THE MINCING MACHINE REVISITED

HANS BÜHLMANN
Zürich

1. PROLOGUE

For the 4th ASTIN colloquium in Trieste (1963) Robert E. Beard [1] had initiated the topic of extreme value theory and its application to actuarial problems. The impetus came from Gumbel's book [2], published in 1958, and the idea of applying extreme value theory ran about as follows:

Problem: *Given*: The size of the largest claim observed $= X_N$
The number of observed claims $= N$.
Find: The excess of loss premium for a portfolio of independent risks identical with those under observation.

The method

The idea was to use Gnedenko's limit distribution for the wide class of distribution functions with an unlimited tail and finite moments, the "exponential type" in Gumbel's terminology, and to calculate the excess of loss premium on the basis of this limit distribution.

The discussion in Trieste was very lively and culminated in Jan Jung's citation of Jan Jung [3]: "There is a natural law which states that you can never get more out of a mincing machine, than what you have put into it. That is, if the reinsurance people want actuarially sound premiums, they must get a decent information about the claim distribution". In more mathematical language: The brilliant idea of Bobbie Beard is unfortunately leading to a non-robust procedure. A deviation in the true (but unknown) underlying distribution may lead to completely different results.

In any case the mincing machine argument seemed so powerful that we have had little reconsideration of the above problem at subsequent ASTIN colloquia. It may therefore be worthwhile to have another look at the problem now, i.e. more than ten years after the Trieste colloquium.

2. The Problem of Herbert Robbins

In June 1967 Herbert Robbins visited our University and he gave us another problem which, however, reminded me of the Trieste discussion. Here is the problem:

Given: A language (unknown to us) consisting of N words W_i each of them appearing independently with probability p_i in a given text.

Suppose: That you note all words which you see while reading, and suppose that you have noted n words (n is arrived at by counting each word as often as you have seen it).

Estimate: The total probability of those words which you have not yet seen.

(You will understand that this problem reminded me of the Trieste problem how to find the excess of loss premium in excess of the observed largest claim).

The following is of course a wrong answer to the problem of Herbert Robbins: "Since we have not seen these words their relative frequency in the sample is zero. Using the relative frequency we therefore arrive at the estimate zero for the unknown probability". The error of reasoning consists in the fact that we have used the same observations to determine the event "set of those words which we have not seen" as well as the relative frequency of its occurrence. The same reasoning error can of course be found over and over again in excess of loss rate-making, that is why I have thought it worthwhile to produce a wrong answer first.

But here is now the right answer. The interesting fact is that the problem *can* be solved reasonably:

As stated the words W_i, $i = 1, 2, \ldots . N$ appear independently with probability p_i.

Define the random variables:

$$X_i = \begin{array}{l} 1 \text{ if } W_i \text{ has appeared \textit{exactly once} in the first } n+1 \text{ recordings} \\ 0 \text{ otherwise} \end{array}$$

$$X = \frac{\sum_{i=1}^{N} X_i}{n+1} = \text{frequency of words which have appeared exactly once in the first } n+1 \text{ recordings}$$

and analogously

$$Y_i = \begin{matrix} 1 & \text{if } W_i \text{ has } \textit{never} \text{ appeared in the first } n \text{ recordings} \\ 0 & \text{otherwise} \end{matrix}$$

$$Y = \sum_{i=1}^{N} p_i Y_i = \text{the unobserved probability}$$

X can be observed and is a reasonable estimate for Y since

i) $E[X - Y] = 0$
ii) $\operatorname{Var}[X - Y] \to 0$ as $n \to \infty$.

This is easily shown since

$$E[X] = \frac{1}{n+1} \sum_{i=1}^{N} E[X_i] = \frac{1}{n+1} \sum_{i=1}^{N} \binom{n+1}{1} p_i (1 - p_i)^n =$$

$$= \sum_{i=1}^{N} p_i (1 - p_i)^n$$

$$E[Y] = \sum_{i=1}^{N} p_i E[Y_i] = \sum_{i=1}^{N} p_i (1 - p_i)^n$$

and hence

$$E[X - Y] = 0$$

$$E[(X - Y)^2] \leq E[X^2] + E[Y^2] \leq E[X] + E[Y] =$$

$$2 \sum_{i=1}^{N} p_i (1 - p_i) \leq 2(1 - p_{min})^n$$

Apparently in this example we have been able to estimate something that we have not seen. Why should we then not rediscuss the question of the Trieste colloquium? On the other hand the example of Robbins suggests also that we distinguish between two problems:

a) the estimation of the *probability* of the occurrence of an excess claim
b) the estimation of the *size* of an excess claim.

I should like to show today that while we cannot solve b) without a reasonable information regarding the claim's distribution we can indeed solve a). Jan Jung in his 1963 paper has already given the hint: use of order statistics.

3. Estimating the Probability of the Occurrence of a Claim of a Given Risk

Let us consider the following problem:

Assume: $X_1, X_2, \ldots\ldots X_N, X$ independent and identically distributed with distribution function F

Problem: Assume $X_1, X_2, \ldots\ldots X_N$ to be observed and rank them as follows

$$X^{(1)} \leq X^{(2)} \leq \ldots \leq X^{(s)} \leq \ldots \leq X^{(r)} \leq \ldots \leq X^{(N)}$$

Find an estimate for $P[X^{(s)} \leq X \leq X^{(r)}]$.

It is clear that this problem includes in particular the problem of how to estimate $P[X \geq X^{(N)}]$ i.e. the problem of estimating the probability of a claim of unobserved high amount. If we suppose that F is continuous then $F(X)$ is uniformly distributed in $[0, 1]$ and since

$P[X^{(s)} \leq X \leq X^{(r)}] = P[F(X^{(s)}) \leq F(X) \leq F(X^{(r)})]$ we may

estimate the unknown probability from the uniform distribution

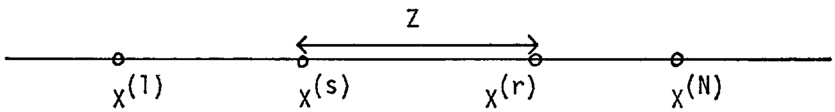

Fig. 1

Since in the uniform distribution the probability to lie between $X^{(s)}$ and $X^{(r)}$ is $Z = X^{(r)} - X^{(s)}$ this is the random variable we have to estimate.

Any text on order statistics e.g. [4] will explain to you that Z has a Beta distribution with density

$$f_Z(x) = \frac{N!}{(r-s-1)!\,(N+s-r)!}\, x^{r-s-1}\,(1-x)^{N+s-r}$$

$$= \frac{(\alpha+\beta+1)!}{\alpha!\,\beta!}\, x^{\alpha}(1-x)^{\beta} \qquad \begin{aligned} \alpha &= r - s - 1 \\ \beta &= N + s - r \end{aligned}$$

with mean

$$\frac{\alpha + 1}{\alpha + \beta + 2} = \frac{r - s}{N + 1} = \frac{\Delta}{N + 1} \qquad \text{where } \Delta = r - s$$

variance

$$\frac{(\alpha + 1)(\beta + 1)}{(\alpha + \beta + 2)^2 (\alpha + \beta + 3)} = \frac{(r - s)(N + s - r + 1)}{(N + 1)^2 (N + 2)} =$$

$$= \frac{\Delta(N - \Delta + 1)}{(N + 1)^2 (N + 2)}$$

For the special case $P[X \geq X^{(N)}]$ the random variable Z has the density

$$f_Z(x) = N(1 - x)^{N-1}$$

with mean

$$\frac{1}{N + 1}$$

variance

$$\frac{N}{(N + 1)^2 (N + 2)}$$

How then shall we estimate Z? There is a fair and there is a cautious way to proceed with estimation

fair estimation: Estimate Z by its expected value $\frac{\Delta}{N + 1}$

cautious estimation: Estimate Z by the $(1 - \varepsilon)$ confidence intervall of minimum length $[m_\varepsilon, M_\varepsilon]$.

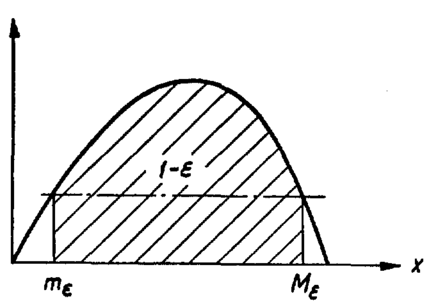

Fig. 2

The cautious estimate for $P[X \geq X^{(N)}]$ can be calculated particularly easy:

$$\int_0^{M_\varepsilon} fz(x)\,dx = -(1-x)^N \Big|_0^{M_\varepsilon} = 1-(1-M_\varepsilon)^N = 1-\varepsilon$$

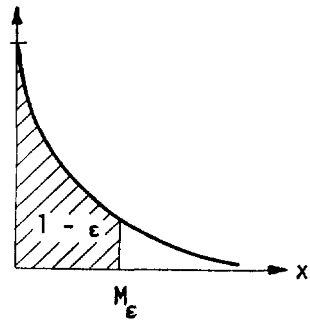

Fig. 3

Hence $(1-M_\varepsilon)^N = \varepsilon$

$$M_\varepsilon = 1 - \sqrt[N]{\varepsilon}$$

Here are some numerical examples for M_ε

N	$\varepsilon = 5\%$	$\varepsilon = 10\%$	$\varepsilon = 20\%$	$E(Z)$
2	0,78	0,68	0,55	0,333
4	0,53	0,44	0,33	0,200
8	0,31	0,25	0,18	0,111
16	0,17	0,13	0,10	0,059
32	0,09	0,07	0,05	0,030
64	0,05	0,04	0,02	0,015
128	0,02	0,02	0,01	0,008
256	0,01	0,01	0,006	0,004

It is quite obvious that the cautious and the fair estimates always will deviate considerably. Nevertheless it is also worthwhile noting that the deviation is not exceeding the ratio 1:3, whereas quotations in practice are quite often more apart.

4. Estimation in a Given Collective

The random variable Z measuring the probability between claim no. s and claim no. r has lead us to an estimate based upon observations from the individual risk. However we have learnt that estimates based upon observations from the individual risk should be combined with data from the whole collective in order to get practicable results for actuarial problems. Can we do this also in the context of our estimation problem?

Let $X_1, X_2, \dots X_N, X$ be distributed according to (continuous) $F_\theta(x)$ with unknown θ

where: i) given θ the random variables are independent and identically distributed

ii) θ is a random variable with distribution $U(\theta)$

iii) $\int F_\theta(y)\, dU(\theta) = F(y)$ is the distribution of the risk in the collective (supposed to be known from statistics about the collective).

Estimate: $P[X^{(s)} \leq X \leq X^{(r)}/X_1, X_2, \dots X_N; \theta]$

Remark: Observe the slightly more elaborate notation in comparison to that used in 3. The meaning is nevertheless the same.

The solution of this estimation problem in the collective is suggested by the fact that the quantity to be estimated can be interpreted in two manners:

First interpretation

$P[X^{(s)} \leq X \leq X^{(r)}/X_1, \dots X_N; \theta] = Z$ is a random variable with known distribution *independent* of θ

Second interpretation

$$P[X^{(s)} \leq X \leq X^{(r)}/X_1, \dots X_N; \theta] = F_\theta(x^{(r)}) - F_\theta(x^{(s)})$$

$x^{(r)}$ = observed value of $X^{(r)}$

$x^{(s)}$ = observed value of $X^{(s)}$

is a function of θ and hence can be interpreted as another random variable with a distribution over the collective.

Of course as we do neither know the values of Z nor of $F_\theta(x^{(r)}) - F_\theta(x^{(s)})$ we can estimate the probability (following the fair method as described in 3) by the expected value of either. This means that we have the two possible estimates:

$\hat{P}_1 = E[Z]$

$\hat{P}_2 = F(x^{(r)}) - F(x^{(s)})$ (distribution over the collective).

Inspired by the credibility techniques we shall try

$$\alpha E[Z] + (1 - \alpha)\,[F(x^{(r)}) - F(x^{(s)})]$$

where α is to be determinated by the least square method, i.e. minimizes

$$E[P[X^{(s)} \leqslant X \leqslant X^{(r)}/X_1, \ldots X_N;\theta] - \alpha E[Z] - \\ - (1 - \alpha)\,(F(x^{(r)}) - F(x^{(s)}))]^2$$

$$= E[\alpha(Z - E(Z)) + (1 - \alpha)\,(F_\theta(x^{(r)}) - F_\theta(x^{(s)}) - \\ - F(x^{(r)}) + F(x^{(s)}))]^2$$

$$= \alpha^2 \operatorname{Var}[Z] + (1 - \alpha)^2 \operatorname{Var}[F_\theta(x^{(r)}) - F_\theta(x^{(s)})].$$

From this finally

$$\alpha = \frac{\operatorname{Var}[F_\theta(x^{(r)}) - F_\theta(x^{(s)})]}{\operatorname{Var}[Z] + \operatorname{Var}[F_\theta(x^{(r)}) - F_\theta(x^{(s)})]}.$$

In the special case where we want to estimate

$$P[X \geq X^{(N)} / X_1, \ldots X_N; \theta],$$

we find

$$\alpha = \frac{\operatorname{Var} F_\theta(x^{(N)})}{\dfrac{N}{(N+1)^2(N+2)} + \operatorname{Var} F_\theta(x^{(N)})}.$$

Observe that all these estimates are different from those obtained by Bill Jewell in his paper on the credible distribution [5]. He estimates the distribution at a *given* value x whereas we have estimated the distribution at a value that *itself was defined on the observations*. Here again we have the same difference of definitions as in the problem suggested by Herbert Robbins. In practical applications one should therefore carefully discuss which of the two definitions applies.

I believe, as Jan Jung already suggested in his Trieste paper, that using order statistics' techniques we can reasonably estimate probabilities and hence frequencies of high excess layers. However estimates of sizes of excess claims will always be questionable in the sense that they depend too much on our choice of the underlying mathematical model. I therefore also believe that what I was telling you today are really the ultimate limits of the mincing machine.

5. Bibliography

[1] "4th ASTIN colloquium. Subjects for discussion", *The ASTIN Bulletin*, Vol. 2, pp. 313-314, (1963).
[2] GUMBEL, E. J., *Statistics of Extremes*, Columbia University Press, N.Y., (1958).
[3] JUNG, JAN, "On the Use of Extreme Values to Estimate the Premium for an Excess of Loss Reinsurance" (1964), *The ASTIN Bulletin*, Vol. 3, pp. 178-184.
[4] SIEGEL, S., *Nonparametric Statistics for Behavioural Sciences*, McGraw-Hill, N.Y., (1956).
[5] JEWELL, WILLIAM S., "The credible distribution", *The ASTIN Bulletin*, Vol. 7, pp. 237-269, (1974).