IT SEEMS DESIRABLE TO IGNORE DATA. (U)

H. CHERNOFF
TR-ONR-24

N00014-75-C-0555
NL
WHEN IT SEEMS DESIRABLE TO IGNORE DATA

by

HERMAN CHERNOFF

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

TECHNICAL REPORT NO. ONR 24
SEPTEMBER 1982

PREPARED UNDER CONTRACT
N00014-75-C-0555 (NR-609-001)
FOR THE OFFICE OF NAVAL RESEARCH

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

STATISTICS CENTER
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

82 10 08 011
When it seems desirable to ignore data

by

AXEL CHERNOFF
Massachusetts Institute of Technology

ABSTRACT

An experiment designed to detect the relative motion of two astronomical
objects raised the problem of testing, against shift alternatives, the
null hypothesis $H_0$ that the energy distributions are equivalent. The relevant
data consist of independent Poisson counts $N_{ij}$ with mean $\lambda_{ij}$ where $\lambda_{ij}$
is the intensity of radiation from the $j$-th object. $N_{ij}$ is the

probability that a random photon from the $j$-th object has energy in a small

interval centered about $E_x$ and $E_y$ in the time duration allocated to the

count $N_{ij}$. The hypothesis $H_0$ states that $\lambda_{ij} = \lambda_{ij}$ for $i = 1, 2, ..., m$.

A natural test uses the statistic $e_{ij}^2 = \lambda_{ij} - \lambda_{ij}$ where the $\lambda_{ij}$ are

estimates of $\lambda_{ij}$. For intervals where the $\lambda_{ij}$ were anticipated to be

small, the requirement of these small $\lambda_{ij}$ values and hence those $\lambda_{ij}$ were

highly variable, comparatively, common sense suggests that the corresponding

$e_{ij}$ and $E_x$ and $E_y$ be omitted in the above statistic, a practice which may be

regarded as valid by statistical dogma. This issue and others raised by the

effects of small $\lambda_{ij}$ lead to the consideration of alternative test

statistics and their relative efficiencies as well as the design problem

of selecting $\lambda_{ij}$.

Key Words: Hypothesis testing, optimal design, Pitman efficiency, Poisson

Arch 1982 Subject Classifications: Primary 62F05; Secondary 62G10;

A satellite based experiment designed to detect a Doppler effect

measuring the relative motion of two neighboring extragalactic objects

can be an instrument which can count the number of photons received from

either one of these objects within a narrow energy band for a specified
duration of time (T). The total allocated time is divided into a

length intervals, each assigned to a distinct energy band and to one

of the astronomical objects. A straightforward analysis of the recorded

data raises some puzzling questions which will be addressed here.

The above analysis compares the estimated mean intensity in energy

levels of the two objects with its estimated standard deviation. If one

is a significant difference, however, since the experiment design

allocated relatively little time to energy bands in which the frequency

of photons were anticipated to be small, these are some bands with

small or zero counts. With the analysis used in such problems

intuitively, to ignore the data in some of these short time low count

intervals. Moreover, with the inclusion of one of these, the

existence of a single observed zero count to a count of one would have a

medium effect on the significance of the results.

Statistical dogma regards the ignoring of data as sinful, yet common

sense seems to urge us to commit this sin. The fact that the analysis

of an extremely expensive experiment involving hundreds of counts in

"the action is", should be greatly affected by the absence of presence of

a single count in our data. Given another opportunity to repeat this

experiment, how should the time intervals be allocated? Most of these

issues are addressed in this paper.
This situation does not qualify as a paradox, since procedures that do not make optimal use of the data may perform better by ignoring noisy data liable to be assigned too great a weight. A minor adjustment of the procedure can reduce the effect of the single count, an effect which is, to some extent, a small sample phenomenon. However, asymptotic analysis reveals that the main issues are not of a small sample nature and that the "natural" analysis of the data makes considerably less than full use of the available information.

The experiment is described in Section 2. The "natural" analysis and its difficulties, together with some mitigating modifications, are presented in Section 3. The discussion in Section 4 of a parametric one-sample version of the problem contributes some insights and bounds on efficiency. Finally, in Section 5 several alternative approaches are described and evaluated for the original two-sample problem.

In Section 6, the efficiencies of several approaches are calculated. The significance levels achieved using these methods are presented for four observed data sets. Finally, there is an Appendix where detailed derivations are presented for some of the less obvious results presented in the first five sections.

I wish to thank Joseph Gastwirth for the benefit of some illuminating discussion and the suggestion to use the Mann-Whitney approach discussed in Section 5.

2. Experiment and Notation

Let $X_i$ represent the number of photons observed during the $i$-th time period of length $T_i$ from a source of intensity $I$ in a narrow band of energy, centered about $t$, which contributes a proportion $p_i$ to the intensity. Thus we have independent observations $X_1, X_2, \ldots$ where

$$L(X_i) = \Theta(p_i T_i)$$

and where $L(x)$ represents the distribution of $X_i$, $\Theta(-1)$ represents the Poisson distribution with mean $1$, and $T_i = 1$. Let

$$p = \sum_{i=1}^{n} p_i$$

be the mean energy of photons from the source (neglecting the effect due to variation of energy within a band).

The parameters $I, p_i$, and $p$ may be estimated by

$$\hat{I} = \sum_{i=1}^{n} \frac{X_i}{p_i T_i}$$

where

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} T_i}$$

and

$$\hat{p}_i = \frac{X_i}{T_i}$$
Asymptotic analysis indicates that for large $n$, the distribution of $\bar{y}$ is approximately normal with mean $\mu$ and variance

$$\sigma^2 = \frac{1}{n} \sum (y_i - \bar{y})^2 = \frac{1}{n} \sum (y_i - \mu)^2$$

where

$$w_i = \frac{2}{p_1^{-1} + p_2^{-1}}$$

For the two-sample problem, we introduce $n_1$, $n_2$, $T_{11}$, $T_{12}$, $T_{21}$, $T_{22}$, $y_1$, $y_2$, $y_1^2$, $y_2^2$, $w_{11}$, and $w_{12}$. To test the hypothesis $H_0$ that there is no difference in energy distributions, we apply

$$Z = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $\bar{y}_1^2$ is derived from $\bar{y}_1$ by replacing $1$, $n_1$, and $y_1$ by $\frac{1}{n_1}$, $\frac{1}{n_1}$, and $\bar{y}_1$. Under the null hypothesis of no difference, $Z$ should be approximately normally distributed with mean 0 and variance 1.

### 3. Data and Analysis

The experiment consisted of four separate parts, each of which contributed to rejecting the null hypothesis $H_0$. We present in Table I, the data from one of the parts which illustrates the problem raised in the introduction. The analysis presented there is based on the use of only the data between $n_1 = 2580$ to $n_1 = 2660$.

The above analysis is the first performed by the author on this data set. Here three values of $f$ which seemed intuitively whole were 0.9, 0.0, and 0.6. It was compared with an independent previous analysis in the experiment which turned out to be equivalent in formula, but different in result. We had included the case $n_1 = 2560$. Then obtain

$$\bar{y}_1 = 2601.2, \quad \bar{y}_2 = 2624.5, \quad \bar{y}_3 = 2648.8$$

The substantial difference between the two results in the above to the discrepant weights given to $f = 0$. Indeed, these weights are

$$w_{11} = 0.16 \quad \text{and} \quad w_{12} = 0.08.\quad \text{Moreover, if the ratio of } w_{11} \text{ to } w_{12} \text{ is equal to the ratio of } n_1 \text{ to } n_2, \text{ there is another dramatic change with the resulting } Z = ' - f' \text{.}$$

To a substantial extent these effects can be regarded as null while effects and they can be reduced considerably by the simple device described below.

Under the null hypothesis $H_0$, $\bar{y}_1$ and $\bar{y}_2$ have a common value $\bar{y}_1$.

The high variability of the contributions to $\bar{y}_1^2$ and $\bar{y}_2^2$ of the estimates $\tilde{y}_{11}$ and $\tilde{y}_{12}$ would be reduced considerably if the estimates of variance $\tilde{y}_{11}^2$ and $\tilde{y}_{12}^2$ used the simple pooled estimate of $\tilde{y}_{11}$.
TABLE 1. Data and Estimates of Parameters and Weights
Based on Data for a Mening from 2580 to 2660
Electron Volts, and a Unit of \( T_{12} \) is 0.32 Seconds

<table>
<thead>
<tr>
<th>( T_{11} )</th>
<th>( T_{12} )</th>
<th>( T_{13} )</th>
<th>( T_{21} )</th>
<th>( T_{22} )</th>
<th>( T_{23} )</th>
<th>( W_{1} )</th>
<th>( W_{2} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2580</td>
<td>3000.4</td>
<td>3</td>
<td>3.4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2560</td>
<td>9017.0</td>
<td>8</td>
<td>1012.4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2580</td>
<td>9413.9</td>
<td>12</td>
<td>5831.1</td>
<td>4</td>
<td>0.28</td>
<td>0.38</td>
<td>0.26</td>
</tr>
<tr>
<td>2600</td>
<td>9261.8</td>
<td>13</td>
<td>9600.4</td>
<td>4</td>
<td>0.31</td>
<td>0.10</td>
<td>0.29</td>
</tr>
<tr>
<td>2620</td>
<td>9137.5</td>
<td>6</td>
<td>9518.5</td>
<td>13</td>
<td>0.15</td>
<td>0.32</td>
<td>0.14</td>
</tr>
<tr>
<td>2640</td>
<td>8731.2</td>
<td>7</td>
<td>9281.2</td>
<td>9</td>
<td>0.18</td>
<td>0.22</td>
<td>0.18</td>
</tr>
<tr>
<td>2660</td>
<td>9425.4</td>
<td>2</td>
<td>9082.3</td>
<td>8</td>
<td>0.68</td>
<td>0.20</td>
<td>0.13</td>
</tr>
<tr>
<td>2700</td>
<td>0.0</td>
<td>0</td>
<td>8402.2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2760</td>
<td>0.0</td>
<td>0</td>
<td>5122.2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2800</td>
<td>0.0</td>
<td>0</td>
<td>4303.8</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( \lambda = 0.00449 \)

\( \lambda = 2609.7 \)

\( \lambda = 4.43 \)

\( \lambda = 2.33 \)

Another possibility would be to use the normalized pooled estimates

\[ \hat{\beta}_{14} = \frac{\hat{\beta}_{11} + \hat{\beta}_{13}}{1 + \hat{\beta}_{12}} \]

where

\[ \hat{\beta}_{14} = \frac{\lambda_{11} + \lambda_{13}}{1 + \lambda_{12}} \]

In Table 2 we summarize the results where \( \alpha = 2540.10 \), i.e. \( \lambda = 4.43 \), is included and excluded for each of the approaches and for both \( \lambda_{1} \) and \( \lambda_{2} \) when \( \lambda = 0 \) is included.

The pooled prescriptions ameliorate considerably one of the problems we observed, but they fail to address the philosophical question of what right we have to omit data. It is clear from Equation (2.46) that it is wise to include \( \alpha \) values for which the \( \lambda_{1} \) are relatively small even if these \( \lambda_{1} \) are absolutely large. The problem is not merely a small
sample problem for which our prescription of pooling or any other ad hoc expedient would be adequate. On the other hand, as indicated in the introduction, it hardly qualifies to be called a paradox, since there never was any claim of optimality, asymptotic or otherwise, for the procedure used. What a suboptimal procedure is used it is not surprising to find that suppressing data, which may be given little or no weight by the procedure, may be desirable.

<table>
<thead>
<tr>
<th>TABLE 2.</th>
<th>$n_1$</th>
<th>$n_2$</th>
<th>$n_3$</th>
<th>$n_4$</th>
<th>$n_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Exposed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2560 excluded</td>
<td>2609.3</td>
<td>4.43</td>
<td>2624.4</td>
<td>4.68</td>
<td>2.33</td>
</tr>
<tr>
<td>2560 included</td>
<td>2601.2</td>
<td>4.65</td>
<td>2624.4</td>
<td>4.68</td>
<td>3.51</td>
</tr>
<tr>
<td>2560 included, $n_{24}$</td>
<td>2601.2</td>
<td>4.65</td>
<td>2612.4</td>
<td>10.52</td>
<td>0.97</td>
</tr>
<tr>
<td>2. Simple Pooled</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2560 excluded</td>
<td>2609.3</td>
<td>5.03</td>
<td>2624.4</td>
<td>5.12</td>
<td>2.09</td>
</tr>
<tr>
<td>2560 included</td>
<td>2601.2</td>
<td>5.25</td>
<td>2624.4</td>
<td>10.51</td>
<td>1.97</td>
</tr>
<tr>
<td>2560 included, $n_{24}$</td>
<td>2601.2</td>
<td>5.32</td>
<td>2612.4</td>
<td>10.73</td>
<td>0.97</td>
</tr>
<tr>
<td>3. Normalized Pooled</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2560 excluded</td>
<td>2609.3</td>
<td>5.16</td>
<td>2624.4</td>
<td>5.19</td>
<td>2.06</td>
</tr>
<tr>
<td>2560 included</td>
<td>2601.2</td>
<td>5.41</td>
<td>2624.4</td>
<td>12.43</td>
<td>1.71</td>
</tr>
<tr>
<td>2560 included, $n_{24}$</td>
<td>2601.2</td>
<td>5.39</td>
<td>2612.4</td>
<td>9.98</td>
<td>0.99</td>
</tr>
</tbody>
</table>
4. A One-Sample Version of the Problem

While we may not be dealing with a paradox, the theoretical problem still remains of how one should properly analyze the data so that the sin of ignoring data can be avoided. Of course, this does not vitiate the previous analysis. It provides an opportunity to compare the rather straightforward analysis with some sort of optimal alternative to see if there has been a serious loss of efficiency. A further consequence of the solution of the theoretical problem is that it casts light on the problem of the optimal choice of the time intervals $T_{ij}$.

Let us assume that the $p_i$ are of a prescribed functional form, uniquely determined except for a one dimensional parameter $\nu$ representing the velocity of the astronomical object studied. Then we have a problem involving 4 parameters $\nu$, $\gamma_1$, $\gamma_2$, and $\gamma_3$ where $\gamma_3$ determines the density $\pi(x) \sim p_i \pi_{k_j}(x)$, $j = 1, 2$. For asymptotic considerations we are concerned mainly with the estimate of $\gamma_3 - \nu$.

We present a more or less standard analysis deriving score functions and information matrices using the likelihood function. For simplicity let us first consider a one-sample version of our problem with parameters $\nu$ and data $T_1, P_1$ where $T_1$ is assumed to be large. The likelihood function is

$$L = \prod_{i=1}^{\nu} P_{k_j}(x_i^0) \prod_{i=1}^{\nu} P_{k_j}(x_i^0)$$

and

$$\log L = -\sum_{i=1}^{\nu} \log P_{k_j}(x_i^0) + \sum_{i=1}^{\nu} \log \left[ \prod_{j=1}^{\nu} P_{k_j}(x_i^0) \right] - \log R_{k_j}$$

The score function is $T = (T_1, T_2)'$ where

$$T_1 = \frac{2\log P_{k_j}(x_i^0)}{\gamma_3 - \nu} \left( \sum_{j=1}^{\nu} P_{k_j}(x_i^0) - \prod_{j=1}^{\nu} P_{k_j}(x_i^0) \right)$$

and the Fisher information matrix is

$$\frac{1}{2} \left( \begin{array}{cc} \frac{2\log P_{k_j}(x_i^0)}{(\gamma_3 - \nu)^2} & -\frac{2\log P_{k_j}(x_i^0)}{\gamma_3 - \nu} \\ -\frac{2\log P_{k_j}(x_i^0)}{\gamma_3 - \nu} & 2 \end{array} \right)$$

where

$$x = \left( \begin{array}{c} T_1P_{k_j}(x_i^0) \\ -\frac{2\log P_{k_j}(x_i^0)}{\gamma_3 - \nu} \end{array} \right)$$

$$\gamma = \left( \begin{array}{c} \frac{1}{2} \sum_{j=1}^{\nu} P_{k_j}(x_i^0) \log P_{k_j}(x_i^0) \\ -\frac{1}{2} \log P_{k_j}(x_i^0) \end{array} \right)$$

Here $E_{i,j}$ is used to represent a formal expectation with respect to the distribution determined by $p_j(x)$ and with the index $i$ at the conditioning random variable.
If the problem were of testing $H_0: \theta = \theta_0$ an asymptotically locally optimal procedure would use the test statistic $W_2 = \bar{X} - \theta_0$, where $\bar{X}$ is the maximum likelihood estimate of $\theta$. This has Pitman efficiency $\sqrt{C_{\theta, \theta_0}} \rho^{-1}$ where $C_{\theta, \theta_0} = \rho^{-2}$ and efficiency is measured per unit time allocated to the experiment. Pitman efficiency is defined in (1). An asymptotically equivalent test statistic is the second component of $W^2(\theta, \theta_0)$ where $\hat{\theta}$ is the maximum likelihood estimate of $\theta$. Note that this statistic is

$$W_2 = \frac{1}{\gamma} W_1(\hat{\theta}, \theta_0) + W_2(\hat{\theta}, \theta_0)$$

where $\gamma$, $\hat{\theta}$, and $\gamma$ are derived from $x$, $\hat{\theta}$, and $\gamma$ by substituting $(\hat{x}, \hat{\theta})$ for $(x, \theta)$, Moreover if $\hat{\theta}$ were replaced by $\hat{\theta}^*$ the maximum likelihood estimate of $\hat{\theta}$ when $\theta$ is assumed to be $\theta_0$, the resulting statistic $W_2$ is also asymptotically locally equivalent to $W_2$ and $\hat{\theta} - \theta_0$. Incidentally, it is easy to see that $\hat{\theta}^* = \hat{\theta}$ if $\hat{\theta}$ is the solution of $\gamma(\hat{\theta}, \theta_0) = 0$.

If $\theta$ were known, one could use $W_2 = W_2(\theta_0)$ with the improved efficiency $\sqrt{C_{\theta_0, \theta_0}} \rho^{-1}$. Since $W_2$ is almost a linear function of the $(\theta_0 - \theta, \hat{\theta})$ and $W_2$ is a linear function of the $(\theta_0 - \theta, \hat{\theta})$, they are much less sensitive to the variance enhancing effects of small $\gamma$, than are expressions which are linear in $\gamma(\hat{\theta} - \theta_0)$. This suggests how information involving $W_2$ for relatively low values of $\gamma$ may be incorporated without degrading performance. Thus the issue of ignoring data has been confronted in this one-sample problem.

The conclusions in the one-sample problem lead to analogous results handling the two-sample problem. In brief it seems that shifts in the distribution may be more effectively detected by estimating other linear functions of the $Y_i$ than $\hat{\theta}(x) = \hat{\theta}_0^{x_1.}$.
5. The Two-Sample Problem

We return to the two-sample problem. If the $p_j(8)$ were known functions of $\theta$, then our problem would be simple. It would suffice to use

$$W_4 = \bar{T}_1 - \bar{T}_2$$

as a test statistic where $\bar{T}_j$ determines the $p_j = p_j(\theta)$. Alternatively one could use $W_3$, the difference of two statistics for each sample. Each of these test statistics has optimal efficacy (per unit time) of

$$E_0 = \left\{ \left( T_1 + T_2 \right) \left[ \frac{1}{2} - \frac{1}{2} \right] - \frac{1}{2} \left[ \frac{1}{2} - \frac{1}{2} \right] \right\}^{-1}$$

where the subscripts refer to the sample. Incidentally one may attach the design problem of allocating the $T_1$ and $T_2$ to optimize the above efficacy.

The above resolution is inadequate for the original version of the problem where it was not assumed that the $p_j(8)$ were well known functions of the parameter $\theta$. Nevertheless it does suggest the possibility of using a test statistic of the form

$$W_6 = \left[ \frac{(T_1 - X_1)}{\sigma_1} \right] \left[ \frac{(T_2 - X_2)}{\sigma_2} \right]$$

An alternative suggestion is to use a statistic of the form

$$W_6^* = \left[ \frac{T_1 - X_1}{\sigma_1} \right] \left[ \frac{T_2 - X_2}{\sigma_2} \right]$$

since the term which $u_1$ multiplies in (5.3) seems to allow for an estimate of $p_1 - p_1$. However these two statistics are effectively equivalent. More precisely

$$W_6 = \frac{T_1 - X_1}{\sigma_1} - \frac{T_2 - X_2}{\sigma_2}$$

where

$$T_1 = \frac{1}{2} T_{11} + \frac{1}{2} T_{12}$$

$$T_2 = \frac{1}{2} T_{11} + \frac{1}{2} T_{22}$$

$$W_6^* = \frac{T_1 - X_1}{\sigma_1} - \frac{T_2 - X_2}{\sigma_2}$$

A second alternative which does not appear to be effective equivalent to $W_6$ is to use statistic of the form

$$W_7 = \frac{T_{11} - T_{12}}{\sigma_1}$$

which would be effectively equivalent to one of the form

$$W_7^* = \frac{T_{11} - T_{12}}{\sigma_1}$$

Note that $W_7^*$ is a minor generalization of $W_7 = K_2$. 


Still another approach consists of initiating a nonparametric test statistic such as that of Mann-Whitney. Thus we could use an estimate of $F_{ij} = E_j - P_i(1 - P_j)$ where $E_j$ and $P_i$ are the energies of independent random processes from the two astronomical objects. For this we use

$$w_1 = \left( \frac{1}{P_i} \right)_{i=1} \left( \frac{1}{P_j} \right)_{j=1} \left( \frac{1}{P_j} \right)_{i=1}$$

It isn't exactly clear what the natural generalization of nonparametric test statistics is in this context. One proposal is to use

$$w_2 = \left( \frac{1}{P_i} \right)_{i=1} \left( \frac{1}{P_j} \right)_{j=1} \left( \frac{1}{P_j} \right)_{i=1}$$

where $A = [a_{ij}]$ is a skew symmetric matrix. The statistic $w_2$ is a special case of $w_2$ where $a_{ij} = \pm 1$ if $i = j$ and $a_{ij} = -1$ if $i > j$. For testing $P_1 = P_2$, we refer to an equivalent statistic $w_{ij}$ as

$$w_{ij} = \frac{1}{P_i} \left( \frac{1}{P_j} \right)_{i=1} \left( \frac{1}{P_j} \right)_{i=1}$$

To evaluate the effectiveness of these statistics we calculate the means and variances of their asymptotic (normal) distributions. These are, for $w_i$

$$w_i = \frac{2}{P_i} \left( \frac{1}{P_i} \right)_{i=1} \left( \frac{1}{P_j} \right)_{j=1}$$

for $w_{ij}$

$$w_{ij} = \frac{1}{P_i} \left( \frac{1}{P_j} \right)_{i=1} \left( \frac{1}{P_j} \right)_{i=1}$$

where $b_i$ and $b_j$ are described below.

Let us assume that $p_{ij} = P_i(1 - P_j)$ where $P_i$ is a translation parameter close to $P_j$. Then $w_i$ and $w_{ij}$ can be approximated by expressions of the form $w_i = \frac{1}{2} t^2 b_i$ and $w_{ij} = \frac{1}{2} t^2 b_j$ by $p_{ij}$. For the statistic $w_i$ and $w_{ij}$, the efficiency for testing $H_0 : \theta = \theta_0$ is given by $\left( \frac{1}{P_i} \right)_{i=1} \left( \frac{1}{P_j} \right)_{j=1}$. The optimal choice of $\theta$ would be $P_i$ and the corresponding statistic $t^2 = b_i$.

If $b_i$ were nonsingular, however $b_j$ is singular since there is a vector $c$ with both $b_j = 0$ and $J^t c = 0$. Then $(b_{ij})^{\frac{1}{2}}$ and $P_i(1 - P_j)$ proceed.
the optimal $q$ and efficacy. The statistic $s_{q}^{*}$ must be treated differently. There $s_{q}^{*}$ is approximated by an expression of the form
\[(s_{q}^{*}-h_{q})=(p_{1}(q_{1})d_{1}^{2}+p_{2}(q_{2})d_{2}^{2})^{-1}\]
and $s_{q}^{*}$ by $h_{q}=(p_{1}(q_{1})d_{1}^{2}+p_{2}(q_{2})d_{2}^{2})^{-1}$, where

\[(5.14)\]

\[s_{q}^{*} = \frac{1}{2} p_{1}(q_{1})d_{1}^{2} + \frac{1}{2} p_{2}(q_{2})d_{2}^{2}\]

The efficacy of $s_{q}^{*}$ is $s_{q}^{*} = \sum_{i=1}^{n} [p_{i}(q_{i})d_{i}^{2}]^{-1} s_{i}$ and $s_{q}^{*}$ can be maximized easily by the method of Lagrange multipliers when we note that the condition that $s_{i}^{*}$ be skew symmetric is equivalent to the condition on $g$ that $g'(q_{i}) = 0$. Then an optimal choice of $s_{i}$ is such that

\[(5.15)\]

\[s_{i}^{*} = -q_{i}^{1/2} y_{i} + \gamma_{i}\]

where $q_{i}$ and $y_{i}$ are Lagrange multipliers. More details on these results follow.

For $s_{q}$ we have $s_{q}^{*}$ and $s_{q}^{*}$ given by

\[(5.16)\]

\[s_{q}^{*} = \frac{\log p_{1}(q_{1})}{d_{1}^{2} + d_{2}^{2}} - y_{i}\]

where

\[(5.17)\]

\[s_{q}^{*} = [(y_{1})^{1/2} + (y_{2})^{1/2}]^{-1}\]

and

\[(5.18)\]

\[s_{q}^{*} = \frac{1}{d_{1}^{2} + d_{2}^{2}} - y_{i}\]

where

\[(5.19)\]

\[s_{q}^{*} = d_{1}^{2} - d_{2}^{2} - y_{i}^{2} - d_{1}^{2} - d_{2}^{2}\]

with $d_{1} = p_{1}(q_{1})d_{1}^{2}$, $d_{2} = \text{diag}(d_{1})$ and $s_{q}^{*} = \text{diag}(d_{2})d_{2}^{-1}$, note that $s_{q}^{*} = 0$ and $s_{q}^{*} = 0$ where $y_{1} = y_{1}^{-1}$.

For $s_{q}$ we have $s_{q}^{*}$ and $s_{q}^{*}$ given by

\[(5.20)\]

\[s_{q}^{*} = \frac{1}{d_{1}^{2} + d_{2}^{2}} - y_{i}^{-1} - d_{1}^{2} - d_{2}^{2}\]

and

\[(5.21)\]

\[s_{q}^{*} = d_{1}^{2} - d_{2}^{2} - y_{i}^{-1} - d_{1}^{2} - d_{2}^{2}\]

with $d_{1} = \text{diag}(d_{1})$, $d_{2} = p_{1}(q_{1})d_{1}^{2}$ and $s_{q}^{*} = \text{diag}(d_{2})$ in $s_{q}^{*}$ as above. Here we note that $s_{q}^{*} = 0$ and $s_{q}^{*} = 0$ for $y_{1} = 1$.

For $s_{q}$ we have $s_{q}^{*} = \log p_{1}(q_{1})/d_{1}^{2}$ and $y_{i}$ as above. Applying the method of Lagrange multipliers, we derive
Thus, while deviations from \( H_0 \) may be detected by observing, for large \( N \), that \( H_{11} \neq H_{12} \), the extent of the translation cannot exactly be determined without assuming a functional form for the density of the energy. Nevertheless, it is possible to check that the failure of \( H_0 \) corresponds to a translation of the energy distribution.

This problem is relatively minor if there are many narrow energy bands and if there is little density in those bands near the boundary of the region to which substantial time \( T_{11} \) has been allocated. However, even \( [H_{12} - H_{11}] \) is likely to be a poor estimate of the shift. If the bands are wide, \( H_0 \) is not adequately represented by the average energy in the band. If there is substantial density at the boundaries of the region studied, the probabilities being estimated are really weighted by a function given the energy, to the prescribed region. Then \( -\ln (P_1 / P_2) \) is a shift in the energy distribution or an uncorrelated estimate of a translation shift in the conditional distribution.

If these difficulties are negligible, we could estimate the translation parameters by maximizing each of these \( P_i \) from \( T_{11} \) and \( T_{12} \) without \( H_0 \) would be required for one of the resulting statistics \( H_{11}^*, H_{12}^* \) to have zero mean.

Another question that has not been carefully examined is the relevance of Pitman efficiency, a measure designed for dealing with small shifts. If the shift is substantial compared to the standard deviations of the components of the density corresponding to the modes of the trial bands studied.
Table of Efficiencies and Significance

The experiment yielded four data sets, the fourth of which appears in Table 1. The others are presented in Table 3. These data suggest models for $p(x)$ of the form

$$P_1(x) = \int_{-\infty}^{x} f(x - \theta)dx$$

where

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

and $\theta$ is the standard normal density. These models were used for our theoretical evaluations. This method of choosing models, loosely fitting the data, may bias our results to yield apparent efficiencies somewhat larger than deserved for methods whose coefficients are "tuned" to optimize with respect to these "fitted" models. The parameter of the models are presented in Table 4. The efficiencies of various methods are presented in Table 5. These methods are applied to the data sets, and in Table 6, the corresponding $z$ values are presented. Since the significance level or $P$ value are given approximately by

$$P = \Phi(z)$$

where $\Phi$ is the standard normal c.d.f., these levels are not presented explicitly. More details follow.
The efficiencies presented in Table 5 are for the statistic testing $H_0: H_{12} = H_{13}$; $1 < i < m$ against local shift alternatives to $H_0$; the model applies with $q_1 = p_2$ (approximately 0).

1. $E_{06}$ is the efficiency of the optimal parametric test.
2. $E_{06}$, $E_{07}$, and $E_{08}$ are the efficiencies of $W_6$, $W_7$, and $W_8$ using the optimal coefficients.
3. $E_{06}$ is the efficiency of the "natural" test based on the statistic $W = D_6^2 (H_{12} - H_{13})$ or the equivalent $W_6^2 = D_6^2 (H_{12} - H_{13})$.
4. $E_{09}$ is the efficiency of $W_9$, the non-Bliss version of $W_9$.

For each model, the data set is relevant only in the values of $T_{14}, T_{15}$ (design parameters) used, and the observed counts $Y_{ij}$ are not relevant.

For each model and data set combination, we consider several alternative subsets of the available counts for inclusion in the analysis. Thus if we take $T_{14}, T_{15}$, $T = T_{14} + T_{15} + T_{16}$ and $T_1$ is of interest as well as $T$. Computations show that $E_{06} = E_{07}$, $E_{09}$ seems to be true in general. This is not very surprising and should not be too difficult to establish.

In particular the equality of $E_{06}$ and $E_{08}$ was anticipated.

Clearly we should have $E_{06} < E_{08}$. Possibly because of the loose fitting of the models to the observed data, $E_{08}$ is very close to $E_{06}$. On the other hand $E_{08}$ is sometimes poor and sometimes very sensitive to the choice of data to be included in the analysis. Generally $E_{08}$ does better than $E_{06}$ and is less sensitive to the choice of data to be included.

In Table 6 we present the values of $Z = W_{10}^*$ corresponding to various estimators applied to the actual data sets. For each estimator two $Z$ values...
Table 5. Efficacies of Tests for $R_{01} / R_{02}$; $1 < m$  

$$E = 10^{-4}, I = 10^{-3}$$  

<table>
<thead>
<tr>
<th>$d$</th>
<th>$n$</th>
<th>$T_0$</th>
<th>$T_1$</th>
<th>$T_2$</th>
<th>$T_3$</th>
<th>$T_4$</th>
<th>$T_{50}$</th>
<th>$T_{51}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.67</td>
<td>1.33</td>
<td>1.37</td>
<td>1.31</td>
<td>1.30</td>
<td>1.25</td>
<td>1.20</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>2.78</td>
<td>1.33</td>
<td>1.65</td>
<td>1.35</td>
<td>1.34</td>
<td>1.25</td>
<td>1.20</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>2.70</td>
<td>2.12</td>
<td>2.30</td>
<td>0.50</td>
<td>0.75</td>
<td>1.20</td>
<td>1.20</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>6.90</td>
<td>6.53</td>
<td>6.77</td>
<td>0.81</td>
<td>5.51</td>
<td>5.13</td>
<td>5.13</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>6.65</td>
<td>5.22</td>
<td>6.04</td>
<td>0.96</td>
<td>5.31</td>
<td>4.87</td>
<td>4.87</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>6.50</td>
<td>6.00</td>
<td>6.39</td>
<td>0.94</td>
<td>5.36</td>
<td>4.87</td>
<td>4.87</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td>7.31</td>
<td>7.16</td>
<td>7.08</td>
<td>0.74</td>
<td>5.76</td>
<td>5.17</td>
<td>5.17</td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>2.10</td>
<td>2.09</td>
<td>1.99</td>
<td>1.39</td>
<td>2.38</td>
<td>3.31</td>
<td>3.31</td>
</tr>
<tr>
<td>9</td>
<td>14</td>
<td>2.33</td>
<td>2.23</td>
<td>1.75</td>
<td>1.64</td>
<td>2.44</td>
<td>3.87</td>
<td>3.87</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>1.69</td>
<td>1.48</td>
<td>1.01</td>
<td>1.13</td>
<td>2.11</td>
<td>2.13</td>
<td>2.13</td>
</tr>
<tr>
<td>11</td>
<td>16</td>
<td>1.97</td>
<td>1.67</td>
<td>1.12</td>
<td>1.41</td>
<td>0.95</td>
<td>1.30</td>
<td>1.05</td>
</tr>
<tr>
<td>12</td>
<td>17</td>
<td>2.52</td>
<td>1.91</td>
<td>0.01</td>
<td>0.21</td>
<td>0.39</td>
<td>1.67</td>
<td>0.01</td>
</tr>
<tr>
<td>13</td>
<td>18</td>
<td>1.73</td>
<td>1.71</td>
<td>1.69</td>
<td>1.68</td>
<td>0.71</td>
<td>1.71</td>
<td>1.20</td>
</tr>
<tr>
<td>14</td>
<td>19</td>
<td>1.42</td>
<td>1.87</td>
<td>1.84</td>
<td>1.84</td>
<td>0.85</td>
<td>1.30</td>
<td>1.50</td>
</tr>
</tbody>
</table>

$d = $ data set number  
$m = $ model number

Table 6. Significance Levels and Estimated Shifts

$I$ values corresponding to various test statistics and data sets. $P = I - k$  
where $I$ is the standard normal c.d.f. Data set 5 is data set 4 with $z_{22}$ replaced by $I$.

$I$ values are estimated shifts in energy distribution normalized by dividing by energy.

<table>
<thead>
<tr>
<th>$d$</th>
<th>$l_1$</th>
<th>$l_2$</th>
<th>$l_3$</th>
<th>$l_4^0$</th>
<th>$l_5^2$</th>
<th>$l_6$</th>
<th>$l_7$</th>
<th>$l_8$</th>
<th>$l_9$</th>
<th>$l_{10}$</th>
<th>$l_{11}$</th>
<th>$l_{12}$</th>
<th>$l_{13}$</th>
<th>$l_{14}$</th>
<th>$l_{15}$</th>
<th>$l_{16}$</th>
<th>$l_{17}$</th>
<th>$l_{18}$</th>
<th>$l_{19}$</th>
<th>$l_{20}$</th>
<th>$l_{21}$</th>
<th>$l_{22}$</th>
<th>$l_{23}$</th>
<th>$l_{24}$</th>
<th>$l_{25}$</th>
<th>$l_{26}$</th>
<th>$l_{27}$</th>
<th>$l_{28}$</th>
<th>$l_{29}$</th>
<th>$l_{30}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

$d = $ data set number  
$m = $ model number
were calculated corresponding to the use, respectively of \( \hat{\Pi}_{12} \) and \( \hat{\Pi}_{12} \) in the estimation of the standard deviation of the \( W \) statistic. Since these \( 2 \) values were almost always very close to one another, only those corresponding to \( \hat{\Pi}_{12} \) are presented. For \( \hat{\Pi}_{12} \), three versions of \( 2 \) were calculated corresponding to the related statistics \( \hat{\Pi}_{03} \), \( \hat{\Pi}_{13} \) and \( \hat{\Pi}_{12} \). The first, \( \hat{\Pi}_{03} \) uses the coefficients \( \hat{k}_{12} \) selected so that

\[
\left[ k_{12} \right]_{03} = \hat{k}_{12}
\]

where \( \hat{k}_{12} \) are the optimal values of \( k_{12} \) according to the appropriate model. Since the value of the \( \hat{k}_{12} \) is not unique, they were selected so that

\[
\hat{k}_{12} = 1 \times 10^{-6} \times \hat{q}_{12} \quad \cdots .
\]

The statistic \( \hat{\Pi}_{13} \) uses similar coefficients subject to

\[
\left[ k_{12} \right]_{13} = \hat{k}_{12}
\]

and \( \hat{\Pi}_{12} \) is derived similarly. We present \( 2 \) values for \( \hat{\Pi}_{03} \) and \( \hat{\Pi}_{13} \), i.e. \( \hat{\Pi}_{03} \) and \( \hat{\Pi}_{13} \). Note that Table 6 involves the model and \( \hat{\Pi}_{13} \) and \( \hat{\Pi}_{12} \) since these determine the coefficients of \( \hat{\Pi}_{13} \) and the \( \hat{\Pi}_{12} \) statistic.

Finally, estimates of the mean shift \( \hat{\theta}_1 = \hat{\theta}_2 \) were calculated. To be more specific, \( \hat{\theta}_2 = \hat{\theta}_1 = \hat{w} \), estimates \( \hat{\theta}_1 = \hat{w} = \hat{w}_1 \), \( \hat{w}_1 = \hat{w}_1 \), the shift in the distribution of energy. The relative velocity of the two astronomical objects is approximately proportional to \( \left( w_1 - w_2 \right) / w_0 \), which is estimated by \( \hat{w}_2 = \hat{w}_1 / \hat{w}_0 \). To estimate this same parameter using \( \hat{w}_1 \) the following coarse technique was used. Compute

\[
\hat{w}_1 = \left[ \begin{array}{c} -1 \hat{w}_1 \\
1 \hat{w}_1 \\
2 \hat{w}_1
\end{array} \right]
\]

and let the corresponding values be estimated by the normalized shift:

\[
\hat{w}_1 = \hat{w}_1 / \hat{w}_0 = \hat{w}_1 / \hat{w}_0 \left( \hat{w}_1 - \hat{w}_0 \right)
\]

The care must be noted that there is a small difference between the \( 2 \) values in Tables 2 and 6 corresponding to data set 6. This difference is due to the fact that in Table 2, Equation 20 was applied with \( \hat{w}_2 = \hat{w}_1 \), substituting for \( \hat{w}_2 \) when \( j = 1 \) and \( \hat{w}_1 = \hat{w}_1 \) when \( j = 2 \). On the other hand, for the calculation in Table 6, \( \hat{w}_2 = \hat{w}_1 / \hat{w}_0 \) was substituted for \( \hat{w}_2 \) in both cases.
BIBLIOGRAPHY


A. APPENDIX

A.1 Equation (2.6)

\[ \hat{\nu} = \left( \frac{1}{n} \sum_{i=1}^{n} x_i \right) - \frac{\hat{\nu}}{\hat{\sigma}} \]

Using the δ method, the (asymptotic) variance of \( \hat{\nu} \) (as \( \hat{\nu} \rightarrow \nu \)) is the variance of

\[ T = \frac{X(n) - E(X)}{E(X)^{1/2}} \]

Since \( E(X) = \lambda \) and \( E(X)^{1/2} = 1 \),

\[ T = 1 - \hat{\nu} (x_i - \mu) \]

with variance

\[ \hat{\nu}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \]
A.2 Efficacy of \( \hat{e} = \hat{e}_0, \hat{W}_2, \hat{W}_2 \)

(a) \( \hat{e} = \hat{e}_0 \). For \( \hat{e} \) close to \( \hat{e}_0 \)

\[
\Delta \hat{e}(\hat{e}_0, \hat{W}_2) = -30(72\hat{I}(1, \hat{e}_0))
\]

Hence the efficacy of \( \hat{e} = \hat{e}_0 \) (per unit time) is

\[
(1272\hat{I}(1, \hat{e}_0))^{\frac{1}{2}}
\]

(dividing the square of the derivative of the mean by the variance)

(b) \( \hat{W}_2 \).

Let \( \hat{e} = (\hat{e}, \hat{e}_0) \) with \( \hat{e} - \hat{e}_0 = O(T^{-4}) \)

\[
\hat{W} = \hat{I}^{-1}(\hat{e}, \hat{e}_0)P(\hat{e}, \hat{e}_0)
\]

where \( \hat{I} = \hat{I}_1(\hat{e}_0, \hat{W}_2) \hat{I}_2(\hat{e}_0, \hat{W}_2) \) and \( \hat{I}_1 \) and \( \hat{I}_2 \) are \( P(1) \). and

\[ \hat{I} = \hat{I}_1(\hat{e}, \hat{W}_2) \hat{I}_2(\hat{e}, \hat{W}_2) \]

Then

\[
\hat{I}^{-1}(\hat{e}_0, \hat{W}_2) = \hat{I}^{-1}(\hat{e}_0, \hat{W}_2) + \hat{I}_2(\hat{e}_0, \hat{W}_2) + \hat{I}_1(\hat{e}, \hat{W}_2)
\]

\[
\hat{W}(\hat{e}_0, \hat{W}_2) = \hat{I}(\hat{e}, \hat{W}_2) \left[ \hat{I}^{-1}(\hat{e}_0, \hat{W}_2) \right] + \hat{I}_1(\hat{e}, \hat{W}_2) + \hat{I}_2(\hat{e}, \hat{W}_2)
\]

\[
\hat{I} = \hat{I}_1(\hat{e}_0, \hat{W}_2) \hat{I}_2(\hat{e}_0, \hat{W}_2)
\]

where the arguments of \( \hat{I} \) can be taken to be \( (\hat{e}, \hat{e}_0) \). It follows that

the efficacy of \( \hat{W}_2 \) is \((1272(1, \hat{e}_0))^{\frac{1}{2}}\)

(c) \( \hat{W}_2 \).

We shall show that \( \hat{n} - 1 = O(T^{-4}) \) after which the answers need
for \( \hat{W}_2 \) can be reproduced without change. The key reason is that
\( \hat{n} - 1 \) appears in the first but not in the second component of \( \hat{n} \) in (A2.4).

The fact that \( \hat{n} - 1 = O(T^{-4}) \) follows directly from the expression
for \( \hat{n} \). A more general approach, not confined to this particular
problem, follows

\[
0 = \hat{I}_1(\hat{e}_0, \hat{W}_2) = \hat{I}_1(\hat{e}_0, \hat{W}_2) + \hat{I}_1(\hat{e}, \hat{W}_2) + \hat{I}_2(\hat{e}, \hat{W}_2)
\]

\[
\hat{I}_1 = \hat{I}_1(\hat{e}_0, \hat{W}_2) + \hat{I}_1(\hat{e}, \hat{W}_2)
\]

\[
\hat{I}_2 = \hat{I}_2(\hat{e}_0, \hat{W}_2) + \hat{I}_2(\hat{e}, \hat{W}_2)
\]

\[
\hat{I}_1 = \hat{I}_1(\hat{e}_0, \hat{W}_2)
\]

\[
\hat{I}_2 = \hat{I}_2(\hat{e}_0, \hat{W}_2)
\]

\[
\hat{I}_1 = \hat{I}_1(\hat{e}_0, \hat{W}_2) + \hat{I}_1(\hat{e}, \hat{W}_2)
\]

\[
\hat{I}_2 = \hat{I}_2(\hat{e}_0, \hat{W}_2) + \hat{I}_2(\hat{e}, \hat{W}_2)
\]
Using the facts that \(Y(1,1) = H(1)\), \(e - e = 0\) \(\frac{2F(1,1)}{2\pi} = H(1)\)

and \(2J_2(1,1)\) which is of the order of magnitude of \(T\), the desired result follows.

A.2 The asymptotic distributions of \(W_0, W_0^*,\) and \(W_0^*\)

(a) The statistics \(W_0, W_0^*\) and \(W_0^*\) are clearly asymptotically normal.

We shall calculate the means and variances of the approximating distributions by applying the method which consists of expanding in terms of \(X_1 = \frac{1}{2}X_{11}T_{11}\) and \(X_2 = \frac{1}{2}X_{12}T_{12}\). We use \(u_0\) to represent statistics evaluated at \((p_{11}, p_{12}) = (\frac{1}{2}X_{11}T_{11}, \frac{1}{2}X_{12}T_{12})^2\).

Note also that

\[W_0^* = \frac{1}{2} u_0 - \frac{X_{12}T_{12}}{T_2}\]

where

\[u_1 = \frac{1}{2}X_{11}T_{11}^2 (p_{12} - p_{11})\]

\[u_2 = \frac{1}{2}X_{12}T_{12}^2 (p_{12} - p_{11})\]

Then the mean for \(W_0^*\) is

\[\mu_0 = u_0^* = \frac{p_{12} - p_{11}}{\left(X_{12}T_{12}^{-1} + X_{12}^{-1}\right)}\]

and the variance is

\[\sigma_0^2 = \frac{1}{2} \text{Var}(X_{12}) \left(\frac{W_0^*}{X_{12}}\right)^2 = \frac{\sigma_0^2}{X_{12}} + \frac{\sigma_0^2}{X_{12}^2}\]
\[
\mathbf{x}_{11} = \mathbf{A}_{11} \mathbf{x}_{11} + \mathbf{B}_{11} u_{11}
\]

\[
\mathbf{x}_{12} = \mathbf{A}_{12} \mathbf{x}_{12} + \mathbf{B}_{12} u_{12}
\]

\[
\mathbf{z}_{i} = \mathbf{H}_{i} \mathbf{x}_{i} + \mathbf{w}_{i}
\]

where

\[
\mathbf{A}_{ij} \in \mathbb{R}^{n_i \times n_j}, \quad \mathbf{B}_{ij} \in \mathbb{R}^{n_i \times 1}, \quad \mathbf{H}_{i} \in \mathbb{R}^{1 \times n_i}
\]

and

\[
\mathbf{w}_{i} \sim \mathcal{N}(0, \Sigma_{w})
\]

Let \( \mathbf{A}_{ij} = \mathbf{I}_{n_i} \), \( \mathbf{B}_{ij} = \mathbf{0} \), and \( \mathbf{H}_{i} = \mathbf{1} \). Then

\[
\mathbf{x}^* = \mathbf{x}_{0} + \mathbf{z}^* - (\mathbf{z}^*)^T \mathbf{z}^* - \mathbf{z}^* \mathbf{z}^* - \mathbf{z}^* \mathbf{A}_{0}^{-1} \mathbf{z}^*
\]

where

\[
\mathbf{x}_{0} = \text{diag}(\mathbf{d})
\]

\[
\mathbf{z}^* = \begin{bmatrix} \mathbf{z}_{11}^* \\ \mathbf{z}_{12}^* \\ \vdots \\ \mathbf{z}_{n_i}^* \end{bmatrix}
\]

\[
\mathbf{A}_{0}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} \\ \mathbf{A}_{12}^{-1} \\ \vdots \\ \mathbf{A}_{n_i}^{-1} \end{bmatrix}
\]

and \( \mathbf{A}_{ij} = \mathbf{I}_{n_i} \). Then

\[
\mathbf{A}_{ij} = \mathbf{I}_{n_i}
\]

\[
\mathbf{B}_{ij} = \mathbf{0}
\]

and

\[
\mathbf{H}_{i} = \mathbf{1}
\]

so

\[
\mathbf{x}^* = \mathbf{x}_{0} + \mathbf{z}^* - (\mathbf{z}^*)^T \mathbf{z}^* - \mathbf{z}^* \mathbf{A}_{0}^{-1} \mathbf{z}^*
\]

where

\[
\mathbf{x}_{0} = \text{diag}(\mathbf{d})
\]

\[
\mathbf{z}^* = \begin{bmatrix} \mathbf{z}_{11}^* \\ \mathbf{z}_{12}^* \\ \vdots \\ \mathbf{z}_{n_i}^* \end{bmatrix}
\]

\[
\mathbf{A}_{0}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} \\ \mathbf{A}_{12}^{-1} \\ \vdots \\ \mathbf{A}_{n_i}^{-1} \end{bmatrix}
\]
\[
\begin{align*}
\mathbf{v}_0 & = \mathbf{b}_1' \mathbf{N}_2 \\
\text{where } A & \text{ is skew symmetric.}
\end{align*}
\]
\[
\begin{align*}
\mathbf{v}_0^* & = [1 \ 0] \mathbf{y}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \\
\mathbf{v}_0^* & = [1 \ 0] \mathbf{y}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \\
\end{align*}
\]
\[
\begin{align*}
\begin{bmatrix} \mathbf{v}_0^* \\ \mathbf{w}_0^* \\ \mathbf{y}_0^* \\ \mathbf{u}_0^* \\ \mathbf{v}_0^* \end{bmatrix} & = \begin{bmatrix} \mathbf{v}_0^* \\ \mathbf{w}_0^* \\ \mathbf{y}_0^* \\ \mathbf{u}_0^* \\ \mathbf{v}_0^* \end{bmatrix} \\
\end{align*}
\]
and
\[
\begin{align*}
\mathbf{v}_0^* & = \mathbf{b}_1' \mathbf{N}_2 \\
\text{and then}
\end{align*}
\[
\begin{align*}
\mathbf{v}_0^{*2} & = \mathbf{v}_0^* \mathbf{v}_0^* = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}
\end{align*}
\]
Note that if \( g^{(1)} \) and \( g^{(2)} \) are close to \( g \),

\[
\frac{1}{n} \sum_{i=1}^{n} \left[ g_i^{(1)} - g_i^{(2)} \right] A_i\frac{\partial p_i}{\partial s}.
\]

If \( g^{(1)} = g(p_{i}, q_{i}, r, s) \) close to \( g_{o} \) it follows that:

\[
\frac{\partial g_i}{\partial s} \frac{\partial p_i}{\partial s} \Bigg|_{s=s_{o}} = -\log p_i(s_{o})/\partial s.
\]

A.4. Two maximization problems.

(a) The maximum value of \( (g')^2/(g' \theta) \) is obtained by minimizing \( g' \theta \)

subject to \( g' = 0 \). Applying the method of Lagrange multipliers

\[
\nabla g = \lambda \\
\lambda = -g^{-1}x
\]

and

\[
\frac{(g')^2}{(g' \theta)} = \lambda x^{-1}
\]

provided \( \lambda \) is nonsingular. However, in our application \( \theta = 0 \) and \( g' = 0 \).

If we replace \( \lambda \) by \( \lambda = \theta = 0 \) and \( g \) by \( g + h \) for \( h = 0 \) with \( g' = 0 \), then

\[
\frac{(g + h')^2}{(g' \theta)} = \lambda x^{-1}
\]

and

\[
\frac{(g + h')^2}{(g' \theta)} = \lambda x^{-1} + \lambda x^{-1} h.
\]

It is clear that the minimizing \( g + h \) for the new problem coincides with

that of the original problem. This is the minimizing value of \( g \) in \( f^{(1)} \), and

the maximum value of \( (g')^2/(g' \theta) \) is \( f^{(2)} \), provided that \( f \) is nonsingular.

(b) To minimize \( |g|^{2} \) subject to \( |g|^{2} = \theta \) where \( g = g(p, q) \) w.r.t. the

condition that \( g_{ij} \) be skew symmetric. The condition of skew symmetry implies

\[
\frac{\partial g_{ij}}{\partial q_{ij}} = 0.
\]

Moreover, given any vector for which \( g' = 0 \), there is a skew symmetric \( g \) for which \( g = 0 \). Hence, for our minimization problem, skew

symmetry is equivalent to \( g' = 0 \), and we may apply the method of Lagrange

multipliers to minimize with respect to \( g \).
\( p_{e_1}^2 e_2 = 2 p_{e_1}^2 + 2 p_{e_2}^2 \)

\( q_2 = 2 p_{e_1}^2 + 2 p_{e_2}^2 \)

The restrictions yield:

\( v_1 [p_{e_1}^2 e_2] + v_2 [p_{e_2}^2 e_2] = 0 \)

\( v_1 [p_{e_1}^2 e_2] + v_2 [p_{e_2}^2 e_2] = 2 \)

\( v_2 = 2 \times \left( \frac{E(e_1)}{E(e_2)} \right) \times \left( \frac{E(e_2)}{E(e_1)} \right) \)

\( v_1 = \frac{E(e_1) \times (E(e_2)^2) - (E(e_1) e_2)^2}{E(e_1) \times (E(e_2)^2) - (E(e_1) e_2)^2} \)

\( v_2 = \frac{-E_2(e_2) \times (E(e_2)^2) - (E(e_1) e_2)^2}{E(e_1) \times (E(e_2)^2) - (E(e_1) e_2)^2} \)

The maximum value of \( (p_{e_1} p_{e_2})^2 / p_{e_1}^2 e_2 \) is \( E^2 / E_{e_1} \) which is

\( E(e_1) = \frac{(E(e_2)^2)}{E(e_1)} + E_{p_{e_1}^2 e_2} - \frac{(E_{p_{e_1}^2 e_2})^2}{(E_{p_{e_1}^2 e_2})} \)
**Title:** When It Seems Desirable To Ignore Data

**Author:** Herman Chernoff

**Performing Organization:**
- **Name:** Statistics Center, Massachusetts Institute of Technology
- **Address:** Cambridge, MA 02139

**Contract or Grant Number:** N00014-75-C-0555

**Report Date:** Sept. 24, 1982

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

**Keywords:** Hypothesis testing, optimal design, Pitman efficacy, Poisson

**Abstract:** See reverse side.
ABSTRACT

An experiment designed to detect the relative motion of two astronomical objects raised the problem of testing, against shift alternatives, the hypothesis $H_0$ that two energy distributions are equivalent. The relevant data consist of independent Poisson counts $X_{ij}$ with means $\lambda_{ij}$, where $X_j$ is the intensity of radiation from the $j$-th object, $p_{ij}$ is the probability that a random photon from the $j$-th object has energy in a small interval centered about $\epsilon_i$ and $T_{ij}$ is the time duration allocated to the count $X_{ij}$. The hypothesis $H_0$ implies that $p_{1j} = p_{2j}$ for $i = 1, 2, \ldots, m$.

A natural test uses the statistic $\sum (\hat{p}_{1j} - \hat{p}_{2j})$ where the $p_{ij}$ are estimates of $p_{ij}$. For intervals where the $p_{ij}$ were anticipated to be small, the experimenter chose small $T_{ij}$ values and hence those $p_{ij}$ were highly variable. Consequently, common sense suggests that the corresponding $X_j$ and $X_{ij}$ be omitted in the above statistic, a practice which may be regarded as sinful by statistical dogma. This issue and others raised by the effects of small $T_{ij}$ lead to the consideration of alternative test statistics and their relative efficiencies as well as the design problem of selecting $T_{ij}$. 
END
DATE
FILMED
11-82
DTI