GROUPING CORRECTIONS AND MAXIMUM LIKELIHOOD EQUATIONS

By D. V. LINDLEY

Communicated by J. Wishart

Received 20 April 1949

1. Any mention of the word 'grouping' immediately brings to a statistician's mind the Sheppard corrections. These are usually used to make inferences about the underlying ungrouped population from observations made on the grouped population, but it is important to realize that, as stated and proved, they have nothing to do with sampling or inference and are merely expressions for the moments of one population in terms of the moments of another population derived from it. They can only be used for the inference problem when allied to the method of moments. This method, as formulated by K. Pearson, consists in taking for θ^* the estimate of the population parameter θ, the same function of the sample moments m_i that θ is of the population moments μ_i, each m_i being an estimate of the corresponding μ_i. If the population is grouped the m_i are estimates of the $\overline{\mu}_i$, the grouped population moments, so we require θ as a function of the $\overline{\mu}_i$ to apply Pearson's method. This can be done since θ is known as a function of the μ_i and the μ_i are known as functions of the $\overline{\mu}_i$ by the corrections. Use of the Sheppard corrections with any other inference method, even when this method, when applied to the continuous population, yields an estimate which is a sample moment, so far as I am aware, has not been examined except for the normal curve.

Now the only justifications for the method of moments are that it is easy to apply and the estimate obtained is consistent. The modern ideas of sufficiency and efficiency have shown that maximum likelihood is perhaps the most satisfactory method of providing an estimator, and we are therefore led to inquire as to the effect of grouping on this method of estimation. The grouping corrections we shall obtain will, unlike Sheppard's corrections, apply to sample values (estimators) and not to population values (parameters). They are therefore of direct use in inference problems and do not involve additional assumptions to make them applicable.

2. *The maximum likelihood corrections, single parameter.* Consider a continuous population with distribution function $F(x, \theta)$ and frequency function $f(x, \theta) = F'(x, \theta)$, where a dash denotes differentiation with respect to x and θ is the parameter we wish to estimate from the sample $x_1, x_2, \ldots, x_k$. The maximum likelihood estimate of θ, $\hat{\theta}$, is the solution of the equation

$$\sum_{i=1}^{i=k} \frac{\partial}{\partial \theta} \log f(x_i, \theta) = 0.$$

Suppose that the population is grouped into intervals of length h, and let one of these intervals be centred at the origin. Let

$$p(nh, \theta) = \int_{(n-\frac{1}{2})h}^{(n+\frac{1}{2})h} f(x, \theta)\, dx = F(\overline{n+\tfrac{1}{2}}h, \theta) - F(\overline{n-\tfrac{1}{2}}h, \theta),$$

https://doi.org/10.1017/S0305004100025512 Published online by Cambridge University Press

where n is an integer, be the probability of obtaining a single value nh. The new maximum likelihood equation for θ is

$$\sum_{i=1}^{i=k} \frac{\partial}{\partial\theta} \log p(n_i h, \theta) = 0,$$

where x_i lies in the interval $[(n_i - \frac{1}{2})h, (n_i + \frac{1}{2})h]$; let its solution be θ_1. It is this value θ_1 that we wish to find, and we have an approximation to it in the estimate

$$\hat{\theta}(n_1 h, n_2 h, \ldots, n_k h) = \theta_0,$$

say, which ignores the grouping, provided this latter exists†. θ_0 may be compared to the uncorrected moment and θ_1 to the corrected moment; on letting $\theta_1 = \theta_0 + \Delta$, Δ can be regarded as a 'Sheppard correction' to be applied to the usual estimate θ_0 to obtain the true maximum likelihood estimate. We obtain an approximation for Δ.

By use of Taylor's theorem, assuming the existence of $f'''(x, \theta)$, we have

$$p(nh, \theta) = hf + \frac{h^3}{24} f'' + h^4 R,$$

where the arguments of f are nh and θ, and R is a remainder term involving the third derivative. We assume that f does not vanish, so that

$$\log p(nh, \theta) = \log hf + \log\left[1 + \frac{h^2}{24}\frac{f''}{f} + h^3 \frac{R}{f}\right]$$

and

$$\frac{\partial}{\partial\theta} \log p(nh, \theta) = \frac{\partial}{\partial\theta} \log f + \frac{h^2}{24} \frac{\partial}{\partial\theta}\left(\frac{f''}{f}\right) + O(h^3).$$

We have to solve the equation obtained by summing the right-hand side over the sample values and equating to zero. This we do by Newton's method using θ_0 as the first approximation obtaining, since

$$\sum_{i=1}^{i=k} \frac{\partial}{\partial\theta} \log f(n_i h, \theta_0) = 0,$$

the result

$$\Delta = -\frac{h^2}{24} \left\{\sum_{i=1}^{i=k} \frac{\partial}{\partial\theta}\left(\frac{f''}{f}\right)_{\theta_0} \Big/ \sum_{i=1}^{i=k} \frac{\partial^2}{\partial\theta^2} (\log f)_{\theta_0}\right\} + O(h^3).$$

If we assume the existence of f^{iv}‡ the additional term is $O(h^4)$ and, by retaining this term and re-using Newton's rule, it is possible to get better approximations, but they are of considerable complexity and, in general, the above expression will suffice.

Example I. The variance of a normal population

$$f = \frac{1}{\sqrt{(2\pi\theta)}} \exp\left\{-\frac{x^2}{2\theta}\right\}$$

We easily obtain

$$f'' = \left(\frac{x^2}{\theta^2} - \frac{1}{\theta}\right) f,$$

so that

$$\frac{\partial}{\partial\theta}\left(\frac{f''}{f}\right) = \left(\frac{1}{\theta^2} - \frac{2x^2}{\theta^3}\right);$$

† It may not exist, for example, if $f(n_r h, \theta) = 0$, i.e. if the likelihood, ignoring the grouping, of a sample value is zero. This does not contradict the hypothesis since $p(n_r h, \theta)$ may not vanish. We return to this point later.

‡ This we shall do throughout the rest of this paper.

https://doi.org/10.1017/S0305004100025512 Published online by Cambridge University Press

also $$\frac{\partial^2}{\partial\theta^2}(\log f) = \frac{1}{2\theta^2} - \frac{x^2}{\theta^3}$$

which gives $$\Delta = -h^2/12.$$

3. *Extension to two parameters.* The above reasoning may easily be extended to the case where two parameters θ and ϕ have to be estimated. If $\phi_1 = \phi_0 + \Delta'$, the simultaneous equations for Δ and Δ' to order h^4 are

$$\Delta\Sigma\frac{\partial^2}{\partial\theta^2}(\log f) + \Delta'\Sigma\frac{\partial^2}{\partial\theta\partial\phi}(\log f) = -\frac{h^2}{24}\Sigma\frac{\partial}{\partial\theta}\left(\frac{f''}{f}\right),$$

$$\Delta\Sigma\frac{\partial^2}{\partial\phi\partial\theta}(\log f) + \Delta'\Sigma\frac{\partial^2}{\partial\phi^2}(\log f) = -\frac{h^2}{24}\Sigma\frac{\partial}{\partial\phi}\left(\frac{f''}{f}\right),$$

where the summation is over the sample values and the arguments for θ and ϕ are θ_0 and ϕ_0. The extension to n parameters is obvious.

Example II. The mean and variance of a normal population. Simple but lengthy computation gives for

$$f = \frac{1}{\sqrt{(2\pi\theta)}}\exp\left\{-\frac{(x-\phi)^2}{2\theta}\right\}, \quad \Delta' = 0, \quad \Delta = -h^2/12.$$

4. *Comparison with the method of moments.* We have seen in the two examples above, where the estimators are the moments, that the corrections are exactly the same as the Sheppard corrections which one would, in practice, apply to the estimates; i.e. the method of moments and maximum likelihood give the same estimators to $O(h^2)$ for the case of a *grouped* normal population†. However, it is easy to see that the correction Δ, when applied to a moment statistic, need not always be the same as the Sheppard correction to order h^2.

Example III. Consider the type III distribution with end-point at the origin

$$f(x,\theta) = \left(\frac{\lambda}{\theta}\right)^\lambda \frac{1}{\Gamma(\lambda)} x^{\lambda-1} e^{-\lambda x/\theta}.$$

It is easy to verify that $\hat{\theta}$ is the arithmetic mean of the x_i, and that‡

$$\Delta = \frac{h^2}{12}\left(\frac{\lambda-1}{\bar{x}_H} - \frac{\lambda}{\bar{x}_A}\right),$$

where $\bar{x}_H$ is the harmonic mean and $\bar{x}_A$ the arithmetic mean of the observations $n_i h$. The Sheppard correction is, of course, zero.

Thus in a particular case the correction Δ may not agree with Sheppard's correction, but since, if our estimate is the nth sample moment, θ_0 converges in probability (c.i.pr.) to $\bar{\mu}_n$ and θ_1 c.i.pr. to μ_n, then $\Delta = \theta_1 - \theta_0$ c.i.pr. to $\mu_n - \bar{\mu}_n$, the Sheppard correction. Since maximum likelihood methods are only known to be optimum in large samples, it follows that we can only claim merit for Δ in large samples, and therefore we have,

† This result has been discussed by R. A. Fisher, 'On the mathematical foundations of theoretical statistics,' *Philos. Trans.* A, 222 (1922), 362.

‡ It is assumed that no value of n_i is zero; we return to this later.

https://doi.org/10.1017/S0305004100025512 Published online by Cambridge University Press

in effect, justified Sheppard's corrections being used with this method. If the maximum likelihood estimator is not a moment function, then Δ provides a completely new tool (e.g. the estimation of λ in Example III).

5. *Loss of efficiency in grouping.* We know that in the limit the maximum likelihood estimate is distributed normally with variance

$$\left[-nE\left(\frac{\partial^2}{\partial\theta^2}\log f\right)\right]^{-1},$$

where E denotes the mathematical expectation of the variable in brackets, and it is most efficient. Accordingly we can find the loss of efficiency caused by grouping. Expressed as a fraction of the optimum efficiency this is $\operatorname{var}\hat{\theta}/\operatorname{var}\theta_1$. Now

$$E\left(\frac{\partial^2}{\partial\theta^2}\log p(nh,\theta)\right) = \sum_{n=-\infty}^{n=+\infty} p(nh,\theta)\,\frac{\partial^2}{\partial\theta^2}\log p(nh,\theta)\dagger$$

$$= \sum_{n=-\infty}^{n=+\infty}\left[\frac{\partial^2}{\partial\theta^2}\log f + \frac{h^2}{24}\frac{\partial^2}{\partial\theta^2}\left(\frac{f''}{f}\right)\right]\left[hf+\frac{h^3}{24}f''\right]+O(h^4).$$

If the Euler-MacLaurin formula is assumed to be valid, this equals

$$\int f\frac{\partial^2}{\partial\theta^2}\log f\,dx+\frac{h^2}{24}\int\left[f\frac{\partial^2}{\partial\theta^2}\left(\frac{f''}{f}\right)+f''\frac{\partial^2}{\partial\theta^2}\log f\right]dx,$$

whence the efficiency is

$$\frac{\operatorname{var}\hat{\theta}}{\operatorname{var}\theta_1} = 1+\frac{h^2}{24}\frac{E\left[\frac{\partial^2}{\partial\theta^2}\left(\frac{f''}{f}\right)+\frac{f''}{f}\frac{\partial^2}{\partial\theta^2}\log f\right]}{E\left[\frac{\partial^2}{\partial\theta^2}\log f\right]}+O(h^4).$$

Example IV. The variance of a normal population. Using the results given in example I we can easily show that the efficiency is

$$1-\frac{1}{6}\frac{h^2}{\theta},$$

which result has been obtained by R. A. Fisher, op. cit.

6. *Corrections for abruptness.* One does not need to be a pure mathematician to see that the conditions under which the corrections have been obtained are not always satisfied. In particular, they fail when the frequency curve has a finite terminal within a grouping interval. We shall consider the special case where the variate is positive so that there is discontinuity in some derivative at the origin; other cases can be treated analogously. Instead of considering the interval as extending from $-\frac{1}{2}h$ to $+\frac{1}{2}h$ it is better to use the half from 0 to $\frac{1}{2}h$, within which the Taylor expansion will be valid, and obtain

$$\frac{\partial}{\partial\theta}\log p(0,\theta) = \frac{\partial}{\partial\theta}\log f\left(\frac{h}{4},\theta\right)+\frac{h^2}{4.24}\frac{\partial}{\partial\theta}\left(\frac{f''}{f}\right)_{\frac{1}{4}h,\theta}+O(h^4).$$

We then use, instead of

$$\theta_0 = \hat{\theta}(n_1h,\ \ldots,\ n_kh) \quad \text{(which may not exist)},$$

$$\theta_0 = \hat{\theta}(n_1h,\ \ldots,\ n_ih,\ \tfrac{1}{4}h,\ \ldots,\ \tfrac{1}{4}h),$$

† We assume the range is infinite; the general case follows similarly.

https://doi.org/10.1017/S0305004100025512 Published online by Cambridge University Press

where we have supposed the sample numbered so that $n_i \neq 0$ for $i \leqslant l \leqslant k$ and $n_i = 0$ for $i > l$. We obtain

$$\Delta = -\frac{h^2}{24} \frac{\sum_{i=1}^{i=l} \frac{\partial}{\partial\theta}\left(\frac{f''}{f}\right)_{n_i h, \theta_0} + \frac{1}{4}(k-l)\frac{\partial}{\partial\theta}\left(\frac{f''}{f}\right)_{\frac{1}{4}h, \theta_0}}{\sum_{i=1}^{i=l} \frac{\partial^2}{\partial\theta^2}\log f(n_i h, \theta_0) + (k-l)\frac{\partial^2}{\partial\theta^2}\log f\left(\frac{h}{4}, \theta_0\right)}.$$

This method can obviously be extended to the case where both terminals are finite. We do not consider the case where the end-point itself has to be estimated: maximum likelihood methods themselves present difficulties in this case.

Example V. The type III distribution given in example III has a positive variate and is an example of the type we have just discussed. We saw in that example that Δ would be infinite if any $n_i = 0$. The modified correction is easily found to be

$$\Delta = \frac{h^2}{12}\left\{\frac{\lambda - 1}{\bar{x}_H} - \frac{\lambda}{\bar{x}_A}\left(\frac{l + \frac{1}{4}(k-l)}{k}\right)\right\},$$

where $\bar{x}_H$ is the harmonic mean of $(n_1 h, \ldots, n_l h, h, \ldots, h)$ and $\bar{x}_A$ is the arithmetic mean of $(n_1 h, \ldots, n_l h, \frac{1}{4}h, \ldots, \frac{1}{4}h)$.

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

STATISTICAL LABORATORY
CAMBRIDGE

https://doi.org/10.1017/S0305004100025512 Published online by Cambridge University Press