

AN INVESTIGATION OF THE USE OF WEIGHTED
AVERAGES IN THE ESTIMATION OF THE MEAN OF
A LONG-TAILED CLAIM SIZE DISTRIBUTION

G. C. TAYLOR
Sydney, Australia.

ABSTRACT

The paper discusses the problem of estimating the mean of a long-tailed claim size distribution when the investigator's knowledge of the distribution is only vague

One method of dealing with this problem, the method developed by Johnson and Hey, is examined and found to produce strongly biased estimators

The situation in which a sufficient statistic (but nothing else) for the claim size distribution is known is examined, and an approximately unbiased estimator developed. This estimator is substantially more efficient than the arithmetic mean in some cases. It appears to be quite successful when the sufficient statistic is real-valued. It is of limited use when the sufficient statistic is vector-valued.

I. THE PROBLEM OF LONG-TAILED CLAIM SIZE DISTRIBUTIONS

For the purposes of this paper we can take a *long-tailed distribution* to be one whose density converges to zero less rapidly than the simple exponential family. Such distributions occur relatively frequently in the field of nonlife insurance. They are particularly prevalent among the distributions of individual claim sizes in respect of fire policies and liability policies.

Since the mean of a distribution is one of its most important properties—and indeed in the context of claim size distributions, usually the most important property—it is necessary that one have as reliable a method as possible for the estimation of this parameter.

In nonlife insurance this estimating problem can prove quite troublesome, because of the fact that standard statistical techniques are of limited applicability. This statement deserves some explanation particularly as the majority of this paper is concerned with methods which lie outside the scope of "standard" methods.

The statistician faced with the problem of estimating the mean of a long-tailed (or any other) distribution would begin by defining the family of likelihoods which are admissible as a representation of the distribution under consideration. He would then select estimates of the unknown parameters according to some optimization criterion, e.g. maximum likelihood, minimum-variance unbiasedness, etc.

The difficulty for the actuary involved with nonlife insurance arises at the very first stage, i.e. in deciding the admissible likelihood functions. In practice, he may have only the vaguest notion of the shape of the distribution. For example, he may be prepared to assert that it is within the exponential family of likelihoods. The exponential family is an extremely large one, so that although the requirement of delimiting the admissible likelihoods has been satisfied technically, the practical benefit of this stage of the procedure is doubtful.

It is basically for this reason that alternative methods of approaching the estimation of mean claim size are necessary. Of course, one can estimate this parameter with the sample mean. This has the advantage of ensuring unbiasedness, but, as is well-known, the sample mean from some long-tailed distributions has rather a large variance. Since unbiasedness and small variance are properties which one would usually like an estimator to possess simultaneously, the need for considering estimators other than the sample mean is immediate.

2. THE JOHNSON-HEY METHOD OF WEIGHTED AVERAGES

Hey (1970), concerned by the disturbance to the sample mean of claim sizes resulting from a few but substantial large claims, suggested that the difficulty might be alleviated by using a *weighted average* of the sample claim sizes, the weights tending to decrease with increasing claim size. This suggestion was followed up by Johnson and Hey (1972).

To state this in mathematical terms, they were concerned that the sample mean claim size, though an unbiased estimator of the true mean, had too large a variance. Their solution was to estimate the true mean claim size m by means of the statistic:

$$M = \left(\sum_{i=1}^n S(C_i) \right) / n \times G, \quad (1)$$

where

- C_1, C_2, \dots, C_n are the sample values of claim size;
- $S(\cdot)$ is a weight function which is nondecreasing but whose first derivative is nonincreasing;
- G is a "grossing-up factor" which is so chosen that M is an unbiased estimate of m .

3. PURPOSE OF THE PAPER

The purpose of the present paper is three-fold:

- (i) to indicate certain dangers arising from use of the Johnson-Hey (*J-H*) method;
- (ii) to point out that there are sound theoretical reasons for introducing the transformation S ;
- (iii) to investigate ways other than the *J-H* method of producing an estimate of m from the statistics C_1, \dots, C_n .

4. SOME COMMENTS ON THE JOHNSON-HEY METHOD

It is clear from a brief scrutiny of formula (1) that the problematic factor is G . Hey himself (1970, p. 81) noted that "we have no knowledge of the sensitivity of the grossing-up factor". Other difficulties arising from the manner in which G is estimated are mentioned by Johnson and Hey (1972, pp. 227-8).

In this section, however, we shall ignore these difficulties by assuming that, for a single given m , it is possible to choose G exactly correctly. We shall see that difficulties still arise in the use of estimator M .

Let us deal with nonzero claims only and assume that their sizes are sampled from a lognormal distribution. It is to be emphasised that this particular distribution has been chosen for illustrative purposes only, though, as Hey (1970, pp. 62-3) and others remark, it is not far from the truth for some classes of motor insurance.

Thus, we assume that C_1, C_2, \dots, C_n is a random sample in which each $\log C_i$ has a normal distribution with mean μ and variance σ^2 . Then, as in well-known (see e.g. Kendall and Stuart, 1961, p. 68),

$$m = E[C_i] = \exp \left\{ \mu + \frac{1}{2} \sigma^2 \right\}. \quad (2)$$

Also

$$E \left[\sum_{i=1}^n \log C_i / n \right] = \mu. \quad (3)$$

Thus, if we choose $S(\cdot)$ as $\log(\cdot)$, then it follows from (1), (2) and (3) that, for M to be unbiased, it is necessary that

$$G = \mu^{-1} \exp \left\{ \mu + \frac{1}{2} \sigma^2 \right\}. \quad (4)$$

A difficulty arises here due to the fact that G is dependent (often quite strongly) on μ and σ^2 . This means that, if G is appropriate to some particular μ and σ^2 , it may not be appropriate to some other choice of these parameters. This is the reason for the phenomenon

noted by Johnson and Hey (1972, p. 228) that G appears to vary between different risk categories.

In order to appreciate the extent of the difficulty, it is necessary to understand that the $J-H$ method provides that G be calculated from the *aggregation of data from all risk categories* in such a way that the estimate of m for the risk category of *an individual chosen at random from the whole portfolio* is unbiased. Note that, despite this type of unbiasedness, the resulting estimators may be biased in respect of each separate risk category, and the bias will of course be worse for the more extreme categories.

A number of simulations were carried out to illustrate this point and some of the results are given in Tables 1 and 2. The sampling distribution for claim size was taken to be lognormal with parameters μ and σ^2 , though, as is fairly obvious, the point being illustrated here is valid for other distributions too. This was confirmed by other simulations whose results are not reproduced here. The portfolio was assumed to consist of five different risk categories. In each case $S(\cdot)$ was taken as $\log(\cdot)$.

TABLE 1

Risk Cate- gory	True Mean		Arithmetic Mean			$J-H$ estimate			
	μ	σ^2	$\exp\{\mu + \frac{1}{2}\sigma^2\}$	*sample size = 10	sample s. = 50	sample s. = 250	sample s. = 10	sample s. = 50	sample s. = 250
1	4.000	1.20	.99	94	99	99	128	128	130
2	4.250	1.10	1.22	127	123	120	137	136	138
3	4.500	1.00	1.48	139	146	149	144	144	146
4	4.625	0.95	1.64	164	162	165	148	149	150
5	4.750	0.90	1.81	188	179	183	154	152	154

* sample size in each risk category

TABLE 2

Risk Cate- gory	True Mean		Arithmetic Mean			$J-H$ estimate			
	μ	σ^2	$\exp\{\mu + \frac{1}{2}\sigma^2\}$	*sample size = 10	sample s. = 50	sample s. = 250	sample s. = 10	sample s. = 50	sample s. = 250
1	4.48	1.00	1.45	149	147	147	151	149	148
2	4.49	1.00	1.47	147	147	148	150	149	148
3	4.50	1.00	1.48	157	150	148	153	149	149
4	4.51	1.00	1.50	152	150	151	152	149	149
5	4.52	1.00	1.51	152	151	151	151	150	149

1 sample size in each risk category

The main effect of the $J-H$ method appears clearly in Table 1 where it can be seen that, although the true mean varies over risk categories by a factor of 1.82, the $J-H$ estimates vary by a factor of the order of only 1.2 approximately. Generally, the $J-H$ estimates

for the various risk categories are "squashed together", with high-risk categories underestimated and low-risk categories overestimated.

This "squashedness" of the J - H estimates has obvious implications for tariff-splitting.

The same phenomenon becomes apparent upon a scrutiny of Johnson and Hey's own results presented on p. 227 of their paper. However, it is not quite so obvious there because their simulated portfolio is rather like that represented in Table 2 of this paper, i.e. risk categories are all quite close together.

Thus, as the portfolio becomes more homogeneous, so is the bias in the J - H method reduced. But then so also is the need for recognizing different risk categories. Regrettably, we must conclude that the J - H method attains reasonable effectiveness only when it is least needed.

5. THEORETICAL JUSTIFICATION FOR WEIGHTED AVERAGE

Let us consider the family of likelihoods, dependent upon some parameter θ , which have the form:

$$f(x | \theta) = c(\theta) h(x) \exp \left[\sum_{j=1}^p \pi_j(\theta) t_j(x) \right]. \quad (5)$$

This is the so-called *exponential family of likelihoods*. It is very rich in the sense that, for most of our practically occurring distributions, we can find a member of the family which will serve as a good approximation.

Moreover, the exponential family has a number of attractive properties which make it relatively easy to work with. In particular (see e.g. Ferguson (1967, pp. 125-37)):

1. The statistic $T = \left(\sum_{i=1}^n t_1(X_i), \dots, \sum_{i=1}^n t_p(X_i) \right)$ is a sufficient statistic, i.e. contains just as much information as does the whole vector of observations X_1, \dots, X_n in a sample of size n .
2. The likelihood of T is also a member of the exponential family, with the same π_j 's as in $f(x | \theta)$.
3. Under rather weak conditions which will usually be met by an insurance portfolio, it is possible to conclude that, if $g(T)$ is an unbiased estimator of a function of θ , then it has the smallest variance among all unbiased estimators.

Since the object of Johnson and Hey's quest was stability of the estimator, Property 3 is particularly suggestive, although it must

be mentioned that this property does not preclude the existence of *more stable but biased* estimators.

Now if a claim size distribution is a member of the exponential family, then, by Property 1,

$$T = \left(\sum_{i=1}^n t_1(C_i), \dots, \sum_{i=1}^n t_p(C_i) \right)$$

is a sufficient statistic. We thus have in Property 3 a theoretical justification for basing our estimate of θ on the average (or, equivalently, the sum) of *transformed claim sizes*. Furthermore, the transformation to be used is by no means arbitrary, but is determined by (5).

The usefulness of this observation is seen fully when viewed against the background of the actuary's vague knowledge of the shape of the distribution, as described in Section 1. If the situation is slightly better than described there and the actuary is willing to assert that $p = 1$ and $t_1(\cdot) = \log(\cdot)$, then from this none too definitive assertion, we may deduce that θ should be estimated by some function of $\sum_{i=1}^n \log C_i$.

6. AN EXAMPLE OF THE USE OF TRANSFORMED CLAIM SIZES

Suppose that C has a lognormal distribution with parameter $\theta = (\mu, \sigma^2)$, then

$$\begin{aligned} f(C | \theta) &= (\sqrt{2\pi} \sigma C)^{-1} \exp [- (\log C - \mu)^2 / 2\sigma^2] \\ &= c(\theta) h(C) \exp [\pi_1(\theta) t_1(C) + \pi_2(\theta) t_2(C)], \\ c(\theta) &= (\sqrt{2\pi} \sigma)^{-1} \exp [- \mu^2 / 2\sigma^2], \\ h(C) &= C^{-1}, \\ \pi_1(\theta) &= \mu / \sigma^2, \quad t_1(C) = \log C, \\ \pi_2(\theta) &= - 1 / 2\sigma^2, \quad t_2(C) = (\log C)^2. \end{aligned}$$

Thus we lose no information from our claim size observations if we reduce them to the two values,

$$T_1 = \frac{1}{n} \sum_{i=1}^n \log C_i \quad \text{and} \quad T_2 = \frac{1}{n} \sum_{i=1}^n (\log C_i)^2 - T_1^2.$$

It is not immediately clear how an unbiased estimator is to be constructed from T_1 and T_2 . However, in the case of the lognormal distribution, it was shown by Finney (1941) that an unbiased estimator of $E[C]$ is

$$\exp(T_1) g\left(\frac{1}{2} T_2\right), \tag{6}$$

where

$$g(x) = 1 + x + \frac{n-1}{n+1} \frac{x^2}{2!} + \frac{(n-1)^2}{(n+1)(n+3)} \frac{x^3}{3!} + \dots \quad (7)$$

For large n , $g(x)$ does not differ by too much from e^x , so that (6) becomes approximately:

$$\left\{ \prod_{i=1}^n C_i \right\}^{1/n} \exp \left[\frac{1}{2} T_2 \right]. \quad (8)$$

This is approximately unbiased, and so, by Property 3 above, has small variance.

We have thus constructed an unbiased estimator with small variance in terms of transformed claim size, where the transformation is:

$$C \rightsquigarrow (\log C, (\log C)^2).$$

7. FURTHER DEVELOPMENT OF THE USE OF TRANSFORMED CLAIM SIZES

It is apparent that the method used in the previous section for estimating $E[C]$ when C is lognormally distributed differs considerably from the J - H method. It was also pointed out that the methods used there lead to minimum-variance unbiased estimators.

Unfortunately, however, the actuary may not be in a position to make as strong an assertion as that claim size is lognormally distributed. Possibly the strongest assertion he can make with any confidence is that claim sizes, after some prescribed transformation (e.g. log) are roughly exponentially distributed. This really amounts to asserting something like the order of convergence of the probability density of claim size.

Under these circumstances, it is natural to seek some extension of the method used in Section 6. This aim is pursued in this section, but it should be stated at the outset that the success achieved in this direction is limited, and perhaps the main result emerging from the study is that, when knowledge of the claim size distribution is as vague as above, the simple arithmetic mean is surprisingly efficient.

Let us suppose that the sample of claim sizes, C_1, C_2, \dots, C_n , is drawn from a distribution belonging to the exponential family with t_1 a one-to-one transform. Henceforth we denote t_1 by just t . The statistic,

$$T_n = \frac{1}{n} \sum_{i=1}^n t(C_i), \quad (9)$$

is a minimum-variance unbiased estimator of $E[t(C)]$, by Properties 2 and 3 given in Section 5. It is therefore reasonable to assume that the statistic $t^{-1}(T_n)$, after approximate correction for bias will provide an estimator of $E[C]$ of relatively small variance.

Let us write

$$\mu = E[t(C)], \quad \sigma^2 = \text{Var} [t(C)].$$

From Section 2,

$$m = E[C].$$

Now, we know that

$$E[t^{-1}(T_1)] = E[t_1^{-1}(t_1(C))] = m. \quad (10)$$

We therefore need to estimate the difference,

$$E[t^{-1}(T_n)] - E[t^{-1}(T_1)],$$

occasioned by increase of sample size from 1 to n . This change represents the bias in $t^{-1}(T_n)$ as an estimator of m .

Let us now write Z_n for the standardized version of T_n , i.e.

$$Z_n = \frac{T_n - \eta}{\sigma\sqrt{n}}.$$

Let the d.f. of Z_n be expanded in an Edgeworth series,

$$\sum_{k=0}^{\infty} c_k^n \Phi^{(k)}(z),$$

where, as usual, $\Phi^{(k)}$ is the k -th derivative of the standard normal d.f. Then

$$E[t^{-1}(T_n)] = \sum c_k^n E^{(k)} [t^{-1}(n^{-1/2} \sigma Z_n + \eta)], \quad (11)$$

where $E^{(k)}$ [function of Z_n] denotes the expected value of the argument on the assumption that Z_n has "distribution function" $\Phi^{(k)}$.

Now, if D denotes the differentiation operator, repeated integration by parts gives

$$E^{(k)} [t^{-1}(\eta + n^{-1/2} \sigma Z_n)] = n^{-k/2} (-\sigma)^k E^{(0)} [D^k t^{-1}(\eta + n^{-1/2} \sigma Z_n)], \quad (12)$$

under obvious regularity conditions on the functions t^{-1} , Dt^{-1} , D^2t^{-1} , etc.

Thus, by (11) and (12),

$$E[t^{-1}(T_n)] = \sum_{k=0}^{\infty} c_k^n n^{-k/2} (-\sigma)^k E^{(0)} [D^k t^{-1}(\eta + n^{-1/2} \sigma Z_n)]. \quad (13)$$

It is apparent from (10) and (13) that

$$E[t^{-1}(T_n)] = m - \sum_{k=0}^{\infty} (-\sigma)^k \{c_k^1 E^{(0)} [D^k t^{-1} (\eta + \sigma Z)] - n^{-k/2} c_k^n E^{(0)} [D^k t^{-1} (\eta + n^{-1/2} \sigma Z)]\},$$

where the subscripts on Z have been suppressed since they are made irrelevant by the distributional assumptions implied by $E^{(0)}$. Hence an unbiased estimator of m is

$$U_n = t^{-1}(T_n) + \sum_{k=0}^{\infty} (-\sigma)^k \{c_k^1 E^{(0)} [D^k t^{-1} (\eta + \sigma Z)] - n^{-k/2} c_k^n E^{(0)} [D^k t^{-1} (\eta + n^{-1/2} \sigma Z)]\}, \quad (14)$$

Now it is known that

$$c_0^n = 1, c_1^n = c_2^n = 0, c_3^n = -\frac{1}{6} n^{-1/2} \gamma_1, c_4^n = \frac{1}{24} n^{-1} \gamma_2,$$

where γ_1 and γ_2 are the coefficients of skewness and excess respectively of T_n . Moreover,

$$\begin{aligned} \gamma_1 \sigma^3 &= K_3, \\ \gamma_2 \sigma^4 &= K_4, \end{aligned}$$

where K_j is the j -th cumulant of $t(C)$.

Using these facts, we can simplify (14) somewhat to give:

$$\begin{aligned} U_n &= t^{-1}(T_n) \\ &+ \{E^{(0)} [t^{-1} (\eta + \sigma Z)] - E^{(0)} [t^{-1} (\eta + n^{-1/2} \sigma Z)]\} \\ &+ \frac{1}{6} K_3 \{E^{(0)} [D^3 t^{-1} (\eta + \sigma Z)] - n^{-2} E^{(0)} [D^3 t^{-1} (\eta + n^{-1/2} \sigma Z)]\} \\ &+ \frac{1}{24} K_4 \{E^{(0)} [D^4 t^{-1} (\eta + \sigma Z)] - n^{-3} E^{(0)} [D^4 t^{-1} (\eta + n^{-1/2} \sigma Z)]\} \\ &+ \dots \end{aligned}$$

Since we do not have true values of μ , σ , K_3 and K_4 , we replace them by estimates. The obvious choices are (see Cramer, 1946, 352)

$$\begin{aligned} \hat{\eta} &= T_n, \quad \hat{\sigma} = \sqrt{\frac{n}{n-1} a_2}, \\ \hat{K}_3 &= \frac{n^2}{(n-1)(n-2)} a_3, \\ \hat{K}_4 &= \frac{n^2}{(n-1)(n-2)(n-3)} \left[(n+1) \frac{a_4}{a_2^2} - 3(n-1) \right], \end{aligned}$$

where a_ν is the ν -th sample central moment of $t(C)$.

Thus we finally adopt as our estimator of n the statistic:

$$\begin{aligned} \hat{m} = & t^{-1}(T_n) \\ & + \{E^{(0)} [t^{-1}(T_n + \hat{\sigma}Z)] - E^{(0)} [t^{-1}(T_n + n^{-1/2}\hat{\sigma}Z)]\} \\ & + \frac{1}{6} \hat{K}_3 \{E^{(0)} [D^3 t^{-1}(T_n + \hat{\sigma}Z)] - n^{-2} E^{(0)} [D^3 t^{-1}(T_n + n^{-1/2}\hat{\sigma}Z)]\} \\ & + \frac{1}{24} \hat{K}_4 \{E^{(0)} [D^4 t^{-1}(T_n + \hat{\sigma}Z)] - n^{-3} E^{(0)} [D^4 t^{-1}(T_n + n^{-1/2}\hat{\sigma}Z)]\} \\ & + \dots \end{aligned} \tag{15}$$

It is of course apparent that \hat{m} is not in general unbiased. However, the inclusion of the corrective terms should remove the majority of the bias which would be present if $t^{-1}(T_n)$ alone were taken as estimator of m .

8. NUMERICAL RESULTS

Although the development of \hat{m} as an estimator of m began with considerations which rested on sound theory (see Section 5), a number of subsequent approximations have led to the position in which the bias and stability of \hat{m} are not entirely clear. For this reason, a number of simulations were carried out in order to compare the estimator \hat{m} with the simple arithmetic mean for bias and stability. The most informative results are summarized in Tables 3 and 4 below.

In Table 3 the sampling distribution for claim size was taken to be log-Laplacian, i.e. $\log C (= L, \text{ say})$ was taken to have a likelihood function, dependent upon parameter k , equal to

$$\frac{1}{2} k \exp[-k |L|], \quad -\infty < L < \infty.$$

TABLE 3

Risk Category	k	True Mean $k^2/(k^2 - 1)$	Arithmetic Mean				\hat{m}	
			sample size = 10	sample size = 50	sample size = 250	sample size = 10	sample size = 50	sample size = 250
1	1.10	5.8	2.3(4.2)	3.7(40.1)	4.0(27.5)	6.3(88.4)	3.5(10.2)	3.3(1.8)
2	1.30	2.4	2.0(46.8)	2.2(1.7)	2.4(1.4)	5.1(67.9)	2.2(1.4)	2.2(0.2)
3	1.49	1.8	1.7(1.8)	1.8(1.0)	1.8(0.1)	2.1(10.3)	1.8(0.8)	1.7(0.1)
4	1.70	1.5	1.5(0.9)	1.6(0.3)	1.5(0.02)	1.7(3.9)	1.6(0.2)	1.5(0.02)
5	1.89	1.4	1.4(0.3)	1.4(0.1)	1.4(0.02)	1.5(0.5)	1.4(0.1)	1.4(0.01)

*sample size in each risk category

In Table 4, the sampling distribution was taken to be lognormal as in Tables 1 and 2. As in Tables 1 and 2, the portfolio is assumed to consist of five risk categories, and $t(\cdot)$ is taken to be $\log(\cdot)$. The figures for "arithmetic mean" and \hat{m} are simulated values of these estimators. The figures in parentheses are the corresponding simulated values of the variances of the estimates.

TABLE 4

Risk Category	True Mean		Arithmetic Mean				\hat{m}		
	μ	σ^2	$\exp\{\mu + \frac{1}{2}\sigma^2\}$	sample *size = 10	sample size = 50	sample size = 250	sample size = 10	sample size = 50	sample size = 250
1	4.000	1.20	99	94(1800)	99(550)	100(110)	98(2200)	101(630)	100(110)
2	4.250	1.10	122	127(3100)	123(700)	120(100)	135(5100)	124(750)	121(110)
3	4.500	1.00	148	139(3200)	146(750)	149(160)	143(4100)	147(800)	149(160)
4	4.625	0.95	164	164(5600)	162(740)	165(150)	173(10000)	163(780)	165(150)
5	4.750	0.90	181	188(5900)	179(870)	183(140)	195(9600)	181(930)	183(140)

sample size in each risk category

9. CONCLUSIONS

The theme of the paper has been the estimation of mean claim size in the light of only vague information about the claim size distribution. When this information includes knowledge of a sufficient statistic, it is tempting to base the estimator on this statistic.

One such estimator is provided by the Johnson-Hey method, but Section 4, and particularly Table 1 therein, reveals that there are quite common situations in which this estimator gives poor results.

The estimator \hat{m} developed in Section 7 attempts to improve on the *J-H* method. Indeed, Table 3 indicates that for some long-tailed claim size distributions, this estimator is largely unbiased and achieves a significant reduction in variance as compared with a simple arithmetic mean. The longer the tail, the larger is the reduction in variance.

The usefulness of \hat{m} as an estimator is limited, however, as is evidenced by Table 4 where the variance of \hat{m} is slightly *greater* than the variance of the arithmetic mean. The reason for this is, presumably, that the sufficient statistic for the distribution involved here is an ordered pair rather than a single real value (as in the case of Table 3), and in such a case the transformation (9) makes only partial use of our knowledge of the sufficient statistic.

Perhaps the estimator \hat{m} can be refined to make fuller use of the sufficient statistic?

Perhaps also the main conclusion to be drawn from this investigation is that, in the possession of only the vague knowledge outlined in Section 1, it is often very difficult to improve upon the simple arithmetic mean as an estimator of mean claim size.

10. ACKNOWLEDGEMENT

The author gratefully acknowledges the use of computing and other facilities provided by the University of Essex, Colchester, England and Heriot-Watt University, Edinburgh, Scotland.

REFERENCES

- CRAMÉR, H., 1946. *Mathematical methods in statistics* Princeton University Press.
- FERGUSON, J. S., 1967 *Mathematical statistics: a decision-theoretic approach* Academic Press.
- FINNEY, D. J., 1941. On the distribution of a variate whose logarithm is normally distributed. *Supplement of the Journal of the Royal Statistical Society*, 7, 155.
- HEY, G. B., 1970. Statistics in non-life insurance *Journal of the Royal Statistical Society (Series A)*, 133, 56-85
- JOHNSON, P. D. and HEY, G. B., 1972 Statistical review of a motor insurance portfolio. *Astin Bulletin*, VI, 222-32.
- KENDALL M. G. and STUART, A., 1963. *The theory of advanced statistics*, Vol. I (2nd. ed.). Charles Griffin and Co Ltd.