EXPERIENCE RATING IN SUBSETS OF RISKS

FRITZ BICHSEL
Berne

0. — PRELIMINARY

00. — *Summary*

We consider a set of risks B, which is divided into subsets B_i according to some property as profession, region etc. The subsets B_i are thought to be too small to permit a calculation of a premium from their own experience. On the other hand we assume that there are differences between the B_i. This paper tries to give a solution to the problem how the experience of the single B_i should be combined with that of the whole B.

The problem is treated only with respect to frequencies of claims. The amounts of the claims are not taken into consideration. I suppose that an extension of the theory to amounts is possible and highly desirable for practical applications. Lack of time did not permit me however to proceed to such an extension in this paper.

01. — *Practical application*

I think that the methods developed in the sequel might be applied to the following practical problems:

— rate making
— experience rating in the restrictive sense of the word, i.e. adjusting periodically the premium or granting premium refunds according to the experience of every single B_i
— judging the experience of a B_i.

02. — *References*

I do not give a list of references which would have to be very long. The basic idea was, as far as I know, first developed by Lundberg.

In order to give a complete and systematic account, this paper

Downloaded from https:/www.cambridge.org/core. University of Basel Library, on 11 Jul 2017 at 14:03:33, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/S0515036100009004

includes the derivation of some basic results which have been known for a long time.

What is new, are, as far as I know, the results of paragraphs 11, 211 and 221-225.

03. — *Accuracy*

It is a principle of this paper to take into consideration only the mean and variance of the occurring distributions, neglecting moments of higher order. Thus we shall, whenever this will prove convenient, replace any distribution by any other distribution having the same mean and standard deviation.

I think that this manner of proceeding is adequate to the nature of the problem, at least for a first approach.

04. — *Subjective probability*

The terminology used, especially in paragraph 1, is that of the so-called subjective probability theory. I think that the more extensive interpretation which the "subjectivists" give to the concept of probability, is more adequate to the problems treated in this paper than that of the "frequentists".

The proof of the pudding lies in the eating. I hope that the pieces offered in the sequel will be to the taste of some of the readers.

05. — *Notation*

050. — We shall always carefully distinguish between a random variable ξ, η etc. and the variables x, y etc. or n, m etc. occurring in the mathematical expression for its distribution or frequency functions.

Thus we shall write

$$P(\xi = n) = \pi(n; a)$$

(compare 060 for the meaning of $\pi(n; a)$)

for a random variable ξ following the simple Poisson-distribution.

The meaning of ξ, η etc. is related to the problem treated whereas $x, y, \ldots, n, m \ldots$ only have a purely mathematical meaning.

051. — In order to give the frequency function of a continuous random variable we shall write

$$P(\xi \approx x) = f(x).$$

Downloaded from https:/www.cambridge.org/core. University of Basel Library, on 11 Jul 2017 at 14:03:33, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/S0515036100009004

Thus if ξ follows a Γ-distribution with parameters a and b we shall write (compare 061)

$$P(\xi \approx x) = \gamma(x; a, b)$$

I apologize for introducing this new notation, but it has proved useful to treat the problem.

Note that $P(\xi \approx x)$ may be greater than 1.

052. — As usual we shall write

$$P(\xi = n \mid \eta = y)$$

to denote the conditional probability that the random variable ξ assumes the value n, granted that another random variable η has the value y.

I think that the notation

$$P(\xi \approx x \mid \eta = y)$$

derived from that of 051 needs no further explanation.

053. — According to the principle exposed in 03, we shall occasionally write

$$F(x; a, b)$$

to denote any function being the distribution function of a random variable with mean a and variance b.

06. — *Abbreviations*

We shall use the following abbreviations:

060 $$\pi(n; a) = \frac{a^n e^{-a}}{n!}$$

This is the frequency function of the (simple) Poisson-distribution.

061 $$\gamma(x; a, b) = \left(\frac{1}{b}\right)^{\frac{a}{b}} \frac{x^{\frac{a}{b}-1}}{\Gamma(a/b)} e^{-x/b}$$

This is the frequency function of the Γ-distribution.

062 $$\psi(n; a, b) = \left(\frac{1}{1+b}\right)^{a/b} \left(\frac{b}{1+b}\right)^{n} \frac{\Gamma\left(\frac{a}{b}+n\right)}{\Gamma(a/b)n!}$$

Downloaded from https:/www.cambridge.org/core. University of Basel Library, on 11 Jul 2017 at 14:03:33, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/S0515036100009004

This is the frequency function of a composed Poisson-distribution, having the distribution 061 as structural function (see 073).

07. — *Lemmas*

070. — If $P(\xi = n) = \pi(n; a)$,
then $\mu(\xi) = a$,
$\sigma^2(\xi) = a$.

071. — If $P(\xi \approx x) = \gamma(x; a, b)$,
then $\mu(\xi) = a$
$\sigma^2(\xi) = ab$.

(Note the difference to the notation of 053).

072. — If $P(\xi \approx x) = \gamma(x; a, b)$
then $P(c\,\xi \approx x) = \gamma(x; ca, cb)$.

073. — If $P(\xi = n | \eta = y) = \pi(n; y)$
and $P(\eta \approx y) = \gamma(y; a, b)$,

$$\text{then } P(\xi = n) = \int_0^\infty \pi(n; y)\,\gamma(y; a, b)\,dy$$
$$= \psi(n; a, b)$$

with $\mu(\xi) = a$
$\sigma^2(\xi) = a(1 + b)$

Note that $\mu(\xi)$ and $\sigma^2(\xi)$ also follow from 070 and 071 by an application of Lemma 075.

074. — $$\frac{\pi(n; x)\,\gamma(x; a, b)}{\psi(n; a, b)} = \gamma\left(x; \frac{a + nb}{1 + b}, \frac{b}{1 + b}\right)$$

075. — We use the notation explained in 053, writing F_1, F_2 etc., instead of F in order to distinguish different functions.

If $P(\xi \leqslant x \mid \eta = y \wedge \zeta = z) = F_1(x; y, z)$
and $P(\eta \leqslant y) = F_2(y; \mu_y, \sigma^2{}_y)$
$P(\zeta \leqslant z) = F_3[z; \sigma^2{}_x(y), \sigma^2{}_z(y)]$

$\sigma^2{}_x(y)$ and $\sigma^2{}_z(y)$ being allowed to depend on y,

then $\mu(\xi) = \mu_y$
$\sigma^2(\xi) = \bar{\sigma}^2{}_x + \sigma^2{}_y$

with $$\bar{\sigma}^2{}_x = \int_0^\infty \sigma^2{}_x(y)\,dF_2(y, \mu_y, \sigma^2{}_y)$$

Downloaded from https:/www.cambridge.org/core. University of Basel Library, on 11 Jul 2017 at 14:03:33, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/S0515036100009004

and thus

$$P(\xi \leqslant x) = F_4(x; \mu_y, \bar{\sigma}^2{}_x + \sigma^2{}_y)$$

Proof for $\sigma^2(\xi)$:

$$\sigma^2(\xi) = \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty (x - \mu_y)^2 \, dF_1(x; y, z) \, dF_2(y; \mu_y, \sigma^2{}_y) \, dF_3(z; \sigma^2{}_x, \sigma^2{}_z)$$

$$= \int_0^\infty \int_{-\infty}^\infty [z + (y - \mu_y)^2] \, dF_2(y; \mu_y, \sigma^2{}_y) \, dF_3(z; \sigma^2{}_x, \sigma^2{}_z)$$

$$= \int_{-\infty}^\infty \int_0^\infty z dF_3[z; \sigma^2{}_x(y), \sigma^2{}_z(y)] \, dF_2(y; \mu_y, \sigma^2{}_y)$$

$$+ \int_{-\infty}^\infty (y - \mu_y)^2 \, dF_2(y; \mu_y, \sigma^2{}_y) = \bar{\sigma}^2{}_x + \sigma^2{}_y$$

1. — Single Risks

In this paragraph we treat single risks or sets of risks which are considered as one single risk. We derive two theorems which are fundamental for the sequel. The first has been known for a long time, the second is, as far as I know, new.

In both theorems we assume that the risk under consideration has been observed during time unity and that k claims occurred. (Throughout the whole paper we take the period of observation as time unity. Every result may, by simple transformations, be extended to other time scales).

From this observation, we wish to determine the "true" claims frequency λ of the risk, which we do not know exactly, but of which we presume to have some a priori knowledge. We assume that this a priori knowledge can be expressed by saying that λ is a random variable following a certain distribution of which we know the mean and the standard deviation. In this paragraph we do not ask wherefrom we know this parameters. We simply consider them as given. According to the principle of 03, we shall assume that λ follows a Γ-distribution.

In the first theorem, λ is supposed to be constant though unknown. In the second theorem we assume that λ changes at random from one time interval to the other following the distribution:

$$P(\lambda \approx x \mid \bar{\lambda} = y) = \gamma(x; y, b_2)$$

of which we know the parameter b_2.

In both theorems we denote by S the number of claims in time unity, considered as a random variable.

Downloaded from https:/www.cambridge.org/core. University of Basel Library, on 11 Jul 2017 at 14:03:33, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/S0515036100009004

10. — *Theorem* 1

If
$$P(S = n \mid \lambda = x) = \pi(n; x),$$
$$P(\lambda \approx x) = \gamma(x; a, b)$$

then
$$E(\lambda \mid S = k) = \frac{a + bk}{1 + b}$$

$E(\lambda \mid S = k)$ denoting the estimate of λ after the observation of k claims.

Proof:

By Bayes' theorem we have
$$P(\lambda \approx x \mid S = k) = \frac{P(S = k \mid \lambda = x)\, P(\lambda \approx x)}{P(S = k)}$$

This gives, applying 073 to the denominator,
$$P(\lambda \approx x \mid S = k) = \frac{\pi(k; x)\, \gamma(x; a, b)}{\psi(k; a, b)}$$

from which we get by 074
$$P(\lambda \approx x \mid S = k) = \gamma\left(x; \frac{a + kb}{1 + b}, \frac{b}{1 + b}\right)$$

For the purposes of insurance we have to take the mean as the best estimate, so that we get
$$E(\lambda \mid S = k) = \frac{a + kb}{1 + b} \qquad \text{q.e.d.}$$

11. — *Theorem* 2

We denote by $\bar{\lambda}$ the mean value of λ for all time intervals, and we look for an estimate $E(\bar{\lambda} \mid S = k)$ on the basis of the observation of one interval.

If
$$P(S = n \mid \lambda = x) = \pi(n; x),$$
$$P(\lambda \approx x \mid \bar{\lambda} = y) = \gamma(x; y, b_2),$$
$$P(\bar{\lambda} \approx y) = \gamma(y; a, b_1).$$

Downloaded from https:/www.cambridge.org/core. University of Basel Library, on 11 Jul 2017 at 14:03:33, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/S0515036100009004

then $$E\ (\bar{\lambda} \mid S = k) = \frac{a + c\, b_1\, k}{1 + cb_1}$$

with $$c = \frac{1}{1 + b_2}$$

Proof:

By 073 we have

112 $$P\ (S = n \mid \bar{\lambda} = x) = \psi\ (n;\ x,\ b_2)$$

with $$\mu\ (S \mid \bar{\lambda} = x) = x,$$
$$\sigma^2\ (S \mid \bar{\lambda} = x) = x\ (1 + b_2).$$

We replace this by another distribution with the same mean and variance, namely by the distribution defined by

113 $$P\ (c\ S = n \mid \bar{\lambda} = x) = \pi\ (n,\ cx)$$

where $c = \dfrac{1}{1 + b_2}$, as already defined above.

Here we make the fiction that S assumes only values such that $c\ S$ is an integer. This fiction is allowed because of the principle of 03.

We have by 070

$$\mu\ (c\ S \mid \bar{\lambda} = x) = cx$$
$$\sigma^2\ (c\ S \mid \bar{\lambda} = x) = cx.$$

from which we get

$$\mu\ (S \mid \bar{\lambda} = x) = x$$
$$\sigma^2\ (S \mid \bar{\lambda} = x) = x/c = x\ (1 + b_2)$$

so that the distribution defined by 113 really has the same mean and variance as that given by 112.

Now we have

$$P\ (c\ S = n \mid \bar{\lambda} = x) = \pi\ (n;\ cx)$$

For $y = cx$, this gives

114 $$P\ (c\ S = n \mid c\ \bar{\lambda} = y) = \pi\ (n,\ y)$$

Downloaded from https:/www.cambridge.org/core. University of Basel Library, on 11 Jul 2017 at 14:03:33, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/S0515036100009004

On the other hand, by 072 we get

115 $$P\,(c\,\bar{\lambda} \approx y) = \gamma\,(y,\, ca,\, cb_1).$$

Now we apply theorem 1 to 114 and 115, which gives

$$E\,(c\,\bar{\lambda} \mid c\,S = c\,k) = \frac{ca + c^2\,b_1\,k}{1 + cb_1}$$

from which follows

$$E\,(\bar{\lambda} \mid S = k) = \frac{a + cb_1 k}{1 + cb_1}, \qquad \text{q.e.d.}$$

2. — Sets of risks

21. — *Set B consisting of single risks B_i*

We suppose that the true claims frequencies λ_i of the risks B_i follow a certain distribution, and we estimate the a posteriori value of a λ_i after the observation of the corresponding B_i during a time interval of length unity. As mentioned before, the result may easily be extended to intervals of any length.

The result of this paragraph may for instance be applied to merit rating in automobile insurance.

Notation:

r = number of risks in B
k_i = number of claims in B_i observed
$k = k_1 + \ldots + k_r$
S_i = number of claims in B_i, considered as a random variable.

210. — *No change of the λ_i in time*

We suppose that the λ_i are, though unknown, constant in time. According to the principle of 03, we presume that the a priori-distribution of the λ_i is a Γ-distribution.

The application of theorem 1 gives directly

$$E\,(\lambda_i \mid S_i = k_i) = \frac{a + b\,k_i}{1 + b}$$

Downloaded from https:/www.cambridge.org/core. University of Basel Library, on 11 Jul 2017 at 14:03:33, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/S0515036100009004

Estimation of the parameters a and b:

We have

$$P(S_i = k_i) = \psi(k_i; a, b)$$
$$\mu(S_i) = a$$
$$\sigma^2(S_i) = a(1 + b)$$

From all the observed k_i, we estimate $\mu(S_i)$ and $\sigma^2(S_i)$ as follows:

$$\mu(S_i) = k/r$$
$$\sigma^2(S_i) = \frac{1}{r-1} \sum \left(k_i - \frac{k}{r}\right)^2$$

This gives

$$a = k/r$$
$$b = \frac{\Sigma (k_i - a)^2}{(r-1)\, a} - 1$$

211. — *The λ_i change at random in time*

We suppose that the λ_i change from one time interval to the other following the distribution given by

$$P(\lambda_i \approx x \mid \bar{\lambda}_i = y) = \gamma(x; y, b_2) \text{ *}$$

b_2 being the same for all i. We wish to estimate $\bar{\lambda}_i$, the mean of the λ_i over all time intervals.

The application of theorem 2 gives directly

2110 $$E(\bar{\lambda}_i \mid S_i = k_i) = \frac{a + c\, b_1\, k_i}{1 + c\, b_1}$$

with $c = \dfrac{1}{1 + b_2}$

Estimation of the parameters:

According to the conditions of theorem 2, we have

$$P(S_i = n \mid \lambda_i = x) = \pi(n; x)$$
$$P(\lambda_i \approx x \mid \bar{\lambda}_i = y) = \gamma(x; y, b_2)$$
$$P(\bar{\lambda}_i \approx y) = \gamma(y; a, b_1)$$

*) These distributions are supposed to be mutually independent.

Downloaded from https:/www.cambridge.org/core. University of Basel Library, on 11 Jul 2017 at 14:03:33, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/S0515036100009004

By a repeated application of 075, we get from this

$$P\,(S_i \leqslant n) = F\,(n;\, a,\, a + ab_1 + ab_2)$$

If we put $$b_1 + b_2 = b,$$

we may determine a and b in the same way as in 210.

In order to estimate b_2, we have, of course, to use the observations of two different time intervals.

If we denote by ${}_1S_i$ the number of claims in a first and by ${}_2S_i$ that number in a second interval, the difference

$$({}_1S_i \mid \bar{\lambda}_i = y_i) - ({}_2S_i \mid \bar{\lambda}_i = y_i)$$

is, by 073, a random variable with mean 0 and variance

$$2\, y_i\, (1 + b_2)$$

Thus we have

$$P\left[\frac{({}_1S_i \mid \bar{\lambda}_i = y_i) - ({}_2S_i \mid \bar{\lambda}_i = y_i)}{\sqrt{y_i}} \leqslant x\right] = F\,[x;\, 0,\, 2(1 + b_2)]$$

Using all the observed values ${}_1k_i$ and ${}_2k_i$ and summing over i, we get from this

$$\sum \frac{({}_1k_i - {}_2k_i)^2}{y_i} = 2r\,(1 + b_2)$$

or

2111 $$b_2 = \frac{1}{2r} \sum \left[\frac{({}_1k_i - {}_2k_i)^2}{y_i}\right] - 1$$

Now we do not know the y_i. To overcome this difficulty, we may use $a = k/r$ as a first approximation for the y_i and determine a first approximation of b_2 from 2111. By 2110 we get now a second approximation of the $y_i = \bar{\lambda}_i$, from which we get a second approximation of b_2, etc.

22. — *Set B consisting of subsets B_i*

We are coming now to the main problem of this paper.

The results of this paragraph are subject to two very important conditions:

Downloaded from https:/www.cambridge.org/core. University of Basel Library, on 11 Jul 2017 at 14:03:33, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/S0515036100009004

A. We presume that we have no a priori conjectures about the values of a particular λ_i. If for instance the B_i correspond to different regions in automobile insurance and if we know that in one particular B_i, circulation is much denser than in other regions, we must not apply the following formulae to that B_i.

B. If the B_i follow a certain order and if the observed values of the claims frequencies of the B_i follow an order which shows a connection with the order of the B_i, the following results must not be applied. Thus the theory cannot be used for the calculation of premium rates for different classes of horse power in automobile insurance.

These two conditions are in accordance with the subjective interpretation of the concept of probability.

For practical purpose they may be neglected for experience rating in the restrictive sense of the word as mentioned in 01.

Of course the two conditions also apply to the results of paragraph 21. If they have not been mentioned already there, it was because the results of that paragraph will, in practice, only be used for experience rating in the restrictive sense.

Notation:

B = set of risks
B_i = subsets of B
B_{ij} = single risk belonging to B_i
S, S_i and S_{ij} = numbers of claims in time unity in B, B_i and B_{ij}, conceived as random variables.
k, k_i and k_{ij} = numbers of claims observed in time unity
λ, λ_i and λ_{ij} = claims frequencies per risk.
r_i = number of risks in B_i
s = number of subsets B_i
$r = r_1 + \ldots + r_s$

220. — λ_{ij} *constant in time*

We suppose that the λ_i follow a certain distribution. According to the principle of 03, we presume that this is a Γ-distribution:

$$P\,(\lambda_i \approx x) = \gamma\,(x;\, a,\, b)$$

The parameters a and b are considered as unknown. Their value will be estimated from the observed k_i.

Downloaded from https:/www.cambridge.org/core. University of Basel Library, on 11 Jul 2017 at 14:03:33, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/S0515036100009004

By 072 we have

$$P\ (r_i\,\lambda_i \approx x) = \gamma\ (x;\, r_i\, a,\, r_i\, b)$$

Now we apply theorem 1 with the following substitutions:

$$\begin{aligned} S &\rightarrow S_i \\ k &\rightarrow k_i \\ \lambda &\rightarrow r_i\,\lambda_i \\ a &\rightarrow r_i\, a \\ b &\rightarrow r_i\, b \end{aligned}$$

This gives

$$E\ (r_i\,\lambda_i \mid S_i = k_i) = \frac{r_i a + r_i b k_i}{1 + r_i b}$$

or

$$\underline{\underline{E\ (\lambda_i \mid S_i = k_i) = \frac{a + b k_i}{1 + r_i b}}}$$

Estimation of parameters:

We have

$$\begin{aligned} &P\ (S_i = n) = \psi\ (n;\, r_i a,\, r_i b) \\ &\mu\ (S_i) = r_i a \\ &\sigma^2\ (S_i) = r_i a\ (1 + r_i b) \end{aligned}$$

Substituting for the S_i the observed k_i and summing over all i, we get

$$\begin{aligned} &k = ra \\ &\Sigma\ (k_i - r_i a)^2 = \Sigma\ r_i a\ (1 + r_i b) \end{aligned}$$

From this we obtain

$$\begin{aligned} a &= k/r \\ b &= \frac{\Sigma\ (k_i - r_i a)^2 - k}{a\ \Sigma\ r_i^2} \end{aligned}$$

(We neglect the fact that the a of the second relation is calculated from the first relation).

221. — λ_{ij} *constant in time, with complete exchange of risks*

In this paragraph we suppose that every B_i is replaced, from one

Downloaded from https:/www.cambridge.org/core. University of Basel Library, on 11 Jul 2017 at 14:03:33, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/S0515036100009004

time interval to the other, by a new set B_i^* containing the same number of risks with the same characteristic i.

After observation of one B_i, we do not wish to determine the λ_i of this particular B_i, but the average claims frequency of sets B_i with the property i, which we denote by Λ_i and which is supposed to follow the distribution

$$P\ (\Lambda_i \approx y) = \gamma\ (y;\ a,\ b_1)$$

The exchange of the risks from one time interval to the other has the effect of a random fluctuation of the λ_i, so that we can apply theorem 2.

We perform the following substitution in that theorem

$$\begin{aligned} S &\rightarrow S_i \\ k &\rightarrow k_i \\ \lambda &\rightarrow r_i\,\lambda_i \\ \bar{\lambda} &\rightarrow r_i\,\Lambda_i \\ b_2 &\rightarrow r_i b_{2,i} \\ a &\rightarrow r_i a \\ b_1 &\rightarrow r_i b_1 \end{aligned}$$

so that it reads:

If

$$\begin{aligned} P\ (r_i\,\lambda_i \approx x \mid r_i\,\Lambda_i \quad &= y) = \gamma\ (x;\ y,\ r_i b_{2,i}) \\ P\ (r_i\,\Lambda_i \approx \quad y) &= \gamma\ (y;\ r_i a,\ r_i b_1) \\ P\ (S_i = n \mid r_i\,\lambda_i \quad &= x) = \pi\ (n;\ x) \end{aligned}$$

then

$$E\ \ (r_i\ \Lambda_i \mid S_i = k_i) = \frac{r_i a + c_i r_i b_1 k_i}{1 + c_i r_i\ b_1}$$

From this we get

$$\underline{\underline{E\ (\Lambda_i \mid S_i = k_i) = \frac{a + c_i\ b_1\ k_i}{1 + c_i\ r_i\ b_1}}} \qquad \text{with } c_i = \frac{1}{1 + r_i b_{2,i}}$$

Estimation of the parameters:

We put

2210 $$P\ (\lambda_{ij} \approx x \mid \lambda_i = y) = \gamma\ (x;\ y,\ d_i)$$

Downloaded from https:/www.cambridge.org/core. University of Basel Library, on 11 Jul 2017 at 14:03:33, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/S0515036100009004

The d_i can be calculated as in 210:

2211 $$d_i = \frac{\sum_j (k_{ij} - k_i/r_i)^2}{(r_i - 1)\, k_i/r_i} - 1$$

Now we have approximately

2212 $$P(\lambda_{ij} \approx x \mid \Lambda_i = y) = P(\lambda_{ij} \approx x \mid \lambda_i = y) = \gamma(x;\, y,\, d_i)$$

and thus

$$P(\lambda_i \approx x \mid \Lambda_i = y) = \gamma(x;\, y,\, d_i/r_i)$$

and

$$P(r_i \lambda_i \approx x \mid r_i \Lambda_i = y) = \gamma(x;\, y,\, d_i)$$

On the other hand, we have

$$P(r_i \lambda_i \approx x \mid r_i \Lambda_i = y) = \gamma(x;\, y,\, r_i b_{2,i})$$

From the last two relations follows

$$b_{2,i} = d_i/r_i$$

so that we get

2213 $$b_{2,i} = \frac{\sum_j (k_{ij} - k_i/r_i)^2}{(r_i - 1)k_i} - \frac{1}{r_i}$$

In order to determine a and b_1, we use the relation

$$P(S_i \leqslant n) = F(n;\, r_i a,\, r_i a + r^2_i\, ab_1 + r_i^2\, ab_{2,i})$$

following from 075.

Substituting for the S_i the observed k_i and summing over all i, we get

$$a = k/r$$

$$b_1 = \frac{\sum (k_i - r_i a)^2 - a \sum r_i^2\, b_{2,i} - k}{a \sum r_i^2}$$

222. — λ_{ij} *constant in time, with partial exchange of risks*

We suppose that we have observed B_i during a time interval 1, and we wish to estimate the claims frequency in a subsequent

Downloaded from https:/www.cambridge.org/core. University of Basel Library, on 11 Jul 2017 at 14:03:33, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/S0515036100009004

interval 2, of a set B_i^* which is formed of a part of B_i and new risks with the same characteristic i.

This case can evidently be treated as a combination of 220 and 221.

223. — λ_{ij} *variable in time, no exchange of risks*

We denote by $\bar{\lambda}_{ij}$ and $\bar{\lambda}_i$ the means of the λ_{ij} and λ_i over all time intervals and put

2231 $$P\,(\lambda_{ij} \approx x \mid \bar{\lambda}_{ij} = y) = \gamma\,(x;\, y,\, b_{3,i})$$

$b_{3,i}$ being the same for all j, and supposing that these distributions are mutually independent for all λ_{ij}.

Substitutions similar to those of 221 give the following result for the estimation of $\bar{\lambda}_i$:

$$\underline{\underline{E\,(\bar{\lambda}_i \mid S_i = k_i) = \frac{a + c_i\, b_1\, k_i}{1 + c_i\, r_i\, b_1}}}$$

$$\text{with } c_i = \frac{1}{1 + r_i b_{2,i}}$$

Estimation of the parameters:

First we determine the $b_{3,\,i}$ of 2231 according to the method of 211. On the other hand, summing 2231 over all j gives

$$P\,(r_i\,\lambda_i \approx x \mid \bar{\lambda}_{i,1} = y_1\,\Lambda \ldots \Lambda \bar{\lambda}_{ir_i} = y_{r_i})$$
$$= \gamma\,(x;\, y_1 + \ldots + y_{r_i},\, b_{3,i})$$

As the distribution on the right side depends only on the sum of the y_j and not on their individual values, the condition on the left side can be replaced by

$$r_i\,\bar{\lambda}_i = y_1 + \ldots + y_{r_i}$$

Putting $$y_1 + \ldots + y_{r_i} = y,$$

this gives

$$P\,(r_i\,\lambda_i \approx x \mid r_i\,\bar{\lambda}_i = y) = \gamma\,(x;\, y,\, b_{3,i})$$

Comparing this with the conditions for the application of theorem 2, we get

$$b_{2,i} = b_{3,i}/r_i$$

The estimation of a and b now is exactly the same as in 221.

Downloaded from https:/www.cambridge.org/core. University of Basel Library, on 11 Jul 2017 at 14:03:33, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/S0515036100009004

224. — λ_{ij} *variable in time, complete exchange of risks*

This is the general case.

Notation for claims frequencies:

λ_{ij} risk observed, time interval observed
$\bar{\lambda}_{ij}$ risk observed, mean over all intervals
λ_i set observed, interval observed
$\bar{\lambda}_i$ set observed, mean over all intervals
Λ_i arbitrary set of risks with characteristic i, arbitrary time interval

This case is perfectly analagous to that of 221. We only have to replace λ_i by $\bar{\lambda}_i$ in 2210 and 2212. Consequently, we can use the formulae of 221 for the estimation of the parameters a, b_1 and $b_{2,i}$.

This may seem paradoxical at first sight. The explanation lays in the fact the the k_{ij} in 2211 and 2213, if there is a variation of the λ_{ij} in time, are influenced both by that variation and the variation of the λ_{ij} about the λ_i.

As the case of the present paragraph and that of 221 can be treated exactly in the same way, it is not necessary — in the case of complete exchange of risks—to know whether there exists a variation in time of the λ_{ij} or not.

225. — λ_{ij} *variable in time, partial exchange of risks*

This is a combination of 223 and 224.

Remarks added before printing

A) Jan Jung told me that, instead of applying the principle of accuracy of 03, I could have obtained my results by linear regression.

Thus, my theorem 2 can be derived as a linear regression of $\bar{\lambda}$ on S as follows:

$$\bar{\lambda}^* = E(\bar{\lambda}) + \frac{\operatorname{cov}(\bar{\lambda}, S)}{\operatorname{var} S}(k - E(S))$$

$$\begin{aligned}\operatorname{cov}(\bar{\lambda}, S) &= E[(\bar{\lambda} - a)(S - a)] \\ &= E[(\bar{\lambda} - a)(S - \bar{\lambda} + \bar{\lambda} - a)] \\ &= E[(\bar{\lambda} - a)(S - \bar{\lambda})] + E(\bar{\lambda} - a)^2\end{aligned}$$

Downloaded from https:/www.cambridge.org/core. University of Basel Library, on 11 Jul 2017 at 14:03:33, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/S0515036100009004

$E\,[(\bar{\lambda} - a)\,(S - \bar{\lambda})] = 0$

(We integrate first over S with fixed $\bar{\lambda}$. As $E\,(S \mid \bar{\lambda}) = \bar{\lambda}$, this gives 0.)

Thus

$\text{cov}\,(\bar{\lambda}, S) = E\,(\bar{\lambda} - a)^2 = ab_1.$

Further, by 075:

$\text{var}\,S = a\,(1 + b_1 + b_2)$

Finally

$$\bar{\lambda}^* = a + \frac{ab_1}{a\,(1 + b_1 + b_2)}\,(k - a)$$

$$= \frac{a + cb_1\,k}{1 + cb_1} \text{ with } c = \frac{1}{1 + b_2}$$

which corresponds to my theorem 2. This derivation is much simpler than that given in my paper. For judging the practical applicability in concrete cases, I think it is useful to know that the theorem can be based either on linear regression or on my principle of accuracy.

There might be some doubt whether it makes sense to speak of a linear regression of $\bar{\lambda}$ on S. I think it makes sense. The case is exactly that of an urn containing slips with a pair of values of $\bar{\lambda}$ and S inscribed on each. The random experiment consists in drawing a slip, and the task is to estimate $\bar{\lambda}$ from S after having read S but not $\bar{\lambda}$. It is, however, a peculiarity of our problem that $\bar{\lambda}$ never will be directly observable, and that therefore special methods must be developed to determine the parameters of the linear regression. This is done in my paper.

B) My theorem 2 as well as, of course, theorem 1, can also be derived as a special case of the credibility relation

$$(1 - b).\ E\,[\mu(\theta)] + b.\ \overline{X}$$

derived in the paper "Experience Rating and Credibility" by Hans Bühlmann, appearing in this same number of the Astin Bulletin.

Downloaded from https:/www.cambridge.org/core. University of Basel Library, on 11 Jul 2017 at 14:03:33, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/S0515036100009004

The credibility relation of Bühlmann too can be derived from the formula for linear regression.

C) In discussions about my paper I realized that the "exchange of risks" treated in paragraphs 221, 222, 224 and 225 was not easily understood. It may therefore be useful to give a concrete example for these cases.

Such an example is workmen's compensation with

B = portfolio of policies
B_i = single policy i, i.e. single factory i
B_{ij} = single worker j of single factory i

The exchange of risks here means the dismissal of old workers and their replacement by new ones. It is obvious that the weight of past experience must be diminished by such an exchange.

Another example would be a class of risks where some of the policies expire and new ones are contracted.

April 1965

Downloaded from https:/www.cambridge.org/core. University of Basel Library, on 11 Jul 2017 at 14:03:33, subject to the Cambridge Core terms of use, available at https:/www.cambridge.org/core/terms. https://doi.org/10.1017/S0515036100009004