ON THE ESTIMATION OF MEANS AND VARIANCES IN THE CASE OF UNEQUAL COMPONENTS

ESA HOVINEN
Finland

The problem

In practical statistical work one frequently meets certain problems. For instance, we may have the following data about loss ratios in certain insurance companies and corresponding numbers of insurance in force:

Company	Loss ratio, ‰	insurance in force, 1000 (e.g.)
i	p_i	t_i
1	5	20
2	10	30
3	5	80
4	4	75
5	40	5
6	20	15
		225

I assume further that we have no reason to believe that the companies, their loss ratios and their structure of insurances in force differ in any other way than by the size of companies. The problem is how to get quick estimates of mean losses and their variances in different companies?

A straitforward way to estimate the mean loss ratio would be to compute the usual mean of numbers p_i, $(\Sigma p_i/6) = 14$; its standard deviation is 6,5. As this procedure of the "first statistician" seems to be too simple and naive, a "more cautious" statistician would compute the weighted mean loss ratio

$$\bar{p} = \frac{\Sigma t_i p_i}{\Sigma t_i} = 7{,}1.$$

The "more cautious" statistician would arque that his result is much better than the other result 14. But what would be the variance of the estimate 7,1, and what is the variance in the different companies?

https://doi.org/10.1017/S0515036100011259 Published online by Cambridge University Press

The purpose of this paper

In this paper I try to solve the problem of the two statisticians. Thus, I try to throw a little light on the methodology in estimation of means and variances in the different companies.

In the example we had certain numbers. To begin a reasonable computation, one necessarily needs specific assumptions of the situation or process, which can lie behind the observed numbers. The assumptions determine the mathematical formulas to be used and the meaning of the computed results.

The first statistician apparently used the assumption that the stochastic variables p_i were equally distributed. The other statistician did not believe in this, hence he used the weights t_i. However, the "more cautious" statistician seems to have assumed that the expectations of p_i be the same.

In the example we have six observations of the numbers p_i, certain arguments according to which the expectations of p_i should not differ, and numbers t_i which measure the "size" of the "observations" namely the companies. To get further, one apparently needs assumptions of the distributions of p_i. The theory of stochastic processes and the theory of risk seem to give possibilities to formulate assumptions of the processes, which could generate the numbers p_i for the different companies.

In stochastic processes there generally is a kernel variable x with a distribution $f(x)$. The process yields from x other variables x_i depending on a parameter t_i. If $f(x)$ and the process are known, one can compute the distribution $F_i(x)$ of x_i and thus all of its statistical characteristics, e.g. the moments $A_{r,i}$ of $F_i(x)$, which apparently depend on certain characteristics of $f(x)$ and the parameter t_i. In the example one can make different assumptions of the underlying process which yields the total loss l_i of the company i. The loss l_i apparently depends on the size t_i of the company i. However, it seems reasonable to assume, that the *intensity variables*

$$p_i = \frac{l_i}{t_i}$$

have the same expectation. We see, that the "more cautious" statistician can find good arguments.

https://doi.org/10.1017/S0515036100011259 Published online by Cambridge University Press

The idea behind this paper is that it should be possible to formulate estimators α_r of certain characteristics of $f(x)$, using all the observed results l_i or p_i and some assumptions about the underlying process. If the estimators α_r are known, the assumed process gives corresponding estimators for e.g. the means and variances of p_i in the different companies, naturally depending on the parameter t_i.

In the few examples treated later, I will show that using quite reasonable assumptions we can get estimators

$$\alpha_1 = \frac{\Sigma l_i}{\Sigma t_i} = \frac{\Sigma p_i t_i}{\Sigma t_i} = \bar{p}$$

and

$$\alpha_2 = \frac{1}{n} \sum_i \frac{t_j}{1 - \frac{t_j}{\Sigma t_j}} (p_j - \alpha_1)^2, \qquad (1)$$

where n is the number of observations, i.e. the number of the companies in the example. The corresponding estimators for the means and variances of the loss ratios p_i for each of the companies are

$$E(p_i) \approx \alpha_1 = \bar{p}$$

$$V(p_i) \approx \frac{\alpha_2}{t_i}.$$

Naturally, the estimated variance of $\bar{p}$ is then $V(\bar{p}) \approx (\alpha_2/\Sigma t_i)$.

The problem is by its very nature much more general than a problem of estimating loss ratios. In mathematical statistics one often meets situations where the components (here the loss ratios of the different companies) are differently distributed. Knowing very few textbooks on the subject I have a feeling that something is lacking: when the components are different, the armament of statisticians fades away. If expression (1) is right, we could further develop methods for testing, enlarge the analysis of variance and so on. The most ambitious possibility would be to abandon the requirement of equal components in statistical inference quite generally.

The problem-solving has many influences from the theories and practical data which the ASTIN group has produced. Therefore I

https://doi.org/10.1017/S0515036100011259 Published online by Cambridge University Press

dare to present this paper to this forum, because I believe the ASTIN members need more powerful statistical armaments. Because I am not very well acquainted with the modern mathematical way of expressing different things I apologize for the possibly antique expressions and developments in this paper.

The problem solving structure

It is known, that if stochastic variables y_i have

distributions $F_i'(y)$
equal expectations $y = \int y_i dF_i'(y)\,;\, i = 1, 2, \ldots$
variances $\sigma_i^2 = \int (y_i - y)^2\, dF_i'(y)$

and if the variables y_i are mutually independent, then the linear estimate

$$\bar{y} = \frac{1}{\sum_i \frac{1}{\sigma_i^2}} \sum_i \frac{y_i}{\sigma_i^2} = \frac{1}{\sum_i \frac{1}{k\sigma_i^2}} \sum \frac{y_i}{k\sigma_i^2} \qquad (2)$$

from a random sample $Y = (Y_1, Y_2, \ldots)$ is an unbiased least mean square estimate of y. The number k is an arbitrary constant $\neq 0$. See, e.g. Hald, Statistical Theory with Engineering Applications, page 243.

If, as in our example, we can make assumptions about the process and its underlying characteristics depending on a parameter t_i, we may perhaps construct variables α, which characterize the function $f(x)$ and which can be estimated for all of the observations i. The variances of these estimators can then be calculated, whereafter (2) gives us the "best overall estimate".

In the example we may assume that each policy is an independent unit, the behaviour of company i can be expressed as a t_i-fold convolution of the behaviour of one policy. Further, we may assume that we know the number t_i exactly, and that the underlying basic probability of loss is a constant (*simple convolution*), SC.

Another situation would be, that we know the numbers of expected losses, which are then our parameters. Naturally, this situation is the same as SC, with the numbers of SC multiplied by an overall assumed loss frequency. But the process is different. We have the case of *generalized Poisson process*, GPP.

Further, we may assume that the process is a generalized Poisson process, but that the underlying probability of a claim varies according to a distribution $S(\lambda)$, which has the expectation $E\{\lambda\} = 1$ and variance $V\{\lambda\} = B$. We have the case of *compound Poisson process*, CPP.

Thus we have got three different assumptions of the process which can lie behind the numbers in our example. The essential differences in these cases lie in the assumptions of the nature of the whole process. How to proceed?

The processes

Above we have found three different kinds of "underlying" processes. In each of them an underlying distribution $f(x)$ of a variable x can be assumed to exist. The processes are, correspondingly

$$F(x) = f(x) * f(x) * \ldots f(x) = f(x)^{t_i *}, \text{ SC} \tag{3}$$

$$F(x) = \sum_k e^{-t_i} \frac{t_i^k}{k!} f(x)^{k*}, \text{ GPP} \tag{4}$$

$$F(x) = \int \sum_k e^{-\lambda t_i} \frac{(\lambda t_i)^k}{k!} f(x)^{k*} dS(\lambda), \text{CPP.} \tag{5}$$

Here and in the sequel I have partially omitted the index i, which should be understood to belong to all of the notations $F(x)$, A_r etc.

A somewhat clearer picture of the processes can be found by exploring the corresponding characteristic functions and moments of the variables x_i. The characteristic functions are, when the characteristic function of the underlying distribution is φ_f,

$$\varphi_F = \varphi_f^{t_i} \text{ (SC)} \tag{6}$$

$$\varphi_F = e^{t_i(\varphi_f - 1)} \text{ (GPP)} \tag{7}$$

$$\varphi_F = \int e^{t\lambda_i(\varphi_f - 1)} dS(\lambda) \text{ (CPP) respectively.} \tag{8}$$

The corresponding moments A_r of F and a_r of f are

$$A_r = i^{-r} \varphi_F^{(r)} \ (u = 0) \tag{9}$$

$$a_r = i^{-r} \varphi_f^{(r)} \ (u = 0). \tag{10}$$

In each of the different cases the characteristic function φ_F is of the form

$$\varphi_F = \chi(\varphi_f, t_i),$$

https://doi.org/10.1017/S0515036100011259 Published online by Cambridge University Press

which gives

$$\begin{cases} A_1 = \chi_\varphi'(1, t_i)\, a_1 \\ A_2 = \chi_\varphi''(1, t_i) a_1^2 + \chi_\varphi'(1, t_i) a_2 \\ A_3 = \chi_\varphi'''(1, t_i) a_1^3 + 3\chi_\varphi''(1, t_i) a_1 a_2 + \chi_\varphi'(1, t_i) a_3 \\ A_4 = \chi_\varphi^{(4)}(1, t_i) a_1^4 + 6\chi_\varphi'''(1, t_i) a_1^2 a_2 + 4\chi_\varphi''(1, t_i) a_1 a_3 \\ \qquad + 3\chi_\varphi''(1, t_i) a_2^2 + \chi_\varphi'(1, t_i) a_4 \end{cases} \tag{11}$$

In the three different cases we have, correspondingly,

$$\begin{cases} \chi_\varphi'(1, t_i) = t_i \\ \chi_\varphi''(1, t_i) = t_i^2 - t_i \\ \chi_\varphi'''(1, t_i) = t_i^3 - 3t_i^2 + 2t_i \\ \chi_\varphi^{(4)}(1, t_i) = t_i^4 - 6t_i^3 + 11t_i^2 - 6t_i \end{cases} \quad \text{SC} \tag{12}$$

$$\begin{cases} \chi_\varphi'(1, t_i) = t_i \\ \chi_\varphi''(1, t_i) = t_i^2 \\ \chi_\varphi'''(1, t_i) = t_i^3 \\ \chi_\varphi^{(4)}(1, t_i) = t_i^4 \end{cases} \quad \text{GPP} \tag{13}$$

$$\begin{cases} \chi_\varphi'(1, t_i) = t_i \\ \chi_\varphi''(1, t_i) = Bt_i^2 + t_i^2 - t_i \\ \cdots\cdots\cdots\cdots\cdots \end{cases} \quad \text{CPP} \tag{14}$$

Together, the equations (11) ... (14) determine the moments or expectations of the r^{th} power of the loss in the different cases.

The mean

The problem of the mean seems to be quite simple. Following the thought behind the formula (2) we must seek expressions which have the mean of p_i as expectation, then find the weights $1/\sigma^2(p_i)$ and use the formula (2).

Remembering the formulas (11)-(14) we can see, that the expectations

$$E(p_i) = E\left(\frac{l_i}{t_i}\right) = \frac{A_{1,i}}{t_i} = a_1 \tag{15}$$

are the same. The corresponding weights σ_i^2 are, consequently

$$\sigma_i^2 = E(p_i - a_1)^2 = E(p_i^2) - a_1^2.$$

https://doi.org/10.1017/S0515036100011259 Published online by Cambridge University Press

Using the formulas (11)-(14) we get

$$\sigma_i^2 = \frac{1}{t_i^2} E(l_i^2) - a_1^2 = \frac{A_2}{t_i^2} - a_1^2, \text{ which gives}$$

$$\begin{cases} \sigma_i^2 = \dfrac{t_i^2 a_1^2 - t_i a_1^2 + t_i a_2}{t_i^2} - a_1^2 = \dfrac{1}{t_i}(a_2 - a_1^2) = \dfrac{k}{t_i} & \text{(SC)} \\ \sigma_i^2 = \dfrac{t_i^2 a_1^2 + t_i a_2}{t_i^2} - a_1^2 = \dfrac{1}{t_i} a_2 = \dfrac{k}{t_i} & \text{(GPP)} \\ \sigma_i^2 = \dfrac{B t_i^2 a_1^2 + t_i^2 a_1^2 - t_i a_1^2 + t_i a_2}{t_i^2} - a_1^2 = B a_1^2 + \dfrac{a_2 - a_1^2}{t_i} & \text{(CPP)} \end{cases} \tag{16}$$

Thus the first problem is solved: the mean loss is in the cases of SC and GPP

$$\bar{p} = \frac{1}{\Sigma (1/k\sigma_i^2)} \sum \frac{p_i}{k\sigma_i^2} = \frac{1}{\Sigma t_i} \sum t_i p_i = \frac{\Sigma t_i p_i}{\Sigma t_i}; \tag{17}$$

this is the same as the formula of the "more cautious" statistician. In the case of CPP we get

$$\bar{p} = \frac{1}{\sum \dfrac{t_i}{B a_1^2 t_i + (a_2 - a_1^2)}} \cdot \sum \frac{t_i p_i}{B a_1^2 t_i + (a_2 - a_1^2)}. \tag{18}$$

If $B = 0$, this expression reduces to (17). If, on the other hand, $B >> \dfrac{a_2 - a_1^2}{a_1^2 t_i}$, the expression (18) reduces to

$$\bar{p} \approx \frac{1}{n} \Sigma p_i, \tag{19}$$

which was the formula of our first statistician! Here the variance B outweighs the other parameters. Thus both of the statisticians are right, under different assumptions.

The problem of the mean losses l_i can thus be regarded as solved. The problem of variances still remains. In the sequel I shall concentrate on the problem of the "more cautious" statistician only and omit the question of the CPP.

The variance

The mean of the variance

Concentrating on the question of finding estimators of α_2 which could generate estimators for the variances of p_i we can study the expression

$$V(p_i) = V\left(\frac{l_i}{t_i}\right) \tag{20}$$

in the cases of SC and GPP. In both of these situations we have

$$V\left(\frac{l_i}{t_i}\right) = E\left(\frac{l_i}{t_i} - a_1\right)^2 = E\left(\left(\frac{l_i}{t_i} - \bar{p}\right) + (\bar{p} - a_1)\right)^2 =$$

$$E\left(\frac{l_i}{t_i} - \bar{p}\right)^2 + E(\bar{p} - a_1)^2 + 2E\left(\left(\frac{l_i}{t_i} - \bar{p}\right)(\bar{p} - a_1)\right). \tag{21}$$

By virtue of the relations (11)-(13), the third term in the expression (21) becomes zero. In verifying this, the mean

$$\bar{p} = \frac{\Sigma\, t_i p_i}{\Sigma\, t_i}$$

is partitioned into parts

$$\bar{p} = \frac{\Sigma\, t_i p_i - p_i t_i}{\Sigma\, t_i} + \frac{p_i t_i}{\Sigma\, t_i} = \frac{\Sigma\, l_i - l_i}{\Sigma\, t_i} + \frac{l_i}{\Sigma\, t_i}$$

and the relation

$$E(xy) = E(x)E(y)$$

for independent variables x and y is used. The second term is similarly

$$E(\bar{p} - a_1)^2 = \frac{t_i}{\Sigma\, t_i}\, E\left(\frac{l_i}{t_i} - a_1\right)^2.$$

Thus

$$V\left(\frac{l_i}{t_i}\right) = E\left(\frac{l_i}{t_i} - \bar{p}\right)^2 + \frac{t_i}{\Sigma\, t_i}\, V\left(\frac{l_i}{t_i}\right).$$

https://doi.org/10.1017/S0515036100011259 Published online by Cambridge University Press

In case of SC

$$V\left(\frac{l_i}{t_i}\right) = \frac{1}{t_i}(a_2 - a_1^2)$$

and in case of GPP

$$V\left(\frac{l_i}{t_i}\right) = \frac{1}{t_i} a_2.$$

We see, that the expressions

$$\boxed{Z_i = t_i V\left(\frac{l_i}{t_i}\right) = E\left[\frac{t_i}{1 - \frac{t_i}{\Sigma t_i}}(p_i - \bar{p})^2\right]} \tag{22}$$

all have the same expectation, which is in SC $= a_2 - a_1^2$, and in GPP $= a_2$. Thus expressions with common means have been found and the first stage in our search for variance is accomplished.

The weights

In using the adopted problem solving structure, we further need the inverse weights

$$k \cdot V(Z_i)$$

as functions of t_i. Let us denote

$$n_i = \frac{1 - \frac{t_i}{\Sigma t_i}}{t_i}$$

$$m = \frac{1}{\Sigma t_i} \text{ and}$$

$$y_i = \sum_{j \neq i} l_j = \Sigma p_j t_j - p_i t_i.$$

Then

$$V(Z_i) = \frac{1}{n_i^2} V(p_i - \bar{p})^2$$

$$= \frac{1}{n_i^2} E(p_i - \bar{p})^4 - (E(Z_i))^2$$

$$= \frac{1}{n_i^2} E(n_i l_i - m y_i)^4 - (E(Z_i))^2.$$

https://doi.org/10.1017/S0515036100011259 Published online by Cambridge University Press

Since the variables l_i and y_i are independent, we get further

$$V(Z_i) = \frac{1}{n_i^2}\left[n_i^4 E(l_i^4) - 4n_i^3 m E(l_i^3)E(y_i) + 6n_i^2 m^2 E(l_i^2)E(y_i^2) - 4n_i m^3 E(l_i)E(y_i^3) + m^4 E(y_i^4)\right] - (E(Z_i))^2.$$

This expression can be worked out further, remembering that $E(x^r) = A_r$, using the formulas (11)-(13) and seeing that the parameters t_i and $\Sigma t - t_i$ correspond to variables l_i and y_i respectively. The computation can be shortened remarkably if one uses the line of thought that Kendall uses in his book "The Advanced Theory of Statistics", page 256. In the case of SC the variance and higher central moments must be independent of the "location parameter" a_1. Thus all the expectations of odd powers of l_i and y_i can be neglegted as well as all terms containing a_i. The computation yields in the case of simple convolution the result

$$V(Z_i) = c_2^2\left(2 + \gamma^2\left(\frac{1}{t_i} + \frac{1}{\Sigma t_i - t_i} - \frac{3}{\Sigma t_i}\right)\right). \tag{23}$$

where $c_2 = a_2 - a_1^2$ is the variance of $f(x)$ and

$$\gamma_2 = \frac{c_4}{c_2^2} - 3 = \frac{\chi_4}{\chi_2^2}$$

is the excess of $f(x)$. The α:s are the corresponding cumulants of $f(x)$.

The case of CPP yields naturally the same result as a convolution process, where

$$c_2 = a_2$$

and

$$\chi_4 = a_4.$$

The variance

According to (2) we can now write the unbiased least mean square estimate

$$V(p_i) \approx \frac{1}{t_i}\frac{1}{\sum \frac{1}{2 + T_i\gamma_2}}\sum \frac{t_i}{\left(1 - \frac{t_i}{\Sigma t_i}\right)(2 + T_i\gamma_2)}(p_i - \bar{p})^2, \tag{24}$$

https://doi.org/10.1017/S0515036100011259 Published online by Cambridge University Press

where

$$T_i = \frac{1}{t_i} + \frac{1}{\Sigma t_i - t_i} - \frac{3}{\Sigma t_i} \; (> 0). \tag{25}$$

The expression (24) is computable and without difficulties programmable on computers if assumptions about excess γ_2 can be made. If all the parameters t_i are equal, (24) reduces to the classical result

$$V(p_i) = \frac{1}{n-1} \Sigma (p_i - \bar{p})^2.$$

If the excess $\gamma_2 = 0$ (e.g. normal distributions), we get

$$\boxed{V(p_i) \approx \frac{1}{t_i} \frac{1}{n} \sum \frac{t_i}{1 - \frac{t_i}{\Sigma t_i}} (p_i - \bar{p})^2.} \tag{26}$$

If, on the other hand, we let $\gamma_2 \to \infty$, the "best" estimate becomes

$$V(p_i) = \frac{1}{t_i} \frac{1}{\Sigma t_i} \sum \frac{t_i^2}{1 - \frac{t_i}{\Sigma t_i}} (p_i - \bar{p})^2. \tag{27}$$

In GPP the excess

$$\gamma_2 = \frac{a_4}{a_2^2} \geqq 1;$$

where $\gamma_2 = 1$ if all the possible "claims" are equal.

Discussion

1. *On the basis of estimation*

The whole line of thought in this paper seems to be in connection with the concept of *infinite divisibility*, as, treated e.g. by Feller in his book "An Introduction to Probability Theory and its Applications, Vol. II"[1]). Taking the variable t_i as a significant factor in estimation, means that the variables p_i belong to the same "family" of distributions, where t_i or its multiples define the exact situation of p_i in the family.

[1]) In his book Feller uses vocabulary differing from the one used here. The most important difference is between GPP and CPP; with CPP Feller means GPP used here.

https://doi.org/10.1017/S0515036100011259 Published online by Cambridge University Press

If we have as in (6)-(8)

$$\varphi_F = \chi(\varphi_f, t_i),$$

the cumulants

$$\varkappa_k = i^{-k} D^{(k)} \log \chi_{(u-0)} \tag{28}$$

equal t_i — times certain characteristics of $f(x)$,

$$D^{(r)}_{\varphi_f} \log \chi(\varphi_f, t_i) = t_i D^{(r)}_{\varphi_f} \log \mathcal{Q}(\varphi_f), \tag{29}$$

if χ is of the form

$$\chi = e^{t_i \mathcal{Q}(\varphi_f)}, \tag{30}$$

where $\mathcal{Q}$ is such a function of φ_f only, that $e^{\mathcal{Q}(\varphi_f)}$ is characteristic function of a distribution. In the case of simple convolution we have $\mathcal{Q}(\varphi) = \ln \varphi$ and in the case of GPP $\mathcal{Q}(\varphi) = \varphi - 1$. In SC the cumulants correspond cumulants of $f(x)$, in GPP the cumulants of $F(x)$ correspond moments of $f(x)$.

2. *The variance of* $V(p_i)$

The variance of (24)

$$V(V(p_i)) = \frac{1}{t_i^2} \frac{1}{\sum \frac{1}{V(Z_i)}}.$$

If $\gamma_2 = 0$, then

$$V(V(p_i)) = \frac{1}{t_i^2} \cdot \frac{2c_2^2}{n} \qquad \text{(SC)}.$$

If $\gamma_2 \to \infty$, $V(V(p_i)) \to \infty$.

If the term $T_i\gamma_2 >> 2$ in (24), and all of the t_i:s are small as compared to Σt_i ("Lindeberg case"), then

$$V(V(p_i)) \to \frac{1}{t_i^2} \cdot \frac{\gamma_2 \cdot c_2^2}{\Sigma t_i}.$$

3. *The original example*

The problem was to solve the problem of the statisticians. The first statistician found $\bar{p} = 14$, $V(\bar{p}) = 42$. We saw that he used assumption CPP with the variation of the loss ratio outweighing

https://doi.org/10.1017/S0515036100011259 Published online by Cambridge University Press

the influence of t_i:s. The variance of p_i should thus be 255 and standard deviation 16.

The other statistician had more difficulties. The variance of p_i should be (24). Making different assumptions of process and its parameters he can get different answers. Taking e.g. the GPP and assuming that for the smallest company the excess ≈ 3 he gets $\gamma_2 = 15$. For the standard deviation of the mean $\bar{p} = 7{,}1$ (Σt_i) he gets $\sigma = 2{,}4$. The application of (26) gives correspondingly $\sigma = 2{,}7$ and the application of (27) the result $\sigma = 2{,}0$.

The answers differ a little of each other and quite strikingly from the answers of the first statistician, $\bar{p} = 14$, $\sigma = 6{,}5$. As the latter statistician was "cautious", I would believe he would generally use formula (26) in other applications!

4. *Other applications*

As another example we can study the problem of smoothing: we have observed e.g. mortality data in a given group where we have observed numbers of deaths in different age groups; the observed cases in these groups naturally being different. It may be difficult to arrange the situation so that the age groups were stochastiqually equal. How to make sound estimates?

This paper has tried to give framework and theory for the analyses of those very different problems we have in everyday life. I believe that we have very many problems of this kind.

https://doi.org/10.1017/S0515036100011259 Published online by Cambridge University Press