Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 6435–6446
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

A filter for "confidence interval P-values"☆

Mary C. Phipps*, Peter M. Byron[1]

School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia

Received 7 December 2006; received in revised form 9 February 2007; accepted 14 February 2007
Available online 28 February 2007

Abstract

When a statistical test involves a nuisance parameter, an exact unconditional P-value is found from a supremum search over the parameter space of the nuisance parameter. The result is less conservative than the corresponding (conditional) exact test. Restricting the region of the supremum search to a confidence interval gives a "confidence interval P-value" which, after an appropriate adjustment, is also exact. We provide a filter to help identify cases for which this exact procedure reduces conservativeness even further, and we illustrate with numerical examples. These examples are also used to address questions about the optimum choice of confidence interval for the restricted supremum search, and to demonstrate the reduction in conservativeness attained by quasi-exact methods.

MSC: 62F03; 62-04; 62-09

Keywords: Confidence interval P-values; Quasi-exact P-values; Fisher's Exact Test; Lancaster's mid-P; Liebermeister's measure

1. Introduction

A test for equal success probabilities in two independent binomial experiments is an example of a problem which involves handling a nuisance parameter, θ. In this context θ is the common success probability under the null hypothesis. For large samples the accepted solution is to use a chi-square approximation to calculate the P-value. When the samples are too small to use this asymptotic approximation, the choice between different procedures for removing dependence on the nuisance parameter has led to much debate over the years. Extensive reviews of the literature are available in Martín Andrés (1991, 1997) and Sahai and Khurshid (1995), clearly outlining the contentious issues and providing extensive references. The two standard exact methods are conditional (Fisher, 1934), or unconditional (Barnard, 1945, 1947). Both approaches are conservative, but for different reasons. Fisher's Exact Test is more conservative than the various unconditional exact methods proposed in the literature, but the latter are computationally intensive, involving a supremum search which must be done numerically as there is no analytical solution.

In a refreshing approach, Berger and Boos (1994) show that it is sometimes possible to reduce conservativeness even further in the unconditional method while maintaining exactness, using a restricted supremum search with an

☆ Expansion of a presentation at the 6th Annual Hawaii International Conference on Statistics, Hawaii, January 2007.

* Corresponding author. Tel.: +61 02 93515763.
E-mail addresses: maryp@maths.usyd.edu.au (M.C. Phipps), p.m.byron@abs.gov.au (P.M. Byron).
[1] Present address: Australian Bureau of Statistics, ACT, Australia.

0167-9473/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2007.02.015

adjustment. We provide a useful filter to help identify cases which are improvable in this way, and we illustrate with numerical examples. This technique of Berger and Boos is not confined to the two binomial problem. Additional applications are discussed in Berger and Sidik (2003), and we use the filter for one of these applications as well.

Mention is made also of quasi-exact measures which are at times mildly anti-conservative, but otherwise behave in a way similar to exact unconditional P-values, without requiring the same level of computation.

2. Notation and definitions

We will use the variables X and Y to represent the number of successes in m and n independent Bernouilli trials with success probabilities p_x and p_y, respectively. A typical bivariate outcome of (X, Y) is (x, y), with $z = x + y$ being a typical outcome of $Z = X + Y$, as summarised in this 2×2 table:

	Success	Failure	Total
Population 1	x	$m - x$	m
Population 2	y	$n - y$	n
Total	z	$m + n - z$	$m + n$

Consider the situation where m and n are small, and where we want to test the null hypothesis H_0: $p_x = p_y$ against the alternative H_1: $p_x > p_y$. Writing θ for the common success probability under H_0, the joint probability is given by (for $0 \leqslant x \leqslant m,\ 0 \leqslant y \leqslant n$):

$$P(X = x, Y = y; \theta) = \binom{m}{x}\binom{n}{y}\theta^{x+y}(1-\theta)^{m+n-x-y}, \tag{1}$$

which depends on θ. Because $Z = X + Y$ is sufficient for the nuisance parameter, θ, it is more convenient to express (1) in terms of (x, z):

$$P(X = x, Z = z; \theta) = \binom{m}{x}\binom{n}{z-x}\theta^{z}(1-\theta)^{m+n-z}. \tag{2}$$

Three exact P-values based on the observation $w_0 = (x_0, z_0)$ of $W = (X, Z)$ are outlined below. The first is $p_{\mathrm{C}}(w_0)$, the conditional P-value, the second is $p_{\mathrm{U}}(w_0)$, the standard unconditional P-value, and the last, $p_\beta(w_0)$, is the Berger–Boos confidence interval P-value.

- *Exact conditional P-value*, $p_{\mathrm{C}}(w_0)$: It is clear from (2) that Z is a sufficient statistic for θ since, with z fixed, and for $\max(0, z-n) \leqslant x \leqslant \min(m, z)$,

$$P(X = x | Z = z) = \frac{\binom{m}{x}\binom{n}{z-x}}{\binom{m+n}{z}}.$$

  The upper tail P-value based on the observation $w_0 = (x_0, z_0)$ and conditioned on $Z = z_0$ is the Fisher-Exact P-value, $p_{\mathrm{C}}(w_0)$. For given m and n, this is defined as follows:

$$p_{\mathrm{C}}(w_0) = P_{\mathrm{H}_0}(X \geqslant x_0 | Z = z_0) = \sum_{x \geqslant x_0} \frac{\binom{m}{x}\binom{n}{z_0-x}}{\binom{m+n}{z_0}}.$$

  Calculation of $p_{\mathrm{C}}(w_0)$ can be achieved easily with a hand calculator for small samples, or using commercial software packages.