DIGITALCOMMONS
— @WAYNESTATE —

Journal of Modern Applied Statistical Methods

Volume 1 | Issue 2 Article 41

11-1-2002

On the Estimation of Binomial Success Probability With Zero Occurrence in Sample

Mehdi Razzaghi
Bloomsburg University, Bloomsburg, PA

95% conf that $\hat{p}=0$ given by $(.05)^{1/n}$

which is obtained by setting the probability of observing n failures equal to 0.5 and solving for p. Bailey noted that this estimator is nearly identical to the median of the Bayesian posterior distribution for p, derived with respect to a uniform distribution using the absolute error loss (AEL) function.

The problem of Bayesian estimation of p with respect to the more general class of a conjugate beta prior distribution but using the squared error loss (SEL) was considered by Basu et at. (1996). By comparing (2) with a few other estimates, Bailey (1997) concluded that $\hat{p}$ performs relatively well in practice and can be used in certain circumstances. It is also worth noting that because the upper 100(1 - α)% confidence limit for p is (see Bickel & Doksum, 2001) given by

$$u = 1 - \alpha^{1/n}$$

then (2) can be interpreted as the median of the sampling distribution of the random variable X/n. Moreover, as mentioned in Louis (1981), u may be thought of as the proportion of the number of successes in a future experiment of the same size and it is the upper 100(1 - α)% Bayesian prediction interval based on a uniform prior distribution.

In this paper, the problem of point estimation of p when a sample shows no occurrence is considered from a more general viewpoint. Several potential estimates based on statistical methods in addition to those suggested in Bailey (1997) and Basu et al. (1996) will be proposed and their properties will be discussed. Next, I review the Bayesian approach and consider the use of other loss functions, and then discuss the properties of an estimate derived from information theory. The next section is devoted to the discussion of a decision theoretic approach for estimating p, and the use of minimax estimation of p is considered. In the final section of this article, I give an example from teratology to provide further illustration of the results.

Bayesian Estimation

It is well known that when the prior distribution of p belongs to the family of a beta distribution β(a, b),

$1 - CL_{\hat{p}} = \alpha^{\frac{1}{n}}$ $\quad \ln(1 - CL_{\hat{p}}) = \alpha\left(\frac{1}{n}\right)$

$n = \frac{\alpha}{\ln(1 - CL_{\hat{p}})}$ $\quad CL_{\hat{p}} = .01$

$\alpha = .01$

$n =$

$$g(p) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1} \quad a, b > 0, 0 < p < 1 \tag{3}$$

where

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

then the posterior distribution of p belongs to the beta family β(a + x, b + n − x) and the Bayes estimate p^* of p based on the SEL function $L(p,p^*) = (p - p^*)^2$, is given by (Basu et al., 1996)

$$p^* = \frac{(a+x)}{(a+b+n)} \tag{4}$$

Thus, if x = 0, then the Bayes estimator for a zero occurrence is

$$p^* = \frac{a}{a+b+n} \tag{5}$$

and in particular if a = b = 1, then the Bayes estimator under a uniform prior is derived. Also, when Jeffreys' non-informative prior, for which a = b = 0.5 is used, then the Bayes estimator of no response is given by

$$p^*_{ni} = \frac{1}{2(n+1)} \tag{6}$$

Basu et al. (1996) compared (5) and (6) with the classical approach based on upper confidence limits and conclude that the Bayes estimate under an informative prior is best. Both estimates (5) and (6), however, are derived using the SEL function which is but one of several possible loss functions that may be used to derive the Bayes estimate of p. In practice, there are many instances that other functions may be preferred.

Actually the SEL is a special case of a larger class of weighted quadratic loss functions

$$L(p, p^*) = w(p)(p - p^*)^2 \tag{7}$$

U = upper 100(1−α)% conf limit on $\hat{p}=0$

~~we want~~ ~~U = .01~~

$CL_{\hat{p}} = 1 - \alpha^{\frac{1}{n}}$

which amounts to no loss if the estimate p^* is within a distance ε from p. For this loss function, the expected posterior is given by

$$P(|p-p^*|) > \varepsilon \mid x) = 1 - P(|p-p^*| \le \varepsilon \mid x).$$

Consequently, if a modal interval of length 2ε is defined as an interval with center at the mode of the distribution, then as $\varepsilon \to 0$, the Bayes estimate with respect to the zero-one loss approaches the mode of the posterior distribution, provided that a mode exists. This in turn implies that the Bayes estimate in this case becomes the maximum likelihood estimate.

Maximum Information Estimation

Good(1965) and Typlados and Brimley (1962) showed that Shannon's information content of the observation x from the binomial distribution (1) is given by

$$I(p) = -p\ln(p) - (1-p)\ln(1-p) + \ln\left[\binom{n}{x} p^x (1-p)^{n-x}\right] \tag{15}$$

By maximizing I(p), one obtains the maximum information (MIE) estimate p_{MIE} of p as the solution of the equation

$$\ln\left(\frac{p}{1-p}\right) = \frac{x}{p} - \frac{n-x}{1-p} \tag{16}$$

In particular when $x = 0$, the MIE of p is the solution of

$$\sqrt[n]{\frac{p}{1-p}} = \exp\left(-\frac{1}{1-p}\right). \tag{17}$$

Chew (1971) pointed out that for $n > 7$, the solution of (17) is up to 3 decimals equal to zero and, once again, it is seen that this method fails to produce a reasonable estimate for p.

Minimax Estimation

The minimax criterion stems from the general theory of two-person zero-sum games of von Neuman and Morgenstern (1944). Loosely, instead of averaging the risk as in Bayesian estimation, one looks at the least favorable scenario for each decision, that is the worst possible risk for that decision, and chooses a decision which gives the least value of the worst risk. Thus, the minimax rule minimizes the maximum risk. Although the methodology ignores all references to prior knowledge, but in the absence of any information regarding p, the minimax estimator provides a Bayesian estimate without knowing the prior distribution. As pointed out by Cox and Hinkley (1974), the minimax rule is defensible when the risk is small, since it ensures that, whatever the true parameter value, the expected loss is small. Although there may be an apparently better rule, any improvement can only be small and may carry with it the danger of a seriously bad performance for some values of the parameter.

Now, for the binomial parameter p in (1), it can be shown that the minimax decision rule, based on the SEL function, is given by (Bickel and Doksum, 2001)

$$\tilde{p} = \frac{x + \sqrt{n}/2}{n + \sqrt{n}} \tag{18}$$

with variance bounded by

$$v = \left[2(1+\sqrt{n})\right]^{-2} \tag{19}$$

The minimax estimator (18) is Bayes with respect to a beta prior with parameters

$\sqrt{n}/2$ and $\sqrt{n}/2$. If $x = 0$, then from (18),

$$\tilde{p} = \left[2(1+\sqrt{n})\right]^{-1} \tag{20}$$

which can be used to estimate the probability of a rare event. In order to compare the minimax estimator given in (20) with those considered in Bailey (1997), $\tilde{p}$ was evaluated for several values of n. Table 2 presents these numerical values, where for comparison, the values of $\hat{p}$ in (2), the estimator suggested by Bailey and the Bayes estimator p^*_{ni} based on a noninformative prior given in (6) are also included. As the sample size chemical hormesis we refer to Calabrese and Baldwin (2000). However, as shown in Razzaghi and Loomis (2001), in developmental toxicology, more than a single replication of an experiment must be considered before a chemical can be declared as being hormetic. For the present data, therefore, in order to fit a monotonic dose-response function, one might consider replacing the observed incidence of zero by an estimate of it. In such a situation, it would seem unreasonable to estimate the probability of response in the 5 mg/kg dose group as 0, as given by the maximum likelihood method. In this case, because n = 98, from (2), (6), (14) and (20),

$$\hat{p} = .007, \quad p^*_{ni} = .005, \quad p^*_{1,ni} = .021, \quad \tilde{p} = 0.046$$

are four different point estimates for the probability of response at the first nonzero dose level.

In order to further investigate the properties of these estimates, a probit model was used to fit the response probability p as a function of the natural logarithm of dose, i.e.

$$p = \Phi(a + b \log d) \tag{23}$$

Using PROC PROBIT in SAS (1996), it was found that the maximum likelihood estimates of the model parameters are $\hat{a} = 03.601$ and $\hat{b} = 0.987$. Using these parameter estimates, it is found that the point estimate of p when d = 5 mg/kg is .022. Furthermore, the standard deviation of $\hat{a} + \hat{b} \log 5$ is 0.163. Based on these quantities, if the 95% confidence interval is evaluated for the predicted proportion, one finds that this range is (.010, .046). Interestingly, although the minimax estimator $\tilde{p}$ is equal to the upper bound in this range, both the Bailey estimator $\hat{p}$ and the Bayesian estimator p^*_{ni} are outside this range and far too small to be plausible. Therefore, in this instance, $p^*_{1,ni}$ and the minimax procedure appear to produce more realistic estimates of p compared to other methods.

Discussion

Lack of occurrence of rare events in biological and physical experiments is not uncommon. In such situations, the maximum likelihood estimate becomes unusable and one needs to resort to alternative statistical methods. Here, I have considered this problem and investigated the use of several other statistical techniques and the minimax estimator.

It is immediately noted from (2) that for the Bailey estimator, $\hat{p} = 0\left(\frac{1}{n}\right)$. This property also holds for the Bayesian estimator considered by Basu et al. (1996). However, for the minimax estimator, from (18) $\tilde{p} = 0\left(\frac{1}{\sqrt{n}}\right)$. This means that for relatively large values of n, both $\hat{p}$ and the Bayes estimate lead to numerically smaller values than the minimax estimator. Actually, it can be shown (Roussas, 1997) that the Bayes estimate for the family of beta prior and SEL has the same asymptotic distribution as the maximum likelihood estimate for arbitrary fixed values of α and β, while the asymptotic distribution of $\sqrt{n}(\tilde{p} - p)$ is normal with mean $\frac{1}{2} - p$ and variance $p(1-p)$. Thus, I can say that the minimax estimator is comparatively more conservative.

However, as discussed by Carlin and Louis (1996), although informative priors enable more precise estimation, extreme care must be taken in their use because they also carry the risk of disastrous performance when their informative content is in error. Although using a non-informative prior leads to a more conservative Bayes estimate, there may be situations when Bayes and other methods underestimate the value of this rare event. This result is demonstrated through an example in developmental toxicology.

The conclusion of this paper is not necessary to recommend the minimax or any other estimator in all situations when there is a zero response. Rather, the goal is to increase awareness and recommend that more caution should be taken when any single method is used to estimate the success probability when sample shows zero occurrence. The choice of the estimate should to a large extent depend on which kind of optimality is judged to be most appropriate for the case in question.