Analytic Estimator Densities for Common Parameters under Misspecified Models

P.J. Hingley

Abstract. An expression is given for the exact probability density function of the parameter values that maximize the likelihood of a statistical model, where the data generating model is allowed to differ from the estimation model. The density can be used to study the robustness of estimation of alternative hypothetical models. It is described for curved exponential families, then specifically for gamma distribution models and for nonlinear regression models. An example is given in the context of alternative models for data from the biochemical ELISA test method. Finally an indication is given of how a robustness index can be calculated to assess the effects of estimation of a common parameter vector under a wrong model.


Keywords. ELISA, exact estimator density, model robustness, nonlinear regression, small sample properties.

1. Introduction

How can the joint density of the maximum likelihood estimates (MLE) from an estimation model be specified, when the data generating model is different? In this case the MLE are termed quasi maximum likelihood estimates (QMLE). The QMLE still maximize the likelihood with respect to the estimation model, but also minimize the Kullbach-Leibler information criterion between the estimation model and the data generating model (White, 1994). Here an exact analytic expression will be developed for the probability density function of the QMLE, that can be applied to models that are nonlinear with respect to their parameters. It can be used for any sample size, because its derivation does not depend on asymptotic properties.

For estimation models with normal error structures, that are linear with respect to their parameters, QMLE have an exact multivariate normal distribution.
But for nonlinear models, even where estimation and data-generating models are identical, the vector of QMLE only has an asymptotic multivariate normal distribution. Experimental designs with finite numbers of sampling points often lead to markedly non-normal distributions (Bates and Watts, 1988). The proposed method is an alternative to carrying out simulation studies.

The method that will be developed here may be of particular use to scientists who need to take account of the possibility that several possible underlying mathematical models could be generating the data in their experiments.

2. Technique for Estimator Densities (TED)

Consider the $n$ members of a sample as a $(n \times 1)$ vector $w$, $g_0(w)$ as the true density of $w$, and $g_1(w|\theta)$ as a presumed density with $(p \times 1)$ parameter vector $\theta$ to be estimated. The log likelihood corresponding to $g_1(w|\theta)$ is $l(\theta|w)$. The space of $w$ is $W$, and the space of $\theta$ is $\Theta$. A vector of independent variables $z$ can, if necessary, be introduced to cope with the regression situation.

Regularity conditions are (with respect to $\theta$ in each case, and throughout $W$ and $\Theta$): $g_1(w|\theta)$ is continuous; $l(\theta|w)$ is twice differentiable, possesses a single maximum and has no other turning points. The QMLE $\hat{\theta}$ is then given by the same expression as that for the MLE in the usual situation, $l'(\theta, w)|_{\theta=\hat{\theta}} = 0$, (where $'$ indicates differentiation with respect to $\theta$). The space of $\hat{\theta}$ is $\hat{\Theta}$, a subspace of $\Theta$.

Consider a $(p \times 1)$ vector $T$.

$$T(\theta, \theta^*, w) = l'(\theta^*, w) - l'(\theta, w), \quad (2.1)$$

The parameter $\theta^*$ is fixed at an arbitrary value (it is not the asymptotic limit of $\hat{\theta}$, e.g., as represented by White, 1994, as $\theta^*$).

The following Theorem 2.1 and Corollary 2.2 show that $W$ can be divided into distinct subspaces $W_{\hat{\theta}}$, corresponding to the different values of $\hat{\theta}$, so that $w$ determines $T(\hat{\theta}, \theta^*, w)$ within each subspace. For continuously distributed data, $W$ is made up of an uncountably large number of such subspaces $W_{\hat{\theta}}$.

**Theorem 2.1.** For fixed $\theta^*$, each possible value of $\hat{\theta}$ within $\hat{\Theta}$ corresponds to a distinct subspace $W_{\hat{\theta}}$.

**Proof.** It will be shown that the converse of the theorem implies a contradiction. For fixed $\theta^*$, consider two unequal values of $\hat{\theta}$, denoted $\hat{\theta}_1$ and $\hat{\theta}_2$, with overlapping subspaces $W_{\hat{\theta}_1}$ and $W_{\hat{\theta}_2}$. Call the common shared part of these subspaces $W_{\hat{\theta}_1+2}$. Pick any value of $w$ from $W_{\hat{\theta}_1+2}$, and call it $w_c$. Then, from (2.1), $T(\hat{\theta}_1, \theta^*, w_c) = l'(\theta^*, w_c) = T(\hat{\theta}_2, \theta^*, w_c)$. According to the regularity conditions, $l(\theta, w_c)$ possesses a single maximum over $W$ at $\hat{\theta}$, at which point $l'(\theta, w_c)|_{\theta=\hat{\theta}} = 0$ and $T = l'(\theta^*, w_c)$. Therefore $\hat{\theta}_1$ and $\hat{\theta}_2$ are both equal to the unique $\hat{\theta}$. This contradicts inequality for $\hat{\theta}_1$ and $\hat{\theta}_2$, and so $W_{\hat{\theta}_1}$ and $W_{\hat{\theta}_2}$ are distinct. $\square$

**Corollary 2.2.** If $\theta^*$ and $w$ are fixed, $T(\hat{\theta}, \theta^*, w)$ is also fixed.