D. Kurzmitteilungen

HANS BÜHLMANN, Zurich, ALOIS GISLER, Winterthur,
and WILLIAM S. JEWELL, Berkeley

Addendum to "Excess Claims and Data Trimming in the Context of Credibility Rating Procedures", BASA, vol. 1, 1982

In our paper, section 2) – The Basic Model – it should be noted that there are two possible interpretations:

(a)	(b)
The random vector $\underline{X} = (X_1, X_2, \ldots, X_n)$ represents the *annual experience* of a given risk in the years 1, 2, ..., *n*.	The random vector $\underline{X} = (X_1, X_2, \ldots, X_n)$ represents the individual *claim amounts* of a given risk, where $n =$ number of observed claims for that risk.
$f_\theta(x)$ is the density of the annual experience,	$f_\theta(x)$ is the density of the individual claim amount,
$p_o(x/\theta)$ is the density of the ordinary annual experience,	$p_o(x/\theta)$ is the density of the ordinary individual claim amount,
$p_e(x)$ is the density of the excess annual experience,	$p_e(x)$ is the density of the excess individual claim amount,
$\mu(\theta) \cong$ pure annual risk premium for the risk characterized by θ.	$\mu(\theta) \cong$ expected individual claim amount for the risk characterized by θ.

Correspondingly the quantities appearing later on in the paper have the following interpretation:

$\mu_o(\theta) \cong$ pure ordinary annual risk premium.	$\mu_o(\theta) \cong$ expected ordinary individual claim amount.

$P[\underline{X}]$ is the experience rated *total* annual risk premium, i.e. the best estimator of $\mu(\theta)$.

$g(\underline{X})$ is the experience rated *ordinary* annual risk premium, i.e. the best estimator of $\mu_0(\theta)$.

$P[\underline{X}]$ is the experience rated expected *total* individual claim amount, i.e. the best estimator of $\mu(\theta)$.

$g(\underline{X})$ is the experience rated expected *ordinary* individual claim amount, i.e. the best estimator of $\mu_0(\theta)$.

In the verbal description of the model we have not been careful enough to stick consequently with one of these two interpretations.

We apologize.

HILARY L. SEAL, Apples

Mixed Poisson – an Ideal Distribution of Claim Numbers?

The recent paper in this Bulletin by A.-M. Gossiaux & J. Lemaire ("Bulletin" 1981, vol. 1) showing the excellent fit of one or more mixed Poisson distributions, namely

(1) Negative binomial,
(2) "Generalized" geometric (a constant times the negative binomial for $h = 1$ with an adjusted zero cell),
(3) Double Poisson,

to six observed sets of automobile claim frequencies published in the actuarial literature, may be supplemented by the following analysis of some data obtained from California.

J. Ferreira Jr. provides a six consecutive license year record for 7,842 California drivers during the period November 1959 to February 1968. The accidents were recorded in each year of driving license by the University of California's Institute for Transportation and Traffic Engineering (I.T.T.E.). The data are available from the author.

Given that n accidents have occurred in an interval of length t the distribution of these accidents over the interval is uniform and thus the probability distribution of the n occurred accidents is the multinomial, namely.

$$p(n_1, n_2, n_3, n_4, n_5, n_6) = \frac{n!}{\prod_{j=1}^{6} n_j!} \prod_{j=1}^{6} p_j^{n_j}$$

$$n_1 + n_2 + n_3 + n_4 + n_5 + n_6 = n$$

$$p_j = \frac{1}{6} \quad j = 1, 2, \ldots, 6$$

Counting the configurations from Ferreira's figures we get:

n	Six-year accident combination (0's suppressed)	Number of drivers	Probability of combination given n	Expected numbers of drivers	$(A-E)^2/E$	
0	–	5 147				
1	1	1 859				
2	1 1	485	30×6^{-2}	495.8	.24	
	2	110	6×6^{-2}	99.2	1.18	
		595		595.0	1.42	$\nu = 1$
3	1 1 1	83	120×6^{-3}	92.8	1.03	
	1 2	77	90×6^{-3}	69.6	.79	
	3	7	6×6^{-3}	4.6	1.25	
		167		167.0	3.07	$\nu = 2$
4	1 1 1 1	10	360×6^{-4}	15.0	1.67	
	1 1 2	32	720×6^{-4}	30.0	.13	
	1 3	7	120×6^{-4}	5.0	.80	
	2 2	5	90×6^{-4}	3.8 }	.25	
	4	–	6×6^{-4}	.2 }		
		54		54.0	2.85	$\nu = 3$
5	1 1 1 1 1	3	720×6^{-5}	1.3	2.22	
	1 1 1 2	6	$3{,}600 \times 6^{-5}$	6.5	.04	
	1 2 2	3	$1{,}800 \times 6^{-5}$	3.2	.01	
	1 1 3	1	$1{,}200 \times 6^{-5}$	2.2	.65	
	2 3	1	300×6^{-5}	.5 }	.05	
	2 other	–	156×6^{-5}	.3 }		
		14		14.0	2.97	$\nu = 4$
6	1 1 1 1 2	2	$10{,}800 \times 6^{-6}$	1.2	.53	
	1 1 2 2	1	$16{,}200 \times 6^{-6}$	1.7	.29	
	1 1 1 3	2	$7{,}200 \times 6^{-6}$	.8 }	.00	
	8 other	–	$12{,}456 \times 6^{-6}$	1.3 }		
		5		5.0	0.82	$\nu = 2$
11		1				
		7842				

The expected numbers of accidents given $n = 2, 3, 4, 5$ or 6, respectively, are calculated in the table and chisquare for "fit" is computed with ν, the degrees of freedom, one less than the number of cells used. All five n-values provide a good fit for the mixed Poisson hypothesis. A somewhat similar comparison for workers' accidents has been made by Hofmann ("Bulletin" 3/1955).

Prof. Hilary L. Seal
La Mottaz
1143 Apples

Bjørn Sundt, Oslo/Zurich[1]

Minimum Entropy in Risk Theory

In the last few years there have been some attempts to introduce the maximum entropy principle in risk theory, most recently in your journal (Maeder, 1982). It seems now to be high time to ask the question: Has this principle anything to do in risk theory? Apparently no one has given convincing arguments for an affirmative answer to this question so far, and based on the papers that have appeared till now, it seems that the answer should rather be no.

The model of Maeder (1982) is as follows: Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with distribution defined by

$$p_i = Pr(X_t = x_i) \quad (i = 0, 1, \ldots, n)$$

with $x_i = ih$; the parameter $h > 0$.

Let

$$U_t = \begin{cases} U & (t=0) \\ U_{t-1} + \Pi - X_t, & (t = 1, 2, \ldots) \end{cases}$$

where

$$\Pi = E(X_i) + \eta$$

with $\eta \geqq 0$. The ruin probability at time t is defined as

$$_{t-1|}q = Pr(U_t < 0 \mid \bigcap_{i<t} (U_i \geqq 0)).$$

Maeder wants to study the effect on these probabilities when varying the third moment of X_t when the first two moments are fixed. As is well known, a distribution is not at all determined by its first three moments, and to get a uniquely determined distribution, Maeder proposes the maximum entropy distribution, found by maximizing the entropy

$$H(p_1, \ldots, p_n) = - \sum_{i=1}^{n} p_i \log p_i \qquad (1)$$

[1] The present note was submitted for publication as a "Letter to the Editor".

under the side conditions that the first three moments of X_t are fixed. As an argument for maximizing this rather unusual quantity, Maeder refers to Berliner & Lev (1980); we shall return to that paper later.

By maximizing (1) one gets

$$p_i = e^{-\lambda_0 - \lambda_1 x_i - \lambda_2 x_i^2 - \lambda_3 x_i^3}. \quad (i = 0, 1, \ldots, n). \tag{2}$$

The parameters λ_k are determined by the side conditions. As an argument that the maximum entropy distribution (2) is a reasonable approximation to a given distribution of X_t, Maeder gives two examples. These examples seem very convincing. There are two reasons for this:

i) A distribution is relatively well determined by its first three moments.
ii) The given distributions lie relatively close to the maximum entropy distribution.

Without giving further arguments for using the maximum entropy principle, the author states in his conclusion that it opens new possibilities to actuarial investigations. Why should it open new possibilities, which possibilities, and how?

For the above mentioned two examples Maeder computes the ruin probabilities both with the given distribution and with the maximum entropy approximation. In the first example the discrepancy between the two sets of ruin probabilities is striking. This shows that ruin probabilities can vary much within a class of distributions with the first three moments fixed. Furthermore, the maximum entropy distribution cannot be considered as a particularly interesting member of this class.

We conclude that Maeder has done an interesting investigation on how the ruin probabilities with maximum entropy distributions with the first two moments fixed, depend on the third moment. But nothing more.

Now to the paper by Berliner & Lev (1980). As Maeder uncritically refers to that paper, it seems necessary to look a bit into it.

I have not the prerequisites for discussing the applicability of the maximum entropy principle in statistical mechanics and thermodynamics. However, these disciplines seem to have very little to do with risk theory. One should therefore not uncritically transfer a concept from these disciplines to risk theory.

The authors give three axioms from which they deduce the entropy principle. In particular axiom (c) (the composition law) does not seem to be clearly motivated within the framework of risk theory.

Let me just point out some of the inaccuracies in the paper.

1. In Section 5 the authors argue that the maximum entropy principle belongs to the subjective school of probability. They say, "This concept helps us to form plausible conclusions when we have only partial subjective knowledge." This partial knowledge is, apparently, the side conditions by the maximization of the entropy. But such arguing has nothing to do with subjective probability as described by e.g. De Groot (1970). Our subjective knowledge may always be expressed by a probability distribution (if we have vague knowledge, by an improper distribution), but it is not in the philosophy of the subjective probability theory to maximize some function to find the distribution of our subjective knowledge.

2. In Section 7 one studies the density $f(t)$ of the time of the first event to appear after time 0. One assumes that

$$\int_0^\infty tf(t)\,dt < \infty, \tag{3}$$

and from this one argues that

$$\int_0^\infty tf(t)\,dt = \bar{t} \tag{4}$$

with $\bar{t} < \infty$. Formula (4) is used as side condition with $\bar{t}$ as a given parameter. But this is a much stronger assumption than (3). In (3) it is said that the expectation of the first occurrence is finite, but now it is said that this expectation is equal to $\bar{t}$. It seems impossible to find a proper probability distribution maximizing the entropy among the distributions satisfying (3). One could of course argue that one is willing to use the stronger assumption (4), that the expected time for the first occurrence is given. But why then this expectation? Why not the expected number of occurrences in one time unit? Then one gets a completely different distribution for the first occurrence. It thus seems to be completely arbitrary to which distributions one gets by the maximum entropy principle.

3. In Section 9 the authors introduce a prior distribution for $\bar{t}$. It is argued that "since $\bar{t}$ is always finite, its expected value must be finite." However, there exist many distributions with infinite expectation. The most well-known example is perhaps the Pareto distribution with density

$$f(x;\alpha,\beta) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}} \qquad (x > \beta)$$

with $0 < \alpha \leq 1$ and $\beta > 0$.

References

Berliner, B. & Lev, B. (1980): On the use of the maximum entropy concept in insurance. *Transactions of the 21st International Congress of Actuaries 1*, 47–62.

De Groot, M.H. (1970): *Optimal statistical decisions.* McGraw-Hill, New York.

Maeder, P. (1982): Asymétrie de la distribution collective des sinistres et probabilités de ruine. *Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker*, 91–115.

Bjørn Sundt
Forsikringsrådet
Postboks 4301 Torshov
Oslo 4
Norwegen

1. In Section 5 the authors argue that the maximum entropy principle belongs to the subjective school of probability. They say, "This concept helps us to form plausible conclusions when we have only partial subjective knowledge." This partial knowledge is, apparently, the side conditions by the maximization of the entropy. But such arguing has nothing to do with subjective probability as described by e.g. De Groot (1970). Our subjective knowledge may always be expressed by a probability distribution (if we have vague knowledge, by an improper distribution), but it is not in the philosophy of the subjective probability theory to maximize some function to find the distribution of our subjective knowledge.

2. In Section 7 one studies the density $f(t)$ of the time of the first event to appear after time 0. One assumes that

$$\int_0^\infty tf(t)\,dt < \infty, \tag{3}$$

and from this one argues that

$$\int_0^\infty tf(t)\,dt = \bar{t} \tag{4}$$

with $\bar{t} < \infty$. Formula (4) is used as side condition with $\bar{t}$ as a given parameter. But this is a much stronger assumption than (3). In (3) it is said that the expectation of the first occurrence is finite, but now it is said that this expectation is equal to $\bar{t}$. It seems impossible to find a proper probability distribution maximizing the entropy among the distributions satisfying (3). One could of course argue that one is willing to use the stronger assumption (4), that the expected time for the first occurrence is given. But why then this expectation? Why not the expected number of occurrences in one time unit? Then one gets a completely different distribution for the first occurrence. It thus seems to be completely arbitrary to which distributions one gets by the maximum entropy principle.

3. In Section 9 the authors introduce a prior distribution for $\bar{t}$. It is argued that "since $\bar{t}$ is always finite, its expected value must be finite." However, there exist many distributions with infinite expectation. The most well-known example is perhaps the Pareto distribution with density

$$f(x;\alpha,\beta) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}} \qquad (x > \beta)$$

with $0 < \alpha \leqq 1$ and $\beta > 0$.

References

Berliner, B. & Lev, B. (1980): On the use of the maximum entropy concept in insurance. *Transactions of the 21st International Congress of Actuaries 1*, 47–62.
De Groot, M.H. (1970): *Optimal statistical decisions.* McGraw-Hill, New York.
Maeder, P. (1982): Asymétrie de la distribution collective des sinistres et probabilités de ruine. *Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker*, 91–115.

Bjørn Sundt
Forsikringsrådet
Postboks 4301 Torshov
Oslo 4
Norwegen

Philippe Maeder, Lausanne

Reply to Bjørn Sundt's Preceding Remarks

Part of the remarks expressed by Mr. Sundt about a recently published paper (Maeder, 1982) take their origin in a wrong interpretation of the example 1 shown there (Poisson/exponential case).

In the author's mind, this example should just demonstrate that the risk, as measured in continuous time (Seal's figures, 1972) or in discrete time (maximum entropy principle), has the same features. But of course by no means one can expect to get ruin probabilities of equivalent order of magnitude! If the comparison had been made between exact ruin probabilities in discrete time and approximate values found with the maximum entropy model, one would not have seen «striking differences» between the two sets, but rather a similarity as convincing as example 2 (about which Mr. Sundt does not say a word). This can be shown in considering just the probability of ruin for the first year, that is equal to:

$$|_1\, q = 1 - F(U + \Pi, 1)$$

and which can be computed rather easily.

These considerations suggest that the results of the above mentioned paper might be more general than what Mr. Sundt's note asserts.

Philippe Maeder
Avenue du Temple 19
1012 Lausanne

References

P. Maeder: «Asymétrie de la distribution collective des sinistres et probabilités de ruine.» Bull. Ass. Act. suisses, 1982, pp. 91–115.

H.L. Seal: «Numerical calculation of the probability of ruin in the Poisson/Exponential case.» Bull. Ass. Act. suisses, 1972, pp. 77–98.