Nuclear Instruments and Methods in Physics Research A 664 (2012) 336–340

Contents lists available at SciVerse ScienceDirect

Nuclear Instruments and Methods in Physics Research A

journal homepage: www.elsevier.com/locate/nima

Calculating a confidence interval on the sum of binned leakage

I. Ruchlin [a,b,*,1], R.W. Schnee [a]

[a] Department of Physics, Syracuse University, Syracuse, NY 13244, USA
[b] Center for Computational Relativity and Gravitation, Rochester Institute of Technology, 85 Lomb Memorial Drive, Rochester, NY 14623, USA

ARTICLE INFO

Article history:
Received 30 June 2011
Received in revised form
29 October 2011
Accepted 9 November 2011
Available online 18 November 2011

Keywords:
Confidence interval
Likelihood ratio
Binomial distribution

ABSTRACT

Calculating the expected number of misclassified outcomes is a standard problem of particular interest for rare-event searches. The Clopper–Pearson method allows calculation of classical confidence intervals on the amount of misclassification if data are all drawn from the same binomial probability distribution. However, data is often better described by breaking it up into several bins, each represented by a different binomial distribution. We describe and provide an algorithm for calculating a classical confidence interval on the expected total number of misclassified events from several bins, based on calibration data with the same probability of misclassification on a bin-by-bin basis. Our method avoids a computationally intensive multidimensional search by introducing a Lagrange multiplier and performing standard root finding. This method has only quadratic time complexity as the number of bins, and produces confidence intervals that are only slightly conservative.

1. Introduction

Many real-world processes can assume one of two possible outcomes; each independent trial or observation can be classified as either "success" or "failure," with the probability of success p. All such trials can be bundled together to form a single experiment with x successes out of a total of n trials. If the experiments are repeated many times, the relative frequency of successes in each experiment follows the binomial distribution (*e.g.* [1]).

For a measurement of x and n, the best estimate $P=x/n$ of the true success probability p can be calculated. Since the ratio $P/(1-P)$ is the best estimate of the expected ratio of successes to failures, the best estimate of the number of successes Y of a second experiment that has the same success probability and a known number of failures b is

$$Y=\frac{P}{1-P}b. \tag{1}$$

Furthermore, methods such as Clopper–Pearson's [2] provide a (frequentist or classical) confidence interval $[P_{\text{low}},P_{\text{high}}]$ with probability content β such that the fraction of experiments with $P_{\text{low}}\le p\le P_{\text{high}}$ is $\approx\beta$ (with the lack of exact equality due to the discrete nature of binomial distribution; see *e.g.* [3] for comparisons of various methods). By extension, these methods may also be used to calculate the confidence interval $[Y_{\text{low}},Y_{\text{high}}]$ on the expected number of successes of the second experiment.

Such estimates may be particularly useful for characterizing backgrounds for rare-event searches. A given background event may have some probability p to be misclassified as a signal event. First, a "calibration" experiment may allow estimation of p based on the number of events n and the number x misclassified as signal (the "leakage"). A second, "search" experiment may provide a measurement of the number of correctly identified background events b. If background events in both experiments have the same probability of correct classification, the expected number of misclassified events Y and a confidence interval $[Y_{\text{low}},Y_{\text{high}}]$ on the expected number may be determined.

Often, however, in order for events in the calibration and search both to have the same probability of misclassification p, events with different characteristics (*e.g.* energy, position, detector, or pixel) must be considered separately, resulting in m separate bins of events for both calibration and search. In the ith calibration bin there are x_i misclassified events out of the total n_i calibration events, resulting in a best estimate $P_i=x_i/n_i$ of the misclassification probability for events in that bin. For the search data, the number of correctly classified events in the ith bin, b_i, is known. If the true misclassification probability, p_i, of an event in the ith bin is the same for both calibration and search data, the best estimate for the total expected number of misclassified events Y is

$$Y=\sum_i^m\frac{P_i}{1-P_i}b_i\equiv f(\mathbf{P}) \tag{2}$$

where $\mathbf{P}\equiv\{P_i\}$.

* Corresponding author at: Center for Computational Relativity and Gravitation, Rochester Institute of Technology, 85 Lomb Memorial Drive, Rochester, NY 14623, USA. Tel.: +1 585 475 2498; fax: +1 585 475 7340.
E-mail addresses: ixr5289@rit.edu (I. Ruchlin), rwschnee@phy.syr.edu (R.W. Schnee).
[1] Current address is at Rochester Institute of Technology.

0168-9002/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.nima.2011.11.031

The likelihood $\mathcal{L}$ that x_i events out of the total n_i calibration events in each bin are misclassified is simply the product of the binomial probabilities, with

$$\mathcal{L} \propto \prod_i^m P_i^{x_i}(1-P_i)^{n_i-x_i}. \tag{3}$$

The global maximum $\hat{\mathcal{L}}$ of the likelihood is trivially given by the set $\hat{\mathbf{P}} \equiv \{\hat{P}_i\} = \{x_i/n_i\}$ for all i. Note that it is possible to estimate the expected leakage only if no calibration bin has zero total events (*i.e.* $n_i \neq 0$ for all i). Substitution of the set $\{\hat{P}_i\}$ in Eq. (2) yields the most likely value of the total expected leakage $\hat{Y} = f(\hat{\mathbf{P}})$.

Unfortunately, for the case with multiple bins, most existing methods cannot be used to calculate a confidence interval on the total expected leakage. Here we describe a method and provide a practical algorithm for this problem. We use the "Unified Approach" described by Feldman and Cousins [4] (see also *e.g.* [5]) extended to deal with nuisance variables by means of the profile likelihood [6] without the large-sample approximation used in *e.g.* [1,7]; here $\mathbf{P}$ are nuisance variables since they are unknown variables for which we are not setting a confidence interval.

2. Method

For every considered value of Y_0, we calculate the profile likelihood

$$\Lambda \equiv \frac{\mathcal{L}(Y_0\,|\,\mathbf{n},\mathbf{x},\mathbf{b},\hat{\hat{\mathbf{P}}})}{\hat{\mathcal{L}}(\hat{Y}\,|\,\mathbf{n},\mathbf{x},\mathbf{b},\hat{\mathbf{P}})} \tag{4}$$

where $\mathbf{n} = \{n_i\}$, $\mathbf{x} = \{x_i\}$, and $\mathbf{b} = \{b_i\}$ are the data, and $\hat{\hat{\mathbf{P}}}$ is the combination of P_i, found by a search described in Section 2.1, that maximizes the likelihood for the value of Y_0 under test. Asymptotically (and far from physical boundaries), the distribution of $-2\ln(\Lambda)$ is χ^2-distributed with one degree of freedom [8], but more accurate results may be obtained by determining the expected distribution by Monte Carlo simulation. For each simulated experiment, $\mathbf{x}$ is randomly determined based on the $\hat{\hat{\mathbf{P}}}$ found above. For each, the best-fit values $\hat{\mathbf{P}}_{\mathrm{MC}}$ and $\hat{\hat{\mathbf{P}}}_{\mathrm{MC}}$ are found, and then the ratio

$$\Lambda_{\mathrm{MC}} \equiv \frac{\mathcal{L}(Y_0\,|\,\mathbf{n},\mathbf{x},\mathbf{b},\hat{\hat{\mathbf{P}}}_{\mathrm{MC}})}{\hat{\mathcal{L}}(\hat{Y}\,|\,\mathbf{n},\mathbf{x},\mathbf{b},\hat{\mathbf{P}}_{\mathrm{MC}})}$$

is calculated. If Λ is larger than $1-\beta$ of the simulated Λ_{MC} ratios, then Y_0 is included in the confidence interval of probability content β. Since the distributions follow the binomial distribution, the uncertainties $\propto 1/\sqrt{N_{\mathrm{MC}}}$, the inverse of the square root of the number of experiments. Thus, to achieve a relative tolerance t, conduct $N_{\mathrm{MC}} = t^{-2}$ Monte Carlo simulations. A root-finding algorithm hunts for the smallest and largest values of Y_0 that are allowed in order to return the desired confidence interval $[Y_{\mathrm{low}}, Y_{\mathrm{high}}]$.

2.1. Formulation

A multiparameter function minimizer, such as MINUIT [9], could be implemented to hunt for the combination of probabilities P_i that maximize Eq. (3) subject to the constraint of Eq. (2). However, this method may have exponential time complexity in the worst case [10], making it unfeasible for the analysis of more than a few bins. Furthermore, there would be some risk of missing the global maximum. Instead, the combination of binomial probabilities that maximize Eq. (3) subject to the constraint of Eq. (2) can be found by introducing a Lagrange multiplier, λ, and solving

$$\frac{\partial}{\partial P_i}[\ln(\mathcal{L}(P_i)) + \lambda(f(\mathbf{P}) - Y_0)] = 0$$

where Y_0 is a given constant. A little algebra yields the solution to this equation for each bin i

$$P_i = \frac{n_i + x_i - \lambda b_i \pm \sqrt{(\lambda b_i - n_i - x_i)^2 - 4 n_i x_i}}{2n_i} \tag{5}$$

while substituting back into Eq. (2) yields an equation for the Lagrange multiplier

$$Y_0 = \sum_i^m b_i \frac{n_i + x_i - \lambda b_i \pm \sqrt{(\lambda b_i - n_i - x_i)^2 - 4 n_i x_i}}{n_i - x_i + \lambda b_i \mp \sqrt{(\lambda b_i - n_i - x_i)^2 - 4 n_i x_i}} \equiv \sum_i^m b_i Y_{0i}. \tag{6}$$

Eq. (6) is really 2^m separate equations, depending on the signs of *each* $\pm$ term. One of the 2^m solutions yields the value of λ that gives the most likely combination of binomial probabilities (*i.e.* $\hat{\hat{\mathbf{P}}}$) for the desired total expected leakage Y_0. Fortunately, further analysis reveals a significant reduction in the number of viable solutions.

For any bin with $b_i \neq 0$

$$\lambda > c_i \equiv \frac{n_i + x_i - 2\sqrt{n_i x_i}}{b_i}$$

is unphysical, producing imaginary or negative probabilities. Since λ must be physical for all bins, λ is required to be less than or equal to the smallest c_i, *i.e.* $\lambda \leq \inf\{c_i\} \equiv \lambda_c$. For any bin with $b_i = 0$, $c_i \to \infty$ so that it places no constraint on λ.

Table 1 lists the different limiting values of λ and their corresponding values of P_i from Eq. (5). The lower limit on the confidence interval must have $P_i \leq x_i/n_i$ for each bin. Therefore, Table 1 indicates that the solution must use the negative root for each bin, reducing the problem from searching among 2^m solutions to solving a single equation.

It is easiest to understand the viable solutions for the confidence interval's upper bound by first noting that, other than the constraint of Eq. (2), each term in

$$\ln(\mathcal{L}) = \sum_i^m \ln(\mathcal{L}_i)$$

is independent. For each term, $\ln(\mathcal{L}_i)$ decreases monotonically with increasing $P_i > \hat{P}_i$, and there is an inflection point at $P_i = P_i(\lambda = c_i)$, with $\ln(\mathcal{L}_i)$ decreasing ever more slowly for larger P_i.

For any bins i and j, it can be shown that

$$\left.\frac{\partial \ln(\mathcal{L})}{\partial Y_{0i}}\right|_{P_i = \hat{\hat{P}}_i} = \left.\frac{\partial \ln(\mathcal{L})}{\partial Y_{0j}}\right|_{P_j = \hat{\hat{P}}_j}.$$

This relation is to be expected. If, instead, the left term were larger (smaller) than the right term, a more likely combination with the same total Y_0 could be found by decreasing Y_{0i} (Y_{0j}) and increasing the other by the same amount.

In a similar way, it may be shown that the combination of probabilities $\hat{\hat{\mathbf{P}}}$ that maximize the likelihood for a given total expected leakage Y_0 never includes more than one bin with $P_i > P_i(\lambda = c_i)$, and hence more than one bin using the positive root of Eq. (2). Fig. 1 helps visualize the reasoning. Consider any two bins i and j that both use

Table 1
Summarized analysis of the behavior of Eq. (5).

Lagrange multiplier	Positive root	Negative root
$\lambda < 0$	$P_i > 1$	$0 < P_i < x_i/n_i$
$\lambda = 0$	$P_i = 1$	$P_i = x_i/n_i$
$0 < \lambda < c_i$	$\sqrt{x_i/n_i} < P_i < 1$	$x_i/n_i < P{-}i < \sqrt{x_i/n_i}$