JIA 78 (1952) 0115-0121

THE MAXIMUM LIKELIHOOD FITTING OF THE DISCRETE PARETO LAW

By H. L. SEAL, B.Sc., Ph.D., F.F.A.
Consulting Actuary, New York

THERE is a considerable statistical literature on the distribution of income among individuals living in specified countries at given epochs. Although there has been criticism of the use of Pareto's (first) law which ascribes to incomes of size x a frequency proportional to

$$x^{-\beta} \quad (0<a<x<\infty,\ \beta>1),$$

it is generally agreed that it offers a useful first approximation at least for incomes above a certain level. In fact, recently Kaiser (1950) has shown that if income is regarded as a function of the attained age many of the arguments against Pareto's law can be answered. This point of view is important in what follows.

It was suggested by Meidell at the Seventh International Congress of Actuaries (Amsterdam, 1912) and later by Hagström that the distribution of sums assured according to size might be roughly proportional to income, and that accordingly the distribution law of such sums might approximate to Pareto's law. Cramér (1926) gave two numerical examples based on this suggestion, and found that although the fit was not perfect the results were evidence that Pareto's law was at least a satisfactory first approximation.

A few years ago it occurred to the author that the discrete analogue of Pareto's law, namely,

$$j^{-\beta}/\zeta(\beta) \quad (j=1, 2, 3, \ldots;\ \beta>1),$$

where $\zeta(\beta)$ is the Riemann Zeta function with real argument β so that $\zeta(\beta)=\sum_{j=1}^{\infty} j^{-\beta}$, might be successfully applied to the distribution of 'duplicate' policies among the policyholders of one or more life offices. Mr W. L. Mayhew, then a colleague of the author, offered to take a random sample of 2000 from the alphabetical list of male lives assured in a British life office, a list which was maintained for administrative purposes and included a statement of the policy number of each policy written on that life. Every tenth name in this list was included until the total of 2000 was reached. Each of the lives assured thus sampled was scheduled according to his year of birth and his policies in force were enumerated. Concurrent duplicates were counted as one policy, but incremental policies under pension schemes were each included separately. It may be mentioned that at the time the sample was taken the office in question was underwriting relatively few individual-policy pension schemes, and in the majority of such schemes the policies were of the deferred annuity type and were not included in the sample.

The resulting frequency distributions of numbers of policies per life assured—twelve distributions corresponding to the central ages $17\frac{1}{2}$, $22\frac{1}{2}$, $27\frac{1}{2}, \ldots, 72\frac{1}{2}$ respectively—were published (1947)* and a somewhat crude method of fitting the Pareto law was employed; the fitting was criticized by

*By an unfortunate slip it was there stated that 2000 *policies* were involved.

Beard and Perks (1949). It has since become apparent that a relatively simple method of fitting is available, utilizing the method of maximum likelihood, and the resulting fit of the distributions has been greatly improved. It is thought that this method of fitting and the distribution of duplicate policies—the only one available in print—are of sufficient general interest to warrant their separate publication. It is not, of course, claimed that the successful application of the Pareto law to this relatively small set of observations implies its validity as a general 'law of nature'.

Suppose, first of all, that it is desired to fit the discrete Pareto law, viz.

$$\pi_j = j^{-\beta}/\zeta(\beta) \quad (j=1, 2, 3, \ldots; \beta > 1)$$

to a series of observations, n_j being the frequency assigned to the value j. The likelihood of the given series is proportional to

$$\prod_{j=1}^{\infty} (\pi_j)^{n_j}.$$

Taking natural logarithms we write

$$L = \sum_{j=1}^{\infty} n_j \log \pi_j = -\beta \sum_{j=1}^{\infty} n_j \log j - \log \zeta(\beta) \sum_{j=1}^{\infty} n_j.$$

On differentiating with respect to β, the only parameter,

$$-\frac{\partial L}{\partial \beta} = \sum_{j=1}^{\infty} n_j \log j + N\frac{\zeta'(\beta)}{\zeta(\beta)}, \quad N = \sum_{j=1}^{\infty} n_j,$$

and this equals zero when

$$-\frac{\zeta'(\beta)}{\zeta(\beta)} = \frac{1}{N} \sum_{j=1}^{\infty} n_j \log j. \tag{1}$$

The right-hand side of (1) is fixed, and it is thus easy to determine β by inverse interpolation in a table of $\zeta'(\beta)/\zeta(\beta)$. Such a table to seven decimal places subject to fourth difference interpolation is available in an article by Walther (1926).

If it were just a question of fitting a single series of observations by the discrete Pareto law the above development would be adequate. However, consideration of the circumstances under which individuals purchase policies indicates that the number of policies on any one life is likely to be an increasing function of the attained age. A simple assumption is that the β's of successive distributions of duplicate policies are connected by the straight line

$$\beta_x = a' + b'x.$$

Suppose that n_{rj} lives with j policies have been observed at age $\alpha + rt$ $(r=0, 1, 2, \ldots, k-1)$ and write

$$\pi_{rj} = j^{-\overline{a+br}}/\zeta(a+br), \quad \text{where } a = a' + b'\alpha,\ b = tb'.$$

The natural logarithm of the likelihood is then proportional to

$$\begin{aligned} L &= \sum_{r=0}^{k-1} \sum_{j=1}^{\infty} n_{rj} \log \pi_{rj} \\ &= -\sum_{r=0}^{k-1} \sum_{j=1}^{\infty} n_{rj}(a+br)\log j - \sum_{r=0}^{k-1} \sum_{j=1}^{\infty} n_{rj} \log \zeta(a+br) \\ &= -\sum_{r=0}^{k-1} (a+br) M_r - \sum_{r=0}^{k-1} N_r \log \zeta(a+br), \end{aligned}$$

where $$N_r = \sum_{j=1}^{\infty} n_{rj}, \quad M_r = \sum_{j=1}^{\infty} n_{rj} \log j.$$

Hence $$-\frac{\partial L}{\partial a} = \sum_{r=0}^{k-1} M_r + \sum_{r=0}^{k-1} N_r \frac{\zeta'(a+br)}{\zeta(a+br)}, \tag{2}$$

$$-\frac{\partial L}{\partial b} = \sum_{r=0}^{k-1} rM_r + \sum_{r=0}^{k-1} rN_r \frac{\zeta'(a+br)}{\zeta(a+br)}. \tag{3}$$

On equating these last two relations to zero a pair of simultaneous equations results from which to determine a and b.

The numerical procedures necessitated, respectively, by (1) and by the simultaneous equations just derived are illustrated by fitting the discrete Pareto law to the frequencies in the columns headed A in Table 1. These frequencies are those deriving from the sampling experience already mentioned.

In order to apply (1), for example, to the observations at attained ages 45–49 (actually, calendar years of birth 1894–8) supposed to be centred at age $47\frac{1}{2}$, we first calculate $\frac{1}{N}\sum_{j=1}^{\infty} n_j \log j = \cdot 148111$ with $N = 282$. Using Walther's table of $\zeta'(\beta)/\zeta(\beta)$ we find that $10^6 \left[\frac{1}{N}\sum_{j=1}^{\infty} n_j \log j + \frac{\zeta'(\beta)}{\zeta(\beta)}\right]$ equals -534 for $\beta = 3\cdot 1$ and equals 13,750 for $\beta = 3\cdot 2$. Linear interpolation indicates that the value zero would be obtained at $\beta = 3\cdot 104$, and recalculation of the function produces 73 for that β value. A second linear interpolation between $\beta = 3\cdot 1$ and $\beta = 3\cdot 104$ confirms that $\beta = 3\cdot 104$ to three decimal places. This procedure was adopted for each of the twelve distributions of duplicate policies provided in Table 1, and the theoretical Pareto frequencies obtained from the resulting β values are given in that table in the columns headed T_1.

The solution of the simultaneous equations required on the assumption of a linear progression of β's with attained age was more intricate. After some preliminary trials the values of (2) and (3) were calculated at the nine points determined by associating each of the a values 3·8, 3·85, 3·9 with each of the b values $-\cdot 12$, $-\cdot 125$, $-\cdot 13$. If the values of a and b required to make (2) and (3) zero simultaneously are written $3\cdot 8 + \cdot 05p$ and $-\cdot 12 - \cdot 005q$, respectively, the relation

$$\begin{aligned} 0 = {} & [1 + p\Delta_1 + \tfrac{1}{2}p(p-1)\Delta_1^2]\, u_{00} \\ & + q\,[\Delta_2 + p\Delta_1\Delta_2 + \tfrac{1}{2}p(p-1)\Delta_1^2\Delta_2]\, u_{00} \\ & + \tfrac{1}{2}q(q-1)\,[\Delta_2^2 + p\Delta_1\Delta_2^2 + \tfrac{1}{2}p(p-1)\Delta_1^2\Delta_2^2]\, u_{00} \end{aligned}$$

(which is written in an obvious notation) was used for both (2) and (3) to obtain a pair of simultaneous equations in p and q. These were solved by successive approximation with the result

$$a = 3\cdot 843, \quad b = -\cdot 122.$$

The frequencies obtained from β values derived in this manner are given in Table 1 in the columns headed T_2.

The results in Table 1 are satisfactory. In particular, the comment, made in the 1947 paper, that the theoretical distributions show a tendency to longer tails than the observations imply has now considerably less force. This is

Table 1. Distribution of policies among 2000 assured lives

Comparison of actual distribution, A, with fittings T_1 by equation (1) and T_2 by equations (2) and (3)

	Attained age 17½			22½			27½			32½			37½			42½			
j	T_1	T_2	A	T_1	T_2	A	T_1	T_2	A	T_1	T_2	A	T_1	T_2	A	T_1	T_2	A	j
1	36·5	35·6	36	58·4	55·2	58	101·9	99·4	101	242·7	243·5	241	285·3	286·6	283	305·2	310·7	307	1
2	2·0	2·5	3	2·2	4·2	3	6·8	8·2	8	22·3	21·9	26	28·6	28·0	35	35·5	33·1	31	2
3	·4	·5	.	·3	·9	.	1·4	1·9	2	5·5	5·3	3	7·5	7·2	4	10·1	8·9	12	3
4	·1	·2	.	·1	·3	.	·5	·7	.	2·0	2·0	3	2·9	2·7	2	4·1	3·5	6	4
5	.	·1	.	.	·1	.	·2	·3	.	·9	·9	.	1·4	1·3	.	2·1	1·7	.	5
6	.	.	.	.	·1	.	·1	·2	.	·5	·5	2	·8	·7	2	1·2	·9	1	6
7	.	.	.	.	.	.	·1	·1	.	·3	·3	.	·4	·4	.	·7	·6	2	7
8	.	.	.	.	.	.	.	·1	.	·2	·2	.	·3	·3	1	·5	·4	.	8
9	.	.	.	.	.	.	.	.	.	·1	·1	.	·2	·2	1	·3	·3	.	9
10	.	.	.	.	.	.	.	.	.	·1	·1	.	·1	·1	.	·2	·2	.	10
11	.	.	.	.	.	.	.	.	.	·1	·1	.	·1	·1	.	·2	·1	2	11
12	.	.	.	.	.	.	.	.	.	.	.	.	·1	·1	.	·1	·1	.	12
13	.	.	.	.	.	.	.	.	.	.	.	.	·1	·1	.	·1	·1	.	13
14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	·1	·1	.	14
15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	·1	·1	.	15
16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	·1	.	.	16
17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	17
Totals	39·0	38·9	39	61·0	60·8	61	111·0	110·9	111	274·7	274·9	275	327·8	327·8	328	360·6	360·8	361	

Table 1 (*continued*)

	Attained age 47½			52½			57½			62½			67½			72½			
j	T_1	T_2	A	T_1	T_2	A	T_1	T_2	A	T_1	T_2	A	T_1	T_2	A	T_1	T_2	A	j
1	235·9	238·6	233	199·3	198·4	200	110·8	113·8	108	72·0	64·2	69	35·1	34·6	33	27·4	28·3	26	1
2	28·8	27·7	35	24·7	25·0	24	16·6	15·6	20	6·3	9·6	10	5·5	5·6	7	5·2	5·0	5	2
3	8·4	7·8	4	7·3	7·5	7	5·5	4·9	5	1·5	3·2	1	1·9	1·9	4	2·0	1·8	4	3
4	3·5	3·2	5	3·0	3·2	2	2·5	2·1	2	·6	1·4	.	·9	·9	1	1·0	·9	1	4
5	1·8	1·6	3	1·6	1·6	.	1·4	1·1	4	·3	·8	.	·5	·5	.	·6	·5	2	5
6	1·0	·9	.	·9	·9	3	·8	·7	.	·1	·5	.	·3	·3	.	·4	·3	.	6
7	·6	·6	1	·6	·6	1	·5	·4	.	·1	·3	.	·2	·2	.	·3	·2	.	7
8	·4	·4	.	·4	·4	1	·4	·3	1	.	·2	.	·1	·1	.	·2	·2	.	8
9	·3	·3	.	·3	·3	.	·3	·2	.	.	·2	.	·1	·1	.	·1	·1	.	9
10	·2	·2	.	·2	·2	1	·2	·2	.	.	·1	.	·1	·1	.	·1	·1	.	10
11	·2	·1	.	·1	·2	.	·2	·1	.	.	·1	.	·1	·1	.	·1	·1	.	11
12	·1	·1	.	·1	·1	.	·1	·1	.	.	·1	and	.	·1	.	·1	·1	.	12
13	·1	·1	1	·1	·1	.	·1	·1	.	.	·1	1 at	.	.	.	·1	.	.	13
14	·1	·1	.	·1	·1	.	·1	·1	.	.	.	$j=18$	.	.	.	.	.	.	14
15	·1	·1	.	·1	·1	.	·1	.	.	.	.	.	.	.	.	.	.	.	15
16	·1	.	.	.	·1	.	·1	.	.	.	.	.	.	.	.	.	.	.	16
17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	17
Totals	281·6	281·8	282	238·8	238·8	239	139·7	139·7	140	80·9	80·8	81	44·8	44·5	45	37·6	37·6	38	

seen clearly in Table 2 where the twelve distributions of Table 1 (taken to further decimal places than there shown) have been collected together and show no signs of systematic discrepancies between theory and observation. Without resorting to probability tests of the deviations between T_1 and A and T_2 and A, respectively, it is reasonable to conclude that the two-parameter

Table 2. Comparison of combined distributions of Table 1 for all ages

j	T_1	T_2	A	j
1	1710·4	1708·9	1695	1
2	184·5	186·4	207	2
3	51·7	51·9	46	3
4	21·2	21·1	22	4
5	10·7	10·6	9	5
6	6·1	6·0	8	6
7	3·8	3·8	4	7
8	2·6	2·5	3	8
9	1·8	1·7	1	9
10	1·3	1·3	1	10
11	1·0	1·0	2	11
12	·8	·7	.	12
13	·6	·6	1	13
14	·5	·5	.	14
15	·4	·4	.	15
16	·3	·3	.	16
17	·3	·3	.	17
18	·2	·2	1	18
	1998·2	1998·2	2000	

Table 3. Comparison of the values of β by various methods

Age	Method of obtaining the value of β			
	Relation (1)	Relations (2) and (3)	Ratio method	First moment
$17\frac{1}{2}$	4·204	3·843	3·585	4·372
$22\frac{1}{2}$	4·738	3·720	4·273	4·866
$27\frac{1}{2}$	3·897	3·598	3·658	4·023
$32\frac{1}{2}$	3·446	3·476	3·212	3·523
$37\frac{1}{2}$	3·316	3·354	3·015	3·361
$42\frac{1}{2}$	3·104	3·231	3·308	3·131
$47\frac{1}{2}$	3·033	3·109	2·735	3·114
$52\frac{1}{2}$	3·015	2·987	3·059	3·057
$57\frac{1}{2}$	2·738	2·864	2·433	2·913
$62\frac{1}{2}$	3·513	2·742	2·787	3·019
$67\frac{1}{2}$	2·670	2·620	2·237	2·946
$72\frac{1}{2}$	2·397	2·498	2·379	2·684

series of distributions is an improvement on the twelve separate distributions with twelve parameters which are necessary on the assumption of independent β's.

We close this note by mentioning two other possible methods of fitting the discrete Pareto law. The first of these, which was the basis of the fitting adopted in the 1947 paper, makes use of the fact that, on the assumption of the Pareto law, the ratio of the frequency of individuals with one policy to that of individuals with two policies is equal on the average to 2^β. The second method is that of moments. Only the first moment is necessary to determine β since the expectation of the mean number of policies per life assured is

$$\sum_{j=1}^{\infty} j\pi_j = \frac{1}{\zeta(\beta)} \sum_{j=1}^{\infty} j^{-\beta+1} = \frac{\zeta(\beta-1)}{\zeta(\beta)},$$

and β may thus be found by trial and error using a table of $\zeta(\beta)$.

Table 3 compares the β values obtained from the previous data using (*a*) relation (1), (*b*) relations (2) and (3), (*c*) the ratio of first and second frequencies, and (*d*) the first moment. It will be noticed that only one of the twelve β values determined by the method of moments is less than the 'maximum-likelihood' value, whereas only two of the 'ratio' method's values exceed the corresponding 'maximum-likelihood' values.

REFERENCES

Beard, R. E. and Perks, W. (1949). The relation between the distribution of sickness and the effect of duplicates on the distribution of deaths. *J.I.A.* LXXV, 75–86.

Cramér, H. (1926). Återförsäkring och maximum på egen risk. *Sjunde nord. Lifförsäkringskong*, Oslo, pp. 64–83.

Kaiser, E. (1950). La distribution des revenus dans la technique mathématique de la sécurité sociale. *Mitt. Ver. schweiz. VersichMath.* L, 249–335.

Seal, H. L. (1947). A probability distribution of deaths at age x when policies are counted instead of lives. *Skand. AktuarTidskr.* XXX, 18–43.

Walther, A. (1926). Anschauliches zur Riemannschen Zetafunktion. *Acta Math.* XLVIII, 393–400.