Canadian Nuclear Safety Commission | Commission canadienne de sûreté nucléaire

CMD 21-H113.A

File/dossier : 6.01.07
Date : 2021-11-12
e-Docs pdf : 6679990

Supplementary Information

Written submission from CNSC Staff

In the Matter of

Request for authorize Bruce Power Inc. to restart Bruce Nuclear Generating Station A Unit 4 and Bruce NGS B Units 5, 7, and 8 following future outages

Public Hearing - Hearing in writing based on written submissions

November 2021

Renseignements supplémentaires

Mémoire du Personnel de la CCSN

À l’égard de

Demande de Bruce Power Inc. afin d’obtenir l’autorisation de redémarrer la tranche 4 de la centrale nucléaire deBruce-A et les tranches 5, 7 et 8 de Bruce-B après tout arrêt futur

Audience Publique - Audience fondée sur des mémoires

Novembre 2021

Canada

Estimating Event Probabilities using Zero Failure Data

Karl Breitung
Researcher, TU Munich, Germany

Marc A. Maes
Professor, University of Calgary, Canada, and UGent, Belgium

ABSTRACT: In a rather well publicized June 2013 CBC broadcast, the safety of a specific series of Canadian onshore gas pipeline joints was declared "absolute" by one proponent, since, historically, no incidents or failures had ever been reported for that site. An opponent then argued that the face value of this risk could never be zero, but "very small". The objective of the present paper is to review just how small, and at which confidence level, one can sensibly consider the actual incidence rate to be. A comparison of the most popular approaches and a comprehensive test for consistency, point to the superiority of the Bayesian estimator together with a non-central posterior probability interval.

1. INTRODUCTION

A common scenario in the context of data mining and explorative statistics of accidents/failures, is that the analyst runs into data segments or subsets that have zero reported incidents. Is it legitimate to construct, with a specified confidence, an upper limit p_U for which we can claim that the "true" incident probability p will not exceed this value?

A typical example concerns pipeline rupture and leak statistics (TSB, 2013): out of thousands of pipeline joints n a large majority of joints are observed to have no recorded incidents. There is often confusion about which confidence intervals apply in a case like this. Some analysts replace the zero incident case $X = 0$ by $X = 1$ because it is "easier" to analyze and because it is after all "conservative".

Others suggest that if, in the case of a binomial model, n consecutive trials have not resulted in any failure ($X = 0$), then we should prudently assume that the (n+1)-th will; subsequently this pessimistic estimate is used for prediction.

2. ESTIMATION FOR ZERO INCIDENTS

One fundamental problem is related to the setting of the problem and to its formulation. In this paper we consider a binomial model originating from a setup where n units are monitored and incidents are observed within each unit. It also applies to the case where n sections of a continuous system such as a pipeline are tested and X are found to be faulty. The analysis for a Poisson model is entirely similar.

Suppose we are to estimate for a sequence of n independent Bernoulli trials, the probability of failure p if X failures have been observed in this sequence. The (frequentist) estimator for p is:

$$\hat{p} = \frac{X}{n} \tag{1}$$

which suggests that, if no failures were observed, the estimator is zero.

In a Bayesian setting, one starts from a prior distribution for the parameter p. Usually, the conjugate prior Beta distribution f_{pr} is used with pdf:

$$f_{\mathrm{pr}}(p) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1} \tag{2}$$

where $B(a,b) = (\Gamma(a)\Gamma(b))/\Gamma(a+b)$ is the Beta function. The random variable with this pdf has a mean equal to $a/(a+b)$ and a standard variance equal to $ab/(a+b)^2(a+b+1)$. Its shape is determined by the parameters a and b. Taking $a = b = 1$ gives a uniform distribution over the unit interval. Sometimes also the Jeffrey's prior with

$a = b = 0.5$ is used, but here we will consider only the uniform prior, since it has some optimality properties as shown in the following. From this we then calculate the posterior using the likelihood of observations:

$$\ell(X|p) = \binom{n}{X} p^X (1-p)^{n-X} \tag{3}$$

yielding:

$$f_{\text{post}}(p|X,n) = \\ = \frac{1}{B(X+1,n-X+1)} p^X (1-p)^{n-X} \tag{4}$$

The Bayes estimator for a parameter is the mean of the posterior distribution f_{post}. Then the Bayes estimator for the probability p is in the case $X = 0$:

$$\hat{p} = \frac{1}{n+2} \tag{5}$$

These results can be found in any standard text about Bayesian methods, e.g. in Press (1989), p. 40.

So in the Bayesian setting the estimator is not equal to zero for $X = 0$ as in the frequentist case.

The next issue concerns the derivation of sensible confidence intervals for the probability p. This poses also significant problems in the case $X = 0$.

3. THE CLOPPER-PEARSON CONFIDENCE INTERVALS

The standard method for determining such intervals is described in detail in Clopper and Pearson (1934). They derive confidence intervals for the probability p in a frequentist setting. Unfortunately, this procedure does not give very satisfactory results for the case $X = 0$. In this case the confidence interval for the level α is in fact a confidence interval for the level $\alpha/2$, i.e. it is too large.

The original objective (Clopper and Pearson, 1934) is to calculate for a given confidence level α and an observed number of failures X as the lower bound p_L for the confidence interval, the value p_L for which:

$$\sum_{j=X}^{n} \binom{n}{j} p_L^j (1-p_L)^{n-j} = \frac{(1-\alpha)}{2} \tag{6}$$

and as the upper bound p_U the value for which:

$$\sum_{j=0}^{X} \binom{n}{j} p_U^j (1-p_U)^{n-j} = \frac{(1-\alpha)}{2} \tag{7}$$

One has for the sums using integration by parts:

$$\sum_{j=X}^{n} \binom{n}{j} p^j (1-p)^{n-j} = \\ = \int_0^p \frac{t^{X-1}(1-t)^{n-X}}{B(X,n-X+1)} dt \tag{8}$$

as well as

$$\sum_{j=0}^{X} \binom{n}{j} p^j (1-p)^{n-j} = \\ = \int_p^1 \frac{(1-t)^{n-X-1} t^X}{B(X+1,n-X)} dt \tag{9}$$

So both bounds fulfilling the equations above can be found by inverting the incomplete Beta function with the respective parameters k and $n - k + 1$.

The authors do not consider in the paper the case of $X = 0$ separately, but from the diagrams one sees that the lower bound is simply set to zero, since the sum on the left side of Eq. (6) is equal to unity, so the equation cannot be fulfilled. Therefore the confidence level is determined by the second equation only resulting in an interval with confidence level $\alpha/2$. So its probability content is larger and it is longer.

The Bayesian probability interval $[p_L, p_U]$ for a specified “confidence” level α can be retrieved from the posterior pdf (4):

$$\Pr(p_L < p < p_U|X,n) = \\ = \int_{p_L}^{p_U} f_{\text{post}}(p|X,n) dp = \alpha \tag{10}$$

For $X = 0$, the lower bound p_L can be set to zero, without distorting the highest posterior density interval too much, so that p_U can be found from inverting the incomplete beta pdf with parameters $X + 1$ and $n - X + 1$ (Box and Tiao, 1992):

$$\int_0^{p_U} f_{\text{post}}(p|X,n)dp =$$
$$= \int_0^{p_U} \frac{p^X(1-p)^{n-X}}{B(X+1,n-X+1)}dp = \alpha \qquad (11)$$

4. COMPARISON OF METHODS

In Pires and Amado (2008) a whole menagerie of about twenty different methods for calculating confidence intervals for the binomial proportions are presented. They are then compared with respect to criteria such as mean coverage probability and expected length.

For a given method with number *i*, which produces the confidence interval $[L_i(j), U_i(j)]$ for *j* successes in *n* experiments, the coverage probability CP for given *n* and given *p* is given by the summation:

$$\text{CP}(p,n,i) =$$
$$= \sum_{j=0} \binom{n}{j} p^j (1-p)^{n-j} I_{[L_i(j),U_i(j)]}(p) \qquad (12)$$

with I_A the indicator function for the interval *A*. For a fixed *n*, the mean coverage probability is found as:

$$\mathbb{E}\big(\text{CP}(n,i)\big) = \int_0^1 \text{CP}(p,n,i)dp \qquad (13)$$

Then the expected length of an interval calculated with method *i*, given *p* and *n* is:

$$\text{E}\ell(p,n,i) =$$
$$= \sum_{j=0} \binom{n}{j} p^j (1-p)^{n-j} \big(U_i(j) - L_i(j)\big) \qquad (14)$$

Integrating over the parameter *p* gives then the overall expected length:

$$\mathbb{E}\big(\text{E}\ell(p,n,i)\big) = \int_0^1 \text{E}\ell(p,n,i)\, dp =$$
$$= \sum_{j=0} \binom{n}{j} \int_0^1 p^j (1-p)^{n-j} \big(U_i(j) - L_i(j)\big)\, dp$$
$$(15)$$

The coverage probabilities CP are shown for $n = 40$ and $n = 400$ in Pires and Amado (2008). These functions show quite erratic behavior as function of *p*.

The twenty methods given in Pires and Amado (2008) are compared considering mean coverage probability, maximum and minimum coverage probability and then also mean, minimum and maximum length of the intervals. A method which is optimal for all possible criteria cannot be found, so mean coverage probability and mean length seem to be the most useful ones. The mean coverage should be as close as possible to the nominal coverage for which the intervals are constructed and the length of the intervals should be minimal.

Consider now these two quality criteria for possible confidence intervals:

1. the mean coverage probability is to be as near as possible to the nominal coverage probability.
2. its length should be as short as possible.

Due to the definition of the Bayesian confidence interval its mean coverage probability is always equal to the nominal. Considering the second criterion of mean length the Bayesian highest posterior density interval is always optimal, see p. 190 in Pires and Amado (2008).

So the Bayesian approach using a posterior based on a uniform prior distribution gives “optimal” intervals if the criteria selected are the mean coverage probability and the mean length of the interval.

In their conclusions Pires and Amado (2008) consider only central intervals thereby excluding the Bayesian intervals citing computational challenges related to non-central intervals. The exclusion of non-central intervals is not really justifiable, since nowadays the computational effort in computing such intervals has largely become irrelevant. For these non-central intervals no clear optimal methods are singled out.

5. NEED FOR CONSISTENT METHODS

Consider once more the case where no failure has been observed till now. Let us assume two different methods for estimating the probability of such a failure have been applied, resulting in two different estimates of the probability:

$$\hat{p}_1 > \hat{p}_2 \tag{16}$$

From this follows that we can use for estimate of *p*:

$$p \approx \hat{p}_1 \tag{17}$$

$$p \approx \hat{p}_2 \tag{18}$$

Subtracting the equations and dividing by $\hat{p}_1 - \hat{p}_2$ gives:

$$0 \approx 1 \tag{19}$$

So one can derive from such inconsistent estimation methods false statements and *ex falso sequitur quodlibet*. How can we improve this situation?

There have been proposals to improve the estimator in the case of zero observations by adding "virtual" observations. For example, if in *n* trials no failure has been observed, then we might assume that the first failure is just around the corner, i.e. it happens at the next trial. So the estimator for *p* is then:

$$\hat{p}_1 = \frac{1}{n+1} = \frac{1}{n} - \frac{1}{n(n+1)} \tag{20}$$

But for large *n* this is approximately equal to 1/*n*, so we here give almost the same probability weight to a case where no failures have been observed as to a case where one failure has been observed in *n* trials for which the estimator is:

$$\hat{p}_2 = \frac{1}{n} \tag{21}$$

If this is now plugged into a series system with *k* identical components, we would have as estimate for the failure F_k of the system in case 1:

$$\widehat{\Pr(F_k^1)} \approx \frac{k}{n+1} = \frac{k}{n} - \frac{k}{n(n+1)} \tag{22}$$

whereas in the second case:

$$\widehat{\Pr(F_k^2)} \approx \frac{k}{n} \tag{23}$$

which for large *n* is almost the same. So for small *k* we see that the probability of failure here is estimated as almost the same in spite of the fact that we have no failure observed in the first case. This leads to an unjustified and "conservative" shift of resources.

The only reasonable recourse is to apply a fully Bayesian setting (Eqs. (2) to (5) above). Furthermore, one should also consider the actual context of the problem. For instance, if the data base is spatial in its structure – as in the case of pipeline incidents Canada-wide (TSB, 2013) – then a spatially hierarchical analysis can be performed where evidence of reported incidents is to a certain optimal extent "shared" over all assets resulting in a more relaxed treatment of "zero" incidents.

6. REFERENCES

Box G.E.P. and Tiao G.C. (1992) "*Bayesian Inference in Statistical Analysis*", Wiley, New York, 1992.

Clopper C.J. and Pearson E.S. (1934) "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial", *Biometrika*, 26:404–423, 1934.

Pires A. and Amado, C. (2008) "Interval Estimators for a Binomial Proportion: Comparison of Twenty Methods", *REVSTAT – Statistical Journal*, 6(2):165–197, June 2008.

Press S.J. (1989) "*Bayesian Statistics: Principles, Models, and Applications*", Wiley, New York, 1989.

TSB (2013) "*Statistical Summary Pipeline Occurrences 2013*", Transportation Safety Board of Canada, http://www.tsb.gc.ca

Risk Analysis, Vol. 31, No. 7, 2011 DOI: 10.1111/j.1539-6924.2010.01568.x

Estimating the Probability of Rare Events: Addressing Zero Failure Data

John Quigley[1,*] and Matthew Revie[1]

Traditional statistical procedures for estimating the probability of an event result in an estimate of zero when no events are realized. Alternative inferential procedures have been proposed for the situation where zero events have been realized but often these are *ad hoc*, relying on selecting methods dependent on the data that have been realized. Such data-dependent inference decisions violate fundamental statistical principles, resulting in estimation procedures whose benefits are difficult to assess. In this article, we propose estimating the probability of an event occurring through minimax inference on the probability that future samples of equal size realize no more events than that in the data on which the inference is based. Although motivated by inference on rare events, the method is not restricted to zero event data and closely approximates the maximum likelihood estimate (MLE) for nonzero data. The use of the minimax procedure provides a risk adverse inferential procedure where there are no events realized. A comparison is made with the MLE and regions of the underlying probability are identified where this approach is superior. Moreover, a comparison is made with three standard approaches to supporting inference where no event data are realized, which we argue are unduly pessimistic. We show that for situations of zero events the estimator can be simply approximated with $\frac{1}{2.5n}$, where n is the number of trials.

KEY WORDS: Binomial; minimax; zero event

1. INTRODUCTION

as low as reasonably practicable

Point estimates of the probability of rare events inform risk mitigation strategies and ALARP analysis.(1) Classical inferential approaches to estimating the probability of events occurring, namely, the ratio of occurrences to opportunities, provides an optimistic estimate when zero events have been realized. The resulting point estimate of 0 for the probability of an event being realized is problematic for risk analysis as it removes all resulting consequences from a risk analysis study. As such, alternative inference procedures have been proposed.

Using expert judgment for supporting inference is appropriate if expertise is available (see Ref. 2 for review). However, there are two shortcomings with this approach. First, for low-frequency events, providing a subjective probability distribution on the probability of an event being realized requires a level of precision to be meaningful that is cognitively burdensome. Second, to utilize a Bayesian inference mechanism requires the prior distribution to be elicited with no knowledge of the data upon which the prior assessment will be updated. If the prior has been elicited at the point when the inference is required, there can be no updating. In this case, the inference must rely entirely on the subjective assessment. Expert judgment can provide meaningful insight but with such shortcomings for rare events, data-driven methods are also desirable.

[1] Department of Management Science, University of Strathclyde, Glasgow G1 1QE, Scotland.

*Address correspondence to Dr. John Quigley, Department of Management Science, University of Strathclyde, Glasgow G1 1QE, Scotland; tel: +44-141-548-3152; j.quigley@strath.ac.uk.

0272-4332/11/0100-1120$22.00/1

There have been several data-driven approaches proposed for assessing the probability of an event being realized given zero events have been realized. However, each suffers from *ad hockery*, in so far as they are proposed to be applied given the situation that zero events have been realized. This presents two shortcomings hampering the confidence of the decisionmakers. First, switching between methods depending on the data can result in incoherent inference. Second, without a transparent methodological basis for an approach, it cannot be adequately critiqued. As such, they require more faith on the part of the decisionmakers.

In Section 2, we provide a brief summary of the data-driven methods for inference proposed for background. Section 3 provides a summary of the minimax criteria for inference, which will form the basis of the proposed method within this article. Section 4 provides the derivation of the proposed approach. Section 5 provides a comparison with the maximum likelihood estimate (MLE) whereas in Section 6 a comparison is provided with alternatives given zero events are realized. Section 7 provides a summary and conclusion.

2. BACKGROUND

There are alternative methods to the MLE, that is, the ratio of events realized to opportunities, that have been proposed for the situation of zero event data and within this section we provide a brief critique. A common criticism we have with many of these approaches is that they have been developed for the situation of zero event data only and as such would require the analyst to switch to an alternative procedure if at least one event were realized. Some approaches are derived with a Bayesian methodology using an uninformative prior distribution, typically a uniform distribution over [0,1] on the probability of an event being realized. However, such an assumption is questionable, particularly in low-frequency events, resulting in overly conservative estimates as the estimate is pulled toward the mean of the prior, that is, $\frac{1}{2}$.

The International Standard IEC[3] propose a lower one-sided confidence limit for zero event data under an exponential distribution. Confidence intervals are derived from the principle that the data are random and once realized will be used within a formula to provide an interval. This interval will contain the true underlying parameter value with a given probability. The intervals are not known *a priori*. However, if we only apply this procedure when we have zero event data then the intervals are always known *a priori*. As such it is not clear what criteria are being used to derive this estimate and its meaning is questionable.

Bailey[4] presented a review of five methods for inferring the probability of an event given zero events have been realized. Two were based on confidence interval methodology and one on using a uniform prior, which have similar criticisms as previously discussed. One proposed using the MLE assuming there had been one event realized, thus providing an overestimate, but without any rigorous derivation we are not provided with guidance as to the extent of the pessimism. The approach recommended was Equation (1):

$$\delta_{\mathrm{B}} = 1 - 0.5^{\frac{1}{n}}, n = 1, 2, \ldots \tag{1}$$

This method is applied in explosives testing for zero event data (Ref. 5, as cited by Ref. 4). The reasoning underpinning this approach is that given zero events have been observed in a sample size of n, the probability that this observed outcome would be realized as opposed to all other outcomes is assigned a probability of 0.5. The purpose of Bailey's review was to compare Equation (1) with the four alternative methods. He concluded that there was little difference in the resulting estimates and that Equation (1) could be viewed as an approximation of the median of the posterior distribution obtained through applying Bayesian methodology with an uninformative prior. In concluding, he generalized Equation (1) to facilitate using alternative values rather than 0.5.

A shortcoming with Equation (1) is that the value 0.5 is obtained through a belief that the probability of realizing the same outcome from a repeated experiment is 0.5. Although we can substitute different probability values, we are applying a subjective probability approach to addressing this situation and although the use of expert judgment is laudable it does mean that it is not a data-driven approach. Moreover, its evaluation should be concerned with an accompanying elicitation protocol.

An alternative method for estimating the probability when zero events have been realized is the Rule of Three also known as Hanleys Rule,[6,7] which recommends Equation (2) as an estimator:

$$\delta_{\mathrm{Rof3}} = \frac{3}{n}, \quad n \geq 30. \tag{2}$$

This approach can be derived as an approximation to a Bayesian approach where a uniform prior is assumed for the underlying probability and the 95th percentile is derived from the posterior distribution, as well as from a frequentist approach, where the probability of observing the data realized is assigned 0.05, both resulting in Equation (2) as an estimate of the underlying probability. As such, this approach is used to obtain an upper bound of the probability and as a point estimate will be overly conservative.

The notion of developing inference on the probability of the same data being realized in a repeated experiment was also considered more rigorously by Dewoody *et al.*[8]. The underlying method in this approach was to apply Bayesian methodology with an uninformative prior. We will develop an inferential procedure from a frequentist perspective for inferring the probability of no more events being realized in a repeated experiment of identical size than that realized in the experiment upon which the inference is based. As such we will not restrict our study to zero events and develop procedures for the entire sample space.

3. REVIEW OF MINIMAX ESTIMATION

We concern this study with estimating the probability of an event being realized from data generated from a binomial distribution, which assumes that there will be n independent trials or opportunities for the event to be realized, each statistically independent and identically distributed with probability p. We will develop our procedures about a quadratic loss function and as such the accuracy of the estimate will be measured by the risk function of Equation (3):

risk function

$$R(g(p), \delta) = E[(g(p) - \delta(X))^2], \tag{3}$$

where p is the probability of an event being realized, $0 \leq p \leq 1$; $g(p)$ is the real-valued function whose value at p is to be estimated, that is, the estimand; n is the sample size, that is, number of observations; X is the random variable taking on values in a sample space according to a binomial distribution with cumulative distribution function (CDF) $F(X; n, p)$; and δ is a real-valued function of the sample space used to estimate $g(p)$.

If our criteria was to choose a procedure, that is, δ, to minimize the expected risk function for estimating the probability of failure, that is, p, then $g(p)$ would equal p and δ could be shown to equal the MLE Equation (4):

$$\delta_{\text{MLE}}(X) = \frac{X}{n}. \tag{4}$$

The risk function, Equation (3), for this situation can be shown to be Equation (5). This corresponds to the variance of the MLE as it is an unbiased estimate of p:

For p=0.5, R=1/(4n)

$$R(p, \delta_{\text{MLE}}) = \frac{p(1-p)}{n}. \tag{5}$$

Upon inspection of Equation (5), we see that the risk is not constant for all p. This risk is maximized at 0.5.

The minimax estimator can be obtained through applying a more risk adverse criteria. For this approach we seek the estimator with the best worst case. Formally, we express this in Equation (6):

$$\delta_{\text{MM}} = \arg\min_{\delta} \max_{p} R(g(p), \delta). \tag{6}$$

The minimax estimator for the parameter p can be show to be:[9]

$$\delta_{\text{MM}}(X) = \frac{X}{n}\frac{\sqrt{n}}{1+\sqrt{n}} + \frac{1}{2}\frac{1}{1+\sqrt{n}}. \tag{7}$$

The risk function for this situation can be shown to be Equation (8):

$$R(p, \delta_{\text{MM}}) = \frac{1}{4}\frac{1}{(1+\sqrt{n})^2}. \tag{8}$$

Comparing Equation (5) with Equation (8) it can be shown that the minimax procedure outperforms the MLE around $p = 0.5$. However, the minimax introduces a bias, pulling the estimate toward 0.5. In addition, for large sample sizes the improvement achieved in this region, i.e., 0.5, is negligible. Although the minimax procedure would overcome the problem of a non-zero estimate when confronted with the zero event data issue, the performance is poor with small values of p. *

Encouraged with its ability to outperform the MLE in certain regions as well as maintaining a risk adverse inferential procedure we will develop a minimax approach for estimating the value of the CDF rather than the probability of an event being realized and determine the value of p associated with the estimate of the CDF.

4. MINIMAX FOR THE CDF

In keeping with the work of Dewoody *et al.* we will initially seek to estimate the probability of no

* $1+2n^{0.5} > 0$

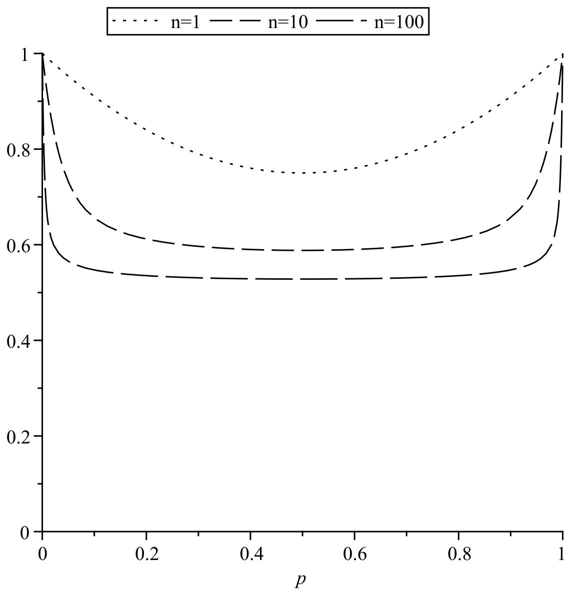

Fig. 1. Expectation of the CDF of a binomial showing sensitivity to changes when p is small.

more than the observed number of events being realized, that is, CDF. An important distinction is that we do not condition on there being zero events realized in the data. As such we are not considering the situation of zero failures; however, we are motivated by high-reliable/low-frequency events/items.

Pursuing criterion of minimizing the expected risk for the CDF would result in using the expected CDF as the estimate. Fig. 1 illustrates the expectation of the CDF for $n = 1, 10, 100$. We see this varies with p, which means that we need to know p before we could estimate it. Asymptotically, these curves approach 0.5, which would result in an estimate for the underlying probability as with Equation (1). However, the convergence is slow and we are concerned with low-frequency events. Note that for small values of p, the mean of the CDF is high and sensitive to changes in p.

The approach we propose to estimate p is to first estimate the CDF associated with either X or $(n - X)$, depending on which is smaller. This is to avoid the situation where all events have been realized as the CDF is 1. The resulting risk function is expressed in Equation (9):

$$R = E\left[\left(F(Y;n,p) - \hat{F}_0\right)^2 \middle| Z = X\right] E[I(Z = X)] + E\left[\left(F(Y;n,1-p) - \hat{F}_1\right)^2 \middle| Z = n - X\right] \times E[1 - I(Z = X)] , \quad (9)$$

where

$$Z = \min(X, n - X),$$

$$I(Z = X) = \begin{cases} 1, & \text{if } Z = X, \\ 0, & if\ Z \neq X, \end{cases}$$

$$F(Y;n,p) = \sum_{i=0}^{Y} \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i},$$

$$0 \le p \le 1,\ n \in \mathbb{N},\ Y = 0, 1, \ldots, n.$$

We then apply a minimax criteria to R for the estimators $(\hat{F}_0, \hat{F}_1)$ and solve for p.

For the remainder of the article we focus on $\hat{F}_0$; however, all of the analysis can be easily adapted to be expressed for $\hat{F}_1$. Fig. 2 illustrates the relationship between the underlying parameter p, candidate estimates of the CDF, that is, $\hat{F}_0$, and the risk function for a sample of 10 for the situation where $0 \le X \le [\frac{n}{2}]$, that is, $Z = X$. This figure shows that for values of p at 0 the risk is very high with a low $\hat{F}_0$, but low when $\hat{F}_0$ is high. Alternatively, when p is 0.5 a high value of $\hat{F}_0$ results in a large risk. The risk associated with high values of p are very low, reflecting a low probability of realizing a number of events in the interval $0 \le X \le [\frac{n}{2}]$.

The risk when p equals 0 is $(1 - \hat{F}_0)^2$, which is a decreasing function of $\hat{F}_0$. When we consider values of p that are nearer to 0.5, we see from Fig. 2 that the risk has a local minimum about $\hat{F}_0$ equal to 0.4. For a minimax criteria we seek a value of $\hat{F}_0$ that minimizes the maximum risk. We find such a value by equating the risk when p equals 0 to the risk when p equals values greater than 0 and solve for $\hat{F}_0$. An expression for this is provided in Theorem 1, as well as showing that the minimax estimate will be the minimum of these iso-risk points with respect to p. This is due to the risk continuously decreasing as a function of $\hat{F}_0$ when p equals 0 so after it hits its first iso-risk point, the risk associated with that particular value of p will be increasing as a function of $\hat{F}_0$ so the maximum risk will also be rising. In Theorem 2, we show that as the sample sizes increases toward infinity the minimax estimator of the CDF approaches $\frac{2}{3}$.

THEOREM 1: *The minimax estimate of the CDF of the binomial distribution evaluated at a random event X is:*

$$\hat{F}_0 = \min_p \hat{F}_p,$$

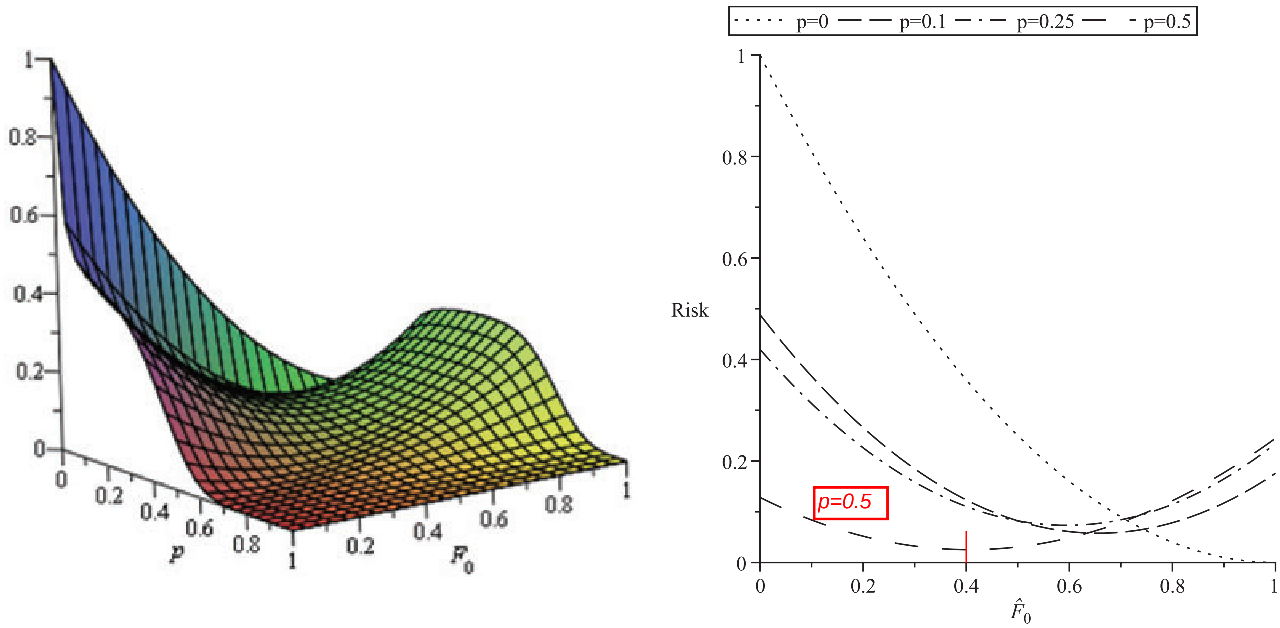

Fig. 2. Illustrates the risk in relationship to the underlying probability p and the estimate of the CDF $\hat{F}_0$ for a sample size of 10 showing a minimum maximum risk at $\hat{F}_0$ equal to 0.7.

where

$$\hat{F}_p = \frac{A - B}{C}, \tag{10}$$

$$A = 1 - \sum_{y=0}^{[\frac{n}{2}]} F(y;n,p)\, f(y;n,p)$$

$$B = \sqrt{\left(1 - \sum_{y=0}^{[\frac{n}{2}]} F(y;n,p)\, f(y;n,p)\right)^2 - \left(1 - F\left(\left[\frac{n}{2}\right];n,p\right)\right)\left(1 - \sum_{y=0}^{[\frac{n}{2}]} \left(F(y;n,p)^2\right) f(y;n,p)\right)}$$

$$C = 1 - F\left(\left[\frac{n}{2}\right];n,p\right).$$

Proof. See Appendix A.

Using numerical methods we evaluated the minimax estimate $\hat{F}_0$ for sample sizes from 1 to 50. These are presented in Table I.

Upon inspection, there are four small but curious features of the estimates. At sample sizes 9, 11, 13, 15, and 17, $\hat{F}_0$ experiences a small local maximum. This is very slight, not even noticeable to four decimal places for sample sizes 15 and 17.

We see from Theorem 2 that the minimax estimate of the CDF, that is, $\hat{F}_0$, approaches $\frac{2}{3}$ as n goes to infinity.

THEOREM 2:

$$\lim_{n\to\infty} \hat{F}_p = \begin{cases} \frac{2}{3}, & p < \frac{1}{2}, \\ \frac{7}{4} - 2\sqrt{\frac{55}{192}} = 0.67956, & p = \frac{1}{2} \\ 1, & p > \frac{1}{2}. \end{cases} \tag{11}$$

Table I. Minimax Estimates of the CDF of a Binomial Distribution $\hat{F}_0$

n	$\hat{F}_0$	n	$\hat{F}_0$	n	$\hat{F}_0$	n	$\hat{F}_0$	n	$\hat{F}_0$
1	0.7500	11	0.7065	21	0.6925	31	0.6868	41	0.6836
2	0.7413	12	0.7021	22	0.6910	32	0.6860	42	0.6831
3	0.7340	13	0.7022	23	0.6910	33	0.6860	43	0.6831
4	0.7276	14	0.6989	24	0.6897	34	0.6853	44	0.6827
5	0.7236	15	0.6989	25	0.6897	35	0.6853	45	0.6827
6	0.7195	16	0.6963	26	0.6886	36	0.6847	46	0.6823
7	0.7172	17	0.6963	27	0.6886	37	0.6847	47	0.6822
8	0.7120	18	0.6942	28	0.6877	38	0.6841	48	0.6819
9	0.7123	19	0.6942	29	0.6876	39	0.6841	49	0.6819
10	0.7063	20	0.6925	30	0.6868	40	0.6836	50	0.6815

Proof. See Appendix B.

Simply, when $X \leq \frac{n}{2}$ we propose Equation (12) for obtaining inference on p where $\hat{F}_0$ is obtained from Table I or asymptotically estimated as $\frac{2}{3}$:

$$\hat{p} = \left\{ p \,\middle|\, F(X; n, p) = \hat{F}_0,\ X \leq \left[\frac{n}{2}\right] \right\}. \tag{12}$$

And when $X > \frac{n}{2}$ we propose the following estimate for p:

$$\hat{p} = 1 - \left\{ p \,\middle|\, F(n - X; n, p) = \hat{F}_0,\ X > \left[\frac{n}{2}\right] \right\}. \tag{13}$$

Therefore, for situations where there are zero events realized we propose Equation (14):

$$\hat{p} = \hat{F}_0^{\frac{1}{n}}. \tag{14}$$

In Section 5, we will compare the inference procedure of Equation (12) with the MLE of Equation (5) generally. In Section 6, we will compare the inference when zero events are realized, that is, (Equation 14) with alternatives such as, Equation (1) as well as the uninformative prior.

5. COMPARISON WITH MLE

In this section, a comparison will be made between the estimators for p obtained using the approach described with Equation (12) and δ_{MLE} expressed in Equation (1). The comparison is in three parts. First, we compare the difference between the two estimators. Second, we evaluate for bias. Finally, we make a comparison based on accuracy.

5.1. Relationship with the MLE

We explored the relationship between the estimate for p using $\hat{p}$ and δ_{MLE} for sample sizes ranging up to 100. The relationships appeared linear and so we pursued fitting linear models. As $\hat{p}$ is nonzero when δ_{MLE} is 0, we had an intercept and we used least squares to estimate the slope to fit a model of the form in Equation (15):

$$\hat{p} \approx \hat{F}_0^{\frac{1}{n}} + b_n \frac{X}{n},\ X = 0, 1, \ldots, \left[\frac{n}{2}\right]. \tag{15}$$

Table II. Estimated Slope and Associated R^2 from Estimating Relation Between δ_{MLE} and $\hat{p}$ Showing Close Linear Relationship

Sample size (n)	$\hat{F}_0^{\frac{1}{n}}$ (i.e., intercept)	b_n	R^2
10	0.0342	0.8581	0.9956
20	0.0182	0.8886	0.9972
30	0.0124	0.9047	0.9978
40	0.0095	0.9152	0.9982
50	0.0076	0.9228	0.9985
60	0.0064	0.9286	0.9987
70	0.0055	0.9301	0.9998
80	0.0048	0.9369	0.9989
90	0.0043	0.9401	0.9990
100	0.0039	0.9423	0.9991

We calculated 10 regression models for sample sizes $10, 20, \ldots, 100$. The results are summarized in Table II. Note the R^2 is very high, suggesting a good fit.

We see from Table II that as the sample size increases the slope increases toward 1, the intercept decreases, and the difference between the two estimates decreases.

We explored the relationship between the slope estimates in Table II with sample size and derived the following relationship empirically:

$$\frac{1}{b_n} \approx 1 + \frac{0.4622}{n^{0.439}},\ 10 \leq n \leq 100$$
$$R^2 = 0.999.$$

Table III. Summary of Difference Between $\hat{p}$ and δ_{MLE} for Nonzero Data

Sample Size (n)	Max	Min	Median
10	0.0190	−0.0378	−0.0061
20	0.0258	−0.0270	−0.0036
30	0.0253	−0.0258	−0.0025
40	0.0240	−0.0243	−0.0019
50	0.0228	−0.0230	−0.0015
60	0.0216	−0.0218	−0.0013
70	0.0206	−0.0207	−0.0011
80	0.0198	−0.0198	−0.0010
90	0.0190	−0.0190	−0.0009
100	0.0183	−0.0183	−0.0008

The R^2 is high, suggesting a good approximation to this relationship over this range of data showing convergence at a rate $o(\sqrt{n})$.

Asymptotically, we have the following result.

THEOREM 3:

$$\lim_{n\to\infty} \hat{p} = \delta_{\mathrm{MLE}}.$$

Proof. See Appendix C.

Finally, we made a comparison between the differences obtained by sample size for nonzero data, that is, data in which at least one event is realized. Table III provides a summary of the difference between the two estimators for nonzero data. Fig. 3 illustrates the absolute relative difference between the two estimators expressed in Equation (16) for nonzero data:

$$\text{Relative difference} = \left|\frac{\hat{p} - \delta_{\mathrm{MLE}}}{\delta_{\mathrm{MLE}}}\right| \qquad (16)$$

In summary, we have established evidence of a linear relationship between $\hat{p}$ and δ_{MLE} with $\hat{p}$ greater than δ_{MLE} when $X = 0$ and less than δ_{MLE} when $X = n$. Asymptotically, the two estimators converge. For nonzero data, the difference expressed relative to δ_{MLE} is greatest when there is one event. For the sample sizes considered, this relative difference was about 0.15.

5.2. Bias

The MLE is the only unbiased estimator for p.[9] Fig. 4 is an illustration of the bias associated with $\hat{p}$ expressed as a function of p; that is, we express bias as $E[\hat{p} - p]$. We see that for a sample size of 10 the bias is between ±0.04 and with a sample size of 50 it is ±0.02.

5.3. Accuracy

For the purpose of comparing accuracy, we compare the ratio of the root mean squared error (RMSE) between δ_{MLE} and $\hat{p}$. The RMSE is defined in the following:

$$\text{Ratio of RMSE} = \sqrt{\frac{E\left[(\delta_{\mathrm{MLE}} - p)^2\right]}{E\left[\left(\hat{p} - p\right)^2\right]}}.$$

Fig. 5 illustrates the ratio of the RMSE for the δ_{MLE} and $\hat{p}$ for various sample sizes and parameter values for p.

The results in Fig. 5 are not surprising. With very low values of p relative to the sample size, the MLE will assign a value of 0 as its point estimate. As the sample size increases, the accuracy of $\hat{p}$ improves relative to the δ_{MLE}. Table IV provides a summary of the range of values for p where $\hat{p}$ outperforms δ_{MLE} as well as the maximum of the ratio, that is, where δ_{MLE} performs relatively the poorest, as well as the local minimum. We use a local minimum as the absolute minimum occurs at p equal to 0 or 1, where the RMSE of the MLE is 0.

5.4. Summary of Comparison

For many applications, the difference between $\hat{p}$ and δ_{MLE} when nonzero data have been realized will be considered negligible and as such each could be considered an approximation of the other. However, $\hat{p}$ is a biased estimator of p; its bias is greatest when p is 0 (or 1) as it is assigning estimates that are greater than 0 (less than 1).

6. COMPARISON WITH ALTERNATIVES TO ZERO EVENT DATA

Assuming zero event data are realized, we compare the different estimates obtained with $\hat{p}$ as compared to the method recommended by Bailey, that is, δ_B (Equation (1)), the method of applying a noninformative uniform prior and applying Bayes Theorem, denoted here by $\delta_U = \frac{1}{n+2}$, and the Rule of Three method denoted by δ_{Rof3} (Equation (2)). We

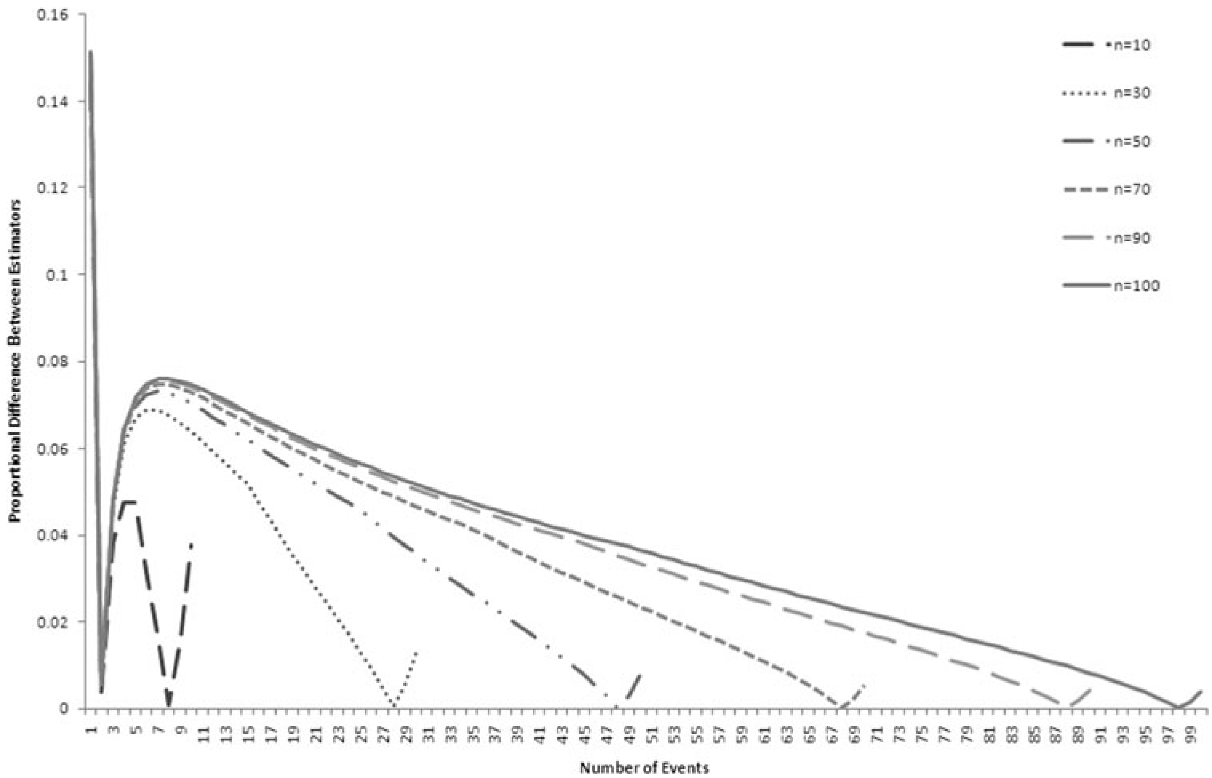

Fig. 3. Difference between $\hat{p}$ and δ_{MLE} relative to δ_{MLE} for nonzero events illustrating difference in relation to sample size.

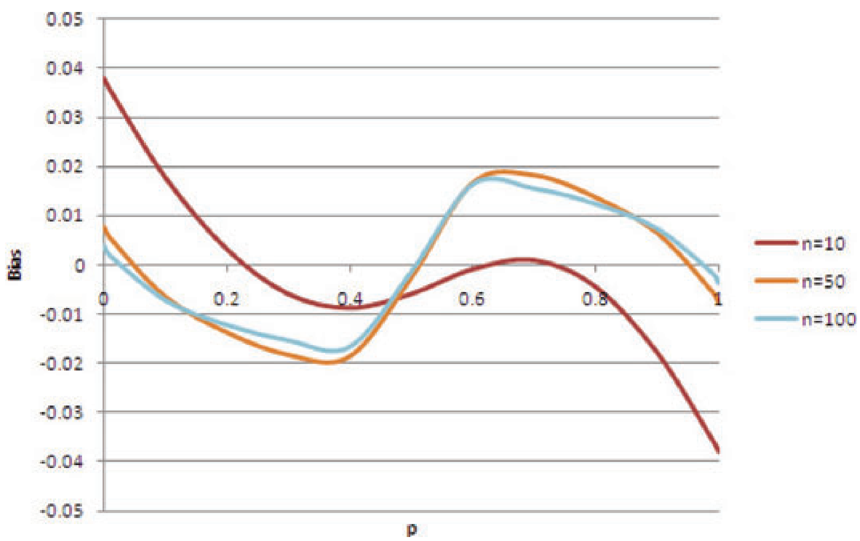

Fig. 4. Bias associated with $\hat{p}$ illustrating intervals of over- and underestimation.

establish the limiting relationship between the estimators:

$$\lim_{n\to\infty}\frac{\delta_B}{\hat{p}}=\frac{1-\left(\frac{1}{2}\right)^{\frac{1}{n}}}{1-\left(\frac{2}{3}\right)^{\frac{1}{n}}}=\frac{\ln(2)}{\ln(3)-\ln(2)}=1.7095, \tag{17}$$

$$\lim_{n\to\infty}\frac{\delta_U}{\hat{p}}=\frac{\frac{1}{n+2}}{1-\left(\frac{2}{3}\right)^{\frac{1}{n}}}=\frac{1}{\ln(3)-\ln(2)}=2.4663, \tag{18}$$

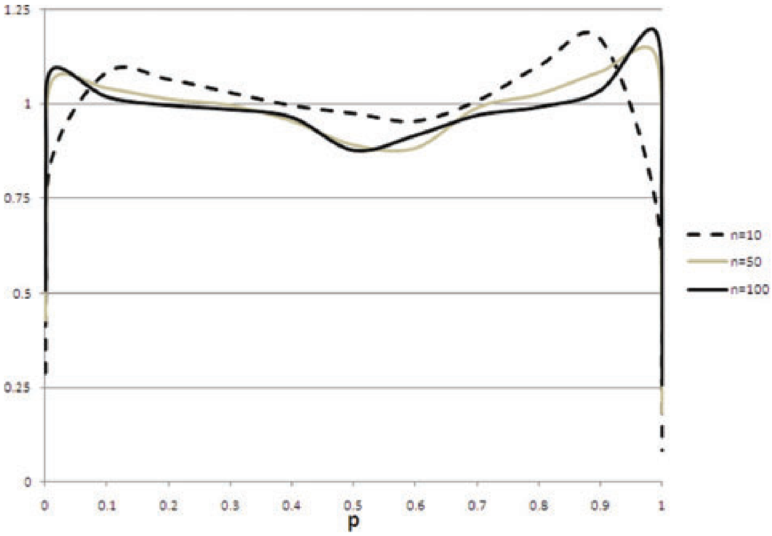

Fig. 5. Comparison of ratio of RMSE of $\hat{p}$ to δ_{MLE} showing neither estimator consistently outperforming the other.

Table IV. Summary of Comparison Between $\hat{p}$ and δ_{MLE}

Sample Size (n)	Interval Where $\hat{p}$ Outperforms δ_{MLE}	Length of Interval	(Local) Minimum of Ratio	Maximum of Ratio
10	(0.0271,0.3844) (0.6891,0.9729)	0.6412	0.9456	1.1710
20	(0.0136,0.3256) (0.7085,0.9864)	0.5899	0.8565	1.1853
30	(0.0091,0.3152) (0.7054,0.9909)	0.5917	0.8234	1.1878
40	(0.0068,0.3036) (0.7083,0.9932)	0.5817	0.8053	1.1890
50	(0.0055,0.2814) (0.7224,0.9945)	0.5481	0.7936	1.1896
60	(0.0045,0.2516) (0.7486,0.9954)	0.4939	0.7853	1.1901
70	(0.0039,0.2255) (0.7745,0.9961)	0.4431	0.7791	1.1904
80	(0.0034,0.2042) (0.7958,0.9966)	0.4015	0.7741	1.1906
90	(0.0030,0.1865) (0.8135,0.9970)	0.3669	0.7701	1.1908
100	(0.0027,0.1717) (0.8283,0.9973)	0.3379	0.7668	1.1910

$$\lim_{n\to\infty} \frac{\delta_{\mathrm{Rof3}}}{\hat{p}} = \frac{\frac{3}{n}}{\left(\frac{2}{3}\right)^{\frac{1}{n}}} = \frac{3}{\ln(2) - \ln(3)} = 7.3989. \quad (19)$$

We have demonstrated in Section 4 that estimates of the CDF less than that associated with $\hat{p}$ will increase the risk for small values of p and as the estimator $\hat{p}$ will provide a smaller estimate of the probability of failure compared with the other three methods for all values of sample size we propose that these alternative methods are overly conservative.

see comparison between 1/2.5n and 1/n+2 at the end of the article

The limit (Equation (18)) of the ratio between uniform prior, that is, δ_U and $\hat{p}$ of 2.4663, suggests a simple approximation to $\hat{p}$ for zero event data expressed in (Equation (20)):

$$\hat{p} \approx \frac{1}{2.5n}, \text{ when zero events are realized.} \quad (20)$$

In Table V, we provide a comparison between this approximation and the actual estimator $\hat{p}$, showing the error is less than 0.01 for sample sizes of 8 or more.

Table V. Comparison of Error Associated with Simple, Approximation of Estimator for Zero Events Realized

Sample Size (n)	$\hat{p}$	$\frac{1}{2.5n}$	Error $(\frac{1}{2.5n} - \hat{p})$
1	0.2500	0.4000	0.1500
2	0.1390	0.2000	0.0610
3	0.0979	0.1333	0.0354
4	0.0764	0.1000	0.0236
5	0.0627	0.0800	0.0173
6	0.0534	0.0667	0.0133
7	0.0464	0.0571	0.0108
8	0.0416	0.0500	0.0084
9	0.0370	0.0444	0.0074
10	0.0342	0.0400	0.0058

7. SUMMARY AND CONCLUSIONS

This work is motivated by risk analysis involving rare events where objective data-driven approaches are desired by decisionmakers and standard procedures such as MLE provide undesirable point estimates of 0. We argue that standard approaches for dealing with the zero event failure data are unconvincing as they are *ad hoc* and lack a rational basis. We derived and compared a conservative inferential approach using minimax criteria, which resulted in nonzero point estimates when 0 events are realized but could be applied for any realization of data. With Theorems 1 and 2 we showed convergence of the estimator with respect to sample size and provided a simple approach to finding the estimate. In short, simple applications of spreadsheets such as Excel will be able to solve for the estimator.

The comparison between the proposed estimator, that is, $\hat{p}$, and the MLE, that is, δ_{MLE}, showed little difference in the actual estimate except when there are 0 events realized. Thus, strengthening the recommendation of the approach more generally, as the MLE can be considered an approximation to this estimate for nonzero events data.

APPENDIX A

Proof. Throughout the following we will denote $F(y;\, n,\, p)$ as $F(y)$ and $f(y;\, n,\, p)$ as $f(y)$ for succinctness.

In Part A we show that for fixed values of $p > 0$ the expected risk has a local minima with respect to the estimate of the CDF, i.e. F_p, in the interval (0, 1) but for $p = 0$, risk is minimised at $F_0 = 1$. In Part B we establish that the expected risk for any fixed value of $p > 0$ is equal to the expected risk for $p = 0$ if we use the estimator $\hat{F}_p$ as expressed in Theorem 1. In Part C we show that the equality occurs only once in the interval $\hat{F}_p \in [0, 1]$ and at a point when the expected risk is rising for $p > 0$ but decreasing for $p = 0$. Finally, to establish Theorem 1 we argue that as risk is continuously dropping when $p = 0$ the Minimax estimate will be realised at the minimum $\hat{F}_p$.

Part A

The expected risk function (equation) is minimised at (A.1), which can be shown through standard application of differential calculus.

$$F_{\min,p} = \frac{\sum_{y=0}^{\left[\frac{n}{2}\right]} F(y)\, f(y)}{F\left(\left[\frac{n}{2}\right]\right)} = E\left[\, F(Y) \middle|\, Y \le \left[\frac{n}{2}\right]\right] \tag{A.1}$$

As this is an expectation of a CDF it is in the interval [0, 1] and will equal 1 if $p = 1$.

Part B

We equate the expected risk for a fixed value of $p > 0$ with the expected risk when $p = 0$, which results in a quadratic equation in terms of the estimate of the CDF, i.e. $\hat{F}_p$.

$$\left(1 - \hat{F}_p\right)^2 = \sum_{y=0}^{\left[\frac{n}{2}\right]} \left(F(y) - \hat{F}_p\right)^2 f(y)$$

Using the standard equation to solve a quadratic, we obtain the expression for $\hat{F}_p$ (10). We choose the negative of the square root as the positive root would correspond to an estimate of $\hat{F}_p$ that would exceed 1.

Part C

We establish that $\hat{F}_p \ge F_{min,p}$.

First we re-parameterise $\hat{F}_p$ expressing it in terms of $F_{\min,p}$.

$$\hat{F}_p - F_{p,\min} = \frac{\left(1 - F\left(\left[\frac{n}{2}\right]\right) F_{p,\min}\right) - \sqrt{\left(1 - F\left(\left[\frac{n}{2}\right]\right) F_{p,\min}\right)^2 - \left(1 - F\left(\left[\frac{n}{2}\right]\right)\right)\left(1 - \sum_{y=0}^{\left[\frac{n}{2}\right]} \left(F(y)^2\right) f(y)\right)}}{1 - F\left(\left[\frac{n}{2}\right]\right)} - F_{p,\min}.$$

A little arithmetical manipulation we express the difference in the following form.

$$\hat{F}_p - F_{p,\min} = \frac{1 - F_{p,\min} - \sqrt{\left(1 - F\left(\left[\frac{n}{2}\right]\right) F_{p,\min}\right)^2 - \left(1 - F\left(\left[\frac{n}{2}\right]\right)\right)\left(1 - \sum_{y=0}^{\left[\frac{n}{2}\right]} \left(F(y)^2\right) f(y)\right)}}{1 - F\left(\left[\frac{n}{2}\right]\right)}$$

We will now establish that there is no real root to this difference.

$$\hat{F}_p - F_{p,\min} = 0$$

$$\rightarrow 1 - F_{p,\min} = \sqrt{\left(1 - F\left(\left[\frac{n}{2}\right]\right) F_{p,\min}\right)^2 - \left(1 - F\left(\left[\frac{n}{2}\right]\right)\right)\left(1 - \sum_{y=0}^{\left[\frac{n}{2}\right]} \left(F(y)^2\right) f(y)\right)}$$

This implies

$$1 - 2F_{p,\min} + (F_{p,\min})^2 > 1 - \left(1 - F\left(\left[\frac{n}{2}\right]\right)\right) \times \left(1 - \sum_{y=0}^{\left[\frac{n}{2}\right]} F(y)^2 f(y)\right) - 2F\left(\left[\frac{n}{2}\right]\right) F_{p,\min} + \left(F\left(\left[\frac{n}{2}\right]\right) F_{p,\min}\right)^2$$

$$\rightarrow \left(1 - F\left(\left[\frac{n}{2}\right]\right)\right)\left(1 - \sum_{y=0}^{\left[\frac{n}{2}\right]} F(y)^2 f(y)\right) - 2\left(1 - F\left(\left[\frac{n}{2}\right]\right)\right) F_{p,\min} + (F_{p,\min})^2 \left(1 - F\left(\left[\frac{n}{2}\right]\right)^2\right) > 0$$

$$\rightarrow \left(1 - \sum_{y=0}^{\left[\frac{n}{2}\right]} F(y)^2 f(y)\right) - 2F_{p,\min} + (F_{p,\min})^2 \left(1 + F\left(\left[\frac{n}{2}\right]\right)\right) > 0$$

The limits of the left hand side of this expression for p are:

$$\lim_{p\to 0} \left(1 - \sum_{y=0}^{\left[\frac{n}{2}\right]} F(y)^2 f(y)\right) - 2F_{p,\min} + (F_{p,\min})^2 \left(1 + F\left(\left[\frac{n}{2}\right]\right)\right) = 0$$

$$\lim_{p\to 1} \left(1 - \sum_{y=0}^{\left[\frac{n}{2}\right]} F(y)^2 f(y)\right) - 2F_{p,\min} + (F_{p,\min})^2 \left(1 + F\left(\left[\frac{n}{2}\right]\right)\right) = 1$$

This expression can be re-expressed in terms of expectations with the following inequalities guaranteeing the expression is positive in the interval $p \in [0, 1]$

$$1 - F\left(\left[\frac{n}{2}\right]\right) E\left[F(Y)^2 \middle| Y \le \left[\frac{n}{2}\right]\right] - 2E\left[F(Y) \middle| Y \le \left[\frac{n}{2}\right]\right] + \left(E\left[F(Y) \middle| Y \le \left[\frac{n}{2}\right]\right]\right)^2 \times \left(1 + F\left(\left[\frac{n}{2}\right]\right)\right) = 1 - 2E\left[F(Y) \middle| Y \le \left[\frac{n}{2}\right]\right] + \left(E\left[F(Y) \middle| Y \le \left[\frac{n}{2}\right]\right]\right)^2 - F\left(\left[\frac{n}{2}\right]\right) \times E\left[F(Y)^2 \middle| Y \le \left[\frac{n}{2}\right]\right] - \left(E\left[F(Y) \middle| Y \le \left[\frac{n}{2}\right]\right]\right)^2$$

$$= \left(1 - E\left[F(Y) \middle| Y \le \left[\frac{n}{2}\right]\right]\right)^2 - F\left(\left[\frac{n}{2}\right]\right) \times Var\left[F(Y) | Y \le \left[\frac{n}{2}\right]\right] > 0$$

□

APPENDIX B

Proof. By the central limit theorem,[10] we know the CDF of a binomial distribution with a mean of np and a standard deviation of $\sqrt{np(1-p)}$ appropriately scaled converges to the normal distribution. More precisely:

$$\lim_{n\to\infty} F(y; n, p) = \Phi\left(\frac{y - np}{\sqrt{np(1-p)}}\right),$$

$$\lim_{n\to\infty} f(y; n, p) = \lim_{n\to\infty} F(y; n, p) - F(y-1; n, p) = d\Phi\left(\frac{y - np}{\sqrt{np(1-p)}}\right),$$

where $\Phi(.)$ is the CDF of a standard normal distribution.

As such

$$\sum_{y=0}^{\left[\frac{n}{2}\right]} F(y; n, p)\, f(y; n, p) \rightarrow \int_0^{\Phi\left(\left[\frac{n}{2}\right]\right)} \Phi \times \left(\frac{x - np}{\sqrt{np(1-p)}}\right) d\Phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right) = \frac{\Phi\left(\left[\frac{n}{2}\right]\right)^2}{2}$$

and

$$\sum_{y=0}^{\left[\frac{n}{2}\right]} \left(F(y;n,p)^2\right) f(y;n,p) \to \int_0^{\Phi\left(\left[\frac{n}{2}\right]\right)} \Phi \times \left(\frac{x-np}{\sqrt{np(1-p)}}\right)^2 d\Phi\left(\frac{x-np}{\sqrt{np(1-p)}}\right) = \frac{\Phi\left(\left[\frac{n}{2}\right]\right)^3}{3}$$

Moreover,

$$\lim_{n\to\infty} \Phi\left(\frac{\left[\frac{n}{2}\right]-np}{\sqrt{np(1-p)}}\right) = \begin{cases} 1, & p < \frac{1}{2}, \\ \frac{1}{2}, & p = \frac{1}{2}, \\ 0, & p > \frac{1}{2}. \end{cases}$$

Therefore, applying Slutzky's Theorem[10] we substitute the limits into the equation and apply L'Hopitals rule for the situation where $p < \frac{1}{2}$ and obtain the following:

$$\lim_{n\to\infty} F_p = \lim_{n\to\infty} \frac{\left(1-\sum_{y=0}^{\left[\frac{n}{2}\right]} F(y;n,p)\, f(y;n,p)\right) - \sqrt{\left(1-\sum_{y=0}^{\left[\frac{n}{2}\right]} F(y;n,p)\, f(y;n,p)\right)^2 - \left(1-F\left(\left[\frac{n}{2}\right];n,p\right)\right)\left(1-\sum_{y=0}^{\left[\frac{n}{2}\right]}\left(F(y;n,p)^2\right) f(y;n,p)\right)}}{1-F\left(\left[\frac{n}{2}\right];n,p\right)}$$

$$= \begin{cases} \frac{2}{3}, & p < \frac{1}{2}, \\ \frac{7}{4} - 2\sqrt{\frac{55}{192}}, & p = \frac{1}{2}, \\ 1, & p > \frac{1}{2}. \end{cases}$$

□

APPENDIX C

Proof. Asymptotically, we have the following:

$$\lim_{n\to\infty} P\left(\frac{X-np}{\sqrt{np(1-p)}} < z\right) = \Phi(z).$$

Using the asymptotic limit established in Theorem 2, we know the z value that corresponds to the $\frac{2}{3}$ percentile, that is, $z = 0.4307$. As such we have the following:

$$\frac{X-n\hat{p}}{\sqrt{n\hat{p}(1-\hat{p})}} = 0.4307$$

$$\to (\hat{p})^2\left(n^2 + 0.1855n\right) - \hat{p}n(2X + 0.1855) + X^2 = 0$$

$$\to \hat{p} = \frac{(2X+0.1855) \pm \sqrt{0.7421X + 0.0344 - \frac{0.7421}{n}X^2}}{2(n+0.1855)}$$

$$= \frac{\delta_{MLE} + \frac{0.0928}{n}}{1 + \frac{0.1855}{n}} \pm \frac{\sqrt{n\left(0.7421\delta_{MLE}(1-\delta_{MLE}) + \frac{0.0344}{n}\right)}}{2(n+0.1855)}.$$

Therefore,

$$\lim_{n\to\infty} \hat{p} = \delta_{\mathrm{MLE}}.$$

□

REFERENCES

1. Bedford T, Cooke R. Probabilistic Risk Analysis: Foundations and Methods. Cambridge University Press, 2001.
2. Bedford T, Quigley J, Walls L. Expert elicitation for reliable system design. Statistical Science, 2006; 21(4):428–450.
3. British Standard. Equipment reliability testing—BS IEC: 60605-4:2001, 2001.
4. Bailey RT. Estimation from zero-failure data. Risk Analysis, 1997; 17(3):375–380.
5. Potter JL, Mughal TM. Sensitivity Testing of Explosives and Propellants at Cryogenic Temperatures, h241-12-20-073. Magna, UT: Hercules Aerospace Co., 1992.

6. Hanley J, Lippman-Hand A. If nothing goes wrong, is everything all Right? Interpreting zero numerators. Journal of the American Medical Association, 1983; 249(13):1743–1745.
7. Eypasch E, Lefering R, Kum C, Troidl H. Probability of adverse events that have not yet occurred: A statistical reminder. British Medical Journal, 1995; 311:619–620.
8. Dewoody Y, Gururaj VT, Martin C. Assessing risk for rare events. Journal of Applied Statistics, 1999; 26(6):681–687.
9. Lehmann EL, Casella G. Theory of Point Estimation (Springer Texts in Statistics). New York: Springer, 1998.
10. Arnold SF. Mathematical Statistics. New Jersey: Prentice-Hall, 1990.