On the (weak) dependence between risk profiles in insurance data analysis

Vázquez–Polo, Francisco–José (fjvpolo@dmc.ulpgc.es)
Dpt. of Quantitative Methods and TiDES Institute
University of Las Palmas de Gran Canaria
Martel, M. (mmartel@dmc.ulpgc.es)
Dpt. of Quantitative Methods and TiDES Institute
University of Las Palmas de Gran Canaria
Hernández–Bastida, A. (bastida@goliat.ugr.es)
Dpt. of Quantitative Methods, University of Granada

ABSTRACT

A common assumption in the statistical model for Bayes premium in the insurance context, is the independence between risk profiles associated with random quantities considered. In this communication we consider the compound collective risk model in which the primary distribution is comprised of the Poisson–Lindley distribution with a λ parameter, and where the secondary distribution is an exponential one with a θ parameter. We consider the case of dependence between risk profiles (i.e., the parameters λ and θ), where the dependence is modelled by a Farlie–Gumbel–Morgenstern family. Statistical properties and some consequences on the Bayes premium of the structure dependence chosen are studied.

Keywords: Bayesian premium, Dependence, Farlie–Gumbel–Morgenstern family of distributions, Poisson–Lindley distribution, risk profile.
Topic: A3 Statistical Methods

On the (weak) dependence between risk profiles in insurance data analysis

ABSTRACT

A common assumption in the statistical model for Bayes premium in the insurance context, is the independence between risk profiles associated with random quantities considered. In this communication we consider the compound collective risk model in which the primary distribution is comprised of the Poisson–Lindley distribution with a λ parameter, and where the secondary distribution is an exponential one with a θ parameter. We consider the case of dependence between risk profiles (i.e., the parameters λ and θ), where the dependence is modelled by a Farlie–Gumbel–Morgenstern family. Statistical properties and some consequences on the Bayes premium of the structure dependence chosen are studied.

1 INTRODUCTION

The main random variables presented in the collective risk model are: the frequency distribution for the number of claims N and a sequence of independent and identically distributed random variables representing the size of the single claims X_i. Frequency N and Severity X_i are assumed to be independent, conditional on distribution parameters. Assumed a given model for these variables (Poisson and exponential, respectively for instance), the interest is then focused in $S = X_1 + ... + X_N$ which denotes the aggregate losses or the total cost over a period.

The computations required to obtain S under the Poisson–exponential model above cited are difficult to perform without the independence hypothesis. Peters *et al.* (2008) proposed that this kind of independence assumption in operational risk models should be investigated further.

On the other hand, actuarial data often present positively skewed and overdispersion. Practitioners need more statistical models which are flexible for fitting data and empirically fits many kinds of loss and/or actuarial data with a strong asymmetry presence (Ghitany et al., 2008) and where some other properties as overdispersion and zero–inflated are usually present in sample observations.

Using the Poisson–Lindley distribution as a primary distribution, we carry out an easy implementable statistical procedure to investigate the importance of the independence assumption and to produce an application that involves computational aspects and a simulated data analysis based on this procedure. Recently, Hernández–Bastida et al. (2011) derived Bayesian premium under the collective risk model using Poisson–Lindley and exponential distributions. Ghitany and Al–Mutairi (2009) provided a comprehensive treatment of statistical behavior of the Poisson–Lindley distribution and its parameter estimation.

In this paper, we propose a model of (weak) dependence between the prior densities of these risk profiles, including the case of independence as a particular case using the Farlie–Gumbel–Morgenstern (FGM) family of distributions (Morgenstern, 1956). By means of these tools, it is a straightforward matter to study how the independence hypothesis affects actuarial decisions. By setting a measure of comparison (for example, the Bayes premium), it suffices to compare this measure over the entire class under consideration with the one that would be obtained under independence.

The paper is organized as follows. In Section 2 we obtain the likelihood derived from the choice of a Poisson–Lindley count distribution and exponential severities, and present the class of priors considered to develop a Bayesian analysis jointly with some interesting properties of these priors distributions. In Section 3 we describe how the models react to variations in the independence of the risk profile priors with respect to the Bayes premium, and how the results obtained can be used in practice. Finally, some conclusions are presented in Section 4.

2 THE MODEL

Sankaran (1970) introduced the discrete Poisson–Lindley distribution by combining the Poisson and Lindley distributions obtaining that a random variable N follows a discrete Poisson–Lindley distribution with parameter λ, when its probability density function is given by

$$\Pr(N = k|\lambda) = \frac{\lambda^2(\lambda + 2 + k)}{(\lambda + 1)^{k+3}}, \qquad k = 0, 1, 2, ... \quad \lambda > 0. \tag{1}$$

Then, it follows from (1) that the moment generating function is given as following:

$$M_1(t;\lambda) = \frac{\lambda^2 (\exp(t) - \lambda - 2)}{(\lambda + 1)^2(\exp(t) - \lambda - 1)}. \tag{2}$$

Some descriptive quantities are now immediately deduced:

$$\mathbb{E}[N] = \frac{\lambda + 2}{\lambda(\lambda + 1)}, \quad \mathrm{Var}(N) = \frac{\lambda^3 + 4\lambda^2 + 6\lambda + 2}{\lambda^2(\lambda + 1)^2},$$

concluding then that the Poisson–Lindley distribution presents overdispersion.

2.1 The likelihood

Suppose now that severities random variables X_i, $i = 0, 1, ...$ follow an Exponential distribution of parameter $\theta \geq 0$,

$$f_2(x_i|\theta) = \theta e^{-\theta x_i}, \qquad x_i > 0. \tag{3}$$

Its moment generating function is given by $M_2(t;\theta) = \dfrac{\theta}{\theta - t}$. The mean and variance, respectively, for each i are then $\mathbb{E}[X_i] = \dfrac{1}{\theta}$, and $\mathrm{Var}(X_i) = \dfrac{1}{\theta^2}$.

We assume conditional independence between claim amounts and claim numbers. Then, in the compound collective model our interest is focused on the random variable "total cost or aggregate loss", S, where its probability density function is defined by

$$f(s|\lambda,\theta)=\sum_{n=0}^{\infty}\Pr(N=n|\lambda)\cdot f_2^{n\star}(x|\theta),$$

where $\Pr(N=n|\lambda)$ denotes the probability that n claims have occurred and $f_2^{n\star}$ is the n–th convolution of the $f_2(x|\theta)$ function in (3).

In order to obtain a closed expression for $f(s|\lambda,\theta)$, we consider the following cases:

1. If $s=0$, then
$$f(0|\lambda,\theta)=\Pr(N=0|\lambda)=\frac{\lambda^2(\lambda+2)}{(\lambda+1)^3}.$$
2. If $s>0$, then after some straightforward calculation we obtain

$$f(s|\lambda,\theta)=\lambda^2(\lambda+1)^{-5}\cdot\theta\cdot(\theta s+(\lambda+1)(\lambda+3))\exp\left(-\frac{\lambda\theta}{\lambda+1}s\right).$$

As consequence of the above result, the mean and variance of variable S are given by

$$\mathbb{E}\left(S\right)=\frac{\lambda+2}{\theta\lambda(\lambda+1)},\qquad\text{and}\qquad \mathrm{Var}\left(S\right)=\frac{2\lambda^3+7\lambda^2+8\lambda+2}{\theta^2\lambda^2(\lambda+1)^2}.$$

2.2 Introducing dependence in the risk profiles: the FGM family of priors

In actuarial literature it is normally assumed that the parameters λ and θ are independent. Although there has been significant theoretical development of the sensitivity procedures in Bayesian statistics for prior independence (Lavine et al., 1991; Wasserman et al., 1993; Berger and Moreno, 1994) relevant applications have been less forthcoming. In this paper we propose to introduce some dependence between the risk profiles θ and λ through the Farlie–Gumbel–Morgenstern (FGM) system. The FGM family has been used recently in Cossette et al. (2008) as a tool to introduce dependence between claim amounts and the interclaim time, but in the context of copula theory.

The bivariate FGM family of distributions (Morgenstern, 1956) has a joint density function of the form

$$\begin{aligned} f(x,y) &= f(x)g(y)\left[1+\omega\left(1-2F(x)\right)\left(1-2G(y)\right)\right] \\ &= f(x)g(y)+\omega\left[f(x)\left(2F(x)-1\right)\right]\left[g(y)\left(2G(y)-1\right)\right]. \end{aligned} \quad (4)$$

It is well known, that the parameter ω is directly proportional to the correlation coefficient and the product moment correlation coefficient for all the FGM distributions with continuous marginals can never exceed $\frac{1}{3}$, corresponding to the case of uniform marginals. In this sense, we refers this situation as relatively weak dependence.

Assuming that the marginal prior distributions are $\lambda \sim \mathcal{G}(\alpha_\lambda, \beta_\lambda)$, and $\theta \sim \mathcal{G}(\alpha_\theta, \beta_\theta)$, i.e. the natural conjugate priors under Poisson or exponential sampling, we proceed now to obtain the correlation coefficient over the FGM family of priors in (4) replacing $f(x)$ by $\pi(\lambda)$ and $g(y)$ by $\pi(\theta)$, respectively.

Observe that

$$\mathbb{E}\left(\lambda \cdot \theta\right) = \frac{1}{\beta_\lambda \beta_\theta} \mathbb{E}\left(\xi_1 \xi_2\right), \quad (5)$$

where $\xi_1 = \beta_\lambda \lambda \sim \mathcal{G}(\alpha_\lambda, 1)$ and $\xi_2 = \beta_\theta \theta \sim \mathcal{G}(\alpha_\theta, 1)$. Then, similarly to D'Este (1981) we obtain

$$\mathbb{E}\left(\lambda \cdot \theta\right) = \beta_\lambda \cdot \beta_\theta \cdot \mathbb{E}\left(\xi_1\right) \cdot \mathbb{E}\left(\xi_2\right) \cdot \left(1 + \omega \left\{ 2\frac{I(\alpha_\lambda)}{B(\alpha_\lambda)} - 1 \right\} \left\{ 2\frac{I(\alpha_\theta)}{B(\alpha_\theta)} - 1 \right\} \right), \quad (6)$$

where $I(\upsilon) = \int_0^1 \frac{z^{\upsilon-1}}{(1+z)^{2\upsilon+1}} dz$ and $B(\upsilon) = \frac{\Gamma(\upsilon)\Gamma(\upsilon+1)}{\Gamma(2\upsilon+1)}$ and function $I(\cdot)$ satisfy the relation

$$2\frac{I(\upsilon)}{B(\upsilon)} - 1 = \frac{1}{\upsilon} \cdot 2^{-2\upsilon} \cdot \frac{\Gamma(2\upsilon+1)}{\Gamma(\upsilon)\Gamma(\upsilon+1)}.$$

Therefore, from (6) we have the following

$$\mathbb{E}\left(\lambda \cdot \theta\right) = \frac{\alpha_\lambda \alpha_\theta}{\beta_\lambda \beta_\theta}\left(1 + \frac{\omega 2^{-2(\alpha_\lambda+\alpha_\theta)}}{\alpha_\lambda \alpha_\theta} \cdot \frac{\Gamma(2\alpha_\lambda+1)\Gamma(2\alpha_\theta+1)}{\Gamma(\alpha_\lambda)\Gamma(\alpha_\lambda+1)\Gamma(\alpha_\theta)\Gamma(\alpha_\theta+1)}\right), \quad (7)$$

and the (a priori) structure of dependence between the risk profiles λ and θ measured by the correlation coefficient is given by

$$\begin{aligned} \mathrm{Corr}(\lambda,\theta) &= \frac{\mathrm{Cov}(\lambda,\theta)}{\sqrt{\mathrm{Var}(\lambda)\mathrm{Var}(\theta)}} \\ &= \frac{\omega}{\sqrt{\alpha_\lambda\alpha_\theta}} \cdot \frac{1}{2^{2(\alpha_\lambda+\alpha_\theta)} \cdot B(\alpha_\lambda,\alpha_\lambda+1) \cdot B(\alpha_\theta,\alpha_\theta+1)}. \end{aligned} \quad (8)$$

For the special case, $\alpha_\lambda = \alpha_\theta = 1$ and since $|\omega| \leq 1$, it follows that

$$|\mathrm{Corr}(\lambda,\theta)| = \frac{|\omega|}{4} \leq \frac{1}{4}. \quad (9)$$

3 Sensitivity of the independence hypothesis

Firstly, let us consider the class of priors given in (4) partitioned into two subclasses of priors reflecting positive or negative dependence each one, respectively:

$$\Pi = \Pi_{(\omega>0)} \cup \Pi_{(\omega<0)},$$

where

$$\Pi_{(\omega>0)} = \left\{\pi(\lambda,\theta) = (1-\omega)\pi_I(\lambda,\theta) + \omega\pi_{(\omega=1)}(\lambda,\theta), \omega \in [0,1]\right\},$$

$$\Pi_{(\omega<0)} = \left\{\pi(\lambda,\theta) = (1-\omega)\pi_I(\lambda,\theta) + \omega\pi_{(\omega=-1)}(\lambda,\theta), \omega \in [0,1]\right\},$$

$\pi_I(\lambda,\theta) = \pi_1(\lambda) \cdot \pi_2(\theta)$ is the prior density obtained under independence and $\pi_{(\omega=1,-1)}(\lambda,\theta)$ in the FGM family with $\omega = 1$ (-1) are fixed densities with marginals $\pi_1(\lambda)$ and $\pi_2(\theta)$, representing the larger of positive (negative) dependence (De la Horra and Fernándes, 1995). In order to test the influence of the independence hypothesis on posterior decisions, we focus the problem in the following way.

The Bayesian premium (i.e., the posterior mean of the true individual premium) plays an important role in ratemaking. As we know, $\mathbb{E}(S) = \frac{\lambda+2}{\lambda(\lambda+1)\theta}$. Then, the Bayes premium is obtained as the (posterior) expected value,

$$\mathbb{E}_{\pi_\omega(\cdot|s)}\left(\frac{\lambda+2}{\lambda(\lambda+1)}\frac{1}{\theta}\right) = \int_0^\infty\int_0^\infty \frac{\lambda+2}{\lambda(\lambda+1)}\frac{1}{\theta}\pi_\omega(\lambda,\theta|s)d\lambda d\theta. \tag{10}$$

We present a particular way of determining whether there are large departures from premium measures when the assumption of prior independence is relaxed, and we find a method to account for such consequences in several common situations. That is, once the data are observed, we are interested in upper and lower bounds of these posterior quantities in (10) over class Π.

As in De la Horra and Fernández (1995), differentiating with respect to ω, the above bounds are calculated comparing only the three following quantities:

$$\frac{\int\int h(\lambda,\theta)\cdot f(s|\lambda,\theta)\cdot \pi_{(\omega=i)}(\lambda,\theta)d\lambda d\theta}{\int\int f(s|\lambda,\theta)\cdot \pi_{(\omega=i)}(\lambda,\theta)d\lambda d\theta}, \quad i=-1,1, \tag{11}$$

$$\text{and} \qquad \frac{\int\int h(\lambda,\theta)\cdot f(s|\lambda,\theta)\cdot \pi_I(\lambda,\theta)d\lambda d\theta}{\int\int f(s|\lambda,\theta)\cdot \pi_I(\lambda,\theta)d\lambda d\theta}, \tag{12}$$

where $h(\lambda,\theta) = \frac{\lambda+2}{\lambda(\lambda+1)}\frac{1}{\theta}$ and $f(s|\lambda,\theta)$ is the likelihood function given in (3).

The difference between the upper and lower bound obtained from (11) and (12), denoted by $U-L$, is a measure of the robustness (or its absence, i.e. sensitivity) of the prior independence, for different values of s. In order to standardize this measure, we use a slight modification of the RS factor (Sivaganesan, 1991) defined by

$$RS = 100\frac{U-L}{2\mathbb{E}_{\pi_I}(h(\lambda,\theta))}. \tag{13}$$

RS is a standardized factor which can be thought of as the percentage variation in the Bayes premium as π_α varies over Π on either side of $\mathbb{E}_{\pi_I}(h(\lambda,\theta)|s)$, which is used as a pattern (independence scene), like the centre of the variation interval (L,U).

Example *Consider an insurance business where the number of claims N has a Poisson–Lindley distribution with the parameter λ. Suppose also that each single claim size distribution is exponential with parameter θ. As commented in previous sections, one of the most useful compound collective risk models consists in assuming a Gamma prior distribution over λ (and θ). This is reasonable, since the shape of the Gamma density is very flexible (Miller and Hickman, 1974; Scollnik, 1995; among others). Let us now consider a numerical illustration. Supposing the actuary assumes the expected frequency to be $\mathbb{E}(\lambda) = 1$ with "no claims" as the most frequent event. Hence, with these two items of partial prior information, it is reasonable to assume that the base prior $\pi_1(\lambda)$ is $\mathcal{G}(1,1)$ (with this elicitation the actuary knows that the mode is around 0). Using similar reasoning, suppose that the prior density for θ is also $\mathcal{G}(1,1)$, (i.e., the expert expects a claim size of 1 monetary unit). Due to the exponential behaviour of the above priors, we can consider this elicited prior within a weak prior information context.*

The ranges of the Bayes premium and the RS factor, for various values of s (from 0 to 10 by steps of 0.01) are shown in Figure 1. Similar conclusions may be obtained with greater values of s. The computations involved in (10)–(12) were carried out using `Mathematica` software. Several minutes of CPU time were needed to complete the calculations. The sensitivity of the answer to independence departures was measured by considering the RS factor in (13).

Some interesting general points emerge from Figure 1. First, observe that for the weak prior information ($\alpha_\lambda = \beta_\lambda = \alpha_\theta = \beta_\theta = 1$) scene considered, the correlation between the risk profiles in (9) is bounded by $\frac{1}{4}$. Then, we would expect a similar behaviour of the RS–factor over the class. If ω is used to control the confidence level of practitioners concerning the independence assumption, similar behaviour would be expected for the RS–factor, for example, but this does not hold.

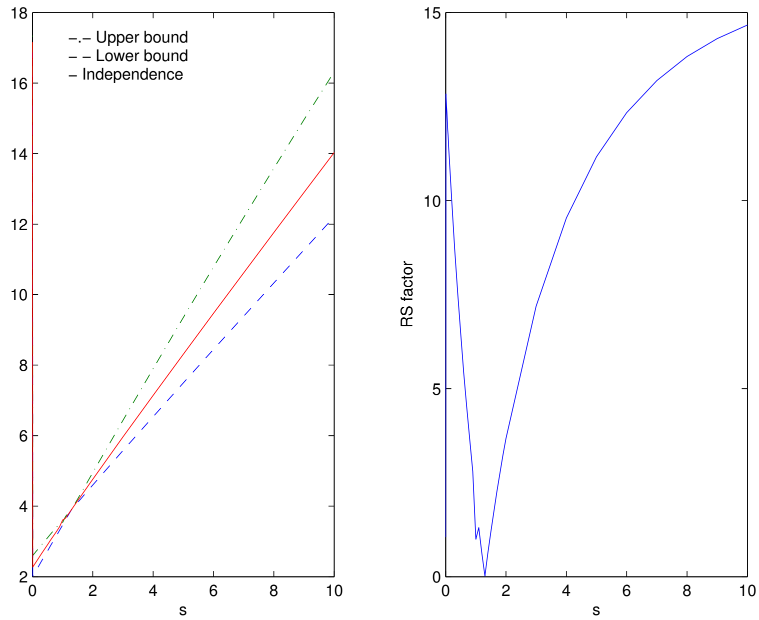

Figure 1: Left: Upper (dash–point line) and Lower (dashed line) bounds for the Bayes premium in the FGM family with both fixed marginals $\mathcal{G}(1,1)$. Continuous line represents the Bayes premium under independence hypothesis. Right: RS factor.

4 CONCLUSIONS

In this paper we have examined the hypothesis of independence between the risk profiles (the parameters of the problem). To do this, firstly we propose an alternative close–fitting collective risk model where the primary and secondary distributions are Poisson–Lindley and exponential, respectively. Secondly, dependence was modelled using the Farlie–Gumbel–Morgenstern family, and the coefficient of linear correlation was determined with respect to the prior bidimensional distribution, in the case of independence, which is also known as the index of mutual dependence. Subsequently, we set out to analyze the robustness of a posterior magnitude of interest, the Bayes premium, with respect to variations from independence. An analytic path was developed, and various specific contexts examined. The numerical conclusions obtained reveal considerable, and very marked, differences between the values of the Bayes premiums within contexts of equal linear correlation.

Acknowledgments

Research partially funded by grant ECO2009–14152 (MICINN, Spain).

5 REFERENCES

- Berger, J.O. and Moreno, E. (1994). "Bayesian robustness in bidimensional models: prior independence (with discussion)". Journal of Statistical Planning and Inference, 40, pp. 161–176.

- Cossette, H.; Marceau, E. and Marri, F. (2008). "On the compound Poisson risk model with dependence based on a generalized Farlie–Gumbel–Morgenstern copula". Insurance: Mathematics and Economics, 43 (3), pp. 444–455.

- D'Este G.M. (1981). "A Morgenstern–type bivariate gamma distribution". Biometrika, 68, pp. 339–340.

- De la Horra, J. and Fernández, C. (1995). "Sensitivity to prior independence via Farlie–Gumbel–Morgenstern model". Communications in Statistics: Theory and Methods, 24 (4), pp. 987–996.

- Ghitany, M.E.; Al–Mutairi, D.K. and Nadarajah, S. (2008). "Zero–truncated Poisson–Lindley distribution and its applications". Mathematics and Computers in Simulation, 79, pp. 279–287.

- Ghitany, M.E. and Al–Mutairi, D.K. (2009). "Estimation methods for the discrete Poisson–Lindley distribution". Journal of Statistical Computation and Simulation, 79, pp. 1–9.

- Hernández–Bastida, A.; Fernández–Sánchez, M.P. and Gómez–Déniz, E. (2011). "Collective risk model: Poisson–Lindley and exponential distributions for Bayes premium and operational risk". Journal of Statistical Computation and Simulation, 81 (6), pp. 759–778.

- Lavine, M.; Wasserman, L. and Wolpert, R. (1991). "Bayesian inference with specified prior marginals". Journal of the American Statistical Association, 86, pp. 964–971.

- Miller, R. and Hickman, J. (1974). "Insurance credibility theory and Bayesian estimation". In Kahn, P. M. (ed.) *Credibility theory and applications*, pp. 249–270. New York: Academic Press.

- Morgenstern, D. (1956). "Einfache beispiele zweidimensionaler verteilungen". Mitt. Math. Statistik, 8, pp. 234–235.

- Peters, G.W., Shevchenko, P.V. and Wüthrich, M.V. (2008). "Dynamic operational risk: modelling dependence and combining different sources of information". Preprint, CSIRO, call number CMIS2741.

- Sankaran, M. (1970). "The discrete Poisson–Lindley distribution". Biometrics, 26, pp. 145–149.

- Scollnik, D.P.M. (1995). "Bayesian analysis of two overdispersed Poisson regression models". Communications in Statistics: Theory and Methods, 24 (11), pp. 2901–2918.

- Sivaganesan, S. (1991). "Sensitivity of some posterior summaries when the prior is unimodal with specified quantiles". Canadian Journal of Statiscis, 19, pp. 57–65.

- Wasserman, L.; Lavine, M. and Wolpert, R. (1993). "Linearization of Bayesian robustness problems". Journal of Statistical Planning and Inference, 37, pp. 307-316.