[27] Sophister (1928). *Biometrika*, **20**, 389–423.

[28] Tchouproff, A. A. (1918). *Biometrika*, **12**, 140–169.

[29] Tchouproff, A. A. (1919). *Biometrika*, **12**, 185–210.

[30] Thiele, T. N. (1903). *Theory of Observations*. C. and E. Layton, London. (Some of the basic concepts relating to cumulants (semi-invariants) are due to Thiele—see recent comments by Hald [15].)

[31] Wishart, J. (1929). *Proc. Lond. Math. Soc., Ser. 2*, **29**, 309–321.

[32] Wishart, J. (1929). *Proc. R. Soc. Edinburgh*, **49**, 78–90.

[33] Wishart, J. (1930). *Biometrika*, **22**, 224–238.

[34] Wishart, J. (1933). *Biometrika*, **25**, 52–60.

Acknowledgment

Research sponsored by the Applied Mathematical Sciences Research Program, Office of Energy Research, U.S. Department of Energy under contract W-7405-eng-26 with the Union Carbide Corporation. (K.O.B.)

(ASYMPTOTIC NORMALITY
EFFICIENCY
ESTIMATION, POINT
FISHER'S *k*-STATISTICS)

K. O. BOWMAN
L. R. SHENTON

METHOD OF SIEVES

Grenander's method of sieves is a general technique through which parametric approaches to *estimation* can be applied to nonparametric problems. Typically, classical approaches such as maximum likelihood* and least squares* fail to produce consistent estimators when applied to nonparametric (infinite dimensional) problems. Thus, for example, the unconstrained maximum likelihood estimator for a density function is not consistent (not even well defined) in the nonparametric case (see Examples 1 and 2 below), and direct application of least squares similarly fails for the nonparametric estimation of a regression function (see Examples 3 and 4 below). Speaking loosely, it might be said that in each case the parameter space (a space of functions) is too large. Grenander [11] suggests the following remedy: perform the optimization* (maximization of the likelihood, minimization of the sum of squared errors, etc.) within a subset of the parameter space, choosing increasingly dense subsets with increasing sample sizes. He calls this sequence of subsets from which the estimator is drawn a *sieve*, and the resulting estimation procedure is his method of sieves. It leads to consistent nonparametric estimators, with different sieves giving rise to different estimators.

The details and versatility of the method are best illustrated by examples; other applications can be found in Grenander [11], wherein the method was first introduced, and in some of the other references.

Example 1. Histogram. Let $x_1, \dots, x_n$ be an independent and identically distributed (i.i.d.) sample from an absolutely continuous distribution with unknown probability density function (p.d.f.) $\alpha_0(x)$. The maximum likelihood estimator for α_0 maximizes the likelihood function

$$\prod_{i=1}^{n} \alpha(x_i). \tag{1}$$

But the maximum of (1) is not achieved within any of the natural parameter spaces for the nonparametric problem (e.g., the collection of all nonnegative functions with area 1). Thus unmodified maximum likelihood is not consistent for nonparametric density estimation.

A sieve is a sequence of subsets of the parameter space indexed by sample size. For each $\lambda > 0$ let us define

$$S_\lambda = \left\{ \alpha : \alpha \text{ is a p.d.f. which is constant on } \left[\frac{k-1}{\lambda}, \frac{k}{\lambda} \right), k = 0, \pm 1, \pm 2, \dots \right\},$$

and allow $\lambda = \lambda_n$ to grow with sample size. $\{S_{\lambda_n}\}$ constitutes a sieve, and the associated (maximum likelihood) method of sieves estimator solves the problem:

$$\text{maximize} \prod_{i=1}^{n} \alpha(x_i) \qquad \text{subject to} \quad \alpha \in S_{\lambda_n}.$$

The well-known solution is the function

$$\hat{\alpha}(x) = \frac{\lambda}{n} \# \left\{ x_i : \frac{k-1}{\lambda_n} \leqslant x_i < \frac{k}{\lambda_n} \right\}$$

$$\text{for} \quad x \in [(\frac{k-1}{\lambda_n}, \frac{k}{\lambda_n})],$$

i.e., the histogram* with bin width λ_n^{-1}. If $\lambda_n \uparrow \infty$ sufficiently slowly, then $\hat{\alpha}$ is consistent, e.g., in the sense that $\int |\hat{\alpha}(x) - \alpha_0(x)|\, dx \to 0$ a.s.

Example 2. Convolution Sieve for Nonparametric Density Estimation*. For the same problem, a different and more interesting sieve is the convolution sieve:

$$S_{\lambda_n} = \left\{ \alpha : \alpha(x) = \int \frac{\lambda_n}{\sqrt{2\pi}} \exp\left[-\frac{\lambda_n^2}{2}(x-y)^2 \right] F(dy), \quad F \text{ an arbitrary c.d.f.} \right\},$$

where λ_n is a nonnegative sequence increasing to infinity. The method of sieves estimator $\hat{\alpha}$ maximizes (1) within the sieve S_{λ_n}. It can be shown [10] that $\hat{\alpha}$ has the form

$$\hat{\alpha}(x) = \sum_{i=1}^{n} p_i \frac{\lambda_n}{\sqrt{2\pi}} \exp\left\{ -\frac{\lambda_n^2}{2}(x-y_i)^2 \right\}$$

for some $y_1, \dots, y_n$ and $p_1, \dots, p_n$ satisfying $p_i \geqslant 0$, $1 \leqslant i \leqslant n$, $\sum_{i=1}^{n} p_i = 1$. It can also be shown that $\{y_1, \dots, y_n\} \neq \{x_1, \dots, x_n\}$ (with probability 1). Thus the convolution sieve defines an estimator closely related to, but distinct from, the Parzen-Rosenblatt Gaussian kernel estimator. Observe that the latter is in the sieve S_{λ_n}: take F to be the empirical distribution function. But the maximum of the likelihood is achieved by using a different distribution. As with the Parzen-Rosenblatt estimator, if $\lambda_n \uparrow \infty$ sufficiently slowly (i.e., the "window width" is decreased sufficiently slowly), then the estimator is consistent. For details see Geman and Hwang [9], and for an interesting discussion of this and related estimators from a different point of view, see Blum and Walter [2].

Example 3. Splines* for Nonparametric Regression. Let X and Y be random variables and let $(x_1, y_1), \dots, (x_n, y_n)$ be an i.i.d. sample from the bivariate distribution of (X, Y). The least squares estimator of the regression function $E(Y \mid X = x)$ minimizes

$$\sum_{i=1}^{n} \{ y_i - \alpha(x_i) \}^2. \tag{2}$$

Observe that the minimum is zero and is achieved by any function that passes through all of the points of observation, $(x_1, y_1), \dots, (x_n, y_n)$. Excepting some very special cases, this set does not in any useful sense converge to the true regression.

For any nonnegative sequence $\lambda_n \uparrow \infty$ define a sieve $\{S_{\lambda_n}\}$ as follows:

$$S_{\lambda_n} = \left\{ \alpha : \alpha \text{ absolutely continuous}, \int \left| \frac{d}{dx} \alpha(x) \right|^2 dx \leqslant \lambda_n \right\}.$$

The least squares method of sieves estimator, $\hat{\alpha}$, for the regression function is the function in S_{λ_n} minimizing (2). The unique minimum is a first-degree polynomial smoothing spline, i.e., $\hat{\alpha}$ is continuous and piecewise linear with discontinuities in $d\hat{\alpha}/dx$ at $x_1, \dots, x_n$ (see ref. 15). It is possible to show that if λ_n increases sufficiently slowly, then the estimator is strongly consistent for $E(Y|X = x)$ in a suitable metric (details are in ref. 8).

Example 4. Dirichlet Kernel for Nonparametric Regression. Recall the nonparametric regression problem discussed in the previous example. Let us here take x, the "independent" variable, to be deterministic. We then think of the distribution of Y as being an unknown function of x, $F_x(\cdot)$. For this example, we assume $x \in [0, 1]$. The problem is then to estimate

$$\alpha_0(x) = E_x[Y] \equiv \int_{-\infty}^{\infty} y F_x(dy), \quad x \in [0, 1],$$

from independent observations $y_1, \dots, y_n$, where $y_i \sim F_{x_i}$, and $x_1, \dots, x_n$ is a deterministic, so-called design, sequence. For example, assume that the design sequence for fixed n is equally spaced on the interval [0, 1] with

$$x_i = \frac{i}{n}, \qquad i = 1, 2, \ldots, n.$$

As with the previous example, unconstrained minimization of the sum of squares of errors, (2), does not produce a useful estimator. Introduce the Fourier sieve

$$S_m = \left\{ \alpha(x) : \alpha(x) = \sum_{k=-m}^{m} a_k e^{2\pi ikx} \right\};$$

S_m is particularly tractable and makes for a good illustration of the method in this setting. The sieve size is governed by the parameter m, which is allowed to increase to infinity with n. If we restrict m_n so that $m_n \leqslant n$ for all n, then $\hat{\alpha}$ is uniquely defined by requiring that it minimize (2) subject to $\alpha \in S_{m_n}$. A simple calculation gives the explicit form:

$$\hat{\alpha}(x) = \frac{1}{n} \sum_{i=1}^{n} y_i D_{m_n}(x - x_i)$$

where D_m is the Dirichlet kernel

$$D_m(x) = \frac{\sin \pi(2m+1)x}{\sin \pi x}.$$

Kernel estimators for nonparametric regression have been widely studied, although from a somewhat different point of view. See refs. 1, 4, 6, 16, and 17 for some recent examples. It is not difficult to exploit this simple form for $\hat{\alpha}$. Depending on the rate at which $m_n \uparrow \infty$, and depending on assumptions about α_0, consistency, rates of convergence, and asymptotic distribution can be established [8].

What makes this example particularly tractable is that the estimator is based on a sieve that consists of increasing subspaces of a Hilbert space. Nguyen and Pham [14] used sieves of this type to estimate the drift function of a repeatedly observed nonstationary diffusion.

Example 5. Nonparametric Estimation of the Drift Function of a Diffusion. From an observation of a sample path of a diffusion process* one can construct consistent estimators for the diffusion drift. If the form of the drift function is known up to a finite collection of parameters, then it is possible to use maximum likelihood and obtain consistent and asymptotically normal estimators (see Brown and Hewitt [3], Feigin [5], Lee and Kozin [12], and Lipster and Shiryayev [13]). But unconstrained maximum likelihood fails in the nonparametric case.

More precisely, let us consider a diffusion process x_t defined by

$$dx_t = \alpha_0(x_t)\,dt + \sigma\,dw_t, \qquad x_0 = x_0,$$

with w_t a standard (one-dimensional) Brownian motion* and x_0 a constant. α_0 and σ are assumed to be unknown; we wish to estimate α_0 from an observation of a sample path of x_t. It is well known that the distribution of x_s, $s \in [0, t]$, is absolutely continuous with respect to the distribution of σw_s, $s \in [0, t]$ (assuming some mild regularity condition on α_0). A likelihood function for the process x_s, $s \in [0, t]$ is the Radon-Nikodym derivative:

$$\exp\left\{ \int_0^t \alpha_0(x_s)\,dx_s - \frac{1}{2} \int_0^t \alpha_0(x_s)^2\,ds \right\}. \qquad (3)$$

The maximum likelihood estimator for α_0 maximizes (3) over a suitable parameter space, most appropriately the space of uniformly Lipschitz continuous functions. But the maximum of the likelihood is not attained, either in this or in any other of the usual function spaces. In a manner analogous to the previous examples, a sieve S_t can be introduced (here indexed by time) and an estimator $\hat{\alpha}$ defined to maximize (3) subject to $\alpha \in S_t$. Provided that the sieve growth is sufficiently slow with respect to t, this method of sieves estimator can be shown to be consistent: $\hat{\alpha} \to \alpha_0$, in a suitable norm, as $t \to \infty$. Details are in Geman [7].

References

[1] Ahmad, I. A. and Lin, P. E. (1976). *Bull. Math. Statist.*, **17**, 63–75.

[2] Blum, J. and Walter, G. (1982). *Technical Report Series of the Intercollege Division of Statistics*, No. 43, University of California, Davis, Calif.

[3] Brown, B. M. and Hewitt, J. I. (1975). *J. Appl. Prob.*, **12**, 228–238.

[4] Devroye, L. P. and Wagner, T. J. (1980). *Ann. Statist.* **8**, 231–239.

[5] Feigin, P. D. (1976). *Adv. Appl. Prob.*, **8**, 712–736.

[6] Gasser, T. and Muller, H. G. (1979). *Smoothing Techniques for Curve Estimation*. T. Gasser and Rosenblatt, eds. Lecture Notes in Mathematics, Springer-Verlag, New York.

[7] Geman, S. (1980). *Colloquia Mathematica Societatis Janos Bolyai*, Vol. *32*. Nonparametric Statistical Inference. North-Holland, Budapest.

[8] Geman, S. (1981). *Reports on Pattern Analysis*, No. 99, Division of Applied Mathematics, Brown University, Providence, RI.

[9] Geman, S. and Hwang, C. R. (1982). *Ann. Statist.*, **10**, 401–414.

[10] Geman, S. and McClure, D. E. (1982). *Proceedings of the NASA Workshop on Density Estimation and Function Smoothing*. L. F. Guseman, Jr., ed. Texas A & M University, College Station, TX, 38–47.

[11] Grenander, U. (1981). *Abstract Inference*. Wiley, New York.

[12] Lee, T. S. and Kozin, F. (1977). *J. Appl. Prob.*, **14**, 527–537.

[13] Lipster, R. S. and Shiryayev, A. N. (1978). *Statistics of Random Processes* Vol II: *Applications*. Springer-Verlag, New York, Chap. 17.

[14] Nguyen, H. T. and Pham, T. D. (1982). *SIAM J. Control and Optimization*.

[15] Schoenberg, I. J. (1964). *Proc. Nat. Acad. Sci.*, **52**, 947–950.

[16] Schuster, E. and Yakowitz, S. (1979). *Ann. Statist.*, **7**, 139–149.

[17] Spiegelman, C. and Sacks, J. (1980). *Ann. Statist.*, **8**, 240–246.

(DENSITY ESTIMATION
ESTIMATION, POINT
KERNEL ESTIMATORS
LEAST SQUARES
MAXIMUM LIKELIHOOD ESTIMATION)

STUART GEMAN

METRIC NUMBER THEORY *See* PROBABILISTIC NUMBER THEORY

METRICS AND DISTANCES ON PROBABILITY SPACES *See* PROBABILITY SPACES, METRICS AND DISTANCES ON

METRICS, IDEAL

This concept was introduced by Zolotarev [1], who discussed applications to mathematical statistics in some detail [2], and later presented further developments [3]. The notion is useful in problems of approximating distributions of random variables obtained from independent random variables by successive application of addition, multiplication, taking maxima, or some other "group operations."

References

[1] Zolotarev, V. M. (1976). *Mat. Sbornik*, **101 (144)**, No. 3 (11), 416–454 (in Russian).

[2] Zolotarev, V. M. (1979). *Austr. J. Statist.*, **21**, 193–208.

[3] Zolotarev, V. M. (1981). In *Stability Problems for Stochastic Models, Proc. Sem. Moscow Inst. System Studies*, 1981, pp. 30–39.

(APPROXIMATIONS TO DISTRIBUTIONS)

METRIKA

The journal *Metrika* bears the subtitle *International Journal for Theoretical and Applied Statistics*. It appears quarterly, starting with volume 1 in 1958. In the course of time the number of pages has increased up to nearly 300 per volume (= 4 fasc.). There are no auxiliary publications.

Research papers and, very rarely, survey papers are published. As expressed in the title, published articles belong to the field of mathematical statistics (see Fig. 1). During the starting years this concept was understood in a wider sense, but now, because of the large number of submitted manuscripts, only articles on statistics in a narrower sense are accepted, i.e., only those on statistical methods and mathematical statistics. Great importance is attached to applicability of proposed and investigated methods.

Articles written in German or in English are acceptable. Far more than half the papers are submitted in English. Besides the actual articles each volume also contains book reviews. Whereas formerly there was a large number of brief reviews, future issues will review fewer books in greater detail.