SOME MODELS OF INFERENCE IN THE RISK THEORY FROM A BAYESIAN VIEWPOINT

LUCIANO DABONI
Trieste, Italy

1. FOREWORD

Our purpose is to introduce some models of inference for risk processes. The bayesian viewpoint is adopted and for our treatment the concepts of exchangeability and partial exchangeability (due to B. de Finetti, [6], [7]) are essential.

We recall the definitions:

The random variables of a sequence $(X_1, X_2 \ldots)$ are exchangeable if, for every n, the joint distribution of n r.v. of the sequence is always the same, whatever the n r.v. are and however they are permuted.

From a structural point of view an exchangeable process X_1, X_2 $\ldots$ can be intended as a sequence of r.v. equally distributed among which a "stochastic dependence due to uncertainty" exists. More precisely the X_i are independent conditionally on any of a given set (finite or not) of exhaustive and exclusive hypothesis. These hypotheses may concern, for instance, the values of a parameter (number or vector) on which the common distribution, of known functional form, of X_i depends. We shall restrict ourselves to this case. Therefore, we shall assume that, conditionally on each possible value θ of a parameter Θ, the X_i are independent with $F(x/\theta)$ as known distribution function. According to the bayesian approach, a probability distribution on Θ must be assigned.

If we denote this d.f. as $U(\theta)$, the distribution of the X_i is the mixture:

$$F(x) = \int F(x|\theta) dU(\theta)$$

If the observations relevant to n of the r.v. X_i (e.g. to the first n $X_1 = x_1, X_2 = x_2, \ldots X_n = x_n$) are available, the distribution on Θ is modified according to

$$u(\theta \,|\, x_1, x_2, \ldots, x_n) \propto L(x_1, x_2, \ldots, x_n \,|\, \theta)\, u(\theta)$$

https://doi.org/10.1017/S0515036100009168 Published online by Cambridge University Press

where $u(\theta)$ denotes the density of the distribution (supposing that it exists. The modification in a discrete case is obvious).

$L(x_1, x_2, \ldots, x_n|\theta)$ is the likelihood of θ, after the observation $\mathbf{x} = x_1, x_2, \ldots x_n$.

We have

$$L(x_1, x_2, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

if $f(x/\theta)$ is the density of the X distributions (in the discrete case the modification is obvious).

The d.f. of the X_i after the n observations is

$$F_n(x) = \int F(x|\theta)dU(\theta|x_1, x_2, \ldots, x_n) \tag{1}$$

In the bayesian formulation the exchangeability can be a substitute for the randomness concept of the sampling theory. According to this viewpoint, the partial exchangeability can be defined assuming that, also in this case, the X_i are independent conditionally on each hypothesis about the value of a parameter Θ, but now their distributions are different and depend, as well as on θ, on another observable entity φ_i, relevant to each X_i. If, for instance, the modalities of φ_i are only two, two subsequences are individualized

$$X_1^{(1)}, X_2^{(1)}, \ldots \text{ and } X_1^{(2)}, X_2^{(2)}, \ldots$$

In each of them the X_i are exchangeable and have as d.f.:

$$F^{(h)}(x) = \int F^{(h)}(x|\theta)dU(\theta), \quad h = 1, 2.$$

After n observations relevant to n variables

$$X_1^{(1)} = x_1^{(1)}, X_2^{(1)} = x_2^{(1)}, \ldots, X_{n_1}^{(1)} = x_{n_1}^{(1)};$$
$$X_1^{(2)} = x_1^{(2)}, \quad X_2^{(2)} = x_2^{(2)} \ldots, X_{n_2}^{(2)} = x_{n_2}^{(2)}$$

$(n_1 + n_2 = n)$, the distributions of the $X_i^{(h)}$ become

$$F_{n_1, n_2}^{(h)}(x) = \int F^{(h)}(x|\theta)dU_n(\theta), \quad h = 1, 2$$

where

$$dU_n(\theta) \propto \prod_{i=1}^{n_1} f^{(1)}(x_i^{(1)}|\theta) \prod_{j=1}^{n_2} f^{(2)}(x_j^{(2)}|\theta)dU(\theta)$$

Referring to one risk process, we shall now assume to collect—for each of the subsequent equally sized intervals of time (e.g. years)—

the observations relevant to the number of claims and the amount of each of them (and then the cumulated claim per year).

Thus, schematically, the observation in the period i supplies the data

$$n_i; x_1^{(i)}, x_2^{(i)}, \ldots, x_{n_i}^{(i)}; \quad S_i = \sum_{n=1}^{n_i} x_n^{(i)}; i = 1, 2, \ldots$$

Instead of limiting the observation to one risk, we could enlarge it to more risks which, although "similar", must be differentiated, from the very beginning, owing to some feature peculiar to each of them. That is, for instance, in the observation of the claim number process in motor insurance to consider cars with different H.P. as "similar" risks, or, in fire insurance, to consider commercial buildings with different kind of services as partially exchangeable risks.

Then, we can resort to the partial exchangeability in order to set up an inference procedure which allows us to specify our opinions about the behaviour of a risk also by means of observations on a similar but differentiated one. This is particularly useful when the observations on a single risk are few.

2. Inference for number of claims

a) Referring to a single risk, be it of interest to make inference for the number of claims per year.

The sampling variables are, in this case, the r.v. of the sequence $N_1, N_2, \ldots$ and we can consider them as the subsequent increments (relevant to equal periods of time) of a claim number process M_t (number of claims in 0, t) $N_i = M_i - M_{i-1}$, $i = 1, 2, \ldots$

In this approach, it is spontaneous to consider a "weighted Poisson process" where

$$P\{M_t = n\} = \int_0^\infty \frac{(\theta t)^n}{n!} e^{-\theta t} dU(\theta), \quad n = 0, 1, \ldots$$

This process, introduced in 1940 by O. Lundberg [10], under the name of "compound Poisson process" has been studied in terms of a process with exchangeable increments by H. Bühlmann [1] in 1960. Inference procedures for this process have already been treated by several Authors (besides O. Lundberg himself in the quoted book [10]). However we deem it interesting to expose some considerations in order to show how the above illustrated formulation can be applied to this fundamental model.

First of all, we can prove that one arrives to a weighted Poisson process from a Markov claim number process assuming that the interarrival times are exchangeable.

The question is to identify the intensities for occurence, $\lambda_n(t)$, $n = 0, 1, \ldots$, of the Markov process according to the required condition of exchangeability.

These intensities appear in the differential system

$$\frac{\partial p_{ii}(\tau,t)}{\partial t} = -\lambda_i(t) p_{ii}(\tau, t) \qquad i = 0, 1, \ldots$$

$$\frac{\partial p_{ij}(\tau,t)}{\partial t} = -\lambda_j(t) p_{ij}(\tau,t) + \lambda_{j-1}(t) p_{ij-1}(\tau,t) \; j > i = 0, 1, \ldots \qquad (2)$$

which, under the initial conditions $p_{ij}(\tau, \tau) = \delta_i^j$, gives the conditional probabilities of transition. Let $T_1, T_2 \ldots$ be the subsequent interarrival times, we have

$$F_{T_1}(t) = P\{T_1 \leq t\} = 1 - p_{00}(0, t) = 1 - e^{-\int_0^t \lambda_0(x)\,dx}.$$

We have also

$$P\{T_2 > t | T_1 = \tau\} = p_{11}(\tau, t + \tau) = e^{-\int_0^t \lambda_1(x+\tau)\,dx}.$$

So the density of the joint distribution of T_1, T_2, if $\lambda_0(t)$, $\lambda_1(t)$ continuous, is

$$f_{T_1,T_2}(t_1, t_2) = \lambda_0(t_1) e^{-\int_0^{t_1} \lambda_0(x)\,dx} \lambda_1(t_1 + t_2) e^{-\int_{t_1}^{t_1+t_2} \lambda_1(x)\,dx}$$

and an analogous formula can be written for the joint distribution of the first n inter-arrival times.

Let us suppose, now, that $\lambda_n(t) > 0$ when t is finite and for each integer n. If $\lambda_n(t) \equiv 0$ for $n > N$, only N arrivals are possible in the Markov process. Let us assume $\lambda_n(t)$ continuous for each n and $t \geq 0$ [1]). Then let us denote $l(t) = e^{-\int_0^t \lambda_0(x)\,dx}$ and $\varphi(t) = e^{-\int_0^t \lambda_1(x)\,dx}$.

[1]) These conditions imply that the joint densities of T_i are continuous and strictly positive for finite values of their arguments. For further considerations on the nature of the $\lambda_n(t)$ see L. Daboni [5], L. Crisma [4] and M. Strudthoff [12].

The functions $l(t)$ and $\varphi(t)$ are derivable and strictly positive for $t \geq 0$.

The condition of exchangeability

$$f_{T_1T_2}(t_1, t_2) = f_{T_2T_1}(t_1, t_2) \quad \forall\ t_1, t_2 \geq 0$$

implies that

$$l'(t_1)\,\frac{\varphi'(t_1+t_2)}{\varphi(t_1)} = l'(t_2)\,\frac{\varphi'(t_1+t_2)}{\varphi(t_2)} \quad \forall\ t_1, t_2 \geq 0$$

namely

$$\varphi(t) \propto l'(t).$$

So $l(t)$ is twice derivable. Now we have

$$\lambda_0 = -\frac{l'(t)}{l(t)} \quad \text{and} \quad \lambda_1(t) = -\frac{\varphi'(t)}{\varphi(t)}$$

and therefore

$$\lambda_1(t) = -\frac{l''(t)}{l'(t)}.$$

Reasoning in the same way upon the distribution of T_1, T_2, T_3, we obtain $\lambda_3(t) = -\dfrac{l'''(t)}{l''(t)}$. And so on for $\lambda_n(t)$, $n > 3$.

Necessarily the $l(t)$ must be completely monotone: $(-1)^i l^{(i)}(t) > 0$ and, generally, it follows that $\lambda_n(t) = -\dfrac{l^{(n+1)}(t)}{l^{(n)}(t)}$.

Now we have $l(0) = 1$ and therefore (Bernstein's theorem) $l(t)$ satisfies our conditions iff

$$l(t) = \int_0^\infty e^{-\theta t}\, dU(\theta)$$

with $U(\theta)$ as d.f. of a r.v. $\Theta \geq 0$.

We have, by now, established that the r.v. T_1, T_2, ..., inter-arrival times in a Markov claim number process are exchangeable iff

$$F_T(t) = 1 - l(t) = 1 - \int_0^\infty e^{-\theta t}\, dU(\theta).$$

In the hypothesis that θ^{-1} and θ^{-2} are integrable with respect to $U(\theta)$, we obtain easily

$$E(T) = \int_0^\infty [1 - F(t)]dt = \int_0^\infty \theta^{-1} dU(\theta) = E(\Theta^{-1}),$$

$$\operatorname{var}(T) = 2E(\Theta^{-2}) - \overline{E(\Theta^{-1})}^2,$$

$$\operatorname{cov}(T_i, T_j) = E(\Theta^{-2}) - \overline{E(\Theta^{-1})}^2, \tag{3}$$

where in the right-hand sides E is the expectation with respect to the distribution $U(\theta)$.

From the (2), by recurrent integrations we find

$$p_{ii+n}(\tau, t) = (-1)^n \frac{(t-\tau)^n}{n!} \frac{l^{(i+n)}(t)}{l^{(i)}(\tau)}, \quad \begin{matrix} i = 0, 1, \ldots \\ n = 0, 1, \ldots \end{matrix}$$

It is easy to verify that

$$P\{M_t - M_\tau = n\} = \sum_i p_{ii+n}(\tau, t)\, p_{0i}(0, \tau) =$$

$$= (-1)^n \frac{(t-\tau)^n}{n!} \sum_i (-1)^i \frac{\tau^i}{i!} l^{(i+n)}(t) =$$

$$= (-1)^n \frac{(t-\tau)^n}{n!} l^{(n)}(t-\tau) = \int_0^\infty \frac{(t-\tau)^n}{n!} \theta^n e^{-\theta(t-\tau)} dU(\theta).$$

Analogously it is proved that for $\tau < t \leq \tau + h$,

$$P\{M_t - M_\tau = n) \wedge (M_{t+h} - M_{\tau+h} = m)\} =$$

$$= (-1)^{n+m} \frac{h^{n+m}}{n!\, m!} l^{(n+m)}(2h), \qquad n, m = 0, 1, \ldots .$$

Thus, for a fixed h, the joint distribution of two increments relevant to equal and disjoint intervals is symmetric. And this is for as many increments as we want.

Then put $N_h(s) = M_{hs} - M_{/h-1/s}$, $h = 1, 2, \ldots, s > 0$, the r.v. of the sequence $N_1(s), N_2(s), N_3(s), \ldots$ are exchangeable with distributions

$$P\{N_h(s) = n\} = \frac{s^n}{n!} \int_0^\infty \theta^n e^{-\theta s} dU(\theta) \qquad h = 1, 2, \ldots$$

https://doi.org/10.1017/S0515036100009168 Published online by Cambridge University Press

It is easy to state

$$E(N_h(s)) = sE(\Theta),$$
$$\text{var}\,(N_h(s)) = s^2\text{var}\,(\Theta) + sE(\Theta),$$
$$\text{cov}\,(N_h(s), N_k(s)) = s^2\text{var}\,(\Theta). \tag{4}$$

Later on we will assume $s = 1$ and denote the number of arrivals in the h:th unitary interval of time (h:th year) by N_h.

In conclusion the r.v.

$$N_1, N_2, N_3, \ldots$$

and

$$T_1, T_2, T_3, \ldots,$$

whose distributions depend on the parameter θ, are exchangeable.

Our initial distribution for Θ is the $U(\theta)$. Now by means of the observations relevant to t r.v.

$$N_1 = n_1, N_2 = n_2, \ldots, N_t = n_t$$

we can specify our knowledge about Θ according to what stated sub 1.

Here we have

$$L(n_1, n_2, \ldots, n/|\theta) \propto \theta^{n_1+n_2+\ldots+n_t}\, e^{-t\theta} = \theta^n e^{-t\theta}$$

with
$$n = \sum_{i=1}^{t} n_i.$$

Hence

$$U_{t,n}(\theta) = U(\theta|n_1, n_2, \ldots, n_t) = \frac{\int_0^\theta x^n e^{-tx} dU(x)}{\int_0^\infty \theta^n e^{-t\theta} dU(\theta)}.$$

Then after the observations we have

$$E_{t,n}(\Theta) = \int_0^\infty \theta dU_{n,t}(\theta) = \frac{\int_0^\infty \theta^{n+1} e^{-t\theta}\, dU(\theta)}{\int_0^\infty \theta^n e^{-t\theta}\, dU(\theta)} = \lambda_n(t)$$

and

$$\text{var}_{t,n}(\Theta) = \lambda_n(t)[\lambda_{n+1}(t) - \lambda_n(t)].$$

This is a positive quantity unless $U_{t,n}(\theta)$ is concentrated in one point. So that for the increments N_{t+i}, $i = 1, 2, \ldots$ and by (4) we have

$$E_{t,n}(N_h) = \lambda_n(t),$$
$$\mathrm{var}_{t,n}(N_h) = \lambda_n(t)[\lambda_{n+1}(t) - \lambda_n(t) + 1],$$
$$\mathrm{cov}_{t,n}(N_h, N_k) = \lambda_n(t)[\lambda_{n+1}(t) - \lambda_n(t)] > 0.$$

If the observations concern the intervals $T_1 = t_1$, $T_2 = t_2$, ... $T_n = t_n$, we have $L(t_1, t_2, \ldots, t_n|\theta) = \theta^n e^{-\theta t}$, with $t = \sum_{i=1}^{n} t_i$.

Then, the conclusions are the same we have reached for the process $\{N_t\}$ except that t is now real, while before it was integer.

In both cases the "sufficient statistic" is the total number, n, of arrivals and the length, t, of the whole interval of observations.

After the n observations and taking into account the (3), we get

$$E_{t,n}(T_i) = E(T_i|t_1, t_2, \ldots, t_n) = \int_0^\infty \theta^{-1} dU_{t,n}(\theta) =$$
$$= \frac{\int_0^\infty \theta^{n-1} e^{-\theta t} dU(\theta)}{\int_0^\infty \theta^n e^{-\theta t} dU(\theta)} = \frac{1}{\lambda_{n-1}(t)}$$

and, if $n \geq 2$,

$$\mathrm{var}_{t,n}(T_i) = \frac{2\,\lambda_{n-1}(t) - \lambda_{n-2}(t)}{\lambda_{n-1}(t) \cdot \lambda_{n-2}(t)},$$
$$\mathrm{cov}_{t,n}(T_i, T_j) = \frac{\lambda_{n-1}(t) - \lambda_{n-2}(t)}{\lambda_{n-1}(t) \cdot \lambda_{n-2}(t)}, \qquad i, j = n + 1, n + 2, \ldots.$$

Finally we note that $\lambda_n(t)$ is a function of the sufficient statistic (n, t), whose functional form depends on the initial distribution $U(\theta)$. In the statistic, the integer n (number of arrivals in 0, t) does not decrease when t increases. Let us suppose, now, that for a fixed $U(\theta)$,

$$\lim_{t \to +\infty} \lambda_n(t; U) = \lambda_u$$

then both covariances of the r.v. $\{N_t\}$ and $\{T_i\}$ converge to 0 when $t \to +\infty$.

The stochastic dependence due to the incertainty fades out and we are, asymptotically, in conditions of independence. Then we can expect that the limit $\lambda_u = \lambda$ does not depend on the distribution $U(\theta)$ unless it is not constant in a neighbourhood of λ. In the quoted paper of O. Lundberg it is proved that if $n/t = \chi$, when t varies, then $\lim_{t \to +\infty} \lambda_n(t) = \chi$ (if $U(\theta)$ is not constant in a neighbourhood of χ).

After treating the inference problem in general conditions, let us try to detail the initial distribution $U(\theta)$.

It is usual to assume that it belongs to the "Gamma" class (conjugated to the likelihood function). It is known that in this case the weighted process is a Pólya process with $\lambda_n(t) = \dfrac{\nu + n}{\lambda + t}$ if

$$\mu(\theta) = U'(\theta) \propto \theta^{\nu - 1} e^{-\lambda\theta}.$$

In fact, however, the choice of such a $U(\theta)$ means that we already have some knowledge about Θ, coming for instance from past experience. With weak knowledge, we could choose $\mu(\theta) \propto \theta^{-1}$ (improper distribution) [2]). The $U_{t,n}(\theta)$ would be proper and of the Gamma type (and even erlangian) as follows from

$$u_{t,n}(\theta) \propto \theta^n e^{-\theta t}\, \theta^{-1} = \theta^{n-1} e^{-\theta t}$$

when $n = \sum_{i=1}^{t} n_i \geq 1$. Here n and t are integer. Considering a Gamma function with positive and real parameters ν, λ, this means that we make a choice in the larger class of the Gamma distribution (with $\nu/\lambda \backsimeq n/t$). As already recalled, under these conditions the claim number process is a Pólya process, where

$$P\{N_h = k\} = \frac{\lambda^\nu}{\Gamma(\nu) k!} \int_0^\infty \theta^{\nu + k - 1} e^{-(1+\lambda)\theta} d\theta, \quad k = 0, 1, \ldots$$

that is

$$P\{N_h = k\} = \frac{(\nu)_k}{k!} p^\nu (1 - p)^k, \quad k = 0, 1, \ldots \tag{5}$$

where $(\nu)_k = \nu(\nu + 1) \ldots (\nu + k - 1)$, $(\nu)_0 = 1$ and $p = \lambda/1 + \lambda$.

The (5) are negative binomial distributions.[2])

---

[2]) The adoption of such improper distributions, failing past experience, has been suggested by H. Jeffreys. See, on this subject, also D. V. Lindley [8], [9] and A. Zellner [14].

https://doi.org/10.1017/S0515036100009168 Published online by Cambridge University Press

b) Then we can approach the inference for the number of claims in the following terms. We have good grounds for considering the arrival process as Poisson's with a random intensity Θ. If the risk is observed from the beginning (i.e. from the starting moment of the possible observations), and then an improper distribution $u(\theta) \propto \theta^{-1}$ is assigned to the parameter Θ, after t years (and with n claims in $0, t$) we should obtain a negative binomial distribution of the subsequent increments. That is the (5) with $\nu = n$ and $p = t/1 + t$.

Starting from the beginning of the $(t + 1)$:th year and thinking that the past observation is not available, we could adopt the method previously illustrated (weighted Poisson process with $u(\theta) \propto \theta^{-1}$) or, otherwise, assume that the N_h are exchangeable with the distribution (5) whose parameters are now random and infer for them.

c) Apart from the previous consideration, we now intend to infer for the process of the exchangeable r.v. $N_1, N_2, \ldots$ with distributions

$$p_k = P\{N_i = k\} = \int P\{N_i = k|\xi, \theta\}\, dU(\xi, \theta), \quad k = 0, 1, \ldots$$

where

$$P\{N_i = k|\xi, \theta\} = \frac{(\xi)_k}{k!}\, \theta^{\xi}(1 - \theta)^k, \qquad 0 < \xi < +\infty, \qquad 0 < \theta < 1, \tag{5'}$$

and, as usual, $U(\xi, \theta)$ is the initial distribution, that is supposed to be provided with density $u(\xi, \theta)$, on the parameters Ξ and Θ. The N_i can, as previously, count the number of claims relevant to a given risk in the i:th year ($i = 1, 2 \ldots$). The model we are going to work out seems to be interesting especially when the observation for every period is slight (see § 3).

If in t subsequent periods, $n_1, n_2, \ldots, n_t$ arrivals respectively have been observed, the distribution of the N_i becomes

$$p_k|n_1, n_2, \ldots, n_t = \int_0^\infty d\xi \int_0^1 P\{N_i = k|\xi, \theta\}\, u(\xi, \theta|n_1, n_2, \ldots, n_t)\, d\theta$$

and, in particular, we have

$$E(N_i|n_1, n_2, \ldots, n_t) = \int_0^\infty d\xi \int_0^1 \xi\, \frac{1 - \theta}{\theta}\, u\,(\xi, \theta|n_1, n_2, \ldots, n_t)\, d\theta.$$

The basic problem now lies in choosing the initial distribution $U(\xi, \theta)$. It is to be noticed for this purpose that the likelihood of the observation is

$$L(n_1, n_2, \ldots, n_t|\xi, \theta) \propto (\xi)_{n_1} \cdot (\xi)_{n_2} \cdot \ldots \cdot (\xi)_{n_t} \cdot \theta^{t\xi} (1-\theta)^n$$

with $n = \sum_{i=1}^{t} n_i$. Then it is sensible for mathematical convenience to choose, for instance, a density of the type

$$u(\xi, \theta) \infty R(\xi)\theta^{\gamma\xi+\alpha-1}(1-\theta)^{\beta-1}, 0 < \xi < +\infty, 0 < \theta < 1, \quad (6)$$

where $R(\xi)$ is a suitable rational function of ξ and α, β, γ are real, non negative, numbers.

Note that, for a fixed ξ and when $\gamma\xi + \alpha - 1 > 0$, $\beta > 1$, the density $u(\xi, \theta)$ reaches its maximum value at the point

$$\theta_\xi = \frac{\gamma\xi + \alpha - 1}{\gamma\xi + \alpha + \beta - 2}$$

and this quantity converges to 1 if $\xi \to +\infty$.

If, in our opinion, the expectation

$$E(N|\xi, \theta) = \xi \frac{1-\theta}{\theta}$$

is near to a value c we could choose α, β, γ, $R(\xi)$ so that the density becomes concentrated in a neighbourhood of the curve $\xi(1-\theta) - c\theta = 0$ in the strip $0 < \theta < 1$, $\xi > 0$ of the (ξ, θ) plane.

We shall assume $u(\xi, \theta) \propto \xi^{-1} \theta^{\gamma\xi+\alpha-1} (1-\theta)^{\beta-1}$ with $\alpha > 0$, $\beta > 0$, $\gamma \geq 0$. We have

$$u(\xi, \theta|n_1, n_2, \ldots, n_t) \propto \xi^{-1}(\xi)_{n_1}(\xi)_{n_2} \ldots (\xi)_{n_t}.\theta^{(t+\gamma)\xi+\alpha-1}(1-\theta)^{n+\beta-1}.$$

Such distribution is proper (for $n \geq 1$). In fact $Q_{n-1}(\xi) = \xi^{-1} \prod_{i=1}^{t} (\xi)_{n_i}$ is a polinomial of the $(n-1)$ degree. On the other hand, we have

$$\int_0^1 \theta^{(t+\gamma)\xi+\alpha-1}(1-\theta)^{n+\beta-1}d\theta = B[(t+\gamma)\xi+\alpha, n+\beta].$$

Now we know that, if q remains fixed, we have

$$\lim_{p\to\infty} \frac{\Gamma(p+q)}{\Gamma(p)} p^{-q} = 1,$$

so that for $\xi > \xi_\varepsilon$

$$B[(t+\gamma)\xi+\alpha, n+\beta] < (1+\varepsilon)\frac{\Gamma(n+\beta)}{[(t+\gamma)\xi+\alpha]^{n+\beta}}.$$

Hence, there exists the integral

$$C = \int_0^\infty Q_{n-1}(\xi)\, B[(t+\gamma)\xi+\alpha, n+\beta]\, d\xi$$

and it is finite when $\alpha > 0$, $\beta > 0$ and $n \geq 1$.

Note that also the initial distribution is proper if in the (6) the function $R(\xi)$ is finite for every $\xi > 0$ and, for $\xi \to +\infty$, $R(\xi) = o(\xi^{-1})$.

For the distributions of the N_i, after the observations, we now have

$$\begin{aligned} p_k|n_1, n_2, \ldots, n_t &= P\{N_i = k|n_1, n_2, \ldots, n_t\} = \\ &= \int_0^\infty d\xi \int_0^1 \frac{(\xi)_k}{k!}\,\theta^\xi(1-\theta)^k\, u(\xi, \theta|n_1, n_2, \ldots, n_t)\, d\theta = \\ &= \frac{C^{-1}}{k!}\int_0^\infty (\xi)_k\, Q_{n-1}(\xi)\, B[(t+\gamma+1)\xi+\alpha, n+\beta+k]\, d\xi, \\ &\qquad k = 0, 1, \ldots;\ i = t+1, t+2, \ldots . \end{aligned}$$

Namely, taking into account the

$$B(p, q+k) = B(p, q)\frac{(q)_k}{(p+q)_k}, \quad k = 0, 1, 2, \ldots,$$

$$p_{k|n_1, n_2, \ldots, n_t} = \frac{(n+\beta)_k}{k!}\,\frac{\displaystyle\int_0^\infty \frac{(\xi)_k Q_{n-1}(\xi)}{[(t+\gamma+1)\xi+\alpha+n+\beta)]_k} B[(t+\gamma+1)\xi+\alpha, n+\beta]\, d\xi}{\displaystyle\int_0^\infty Q_{n-1}(\xi)\, B[(t+\gamma)\xi+\alpha, n+\beta]\, d\xi}.$$

If, in particular, β is integer, taking into account the

$$B(p, n+\beta) = \frac{\Gamma(n+\beta)}{(p)_{n+\beta}}$$

https://doi.org/10.1017/S0515036100009168 Published online by Cambridge University Press

and putting

$$\Pi_m(\xi;t) = \prod_{i=0}^{m-1} \left(\xi + \frac{\alpha+i}{t+\gamma}\right)$$

we have

$$p_{k|n_1,n_2,\ldots,n_t} = \frac{(n+\beta)_k}{k!}\left(\frac{1}{t+\gamma+1}\right)^k\left(\frac{t+\gamma}{t+\gamma+1}\right)^{n+\beta}\cdot\mu_k(\alpha,\beta,\gamma;n_1,n_2,\ldots,n_t) \tag{7}$$

with

$$\mu_k = \frac{\int_0^\infty \nu_k(\xi)\,\frac{Q_{n-1}(\xi)}{\Pi_{n+\beta}(\xi;t)}\,d\xi}{\int_0^\infty \frac{Q_{n-1}(\xi)}{\Pi_{n+\beta}(\xi;t)}\,d\xi}.$$

Here is

$$\nu_k(\xi) = \frac{(\xi)_k}{\prod_{i=0}^{k-1}\left(\xi + \frac{n+\beta+\alpha+i}{t+\gamma+1}\right)}\cdot\frac{\Pi_{n+\beta}(\xi;t)}{\Pi_{n+\beta}(\xi;t+1)}, \quad k = 1, 2, \ldots$$

and

$$\nu_0(\xi) = \frac{\Pi_{n+\beta}(\xi;t)}{\Pi_{n+\beta}(\xi;t+1)}.$$

Apart from perturbing factors μ_k, the (7) are the probabilities of a negative binomial distribution with parameters $n+\beta, \frac{t+\gamma}{t+\gamma+1}$. Note that the μ_k is the weighted average of the function $\nu_k(\xi)$ with the weighting density $Q_{n-1}(\xi)/\Pi_{n+\beta}(\xi;t)$.

For every ξ in $(0, +\infty)$ we have $\nu_0(\xi) < 1$ and $\nu_k(\xi) \gtreqless \nu_{k+1}(\xi)$ according to whether $k \lesseqgtr \frac{n+\alpha+\beta}{t+\gamma}$.

Therefore also the sequence μ_k at first increases (and $\mu_0 < 1$) and then decreases $\left(\text{if there exists } k > \frac{n+\alpha+\beta}{t+\gamma}\right)$.

These results (valid if β is integer and $\alpha > 0$), however, do not allow us to evaluate the expectation $E(N_i|n_1, n_2, \ldots, n_t)$ of the N_i after the observations.

It is easy to state that $E(N_i|\ldots.)$ exists if $\alpha \geq 1$ (and $\beta > 0$). Under these conditions we have

$$E(N_i|n_1, n_2, \ldots, n_t) = \int_0^\infty d\xi \int_0^1 \xi\theta^{-1}(1-\theta)\, u(\xi, \theta|n_1, n_2, \ldots, n_t)\, d\theta =$$

$$= \frac{\int_0^\infty \xi Q_{n-1}(\xi)\, B[(t+\gamma)\xi + \alpha - 1, n + \beta + 1] d\xi}{\int_0^\infty Q_{n-1}(\xi)\, B[(t+\gamma)\xi + \alpha, n + \beta]\, d\xi}$$

and since, if $p > 0$,

$$B(p, q+1) = \frac{q}{p} B(p+1, q)$$

results

$$E(N_i|n_1, n_2, \ldots, n_t) = \frac{n+\beta}{t+\gamma} \text{ if } \alpha = 1, \text{ and}$$

for $\alpha > 1$

$$E(N_i|n_1, n_2, \ldots, n_t) =$$

$$= \frac{\int_0^\infty \frac{(n+\beta)\xi}{(t+\gamma)\xi + \alpha - 1} Q_{n-1}(\xi)\, B[(t+\gamma)\xi + \alpha, n + \beta]\, d\xi}{\int_0^\infty Q_{n-1}(\xi)\, B[(t+\gamma)\xi + \alpha, n+\beta]\, d\xi}.$$

The ratio in the right side is the average of the function

$$\mu(\xi) = \frac{(n+\beta)\xi}{(t+\gamma)\xi + \alpha - 1}$$

with the density $Q_{n-1}(\xi)\, B[(t+\gamma)\xi + \alpha, n+\beta]$ and, in $0 < \xi < +\infty$, we have $0 < \mu(\xi) < \dfrac{n+\beta}{t+\gamma}$.

If the ratio n/t converges to a limit λ, when t (and n) diverge, then asymptotically we have

$$\mu(\xi) \sim \frac{n+\beta}{t+\gamma} \sim \lambda$$

for $\xi \neq 0$ and $E(N_i|n_1, n_2, \ldots, n_t) \sim \lambda$.

https://doi.org/10.1017/S0515036100009168 Published online by Cambridge University Press

Note that this conclusion and the one relevant to the whole distribution, are valid in the more general class of initial distribution (6) if in $R(\xi) = \frac{P_i(\xi)}{P_j(\xi)}$, the degree, i, of the polinomial at the numerator, $P_i(\xi)$, is smaller than j of the denominator. Instead of $Q_{n-1}(\xi)$, there is, in this case, the rational function $R(\xi)\ \Pi\ (\xi)n_i$, so that only the weighting densities above considered vary.

3. Inference for the number of claims: partial exchangeability

We propose the following model of partial exchangeability, whose calculations will not be given.

Let us observe the numbers of claims, per year, relevant to two "similar" but differentiated risks. Let us assume that for both, the distributions of the number of claims in a year are negative binomial, namely of the form (5′), but that the parameters ξ_i, θ_i are different for the two risks.

We are under conditions of partial exchangeability, if, for instance, the parameter ξ is random and common to both risks, while the parameters θ_1 and θ_2 are different and known.

Alternatively we can consider θ random and common to both risks and differentiate them by means of the known value of the parameters ξ_1 and ξ_2.

More generally, we could assume that the common parameter is unknown too; we will deal with this case afterwards.

a) Let then $N_1^{(1)}, N_2^{(1)}, \ldots.$
and $N_1^{(2)}, N_2^{(2)}, \ldots.$

be the r.v. which count the numbers of claims per year of the two risks and assume that

$$P\{N_i^{(1)} = k\} = \iiint \frac{(\xi)_k}{k!}\, \theta_1^{\xi}\, (1-\theta_1)^k\, u(\xi, \theta_1, \theta_2)\, d\xi\, d\theta_1\, d\theta_2$$

and

$$P\{N_i^{(2)} = k\} = \iiint \frac{(\xi)_k}{k!}\, \theta_2^{\xi}\, (1-\theta_2)^k\, u(\xi, \theta_1, \theta_2)\, d\xi\, d\theta,\, d\theta_2\,.$$

Here the $u(\xi, \theta_1, \theta_2)$ denotes the "initial" joint density of the r.v. Ξ, Θ_1, Θ_2. When ξ is certain the parameters Θ_1, Θ_2 must be stochastically dependent; in the opposite case we shall have, obviously, two different exchangeable processes. On the other hand, we could assume that the parameter Ξ is random and independent of the other two and then (according to our opinion and information) choose a density of the form

$$u(\xi, \theta_1, \theta_2) = v(\xi) \cdot u(\theta_1, \theta_2)$$

The $v(\xi)$ is definite (proper or improper) for $0 < \xi < +\infty$ and the $u(\theta_1, \theta_2)$ (proper or improper) on the square $0 < \theta_1 < 1$, $0 < \theta_2 < 1$.

If—in t subsequent years—$n_1, n_2, \ldots, n_t$ claims for the first risk and $\nu_1, \nu_2, \ldots, \nu_t$ for the second one respectively have been observed (on the whole $n + \nu = \Sigma n_i + \Sigma \nu_i$ claims) the likelihood for the hypothesis ξ, θ_1, θ_2 is the function $L(n_1, n_2, \ldots, n_t; \nu_1, \nu_2, \ldots, \nu_t | \xi, \theta_1, \theta_2) \propto (\xi)_{n_1} \cdot (\xi)_{n_2} \cdot \ldots \cdot (\xi)_{n_t} \cdot (\xi)_{\nu_1} (\xi)_{\nu_2} \ldots (\xi_{\nu_t}) \, \theta_1^{t\xi} \cdot \theta_2^{t\xi} (1 - \theta_1)^n (1 - \theta_2)^\nu = Q_{n+\nu}(\xi) \, (\theta_1 \theta_2)^{t\xi} (1 - \theta_1)^n (1 - \theta_2)^\nu$, where $Q_{n+\nu}(\xi)$ is a polinomial in ξ of the $(n + \nu)$ degree.

Also taking into account what said sub 2), a suitable choice of the $u(\xi, \theta_1, \theta_2)$ is the

$$u(\xi, \theta_1, \theta_2) \propto R \, \xi \, P(\theta_1, \theta_2)$$

where $P(\theta_1, \theta_2)$ is a polinomial with non-negative values in the square domain of θ_1, θ_2; when P is set, the necessary correlation between the two parameters must be taken into account.

Finally, in order to infer, for instance, for $N_i^{(1)}$ we need the marginal final distributions, whose the d.f. are

$$U(\xi, \theta_1 | n_1, n_2, \ldots, n_t; \nu_1, \nu_2, \ldots, \nu_t) = \int_0^\xi dx \int_0^{\theta_1} d\tau \int_0^1 u(x, \tau, \theta_2 | \ldots) \, d\theta_2.$$

When we are interested only in evaluating the expectations, we have

$$E(N_i^{(h)} | \ldots ; \ldots) = \int_0^\infty d\xi \int_0^1 \xi \, \theta_h^{-1} (1 - \theta_h) \, u(\xi, \theta_h | \ldots ; \ldots) \, d\theta_h, \; h = 1,2.$$

The other case can be dealt with formally in the same way.

4. Inference for cumulated claims

With reference to one risk, it is now of interest to infer as well as for the claim frequency also for the distribution of a single claim.

We now assume that the r.v. $X_1, X_2 \ldots$ measuring the subsequent claims are both independent conditionally on any hypothesis about the arrival process and exchangeable (that is, even here, equally distributed conditionally on every hypothesis about a parameter Θ on which their common density $f(x|\theta)$ depends).

If a Poisson mixture process (with random arrival intensity Λ) is chosen as arrival process and an initial distribution on Λ and Θ with joint density $u(\lambda, \theta)$ is assigned, then the distribution of cumulated claim in a year is the *mixture*, weighted with that density, of the classical distribution of a compound Poisson process. In such way, the cumulated claim per year $S_1, S_2, \ldots$, with $S_i = \sum_{h=1}^{N_i} X_h$ are exchangeable.

If the observations in a year are the $X_1 = x_1, X_2 = x_2, \ldots\ldots X_n = x_n$, the distribution of the cumulated claim is the mixture of the compound Poisson process according to the final density $u(\lambda, \theta|x_1, x_2, \ldots, x_n)$, product of the likelihood and of the initial density.

The likelihood of the hypotheses λ, θ for that observation is

$$L(x_1, x_2, \ldots, x_n|\lambda, \theta) \propto \lambda^n e^{-\lambda} \prod_{i=0}^{n} f(x_i|\theta)$$

and a sufficient statistic of finite dimension exists if $f(x|\theta)$ is chosen in the exponential family (Gamma, Pareto, etc.).

When choosing, then $u(\theta, \lambda)$ we can assume either the independence of the two parameters or, what seems more sensible, a stochastic dependence of them (which makes the claim distributions depend on the arrival intensity).

Also because of the imaginable difficulties in the calculation, we can generally restrict ourselves to infer for $E(S_i)$ and var (S_i) instead of the whole distribution.

Under our hypotheses, after the observations $\underline{x} = (x_1, x_2, \ldots, x_n)$ we have

$$E(S|\underline{x}) = \iint \lambda\, E(X|\theta)\, u(\lambda, \theta|\mathbf{x})\, d\lambda d\theta$$

and

$$\mathrm{var}\,(S|\underline{x}) = E\{\mathrm{var}\,(S|\lambda, \theta)|\underline{x}\} + \mathrm{var}\{E(S|\lambda, \theta)|\underline{x}\}$$

where var $(.|\underline{x})$ means the variance relevant to the final distribution $u(\lambda, \theta|\underline{x})$ and

$$E[\text{var}\ (S|\lambda, \theta)] = E\{\lambda[\text{var}\ (X|\lambda, \theta) + E(X|\lambda, \theta)^2]\}.$$

As a very simple example, that we only mention without performing the easy calculations, we could state $f(x|\theta) = \theta\, e^{-\theta x}$ and assume that the density of the initial distribution is the product of two Gamma densities relevant to λ and θ respectively (independence of Λ and Θ). The use of a Gamma bivariate distribution would be formally more complex but would allow us to introduce an a priori dependance between the two parameters.

Conclusive remarks

In conclusion, we would point out that, basically, our approach concerns the use of a bayesian adaptive process which, at least conceptually, seems to be worthy with respect to the theory of experience rating. And so, essentially, because it allows us to face different problems by means of a unitary approach.

There would be no difficulties with respect to the assumption of exchangeability which, after all, induces us to suppose that some increments of the risk processes are mixtures of independent r.v. and the mixture varies according to the information. The partial exchangeability concerns the possible heterogeneity of the risks individually observed.

The very difficulty arises when we choose, case by case, both the initial distribution and (in the parametric analysis we have treated) the conditional distribution of the r.v. we are interested in (number of claims or cumulated claim per period).

The greater the information based on the past experience is, the less difficult such choices are. On the other hand, when the observation increases, the influence of the choice of the initial distribution (generally) fade out. Finally, of course, the unavoidable difficulties in calculation must be taken into account.

References

[1] Bühlmann, H., *Optimale Prämienstufensysteme*, Mitteilung der Vereinigung schweizerischer Versicherungsmathematiker, Bd 64, 1964.
[2] Bühlmann, H., *Kollektive Risikotheorien*, ibidem, Bd 67, 1967.

[3] BUHLMANN, H., *Austauschbare stochatısche Varıabeln und ıhre Grenzwertsaetze*, 1960 (Unıv of Calıfornıa Press)
[4] CRISMA, L *Una proprıetà caratterıstıca deı processı dı Pólya*, 1970 (Studı e rıcerche, VII, Unıv Parma)
[5] DABONI, L , *Processı markoffıanı dı puro ıngresso con ıntervallı dı attesa scambıabılı*, 1969 (Studı e rıcerche, VI, Unıv Parma)
[6] DE FINETTI, B , *La prevısıon ses loıs logıques, ses sources subjectıves*, 1937 (An Inst H Poıncaré)
[7] DE FINETTI, B , *Teorıa delle probabılıtà*, 1970, Eınaudı, Torıno
[8] LINDLEY, D V., *Bayesıan statıstıcs A revıew* Publıshed by Sıam Phıladelphıa, 1971
[9] LINDLEY, D V , *Introductıon to Probabılıty and Statıstıcs from a Bayesıan Vıewpoınt* Pt I *Probabılıty* Pt 2 *Inference* 1965 Unıv Press Cambrıdge
[10] LUNDBΓRG, O , *On random processes and theır applıcatıon to sıckness and accıdent statıstıcs*, 1940 Unıv of Stockholm (2nd edıt Uppsala 1964)
[11] SEAL, H , *Stochastıc theory of a Rısk Busıness*, 1969 J Wıley
[12] STRUDTHOFF, M , *Un processo markoffıano dı puro ıngresso a parametro contınuo ottenuto come estensıone del processo dıscreto della frequenza dı successo nel modello ıpergeometrıco* 1973 Ist Mat Unıv Trıeste Vol V
[13] WEDLIN, A , *Osservazıonı su alcunı processı ınferenzıalı ın ıpotesı dı scambıabılıtà parzıale* Unıversıtà dı Trıeste, 1972
[14] ZELLNER, A , *An ıntroductıon to Bayesıan Inference ın Econometrıcs*, 1971, J Wıley N Y

https://doi.org/10.1017/S0515036100009168 Published online by Cambridge University Press