Loss Due to Missing Data in Efficiency of a Locally Optimal Test for Homogeneity with Respect to Very Rare Events*,**

by

Prem S. Puri

Department of Statistics
Division of Mathematical Sciences

Mimeograph Series No. 234
August, 1970

---

*This investigation was supported in part by research grants GM 10525-07 at University of California, Berkeley, from the National Institutes of Health, Public Health Service, and NONR 1100(26) from the Office of Naval Research at Purdue University.

The author is greatly indebted to Professor J. Neyman for inspiring him to investigate the above problem and for some very rewarding discussions.

**To appear in the proceedings of the National Academy of Science (1970).

Loss Due to Missing Data in Efficiency of a Locally Optimal Test for Homogeneity with Respect to Very Rare Events*

by

Prem S. Puri
Purdue University

Abstract. With reference to observations of supernovae in galaxies, a locally optimal test for the hypothesis of homogeneity of the observational units with respect to occurrence of supernovae (treated as a very rare event) is obtained for the case where the data are available only for those galaxies where at least one supernova is observed. It is shown that the loss in the asymptotic efficiency of the test due to lack of reporting of galaxies with no supernovae is very heavy and in fact is infinite for the case of very rare events.

1. Introduction. This paper may be regarded as a sequel to an earlier paper by Bühler at el[2], where the authors consider the notion of very rare events, arising in experimental situations such as the observation of supernovae in galaxies. The specific problem that was dealt with there was to construct a locally optimal test of the hypothesis of homogeneity of observational units with respect to occurrence of very rare events. Let N denote the number of units of observations for each of which the number of occurrences of an event E is observed during a fixed period of observation. Let M denote the total number of events E observed. An event was defined[2] as 'rare' if for a large N, the quotient M/N is of the order of unity. A 'very rare' event was defined to be the one where this

*This investigation was supported in part by research grants GM 10525-07 at University of California, Berkeley, from the National Institutes of Health, Public Health Service, and NONR 1100(26) from the Office of Naval Research at Purdue University.
The author is greatly indebted to Professor J. Neyman for inspiring him to investigate the above problem and for some very rewarding discussions.

quotient is smaller in order than unity. More specifically M^2/N instead was assumed to be of the order of unity. In particular, the problem was considered there in reference to occurrence of supernovae in galaxies and mathematically it can be described as follows:

Let U_{ij} denote the jth observational unit for the ith observatory with $j=1,2,\cdots,N_i$ and $i=1,2,\cdots,s$. Each of the units U_{ij}, $j=1,2,\cdots,N_i$ is observed for the same period t_i of time. The number of events observed in U_{ij} over time t_i is treated as a random variable and is denoted by μ_{ij}. It is assumed that all these random variables are mutually independent, with μ_{ij} distributed as a Poisson variable with expectation $\lambda_{ij} t_i$, where λ_{ij} is the expected number of events E in U_{ij} per unit time. The hypothesis H_0 of homogeneity of the observational units with respect to the very rare event E, is then equivalent to the statement that all the λ_{ij} have the same unspecified value λ. The nature of the hetrogeneity of λ's used as hypothesis alternative to H_0, was exhibited by assuming that each observational unit takes a particular value of a random variable Λ as the value corresponding to its λ_{ij}. Thus the hypothesis H_0 of homogeneity would correspond to the case where $\Lambda=\lambda$, a constant, with probability one. With a slight modification of what was assumed before[2], we assume here that

$$\Lambda = \lambda_0 \exp[X \xi^{1/2}] \quad , \tag{1}$$

where $\lambda_0 > 0$, $\xi \geq 0$ are two unspecified parameters; X is treated as a random variable taking values in an interval $[c,d]$ where c and d are two arbitrary but fixed finite constants. Let $F(x)$ be the distribution function of X . About $F(x)$ all we shall assume that

$$\int x dF(x) = \int x^3 dF(x) = 0 \ ; \ \int x^2 dF(x) = \int dF(x) = 1 \quad . \tag{2}$$

Thus, with $F(x)$ otherwise arbitrary, the random variable μ_{ij} observed during controlled time t_i in a randomly selected jth galaxy at the ith observatory would be distributed as a mixture of Poisson variables and its distribution may be described by

$$\Pr(\mu_{ij}=m_{ij}) = \frac{(\lambda_0 t_i)^{m_{ij}}}{m_{ij}!} \int \exp\{-\lambda_0 t_i \exp[x \xi^{1/2}] + m_{ij} x \xi^{1/2}\} dF(x), \tag{3}$$

$j=1,2,\cdots,N_i$; $i=1,2,\cdots,s$. While λ_0 is a nuisance parameter, the hypothesis of homogeneity by virtue of (1) is described by H_0: $\xi = 0$. Thus given (3), the problem reduces to developing a test of H_0: $\xi = 0$ against the alternative H_1: $\xi > 0$. Earlier[2] a locally optimal test was constructed for this problem. Also, an asymptotic Poisson approximation to the distribution of the test statistic was realized for large N under the assumptions that $N_i/N \to \nu_i$ as $N \to \infty$ with $0 < \nu_i < 1$, and that $\lambda_0 = O(N^{-1/2})$; the latter assumption being due to the very rarity of the events in question.

We wish to point out here that it is a common practice among the astronomers that while looking for occurrences of supernovae, they tend to report only those cases where they succeed in observing at least one supernova; the galaxies showing no supernova are left unreported. This being the case, our random variables μ_{ij} are observed as truncated only to the positive integer values, so that one needs to construct tests appropriate for the situation, given that $\mu_{ij} \geq 1$. Thus, subject to this we have modified (3) as given by

$$\Pr(\mu_{ij} = m_{ij} \mid \mu_{ij} \geq 1)$$

$$= \frac{(\lambda_0 t_i)^{m_{ij}} \int \exp\{-\lambda_0 t_i \exp[x\,\xi^{1/2}] + m_{ij}\, x\, \xi^{1/2}\} dF(x)}{m_{ij}!\,\{1 - \int \exp[-\lambda_0 t_i \exp(x\,\xi^{1/2})] dF(x)\}} \qquad (4)$$

for $m_{ij} = 1, 2, \cdots$; $j = 1, 2, \cdots, n_i$; $i = 1, 2, \cdots, s$. Here we have used letter n_i instead of N_i to differentiate between the observations truncated as in (4) and the nontruncated ones as in (3). Let $n = \Sigma_{i=1}^{s} n_i$. Essentially for each i , n_i is the random variable representing the number of those galaxies observed with at least one supernova out of a total of N_i galaxies. As such, it follows from the law of large numbers that

$$n_i/N_i \xrightarrow{P} 1 - \int \exp\{-\lambda_0 t_i \exp[x\,\xi^{1/2}]\}dF(x) \ , \qquad (5)$$

as $N_i \to \infty$. We shall need this result later.

In section 2, we apply the theory of optimal $C(\alpha)$-tests developed by Neyman[3], and as generalized later by Bartoo and Puri[1], for constructing a locally optimal test of the hypothesis H_0 based on the observations subject to (4), treating λ_0 as a fixed nuisance parameter. Then, keeping with the spirit of very rare events, the distribution of the relevant test statistic is approximated for large n to a Poisson distribution with λ_0 tending to zero in an appropriate manner as $n \to \infty$.

A similar optimal $C(\alpha)$-test of H_0 based on the non-truncated observations subject to (3), is already given elsewhere[1]. In section 3, performances of the two tests, one based on truncated data subject to (4) and the other based on complete data subject to (3), are compared. It is shown that the loss in asymptotic efficiency of the test, because of being unable to use the part of data of galaxies with no supernova, is very heavy and in fact is infinite for the case of very rare events.

2. <u>An Optimal $C(\alpha)$-test based on truncated Data</u>. In this section, even though for a given N_i , n_i is a random variable,

nevertheless we shall consider n_i as nonrandom and fixed. This is reasonable in view of (5) since we shall concentrate ourselves to the asymptotic case with $N_i \to \infty$, for each i. The theory for constructing an optimal test of the hypothesis $H_0: \xi = 0$, based on the observations μ_{ij} which are independently but not identically distributed, is dealt with by Bartoo and Puri[1]. Following their notation, the construction of the test involves first with finding the partial derivatives $\varphi_{ij(\xi)}$ and $\varphi_{ij(\lambda_0)}$ of the log-likelihood for the observation μ_{ij}. Here both the derivatives are evaluated at $H_0: \xi = 0$. Next, one obtains a regression constant a such that the variance of

$$\psi(\underset{\sim}{\mu}) = \Sigma_{i=1}^{s}\Sigma_{j=1}^{n_i}[\varphi_{ij(\xi)}(\mu_{ij};\lambda_0) - a\ \varphi_{ij(\lambda_0)}(\mu_{ij};\lambda_0)] \quad , \tag{6}$$

obtained under H_0, is minimized. Using this constant, let

$$f_{ij}^{*} = \varphi_{ij(\xi)}(\mu_{ij};\lambda_0) - a\ \varphi_{ij(\lambda_0)}(\mu_{ij};\lambda_0) \quad . \tag{7}$$

In the present case, using (4) we have

$$\begin{cases} \varphi_{ij(\xi)} = \frac{1}{2}\{\mu_{ij}(\mu_{ij}-1) - \dfrac{(\lambda_0 t_i)^2}{1-\exp(-\lambda_0 t_i)}\} + (\frac{1}{2} - \lambda_0 t_i)\lambda_0 \varphi_{ij(\lambda_0)} \\ \\ \varphi_{ij(\lambda_0)} = \dfrac{1}{\lambda_0}\{\mu_{ij} - \dfrac{\lambda_0 t_i}{1-\exp(-\lambda_0 t_i)}\} \quad . \end{cases} \tag{8}$$

The desired constant a is given by

$$a = \lambda_0(1+A)/2 \quad , \tag{9}$$

where

$$A = \frac{\sum_{i=1}^{s} n_i[(\lambda_0 t_i)^3 \exp(-\lambda_0 t_i)]/[1 - \exp(-\lambda_0 t_i)]^2}{\sum_{i=1}^{s} n_i[\lambda_0 t_i - \lambda_0 t_i(1+\lambda_0 t_i)\exp(-\lambda_0 t_i)]/[1 - \exp(-\lambda_0 t_i)]^2} \quad . \tag{10}$$

Also, the variance of (6) under H_0 is given by

$$\Sigma_{i=1}^{s}\Sigma_{j=1}^{n_i}\mathrm{Var} f_{ij}^{*} = \frac{1}{4}\left\{\Sigma_{i=1}^{s} n_i\left[\frac{2(\lambda_0 t_i)^2 - (\lambda_0 t_i)^2[2+(\lambda_0 t_i)^2]\exp(-\lambda_0 t_i)}{(1-\exp(-\lambda_0 t_i))^2}\right]\right.$$

$$\left. - \frac{\left[\sum_{i=1}^{s} n_i \dfrac{(\lambda_0 t_i)^3 \exp(-\lambda_0 t_i)}{(1-\exp(-\lambda_0 t_i))^2}\right]^2}{\sum_{i=1}^{s} n_i\left[\dfrac{\lambda_0 t_i - \lambda_0 t_i(1+\lambda_0 t_i)\exp(-\lambda_0 t_i)}{(1-\exp(-\lambda_0 t_i))^2}\right]}\right\} . \tag{11}$$

Following the standard lines (see an earlier paper[1] for details), in view of the condition (2) satisfied by F , it can be easily demonstrated that the sequence $\{f_{ij}^{*}\}$, as defined in (7), is a Cramér sequence. Finally the test criterion for the optimal $C(\alpha)$-test for the hypothesis H_0 , is to reject H_0 if $Z^{*}(\lambda_0) > \nu(\alpha)$, where $\nu(\alpha)$ satisfies $\Phi(\nu(\alpha)) = 1-\alpha$, Φ being the cumulative standard normal distribution function, and where and where

$$Z^*(\lambda_0) = [\Sigma_{i=1}^{s}\Sigma_{j=1}^{n_i} f_{ij}^*]/[\Sigma_{i=1}^{s}\Sigma_{j=1}^{n_i}\mathrm{Var}\ f_{ij}^*]^{1/2} \quad . \qquad (12)$$

The test statistic Z^* is asymptotically $\mathcal{N}(0,1)$ under H_0 as $n \to \infty$, and $n_i \to \infty$ for each i such that $n_i/n \to \delta_i$ with $0 < \delta_i < 1$. λ_0 in (12) may be replaced by some suitable root n consistent estimator. The asymptotic power of the above test for alternatives closely to $H_0: \xi = 0$, is given by

$$1 - \Phi(\nu(\alpha) - \xi[\Sigma_{i=1}^{s}\Sigma_{j=1}^{n_i}\mathrm{Var}\ f_{ij}^*]^{1/2}) \quad . \qquad (13)$$

In the above, the asymptotic normality approximation for the test criterion (6) or equivalently (12) holds only if λ_0 is fixed and is not too small. Unfortunately, in the case of very rare events this is not the case. As such, the normal distribution approximation is very poor even for large n_i and needs a further approximation when λ_0 is assumed to be extremely small. At this point, we need to add few words concerning the order of values for λ_0 that may be considered appropriate for this approximation. Earlier[2] an optimal test was constructed based on complete data including even those galaxies which show no supernova. There, the approximation for the distribution of the test statistic, owing to the very rarety of the event, was achieved by assumign that $\lambda_0 = \delta N^{1/2}$. In the present case, it is easily seen that as $N \to \infty$ and $N_i \to \infty$ such that $N_i/N \to \nu_i$ with $0 < \nu_i < 1$, for $i=1,2,\cdots,s$, the law of large number yields approximately

$$n_i \approx N\nu_i\{1 - \int \exp[-\lambda_0 t_i \exp(x\,\xi^{1/2})]\,dF(x)\} \tag{14}$$

and hence

$$n/N \approx \Sigma_{i=1}^{s}\nu_i\{1 - \int \exp[-\lambda_0 t_i \exp(x\,\xi^{1/2})]\,dF(x)\} \,. \tag{15}$$

Here $\approx$ implies that the ratio of the two sides tends to one in probability as $N \to \infty$. The fact that $\lambda_0 = 0(N^{-1/2})$ and (15), immediately yield that in terms of n, $\lambda_0 = 0(n^{-1})$. Unfortunately, by taking $\lambda_0 = 0(n^{-1}) = \Delta/n$ for some $\Delta > 0$, the distribution of $\Sigma_{i=1}^{s}\Sigma_{j=1}^{n_i} f_{ij}^{*}$ tends to be degenerate as $n_i \to \infty$ and $n \to \infty$ such that $n_i/n \to \delta_i$. This is not unexpected however, for the simple reason that if we are going to base our decision only on the data truncated at zero, then in order to make any headway in our problem of testing the homogeneity of occurrence of supernovae, the events, although are very rare, better not be considered as rare as to make $\lambda_0 = 0(n^{-1}) = 0(N^{-1/2})$, remembering that this order of λ_0 was considered very rare earlier[2] only for the nontruncated data. Thus, in order that we obtain a reasonable approximation for the asymptotic distribution of our test statistic (6), in the sense that it be nondegenerate and that the variance (11) tend to a nonzero but a finite quantity as $n \to \infty$, an appropriate order of λ_0 turns out to be $\lambda_0 = O(n^{-1/2})$. Following this analysis in terms of N, it can be easily seen that this amounts to taking $\lambda_0 = O(N^{-1/3})$ instead of $\lambda_0 = O(N^{-1/2})$ as was the case before[2]. Our next step then is to approximate the distribution of (6) or equivalently of $\Sigma\Sigma f^*_{ij}$ with $\lambda_0 = \Delta n^{-1/2}$, and $\Delta > 0$. Rewriting (6) we have

$$\psi(\underset{\sim}{\mu}) = T_n(\underset{\sim}{\mu}) - \mathcal{S}_n(\underset{\sim}{\mu}) + B_n \tag{16}$$

where

$$T_n(\underset{\sim}{\mu}) = \frac{1}{2}\Sigma_{i=1}^{s}\Sigma_{j=1}^{n_i}(\mu_{ij}-1)(\mu_{ij}-2) \quad , \tag{17a}$$

$$\mathcal{S}_n(\underset{\sim}{\mu}) = \Sigma_{i=1}^{s}(A/2+\lambda_0 t_i - 1)\left\{\mu_{i\cdot} - \frac{n_i\lambda_0 t_i}{1-\exp(-\lambda_0 t_i)}\right\}, \tag{17b}$$

$$B_n = \Sigma_{i=1}^{s}\left\{\frac{n_i\lambda_0 t_i}{1-\exp(-\lambda_0 t_i)} - n_i - \frac{n_i(\lambda_0 t_i)^2}{2[1-\exp(-\lambda_0 t_i 0]}\right\}, \tag{17c}$$

$$\mu_{i\cdot} = \sum_j \mu_{ij} \quad , \tag{17d}$$

and A is as defined in (10). We need now the following lemma, the proof of which, being elementary, is omitted.

<u>Lemma</u>. <u>Subject to</u> (4), <u>and with</u> $\lambda_0 = \Delta n^{-1/2} = \Delta\delta_i^{1/2} n_i^{-1/2}$, <u>for every</u> $i=1,2,\cdots,s$, as $n_i \to \infty$,

$$\mu_{i\cdot}/n_i \xrightarrow{P} 1 \quad , \tag{18}$$

and

$$n_i^{-1/2}\left(\mu_{i\cdot} - \frac{n_i\lambda_0 t_i}{1-\exp(-\lambda_0 t_i)}\right) \xrightarrow{P} \frac{\Delta\delta_i^{1/2} t_i}{2} \left\{\frac{E[\exp(2X\xi^{1/2})]}{E[\exp(X\xi^{1/2})]} - 1\right\}, \tag{19}$$

the last limit being zero under the hypothesis $H_0: \xi = 0$.

We may recall here that by assumption X takes values in a finite interval $[c,d]$ with probability one, so that it does have a moment generating function.

Now it can be shown that as $n_i \to \infty$,

$$A/2 + \lambda_0 t_i - 1 = \Delta\delta_i^{1/2} n_i^{-1/2}\left[t_i - \frac{\Sigma\delta_i t_i^2}{3\Sigma\delta_i t_i}\right] + o(n_i^{-1/2}) \ , \tag{20}$$

and that

$$\lim_{n\to\infty} B_n = -\frac{\Delta^2}{6}\Sigma_{i=1}^s \delta_i t_i^2 \ . \tag{21}$$

By virtue of (19) and (20), it then follows that as $n \to \infty$,

$$\mathcal{T}_n(\underset{\sim}{\mu}) \xrightarrow{P} \Delta^2/2 \cdot \left(\frac{E[\exp(2X\xi^{1/2})]}{E[\exp(X\xi^{1/2})]} - 1\right) \cdot \sum_{i=1}^{s} \delta_i t_i\left(t_i - \frac{\Sigma\delta_i t_i^2}{3\Sigma\delta_i t_i}\right), \tag{22}$$

which is zero when $\xi = 0$. Thus rejection of H_0 for large values of $\Sigma\Sigma f_{ij}^*$ is equivalent asymptotically to its rejection for large values of the statistic $T_n(\underset{\sim}{\mu})$. As such the optimal $C(\alpha)$-test for very rare events in the present case reduces to the rule of rejecting H_0 whenever

$$T_n(\underset{\sim}{\mu}) > \nu^* \ , \tag{23}$$

where ν^* is a constant determined from the asymptotic distribution of $T_n(\underset{\sim}{\mu})$ for large n, such that under H_0, the probability of $T_n(\underset{\sim}{\mu}) > \nu^*$ is acceptable as the level of significance. Note that this test remains an optimal $C(\alpha)$-test no matter what the mixing distribution F of (4) is, except that it satisfied condition (2). What remains to be done now is to find the limiting distribution of $T_n(\underset{\sim}{\mu})$ as $n \to \infty$. This is achieved in the following theorem.

Theorem. Subject to (4) and $\lambda_0 = \Delta n^{-1/2} = \Delta\delta_i^{1/2} n_i^{-1/2}$, whatever be the mixing distribution F satisfying (2) and that it be concentrated on an arbitrary finite interval $[c,d]$, the limiting distribution of $T_n(\underset{\sim}{\mu})$ as $n \to \infty$ is Poisson with expectation

$$\tau(\xi) = \frac{E[\exp(3X\xi^{1/2})]}{E[\exp(X\xi^{1/2})]} \cdot \frac{\Delta^2}{6}\left[\sum_{i=1}^{s} \delta_i t_i^2\right] \quad . \tag{24}$$

The proof of this theorem is based on taking the limit of the characteristic function of $T_n(\underset{\sim}{\mu})$ in a straightforward manner. Thus under H_0, the constant ν^* is obtained by finding the maximum k satisfying $\Pr(T_n(\underset{\sim}{\mu}) > k) \leq \alpha$, where the distribution of $T_n(\underset{\sim}{\mu})$ is taken to be Poisson with expectation $\tau(0) = (\Delta^2/6)(\Sigma_{i=1}^{s}\delta_i t_i^2)$, and α is the prefixed level of significance. Here since $\delta_i = n_i/n$,

$$\tau(0) = \frac{\lambda_0^2}{6} \Sigma_{i=1}^{s} n_i t_i^2 \quad .$$

Also, λ_0 should be replaced by a suitable root n consistent estimator such as

$$\hat{\lambda} = [\Sigma_{i=1}^{s}\Sigma_{j=1}^{n_i}\mu_{ij}(\mu_{ij}-1)]/[\Sigma_{i=1}^{s}\Sigma_{j=1}^{n_i}n_{ij}t_i] \quad . \tag{26}$$

Using a table of Poisson distribution one can easily find out the constant ν^*. For asymptotic power calculations again one can proceed as in Bühler et al[2]. For this one needs to calculate the power for different values of the constant $E[\exp(3X\xi^{1/2})]/E[\exp(X\xi^{1/2})]$ in (24) for $\xi > 0$, while using the Poisson approximation for the distribution of $T_n(\underset{\sim}{\mu})$ as suggested by the above theorem.

3. Performance Comparison of the Two tests. Let us now consider the situation[2] where all the observations including the ones with no supernova are reported. Let T_{ij} denote for this case the number of supernovae reported in the jth galaxy observed for a period of time t_i at the ith observatory with $j=1,2,\cdots,N_i$; $i=1,2,\cdots,s$. Let $N_i/N = \nu_i$ where $N = \Sigma N_i$. Here T_{ij} follows the distribution as given in (3) and they are assumed to be mutually independent. For this case, the optimal $C(\alpha)$-test for the hypothesis of homogeneity $H_0: \xi = 0$ has been given elsewhere[1]. This involves calculation of an expression analogous to (6) given by

$$\tilde{\psi}(\underset{\sim}{T}) = \frac{1}{2}\Sigma_{i=1}^{s}\Sigma_{j=1}^{N_i}T_{ij}(T_{ij}-1) - \lambda_0\Sigma_{i=1}^{s}t_iT_{i\cdot} + \frac{1}{2}\lambda_0\Sigma_{i=1}^{s}N_it_i^{\,2} \quad , \tag{27}$$

where $T_{i\cdot} = \Sigma_{j=1}^{s} T_{ij}$. Variance of (27) under H_0 is given by

$$\text{Var } \tilde{\psi} = \frac{\lambda_0^2}{2} \sum_{i=1}^{s} N_i t_i^2 \quad . \tag{28}$$

The optimal $C(\alpha)$-test is the rule of rejecting H_0 whenever

$$\tilde{Z}(\lambda_0) = \tilde{\psi}(\underset{\sim}{T}) / [\text{Var } \tilde{\psi}]^{1/2} > \nu(\alpha) \quad , \tag{29}$$

where $\nu(\alpha)$ is same as defined before. Here the asymptotic distribution of $\tilde{Z}(\lambda_0)$ is $\mathfrak{N}(0,1)$ for large N . For fixed λ_0, the asymptotic local efficiency of an optimal $C(\alpha)$-test based on complete data as against the one based on truncated data, is given by the ratio of the variance (28) to the variance (11), with n_i in (11) replaced by $N_i[1-\exp(-\lambda_0 t_i)]$, which is approximately valid for large N . This efficiency is considerably large in general. In fact in the present case of very rare events with $\lambda_0 = O(N^{-1/3})$, this efficiency tends to ∞ . This can be easily verified by noticing that the variance (11), in view of the theorem of section 2, tends under H_0 to $\tau(0)$, while the variance (28) tends to ∞ . Thus the practice of not reporting of galaxies with no supernova results in a considerable loss in efficiency of the appropriate test used for detecting any hetrogeneity in the occurrence of supernovae over various galaxies.

Finally, we close with the remark that for the nontruncated case treated earlier[2], the approximation of the test statistic resulted into a Poisson distribution, where we had taken essentially $\lambda_0 = O(N^{-1/2})$. By contrast, under the present conditions on λ_0, the corresponding test statistic such as $\tilde{Z}$ of (29) is asymptotically $\mathcal{N}(0,1)$. This is because of the fact that we have taken $\lambda_0 = O(N^{-1/3})$ here, which implies that the events in question, although are very rare, but are not as rare as were treated earlier[2]. If, on the other hand, in reality the rarety is really to the extent that $\lambda_0 = O(N^{-1/2})$, then, as mentioned earlier, the corresponding optimal $C(\alpha)$-test statistic based on the truncated data and given by (6), degenerates in the limit, making thereby the complete reporting even more essential.

[1] Bartoo, J. B. and Puri, P.S., "On optimal asymptotic tests of composite statistical hypotheses", Ann. Math. Statist., **38**, 1845-1852 (1967).

[2] Bühler, W.,at el, "Locally optimal test for homogeneity with respect to very rare events," Proc. Nat. Acad. Sc., **54**, 673-680 (1965).

[3] Neyman, J., "Optimal asymptotic tests of composite statistical hypotheses," probability and statistics (The Herald Cramér volume) Almquist and Wiksells, Uppsala. (1959).

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Purdue University	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

Loss Due to Missing Data in Efficiency of a Locally Optimal Test for Homogeneity with Respect to very Rare Events.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report, August, 1970

5. AUTHOR(S) (Last name, first name, initial)

Puri, Prem S.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
August, 1970	17	None

8a. CONTRACT OR GRANT NO. NONR 1100(26) b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Mimeo Series # 234 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Office of Naval Research Washington, D. C.

13. ABSTRACT

With reference to observations of supernovae in galaxies, a locally optimal test for the hypothesis of homogeneity of the observational units with respect to occurrence of supernovae (treated as a very rare event) is obtained for the case where the data are available only for those galaxies where at least one supernova is observed. It is shown that the loss in the asymptotic efficiency of the test due to lack of reporting of galaxies with no supernovae is very heavy and in fact is infinite for the case of very rare events.

DD FORM 1 JAN 64 1473

Unclassified
Security Classification