) = l_{n-1}, \dots, L_0(\nu) = l_0] > 0$.

1.6Transition Matrix

Further, $P(\nu)$ is the one-step transition matrix, i.e.

$$\mathrm{P}(\nu) = \begin{pmatrix} \mathrm{P}_{00}(\nu) & \cdots & \mathrm{P}_{0s}(\nu) \\ \vdots & \ddots & \vdots \\ \mathrm{P}_{s0}(\nu) & \cdots & \mathrm{P}_{ss}(\nu) \end{pmatrix}.$$

As already mentioned the future level of a policyholder is independent of its past levels and only depends on its present level and also on the number of claims reported during the present year. In matrix form, we can write $P(\nu)$ as

$$P(\nu) = \sum_{k=0}^{\infty} \frac{\nu^k}{k!} \exp(-\nu)\, T(k)$$

provided the $N_t s$ are independent and Poisson (ν) distributed.

The Multi-Step Transition Probability evaluates the likelihood of being transferred from level *i* to level *j* in n steps.

$$P_{ij}^{(n)}(\nu) = P[L_{k+n}(\nu) = j | L_k(\nu) = i]$$

Note that this is the probability that $L_{k+n}(\nu) = j$ given $L_k(\nu) = i$ for any k. The process describing the trajectory of the policyholder across the levels is thus stationary. From

$$P_{ij}^{(n)}(\nu) = \sum_{i_1=0}^{s}\sum_{i_2=0}^{s} \dots \sum_{i_{n-1}=0}^{s} P_{ii_1}(\nu)P_{i_1 i_2}(\nu) \dots P_{i_{n-1}j}(\nu),$$

It is clear that it includes all the possible paths from *i* to *j* and the probability of their occurrence. This is the n-step transition probability $P_{ij}^{(n)}(\nu)$Therefore, the matrix

$$P^{(n)}(\nu) = \begin{pmatrix} P_{00}^{(n)}(\nu) & \cdots & P_{0s}^{(n)}(\nu) \\ \vdots & \ddots & \vdots \\ P_{s0}^{(n)}(\nu) & \cdots & P_{ss}^{(n)}(\nu) \end{pmatrix}$$

is called the n-step transition matrix corresponding to$P(\nu)$. The following result shows that $P^{(n)}(\nu)$ is a stochastic matrix, being the *n*-th power of the one-step transition matrix $P(\nu)$.

Property 1.1 For all $n, m = 0,1, \dots,$

$$P^{(n)}(\nu) = P^{(n)}(\nu).$$

And hence

$$P^{(n+m)}(\nu) = P^{(n)}(\nu)P^{(m)}(\nu),$$

See. Denuit et al. (2007) for more details.

1.7 Poisson distribution

Poisson distribution plays prominent roles in modeling discrete count data, mainly because of its descriptive adequacy as a model when only randomness is present and the underlying population is homogeneous. Unfortunately, this is not a realistic assumption to make in modeling many real insurance data sets. Poisson mixtures are well-known counterparts to the simple Poisson distribution for the description of inhomogeneous populations. The problem of unobserved heterogeneity arises because differences in driving behavior among individuals cannot be observed by the actuary. One of the well-known consequences of unobserved heterogeneity in count data analysis is overdispersion: the variance of the count variable is larger than the mean. Apart from its implications for the low-order moment structure of the counts, unobserved heterogeneity has important implications for the probability structure of the ensuing mixture model.

The Poisson distribution was discovered by Siméon-Denis Poisson (1781–1840). A Poisson random variable is a count of the number of events that occur in a certain time interval or spatial area. For example, the number of cars passing a fixed point in a five-minute interval, or the number of claims reported to an insurance company by an insured driver in a given period. A typical characteristic associated with the Poisson distribution is certainly equidispersion: the variance of the Poisson distribution is equal to its mean. Assume x is identically independent distributed random variable (iid). And x is distributed as Poisson distribution with parameter λ, so the probability density function is

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}.$$

The moment and the maximum likelihood estimators for parameter λ would be the mean of sample. On the other hand $\hat{\lambda} = \bar{x}$.

Poisson-gamma model

The simplest random effects model for count data is the Poisson distribution is often reasonable description for events which occur both “randomly and independently” in time. The gamma distribution can be used to express the joint distribution in a closed form. The reason that why we choose gamma distribution as prior distribution is that: The gamma distribution family is a flexible family that allows us to represent any prior knowledge and its support coincides with the parameter space of α’s. Given $\Theta_i = \theta_i$, the annual numbers of claims N_i is given by a Poisson distribution with parameter $\lambda_i\theta_i$. θ_i Superposed to the annual claims frequency and $E(\theta_i) = 1$. So we have

$$P\left(N_i = k \middle| \Theta_i = \theta_i\right) = \exp(-\lambda_i\theta_i)\frac{(\lambda_i\theta_i)^k}{k!} \quad , \ k \in N \ , \qquad i = 1,2,\dots,m.$$

All the θ_i’s are assumed to be independent and follow a standard Gamma distribution with probability density function:

$$f_{\Theta}(\theta) = \frac{1}{\Gamma(a)} a^{a}\theta^{a-1}\exp(-a\theta) \quad , \theta \in R^{+}.$$

From this, we conclude that the number of claims of a randomly picked up policyholder from the portfolio is Negative Binomial distributed:

Multivariate zero-inflated Poisson-gamma

Count data frequently over dispersion and excess zeros, which motivates zero-inflated count models (Lambert 1992; Greene 1994). Zero-inflated count models offer a way of modeling the excess zero in addition to allowing for over dispersion in a standard parametric model. Zero inflation arises when one mechanism generates only zeros and the other process generates both zero and nonzero counts. A (discrete) random variable Y_i is said to have a zero-inflated distribution if it has value 0 with probability Φ, otherwise it has some other distribution with P(Y=0)>0. Hence P(Y=0) comes from two sources, and the Φ sources can sometimes be thought of as a structural zero. The most common example of a zero-inflated distribution is the zero-inflated Poisson. It has value 0 with probability Φ else is Poisson (λ) distributed. Zero-inflated count data models have generic probability function:

$$f(y) = \frac{(1-\Phi)e^{-\lambda}\lambda^{y}}{y!} \qquad , y = 1,2, \ldots.$$

$$f(0) = \Phi + (1+\Phi)\mathrm{e}^{-\lambda} \qquad \mathrm{y} = 0$$

ZIP models can be considered as a mixture of a zero point mass and Poisson distribution and where first use to study soldering defects on print wiring boards (Lambert 1992) count data may also exhibit a great number of zeros than expected from the Poisson model. The zero inflated Poisson (ZIP) model is commonly used in modeling data with excess zero. It is a mixture of Poisson and a degenerate distribution at zero. Lambert (1992) used the ZIP in modeling a manufacturing process. However, count data usually exhibit the joint presence of excess zero counts and over dispersion. In this event, the zero inflated negative binomials (ZINB) distribution provides a better fit. . Gupta et al. (1996) introduced zero adjusted generalized Poisson distribution. Hall (2000) described the zero-inflated binomial (ZIB) regression model and incorporated random effects into ZIP and ZIB models. And Lee et al. (2001) generalized the ZIP model to accommodate the extent of individual exposure. See Yip & Yau (2005) for an application to insurance claim count data. Yau and Yip presented the ZIP, ZINB, zero inflated generalized double Poisson (ZIDP) to accommodate the excess zero for insurance claim data

1.8 Poisson regression

The Poisson (log linear) regression model is the most basic model that explicitly takes into account the nonnegative integer-valued aspect of the dependent count variable. In this model, the probability of an event count y_i, given the vector of covariates x_i, is given by the Poisson distribution:

$$\Pr(Y_i = y_i|x_i) = \frac{e^{-\mu_i}{\mu_i}^{y_i}}{y_i!} \quad y_i = 0,1,2, \dots$$

The mean parameter μ_i(the conditional mean number of events in period i) is a function of the vector of covariates in period i:

$$E(y_i|x_i) = \mu_i = \exp\,(x_i\acute{B}).$$

The name log-linear is also used for the Poisson regression model because the logarithm of the conditional mean is linear in the parameters:

$$Ln[E(y_i|x_i)] = \ln(\mu_i) = \acute{x}_iB.$$

The Poisson regression model assumes that the data are equally dispersed- that is, that the conditional variance equals the conditional mean. Poisson regression models are generalized linear models with the logarithm as the (canonical) link function, and the Poisson distribution function.

1.9 Generalized linear model

Generalized linear models (GLM) have been used for actuarial purposes by several authors (see, e.g. Haberman & Renshaw (1996) and references therein). In the framework of actuarial applications, there are several attempts to use GLM in order to describe the claim frequency, the claim size or other characteristics of a portfolio. Generalized linear modeling is used to assess and quantify the relationship between a response variable and explanatory variables. The modeling differs from ordinary regression modeling in two important respects:

- The distribution of the response is chosen from the exponential family. Thus the distribution of the response need not be normal or close to normal and may be explicitly non-normal.
- A transformation of the mean of the response is linearly related to the explanatory variables.

The insurance company achieves risk classification using generalized linear models (Poisson or logistic regressions, for instance). Generalized linear models are important in the analysis of insurance data. With insurance data, the assumptions of the normal model are frequently not applicable.

Whenever the counts are small, which is typically the case in automobile insurance, the Normal approximation is poor and fails to account for the discreteness of the data. Normal regression should be avoided in this case. Generalized linear models provide an appropriate framework for the analysis of count data. A linear model for the logarithm of the λ_i ‘s is often used in actuarial science (see e.g. Pinquet, 1997). This provides a regression model for count data analogous to the usual normal regression for continuous data According to standard methodology of generalized linear models, the logarithmic function is also the natural link for the Poisson distribution (see e.g. Dobson, 1990). The general linear model (GLM) is a statistical linear model. It may be written as

$$Y = XB + U$$

where **Y** is a matrix with series of multivariate measurements, **X** is a matrix that might be a design matrix, **B** is a matrix containing parameters that are usually to be estimated and **U** is a matrix containing errors or noise. The errors are usually assumed to follow a multivariate normal distribution. If the errors do not follow a multivariate normal distribution, generalized linear models may be used to relax assumptions about **Y** and **U**. In a GLM, each outcome of the dependent variables, **Y**, is assumed to be generated from a particular distribution in the exponential family, a large range of probability distributions that includes the normal, binomial and Poisson distributions, among others. The mean, $\boldsymbol{\mu}$, of the distribution depends on the independent variables, **X**, through:

$$W(Y) = \mu = g^{-1}(X\boldsymbol{\beta})$$

where E(**Y**) is the expected value of **Y**; **X**$\boldsymbol{\beta}$ is the *linear predictor*, a linear combination of unknown parameters, $\boldsymbol{\beta}$; g is the link function. The link function provides the relationship between the linear predictor and the mean of the distribution function.

1.10 Loss functions

Whatever the model selected for the number of claims, the *a posteriori* premium correction is derived from the application of a loss function. The standard choice is a quadratic loss. In this case, the credibility premium is the function of past claim numbers that minimizes the expected squared difference with the next year claim number. It is well known that the solution is given by the *a posteriori* expectation.

The penalties obtained in a credibility system calling upon a quadratic loss function are often so severe that it is almost impossible to implement them in practice, mainly for commercial reasons. In order to avoid this problem, some authors have proposed resorting to an exponential loss function: the hope is that breaking the symmetry between the overcharges and the undercharges leads to reasonable penalties. This reduces the *maluses* and the *bonuses*, and results in a financially balanced system.

Early references about the use of this kind of loss function include Ferreira (1977) and Lemaire (1979). Adopting the semiparametric model proposed in Young (1997, 2000) but considering that the piecewise linear function has better characteristics in simplicity and intuition than the kernel, Huang, Song & Liang (2003) used the piecewise linear function as the estimate of the prior distribution and to obtain the estimates for the credibility formula.

Young (1998a) uses a loss function that is a linear combination of a squared-error term and a second-derivative term. The squared-error term measures the accuracy of the estimator, while the second-derivative term constrains the estimator to be close to linear.

Young (2000) resorts to a loss function that can be decomposed into a squared-error term and a term that encourages the credibility premium to be constant. This author shows that by using this loss function, the problem of upward divergence noted in Young (1997) is reduced. See also Young (1998b). Young (2000) also provides a simple routine for minimizing the loss function, based on the discussion of De Vylder in Young (1998a). See also Young & De Vylder (2000), where the loss function is a linear combination of a squared-error term and a term that encourages the estimator to be close to constant, especially in the tails of the distribution of claims, where Young (1997) noted the difficulty with her semiparametric approach. The quadratic loss function is by far the most widely used in practice. The results with the exponential loss function are taken from Bermúdez, et al. (2001).

Morillo & Bermúdez (2003) used an exponential loss function in connection with the Poisson-Inverse Gaussian model. Other loss functions can be envisaged.

LINEX loss functions

One advantage of using the squared or exponential loss function is that is penalizes overestimation or underestimation. Overestimate of a parameter can lead to more or less severe consequences than underestimation. The use of an asymmetrical loss function, which ascribes greater importance to overestimation or underestimation, can be considered for the estimation of the parameters. The asymmetric LINEX loss function put forth by Varian (1975). The loss function is defined, for b>0 and a≠0, by:

$$L_{Linex} = exp\{a(\hat{\theta} - \theta)\} - a(\hat{\theta} - \theta) - 1.$$

This loss function is depicted in figures 1-1a and b. for $\hat{\theta} - \theta = 0$, the loss is zero. For a>0, the loss declines almost exponentially for $(\hat{\theta} - \theta) > 0$, and rises approximately linearly when $(\hat{\theta} - \theta) < 0$. For a<0, the reverse is true. It is straightforward to find that the Bayes estimator is given by

$$\hat{\theta} = \theta_{Bayes} = \frac{-1}{a} logE(e^{-a\theta}).$$

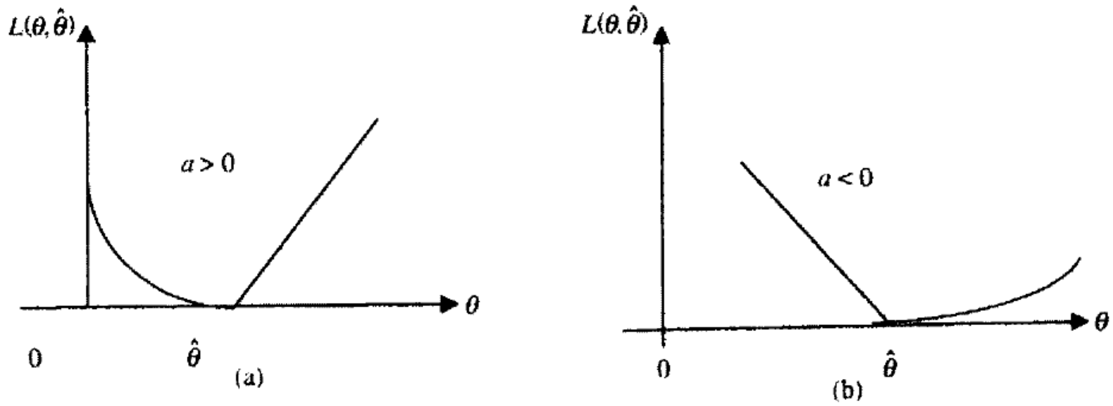

Figure 1-1 (a) LINEX loss function for a>0 (b) LINEX loss function for a<0

In decision theory, LINEX loss function (with a>0) is popular loss which consider in situation that overestimation is more considerable than underestimation. Meanwhile, in the reverse situation (underestimation is more considerable than overestimation) LINEX loss function given by a<0 is more applicable loss. (Press, 2003))

Gomez, et al. (2006) described the Bayes premiums in a BMS are obtained using weighted LINEX loss function under the Poisson-Gamma model. They believe that the LINEX loss function can be applied in BMS because it solves the problem of overcharges.

Balance loss function

Balance loss function is the form

$$L_{\rho,w,\delta_0}(\theta, \delta) = w\rho(\delta_0, \delta) + (1 - w)\rho(\theta, \delta).$$

That ρ is an arbitrary loss function, while δ_0 is a chosen a priori "target" estimator of ɵ, obtained for instance from the criterion of maximum likelihood estimator, least-square, or unbiasedness among others. We consider here Bayes estimation under balanced loss function L_{ρ,w,δ_0}as in previous formula. When w=0, we simply use L_0 instead of $L_{\rho,0,\delta_0}$unless we want to emphasize the role of ρ.

1.11 Credibility

Credibility theory is the art of combining different collection of data to option and accurate overall estimate. It provides actuaries with techniques to determine insurance premiums for contacts that belong to a (more or less) heterogeneous portfolio, where there is limited or irregular claim experience for each contract but ample claim experience for the portfolio.

Keffer (1929) initially suggested using a Bayesian perspective for experience rating in the context of group life insurance. Subsequently, Bailey (1945, 1950) showed how to derive the linear credibility form from a Bayesian perspective as the mean of a predictive distribution.

Bühlmann (1967) described a fundamental model containing latent (unobserved) effects that are common to all claims from a risk class. Bühlmann called these "structure effects."

Miller & Hickman (1975) examined credibility in the context of aggregate loss distributions. Pinquet (1997) was also interested in automobile claims; he considered collision claims arising from two lines, at fault and no fault coverage. Both of these papers assumed parametric distributions for the claims number and amount distributions and used Bayesian techniques to develop estimators.

Credibility theory can be seen as a set of quantitative tools that allows the insurers to perform experience rating, that is, to adjust future premiums based on past experience. In many cases, a compromise estimator is derived from a convex combination of a prior mean and the mean of the current observations. The weight given to the observed mean is called the credibility factor (since it fixes the extent to which the actuary may be confident in the data). Credibility theory has a long history in actuarial science Bühlmann's basic model formulation extends readily to encompass a large class of models for a review that is oriented towards linear regression and longitudinal data models (Frees, Young & Luo (1999)). To account for the entire distribution of claims, a common approach used in credibility is to adopt a Bayesian perspective.

Credibility theory is a branch of actuarial science. It was developed originally as method to calculate the risk premium by combining the individual risk experience with class risk