Computational Statistics and Data Analysis 61 (2013) 38–49

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Fiducial-based tolerance intervals for some discrete distributions

Thomas Mathew[a,b], Derek S. Young[b,*]

[a] Department of Mathematics and Statistics, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, United States
[b] US Census Bureau, 4600 Silver Hill Road, Washington, DC 20233, United States

ARTICLE INFO

Article history:
Received 8 August 2012
Received in revised form 19 November 2012
Accepted 19 November 2012
Available online 28 November 2012

Keywords:
Binomial distribution
Coverage probability
Fiducial inference
Inverse sampling
Negative binomial distribution
Poisson distribution
`tolerance` package

ABSTRACT

Fiducial quantities are proposed to construct approximate tolerance limits and intervals for functions of some discrete random variables. Using established fiducial quantities for binomial proportions, Poisson rates, and negative binomial proportions, an approach is demonstrated to handle functions of discrete random variables, whose distributions are either not available or are intractable. The construction of tolerance intervals using fiducial quantities is straightforward and, thus, amenable to numerical computation. An extensive numerical study shows that for most settings of the cases considered, the coverage probabilities are near the nominal levels. The applicability of the method is further demonstrated using four real datasets, including a discussion of the corresponding software that is available for the R programming language.

Published by Elsevier B.V.

1. Introduction

Tolerance intervals are intervals that capture a specified proportion (say, γ) or more of a population with a given confidence level (say, $1-\alpha$). Such an interval is referred to as a $(\gamma, 1-\alpha)$ tolerance interval, where γ is called the *content* of the tolerance interval. A tolerance interval is computed using a random sample and the confidence level $1-\alpha$ reflects the sampling variability. The topic of tolerance intervals has received considerable attention for continuous distributions. Limited results are available in the discrete case, most notably for the binomial and Poisson distributions. We refer to the book by Krishnamoorthy and Mathew (2009) for a detailed treatment of tolerance intervals for discrete as well as for continuous distributions.

As far as we are aware, the tolerance intervals available for the binomial and Poisson distributions deal with the construction of such intervals for a single binomial distribution or for a single Poisson distribution (see Hahn and Chandra, 1981; Wang and Tsung, 2009; Krishnamoorthy et al., 2011), but are currently unavailable for comparing two binomial distributions or for comparing two Poisson distributions. There is also some tolerance interval work for a single negative binomial distribution (see Zacks, 1970; Cai and Wang, 2009; Young, in press), but again, there are no available methods for comparing two negative binomial proportions. The goal of this article is to derive tolerance intervals that will facilitate such comparisons. In fact, the approach we propose can be used to construct tolerance intervals for the distribution of any meaningful function of two binomial random variables, two Poisson random variables, or two negative binomial random variables.

The motivation for our investigation is as follows. A tolerance interval for the distribution of a random variable is an interval for the entire distribution and values outside a tolerance interval are rarely observed. For example, if we have a

* Corresponding author. Tel.: +1 301 763 6347; fax: +1 301 763 8399.
E-mail addresses: thomas.mathew@census.gov (T. Mathew), derek.young@census.gov (D.S. Young).

0167-9473/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.csda.2012.11.015

tolerance interval for the distribution of the difference between two sample proportions, the interval provides us information on the likely values of the difference between the sample proportions in a future study. In particular, if the tolerance limits (i.e., the end points) of a two-sided tolerance interval are both positive, then we can conclude that the difference between the two sample proportions arising from a future study is expected to be positive, unless the conditions of the study have changed. In fact, an observed value outside the tolerance interval may be an indication that the study conditions have changed. Contrasting this with a confidence interval for the difference between two population proportions, we note that if a two-sided confidence interval has positive end points, this does not preclude the possibility of the difference between two sample proportions in a future study from assuming negative values. In short, a tolerance interval provides us with valuable information on the most likely scenarios in future studies carried out under the same conditions.

It is well known that for any distribution, an upper tolerance limit having content γ and confidence level $1-\alpha$ is simply a $100(1-\alpha)\%$ upper confidence limit for the γth percentile of the distribution. Similarly, a lower tolerance limit having content γ and confidence level $1-\alpha$ is simply a $100(1-\alpha)\%$ lower confidence limit for the $(1-\gamma)$th percentile of the distribution. We shall use this property to derive upper and lower tolerance limits for the distribution of the difference between two independently estimated binomial proportions, the distribution of the ratio of two independently estimated Poisson rates, and the distribution of the estimated odds ratio in a negative binomial setting. We note that each of the above distributions involves unknown parameters, so we use a fiducial approach to compute an upper (or lower) confidence limit for the appropriate percentile. Since the percentiles have no explicit expressions, we implement the fiducial approach numerically. We also develop an approximation for computing two-sided tolerance intervals.

We will not give detailed background information on the fiducial approach here, but rather refer to Hannig (2009) for a thorough treatment and history of the topic as well as further references. We note that Zacks (1971) presents a theoretical framework for $(\gamma, 1-\alpha)$ fiducial tolerance regions by defining them as $(\gamma, 1-\alpha)$ Bayes tolerance regions against an improper and invariant prior. We also refer to Krishnamoorthy and Lee (2010) where the fiducial approach is used to compute confidence intervals for binomial and Poisson parameters and Wang et al. (2012) where prediction limits are derived using the fiducial approach.

This paper is organized as follows. In Section 2, we present a general fiducial approach for computing tolerance intervals. In Section 3, the fiducial approach is used to compute tolerance limits for the distribution of functions of two binomial, Poisson, or negative binomial random variables. The necessary computation is explained in Section 4 and in Section 5, we present a coverage study to validate our approach. We apply our method to real data examples in Section 6 and conclude with a brief discussion in Section 7.

2. Fiducial quantities and tolerance intervals

A fiducial quantity is a random variable whose distribution provides a fiducial distribution for a parameter of interest. Fiducial quantities for parameters under various discrete and continuous distributions are given in Hannig (2009). To quote Hannig (2009), "The main aim of fiducial inference is to define a distribution for parameters of interest that captures all of the information that the data contains about these parameters." Our interest in this article concerns the binomial, Poisson, and negative binomial distributions.

2.1. Approximate fiducial quantities

For parameters in discrete distributions, the fiducial quantity is not unique, in general. We now introduce the approximate fiducial quantities used in this paper; see Krishnamoorthy and Lee (2010) for the binomial and Poisson settings and Cai and Krishnamoorthy (2005) for the negative binomial setting.

1. Suppose x is an observed value of the random variable $X \sim \text{Binomial}(n, \pi)$. An approximate fiducial quantity for π is

$$Q_\pi = \text{Beta}(x + 1/2, n - x + 1/2). \tag{1}$$

2. Let $X_1, \ldots, X_n$ be independent and identically distributed according to a Poisson(λ) distribution and let $X = \sum_{i=1}^{n} X_i$ so that $X \sim \text{Poisson}(n\lambda)$. Suppose that x is an observed value of the random variable X. Then an approximate fiducial quantity for λ is

$$Q_\lambda = \chi^2_{2x+1}/2n, \tag{2}$$

where χ^2_k denotes the chi-square distribution with k degrees of freedom.

3. Suppose x is an observed value of the random variable $X \sim \text{NegBin}(n, \nu)$, which is the number of failures that occur in a sequence of Bernoulli trials before n successes (or events) are observed. The probability of success on a given trial is given by ν. An approximate fiducial quantity for ν is given by

$$Q_\nu = \text{Beta}(x, x + 1/2). \tag{3}$$