

COMMENTS ON DIFFERENT DEDUCTIONS OF  
EXPRESSIONS FOR CONDITIONAL EXPECTATIONS

CARL PHILIPSON

Stockholm

1. In a paper presented to the fifth Astin Colloquium (Lucerne, 1965) Bühlmann has given some propositions with regard to Experience Rating understood as a sequence of estimates of the expectation with respect to the distribution function of  $\theta$ ,  $H(\theta)$  say, of  $\mu(\theta)$  which is the mean for fixed  $\theta$  with respect to the distribution function  $G(x; \theta)$  of the variables  $x_v$ , for each value of  $v = 1, 2 \dots n$ . In such problems the estimator function for  $\mu(\theta)$  is generally chosen to be the conditional mean of  $\mu(\theta)$  for a given set of observed values of  $x_v$ ,  $v = 1, 2 \dots n$ . This is generally justified by the principle of least square deviation. According to Bühlmann this justification is not sufficient. Therefore, he bases the choice of this estimator function upon a postulate of equilibrium, described in the following lines.

Let  $X'$  be a subset of  $X$ , and  $C(X')$  a cylinder with the base  $X' \subset X$  in the product space  $X \times \Theta$ , where  $X$  is the set of all possible  $\{x_v\}$  and  $\Theta$  the set of all possible  $\theta$ , then the postulate of equilibrium implies the equality between the expectations of  $\mu(\theta)$  and of a function of  $\{x_v\}$  on each cylinder  $C(X')$ . This is exactly Kolmogoroff's definition of the conditional expectation:

$$E [\mu(\theta) \mid x_1, x_2 \dots x_n] \tag{1}$$

Bühlmann states, further, that the best linear estimate of (1) based on the arithmetic mean  $\bar{x}$  of  $n$  sample values  $x_v$  of  $x_v$ ,  $v = 1, 2 \dots n$  can be written in the form of the Credibility Formula:

$$(1-\alpha) E [\mu(\theta)] + \alpha \bar{x}, \tag{2}$$

where the symbol  $E[\cdot]$  denotes the expectation over the product space  $X \times \Theta$  and

$$\alpha = \frac{\text{Var} [\mu(\theta)]}{\text{Var} [\mu(\theta)] + E(\bar{x} - \mu(\theta))^2}$$

If the  $x_v$ 's are independent and identically distributed, this leads to  $\alpha = \frac{n}{n+k}$ , where  $k = E[\sigma^2(\theta)] / \text{Var}[\mu(\theta)]$ , and  $\sigma^2(\theta)$  is the variance for a fixed  $\theta$  of  $G(x; \theta)$ . This case of the credibility formula is, generally, applied in American practice. (The proposition has later been proved for more general conditions).

2. Bühlmann's deduction amounts to a distribution-free credibility formula, (2). As a particular case of Bühlmann's proposition we now make the following assumptions.

Let  $x_v$  take only the values 0 and 1, distributed for each fixed value of  $\theta$  with an exponential probability function fulfilling the conditions for the sum  $x^* = \sum_{v=1}^m x_v$  being distributed with the Poisson probability distribution for every fixed value of  $\theta$

$$G(x^*; \theta) = \frac{\theta^{x^*}}{x^{*!}} e^{-\theta}; \mu(\theta) = \theta. \tag{3}$$

If the parameter  $\theta$  is a positive random variable of mean  $t$ , say, then the marginal probability for  $x^*$  being equal to  $n$ , for a given value of  $t$ ,  $P_n(t)$  say, is given by the following relation, where the last member is obtained by the introduction of the new variable of integration  $v = \theta/t$ .

$$P_n(t) = \int_0^\infty \frac{\theta^n}{n!} e^{-\theta} dH(\theta) = \int_0^\infty \frac{(vt)^n}{n!} e^{-vt} dH(vt) \tag{4}$$

Evidently  $\int_0^\infty v d_v H(vt) = 1$ ; let  $\int_0^\infty (v-1)^2 d_v H(vt)$  be designated by  $b$ .

In this case  $E[x^* - \mu(\theta)]^2 = E(\theta) = t$ ,  $\text{Var}[\mu(\theta)] = E(\theta-t)^2 = bt^2$  and, consequently, by a deduction similar to Bühlmann's proof the best linear estimate for  $\mu(\theta)$  based on  $x^* = n$  is given by

$$t \cdot \frac{1 + bn}{1 + bt}, \tag{5}$$

which implies that the best linear estimate of the conditional mean of  $\theta/t = v$ , given  $(n, t)$ , is equal to

$$\frac{1 + bn}{1 + bt} \tag{6}$$

(6) can be used in practice only if  $b$  is either known or can be estimated (see section 6).

3. Ove Lundberg (Astin Bull. Vol IV Part I) presented to the same colloquium a paper on rating systems based on compound Poisson processes, which are defined by probability distributions of the number  $n$  of events in the interval  $(0, t)$  in a form obtained from (4) by replacing the notation  $H(vt)$  with  $U(v)$ , say. Lundberg derives an exact expression for the conditional mean of  $v$  given  $(n, t)$  in the following form.

$$E(v|n) = - \frac{P_o^{(n+1)}(t)}{P_o^{(n)}(t)} = p_n(t), \quad P_o^{(n)}(t) = \frac{d^n P_o(t)}{dt^n} \quad (7)$$

where  $p_n(t)$  is the intensity function of the compound Poisson process (in the narrow sense). (7) was already given by Lundberg in 1940 (On Random Processes and Their Application to Sickness and Accident Statistics, Uppsala).

For a particular case of (7) Lundberg also deduces the best linear estimate of  $E(v|n)$  based on  $(n, t)$  and finds an expression coinciding with (6), derived in the previous section from Bühlmann's proof. (6) is the canonical form of the intensity function of a Polya process for which  $U(v)$  is defined by a central Pearson Type III density function. Delaporte (Trans. Int. Congr. Act. London, Vol. III, 1964 and other papers, quoted there) has used an expression equal to the intensity function of a modified process, for which  $U(v)$  is defined by a non-central Type III density function. Delaporte has published numerical data for the risk in motor insurance which show an excellent agreement between the empirical frequencies and the probabilities calculated from his model. Bichsel's paper to the colloquium in Lucerne is also based on expressions in the form of (6). He accounts in his developments for a random variation in time in the parameter  $\theta$ . The present author in an earlier note (Mitteil. Ver. schweiz. Vers. Math., 64, 1964) has pointed to the necessity of accounting for a time variation in  $\theta$ , including not only random but also non-random variation. This was also stressed by Bühlmann.

4. Robbins (Rev. de l'Inst. Int. de Stat., 31, 1963) has given a theory of Bayesian estimation in a quite general form. He uses a particular test function which he calls the Bayesian loss function, the estimates being chosen so as to minimize this function. As an

example of the application of his method, he deduces the best Bayesian estimate in this sense for  $E[\mu(\theta)]$  in the case where  $G(x^*; \theta)$  is defined by (3) and arrives at the following expression, given here in our notation

$$\frac{(n + 1) P_{n+1}(t)}{P_n(t)} \tag{8}$$

By using the well-known identity  $P_n(t) = \frac{(-t)^n}{n!} P_0^{(n)}(t)$  in the theory of compound Poisson processes, it is seen that (8) is consistent with (5). Robbins remarks that this result can easily be extended to cases, where  $G(x^*; \theta)$  is in the general form  $\theta x^* u(x^*) v(\theta)$ ,  $u$  and  $v$  being functions of a single argument, including the negative binomial. He has also exemplified the application to some commonly used continuous distributions—the normal and negative exponential distribution, and the  $\Gamma$ -distribution—and to a modified binomial distribution. For the risk theory the compound exponential distributions have a great interest (cf. Thyron, Bull. Ass. Roy. Act. Belges, 1964).

5. As a final remark on the problem considered in the previous sections it shall be pointed out that distributions of the kind treated here play an important role in decision theory. In this case  $\theta$  is considered as a measure of factors of influence on the result of a decision. A person comes to a decision with regard to a more or less intuitive knowledge of  $\theta$ .

6. In cases where  $b$  and  $t$  are unknown, these quantities may often be estimated from available material. Grenander (Skand. Akt. Tidskr., 1957) has given formulae for unbiased estimation of the mean and variance of  $H(\theta)$ , which leads to the following relations for a statistical group containing  $m$  units divided in subgroups of  $m_v$  units,  $v$  being the number of events in  $(0, t)$  for each subgroup.

$$t^* = \frac{1}{m} \sum_{v=0}^{\infty} m_v v \tag{9}$$

$$b^* = \frac{1}{t^{*2} (m-1)} \sum_{v=0}^{\infty} m_v v (v-1) - \frac{m}{m-1} + \frac{1}{t^* (m-1)}$$

$b^*$  is asymptotically normal with zero mean and asymptotic variance  $\frac{2}{t^{*2} m}$ .

In cases where the corresponding magnitudes are associated with an individual risk within the statistical group, the mean of the magnitudes for an individual may be obtained from (9) by division with  $m$  (cf. a paper by the present author, Trans. Int. Congr. Act. Vol. III, London 1964). Bühlmann remarks in the paper quoted in section 1, that the direct estimates of the magnitudes appearing in  $\alpha$  in (2) would be of great interest and this problem has not yet been attacked. Evidently, it has been treated by Grenander, as quoted above, and also by Anscombe to be quoted in the next section.

7. In the paper quoted Grenander has discussed the difficulties of estimating the distribution  $H(\theta)$ . Anscombe (Biometrika, 1950) has, however, given a method for the choice between some well-known forms for this distribution, based on the third factorial moment of  $P_n(t)$ . A better estimate of (7) than the linear estimate based on the Polya process may be found from data, which are sufficient for giving a satisfactory estimate of the third factorial moment, by using the intensity function of a process chosen among Anscombe's type processes with the aid of this moment. For the calculation of the intensity function the estimates of  $b$  and  $t$  according to the previous section may be used.

Stockholm den 24 maj 1966