Prediction and estimation for the compound Poisson distribution

(empirical Bayes/accident proneness)

HERBERT ROBBINS

Department of Mathematical Statistics, Columbia University, New York, New York 10027

Contributed by Herbert Robbins, May 2, 1977

ABSTRACT An empirical Bayes method is proposed for predicting the future performance of certain subgroups of a compound Poisson population.

Suppose that last year a randomly selected group of n people experienced, respectively, $x_1, \dots, x_n$ "accidents" of some sort. Under similar conditions during next year the same n people will experience some as yet unknown number of accidents $y_1, \dots, y_n$. Let i be any positive integer. We are concerned with predicting, from the values $x_1, \dots, x_n$ alone, the value of

$$S_{i,n} = \{\text{sum of the } y_j\text{'s for which the corresponding } x_j\text{'s are } < i\} \quad [1]$$

that is, the total number of accidents that will be experienced next year by those people who experienced less than i accidents last year.

The quantity

$$\{\text{sum of the } x_j\text{'s that are } < i\} \quad [2]$$

is a poor predictor of [1], since it will usually underestimate [1] by a considerable amount. Thus, for $i = 1$ the value of [2] is 0, but the people who were lucky last year (perhaps only a small fraction of n) are not likely to remain so next year, and hence $S_{1,n}$ will be rather large if people in general are accident prone.

We shall show that, in contrast to [2], a good predictor of $S_{i,n}$ is the quantity

$$E_{i,n} = \{\text{sum of the } x_j\text{'s that are } \leq i\} \quad [3]$$

In particular, for $i = 1$, a good predictor of $S_{1,n} = \{\text{sum of the } y_j\text{'s for which the corresponding } x_j\text{'s are } 0\}$ is the quantity $E_{1,n} = \{\text{number of people who had exactly 1 accident last year}\}$.

The main theorem

Let (λ, x, y) be a random vector ($0 < \lambda < \infty$; $x, y = 0, 1, \dots$) such that

Given λ , x and y are independent

Poisson random variables with mean λ . [4]

We denote the distribution function of λ by G , but make no assumptions about it.

The conditional probability function of x , given λ , is the Poisson

$$f(x|\lambda) = e^{-\lambda} \lambda^x / x! \quad [5]$$

with a similar expression for $f(y|\lambda)$. The unconditional probability function of x is the compound Poisson

$$f(x) = \int_0^\infty (e^{-\lambda} \lambda^x / x!) dG(\lambda) \quad [6]$$

with a similar expression for $f(y)$. It follows from Eq. 5 that

$$E(x|\lambda) = \lambda, \quad E(x^2|\lambda) = \lambda + \lambda^2, \quad \text{Var}(x|\lambda) = \lambda, \\ E x = E \lambda, \quad E(x^2) = E \lambda + E(\lambda^2), \\ \text{Var } x = E \lambda + \text{Var } \lambda \quad [7]$$

and similarly for y .

We now define the random variables

$$v = \begin{cases} 1 & \text{if } x < i \\ 0 & \text{if } x \geq i \end{cases}, \quad w = vy = \begin{cases} y & \text{if } x < i \\ 0 & \text{if } x \geq i \end{cases} \quad [8]$$

and observe that by [4]

$$E(w|\lambda) = E(vy|\lambda) = E(v|\lambda) \cdot E(y|\lambda) = \lambda E(v|\lambda) \\ = \sum_{x=0}^{i-1} (e^{-\lambda} \lambda^{x+1} / x!) \quad [9]$$

so that

$$Ew = \sum_{x=0}^{i-1} \int_0^\infty (e^{-\lambda} \lambda^{x+1} / x!) dG(\lambda) \\ = \sum_{x=0}^{i-1} (x+1)f(x+1) = \sum_{x=0}^i xf(x) \quad [10]$$

Defining

$$u = \begin{cases} x & \text{if } x \leq i \\ 0 & \text{if } x > i \end{cases} \quad [11]$$

we see that

$$Eu = Ew \quad [12]$$

Let (λ_j, x_j, y_j) for $j = 1, \dots, n$ be independent random vectors with the same distribution as (λ, x, y) , and define

$$S_{i,n} = \sum_1^n w_j = \text{sum of the } y_j\text{'s for which the corresponding } x_j\text{'s are } < i \quad [13]$$

$$E_{i,n} = \sum_1^n u_j = \text{sum of the } x_j\text{'s that are } \leq i \quad [14]$$

$$N_{i,n} = \sum_1^n v_j = \text{number of } x_j\text{'s that are } < i \quad [15]$$

Then, as $n \rightarrow \infty$ the law of large numbers and the central limit theorem combine to show that

$$\sqrt{n} \left\{ \frac{E_{i,n}}{N_{i,n}} - \frac{S_{i,n}}{N_{i,n}} \right\} = \frac{\sum_1^n (u_j - w_j) / \sqrt{n}}{\sum_1^n v_j / n} \rightarrow N(0, A_i^2) \quad [16]$$

in distribution, where

$$A_i^2 = E(u - w)^2 / E^2 v \quad [17]$$

In order to evaluate Eq. 17 in terms of Eq. 6 it is convenient to define

$$a_i = \sum_{x=0}^{i-1} f(x), \quad b_i = \sum_{x=0}^i xf(x), \quad c_i = \sum_{x=0}^i x^2 f(x) \quad [18]$$

so that

$$Eu = Ew = b_i, \quad Ev = E(v^2) = a_i, \quad E(u^2) = c_i \quad [19]$$

A little computation shows that

$$E(uw) = c_i - b_i, \quad E(w^2) = b_i - b_{i+1} + c_{i+1} \quad [20]$$

Hence, from Eq. 17

$$A_i^2 = [c_i - 2(c_i - b_i) + b_i - b_{i+1} + c_{i+1}]/a_i^2 \quad [21]$$

$$= [3b_i - b_{i+1} + c_{i+1} - c_i]/a_i^2$$

We note that as $i \rightarrow \infty$, $A_i^2 \rightarrow 2Ex = 2E\lambda$.

The random variables

$$a_{i,n} = [\text{number of } x_1, \dots, x_n \text{ that are } < i]/n$$

$$b_{i,n} = [\text{sum of those } x_1, \dots, x_n \text{ that are } \leq i]/n \quad [22]$$

$$c_{i,n} = [\text{sum of squares of those } x_1, \dots, x_n \text{ that are } \leq i]/n$$

$$A_{i,n}^2 = [3b_{i,n} - b_{i+1,n} + c_{i+1,n} - c_{i,n}]/a_{i,n}^2$$

converge as $n \rightarrow \infty$ to a_i , b_i , c_i , and A_i^2 , respectively. Hence, from Eqs. 16 and 21 we have the

THEOREM. For any fixed $i = 1, 2, \dots$, as $n \rightarrow \infty$

$$\frac{\sqrt{n}(E_{i,n} - S_{i,n})}{A_{i,n} \cdot N_{i,n}} \rightarrow N(0,1) \text{ in distribution.} \quad [23]$$

As $n \rightarrow \infty$, the probability tends to 0.95 that

$$\frac{E_{i,n}}{N_{i,n}} - \frac{(1.96)A_{i,n}}{\sqrt{n}} < \frac{S_{i,n}}{N_{i,n}} < \frac{E_{i,n}}{N_{i,n}} + \frac{(1.96)A_{i,n}}{\sqrt{n}} \quad [24]$$

The endpoints of the confidence interval [24] for the ratio $S_{i,n}/N_{i,n}$ are functions of $x_1, \dots, x_n$ alone, and do not involve any knowledge of or assumption about the nature of the underlying mixing distribution function G . Of course, for finite n the exact probability that [24] will hold does depend on G .

We can also use $E_{i,n}$ to estimate the parameter (i.e., functional of G)

$$G_i = E(\lambda | x < i) = \frac{\sum_{x=0}^{i-1} f(x)E(\lambda | x)}{\sum_{x=0}^{i-1} f(x)} \quad [25]$$

Since

$$E(\lambda | x) = \frac{\int_0^\infty e^{-\lambda} \lambda^{x+1} dG(\lambda)}{\int_0^\infty e^{-\lambda} \lambda^x dG(\lambda)} = \frac{(x+1)f(x+1)}{f(x)} \quad [26]$$

we can write

$$G_i = \frac{\sum_{x=0}^{i-1} (x+1)f(x+1)}{\sum_{x=0}^{i-1} f(x)} = \frac{\sum_{x=0}^i xf(x)}{\sum_{x=0}^{i-1} f(x)} = \frac{b_i}{a_i} = \frac{Eu}{Ev} \quad [27]$$

Hence, as $n \rightarrow \infty$

$$\sqrt{n} \left\{ \frac{E_{i,n}}{N_{i,n}} - G_i \right\} = \frac{\sum_1^n \left(u_j - \frac{Eu}{Ev} \cdot v_j \right) / \sqrt{n}}{\sum_1^n v_j / n} \rightarrow N(0, B_i^2) \quad [28]$$

in distribution, where by computation

$$B_i^2 = [a_i c_i - 2b_i b_{i-1} + b_i^2]/a_i^3 \quad [29]$$

As $i \rightarrow \infty$, $B_i^2 \rightarrow \text{Var } x = E\lambda + \text{Var } \lambda$. An approximately 95% confidence interval for G_i is

$$\frac{E_{i,n}}{N_{i,n}} \pm \frac{(1.96)B_{i,n}}{\sqrt{n}} \quad [30]$$

where $B_{i,n}^2$ is defined by Eq. 29 with a_i replaced by $a_{i,n}$, etc.

Remarks

(A) The parameter G_i defined by Eq. 25 is equal to $E(y|x < i)$. This follows from [4] and the fact that $E(y|\lambda) = \lambda$.

(B) It is easy to show (Schwarz inequality) that

$$G_1 \leq G_2 \leq \dots \rightarrow E\lambda \quad [31]$$

Since the sample estimates $E_{i,n}/N_{i,n}$ of the G_i are not necessarily increasing in i , this suggests that for some purposes it should be possible to improve on them by smoothing (isotonic regression).

(C) It would be interesting to compare Eq. 28 with the behavior of the maximum likelihood estimator of G_i , either with respect to all possible G s or within some parametric class such as those with the probability density function

$$g_{\alpha, \theta}(\lambda) = \frac{e^{-\lambda/\theta} \cdot \lambda^{\alpha-1}}{\theta^\alpha \Gamma(\alpha)} \quad [32]$$

with two unknown positive parameters α, θ .

(D) Many other quantities can be predicted or estimated from $x_1, \dots, x_n$ by similar methods. For example, to estimate

$$E(\lambda | x \geq i) = E(y | x \geq i) = \frac{E\lambda - b_i}{1 - a_i} \quad [33]$$

we can use

$$\frac{\text{sum of the } x_j\text{'s that are } \geq i}{\text{number of } x_j\text{'s that are } \geq i} \quad [34]$$

As $n \rightarrow \infty$ it can be shown that

$$\sqrt{n}((34) - (33)) \rightarrow N(0, C_i^2) \text{ in distribution} \quad [35]$$

where

$$C_i^2 = [(1 - a_i)(E(x^2) - c_i) - (Ex - b_i)^2]/(1 - a_i)^3 \quad [36]$$

The sequence $E(\lambda | x \geq i)$ is also increasing in i .

(E) Analogous results hold for mixtures of some important parametric families other than the Poisson; e.g., negative exponential, negative binomial, and normal. The binomial case is interesting as an example of nonidentifiability.

This work was supported by the Office of Naval Research Grant N00014-75-C-0560-P00001.

The costs of publication of this article were defrayed in part by the payment of page charges from funds made available to support the research which is the subject of the article. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.