Selecting the Better Bernoulli Treatment Using a Matched Samples Design

BY AJIT C. TAMHANE

Northwestern University, Evanston, Illinois

[Received June 1978. Final revision February 1979]

SUMMARY

The problem of selecting the better Bernoulli treatment using a matched samples design is considered in the framework of the indifference-zone approach. A single-stage procedure is proposed and its properties are studied. Tables of sample sizes for implementing the proposed procedure are given. A comparison is made with an independent samples design and the associated Sobel–Huyett selection procedure.

Keywords: MATCHED SAMPLES DESIGN; INDEPENDENT SAMPLES DESIGN; BERNOULLI TREATMENTS; RANKING AND SELECTION; INDIFFERENCE-ZONE APPROACH

1. INTRODUCTION

In this paper we consider the problem of selecting the “better” of the two Bernoulli populations (i.e. the one having the larger success probability) when a matched samples design is used. The corresponding problem when the independent samples design is used has been considered by Sobel and Huyett (1957). It should be noted that although considerable literature exists on the problem of comparing matched proportions (see McNemar, 1947; Cochran, 1950; Bennet, 1967, 1968; Bhapkar, 1973; Bhapkar and Somes, 1977), mostly it deals with tests of homogeneity. However, in many practical situations the experimenter’s goal is to select the “best” treatment; a test of homogeneity does not provide the information which the experimenter truly seeks in such situations. This paper provides an appropriate formulation of the selection problem and gives a procedure for attaining this goal.

2. ASSUMPTIONS, NOTATION AND PROBLEM FORMULATION

Consider two treatments \( T_1 \) and \( T_2 \) and let \( \pi_{ij} \) denote the probability that a matched observation on \( T_1 \) and \( T_2 \) results in outcome \( i \) with \( T_1 \) and outcome \( j \) with \( T_2 \) (\( i, j = 0, 1 \)) where \( 1 \) denotes success and \( 0 \) denotes failure. We have \( \sum \sum \pi_{ij} = 1 \). We assume that the \( \pi_{ij} \) remain constant throughout the trial. Thus each matched observation can be thought of as a realization from a fixed multinomial distribution with four cells: \( (1,1), (1,0), (0,1) \) and \( (0,0) \); the corresponding probabilities are \( \pi_{11}, \pi_{10}, \pi_{01} \) and \( \pi_{00} \), respectively. Let \( p_1 = \pi_{11} + \pi_{10} \) and \( p_2 = \pi_{11} + \pi_{01} \) be the success probabilities of \( T_1 \) and \( T_2 \) respectively and let \( p_{[1]} \leq p_{[2]} \) denote the ordered values of the \( p_i \). We assume that the \( \pi_{ij} \) are unknown, but the experimenter is able to specify an upper limit \( \pi^*(0 < \pi^* \leq 1) \) on \( \pi_{10} + \pi_{01} = \mu \) (say). (In general, if matching is properly done and \( T_1 \) and \( T_2 \) are comparable to each other then \( \pi_i \), the probability of different outcomes on \( T_1 \) and \( T_2 \) with the same matched observation, will be small; see Section 5 for further discussion.) The experimenter’s goal is to select the treatment associated with \( p_{[2]} \); such a selection is referred to as a correct selection (CS) and the corresponding probability is denoted by PCS. The experimenter restricts consideration to procedures which guarantee the probability requirement:

\[
\text{PCS} \geq P^* \quad \text{whenever} \quad p_{[2]} - p_{[1]} = \delta \geq \delta^*, \quad \text{and} \quad \pi_{10} + \pi_{01} = \mu \leq \pi^*, \tag{2.1}
\]

where \( \{\pi^*, \delta^*, P^*\} \) are constants specified before experimentation starts; \( 0 < \pi^* \leq 1, 0 < \delta^* \leq P^* \) and \( \frac{1}{2} < P^* < 1 \).
3. THE PROPOSED PROCEDURE AND ITS PCS

We propose a natural selection procedure which takes \( n \) matched observations on \( T_1 \) and \( T_2 \). Let the outcomes be represented in a \( 2 \times 2 \) table:

\[
\begin{array}{cc}
T_2 & \text{Success} & \text{Failure} \\
T_1 & x_{11} & x_{10} \\
& x_{01} & x_{00} \\
\end{array}
\]

The decision rule is: select \( T_1 \) if \( x_{10} > x_{01} \); select \( T_2 \) if \( x_{10} < x_{01} \); and select \( T_1 \) or \( T_2 \) at random assigning equal probability to each one if \( x_{10} = x_{01} \). It might be noted that \( x_{11} \) and \( x_{00} \) play no role in the selection procedure. Our main problem is to find the minimum value of \( n \) which guarantees (2.1).

Without loss of generality assume that \( T_1 \) is the better treatment, i.e. \( p_1 > p_2 \) or equivalently \( \pi_{10} > \pi_{01} \). Then

\[
P_{\text{CS}} = P(X_{10} > X_{01}) + \frac{1}{2}P(X_{10} = X_{01})
\]

\[
= \sum_{x=0}^{n} [P(X_{10} > X_{01} | X_{10} = x) + \frac{1}{2}P(X_{10} = X_{01} | X_{10} + X_{01} = x)] \left( \begin{array}{c} n \\ x \end{array} \right) \pi^x (1-\pi)^{n-x}. \quad (3.1)
\]

Note that the quantity inside the square brackets in (3.1) is just \( P(Y > \frac{1}{2}x) + \frac{1}{2}P(Y = \frac{1}{2}x) \) if \( x \) is even, and \( P(Y > \frac{1}{2}(x+1)) \) if \( x \) is odd, where \( Y \) has a binomial distribution with parameters \( x \) and \( \lambda = \pi_{10}/\pi \). Denoting it by \( g(x, \lambda) \) we have

\[
g(x, \lambda) = \begin{cases} 
I_\lambda(x+1, \frac{1}{2})(x+1) & \text{for odd } x \geq 1, \\
I_\lambda(\frac{1}{2}x, \frac{1}{2}x) & \text{for even } x \geq 2, \\
\frac{1}{2} & \text{for } x = 0,
\end{cases} \quad (3.2)
\]

where

\[
I_p(a, b) = \frac{\Gamma(a+b)}{\Gamma(a) \Gamma(b)} \int_0^p u^{a-1}(1-u)^{b-1} \, du
\]

denotes the usual incomplete beta function.

To guarantee (2.1) it is necessary to find the infimum of the PCS over the region \( p_1 - p_2 = \pi_{10} - \pi_{01} \geq \delta^* \), \( \pi_{10} + \pi_{01} \leq \pi^* \); the minimum value of \( n \) which makes this infimum \( \geq P^* \) will be the desired sample size. To find the infimum, represent \( P_{\text{CS}} = E_{\pi}(g(X, \lambda)) \) where \( X \) is a binomial random variable with parameters \( n \) and \( \pi \) and \( E_{\pi} \) denotes the expectation evaluated at parameter value \( \pi \). Intuitively the infimum of the PCS over the specified region will occur when the \( p_i \) are as close as possible and the matching is as ineffective as possible, i.e. when \( p_1 - p_2 = \pi_{10} - \pi_{01} = \delta^* \) and \( \pi_{10} + \pi_{01} = \pi^* \) which is referred to as the least favourable configuration (LFC). However, note that it is not completely obvious that the PCS decreases with increasing \( \pi \) since this also corresponds to an increase in the number of "effective" observations \( x_{10} \) and \( x_{01} \). A formal proof of the LFC is thus needed and is given in the Appendix. Note that at the LFC, we have

\[
P_{\text{CS}_{\text{LFC}}} = E_{\pi^*}(g(X, \frac{1}{2} + \delta^*/2\pi^*))
\]

\[
= \sum_{x=0}^{n} g(x, \frac{1}{2} + \delta^*/2\pi^*) \binom{n}{x} (\pi^*)^x (1-\pi^*)^{n-x}. \quad (3.3)
\]

It is fairly straightforward to verify that the right-hand side of (3.3) is increasing in \( n \) and tends to 1 as \( n \) tends to \( \infty \). Thus any desired value of \( P^* \) can be attained by choosing \( n \) large enough. It should be noted that, when \( \delta^* = \pi^* \), (3.3) simplifies to \( 1 - \frac{1}{2}(1-\pi^*)^n \) and when \( \pi^* = 1 \), (3.3) simplifies to \( g(n, \frac{1}{2} + \frac{1}{2}\delta^*) \).
The values of $n$ which guarantee (2.1) were found using (3.3) for $n \leq 35$. For $n > 35$, the following normal approximation was used. Note that since $X_{10}, X_{01}$ are multinomial frequencies, for large $n$, $(X_{10} - X_{01})$ can be regarded as a normal random variable with mean $n(\pi_{10} - \pi_{01})$ and variance $n(\pi_{10} + \pi_{01} - (\pi_{10} - \pi_{01})^2)$. Therefore the PCS under the LFC can be written as $\text{PCS}_{\text{LFC}} \approx \Phi(\delta^* \sqrt{n/(\pi^* - \delta^*)^2})$ where $\Phi(.)$ denotes the standard normal distribution function. From this we obtain

$$
\begin{align*}
\hat{n} & \approx \frac{(\pi^* - \delta^*)^2 (\Phi^{-1}(P^*))^2}{\delta^*^2}. \\
\end{align*}
$$

This approximation is useful when $P^*$ is large and/or $\delta^*$ is small and/or $\pi^*$ is large. The values of $n$ obtained from (4.1) were rounded upwards. The calculations for $P^* = 0.90$ and 0.95 appear in Table 1; the values of $n$ for $P^* = 0.99$ are almost exactly double those for $P^* = 0.95$ and hence are not given here.

### Table 1

**Values of $n$**

<table>
<thead>
<tr>
<th>$P^* = 0.90$</th>
<th>0.05</th>
<th>0.10</th>
<th>0.15</th>
<th>0.20</th>
<th>0.25</th>
<th>0.30</th>
<th>0.35</th>
<th>0.40</th>
<th>0.45</th>
<th>0.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta^*$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\pi^*$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>65</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>130</td>
<td>32</td>
<td>14</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>196</td>
<td>48</td>
<td>21</td>
<td>12</td>
<td>7</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>262</td>
<td>65</td>
<td>29</td>
<td>16</td>
<td>10</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>327</td>
<td>81</td>
<td>35</td>
<td>20</td>
<td>13</td>
<td>9</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>0.6</td>
<td>393</td>
<td>97</td>
<td>43</td>
<td>24</td>
<td>15</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>0.7</td>
<td>459</td>
<td>114</td>
<td>50</td>
<td>28</td>
<td>18</td>
<td>12</td>
<td>9</td>
<td>7</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>0.8</td>
<td>524</td>
<td>130</td>
<td>57</td>
<td>32</td>
<td>21</td>
<td>14</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>0.9</td>
<td>590</td>
<td>147</td>
<td>65</td>
<td>36</td>
<td>23</td>
<td>16</td>
<td>12</td>
<td>9</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>1.0</td>
<td>656</td>
<td>163</td>
<td>72</td>
<td>40</td>
<td>25</td>
<td>17</td>
<td>13</td>
<td>9</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$P^* = 0.95$</th>
<th>0.05</th>
<th>0.10</th>
<th>0.15</th>
<th>0.20</th>
<th>0.25</th>
<th>0.30</th>
<th>0.35</th>
<th>0.40</th>
<th>0.45</th>
<th>0.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta^*$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\pi^*$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>106</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>214</td>
<td>52</td>
<td>21</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>322</td>
<td>79</td>
<td>34</td>
<td>18</td>
<td>11</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>431</td>
<td>106</td>
<td>46</td>
<td>25</td>
<td>16</td>
<td>10</td>
<td>7</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>539</td>
<td>133</td>
<td>58</td>
<td>32</td>
<td>20</td>
<td>14</td>
<td>10</td>
<td>7</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>0.6</td>
<td>647</td>
<td>160</td>
<td>70</td>
<td>38</td>
<td>25</td>
<td>17</td>
<td>12</td>
<td>9</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>0.7</td>
<td>755</td>
<td>187</td>
<td>82</td>
<td>45</td>
<td>29</td>
<td>20</td>
<td>14</td>
<td>11</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>0.8</td>
<td>864</td>
<td>214</td>
<td>94</td>
<td>52</td>
<td>34</td>
<td>23</td>
<td>17</td>
<td>13</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>0.9</td>
<td>972</td>
<td>241</td>
<td>106</td>
<td>59</td>
<td>37</td>
<td>26</td>
<td>19</td>
<td>14</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>1.0</td>
<td>1080</td>
<td>268</td>
<td>118</td>
<td>65</td>
<td>41</td>
<td>29</td>
<td>21</td>
<td>17</td>
<td>13</td>
<td>9</td>
</tr>
</tbody>
</table>

To check the accuracy of the normal approximation we computed exact and approximate $n$-values for $20 \leq n \leq 35$ and found that the approximate $n$ is always within $\pm 1$ of the exact $n$; the accuracy of the approximation improves with increasing $n$. Thus the normal approximation should be very good for $n > 35$. 

5. Comparison of Matched Samples Design with Independent Samples Design

In the case of independent samples design the goal of the experimenter is the same as before, namely to select the treatment having the larger success probability. However, since no $\pi_{ij}$ are present here, the probability requirement simply reads

$$\text{PCS} \geq P^* \quad \text{whenever } p_{(a)} - p_{(b)} = \delta \geq \delta^*,$$

(5.1)

where $\{\delta^*, P^*\}$ are constants specified before experimentation starts; $0 < \delta^* < 1$ and $\frac{1}{2} < P^* < 1$. Sobel and Huyett (1957) proposed a single-stage procedure which guarantees (5.1) and showed that the optimal sample size per population, in large samples, is

$$n \approx \frac{(1 - \delta^*)^2}{2 \delta^*}. \quad (5.2)$$

The relative efficiency (RE) of the matched samples design in terms of the ratio of sample sizes obtained from (4.1) and (5.2) is

$$\text{RE} \approx \frac{1 - \delta^*}{2(\pi^* - \delta^*)}. \quad (5.3)$$

Note that we assume the values of $\delta^*$ and $P^*$ specified by (2.1) and (5.1) are the same and that RE does not depend on $P^*$. Furthermore, RE > 1 if $\pi^* < \frac{1}{2}(1 + \delta^*)$. Also if $\pi^* = 1$ (i.e. $\pi = \pi_{10} + \pi_{01}$ is not constrained by any prior knowledge about its value) then RE = 0.5. Thus for $\pi^* > \frac{1}{2}(1 + \delta^*)$, the matched samples design is less efficient in the large sample case than the independent samples design. If the experimenter does not assume any prior knowledge concerning the value of $\pi$, then in the large sample case the matched samples design requires twice as many observations as the independent samples design to guarantee the same probability requirement.

These results give a quantitative measure of how effective the matching must be (in our notation how small $\pi$ must be) so that the matched samples design is more efficient than the independent samples design. The main conclusion to be drawn from this discussion is that, in the design of a matched samples experiment a high level of matching should be ensured. This can be achieved by choosing the matching variables so that they are highly correlated with the outcome variables. If the matching is ineffective then there can be considerable loss in efficiency relative to the independent samples design.

The results obtained here are in broad agreement with the similar work done for the testing problem in $2 \times 2$ tables by several authors, see, for example, Youkeles (1963), Worcester (1964) and Miettinen (1968). These authors also reach the conclusion that for testing the homogeneity of two proportions, matched samples design can be disadvantageous if the matching is not effective and the advantage is not substantial unless the matching is highly effective. For additional references and also for some practical aspects of matching see McKinlay (1977).

Acknowledgements

The author is thankful to three referees for making many useful suggestions for improvement of the presentation. This research was supported by NSF Grant No. ENG77–06112.

References

Corresponding where this function and Gupta each at for YOUKELES, SOBEL, WORCESTER, COCHRAN, McKINLAY, GUPTA, MIETTINEN, McNEMAR, (a1/a,)g(x, 30 even Ef,(g(X, all-or-none percentages. Tech. Theorem. With the algebraic manipulations it can be shown that, for x odd, g(x, λ*) − g(x − 1, λ*) = −(π/π) (d/dπ) g(x, λ*).

Thus, we find that the left-hand side of (A.3) is 0 for x odd. For x even ≥ 2 it follows from (3.2a) and (3.2b) that g(x, λ*) − g(x − 1, λ*) = 0. Furthermore, it can be easily verified that (d/dπ) g(x, λ*) < 0. Thus (A.3) is verified in all the cases. Because of the strict inequality for x even ≥ 2, it follows that Eπ{g(X, λ*)} is strictly decreasing in π.

Appendix

To prove the assertion regarding the LFC, we first keep π fixed and regard the PCS as a function of λ = π10/π. To show that, subject to π10 + π01 = π (fixed), the PCS is minimized at π10 − π01 = δ*, it suffices to show that g is a non-decreasing function of λ for each x. But this follows immediately from (3.2); in fact g is a strictly increasing function of λ > δ/2 for each x ≥ 1.

Our next task is to find the infimum over π ≤ π* of

\[ \inf_{\pi_{10} - \pi_{01} = \delta^*} \text{PCS} = E_\pi \{g(X, \lambda^*)\}, \] (A.1)

where λ* = 1/2(π + δ*/π). To show that this infimum occurs at π = π* we prove the following.

**Theorem.** \( E_\pi \{g(X, \lambda^*)\} \) is a decreasing function of π.

**Proof.** Denote the binomial probability function \( \binom{n}{x} p^x (1 - p)^{n-x} \) by \( b(x; n, p) \) and the corresponding distribution function by \( B(x; n, p) \). A “discrete analog” of Theorem 2.1 of Gupta and Panchapakesan (1972) shows that the condition to be verified for the monotonicity of \( E_\pi \{g(X, \lambda^*)\} \) relative to π is

\[ (\partial/\partial \pi) g(x, \lambda^*) b(x; n, \pi) - \{g(x, \lambda^*) - g(x - 1, \lambda^*)\} (\partial/\partial \pi) B(x - 1; n, \pi) \leq 0 \] (A.2)

for 1 ≤ x ≤ n; for x = 0 the left-hand side of (A.2) is 0. Substitute in (A.3),

\[ (\partial/\partial \pi) B(x - 1; n, \pi) = -(x/\pi) b(x; n, \pi) \] (A.3)

and find that the condition to be verified becomes

\[ (\partial/\partial \pi) g(x, \lambda^*) + (x/\pi) \{g(x, \lambda^*) - g(x - 1, \lambda^*)\} \leq 0. \] (A.3)

With some algebraic manipulations it can be shown that, for x odd,

\[ g(x, \lambda^*) - g(x - 1, \lambda^*) = -(\pi/\pi) (\partial/\partial \pi) g(x, \lambda^*). \]