Accid. Anal. & Prev. Vol. 25, No. 1, pp. 91–97, 1993
Printed in the U.S.A.

0001-4575/93 $6.00 + .00
© 1993 Pergamon Press Ltd.

ARE ACCIDENTS POISSON DISTRIBUTED? A STATISTICAL TEST

ALAN NICHOLSON* and YIIK-DIEW WONG†

*Department of Civil Engineering, University of Canterbury, Christchurch, New Zealand; †School of Civil and Structural Engineering, Nanyang Technological University, Singapore

(*Received* 2 *April* 1991; *in revised form* 22 *January* 1992)

Abstract—The common and convenient assumption in accident count analysis, that accidents are Poisson-distributed, is reexamined. Two statistical tests, for evaluating the assumption are described and compared. It is shown that a test based upon a combinatorial analysis is much more accurate than the alternative chi-square test when accident counts are expected to be small. The more accurate test is used to reinterpret data on accident count variability, the results indicating that the Poisson distribution is appropriate for the analysis of accidents at individual sites.

INTRODUCTION

Statistical analysis of accidents generally involves assuming some probability distribution to describe the randomness of accident occurrence and it is common practice to assume that road accidents at a location are governed by a stationary Poisson process (i.e. annual accident counts are Poisson distributed). A major attraction of the Poisson distribution is its ease of use. In fact, there is little empirical data to support the assumption, and one previous study (Nicholson 1985), involving the analysis of annual accident count data, concluded that "at many locations, the pattern of accident occurrence is either too regular or too irregular to be well described by the (stationary) Poisson process". That study involved use of an approximate statistical test, entailing checking whether the variance of the observed annual accident counts was sufficiently different from their mean to justify rejecting the Poisson distribution. Given the convenience of the Poisson distribution, it seems appropriate to require clear evidence that its use leads to unsound conclusions before rejecting it.

It should be noted that rejection of the Poisson distribution may occur if accidents are governed by a nonstationary and/or non-Poisson process. For instance, if the number of accidents in a year is governed by a Poisson process, but the expected number of accidents varies randomly from year to year according to a gamma distribution, then the annual accident counts will be negative-binomial distributed (Greenwood and Yule 1920).

The expected number of accidents may even vary nonrandomly. If there appears to be a trend, discontinuity, or pattern of overcorrection in the accident count data, one can test the statistical significance of the apparent deviation from a random sequence of accident counts (Nicholson 1986).

This paper describes an exact statistical test of whether accident counts are governed by a stationary, random Poisson process. The relationship between the exact and approximate tests is described, and the data used in the previous study (Nicholson 1985) are reinterpreted using the exact test.

DERIVATION OF EXACT TEST

Consider the case of A accidents having occurred at a location in N years, and let X_j ($j = 1,2,\ldots,N$) be the number of accidents in the jth year. Any sequence of annual accident counts $X_1,X_2,\ldots,X_N$ is possible, so long as

$$X_j \geq 0 \qquad (j = 1,2,\ldots,N) \tag{1}$$

$$\sum_{j=1}^{N} X_j = A. \tag{2}$$

The ratio of the variance to the mean for such a sequence is

$$\left\{\sum_{j=1}^{N} [X_j - (A/N)]^2\right\} \bigg/ \{(N-1)(A/N)\}, \tag{3}$$

and the expected value of the ratio is unity if accidents are Poisson-distributed.

If accident occurrence is governed by a random process, then the generation of a sequence of annual accident counts is analogous to the random placement of A indistinguishable objects into N containers. Although each accident is distinguishable from each other accident in reality, in statistical studies of the distribution of accidents among years, one is interested only in the number of accidents in each year and not in the year of occurrence of each individual accident. Hence it is appropriate to treat the accidents as being indistinguishable. The sequence of accident counts $X_1,X_2,\ldots,X_N$ is analogous to the number of objects in each container.

Two sequences are distinguishable only if the corresponding elements are not identical. For the case of two accidents in three years, there are six distinguishable sequences:

0,0,2; 0,2,0; 2,0,0; 0,1,1; 1,0,1; 1,1,0.

It can be shown (Feller 1968) that the number of distinguishable sequences satisfying the two conditions above is

$$(N + A - 1)!/[A!(N - 1)!], \tag{4}$$

and it can be seen that as A and N get large, then the number of such sequences gets very large. For the case of 30 accidents in 10 years, say, the number of distinguishable sequences exceeds 200×10^6.

A sequence $X_1,X_2,\ldots,X_N$ is the result of A independent trials with N possible outcomes, each trial involving the random allocation of one of the A accidents to one of the N years. Hence the probability of a sequence $X_1,X_2,\ldots,X_N$ is given by the multinomial distribution

$$f[X_1,X_2,\ldots,X_N] = A! \quad \prod_{j=1}^{N} (P_j)^{X_j} \Big/ \prod_{j=1}^{N} X_j! \tag{5}$$

Since the probability P_j of each outcome is $(1/N)$, then

$$f[X_1,X_2,\ldots,X_N] = A! \quad \prod_{j=1}^{N} (1/N)^{X_j} \Big/ \prod_{j=1}^{N} X_j! \tag{6}$$

Since we are interested in whether the variance-to-mean ratio for an observed sequence is sufficiently larger or smaller than unity to reject the traditional Poisson model of accident occurrence, we need to know the probability of the ratio's being larger or smaller than the observed value when the accidents are governed by a Poisson process. It is therefore sensible to group together those sequences having the same variance-to-mean ratio. For the case of two accidents in three years, three of the six distinguishable sequences have a variance-to-mean ratio equal to 0.5 while the other three have a ratio equal to 2.

Assuming that the order of the observations is not important (ie. that they may be treated as exchangeable), we can consider the sequence $K_0,K_1,\ldots,K_A$ where K_i = the number of years in which there are i accidents.

Clearly any such sequence is possible, so long as

$$K_i \geq 0 \qquad (i = 0,1,\ldots,A) \tag{7}$$

$$\sum_{i=0}^{A} K_i = N \tag{8}$$

$$\sum_{i=0}^{A} i(K_i) = A. \tag{9}$$

The ratio of the variance to the mean for such a sequence is

$$\left\{\sum_{i=0}^{A} K_i[i - (A/N)]^2\right\} \Big/ \{(N - 1)(A/N)\}, \tag{10}$$

and the expected value of the ratio is unity if accidents are Poisson-distributed.

For the case of two accidents in three years, there are two such sequences:

1,2,0 (variance/mean = 0.5), and 2,0,1 (variance/mean = 2.0).

The sequences $K_0,K_1,\ldots,K_A$ are very similar to the partitions of combinatorial analysis. Any positive integer can be partitioned (without regard to order) into a group of positive integers that sum to the original positive integer. For instance, the seven partitions of the number five are: (5); (1,4); (2,3); (1,1,3); (1,2,2); (1,1,1,2) and (1,1,1,1,1). A sequence $K_0,K_1,\ldots,K_A$ differs from a partition of A only in that the sequence includes a specific number of zero counts, such that eqn (8) is satisfied.

For a partition, however, there is no such constraint upon the number of elements of the partitions, which do not include zeros. For N equal to or greater than A, the number of possible sequences $K_0,K_1,\ldots K_A$ is the same as the number of partitions of A. For N less than A, the number of sequences is less than the number of partitions (i.e. the set of sequences is a subset of the set of partitions).

Expressions for (and tables of) the number of partitions of a positive integer are given in Abramowitz and Stegun (1964). Now, the number of partitions of A is very much less than the number of sequences $X_1,X_2,\ldots,X_N$ that sum to A. For instance, for $A = N = 10$, there are 92,378 possible sequences $X_1,X_2,\ldots,X_N$, but only 42 possible sequences $K_0,K_1,\ldots,K_A$ (or partitions). Since the exact test entails generating either all possible sequences $X_1,X_2,\ldots,X_N$ or $K_0,K_1,\ldots,K_A$, then it is clearly advantageous to use the latter.

A sequence $K_0,K_1,\ldots,K_A$ is the result of N independent trials with $(A + 1)$ possible outcomes, each trial involving the random allocation of one of the N observations to one of the $(A + 1)$ possible annual counts. Hence the probability of a sequence $K_0,K_1,\ldots,K_A$ is obtained from the multinomial distribution

$$f[K_0,K_1\ldots,K_A] = N!\left[\prod_{i=0}^{A}(P_i)^{K_i}\right]\Big/\left[\prod_{i=0}^{A}K_i!\right] \quad (11)$$

where P_i is the probability of i accidents in a year. If accidents are governed by a stationary Poisson process with parameter M,

$$P_i = (e)^{-M}(M)^i/i!, \quad (12)$$

there is no theoretical upper limit on i. Thus the upper limits of the product sums in the multinominal distribution should theoretically be $i = \infty$. For given A and N, however, only sequences with $K_i = 0$ for $i > A$ are possible. For $i > A$, the terms in the product sums are all equal to unity and do not affect the product sums. It follows (Fisher 1950) that the probability of a sequence $K_0,K_1,\ldots,K_A$ is

$$= A!N!\Big/\left\{(N)^A\left[\prod_{i=0}^{A}K_i!\right]\left[\prod_{i=0}^{A}(i!)^{K_i}\right]\right\}. \quad (13)$$

To identify critical values for the variance-to-mean ratio, a complete set of all possible sequences $K_0,K_1,\ldots,K_A$ was generated, for each distinct combination of values for A and N. Then the probability of each such sequence, along with its variance-to-mean ratio, was calculated. After ordering the sequences with respect to the variance-to-mean ratio, a cumulative distribution function for the ratio was obtained, from which lower and upper confidence limits for the variance-to-mean ratio were found (Tables 1–4).

Given a particular sequence of annual accident counts at a location (i.e. A and N are known), the variance-to-mean ratio for the sequence can be calculated and checked to see whether it lies outside the appropriate confidence limits. If it does, then one may reject the use of the Poisson distribution for analysing the accident data for that location.

AN APPROXIMATE TEST

It can be seen above that the use of the exact test involves a considerable effort, due to the very large number of possible accident count sequences for typical values of A and N, and the form of the probability expression. A simple alternative test involves assuming that the statistic

$$\sum_{i=1}^{N}[X_i - (A/N)]^2\Big/(A/N) \quad (14)$$

is chi-square distributed with $(N - 1)$ degrees of freedom. This statistic is very similar to the variance-to-mean ratio defined previously, the only difference being the omission of the $(N - 1)$ term in the denominator.

Using tables of the chi-square distribution, it is possible to prepare a chart (Fig. 1) showing the confidence limits for the variance-to-mean ratio for varying values of the series duration, N. It can be seen that the confidence interval decreases in width as the series duration increases. In addition, the rate of narrowing decreases, and the interval becomes more symmetrical about the expected value (unity), as the series duration increases.

The accuracy of this test has been discussed by several researchers. For instance, Fisher (1950) noted that this test is suspect when the expected annual accident count (i.e. A/N) is low. Rao and Chakravarti (1956) suggested that the accuracy of the approxi-

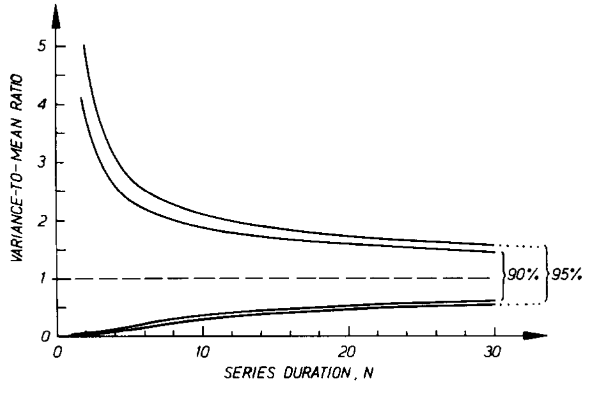

Fig. 1. Confidence limits using chi-square test.

mate test improves as (A/N) and N increase. In applying the test to accident count analysis, Nicholson (1985) relied upon the conclusion of Sukhatme (1938) that the test may be applied when (A/N) is as small as unity, when N is equal to 15 or more.

COMPARISON OF EXACT AND APPROXIMATE TESTS

Tables 1 and 2 show the lower and upper bounds, respectively, of the 90% confidence interval, based upon the exact test. Tables 3 and 4 show the bounds of the 95% confidence interval. For the case of 30 accidents in 10 years, say, one can be 90% confident that the variance-to-mean ratio will lie in the range 0.370 (Table 1) to 1.852 (Table 2), and 95% confident that it will lie in the range 0.296 (Table 3) to 2.074 (Table 4).

It should be noted that it is not generally possible to find the critical value of the variance-to-mean ratio, such that there is exactly a 2.5% or 5% probability of getting a larger or smaller value by chance, when accidents are truly Poisson-distributed. The bounds shown in Tables 1 to 4 are therefore not absolutely exact. For some combinations of total accidents (A) and number of years (N) there is no meaningful critical value. For small A and/or small N, the smallest (or largest) possible value of the variance-to-mean ratio may have a probability much greater than 2.5% or 5%, and in such cases, no critical values are given.

Tables 1 to 4 also show the critical values based upon the chi-square, approximate test. It is notable that these critical values do not depend upon the total number of accidents (A). It can be seen that for a given number of years (N), the critical value from the exact test tends toward the critical value from the approximate test as A increases.

The convergence characteristics depend upon the value of N, and as shown in Fig. 2, those characteristics improve as N increases. Fig. 3 shows that the level of agreement between the critical values from the two tests improves as the accident rate (A/N) increases. The level of agreement between the tests is also shown in Fig. 4, and it seems that there is a fairly good agreement between the exact and approximate tests so long as A is about 15 or more and N is about 6 or more.

It is worth noting that the approximate (chi-square) confidence interval does not lie outside the exact confidence interval for all values of A and N. That is, the approximate (asymptotic) test is not conservative, and it may lead to rejection of the hypothesis (that accidents are Poisson-distributed) when it should not be rejected. This seems inappropriate, given the convenience of the Poisson distribution.

The exact test has been used to reinterpret the data (i.e. 66 annual accident count sequences) obtained by Nicholson (1985), the result being that about 15% of the observed variance-to-mean ratios lay outside the 90% confidence interval. This proportion is much less than the 26% lying outside the cor-

Table 1. Lower bound for variance-to-mean ratio (90% confidence interval)

A \ N	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
2	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
3	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
4	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
5	---	---	---	0.500	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
6	---	---	---	---	0.400	0.556	---	---	---	---	---	---	---	---	---	---	---	---	---
7	---	---	---	---	---	0.333	0.469	0.571	---	---	---	---	---	---	---	---	---	---	---
8	---	---	0.333	---	---	0.417	0.286	0.406	0.500	0.575	0.636	---	---	---	---	---	---	---	---
9	---	---	---	---	---	0.444	0.365	0.500	0.358	0.444	0.515	0.574	0.624	0.667	---	---	---	---	---
10	---	---	---	0.250	---	---	0.400	0.325	0.444	0.540	0.400	0.467	0.523	0.571	0.613	0.650	0.682	---	---
11	---	---	---	0.318	0.309	0.394	0.403	0.364	0.495	0.400	0.488	0.561	0.427	0.481	0.527	0.568	0.604	0.636	0.665
12	---	---	0.222	---	0.200	0.333	0.381	0.375	0.333	0.450	0.545	0.444	0.513	0.571	0.622	0.490	0.529	0.565	0.596
13	---	---	---	---	0.262	0.256	0.341	0.365	0.350	0.477	0.413	0.500	0.574	0.473	0.528	0.577	0.620	0.658	0.530
14	---	---	---	0.250	0.286	0.333	0.286	0.339	0.349	0.486	0.442	0.536	0.462	0.531	0.590	0.491	0.538	0.579	0.617
15	---	---	---	0.167	0.280	0.222	0.371	0.300	0.333	0.480	0.455	0.411	0.497	0.571	0.493	0.550	0.600	0.504	0.544
16	---	---	0.167	0.219	0.250	0.250	0.286	0.391	0.444	0.462	0.455	0.427	0.519	0.464	0.533	0.594	0.515	0.563	0.605
17	---	---	0.216	0.235	0.200	0.255	0.328	0.324	0.399	0.435	0.444	0.431	0.529	0.487	0.561	0.500	0.557	0.608	0.529
18	---	0.167	---	0.222	0.267	0.241	0.349	0.375	0.346	0.400	0.424	0.426	0.530	0.500	0.578	0.528	0.588	0.525	0.573
19	---	---	0.193	0.184	0.179	0.333	0.353	0.408	0.404	0.474	0.397	0.412	0.522	0.504	0.474	0.546	0.498	0.556	0.607
20	---	---	0.133	0.125	0.200	0.283	0.343	0.312	0.444	0.420	0.473	0.500	0.508	0.500	0.480	0.556	0.518	0.578	0.526
21	---	0.143	0.175	0.167	0.200	0.333	0.320	0.321	0.365	0.467	0.429	0.468	0.487	0.490	0.479	0.560	0.529	0.593	0.549
22	---	---	---	0.182	0.291	0.258	0.286	0.318	0.384	0.400	0.479	0.432	0.462	0.474	0.473	0.557	0.535	0.505	0.565
23	---	---	0.159	0.174	0.252	0.275	0.342	0.402	0.391	0.426	0.423	0.486	0.431	0.453	0.554	0.549	0.535	0.512	0.574
24	---	0.125	0.111	0.146	0.200	0.278	0.286	0.375	0.389	0.442	0.455	0.438	0.487	0.518	0.533	0.536	0.529	0.514	0.579
25	---	---	0.147	0.200	0.232	0.267	0.314	0.340	0.378	0.448	0.476	0.473	0.446	0.486	0.509	0.520	0.520	0.511	0.579
26	---	---	0.154	0.135	0.246	0.333	0.330	0.385	0.359	0.446	0.406	0.417	0.485	0.451	0.482	0.500	0.507	0.585	0.575
27	---	0.111	0.136	0.148	0.244	0.296	0.333	0.333	0.416	0.437	0.414	0.438	0.436	0.492	0.531	0.556	0.569	0.572	0.567
28	---	---	0.095	0.232	0.229	0.250	0.327	0.357	0.381	0.421	0.416	0.452	0.462	0.526	0.495	0.527	0.546	0.556	0.556
29	---	---	0.126	0.207	0.283	0.276	0.310	0.371	0.418	0.400	0.411	0.460	0.480	0.478	0.531	0.496	0.521	0.536	0.543
30	---	0.100	0.133	0.167	0.240	0.289	0.286	0.375	0.370	0.373	0.400	0.461	0.492	0.500	0.489	0.533	0.565	0.515	0.526
31	---	---	0.118	0.194	0.265	0.290	0.327	0.371	0.391	0.413	0.455	0.457	0.499	0.516	0.514	0.496	0.533	0.559	0.576
32	---	---	0.083	0.203	0.200	0.281	0.286	0.359	0.403	0.375	0.432	0.448	0.500	0.460	0.533	0.523	0.500	0.531	0.553
33	---	0.091	0.111	0.197	0.200	0.263	0.307	0.341	0.407	0.400	0.405	0.434	0.497	0.468	0.483	0.545	0.529	0.566	0.528
CHI	0.004	0.051	0.117	0.178	0.229	0.273	0.310	0.342	0.369	0.394	0.416	0.436	0.453	0.469	0.484	0.498	0.510	0.522	0.532

Table 2. Upper bound for variance-to-mean ratio (90% confidence interval)

N	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A																			
1	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
2	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	0.947
3	---	---	---	1.333	1.400	1.444	1.476	1.500	1.519	1.533	1.545	1.556	1.564	1.571	1.578	1.583	1.588	1.593	1.596
4	---	1.750	2.000	2.125	2.200	2.250	2.286	1.750	1.778	1.800	1.273	1.292	1.308	1.321	1.333	1.344	1.353	1.361	1.368
5	---	2.600	2.867	2.000	2.120	1.733	1.800	1.850	1.889	1.920	1.945	1.967	1.554	1.571	1.587	1.600	1.612	1.622	1.632
6	2.667	3.500	2.444	2.250	2.400	1.722	1.810	1.875	1.926	1.600	1.636	1.667	1.692	1.714	1.733	1.750	1.765	1.778	1.439
7	3.571	2.714	2.810	2.000	2.200	2.000	2.102	1.857	1.603	1.657	1.701	1.738	1.769	1.796	1.514	1.536	1.555	1.571	1.586
8	4.500	3.125	2.333	2.375	2.000	1.875	2.000	1.812	1.889	1.950	1.727	1.500	1.538	1.571	1.600	1.625	1.647	1.667	1.684
9	2.778	3.000	2.185	2.333	2.067	2.000	1.889	1.750	1.840	1.911	1.727	1.778	1.581	1.619	1.652	1.681	1.471	1.494	1.515
10	3.600	3.100	2.533	2.250	2.080	2.067	2.000	1.900	1.778	1.640	1.709	1.767	1.815	1.643	1.680	1.712	1.529	1.556	1.579
11	4.455	2.818	2.515	2.364	2.055	2.091	2.065	1.795	1.909	1.800	1.678	1.742	1.601	1.649	1.691	1.727	1.567	1.596	1.622
12	3.000	3.000	2.444	2.208	2.200	1.889	1.905	1.875	1.815	1.733	1.818	1.708	1.590	1.643	1.689	1.552	1.588	1.620	1.649
13	3.769	2.846	2.538	2.231	2.108	2.051	1.923	1.923	1.889	1.831	1.755	1.667	1.734	1.626	1.677	1.558	1.597	1.632	1.502
14	4.571	3.071	2.571	2.393	2.171	2.000	1.918	1.946	1.778	1.743	1.844	1.774	1.692	1.602	1.657	1.554	1.597	1.635	1.519
15	3.267	2.600	2.556	2.167	2.200	2.089	1.895	1.950	1.815	1.800	1.764	1.711	1.646	1.714	1.631	1.683	1.588	1.630	1.526
16	4.000	3.062	2.500	2.406	2.200	2.000	2.000	1.937	1.833	1.837	1.818	1.781	1.731	1.670	1.600	1.656	1.574	1.618	1.526
17	2.882	2.882	2.569	2.294	2.176	2.039	1.941	1.912	1.837	1.859	1.727	1.706	1.670	1.622	1.690	1.625	1.554	1.601	1.520
18	3.556	3.167	2.593	2.306	2.133	2.056	2.000	1.875	1.827	1.744	1.758	1.750	1.726	1.690	1.644	1.590	1.647	1.580	1.626
19	4.263	2.737	2.579	2.421	2.200	2.053	1.917	1.947	1.807	1.863	1.775	1.781	1.656	1.632	1.596	1.664	1.613	1.556	1.604
20	3.200	3.050	2.533	2.375	2.120	2.033	1.943	1.887	1.889	1.850	1.782	1.692	1.692	1.679	1.653	1.619	1.576	1.633	1.579
21	3.857	3.000	2.587	2.310	2.143	2.000	1.952	1.929	1.847	1.829	1.779	1.706	1.718	1.714	1.698	1.673	1.639	1.598	1.551
22	4.545	3.045	2.485	2.341	2.145	2.061	1.948	1.955	1.899	1.800	1.769	1.712	1.734	1.643	1.636	1.619	1.594	1.561	1.617
23	3.522	2.913	2.594	2.348	2.130	2.101	1.932	1.870	1.841	1.765	1.751	1.710	1.742	1.665	1.667	1.658	1.639	1.614	1.581
24	4.167	2.625	2.556	2.333	2.200	2.028	2.000	1.875	1.870	1.817	1.727	1.701	1.744	1.679	1.689	1.599	1.588	1.569	1.544
25	3.240	2.920	2.600	2.300	2.152	2.040	1.960	1.870	1.889	1.768	1.785	1.773	1.738	1.686	1.619	1.625	1.621	1.609	1.589
26	3.846	2.923	2.615	2.346	2.185	2.038	2.000	1.942	1.897	1.800	1.748	1.750	1.728	1.687	1.631	1.644	1.647	1.560	1.547
27	4.481	3.111	2.605	2.278	2.200	2.025	1.942	1.917	1.815	1.822	1.788	1.722	1.712	1.683	1.617	1.657	1.588	1.588	1.581
28	3.571	2.821	2.571	2.375	2.200	2.083	1.959	1.884	1.889	1.836	1.740	1.768	1.692	1.673	1.638	1.665	1.605	1.611	1.609
29	4.172	3.138	2.609	2.362	2.186	2.046	1.966	1.922	1.874	1.841	1.765	1.730	1.743	1.660	1.634	1.668	1.617	1.628	1.559
30	3.333	2.800	2.622	2.333	2.160	2.078	1.962	1.950	1.852	1.840	1.782	1.761	1.713	1.643	1.627	1.667	1.624	1.570	1.579
31	3.903	2.935	2.613	2.371	2.200	2.097	1.949	1.895	1.896	1.832	1.792	1.715	1.680	1.691	1.684	1.661	1.626	1.581	1.594
32	4.500	3.031	2.583	2.312	2.225	2.104	2.000	1.906	1.861	1.819	1.795	1.734	1.712	1.665	1.667	1.652	1.625	1.587	1.605
33	3.667	2.818	2.535	2.318	2.164	2.101	1.970	1.909	1.889	1.800	1.793	1.747	1.737	1.701	1.646	1.640	1.620	1.589	1.549
CHI	3.841	2.996	2.605	2.372	2.214	2.099	2.010	1.938	1.880	1.831	1.789	1.752	1.720	1.692	1.666	1.644	1.623	1.604	1.586

Table 3. Lower bound for variance-to-mean ratio (90% confidence interval)

N	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A																			
1	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
2	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
3	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
4	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
5	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
6	---	---	---	---	0.400	---	---	---	---	---	---	---	---	---	---	---	---	---	---
7	---	---	---	---	---	0.333	0.469	---	---	---	---	---	---	---	---	---	---	---	---
8	---	---	---	---	---	0.417	0.286	0.406	0.500	---	---	---	---	---	---	---	---	---	---
9	---	---	---	---	---	---	0.365	0.250	0.358	0.444	0.515	0.574	---	---	---	---	---	---	---
10	---	---	---	0.250	---	---	0.400	0.325	0.444	0.320	0.400	0.467	0.523	0.571	---	---	---	---	---
11	---	---	---	---	0.309	---	---	0.364	0.293	0.400	0.488	0.364	0.427	0.481	0.527	0.568	0.604	---	---
12	---	---	0.222	---	0.200	0.333	---	0.375	0.333	0.450	0.364	0.444	0.513	0.393	0.444	0.490	0.529	0.565	0.596
13	---	---	---	---	0.262	0.256	0.341	0.365	0.350	0.308	0.413	0.500	0.408	0.473	0.528	0.577	0.457	0.496	0.530
14	---	---	---	---	---	0.167	0.286	0.339	0.349	0.329	0.442	0.381	0.462	0.531	0.438	0.491	0.538	0.579	0.466
15	---	---	---	0.167	---	0.222	0.219	0.300	0.333	0.333	0.455	0.411	0.497	0.429	0.493	0.550	0.459	0.504	0.544
16	---	---	0.167	0.219	0.250	0.250	0.286	0.250	0.306	0.325	0.318	0.427	0.385	0.464	0.400	0.461	0.515	0.563	0.474
17	---	---	---	---	0.200	0.255	0.328	0.324	0.268	0.306	0.444	0.431	0.403	0.487	0.435	0.500	0.433	0.484	0.529
18	---	---	---	---	0.133	0.241	0.222	0.250	0.346	0.400	0.424	0.426	0.410	0.381	0.459	0.528	0.471	0.525	0.456
19	---	---	---	0.184	0.179	0.211	0.233	0.289	0.287	0.358	0.397	0.412	0.409	0.391	0.474	0.434	0.498	0.444	0.496
20	---	---	0.133	0.125	0.200	0.167	0.229	0.312	0.333	0.310	0.364	0.392	0.400	0.393	0.480	0.450	0.518	0.472	0.526
21	---	---	---	0.167	0.200	0.222	0.211	0.321	0.365	0.362	0.325	0.365	0.385	0.388	0.479	0.458	0.429	0.492	0.549
22	---	---	---	0.182	0.182	0.258	0.286	0.318	0.283	0.300	0.380	0.432	0.364	0.474	0.473	0.460	0.439	0.505	0.469
23	---	---	---	0.174	0.148	0.174	0.242	0.304	0.295	0.330	0.328	0.391	0.431	0.453	0.461	0.457	0.442	0.512	0.483
24	---	---	0.111	0.146	0.200	0.181	0.286	0.281	0.296	0.350	0.364	0.347	0.397	0.429	0.444	0.448	0.441	0.514	0.491
25	---	---	---	0.100	0.136	0.267	0.223	0.250	0.289	0.360	0.389	0.387	0.446	0.400	0.424	0.435	0.435	0.511	0.495
26	---	---	---	0.135	0.154	0.244	0.242	0.298	0.274	0.362	0.406	0.417	0.402	0.451	0.482	0.418	0.507	0.504	0.494
27	---	---	---	0.148	0.156	0.210	0.249	0.250	0.333	0.356	0.333	0.358	0.436	0.413	0.452	0.477	0.490	0.494	0.489
28	---	---	0.095	0.143	0.143	0.250	0.245	0.277	0.302	0.343	0.338	0.375	0.385	0.449	0.419	0.451	0.471	0.480	0.481
29	---	---	0.126	0.121	0.200	0.195	0.232	0.293	0.341	0.324	0.335	0.385	0.406	0.404	0.457	0.422	0.448	0.464	0.470
30	---	---	---	0.083	0.160	0.211	0.286	0.300	0.296	0.373	0.327	0.389	0.421	0.429	0.418	0.463	0.494	0.515	0.526
31	---	---	0.118	0.113	0.187	0.215	0.253	0.298	0.319	0.342	0.384	0.387	0.429	0.447	0.445	0.427	0.465	0.491	0.508
32	---	---	0.083	0.125	0.200	0.208	0.286	0.289	0.333	0.306	0.364	0.380	0.433	0.393	0.467	0.457	0.434	0.465	0.487
33	---	0.091	0.111	0.121	0.200	0.192	0.238	0.273	0.340	0.333	0.339	0.369	0.431	0.403	0.418	0.481	0.465	0.502	0.464
CHI	0.001	0.025	0.072	0.121	0.166	0.206	0.241	0.273	0.300	0.325	0.347	0.367	0.385	0.402	0.417	0.432	0.445	0.457	0.469

Table 4. Upper bound for variance-to-mean ratio (90% confidence interval)

A \ N	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
2	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
3	---	---	---	---	---	1.444	1.476	1.500	1.519	1.533	1.545	1.556	1.564	1.571	1.578	1.583	1.588	1.593	1.596
4	---	---	2.000	2.125	2.200	2.250	2.286	2.312	2.333	2.350	2.364	1.833	1.846	1.857	1.867	1.344	1.353	1.361	1.368
5	---	2.600	2.867	3.000	2.120	2.200	2.257	1.850	1.889	1.920	1.945	1.967	1.985	2.000	2.013	2.025	2.035	2.044	1.632
6	---	3.500	2.444	2.667	2.400	2.500	2.571	1.875	1.926	1.967	2.000	2.028	1.692	1.714	1.733	1.750	1.765	1.778	1.789
7	3.571	2.714	2.810	2.714	2.200	2.333	2.102	2.179	2.238	1.971	1.701	1.738	1.769	1.796	1.819	1.839	1.857	1.873	1.586
8	4.500	3.500	3.000	2.687	2.600	2.458	2.000	2.094	2.167	1.950	2.000	2.042	2.077	1.839	1.867	1.625	1.647	1.667	1.684
9	5.444	3.000	3.074	2.611	2.333	2.519	2.143	2.250	2.086	1.911	1.970	2.019	1.821	1.857	1.889	1.917	1.706	1.728	1.749
10	3.600	3.100	2.800	2.750	2.320	2.300	2.229	2.125	2.000	2.080	1.927	1.767	1.815	1.857	1.893	1.925	1.741	1.767	1.789
11	4.455	3.909	3.000	2.591	2.491	2.303	2.273	2.205	2.111	2.000	2.074	1.939	1.797	1.844	1.885	1.727	1.759	1.788	1.813
12	5.333	3.250	2.889	2.625	2.400	2.472	2.095	2.062	2.000	2.100	2.000	1.889	1.949	1.821	1.867	1.729	1.765	1.796	1.649
13	3.769	3.769	3.154	2.808	2.477	2.410	2.275	2.096	2.060	2.000	1.923	2.000	1.899	1.791	1.841	1.885	1.760	1.795	1.664
14	4.571	3.500	2.762	2.750	2.514	2.333	2.245	2.107	2.095	2.057	2.000	1.929	1.846	1.908	1.810	1.857	1.748	1.786	1.669
15	5.400	3.800	2.911	2.833	2.520	2.400	2.200	2.100	2.111	2.093	1.909	1.856	1.933	1.857	1.773	1.825	1.729	1.770	1.667
16	4.000	3.250	3.000	2.719	2.500	2.292	2.286	2.219	2.111	1.975	1.955	1.917	1.865	1.804	1.867	1.789	1.838	1.750	1.789
17	4.765	3.765	3.039	2.735	2.459	2.314	2.210	2.176	2.098	1.988	1.984	1.961	1.923	1.874	1.816	1.750	1.803	1.725	1.768
18	5.556	3.500	3.037	2.722	2.533	2.315	2.254	2.125	2.074	1.989	2.000	1.991	1.846	1.810	1.881	1.826	1.765	1.698	1.743
19	4.263	3.842	3.000	2.684	2.453	2.421	2.278	2.184	2.041	1.979	2.005	1.895	1.883	1.857	1.821	1.776	1.724	1.778	1.715
20	5.000	3.650	3.067	2.750	2.480	2.383	2.286	2.112	2.111	2.070	2.000	1.908	1.908	1.893	1.867	1.831	1.788	1.739	1.684
21	5.762	3.857	3.095	2.786	2.486	2.333	2.279	2.143	2.058	2.038	1.987	1.913	1.923	1.816	1.800	1.774	1.739	1.799	1.752
22	4.545	3.591	3.091	2.682	2.473	2.379	2.260	2.159	2.101	2.000	1.967	1.909	1.930	1.838	1.830	1.812	1.786	1.753	1.713
23	5.261	3.435	3.058	2.783	2.548	2.406	2.230	2.163	2.130	2.052	1.941	1.899	1.930	1.851	1.852	1.842	1.731	1.797	1.764
24	4.167	3.500	3.000	2.750	2.500	2.417	2.286	2.156	2.056	2.000	2.000	1.972	1.923	1.857	1.867	1.776	1.765	1.745	1.719
25	4.840	3.640	3.027	2.700	2.536	2.413	2.234	2.140	2.067	2.032	1.960	1.947	1.911	1.857	1.789	1.795	1.791	1.778	1.758
26	5.538	3.500	3.026	2.731	2.554	2.397	2.264	2.202	2.068	2.054	2.000	1.917	1.893	1.852	1.795	1.808	1.810	1.722	1.709
27	4.481	3.444	3.000	2.741	2.556	2.370	2.280	2.167	2.062	2.067	1.949	1.963	1.872	1.841	1.795	1.815	1.745	1.745	1.737
28	5.143	3.679	3.143	2.732	2.543	2.333	2.286	2.205	2.127	2.071	1.974	1.923	1.923	1.827	1.867	1.817	1.756	1.762	1.759
29	4.172	3.552	3.069	2.707	2.517	2.368	2.281	2.155	2.103	2.069	1.991	1.954	1.891	1.882	1.855	1.815	1.763	1.774	1.704
30	4.800	3.700	3.067	2.750	2.560	2.389	2.267	2.175	2.074	2.060	2.000	1.906	1.928	1.857	1.840	1.808	1.765	1.781	1.719
31	5.452	3.613	3.129	2.694	2.510	2.398	2.244	2.185	2.111	2.045	2.003	1.925	1.888	1.829	1.822	1.798	1.763	1.785	1.730
32	4.500	3.781	3.083	2.703	2.525	2.396	2.286	2.187	2.069	2.025	2.000	1.937	1.913	1.866	1.800	1.785	1.757	1.785	1.737
33	5.121	3.545	3.020	2.773	2.527	2.384	2.247	2.182	2.091	2.067	1.992	1.944	1.867	1.831	1.840	1.833	1.749	1.781	1.740
CHI	5.024	3.689	3.116	2.786	2.567	2.408	2.288	2.192	2.114	2.048	1.993	1.945	1.903	1.866	1.833	1.803	1.771	1.751	1.729

responding approximate confidence interval, despite the apparently good level of agreement for A and N being equal or greater than 15 and 6, respectively. This indicates that the exact test should be used whenever practicable (i.e. for N less than 21 and/or A less than 34). For larger values of A and N, the discrepancy between the exact and approximate tests is very small (see Fig. 4), and the approximate test may be used.

CONCLUSION

Because road layout and traffic flows do not remain fairly constant for long periods, say, accident count analysis frequently entails using data for a small number of years, during which a small number of accidents have occurred. However, the approximate, asymptotic test is appropriate only when there has been a large number of accidents during a large number of years. Hence the exact statistical test, based upon a combinatorial analysis, will generally be the more appropriate test. The exact test can readily be applied, using the tables which have been derived.

Although it might appear that there is a good agreement between the approximate and exact tests when the number of accidents and years exceed about 15 and 6, respectively, the discrepancy between the tests can give rise to serious error. This has been demonstrated by the result of reinterpreting data that, when subjected to the approximate test indicated that

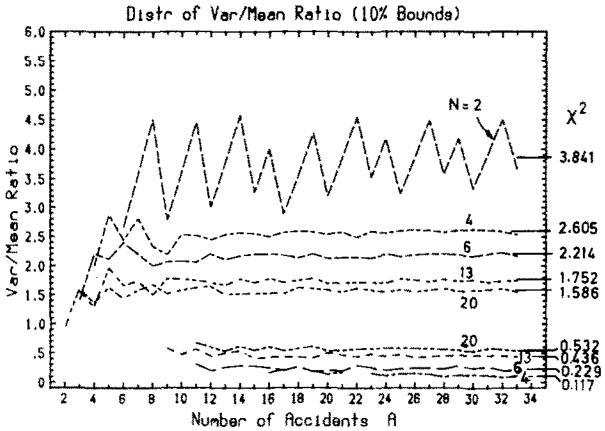

Fig. 2. Convergence characteristics for varying N and A.

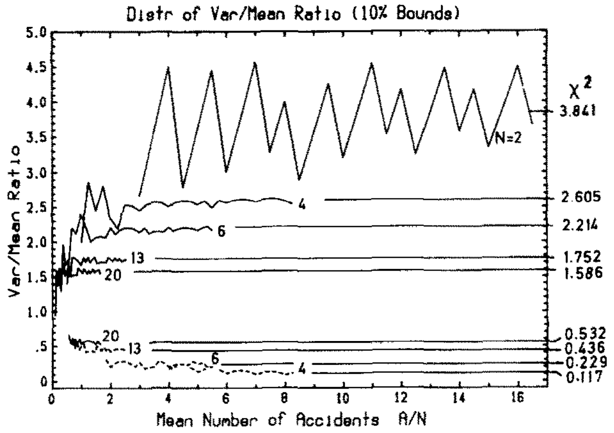

Fig. 3. Convergence characteristics for varying N and A/N.

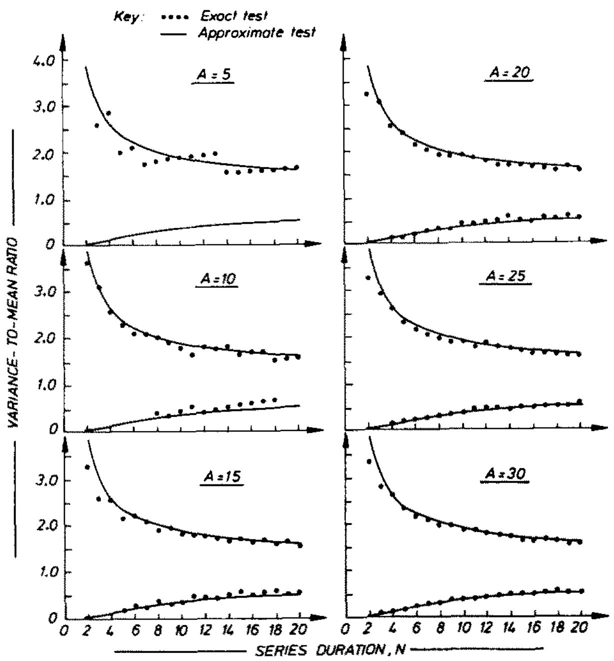

Fig. 4. comparison of exact and approximate tests.

the Poisson distribution is not universally appropriate, but when subjected to the exact test indicate that there is little (if any) justification for rejecting the Poisson distribution.

It therefore seems reasonable that the analysis of annual road accident counts for individual sites should generally be done on the basis that they are Poisson distributed. It is a simple task to check the assumption, using the tables given in this paper, and this should be done whenever the annual accident counts for the location being investigated seem to exhibit an unusually large or small variance-to-mean ratio.

Acknowledgements—The research described in this paper was financially supported by the Road Research Unit of the New Zealand National Roads Board. The comments of Siem Oppe on an earlier version of the paper are gratefully acknowledged.

REFERENCES

Abramowitz, M.; Stegun, I. A. Handbook of mathematical functions (Ch. 24). New York: Dover; 1964.

Feller, W. An introduction to probability theory and its applications (Vol. 1, 3rd Ed., p. 38). Wiley, 1968.

Fisher, R. A. The significance of deviations from expectation in a Poisson series. Biometrics 6:17–24; 1950.

Greenwood, M.; Yule, G. U. An inquiry into the nature of frequency distributions. Journal of Royal Statistical Society, A, 83:255–279; 1920.

Nicholson, A. J. The variability of accident counts. Accid. Anal. Prev. 17:47–56; 1985.

Nicholson, A. J. The randomness of accident counts. Accid. Anal. Prev. 18:193–198; 1986.

Rao, C. R.; Chakravarti, I. M. Some small sample tests of significance for a Poisson distribution. Biometrics 12:264–282; 1956.

Sukhatme, P. V. On the distribution of chi-square in samples of the Poisson process. Supplement to Journal of Royal Statistical Society 5(2):75–79; 1938.