Online supplement for "Market-based insurance ratemaking: application to pet insurance"

Pierre-Olivier Goffard [*1], Pierrick Piette[†2,3], and Gareth W. Peters[‡4]

[1] Université de Strasbourg, Institut de Recherche Mathématique Avancée, Strasbourg, France

[2] Univ Lyon, Université Claude Bernard Lyon 1, Institut de Science Financière et d'Assurances (ISFA), Laboratoire SAF EA2429, F-69366, Lyon, France

[3] Seyna, 10 Rue du Faubourg Montmartre 75009 Paris

[4] University of California Santa Barbara, Department of Statistics and Applied Probability, Santa Barbara CA 93106-3110, USA

January 21, 2025

The goal of this supplementary material is to discuss the identifiability issue associated to finding the parameters of the risk model. Let X be a random variable that represents the total health expenses over a given time period, say one year, associated to a pet. Our goal is to find the parameter θ that best explains our market data made of insurance quotes

$$\widetilde{p}_i = f_i\left[\mathbb{E}(g_i(X))\right],\ i = 1,\dots,n,$$

where g_i are the coverage functions and f_i are the loading function. The coverage functions are known and of the form

$$g(x) = \min(\max(r\cdot x - d, 0), l),$$

where r is the rate of coverage, d is the deductible and l is the limit.The loading functions are unknown and will be approximated by a generic function f.

In Section 1, we consider the problem of finding θ if we know access to the pure premiums

$$p_i = \left[\mathbb{E}(g_i(X))\right],\ i = 1,\dots,n,$$

[*]Email: goffard@unistra.fr.

[†]Email: pierrick.piette@gmail.com.

[‡]Email: garethpeters@ucsb.edu.

which is not a realistic situation in practice. We look into the actual problem in Section 2. An isotonic regression model is used to approximate f. This choice is compared to a linear fit $f(x) = b \cdot x$. We consider the model

$$X = \sum_{k=1}^{N} U_k,$$

where $N \sim \text{Pois}(\lambda = 3)$ and the U_k are IID and lognormally distributed $U \sim \text{LogNormal}(\mu = 0, \sigma = 1)$. Our aim is to estimate the parameters $\theta = (\ \lambda \quad \mu \quad \sigma\ )$ based on three insurance quotes associated to the following coverage functions

$$g_1 = 0.85 \cdot X,\ g_2 = \max(X - 1.8, 0), \text{ and } g_3 = \min(X, 6).$$

In terms of coverage rate, deductible and limit, it is equivalent to $(r_1 = 0.85, d_1 = 0, l_1 = \infty)$, $(r_2 = 1, d_2 = 1.8, l_2 = \infty)$, and $(r_3 = 1, d_3 = 0, l_3 = 6)$ respectively. The pure premium associated to these coverages are provided in Table 1.

r	d	l	Pure premium
0.75	0.00	Inf	3.71
1.00	1.80	Inf	3.40
1.00	0.00	6.00	3.63

Table 1: Pure premium associated to g_1, g_2 and g_3 of the $\text{Pois}(\lambda = 3) - \text{LogNormal}(\mu = 0, \sigma = 1)$ risk model.

1 Optimization problem 1

The first optimization that we consider reads as follows:

Problem 1. *Let*

$$p_i^\theta = \mathbb{E}_\theta\left[g_i(X)\right],\ \textit{for } i = 1, \ldots, n,$$

the pure premium associated to a risk X parametrized by $\theta \in \Theta \subset \mathbb{R}^d$. We wish to find $\theta \in \Theta \subset \mathbb{R}^d$ to minimize $d\left[p_{1:n}, p_{i:n}^\theta\right]$, where $d(\cdot,\cdot)$ denotes a distance function over the observation space.

We measure the discrepancy between observed and model-generated pure premiums using the root mean square error (RMSE) defined as

$$\text{RMSE}\left[p_{1:n}, p_{i:n}^\theta\right] = \sqrt{\sum_{i=1}^{n} w_i^{\text{RMSE}} [p_i - p_i^\theta]^2}. \tag{1}$$

The statistical framework is that of minimum distance estimation. We do not have access to the full shape of the data distribution. We must base our inference on specific moments, just as in the generalized method of moments, a popular method among econometricians (see Hansen [1]). The model is identifiable if there exists a unique θ^* such that

$$\theta^* = \underset{\theta \in \Theta}{\arg\min}\, \text{RMSE}(p_{1:n}, p^{\theta}_{1:n}). \tag{2}$$

Existence stems from the fact that the parameter space Θ is compact and the map $\theta \mapsto \text{RMSE}(p_{1:n}, p^{\theta}_{1:n})$ is continuous. Uniqueness is more difficult to verify as it depends on the functional g_i's. Given the model for X and the insurance coverages, the pure premium do not have an analytical expression making it difficult to show the convexity of (1). A simple necesssary condition is that the number of parameters must be smaller than n, the number of moments considered. We consider here a situation where we have only three pure premiums ($n = 3$) and we know that

$$\text{RMSE}\left[p_{1:n}, p^{\theta^*}_{1:n}\right] = 0,$$

Figure 1 shows the plot of the functions

$$\lambda \mapsto d\left(p_{1:n}, p^{\theta}_{1:n}\right)\Big|_{\mu=0,\sigma=1}, \ \mu \mapsto d\left(p_{1:n}, p^{\theta}_{1:n}\right)\Big|_{\sigma=1,\lambda=3}, \text{ and } \sigma \mapsto d\left(p_{1:n}, p^{\theta}_{1:n}\right)\Big|_{\mu=0,\lambda=3}.$$

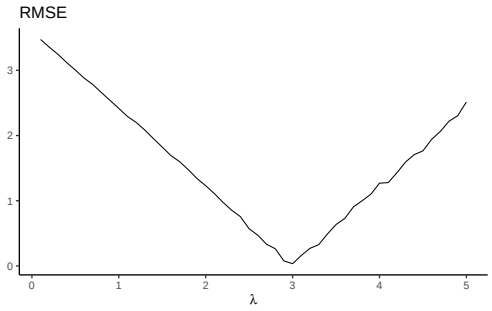
By taking the parameters separately, it looks like the model is identifiable. It does not mean that we can identify a unique parametrization if we do not fix two parameters out of three. We cannot use a grid search procedure to explore the three dimensional parameter space and so we use the optimization procedure described in Section 3.3 of our paper. The parameters of the algorithm are set as follows:

$$J = 500, \ R = 500, \text{ and } \epsilon_{\text{min}} = 0.02.$$

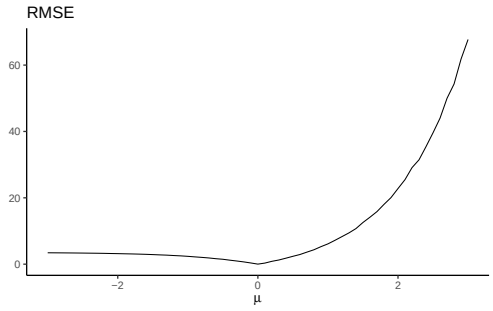
Recall that J is the population size of the clouds of particles, R is the number of Monte Carlo replications and ϵ_{min} is a threshold for the tolerance level. The algorithm will stop if the tolerance threshold reaches ϵ_{min}. The prior distribution of the parameters are given by

$$\lambda \sim \text{Unif}([0, 10]), \ \mu \sim \text{Unif}([-3, 3]) \text{ and } \sigma \sim \text{Unif}([0, 2]).$$

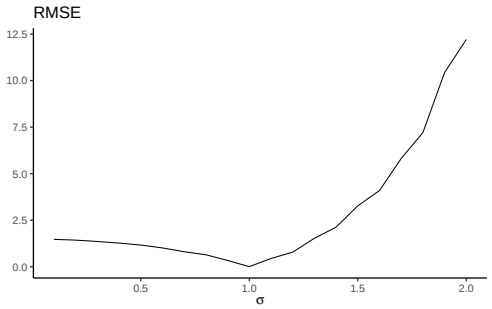
The choice of the level of $\epsilon_{\text{min}} = 0.02$ follows from an estimation of the error on the RMSE resulkting from the use of a Monte Carlo approximation to calculate the pure premium. If we were able to compute exactly the pure premium then our algorithm should return some dirac distributions pointing to the true parameter value. The fact that we must use a Monte Carlo approximation justifies the use of an Approximate Bayesian Computation procedure as we encapsulate in the posterior distribution the uncertainty associated to the Monte



(a) $\lambda \mapsto d\left(p_{1:n}, p^{\theta}_{1:n}\right)\Big|_{\mu=0,\sigma=1}$



(b) $\mu \mapsto d\left(p_{1:n}, p^{\theta}_{1:n}\right)\Big|_{\sigma=1,\lambda=3}$



(c) $\sigma \mapsto d\left(p_{1:n}, p^{\theta}_{1:n}\right)\Big|_{\mu=0,\lambda=3}$

Figure 1: Plots of the RMSE when looking into each parameter separately.

Carlo error. Here we conclue that with these three pure premium allows us to identify the true parameters of the model. The situation is more difficult when we only have access to thecommercial premium as in the forthcoming section.

2 Optimization problem 2

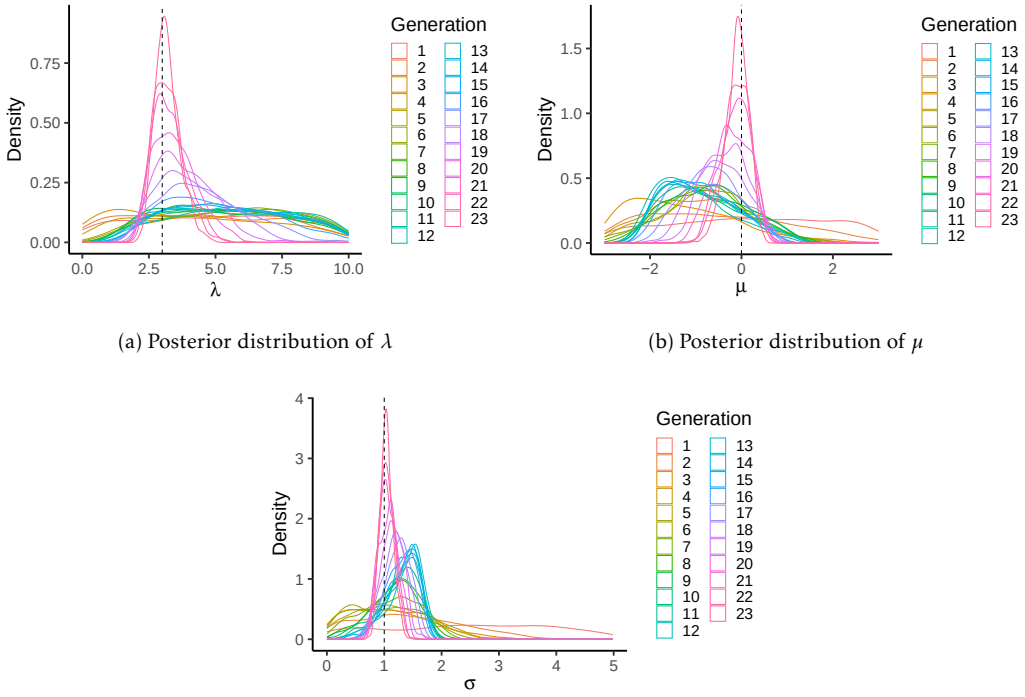
As mentionned earlier, we do not have access to the pure premiums. Instead we have a collection of commercial rates $\widetilde{p}_{1:n} = \{\widetilde{p}_1, \ldots, \widetilde{p}_n\}$ defined as

$$\widetilde{p}_i = f_i(p_i) = f_i\left\{\mathbb{E}_{\theta_0}\left[g_i(X)\right]\right\},\ i = 1, \ldots, n.$$

It leads us to formulate the following optimization problem:

Problem 2. *Denote by*

$$p^{\theta}_i = \mathbb{E}_{\theta}\left[g_i(X)\right],\ \textit{for } i = 1, \ldots, n,$$



(a) Posterior distribution of λ

(b) Posterior distribution of μ

(c) Posterior distribution of σ

Figure 2: Posterior distributions of the parameters of the Pois(λ) – LogNormal(μ, σ) risk model.

the pure premium associated to a risk X parametrized by $\theta \in \Theta \subset \mathbb{R}^d$. Further denote by $p^{\theta}_{1:n}$ and $\widetilde{p}_{1:n}$ the collections of pure and commercial premiums.

We wish to find $\theta \in \Theta \subset \mathbb{R}^d$ and $f : \mathbb{R}_+ \mapsto \mathbb{R}_+$ to minimize $d\left[\widetilde{p}_{1:n}, f(p^{\theta}_{i:n})\right]$,

where the function f is applied elementwise on $p^{\theta}_{i:n}$ and $d(\cdot,\cdot)$ denotes a distance function over the observation space, subject to

$$\widetilde{p}_i \geq p^{\theta}_i \tag{3}$$

and

$$f(p^{\theta}_i) \geq p^{\theta}_i, \tag{4}$$

for $i = 1, \ldots, n$.

Our first task is to find a generic function f to represent the safety loading functions f_i's used by the competitors. For this, we use isotonic regression. It is a statistical technique used for fitting a non-decreasing function

to a set of data points. The idea is that if two pure premium satisfy $p_i \leq p_j$ then the commercial premium should also verify $\widetilde{p}_i \leq \widetilde{p}_j$. Consider a collection of candidate pure premiums $p^{\theta}_{i:n}$, associated to a candidate risk parameter θ. Our datapoints are therefore pairs of pure and commercial premiums $(p^{\theta}_i, \widetilde{p}_i)_{i=1,\ldots,n}$. Suppose the pure premium have been ordered such that $p^{\theta}_i \leq p^{\theta}_j$ for $i \leq j$, isotonic regression seeks a least square fit $\widetilde{p}^{\theta}_i$ for the $\widetilde{p}_i$'s such that $\widetilde{p}^{\theta}_i \leq \widetilde{p}^{\theta}_j$ for $p^{\theta}_i \leq p^{\theta}_j$. It reduces to find $\widetilde{p}^{\theta}_1, \ldots, \widetilde{p}^{\theta}_n$ that minimize

$$\sum_{i=1}^{n} w^{\text{ISO}}_i (\widetilde{p}^{\theta}_i - \widetilde{p}_i)^2, \text{ subject to } \widetilde{p}^{\theta}_i \leq \widetilde{p}^{\theta}_j \text{ whenever } p^{\theta}_i \leq p^{\theta}_j,$$

where $\left(w^{\text{ISO}}_i\right)_{i=1,\ldots,n}$ denotes the weights associated to each pair $(p^{\theta}_i, \widetilde{p}_i)_{i=1,\ldots,n}$. Since the p^{θ}_i's fall in a totally ordered space, a simple iterative procedure called the Pool Adjacent Violators Algorithm (PAVA) can be used. Here's a high-level overview of how it works:

1. Initialize the sequence of values to be the same as the data points $\widetilde{p}^{*}_i = \widetilde{p}_i$.

2. Iterate through the sequence and identify "violations," which occur when the current value is greater than the next value, that is

   $$\widetilde{p}^{*}_i > \widetilde{p}^{*}_{i+1} \text{ for some } i = 1, \ldots, n.$$

   When a violation is found, adjust the values in the associated segment of the sequence to be the average of the values,

   $$\widetilde{p}^{*}_i \leftarrow (\widetilde{p}^{*}_i + \widetilde{p}^{*}_{i+1})/2,$$

   ensuring monotonicity.

3. Repeat Step 2 until no violations are left.

We use the `isoreg` function from *R* to get the fitted values $\widetilde{p}^{\theta}_i$, $i = 1, \ldots, n$. To complete the isotonic regression task we shall find a function f such that $f(p^{\theta}_i) = \widetilde{p}^{\theta}_i$. A common choice is a piece-wise constant function that interpolates the $\widetilde{p}^{\theta}_i$'s. We now turn to the definition of a distance. Our starting point to measure the discrepancy between observed and model-generated commercial rates is the root mean square error (RMSE) defined as

$$\text{RMSE}\left[\widetilde{p}_{1:n}, f(p^{\theta}_{i:n})\right] = \sqrt{\sum_{i=1}^{n} w^{\text{RMSE}}_i [\widetilde{p}_i - f(p^{\theta}_i)]^2}, \tag{5}$$

for a candidate risk parameter θ and an isotonic fit f. We seek

$$\theta^{*} = \underset{\theta \in \Theta}{\arg\min}\, \text{RMSE}\left[\widetilde{p}_{1:n}, f(p^{\theta}_{i:n})\right].$$

The existence of such θ^* is guaranteed because $\theta \mapsto \text{RMSE}\left[\widetilde{p}_{1:n}, f(p^{\theta}_{i:n})\right]$ only takes a finite number of values. Indeed, to each $\theta \in \Theta$ is associated a unique permutation $s^{\theta} \in S_n$, where S_n denotes the set of all the permutations of $\{1,\dots,n\}$, such that

$$p^{\theta}_{s^{\theta}(1)} \leq \dots \leq p^{\theta}_{s^{\theta}(n)}.$$

This permutation s^{θ} defines a unique isotonic fit f based on

$$\widetilde{p}^{\theta}_{s^{\theta}(1)} \leq \dots \leq \widetilde{p}^{\theta}_{s^{\theta}(n)},$$

leading to a given RMSE value $\text{RMSE}\left[\widetilde{p}_{1:n}, f(p^{\theta}_{i:n})\right]$. Concretely, for $\theta_1, \theta_2 \in \Theta$, if it holds that $s^{\theta_1} = s^{\theta_2}$ then $\text{RMSE}\left[\widetilde{p}_{1:n}, f(p^{\theta_1}_{i:n})\right] = \text{RMSE}\left[\widetilde{p}_{1:n}, f(p^{\theta_2}_{i:n})\right]$. The application $\theta \mapsto s^{\theta}_n$ is surjective since $S^{\Theta}_n = \{s^{\theta}_n \, ; \, \theta \in \Theta\}\}$ is finite. The fact that Θ is a continuous space implies that θ^* cannot be unique. Our problem is an ill-posed inverse problem. Ill-posedness is usually dealt with by adding a regularization to the objective function that one wants to minimize. The ratio of $p/\widetilde{p}$ corresponds to what practitioners would call the expected Loss Ratio (LR). Our solution is based on targeting a given loss ratio. The loss ratio is a standard measure to assess the profitability of insurance lines of business. An insurance company that enters a new market is likely to have insights on the loss ratio relative to this market, for example by having informal discussions with reinsurers, brokers or competitors. These insights may translate into the definition of a lower and upper bound denoted by LR_{low} and LR_{high}, respectively. We can then assume that the loss ratios $\text{LR}_i = p_i/\widetilde{p}_i$, for $i = 1,\dots,n$, should fall in the range $[\text{LR}_{\text{low}}, \text{LR}_{\text{high}}]$, which we refer to as the loss ratio corridor. Assuming that $\text{LR}_{\text{high}} < 1$, we may ensure both constraint (3) and $\text{LR}_i \in [\text{LR}_{\text{low}}, \text{LR}_{\text{high}}]$ by adding to our distance (5) two regularization terms defined as

$$\text{Reg}_{\text{low}}\left(\widetilde{p}_{1:n}, p^{\theta}_{1:n}\right) = \sqrt{\sum_{i=1}^{n} w_i^{\text{RMSE}} \left(\widetilde{p}_i - p^{\theta}_i \cdot \text{LR}^{-1}_{\text{low}}\right)^2_+},$$

and

$$\text{Reg}_{\text{high}}\left(\widetilde{p}_{1:n}, p^{\theta}_{1:n}\right) = \sqrt{\sum_{i=1}^{n} w_i^{\text{RMSE}} \left(p^{\theta}_i \cdot \text{LR}^{-1}_{\text{high}} - \widetilde{p}_i\right)^2_+},$$

where $(x)_+ = \max(x, 0)$ denotes the positive part of x. The distance we consider within Problem 2 is now given by

$$d\left[\widetilde{p}_{1:n}, f(p^{\theta}_{1:n})\right] = \text{RMSE}\left[\widetilde{p}_{1:n}, f(p^{\theta}_{1:n})\right] + \text{Reg}_{\text{low}}\left(\widetilde{p}_{1:n}, p^{\theta}_{1:n}\right) + \text{Reg}_{\text{high}}\left(\widetilde{p}_{1:n}, p^{\theta}_{1:n}\right).$$

We take again our three insurance coverages g_1, g_2 and g_3 and we distinguish two safety loading situations. The well-specified case in Section 2.1 and the miss-specified on in Section 2.2

2.1 Well-specfifed safety loading

Consider the following loading functions

$$f_1(x) = 1.38 \cdot x,\ f_2(x) = 1.1 \cdot x, \text{ and } f_3(x) = 1.4 \cdot x.$$

The isotonic model is well specified in this case because the pure and commercial premium are ordered in the same way. In that case the isotonic regression exactly interpolates the datapoints as shown on Figure 3.

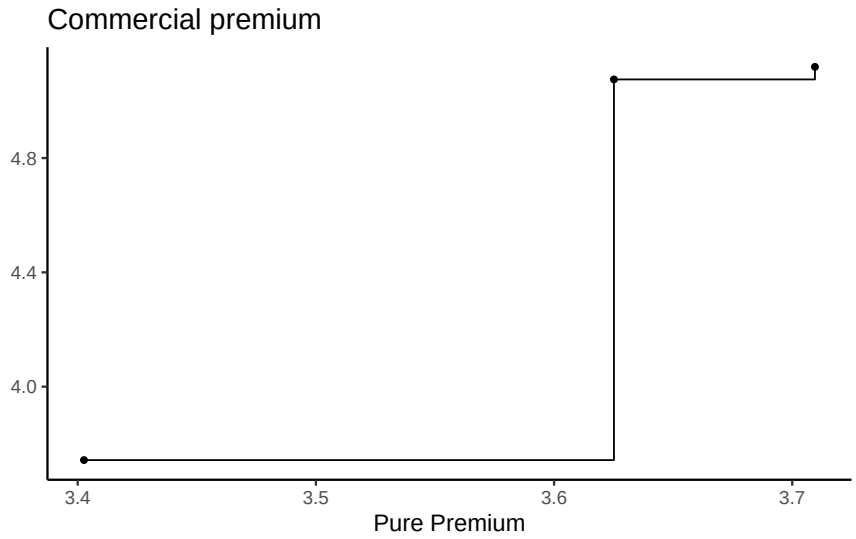


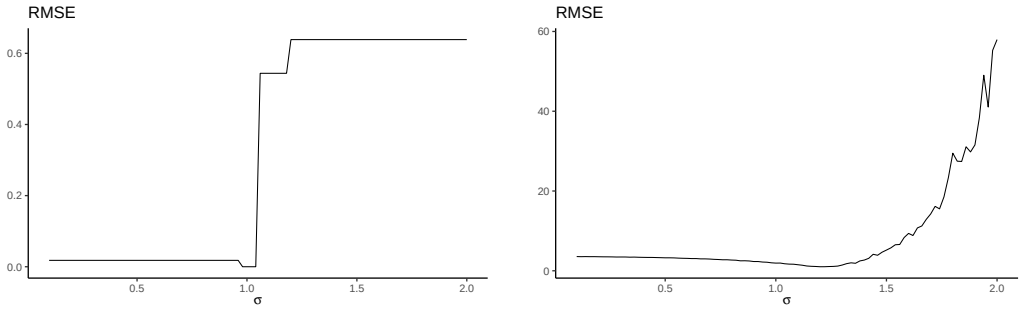
Figure 3: Intermediate posterior distributions for λ and σ.

Figure 1c shows the plot of the $\sigma \mapsto d\left(p_{1:n}, p^{\theta}_{1:n}\right)\Big|_{\lambda=3,\mu=0}$ when f is isotonic (Figure 4a) and when f is linear (Figure 4b).

Note that in this example we do not cancrel the regularization terms. When f is isotonic, the values of σ for which the RMSE is null are associated to the same ordering of the pure premium than that of the true parameter $\sigma = 1$, see Figure 4a. Several values of σ are therefore optimal including the true one. On the other hand, when f is linear, the RMSE is minimal for the parameter value that manage to align the commercial premium. It looks like one value of σ however it is far from the true value, see Figure 4b. When looking at Figure 4a, one might expect that by running our ABC algorithm we shall get a uniform distribution as posterior distribution. We set the parameters of the algorithm as

$$J = 2000,\ R = 500, \text{ and } \epsilon_{\min} = 10^{-}8.$$

We set a prior distribution on σ as $\sigma \sim \mathsf{Unif}([0, 2])$. The tolerance level is very low as we are after the values of



(a) f is isotonic

(b) f is linear

Figure 4: $\sigma \mapsto d\left(p_{1:n}, p^{\theta}_{1:n}\right)\Big|_{\mu=0,\sigma=1}$.

sigma that yields the right ordering associated to a nearly null RMSE. The posterior distribution is provided on Figure 5.

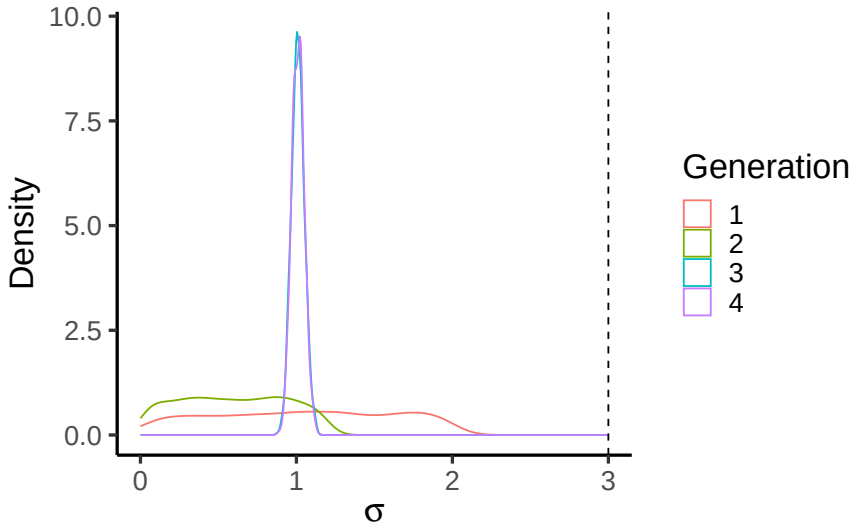


Figure 5: Intermediate posterior distributions for σ.

It turns out that the posterior distribution is not all that uniform, this is explained by the Monte Carlo approximation we use which might change the pure premium ordering when getting close to the edge of the range of admissible σ values. The posterior distribution displayed on Figure 5 could look good enough to the naive eye and one could argue that regularization is not required here. When looking for λ, μ and σ at the same time the parameters can offset each other and the right ordering of the pure premium does not allow us

to extract relevant information. We run our ABC algorithm with the following settings

$$J = 500,\ R = 500, \text{ and } \epsilon_{\min} = 10^{-}8.$$

The prior distributions on the model parameters are as follows:

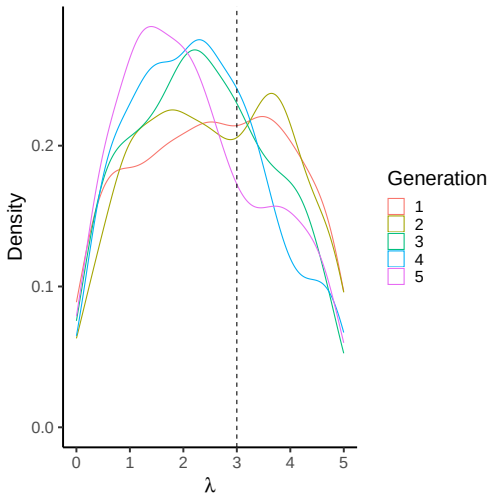
$$\lambda \sim \mathsf{Unif}([0, 10]),\ \mu \sim \mathsf{Unif}([-3, 3]) \text{ and } \sigma \sim \mathsf{Unif}([0, 2]).$$

The posterior distribution are provided on Figure 1c

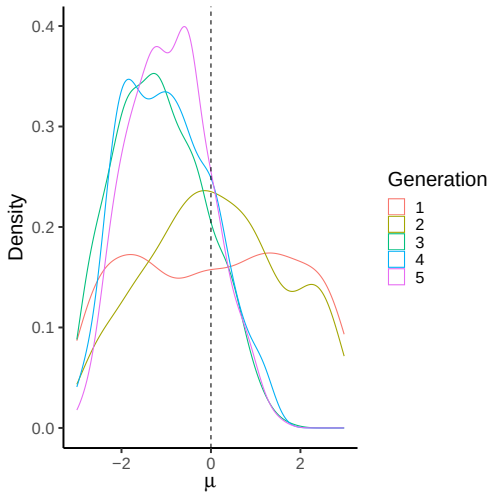
2.2 Misspecified safety loading

References

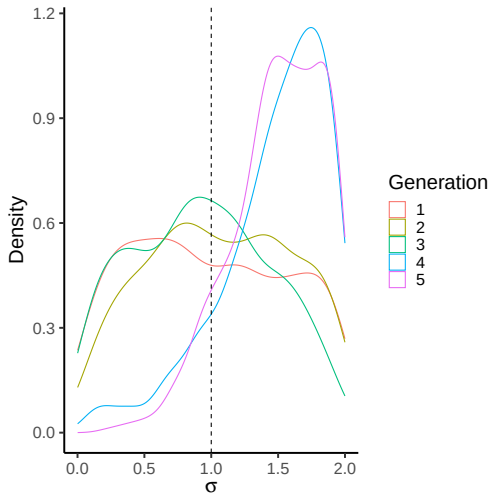
[1] Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029, July 1982. ISSN 0012-9682. doi: 10.2307/1912775.



(a) Posterior distribution of λ



(b) Posterior distribution of μ



(c) Posterior distribution of σ

Figure 6: Posterior distributions of the parameters of the Pois(λ) – LogNormal(μ, σ) risk model.