AD-A267 146

Bounds on Inconsistent Inferences for Sequences of Trials with Varying Probabilities

Herman Chernoff and Yingnian Wu

Harvard University
Cambridge, MA 02139

Technical Report No. ONR-C-13

DTIC
ELECTE
JUL 22 1993
S B D

June 21, 1993

Reproduction in whole or in part is permitted for any purpose of the United States Government.

This document has been approved for public release and sale, its distribution is unlimited.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) ONR-C-13	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Department of Statistics Harvard University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Department of Statistics, Rm. SC713 Harvard University Cambridge, MA 02138	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING / SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable) Code 1111	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) Office of Naval Research Arlington, VA 22217-5000	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)
Bounds on Inconsistent Inferences for Sequences of Trials with Varying Probabilities

12. PERSONAL AUTHOR(S) Herman Chernoff and Yingnian Wu

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical Report	FROM _____ TO _____	June 21, 1993	17

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

See reverse side.

93-16544

93 7 21 050

20pl

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION
22a NAME OF RESPONSIBLE INDIVIDUAL Herman Chernoff	22b TELEPHONE (Include Area Code) 617-495-5462 / 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

Bounds on Inconsistent Inferences for Sequences of Trials with Varying Probabilities

Herman Chernoff and Yingnian Wu

Harvard University
Cambridge, MA 02139

ABSTRACT

Consider independent pairs of Bernoulli trials on two unknown sequences of probabilities $\mathbf{p}^{(1)} = \{p_i^{(1)} : 1 \le i \le n\}$ and $\mathbf{p}^{(2)} = \{p_i^{(2)} : 1 \le i \le n\}$. The data are the numbers of pairs which consist of (0,0), (0,1), (1,0), and (1,1) and can be summarized in a two-way table with entries n_{00}, n_{01}, n_{10} and n_{11} adding up to n. The two problems of estimating the mean and variance of the number of discordant pairs $n_{01} + n_{10}$ when $H_0 : \mathbf{p}^{(1)} = \mathbf{p}^{(2)}$ is true, and of testing H_0 using the number of discordant pairs as a test statistic are considered. Two novel issues arise. While relevant parameters can not be estimated consistently from the available data, useful bounds on these can be derived. While the test is poor for alternatives typically considered in the literature, it may be effective for detecting the presence of unknown explanatory factors which discriminate between supposedly matched pairs.

AMS 1991 subject classifications. Primary 62E17, secondary 62F03.

Key words and phrases. Two-way tables, matched pairs, odds ratio, discordant pairs.

Bounds on Inconsistent Inferences for Sequences of Trials with Varying Probabilities

Herman Chernoff and Yingnian Wu

Harvard University
Cambridge, MA 02139

ABSTRACT

Consider independent pairs of Bernoulli trials on two unknown sequences of probabilities $\mathbf{p}^{(1)} = \{p_i^{(1)} : 1 \le i \le n\}$ and $\mathbf{p}^{(2)} = \{p_i^{(2)} : 1 \le i \le n\}$. The data are the numbers of pairs which consist of (0,0), (0,1), (1,0), and (1,1) and can be summarized in a two-way table with entries n_{00}, n_{01}, n_{10} and n_{11} adding up to n. The two problems of estimating the mean and variance of the number of discordant pairs $n_{01} + n_{10}$ when $H_0 : \mathbf{p}^{(1)} = \mathbf{p}^{(2)}$ is true, and of testing H_0 using the number of discordant pairs as a test statistic are considered. Two novel issues arise. While relevant parameters can not be estimated consistently from the available data, useful bounds on these can be derived. While the test is poor for alternatives typically considered in the literature, it may be effective for detecting the presence of unknown explanatory factors which discriminate between supposedly matched pairs.

AMS 1991 subject classifications. Primary 62E17, secondary 62F03.

Key words and phrases. Two-way tables, matched pairs, odds ratio, discordant pairs.

DTIC QUALITY INSPECTED 5

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

1. Introduction.

The following two problems were posed in response to a question verbally posed by two geneticists. It turned out that the problems did not address their questions which could be answered by reference to U statistics. This left two problems of some theoretical interest, but with no apparent application. The recall of previous unpublished work by Cornfield and Greenhouse (1975) led to subsequent discussions with S. Greenhouse and J. L. Gastwirth which suggest potential applications of these problems to issues in discrimination.

Consider independent pairs of independent Bernoulli observations on two sequences of probabilities $\mathbf{p}^{(1)} = \{p_i^{(1)} : 1 \le i \le n\}$ and $\mathbf{p}^{(2)} = \{p_i^{(2)} : 1 \le i \le n\}$. In biostatistical and discrimination applications one is often interested in knowing whether the $p_i^{(1)}$ tend to be greater than the $p_i^{(2)}$. The data provide only four useful items of information for this situation involving $2n$ parameters. These are the numbers of pairs which consist of (0,0), (0,1), (1,0) and (1,1) respectively, and can be summarized in a two-way table with entries n_{00}, n_{01}, n_{10} and n_{11} adding up to n. In many such applications it is reasonable to formulate a test of the null hypothesis $H_0 : p_i^{(1)} = p_i^{(2)}$, $1 \le i \le n$ by postulating that the odd-ratios

$$\psi_i = \frac{p_i^{(1)}/(1-p_i^{(1)})}{p_i^{(2)}/(1-p_i^{(2)})}, \quad i = 1, 2, \ldots, n$$

are all equal to a common value ψ and to test whether $\psi = 1$. Several recent examples are Gastwirth and Greenhouse (1987) and Yu (1993).

An interesting case for analysis is that where $n_{00} = n_{11} = 1,000$, $n_{10} = 20$ but $n_{01} = 5$. While the overall success rates for both cases are almost equal, it is clear that the discrepancy between n_{10} and n_{01} is statistically significant and could lead to rejecting H_0 if that hypothesis were seriously intended. While this example would fail to prove that one treatment is much better than another (in a case where two treatments were applied to n matched pairs of individuals) the McNemar test would clearly demonstrate that there is a small subpopulation on which the treatments have a decidedly different effect. In other words, it would be evidence of the presence in the population of an explanatory factor discriminating among supposedly matched pairs, and which may or may not be important to uncover.

Another example, related to the problems we shall pose, is that where the data consist of $n_{00} = n_{11} = 0$ and $n_{10} = 100 = n_{01}$. In this case there is no indication that there is an overall tendency for the $p_i^{(1)}$ to exceed the $p_i^{(2)}$. Nevertheless in discrimination cases the data clearly show that subjects were treated differently, depending on the existence of some hidden factors.

The problems originally posed are the following.

Problem 1. Assuming that $H_0 : p_i^{(1)} = p_i^{(2)} = p_i$ for $1 \le i \le n$, where $\mathbf{p} = \{p_i : 1 \le i \le n\}$ is unspecified, what can be said about the mean and variance of

$$D = (n_{01} + n_{10})/n \ ?$$

Problem 2. Test the hypothesis H_0 using D as a test statistic.

Because of the paucity of data, it is unlikely that D will be an effective test statistic for testing H_0 in most applications. Nevertheless, as the second example indicates, there may be situations where D leads to rejecting H_0 and reveals the existence of an effective explanatory factor which may be worth discovering. We shall see that while it is impossible to estimate accurately the variances of estimates or the significance levels of tests, useful bounds on these may be derived. In Section 4 we shall generalize to the use of $(a_1\ n_{10} + a_2\ n_{01})/n$ as a test statistic. Note that as long as a_1 and a_2 are positive or have the same sign, the use of this generalization attacks the side issue of hidden factors rather than the usual issue of whether the $p_i^{(1)}$ tend to exceed the $p_i^{(2)}$.

In Section 5 we shall consider the case where there are three observations on $\mathbf{p}$ and that where there are two observations on $\mathbf{p}^{(1)}$ and one on $\mathbf{p}^{(2)}$.

Almost all of the derivations will appear in the appendices which will make extensive use of the Geometry of Moments presented in Karlin and Shapley (1953). Certain aspects of the geometry of the space of $(n_{00},\ n_{01},\ n_{10},\ n_{11})$ are discussed in Fienberg and Gilbert (1970) and in Diaconis (1977). Fienberg and Gilbert discuss, among others, the set on which there is a common odds ratio. Diaconis is interested in aspects relevant to exchangeability.

2. Problem 1.

We may regard D as the average of n Bernoulli random variables D_i where $D_i = 1$ if the i-th pair doesn't match, i.e., the pair consists of (1,0) or (0,1). Then the expectation and variance under H_0 are given by

$$E_0 D_i = 2p_i(1 - p_i) = d_i^{(0)}$$

and

$$\mathrm{Var}_0(D_i) = d_i^{(0)}(1 - d_i^{(0)}).$$

Thus

$$\Delta_0 = E_0 D = n^{-1} \sum d_i^{(0)} = \mathcal{E}(d^{(0)}) = 2\mathcal{E}\{p(1-p)\} \tag{2.1}$$

where $\mathcal{E}$ stands for the average over the n subscripted values. Similarly

$$\sigma_0^2 = n\mathrm{Var}_0(D) = \mathcal{E}\{d^{(0)}(1 - d^{(0)})\} = \mathcal{E}\{2p(1-p)(1 - 2p(1-p))\} \tag{2.2}$$

While D can be used as an estimate of Δ_0, σ_0^2 can not be estimated consistently from the available data. On the other hand, it is easy to see that $0 \leq \Delta_0 \leq 2\mathcal{E}(p)[1-\mathcal{E}(p)] \leq 1/2$ and we shall show in Appendix A3 that for a given Δ_0,

$$\Delta_0/2 \leq \sigma_0^2 \leq \Delta_0(1-\Delta_0) \tag{2.3}$$

Note that the ratio

$$\frac{\Delta_0(1-\Delta_0)}{\Delta_0/2} = 2(1-\Delta_0)$$

ranges from 2 to 1 as Δ_0 ranges from 0 to 1/2 and that the difference

$$\Delta_0(1-\Delta_0) - \Delta_0/2 = \Delta_0(1/2-\Delta_0)$$

ranges from 0 to 1/16 and back to zero, peaking at $\Delta_0 = 1/4$.

Treating the D_i as i.i.d. Bernoulli random variables with common probability Δ_0 would give the correct mean for D but could possibly overestimate the variance by a factor of $2(1-\Delta_0)$ which is close to 2 if Δ_0 is small. That means that a naive confidence interval for Δ_0 based on the assumption of a common probability would be conservative, and possibly by as much as a length factor of $\sqrt{2}$.

More precise bounds are derived in Appendix A3 making use of $\pi = \mathcal{E}(p)$ which can also be estimated from the data. Using π leads to relatively minor improvement of the upper bound. It has no effect on the lower bound in the triangle of (π, Δ_0) values with vertices $(0,0)$, $(1,0)$, and $(1/2, 1/2)$, but it leads to substantial improvement near the upper boundary where $\Delta_0 = 2\pi(1-\pi)$.

3. Problem 2.

Since $\Delta_0 = E_0 D$ in Problem 1 can range from 0 to 1/2, it follows that D can be used to reject H_0 only if D is significantly greater than 1/2. However if $\mathcal{E}(p)$ were known, then $\Delta_0 = E_0 D = 2\mathcal{E}(p(1-p)) \leq 2[\mathcal{E}(p) - [\mathcal{E}(p)]^2]$ and we could reject H_0 for values of $D < 1/2$ provided they were significantly greater than $2\mathcal{E}(p)[1-\mathcal{E}(p)]$. Not knowing $\mathcal{E}(p)$, we could estimate it and use as our test statistic

$$T = D - 2\hat{\pi}(1-\hat{\pi}) \tag{3.1}$$

where $\hat{\pi} = (\hat{\pi}^{(1)} + \hat{\pi}^{(2)})/2$, $\hat{\pi}^{(1)} = (n_{10} + n_{11})/n$, and $\hat{\pi}^{(2)} = (n_{01} + n_{11})/n$. Under the general assumptions where $\mathbf{p}^{(1)}$ is not necessarily equal to $\mathbf{p}^{(2)}$, we define $\mathbf{p} = (\mathbf{p}^{(1)} + \mathbf{p}^{(2)})/2$ and then $\hat{\pi}$, $\hat{\pi}^{(1)}$ and $\hat{\pi}^{(2)}$ are estimates of $\pi = \mathcal{E}(p)$, $\pi^{(1)} = \mathcal{E}(p^{(1)})$ and $\pi^{(2)} = \mathcal{E}(p^{(2)})$ respectively.

We see in Appendix A4 that

$$E(T) = [\Delta - 2\pi(1-\pi)] + \frac{1}{2n}\mathcal{E}\{p^{(1)}(1-p^{(1)}) + p^{(2)}(1-p^{(2)})\} \tag{3.2}$$

where

$$\Delta = E(D) = \mathcal{E}(d) \tag{3.3}$$

and $d_i = E(D_i) = p_i^{(1)}(1 - p_i^{(2)}) + p_i^{(2)}(1 - p_i^{(1)})$ for $1 \le i \le n$. The expression for $E(T)$ may be regarded as the sum of two terms, the second of which is $O(n^{-1})$ and is bounded from above by $[\pi^{(1)}(1 - \pi^{(1)}) + \pi^{(2)}(1 - \pi^{(2)})]/2n$. Under the hypothesis, the main term of $E_0(T)$ is $\Delta_0 - 2\pi(1 - \pi) = -2\sigma_p^2$ where $\sigma_p^2 = \mathcal{E}(p^2) - \pi^2$.

Neglecting terms of higher order, the variance of T is seen to be approximated by $n^{-1}\tau^2$ where

$$\tau^2 = \mathcal{E}\{4\pi^2 d - 4(2\pi - 1)pd - d^2\} - (2\pi - 1)^2\mathcal{E}((p^{(1)} - p^{(2)})^2) \tag{3.4}$$

Under the hypothesis H_0, this variance becomes

$$\tau_0^2(\pi,\ \Delta_0) = 4\pi^2\Delta_0 - 12\pi^4 + (4 - 16\pi + 24\pi^2)(\pi - \Delta_0/2) - 4\mathcal{E}(p - \pi)^4. \tag{3.5}$$

To study the range of the main term in $E(T)$, $\Delta - 2\pi(1 - \pi)$, we demonstrate in appendix A5 that

$$0 \le |\pi^{(1)} - \pi^{(2)}| \le \Delta \le 2\pi \qquad \text{if } 0 \le \pi \le 1/2 \tag{3.6a}$$

and

$$0 \le |\pi^{(1)} - \pi^{(2)}| \le \Delta \le 2(1 - \pi) \qquad \text{if } 1/2 \le \pi \le 1. \tag{3.6b}$$

and that these inequalities are sharp given $\pi^{(1)}$ and $\pi^{(2)}$. Without specifying $\pi^{(1)}$ and $\pi^{(2)}$, which can be estimated from the margins, we see that $(\pi,\ \Delta)$ lies in the triangle with vertices $(0,0)$, $(1/2,\ 1)$, and $(1,0)$. Under the hypothesis $(\pi,\ \Delta_0)$ is restricted to the subset of the triangle under the parabola $\Delta_0 = 2\pi(1 - \pi)$. Where $(\pi,\ \Delta)$ lies in the triangle depends on the value of $\sigma_{12} = \mathcal{E}(p^{(1)}p^{(2)}) - \pi^{(1)}\pi^{(2)}$. When $(\pi,\ \Delta)$ lies above the triangle, $E(T)$ is positive, the hypothesis is not true, and we will be able to reject H_0 with enough data. If $(\pi,\ \Delta)$ lies below the parabola, $E(T) \le 0$ and the hypothesis may or may not be true, but we will not be able to use T to reject the hypothesis. Of course other test statistics could be effective if we were aiming seriously at testing H_0. In particular it would be easy to detect deviations from $\pi^{(1)} = \pi^{(2)}$.

To maximize τ_0 subject to given values of $\Delta_0 = \mathcal{E}(2p(1 - p))$ and $\pi = \mathcal{E}(p)$ one must minimize

$$V_1 = \mathcal{E}(p - \pi)^4.$$

It is easy to see that the minimum of V_1, unrestricted by the condition $0 \le p \le 1$, is achieved by $p = \pi \pm \sigma_p$ each with probability $1/2$. When $\pi - \sigma_p < 0$ or $\pi + \sigma_p > 1$,

we see in Appendix A3 that the restricted minimum is achieved by one of the 2 point distributions assigning probability θ at q and $1-\theta$ at 0 or 1, for appropriate values of q and θ. In each case π and Δ_0 determine q and θ. Let

$$V_2(\pi,\sigma) = \sigma^2(\sigma^6+\pi^6)/[\pi^2(\sigma^2+\pi^2)] \tag{3.7}$$

The minimum of V_1 is

$$V_{1m} = \sigma_p^4 \qquad \text{if} \qquad \sigma_p < \pi_m = \min(\pi, 1-\pi), \tag{3.8a}$$

and otherwise

$$V_{1m} = V_2(\pi_m, \sigma_p). \tag{3.8b}$$

The maximum of V_1 is also attained by a two point distribution involving 0 or 1 if $0 \le \pi_m \le 1/4$, and is then

$$V_{1M} = V_2(1-\pi_m, \sigma_p) \tag{3.8c}$$

If $1/4 < \pi_m \le 1/2$ then V_M may be $V_2(1-\pi_m,\ \sigma_p)$ or may be attained by a 3 point distribution involving 0, 1, and $q = \text{med}(q_1, q_2, 2\pi - 1/2)$ where $q_1 = [\pi - \mathcal{E}(p^2)]/(1-\pi)$ and $q_2 = \mathcal{E}(p^2)/\pi$.

Then τ_0 is bounded above and below by $\tau_{0M}(\pi,\Delta_0)$ and $\tau_{0m}(\pi,\Delta_0)$ where these are derived from τ_0^2 by replacing $\mathcal{E}(p-\pi)^4$ by V_{1m} and V_{1M} respectively, and where

$$\Delta_0 = 2[\pi(1-\pi) - \sigma_p^2]\ . \tag{3.9}$$

For large n

$$Z = \frac{n^{1/2}[D - 2\hat{\pi}(1-\hat{\pi})]}{\tau_0(\hat{\pi}, D)} \tag{3.10}$$

should be approximately normally distributed with mean less than or equal to 0 and variance 1 when the hypothesis H_0 is true. The expectation of T and the bounds on τ_0 provide corresponding approximate bounds to the probability of rejection, when the hypothesis is true, for a test using T as the test statistic.

For a given joint distribution for $(p_i^{(1)}, p_i^{(2)})$ it is possible to calculate $E(T)$ and τ and to estimate the corresponding noncentrality parameter and the power of the test of H_0. For illustrative and computational purposes a mixture of independent beta distributions of the form $f(q_1,q_2) = \sum_{i=1}^{k} w_i Be(q_1;\alpha_{1i},\beta_{1i})Be(q_2;\alpha_{2i},\beta_{2i})$ might be suitable. To calculate bounds on the power function of the test without assuming a proposed distribution, we

should calculate bounds on τ^2 for given $\pi^{(1)}, \pi^{(2)}$ and Δ, which may be estimated from the data. Of course if $\pi^{(1)}$ and $\pi^{(2)}$ are not close, their estimates would clearly indicate that H_0 is not true. But our use of Z is directed more at detecting hidden explanatory factors than at testing the validity of H_0. In any case, bounding τ^2 involves minimizing and maximizing the variance of $(p^{(1)} - \pi^{(1)})(p^{(2)} - \pi^{(2)})$ subject to specified values of $\pi^{(1)}, \pi^{(2)}$ and $\mathcal{E}(p^{(1)}p^{(2)})$. This problem is discussed in Appendix 6.

4. Generalization of T.

The test statistic T treats the pair (1,0) the same as (0,1). To direct the test toward detecting specific alternatives where one of these pairs is more likely to occur than another, we may apply the test statistic

$$T_1 = (a_1 n_{10} + a_2 n_{01})/n - (a_1 + a_2)\hat{\pi}(1 - \hat{\pi}) \tag{4.1}$$

Then, we see in Appendix A4, that

$$\begin{aligned} ET_1 = -(a_1 + a_2)\sigma_{12} &+ \frac{a_1 - a_2}{2}(\pi^{(1)} - \pi^{(2)}) + \frac{a_1 + a_2}{4}(\pi^{(1)} - \pi^{(2)})^2 \\ &+ \frac{a_1 + a_2}{4n}\mathcal{E}[p^{(1)}(1 - p^{(1)}) + p^{(2)}(1 - p^{(2)})] \end{aligned} \tag{4.2}$$

where

$$\sigma_{12} = \mathcal{E}(p^{(1)}p^{(2)}) - \pi^{(1)}\pi^{(2)} = \frac{1}{2}[\pi^{(1)}(1 - \pi^{(2)}) + \pi^{(2)}(1 - \pi^{(1)}) - \Delta] \tag{4.3}$$

Also $\text{Var}\,(T_1) = n^{-1}\tau_1^2$ plus higher order terms where

$$\begin{aligned} \tau_1^2 = \mathcal{E}\Big\{ & b_1^2 p^{(1)}(1 - p^{(1)}) + b_2^2 p^{(2)}(1 - p^{(2)}) + (a_1 + a_2)^2 p^{(1)}p^{(2)}(1 - p^{(1)}p^{(2)}) \\ & - 2(a_1 + a_2)p^{(1)}p^{(2)}\big[b_1(1 - p^{(1)}) + b_2(1 - p^{(2)})\big]\Big\} \end{aligned} \tag{4.4}$$

where

$$b_1 = (a_1 + a_2)\pi + (a_1 - a_2)/2 \tag{4.5a}$$

and

$$b_2 = (a_1 + a_2)\pi - (a_1 - a_2)/2. \tag{4.5b}$$

Incidentally $b_1 + b_2 = 2\pi(a_1 + a_2)$, $b_1 - b_2 = a_1 - a_2$, $b_1 b_2 = \pi^2(a_1 + a_2)^2 - [(a_1 - a_2)/2]^2$.

5. Multiple Observations.

The difficulties in bounding the basic parameters in the inferences in our problems are mitigated when more observations are available on each p_i.

5.1 Problem 1 With 3 Observations.

Suppose that in Problem 1, we had 3 observations leading to the data n_0, n_1, n_2, n_3 where n_j is the number of i values (trials) for which we observe j successes. Then

$$E_0(n_j/n) = \binom{3}{j}\mathcal{E}\{p^j(1-p)^{3-j}\},\ j = 0,1,2,3 \tag{5.1}$$

and we can estimate $\mathcal{E}(p)$, $\mathcal{E}(p^2)$ and $\mathcal{E}(p^3)$, since

$$\begin{aligned} E_0[(3n_3 + 2n_2 + n_1)/3n] &= \mathcal{E}(p) \\ E_0[(3n_3 + n_2)/3n] &= \mathcal{E}(p^2) \\ E_0[n_3/n] &= \mathcal{E}(p^3) \end{aligned} \tag{5.2}$$

Thus we may estimate $\Delta_0 = 2\mathcal{E}\{p(1-p)\}$ by

$$\tilde{D} = 2(n_1 + n_2)/3n \tag{5.3}$$

for which

$$\tilde{\sigma}_0^2 = n\mathrm{Var}_0(\tilde{D}) = \mathcal{E}\{4p(1-p)[1-3p(1-p)]/3\}. \tag{5.4}$$

To bound $\tilde{\sigma}_0^2$ using our estimates, we need to bound $\mathcal{E}(p^4)$ given $\mathcal{E}(p)$, $\mathcal{E}(p^2)$ and $\mathcal{E}(p^3)$. That problem is addressed in Appendix A7.

With 4 observations on $\mathbf{p}$, we will have estimates of $\mathcal{E}(p)$, $\mathcal{E}(p^2)$, $\mathcal{E}(p^3)$ and $\mathcal{E}(p^4)$ and the variance of the natural estimate of Δ_0 may be estimated consistently form the available data.

5.2 Multiple Observations for Problem 2.

Suppose that for the test of H_0 we have two observations on $\mathbf{p}^{(1)}$ and one for $\mathbf{p}^{(2)}$. We may label our observations by n_{jk} = number of trials with j successes on $p_i^{(1)}$ and k successes on $p_i^{(2)}$ with $j = 0,1,2$ and $k = 0,1$.

One test statistics that may be used would be based on $(n_{20} + n_{01})/n$ which has expectation $\Delta_{30} = \mathcal{E}(p - p^3)$ under the hypothesis H_0. From the observations on $\mathbf{p}^{(1)}$ we can estimate $\mathcal{E}(p^{(1)})$ and $\mathcal{E}(p^{(1)2})$. From those on $\mathbf{p}^{(2)}$ we can estimate $\mathcal{E}(p^{(2)})$. Given $\mathcal{E}(p)$ and $\mathcal{E}(p^2)$, the bounds on Δ_{30} are derived in Appendix A8. In

particular the maximum value is $\pi - [\mathcal{E}(p^2)]^2/\pi$ which may be estimated by substituting $\hat{\pi} = (2n_{20} + n_{10} + n_{01})/3n$ for π and (n_{20}/n) for $\mathcal{E}(p^2)$. Thus a natural test statistic is

$$T_2 = \frac{1}{n}\left\{(n_{20} + n_{01}) - [\frac{2n_{20} + n_{10} + n_{01}}{3} - \frac{3n_{20}^2}{2n_{20} + n_{10} - n_{01}}]\right\}$$

$$= \frac{1}{n}\left\{\frac{n_{20} - n_{10} + 2n_{01}}{3} - \frac{3n_{20}^2}{2n_{20} + n_{10} + n_{01}}\right\} \tag{5.5}$$

We shall not elaborate on bounds on the variance of T_2 here. In a personal communication, K.F. Yu has pointed out that with Σ observations on each of $\mathbf{p}^{(1)}$ and $\mathbf{p}^{(2)}$, the statistic $(n_{02} + n_{20}) - n_{11}/2$ has mean 0 and variance estimated by $n_{02} + n_{20} + n_{11}/4$ under the hypothesis. Thus bounds are no longer required.

Appendix.

The following remarks represent a brief summary of the Geometry of Moments which is a major tool in deriving many of the bounds in this appendix. Let $\mathbf{h}(X)$ be a k-dimensional vector valued function of a random variable X. As the distribution F of X varies over a convex set of distributions, the range of $E\mathbf{h}(X)$ is a convex set. If the class of distributions is the set of all distributions over a closed bounded interval I and $\mathbf{h}$ is continuous on I, the range of $E\mathbf{h}(X)$ is the convex set generated by $\{\mathbf{h}(x) : x \in I\}$ and is closed and bounded.

To maximize one coordinate of $E\mathbf{h}(X)$ when the others are specified involves a boundary point of the convex set which can be represented in terms of a k point distribution (involving at most k points of I). Moreover there is a supporting hyperplane at this boundary point which maximizes some linear function $E\mathbf{a}^T\mathbf{h}(X)$, and every one of the k or fewer points of I maximizes $\mathbf{a}^T\mathbf{h}(x)$ for $x \in I$. Finally the coefficient of the coordinate being maximized can be taken to be one if the specified expectations lie in an interior point of their $k-1$ dimensional convex range.

As a simple example consider the range of $(E(X), E(X^2))$ over the class of all distributions on $[0,1]$. This is the convex set generated by $A = \{(x_1, x_2) : x_2 = x_1^2,\ 0 \le x_1 \le 1\}$ and is the set bounded by A and $B = \{(x_1, x_2) : x_2 = x_1,\ 0 \le x_1 \le 1\}$. It follows that, subject to $EX = \mu$, $\mu^2 \le EX^2 \le \mu$. Moreover, it is clear that these bounds may be achieved by the one point distribution at μ and a two point distribution which gives probability μ to 1 and $1-\mu$ to 0.

Given any point for which $\mu^2 \le \mu_2' \le \mu$, the class of distributions for which $EX = \mu$ and $EX^2 = \mu_2'$ must have support on a subset of A, the convex hull of which contains (μ, μ_2'). It follows that there are two points q_1 and q_2 in [0,1] such that no distribution for which $P\{X > q_1\} = 1$ or for which $P\{X < q_2\} = 1$ will yield the given values of

(EX, EX^2). These two points are obtained by observing where the lines for (1,1) and (0,0) through (μ, μ_2') intersect the generating parabola segment A. Thus, for $0 < \mu < 1$

$$(A.1) \qquad q_1 = (\mu - \mu_2')/(1 - \mu)$$

and

$$(A.2) \qquad q_2 = \mu_2'/\mu.$$

A1. Special 2 Point and 3 Point Distributions with Specified Mean and Variance.

We will have occasion to consider several special two point and 3 point distributions. First we consider the two point distribution on 0 and q with specified values of $(EX,\ EX^2) = (\mu, \mu_2')$ where $0 < \mu < 1$, $0 < q \leq 1$, and q is assigned probability θ. Then $\theta q = \mu$ and $\theta q^2 = \mu_2'$ and

$$(A1.1) \qquad q = \mu_2'/\mu = q_2 \text{ and } \theta = \mu^2/\mu_2' \ .$$

Incidentally, for this distribution

$$(A1.2a) \qquad \mu_{30}' = EX^3 = (\mu_2')^2/\mu$$

and

$$(A1.2b) \qquad \mu_{40}' = EX^4 = (\mu_2')^3/\mu^2$$

Also,

$$(A1.3a) \qquad \mu_{30} = E(X - \mu)^3 = \frac{\sigma^2}{\mu}(\sigma^2 - \mu^2)$$

and

$$(A1.3b) \qquad \mu_{40} = E(X - \mu)^4 = \frac{\sigma^2}{\mu^2}\frac{\sigma^6 + \mu^6}{\sigma^2 + \mu^2} = V_2(\mu, \sigma)$$

where $\sigma^2 = \mu_2 = \mu_2' - \mu^2$ is the variance of X.

Next we consider the two point distribution on 1 and q with specified values of $(EX, EX^2) = (\mu, \mu_2')$ where $0 \leq \mu \leq 1$ and $0 \leq q < 1$ and q is assigned probability θ. Then, consideration of the transformation $Y = 1 - X$, yields

$$(A1.4) \qquad q = (\mu - \mu_2')/(1 - \mu) = q_1 \text{ and } \theta = (1 - \mu)^2/[(1 - \mu)^2 + \sigma^2],$$

and

$$(A1.5a) \qquad \mu_{31} = E(X-\mu)^3 = \frac{-\sigma^2}{(1-\mu)}(\sigma^2 - (1-\mu)^2)$$

and

$$(A1.5b) \qquad \mu_{41} = E(X-\mu)^4 = V_2(1-\mu,\sigma)$$

A more general two point distribution with specified (μ, μ_2') will assign probability θ to $\mu + r(1-\theta)$ and $1-\theta$ to $\mu - r\theta$ for $r > 0$, and $0 < \theta < 1$. For this distribution r and θ are connected by

$$(A1.6) \qquad \sigma^2 = r^2\theta(1-\theta).$$

If we drop the restriction that $\mu - r\theta$ and $\mu + r(1-\theta)$ be in the interval [0,1] we have $\theta = 1/2$ when $r = 2\sigma$. Then we will have use for $dr/d\theta$ and $d(r\theta)/d\theta$. It is easy to see that $dr/d\theta = r(\theta - 1/2)$

$$(A1.7) \qquad d(r\theta)/d\theta = r(1 - \theta/2 + \theta^2) > 0.$$

and

$$(A1.8) \qquad d(r(1-\theta))/d\theta = r(-3/2 + 3\theta/2 - \theta^2) < 0.$$

Finally, consider the 3 point distribution which assigns probability ϕ to 1, θ to q and $1 - \theta - \phi$ to 0 where $0 < q < 1$. For the convex hull of (0,0), (1,1) and (q, q^2) to contain (μ, μ_2'), where $0 < \mu^2 \le \mu_2' \le \mu < 1$, we must have $q_1 \le q \le q_2$. Then it is easy to derive $q = (\mu_2' - \phi)/(\mu - \phi)$, and

$$(A1.9) \qquad \phi = \frac{\mu_2' - \mu q}{1-q} = \mu - \frac{\mu - \mu_2'}{1-q}$$

and

$$(A1.10) \qquad \theta = (\mu - \mu_2')/q(1-q)$$

A2. Bounds on $E(X-\mu)^4$ and $E(X-1/2)^4$.

We derive upper and lower bounds on $\mu_4 = E(X-\mu)^4$ and $E(X-1/2)^4$ subject to $P\{0 \le X \le 1\} = 1$ and specified values of μ and $\mu_2' = \sigma^2 + \mu^2$. The trivial cases where $\mu_2' = \mu^2$ and $\mu_2' = \mu$ are bypassed.

Since $\mu_4 = E(X-\mu)^4 = EX^4 - 4\mu EX^3 + 6\mu^2\mu_2' - 3\mu^4$, we may consider optimizing $E(X^4 - 4\mu X^3)$. The function $g(x) = x^4 - 4\mu x^3 - \lambda_1 x - \lambda_2 x^2$ has at most one local maximum and two local minima. It follows that the maximum of $g(x)$ over [0,1] can be attained on at most 3 points, of which only one can be an interior point. The minimum can be attained on at most two points.

To attack the maximization problem, we first apply the 3 point distribution of Appendix A1, and

$$\begin{aligned} E(X^4 - 4\mu X^3) &= \phi(1-4\mu) + \theta(q^4 - 4\mu q^3) \\ &= \mu(1-4\mu) - (\mu - \mu_2')[(1-4\mu)(1+q) + q^2] \end{aligned}$$

which attains its maximum at $q = 2\mu - 1/2$. But we are restricted to $q_1 \le q \le q_2$ by the argument in A1. Hence the restricted maximum of $E(X-\mu)^4$ occurs when

$$(A2.1) \qquad q = q_0 = \mathrm{med}(q_1, q_2, 2\mu - 1/2).$$

This implies that we have a 2 point distribution when $2\mu - 1/2 \le q_1$ or when $2\mu - 1/2 \ge q_2$. In particular, whenever $\mu \le 1/4$, we have a 2 point distribution. For $1/4 < \mu < 3/4$, we may have a 2 or 3 point distribution depending on the value of μ_2'.

To maximize $E(X-1/2)^4 = E(X^4 - 2X^3) + 3\mu_2'/2 - \mu/2 + 1/16$, we again apply the 3 point distribution to

$$\begin{aligned} E(X^4 - 2X^3) &= -\phi + \theta(q^4 - 2q^3) \\ &= -\mu + (\mu - \mu_2')(1 + q - q^2) \end{aligned}$$

which is maximized at $q = 1/2$. Thus the restricted maximum occurs when

$$(A2.2) \qquad q = q_0^* = \mathrm{med}(q_1, q_2, 1/2).$$

For the minimization problem for $E(X-\mu)^4$, consider first the 2 point distribution which minimizes $E(X-\mu)^4$ without the restriction of X to [0,1]. That is clearly the distribution which attaches probability $1/2$ to each of $\mu \pm \sigma$ and yields the value σ^4. If $0 \le \mu - \sigma < \mu + \sigma \le 1$, this distribution solves the restricted minimization problem.

Since $\sigma^2 = \mu_2' - \mu^2 \le \mu - \mu^2 \le 1/4$, $\mu - \sigma < 0$ implies $\mu < \sigma \le 1/2$. Similarly $\mu + \sigma > 1$ implies $\mu > 1/2$. If $\mu - \sigma < 0$, we refer to the two point distribution of A1 at $\mu - r\theta$ and $\mu + r(1-\theta)$. Then

$$v_1(\theta) = E(X-\mu)^4 = r^4\theta(1-\theta)[1 - 3\theta(1-\theta)] = \sigma^4\left[\frac{1}{\theta(1-\theta)} - 3\right].$$

Since $d(r\theta)/d\theta > 0$, it follows that as $\mu - r\theta$ increases from $\mu - \sigma$ where $\theta = 1/2$, θ decreases and $v_1(\theta)$ increases. Thus the minimum value of $v_1(\theta)$, subject to the

restrictions, occurs when $\mu - r\theta = 0$, i.e., for the 2 point distribution at 0 and q_2 and the minimum values of $E(X-\mu)^4$ is

$$(A2.3) \qquad V_2(\mu,\sigma) = \sigma^2(\mu^6+\sigma^6)/\mu^2(\mu^2+\sigma^2).$$

A symmetric argument for the case $\mu+\sigma > 1$ yields the two point distribution at q_1 and 1 with the minimum value of $V_2(1-\mu,\sigma)$.

The minimization problem for $E(X-1/2)^4$ is somewhat more complicated. We note that $E(X-1/2)^4 = E[(X-\mu)^4 + 4(\mu-1/2)(X-\mu)^3] + 6(\mu-1/2)^2\sigma^2 + (\mu-1/2)^4$ and that it suffices to minimize

$$\begin{aligned} v_2 &= E\{(X-\mu)^4 + 4(\mu-1/2)(X-\mu)^3\} \\ &= \sigma^2[(r^2-3\sigma^2) + (2\mu-1)r(1-2\theta)] \end{aligned}$$

subject to the restrictions.

Suppose that $0 < \mu < 1/2$. Since r takes on the same value for θ and $1-\theta$, it is clear that the minimizing value of θ will be less than $1/2$. Ignoring the restriction $0 \le \mu - r\theta < \mu + r(1-\theta) \le 1$, we have

$$(A2.4) \qquad \frac{dv_2}{d\theta} = r\sigma^2(2\theta-1)[r - r_0(\theta)]$$

where

$$(A2.5) \qquad r_0(\theta) = (2\mu-1)[(\theta-1/2) + (\theta-1/2)^{-1}].$$

Then, as θ goes from 0 to $1/2$, r decreases form ∞ to 2σ, θr increases from 0 to σ and $r_0(\theta)$ increases from $5(1/2-\mu) > 0$ to ∞. Thus, there is a unique value of θ_0 of θ for which $r_0(\theta_0) = r$ and $\theta_0 < 1/2$.

If $0 \le \mu - \theta_0 r_0(\theta_0) < \mu + (1-\theta_0)r_0(\theta_0) \le 1$, θ_0 and $r_0(\theta_0)$ define the minimizing two point distribution. If $\mu - \theta_0 r_0(\theta_0) < 0$, we see that as θ decreases from θ_0, $r\theta$ decreases and $\mu - r\theta$ increases. At the same time r increases and $r_0(\theta)$ decreases. Thus $dv_2/d\theta < 0$ and v_2 increases. Then the minimizing value of v_2 subject to the restrictions will occur when $\mu - r\theta = 0$, i.e., for the two point distribution at 0 and q_2.

If $0 < \mu - \theta_0 r_0(\theta_0) < 1 < \mu + (1-\theta_0)r_0(\theta_0)$, then as θ increases from θ_0 toward $1/2$, $r(1-\theta)$ decreases, r decreases, $r_0(\theta)$ increases, and hence v_2 increases. Then the minimizing value, as long as $\theta < 1/2$, occurs at the two point distribution at 1 and q_1. Since we showed above that the minimizing value of θ subject to the restrictions is less than $1/2$ we have demonstrated, for $\mu < 1/2$, that the minimizing distribution is one of three two point distributions depending on θ_0 and $r_0(\theta_0)$.

The case of $\mu > 1/2$ follows by symmetry. When $\mu = 1/2$ we have $\sigma < 1/2$ and the two point distribution on $\mu \pm \sigma$ is the minimizing distribution.

A3. Bounds on σ_0^2.

Since $\Delta_0 = \mathcal{E}(d^{(0)})$ and $\sigma_0^2 = \mathcal{E}\{d^{(0)}(1-d^{(0)})\}$ and $d_i^{(0)} = 2p_i(1-p_i)$ can vary from 0 to 1/2, the range of (Δ_0, σ_0^2) is the convex hull of $A = \{(x, x(1-x)) : 0 \le x \le 1/2\}$. That convex set is bounded by A and B, the straight line segment from (0,0) to (1/2, 1/4). Thus A and B determine the upper and lower bounds of σ_0^2 for given Δ_0 and indicate how they may be achieved.

The lower bound is attained when some of the p_i are 1/2 and all the others are 0 or 1. The upper bound is attained when all the $d_i^{(0)}$ are equal to Δ_0. Except when $\Delta_0 = 1/2$, there are 2 possible values of p_i which give the same value of $d_i^{(0)} = \Delta_0$.

The bounds can be refined if we are given Δ_0 and π. Then our problem becomes that of minimizing and maximizing $\mathcal{E}\{2p(1-p)[1-2p(1-p)]\}$ subject to specified values of $\mathcal{E}(p)$ and $\mathcal{E}(p^2)$. But that reduces to maximizing and minimizing $\mathcal{E}(p^4 - 2p^3)$ or $\mathcal{E}(p - 1/2)^4$. That problem is treated in A2.

A4. Mean and Variance of T.

Let $T_1 = (a_1 n_{10} + a_2 n_{01})/n - (a_1 + a_2)\hat{\pi}(1 - \hat{\pi})$. Section 3 deals with the case where $a_1 = a_2 = 1$. We represent the outcome for the i-th pair by $(X_i^{(1)}, X_i^{(2)})$ and by $(X_{00i}, X_{10i}, X_{01i}, X_{11i})$ where $X_{jki} = 1$ if the outcome is (j, k) and 0 otherwise. Then $X_i^{(1)} = X_{10i} + X_{11i}$, $X_i^{(2)} = X_{01i} + X_{11i}$,

$$\hat{\pi}_1 = n^{-1} \sum_{i=1}^{n} (X_{10i} + X_{11i}) = n^{-1} \sum_{i=1}^{n} X_i^{(1)} = (n_{10} + n_{11})/n$$

and

$$\hat{\pi}_2 = n^{-1} \sum_{i=1}^{n} (X_{01i} + X_{11i}) = n^{-1} \sum_{i=1}^{n} X_i^{(2)} = (n_{01} + n_{11})/n$$

Let $\epsilon_i = (X_i^{(1)} + X_i^{(2)})/2 - p_i$. Then

$$\hat{\pi}^2 = \left(\pi + \frac{1}{n}\sum_{i=1}^{n} \epsilon_i\right)^2 = \pi^2 + 2\pi n^{-1} \sum_{i=1}^{n} \epsilon_i + n^{-2}\left(\sum_{i=1}^{n} \epsilon_i\right)^2$$

and

$$T_1 = (a_1 + a_2)(\pi^2 - \pi) + n^{-1} \sum_{i=1}^{n} \left[a_1 X_{10i} + a_2 X_{01i} - (a_1 + a_2)(1 - 2\pi)\epsilon_i\right]$$

$$+ (a_1 + a_2) n^{-2} \left(\sum_{i=1}^{n} \epsilon_i\right)^2 \qquad (A4.1)$$

where the last term is $O_p(n^{-1})$ and has mean $(a_1+a_2)\mathcal{E}[p^{(1)}(1-p^{(1)})+p^{(2)}(1-p^{(2)})]/4n$ and the second term is $\mathcal{E}[a_1p^{(1)}(1-p^{(2)})+a_2p^{(2)}(1-p^{(1)})]+O_p(n^{-1/2})$ and has the same variance as $n^{-1}\sum S_i$ where

$$S_i = a_1X_{10i} + a_2X_{01i} - (a_1+a_2)(1-2\pi)(X_i^{(1)}+X_i^{(2)})/2$$

Thus

$$ET_1 = (a_1+a_2)\Big\{-\sigma_{12} + \frac{(\pi^{(1)}-\pi^{(2)})^2}{4} + \frac{a_1-a_2}{a_1+a_2}\frac{\pi^{(1)}-\pi^{(2)}}{2} + \mathcal{E}[p^{(1)}(1-p^{(1)})+p^{(2)}(1-p^{(2)})]/4n\Big\} \tag{A4.2}$$

where $\sigma_{12} = \mathcal{E}p^{(1)}p^{(2)} - \pi^{(1)}\pi^{(2)}$.

Now we rewrite

$$S_i = b_1X_i^{(1)} + b_2X_i^{(2)} - (a_1+a_2)X_{11i}$$

where $b_1 = (a_1+a_2)\pi + (a_1-a_2)/2$ and $b_2 = (a_1+a_2)\pi - (a_1-a_2)/2$, and we observe that $\mathrm{Cov}(X_i^{(1)}, X_{11i}) = p_i^{(1)}p_i^{(2)}(1-p_i^{(1)})$ and $\mathrm{Cov}(X_i^{(2)}, X_{11i}) = p_i^{(1)}p_i^{(2)}(1-p_i^{(2)})$. It follows that, neglecting the $O_p(n^{-1})$ term of T_1, $\mathrm{Var}T_1 \approx n^{-1}\tau_1^2$ where

$$\tau_1^2 = \mathcal{E}\Big\{b_1^2p^{(1)}(1-p^{(1)}) + b_2^2p^{(2)}(1-p^{(2)}) + (a_1+a_2)^2p^{(1)}p^{(2)}(1-p^{(1)}p^{(2)}) - 2(a_1+a_2)p^{(1)}p^{(2)}[b_1(1-p^{(1)})+b_2(1-p^{(2)})]\Big\} \tag{A4.3}$$

To derive Equation (3.2) we set $a_1 = a_2 = 1$ in (A4.2) and note that

$$\begin{aligned}\Delta - 2\pi(1-\pi) &= \mathcal{E}(p^{(1)}+p^{(2)}-2p^{(1)}p^{(2)}) - (\pi^{(1)}+\pi^{(2)})(1-\frac{\pi^{(1)}+\pi^{(2)}}{2}) \\ &= \pi^{(1)}+\pi^{(2)}-2\sigma_{12}-2\pi^{(1)}\pi^{(2)}-\pi^{(1)}-\pi^{(2)}+(\pi^{(1)}+\pi^{(2)})^2/2 \\ &= -2\sigma_{12}-(\pi^{(1)}-\pi^{(2)})^2/2\end{aligned}$$

To derive Equation (3.4), we set $a_1 = a_2 = 1$ in (A4.3). Then $b_1 = b_2 = 2\pi$. The matching of coefficients of $4\pi^2$, 4π and 1 in these two disparate forms involves showing that

$$\begin{aligned}p^{(1)}(1-p^{(1)})+p^{(2)}(1-p^{(2)}) &= d-(p^{(1)}-p^{(2)})^2 \\ -2p^{(1)}p^{(2)}(2-p^{(1)}-p^{(2)}) &= -2pd+(p^{(1)}-p^{(2)})^2 \\ 4p^{(1)}p^{(2)}(1-p^{(1)}p^{(2)}) &= 4pd-d^2-(p^{(1)}-p^{(2)})^2\end{aligned}$$

and this may be facilitated by noticing that $p^{(1)} + p^{(2)} = 2p$, $d = 2p - 2p^{(1)}p^{(2)}$ and $(p^{(1)} - p^{(2)})^2 = 4p^2 - 4p^{(1)}p^{(2)}$.

A5. Bounds on Δ and EXY.

First we consider upper and lower bounds on $E(XY)$ subject to the restrictions that $EX = \mu$, $EY = \nu$, and $0 \le X \le 1$ and $0 \le Y \le 1$ with probability one. We have $0 \le EXY \le EX = \mu$. Similarly $0 \le EXY \le \nu$. Moreover $E(1-X)(1-Y) \ge 0$ and hence $EXY \ge \mu + \nu - 1$. Thus

$$(A5.1) \qquad \max(0,\ EX + EY - 1) \le E(XY) \le \min(EX,\ EY).$$

Moreover these bounds are easily attained using 2 point distributions on adjacent edges of the unit square. For example if $EX < EY$, the distribution which assigns probability ν to $(\mu/\nu, 1)$ and $(1-\nu)$ to (0,0) yields $EXY = \mu$. If $\mu + \nu > 1$, the distribution which assigns probability $(1-\mu)$ to (0,1) and μ to $(1, (\mu+\nu-1)/\mu)$ yields $EXY = \mu + \nu - 1$.

To consider Δ we note that given $\pi^{(1)}$ and $\pi^{(2)}$ with $\pi^{(1)} \le \pi^{(2)}$, it follows that

$$0 \le \mathcal{E}p^{(1)}p^{(2)} \le \pi^{(1)} \qquad \text{if } 0 \le \pi \le 1/2$$

and

$$2\pi - 1 \le \mathcal{E}p^{(1)}p^{(2)} \le \pi^{(1)} \qquad \text{if } 1/2 \le \pi \le 1.$$

Since $\Delta = \mathcal{E}(p^{(1)} + p^{(2)} - 2p^{(1)}p^{(2)}) = 2(\pi - \mathcal{E}p^{(1)}p^{(2)})$ it follows

$$(A5.2a) \qquad 2\pi \ge \Delta \ge |\pi^{(1)} - \pi^{(2)}| \ge 0 \qquad \text{if } 0 \le \pi \le 1/2$$

and

$$(A5.2b) \qquad 2(1-\pi) \ge \Delta \ge |\pi^{(1)} - \pi^{(2)}| \ge 0 \qquad \text{if } 1/2 \le \pi \le 1$$

A6. Bounds on the Variance of $(p^{(1)} - \pi^{(1)})(p^{(2)} - \pi^{(2)})$.

The problem of establishing bounds on $\mathcal{E}\{(p^{(1)} - \pi^{(1)})^2(p^{(2)} - \pi^{(2)})^2\}$ subject to specified values of $\pi^{(1)}$, $\pi^{(2)}$ and σ_{12} may be rephrased as that of minimizing and maximizing EX^2Y^2 or the variance of XY subject to the restrictions $EX = EY = 0$, $EXY = c$, and $(X,Y) \in R = \{(x,y) : -\alpha \le x \le 1-\alpha,\ -\beta \le y \le 1-\beta\}$ where α and β, representing $\pi^{(1)}$ and $\pi^{(2)}$, are between 0 and 1. Applying A5 we see that

$$(A6.1) \qquad -c_2 = -\min(\alpha\beta,\ (1-\alpha)(1-\beta)) \le c \le \min(\alpha(1-\beta),\ \beta(1-\alpha)) = c_1$$

This result can also be derived using the Geometry of Moments by studying where $xy - \lambda_1 x - \lambda_2 y$ is minimized and maximized.

It is possible to demonstrate that the maximum is attained by a *three point distribution*, with two of the points on opposite vertices of R.

The minimization problem reduces to two cases. The easier case is that where $-c_1 \leq c \leq c_2$. In that case the two branches of the hyperbola $xy = c$ have points in R and it is possible to find a two point distribution for which $\mathrm{Var}(XY) = 0$.

For some values of α and β, it is possible to find values of c where $c_2 < c \leq c_1$. In those cases we can show that there is a solution involving at most 4 points, only one of which can be an interior point. The conjecture that there is a two point solution consisting of a vertex and another point (on the line from the vertex through the origin) is supported by numerical calculations.

A7. Bounds on $\tilde{\sigma}_0^2$.

Since $\tilde{\sigma}_0^2 = \mathcal{E}\{(4p - 16p^2 + 24p^3 - 12p^4)/3\}$, minimizing and maximizing $\tilde{\sigma}_0^2$ subject to specified values of $\mathcal{E}(p)$, $\mathcal{E}(p^2)$ and $\mathcal{E}(p^3)$ is equivalent to maximizing and minimizing EX^4 subject to the specified values of the first 3 moments and $0 \leq X \leq 1$. As in Appendix A2, maximizing EX^4 involves at most a 3 point distribution, only one point of which is an interior point of [0,1] and minimizing EX^4 involves at most a 2 point distribution. The three moments uniquely specify such distributions which may be calculated directly.

A8. Bounds on Δ_{30}.

We wish to minimize and maximize $E(X - X^3)$ subject to specified values of EX and EX^2 and $0 \leq X \leq 1$. This is equivalent to maximizing and minimizing EX^3 or $\mu_3 = E(X - \mu)^3$. The function $g(x) = x^3 + \lambda_1 x + \lambda_2 x^2$ has at most one local minimum and one local maximum. It follows that both the minimum and maximum of g on [0,1] can involve at most two points, only one of which can be an interior point of [0,1]. In the maximization case the boundary point has to be 1, and in the minimization case it is zero. Thus the minimum and maximum of μ_3 are μ_{30} and μ_{31} of Appendix A1. In particular the maximum of $E(X - X^3)$ is $\mu - (\mu_2')^2/\mu$.

REFERENCES

Cornfield, J. and Greenhouse, S. W. (1975). "The Estimation of a Common Odds Ratio," George Washington University Technical Report, 1-26.

Diaconis, P., (1977). "Finite Forms of de Finetti's Theorem on Exchangeability," *Synthese*, **36**, 271-281.

Fienberg, S. E. and Gilbert, J. P. (1970). "The Geometry of a Two by Two Contingency Table," *Journal of the American Statistics Association*, **65**, 695-701.

Gastwirth, J. L. and Greenhouse, S. W. (1987). "Estimating a Common Relative Risk: Application in Equal Employment," *Journal fo the American Statistical Association*, **82**, 38-45.

Karlin, S. and Shapley, L. L. (1953). "Geometry of Moment Spaces," *Memoirs of the American Mathematical Society*, **12**, 1-93.

Yu, K. F. (1993). "A Necessary and Sufficient Condition for the Strong Consistency of a Family of Estimators of the Common Odds Ratio. NIH Technical Report 1-17.