Statistica Sinica **6**(1996), 809-829

ASYMPTOTICS FOR A 2 × 2 TABLE WITH FIXED MARGINS

S. G. Kou and Z. Ying

Rutgers University

Abstract: Two coins A and B are tossed N_1 and N_2 times, respectively. Denote by M_1 (M_2) the total number of heads (tails) and X the number of heads from coin A. It is well known that conditional on the M_i and N_i, X has a noncentral hypergeometric distribution which depends only on the odds ratio θ between the success probabilities of the two coins. This model is commonly used in the analysis of a single 2×2 table, in which approximating X and estimating θ are of major concerns. Based on a connection between the probability generating function of X and the classical Jacobi polynomials, we show that X is equal in distribution to a sum of independent, though not identically distributed, Bernoulli random variables. It is then established that the central limit theorem $(X - \mathrm{E}X)/[\mathrm{Var}\,(X)]^{1/2} \to_{\mathcal{L}} N(0,1)$ holds if and only if $M_1M_2N_1N_2/N^3 \to \infty$, where $N = N_1+N_2$. In addition, this minimum condition is shown to be sufficient for (1) the maximum likelihood estimator of θ and the empirical odds ratio to be consistent and asymptotically normal, (2) some classical estimators for the asymptotic variance of the empirical odds ratio, such as those suggested in Cornfield (1956) and Woolf (1955), as well as a new variance estimator to be consistent. A Berry-Esseen-type bound is found, and a necessary and sufficient condition for X to be approximated by the Poisson distribution is established as well.

Key words and phrases: Noncentral hypergeometric distribution, empirical odds ratio, generating function, asymptotic normality, maximum likelihood estimator, Jacobi polynomials, Poisson approximation.

1. Introduction

Suppose an urn contains N_1 white balls and $N_2 = N - N_1$ red balls. Randomly draw M_1 ($\le N$) balls from the urn and denote by X the number of white balls. It is well known that X follows a hypergeometric distribution:

$$P(X=x) = \binom{N_1}{x}\binom{N_2}{M_1-x} \Big/ \binom{N}{M_1}, \quad L \le x \le S, \tag{1.1}$$

where $L = \max(0, M_1 - N_2)$ and $S = \min(N_1, M_1)$ (cf. Feller (1968), §2.6).

The name hypergeometric distribution comes from its relationship with the family of hypergeometric functions (Feller (1968), p.44, footnote). Let $(n)_k$ be the Pochhammer symbol denoting $(n)_k = n(n+1)\cdots(n+k-1)$ and $(n)_0 = 1$. The hypergeometric function with parameters a, b and c is defined by

$$F(a,b,c;z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k k!} z^k \qquad c \neq 0, -1, -2, \ldots$$

In particular, by setting $a = -N_1$, $b = -M_1$ and $c = N_2 - M_1 + 1$ and assuming $N_2 \geq M_1$, the function is reduced to a polynomial with degree at most S:

$$F(-N_1, -M_1, N_2+1-M_1; z) = \sum_{k=0}^{S} \frac{(-N_1)_k (-M_1)_k}{(N_2+1-M_1)_k k!} z^k, \quad N_1+M_1 \leq N. \tag{1.2}$$

It is not difficult to see (Johnson, Kotz and Kemp (1992), pp. 237, 238 and 280) that (1.2) is, when scaled by a constant, the probability generating function for (1.1). On the other hand, it is easy to see that if $N_2 < M_1$ then

$$\begin{aligned} &\frac{\binom{N_1}{M_2}}{\binom{N}{M_2}} z^{M_1-N_2} F(-N_2, -M_2, N_1+1-M_2; z) \\ &= \frac{\binom{N_1}{M_2}}{\binom{N}{M_2}} z^{M_1-N_2} \sum_{k=0}^{\min(N_2,M_2)} \frac{(-N_2)_k (-M_2)_k}{(N_1+1-M_2)_k k!} z^k \end{aligned}$$

becomes the probability generating function for (1.1).

Widespread use of the hypergeometric distributions in statistics is, at least partly, attributed to another method of generating a hypergeometric random variable. Suppose two coins A and B with the same success probability are tossed N_1 and N_2 times, respectively, and let M_1 (M_2) denote the total number of heads (tails) out of the $N = N_1 + N_2$ tosses and X the number of heads from the N_1 tosses of coin A. Then conditional on N_1, N_2, M_1 and M_2, X has the hypergeometric distribution given by (1.1). Thus the hypothesis that the success probabilities for the two coins are the same can be tested at a specified size by using the hypergeometric distribution.

On the other hand, if the success probabilities, denoted by π_A and π_B, for the two coins are different, then the conditional distribution of X given the N_i and M_i is known to follow a *noncentral* hypergeometric distribution, which depends on the success probabilities only through their odds ratio $\theta = \pi_A(1-\pi_B)/[\pi_B(1-\pi_A)]$. More precisely, the noncentral hypergeometric distribution is given by

$$P(X=x) = \binom{N_1}{x}\binom{N_2}{M_1-x}\theta^x \Big/ \sum_{u=L}^{S} \binom{N_1}{u}\binom{N_2}{M_1-u}\theta^u, \quad L \leq x \leq S. \tag{1.3}$$

One of the most common situations in which one needs to condition on the N_i and M_i occurs in epidemiological case-control (retrospective) studies, where

N_1 and N_2 are taken as the numbers of persons in exposed and unexposed groups, respectively, and M_1 and M_2 represent the numbers of diseased and disease-free individuals, respectively (cf. Breslow and Day (1980)). A 2×2 table illustrates the design:

	Diseased	Disease-free	
Exposed	X	$N_1 - X$	N_1
Unexposed	$M_1 - X$	$X + N_2 - M_1$	N_2
	M_1	M_2	N

The hypothesis of interest in this case is usually that the disease rates for the exposed and the unexposed groups are the same. If this hypothesis holds, then the distribution of X, conditioning on N_1, N_2, M_1 and M_2 (marginals), is again the (central) hypergeometric given by (1.1). Otherwise, X has the noncentral hypergeometric distribution given by (1.3). A key issue then becomes how to estimate odds ratio θ.

For both theoretical and practical considerations, it is extremely important to approximate both the central and noncentral hypergeometric distributions, as they are essential to hypothesis testing and parameter estimation. Under certain regularity conditions, we naturally expect that $(X - \mathrm{E}X)/\sqrt{\mathrm{Var}\,(X)}$ converges to the standard normal. In fact, when M_1 is small relative to N and N_1/N is not close to 0 and 1, it is well known that the (central) hypergeometric distribution can be approximated by the binomial, thus also by the normal in view of the classical DeMoivre-Laplace central limit theorem. Lehmann (1975), Appendix 4, gives an interesting proof of the asymptotic normality of X under the assumptions that $\mathrm{Var}\,(X) \to \infty$ and N_1/N is bounded away from 0 and 1. A sharper result, hidden in Vatutin and Mikhailov (1982), states that $\mathrm{Var}\,(X) \to \infty$ is sufficient for X to be asymptotically normal. The approach there is rather involved and is essentially based on a method developed earlier by Harper (1967).

We are concerned in this paper with the family of noncentral hypergeometric distributions as specified by (1.3). Our main goals are (1) to find minimum conditions that guarantee the normal and the Poisson approximations for X and (2) to develop a relatively complete theory for the maximum likelihood estimator of the odds ratio θ and the empirical odds ratio.

In section 2, by observing a connection between the probability generating function of the noncentral hypergeometric random variable X and the classical Jacobi polynomials, we shall show that X of (1.3) is equal in distribution to a sum of independent Bernoulli random variables. This representation can be used for both theoretical and numerical purposes. In particular, it helps us to prove that $N_1M_1N_2M_2/N^3 \to \infty$ is both necessary and sufficient for X to be asymptotically normal. In this connection, a Berry-Esseen type bound is also derived.

A substantial portion of this paper is devoted to the second problem, estimation of θ. Available eatimators are some classical ones, most notably, the maximum likelihood estimator and the empirical odds ratio $\hat{\theta}_e = X(X + N_2 - M_1)/((M_1 - X)(N_1 - X))$. Note that $\hat{\theta}_e$ is the maximum likelihood estimator for the unconditional model, i.e., only the N_i, but not the M_i, are conditioned. In section 3 we shall prove rigorously that both the maximum likelihood estimator and the empirical odds ratio are consistent and asymptotically normal under the minimum condition $N_1M_1N_2M_2/N^3 \to \infty$. To construct a confidence interval for θ, Woolf (1955) argued that one might use $\log \hat{\theta}_e$, which could have limiting normal distribution with variance approximated by

$$X^{-1} + (X + N_2 - M_1)^{-1} + (N_1 - X)^{-1} + (M_1 - X)^{-1}. \tag{1.4}$$

Moreover, Cornfield (1956) gave a heuristic derivation that (1.4) should be used to estimate $1/\mathrm{Var}\,(X)$.

Readers are referred to Breslow and Day (1980), §4.2, §4.5 for detailed discussions on the subject with interesting examples. Results by Woolf (1955) and Cornfield (1956) will be justified rigorously under precise conditions in Section 3.2. Moreover, some bounds are also provided for the maximum likelihood estimator in Section 4.

Approximation of the noncentral hypergeometric distribution by the Poisson distribution will be discussed in Section 4, where a necessary and sufficient condition for such approximation to be valid is also given. Section 5 contains some technical developments.

2. Noncentral Hypergeometric Distributions

Let $G_\theta(z) = \mathrm{E}_\theta(z^X)$ denote the *probability generating function* of noncentral hypergeometric random variable X with parameter θ. Here and in the sequel, subscript θ in E_θ and Var_θ indicates that the expectation and variance are taken with θ being the true parameter. It is readily seen that the probability generating function of the noncentral hypergeometric random variable in (1.3) is given by

$$G_\theta(z) = \frac{\phi(\theta z)}{\phi(\theta)}, \tag{2.1}$$

where

$$\phi(z) = \sum_{u=L}^{S} \binom{N_1}{u}\binom{N_2}{M_1 - u} z^u. \tag{2.2}$$

Notice that $\phi(\theta)$ is exactly the denominator in (1.3). We now introduce the following key lemma.

Lemma 2.1. *All roots of polynomial $\phi(z)$ are real and nonpositive.*

A proof of Lemma 2.1. is hidden in Vatutin and Mikhailov (1982), which is in turn based on Harper (1967). The arguments there are quite involved. By observing a connection between $\phi(z)$ and the family of Jacobi polynomials, we are able to give here a much simpler and clearer proof. The Jacobi polynomials are defined as

$$P_n^{(\alpha,\beta)}(x) = \frac{1}{2^n}\sum_{u=0}^{n}\binom{n+\alpha}{u}\binom{n+\beta}{n-u}(x-1)^{n-u}(x+1)^u, \quad -1<x<1, \tag{2.3}$$

where $\alpha > -1$ and $\beta > -1$. It is well known that for fixed α and β, the sequence $\{P_n^{(\alpha,\beta)},\ n \geq 0\}$ are orthogonal polynomials and, consequently, their roots must all be real. For detailed discussions about the Jacobi polynomials, see Erdélyi et al. (1953), §10.8 and p. 202 and Szegö (1959).

Proof of Lemma 2.1. We only need to show that all the roots are real, since they cannot be positive. Four cases will be considered separately.
Case 1. M_1 is the smallest among M_i, N_i, $i = 1, 2$. Then $S = M_1$ and $L = 0$, recalling that $S = \min(N_1, M_1)$ and $L = \max(0, M_1 - N_2)$. Letting $\alpha = N_1 - M_1 \geq 0$, $\beta = N_2 - M_1 \geq 0$ and $n = M_1$, we conclude from (2.3) that

$$P_n^{(\alpha,\beta)}(x) = \frac{1}{2^{M_1}}\sum_{u=0}^{M_1}\binom{N_1}{u}\binom{N_2}{M_1-u}(x-1)^{M_1-u}(x+1)^u,$$

has M_1 real roots, which must all be inside $(-1, 1)$. Furthermore, letting $y = (x+1)/(x-1)$, we then have

$$\frac{1}{(y-1)^{M_1}}\phi(y) = \frac{1}{(y-1)^{M_1}}\sum_{u=0}^{M_1}\binom{N_1}{u}\binom{N_2}{M_1-u}y^u = P_n^{(\alpha,\beta)}(x).$$

Since $P_n^{(\alpha,\beta)}$ has M_1 roots in $(-1, 1)$ and since, as x goes from -1 to 1, y goes from 0 to $-\infty$, we conclude that $\phi(y)$ must have M_1 roots in $(-\infty, 0)$.
Case 2. N_1 is the smallest. In this case, we just rewrite $\phi(z)$ as

$$\phi(z) = \sum_{u=0}^{N_1}\binom{M_1}{u}\binom{M_2}{N_1-u}z^u,$$

set $\alpha = M_1 - N_1 \geq 0$, $\beta = M_2 - N_1 \geq 0$ and $n = N_1$ in (2.3), and then proceed exactly as in case 1.
Case 3. If N_2 is the smallest, we rewrite

$$\phi(z) = z^{M_1-N_2}\cdot\sum_{u=0}^{N_2}\binom{M_2}{u}\binom{M_1}{N_2-u}z^u,$$

and put $\alpha = M_2 - N_2 \geq 0$, $\beta = M_1 - N_2 \geq 0$ and $n = N_2$ in (2.3), and then proceed as before.

Case 4. M_2 is the smallest. In this case, we use

$$\phi(z) = z^{M_1 - N_2} \cdot \sum_{u=0}^{M_2} \binom{N_2}{u} \binom{N_1}{M_2 - u} z^u,$$

and $\alpha = N_2 - M_2$, $\beta = N_1 - M_2$, $n = M_2$.

We shall denote the roots of $\phi(z)$ by $-\lambda_1, \ldots, -\lambda_S$, $\lambda_i \geq 0$, $1 \leq i \leq S$. In addition, it is clear from (2.2) that there are exactly L of these roots equal to 0. Based on this lemma, we are able to show now that any noncentral hypergeometric random variable X can be expressed as a sum of independent Bernoulli random variables. Therefore, many available techniques developed for sums of independent random variables can be readily applied.

Theorem 2.1. *Consider a noncentral hypergeometric random variable X specified by* (1.3)*. Let $\eta_1, \eta_2, \ldots,$ be a sequence of independent uniform*$(0,1)$ *random variables. Then*

$$X \stackrel{d}{=} \sum_{i=1}^{S} I(\eta_i \leq (1 + \theta^{-1}\lambda_i)^{-1}), \tag{2.4}$$

where "$\stackrel{d}{=}$" denotes equality in distribution and $I(\cdot)$ the indicator function, recalling $S = \min(N_1,\ M_1)$.

Proof. Lemma 2.1 implies that, for some constant c, $\phi(z) = c\prod_{i=1}^{S}(z + \lambda_i)$, which in turn leads to

$$G_\theta(z) = \prod_{i=1}^{S} \frac{z + \theta^{-1}\lambda_i}{1 + \theta^{-1}\lambda_i}, \tag{2.5}$$

in view of (2.1) and $G_\theta(1) = 1$. But the right-hand side of (2.5) is exactly the probability generating function of the right-hand side of (2.4), whence (2.4) holds, since a probability generating function uniquely determines a distribution function.

Theorem 2.1 is useful not only because it decomposes a noncentral hypergeometrical random variable into a sum of independent Bernoulli random variables, but also because the λ_i depends only on the N_i and M_i, not on θ, as $\phi(z)$ does not involve θ. This fact greatly reduces the computational burden in dealing with noncentral hypergeometric distributions. Furthermore, (2.4) provides a convenient way to simulate noncentral hypergeometric random variables, since roots of the Jacobi polynomials are well studied and readily available in many software packages.

From (2.4) it follows that the mean and the variance of X can be expressed in terms of the λ_i and θ:

$$\mathrm{E}_\theta X = \sum_{i=1}^{S} \frac{1}{1+\theta^{-1}\lambda_i}, \qquad \mathrm{Var}_\theta(X) = \sum_{i=1}^{S} \frac{\theta^{-1}\lambda_i}{(1+\theta^{-1}\lambda_i)^2}. \tag{2.6}$$

Since $\theta = 1$ corresponds to the case of a central hypergeometric distribution, whose mean and variance are N_1M_1/N, $N_1N_2M_1M_2/(N^2(N-1))$, respectively, we conclude that

$$\sum_{i=1}^{S} \frac{1}{1+\lambda_i} = \frac{M_1N_1}{N} \quad \text{and} \quad \sum_{i=1}^{S} \frac{\lambda_i}{(1+\lambda_i)^2} = \frac{N_1N_2M_1M_2}{N^2(N-1)}. \tag{2.7}$$

Corollary 2.1. *We have the following bounds for the mean and variance of the noncentral hypergeometric random variable:*

$$\min(1,\theta)M_1N_1/N \le \mathrm{E}_\theta X \le \max(1,\theta)M_1N_1/N, \tag{2.8}$$

$$\min(\theta, \frac{1}{\theta})\frac{N_1N_2M_1M_2}{N^2(N-1)} \le \mathrm{Var}_\theta(X) \le \max(\theta, \frac{1}{\theta})\frac{N_1N_2M_1M_2}{N^2(N-1)}. \tag{2.9}$$

Proof. These inequalities follow immediately from (2.7) and the following elementary inequalities

$$\min(1, \frac{a}{b})\frac{1}{1+b^{-1}\lambda_i} \le \frac{1}{1+a^{-1}\lambda_i} \le \max(1, \frac{a}{b})\frac{1}{1+b^{-1}\lambda_i}, \tag{2.10}$$

$$\min(a/b, b/a)\frac{b^{-1}\lambda_i}{(1+b^{-1}\lambda_i)^2} \le \frac{a^{-1}\lambda_i}{(1+a^{-1}\lambda_i)^2} \le \max(a/b, b/a)\frac{b^{-1}\lambda_i}{(1+b^{-1}\lambda_i)^2}, \tag{2.11}$$

where $0 < a < \infty$ and $0 < b < \infty$.

Theorem 2.2. *Let θ denote the true odds ratio parameter. Then the following three statements are equivalent:* (i) $(X - \mathrm{E}_\theta X)/\sqrt{\mathrm{Var}_\theta(X)} \to_{\mathcal{L}} N(0,1)$; (ii) $\mathrm{Var}_\theta(X) \to \infty$; (iii) $N_1N_2M_1M_2/N^3 \to \infty$.

Proof. The equivalence between (ii) and (iii) follows readily from (2.9). From Theorem 2.1, we know that X is equal in distribution to a sum of independent random variables. Thus, by the Lindeberg central limit theorem, (ii) implies (i). Conversely, note that if (ii) does not hold, then we can find a subsequence such that the support of $(X - \mathrm{E}_\theta X)/\sqrt{\mathrm{Var}_\theta}(X)$ will not be dense on the real line (in fact it is not dense in any proper interval). Hence (i) cannot be true.

Theorem 2.3. *The following Berry-Esseen type result holds:*

$$\sup_x \Big| P_\theta\Big(\frac{X - \mathrm{E}_\theta X}{\sqrt{\mathrm{Var}_\theta(X)}} \le x\Big) - \Phi(x)\Big| \le \frac{\gamma}{\sqrt{\mathrm{Var}_\theta(X)}}$$
$$\le \gamma \max\{\theta^{1/2}, \theta^{-1/2}\} \frac{N(N-1)^{1/2}}{(N_1 N_2 M_1 M_2)^{1/2}}, \quad (2.12)$$

where Φ denotes the standard normal distribution and γ is the usual Berry-Esseen constant.

Proof. According to the Berry-Esseen inequality (Chow and Teicher (1988), p. 304) and the representation (2.4), the left-hand side of (2.12) is less than or equal to

$$\frac{\gamma \sum_{i=1}^S \mathrm{E}_\theta |I(\eta_i \le (1+\theta^{-1}\lambda_i)^{-1}) - (1+\theta^{-1}\lambda_i)^{-1}|^3}{[\sum_{i=1}^S \mathrm{Var}_\theta(I(\eta \le (1+\theta^{-1}\lambda_i)^{-1}))]^{3/2}}.$$

Thus the first inequality holds by noting that

$$\mathrm{E}|I(\eta_i \le (1+\theta^{-1}\lambda_i)^{-1}) - (1+\theta^{-1}\lambda_i)^{-1}|^3 \le \mathrm{Var}_\theta(I(\eta_i \le (1+\theta^{-1}\lambda_i)^{-1})),$$

and the second one follows from (2.9).

3. Asymptotics for Estimators of Odds Ratio

3.1. Maximum likelihood estimator

In epidemiological case-control studies, it is of fundamental interest to estimate the value of the odds ratio $\theta \in (0, \infty)$. An obvious candidate is the maximum likelihood estimator (MLE) of θ, to be denoted henceforth by $\hat\theta$. Because the family of noncentral hypergeometric distributions parametrized by θ forms an exponential family with X being the canonical sufficient statistics, $\hat\theta$ must satisfy

$$X = \sum_{i=1}^S (1+\hat\theta^{-1}\lambda_i)^{-1}. \quad (3.1)$$

Theorem 3.1. *A necessary and sufficient condition guaranteeing the existence and uniqueness of $\hat\theta$ is*

$$L < X < S. \quad (3.2)$$

Proof. We first show that (3.2) is sufficient. Since $L (< S)$ is the number of the λ_i being 0, it is obvious that, $\sum_{i=1}^S (1+\theta^{-1}\lambda_i)^{-1}$ is strictly increasing in θ and goes from L to S as θ goes from 0 to ∞. Thus, in view of (3.2), there exists a unique $\hat\theta$ satisfying (3.1). The same argument also leads to the conclusion that no such $\hat\theta$ exists when either $X = L$ or $X = S$, whence the necessity holds as well.

Theorem 3.2. *Suppose that* $\mathrm{Var}_\theta(X) \to \infty$, *or equivalently,* $N_1N_2M_1M_2/N^3 \to \infty$. *Then* $\hat{\theta}$ *is consistent and asymptotically normal. More precisely, we have* $\hat{\theta} \to \theta$ *in probability and*

$$\theta^{-1}\sqrt{\mathrm{Var}_\theta(X)}(\hat{\theta} - \theta) \to_{\mathcal{L}} N(0,1). \tag{3.3}$$

Remark 3.1. Since we only have one 2×2 table and observations do not consist of a sequence of independent random variables, the usual asymptotic results for maximum likelihood estimators do not apply directly here.

Proof of Theorem 3.2. Let $K = N_1N_2M_1M_2/N^2(N-1)$. To prove consistency, it suffices to show that for any $0 < \epsilon < \theta$, there exists $\delta = \delta(\epsilon) > 0$ such that

$$P\Big(\inf_{\tilde{\theta}:|\tilde{\theta}-\theta|\geq\epsilon} \frac{|X - \sum_{i=1}^S (1+\tilde{\theta}^{-1}\lambda_i)^{-1}|}{K} > \delta\Big) \longrightarrow 1, \tag{3.4}$$

as $\mathrm{Var}_\theta(X) \to \infty$. In view of (2.9), the Markov inequality implies that $(X - \mathrm{E}_\theta(X))/K \xrightarrow{p} 0$. Therefore (3.4) holds if we can show that

$$\inf_{\tilde{\theta}:|\tilde{\theta}-\theta|\geq\epsilon} \frac{|\sum_{i=1}^S (1+\theta^{-1}\lambda_i)^{-1} - \sum_{i=1}^S (1+\tilde{\theta}^{-1}\lambda_i)^{-1}|}{K} \geq 2\delta. \tag{3.5}$$

Now, the monotonicity implies that the "inf" in (3.5) can only be achieved at $\tilde{\theta} = \theta \pm \epsilon$. In addition, by the mean-value theorem

$$\sum_{i=1}^S (1+\theta^{-1}\lambda_i)^{-1} - \sum_{i=1}^S (1+\tilde{\theta}^{-1}\lambda_i)^{-1} = \sum_{i=1}^S \frac{\tilde{\theta}_*^{-2}\lambda_i}{(1+\tilde{\theta}_*^{-1}\lambda_i)^2}(\theta - \tilde{\theta})$$

for some $\tilde{\theta}_*$ between θ and $\tilde{\theta}$. Thus we get

$$\begin{aligned}\text{l.h.s. of (3.5)} \;&\geq \frac{\sum_{i=1}^S (\theta+\epsilon)^{-2}\lambda_i/(1+(\theta-\epsilon)^{-1}\lambda_i)^2}{\sum_{i=1}^S \lambda_i/(1+\lambda_i)^2}\epsilon \\ &\geq (\theta-\epsilon)(\theta+\epsilon)^{-2}\min(\theta-\epsilon, \frac{1}{\theta-\epsilon})\epsilon,\end{aligned}$$

thanks to the elementary inequality (2.11). Taking $\delta = 2^{-1}(\theta-\epsilon)(\theta+\epsilon)^{-2}\min\{\theta-\epsilon, (\theta-\epsilon)^{-1}\}$, we have (3.5) and therefore the consistency of $\hat{\theta}$.

To show the asymptotic normality, we observe, again by the mean-value theorem,

$$X - \mathrm{E}_\theta X = \sum_{i=1}^S \frac{1}{1+\hat{\theta}^{-1}\lambda_i} - \sum_{i=1}^S \frac{1}{1+\theta^{-1}\lambda_i} = \sum_{i=1}^S \frac{\theta_*^{-2}\lambda_i}{(1+\theta_*^{-1}\lambda_i)^2}(\hat{\theta} - \theta) \tag{3.6}$$

for some θ_* between θ and $\hat\theta$. Because $\hat\theta \to \theta$ in probability, (2.11) can be used to show

$$\sum_{i=1}^{S} \frac{\theta_*^{-2}\lambda_i}{(1+\theta_*^{-1}\lambda_i)^2} \Big/ \sum_{i=1}^{S} \frac{\theta^{-2}\lambda_i}{(1+\theta^{-1}\lambda_i)^2} \to 1$$

in probability. Therefore, by (3.6), we have

$$\theta^{-1}\sqrt{\mathrm{Var}_\theta(X)}(\hat\theta - \theta) = \frac{\sum_{i=1}^{S} \theta^{-2}\lambda_i/(1+\theta^{-1}\lambda_i)^2}{\sum_{i=1}^{S} \theta_*^{-2}\lambda_i/(1+\theta_*^{-1}\lambda_i)^2}\Big[\frac{X - \mathrm{E}_\theta X}{\sqrt{\mathrm{Var}_\theta(X)}}\Big],$$

which converges to $N(0,1)$ in view of Theorem 2.2.

Applying the usual delta method, we immediate have the following corollary.

Corollary 3.1. *Suppose that $N_1N_2M_1M_2/N^3 \to \infty$. Then*

$$\sqrt{\mathrm{Var}_\theta(X)}(\log\hat\theta - \log\theta) \to_{\mathcal{L}} N(0,1).$$

Remark 3.2. It should be pointed out that the square of the normalizing term in (3.3) achieves the Fisher's information; more precisely,

$$\frac{\mathrm{Var}_\theta(X)}{\theta^2} = \mathrm{E}\Big[\frac{\partial}{\partial\theta}\log f(\theta, X)\Big]^2 = \mathrm{E}\Big[-\frac{\partial^2}{\partial\theta^2}\log f(\theta, X)\Big],$$

where $f(\theta, X)$ is the likelihood of X, as defined by the right hand side of (1.3) with x replaced by X.

If we are going to use (3.3) to construct an asymptotic confidence interval for the odds ratio parameter θ, then we may need to estimate $\mathrm{Var}(X)$. A natural candidate to estimate it is

$$\widehat{\mathrm{Var}(X)} = \sum_{i=1}^{S} \frac{\hat\theta^{-1}\lambda_i}{(1+\hat\theta^{-1}\lambda_i)^2}.$$

Theorem 3.3. *Under the condition that $N_1N_2M_1M_2/N^3 \to \infty$, we have*

$$\frac{\widehat{\mathrm{Var}(X)}}{\mathrm{Var}_\theta(X)} \xrightarrow{p} 1, \tag{3.7}$$

and therefore,

$$\hat\theta^{-1}\cdot(\widehat{\mathrm{Var}(X)})^{1/2}\cdot(\hat\theta-\theta) \to_{\mathcal{L}} N(0,1). \tag{3.8}$$

Proof. In view of inequality (2.11), we have

$$\min(\hat\theta^2/\theta^2, \theta^2/\hat\theta^2) \le \frac{\widehat{\mathrm{Var}(X)}}{\mathrm{Var}_\theta(X)} \le \max(\hat\theta^2/\theta^2, \theta^2/\hat\theta^2).$$

From this and the fact that $\hat{\theta} \to \theta$ in probability we conclude (3.7).

3.2. Empirical odds ratio

We have shown that the MLE for θ is consistent and asymptotically normal under the minimum condition $N_1M_1N_2M_2/N^3 \to \infty$. An alternative and, perhaps, more popular estimator of θ is the empirical odds ratio

$$\hat{\theta}_e = \frac{X(X+N_2-M_1)}{(N_1-X)(M_1-X)},$$

which is certainly computationally more convenient and appears to be more natural. As we mentioned earlier, $\hat{\theta}_e$ is the maximum likelihood estimator for the odds ratio in the 2×2 table when one of its margins is not fixed. Thus, it is intuitively clear that $\hat{\theta}_e$ should be close to $\hat{\theta}$. For this reason $\hat{\theta}_e$ is also called the asymptotic maximum likelihood estimator (Breslow and Day (1980), p. 130). We shall show that $\hat{\theta}_e$ is, indeed, asymptotically equivalent to $\hat{\theta}$, again under the minimum condition $N_1M_1N_2M_2/N^3 \to \infty$. We shall also give rigorous justifications for some classical methods of constructing confidence intervals for θ. To do so we need some preliminary results.

Lemma 3.1. *For the noncentral hypergeometric random variable X, we have:*
(i) $\mathrm{Var}_\theta(X) \le \min\{\mathrm{E}_\theta(X), \mathrm{E}_\theta(N_1-X), \mathrm{E}_\theta(M_1-X), \mathrm{E}_\theta(X+N_2-M_1)\}$.
(ii) $\mathrm{E}_\theta[X(X+N_2-M_1)]/\mathrm{E}_\theta[(N_1-X)(M_1-X)] = \theta$.
(iii) *Let Y be either X or N_1-X or M_1-X or $X+N_2-M_1$ and θ be the true odds ratio. Then $Y/\mathrm{E}_\theta Y \to 1$ in probability as $N_1M_1N_2M_2/N^3 \to \infty$.*

Proof. From (2.6), $\mathrm{Var}_\theta(X) \le \mathrm{E}_\theta(X)$. So, by symmetry, (i) holds. The identity (ii) is implicitly stated in Harkness (1965, p. 939, (3)) for the case that $N_2 \ge M_1$, and obtained in the current form by Mantel and Hankey (1975). Part (iii) follows readily from Theorem 2.2, the Markov inequality, and part (i).

Lemma 3.2. *For any $\epsilon \in (0,1)$, as $N_1M_1N_2M_2/N^3 \to \infty$,*

$$\sup_{\epsilon\le\theta\le\epsilon^{-1}} \Big| \Big(\frac{\mathrm{E}_\theta(X)\mathrm{E}_\theta(X+N_2-M_1)}{\mathrm{E}_\theta(N_1-X)\mathrm{E}_\theta(M_1-X)} - \theta\Big)\Big(\frac{1}{\mathrm{E}_\theta(N_1-X)} + \frac{1}{\mathrm{E}_\theta(M_1-X)}\Big)^{-1}\Big| = O(1). \tag{3.9}$$

Proof. First of all, note that by (2.8)

$$\frac{1}{\mathrm{E}_\theta(N_1-X)} + \frac{1}{\mathrm{E}_\theta(M_1-X)} \longrightarrow 0.$$

With covariance formulas

$$\begin{aligned} \mathrm{E}_\theta[X(X+N_2-M_1)] &= \mathrm{Var}_\theta(X) + \mathrm{E}_\theta(X)\mathrm{E}_\theta(X+N_2-M_1), \\ \mathrm{E}_\theta[(N_1-X)(M_1-X)] &= \mathrm{Var}_\theta(X) + \mathrm{E}_\theta(N_1-X)\mathrm{E}_\theta(M_1-X), \end{aligned}$$

we have, using Lemma 3.1.(ii),

$$\theta = \frac{\mathrm{E}_\theta[X(X+N_2-M_1)]}{\mathrm{E}_\theta[(N_1-X)(M_1-X)]} = \frac{A+B}{A+1},$$

where

$$A = \frac{\mathrm{Var}_\theta(X)}{\mathrm{E}_\theta(N_1-X)\mathrm{E}_\theta(M_1-X)} \qquad \text{and} \qquad B = \frac{\mathrm{E}_\theta(X)\mathrm{E}_\theta(X+N_2-M_1)}{\mathrm{E}_\theta(N_1-X)\mathrm{E}_\theta(M_1-X)}.$$

Therefore, $B-\theta = A(\theta-1)$. However, uniformly in $\theta \in [\epsilon, \epsilon^{-1}]$,

$$A = O\Big(\frac{1}{\mathrm{E}_\theta(N_1-X)} + \frac{1}{\mathrm{E}_\theta(M_1-X)}\Big),$$

via Lemma 3.1.(i). Hence (3.9) follows.

Define

$$\eta_\theta = [\mathrm{E}_\theta(X)]^{-1} + [\mathrm{E}_\theta(X+N_2-M_1)]^{-1} + [\mathrm{E}_\theta(N_1-X)]^{-1} + [\mathrm{E}_\theta(M_1-X)]^{-1}.$$

We have the following theorem.

Theorem 3.4. *Let θ be the true parameter. Assume $N_1M_1N_2M_2/N^3 \to \infty$. Then*

$$\log\hat{\theta}_e - \log\theta = \eta_\theta(X - \mathrm{E}_\theta X) + O_p(\eta_\theta). \tag{3.10}$$

Proof. It is clear from Lemma 3.2 and then Lemma 3.1 that

$$\begin{aligned}
&\log\hat{\theta}_e - \log\theta \\
&= \log\frac{X}{\mathrm{E}_\theta(X)} + \log\frac{X+N_2-M_1}{\mathrm{E}_\theta(X+N_2-M_1)} - \log\frac{N_1-X}{\mathrm{E}_\theta(N_1-X)} - \log\frac{M_1-X}{\mathrm{E}_\theta(M_1-X)} + O(\eta_\theta) \\
&= \frac{X}{\mathrm{E}_\theta(X)} - 1 + \frac{X+N_2-M_1}{\mathrm{E}_\theta(X+N_2-M_1)} - 1 - \frac{N_1-X}{\mathrm{E}_\theta(N_1-X)} + 1 - \frac{M_1-X}{\mathrm{E}_\theta(M_1-X)} + 1 + O(\eta_\theta) \\
&\quad + O_p\Big(\Big(\frac{X}{\mathrm{E}_\theta(X)} - 1\Big)^2 + \Big(\frac{X+N_2-M_1}{\mathrm{E}_\theta(X+N_2-M_1)} - 1\Big)^2 + \Big(\frac{N_1-X}{\mathrm{E}_\theta(N_1-X)} - 1\Big)^2 \\
&\qquad + \Big(\frac{M_1-X}{\mathrm{E}_\theta(M_1-X)} - 1\Big)^2\Big) \\
&= (X - \mathrm{E}_\theta X)\Big(\frac{1}{\mathrm{E}_\theta(X)} + \frac{1}{\mathrm{E}_\theta(X+N_2-M_1)} + \frac{1}{\mathrm{E}_\theta(N_1-X)} + \frac{1}{\mathrm{E}_\theta(M_1-X)}\Big) + O_p(\eta_\theta),
\end{aligned}$$

where the last equality follows from the Markov inequality and

$$\mathrm{E}_\theta\Big(\frac{X}{\mathrm{E}_\theta(X)} - 1\Big)^2 = \frac{\mathrm{Var}_\theta(X)}{(\mathrm{E}_\theta(X))^2} \le \frac{1}{\mathrm{E}_\theta(X)},$$

etc. Hence (3.10) holds.

Now, since $X - \mathrm{E}_\theta X$ is asymptotically normal by Theorem 2.2, $\log\hat{\theta}_e - \log\theta$ is also asymptotically normal by Theorem 3.4. More precisely, we have the following corollary.

Corollary 3.2. *Let θ be the true parameter and assume $N_1M_1N_2M_2/N^3 \to \infty$. Then*

$$\begin{aligned}\frac{\log\hat{\theta}_e - \log\theta}{\eta_\theta\sqrt{\mathrm{Var}_\theta(X)}} &\to_{\mathcal{L}} N(0,1),\\ \frac{\hat{\theta}_e - \theta}{\theta\eta_\theta\sqrt{\mathrm{Var}_\theta(X)}} &\to_{\mathcal{L}} N(0,1).\end{aligned} \tag{3.11}$$

To construct a confidence interval, Woolf (1955) suggested that $\log\hat{\theta}_e$ might be asymptotically normal with variance being approximately equal to $X^{-1}+(X+N_2-M_1)^{-1}+(N_1-X)^{-1}+(M_1-X)^{-1}$. Cornfield (1956), on the other hand, gave a heuristic derivation for the asymptotic variance of X to be approximated by $[\tilde{x}^{-1}+(\tilde{x}+N_2-M_1)^{-1}+(N_1-\tilde{x})^{-1}+(M_1-\tilde{x})^{-1}]^{-1}$, where $\tilde{x}$ is the mode for the distribution of X. Since X is asymptotically normal, $\tilde{x}/\mathrm{E}_\theta X \to 1$. Therefore, $\tilde{x}$ may be replaced by $\mathrm{E}_\theta X$, which in turn can be approximated by X. In the rest of this section, we shall supply proofs for these asymptotics.

Theorem 3.5. *Suppose that $N_1M_1N_2M_2/N^3 \to \infty$. Then for any $0 < \theta < \infty$,*

$$\mathrm{Var}_\theta(X)\Big(\frac{1}{\mathrm{E}_\theta(X)} + \frac{1}{\mathrm{E}_\theta(X+N_2-M_1)} + \frac{1}{\mathrm{E}_\theta(N_1-X)} + \frac{1}{\mathrm{E}_\theta(M_1-X)}\Big) \longrightarrow 1.$$

To prove Theorem 3.5, we need following bounds on $\hat{\theta}_e$ and $\hat{\theta}$.

Lemma 3.3. *Consider the following inequalities:*

A.1. $\hat{\theta}_e \le \hat{\theta} \le \hat{\theta}_e + \frac{X(M_1N_1-NX)}{M_1N_1(M_1-X)(N_1-X)}$,

A.2. $\hat{\theta}_e \le \hat{\theta} \le \hat{\theta}_e + \frac{(X-M_1+N_2)(M_1N_1-NX)}{M_2N_2(N_1-X)(M_1-X)}$,

B.1. $\hat{\theta}_e\Big(1+\frac{NX-M_1N_1}{M_2N_1(M_1-X)}\Big)^{-1} \le \hat{\theta} \le \hat{\theta}_e$,

B.2. $\hat{\theta}_e\Big(1+\frac{NX-M_1N_1}{M_1N_2(N_1-X)}\Big)^{-1} \le \hat{\theta} \le \hat{\theta}_e$.

(i) *If N_1 is the smallest among M_i, N_i, $i=1,2$, then*

$$\left\{\begin{array}{lll}\text{A.1} & \textit{holds}, & \textit{if}\ \ L < X \le M_1N_1/N \\ \text{B.1} & \textit{holds}, & \textit{if}\ \ M_1N_1/N \le X < M_1\end{array}\right\}.$$

(ii) *If M_1 is the smallest among M_i, N_i, $i=1,2$, then*

$$\left\{\begin{array}{lll}\text{A.1} & \textit{holds}, & \textit{if}\ \ L < X \le M_1N_1/N \\ \text{B.2} & \textit{holds}, & \textit{if}\ \ M_1N_1/N \le X < N_1\end{array}\right\}.$$

(iii) *If M_2 is the smallest, then*

$$\left\{\begin{array}{lll} \text{A.2} & \textit{holds}, & \textit{if} \quad L < X \le M_1N_1/N \\ \text{B.1} & \textit{holds}, & \textit{if} \quad M_1N_1/N \le X < N_1 \end{array}\right\}.$$

(iv) *If N_2 is the smallest, then*

$$\left\{\begin{array}{lll} \text{A.2} & \textit{holds}, & \textit{if} \quad L < X \le M_1N_1/N \\ \text{B.2} & \textit{holds}, & \textit{if} \quad M_1N_1/N \le X < N_1 \end{array}\right\}.$$

When $X = M_1N_1/N$, $\hat{\theta}_e = \hat{\theta} = 1$, and the above bounds become equalities. Lemma 3.3.(i) and the first inequality in (ii) are taken from Harkness (1965) (See also Johnson, Kotz and Kemp (1992), p.281). Note that we have corrected an error in Harkness (1965), equation (13) which is perpetuated in Johnson, Kotz and Kemp (1992), p.281. Since the proof in the original paper of Harkness is also rather sketchy, we shall provide a detailed proof for the preceding lemma in Section 5.

Since Lemma 3.1.(ii) and Corollary 2.1 imply

$$\frac{X(M_1N_1 - NX)}{M_1N_1(X - M_1)(X - N_1)} = O_p(\frac{N^3}{N_1M_1N_2M_2}),$$
$$\frac{NX - M_1N_1}{M_2N_1(M_1 - X)} = O_p(\frac{N^3}{N_1M_1N_2M_2}),$$

among others, we get, from Lemma 3.3 and Theorem 3.2, the following corollary.

Corollary 3.3. *Suppose that $N_1M_1N_2M_2/N^3 \to \infty$. Then for any $0 < \theta < \infty$,*

$$|\hat{\theta}_e - \hat{\theta}| = O_p\Big(\frac{N^3}{N_1M_1N_2M_2}\Big) = O_p([\mathrm{Var}_{\,\theta}(X)]^{-1}), \tag{3.12}$$

and therefore

$$\theta^{-1}\sqrt{\mathrm{Var}_{\,\theta}(X)}(\hat{\theta}_e - \theta) \quad \to_{\mathcal{L}} \quad N(0,1), \tag{3.13}$$
$$\sqrt{\mathrm{Var}_{\,\theta}(X)}(\log\hat{\theta}_e - \log\theta) \quad \to_{\mathcal{L}} \quad N(0,1).$$

Proof of Theorem 3.5. The result follows by combining (3.12) with (3.11).

Corollary 3.4. *With $0<\theta<\infty$ being the true parameter, as $N_1M_1N_2M_2/N^3 \to \infty$,*

$$\Big(\frac{1}{X} + \frac{1}{X{+}N_2{-}M_1} + \frac{1}{N_1{-}X} + \frac{1}{M_1{-}X}\Big)\mathrm{Var}_{\,\theta}(X) \ \xrightarrow{p}\ 1, \tag{3.14}$$

$$\Big(\frac{1}{X} + \frac{1}{X{+}N_2{-}M_1} + \frac{1}{N_1{-}X} + \frac{1}{M_1{-}X}\Big)^{-1/2}(\log\hat{\theta}_e - \log\theta) \ \to_{\mathcal{L}}\ N(0,1), \tag{3.15}$$

$$\hat{\theta}_e^{-1}\Big(\frac{1}{X} + \frac{1}{X{+}N_2{-}M_1} + \frac{1}{N_1{-}X} + \frac{1}{M_1{-}X}\Big)^{-1/2}(\hat{\theta}_e - \theta) \ \to_{\mathcal{L}}\ N(0,1). \tag{3.16}$$

Proof. In view of Lemma 3.1,

$$\frac{[\mathrm{E}_\theta(X)]^{-1}+[\mathrm{E}_\theta(X+N_2-M_1)]^{-1}+[\mathrm{E}_\theta(N_1-X)]^{-1}+[\mathrm{E}_\theta(M_1-X)]^{-1}}{X^{-1}+(X+N_2-M_1)^{-1}+(N_1-X)^{-1}+(M_1-X)^{-1}} \xrightarrow{p} 1.$$

Our conclusion then follows from Theorem 3.5 and Corollary 3.2.

Results (3.14) and (3.15) provide justifications, under the minimum condition, for the variance approximations proposed by Cornfield (1956) and Woolf (1955). Furthermore, (3.15) and (3.16) are two classical ways to construct asymptotic confidence intervals for θ. We are currently studying, both numerically and analytically, the two constructions as well as the one given by (3.8). The results will be reported elsewhere.

4. Poisson Approximation

We next discuss the Poisson approximation to the noncentral hypergeometric distribution. Since $X \geq (M_1-N_2)^+$, where $(M_1-N_2)^+ = \max\{M_1-N_2, 0\}$, we need to consider $X-(M_1-N_2)^+$. To avoid this, we may select the cell for X so that $N_1 \leq N_2$ and $M_1 \leq M_2$, effectively putting $(M_1-N_2)^+=0$. The following theorem gives a simple necessary and sufficient condition on such approximation.

Theorem 4.1. *Let X be noncentral hypergeometric with parameter θ. Assume $N_1 \leq N_2$ and $M_1 \leq M_2$. Then a necessary and sufficient condition for*

$$\sup_{k\geq 0}\Big|P(X=k)-\frac{\alpha_N^k}{k!}e^{-\alpha_N}\Big| \to 0 \quad \textit{as } N\to\infty \tag{4.1}$$

for some $\{\alpha_N\}$ bounded away from 0 and ∞, i.e. $0<\inf_N \alpha_N \leq \sup_N \alpha_N < \infty$, is that

$$\min(M_1,\ N_1)\to\infty \quad \textit{and} \quad M_1N_1/N \textit{ is bounded away from } 0 \textit{ and } \infty. \tag{4.2}$$

Remark 4.1. In the case of $\alpha_N \to \alpha$, (4.1) becomes $X \to_{\mathcal{L}}$ Poisson(α).

Remark 4.2. Theorem 4.1 is useful in considering when it is appropriate to apply the normal approximation or the Poisson approximation. If M_1N_1/N is bounded, then we know that $M_1M_2N_1N_2/N^3 \leq M_1N_1/N$ is also bounded. In this case the necessary and sufficient condition for the normal approximation is violated but the Poisson approximation is appropriate due to the preceding theorem. On the other hand, if $M_1M_2N_1N_2/N^3$ is large, then the normal approximation is appropriate but the Poisson is not.

Remark 4.3. In general, if we do not require $M_1 \leq M_2$ and $N_1 \leq N_2$, then the theorem should be rephrased. The necessary and sufficient condition for $X - (M_1 - N_2)^+$ to be approximately Poisson is that $\min(M_i, N_i, i = 1, 2) \to \infty$ and $M_1 N_1/N - (M_1 - N_2)^+$ is bounded away from 0 and ∞. The same proof as given below with minor variations applies to the general case.

Proof of Theorem 4.1. We first prove the necessity. Suppose (4.1) holds. Then clearly $M_1 \to \infty$ and $N_1 \to \infty$ since $X \leq \min(M_1, N_1)$. Therefore we only need to show that $M_1 N_1/N$ is bounded away from 0 and ∞. Clearly $M_1 N_1/N$ must be bounded away from 0 because otherwise there is a subsequence along which $\mathrm{E}_\theta(X) \to 0$, a contradiction to (4.1). We may assume (again by the subsequence argument) $\alpha_N \to \alpha > 0$. Now if we can show that the summands in (2.4) are uniformly asymptotically negligible (Loève (1977), p. 302), which is equivalent to

$$\max_{1 \leq i \leq S} (1 + \theta^{-1}\lambda_i)^{-1} \to 0, \qquad \text{or} \qquad \min_i \lambda_i \to \infty, \tag{4.3}$$

then the Poisson Convergence Criterion (Loève (1977), p. 329) entails $\mathrm{E}_\theta X \to \alpha$. By (2.8) we then know that $M_1 N_1/N$ is bounded away from ∞.

To show (4.3), rearrange the λ_i so that $0 < \lambda_1 \leq \cdots \leq \lambda_S$. Suppose (4.3) does not hold, we can then, by the subsequence argument, assume $\lambda_1 \to \lambda^* < \infty$. Define $U = I(\eta_1 \leq (1 + \theta^{-1}\lambda_1)^{-1})$ and $V = X - U$. We have $U \to_{\mathcal{L}} U^*$, a Bernoulli random variable with the success probability $p^* = (1 + \theta^{-1}\lambda^*)^{-1}$, and $V \to_{\mathcal{L}} V^*$, for some proper disitribution V^*, because U and V are independent and $U + V = X \to_{\mathcal{L}}\text{Poisson}(\alpha)$. Therefore, we obtain the following equation about the probability generating function:

$$[(z-1)p^* + 1]\mathrm{E}_\theta(z^{V^*}) = e^{(z-1)\alpha},$$

which, however, can not hold at $(z-1)p^* + 1 = 0$ or $z = (p^* - 1)/p^*$. Hence, (4.3) must be true.

We prove next the sufficiency. It follows from (4.2) that $N_2/N \to 1$ and $M_2/N \to 1$. Therefore,

$$N_1 M_1/N - N_1 M_1 N_2 M_2/[N^2(N-1)] \to 0. \tag{4.4}$$

(4.4) and (2.7) give us $\sum 1/(1+\lambda_i)^2 \to 0$, which implies (4.3); and thus the summands in (2.4) are uniformly asymptotically negligible. Let $\alpha_N = \sum_{i=1}^{S} 1/(1 + \theta^{-1}\lambda_i)$. Note that α_N is bounded away from 0 and ∞, as it is also so for $M_1 N_1/N$. We may then assume without loss of generality that $\alpha_N \to \alpha$. Now (4.1) follows readily by checking the two conditions in the Poisson Convergence Criterion (Loève (1977), p. 329).

5. Proof of Lemma 3.3

Let us denote the noncentral hypergeometric distribution by Hyper$(N_1, N_2, M_1, M_2, \theta)$. First of all we point out that only part (i) needs to be shown. Indeed,

(1) if N_1 is the smallest, then we can rewrite $X \overset{d}{=}$ Hyper$(M_1, M_2, N_1, N_2, \theta)$, and both the empirical odds ratio and the MLE will not be altered afterwards;

(2) if N_2 is the smallest, then it can be seen plainly that $X = M_1 - N_2 + Y$, where Y is distributed as Hyper$(M_2, M_1, N_2, N_1, \theta)$, whence, the empirical odds ratio and the MLE of Y are exactly equal to those of X;

(3) similarly, if M_2 is the smallest, then we use $X = M_1 - N_2 + Y$, where Y is distributed as Hyper$(N_2, N_1, M_2, M_1, \theta)$.

The proof of part (1) consists of two parts.

Part 1. The case $0 = L < X < M_1N_1/N$.

In this case, clearly $M_1 \geq 2$ and $N_1 \geq 2$. We get from Lemma 3.1.(ii) that

$$(1-\theta)(\mathrm{E}_\theta(X))^2 + c\mathrm{E}_\theta(X) - \theta N_1M_1 + (1-\theta)\mathrm{Var}_\theta(X) = 0, \tag{5.1}$$

where $c \equiv N - (N_1 + M_1)(1-\theta)$. Therefore,

$$f(\theta) \equiv \mathrm{E}_\theta(X) = \frac{-c + \sqrt{c^2 + 4\theta(1-\theta)N_1M_1 - 4(1-\theta)^2\mathrm{Var}_\theta(X)}}{2(1-\theta)}.$$

So if $0 < \theta < 1$, then

$$f(\theta) \leq \frac{-c + \sqrt{c^2 + 4\theta(1-\theta)N_1M_1}}{2(1-\theta)} = r(\theta) \quad \text{(say)}. \tag{5.2}$$

Note here $f(\theta)$ is a monotone function, and for the MLE $\hat{\theta}$, $f(\hat{\theta}) = X$. Moreover, $r(\theta)$ is exactly the *only* nonnegative root of the quadratic equation of y,

$$(1-\theta)y^2 + cy - \theta N_1M_1 = 0, \tag{5.3}$$

or equivalently, the unique positive root of

$$\frac{y(y + N - (M_1 + N_1))}{(N_1 - y)(M_1 - y)} = \theta. \tag{5.4}$$

It is interesting to see the similarity between (5.4) and the definition of the empirical odds ratio. Denote by $r_{-1}(\theta)$ and $f_{-1}(\theta)$ the positive root of (5.4) and the expectation of X, respectively, with N_1, M_1, N each reduced by 1 while N_2, M_2, θ are kept the same (this is feasible since $M_1, N_1 \geq 2$). Clearly, both $r(\theta)$ and $r_{-1}(\theta)$ are monotone increasing functions. Moreover,

$$r(\hat{\theta}_e) = X. \tag{5.5}$$

Taking the derivatives of (1.2), we get

$$\mathrm{E}_\theta(X) = \frac{\alpha\beta}{\gamma} \cdot \frac{\theta F(\alpha+1,\beta+1,\gamma+1,\theta)}{F(\alpha,\beta,\gamma,\theta)},$$

$$\mathrm{E}_\theta(X^2) - \mathrm{E}_\theta(X) = \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)} \cdot \frac{\theta^2 F(\alpha+2,\beta+2,\gamma+2,\theta)}{F(\alpha,\beta,\gamma,\theta)},$$

from which we find that

$$\mathrm{E}_\theta(X) f_{-1}(\theta) = \mathrm{E}_\theta(X^2) - \mathrm{E}_\theta(X), \qquad N_1 + M_1 \le N, \tag{5.6}$$

where $\alpha = -N_1$, $\beta = -M_1$, $\gamma = N_2 + 1 - M_1$. Thus, (5.6), in conjunction with (5.1), yields a recurrence relation

$$f(\theta) = \mathrm{E}_\theta(X) = \frac{N_1 M_1 \theta}{(1-\theta) f_{-1}(\theta) + c + (1-\theta)}, \quad 0 < \theta < 1, \quad N_1 + M_1 \le N. \tag{5.7}$$

A key fact to be used later on is

$$g(\theta) \le \mathrm{E}_\theta X = f(\theta) \le r(\theta), \qquad 0 < \theta < 1, \quad N_1 + M_1 \le N, \tag{5.8}$$

where $g(\theta)$ is a monotone increasing function defined by

$$g(\theta) = \frac{N_1 M_1 r_{-1}(\theta)}{r_{-1}(\theta) + N_1 M_1 - N_1 - M_1 + 1}.$$

To show (5.8) we only need to verify $f(\theta) \ge g(\theta)$, noting that (5.2) implies $f(\theta) \le r(\theta)$. But since $f_{-1}(\theta) \le r_{-1}(\theta)$, (5.7) yields

$$f(\theta) \ge \frac{N_1 M_1 \theta}{(1-\theta) r_{-1}(\theta) + c + (1-\theta)}. \tag{5.9}$$

Recalling (5.3), we get

$$r_{-1}^2 (1-\theta) + (c + 1 - 2\theta) r_{-1} - \theta(N_1 M_1 - N_1 - M_1 + 1) = 0,$$

from which we see that the right-hand side of (5.9) is exactly equal to $g(\theta)$, and (5.8) is obtained.

Since $X < M_1 N_1 / N$, $\hat\theta < 1$. substituting θ by $\hat\theta$ in (5.8), we get $g(\hat\theta) \le f(\hat\theta) = X \le r(\hat\theta)$, whence $r^{-1}(X) \le \hat\theta \le g^{-1}(X)$, where $r^{-1}(\cdot)$ and $g^{-1}(\cdot)$ are inverse functions of $r(\cdot)$ and $g(\cdot)$, respectively. Immediately, (5.5) implies that $r^{-1}(X) = \hat\theta_e$. Setting

$$D = \frac{X(N_1 - 1)(M_1 - 1)}{N_1 M_1 - X}, \qquad \theta^{(1)} = \frac{D(D + N - (N_1 + M_1) + 1)}{(D - N_1 + 1)(D - M_1 + 1)},$$

we then get

$$\frac{N_1 M_1 D}{D + N_1 M_1 - N_1 - M_1 + 1} = X, \quad r_{-1}(\theta^{(1)}) = D.$$

Therefore, $g(\theta^{(1)}) = X$ or, equivalently, $g^{-1}(X) = \theta^{(1)}$. Hence, we conclude

$$\hat{\theta}_e \le \hat{\theta} \le \theta^{(1)}. \tag{5.10}$$

After some algebra, we see that

$$D - N_1 + 1 = \frac{M_1(N_1 - 1)(X - N_1)}{N_1 M_1 - X},$$
$$D - M_1 + 1 = \frac{N_1(M_1 - 1)(X - M_1)}{N_1 M_1 - X},$$
$$D + N - (N_1 + M_1) + 1 = \frac{N_1 M_1}{N_1 M_1 - X}(X + N - \frac{NX}{M_1 N_1} - N_1 - M_1 + 1),$$

from which we obtain

$$\theta^{(1)} = \hat{\theta}_e + \frac{X(M_1 N_1 - NX)}{M_1 N_1 (X - M_1)(X - N_1)},$$

and the proof for this part is completed by invoking (5.10).

Part 2. The case $N_1 > X > M_1 N_1 / N$.

By symmetry, it is easy to see that $\mathrm{E}(X|N_1, M_1, N, \theta) = N_1 - \mathrm{E}(X|N_1, M_2, N, 1/\theta)$. Therefore, analogous to (5.8),

$$N_1 - \tilde{g}(\theta) \le \mathrm{E}_\theta X = f(\theta) \le N_1 - \tilde{r}(\theta), \quad 1 < \theta < \infty, \quad N_1 + M_2 \le N, \tag{5.11}$$

where $\tilde{r}(\theta)$ is the unique positive root of the equation

$$\frac{1}{\theta} = \frac{y(y + N - (M_2 + N_1))}{(N_1 - y)(M_2 - y)}, \tag{5.12}$$

and we set

$$\tilde{g}(\theta) = \frac{N_1 M_2 \tilde{r}_{-1}(\theta)}{\tilde{r}_{-1}(\theta) + N_1 M_2 - N_1 - M_2 + 1},$$

where $\tilde{r}_{-1}(\theta)$ denotes the positive root of (5.12) with N_1, M_2, N each reduced by 1 while N_2, M_1, θ are kept the same. Note that (5.11) requires that $N_1 + M_2 \le N$, because it comes from (5.8). We point out that this requirement was ignored in Harkness (1965), making a major result there invalid as stated. Clearly, $\tilde{r}(\theta)$, $\tilde{r}_{-1}(\theta)$, $\tilde{g}(\theta)$ are all monotone decreasing functions.

Since $X > M_1 N_1 / N$, $\hat{\theta} > 1$, therefore, we get from (5.11)

$$N_1 - \tilde{g}(\hat{\theta}) \le f(\hat{\theta}) = X \le N_1 - \tilde{r}(\hat{\theta}).$$

Thus we just need to invert the above functions. Set

$$E = \frac{(N_1 - X)(N_1 - 1)(M_2 - 1)}{N_1 M_2 - (N_1 - X)}, \qquad \theta^{(2)} = \frac{(E - N_1 + 1)(E - M_2 + 1)}{E(E + N - (N_1 + M_2) + 1)}.$$

It follows that

$$N_1 - \frac{N_1 M_2 E}{E + N_1 M_2 - N_1 - M_2 + 1} = X, \quad \tilde{r}_{-1}(\theta^{(2)}) = E,$$

whence $N_1 - \tilde{g}(\theta^{(2)}) = X$. This, in conjunction with the fact that $N_1 - \tilde{r}(\hat{\theta}_e) = X$, gives $\theta^{(2)} \le \hat{\theta} \le \hat{\theta}_e$. Some algebra yields

$$\begin{aligned} E - N_1 + 1 &= \frac{M_2(N_1 - 1)((N_1 - X) - N_1)}{N_1 M_2 - (N_1 - X)}, \\ E - M_2 + 1 &= \frac{N_1(M_2 - 1)((N_1 - X) - M_2)}{N_1 M_2 - (N_1 - X)}, \\ E+N-(N_1+M_2)+1 &= \frac{N_1 M_2}{N_1 M_2 - (N_1 - X)}\Big((N_1 - X) + N - \frac{N(N_1 - X)}{M_2 N_1} \\ &\qquad - N_1 - M_2 + 1\Big). \end{aligned}$$

Thus

$$\theta^{(2)} = \hat{\theta}_e \Big(1 + \frac{NX - M_1 N_1}{M_2 N_1 (M_1 - X)}\Big)^{-1},$$

and part 2 of the proof is finished.

Acknowledgement

The authors would like to thank Professor Bruce Levin for the introduction and subsequent discussions about contingency tables. The research of the second author was supported by the National Science Foundation and the National Security Agency.

References

Breslow, N. E. and Day, N. E. (1980). *Statistical Methods in Cancer Research, Vol. I, The Design and Analysis of Case-Control Studies.* IARC, Lyon.

Chow, Y. S. and Teicher, H. (1988). *Probability Theory : Independence, Interchangeability, Martingales*, 2nd edition. Springer-Verlag, New York.

Cornfield, J. (1956). A statistical problem arising from retrospective studies. In *Proceedings of the Third Berkeley Symposium, IV* (Edited by Neyman, J.), p. 133-148 University of California Press, Berkeley.

Erdélyi, A., et al. (1953). *Higher Transcendental Functions, Vol II.* Bateman manuscript project. McGraw-Hill, New York.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications.* 3rd edition. John Wiley, New York.

Harkness, W. L. (1965). Properties of extended hypergeometric distribution. *Ann. Math. Statist.* **36**, 938-945.

Harper, L. H. (1967). Stirling behavior is asymptotically normal. *Ann. Math. Statist.* **38**, 410-414.

Johnson, N. L., Kotz, S. and Kemp, A. W. (1992). *Univariate Discrete Distributions.* 2nd. ed. John Wiley & Sons, New York.

Lehmann, E. L. (1975). *Nonparametrics.* Holden-Day, California.

Loève, M. (1977). *Probability Theory Vol. I.* Springer-Verlag, New York.

Mantel, N. and Hankey, B. F. (1975). The odds ratios of a 2×2 contingency table. *Amer. Statistician* **29**, 143-145.

Szegö, G. (1959). *Orthogonal Polynomials.* Amer. Math. Soc. Colloq. Pub., Vol. **23**, Amer. Math. Soc., Providence, R.I.

Vatutin, V. A. and Mikhailov, V. G. (1982). Limit theorems for the number of empty cells in an equiprobable scheme for group allocation of particles. *Theory of Probab. and its Appl.* **27**, 734-743.

Woolf, B. (1955). On estimating the relationship between blood group and disease. *Ann. Human Genetics* **19**, 251-253.

Department of Statistics, Hill Center, Busch Campus, Rutgers University, Piscataway, NJ 08855, U.S.A.

(Received January 1995; accepted March 1996)