The Korean Communications
in Statistics Vol. 5 No. 3, 1998
pp. 905-926

Improved Mid P-value Method for Statistical Inference in Three-Way Contingency Tables

Donguk Kim[1]

Abstract

We propose a modified mid P-value method to reduce the conservativeness for the inference of conditional associations in three-way contingency tables. This improves the ordinary mid P-value method. For $2 \times 2 \times K$ tables, we propose confidence intervals for an assumed common odds ratio based on inverting two separate one-sided tests using the modified mid P-value. Also, an alternative and usually even better ways of constructing intervals, based on inverting a two-sided test, are presented. The actual probability of coverage of a $100(1-\alpha)\%$ confidence interval is centered about the nominal level, but the modified mid P-value approach gives actual coverage probability even closer to the nominal level than the ordinary mid P-value approach.

1. Introduction

When the exact distribution of the test statistic is discrete, it is known that ordinary exact tests and confidence intervals can be highly conservative. Though exact tests are guaranteed to control the probability of Type I error at any nominal level, we may not achieve a probability of Type I error of the nominal level exactly. The actual probability of Type I error may be considerably smaller. For exact inference about a parameter of interest, we condition on sufficient statistics for unknown parameters to eliminate them. This extra conditioning makes the distribution of the test statistic more highly discrete.

For three-way $I \times J \times K$ tables, consider the hypothesis of conditional independence of two variables, given the third one. Let $N = \{n_{ijk}\}$ denote the cell counts, with expected frequencies $\{m_{ijk}\}$. For this hypothesis, we discuss conditional tests that reduce the conservativeness, generalizing Fisher's exact test for 2×2 tables. We also discuss confidence intervals for odds ratios pertaining to conditional association. Let X, Y, and Z denote the row, column, and layer classification, respectively. The hypothesis of conditional independence of X and Y, given Z, is usually tested against the alternative of no

1) Assistant Professor, Department of Statistics, Sungkyunkwan University, Seoul, 110-745, Korea.

three-factor interaction. This alternative is the loglinear model of form

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}, \quad (1.1)$$

having sufficient statistics $(\{n_{ij+}\}, \{n_{i+k}\}, \{n_{+jk}\})$. The subscript "+" denotes the sum over the index it replaces. The null hypothesis corresponds to the special case of this model in which all $\lambda_{ij}^{XY} = 0$. Exact conditional tests utilize the distribution of $\{n_{ij+}\}$, the sufficient statistics for these parameters, conditional on the other sufficient statistics, that relate to the remaining parameters.

One way to reduce conservativeness is the mid P adjustment. According to Lancaster (1961), the mid P adjustment utilizes half of the probability of the observed statistic. This reduces the conservativeness due to discreteness and does not rely on randomization to eliminate the conservativeness. But one drawback is that it can not guarantee exactness, in the sense that the actual size possibly exceeds the nominal level, and its coverage of the confidence interval at $100(1-\alpha)\%$ level is no longer guaranteed theoretically. It comes from the fact that the mid P approach subtracts half of the probability of the observed statistic from the exact P-value. Mehta and Walsh (1992) conducted an extensive simulation study to study the mid P confidence intervals. Barnard (1990) gave a philosophical discussion of P-values including mid P. Also, Vollset, Hirji, and Afifi (1991) suggested mid P method than the ordinary exact method in a stratified logistic regression model. Kim and Agresti (1995) proposed modified exact inferential methods to reduce the conservativeness and showed that two modified confidence intervals using modified exact P-value maintain at least a fixed confidence level but tend to be much narrower.

In this paper, we discuss a modified mid P-value. The modified mid P-value utilizes the idea of modified P-value in Kim and Agresti (1995). That is, it utilizes both the usual test statistic and, at the observed value of that statistic, a supplementary statistic T' directed toward a broad alternative. By using this approach, we show that the modified mid P method improves the ordinary mid P method, that is, it is less discrete than the ordinary one, and leads to less conservative, and is even closer to the nominal level than the ordinary mid P in the probability of coverage.

Section 2 discusses modified mid P-value for testing conditional independence against the alternative of no three-factor interaction. We show the modified mid P-value is less discrete, and leads to a less conservative test. By inverting results of tests using modified mid P-values, we will obtain a less conservative confidence interval, in the sense that the modified confidence interval has confidence coefficient closer to the nominal level and can be narrower than the ordinary one. In Section 3 we propose a modified confidence interval for a common odds ratio in $2 \times 2 \times K$ tables inverting the test based on a modified one-sided mid P-value. Section 4 presents an alternative and usually even better way of constructing confidence intervals, based on inverting a two-sided test with a modified mid P-value. Computation of coverage functions can be a useful graphical diagnostic tool for assessing the appropriateness of methods. To compare these types of intervals, we calculate actual coverage probability of the confidence intervals based on inverting one-sided or two-sided tests using the ordinary or modified mid P-value. If those coverage probability of the confidence intervals using the modified mid P-value achieves the nominal level closely when the exact test is conservative, then the modified mid P test will be an excellent alternative to the exact test. In Section 5 we discuss general ways of choosing a secondary statistic for testing conditional independence in $I \times J \times K$ contingency tables.

2. Modified Mid P-value

Suppose we would like to conduct an exact conditional test for categorical data using some preassigned size α. Denote by Γ the set of contingency tables having the same marginal counts as the ones that are fixed by the conditioning argument for the exact conditional test. For the test of conditional independence, for instance, Γ is the set of $I \times J \times K$ tables of nonnegative integers, $\Gamma = \{Z : \sum_i z_{ijk} = n_{+jk}, \ \sum_j z_{ijk} = n_{i+k}, \quad \text{for all} \ i, j, k\}$. In general, suppose we have a test statistic T, such as a Wald, likelihood ratio, or score statistic, and suppose t_o is the observed value of T. If large values of T contradict the null, the usual P-value is $P = P_{H_0}(T \geq t_o)$. In the exact conditional approach, one conditions on sufficient statistics for unknown parameters in order to eliminate them. The extra conditioning reduces the set of possible test statistic values, making the distribution more highly discrete. Hence, tests of nominal size α based on the exact conditional P-value can be even more conservative. The actual probability of Type I error can be considerably less than the nominal value unless the sample size is reasonably large.

If an exact test is desired of arbitrary size α, supplementary randomization would be required. A P-value corresponding to a test using supplementary randomization has the form

$$P_U = P_{H_0}(T > t_o) + U P_{H_0}(T = t_o), \tag{2.1}$$

where U denotes a uniform (0,1) random variable. This approximates the tail area by a random proportion of the probability of the observed value of T. It will have actual Type I error probability of the nominal level. Even though such an adjustment would eliminate the conservativeness completely, the inference is based on randomized decision rules and such a randomized test is usually unacceptable.

One modification that aims at reducing the conservative bias of the exact method without supplementary randomization is mid P-value (Lancaster 1961) method. The mid P-value is an alternative to the usual P-value that many statisticians have recommended as a way of compromising between having a conservative test and using supplementary randomization (*e.g.*, Barnard 1990). It is defined as

$$P_{\text{mid}} = P_{H_0}(T > t_o) + (1/2)\, P_{H_0}(T = t_o). \tag{2.2}$$

It subtracts half of the probability of the observed statistic from the usual exact P-value and replaces U in (2.1) by its expectation. The mid P-value has the appealing property that its null expected value for a discrete distribution equals exactly $1/2$, the expected P-value for a continuous distribution. A disadvantage is that a test based on it is no longer exact, the actual size possibly exceeding the nominal value. It is because the mid P-value subtracts half of the probability of the observed statistic from the exact P-value.

By utilizing the modified P-value approach by Kim and Agresti (1995), we can improve the mid P-value so that the modified mid P-value approach reduces the conservativeness and also is even closer to the exactness. The modified mid P-value uses a partition of the sample space that is more refined than we get using T alone, like the modified P-value. That is, T constructs a primary partitioning of all tables that have the sufficient statistics fixed by the conditional test. Then, within fixed values of T, T' generates a secondary partitioning using some other index of the degree to which the data contradict the null hypothesis. The statistic T' is a test statistic directed toward a somewhat broader alternative hypothesis. Let t_o and t_o' denote the observed values of the primary and secondary statistics, respectively. The mid P-value assigns weight $1/2$ to probabilities of all tables comparable to the observed table in the sense that $T = t_o$. For the modified mid P-values, the comparable tables are those with $T = t_o$ and $T' = t_o'$. Thus, we can define the modified mid P-value by

$$P^*_{\text{mid}} = P_{H_0}(T > t_o) + P_{H_0}(T = t_o, T' > t_o') + (1/2) P_{H_0}(T = t_o, T' = t_o'). \tag{2.3}$$

Hence this is the sum of the probability of more extreme values of T and more extreme values of T' at $T = t_o$ and half of $P_{H_0}(T = t_o, T' = t_o')$.

Like the ordinary mid P-value, the modified mid P-value has null expected value equal to $1/2$. The result is easily obtained by noting that the modified mid P-value is a special case of the usual mid P-value using a more refined partitioning of T and T'. The ordinary mid P-value uses a partitioning based on T. The modified mid P-value uses a partitioning based on T and T' within T. We assume that T and T' have positive values. Let $\text{Gap}(T)$ denote the minimum difference between two consecutive values of T. Define a new statistic $T^* = T \times \text{Max}(T')/\text{Gap}(T) + T'$. If $\text{Min}(T')$ equals 0, we transform from T' to $T'+1$ in order to avoid ties in T^*. Then, $T^*(Z_1) > T^*(Z_2)$ for all tables Z_1, Z_2 with $T(Z_1) > T(Z_2)$. Let t^*_o denote the value of T^* for the observed table. Note that a partitioning of the sample space using T and T' within T is equivalent to a partitioning of the sample space using T^*. Since there are no ties, ordering tables using T and T' within T is equivalent to ordering tables using T^*. Then, the sum of half of

$P_{H_0}(T=t_o, T'=t_o')$ and the probability of more extreme values of T' at $T=t_o$ and more extreme values of T is equivalent to the sum of half of $P_{H_0}(T^*=t_o^*)$ and the probability of more extreme values of T^*. That is,

$$P^*_{\text{mid}} = P_{H_0}(T^* > t_o^*) + (1/2)P_{H_0}(T^* = t_o^*). \tag{2.4}$$

Hence, the modified mid P-value is a special case of the mid P-value with a more refined partitioning, and its null expected value is equal to $1/2$.

As the mid P-value subtracts half of the probability of the observed statistic from the ordinary exact P-value, the modified mid P-value and the modified P-value (Kim and Agresti, 1995), denoted by P^*, have the following property :

$$P^*_{\text{mid}} = P^* - \frac{1}{2} P_{H_0}(T=t_o, T'=t_o'). \tag{2.5}$$

The modified mid P-value subtracts half of the probability of tables of $T=t_o$ and $T'=t_o'$ from the modified P-value, and the modified P-value is defined as

$$P^* = P_{H_0}(T > t_o) + P_{H_0}(T = t_o,\ T' \geq t_o').$$

This is less conservative and does not allow randomization to eliminate the conservativeness. Another possibility for the secondary partitioning is to use the null table probability. Let $B=\{Z : Z \in \Gamma,\ T=t_o,\ P(Z) \leq P(N)\}$, where the probabilities are computed under the null. The modified P-value is then $P_p^* = P_{H_0}(T > t_o) + P_{H_0}(B)$. The difference between the modified P-value and modified mid P-value is less than the difference between the ordinary P-value and ordinary mid P-value. That is, $(P^* - P^*_{\text{mid}}) \leq (P - P_{\text{mid}})$. One can calculate this modified mid P-value for any test statistic having a discrete distribution.

2.1 Examples

We consider the test of conditional independence in three-way contingency tables under the assumption of no three-factor interaction. We will illustrate the ordinary and modified mid P-values using $2 \times 2 \times 3$, $2 \times 2 \times 4$, and $2 \times 2 \times 18$ contingency tables. For $2 \times 2 \times K$ tables, the exact test utilizes the score statistic as a test statistic $T=\sum_k n_{11k}$, given $\{n_{1+k}, n_{2+k}, n_{+1k}, n_{+2k}\}$. It assumes homogeneity of the odds ratios in the $2 \times 2 \times K$ contingency tables. Then one could use the score statistic for the general alternative for the secondary partitioning. This is simply $T' = \sum_k X_k^2$, where X_k^2 denotes the Pearson statistic for testing independence in the kth partial table.

We illustrate the modified mid P-values using Table 2.1, taken from Mantel (1963). Let P=penicillin level, D=delay, and C=whether cured. Under the assumption of a constant odds ratio θ between D and C at each level of P, we test $H_0: \theta=1$ against $H_a: \theta>1$. Our alternative is the higher cure rate for immediate injection. For the first and last table, the conditional distribution of n_{11k} is degenerate, hence we conduct the test using the three remaining tables. Table 2.2 is taken from Agresti (1990, p. 256). It classifies poliomyelitis cases by age, paralytic status, and by whether the subject had been injected with Salk vaccine. We take four age groups that seem to be relatively sparse. For another example, Table 2.3 is the "crying babies" data given by Cox (1970, p. 5), a $2 \times 2 \times 18$ table. Only one child was treated on each day.

For asymptotic test using Table 2.1, in order to study the association between D and C, we test conditional independence against no three-factor interaction. The likelihood-ratio chi-squared statistic for testing the fit of no three-factor interaction model equals 7.494, with $d.f. = 2$. The estimated association parameter, which is unconditional maximum likelihood estimate, is 2.550 with $s.e. = 1.175$. The likelihood-ratio chi-squared statistic for testing conditional independence, assuming no three-factor interaction model, $G^2[(PD, PC)|(CD, PD, PC)]$, is $14.294-7.494=6.800$ with $d.f.=1$. The asymptotic P-value, when treated as chi-squared on a single degree of freedom, is 0.009. There seems to be very strong evidence of higher cure rate for immediate injection. However, the data is sparse enough to make large-sample approximations questionable. Likewise, for Tables 2.2 and 2.3, the asymptotic P-values are $0.029, 0.031$, respectively.

For the modified mid P-value, we can use $T'=\sum X_k^2$ or the table probability for the secondary statistic. For Table 2.1, $P_{\text{mid}}=0.011$ and $P^*_{\text{mid}}=0.002$ for both modified mid P-values using $\sum X_k^2$ or the table probability. Likewise, for Table 2.2, we have $P_{\text{mid}}=0.020$, and $P^*_{\text{mid}}=0.022$ for both modified mid P-values using $\sum X_k^2$ or the table probability. For Table 2.3, $P_{\text{mid}}=0.028$, and $P^*_{\text{mid}}=0.024$ with $T'=\sum X_k^2$ and 0.021 with the table probability. Figures 2.1 and 2.2 present the cumulative distribution functions of the modified exact P-value and the modified mid P-value using $T'=\sum X_k^2$, and the corresponding cumulative distribution functions using the table probability for T', respectively, for null conditional distributions based on the margins of Table 2.1. The modified mid P-value jumps and exceeds the nominal value, while the modified P-value jumps closely to the nominal value and never exceeds it. The gap between the actual size and the nominal value is smaller for the modified mid P-value than for the modified P-value. We can see this by drawing a line connecting (0,0) and (1,1) in each Figure.

Figures 2.3 and 2.4 display the cumulative distribution functions of the ordinary mid P-value and the modified mid P-value using $T'=\sum X_k^2$, and the corresponding cumulative distribution functions using the table probability for the modified mid P-value, respectively, for the null conditional distribution based on the margins of Table 2.1. Though tests based on the ordinary and modified mid P-value are not exact, the gap between the actual size and the nominal level tends to be less for the modified mid P-value than for the ordinary mid P-value. One way to measure how close the *cdf* of P-value is to the uniform *cdf* is by the measure

$$M = \int_0^1 |F(x) - G(x)|\, dx,$$

where $F =$ *cdf* of P and $G =$ uniform *cdf*. M is bounded by $1/2$, and large value of M implies the severe discrepancy between two distribution functions. Using Table 2.1 with $T' = \sum X_k^2$, we have $M = 0.055$ for P_{mid}, and $M = 0.022$ for P^*_{mid}. For the exact P-values, we have $M = 0.111$ for P, and $M = 0.045$ for P^*. For Table 2.2, we have $M = 0.038$ for P_{mid}, and $M = 0.003$ for P^*_{mid}. For Table 2.2, the discrepancy between the actual size and the nominal level is negligible for the modified mid P-value, and the amount from using modified mid P is only 7.9% of that from using ordinary mid P. We can see the modified mid P has a distribution that is less discrete.

Table 2.1. Example 1 for analyses.

Penicillin Level	Delay	Response	
		Cured	Died
1/8	None	0	6
	1 1/2 Hour	0	5
1/4	None	3	3
	1 1/2 Hour	0	6
1/2	None	6	0
	1 1/2 Hour	2	4
1	None	5	1
	1 1/2 Hour	6	0
4	None	2	0
	1 1/2 Hour	5	0

Source: Mantel (1963)

Table 2.2. Example 2 for analysis.

Age	Salk Vaccine	Paralysis	
		No	Yes
10-14	Yes	3	2
	No	3	2
15-19	Yes	7	4
	No	1	6
20-39	Yes	12	3
	No	7	5
40+	Yes	1	0
	No	3	2

Source: Agresti (1990)

Table 2.3. Example 3 for analyses.

Day	Group	Response	
		Not Crying	Crying
1	Treated	1	0
	Control	3	5
2	Treated	1	0
	Control	2	4
3	Treated	1	0
	Control	1	4
4	Treated	0	1
	Control	1	5
5	Treated	1	0
	Control	4	1
6	Treated	1	0
	Control	4	5
7	Treated	1	0
	Control	5	3
8	Treated	1	0
	Control	4	4
9	Treated	1	0
	Control	3	2
10	Treated	0	1
	Control	8	1
11	Treated	1	0
	Control	5	1
12	Treated	1	0
	Control	8	1
13	Treated	1	0
	Control	5	3
14	Treated	1	0
	Control	4	1
15	Treated	1	0
	Control	4	2
16	Treated	1	0
	Control	7	1
17	Treated	0	1
	Control	4	2
18	Treated	1	0
	Control	5	3

Source: Cox (1970)

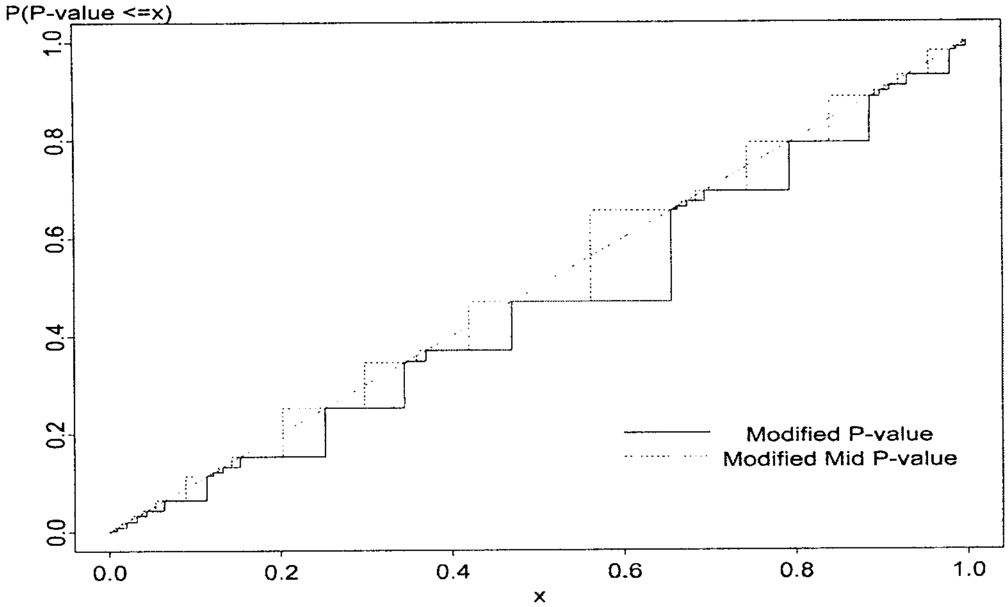

Figure 2.1. Cumulative distribution functions of the modified exact P-value and the modified mid P-value with $T' = \sum X_k^2$, for the margins of Table 2.1.

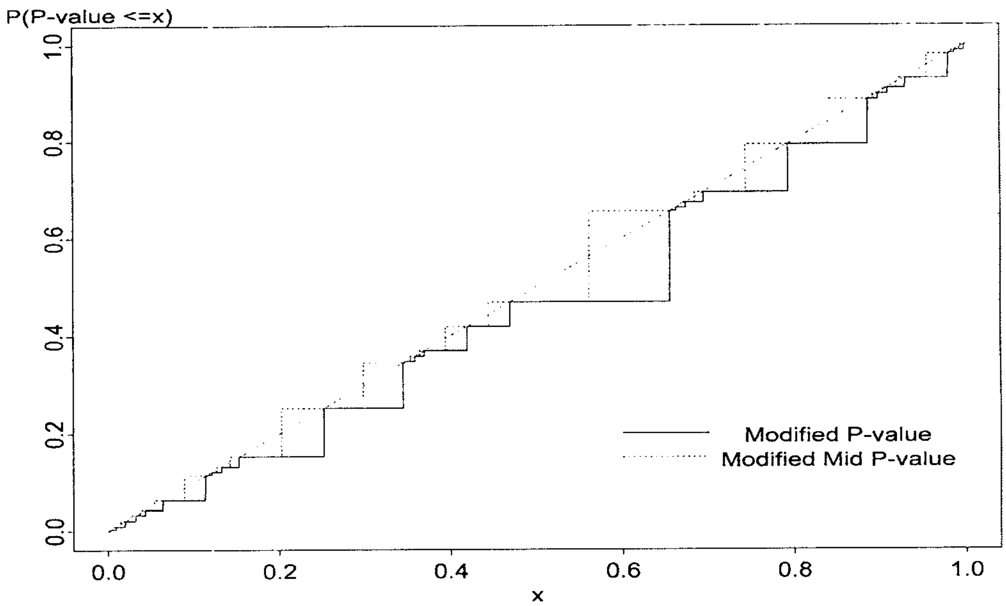

Figure 2.2. Cumulative distribution functions of the modified exact P-value and the modified mid P-value with $T' = P(Z)$, for the margins of Table 2.1.

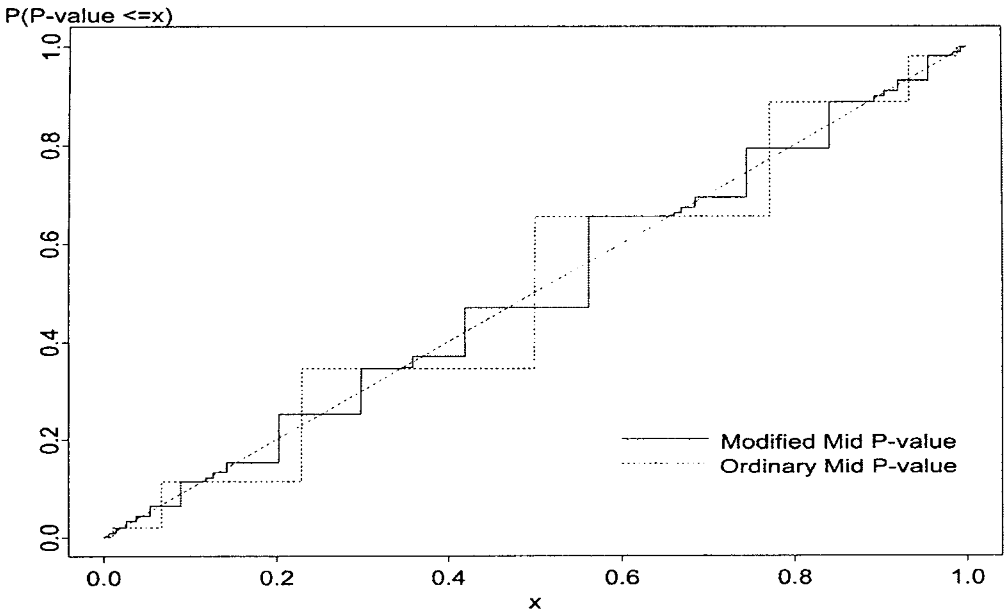

Figure 2.3. Cumulative distribution functions of the ordinary mid P-value and the modified mid P-value with $T' = \sum X_k^2$, for the margins of Table 2.1.

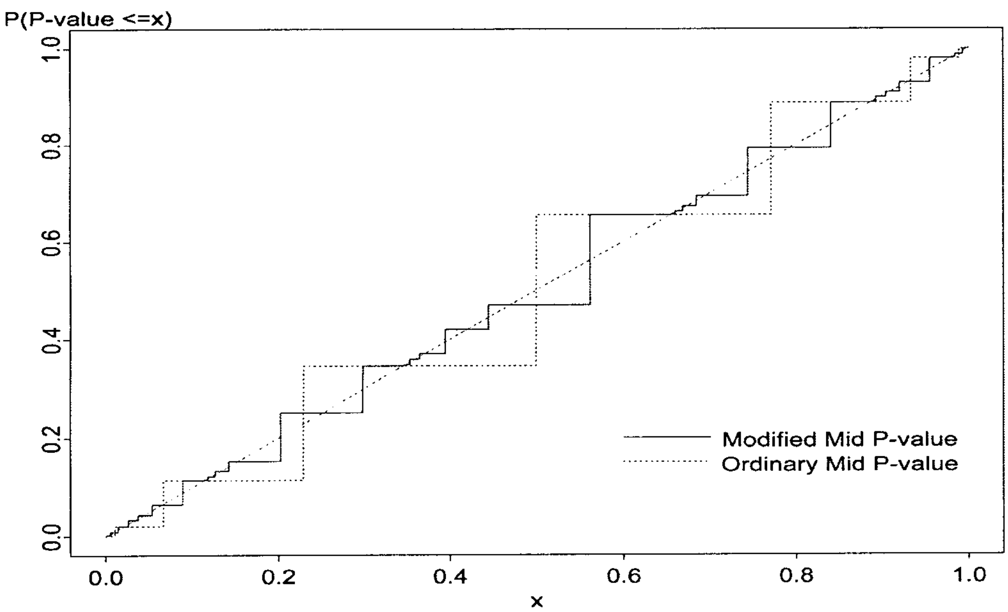

Figure 2.4. Cumulative distribution functions of the ordinary mid P-value and the modified mid P-value with $T' = P(Z)$, for the margins of Table 2.1.

3. The One-sided Modified Mid P Confidence Interval

Discreteness also affects interval estimation. An exact confidence interval for a parameter can be constructed by inverting the exact conditional test. The ordinary confidence interval is based on inverting two separate one-sided tests using the ordinary P-value. Due to discreteness of the test statistic, we get quite conservative as well as unduly wide confidence interval. The actual confidence coefficient is at least the nominal level. In order to overcome these problems, we construct confidence intervals using modified mid P-value based on inverting two separate one-sided tests.

For an exact confidence interval for a parameter, we invert an exact conditional test about that parameter. To illustrate, suppose we want to estimate an assumed common odds ratio, θ, in a $2 \times 2 \times K$ contingency table. The reference set Γ is defined in Section 2, and let $\Gamma_t = \{ Z : Z \in \Gamma, \sum_k n_{11k} = t\}$. Ordinary exact confidence limits for the common odds ratio are constructed from the conditional distribution of $T = \sum_k n_{11k}$, that is,

$$P(T = t, \theta) = \frac{c_t \theta^t}{\sum_{u=t_{\min}}^{t_{\max}} c_u \theta^u}, \tag{3.1}$$

where $c_t = \sum_{Z \in \Gamma_t} \Pi_k \binom{n_{+1k}}{z_k} \binom{n_{+2k}}{n_{1+k} - z_k}$, and where $t_{\min} = \sum_k \max(0, n_{1+k} - n_{+2k})$ and $t_{\max} = \sum_k \min(n_{1+k}, n_{+1k})$.

The ordinary interval (Cox 1970, Gart 1970, Mehta *et al.* 1985, Vollset, Hirji, and Elashoff 1991) is based on inverting two separate one-sided tests. It equals (θ_-, θ_+), where for $t_{\min} \le t_o \le t_{\max}$,

$$\begin{aligned} \text{at} \quad \theta = \theta_- : P_1(\theta) &= \sum_{t \ge t_o} P(t; \theta) = \frac{\alpha}{2}, \\ \text{at} \quad \theta = \theta_+ : P_2(\theta) &= \sum_{t \le t_o} P(t; \theta) = \frac{\alpha}{2}. \end{aligned} \tag{3.2}$$

When $t_o = t_{\min}$, the lower endpoint is 0; if $t_o = t_{\max}$, the upper endpoint is ∞. It is easily shown that $(\theta_-(t), \theta_+(t))$ has a confidence coefficient at least $100(1-\alpha)$ (Mehta *et al.* 1985). Due to discreteness of the distribution of T, we have only a conservative confidence interval, and the actual confidence coefficient is unknown.

For confidence intervals for a common odds ratio based on either inverting two separate one-sided tests or inverting a two-sided test, one can construct even narrower intervals, albeit not exact ones, by inverting the tests based on the modified mid P-value. The ordinary mid P confidence limits based on inverting two separate one-sided tests are found using the functions

$$P_{\text{mid}(1)}(\theta) = P_1(\theta) - \frac{1}{2}P(t_o;\theta)), \qquad (3.3)$$
$$P_{\text{mid}(2)}(\theta) = P_2(\theta) - \frac{1}{2}P(t_o;\theta)).$$

The lower limit, θ_-, is the smallest of all θ's to satisfy $P_{\text{mid}(1)}(\theta) \geq \frac{\alpha}{2}$, and the upper limit, θ_+, is the largest of all θ's to satisfy $P_{\text{mid}(2)}(\theta) \geq \frac{\alpha}{2}$. Though approximate, this type of confidence interval based on the ordinary mid P-value does not guarantee the nominal level $1-\alpha$, but they were shorter on average than the corresponding exact intervals (Mehta and Walsh 1992). We can construct better intervals by inverting the tests based on modified mid P-values.

Following the modified approach based on using a one-sided modified mid P-value, let $B_1(\theta) = \{Z : Z \in \Gamma, T = t_o, T'(\theta) = t_o'(\theta)\}$. The modified mid P confidence interval based on inverting two separate one-sided tests uses

$$P^*_{\text{mid}(1)}(\theta) = P^*_1(\theta) - \frac{1}{2}P(B_1(\theta);\theta), \qquad (3.4)$$
$$P^*_{\text{mid}(2)}(\theta) = P^*_2(\theta) - \frac{1}{2}P(B_1(\theta);\theta),$$

where $P^*_1(\theta)$ and $P^*_2(\theta)$ are defined as

$$P^*_1(\theta) = \sum_{t > t_o} P(t;\theta) + P[B(\theta);\theta], \quad P^*_2(\theta) = \sum_{t < t_o} P(t;\theta) + P[B(\theta);\theta],$$

where

$B(\theta) = \{Z : Z \in \Gamma, T = t_o, T'(\theta) \geq t_o'(\theta)\}$, or $B(\theta) = \{Z : Z \in \Gamma, T = t_o, P(Z;\theta) \leq P(N;\theta)\}$, which is depending on the secondary statistic. The limits are determined by the same method used for the ordinary mid P confidence interval, using $P^*_{\text{mid}(1)}(\theta)$ for the lower limit and $P^*_{\text{mid}(2)}(\theta)$ for the upper limit.

This approach can give narrower intervals than those obtained by inverting the one-sided test with the ordinary mid P-value, depending on the discreteness of the distribution of the test statistic. The modified mid P method gives actual coverage probability closer to the nominal level and it has less undercoverage than the ordinary mid P. We illustrate these confidence intervals for a common odds ratio using Tables 2.1, 2.2, and 2.3. For table 2.1, the 95% confidence interval by inverting one-sided test is $(1.34, 266.54)$ based on the ordinary mid P-values and $(2.22, 56.00)$ based on the modified mid P-values using $\sum X^2_k(\theta)$ or the table probability for T'. Using Table 2.2, the confidence intervals are $(1.05, 10.89)$ using the ordinary mid P-values, $(1.03, 11.19)$ using the modified mid P-values with $\sum X^2_k(\theta)$ or the table probability for T'. For Table 2.3, the confidence intervals are $(0.98, 16.89)$ using the ordinary mid P-values, $(1.01, 13.61)$ using the modified mid P-values with $\sum X^2_k(\theta)$, and

(1.04, 14.85) using the modified mid P-values with the table probability for T'. The corresponding 95% confidence intervals by using large-sample approximations are (1.28, 128.12) for Table 2.1, (1.10, 11.29) for Table 2.2, and (0,99, 17.64) for Table 2.3. For tables 2.1 and 2.3, the modified intervals are narrower than the ordinary one, and they are included in the interval using large-sample approximations. From the intervals using Table 2.2, we think the modified mid P method provides some adjustment on the ordinary mid P method so that it reduces the undercoverage of coverage probability. The accuracy of the normal approximation can be influenced by the skewness, kurtosis, and the discreteness of the distribution of the test statistic. Furthermore, in small samples, or when there are many nuisance parameters, the use of unconditional maximum likelihood estimator can be quite misleading, and the corresponding confidence intervals using large-sample approximations can also be inaccurate.

4. The Two-sided Modified Mid P Confidence Interval

Using a two-sided approach, Sterne (1954) constructed a confidence interval for a single binomial parameter, and Baptista and Pike (1977) constructed a confidence limits for the odds ratio in a 2 × 2 table. This two sided confidence interval also is conservative. As two-sided approach tends to give an interval that is usually narrower than the one based on inverting two separate one-sided tests (Kim and Agresti 1995), we can construct a better interval using two-sided mid P-values. Though these cannot guarantee achieving at least the nominal confidence level, one could define mid P versions of the ordinary two-sided and modified two-sided intervals and the modified mid P approach gives coverage probability even closer to the nominal level than the ordinary mid P approach. For testing a particular value of θ, a two-sided P-value is given by

$$P(\theta) = \sum_{\{t:\, P(t;\,\theta) \le P(t_o;\,\theta)\}} P(t;\, \theta).$$

Then, we define two-sided mid P-value as

$$P_{\text{mid}}(\theta) = P(\theta) - \frac{1}{2} P(\{Z: Z \in \Gamma, P(t;\theta) = P(t_o;\theta)\}). \tag{4.1}$$

The two-sided mid P confidence interval consists of the values of θ for which this two-sided mid P-value equals at least α.

Following the modified approach, one can construct a modified confidence interval based on two-sided tests by using a modified mid P-value. We define a modified two-sided mid P-value for testing a particular value of θ as

$$P^{*}_{\text{mid}}(\theta) = P^{*}(\theta) - \frac{1}{2} P(\{Z: Z \in \Gamma, P(t;\theta) = P(t_o;\theta), T'(\theta) = t_o'(\theta)\}), \tag{4.2}$$

where $P^{*}(\theta) = P(\theta) - P(\{Z: Z \in \Gamma, P(t;\theta) = P(t_{o.}\theta), T'(\theta) < t_o'(\theta)\})$. For the modified two-sided confidence interval, we consider the shortest interval that contains all of the values of θ for which $P^{*}_{\text{mid}}(\theta) \geq \alpha$.

We illustrate these confidence intervals for a common odds ratio using Tables 2.1, 2.2, and 2.3. For Table 2.1, the 95% confidence interval by inverting a two-sided test is $(1.38, 131.51)$ based on the ordinary mid P-values and $(1.38, 35.51)$ based on modified mid P-values using $T' = \sum X_k^2(\theta)$. For Table 2.2, the confidence intervals are $(1.12, 11.04)$ for both ordinary mid P-value and modified mid P-value with $T' = \sum X_k^2(\theta)$. Using Table 2.3, the confidence intervals are $(1.01, 12.58)$ and $(1.01, 10.29)$ using the ordinary and modified mid P-values with $T' = \sum X_k^2(\theta)$, respectively. For Tables 2.1 and 2.3, the confidence interval constructed by using ordinary two-sided mid P-value is shorter than the ordinary one based on two one-sided mid P-values, and for each type of interval, the modified interval is narrower than the ordinary one. Table 4.1 summarizes these 95% confidence intervals using Tables 2.1, 2.2, and 2.3.

The performance of the modified mid P method can be studied by computing the coverage probability of confidence intervals. For the conditional distribution having the fixed marginal counts of Table 2.1, Figure 4.1 shows actual coverage probability as a function of the true log odds ratio, for the 95% confidence intervals based on inverting separate one-sided tests using the ordinary mid P-value or the modified mid P-value with $T' = \sum X_k^2(\theta)$. The exact method yields a coverage exceeding the nominal level, depending on the degree of conservativeness, whereas the coverage of the mid P-value fluctuates about the nominal level. For either approach, for sufficiently large $|\log \theta|$, the actual probability of coverage of a $100(1-\alpha)\%$ confidence interval is centered about $1-\alpha/2$ and that of the modified mid P-value deviates less from $1-\alpha/2$. Figure 4.2 gives an analogous display for the confidence intervals based on inverting two-sided tests using the ordinary mid P-value or the modified mid P-value with $T' = \sum X_k^2(\theta)$. There is an advantage to the interval based on the modified mid P-value. For either approach, the actual probability of coverage of a $100(1-\alpha)\%$ confidence interval is centered about the nominal level, and that of the modified mid P-value is even closer to the nominal level than that of the ordinary mid P-value.

Figures 4.3 and 4.4 show analogous display of actual coverage probability using Table 2.2, that has one more stratum than Table 2.1. We can also see the advantage of modified mid P approach. For both one-sided and two-sided modified approach, actual coverage probability is very close to the confidence level with almost negligible undercoverage. Actually, for the one-sided modified interval, the coverage probability is almost the same as the confidence level except for large $|\log \theta|$, and the two-sided modified approach has coverage probability almost close to the confidence level for most values of θ. We see the modified approach does not have much conservativeness and maintains almost guaranteed level. For intervals using mid P-values, we suggest the use of the confidence interval based on inverting two-sided tests using the modified mid P-value.

Table 4.1. Various 95% confidence intervals for the common odds ratio using mid P-value.

Method	Data set 1	Data set 2	Data set 3
Ordinary 1-sided mid P	(1.34, 266.54)	(1.05, 10.89)	(0.98, 16.89)
Modified 1-sided mid P (P^*)	(2.22, 56.00)	(1.03, 11.19)	(1.01, 13.61)
Modified 1-sided mid P (P_p^*)	(2.22, 56.00)	(1.03, 11.19)	(1.04, 14.85)
Ordinary 2-sided mid P	(1.38, 131.51)	(1.12, 11.04)	(1.01, 12.58)
Modified 2-sided mid P (P^*)	(1.38, 35.51)	(1.12, 11.04)	(1.01, 10.29)

5. General ways of choosing T' in $I \times J \times K$ tables

We discuss a general way of choosing a secondary statistic T' to generate a secondary partitioning of tables having the observed value of T in $I \times J \times K$ contingency tables. For testing conditional independence of X and Y, given Z, assuming no three-factor interaction, we let T_N be the test statistic when both X and Y are nominal, let T_{NO} be the test statistic when X is nominal and Y is ordinal, and let T_O be the test statistic when both X and Y are ordinal. Also, let T_N^*, T_{NO}^* and T_O^* be the corresponding test statistics when we permit three-factor interaction. Kim and Agresti (1997) discussed these statistics and obtained precise estimates of P-values for exact conditional tests in $I \times J \times K$ contingency tables. They also applied tests with homogeneous association alternatives to nearly exact tests of marginal homogeneity for multivariate responses having the same categorical scale for each component. These are score statistics. To form a modified mid P-value, T' is a statistic directed toward a broader alternative. Then T' can catch some information about the validity of the null hypothesis when the assumed alternative for T is not exactly satisfied. We focus on score statistic for T, because inferential analyses using the exact distribution are then computationally much simpler. Ordinary P-values for these six tests correspond to six loglinear models for alternative hypothesis.

The general rule to construct the modified mid P-value is as follows. We use a score statistic for T', in order to have consistency and our principle is to choose a T' from the next most general alternative, while keeping the same assumption as T about three-factor interaction. Then, for example, assuming no three-factor interaction, (T, T') is (T_{NO}, T_N) for the nominal-by-ordinal case, since the nominal-by-nominal case is more general, and it also corresponds to the next most general alternative assuming no three-factor interaction. For the ordinal-by-ordinal case, the next most general alternative corresponds to the nominal-by-ordinal case. Hence (T, T') is (T_O, T_{NO}). Accordingly, for the ordinal-by-ordinal case permitting three-factor interaction, (T, T') is (T^*_O, T^*_{NO}). There might be other principles to form T', for example, to select T' from the most general alternative among all cases, but our principle can be recommended because the modified mid P-values can be defined for most cases using this principle, and it can utilize the ordinality of classification variables. Table 5.1 summarizes test statistics for the construction of ordinary and modified mid P-values for testing conditional independence in $I \times J \times K$ contingency tables using this principle. When $I = J = 2$, three statistics assuming no three-factor interaction, T_N, T_{NO}, and T_O, coincide for unit-spaced scores, and utilize $\sum_k n_{11k}$. Also, when $I = J = 2$, we get $T^*_N = T^*_{NO} = T^*_O$.

Table 5.1. Test statistics for the construction of the ordinary and modified mid P-values for testing conditional independence in $I \times J \times K$ contingency tables.

	Ordinary P-value	Modified Mid P-value P^*_{mid}
	T	(T, T')
Assuming no three-factor interaction		
Nominal-by-Nominal	T_N	(T_N, T^*_N)
Nominal-by-Ordinal	T_{NO}	(T_{NO}, T_N)
Ordinal-by-Ordinal	T_O	(T_O, T_{NO})
Permitting three-factor interaction		
Nominal-by-Nominal	T^*_N	$(T^*_N, P(Z))$
Nominal-by-Ordinal	T^*_{NO}	(T^*_{NO}, T^*_N)
Ordinal-by-Ordinal	T^*_O	(T^*_O, T^*_{NO})

6. Conclusion

For the test of conditional independence for categorical data, available tests are exact or nearly exact test based on the exact distribution of test statistic, such as exact P, mid P, and modified mid P, asymptotic test such as test based on loglinear models using $G^2[(XZ, YZ) \mid (XY, XZ, YZ)]$, and Cochran-Mantel-Haenszel (CMH) test (Mantel and Haenszel 1959). For $2 \times 2 \times K$ contingency tables, both the model based statistic and CMH statistic have approximately a chi-squared distribution with $df = 1$ under the null hypothesis of conditional independence when there is no three-factor interaction. CMH test is inappropriate when the association changes dramatically across the strata (Agresti 1990). Asymptotic results require many observations in each stratum. If data is so sparse that asymptotic theory does not hold, model based tests are not appropriate, and CMH can be applicable. But an exact test is preferred over CMH or model based test when the sample size is small. Uniformly most powerful unbiased test of conditional independence for $2 \times 2 \times K$ tables is based on $\sum n_{11k}$. We can think the mid P approach is some kind of continuity correction for estimating the tail area of a discrete distribution, whereas some asymptotic method such as CMH statistic includes a continuity correction (1/2). The modified mid P method reduces the conservativeness of exact method that comes from the high degree of discreteness of the test statistic and also improves the ordinary mid P method so that it gives nearly exact results.

When a test statistic has a discrete distribution, the exact methods lead to conservative test, and it is generally not possible to have confidence intervals with a specified coverage level. Only exact method guarantees coverage probability not below the specified level. But actual coverage probability may be much higher than the level. We have shown that use of a modified mid P-value leads to tests and confidence intervals that are less conservative than the usual one. The improvement can be considerable when K is large but the sample size is not. We prefer modified mid P tests and confidence intervals over the ordinary ones, because they are less discrete and less conservative than ordinary ones and are closer to the nominal level, while the exact test is conservative. The modified mid P-value is nearly exact. We prefer confidence intervals based on inverting two-sided tests using modified mid P-value over those based on inverting two separate one-sided tests, because they tend to be less conservative and nearly exact. Hence, modified mid P method provides more adjustment for the excessive conservativeness of the ordinary P and are centered about the nominal level with narrower fluctuations without possibility of severe undercoverage than the ordinary mid P. The modified mid P procedure will be an excellent alternative for an exact test when the exact test is conservative and the large-sample approximations is poor, for example, for those tables whose sample size is small or contingency tables are sparse. Some statistical software (e.g., StatXact 1991, Vollset and Hirji 1991) provides only the ordinary one-sided mid P capability in exact procedures by inverting separate one-sided tests. We suggest the modified mid P method, because it provides almost exact results with less conservativeness.

The idea of a modified mid P can be applicable to any contingency tables, and it can be calculated for any test statistic having a discrete distribution. One application can be the exact tests of no three-factor interaction. Zelen (1971) presented an exact test of homogeneity of odds ratios in $2 \times 2 \times K$ tables. For an exact test of no three-factor interaction for $2 \times 2 \times K$ tables, an efficient score statistic against the saturated model is the Pearson statistic for testing the fit of that model (Agresti 1992). We could use this score statistic as a primary statistic and the table probability as a secondary statistic to define the modified mid

P-values. Another example is to consider the modified mid P confidence interval for the linear-by-linear association parameter in linear-by-linear association model (Agresti 1990). Under the alternative, the conditional distribution of test statistic $T = \sum\sum u_i v_j n_{ij}$ has a noncentral hypergeometric distribution. By using a modified mid P confidence interval, we could reduce the conservativeness of the Agresti-Mehta-Patel (1990) interval.

When the exact or modified mid P test is infeasible and the application of large-sample approximations is questionable, one suggestion can be to use the saddlepoint approximations. Saddlepoint approximation provides good approximation to exact test and reduces the degree of conservativeness.

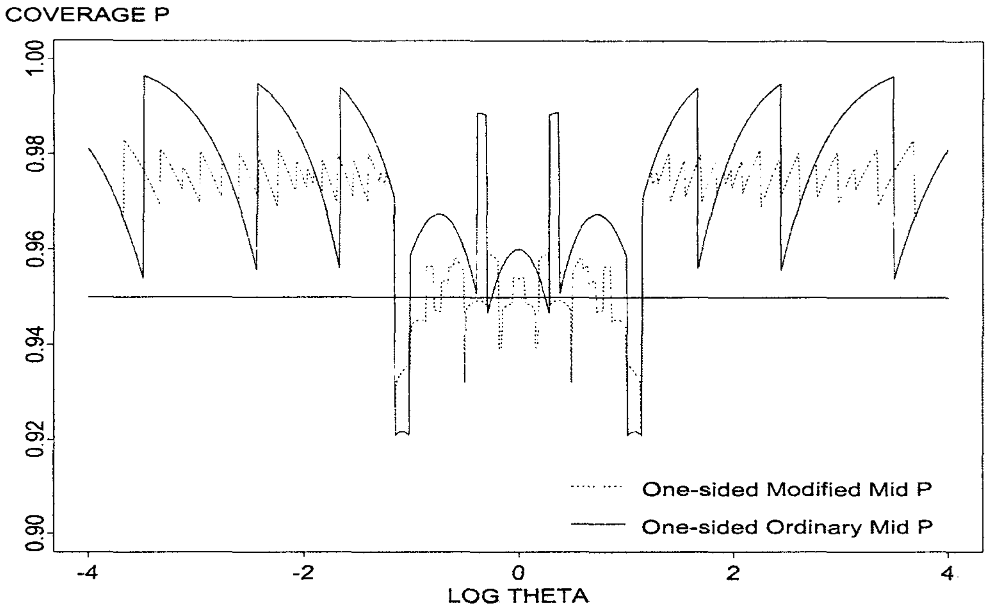

Figure 4.1. Coverage probability for confidence intervals based on inverting one-sided tests using mid P-values with $T' = \sum X_k^2(\theta)$, for conditional distribution based on margins of Table 2.1.

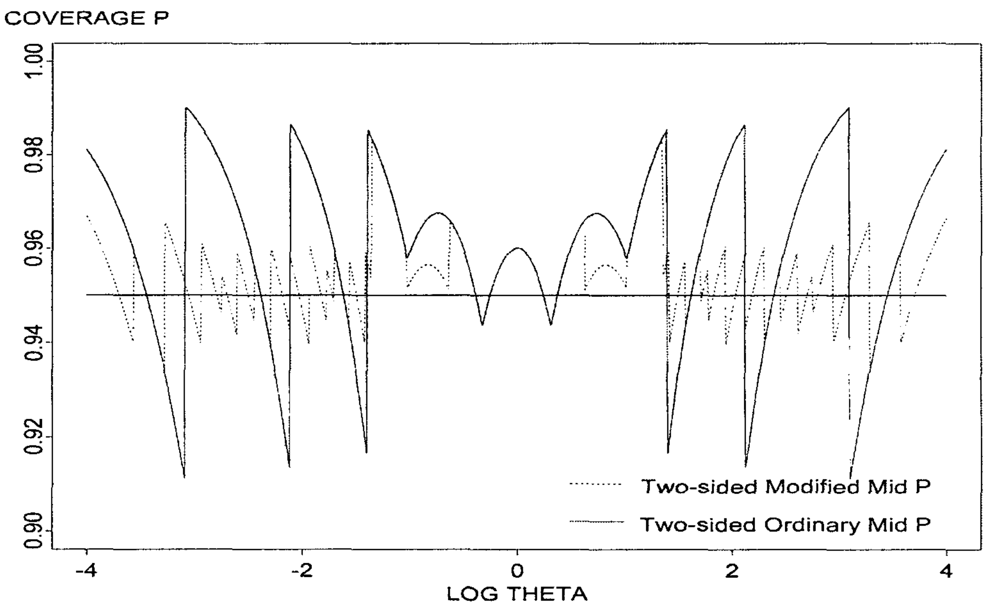

Figure 4.2. Coverage probability for confidence intervals based on inverting two-sided tests using mid P-values with $T' = \sum X_k^2(\theta)$, for conditional distribution based on margins of Table 2.1.

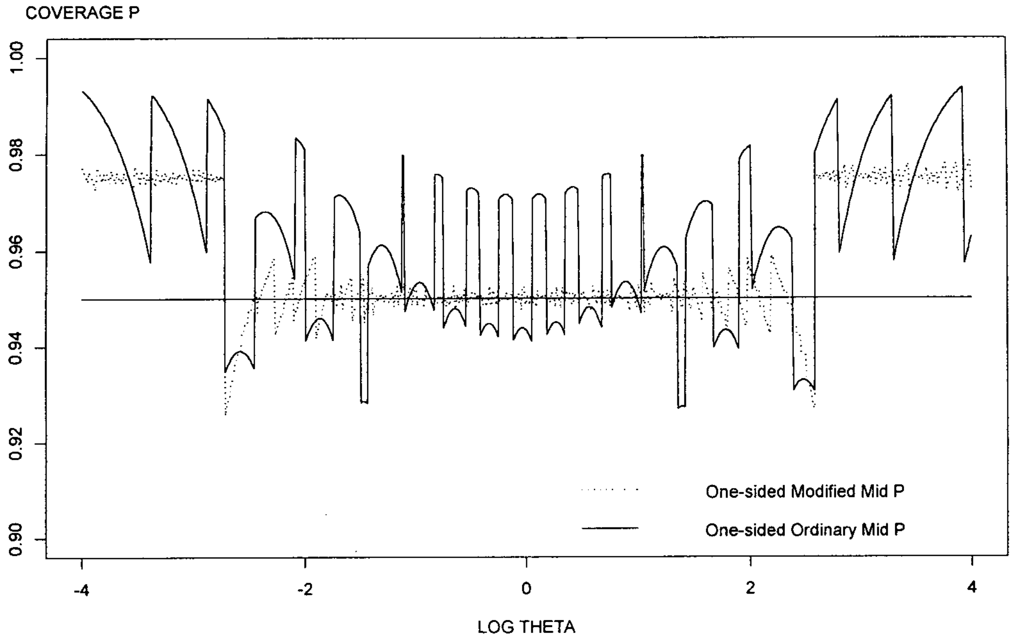

Figure 4.3. Coverage probability for confidence intervals based on inverting one-sided tests using mid P-values with $T' = \sum X_k^2(\theta)$, for conditional distribution based on margins of Table 2.2.

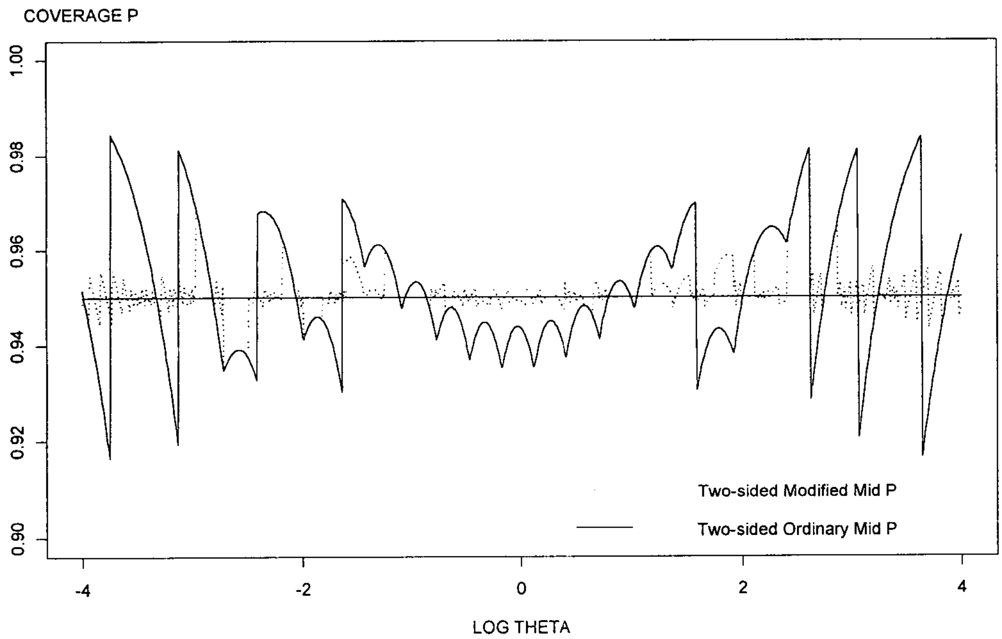

Figure 4.4. Coverage probability for confidence intervals based on inverting two-sided tests using mid P-values with $T' = \sum X_k^2(\theta)$, for conditional distribution based on margins of Table 2.2.

Acknowledgements

The author are grateful to a referee for helpful comments and suggestions.

References

[1] Agresti, A. (1990). *Categorical Data Analysis*, John Wiley & Sons, New York.

[2] Agresti, A. (1992). A survey of exact inference for contingency tables, *Statistical Science*, Vol. 7, 131-177.

[3] Agresti, A., Mehta, C.R. and Patel, N.R. (1990). Exact inference for contingency tables with ordered categories, *Journal of the American Statistical Association*, Vol. 85, 453-458.

[4] Baptista, J. and Pike, M.C. (1977). Exact two-sided confidence limits for the odds ratio in a 2×2 table, *Journal of the Royal Statistical Society, C*, Vol. 26, 214-220.

[5] Barnard, G.A. (1990). Must clinical trials be large ? The interpretation of P-values and the combination of test results, *Statistics in Medicine*, Vol. 9, 601-614.

[6] Cox, D.R. (1970). *Analysis of Binary Data*, Chapman and Hall, London.

[7] Gart, J. J. (1970). Point and interval estimation of the common odds ratio in the combination of 2×2 tables with fixed marginals, *Biometrika*, Vol. 57, 471-475.

[8] Kim, D. and Agresti, A. (1995). Improved exact inference about conditional association in three-way contingency tables, *Journal of the American Statistical Association*, Vol. 90, 632-639.

[9] Kim, D. and Agresti, A. (1997). Nearly exact tests of conditional independence and marginal homogeneity for sparse contingency tables, *Computational Statistics and Data Analysis*, Vol. 24, 89-104.

[10] Lancaster, H. O. (1961). Significance tests in discrete distributions, *Journal of the American Statistical Association*, Vol. 56, 223-234.

[11] Mantel, N. (1963). Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure, *Journal of the American Statistical Association*, Vol. 58, 690-700.

[12] Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease, *Journal of the National Cancer Institute*, Vol. 22, 719-748.

[13] Mehta, C.R., Patel, N.R. and Gray, R. (1985). Computing an exact confidence interval for the common odds ratio in several 2×2 contingency tables, *Journal of the American Statistical Association*, Vol. 80, 969-973.

[14] Mehta, C.R. and Walsh, S.J. (1992). Comparison of exact, mid-p, and Mantel-Haenszel confidence intervals for the common odds ratio across several 2×2 contingency tables, *The American Statistician*, Vol. 46, 146-150.

[15] StatXact. (1991). *StatXact: Statistical software for exact nonparametric inference*, Version

2, Cambridge, MA: Cytel Software.
[16] Sterne, T.E. (1954). Some remarks on confidence or fiducial limits, *Biometrika*, Vol. 41, 275-278.
[17] Vollset, S.E. and Hirji, K.F. (1991). A microcomputer program for exact and asymptotic analysis of several 2×2 tables, *Epidemiology*, Vol. 2, 217-220.
[18] Vollset, S.E., Hirji, K.F. and Afifi, A.A. (1991). Evaluation of exact and asymptotic interval estimators in logistic analysis of matched case-control studies, *Biometrics*, Vol. 47, 1311-1325.
[19] Vollset, S.E., Hirji, K.F. and Elashoff, R.M. (1991). Fast computation of exact confidence limits for the common odds ratio in a series of 2×2 tables, *Journal of the American Statistical Association*, Vol. 86, 404-409.
[20] Zelen, M. (1971). The analysis of several 2×2 contingency tables, *Biometrika*, Vol. 58, 129-137.