APPROXIMATE CONFIDENCE INTERVALS

BY M. S. BARTLETT
University of Manchester

1. INTRODUCTION

It is fairly generally known that asymptotic confidence intervals for an unknown parameter θ may be obtained directly from the likelihood (log) derivative $\partial L/\partial\theta$, where $p \equiv \exp L$ is the probability of the sample (see, for example, M. G. Kendall, 1946, §19·10). Such intervals are asymptotically equivalent to those obtained from the maximum-likelihood estimate $\hat{\theta}$, and have the property of providing asymptotically shortest confidence intervals on the average (see Kendall, §19·12). The direct use of $\partial L/\partial\theta$ has the advantage that its mean and variance are (under the usual differentiation conditions, which will be assumed throughout up to any order required) known *exactly* as 0 and

$$I \equiv E\left\{\left(\frac{\partial L}{\partial\theta}\right)^2\right\} = E\left\{-\frac{\partial^2 L}{\partial\theta^2}\right\},$$

in contrast with the asymptotic properties of $\hat{\theta}$, so that the only approximation introduced as far as the sampling distribution of $\partial L/\partial\theta$ is concerned is in regarding it as normal. This suggests that its use to provide confidence intervals should be approximately valid under fairly wide conditions, and, in so far as we can study the precise distribution of $\partial L/\partial\theta$, may even become exact. The further requirement for a valid confidence interval obtained from any random quantity $T(\theta)$ is that $T(\theta)$ bears a monotonic relation with θ for all samples, yielding a unique and admissible value θ_0 for each critical value T_0.

Now with regard to the normality approximation it is quite feasible to investigate also the higher moments of $\partial L/\partial\theta$, so that a correction to allow, say, for its skewness is not unduly complicated. It is the purpose of this note to develop such a further approximation procedure. It would appear difficult to specify precise and useful conditions under which such an approximation, leading to the use of a quantity $T(\theta)$, say, is necessarily free from possible violation of the further requirements on the relation of $T(\theta)$ with θ, but one or two incidental comments on this point may be helpful. We may expect any correcting terms to have a relatively minor effect except perhaps for very small samples, and thus not to disturb any monotonic relation with θ in the neighbourhood of the critical values of θ. With regard to the intervals which would be obtained from $\partial L/\partial\theta$ if its *exact* distribution were made use of, it is noted that if a sufficient statistic for θ exists the procedure is equivalent to making use of it,* but this does not necessarily imply that an exact confidence interval is possible even in such cases. For example, in small samples a confidence limit for θ, even when the required monotonic relation with θ is present, may yield a value outside the admissible range (this can happen also with the maximum likelihood estimate). In such cases the fact that we cannot always make full use of the sampling distribution means that our admissible confidence limits *over all samples* should have a confidence coefficient *not less than* the level

* This naturally raises the question of how far $\partial L/\partial\theta$ may be regarded in general as a 'theoretical sufficient statistic' for θ (cf. Bartlett, 1952, §6). However, this question is not altogether an unambiguous one, and I think that the answer is strictly only 'yes' if $\partial L/\partial\theta$, as a random variable, is considered *simultaneously* for all possible θ, and is 'no' if it is merely considered for the *single* true value of θ.

claimed. It might also be remarked that the approximation procedure is sometimes convenient even if a sufficient statistic does exist, in cases where the exact distribution of the sufficient statistic is difficult to handle (e.g. if the observations are from truncated distributions).

Only the case of *one* unknown parameter θ will be considered here, but it is hoped to discuss the case of more than one unknown in a later paper. The extension to confidence *regions* is in principle straightforward, but the case of one unknown plus 'nuisance parameters' is more subtle, being much more dependent on the nature of the sufficiency properties of $\partial L/\partial\theta$ (see earlier footnote; cf. also Bartlett, 1936, §6).

2. MOMENTS OF $\partial L/\partial\theta$

For reference a convenient method will be given of expressing the moments of $\partial L/\partial\theta$ in their simplest terms for evaluation, as an extension of the familiar result

$$E\left\{\left(\frac{\partial L}{\partial\theta}\right)^2\right\} = E\left\{-\frac{\partial^2 L}{\partial\theta^2}\right\}.$$

It is desirable to introduce a more convenient notation, and we shall write

$$E\left\{\frac{\partial L}{\partial\theta}\right\} \equiv L_1, \qquad E\left\{\frac{\partial^2 L}{\partial\theta^2}\right\} \equiv L_2,$$

$$\frac{\partial L_1}{\partial\theta} \equiv {}_1L_1, \qquad \frac{\partial^2 L_1}{\partial\theta^2} \equiv {}_2L_1,$$

$$E\left\{\left(\frac{\partial L}{\partial\theta}\right)^2\right\} \equiv L_1^{(2)}, \quad E\left\{\frac{\partial L}{\partial\theta}\frac{\partial^2 L}{\partial\theta^2}\right\} \equiv (L_1 L_2),$$

etc. Now

$$1 = \int p(\theta+\tau) = \int p(\theta)\exp\left\{\tau\frac{\partial L}{\partial\theta} + \tfrac{1}{2}\tau^2\frac{\partial^2 L}{\partial\theta^2} + \dots\right\},$$

where $\int p(\theta)$ is a formal notation for the probability Stieltjes integral over all possible samples. Expanding the last expression in powers of τ, and equating coefficients, we obtain an infinite series of relations of which the first four are

$$\left.\begin{aligned} &L_1 = 0,\\ &L_2 + L_1^{(2)} = 0,\\ &L_3 + 3(L_1 L_2) + L_1^{(3)} = 0,\\ &L_4 + 3L_2^{(2)} + 4(L_1 L_3) + 6(L_1^{(2)} L_2) + L_1^{(4)} = 0.\end{aligned}\right\} \tag{1}$$

The first two are the familiar ones already quoted, and the next two involve the third and fourth moments $L_1^{(3)}$ and $L_1^{(4)}$. These are not yet in their simplest terms; thus it is possible to express $L_1^{(3)}$ in 'linear' terms alone. We may, however, obtain an unlimited number of further relations by differentiating the relations (1). We have by differentiating the second and third relations once, noting that

$${}_1L_1^{(2)} \equiv \frac{\partial}{\partial\theta}\int p(\theta)\left(\frac{\partial L}{\partial\theta}\right)^2 = L_1^{(3)} + 2(L_1 L_2),$$

etc.,

$${}_1L_2 + L_1^{(3)} + 2(L_1 L_2) = 0,$$

$${}_1L_3 + 3\{(L_1^{(2)} L_2) + L_2^{(2)} + (L_1 L_3)\} + \{L_1^{(4)} + 3(L_1^{(2)} L_2)\} = 0.$$

Differentiating the former of these two derived relations again, we obtain further

$${}_2L_2 + \{L_1^{(4)} + 3(L_1^{(2)}L_2)\} + 2\{(L_1^{(2)}L_2) + L_2^{(2)} + (L_1 L_3)\} = 0.$$

We find finally, reverting to the full notation for clarity,

$$\kappa_3\left(\frac{\partial L}{\partial \theta}\right) \equiv E\left\{\left(\frac{\partial L}{\partial \theta}\right)^3\right\} = 3\frac{\partial I}{\partial \theta} + 2E\left\{\frac{\partial^3 L}{\partial \theta^3}\right\}, \tag{2}$$

$$\begin{aligned}\kappa_4\left(\frac{\partial L}{\partial \theta}\right) &\equiv E\left\{\left(\frac{\partial L}{\partial \theta}\right)^4\right\} - 3I^2 \\ &= 6\frac{\partial^2 I}{\partial \theta^2} + 8\frac{\partial}{\partial \theta}E\left\{\frac{\partial^3 L}{\partial \theta^3}\right\} - 3E\left\{\frac{\partial^4 L}{\partial \theta^4}\right\} + 3\sigma^2\left\{\frac{\partial^2 L}{\partial \theta^2}\right\},\end{aligned} \tag{3}$$

where it will be noticed that κ_4 involves one 'non-linear' quantity, the last variance term involving the mean square of $\partial^2 L/\partial\theta^2$. Whether these formulae (2) and (3) are useful will of course depend on the form of L; for example, in the case of the Cauchy distribution it is simpler to evaluate κ_3 and κ_4 directly.

3. Approximate correction for the skewness of $\partial L/\partial\theta$

It seems most useful to consider polynomial transformations of $\partial L/\partial\theta$ as in principle such transformations may be chosen with the aid of the above relations to annihilate the skewness (and, if necessary, any higher non-zero cumulants) exactly, although it is perhaps unlikely in practice that we shall wish to go further than a first further approximation involving κ_3. Consider the theoretical statistic

$$T_\lambda(\theta) \equiv \frac{\partial L}{\partial \theta} + \lambda\left[\left(\frac{\partial L}{\partial \theta}\right)^2 - I\right], \tag{4}$$

which still has zero mean, and variance

$$I + 2\lambda\kappa_3 + \lambda^2\sigma^2\left\{\left(\frac{\partial L}{\partial \theta}\right)^2\right\}. \tag{5}$$

Its skewness is
$$\kappa_3 + 3\lambda\sigma^2\left\{\left(\frac{\partial L}{\partial \theta}\right)^2\right\} + O(\lambda^2). \tag{6}$$

As $\sigma^2\{(\partial L/\partial\theta)^2\} = \kappa_4 + 2I^2$, we therefore choose, to the first order of approximation,

$$\lambda = -\frac{1}{6}\frac{\kappa_3}{I^2}, \tag{7}$$

whence
$$T(\theta) \equiv \frac{\partial L}{\partial \theta} - \frac{1}{6}\frac{\kappa_3}{I^2}\left[\left(\frac{\partial L}{\partial \theta}\right)^2 - I\right]. \tag{8}$$

The variance of T, if we neglect κ_3^2, is still I. More accurate confidence intervals should therefore be obtained if we write
$$T(\theta) = \pm\mu\sqrt{I(\theta)}, \tag{9}$$

where μ is the appropriate normal deviate for any required significance level, and solve for θ.

Of course if $\kappa_3 = 0$, the standard approximation remains unaltered to this order. In this particular case, i.e. *when* $\kappa_3 = 0$, the correction for the next cumulant κ_4 is also recorded. We write
$$T_\nu(\theta) = \frac{\partial L}{\partial \theta} + \nu\left(\frac{\partial L}{\partial \theta}\right)^3 \tag{10}$$

with mean zero, variance
$$\sigma^2 = I+2\nu[\kappa_4+3I^2]+O(\nu^2), \tag{11}$$

skewness
$$E\{T_\nu^3\} = 3\nu E\left\{\left(\frac{\partial L}{\partial\theta}\right)^5\right\}+O(\nu^2)$$
$$= 0+O(\nu^2, \nu\kappa_5), \tag{12}$$

and kurtosis
$$E\{T_\nu^4\}-3\sigma^4 = \kappa_4+3I^2+4\nu E\left\{\left(\frac{\partial L}{\partial\theta}\right)^6\right\}-3\sigma^4+O(\nu^2),$$

or, after reduction,
$$\kappa_4+24\nu I^3+O(\nu^2, \nu\kappa_4, \nu\kappa_6). \tag{13}$$

Hence we choose
$$\nu = -\frac{1}{24}\frac{\kappa_4}{I^3} \tag{14}$$

and
$$T(\theta) \equiv \frac{\partial L}{\partial\theta}-\frac{1}{24}\frac{\kappa_4}{I^3}\left(\frac{\partial L}{\partial\theta}\right)^3, \tag{15}$$

with variance
$$\sigma_T^2 \sim I(1-\tfrac{1}{4}\kappa_4/I^2). \tag{16}$$

4. Example I

To test out this procedure, we shall try it out on a standard problem whose exact solution is well known, namely, confidence limits for a (normal) sample variance. This example is taken because the sample variance s^2 is quite skew for a moderate number of degrees of freedom, and there is some interest in seeing how far the method can cope with this skewness. We have

$$\frac{\partial L}{\partial\theta} = \frac{n}{2\theta^2}(s^2-\theta),$$
$$-\frac{\partial^2 L}{\partial\theta^2} = -\frac{n}{2\theta^2}+\frac{ns^2}{\theta^3}, \qquad I = \frac{n}{2\theta^2},$$
$$\frac{\partial^3 L}{\partial\theta^3} = -\frac{n}{\theta^3}+\frac{3ns^2}{\theta^4}, \quad E\left\{\frac{\partial^3 L}{\partial\theta^3}\right\} = \frac{2n}{\theta^3},$$
$$\frac{\partial I}{\partial\theta} = -\frac{n}{\theta^3}, \qquad \kappa_3 = \frac{n}{\theta^3},$$

the value of κ_3 checking of course with the value inferred from the properties of s^2. The standard approximation for the confidence interval would amount to solving

$$\frac{n}{2\theta^2}(1-\theta) = \pm\mu\sqrt{\left(\frac{n}{2\theta^2}\right)}, \tag{17}$$

where for convenience we have expressed confidence limits θ in terms of s^2, i.e. have put $s^2 = 1$. The more accurate approximation suggested replaces (17) by the equation

$$\frac{n}{2\theta^2}(1-\theta)-\frac{1}{6}\frac{n}{\theta^3}\left[(1-\theta)^2-\frac{2\theta^2}{n}\right] = \pm\mu\sqrt{\left(\frac{n}{2\theta^2}\right)}. \tag{18}$$

This as it stands no longer yields a simple quadratic equation for θ, but to the same order of approximation we may replace the square bracket by its value from (17), viz. $2\theta^2(\mu^2-1)/n$, and (18) then also reduces to a quadratic (i.e. to a linear equation for each sign of μ). A corresponding simplification is available in general, for the crude interval is obtained from

$$\frac{\partial L}{\partial \theta} = \pm \mu \sqrt{I}, \tag{19}$$

and substituting in $T(\theta)$ in (9), we obtain

$$\frac{\partial L}{\partial \theta} - \frac{1}{6}\frac{\kappa_3}{I}(\mu^2 - 1) = \pm \mu \sqrt{I} \tag{20}$$

as the corresponding modification of the general equation (9).†

In arithmetical work where $\partial L/\partial \theta$ is calculated for various values of θ, either (20) or (9) may be used. The original approximation (19) is most conveniently plotted as

$$\frac{\partial L}{\partial \theta}\Big/\sqrt{I} = \pm \mu, \tag{21}$$

for the values of θ for which the left-hand side passes any chosen value of μ can be read off from the same graph. Equation (9) similarly is written

$$\frac{\partial L}{\partial \theta}\Big/\sqrt{I} - \frac{1}{6}\frac{\kappa_3}{I^{\frac{3}{2}}}\left[\left(\frac{\partial L}{\partial \theta}\Big/\sqrt{I}\right)^2 - 1\right] = \pm \mu, \tag{22}$$

and may be used in the same way. The equation equivalent to (20), viz.

$$\frac{\partial L}{\partial \theta}\Big/\sqrt{I} - \frac{1}{6}\frac{\kappa_3}{I^{\frac{3}{2}}}[\mu^2 - 1] = \pm \mu, \tag{23}$$

may only be used at the one chosen value of μ, but has the slight advantage that if interpolation for θ (or some suitable function $g(\theta)$) is approximately linear in (21), it is likely to be so also for (23), rather than for (22).

Table 1

n	Exact values				First or standard approximation				Second or further approximation			
	Lower 0·01	Lower 0·05	Upper 0·05	Upper 0·01	Lower 0·01	Lower 0·05	Upper 0·05	Upper 0·01	Lower 0·01	Lower 0·05	Upper 0·05	Upper 0·01
5	0·331	0·452	4·37	9·03	0·405	0·490	∞*	∞*	0·327	0·441	5·35	8·55
10	0·431	0·546	2·54	3·91	0·490	0·576	3·78	∞*	0·428	0·541	2·65	3·94
20	0·532	0·637	1·84	2·42	0·576	0·658	2·08	3·78	0·531	0·634	1·86	2·43
30	0·589	0·685	1·62	2·01	0·625	0·702	1·74	2·50	0·589	0·684	1·63	2·01

* In this example the approximations break down at the upper limit for small n and P; this occurs with the first approximation even for some of the values in the table.

To return to (17) and (18), the resulting limits are compared in Table 1 with the correct limits obtained from the distribution of s^2, for upper and lower significance limits $P = 0{\cdot}05$ and 0·01, or total significance limits of 0·10 and 0·02, and representative values 5, 10, 20, and 30 for n.

† This further approximation has been used when obtaining the values given in Tables 1 and 2.

It will be seen that even the second approximation is not too good at the upper limit for $n = 5$, and would be inferior to at least one other available approximation for the significance limits for s^2, but the important point to remember is that it is a much more general method, and this example suggests that it should be a considerable improvement over the first approximation, which is still hardly satisfactory in this example for n as large as 30.

5. Example II

As a second example we shall examine a standard discrete distribution, choosing for simplicity the Poisson rather than the binomial. Here we easily find, for θ the unknown mean and x an observation,

$$\frac{\partial L}{\partial \theta} = \frac{x}{\theta} - 1, \quad I = \frac{1}{\theta},$$

$$E\left\{\frac{\partial^3 L}{\partial \theta^3}\right\} = \frac{2}{\theta^2}, \qquad \kappa_3 = \frac{1}{\theta^2},$$

whence the crude approximation gives the equation

$$\frac{x}{\theta} - 1 = \pm\frac{\mu}{\sqrt{\theta}}, \tag{24}$$

and the further approximation (corresponding to (20))

$$\frac{x}{\theta} - 1 - \frac{1}{6\theta}(\mu^2 - 1) = \pm\frac{\mu}{\sqrt{\theta}}. \tag{25}$$

Of course for discrete distributions a confidence interval (at least if obtained from the data directly without introduction of extra randomization devices) can only be given as an inequality for the confidence level. This is perhaps hardly crucial if an approximate method

Table 2

x	Exact values*				First or standard approximation				Second or further approximation			
	Lower 0·01	Lower 0·05	Upper 0·05	Upper 0·01	Lower 0·01	Lower 0·05	Upper 0·05	Upper 0·01	Lower 0·01	Lower 0·05	Upper 0·05	Upper 0·01
0	0	0	3·00	4·61	†	†	3·64	6·37	†	†	3·12	4·93
1	0·01	0·05	4·74	6·64	0·04	0·07	5·28	8·14	†	0·02	4·83	6·86
2	0·15	0·36	6·30	8·41	0·28	0·43	6·79	9·77	0·09	0·31	6·37	8·58
3	0·44	0·82	7·75	10·05	0·64	0·92	8·22	11·33	0·36	0·77	7·81	10·19
5	1·28	1·97	10·51	13·11	1·58	2·11	10·94	14·30	1·21	1·93	10·56	13·22
10	4·13	5·43	16·96	20·14	4·54	5·61	17·35	21·21	4·07	5·40	17·00	20·23
20	11·08	13·25	29·06	33·10	11·58	13·46	29·42	34·08	11·04	13·23	29·09	33·16
30	18·74	21·59	40·69	45·40	19·29	21·82	41·04	46·33	18·71	21·58	40·71	45·45

* Quoted from Garwood (1936).

† The first approximation gives $\theta = 0$ here only if the continuity correction is *not* used. The second approximation no longer gives near $\theta = 0$ a monotonic relation between $T(\theta)$ and θ for small x, and thus breaks down near $\theta = 0$. (The inadequacy of the second approximation for very small θ is hardly unexpected, for the Poisson mean θ in this example effectively takes the place of n, the size of the sample.)

is in any case being used, but it reminds us that a still further approximation is now involved in obtaining limits from normal theory. For a purely discrete distribution as in this example some gain would be expected from the introduction of the usual continuity correction, which implies using $x-\frac{1}{2}$ when obtaining the lower limit for θ and $x+\frac{1}{2}$ for the upper limit. This correction has therefore been used throughout, though from an examination of one or two values the consequent gain for the first crude approximation appeared much more doubtful than for the second.

It will be seen from Table 2 that the second approximation is a considerable improvement over the first (except near $\theta = 0$), and is in fact somewhat better for small x at the upper limit than we might have anticipated from the experience with the first example. The exact values quoted were readily obtained by Garwood from χ^2 significance limits, and, as in Example I, it must be emphasized that the purpose of examining the present approximation in these standard examples is merely to obtain an idea of its accuracy.

6. Example III

The problem which actually initiated this inquiry is being discussed in detail elsewhere, but it may be helpful to indicate it here. From the tracks of certain cosmic ray particles which 'decay' into other fundamental particles, it was required to estimate the 'decay' parameter θ in the lifetime distribution

$$f(t)\,dt = e^{-t/\theta}\,dt/\theta. \tag{26}$$

This is simple enough in the case of unlimited track length, but for tracks in a chamber of finite size, not all particles will decay; the chance of doing so moreover varies with each particle because its time of passage through the chamber depends on its momentum. The situation is not always the same as this and is often complicated by further observational difficulties, but for simplicity here we shall suppose we have just N particles of one type, with effective times T_s ($s = 1 \ldots N$) in the chamber, of which n decay at times t_r ($r = 1 \ldots n$). The likelihood function in this case is a mixture of finite probability factors and densities, but this causes no difficulty. For such data we have, as the probability of the sth undecayed particle is $Q_s = e^{-T_s/\theta}$, and the probability density of the rth decayed particle is $f(t_r) = e^{-t_r/\theta}/\theta$,

$$\frac{\partial L}{\partial \theta} = -\frac{n}{\theta} + \sum_{r=1}^{n} \frac{t_r}{\theta^2} + \sum_{s=n+1}^{N} \frac{T_s}{\theta^2}, \tag{27}$$

giving in this particular case the immediate estimate

$$\hat{\theta} = \frac{1}{n}\left(\sum_{r=1}^{n} t_r + \sum_{s=n+1}^{N} T_s\right). \tag{28}$$

(The record of undecayed particles is not always available; in that case the estimation equation is different and not quite such a simple solution $\hat{\theta}$ is possible from the combined data.) It is not difficult to show that for equation (27)

$$I = \sum_{s=1}^{N} \frac{P_s}{\theta^2}, \tag{29}$$

where $P_s = 1-Q_s$. Thus the first approximation for the confidence interval would be obtained from equation (19), with $\partial L/\partial\theta$ given now by (27) and I by (29). In the limiting case when all T_s become large, the problem reduces to Example I, for each measurement t from the exponential distribution (26) is equivalent to a χ^2 with two degrees of freedom, and the mean t would be equivalent to a χ^2 (or s^2) with $2N$ degrees of freedom. In practice the T_s are often small, but in any case, in any attempt to obtain a confidence interval for θ, it would seem advisable when possible to go to the further approximation (9) or (20).

It is not too difficult, at least in the case covered by equation (27), to obtain the exact moment-generating function $M(\phi)$ of $\theta^2\,\partial L/\partial\theta$, remembering that the chance of decay somewhere in the chamber for each particle is $P_s\,(s = 1 \ldots N)$. The answer comes out to be $M(\phi) = \prod_{s=1}^{N} M_s(\phi)$, with

$$M_s(\phi) = e^{-\phi\theta}\,[1 - Q_s\,e^{\phi T_s}\{1-(1-\theta\phi)\,e^{\theta\phi}\}]/(1-\theta\phi). \tag{30}$$

Expanding $\log M(\phi)$, we find for the cumulants of $\theta^2\,\partial L/\partial\theta$

$$\left.\begin{aligned} \kappa_1 &= 0, \quad \kappa_2 = \theta^2 \sum_{s=1}^{N} P_s, \\ \kappa_3 &= 2\theta^3 \sum_{s=1}^{N} P_s - 3\theta^2 \sum_{s=1}^{N} Q_s T_s, \\ \kappa_4 &= 6N\theta^4 - 8\theta^3 \sum_{s=1}^{N} Q_s T_s - 6\theta^2 \sum_{s=1}^{N} T_s^2 Q_s + \theta^4 \sum_{s=1}^{N} Q_s - 3\theta^4 \sum_{s=1}^{N} Q_s^2. \end{aligned}\right\} \tag{31}$$

However, as the distribution of $\partial L/\partial\theta$ is a mixture of continuous and discrete components (for there is a finite chance that none of the N particles decays), it would be difficult to make use of the exact distribution of $\partial L/\partial\theta$ in this particular case, and the use of the cumulants (31), while first obtained directly by the above method, would be equivalent to using the general method developed earlier in this paper. The formulae in (31) agree of course with the results derived via the general method, and the expressions for κ_3 and κ_4 were checked by this means.

For maximum-likelihood or confidence interval equations not directly soluble, iterative or interpolative methods may usually be used (see equations (21), (22) and (23), §4).

I am indebted to Mrs A. Linnert for assistance with the calculations for Tables 1 and 2.

REFERENCES

BARTLETT, M. S. (1936). Statistical information and properties of sufficiency. *Proc. Roy. Soc.* A, **154**, 124.
BARTLETT, M. S. (1952). The statistical significance of odd bits of information. *Biometrika*, **39**, 228.
GARWOOD, F. (1936). Fiducial limits for the Poisson distribution. *Biometrika*, **28**, 437.
KENDALL, M. G. (1946). *The Advanced Theory of Statistics*, 1st ed. **2**. London: Griffin.

Notes added in proof. (i) The coefficients in the polynomial transformations considered in §3 are equivalent to those in a series expansion given by E. A. Cornish & R. A. Fisher in §8 of their paper 'Moments and cumulants in the specification of distributions,' *Rev. Inst. Int. Statis.* (1937), **4**, which should be consulted if further terms of the expansion are required.

(ii) The problem referred to in §6 I have discussed further in my paper 'On the statistical estimation of mean life-time,' *Phil. Mag.* (7th series), 1953, **44** (in the Press).