December 2020

Mixed Poisson Modeling of Claim Counts

Yidan Guo
Clemson University, yidang@clemson.edu

Mixed Poisson Modeling of Claim Counts

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Mathematical and Statistical Sciences

---

by
Yidan Guo
December 2020

Accepted by:
Dr. Xiaoqian Sun, Committee Chair
Dr. Yuyuan Ouyang
Dr. Mark Cawood

Abstract

In the credibility realm, an insurance company tries to find a way to establish a premium scheme so that the premium is sufficiently large to cover its obligations. On the other hand, an insured will only accept a premium when the policy is fair to him.

The ideal case is to establish an insured's premium based solely on its own experience, this is called the risk premium. We will introduce two estimation mechanisms to estimate risk premium, namely, Bayesian premium and Bühlmann-Straub credibility premium.

To implement the two models, we use a data set from Singapore Driving Experience in 1993. We estimate the average claim counts of each insured when weighted by their exposure of risk, i.e. given the record of claim counts and exposure weights, we build a model to estimate the claim counts of the insured by means of Bayesian premium and Bühlmann-Straub credibility premium.

Table of Contents

List of Figures

Chapter 1

Background

In the credibility realm, an insurer tries to find a way to establish a premium scheme so that the premium is sufficiently large to cover its obligations. On the other hand, an insured will only accept a premium when the policy is fair to him. The ideal case is to establish an insured's premium based solely on its own experience, this is called the risk premium. We will introduce two mechanisms to estimate risk premium, namely, Bayesian Premium and Bühlmann-Straub credibility premium.

Regarding Bayesian premium the Empirical Bayesian inference approach is applied. On the other hand, estimating Bühlmann-Straub Credibility premium refers to the process of estimating structure parameters, which depict insured's risk information. This is the critical part of the premium scheme.

We will introduce the Exact Bayesian Credibility and Bühlmann-Straub Model along with Empirical Bayes approach, and make comparisons.

Before we step further toward Empirical Bayesian approach, let's take a glance at some credibility ideas.

1.1 Definitions

Definition 1.1. *The following random variables are defined*

N*: claim counts of the insured.*

N_i*: claim counts of insured i for i=1,2,...,I.*

Θ_i*: risk parameter for insured i = 1, ..., I. It incorporates every characteristic of the insured.*

Θ_i is unknown and is assumed to be constant throughout the life of the insurance contract. Furthermore, Θ_i is assumed to be drawn at random from the same distribution. (This is a property called homogeneity).

Definition 1.2 (Risk Premium; see [5]). *The risk premium, $\mu(\theta)$, is the correct premium to charge an insured if the insured's risk level, θ, is known. The risk premium is thus the expected value of the insured's aggregate claim amount in one period, given his or her risk level.*

Mathematically, the risk premium is given by

$$\mu(\theta) = E[X|\Theta = \theta] = \int_0^\infty x f(x|\theta) dx.$$

This risk premium accurately gives the rate that an insured should be charged solely on their own experience. This is exactly what we want. But the risk parameter Θ that rates the insured's experience is unobservable, hence the risk premium can't be exactly calculated. We have to estimate it from data. Our ultimate goal throughout is to estimate the risk premium.

The collective premium is the other extreme, where we totally have no knowledge of the insured's risk level.

Definition 1.3 (Collective Premium; see [5]). *The collective premium m is the pure premium charged when nothing is known about the insured's risk level (for example, an insured at the first year). The collective premium is in essence the average value of all possible risk premiums.*

Mathematically, the collective premium is given by

$$m = E[X] = E[E[X|\Theta]] = E[\mu(\Theta)]$$

Another important premium is called Bayesian premium.

Definition 1.4 (Bayesian Premium; see [5]). *Suppose the data for T consecutive periods are $X_1, \ldots, X_T$,* *then the Bayesian premium $\mathcal{B}$ ($X_1, \ldots, X_T$) is given by*

$$\begin{aligned}
\mathcal{B}(X_1, \ldots, X_T) &= \operatorname*{argmin}_{g(\cdot)} E[\mu(\Theta) - g(X_1, \ldots, X_T)]^2 \\
&= E[\mu(\Theta)|X_1, \ldots, X_T], & \mu(\Theta) = E[X_{T+1}|\Theta] = E[X_{T+1}|\Theta, X_1, \ldots, X_T] \\
&= E[E[X_{T+1}|\Theta, X_1, \ldots, X_T]|X_1, \ldots, X_T] \\
&= E[X_{T+1}|X_1, \ldots, X_T], & \textit{by Tower property}
\end{aligned}$$

Here $\mu(\Theta)$ is the risk premium that we want to estimate, $g(X_1, \ldots, X_T)$ is an estimator obtained from the data. The Bayesian premium gives a measurement of how the estimate behaves by minimizing the distance between the two, which in this circumstance is defined as the mean square error(MSE). Under this measurement, Bayesian premium is the closest estimator to the risk premium, hence the best of all.

Definition 1.5 (Credibility Premium; see [5])**.** *A credibility premium $\mathcal{P}$ is a linear function of a special type of observations $X_1, \ldots, X_T$ of an insured: it is a convex combination of the individual experience weighted average ($\bar{X}$) and the collective premium (m), i.e.,*

$$\mathcal{P}(X_1, \ldots, X_T) = z\bar{X} + (1 - z)m,$$

where $0 \leq z \leq 1$ is the credibility factor and (1-z) is the complement of credibility.

To help understand the technique of Bayesian premium, let's first have a review of Bayesian inference.

1.2 Bayesian Inference Review

A random sample $X = (X_1, \ldots, X_n)$ is drawn from a population whose distribution is indexed by P_θ, where the parameter θ is considered to be an unknown but fixed quantity. Bayesian approach assumes θ to be a quantity that is unknown but could be described by a probability distribution, which is called prior distribution. Prior distribution is formulated before data is generated, and then a sample is taken from a population indexed by θ. With information from the sample the believe of the prior is updated by means of Bayes's rule. That is called a posterior distribution.

Hence, with this belief we actually collect a data set of:

$$X_i|\theta \overset{i.i.d}{\sim} P_\theta, i = 1, \dots, n. \text{ where } \theta \sim \pi(\theta)$$

Here P_θ is called the sampling distribution; $\pi(\theta)$ is the prior distribution; and $\theta|X$ follows a distribution called posterior distribution. Bayesian inference is interested in obtaining the posterior distribution. Therefore this could be approached by means of Bayes theorem.

Theorem 1.1 (Bayes's Rule; see [2]). *Let $A_1, A_2, \dots$ be a partition of the sample space, and let B be any set. Then, for each $i = 1, 2, \dots,$*

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{\infty} P(B|A_i)P(A_i)}$$

Further more,

Theorem 1.2 (Law of total probability).

$$P(B) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i)$$

Theorem 1.3. *By Bayes theorem, we have the distribution of $\theta|X$ follows*

$$\begin{aligned} P(\theta|X) &= \frac{P(X|\theta)P(\theta)}{P(X)} \\ &= \frac{P(X|\theta)P(\theta)}{\int P(X|\theta)P(\theta)d\theta}, \qquad \textit{by Law of total probability} \\ &\propto P(X|\theta)P(\theta), \end{aligned}$$

where $P(X|\theta) = \prod_{i=1}^{n} P(X_i|\theta) =: L(\theta|X)$ is called the likelihood function when it's viewed as a function of θ.

Calculating the posterior distribution could be complicated, sometimes impossible. Luckily, good properties arise under certain circumstances, that's the conjugate families.

Definition 1.6 (Conjugate families; see, e.g., [4]). *Consider a class of prior distributions, $p(\theta) \in \mathcal{P}$. We say that the class is conjugate for a sampling model $p(y|\theta)$, if $p(\theta) \in \mathcal{P}$ implies that $p(\theta|Y) \in \mathcal{P}$ for all $p(\theta) \in \mathcal{P}$ and data y. For example, when data follows Binomial distribution, both the prior* *and the posterior follow a Beta distribution. Then, Beta is conjugate for Binomial.*

Here is a table of some conjugate families.

Sampling distributions	Prior/Posterior
Binomial	Beta
Exponential	Gamma
Poisson	Gamma
Normal(known variance)	Normal
Normal(unknown mean/variance)	Normal-Gamma

We will apply the Poisson-Gamma pair in our model.

Remark 1.1. *Note that conjugate families are always available for data models within the exponential family.*

1.3 Exact Bayesian Credibility

Recall the Bayesian premium is found to be

$$\mathcal{B}(X_1, \ldots, X_T) = E[\mu(\Theta)|X_1, \ldots, X_T]$$

or equivalently,

$$\mathcal{B}(X_1, \ldots, X_T) = E[X_{T+1}|X_1, \ldots, X_T]$$

Using the first equation, the Bayesian premium could be calculated by two steps. Firstly, find the posterior distribution of Θ given data, $P(\Theta|X)$, and then calculate the expected value of $\mu(\Theta)$ with respect to this posterior distribution. So,

$$\begin{aligned}\mathcal{B}(X_1, \ldots, X_T) &= E[\mu(\Theta)|X_1, \ldots, X_T] \\ &= \int_\theta \mu(\Theta)P(\Theta|X)\ d\theta\end{aligned}$$

This brings a drawback of Bayesian premium to our attention: the distributions of $X_t|\Theta$ and Θ have to be both known. Another thing worth mentioning is that compared to Credibility premium, which lies between individual experience average $\bar{X}$ and the collective premium m, Bayesian premium doesn't guarantee a linear combination of the two. But under the nice form of distribution combinations from Definition (1.6), Bayesian premiums are exactly credibility premiums. That is,

Remark 1.2. *For Poisson sampling distribution with Gamma prior, Bayesian premium is exactly the posterior mean.*

1.4 Bühlmann-Straub Model

Bühlmann-Straub model is based on the Bayesian approach to credibility with removal of the distributional assumptions so that the calculations are done in the nonparametric setting. Assumptions for Bühlmann-Straub model:

1. The insured's vectors $(X_{i1}, \ldots, X_{iT_i}, \Theta_i), i = 1, \ldots, I$, are mutually independent,
2. The risk parameters $\Theta_i, i = 1, \ldots, I$, are independent and identically distributed,
3. The variablesX_{it} have finite variance; and
4. For $i = 1, \ldots, I$, and $t, u = 1, \ldots, T_i$,

$$\begin{aligned} E[X_{it}|\Theta_i] &= \mu(\Theta_i) \\ Cov(X_{it}, X_{iu}|\Theta_i) &= \delta_{tu}\frac{\sigma^2(\Theta_i)}{w_{it}}, \end{aligned} \tag{1.1}$$

where δ_{tu} is the Kronecker delta, which equals one if $t = u$ and zero otherwise. Note that equation (1.1) states that, given the risk parameter, successive claim records of an insured are uncorrelated. Complete independence is thus not required. It reflects the noncorrelation within the insured's claims experience across the years and the homogeneity in time.

To see how it works, let's introduce some notations first:

$\Theta_i,$ **risk level**

$\mu(\Theta_i) = E[X_{it}|\Theta_i],$ **risk premium**

$m = E[\mu(\Theta_i)],$ **collective premium**

$s^2 = E[\sigma^2(\Theta_i)],$

$a = \text{Var}[\mu(\Theta_i)],$

$$w_{i\cdot} = \sum_{t=1}^{T_i} w_{it},$$

$$w_{\cdot\cdot} = \sum_{i=1}^{I}\sum_{t=1}^{T_i} w_{it} = \sum_{i=1}^{I} w_{i\cdot},$$

$$X_{i\cdot}^{w} = \sum_{t=1}^{T_i} \frac{w_{it}}{w_{i\cdot}} X_{it},$$ **weighted average of claims of insured i**

$$X_{\cdot\cdot}^{w} = \sum_{i=1}^{I}\sum_{t=1}^{T_i} \frac{w_{it}}{w_{\cdot\cdot}} X_{it},$$

$z_i,$ **credibility factor**

$K,$ **credibility constant**

The parameters m, s^2, and a are called the structure parameters. They are functions of the unobserved random variable Θ. Structure parameters are independent of i because of Assumption (3). These structure parameters are generally unknown and must be estimated from the entire portfolio data. The parameter z_i is called the credibility factor.

According to [**Buhlmann2006course**], for $N_i, i = 1, \ldots, I$, as claim counts during a certain period. Our interests are:

$$
\begin{aligned}
\mu(\Theta_i) &:= E[N_i|\Theta_i], && \textbf{Individual risk premium} \\
\sigma^2(\Theta_i) &:= E_i \mathrm{Var}[N_i|\Theta_i], && \textbf{Variance within individual risk} \\
m/\mu_0 &:= E[\mu(\Theta_i)], && \textbf{Collective premium} \\
s^2/\sigma^2 &:= E[\sigma^2(\Theta_i)], && \textbf{Average variance within individual risk} \\
a/\tau^2 &:= \mathrm{Var}[\mu(\Theta_i)], && \textbf{Variance between individual risk premiums} \\
K &:= \frac{s^2}{a}, && \textbf{Credibility constant} \\
& z_i, && \textbf{Credibility factor}
\end{aligned} \tag{1.2}
$$

Remember that our ultimate goal is to estimate risk premium. By minimizing the MSE, Bühlmann-Straub model gives the credibility premium, $\mathcal{P}_i$, as the best estimator. That is,

$$
\begin{aligned}
\mathcal{P}_i(X_{i1}, \ldots, X_{iT_i}) &= \underset{g(\cdot)}{\operatorname{argmin}}\, E[\mu(\Theta_i) - g(X_{i1}, \ldots, X_{iT_i})]^2 \\
&\equiv \hat{X}_{i,T_i+1} \\
&= z_i X_{i\cdot}^w + (1 - z_i)m,
\end{aligned}
$$

where

$$
\begin{aligned}
z_i &= \frac{w_{i\cdot}}{w_{i\cdot} + K} = \frac{w_i}{w_i + K}, \\
T_i &= 1, \text{ and} \\
K &= \frac{s^2}{a}.
\end{aligned}
$$

Clearly, the credibility factor could be estimated from the data by estimating the structure parameters m, s^2, and a. Hence, the credibility premium estimated should be the closest to the insured's risk premium.

Bühlmann-Straub model gives estimations of structure parameters as follows:

$$\hat{m} = X_{\cdot\cdot}^{w} = \frac{\sum_{i=1}^{I} w_i X_i}{\sum_{i=1}^{I} w_i},$$
$$\hat{S}^2 = \frac{1}{N-I}\sum_{i=1}^{I}\sum_{t=1}^{T_i} w_{it}(X_{it} - X_{i\cdot}^{w})^2, \text{ where } N = \sum_{i=1}^{I} T_i,$$
$$\hat{a} = \frac{w_{\cdot\cdot}}{w_{\cdot\cdot}^2 - \sum_i w_{i\cdot}^2}\left(\sum_{i=1}^{I} w_i (X_{i\cdot}^{w} - X_{\cdot\cdot}^{w})^2 - (I-1)S^2\right).$$

The estimators have the following properties:

1. Unbiasedness. i.e. $E[\hat{S}^2] = \sigma^2, E[\hat{a}] = a, E[\hat{m}] = m.$
2. Consistency. i.e. $\hat{S}^2 \to \sigma^2,$ as $I \to \infty,$ and
   $\hat{a} \to a,\ as\ I \to \infty$, if none of the risks is "dominating"

Furthermore, for the case of mixed Poisson sampling distribution Poi($w_{ij}\theta_i$) where $\mu(\Theta_i) = w_{ij}\theta_i$, Bühlmann and Gisler [**Buhlmann2006course**] introduce an alternative approach by one more step of variable transformation:

Let $F_{ij} = \frac{N_{ij}}{w_{ij}}$, under the condition that "Given Θ_i the $N_{ij}(j = 1, 2, \dots)$ are independent and Poisson distributed with Poisson parameter $\mu_{ij}(\Theta_i) = \Theta_i w_{ij}$."

Then our interests are defined as:

$\mu(\Theta_i) := E[F_i|\Theta_i]$, **Individual risk premium** (1.3)

$\sigma^2(\Theta_i) := E_i \text{Var}[F_i|\Theta_i]$, **Variance within individual risk (normalized for weight 1)**

$m := E[\mu(\Theta_i)]$, **Collective premium**

$s^2 := E[\sigma^2(\Theta_i)]$, **Average variance within individual risk (normalized for weight 1)**

$a := \text{Var}[\mu(\Theta_i)]$, **Variance between individual risk premiums**

By definition (1.3), the conditional expectation and variance of F_i is given, thus the credibility factor is expressed as following:

$$
\begin{aligned}
E[F_i|\Theta_i] =& \mu(\Theta_i), \\
\mathrm{Var}[F_i|\Theta_i] =& \frac{\sigma^2(\Theta_i)}{w_{ij}}, \\
\Rightarrow \mathrm{Var}[F_i] =& E[\mathrm{Var}[F_i|\Theta_i]] + \mathrm{Var}[E[F_i|\Theta_i]] = \frac{s^2}{w_{i\bullet}} + a, \\
Cov(\mu(\Theta_i), F_i) =& E[Cov(\mu(\Theta_i), F_i|\Theta_i)] + Cov(\mu(\Theta_i), E[F_i|\Theta_i]) = 0 + Var[\mu(\Theta_i)] = a, \\
z_i =& \frac{Cov(\mu(\Theta_i), F_i)}{\mathrm{Var}[F_i]} = \frac{a}{\frac{s^2}{w_{i\bullet}} + a} = \frac{w_{i\bullet}}{w_{i\bullet} + \frac{s^2}{a}}.
\end{aligned}
$$

On the other hand, by our assumption the conditional expectation and variance is expressed as:

$$
\begin{aligned}
E[F_i|\Theta_i] =& \frac{E[N_i|\Theta_i]}{w_{ij}} = \frac{w_{ij}\Theta_i}{w_{ij}} = \Theta_i, \\
\mathrm{Var}[F_i|\Theta_i] =& \frac{\mathrm{Var}[N_i|\Theta_i]}{E_i^2} = \frac{w_{ij}\Theta_i}{E_i^2} = \frac{\Theta_i}{w_{ij}}.
\end{aligned}
$$

Hence, we conclude:

$$
\begin{aligned}
m/\mu_0 :=& E[\mu(\Theta_i)] = E(w_{ij}\Theta_i) = w_{ij}E(\Theta_i), \\
s^2/\sigma^2 :=& E[\sigma^2(\Theta_i)] = E(w_{ij}\Theta_i) = w_{ij}E(\Theta_i), \\
a/\tau^2 :=& \mathrm{Var}[\mu(\Theta_i)] = \mathrm{Var}(w_{ij}\Theta_i) = w_{ij}^2\mathrm{Var}(\Theta_i), \\
\bar{K} =& \frac{s^2}{a} = \frac{E[\Theta_i]}{\mathrm{Var}[\Theta_i]} = (w_{ij}\mathrm{Var}[\Theta_i])^{-1}, \\
z_i =& \frac{w_{i\bullet}}{w_{i\bullet} + \bar{K}} = \frac{w_{i\bullet}}{w_{i\bullet} + \frac{s^2}{a}}.
\end{aligned}
$$

Recall a consequence of the Poisson assumption is that the structure parameters m and σ^2 are equal. To have this equality also for the estimated structure parameters, we can use the following iterative algorithm.

We first define:

$$\begin{aligned} c :=& \frac{I-1}{I}\left\{\sum_i \frac{w_{i\bullet}}{w_{\bullet\bullet}}(1-\frac{w_{i\bullet}}{w_{\bullet\bullet}})\right\}^{-1}, \\ T :=& \frac{I}{I-1}\cdot\sum_i \frac{w_{i\bullet}}{w_{\bullet\bullet}}(F_i-\bar{F})^2, \\ \text{where } \bar{F} =& \sum_i \frac{w_{i\bullet}}{w_{\bullet\bullet}}F_i \end{aligned}$$

The algorithm of estimating m/s^2 and a is as below. Start with initial values

$$\begin{aligned} m^{(0)} =& \bar{F}, \\ a^{(0)} =& c\left\{T - \frac{I \cdot m^{(0)}}{w_{\bullet\bullet}}\right\}. \end{aligned}$$

The iteration from step n to step $n+1$ is

$$\begin{aligned} \bar{K}^{(n)} =& \frac{m^{(n)}}{a^{(n)}}, \\ z_i^{(n)} =& \frac{w_{i\bullet}}{w_{i\bullet}+\bar{K}^{(n)}}, \\ m^{(n+1)} =& \sum_i \frac{z_i^{(n)}}{z_\bullet^{(n)}}F_i, \\ a^{(n+1)} =& c\left\{T - \frac{I\cdot m^{(n+1)}}{w_{\bullet\bullet}}\right\}. \end{aligned}$$

The iteration stops when $|m^{(n+1)} - m^{(n)}|$ and $|a^{(n+1)} - a^{(n)}|$ are sufficiently small.

Denote the variables with "hat" to be the estimates from the algorithm. The following expressions are obtained:

$$\widehat{m} = \widehat{s^2},\ \widehat{a},\ \widehat{\bar{K}},\ \widehat{z}_i = \frac{w_{i\bullet}}{w_{i\bullet}+\widehat{\bar{K}}}.$$

The Credibility premium is thus:

$$\mathcal{P} = \widehat{z}_i F_i + (1-\widehat{z}_i)\widehat{m}.$$

Example 1.1 (See [1])**:**

The number of claims made by an individual insured in a year follows a Poisson distribution with mean θ. The prior distribution for θ is gamma with $\alpha = 1, \beta = 1.2$. Observe 3 claims in year 1 and no claims in year 2. What's the number of claims in Year 3?

From the question we have $N|\theta \sim \text{Poi}(\theta)$; $\theta \sim \Gamma(\alpha, \frac{1}{\beta}) = \Gamma(1, \frac{1}{1.2})$; $N_1 = 3$; $N_2 = 0$; and $T = 2$. Then we have by definition

$$\begin{aligned}
m &= E[\mu(\Theta)] = E[E[N|\theta]] = E[\theta] = \alpha\beta = 1.2 \\
s^2 &= E[\sigma^2(\Theta)] = E[\text{Var}[N|\theta]] = E[\theta] = \alpha\beta = 1.2 \\
a &= \text{Var}[\mu(\Theta)] = \text{Var}[E[N|\theta]] = \text{Var}[\theta] = \alpha\beta^2 = 1.44 \\
\bar{N} &= \frac{N_1 + N_2}{2} = \frac{3+0}{2} = 1.5 \\
\Rightarrow K &= \frac{s^2}{a} = \frac{1.2}{1.44} (= \frac{1}{\beta}) \\
z &= \frac{T}{T+K} = \frac{2}{2 + 1/1.2} \\
\mathcal{P} = X_3 = z\bar{N} + (1-z)m &= \frac{2}{2+1/1.2} \times 1.5 + (1 - \frac{2}{2+1/1.2}) \times 1.2 = 1.41
\end{aligned}$$

That is, the number of claims in Year 3 (or, the Credibility/Bayesian premium in Year 3) is 1.41.

Example 1.2 (See [1])**:**

The number of claims incurred in a month by any insured has a Poisson distribution with mean θ. Claim frequencies of different insureds are independent. The prior distribution is gamma with pdf

$$f(\theta) = \frac{(100\theta)^6 e^{-100\theta}}{120\theta},$$

i.e., $\Gamma(6, 1/100)$.

Observed data:

month	no.of insureds(w_t)	no. of claims
1	100	6
2	150	8
3	200	11
4	300	?

Determine the Bühlmann-Straub credibility estimate of the number of claims in Month 4.

Let w_t = number of insureds of Month t, and N_t = number of claims of Month t.

Note that from the statement we have $N_t|\theta \sim \text{Poi}(\theta)$, $t = 1, \ldots, T$, where $\theta \sim \Gamma(6, 1/100)$, and $T = 3$.

$$
\begin{aligned}
m &= E[\mu(\Theta)] = E[E[N|\theta]] = E[\theta] = \alpha\beta = .06 \\
s^2 &= E[\sigma^2(\Theta)] = E[\mathrm{Var}[N|\theta]] = E[\theta] = \alpha\beta = .06 \\
a &= \mathrm{Var}[\mu(\Theta)] = \mathrm{Var}[E[N|\theta]] = \mathrm{Var}[\theta] = \alpha\beta^2 = 6 * 10^{-4} \\
N_{i\cdot}^w &= \frac{\sum_{t=1}^3 N_t}{\sum_{t=1}^3 w_t} = \frac{6+8+11}{100+150+200} = \frac{25}{450} = \frac{1}{18} \\
\Rightarrow K &= \frac{s^2}{a} = \frac{1}{\beta} = 100 \\
z &= \frac{w_{i\cdot}}{w_{i\cdot} + K} = \frac{\sum_{t=1}^3 w_t}{\sum_{t=1}^3 w_t + K} = \frac{100+150+200}{(100+150+200)+100} = \frac{9}{11} \\
\mathcal{P} &= zN_{i\cdot}^w + (1-z)m = \frac{9}{11} \times \frac{1}{18} + (1 - \frac{9}{11}) \times .06 = .05636 \\
w_4\mathcal{P} &= 300 \times .05636 = 16.9
\end{aligned}
$$

Hence the number of claims in Month 4 is estimated to be 16.9.

Chapter 2

Numerical Experiment

2.1 Introduction

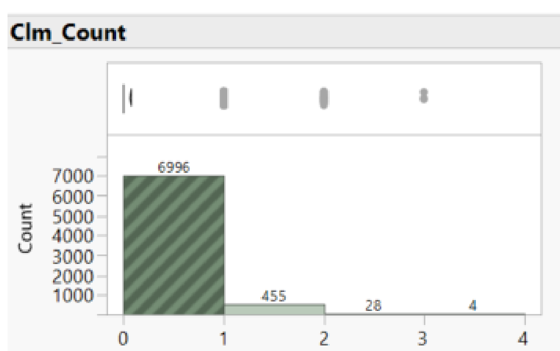
2.1.1 Data set

Our data set called "Singapore Automobile Claims" is a subset of Singapore Driving Experience, that focuses on the number of automobile accidents in 1993, weighted by risk of exposure. The data set gives characteristics including vehicle variables as well as person-level variables on 7483 individuals.

We estimate the average claim counts of each insured when weighted by their exposure of risk. That is, given the record of claim counts and exposure weights, we build up a model to estimate the average claim counts of insured by means of Bayesian premium and Bühlmann-Straub credibility premium.

A summary of the variables is as following:

```
   Clm_Count          Exp_weights
 Min.   :0.00000   Min.   :0.005476
 1st Qu.:0.00000   1st Qu.:0.279261
 Median :0.00000   Median :0.503764
 Mean   :0.06989   Mean   :0.519859
 3rd Qu.:0.00000   3rd Qu.:0.752909
 Max.   :3.00000   Max.   :1.000000
```



We will use two variables:

1. N_i: number of claims, where $N_i \in \{0, 1, 2, 3\}$, for $i = 1, \ldots, I$, and

2. E_i: individual exposure weight, $E_i \in [0.005476, 1.000000]$, for $i = 1, \ldots, I$,

with $I = 7483$ to implement the project. Namely, manipulation of the data set would help implement the model as below.

2.1.2 Model Assumptions

Assume

$$N_i|\theta_i \overset{indep}{\sim} Poi(E_i\theta_i),\ i = 1, \ldots, I$$

where N_i = number of claims for policyholder i during E_i; E_i is the exposure at risk of this policyholder; and θ_i is the risk level for this policyholder.

Also assume

$$\theta_1, \ldots, \theta_I \overset{i.i.d.}{\sim} g(\theta),$$

where $g(\theta)$ is unknown (called the structure distribution). In short, our model is

$$\begin{aligned} &N_i|\theta_i \overset{indep}{\sim} Poi(E_i\theta_i),\ i = 1, \ldots, I \\ &\theta_1, \ldots, \theta_I \overset{i.i.d.}{\sim} g(\theta) \end{aligned}$$

Let's find the Bayesian premium first.

2.2 Bayesian Premium

Recall our model assumption that our sampling distribution is Poisson, and the conjugate family for Poisson is Gamma distribution. Also, recall in Chapter 1 we concluded that the Bayesian premium is exactly the posterior mean, which means the prior distribution matters. Considering $\Gamma(\alpha, \beta)$, i.e. $\theta_i \sim \Gamma(\alpha, \beta), i = 1, \ldots, I$, the posterior function is thus

$$
\begin{aligned}
f(\theta_i|N_i) &= \frac{f(N_i|\theta_i)f(\theta_i)}{\int f(N_i|\theta_i)f(\theta_i)d\theta_i}, \\
&\propto \frac{e^{-E_i\theta_i}(E_i\theta_i)^{N_i}}{N_i!}\frac{1}{\Gamma(\alpha)\beta^\alpha}\theta_i^{\alpha-1}e^{-\frac{\theta_i}{\beta}}, \\
&\propto \theta_i^{N_i+\alpha-1}e^{-(E_i+\frac{1}{\beta})\theta_i}, \\
&= \theta_i^{N_i+\alpha-1}e^{-(\frac{1}{\beta/(E_i\beta+1)})\theta_i}.
\end{aligned}
$$

That is, $\theta_i|N_i \sim \Gamma(N_i+\alpha, \frac{\beta}{E_i\beta+1})$.

Hence, $E[\theta_i|N_i] = \frac{\beta(N_i+\alpha)}{E_i\beta+1}$ should be our Bayesian premium for the i^{th} policyholder.

Observe that

$$
\begin{aligned}
E[\theta_i|N_i] =& \frac{\beta(N_i+\alpha)}{E_i\beta+1}, \\
=& \frac{\beta}{E_i\beta+1}N_i + \frac{1}{E_i\beta+1}\alpha\beta, \\
=& z_i N_{i\bullet}^{E_i} + (1-z_i)m,
\end{aligned}
$$

where $z_i = \frac{\beta}{E_i\beta+1}$, $N_{i\bullet}^{E_i} = N_i$, and $m = \alpha\beta$.

This is exactly the credibility premium.

That is, given prior distribution is known, as is the conjugate family of the sampling distribution, we should have Bayesian premium equals the Credibility premium. This property helps in analyzing our results.

Our next step is to specify the prior parameters α, β.

Define $L_i(\alpha,\beta) = f(N_i|\alpha,\beta)$ and $L(\alpha,\beta) = \prod_{i=1}^{I} L_i(\alpha,\beta)$, let $\ell(\alpha,\beta) = \log L(\alpha,\beta)$.

Then, we have

$$
\begin{aligned}
L_i(\alpha,\beta) =& f(N_i|\alpha,\beta) \\
=& \int_{\theta_i} f(N_i,\theta_i|\alpha,\beta)d\theta_i \\
=& \int_{\theta_i} f(N_i|\theta_i)\pi(\theta_i|\alpha,\beta)d\theta_i \\
=& \int_{\theta_i} \frac{e^{-E_i\theta_i}(E_i\theta_i)^{N_i}}{N_i!}\frac{1}{\Gamma(\alpha)\beta^\alpha}\theta_i^{\alpha-1}e^{-\frac{\theta_i}{\beta}}d\theta_i \\
=& \frac{E_i^{N_i}}{N_i!}\frac{1}{\Gamma(\alpha)\beta^\alpha}\int_{\theta_i}\theta_i^{N_i+\alpha-1}e^{-(\frac{1}{\beta}+E_i)\theta_i}d\theta_i \\
=& \frac{E_i^{N_i}}{N_i!}\frac{\Gamma(N_i+\alpha)(\frac{\beta}{1+E_i\beta})^{N_i+\alpha}}{\Gamma(\alpha)\beta^\alpha} \\
=& \frac{E_i^{N_i}}{N_i!}\frac{\Gamma(N_i+\alpha)}{\Gamma(\alpha)}\frac{\beta^{N_i}}{(1+E_i\beta)^{N_i+\alpha}}
\end{aligned}
$$

$$
\log L_i(\alpha,\beta) = N_i\log E_i - \log N_i! + \log\Gamma(N_i+\alpha) - \log\Gamma(\alpha) + N_i\log\beta - (N_i+\alpha)\log(1+E_i\beta)
$$

$$
\begin{aligned}
\ell(\alpha,\beta) = \log L(\alpha,\beta) &= \sum_{i=1}^{I}\log L_i(\alpha,\beta) \\
&= \sum_{i=1}^{I}\Big(N_i\log E_i - \log N_i! + \log\Gamma(N_i+\alpha) - \log\Gamma(\alpha) + N_i\log\beta - (N_i+\alpha)\log(1+E_i\beta)\Big)
\end{aligned}
$$

Using the nonlinear minimization function "nlm" in R to solve $\ell(\cdot)$ function, we obtained a convergent value of $\hat{\alpha} = 1.500, \hat{\beta} = 0.090$.

Hence our Bayes premium is $E[\theta_i|N_i] = \frac{\beta(N_i+\alpha)}{E_i\beta+1} = \frac{0.09(N_i+1.5)}{0.09E_i+1}$. Running this function in R gives us a table of the Bayes premium.

A graph of Bayesian premium versus Exposure categorized by claim counts shows a negative, linear relationship between Bayesian premium and Exposure and a positive relationship between Bayesian premium and claim counts.

Also note that Bayesian premium has a range of 0.12 to 0.38.

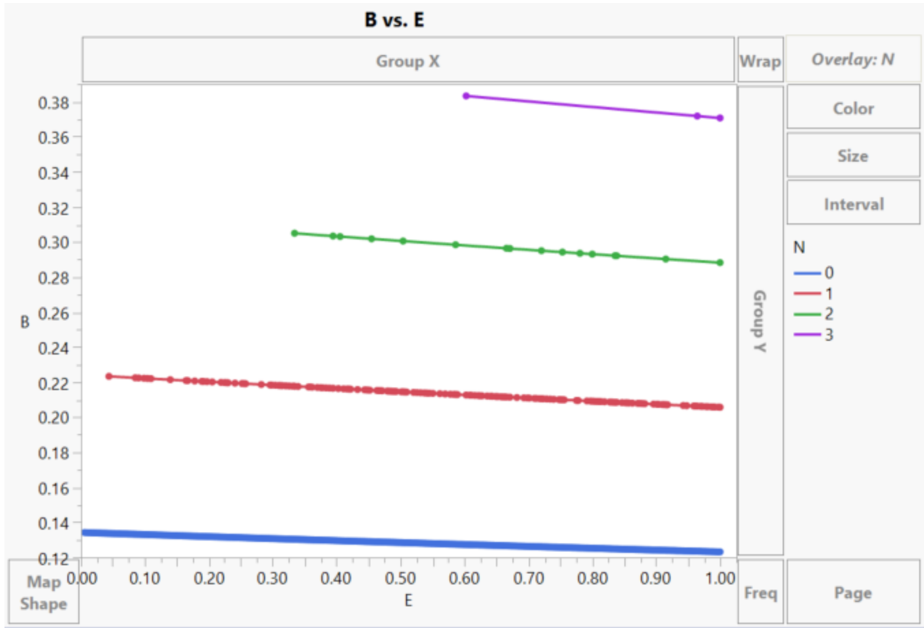
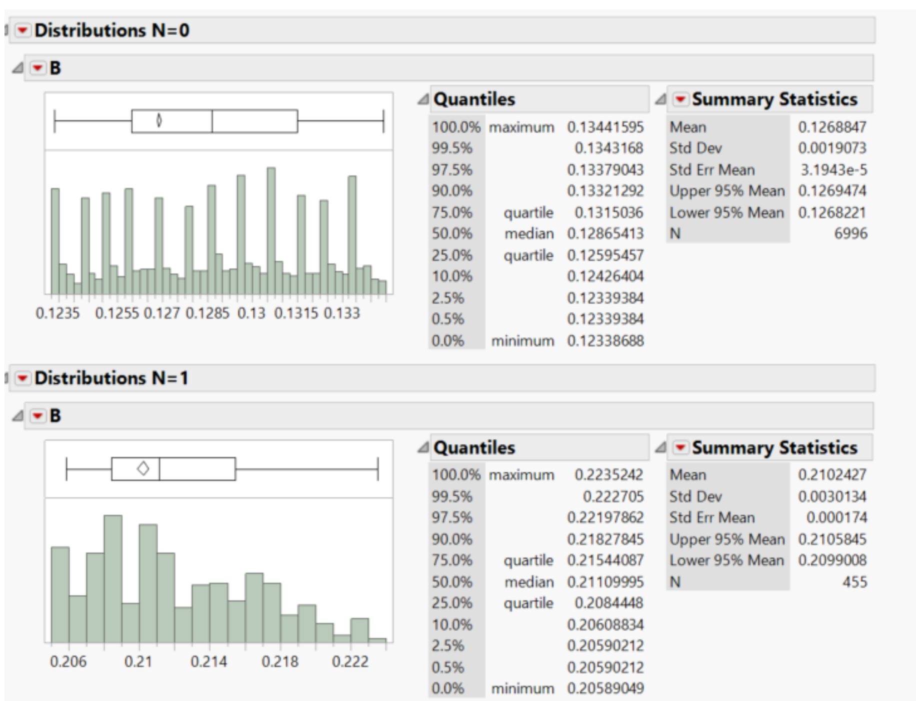


Figure 2.1: Bayesian premium versus Exposure given Claim Counts



Distributions N=0

B

Quantiles

100.0%	maximum	0.13441595
99.5%		0.1343168
97.5%		0.13379043
90.0%		0.13321292
75.0%	quartile	0.1315036
50.0%	median	0.12865413
25.0%	quartile	0.12595457
10.0%		0.12426404
2.5%		0.12339384
0.5%		0.12339384
0.0%	minimum	0.12338688

Summary Statistics

Mean	0.1268847
Std Dev	0.0019073
Std Err Mean	3.1943e-5
Upper 95% Mean	0.1269474
Lower 95% Mean	0.1268221
N	6996

Distributions N=1

B

Quantiles

100.0%	maximum	0.2235242
99.5%		0.222705
97.5%		0.22197862
90.0%		0.21827845
75.0%	quartile	0.21544087
50.0%	median	0.21109995
25.0%	quartile	0.2084448
10.0%		0.20608834
2.5%		0.20590212
0.5%		0.20590212
0.0%	minimum	0.20589049

Summary Statistics

Mean	0.2102427
Std Dev	0.0030134
Std Err Mean	0.000174
Upper 95% Mean	0.2105845
Lower 95% Mean	0.2099008
N	455

Figure 2.2: Summary when Claim Counts = 0,1

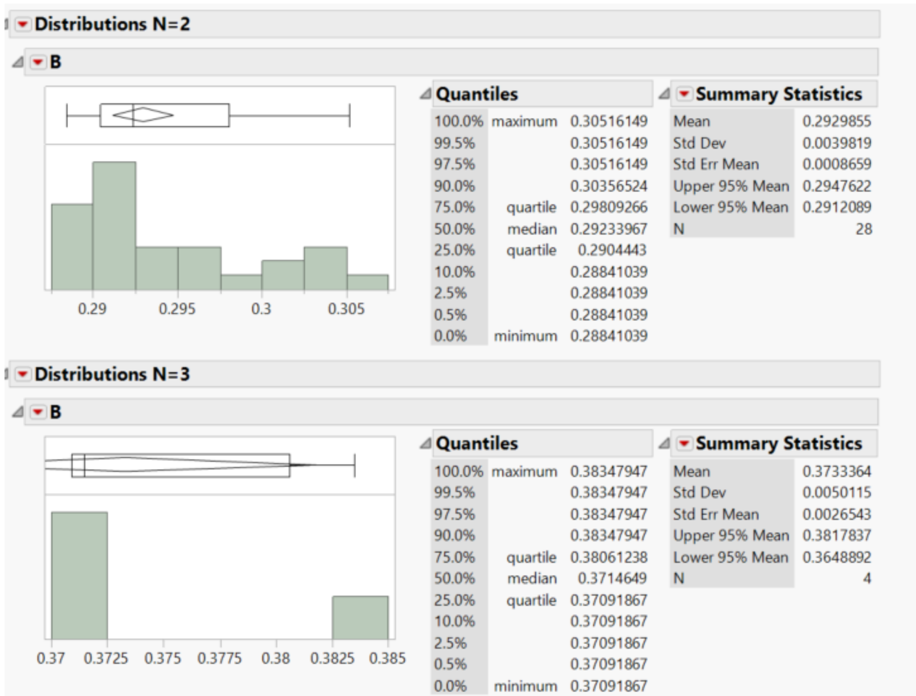


Figure 2.3: Summary when Claim Counts = 2,3

2.3 Credibility Premium

Recall our model assumptions that $N_i|\theta_i \overset{indep}{\sim} \text{Poi}(E_i\theta_i)$, and $\theta_1, \ldots, \theta_n \overset{i.i.d.}{\sim} g(\theta)$. Also note that the weight w in our case is the risk of exposure E. According to [**Buhlmann2006course**] we define $F_i = N_i/E_i$, the claim frequency weighted by exposure.

Recall our interests are:

$\mu(\theta_i) := E[F_i|\theta_i]$, **Individual risk premium**

$\sigma^2(\theta_i) := E_i \text{Var}[F_i|\theta_i]$, **Variance within individual risk (normalized for weight 1)**

$m / \mu_0 := E[\mu(\Theta_i)]$, **Collective premium**

$s^2 / \sigma^2 := E[\sigma^2(\Theta_i)]$, **Average variance within individual risk (normalized for weight 1)**

$a / \tau^2 := \text{Var}[\mu(\Theta_i)]$, **Variance between individual risk premiums**

By our assumptions the conditional expectation and variance are expressed as:

$$
\begin{aligned}
E[F_i|\theta_i] =& \frac{E[N_i|\theta_i]}{E_i} = \frac{E_i\theta_i}{E_i} = \theta_i,\\
\text{Var}[F_i|\theta_i] =& \frac{\text{Var}[N_i|\theta_i]}{E_i^2} = \frac{E_i\theta_i}{E_i^2} = \frac{\theta_i}{E_i}.
\end{aligned}
$$

Hence, our interests become functions of θ_i:

$$
\begin{aligned}
\mu(\theta_i) :=& E[F_i|\theta_i] = \theta_i,\\
\sigma^2(\theta_i) :=& E_i \text{Var}[F_i|\theta_i] = \theta_i,\\
m/\mu_0 :=& E[\mu(\Theta_i)] = E(\theta_i),\\
s^2/\sigma^2 :=& E[\sigma^2(\Theta_i)] = E[\theta_i],\\
a/\tau^2 :=& \text{Var}[\mu(\Theta_i)] = \text{Var}[\theta_i],\\
\bar{K} =& \frac{s^2}{a},\\
z_i =& \frac{E_i}{E_i + \bar{K}} = \frac{E_i}{E_i + \frac{s^2}{a}}.
\end{aligned}
$$

To estimate the structure parameters let's first define:

$$
\begin{aligned}
E. =& \sum_i E_i,\\
c :=& \frac{I-1}{I}\left\{\sum_i \frac{w_{i\cdot}}{w_{\cdot\cdot}}(1-\frac{w_{i\cdot}}{w_{\cdot\cdot}})\right\}^{-1} = \frac{I-1}{I}\left\{\sum_i \frac{E_i}{E.}(1-\frac{E_i}{E.})\right\}^{-1},\\
T :=& \frac{I}{I-1}\cdot\sum_i \frac{w_{i\cdot}}{w_{\cdot\cdot}}(F_i-\bar{F})^2 = \frac{I}{I-1}\cdot\sum_i \frac{E_i}{E.}(F_i-\bar{F})^2,\\
\text{where } \bar{F} =& \sum_i \frac{w_{i\cdot}}{w_{\cdot\cdot}}F_i = \sum_i \frac{E_i}{E.}F_i = \sum_i \frac{E_i}{\sum_i E_i}F_i
\end{aligned}
$$

The algorithm of estimating m/s^2 and a is:

Starting with initial values

$$
\begin{aligned}
m^{(0)} =& \bar{F} \text{ and}\\
a^{(0)} =& c\left\{T - \frac{I\cdot m^{(0)}}{E.}\right\},
\end{aligned}
$$

the iteration from step n to step $n+1$ is

$$\begin{aligned}\bar{K}^{(n)} =& \frac{m^{(n)}}{a^{(n)}},\\ z_i^{(n)} =& \frac{E_i}{E_i + \bar{K}^{(n)}},\\ m^{(n+1)} =& \sum_i \frac{z_i^{(n)}}{z_{\bullet}^{(n)}} F_i,\\ a^{(n+1)} =& c\left\{T - \frac{I \cdot m^{(n+1)}}{E_{\bullet}}\right\}.\end{aligned}$$

The iteration stops when $|m^{(n+1)} - m^{(n)}|$ and $|a^{(n+1)} - a^{(n)}|$ are sufficiently small. The tolerance value is set to be TOL = 1.e-6.

We obtain:

$$\begin{aligned}\widehat{m} =& \widehat{s^2} = 0.1344885,\\ \widehat{a} =& 0.01379349,\\ \widehat{\bar{K}} =& 9.686445,\\ \widehat{z_i} =& \frac{E_i}{E_i + \widehat{\bar{K}}},\end{aligned}$$

The Credibility premium is thus

$$\mathcal{P} = \widehat{z_i} F_i + (1 - \widehat{z_i})\widehat{m}.$$

Similarly, we obtained a group of values very close to Bayesian premiums. A graph of Credibility premium versus Exposure categorized by claim counts shows a negative, linear relationship between Bayesian premium and Exposure and a positive relationship between Bayesian premium and claim counts.

Also note that Credibility premium has a range of 0.12 to 0.42, which is larger than Bayesian premium.

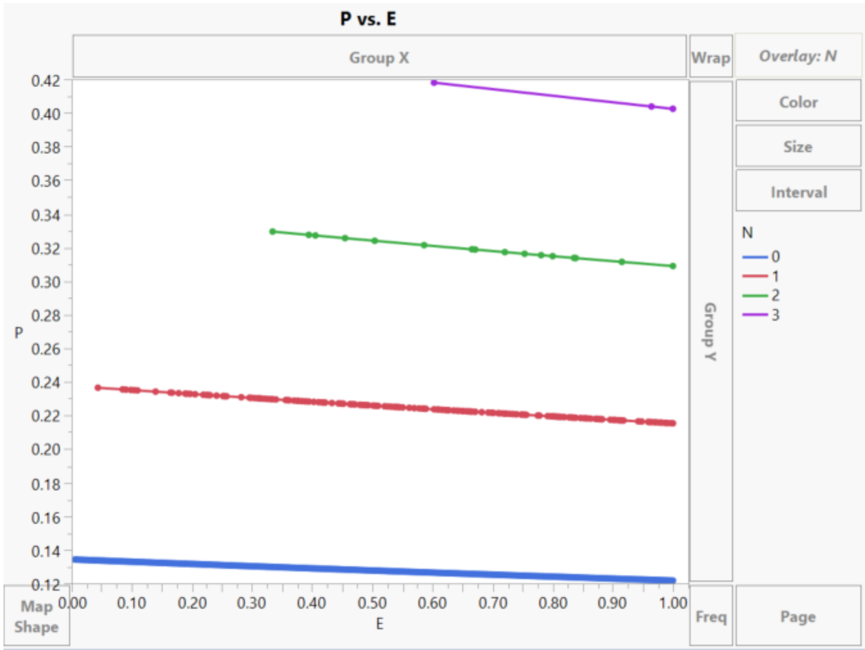


Figure 2.4: Credibility premium versus Exposure given claim counts

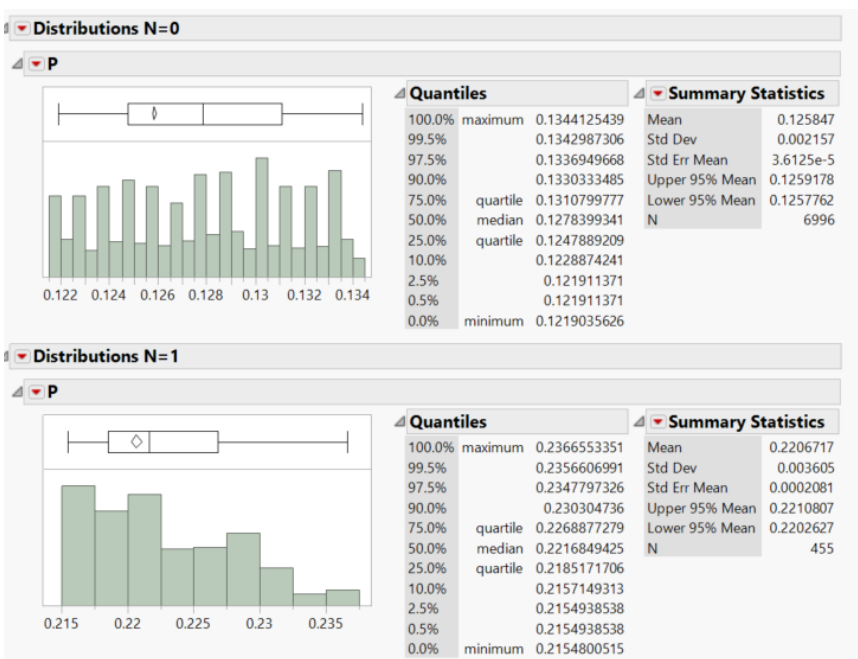


Figure 2.5: Summary when Claim Counts = 0,1

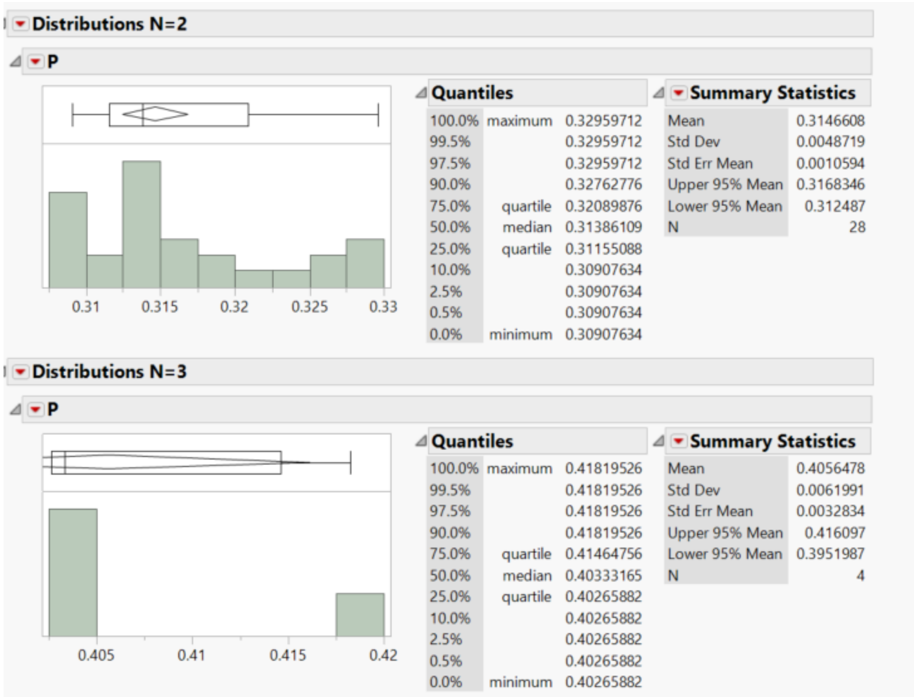


Figure 2.6: Summary when Claim Counts = 2,3

While Bayesian premium assumes the prior distribution of θ_i, Bühlmann-Straub Credibility premium follows the nonparametric method with no prior distribution assumptions. This brings flexibility to Bühlmann-Straub model and the results would be more accurate. Recall the property that given prior distribution is known and conjugate, the two estimates should be exactly the same. The similarity of our outputs could be explained by this property.

2.4 Application

So far, we introduced two methods of estimating an insured's average claim counts, which is also viewed as risk factor, a related definition called relative risk factor is defined as the estimated risk factor divided by the estimated group average risk factor. Relative risk factor helps insurance companies to determine of how much premium(money) will be charged next year for a particular customer. For example, if the current average premium of all insured is \$500/yr, and the relative risk factor for a particular customer is 1.542 with one claim during one year, then the insurance company will charge \$500 $\times$ 1.542=\$771 for next year when renewing.

Graphs of relative risk factors are as below,

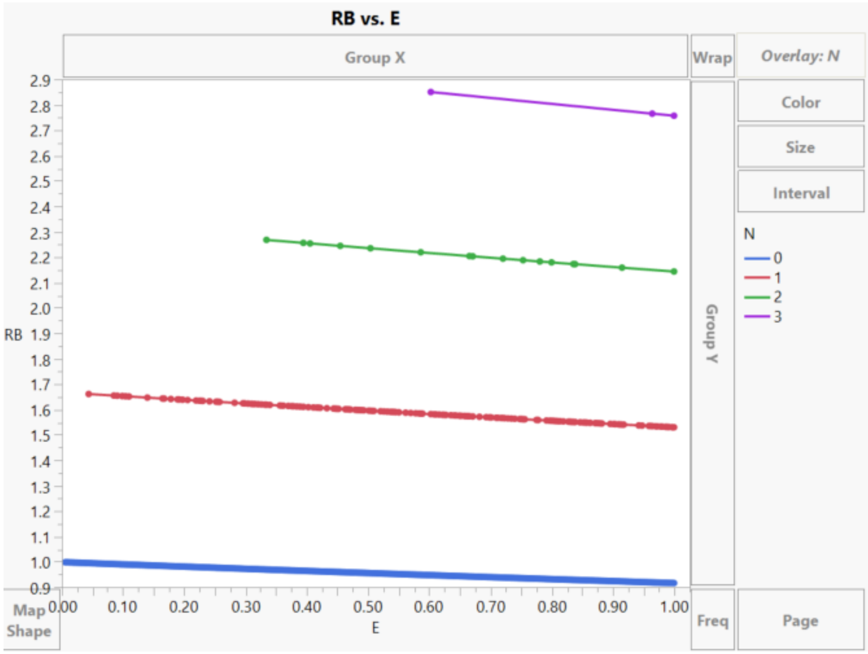


Figure 2.7: Relative Bayesian premium versus Exposure given claim counts

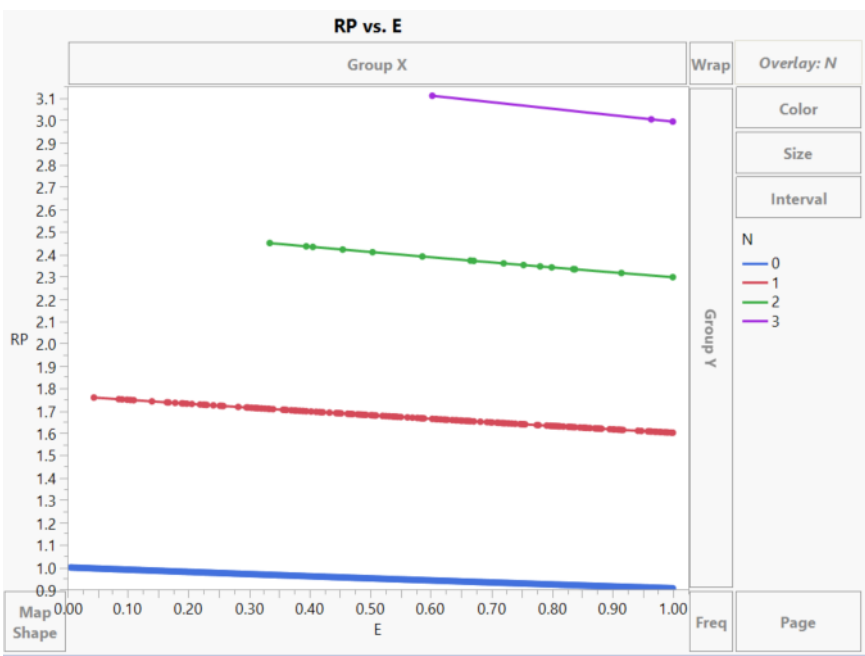


Figure 2.8: Relative credibility premium versus Exposure given claim counts

Chapter 3

Discussion

The Bühlmann-Straub model estimates the parameter Θ_i in the nonparametric scheme by the first-moment method. The Bayesian inference brings up conjecture on prior distribution and works on combinations of prior-sampling distributions. On the contrary, Efron [3] estimates the prior function $g(\theta)$ using deconvolution. He estimates $g(\theta)$ directly as a polynomial function. The analysis on this method should be more accurate.

This topic could be extended by deconvolution approach.

Appendix

The R codes are as below

```
#Step 0, read data, select and rename variables
N <- dat$Clm_Count
E <- dat$Exp_weights
I <- length(N)

#step 2, Bayesian approach with prior dist. Gamma(alpha, beta), need the value of alpha,beta
# use Likelihood function to find the MLE of alpha, beta
cl=dat$Clm_Count
ep=dat$Exp_weights

Lab=function(a,b,cl,ep){
  y=rep(0,length(cl))
  for(i in 1:length(cl)){
    n=cl[i]
    e=ep[i]
    y[i]= log(gamma(n+a))+n*log(b) -log(gamma(a))-(a+n)*log(e*b+1)
  }
  return(sum(y))
}
FF=function(a){
  g= -Lab(a[1],a[2],c=cl,e=ep)
  return(g)
}
```

```
nlm(FF,c(10,.5))#$estimate [1] 1.49553297 0.08992255
nlm(FF,c(1,.5)) #$estimate [1] 1.49553297 0.08992255

#step 2.0, try a different coding to help double check, find the same values of alpha and beta.

n=dat$Clm_Count
e=dat$Exp_weights

Lab=function(a,b){
  y = log(gamma(n+a))+n*log(b) -log(gamma(a))-(a+n)*log(e*b+1)
  return(sum(y))
}
FF=function(a){
  g= -Lab(a[1],a[2])
  return(g)
}
nlm(FF,c(10,.5))#$estimate [1] 1.49553297 0.08992255
nlm(FF,c(1,.5)) #$estimate [1] 1.49553297 0.08992255

#step 2.1, Bayesian premium with prior dist. Gamma(alpha, beta), which is exactly the posterior mean
#Bayes Premium with alpha=1.5, beta=.09 specified.
alpha = 1.49553297;
beta = 0.08992255;
B = vector()
for (i in 1:I){
  Bi = (beta*(N[i]+alpha))/(E[i]*beta +1)
  B = append(B,Bi)
}
print(B)
output <- cbind(N,E,B)
write.csv (output, file = "Bayes.csv")

#step 2.2, B-S Credibility premium with nonparametric setting
```

```
FF = N/E
Fbar = sum(E*FF)/sum(E)
c = ((I-1)/I)*(sum((E/sum(E))*(1-(E/sum(E)))))^(-1)
TT = (I/(I-1))* sum((E/sum(E))*(FF - Fbar)^2)
m = Fbar
ss = m
a = c*(TT - I*ss/sum(E))
k = ss/a
z = E/(E+k)
mnew = sum(z*FF)/sum(z)
anew = c*(TT - I*mnew/sum(E))
while( abs(mnew-m) > 1.e-6 && abs(anew-a) > 1.e-6){
  m = mnew #0.1344885
  a = anew #0.1344885
  par = cbind(mnew, anew,m, a,k)
  print(z)
  print(par)
}
P = z*FF +(1-z)*m
write.csv (outputP, file = "Credibility4.csv")

#step 2.3, Credibility premium with prior dist. Gamma(alpha, beta)
#Credibility Premium with alpha=1.5, beta=.09 specified.
P = vector()
for (i in 1:I){
  Pi =(beta /(E[i]*beta +1))*(N[i]+alpha)
  P = append(P,Pi)
}
print(P)
EP =(E*beta /(E*beta +1))*(N+alpha)
output<-cbind(N,E,P,EP)
write.csv (output, file = "Credibility2.csv")
```

Bibliography

[1] ACTUARYUTEXAS. *Group 8-Buhlmann Straubb*. Youtube. 2012. URL: `https://www.youtube.com/watch?v=gOM4RBkVV9M`.

[2] George Casella and Roger L Berger. *Statistical inference*. Vol. 2. Duxbury Pacific Grove, CA, 2002.

[3] Bradley Efron. "Empirical Bayes deconvolution estimates". In: *Biometrika* 103.1 (2016), pp. 1–20.

[4] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. CRC press, 2013.

[5] Vincent Goulet. "Principles and application of credibility theory". In: (1998).