INTERNATIONAL RESEARCHERS

ANOTHER LOOK AT HYPOTHESIS TESTING

Margot Tollefson

Volume No.7 Issue No.3 September 2018

www.iresearcher.org

ISSN 2227-7471

THE INTERNATIONAL RESEARCH JOURNAL “INTERNATIONAL RESEACHERS”

www.iresearcher.org

This publication is subject to that author (s) is (are) responsible for Plagiarism, the contents, the accuracy of citations, quotations, diagrams, tables and maps. Corresponding author is responsible for the consent of other authors.

INTERNATIONAL RESEARCHERS is peer-reviewed, supported by rigorous processes of criterion-referenced article ranking and qualitative commentary, ensuring that only intellectual work of the greatest substance and highest significance is published.

INTERNATIONAL RESEARCHERS is indexed in wellknown indexing diectories

INDEX COPERNICUS INTERNATIONAL JOURNAL MASTER LIST **with ICV value 5.90**

ULRICHSWEB GLOBAL SERIALS DIRECTORY

Journal Seek

Directory of Research Journals Indexing

and moniter by

THOMSON REUTERS

ANOTHER LOOK AT HYPOTHESIS TESTING

Margot Tollefson, PhD

Consulting Statistician, Vanward Statistics, Stratford, IA
(UNITED STATES OF AMERICA)
margot@vanwardstat.com

ABSTRACT

This paper looks at why traditional hypothesis tests perform so poorly and what to do to improve performance. The hypothesis test is seen as a Bernoulli random variable. The criteria for determining the size of the test is changed from the use of the Bernoulli expected value to the use of the median of the geometric distribution. A table of test sizes based on the median is given.

Keywords: test size; Bernoulli; geometric

1. INTRODUCTION

The development of hypothesis testing was a major accomplishment of twentieth century statistics. As the theory was widely applied, the weaknesses of the method became apparent. Too often, results that tested as unusual could not be replicated. The statistics profession began to move toward testing using Bayesian methods. This paper looks at hypothesis testing from the view of why false positives occur so often and what can be done to improve replicability.

2. SOME BACKGROUND

An early synopsis of hypothesis testing was written in the 1950's by Lehmann (1986). Many types of hypothesis tests have nice statistical properties, as can be seen in Lehmann. This paper will cover tests for which there is a null hypothesis and one alternative hypothesis (the hypotheses can be compound hypotheses). Only unbiased tests are covered, where unbiased tests are defined as those tests for which the probability of rejecting the null hypothesis, given that the null hypothesis is true, is less than or equal to the probability of rejecting the null hypothesis given that the null hypothesis is false; see Lehmann (1986, pp. 12-13, ex. 11). In what follows, no attempt will be made to develop a new test. Instead, the criteria for setting the size of the test will be changed.

3. THE HYPOTHESIS TEST AS A BERNOULLI RANDOM VARIABLE

Assume there is a test statistic used to test a hypothesis. Assume there exists an area for which, if the test statistic falls in the area (the rejection region), the null hypothesis is rejected. Then, the probability that the test statistic falls in the rejection region, if the null hypothesis is true, is the defined as the size of the test (usually denoted by 'α'), and the test result is significant if the test statistic falls in the rejection region and not significant if the test statistic does not fall in the rejection region.

Let a random variable, X, take on the value zero if the test gives a non-significant result, and the value one if the test gives a significant result. Then, such a hypothesis test is a Bernoulli random variable, with the probability of success equal to the size of the test if the null hypothesis is true; see Forbes, Evans, Hastings, and Peacock (2010, pp. 53-54). If the null hypothesis is not true, the assumption is that the probability of success is larger than or equal to the size of the test (unbiasedness).

4. THE HYPOTHESIS TEST USING THE GEOMETRIC DISTRIBUTION

The approach in this paper is to use the expected number of trials before a success occurs to set the size of a test, rather than to use the probability of a success. The distribution of the number of Bernoulli trials until just before the first success is one form of the geometric distribution and is used here; see Forbes, et.al. (2010, pp. 114-116). Because the geometric distribution is the discrete version of the exponential distribution, (see Forbes, et.al. (2010, pp. 114-116)), the distribution is a highly skewed and monotonically decreasing distribution. Since the distribution is skewed, a better measure of the central tendency than the mean (the probability of success for a Bernoulli random variable) is the median. For any geometric distribution, the median is far to the left of the mean. The mean is pulled to the right by occasional large observations. See Tollefson (2014, pp. 138-140).

The expected number of trials before a success for identically and independently distributed Bernoulli random trials is one divided by the Bernoulli probability of success, minus one. For example, if the size of the test equals 0.05, the expected number of trials until just before a success equals nineteen. However, the probability of observing a success in less than 20 trials is 0.6415. (Since the geometric distribution is the discrete version of the exponential distribution, the probability that the number of trials until a success is less than the expected value of the Bernoulli distribution will always be about 0.6321; that is, the number will always be much larger than what would be found with a symmetric distribution. See Appendix 1.) As has been observed, there are too many false positives.
The median of a distribution has the property that the probability of observing an observation less than the median equals the probability of observing an observation greater than the median. For discrete distributions, sometimes the median will fall on a value that occurs with probability greater than zero, which is often true with the geometric distribution. The method below can still be used to find the size for a hypothesis test, since the cumulative distribution function of the geometric distribution is used to find the size and the cumulative distribution function can be solved for the size analytically. The discreteness of the distribution should not be a problem.

5. FINDING THE SIZE OF A HYPOTHESIS TEST USING THE MEDIAN AND THE CUMULATIVE PROBABILITY DISTRIBUTION FUNCTION

The geometric distribution is a discrete distribution with a probability distribution function of:

$$f(x) = \alpha(1-\alpha)^x, \quad x = 0, 1, 2, \cdots; \alpha = \mathrm{P}(\mathrm{X}=1)$$

and a cumulative probability distribution function of:

$$F(x) = 1 - (1-\alpha)^{x+1}, \quad x = 0, 1, 2, \cdots; \alpha = \mathrm{P}(\mathrm{X}=1),$$

where alpha is the size of the test and X is a Bernoulli random variable; see Forbes, et.al (2010, pp. 114-116).

The method for using the median rather than the mean to set the size of the test is straightforward. Let the number of trials in the equation of the cumulative distribution function, x, be equal to one divided by the Bernoulli probability of success, minus one. The cumulative probability distribution function can then be used to find the corresponding size for the test by setting the cumulative probability function equal to 0.5 and solving for α:

$$\begin{aligned} 0.5 &= 1 - (1-\alpha)^{\frac{1}{p}} \\ (1-\alpha)^{\frac{1}{p}} &= 0.5 \\ \frac{1}{p}\log(1-\alpha) &= \log(0.5) \\ 1-\alpha &= \exp(p\log(0.5)) \\ \alpha &= 1-(0.5)^p \end{aligned}$$

For example, for the expected number of trials until just before a success occurs to be equal to nineteen, alpha must be set to 0.034. With alpha equal 0.034, fifty percent of the time a false positive occurs before twenty trials and fifty percent of the time in twenty trials or more. In Table 1., some sizes are found for some common probability levels based the median.

Table 1. Sizes for some Common Probability Levels

p	1/p	size
0.10	10	0.067
0.05	20	0.034
0.025	40	0.017
0.01	100	0.0069
0.001	1000	0.00069

6. CONCLUSIONS

The result in this paper is simple. Because the geometric distribution is highly skewed, too often hypothesis tests give significant results that cannot be replicated. Using the median of the geometric distribution to find the size for a hypothesis test is a more conservative approach than using the mean of the Bernoulli distribution and is more intuitive as to what one would expect with a hypothesis test. There is some question of the effect of the discreteness of the geometric distribution on the median, in whether to use the integer of the median or to include the decimal portion. If the result in the paper is adapted into general use, whether the high level of false positives decreases and traditional hypothesis testing becomes more acceptable again remains to be seen.

APPENDIX 1.

The exponential distribution has the probability distribution function

$$f(x) = \lambda \exp(-\lambda x),\ 0 \leq x < \infty,\ \lambda > 0,$$

and the cumulative distribution function

$$F(x) = 1 - \exp(-\lambda x),\ \ 0 \leq x < \infty,\ \lambda > 0,$$

where one divided by lambda is the expected value of the distribution. See Forbes, et.al. (2010, pp. 88-92). Clearly, if x is set equal to the expected value of the distribution in the formula for the cumulative distribution function, then the result has the value of (one) minus (one divided by the exponential), or 0.6321. The result is independent of the value of lambda. So, 0.6321 is the probability of observing a value less than the expected value of the distribution for any exponential distribution.

REFERENCES

Forbes, C. E. (2010). *Statistical Distributions* (4th ed.). New York, NY: John Wiley & Sons.
Lehmann, E. L. (1986). *Testing Statistical Hypotheses* (2nd ed.). New York, NY: John Wiley & Sons.
Tollefson, M. (2014). *R Quick Syntax Reference.* New York, NY: Apress.