ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 20, Number 1, Winter 1990

MAXIMUM LIKELIHOOD AND BEST APPROXIMATIONS

A. EGGER

ABSTRACT. That least squares approximation is an appropriate method in the presence of normally distributed errors is a consequence of the fact that the Maximum Likelihood Estimate and Best Approximation Problems coincide in this setting. It is shown that such a relationship holds only for exponentially distributed errors and the ℓ^p norms. Thus there is no norm which is similarly suited for curve fitting or data smoothing in the presence of errors distributed according to, for example, the Cauchy distribution.

Natural sources of approximation problems include curve fitting, signal filtering, data smoothing and parameter estimation. In the discrete setting, the object to be approximated is a vector $z \in \mathbf{R}^n$. Given a set $K \subset \mathbf{R}^n$ of approximating vectors, the Best Approximation Problem is to determine $k \in K$ as close as possible to z. When the distance function is given by a norm $|| \, ||$, this is equivalent to minimizing $||k - z||$ over K. We shall say that $x^* \in K$ is a Best Approximation (BA) from K to z if $||x^* - z|| = \inf_{k \in K} ||k - z||$. There are infinitely many norms on $\mathbf{R}^n$ and, although they are all equivalent, they generate distinct Best Approximation Problems. Among the most frequently considered norms are the ℓ^p norms, $||x||_p = (\Sigma |x_i|^p)^{1/p}$, $1 \leq p < \infty$, and $||x||_\infty = \max\{|x_i| : 1 \leq i \leq n\}$. In a specific problem, the choice of a norm is typically influenced by computational considerations. Often, however, it is known or assumed that the residual $r = z - x^*$ is distributed according to some probability density function ρ. In this setting, there is sometimes a norm which is particularly appropriate.

Suppose that the approximating set K is a subspace of $\mathbf{R}^n$ with basis $V^1, V^2, \ldots, V^k$, where $V^i = (v_1^i, v_2^i, \ldots, v_n^i)$. Assume that $z = \overline{x} + \overline{\varepsilon}$, where $\overline{x} \in K$ and $\overline{\varepsilon} = (\overline{\varepsilon}_1, \overline{\varepsilon}_2, \ldots, \overline{\varepsilon}_n)$ are such that the $\overline{\varepsilon}_i$ are independent random errors distributed according to a probability density ρ. Specifying $x \in K$ to approximate z is equivalent to specifying a vector $\nu = (\nu_1, \ldots, \nu_k)$ of coefficients that generates a residual

Received by the editors on April 6, 1987 and in revised form on June 25, 1987.