The Odds Ratio Uniformity Test

Alan Kaminsky  
Department of Computer Science  
B. Thomas Golisano College of Computing and Information Sciences  
Rochester Institute of Technology  
ark@cs.rit.edu  
February 2, 2017

Introduction

The odds ratio uniformity test is an alternative to frequentist statistical tests such as the chi-square test. A strong point of the odds ratio uniformity test is that the results of multiple independent tests can easily be aggregated to yield a single overall result. The odds ratio uniformity test uses the methodology of Bayesian model selection applied to binomial distributions. For more information about Bayesian model selection, see [Kass1995].

Bayes Factors and Odds Ratios

Let $H$ denote a hypothesis, or model, describing some process. Let $D$ denote an experimental data sample, or just sample, observed by running the process. Let $\Pr(H)$ be the probability of the model. Let $\Pr(D|H)$ be the conditional probability of the sample given the model. Let $\Pr(D)$ be the probability of the sample, apart from any particular model. Bayes’s Theorem states that $\Pr(H|D)$, the conditional probability of the model given the sample, is

$$
\Pr(H|D) = \frac{\Pr(D|H) \Pr(H)}{\Pr(D)}.
$$

(1)

Suppose there are two alternative models $H_1$ and $H_2$ that could describe a process. After observing sample $D$, the posterior odds ratio of the two models, $\Pr(H_1|D)/\Pr(H_2|D)$, is calculated from Equation (1) as

$$
\frac{\Pr(H_1|D)}{\Pr(H_2|D)} = \frac{\Pr(D|H_1)}{\Pr(D|H_2)} \cdot \frac{\Pr(H_1)}{\Pr(H_2)},
$$

(2)

where the term $\Pr(H_1)/\Pr(H_2)$ is the prior odds ratio of the two models, and the term $\Pr(D|H_1)/\Pr(D|H_2)$ is the Bayes factor. The odds ratio represents one’s belief about the relative probabilities of the two models. Given one’s initial belief before observing any samples (the prior odds ratio), the Bayes factor is used to update one’s belief after performing an experiment and observing a sample (the posterior odds ratio). Stated simply, posterior odds ratio = Bayes factor × prior odds ratio.

Suppose two experiments are performed and two samples, $D_1$ and $D_2$, are observed. Assuming the samples are independent, it is straightforward to calculate that the posterior odds ratio based on both samples is
\[
\frac{\operatorname{pr}(H_1 | D_2, D_1)}{\operatorname{pr}(H_2 | D_2, D_1)} = \frac{\operatorname{pr}(D_2 | H_1)}{\operatorname{pr}(D_2 | H_2)} \cdot \frac{\operatorname{pr}(H_1 | D_1)}{\operatorname{pr}(H_2 | D_1)} = \frac{\operatorname{pr}(D_2 | H_1)}{\operatorname{pr}(D_2 | H_2)} \cdot \frac{\operatorname{pr}(D_1 | H_1)}{\operatorname{pr}(D_1 | H_2)} \cdot \frac{\operatorname{pr}(H_1)}{\operatorname{pr}(H_2)}.
\]

(3)

In other words, the posterior odds ratio for the first experiment becomes the prior odds ratio for the second experiment. Equation (3) can be extended to any number of independent samples \(D_i\); the final posterior odds ratio is just the initial prior odds ratio multiplied by all the samples’ Bayes factors.

Model selection is the problem of deciding which model, \(H_1\) or \(H_2\), is better supported by a series of one or more samples \(D_i\). In the Bayesian framework, this is determined by the posterior odds ratio (3). Henceforth, “odds ratio” will mean the posterior odds ratio. If the odds ratio is greater than 1, then \(H_1\)’s probability is greater than \(H_2\)’s probability, given the data; that is, the data supports \(H_1\) better than it supports \(H_2\). The larger the odds ratio, the higher the degree of support. An odds ratio of 100 or more is generally considered to indicate decisive support for \(H_1\) [Kass1995]. If on the other hand the odds ratio is less than 1, then the data supports \(H_2\) rather than \(H_1\), and an odds ratio of 0.01 or less indicates decisive support for \(H_2\).

Models With Parameters

In the preceding formulas, the models had no free parameters. Now suppose that model \(H_1\) has a parameter \(\theta_1\) and model \(H_2\) has a parameter \(\theta_2\). Then the conditional probabilities of the samples given each of the models are obtained by integrating over the possible parameter values [Kass1995]:

\[
\operatorname{pr}(D | H_1) = \int \operatorname{pr}(D | \theta_1, H_1) \, \pi(\theta_1 | H_1) \, d\theta_1,
\]

(4)

\[
\operatorname{pr}(D | H_2) = \int \operatorname{pr}(D | \theta_2, H_2) \, \pi(\theta_2 | H_2) \, d\theta_2,
\]

(5)

where \(\operatorname{pr}(D | \theta_1, H_1)\) is the probability of observing the sample under model \(H_1\) with the parameter value \(\theta_1\), \(\pi(\theta_1 | H_1)\) is the prior probability density of \(\theta_1\) under model \(H_1\), and likewise for \(H_2\) and \(\theta_2\). The Bayes factor is then the ratio of these two integrals.

Odds Ratio for Binomial Models

Suppose an experiment performs \(n\) Bernoulli trials, where the probability of success is \(\theta\), and counts the number of successes \(k\), which obeys a binomial distribution. The values \(n\) and \(k\) constitute the sample \(D\). With this as the model \(H\), the probability of \(D\) given \(H\) with parameter \(\theta\) is

\[
\operatorname{pr}(D | \theta, H) = \binom{n}{k} \theta^k (1-\theta)^{n-k} = \frac{n!}{k! (n-k)!} \theta^k (1-\theta)^{n-k}.
\]

(6)

Consider the odds ratio for two particular binomial models, \(H_1\) and \(H_2\). \(H_1\) is that the Bernoulli success probability \(\theta_1\) is a certain value \(p\), the value that the success probability is “supposed” to have. Then the prior probability density of \(\theta_1\) is a delta function, \(\pi(\theta_1 | H_1) = \delta(\theta_1 - p)\), and the Bayes factor numerator (4) becomes

\[
\operatorname{pr}(D | H_1) = \frac{n!}{k! (n-k)!} \, p^k (1-p)^{n-k}.
\]

(7)
The Odds Ratio Uniformity Test

$H_2$ is that the Bernoulli success probability $θ_2$ is some unknown value between 0 and 1, not necessarily the value it is “supposed” to have. The prior probability density of $θ_2$ is taken to be a uniform distribution: $π(θ_2|H_2) = 1$ for $0 ≤ θ_2 ≤ 1$ and $π(θ_2|H_2) = 0$ otherwise. The Bayes factor denominator (5) becomes

$$
pr(D|H_2) = \int_0^1 \frac{n!}{k!(n-k)!} \theta_2^k (1-\theta_2)^{n-k} \, d\theta_2 = \frac{1}{n+1} .
$$

(8)

Putting everything together, the Bayes factor for the two binomial models is

$$
\frac{pr(D|H_1)}{pr(D|H_2)} = \frac{(n+1)!}{k!(n-k)!} p^k (1-p)^{n-k} .
$$

(9)

Substituting the gamma function for the factorial, $n! = \Gamma(n+1)$, gives

$$
\frac{pr(D|H_1)}{pr(D|H_2)} = \frac{\Gamma(n+2)}{\Gamma(k+1) \Gamma(n-k+1)} p^k (1-p)^{n-k} .
$$

(10)

Because the gamma function’s value typically overflows the range of floating point values in a computer program, we compute the logarithm of the Bayes factor instead of the Bayes factor itself:

$$
\log \frac{pr(D|H_1)}{pr(D|H_2)} = \log \Gamma(n+2) - \log \Gamma(k+1) - \log \Gamma(n-k+1) + k \log p + (n-k) \log(1-p) .
$$

(11)

The log-gamma function can be computed efficiently (see [Press2007] page 256), and mathematical software libraries usually include log-gamma.

**Odds Ratio Test**

The above experiment can be viewed as a test of whether $H_1$ is true, that is, whether the success probability is $p$. The log (posterior) odds ratio of the models $H_1$ and $H_2$ is the log prior odds ratio plus the log Bayes factor (11). Assuming that $H_1$ and $H_2$ are equally probable at the start, the log odds ratio is just the log Bayes factor. The test passes if the log odds ratio is greater than 0, otherwise the test fails.

When multiple independent runs of the above experiment are performed, the overall log odds ratio is the sum of all the log Bayes factors. In this way, one can aggregate the results of a series of individual tests, yielding an overall odds ratio test. Again, the aggregate test passes if the overall log odds ratio is greater than 0, otherwise the aggregate test fails.

Note that the odds ratio test is not a frequentist statistical test that is attempting to disprove some null hypothesis. The odds ratio test is just a particular way to decide how likely or unlikely it was that a series of observations came from a Bernoulli($p$) distribution, by calculating a posterior odds ratio. While a frequentist statistical test could be defined based on odds ratios, I am not doing that here.

**Odds Ratio Uniformity Test**

Consider a random variable $X$ with a discrete uniform distribution. The variable has $B$ different possible values (“bins”), $0 ≤ x ≤ B − 1$. An experiment with $n$ trials is performed. In each trial, the
The Odds Ratio Uniformity Test

random variable's value is observed, and a counter for the corresponding bin is incremented. If the variable obeys a discrete uniform distribution, all the counters should end up the same.

The odds ratio uniformity test calculates the odds ratio of two hypotheses: $H_1$, that $X$ obeys a discrete uniform distribution, and $H_2$, that $X$ does not obey a discrete uniform distribution. To do so, first calculate the observed cumulative distribution of $X$ and the expected cumulative distribution of $X$ under model $H_1$. The observed cumulative distribution is

$$F_{\text{obs}}(x) = \sum_{i=0}^{x} \text{counter}[x], \quad 0 \leq x \leq B - 1,$$

and the expected cumulative distribution is

$$F_{\text{exp}}(x) = \frac{(x+1)n}{B}, \quad 0 \leq x \leq B - 1.\quad (13)$$

Let $y$ be the bin such that the absolute difference $|F_{\text{obs}}(y) - F_{\text{exp}}(y)|$ is maximized.* The trials are now viewed as Bernoulli trials, where incrementing a bin less than or equal to $y$ is a success, the observed number of successes in $n$ trials is $k = F_{\text{obs}}(y)$, and the success probability is $p = F_{\text{exp}}(y)/n = (y + 1)/B$ if $H_1$ is true. An odds ratio test for a discrete uniform distribution ($H_1$ versus $H_2$) is therefore equivalent to an odds ratio test for this particular binomial distribution, with Equation (11) giving the log Bayes factor. If the log Bayes factor is greater than 0, then $X$ obeys a discrete uniform distribution, otherwise $X$ does not obey a discrete uniform distribution.

A couple of examples will illustrate the odds ratio uniformity test. I queried a pseudorandom number generator one million times; each value was uniformly distributed in the range 0.0 (inclusive) through 1.0 (exclusive); I multiplied the value by 10 and truncated to an integer, yielding a bin $x$ in the range 0 through 9; and I accumulated the values into 10 bins, yielding this data:

| $x$ | counter[$x$] | $F_{\text{obs}}(x)$ | $F_{\text{exp}}(x)$ | $|F_{\text{obs}}(x) - F_{\text{exp}}(x)|$ |
|-----|--------------|---------------------|---------------------|---------------------------------|
| 0   | 99476        | 99476               | 100000             | 524                             |
| 1   | 100498       | 199974              | 200000             | 26                              |
| 2   | 99806        | 299780              | 300000             | 220                             |
| 3   | 99881        | 399661              | 400000             | 339                             |
| 4   | 99840        | 499501              | 500000             | 499                             |
| 5   | 99999        | 599500              | 600000             | 500                             |
| 6   | 99917        | 699417              | 700000             | 583                             |
| 7   | 100165       | 799582              | 800000             | 418                             |
| 8   | 100190       | 899772              | 900000             | 228                             |
| 9   | 100228       | 1000000             | 1000000            | 0                               |

The maximum absolute difference between the observed and expected cumulative distributions occurred at bin 6. With $n = 1000000$, $k = 699417$, and $p = 0.7$, the log Bayes factor is 5.9596. In other words, the odds are about $\exp(5.9596) = 387$ to 1 that this data came from a discrete uniform distribution.

I queried a pseudorandom number generator one million times again, but this time I raised each value to the power 1.01 before converting it to a bin. This introduced a slight bias towards smaller bins. I got this data:

* This is similar to what is done in a Kolmogorov-Smirnov test for a continuous uniform distribution.
The Odds Ratio Uniformity Test

The maximum absolute difference between the observed and expected cumulative distributions occurred at bin 2. With $n = 1000000$, $k = 303360$, and $p = 0.3$, the log Bayes factor is $-20.057$. In other words, the odds are about $\exp(-20.057) = 514$ million to 1 that this data did not come from a discrete uniform distribution.

The odds ratio test can be applied to any discrete distribution, not just a discrete uniform distribution. Just substitute, for Equation (13), the cumulative distribution function of the expected distribution.

**References**
