Bayesian Credibility for GLMs

Jose Garrido

Department of Mathematics and Statistics
Concordia University, Montreal
and U. Carlos III of Madrid, Spain

Advances in Predictive Analytics
University of Waterloo, December 1–2, 2017

(joint work with Oscar Quijano, Concordia University)

Research funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) and Concordia University, Montreal

Abstract

We revisit the classical credibility results of Jewell (1974, AB) and Bühlmann (1967, AB) to obtain credibility premiums for a GLM severity model using a modern Bayesian approach.

Here prior distributions are chosen from out–of–sample information, without restrictions to be conjugate to the severity distribution.

Then we use the relative entropy between the "true" and the estimated models as a loss function, without restricting credibility premiums to be linear.

A numerical illustration on real data shows the feasibility of the approach, now that computing power is cheap, and simulations software readily available.

Dedication

Dedicated to the memory of Bent Jørgensen (1954–2005);
a true scholar, a gentlemen and a great person.

Overview:

1. 1. Preliminaries
   - 1.1 Exponential dispersion family
   - 1.2 Relative entropy
   - 1.3 Generalized linear models
2. 2. Entropic estimation
3. 3. Bayesian credibility for GLMs
   - 3.1 Linear credibility for GLMs
     - 3.1.1 Type 1 linear credibility is impossible
     - 3.1.2 Type 2 linear credibility is sometimes possible
   - 3.2 Entropic credibility for GLMs
     - 3.2.1 Estimation of the mean
     - 3.2.2 Estimation of the dispersion parameter
   - 3.3 Numerical illustrations

Introduction

In the 60's credibility theory and Bayesian statistics developed in parallel.

Bühlmann (1967, AB) proposed a "Bayesian derivation" to linear credibility premiums:

$$z\,\bar{Y} + (1 - z)\,x_0,$$

where

- $\bar{Y}$ is a risk class sample average of claims,
- x_0 is a manual premium based on out–of–sample data,
- z is a credibility weight that gives more relative importance to in–sample or out–of–sample data.

Hans Bühlmann at ETH

Figure: https://people.math.ethz.ch/~hbuhl/

Bayesian premiums

The "Bayesian" justification was a follows:

- θ is a latent unobservable variable (risk parameter) that affects the claim frequency or severity distributions,
- given θ, the $y_1, \ldots, y_n$ are conditionally iid claim frequencies or severities with known conditional distribution $F_{Y|\theta}$,
- the class mean $\mu(\theta) = \mathbb{E}(Y_r|\theta) = \mathbb{E}(Y_s|\theta)$ and class variance $\sigma^2(\theta) = \mathbb{V}(Y_r|\theta) = \mathbb{V}(Y_s|\theta)$, for $r, s = 1, \ldots, n$,
- $\mathbb{E}(Y_r^2)$ is finite,
- the structural function ("prior distribution") $\pi_{n_0,x_0}(\theta)$ is known (presumably including the parameters n_0 and x_0).

Exact premiums

The Bayesian solution that minimizes the quadratic loss:

$$\mathbb{E}\left\{\left[\mu(\theta) - g(Y_1, \ldots, Y_n)\right]^2\right\},$$

where the expectation is taken over the joint distribution of $Y_1, \ldots, Y_n$ and θ, is

$$g(Y_1, \ldots, Y_n) = \mathbb{E}\left[\mu(\theta) | Y_1, \ldots, Y_n\right],$$

it means the posterior mean (or Bayes rule).

For many combinations of the distribution $F_{Y|\theta}$ and prior $\pi_{n_0,x_0}(\theta)$, the posterior mean involves the calculation of high–dimensional integrals. These non–linear premiums were difficult if not impossible to calculate back in 1967.

Linear approximations

So Bühlmann (1967, AB) proposed a linear approximation to the exact Bayesian premium by minimizing the expected square loss error only among linear estimators of the form:

$$g(Y_1, \ldots, Y_n) = c_0 + c_1 Y_1 + \cdots + c_n Y_n.$$

The solution is $z \bar{Y} + (1 - z) x_0$, where $c_0 = (1 - z) x_0$, $c_i = (n + \phi n_0)^{-1}$, for all $i = 1, \ldots, n$, then $x_0 = \mathbb{E}(Y)$ and $z = n/(n + \phi n_0)$, for $\phi n_0 = \mathbb{E}[\sigma^2(\theta)]/\mathbb{V}[\mu(\theta)]$.

Although these premiums are not exact, they are linear, which means that they were easy to compute and to interpret (for known n_0, x_0).

Exact credibility

To address the issue of accuracy, Jewell (1974, AB) showed that linear credibility premiums are exact for the exponential family; that is, if the claim frequency or severity probability or density of Y is in the exponential dispersion family (EDF):

$$f(y|\theta,\phi) = a(y,\phi)\exp\left(\frac{1}{\phi}\left\{y\theta-\kappa(\theta)\right\}\right),$$

and assuming that ϕ is known, plus the conjugate prior on θ is:

$$\pi_{n_0,x_0}(\theta) \propto \exp\left(n_0[x_0\theta-\kappa(\theta)]\right),$$

for some (presumably known) parameters $n_0 > 0$ and x_0,

Exact credibility (...continued)

then Jewell showed that the posterior mean that minimizes the general quadratic loss problem is given by

$$\begin{aligned}\mathbb{E}\left[\mu(\theta)|y_1,\ldots,y_n\right] &= \frac{\phi n_0}{\phi n_0 + n} x_0 + \frac{n}{\phi n_0 + n} \bar{y} \\ &= (1-z)x_0 + z\bar{y},\end{aligned}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $z = \frac{n}{\phi n_0 + n}$, same as Bühlmann's.

So exact credibility occurs for claim distributions in the EDF.

Research question: since the EDF is the basis for the error distribution in GLMs, is it possible to extend exact/linear credibility to GLM premiums?

William Jewell (1932-2003)

Figure: http://www.berkeley.edu/news/media/releases/2003/01/31_.html

1.1 Exponential dispersion family (EDF)

The unit deviance function plays an important role in the theory of GLMs; the model assessment is done through tests based on its asymptotic behaviour.

It also allows for a mean–value re–parameterization of the EDF, somewhat like for the normal distribution:

$$f(y|\mu,\phi) = c(y,\phi)\,\exp\Big\{-\frac{1}{2\phi}d(y,\mu)\Big\},$$

where the deviance of a regular EDF is given by:

$$d(y,\mu) = 2\Big[y\,\{\dot{\kappa}^{-1}(y) - \dot{\kappa}^{-1}(\mu)\} - \kappa\big(\dot{\kappa}^{-1}(y)\big) + \kappa\big(\dot{\kappa}^{-1}(\mu)\big)\Big],$$

and κ is the cumulant function of Y (for more details and results on d, see Jørgensen, 1997, C&H).

Risk aggregation

There is a useful data aggregation property of reproductive EDFs (used for continuous distributions, GLMs and discrete distributions with $\phi = 1$).

The notation of Jørgensen (1997, C&H) is convenient to express this property: if Y has mean μ and EDF density as above, it is denoted as $Y \sim ED(\mu, \phi/w)$.

The property is then as follows: if $Y_1, \ldots, Y_n$ are independent, and $Y_i \sim ED(\mu, \phi/w_i)$, then

$$\bar{Y} = \frac{w_1 Y_1 + \cdots + w_n Y_n}{w_+} \sim ED(\mu, \phi/w_+), \qquad w_+ = \sum_{i=1}^{n} w_i .$$

1.2 Relative entropy

Instead of minimizing a quadratic loss, here we select optimal credibility premiums that minimize a relative entropy criterion.

Let $m_i(x)$ be probability measures with $dm_i(x) = f_i(x)ds(x)$ for $i = 1, 2$, some density functions f_1, f_2, and some probability measure s, such that $m_1 \equiv m_2 \equiv s$ [1].

Definition

The relative entropy of m_2 from m_1 is defined as

$$D(m_1 \parallel m_2) = \mathbb{E}_{m_1}\left[\log\left(\frac{f_1(X)}{f_2(X)}\right)\right] = \int \log\left(\frac{f_1(x)}{f_2(x)}\right) dm_1(x).$$

[1] absolutely continuous with respect to each other.

This definition was introduced by Kullback and Leibler (1951, AMS) and $D(\cdot \parallel \cdot)$ is often called the Kullback–Leibler divergence.

We prefer the term relative entropy, because what Kullback and Leibler called the divergence between m_1 and m_2 was actually:

$$D(m_1 \parallel m_2) + D(m_2 \parallel m_1).$$

1.3 Generalized linear models (GLMs)

The family of GLMs is an extension of the linear regression model (McCullagh & Nelder, 1989, C&H) that transforms the mean response by a chosen link function.

The log–link is the most popular for insurance data, where the linear predictor gets exponentiated to ensure that premiums are always positive and to preserve the multiplicative structure of the variable relativities.

For claims modeling, the log–link is often used to model both, frequency and severity.

GLMs: parameters

In GLMs, responses $y_1, \ldots, y_n$ are assumed to be independent and linearly related to the predictor variables through a non-linear link function as follows:

$$g\big(\mathbb{E}[y_i \mid \mathbf{x}_i]\big) := g(\mu_i) = \sum_{j=0}^{p} \beta_j x_{ij} \; = \; \eta_i,$$

where $\eta_i := \sum_{j=0}^{p} \beta_j x_{ij}$ is the linear predictor $(x_{0j} \equiv 1)$.

The true mean is given by taking the inverse transformation

$$\mu_i = g^{-1}(\eta_i).$$

The exponential dispersion family (EDF)

The distribution of regression errors in the EDF is assumed to be written as

$$f(y|\theta,\phi) = a(y,\phi)\exp\left(\frac{w}{\phi}\left\{y\theta - \kappa(\theta)\right\}\right),$$

where ϕ has been replaced by ϕ/w with the introduction of a weight $w \geq 0$ so that the mean and variance can be expressed as $\mu = \kappa'(\theta)$ and $\sigma^2 = \phi\,\kappa''(\theta)/w$, respectively.

In applications the weight w is known usually, while the dispersion parameter ϕ needs to be estimated.

Similarly the explanatory variables, $\mathbf{x} = (1, x_1, \cdots, x_p)$, and the link function g are usually known, while the vector of coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)$ needs to be estimated.

Risk classes

The portfolio can be divided into different risk classes according to the values of the explanatory variables; we can group all the observations that share the same values of the explanatory variables $\mathbf{x}$ and aggregate as explained above.

Remark: Note that this grouping does not lead to a loss of information to estimate the mean since $\bar{Y}$ is a sufficient statistic for θ in the EDF (but not for ϕ).

Sample information

After aggregation, let m be the number of risk classes and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ a vector whose entries are the different values of θ over all classes.

The sample (joint) density can be expressed as:

$$f(\mathbf{y}|\boldsymbol{\theta}, \phi) = A(\mathbf{y}, \phi) \exp\left(\frac{\mathbf{y}^T W \boldsymbol{\theta} - \mathbf{1}^T W \boldsymbol{\kappa}(\boldsymbol{\theta})}{\phi}\right), \qquad \mathbf{y} \in \mathbb{R}^m,$$

where $\boldsymbol{\kappa}(\boldsymbol{\theta}) = \big(\kappa(\theta_1), \ldots, \kappa(\theta_m)\big)$, $W = \mathrm{diag}(w_1, \ldots, w_m)$, with w_i being the sum of all the weights in the i-th class, $\mathbf{1} = (1, \ldots, 1)$ and $A(\mathbf{y}, \phi) = \prod_{i=1}^m a(y_i, \frac{w_i}{\phi})$.

Minimization problem

Without getting into details it is useful to re–parameterize the sample density in terms of the mean vector $\boldsymbol{\mu}$ instead of $\boldsymbol{\theta}$:

$$f(\mathbf{y}|\boldsymbol{\mu},\phi) = C(\mathbf{y},\phi)\exp\left(-\frac{1}{2\phi}D(\mathbf{y},\boldsymbol{\mu})\right),$$

where $C(\mathbf{y},\phi) = \prod_{i=1}^{m} c(y_i, \frac{\phi}{w_i})$ and

$$D(\mathbf{y},\boldsymbol{\mu}) = \sum_{i=1}^{m} w_i\, d(y_i,\mu_i).$$

D is called the deviance of the model. It is the objective function to be minimized.

Deviance

Here are some of the properties of the deviance D:

- Given a sample, finding the mle of $\boldsymbol{\theta}$ is equivalent to finding the value of $\boldsymbol{\beta}$ that minimizes the deviance.
- D can be used to estimate the dispersion parameter (although it is not the only method). The deviance estimator of ϕ is given by

$$\hat{\phi} = \frac{D(\mathbf{y}, \hat{\boldsymbol{\mu}})}{m - p}.$$

- The asymptotic distribution of D plays an important role in model assessment and variable selection.

For further details about the use and properties of the deviance we recommend Jørgensen (1992, IMPA).

2. Entropic estimation

A posterior distribution is more informative than a point estimator, reflecting the uncertainty about the true parameter.

Now, to charge a premium the insurer needs to define a point estimator from this posterior distribution, that will become the credibility premium.

General loss/risk function

In Bayesian point estimation one first chooses a loss function $\mathcal{L}(\theta_0, \theta_1)$ representing the cost of estimating θ with θ_1 instead of the true value θ_0.

Since θ_0 is not known, we define a risk function as

$$\mathcal{R}(\theta) = \mathbb{E}\left[\mathcal{L}(\theta, \theta_1)\right],$$

where the expectation is taken with respect to the posterior distribution of θ_1, as if θ were the true parameter.

Then the point estimator $\hat{\theta}$ of θ_0 is the θ that minimizes $\mathcal{R}$:

$$\hat{\theta} = \operatorname*{argmin}_{\theta} \mathcal{R}(\theta).$$

Entropic loss/risk functions

The entropic estimator is the Bayesian point estimator when the loss function $\mathcal{L}$ is the relative entropy D, between the distribution with the true parameter θ_0 and the estimated one.

More precisely, the entropic loss function is defined as:

$$\mathcal{L}(\theta, \theta_0) = \mathbb{E}_{\theta_0}\left[\log\left(\frac{f_{\theta_0}(Y)}{f_\theta(Y)}\right)\right],$$

for $Y \sim f_\theta$. The corresponding entropic risk function is then

$$\mathcal{R}(\theta) = \mathbb{E}\left[\log\left(\frac{f_{\theta_0}(Y)}{f_\theta(Y)}\right)\right] := \mathbb{E}_\pi\left[\mathbb{E}_{\theta_0}\left[\log\left(\frac{f_{\theta_0}(Y)}{f_\theta(Y)}\right)\right]\right],$$

where $\mathbb{E}_{\theta_0}$ is the conditional expectation given a fixed θ_0 and $\mathbb{E}_\pi$ is taken with respect to the posterior distribution of θ.

General entropic estimators

Definition

The entropic estimator is defined as

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, \mathbb{E}\left[\log\left(\frac{f_{\theta_0}(Y)}{f_{\theta}(Y)}\right)\right],$$

where the expectation is taken with respect to the posterior distribution of θ.

Entropic estimators for the EDF

Proposition

Let $\pi(\mu, \phi)$ be a prior distribution for (μ, ϕ), the vector $\mathbf{y} = (y_1, \ldots, y_n)$ be a conditionally iid sample, given (μ, ϕ), and $\pi(\mu, \phi|\mathbf{y})$ the corresponding posterior. The entropic estimator $\hat{\mu}$ of μ is then given by

$$\hat{\mu} = \mathbb{E}[Y \mid \mathbf{y}] = \mathbb{E}_\pi \left[\mathbb{E}_{\mu,\phi}[Y]\right],$$

where $\mathbb{E}_\pi$ and $\mathbb{E}_{\mu,\phi}$ represent the expectations with respect to the posterior distribution and with respect to fixed values of (μ, ϕ), respectively.

In summary

This proposition shows that for univariate EDFs the posterior mean not only minimizes the expected square error risk (Jewell, 1974), but also minimizes the posterior entropic risk.

The following sections show that this property does not generalize to GLMs, due to the difference of dimension between the response vector and the regression coefficients.

A direct consequence of the proposition is that Jewell's estimator for the EDF is an entropic estimator.

3. Bayesian credibility for GLMs
3.1 Linear credibility for GLMs

Remember our research question: Is it possible to extend Jewell's exact/linear credibility result to GLMs?

More precisely: is there a prior on the regression coefficients β that leads to a posterior mean given by a weighted average of an out–of–sample estimate with the sample mean of a GLM?

Two types of linear credibility for GLMs

There are 2 ways to interpret this question:

1. *Type 1:* All m dimensions in $\hat{\boldsymbol{\mu}}$ are given the same credibility weight then the credibility premium is:

$$\hat{\boldsymbol{\mu}}_c = z\,\hat{\boldsymbol{\mu}} + (1 - z)\,\mathbf{M},$$

   where $\hat{\boldsymbol{\mu}}$ contains the GLM mean estimators, $\mathbf{M}$ is a vector of out–of–sample "manual" premiums and $z \in (0, 1)$ the constant credibility factor.

2. *Type 2:* Coordinates have possibly different credibility factors:

$$\hat{\boldsymbol{\mu}}_c = \mathbf{Z}\,\hat{\boldsymbol{\mu}} + (\mathbb{I} - \mathbf{Z})\,\mathbf{M},$$

   where $\hat{\boldsymbol{\mu}}$ and $\mathbf{M}$ are as above, but $\mathbf{Z} = \mathrm{diag}(z_1, \ldots, z_m)$, where z_i is the i-th class credibility weight and $\mathbb{I}$ the identity matrix.

3.1.1 Type 1 linear credibility is impossible

First note that Type 1 linear credibility is a special case of Type 2; as it is more restrictive.

After adapting Jewell's prior to make it conjugate to the GLM sample density, we show that (omitting details):

1. It does not lead to linear credibility, which is actually impossible in this restrictive context with any prior.
2. No general analytic expression exists for $\boldsymbol{\mu}$, although this might be possible for some special choices of π.

Thus numerical methods or MCMC are needed in order to calculate the posterior mean, which defeats the purpose of using a conjugate prior.

3.1.2 Type 2 linear credibility is sometimes possible

We need for the GLM a $\hat{\beta}_c$ such that $\hat{\boldsymbol{\mu}}_c = \mathbf{g}^{-1}(X\hat{\beta}_c)$, that is:

$$\mathbf{g}^{-1}(X\hat{\beta}_c) = \mathbf{Z}\,\hat{\boldsymbol{\mu}} + (\mathbb{I} - \mathbf{Z})\,\mathbf{M}.$$

It turns out that in non–saturated models, that is $\dim(\beta) < \dim(\boldsymbol{\mu})$, the existence of such a $\hat{\beta}_c$ depends on the observed sample.

We illustrate this with a simple example in dimension 2. Here the portfolio is divided in only 2 classes using a binary covariate with no intercept (otherwise we would have a saturated model):

$$X = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{and} \quad \hat{\beta}_c \in \mathbb{R}.$$

A toy example

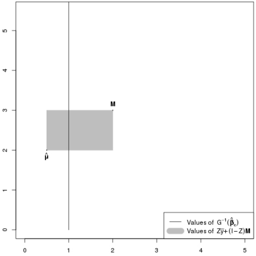
Assuming a log–link function, we need $\hat{\beta}_c$ such that

$$\hat{\boldsymbol{\mu}}_c = \mathbf{g}^{-1}(X\hat{\beta}_c) = \mathbf{g}^{-1}(0, \hat{\beta}_c) = \big(\exp(0), \exp(\hat{\beta}_c)\big) = \big(1, \exp(\hat{\beta}_c)\big).$$

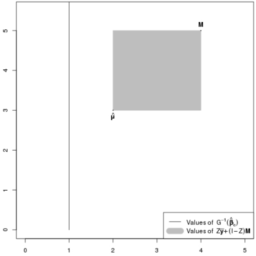
Graphing this equation we see that the left–hand side takes values only on the half upper side of the vertical line $x = 1$.

Imagine 2 scenarios; in Scenario 1, $\hat{\boldsymbol{\mu}} = (0.5, 2)$ and $\mathbf{M} = (2, 3)$, while in Scenario 2, $\hat{\boldsymbol{\mu}} = (2, 3)$ and $\mathbf{M} = (4, 5)$.

The following figure shows the possible values of the left– and right–hand side of the above equation for both scenarios.



(a) Scenario 1



(b) Scenario 2

Figure: Values of the left– and right–hand side in both scenarios

The vertical line = $\hat{\boldsymbol{\mu}}_c$, the rectangle = $Z\,\hat{\boldsymbol{\mu}} + (\mathbb{I} - Z)\,\mathbf{M}$ as the diagonal entries of Z vary from 0 to 1.

3.2 Entropic credibility for GLMs

Computing power is no longer scarce nor expensive, as when Bühlmann and Jewell proposed their credibility premiums.

We propose here a modern, computational Bayesian approach to credibility; an entropic estimator of the mean vector of a GLM as the credibility premium.

An arbitrary prior π is assumed (not necessarily conjugate to the EDF) with corresponding posterior distribution:

$$\pi(\boldsymbol{\beta}, \phi|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\beta}, \phi)\, \pi(\boldsymbol{\beta}, \phi).$$

3.2.1 Estimation of the mean: First good news

Proposition

The entropic estimator $\boldsymbol{\beta}^*$ of the coefficients of a Bayesian GLM are equal to the maximum likelihood estimator of a frequentist GLM with the same covariates, response distribution and weights, but with an observed response vector equal to $\mathbb{E}[\mathbf{Y}]$, the predictive posterior mean.

Corollary

If $\boldsymbol{\beta}^*$ is the entropic estimator of the coefficients of a Bayesian GLM, then the entropic premium is given by

$$\boldsymbol{\mu}^* = \mathbf{g}^{-1}(X\boldsymbol{\beta}^*).$$

3.2.2 Estimation of the dispersion parameter

The entropic credibility estimator also takes into consideration the uncertainty of the dispersion parameter, because the posterior distribution of β depends on the posterior of ϕ (second good news).

This differs from classical credibility where the dispersion parameter is considered known, as in Jewell (1974, AB) or Diaconis and Ylvisaker (1979, AS).

We found only one article that considers a prior distribution for the dispersion parameter: Landsman and Makov (1998, SAJ, see our remarks about this article).

No general procedure

We have not found a general procedure (all EDF/priors) to obtain the entropic estimator of the dispersion parameter.

We present cases where an estimator can be found and the difficulties in obtaining a general solution.

Note that a point–estimator for ϕ is not necessary to obtain the credibility premium, as seen above, or its uncertainty, measured by the posterior distribution.

Proper dispersion model

A case worth mentioning where the estimation minimization simplifies is when the response distribution is a proper dispersion model (see Jorgensen, 1997, C&H, Chap. 5), that is when c in the EDF density can be decomposed as

$$c(y, \phi) = d(y)\, e(\phi).$$

There are only 3 exponential dispersion continnuous models for which this factorization holds: the gamma, inverse Gaussian and normal distributions (see Jorgensen, 1997, C&H, Chap. 5 and Daniels, 1980, Biometrika, for a proof). The following table gives e is for these 3 models.

Distribution	normal	gamma	inverse Gaussian
$e(\phi)$	$\phi^{-1/2}$	$\dfrac{e^{-1/\phi}}{\Gamma(\frac{1}{\phi})\phi^{1/\phi}}$	$\phi^{-1/2}$

Table: $e(\phi)$ for the three proper exponential dispersion families

In the general case, when this factorization does not hold, Markov Chain Monte Carlo (MCMC) is used to simplify the minimization problem of the entropic risk function.

Then the converge of the simulated solutions $\tilde{\phi}_N$ to the true ϕ^* must be verified, as $N \to \infty$.

We do not yet have easy–to–check sufficient conditions to guarantee convergence in all cases; a convergence theorem that can be useful in some cases is given in the paper.

3.3 Numerical illustrations

Algorithm for entropic credibility premiums:

1. Find $\mathbb{E}[\mathbf{Y}]$ from the predictive posterior distribution.
2. Fit a frequentist GLM with the same covariates, response distribution and weights, but with observed response vector $= \mathbb{E}[\mathbf{Y}]$. This gives $\boldsymbol{\beta}^*$, the entropic estimator of the coefficients.
3. Find the entropic mean using $\boldsymbol{\mu}^* = \mathbf{g}^{-1}(X\boldsymbol{\beta}^*)$.

Steps 2 and 3 are simple: can be done in R without major problems (see R Core Team, 2017).

The difficult part is Step 1; one possible solution is MCMC.

Remarks:

- The greater the number of classes m, the more computing is required, both in terms of memory and CPU. It is very useful to first aggregate the data. This can drastically reduce m.
- Continuous variables can make data aggregation useless. Converting their support into intervals transforms them into categorical variables.
- Bayesian methods are time consuming for variable selection (running MCMC simulations for each combination of variables). More practically, select variables using a frequentist GLM (faster to fit). The resulting variables then give the initial Bayesian model.

A car insurance dataset

The R interface to STAN (see Stan Development Team, 2016) was used to find entropic credibility estimators for the severity model of the publicly available dataset that appears in de Jong and Heller (2008, CUP).

It is based on 67,856 one–year policies from 2004 or 2005. It can be down–loaded from the companion site of the book: `http://www.acst.mq.edu.au/GLMsforInsuranceData`, as the dataset called Car.

The following table gives the description of the variables provided on the web–site.

Variable name	Description
veh_value	vehicle value, in $10,000s
exposure	0-1
clm	occurrence of claim (0 = no, 1 = yes)
numclaims	number of claims
claimcst0	claim amount (0 if no claim)
veh_body	vehicle body, coded as BUS CONVT = convertible COUPE HBACK = hatchback HDTOP = hardtop MCARA = motorized caravan MIBUS = minibus PANVN = panel van RDSTR = roadster SEDAN STNWG = station wagon TRUCK UTE - utility
veh_age	age of vehicle: 1 (youngest), 2, 3, 4
gender	gender of driver: M, F
area	driver's area of residence: A, B, C, D, E, F
agecat	driver's age category: 1 (youngest), 2, 3, 4, 5, 6

Table: Vehicle insurance variables

Model fitting

- We fit a GLM with gamma claim severity (*claimcst*0).
- We modified two explanatory variables, dividing the support of the continuous variable *veh_value* into 3 intervals $[0, 1.2)$, $[1.2, 1.86)$ and $[1.86, \infty)$, which we label as $P1$, $P2$ and $P3$, respectively.
- The areas of residence A,B,C and D were also grouped together, thus the variable *area* included in our model takes three values: ABCD, E or F.
- Unfortunately, there is no out–of–sample information here; we use non–informative priors for all parameters: The β's are assumed iid uniform on $(-20, 20)$.
- The dispersion parameter is assumed to follow a uniform distribution on $(0, 1000)$, independently from the betas.

Model fitting (...continued)

- After aggregating the data, the number of observations reduced from 67,856 policies to 101 classes.
- The following table shows the information used for the Bayesian severity GLM. The value between parenthesis on the right of each explanatory variable corresponds to the reference category used in the model.

Model information		MCMC information	
Response distribution	gamma	No. of chains	3
Weight variable	numclaims	Warmup period	2,000
Covariates	agecat(1)		
	gender(F)	Simulations kept	28,000
	area(ABCD)	(per chain)	
	veh_value(P1)		
Prior	betas are iid $U(-20, 20)$ and $\phi \sim U(0, 1000)$		

Table: Severity model

Model fitting (. . . continued)

- In order to find the entropic estimator for the β's, it is first necessary to find the posterior mean $\mathbb{E}[\mathbf{Y}]$ with MCMC.
- For this example the simulations on STAN took about 50 seconds, counting compilation time, on three 2.67GHz processors.
- After, the entropic β's are found by fitting a frequentist GLM with the posterior mean as response vector. The following table shows the entropic coefficients obtained for this data.

Coefficient	Value	Coefficient	Value
(Intercept)	7.784	genderM	0.183
agecat2	-0.207	areaE	0.152
agecat3	-0.303	areaF	0.377
agecat4	-0.301	veh_valueP2	-0.117
agecat5	-0.403	veh_valueP3	-0.156
agecat6	-0.331		

Table: Entropic coefficients

Model fitting (...continued)

- With the above estimators the entropic premium was calculated for each risk class. We do not report here all premiums since there are 101 classes.
- A full table can be found on this article's web–site in the section "Classes Table".
- The following figure shows the entropic premiums (in increasing value) for all classes.

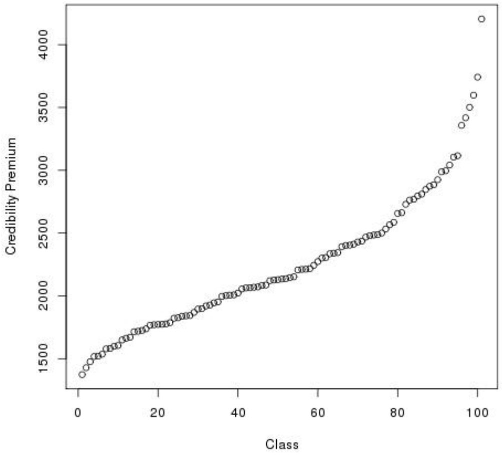


Figure: Entropic credibility premiums (in increasing value)

Conclusions

Linear credibility is artificial as a Bayesian model: Jewell's prior is not chosen for adequacy, but only for ease of computation.

Compared to when Bühlman and Jewell published their work, now computing power is cheap, and sophisticated simulation software readily available to anyone on Internet.

In a modern Bayesian approach to credibility the prior can be based on out–of–sample information, not ease of computation.

The convergence of MCMC simulations is the new limitation on possible priors. We use relative entropy as the loss function.

Compared to classical credibility for the EDF, the advantage here is that the uncertainty on the dispersion parameter is considered. We show that substantial computing is required, but the method is easily applied to insurance datasets.

Links

Preprint: https://arxiv.org/pdf/1710.08553

R code: https://oquijano.net/articles/bayesian_credibility/

Dataset: http://www.acst.mq.edu.au/GLMsforInsuranceData

Thank you for your attention!

Bibliography

Bühlmann, H. (1967) "Experience rating and credibility", *ASTIN Bulletin*, **4**, 3, pp.199–207.

Daniels, H.E. (1980) "Exact saddlepoint approximations", *Biometrika*, **67**(1), pp.59–63.

de Jong, P. and Heller, G.Z. (2008) *Generalized Linear Models for Insurance Data*, Cambridge University Press.

Diaconis, P. and Ylvisaker, D. (1979) "Conjugate priors for exponential families", *The Annals of Statistics*, **7**(2), pp.269–281.

Jewell, W.S. (1974) "Credible means are exact Bayesian for exponential families", *ASTIN Bulletin*, **8**, 1, pp.77–90.

Bibliography (. . . continued)

Jørgensen, B. (1992) "The theory of exponential dispersion models and analysis of deviance", *Instituto de Matemática Pura e Aplicada*, (IMPA), Brazil.

Jørgensen, B. (1997) *The Theory of Dispersion Models*, Chapman & Hall, London.

Landsman, Z.M. and Makov, U.E. (1998) "Exponential dispersion models and credibility", *Scandinavian Actuarial Journal*, 1, pp.89–96.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*, Chapman and Hall, New–York.

Bibliography (. . . continued)

R Core Team (2017) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, `https://www.R-project.org/`.

Stan Development Team (2016) "RSTAN: the R interface to Stan", *R package version 2.14.1*, `http://mc-stan.org/`.