

Optimal Smoothing of Gaussian Periodic Data

GARRETT BIRKHOFF & ARTHUR S. PRIVER

1. Introduction. A basic problem of the “calculus of observations” [18, Ch. XI] consists in *fitting* discrete data by as “smooth” a function as possible, *interpolated* through the data points. This problem can often be solved by standard interpolation formulas, by using broken-line, spline [1], [8], or other piecewise polynomial functions on a physical, semilog, or loglog scale.

A more general problem arises when the data points have statistical scatter, and when they involve errors of observation (human or instrumental) and/or computation (*e.g.*, roundoff). When the data points are widely spaced, and the scatter and probable errors are small, interpolation may still be the most practical solution. However, when the scatter or probable errors are appreciable and many data points are available, it is usually better to use a *smoothing* scheme based on least squares *approximation* and/or averaging to reduce the probable error.

Many “smoothing” schemes have been proposed in the literature. Clearly, the effectiveness of such smoothing schemes depends on the expected nature or “prior distribution” of (i) the sample function being measured and (ii) the probable measurement errors of the observations. Given this prior distribution, one can define a smoothing scheme to be *optimal* when it minimizes the expected mean square error; and one can then try to determine this optimal scheme *theoretically* by Bayesian techniques, using conditional probabilities.

In the present paper, we solve the resulting optimality problem in what is probably the simplest case: that of *periodic data* measured at equally spaced intervals, and having a homogeneous *normal* (Gaussian) prior distribution, under the assumption that measurement errors are also normally distributed and independent; see §2. The *optimal smoothing* scheme is then given by an easily computed convolution kernel, as we show in §5. (In some design problems, the “optimal” solution may combine closeness of fit with smoothness, for esthetic reasons; see [1, §13].)

Unfortunately, the practical value of the preceding theoretical result is very limited, because *the prior distribution is almost never known in practice*, even for periodic data measured at equally spaced intervals. Even when the prior distribution may be presumed to satisfy the conditions described above, one does not know what the relevant variances are.

In §6, we therefore return to the *practical* problem of smoothing periodic data measured at equally spaced intervals, *when the prior distribution is unknown*.