*The Annals of Statistics*
1974, Vol. 2, No. 3, 568–571

# ESTIMATING EQUATIONS IN THE PRESENCE OF A NUISANCE PARAMETER

BY V. P. GODAMBE AND M. E. THOMPSON

*University of Waterloo*

Estimating equations for a real parameter $\theta$ which indexes a family of densities $p(x, \theta)$ were considered in the note by Godambe (*Ann. Math. Statist.* **31** (1960) 1208–1211). An optimality property of the equation $\partial \log p/\partial\theta = 0$ among unbiased estimating equations was established. In this paper an analogous result is proved for estimation of a real parameter $\theta_1$ in the presence of a nuisance parameter $\theta_2$.

Every procedure of point estimation for an unknown real parameter $\theta$ can be viewed as solving for $\theta$ an equation $g(x, \theta) = 0$, $g$ being a real function with arguments $\theta$ and the observed value of the corresponding random variable $x$. The equation $g = 0$ is then called an estimating equation. Denoting by $E_\theta$ the expectation on $\theta$, let

$$\mathscr{G} = \{g : E_\theta(g) = 0 \text{ for all permissible } \theta\}\,. \tag{1}$$

An estimating equation $g = 0$ is called unbiased if $g \in \mathscr{G}$. Restricting ourselves to the class of unbiased estimating equations we define $g^* = 0$ as an *optimum estimating equation* if $g^* \in \mathscr{G}$ and for every other $g \in \mathscr{G}$,

$$E_\theta\left[g^*/E_\theta \frac{\partial g^*}{\partial\theta}\right]^2 \leqq E_\theta\left[g/E_\theta \frac{\partial g}{\partial\theta}\right]^2 \tag{2}$$

for all permissible values of $\theta$. For a proper motivation of this criterion of optimality we refer to Godambe (1960), where it is also proved that under some general regularity conditions to be satisfied by $\mathscr{G}$ above and the underlying frequency function $p(x, \theta)$, (which is supposed to be completely specified up to the unknown parameter $\theta$), $g^*$ in (2) is given by

$$g^* = \frac{\partial \log p}{\partial\theta}\,; \tag{3}$$

thus the maximum likelihood equation $\partial \log p/\partial\theta = 0$ is optimum.

In this article we investigate optimum estimating equations in the presence of a nuisance parameter; that is, now we assume $\theta = (\theta_1, \theta_2)$, $\theta_1$, $\theta_2$ being both real; and we are interested in estimating $\theta_1$ only (ignoring $\theta_2$). Hence we may ask the question, if $\mathscr{G}_1$ is the subclass of $\mathscr{G}$ in (1) consisting of functions which have arguments $x$ and $\theta_1$ only and which satisfy appropriate regularity conditions

Received January 1973; revised July 1973.

*AMS* 1970 *subject classification.* 62.20.

*Key words and phrases.* Estimating equations, maximum likelihood estimation, nuisance parameter.

((ii) to (iv) below), is there a $g^* \in \mathscr{G}_1$ such that for all $g \in \mathscr{G}_1$

$$E_\theta\left[g^*/E_\theta\frac{\partial g^*}{\partial\theta_1}\right]^2 \leqq E\left[g/E_\theta\frac{\partial g}{\partial\theta_1}\right]^2 \tag{4}$$

for all permissible $\theta$? An answer to this question, under the conditions specified below, is given by the theorem to follow.

The frequency function $p(x, \theta)$ is defined on the abstract measurable (measure $\mu$) sample space $\mathscr{X}$ for every value of the parameter $\theta \in \Omega$, the parameter space. Thus the function $p$ is completely specified up to the unknown parameter $\theta$. We assume $\Omega = \Omega_1 \times \Omega_2$ where $\Omega_1 = \{\theta_1\}$, $\Omega_2 = \{\theta_2\}$ and

(a) both $\Omega_1$ and $\Omega_2$ are open intervals of the real line;

(b) for almost all $x(\mu)$, $\partial \log p/\partial\theta_i$, $\partial^2 \log p/\partial\theta_i^2$, $i = 1, 2$ exist for all $\theta \in \Omega$;

(c) $\int p\, d\mu$ and $\int (\partial \log p/\partial\theta_i)p\, d\mu$, $i = 1, 2$ are differentiable under the integral sign for all $\theta \in \Omega$;

(d) $E_\theta(\partial \log p/\partial\theta_i)^2 > 0$, $i = 1, 2$ for all $\theta \in \Omega$.

Next the class of functions $\mathscr{G}_1$ on $\mathscr{X} \times \Omega_1$ referred to above is assumed to satisfy the following conditions: For every $g \in \mathscr{G}_1$

(i) $E_\theta(g) = 0$ for all $\theta \in \Omega$;

(ii) for almost all $x(\mu)$, $\partial g/\partial\theta_1$ exists for all $\theta \in \Omega$;

(iii) $\int gp\, d\mu$ is once differentiable w.r.t. $\theta_1$ and twice w.r.t. $\theta_2$, under the integral sign;

(iv) $[E_\theta(\partial g/\partial\theta_1)]^2 > 0$ for all $\theta \in \Omega$.

With this we have the

THEOREM. *Under the conditions* (a)—(d) *and* (i)—(iv) *above for all* $g \in \mathscr{G}_1$, *a function* $g^* \in \mathscr{G}_1$ *and satisfying* (4) *above is given by*

$$g^* = C_1(\theta_1, \theta_2)\, \partial \log p/\partial\theta_1 + C_2(\theta_1, \theta_2)[(\partial \log p/\partial\theta_2)^2 + (\partial^2 \log p/\partial\theta_2^2)]\,, \tag{5}$$

*provided* $C_1(\theta_1, \theta_2)$ *and* $C_2(\theta_1, \theta_2)$ *in* (5) *are such that the resulting* $g^*$ *is independent of* $\theta_2$, *and satisfies* (ii)—(iv).

PROOF. For all $g \in \mathscr{G}_1$ we have

$$\begin{aligned} E_\theta\left(g/E_\theta\frac{\partial g}{\partial\theta_1}\right)^2 = E_\theta\left[\left(g/E_\theta\frac{\partial g}{\partial\theta_1}\right) - \left(g^*/E_\theta\frac{\partial g^*}{\partial\theta_1}\right)\right]^2 - E_\theta\left[g^*/E_\theta\frac{\partial g^*}{\partial\theta_1}\right]^2 \\ + 2E_\theta\left[\left(g/E_\theta\frac{\partial g}{\partial\theta_1}\right)\left(g^*/E_\theta\frac{\partial g^*}{\partial\theta_1}\right)\right] \end{aligned} \tag{6}$$

where $g^*$ is given by (5). Because of the conditions (i) and (iii) above for $g \in \mathscr{G}_1$,

$$0 = \int gp\, d\mu = \int g[\partial \log p/\partial\theta_1]p\, d\mu + \int (\partial g/\partial\theta_1)p\, d\mu\,;$$

that is, for all $\theta \in \Omega$

$$E_\theta(g[\partial \log p/\partial\theta_1])/E_\theta(\partial g/\partial\theta_1) = -1\,. \tag{7}$$

Further, because of the condition (iii) above we have for $g \in \mathscr{G}_1$

$$0 = \int gp \, d\mu = \int g[\partial \log p/\partial\theta_2]p \, d\mu = \int g[\partial \log p/\partial\theta_2]^2 p \, d\mu + \int g[\partial^2 \log p/\partial\theta_2^2]p \, d\mu \,,$$

i.e. for all $\theta \in \Omega$

$$E_\theta g[(\partial \log p/\partial\theta_2)^2 + (\partial^2 \log p/\partial\theta_2^2)] = 0 \,. \tag{8}$$

Substituting (7) and (8) in (6) we get that $E_\theta(g/E_\theta(\partial g/\partial\theta_1))^2$ is minimized in $\mathscr{G}_1$ for $g = g^*$ given by (5), provided $g^* \in \mathscr{G}_1$. Differentiating twice in the equation

$$1 = \int p \, d\mu$$

and using condition (c) shows that $E_\theta g^* = 0$. Hence the theorem.

EXAMPLE. Let in the above notation $x = (y_1, \cdots, y_n)$ and

$$p(x, \theta) = (1/(2\pi\theta_1)^{\frac{1}{2}})^n \cdot \exp(-\sum (y_i - \theta_2)^2/2\theta_1)$$

for $-\infty < y_i < \infty$, $0 < \theta_1 < \infty$, $-\infty < \theta_2 < \infty$. Then for all $\theta_1$ and $\theta_2$,

$$\partial \log p/\partial\theta_1 = -\frac{n}{2\theta_1} + \frac{\sum (y_i - \theta_2)^2}{2\theta_1^2} = -\frac{n}{2\theta_1} + \frac{(n-1)s^2 + n(\bar{y} - \theta_2)^2}{2\theta_1^2} \,;$$

$$\partial \log p/\partial\theta_2 = \sum (y_i - \theta_2)/\theta_1 = n(\bar{y} - \theta_2)/\theta_1 \,; \qquad \partial^2 \log p/\partial\theta_2^2 = -n/\theta_1 \,;$$

with $\bar{y} = \sum_1^n y_i/n$ and $s^2 = \sum_1^n (y_i - \bar{y})^2/(n-1)$. It is easy to see that substituting $C_2(\theta_1, \theta_2)$ in (5) equal to $-1/2n$ and $C_1(\theta_1, \theta_2) = 1$ we get

$$g^* = \frac{n-1}{2\theta_1^2}(s^2 - \theta_1) \,. \tag{9}$$

Also the verification of the conditions (ii)—(iv) for $g^*$ and (a)—(d) for this example is obvious.

By applying the theorem to this example a second time, this time regarding $\theta_2$ (the mean) as the parameter of interest and $\theta_1$ (the variance) as the nuisance parameter, one can easily show that the equation $\bar{y} - \theta_2 = 0$ is optimal for estimating $\theta_2$.

A method of wide applicability for obtaining a plausible (possibly not optimal) member of $\mathscr{G}_1$ to estimate $\theta_1$ has been proposed by G. A. Barnard (1972). This method can be used, for example, when it is possible to write $\partial \log p/\partial\theta_1$ in the form $\phi_1 + \phi_2$ where (A) $\phi_2$ takes the value 0 when $\theta_2$ is set equal to its maximum likelihood estimate $\hat{\theta}_2$ and (B) $g_1 = \phi_1 + E_\theta \phi_2$ is a function of $x$ and $\theta_1$ only. The estimating equation $g_1 = 0$ seems especially reasonable when it is noted that conditions (A) and (B) hold if appropriate regularity conditions are satisfied and if

$$g_1 = \phi_1 - E_\theta \phi_1 = \frac{\partial \log p}{\partial\theta_1}\bigg|_{\theta_2 = \hat{\theta}_2} - E_\theta\left(\frac{\partial \log p}{\partial\theta_1}\bigg|_{\theta_2 = \hat{\theta}_2}\right)$$

is a function of $x$ and $\theta_1$ only. It is not known whether $g_1$ is optimal in the sense of (4) except in the special case of the example above.

For a different approach to the theory of estimating equations in the multiparameter situation the reader is referred to Bhapkar (1971).

**Acknowledgment.** With pleasure the authors acknowledge that the possibility of some kind of optimality of the estimating equation giving (9) was suggested by G. A. Barnard in a personal conversation. However, the formulation of the optimality criterion (4) and the theorem above is due to the authors.

## REFERENCES

BARNARD, G. A. (1972). Oral communication.

BHAPKAR, V. P. (1972). On a measure of efficiency of an estimating equation. *Sankhyā Ser. A* **34** 467–472.

DURBIN, J. (1960). Estimation of parameters in time series regression models. *J. Roy. Statist. Soc. Ser. B* **22** 139–153.

GODAMBE, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.* **31** 1208–1211.

GODAMBE, V. P. and THOMPSON, M. E. (1973). Estimating equations in the presence of a nuisance parameter (abstract). *Bull. IMS* **2** 57.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WATERLOO
WATERLOO, ONTARIO, CANADA