Likelihood Ratio Test (LRT) Methodology

Document Version Number: 1.0
Document Version Date: September 2014

In statistics, a likelihood ratio test is a statistical test used to compare the fit of two models, one of which is the null model and the other is the alternative model. The test is based on the likelihood ratio, which expresses how many times more likely the data are under one model than the other. This likelihood ratio, or equivalently its logarithm, can then be used to compute a p-value, or compared to a critical value to decide whether to reject the null model in favor of the alternative model. When the logarithm of the likelihood ratio is used, the statistic is known as a log-likelihood ratio statistic, and the probability distribution of this test statistic, assuming that the null model is true, can be approximated.

Usually large drug safety databases, such as FAERS, include thousands of drugs and AEs, which can be presented in a data matrix with *I* rows (AEs) and *J* columns (drugs). For each AE-drug combination (cell (i, j)) in the data matrix, the number of cases reported is defined as n_{ij} . Let nidot and ndotj, ($n_{i.}, n_{.j}$) , be the marginal totals (i.e., the total number of reports) for the ith AE and jth drug; and let ndotdot ($n_{..}$) be the (grand) total number of reports. The large *I×J* matrix is shown below in Table 1

Table 1: I×J data matrix

		Drugs					
AEs		1	…	j	…	J	Row total
	1	n_{11}	…	n_{1j}	…	n_{1J}	$n_{1.}$
	2	n_{21}	…	n_{2j}	…	n_{2J}	$n_{2.}$
	…	…	…	…	…	…	…
	I	n_{i1}	…	n_{ij}	…	n_{iJ}	$n_{i.}$
	…	…	…	…	…	…	…
	I	n_{I1}	…	n_{Ij}	…	n_{IJ}	$n_{I.}$
	Column total	$n_{.1}$	…	$n_{.j}$	…	$n_{.J}$	$n_{..}$

For a (fixed) drug j of interest, the I×J data-matrix can be collapsed into I, 2×2 tables, each corresponding to an AE (as shown in Table 2).

Table 2: One 2×2 table and related notations

	Drug j	Other Drugs	Row total
AE i	n_{ij} =a	$(n_{i.} - n_{ij})$ =b	$n_{i.} = a + b$
Other AEs	$(n_{.j} - n_{ij})$ =c	$(n_{..} - n_{i.} - n_{.j} + n_{ij})$ =d	$n_{..} - n_{i.} = c + d$
Column total	$n_{.j}$ =a+c	$n_{..} - n_{.j}$ =b+d	$n_{..} = a + b + c + d$

The likelihood ratio for AE i and Drug j:

$$LR_{ij} = \frac{\left(\frac{n_{ij}}{n_{i.}}\right)^{n_{ij}} \left(\frac{n_{.j} - n_{ij}}{n_{..} - n_{i.}}\right)^{n_{.j} - n_{ij}}}{\left(\frac{n_{.j}}{n_{..}}\right)^{n_{.j}}},$$

Or on logarithmic scale,

$$\log LR_{ij} =$$
$$a \times [\log(a) - \log(a+b)] + c \times [\log(c) - \log(c+d)] - (a+c) \times [\log(a+c) - \log(a+b+c+d)].$$

The maximum likelihood ratio (MLR) test statistic is $\max_i LR_{ij}$, where the maximum is taken over $i = 1, \ldots, I$.

Since logarithm $\log(LR_{ij})$ is a monotonic (increasing) function of LR_{ij}, it is convenient to work with

$$MLLR = \max_i (\log LR_{ij}).$$

The distribution of $MLLR$ test statistic under the null hypothesis is intractable, and a Monte Carlo procedure, outlined next, is used to obtain its empirical distribution. Under null hypothesis, the conditional distribution of $(n_{1j}, \ldots, n_{Ij})$, conditional on the sum $n_{.j}$, is a multinomial distribution with parameters $n_{.j}$ and probabilities $(n_{1.}/n_{..}, \ldots, n_{I.}/n_{..})$; that is,

$$(n_{1j}, \ldots, n_{Ij}) \mid n_{.j} \sim Mult(n_{.j}, (\frac{n_{1.}}{n_{..}}, \ldots, \frac{n_{I.}}{n_{..}})).$$

The empirical distribution of $MLLR$ under the null hypothesis can now be obtained by generating a large number of Monte Carlo samples for the cell-report counts $(n_{1j},...,n_{Ij})$, for the j^{th} column in the $I \times J$ table on the $Drug_j$, using this multinomial distribution. If the $MLLR$ based on the observed data, $MLLR_{data}$, is greater than the threshold value of $MLLR_{0.05}$ (the upper 5^{th} percentile point of the empirical distribution) the null hypothesis is rejected with alpha=0.05. The AE associated with $MLLR_{data}$ is then the most significant signal detected. This is the AE with the largest value of logLR. We can identify the other AE signals (secondary signals) as well, if their logLR values are greater than $MLLR_{0.05}$. This process controls the type-I error and the false discovery rate (Huang et al. 2011).

More details of the method can be found in:

- *Lan Huang, Jyoti Zalkikar, and Ram Tiwari. A likelihood based method for signal detection with application to FDA's drug safety data. Journal of the American Statistical Association (JASA), 2011*