Osaka Mathematical Journal
Vol. 7, No. 1, June, 1955.

Fit of a Poisson Distribution by the Index of Dispersion

By Masashi OKAMOTO

1. Introduction. To fit a Poisson distribution to a series of small samples we sometimes calculate for each sample $(x_1, \cdots, x_n)$ a statistic, the so-called index of dispersion,

$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\bar{x}}, \tag{1}$$

where $\bar{x}$ stands for the sample mean (cf. R. A. Fisher [1], p. 58). It is easily verified that, if the sample comes in fact from a Poisson distribution, the statistic (1) follows asymptotically a χ^2 distribution with $n-1$ degrees of freedom when the location parameter of the Poisson distribution is sufficiently large. The deviation of the true distribution from the asymptotic one has been investigated by P. V. Sukhatme [2] experimentally. Aiming at clarifying the matter further, this paper gives exact formulas for some low moments of (1), while P. G. Hoel [3] gives expanding forms of them. In the course of evaluating them there arises a necessity to consider negative moments of a positive Poisson variate. These moments may be calculated in the same way that F. F. Stephan [4] proposed concerning negative moments of a positive Bernoulli variate. Recently the first negative moment has been tabulated by E. L. Grab and I. R. Savage [5] for some values of the parameter of the Poisson distribution.

2. Moments of χ^2. We shall first state two well-known lemmas which are required later in calculating moments of (1).

Lemma 1. *Let random variables x_i, $i = 1, \cdots, n$, be distributed independently according to a Poisson distribution with the location parameter λ_i, respectively. Then the joint conditional distribution of $x_1, \cdots, x_n$ given $\sum_{i=1}^{n} x_i = X$ (const.) is the multinomial distribution with probabilities $\{\lambda_i (\sum_{i=1}^{n} \lambda_i)^{-1}, i = 1, \cdots, n\}$ and the total number X of repetitions.*

Lemma 2. *If $(x_1, \cdots, x_n)$ follows the multinomial distribution with probabilities $(p_1, \cdots, p_n)$ and the total number X of repetitions, then*

$$E(x_1^{[r_1]}\cdots x_n^{[r_n]})=X^{[r_1+\cdots+r_n]}p_1^{r_1}\cdots p_n^{r_n},$$

where

$$\begin{aligned} x^{[r]} &= x(x-1)\cdots(x-r+1) \qquad &for \quad r>0, \\ &=1 &for \quad r=0. \end{aligned}$$

Hence it follows that

$$\begin{aligned} &E(x_i)=Xp_i, \\ &E(x_i-Xp_i)^2=Xp_i(1-p_i) \\ &E(x_i-Xp_i)^4=3X^2p_i^2q_i^2+Xp_iq_i(1-6p_iq_i) \\ &E(x_i-Xp_i)^2(x_j-Xp_j)^2=X^2p_ip_j(1-p_i-p_j+3p_ip_j) \\ &\qquad\qquad\qquad\qquad -Xp_ip_j(1-2p_i-2p_j+6p_ip_j) \end{aligned}$$

$i, j=1, \cdots, n,\ i\neq j,$ *where* $q_i=1-p_i$.

Now we must in advance decide how to dispose of the case when all x_i's happen to be zeroes. Since (1) is then indeterminate, we may put it equal to zero or any other value. On the other hand we may exclude that case and consider the conditional distribution given $\bar{x}$ positive. Though it is certain that Sukhatme met with such an instance in his large-scale experiment, it is not stated explicitly in [2] which of these alternatives he accepted. It seems to the author that the last alternative is most suitable not only from the practical point of view, but from the mathematical one, because, as is seen later, the expressions for moments of (1) become simplest under this convention.

Theorem. *If a random sample comes from a Poisson distribution with the parameter* λ, *then under the condition* $\bar{x}>0$ *it holds that*

$$(2)\qquad \begin{aligned} E(\chi^2) &= n-1, \\ V(\chi^2) &= 2(n-1)[1-f(n\lambda)], \end{aligned}$$

where

$$(3)\qquad f(a)=\frac{1}{e^a-1}\sum_{i=1}^{\infty}\frac{a^i}{i!\,i}.$$

Proof. We shall denote by E and E' the expectation with respect to $X=\sum_{i=1}^n x_i$ under the condition $X>0$ and that with respect to the conditional distribution of $(x_1, \cdots, x_n)$ given X, respectively. Though the symbol E is used in two different meanings, there will be no ambiguity. Since by Lemma 1 the conditional distribution of x_i given X is binomial with the probability $1/n$ and the repetition number X, it follows from Lemma 2

$$E'(x_i) = X/n\,,$$
$$E'(x_i - X/n)^2 = Xn^{-1}(1-n^{-1})\,, \qquad i = 1, \cdots, n\,.$$

Then

$$E(\chi^2) = E[E'(\chi^2)] = E\left[\frac{n}{X}\sum_{i=1}^{n} E'\left(x_i - \frac{X}{n}\right)^2\right]$$
$$= E(n-1) = n-1\,.$$

With regard to the variance we have first

$$E(\chi^2)^2 = E[E'(\chi^2)^2]$$

and

$$E'(\chi^2)^2 = \frac{n^2}{X^2}\left[\sum_{i=1}^{n} E'\left(x_i - \frac{X}{n}\right)^4 + \sum_{i \neq j}\sum E'\left(x_i - \frac{X}{n}\right)^2\left(x_j - \frac{X}{n}\right)^2\right].$$

Lemmas 1 and 2 imply

$$E'\left(x_i - \frac{X}{n}\right)^4 = 3X^2\frac{1}{n^2}\left(1-\frac{1}{n^2}\right) + X\frac{1}{n}\left(1-\frac{1}{n}\right)\left[1-\frac{6}{n}\left(1-\frac{1}{n}\right)\right],$$
$$E'\left(x_i - \frac{X}{n}\right)^2\left(x_j - \frac{X}{n}\right)^2 = X^2\frac{1}{n^2}\left(1-\frac{2}{n}+\frac{3}{n^2}\right) - X\frac{1}{n^2}\left(1-\frac{4}{n}+\frac{6}{n^2}\right).$$

Substituting these into the last equation, we have

$$E'(\chi^2)^2 = (n-1)\left(n+1-\frac{2}{X}\right),$$

whence

$$E(\chi^2)^2 = (n-1)\left[n+1-2E\left(\frac{1}{X}\right)\right].$$

Thus it follows that

$$V(\chi^2) = E(\chi^2)^2 - [E(\chi^2)]^2 = 2(n-1)\left[1-E\left(\frac{1}{X}\right)\right].$$

Since, being the sum of n independent Poisson variates, X follows itself a Poisson distribution with the parameter $n\lambda$, and since E denotes the expectation under the condition $X>0$, it holds that

$$E\left(\frac{1}{X}\right) = \frac{1}{1-e^{-n\lambda}}\sum_{i=1}^{\infty} e^{-n\lambda}\frac{(n\lambda)^i}{i\,!\,i} = f(n\lambda)\,,$$

where the function f is defined by (3). The proof is now complete.

The determination of first two moments of the statistic χ^2 is thus not involved, while the third or the fourth moment is somewhat difficult to calculate. We shall therefore give the result without proof:

$$
(4) \quad \begin{aligned}
\mu_3(\chi^2) &= 8(n-1)+4(n-1)(n-8)E(X^{-1})-4(n-1)(n-6)E(X^{-2}), \\
\mu_4(\chi^2) &= 12(n-1)(n+3)+24(n-1)(3n-19)E(X^{-1}) \\
&\quad +4(n-1)(2n^2-81n+285)E(X^{-2})-8(n-1)(n^2-30n+90)E(X^{-3}).
\end{aligned}
$$

3. Negative moments of a positive Poisson distribution. If a Poisson variate is subject to the condition excluding the value zero, it will be designated the positive Poisson variate, as was done by Stephan [4] for a Bernoulli variate. This definition was given also by Grab and Savage [5]. As is seen in equations (2) and (4), the distribution of χ^2 is dependent on negative moments of a positive Poisson variate. These moments tend to zero as the location parameter tends to infinity, so that moments of any order of χ^2 tend to the corresponding moments of the χ^2 distribution with $n-1$ degrees of freedom. Thus they give the extent of the deviation of the true distribution of χ^2 from the approximating χ^2 distribution. Among them, however, the first negative moment $E(X^{-1})$ is most important, because it alone appears in the expression of the variance which is most important of all moments except the mean. (The mean is identically equal to $n-1$ and needs no consideration.) A table of $E(X^{-1})$ for some values of the parameter was given by Grab and Savage [5]. The author performed some computations independently of them for a range 1 (1) 50 (5) 125. Values common to two computations coincide completely. Though Grab and Savage used the defining equation (3) for computing $f(a)$, it will be convenient for large values of a to use the factorial series

$$
\frac{1}{x} = \sum_{i=1}^{t} \frac{(i-1)!\,x!}{(x+i)!} + \frac{t!\,(x-1)!}{(x+t)!},
$$

and corresponding

$$
(5) \qquad f(a) = \sum_{i=1}^{t} \left[\sum_{x=1}^{\infty} \frac{(i-1)!\,x!}{(x+i)!} P(x) \right] + R_t,
$$

where

$$
P(x) = \frac{1}{e^a-1} \cdot \frac{a^x}{x!},
$$

$$
R_t = \sum_{x=1}^{\infty} \frac{t!\,(x-1)!}{(x+t)!} P(x).
$$

The series (5) is perhaps preferable for n larger than or equal to 15, 20, 25 in order to obtain 5, 7, 9 significant figures of $f(a)$, respectively.

Table 1

a	$f(a)$	a	$f(a)$
1	0.76698 83544	36	0.02859 62855
2	0.57659 08853	37	0.02780 05753
3	0.43268 39036	38	0.02704 79867
4	0.32962 63851	39	0.02633 51039
5	0.25776 95370	40	0.02565 88628
6	0.20779 02684	41	0.02501 65069
7	0.17248 62160	42	0.02440 55496
8	0.14688 90650	43	0.02382 37423
9	0.12775 77299	44	0.02326 90458
10	0.11302 14089	45	0.02273 96073
11	0.10135 48155	46	0.02223 37388
12	0.09189 62957	47	0.02174 98999
13	0.08407 21168	48	0.02128 66813
14	0.07748 96415	49	0.02084 27918
15	0.07187 25576	50	0.02041 70456
16	0.06702 11916	55	0.01852 51260
17	0.06278 77256	60	0.01695 42004
18	0.05906 03526	65	0.01562 89430
19	0.05575 28883	70	0.01449 58921
20	0.05279 77880	75	0.01351 60523
21	0.05014 13367	80	0.01266 03106
22	0.04774 02591	85	0.01190 64915
23	0.04555 92941	90	0.01123 74071
24	0.04356 94087	95	0.01063 95288
25	0.04174 64774	100	0.01010 20625
26	0.04007 02838	105	0.00961 62915
27	0.03852 37570	110	0.00917 50989
28	0.03709 23814	115	0.00877 26171
29	0.03576 37344	120	0.00840 39651
30	0.03452 71218	125	0.00806 50494
31	0.03337 32863		
32	0.03229 41739		
33	0.03128 27441		
34	0.03033 28153		
35	0.02943 89370		

4. Comments on Sukhatme's experiment. P. V. Sukhatme's data [2] consist of five tables. Table I represents five samples each of 400 X^2's which are calculated from samples of size $n=5$ taken randomly from Poisson populations with the parameter $\lambda=1$, 2, 3, 4 and In Tables II and III X^2's are calculated from samples of $n=10$, and 15, respectively, instead of 5 in Table I above. Tables IV and V are replicates of Tables II and III, respectively, using new materials, while the latter two make use of samples used already in Table I. Each table gives the frequency distribution of 400 X^2's in contrast with the X^2 distribution of $n-1$ degrees of freedom, whereby the fit is tested by the usual X^2 method. The fit is good except for the first and the second samples in Table I, the first and the fourth in Table II, and the first and the second in Table IV. The first samples in Tables I and II show especially remarkable discrepancy.

Here arise two problems: First, is any of these samples of bad fit not to be considered exceptional as a random sample notwithstanding the fact that it was taken randomly from a Poisson population? Second, is the approximation by the X^2 distribution independent of the sample size n, as Sukhatme asserts?

To give an answer to the first question we calculated the sample means and variances for Sukhatme's data and compared them with the theoretical ones (2). Results are shown in Table II, whence we see that the data conform to (2) very well. This suggests that for small n with small λ as much discrepancy will be inevitable.

Table 2

n	λ	Significance level of fit	Sample mean	Expected mean	Sample variance	Expected variance
5	1	0.00000	3.934	9	5.745	5.938
	2	0.0548	4.071	9	6.424	7.096
10	1 (Sukhatme's Table II)	0.0047	8.897	9	14.761	15.966
	4 (ibid)	0.0540	9.499	9	17.518	17.538
	1 (Sukhatme's Table IV)	0.0416	9.147	9	15.958	15.966
	2 (ibid)	0.0636	8.635	9	13.640	17.050

As to the second problem there may be controversy. In so far as the degree of approximation of a distribution by another is not defined, the decisive answer will be impossible. Keeping, however, the definition intact, we shall rely upon low moments, i.e., the mean and the variance. The means coincide for the true and the theoretical distributions, while the ratio of variances is smaller than unity by a quantity depending only on n, so that the approximation may depend only on n at least in the first approximation. This contradicts Sukhatme's assertion that the approximation is independent of n. His experiment, however, seems to favour us.

(Received March 10, 1955)

References

[1] R. A. Fisher, Statistical Methods for Research Workers, 11th ed. 1951.
[2] P. V. Sukhatme, On the distribution of χ^2 in samples of the Poisson series, Jour. Roy. Stat. Soc., Suppl. **5** (1938), 75–79.
[3] P. G. Hoel, On indices of dispersion, Ann. Math. Stat., **14** (1943), 155–162.
[4] F. F. Stephan, The expected value and variance of the reciprocal and other negative powers of a positive Bernoullian variate, Ann. Math. Stat., **16** (1945), 50–61.
[5] E. L. Grab and I. R. Savage, Tables of the expected value of 1/X for positive Bernoulli and Poisson variables, Jour. Am. Stat. Assoc., **49** (1954), 169–177.