Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 84–103

LEAST ABSOLUTE VALUE AND MEDIAN POLISH

BY J. H. B. Kemperman
University of Rochester

We are interested in best L_p-approximations $\Sigma_r\beta_r g_r(x)$ to a given finite array of numbers $z^{(o)}(x)$, $(x \in X)$. For the case $p > 1$, a natural iterated polishing method is shown to converge to the unique optimal solution. Let $p = 1$. Several conditions are obtained, each of which is necessary and sufficient for a given array of residuals $z(x)$ $(x \in X)$ to be optimal. Detailed results are derived for the case of a two-way $m \times n$ layout, allowing several observations $z^{(o)}_{ijk}$ in cell (i, j). For instance, a set of residuals is optimal if and only if there exists a solution to an associated moment problem with given marginals, which depends only on the signs σ_{ijk} of the residuals z_{ijk}. This criterion leads to an elegant and efficient max-flow-min-cut type of algorithm for calculating a best L_1-approximation. For the case of a single observation in each cell, it is also determined precisely which pairs (m, n) are 'safe' for Tukey's median polish, in the sense that the endproduct of an $m \times n$ polish is necessarily a best L_1-approximation. The answer depends on the type of allowable medians.

1. Introduction. Let the $m \times n$ matrix $\mathbf{Z}^{(o)} = (z^{(o)}_{ij})$ represent a two-way table of observations. An elementary way of arriving at a reasonable additive approximation $\alpha_i + \beta_j$ is by means of median polish, as developed by Tukey (1977); see Section 4 for further details. An algorithm in APL and further comments can be found in Anscombe ((1981) p. 106, 382).

One motivating idea behind median polish is that it *might* minimize the L_1-norm of the matrix $\mathbf{Z} = (z_{ij})$ of residuals $z_{ij} = z^{(o)}_{ij} - \alpha_i - \beta_j$. However, this is not always true as follows already from the well-known fact that the norm of the final endproduct of a median polish or mid-median polish may not be the same when starting with a polishing of the rows as when starting with a polishing of the columns.

These endproducts will be called an EMP or EMMP, respectively. More generally, an $m \times n$ matrix $\mathbf{Z}$ will be called an EMP or EMMP, respectively, when 0 is a median or mid-median, respectively of each row and each column.

The matrix $\mathbf{Z}$ of residuals will be said to be *optimal* if its norm cannot be further reduced. For this it is necessary that $\mathbf{Z}$ be an EMP. It is shown (Theorem 6) that for each choice of (m, n) there exist non-optimal EMP's. There even exist non-optimal EMMP's, unless (m, n) is one of the special pairs $(2, n)$; $(3, 4)$; $(4, 4)$; $(4, 5)$ and $(4, 6)$, (assuming that $2 \leq m \leq n$). Thus, if $m = 4$ and $n = 6$ then the endproduct of a convergent mid-median polish process is always optimal. This is false when $m = 3$ and $n = 3$ or 5.

The main purpose of the present paper is to derive efficient tests for optimality together with explicit procedures for improving a given non-optimal matrix. Many of our results lead to an explicit algorithm, usually safer and faster than median polish, though that algorithm may not be spelled out in any great detail. For, our principal goal is to achieve a good theoretical understanding of the main problem.

Most results are developed for the general regression problem, where one wants to minimize the L_p-norm (2.1) by a suitable choice of the free parameters β_r. Median polish carries over to this general problem in a natural way. We show in Theorem 1 that this

[1] Research supported in part by the National Science Foundation.
AMS 1980 subject classifications. Primary, 62J05; secondary, 41A50, 65D10.
Key words and phrases: Linear regression, least absolute value, best L_1 approximation, criterion for optimality, two-way layout, median polish, given marginals, max-flow-min-cut, algorithm.