Practice in Testing Two Binomial Populations

STAT6750 Statistical Consulting

Hengrui Luo

Department of Statistics
The Ohio State University

November 6, 2016

How to determine the sampling scheme? I

- **Binomial Sampling Scheme**
  - The example of *Lady Tasting Tea*. Milk first or tea first?
  - Conditional model conditioned on total numbers of trials.

Example

(Clinical trials) The number of successes in a sequence of medicine trials for each treatment of a certain disease. The total number of trials is determined before the experiment and hence can be conditioned on.

  - Similar consideration can be applied to Poisson populations under certain circumstances, remember that the Poisson approximation of binomial population when the $np \to \lambda$ as $n \to \infty$. However in that case the total number of trials is not known, instead we study the occurrence in one time unit. [19] p.124

Example

(Radioactivity) The number of particles emitted in one time unit from a radioactive substance can be modeled using a Poisson distribution.

How to determine the sampling scheme? II

- The binomial population can be summarized into a 2×2 contingency table.

Definition

(Contingency Table) In the 2×2 *contingency table* as below, columns are data from different populations and the rows are data from different categories.

	Sample from Population 1	Sample from Population 2	Total
Success	s_1	s_2	$s_1 + s_2$
Failure	f_1	f_2	$f_1 + f_2$
Total	$s_1 + f_1$	$s_2 + f_2$	n

The primitive question we should ask is **the following null hypothesis**
$H_0 : p_1 = p_2 = p$ v.s. $H_1 : p_1 \neq p_2$

How to determine the sampling scheme? III

- **Other Sampling Schemes**

  However, when we summarized the data into a contingency table we actually lost information about how the data is sampled. Possible models for a 2×2 contingency table include [19]p.132:
  Assuming that we are giving different treatments to each subjects from two different populations, namely the treatment group and control group in a design of experiment. And we to test the hypothesis whether the treatment makes difference between populations.

  - **Binomial: Margins $s_1 + f_1, s_2 + f_2$ are fixed; margins $s_1 + s_2; f_1 + f_2$are random OR**
    Margins $s_1 + f_1, s_2 + f_2$ are random; margins $s_1 + s_2; f_1 + f_2$are fixed.
    - ***Each subject receives different treatments randomly.***
  - Poisson: Margins $s_1 + f_1, s_2 + f_2$; margins $s_1 + s_2; f_1 + f_2$; margin n are fixed.
    - *Each subject receives different treatments randomly.*
  - Multinomial: Margins $s_1 + f_1, s_2 + f_2$; margins $s_1 + s_2; f_1 + f_2$are fixed. Margin n is random.
    - *Each subject receives different treatments randomly.*
  - Hypergeometric: Margins $s_1 + f_1, s_2 + f_2$; margins $s_1 + s_2; f_1 + f_2$; margin n are fixed.
    - *Different treatments are randomly assigned to each subject.*
  - Observational Study: No randomness at all.

A History of Binomial Tests I

- From the historical view, we can observe that there are two peaks in the research history of the binomial tests.
  - Around 1900~1920, first tests on binomial proportions are proposed. Pearson and Yates focused on the continuous approximation of binomial populations and based their tests on χ^2 distribution. Fisher focused on randomization/permutation of the data and led to the well-known Fisher's exact test.
    - ★ Pearson χ^2 Test [25, 13]
    - ★ Yates corrected χ^2 Test [29]
    - ★ Fisher exact Test [14]
  - Around 1940~1950, the discussion on the appropriateness of using Fisher's exact test boosted the development of tests on contingency tables. Many new alternative methods were proposed in place of Fisher's exact test.
    - ★ Barnard-Boschloo Test [4]
    - ★ McNemar's Test [22]
    - ★ Cochran Q Test [9]
    - ★ Cochran-Mantel-Haenszel Test [10]
    - ★ Stuart-Maxwell Test [26]
    - ★ Bhapkar Test [7]

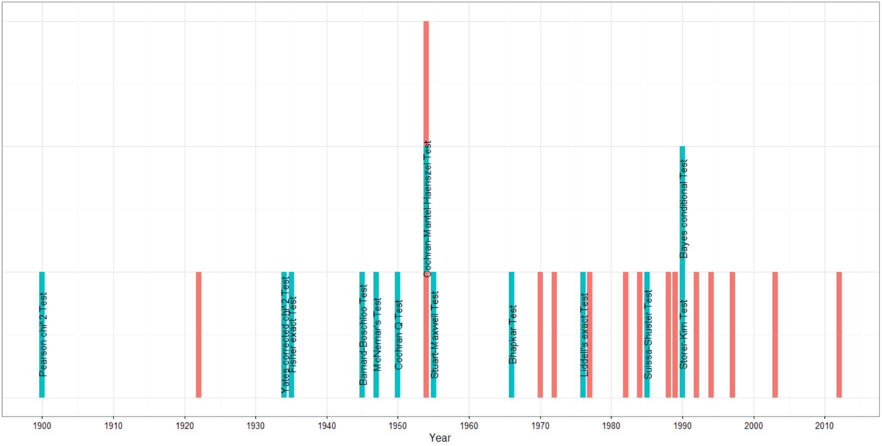
A History of Binomial Tests II

- Around 1980~1990, analyses of power and type I error of the methods proposed earlier and an examination of tests in view of hypothesis testing framework led to more reflections on the basis of earlier tests, the question of appropriateness of the choice of tests for a specific sampling scheme was studied. Also, some Bayes procedure is proposed around this period.
  - Liddell's exact Test [20]
  - Suissa-Shuster Test [27]
  - Storer-Kim Test [24]
  - Bayes conditional Test [2]

A History of Binomial Tests III

- From the hypothesis testing theoretic view, we can categorize all these binomial tests according to how they deal with the nuisance parameters. In setting of contingency tables, the nuisance paramters are the margin sums. [21, 5] This can also be understood in different sampling schemes as we stated previously.
  - Conditional tests: Condition on nuisance parameters.
    - Pearson χ^2 Test [25, 13]
    - Yates corrected χ^2 Test [29]
    - Fisher exact Test [14]
  - Unconditional tests: Discuss the "worst case scenario" for the value of nuisance parameters.
    - Barnard-Boschloo Test [4]
    - Suissa-Shuster Test [27]
    - Storer-Kim Test [24]
  - Bayes tests: Put an prior on nuisance parameters.
    - Bayes conditional Test [2]

Figure: Timeline of researches on binomial tests. Green color represents the fact that a new method is proposed.



Choice of Tests with Examples I

- Conditional tests are suitable for designed experiments. Because in designed experiments the margins are often part of the design and fixed.
  - Pros: The conditional tests have good power and type I error when the underlying sampling scheme is correct. They do not suffer from computation obstacle even if the data is a large number because they based on a known distribution.
  - Cons: It is usually hard to verify whether the underlying sampling scheme is correct and it is hard to maintain the overall error rate when there are a lot of tests. Neither Bonferroni nor FDR adjustment is a good choice.

Choice of Tests with Examples II

Example

(Agriculture Experiments) Two plots were randomly chosen for experiment. One plot is treated with pesticide and another is not treated with pesticide. With pre-planned number of corn seeds planted in each plot, we can get counting data after the corn seeds are all yielded. We analyzed the effect of the pesticide on the growth and yield of corn.

	Plot 1(pesticide) seeds	Plot 2 seeds	Total
Success(grow and yield)	234	101	$s_1 + s_2$
Failure	78	90	$f_1 + f_2$
Total	312(fixed)	191(fixed)	503

The results from each conditional tests:

	Fisher	Pearson	Yates correction
p-value	5.665e-07	3.308e-07	5.512e-07

Choice of Tests with Examples III

- Unconditional tests are generally more powerful than conditional ones [6, 1, 23]. However unconditional tests usually involves a maximization procedure when calculating p-value, which makes them less favorable when a large number of hypothesis tests need to be conducted. Also, its calculation of p-values cannot handle large numbers very easily because of the combinatorial number caused by permutations.
  Therefore it is usually the approximate version of unconditional tests instead of exact version of unconditional tests are used for large datasets like genome-wide data.

Choice of Tests with Examples IV

Example

(Genetic data) The DNA methylation sequence is read from chemical techniques in form of counting data. More counts at a location mean the higher level of methylation is at a certain location. Researchers want to determine whether the methylation level at a specific location varies from person to person. At one specific location the methylation data from two individuals can be represented in a contingency table: (in fact it is not very suitable to be described as a binomial population.)

	Individual 1 DNA	Individual 2 DNA	Total
Success(detected)	35	23	?
Failure(not detected)	3	10	?
Total	?	?	n

Chemical techniques cannot ensure the margins are fixed. Therefore we choose unconditional tests:

	Barnard-Boschloo	Suissa-Shuster	Storer-Kim
p-value	0.02151	0.1129	0.02146102

Choice of Tests with Examples V

- Bayes tests relies on simulation to yield a posterior interval. Only under certain circumstances [2] and particular choice of priors (mixture of Dirichlet distributions) can we use Bayes factor to make inference. This makes such an inference not very desirable and requires consideration about the prior before put into use.

References I

[1] Agresti, Alan. "A survey of exact inference for contingency tables." Statistical Science (1992): 131-153.

[2] Albert, James H. "A bayesian test for a two-way contingency table using independence priors." Canadian Journal of Statistics 18.4 (1990): 347-363.

[3] Andrés, A. Martín, and A. Silva Mato. "Choosing the optimal unconditioned test for comparing two independent proportions." Computational statistics & data analysis 17.5 (1994): 555-574.

[4] Barnard, G. A. "A new test for 2$\times$ 2 tables." Nature 156 (1945): 177.

[5] Basu, Debabrata. "On the elimination of nuisance parameters." Journal of the American Statistical Association 72.358 (1977): 355-366.

[6] Berger, R. "Power comparison of exact unconditional tests for comparing two binomial proportions." Institute of Statistics Mimeo Series 2266 (1994).

[7] Bhapkar, Vasant P. "A note on the equivalence of two test criteria for hypotheses in categorical data." Journal of the American Statistical Association 61.313 (1966): 228-235.

References II

[8] Chernoff, Herman, and E. L. Lehmann. "The use of maximum likelihood estimates in χ^2 tests for goodness of fit." The Annals of Mathematical Statistics 25.3 (1954): 579-586.

[9] Cochran, William G. "The comparison of percentages in matched samples." Biometrika 37.3/4 (1950): 256-266.

[10] Cochran, William G. "Some methods for strengthening the common χ 2 tests." Biometrics 10.4 (1954): 417-451.

[11] Cramér, Harald. Mathematical methods of statistics. Vol. 9. Princeton university press, 1945.

[12] D'agostino, Ralph B., Warren Chase, and Albert Belanger. "The appropriateness of some common procedures for testing the equality of two independent binomial populations." The American Statistician 42.3 (1988): 198-202.

[13] Fisher, Ronald A. "On the interpretation of χ 2 from contingency tables, and the calculation of P." Journal of the Royal Statistical Society 85.1 (1922): 87-94.

References III

[14] Fisher, Sir Ronald Aylmer, et al. "The design of experiments." (1960).

[15] Greenland, Sander. "On the logical justification of conditional tests for two-by-two contingency tables." The American Statistician 45.3 (1991): 248-251.

[16] Howell, David C. "Chi-square test: analysis of contingency tables." International Encyclopedia of Statistical Science. Springer Berlin Heidelberg, 2011. 250-252.

[17] Krishnamoorthy, K., and Jessica Thomson. "Hypothesis testing about proportions in two finite populations." The American Statistician (2012).

[18] Lehmann, Erich Leo, Testing statistical hypotheses. Ed. George Casella. Vol. 150. New York et al: Wiley, 1986.

[19] Lehmann, Erich Leo, and Joseph P. Romano. Testing statistical hypotheses. Ed. George Casella. Vol. 150. New York et al: Wiley, 2008.

[20] Liddell, Douglas. "Practical tests of $2\times$ 2 contingency tables." The Statistician (1976): 295-304.

References IV

[21] Little, Roderick JA. "Testing the equality of two independent binomial proportions." The American Statistician 43.4 (1989): 283-288.

[22] McNemar, Quinn. "Note on the sampling error of the difference between correlated proportions or percentages." Psychometrika 12.2 (1947): 153-157.

[23] Mehrotra, Devan V., Ivan SF Chan, and Roger L. Berger. "A cautionary note on exact unconditional inference for a difference between two independent binomial proportions." Biometrics 59.2 (2003): 441-450.

[24] Storer, Barry E., and Choongrak Kim. "Exact properties of some exact test statistics for comparing two binomial proportions." Journal of the American Statistical Association 85.409 (1990): 146-155.

[25] Pearson, Karl. "X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling." The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 50.302 (1900): 157-175.

References V

[26] Stuart, Alan. "A test for homogeneity of the marginal distributions in a two-way classification." Biometrika 42.3/4 (1955): 412-416.

[27] Suissa, Samy, and Jonathan J. Shuster. "Exact unconditional sample sizes for the $2\times$ 2 binomial trial." Journal of the Royal Statistical Society. Series A (General) (1985): 317-327.

[28] Upton, Graham JG. "A comparison of alternative tests for the $2\times$ 2 comparative trial." Journal of the Royal Statistical Society. Series A (General) (1982): 86-105.

[29] Yates, Frank. "Contingency tables involving small numbers and the χ^2 test." Supplement to the Journal of the Royal Statistical Society 1.2 (1934): 217-235.