Yonsei Medical Journal
Vol. 26, No. 2, 1985

An Application of Discrete Probability Models to the Number of Outpatient Visits

Han Joong Kim, M.D., Dong Kee Kim, M.A., and Il Soon Kim, M.D.

It is hypothesized that the number of outpatient visits can be represented by three different probability models: the truncated Poisson distribution, the Zeta distribution and the logarithmic series distribution. Maximum likelihood estimates of parameters of the above distributions were obtained by using grouped data according to the number of visits. A goodness-of-fit test was also made to compare the fit of the three distributions, and the value of this statistic was classified and compared according to the types of medical care facilities. Based on the likelihood ratio statistic as a test criterion, both the truncated Poisson and Zeta distributions were not appropriate for the model of the number of outpatient visits. However, the logarithmic series distribution provides a good fit to data in the case of university hospitals, general hospitals and hospitals. When we apply this distribution in the 10 most common diseases, the estimates of the parameter vary from 0.39567 to 0.54176 for university hospitals, from 0.45329 to 0.65387 for general hospitals, and from 0.55104 to 0.77625 for hospitals. On the other hand, in the case of clinics, even the logarithmic series distribution cannot be fitted to the data well. A characteristic of clinic utilization with almost homogeneous treatment patterns, in spite of the fact that there are a great many clinics, could be the reason for the above results.

Key Words: Health utilization service, the number of outpatient visits, maximum likelihood estimation, the test of goodness-of-fit, discrete probability distributions, biometry, biostatistics.

Since the beginning of health insurance in 1977, health utilization service research regarding health insurance has been dramatically advanced and analyzed in many aspects. Studies concerned with developing medical fee schedules and reviewing health insurance claim payments should be analyzed by using some advanced statistical methods.

Kim *et al.*, (1984) indicated that the outpatient fee was influenced by such variables as classification of diseases, types of medical care facilities, first visit or not, the number of visits, and the age of patients. The above variables, excluding the number of outpatient visits, have been analyzed through health utilization services research. Also, more theoretical aspect of the number of visits is indispensable for further research.

When a person suffers from a certain disease, he visits a doctor. The number of visits of the patient suffering from a certain disease may vary due to its severity as well as to other variables. The number of visits is not large in a given period and follows the discrete distribution with probabilities specified on positive integers.

Fig. 1 represents the histogram of the number of outpatient visits in the case of acute nasopharyngitis and acute upper respiratory infections of multiple or unspecified sites in general hospitals. It was found that the distribution of the number of visits cannot be determined on 0 visits, has its mode at 1 visit, and decreases rapidly after 2 or more visits. Among several kinds of discrete distributions which satisfy properties of the number of outpatient visits, the following three are to be introduced.

First, we can use the truncated Poisson distribution for the probability model of the number of outpatient visits. Samples from distributions of the number of persons per residence suffering from a contagious disease and of the number of accidents per worker in a factory during specified intervals of time are of this type. The most common form of truncation is the omission of the zero class because the observational apparatus becomes active when only one event occurs. Such samples have been considered by David and Johnson (1952), Cohen (1960) and Selvin (1974). The corresponding truncated Poisson is

Department of Preventive Medicine and Public Health Yonsei University College of Medicine Seoul, Korea

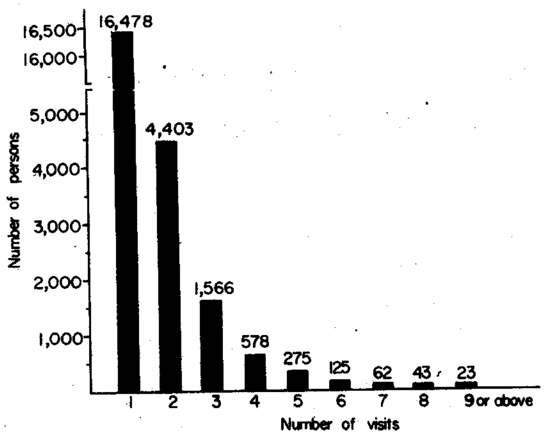


Fig. 1. Histogram of the number of outpatient visits in the case of acute nasopharyngitis and acute URI in general hospitals.

$$P_r(X=k) = \frac{\lambda^k (e^{\lambda}-1)^{-1}}{k!} \quad (k=1, 2, \ldots; \lambda>0)$$

Second, the Zeta distribution can be applicable. It is familiar in a variety of empirical areas, including linguistics, personal income distribution, and the distribution of biological genera and species. Earlier work of interest in this connection was done by McNeil (1973) and Hill (1974). Such distributions are defined by the following equations

$$P_r(X=k) = c \cdot k^{-(\rho+1)} \quad (k=1, 2, \ldots)$$

$$\text{with} \quad c = [\sum_{i=1}^{\infty} k^{-(\rho+1)}]^{-1} = [\zeta(\rho+1)]^{-1}$$

where $\zeta(\cdot)$ denotes the Riemann-Zeta function.

Third, the logarithmic series distribution can be applied to the number of outpatient visits. It was applied to the results of sampling butterflies (Corbet's data), to numbers of moths of different species caught in a light-trap over a given period of time (William's data), and to numbers of bacteria per colony. The history and various applications are summarized by Williamson and Bretherton (1964), Patil and Bildikar (1967). It is found that this distribution is defined by

$$P_r(X=k) = \alpha \cdot \theta^k / k \quad (k=1, 2, \ldots; 0<\theta<1)$$

$$\text{where} \quad \alpha = [-\log(1-\theta)]^{-1}$$

In this study these three discrete distributions were employed in order to represent the probability models of the number of outpatient visits by types of medical care facilities* and to compare the results of testing the models' hypotheses. Also, the appropriate distributions in the case of the 10 most common diseases were applied.

* Medical care facilities are classified into the three categories: general hospitals, hospitals, and clinics according to the Korean medical law. A general hospital should have at least 8 departments including internal medicine, surgery, ob-gy, pediatrics, radioloogy, anesthesiology, clinical pathology, and dentistry (with specialists in each department) and should be equipped with over 80 beds with sufficient doctors. A clinic is classified as a doctor's office. Besides the above classifications, university hospitals affiliated with medical schools are usually subclassified under general hospitals because of their sophisticated characteristics. Thus, we classified medical care facilities into the above four categories.

METHODS

Data

For this study 1.9 million pieces of data claimed and stored from March 1984 through May 1984 in the computer system at the Korea Medical Insurance Corporation were selected to be observed and analyzed.

Estimation of the Parameter

In order to fit these distributions to the observed data, the maximum likelihood estimation procedure was provided for the estimates of the parameter. It was shown that the sample mean, $\bar{x}$, is a sufficient statistic and the maximum likelihood estimate of the population mean. Thus, the estimate of the parameter can be obtained from the sample mean.

Admittedly, maximum likelihood estimates are troublesome to calculate without proper tables since it is necessary to solve a somewhat complicated non-linear equation. Tables and charts studied and presented by Cohen(1960) for the truncated Poisson, Johnson and Kotz(1969), McNeil(1973) for the Zeta distribution, Barton *et al.*, (1963), Williamson and Bretherton (1964) for the logarithmic series distribution can reduce the time-consuming calculation.

1. The truncated Poisson distribution: $\bar{X} = \dfrac{\lambda}{1-e^{-\lambda}}$

   $\bar{x}$; sample mean

2. Zeta distribution: $\dfrac{\sum_{i=1}^{n} \log X_i}{n} = \zeta'(\rho+1)/\zeta(\rho+1)$

$\zeta(\cdot)$: Riemann-Zeta function
$\zeta'(\cdot)$: The first derivative of Riemann-Zeta function

3. The logarithmic series distribution:

$$\bar{X} = \frac{\theta}{-(1-\theta)\log(1-\theta)}$$

Tests of the Model

The compatibility of a set of observed sample values with some distribution can be checked by a goodness-of-fit type of test, where the null and alternative hypotheses are

$$H_0 : F_X(X) = F_O(X) \text{ for all } X$$
$$H_1 : F_X(X) = F_O(X) \text{ for some } X$$

Assuming that the population distribution is completely specified by the null hypothesis, one can calculate the probability that a random observation will be classified into each of the chosen or fixed categories. Since the null hypothesis was assumed to specify the population distribution completely, this hypothesis then is actually concerned only with the values of these parameters and can be equivalently stated as

$$H_0 : \theta_i = \frac{e_i}{n},\ i=1, 2, \ldots, k$$

Here, e_i $(i=1, 2, \ldots, k)$ are expected frequencies.

The likelihood-ratio statistic whose distribution approximates the chi-square distribution with(k-1) degree of freedom can be used as a test criterion for the goodness-of-fit.

$$-2\sum_{i=1}^{k} f_i\left(\log\theta_i - \log\frac{f_i}{n}\right) \sim \chi^2_{(k-1)}$$

where f_i $(i=1, 2, \ldots k)$ are observed frequencies.

RESULTS

The authors considered a sample of the number of outpatient visits with a specific disease classified according to the type of medical care facilities. Table 1 shows the observed and expected frequencies of three probability models in the case of contact dermatitis and other eczema patients in university hospitals. Also, Table 2, Table 3, and Table 4 show these frequencies in the case of acute nasopharyngitis and acute URI patients in general hospitals, other noninfective gastroenteritis and colitis patients in hospitals, and gastritis and duodenitis patients in clinics. The results of testing the goodness-of-fit are shown in these tables.

Based on these results, we reject the truncated Poisson and Zeta distributions for providing an adequate probability model of the number of outpatient visits in the case of university hospitals, general hospitals, hospitals, and clinics. On the other hand, the logarithmic series distribution provides a good fit to the data in the case of university hospitals, general hospitals, and hospitals based on the chi-square test statistic at 5% level of significance. However, in the case of clinics, the logarithmic series cannot provide a good fit since the distribution has a comparatively good fit only up to 7 visits; but from 8 or more, the expected probability cannot reach the observed probability.

Table 5 shows the mean number of outpatient visits by types of medical care facilities. It shows that

Table 1. Frequency distribution of the number of outpatient visits in the case of contact dermatitis and other eozema in university hospitals

Number of visits	Observed frequency	Expected frequency		
		T.Poisson	Zeta	Log-series
1	1,595 (78.49)	1,508.15(74.22)	1,660.90(81.74)	1,585.77(78.04)
2	307 (15.11)	429.36(21.13)	222.71(10.96)	320.65(15.78)
3	90 (4.43)	81.48(4.01)	68.68(3.38)	86.56(4.26)
4	22 (1.08)	11.58(0.57)	29.87(1.47)	26.21(1.29)
5	12 (0.59)	1.22(0.06)	16.87(0.77)	8.53(0.42)
6	4 (0.20)	0.20(0.01)	9.14(0.45)	2.81(0.14)
7 or above	2 (0.10)	0.02(0.00)	18.29(0.90)	1.42(0.07)
Total	2,032 (100.00)	2,032(100.00)	2,032(100.00)	2,032(100.00)
	$\bar{x}=1.31$	$\hat{\lambda}=0.56932$	$\hat{\rho}=1.8986$	$\hat{\theta}=0.40447$
	S.D=1.00	$\hat{\theta}=115.82$	$\chi^2=81.20$	$\chi^2=3.37$
	d.f=5	$p<0.0001$	$p<0.0001$	$p>0.05$

Table 2. Frequency distribution of the number of outpatient visits in the case of acute nasopharyngists and acute upper respiratory infections of multiple or unspecified sites in general hospitals

Number of visits	Observed frequency	Expected frequency		
		T.Poisson	Zeta	Log-series
1	16,478 (69.96)	14,692.36(62.38)	17,751.90(75.37)	16,449.42(69.84)
2	4,403 (18.69)	6,462.94(27.44)	3,054.82(12.97)	4,404.41(18.70)
3	1,566 (6.65)	1,896.02(8.05)	1,102.82(4.68)	1,570.99(6.67)
4	578 (2.45)	416.89(1.77)	525.23(2.23)	631.22(2.68)
5	275 (1.17)	73.01(0.31)	299.12(1.27)	270.86(1.15)
6	125 (0.53)	11.78(0.05)	188.42(0.80)	120.12(0.51)
7	62 (0.26)	2.36(0.01)	127.19(0.54)	56.53(0.24)
8	43 (0.18)	0.24(0.00)	89.50(0.38)	25.91(0.11)
9 or above	23 (0.10)	0.24(0.00)	423.95(1.80)	23.55(0.10)
Total	23,553 (100.00)	23,553(100.00)	23,553(100.00)	23,553(100.00)
	$\bar{x}$=1.51	$\hat{\lambda}$=0.87989	$\hat{\rho}$=1.53909	$\hat{\theta}$=0.53543
	S.D=1.00	χ^2=2563.08	χ^2=1540.47	χ^2=14.86
	d.f=7	p<0.0001	p<0.0001	p>0.05

Table 3. Frequency distribution of the number of outpatient visits in the case of other noninfective gastroerteritis and colitis in hospitals

Number of visits	Observed frequency	Expected frequency		
		T.Poisson	Zeta	Log-series
1	2,209 (70.24)	1,908.07(60.67)	2,354.03(74.85)	2,167.07(68.81)
2	546 (17.36)	885.32(28.15)	412.00(13.10)	596.29(18.96)
3	204 (6.49)	273.93(8.71)	148.76(4.73)	218.89(6.96)
4	88 (2.80)	63.53(2.02)	72.02(2.29)	90.58(2.88)
5	54 (1.72)	11.96(6.38)	41.20(1.31)	39.94(1.27)
6	25 (0.79)	1.89(0.06)	26.10(0.83)	18.24(0.58)
7	10 (0.31)	0.31(0.01)	17.61(0.56)	8.81(0.28)
8 or above	9 (0.28)	0.03(0.00)	72.34(2.30)	8.18(0.26)
Total	3,145 (100.00)	3,145(100.00)	3,145(100.00)	3,145(100.00)
	$\bar{x}$=1.53	$\hat{\lambda}$=0.92809	$\hat{\rho}$=1.5139	$\hat{\theta}$=0.55104
	S.D=1.41	χ^2=519.38	χ^2=168.92	χ^2=13.33
	d.f=6	p<0.0001	p<0.0001	p>0.05

the mean number of visits is the largest in clinics, which means that the distribution has a long tail. The logarithmic series distribution cannot explain this point because of its inability to adequately handle the tail of distribution.

When the logarithmic series distribution is applied in over four diseases, the same results can be obtained as shown in Table 6. Results indicate the logarithmic series distribution is found to provide a good fit to data in the case of university hospitals, general hospitals, and hospitals, but in the case of clinics, even the logarithmic series distribution cannot be fitted to the data very well.

When we applied this distribution in the 10 most common diseases, the estimates of the parameter varied from 0.39567 to 0.54176 (the mean number

Table 4. Frequency distribution of the number of outpatient visits in the case of gastritis and duodenitis in clinics

Number of visits	Observed frequency	Expected frequency		
		T.Poisson	Zeta	Log-series
1	33,803 (61.52)	26,635.47(48.27)	37,810.43(68.78)	13,621.49(61.16)
2	10,784 (19.63)	17,442.93(31.73)	7,894.12(14.36)	11,099.65(20.19)
3	4,696 (8.55)	7,646.74(13.91)	3,155.45(5.74)	4,887.10(8.89)
4	2,374 (4.32)	2,512.27(4.57)	1,649.19(3.00)	2,418.81(4.40)
5	1,240 (2.26)	659.68(1.20)	995.01(1.81)	1,275.37(2.32)
6	740 (1.35)	142.93(0.20)	659.68(1.20)	703.65(1.28)
7	429 (0.78)	27.49(0.05)	467.27(0.85)	401.30(0.73)
8	318 (0.58)	5.50(0.01)	346.33(0.63)	230.89(0.42)
9	200 (0.36)	0:55(0.00)	263.87(0.48)	137.43(0.25)
10	142 (0.36)	0.55(0.00)	1,704.16(3.10)	82.46(0.15)
11 or above	247 (0.66)	0.55(0.66)		126.44(0.23)
Total	54,973 (100.00)	54,973(100.00)	54,973(100.00)	54,973(100.00)
	$\bar{x}$=1.80	$\hat{\lambda}$=1.3147	$\hat{\rho}$=1.2604	$\hat{\theta}$=0.6593
	S.D=1.73	χ^2=17036.07	χ^2=3945.68	χ^2=180.40
	d.f= 9	p<0.0001	p<0.0001	p<0.0001

Table 5. Mean number of outpatient visits by the type of medical care facilities

Types	Number of observations	Mean number of visits (S.D.)
University hospitals	89,598	1.46(1.41)
General hospitals	209,019	1.65(2.23)
Hospitals	132,900	1.90(2.23)
Clinics	1,452,604	2.38(2.64)
Total	1,884,121	2.20(2.23)

of visits goes from 1.30 to 1.52) for university hospitals; from 0.45329 to 0.65387 (the mean number of visits goes from 1.38 to 1.77) for general hospitals; from 0.55104 to 0.77625 (the mean number of visits goes from 1.53 to 2.35) for hospitals (Table 7). This finding might be applicable to health utilization services research; for example, outpatient administration, etc.

DISCUSSION

When the three probability distributions are applied to the number of outpatient visits, the truncated

Table 6. Estimates and test statistics of the four common diseases used by the logarithmic series distribution

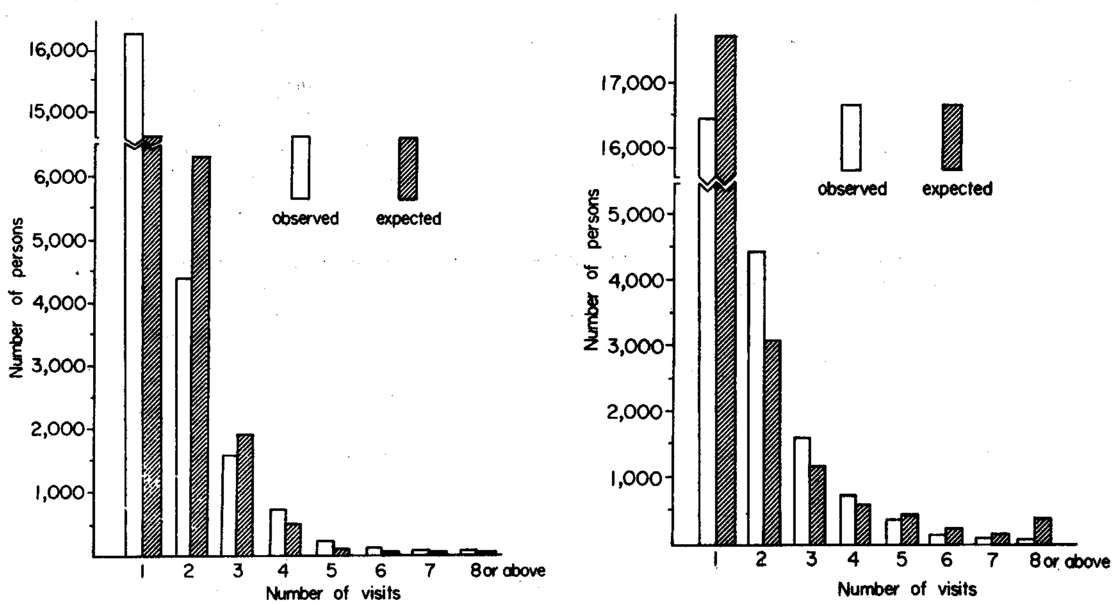
Disease	Estimator($\hat{\theta}$) Test statistic(χ^2)	University Hospitals	General Hospitals	Hospitals	Clinics
Acute nasopharyngitis	$\hat{\theta}$	0.41493	0.53543	0.56927	0.71533
and acute URI	χ^2	8.26	14.86	17.83	1540.0**
Gastritis and duodenitis	$\hat{\theta}$	0.41488	0.51272	0.60502	0.65930
	χ^2	17.23	36.0*	14.10	180.0**
Contact dermatitis and	$\hat{\theta}$	0.45329	0.40447	0.66310	0.80186
other eczema	χ^2	13.42	3.37	28.74*	184.18**
Other noninfective	$\hat{\theta}$	0.39567	0.47220	0.55104	0.66020
gastroenteritis and colitis	χ^2	7.50	4.50	13.33	112.7*

* P<0.005, d.f=9

** P<0.001, d.f=9

Table 7. Mean number of visits and estimates of the parameter of the 10 most common diseases

ICD Code	Disease	Mean(S.D) $\hat{\theta}$	University Hospitals	General Hopitals	Hospitals
460, 465	Acute nasopharyngitis and acute URI of multiple or unspecified sites	$\bar{x}$(S.D.)	1.33(1.00)	1.51(1.00)	1.57(1.00)
		$\hat{\theta}$	0.41493	0.53543	0.56927
466	Acute bronchitis and bronchiolitis	$\bar{x}$(S.D.)	1.52(1.00)	1.77(1.41)	2.08(1.73)
		$\hat{\theta}$	0.52392	0.65387	0.72942
535	Gastritis and duodenitis	$\bar{x}$(S.D.)	1.33(1.00)	1.47(0.91)	1.65(1.73)
		$\hat{\theta}$	0.41488	0.51272	0.60502
463	Acute tonsillitis	$\bar{x}$(S.D.)	1.38(1.00)	1.56(1.00)	1.71(1.41)
		$\hat{\theta}$	0.45903	0.56392	0.62808
462	Acute pharyngitis	$\bar{x}$(S.D.)	1.43(1.00)	1.51(1.00)	1.70(1.41)
		$\hat{\theta}$	0.49338	0.53359	0.62529
692	Contact dermatitis and other eczema	$\bar{x}$(S.D.)	1.31(1.00)	1.38(1.00)	1.81(1.73)
		$\hat{\theta}$	0.40447	0.45329	0.6631
372	Disorders of conjunctiva	$\bar{x}$(S.D.)	1.31(1.00)	1.46(1.00)	1.71(1.73)
		$\hat{\theta}$	0.40462	0.50701	0.62529
616	Inflammatory diseases of the cervix, vagina and vulva	$\bar{x}$(S.D.)	1.32(1.00)	1.77(1.73)	2.35(2.23)
		$\hat{\theta}$	0.41128	0.64990	0.77625
558	Other noninfective gastroenteritis and colitis	$\bar{x}$(S.D.)	1.30(1.00)	1.40(1.00)	1.53(1.41)
		$\hat{\theta}$	0.39567	0.47220	0.55104
300	Neurotic disorders	$\bar{x}$(S.D.)	1.52(1.00)	1.56(1.00)	1.66(1.73)
		$\hat{\theta}$	0.54176	0.56297	0.60683



a) Truncated poisson distribution b) Zeta distribution

Fig. 2. Observed and expected number of persons by number of outpatient visits in the case of acute nasopharyngitis and acute URI in general hospitals.

Poisson and Zeta distributions are not found to be adequate models. The logarithmic series distribution, however, provides an excellent good fit except in the case of clinics.

David and Johnson(1952), Newell(1965), and Selvin(1974) suggest that the truncated Poisson can be derived from such samples as the number of persons per house suffering from an infectious disease like measles, the number of accidents per worker in a factory, and the number of cases of spina bifida with hydrocephalus found per month. In these samples, the zero class has been truncated because the number of persons who weren't in an accident could not be enumerated. When this distribution is used in the number of outpatient visits, the expected probability decreases rapidly as the number increases. Thus, the expected frequencies are smaller than those observed in the tail.

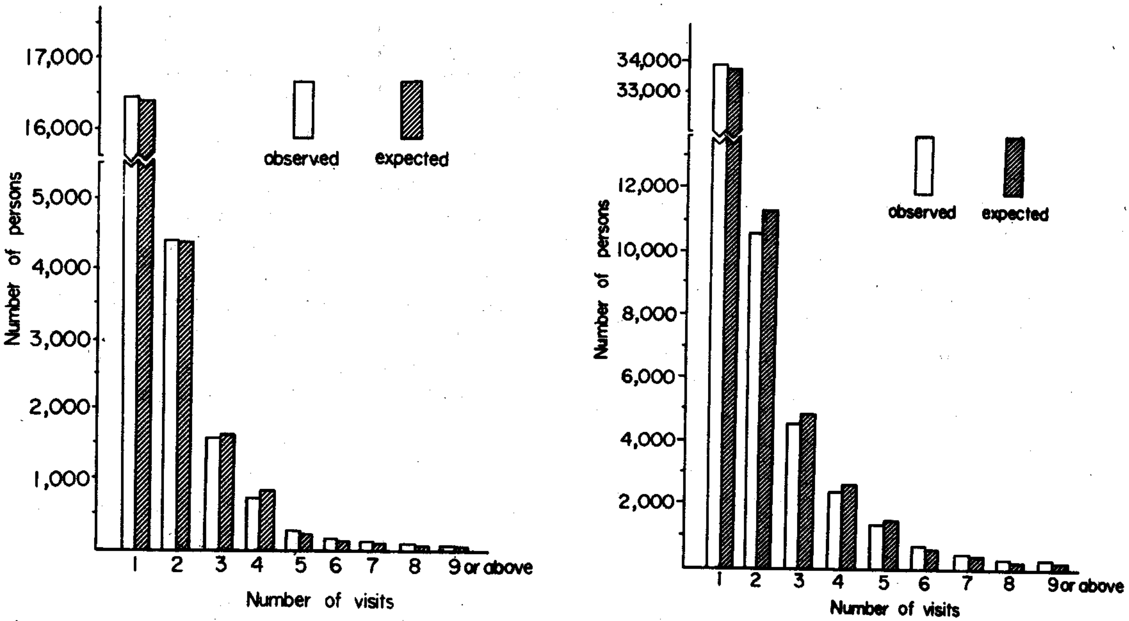
Zeta distribution, which has been known as Zipf-Estoup's Law, was derived from the number of occurrences of the k-th most frequent word by McNeil(1973), and the rank-frequency form by Hill(1974) and Hill and Woodroofe(1975). When this distribution is fitted to the number of visits, the expected frequencies are larger than those observed in the tail as shown in Fig. 2.

The logarithmic series distribution, which is an adequate model for determining the number of outpatient visits in the case of university hospitals, general hospitals, and hospitals was developed by R.A.Fisher in an investigation of the frequency distribution of numbers of species of animals obtained in random samples. Kendall(1948) represented a form of population growth by a discrete Markov process leading to a negative binomial distribution and its limiting form, a logarithmic series distribution. The relationship among the logarithmic series, Poisson, and negative binomial distributions was pointed out by Quenouille (1949). Chatfield *et al.*, (1966) has used this distribution to represent the distribution of numbers of items of a product purchased by a buyer in a specific period of time. The following is the theoretical background indicating that the logarithmic series distribution can be applicable to the number of outpatient visits in university hospitals, general hospitals, and hospitals.

The number of patient visits in a given period of time would be distributed as a Poisson variable with parameter λ.

$$P_r(X=k) = e^{-\lambda} \cdot \lambda^k / k! \quad (k=0, 1, 2, \ldots)$$

In the case of university hospitals, general hospitals and hospitals, each institution belonging to these medical care facilities has different treatment patterns due to varieties of hospital size, equipment and case



a) Acute nasopharyngitis and acute URI (in the case of general hospitals)

b) Gastritis and duodentis (in the case of clinics)

Fig. 3. Observed and expected number of persons by number of outpatient visits used by the logarithmic series distribution.

mix of the patients. Also, the number of total institutions belonging to the above categories in Korea are relatively small compared to the number of clinics. Therefore, the heterogeneity of the treatment pattern between institutions in the above three medical care facilities becomes more severe than in clinics. As a result it is no longer possible to assume that λ is fixed for each institution because of this heterogeneous characteristic. In connection with work by Quenouille (1949), if λ is distributed with a gamma-type probability density function, the frequency with which exact k visits of a patient are observed would be distributed in the negative binomial form. That is,

$$P_r(X=k) = \frac{\Gamma(n+k)}{k!\,\Gamma(n+1)} \cdot \frac{n}{q^n-1} \cdot \left(\frac{p}{q}\right)^k$$

Here we note that only frequencies of observations greater than zero are available, since the collection gives no indication of the number of general hospitals not found in the sample. If it is assumed that a large number of general hospitals is so rare that the probability of their observation is small, a negative binomial distribution, if it could be fitted to the data, would have a value of n approaching zero. We are thus led to consider the limiting form of the negative binomial distribution where n tends to zero and the zero class is ignored. The limiting form is called the logarithmic series distribution, the probability of events being

$$P_r(X=k) = \alpha\cdot\theta^k/k \quad (k=1, 2, \ldots; 0<\theta<1)$$

where $\alpha=-1/\log(1-\theta)$

On the other hand, in the case of clinics, even the logarithmic series distribution cannot be fitted to the data well (Fig.3). Compared with the characteristic of the utilization of university hospitals, general hospitals, and hospitals that are relatively few in number but offer various kinds of treatment patterns, that of clinic utilization with almost homogeneous treatment patterns in spite of the fact that there are a great many clinics could be the reason for the above results.

Further research is needed to construct the compound or modified distribution which would explain the number of outpatient visits in the case of clinics.

REFERENCES

Barton DE, David FN, Merrington M: *Tables for the solution of the exponential equation exp(b)/(1-b)=1. Biometrika 50: 169-172, 1963*

Chatfield C, Ehrenberg AS , Williams CB: *Progress on a simplified model of stationary purchasing behavior (with discussion). J Roy Stat Soc, Series A, 129: 317-367, 1966*

Chiang CL: *An index of health; Mathematical models. U.S. National Center for Health Statistics, Vital and Health Statistics, Series 2, No. 5, 1965*

Cohen AC: *Estimating the parameters of a modified Poisson distribution. J Am Stat Assn 55: 139-144, 1960*

Cohen AC: *Estimation in the truncated Poisson when zeros and some are missing. J Am Stat Assn 55: 342-348, 1960*

David FN, Johnson NL: *The trucated Poisson. Biometrics 8: 275-285, 1952*

Hill BM: *The rank frequency form of Zipf's law. J Am Stat Assn 69: 1017-1026, 1974*

Hill BM , Woodbroofe M: *Stronger forms of Zipf's law. J Am Stat Assn 70: 212-219, 1975*

Johnson NL , Kotz S: *Discrete distributions. New York: Houghton Mifflin, 1969*

Kendall DG: *On some modes of population growth leading to R.A.Fisher's logarithmic series distribution. Biometrika 35: 6-15, 1948*

Kim IS, Kim HJ, et al.: *A study on the use of a computer system for reviewing health insurance claim payments. Inst Pop Hlth Ser Res, Yonsei University, 1984*

McNeil DR: *Estimating an author's vocabulary. J Am Stat Assn 68: 92-96, 1973*

Newell DJ: *Unusual frequency distributions. Biometrics 8: 275-285, 1963*

Patil GP, Bildiker S: *Multivariate logarithmic series distribution as a probability model in population and community ecology and some of its statistical properties. J Am Stat Assn 62: 655-674, 1967*

Quenouille MH: *A relation between the logarithmic, Poisson and negative binomial series. biometrics 5: 162-164, 1949*

Selvin S: *Maximum likelihood estimation in the truncated or censored Poisson distribution. J Am Stat Assn 69: 234-237, 1974*

Steyn HS: *On the multivariate Poisson normal distribution. J Am Stat Assn 71: 233-236, 1976*

Tate RF , Goen RL: *Minimum variance unbiased estimation for the truncated Poisson distribution. Ann Math Stat 29: 755-765, 1958*

Williamson E , Bretherton MH: *Tables of the logrithmic series distribution. Ann Math Stat 35: 284-294, 1964*