

The estimation of the number of mutationally silent loci in saturation-mapping experiments

By JOHN A. BARRETT

*Department of Genetics, University of Cambridge, Downing Street,
Cambridge CB2 3EH*

(Received 12 March 1979)

SUMMARY

When the technique of saturation mapping is employed to estimate the number of loci in a distinct chromosomal region, there is always the possibility that some loci will not be detected. If the number of mutants per locus follows a Poisson distribution, the number of mutationally silent loci can be estimated. This paper describes a method for fitting such data to a Poisson distribution truncated at the zero class and a method for estimating the number of mutationally silent loci. The use of these methods is demonstrated by their application to some published data.

1. INTRODUCTION

The banding patterns of the polytene salivary gland chromosomes of *Drosophila* species have intrigued geneticists for many years (Bridges 1935, 1938). The thesis of 'one function: one chromomere' has recently been revived and investigated by a number of workers (Hochman, 1971, 1973; Judd, Shen & Kaufman, 1972; Judd & Young, 1973; Lim & Snyder, 1974; Liu & Lim, 1975). The method has been to induce and recover mutants from every locus within a small cytologically distinct region of chromosome. The mutants have been characterized by complementation tests and mapped both by cytological and genetic tests. Two regions have principally been used in these experiments; the *zeste-white* region of the X-chromosome (Judd *et al.* 1972; Lim & Snyder, 1974; Liu & Lim, 1975) and chromosome 4 (Hochman, 1971, 1973) of *Drosophila melanogaster*. Even the most efficient mutagens induce mutants at a very low frequency, e.g. 25 mM EMS induces recessive lethals in the *zeste-white* region at a rate of 0.86% for the whole region (Lim & Snyder, 1974). Unless large numbers of flies are to be screened for mutants, there is always the possibility that no mutants will be recovered from some loci.

This paper describes a statistical method which estimates the number of mutationally silent loci. The problems of fitting observations to the statistical model are also discussed.

The number of mutants per locus recovered from the region investigated should follow a Poisson distribution *if* the following conditions are met.

- (1) The probability of a single mutational event is very small.
- (2) All loci are equal in size.

(3) All loci have the same mutability.

(4) There is no locus/mutagen interaction.

The most effective mutagen described by Lim & Snyder (1974) is EMS which induces mutations in the *zeste-white* region at a rate of 0.86%. If there are twenty loci in this region, the average mutation rate per locus is 0.04%. Condition 1 is therefore met. If the remaining conditions can be assumed to be true, the data from such experiments can be fitted to a Poisson distribution. However, the class of loci from which no mutants have been recovered is missing from the data. The observations must, therefore, be fitted to a truncated Poisson distribution.

	Number of Mutants							
	1	2	3	4	5	6	7	...
No. of Loci ...	A	B	C	D	E	F	G	
	A + B + C + D + ... = N.							

The likelihood of obtaining a given sample is:

$$L = \left(\frac{1}{1-p_0} \right)^N \frac{N!}{A!B!C! \dots} p_1^A p_2^B p_3^C \dots, \quad \text{where } p_i = \frac{m^i e^{-m}}{i!(1-e^{-m})}$$

from which we can obtain the maximum likelihood estimate of the Poisson parameter, \hat{m} . (Cohen, 1960).

$$\frac{\hat{m}}{1-e^{-\hat{m}}} = \frac{(A + 2B + 3C + 4D + \dots)}{N} = \bar{x}. \tag{1}$$

This equation can be solved iteratively:

Let
$$m'' = \bar{x}(1-e^{-m'}). \tag{2}$$

The solution \hat{m} is obtained when:

$$m'' = m'.$$

It then follows that the variance of \hat{m} is:

$$\text{var}(\hat{m}) = \frac{m(1-e^{-m})^2}{N(1-e^{-m}(1+\hat{m}))}. \tag{3}$$

The advent of small programmable calculators has made the graphical approach of Cohen (1960) redundant since equations (2) and (3) can be easily solved on these machines.

It then follows that the estimated number in the zero class (\hat{n}_0) is

$$\hat{n}_0 = \frac{N e^{-\hat{m}}}{1-e^{-\hat{m}}}.$$

From a Taylor expansion (see Appendix) it can be shown that:

$$\text{var}(n_0) \cong \frac{N\hat{m} e^{-2\hat{m}}}{(1-e^{-\hat{m}}(1+\hat{m})) (1-e^{-\hat{m}})^2}.$$

An analogous method to Fisher's Index of Dispersion (Fisher, Thornton & MacKenzie, 1922) has been derived by which it is possible to test the fit of a truncated Poisson distribution (see Appendix) by calculating,

$$\chi_{N-1}^2 = \frac{\sum (x_i(1 - e^{-\hat{m}}) - \hat{m})^2}{\hat{m}(1 - e^{-\hat{m}}(1 + \hat{m}))}$$

and comparing it with tabulated χ^2 values on $N - 1$ degrees of freedom.

2. APPLICATION OF THE METHOD

To demonstrate the use of this method the data of von Bortkiewicz (1898) on the number of men killed by horse-kick in ten corps of the Prussian army over a period of twenty years has been used.

	Deaths				
	0	1	2	3	4
Frequency observed ...	109	65	22	3	1

Let us assume that the zero class is missing and we wish to estimate it from the remaining data. The above method yields:

$$\begin{aligned} \hat{m} &= 0.6183 & \text{var}(m) &= 0.0113 \\ \hat{n}_0 &= 106.34 & \text{var}(n_0) &= 600.40 \\ & & \text{Index of dispersion} - \chi_{90}^2 &= 87.1879 \text{ N.S.}, \end{aligned}$$

which compares with the estimates from the full distribution:

$$\begin{aligned} \hat{m} &= 0.6100 & \text{var}(m) &= 0.00305. \\ \hat{n}_0 &= 108.67 \end{aligned}$$

Inevitably the smaller size of the truncated sample produces a larger estimate of the variance of m , than the full sample. This method relies on a Normal approximation to the data and so for the testing procedure to be valid, reasonably large samples must be used. However, most of the published data consist of rather small samples. An alternative approach for these data is to test the goodness-of-fit to a truncated Poisson distribution by use of χ^2 . The value of \hat{m} , estimated by iteration (see (2) above) is used to estimate the expected frequencies in each class. Although the goodness-of-fit χ^2 is more conservative, but is more reliable for small samples, both methods give similar results with large samples, e.g. applying the goodness-of-fit test to the data of von Bortkiewicz when the zero class is missing:

	Deaths		
	1	2	3 and greater
Observed ...	65	22	4
Expected ...	65.7481	20.3260	4.9259
Goodness-of-fit $\chi_1^2 = 0.3204$ N.S.			

With smaller samples, the goodness-of-fit χ^2 test will often fail to reject the null hypothesis even when it is untrue (Type II Error); non-significant results must, therefore, be treated with some caution. In these cases, failure to reject the null hypothesis may merely indicate that the sample size is too small to detect a deviation from a truncated Poisson distribution.

Where a goodness-of-fit χ^2 test has been used, classes have been pooled to give expected values of the order of 3.0; where sample sizes are so small that even with pooling it is not possible to obtain a minimum of three classes each with expected values of about 3.0, no test has been carried out. It is possible with these very small samples to test the fit to a truncated Poisson distribution using the exact method of Fisher (1950) but the method is laborious and even if a non-significant result is obtained, the chances of Type II Error are still large.

Applying these methods to some published data on saturation-mapping experiments the following results have been obtained.

(i) *Hochman's data*

The method was applied to the results of Hochman (1971) for chromosome 4 of *Drosophila melanogaster* (Table 1).

$$\hat{m} = 4.5450 \quad \text{var}(m) = 0.1477.$$

Goodness-of-fit $-\chi^2_5 = 22.41^{***}$; Index of Dispersion $-\chi^2_{31} = 204.53^{***}$.
And Hochman (1973) (Table 2).

$$\hat{m} = 5.1363 \quad \text{var}(m) = 0.1463.$$

Goodness-of-fit $-\chi^2_5 = 31.92^{***}$; Index of Dispersion $-\chi^2_{35} = 269.68^{***}$.

The pooled data used by Hochman obviously do not fit a Poisson distribution and so further estimates from the data are not valid. Hochman's estimates of the zero class are, therefore, in error.

(ii) *Judd's data* (Judd *et al.* 1972; Judd & Young, 1973)

The mapping of the 3A-3C region of the X-chromosome of *Drosophila melanogaster* by Judd *et al.* (1972) produced the following results:

$$\hat{m} = 9.6661 \quad \text{var}(m) = 0.8059$$

$$\text{Goodness-of-fit } -\chi^2_1 = 7.11^{**}.$$

These data obviously do not fit a Poisson distribution. Even if the twenty-six mutants later described by Judd & Young (1973) are added to these data the fit is not improved (Table 4).

$$\hat{m} = 11.6363 \quad \text{var}(m) = 1.0286$$

$$\text{Goodness-of-fit } -\chi^2_1 = 5.51^{**}.$$

Judd however, recognized the limitations of his data and did not attempt to estimate the number of non-mutating loci.

Table 1. *Data from Hochman (1971)*

		No. of mutants per locus										
		1	2	3	4	5	6	8	9	11	12	29
No. of loci	...	7	7	6	2	1	3	2	1	1	1	1
		No. of mutants per locus (pooled data)										
		1-2	3	4	5	6	> 6					
Observed nos.	...	14	6	2	1	3	6					
Expected nos.	...	5.11	5.37	6.11	5.55	4.21	5.65					

Table 2. *Data from Hochman (1973)*

		No. of mutants per locus												
		1	2	3	4	5	7	8	9	10	11	15	35	
No. of loci	...	10	3	5	6	2	2	2	1	1	2	1	1	
		No. of mutants per locus (pooled data)												
		1-2	3	4	5	6	7	> 7						
Observed nos.	...	13	5	6	2	0	2	8						
Expected nos.	...	3.90	4.81	6.17	6.34	5.43	3.98	5.36						

Table 3. *Data from Judd et al. (1972)*

		No. of mutants per locus									
		2	3	4	5	7	8	9	12	20	34
No. of loci	...	1	1	1	2	2	1	1	1	1	1
		No. of mutants per locus (pooled data)									
		1-7	8-10	> 10							
Observed nos.	...	7	2	3							
Expected nos.	...	3.02	4.48	4.50							

Table 4. *Data from Judd and Young (1973)*

		No. of mutants per locus										
		2	3	4	6	8	9	10	11	14	23	41
No. of loci	...	1	1	1	1	2	1	1	1	1	1	1
		No. of mutants per locus (pooled data)										
		1-9	10-12	> 12								
Observed nos.	...	7	2	3								
Expected nos.	...	3.37	4.11	5.52								

Table 5. *Data from Lim & Snyder (1974)*

		No. of mutants per locus						
		1	3	4	5	7	28	35
No. of loci ...	2	1	2	3	2	1	1	
		No. of mutants per locus (pooled data)						
		1-7		8-9		> 9		
Observed nos. ...		10		0		2		
Expected nos. ...		4.25		3.20		4.55		

Table 6. *Data from Liu & Lim (1975)*

		No. of mutants per locus								
		1	2	3	5	6	7	12	18	21
No. of loci ...	2	1	3	2	2	1	2	1	1	
		No. of mutants per locus (pooled data)								
		1-5			6-7		> 7			
Observed nos. ...		8			3		4			
Expected nos. ...		4.51			4.48		6.01			

Table 7

Treatment	No. flies	
	non-mutant	Mutant
25 mM EMS	970	6
25 mM EMS	3385	34
25 mM EMS	5349	50
0.1 mM TEM	5141	7
0.1 mM TEM	4894	7
0.1 mM TEM	6158	16
0.15 mM TEM†	607	1
0.2 mM TEM†	865	3
	1472	4

Brandt and Snedecor's χ^2_6 on whole table 78.09***; for EMS experiments only $\chi^2_2 = 1.21$ n.s.; for TEM experiments only, $\chi^2_3 = 3.39$ n.s.

† Data pooled for analysis.

(iii) *Lim's data*

The data pooled for both TEM and EMS mutagenesis (Lim & Snyder, 1974), give the following results (Table 5).

$$\hat{m} = 8.7486 \quad \text{var}(m) = 0.7299$$

$$\text{Goodness-of-fit } \chi^2_1 = 12.42***.$$

For MMS induction (Liu & Lim, 1975) (Table 6).

$$\hat{m} = 6.9936 \quad \text{var}(m) = 0.4688$$

$$\text{Goodness-of-fit } -\chi^2_1 = 3.85*.$$

The most obvious source of the deviation from a Poisson distribution in Table 5 is that the data are the results of different experiments carried out at different times, under different conditions using mutagens of differing effectiveness; Lim & Snyder (1974) give the induced mutation rates for both TEM and EMS (Table 7). A large source of heterogeneity is the TEM-EMS difference. The three different EMS experiments show no difference in the proportions of mutants recovered. Similarly, the TMS experiments show no significant differences in the proportions of mutants recovered. However, there is a significant difference between the proportions of mutants recovered in the EMS and TMS experiments.

Table 8(a) and (b) classify Hochman's data according to the mutagen used. In all cases the sample sizes are small; so small in many cases that a goodness-of-fit test has not been carried out. In those cases where a test has been carried out, a non-significant deviation from a truncated Poisson distribution has been obtained. However, because of the small sample sizes, further analysis is not justified. If the results for non-EMS and non-X-ray induced mutations are pooled (Table 9), the pooled data do not deviate significantly from a Poisson distribution ($\chi^2_1 = 2.03$). The pooling of results from different experiments in order to estimate the number of mutationally-silent alleles may be invalid since the distributions of mutant produced by different mutagens may not be statistically independent. This same criticism can be made of Hochman's analysis (1973).

The data of Judd *et al.* (1971, 1973) (Tables 3 and 4) are pooled for different mutagens and procedures. The data are re-classified by mutagen in Tables 10 and 11; in Table 10 data are derived from Judd *et al.* (1971) and the data in Table 11 from Judd & Young (1973). Again most of the data do not deviate significantly from Poisson distributions, but the sample sizes are small.

The data presented by Lim & Snyder (1974) and Liu & Lim (1975) consider each mutagen separately (Tables 12 and 6). Where the sample is large enough, i.e. for EMS mutagenesis, the data do not fit a truncated Poisson distribution. That these experiments still show deviations from random production of mutants indicates that one or more of the assumptions outlined above must be wrong.

Condition: (1) Valid

- (2) There is no independent evidence that the loci are equal in size.
- (3) May be invalid. There is no evidence that all of the loci have the same mutability.
- (4) The assumption of no mutagen/locus interaction may be invalid.

The existence of the same hot-spots in the *zeste-white* region of the X-chromosome, notably *zw1* and *zw2*, when alkylating agents are used for mutagenesis has been shown by Lim & Snyder (1974), Liu & Lim (1975), Judd *et al.* (1971) and Judd & Young (1973) in independent experiments. Hot-spots could be caused by any one or any combination of assumptions (2)–(4) being untrue. Until independent evidence is available on the size of loci and their mutability it will not be possible to correct for the hot-spots.

Table 8(a). † *Data of Hochman (1971): no. of loci*

Mutagen	No. of mutants per locus							
	1	2	3	4	5	6	17	
X-rays	5	2	3	0	0	2	0	n.s.
Spontaneous	10	3	3	1	0	0	0	n.s.
EMS	9	5	6	2	0	2	1	n.s.
ICR	4	0	1	1	0	0	0	Not tested
MEL	1	1	0	0	0	0	0	Not tested

Goodness-of-fit χ^2 : X-rays, $\chi_1^2 = 1.6964$; spontaneous, $\chi_1^2 = 1.5575$; EMS, $\chi_3^2 = 5.7986$.

† Curly brackets in the body of the table indicate classes pooled for analysis.

Table 8(b). † *Data of Hochman (1973): number of loci*

Mutagen	No. of mutants per locus									
	1	2	3	4	5	6	9	10	23	
X-rays	5	2	3	0	0	2	0	0	0	n.s.
Spontaneous	11	3	3	1	0	0	0	0	0	n.s.
EMS	10	2	7	5	0	1	2	1	1	n.s.
ICR	4	0	1	1	0	0	0	0	0	Not tested
MEL	1	1	0	0	0	0	0	0	0	Not tested

Goodness-of-fit χ^2 : X-rays, $\chi_1^2 = 1.6960$; spontaneous, $\chi_1^2 = 1.8067$; EMS, $\chi_3^2 = 7.4162$.

† Curly brackets in the body of the table indicate classes pooled for analysis.

Table 9. *Data of Hochman (1971): no. of mutants per locus (excluding X-ray and EMS induced mutants)*

No. of loci ...	No. of mutants per locus					
	1	2	3	4	5	6
...	9	4	3	0	1	2
No. of mutants per locus (pooled data)						
	1	2	> 2			
Observed nos. ...	9	4	6			
Expected nos. ...	6.20	6.00	6.80			

Goodness-of-fit: $\chi_1^2 = 2.0286$ n.s.

3. DISCUSSION

In saturation mapping experiments there is always a possibility that not all of the loci in the region under consideration produce mutants. When small numbers of mutants are recovered, the number of mutationally silent loci may contribute a

Table 10.† *Data of Judd et al. (1971): no. of loci*

Mutagen	No. of mutants per locus											
	1	2	3	4	5	6	7	8	9	10	20	
X-rays	0	4	1	1	1	1	0	0	1	0	1	N.S.
EMS	2	0	1	-	-	-	-	-	-	-	-	Not tested
Ethylenimine	0	1	1	-	-	-	-	-	-	-	-	Not tested
Ethylenimine + X-rays	2	1	1	-	-	-	-	-	-	-	-	Not tested
NGG	3	0	1	1	1	0	0	1	-	-	-	Not tested
DMS + X-rays	0	1	1	-	-	-	-	-	-	-	-	Not tested
ICR-170	2	0	0	-	-	-	-	-	-	-	-	

Goodness-of-fit χ^2 : X-rays, $\chi^2_1 = 2.5397$ n.s.

† Curly brackets in the body of the table indicate classes pooled for analysis.

Table 11. *Data of Judd & Young (1973): no. of loci*

Mutagen	No. of mutants per locus									
	1	2	3	4	5	6	7	8	9	10
EMS	2	1	2	2	0	0	0	0	0	1

Not tested.

Table 12. *Data of Lim & Snyder (1974)*

No. of loci	Mutagen	No. of mutants per locus										
		1	2	3	4	5	6	7	12	23		
	EMS	5	0	3	1	2	0	1	0	2	**	
	TEM	2	2	1	0	1	0	0	1	Not tested		

No. of mutants per locus (pooled data (EMS only))

	1-4	5-6	> 6
Observed nos. ...	9	2	3
Expected nos. ...	4.41	4.58	5.01

Goodness-of-fit: $\chi^2_1 = 7.0253^{**}$

large proportion of the loci in the region. Various investigators (Hochman, 1971, 1972) have attempted to estimate the number of non-mutant sites assuming a Poisson distribution of the number of mutants per locus. However, in most of the data so far published, the number of mutants per locus do not have a Poisson distribution. This fact was recognised by Judd and his co-workers and no attempt was made in their reports to estimate the number of non-mutant sites.

Where estimates have been made (Hochman, 1971, 1973) a method derived by Alikhanian (1937) based on the Poisson distribution was used. However, this method uses only the one-hit and two-hit classes. Consequently a lot of information from the experiments is ignored. For von Bortkiewicz' data Alikhanian's

method gives an estimate of the zero class of 96.02. The method outlined in this paper utilizes all of the available information. Since most of the published data cannot be shown to follow Poisson distributions due to mutagenic hot-spots or insufficient data both estimation procedures are inapplicable; for example, the data of Wright *et al.* (1976) and Gvozdev *et al.* (1975) have not been analysed because the sample sizes are too small.

The systematic departure of much of the published data from random distribution of mutants among the loci cannot be allowed for in the statistical model. Such data cannot be used to estimate the number of non-mutant sites. The only way out of this dilemma is to accumulate such a large number of mutants that the classes with small numbers of hits are absent from the data. This would reduce the probability of non-mutant loci to a very small value. This of course does not rule out the possibility of non-mutant sites remaining if such sites are not affected by the mutagens used or detectable by the genetic screen.

I would like to thank the reviewer of an earlier draft of this paper for many helpful comments and suggestions.

REFERENCES

- ALIKANIAN, S. I. (1937). A study of the lethal mutations in the left end of the sex chromosome in *Drosophila melanogaster*. *Zoologicheskii Zhurnal (Moscow)* **16**, 247–279.
- VON BORTKIEWICZ, L. (1898). *Das Gesetz der Kleinen Zahlen*. Leipzig.
- BRIDGES, C. B. (1935). Salivary chromosome maps, with a key to the banding of the chromosomes of *Drosophila melanogaster*. *Journal of Heredity* **26**, 60–64.
- BRIDGES, C. B. (1938). A revised map of the salivary gland X-chromosome of *Drosophila melanogaster*. *Journal of Heredity* **29**, 11–13.
- COHEN, A. C. (1960). Estimating the parameter in a conditional Poisson distribution. *Biometrics* **16**, 203–211.
- FISHER, R. A., THORNTON, H. G. & MACKENZIE, W. A. (1922). The accuracy of the plating method of estimating the density of bacterial populations. *Annals of Applied Biology* **9**, 325–359.
- FISHER, R. A. (1928). On a property connecting the Chi-square measure of discrepancy with the method of maximum likelihood. *Atti del Congresso Internazionale dei Matematica Bologna*, vol. VI, 95–100.
- FISHER, R. A. (1950). The significance of deviations from expectation in a Poisson series. *Biometrics* **6**, 17–24.
- GVOZDEV, V. A., GOSTIMSKY, S. A., GERASIMOVA, T. I., DUBROVSKAYA, E. S. & BRASLAVSKAYA, O. Y. U. (1975). Fine genetic structure of the 2D3-2F5 region of the X-chromosome of *Drosophila melanogaster*. *Molecular and General Genetics* **141**, 269–275.
- HOCHMAN, B. (1971). Analysis of chromosome 4 in *Drosophila melanogaster*. II. Ethyl methane sulfonate induced lethals. *Genetics* **67**, 235–252.
- HOCHMAN, B. (1973). Analysis of a whole chromosome in *Drosophila*. *Cold Spring Harbor Symposia on Quantitative Biology* **38**, 381–389.
- JUDD, B. H., SHEN, M. W. & KAUFMAN, T. C. (1972). The anatomy and function of a segment of the X-chromosome of *Drosophila melanogaster*. *Genetics* **71**, 139–156.
- JUDD, B. H. & YOUNG, M. W. (1973). An examination of the one cistron:one chromosome concept. *Cold Spring Harbor Symposia on Quantitative Biology* **38**, 573–579.
- LIM, J. K. & SNYDER, L. A. (1974). Cytogenetic and complementation analysis of recessive lethal mutations induced in the X-chromosome of *Drosophila* by three alkylating agents. *Genetic Research (Cambridge)* **24**, 1–10.
- LIU, C. P. & LIM, J. K. (1975). Complementation analysis of methyl methane-sulfonate-induced recessive lethal mutations in the *zeste-white* region of *Drosophila melanogaster*. *Genetics* **79**, 601–611.

WRIGHT, T. R. F., BEWLEY, G. C. & SHERALD, A. F. (1976). The Genetics of DOPA decarboxylase in *Drosophila melanogaster*. II. Isolation and characterization of DOPA decarboxylase deficient mutants and their relationships to the α -methyl-DOPA-hyper-sensitive mutants. *Genetics* **84**, 287-310.

APPENDIX

In a Poisson distribution the probability of the zero class is:

$$P_0 = e^{-m},$$

hence we can estimate the number of the missing zero class of a truncated Poisson distribution by

$$n_0 = \frac{N e^{-m}}{1 - e^{-m}}.$$

Let $\mathcal{E}(x) = a$ and $\mathcal{E}(x-a)^2 = \text{var}(x)$, where $\mathcal{E}(x)$ is the *expected value* of x and $\mathcal{E}(x-a)^2$ is the *expected value* of $(x-a)^2$. Then by Taylor's Series

$$f(x) = f(a) + (x-a) f'(a) + \frac{(x-a)^2}{2!} f''(a) + \dots,$$

then
$$f(x) - f(a) \cong (x-a) f'(a),$$

if the higher derivatives are small.

$$\begin{aligned} \therefore \mathcal{E}(f(x) - f(a))^2 &\cong \mathcal{E}(x-a)^2 [f'(a)]^2 \\ \therefore \text{var}[f(x)] &\cong [f'(a)]^2 \text{var}(x) \\ \therefore \text{var} \left[\frac{e^{-\hat{m}N}}{1 - e^{-\hat{m}}} \right] &\cong \text{var}(\hat{m}) \left[\frac{\partial}{\partial m} \frac{e^{-\hat{m}N}}{1 - e^{-\hat{m}}} \right]^2 \\ \therefore \text{var}(n_0) &\cong \frac{N \hat{m} e^{-2\hat{m}}}{(1 - e^{-\hat{m}}(1 + \hat{m}))} (1 - e^{-\hat{m}})^2 \end{aligned}$$

For a large sample, say size N , it follows from the method of maximum likelihood that (Fisher, 1928):

$$\chi^2_N = \sum_i \frac{[(dL_i/dx) (\mathcal{E}x)]^2}{-(d^2L_i/dx^2) (\mathcal{E}x)}.$$

Where, L_i is the log likelihood function of i th observation, $(dL_i/dx) (\mathcal{E}x)$ the first derivative and $(d^2L_i/dx^2) (\mathcal{E}x)$ the second derivative, both evaluated at the expected value of x .

If the expected value of x is not known, an estimate, \hat{x} , must be used. This estimate is obtained by summing the first derivatives the log likelihood of each observation and equating the sum zero. When an estimate is used, χ^2 is distributed with $N-1$ degrees of freedom. For a truncated Poisson distribution, with an estimate of Poisson parameter, m :

$$\begin{aligned} \frac{dL_i}{dm} (\hat{m}) &= \frac{\hat{m} - x_i (1 - e^{-\hat{m}})}{(1 - e^{-\hat{m}}) \hat{m}}, \\ \frac{d^2L_i}{dm^2} (\hat{m}) &= \frac{(1 - e^{-\hat{m}}(1 + \hat{m}))}{(1 - e^{-\hat{m}})^2 \hat{m}}. \end{aligned}$$

From which it follows that

$$\chi^2_{N-1} = \sum_i \left[\frac{\hat{m} - x_i(1 - e^{-\hat{m}})}{(1 - e^{-\hat{m}})\hat{m}} \right]^2 \frac{(1 - e^{-\hat{m}})2\hat{m}}{1 - e^{-\hat{m}}(1 + \hat{m})},$$

$$\chi^2_{N-1} = \sum_i \left[\frac{(x_i(1 - e^{-\hat{m}}) - \hat{m}^2)}{\hat{m}(1 - e^{-\hat{m}}(1 + \hat{m}))^2} \right]$$

which is analogous to Fisher's Index of Dispersion for a full Poisson distribution.