Cross-Validated Spline Methods for the Estimation of Three-Dimensional Tumor Size Distributions From Observations on Two-Dimensional Cross Sections

DOUGLAS NYCHKA, GRACE WAHBA, STANLEY GOLDFARB, and THOMAS PUGH*

We study the problem of estimating the distribution of the three-dimensional radiuses of a collection of spheres, given measurements of the two-dimensional radiuses of a sample of planar cross sections. This problem arises in the estimation of the tumor size distribution of spherical microtumors induced in mouse livers following injection of a carcinogen. We first convert this problem to a form suitable for the application of cross-validated spline methods for the solution of ill-posed integral equations given noisy data. Then we develop special numerical techniques that will allow the spline methods to be accurately applied to integral equations like those associated with the present problem. We apply the resulting method to some mouse-liver data. The subject mouse liver has been completely dissected, allowing a rare comparison of the estimate with the "truth." The statistical properties of the estimate are explored via Monte Carlo methods. The interplay between statistical and numerical analytic methods for problems like this are explored and the use of eigensequence plots for studying "ill posedness" is described.

KEY WORDS: Random spheres model; Stereology; Cross-validated splines; Tumor size distribution; Ill-posed problem.

1. INTRODUCTION

We have been working with data from experiments in pathology studying the growth of microtumors (hepatocellular foci) in the livers of mice (see Koen, Pugh, and Goldfarb, 1983). Mice are injected at 15 days of age with a carcinogen that induces the formation of malignant tumors in the liver. After a fixed period of time the mice are sacrificed, and samples of liver tissue are stained and embedded in a paraffin block. The matrix of paraffin enables the sample to be sliced thinly, and these slices are mounted on microscope slides. Tumors in the sample will now appear in cross section on these slides, and their cross-sectional area or radiuses, if spherical, can be measured.

It is desired to estimate the number density and three-dimensional size distribution of the liver tumors from the cross-sectional observations. In these particular experiments, a single mouse liver may contain anywhere from a few to several hundred microtumors. Different mathematical models for tumor growth have different implications for the variation of tumor size distribution with mouse age. Thus it is desired to identify tumor size distributions for groups of experimental animals sacrificed at different times after the exposure to the carcinogen. These growth models are important because they might suggest some of the mechanisms that initiate and promote liver cancer. By the limitations of the dissection procedure, tumors can only be identified by their cross sections. Since tumors of different sizes can produce the same size cross sections, there is not a direct correspondence between the cross-sectional data and the distribution of tumor sizes. Although it is possible to take many, closely spaced slices and completely reconstruct each tumor, this procedure is both tedious and costly. What is required is a statistical method that estimates the three-dimensional tumor size distribution from observations of two-dimensional cross sections from a modest number of slices.

The biology of the liver suggests that the tumors will be uniformly distributed throughout the tissue, and examination of successive cross sections has indicated that the tumors are roughly spherical. These assumptions suggest a model from geometric probability. Consider a medium that contains spheres whose centers are distributed according to a Poisson process in space with constant intensity and whose equatorial radiuses are distributed according to the cumulative distribution function $F_3(r)$. It is assumed that the tumor number density is small enough so that distinct spheres do not interfere with one another. Now suppose this medium is sliced in a manner independent of the spheres' sizes and locations. Let $F_2(x)$ denote the cumulative distribution function of the (two-dimensional) cross-sectional radiuses from randomly selected slices. The relationship between F_2 and F_3 was

* Douglas Nychka is Assistant Professor, Department of Statistics, North Carolina State University, Raleigh, NC 27650; Grace Wahba is Professor, Department of Statistics, Stanley Goldfarb is Professor and Thomas Pugh is Project Associate, Department of Pathology and Laboratory Medicine, University of Wisconsin, Madison, WI 53706. This research was sponsored by Office of Naval Research Contract N00014-77-C-0675, Army Research Office Contract DAAG29-80-K-0042, and National Institute of Health Grants CA15664 and CA25522.

© Journal of the American Statistical Association
December 1984, Volume 79, Number 388
Theory and Methods Section

derived by Wicksell (1925) and is

$$F_2(x) = 1 - \frac{1}{\mu}\int_x^R \sqrt{r^2 - x^2}\, dF_3(r), \qquad R \geq x \geq 0, \quad (1.1)$$

where R is an upper bound for the maximum possible value of r and μ is the mean (three-dimensional) radius,

$$\mu = \int_0^R r dF_3(r). \qquad (1.2)$$

Equation (1.1) is obtained by a conditioning argument. If a single sphere of radius r is cut by a particular plane, then the distance from the cutting plane to the center of the sphere is equally likely to be anywhere between 0 and r, and the cdf of the cross-sectional radius is $F_2(x) = 1 - \sqrt{(r^2 - x^2)}/r$ $(0 \leq x \leq r)$. The probability that a sphere of radius r will be cut is proportional to its radius times its relative frequency in the sphere population.

In practice, tissue slices are parallel and uniformly spaced (see Figure 1), and the orientation of the cutting planes is chosen to maximize the cross section of the cuts. For (1.1) to hold, it is only necessary to assume that the sphere centers are distributed so that the preceding conditioning argument holds.

In this work we will usually be acting as though we are sampling from some population of tumors that possess a density f_3. The problem is: Given a sample from F_2, obtain a good estimate for the density $f_3(r) = F_3'(r)$. In practice, tumor cross sections can only be observed if they are larger than some radius ϵ. In this case, clearly the experiment does not provide information concerning $f_3(x)$ for $x \leq \epsilon$. However, an integral relationship between the two-dimensional distribution, conditional on $x \geq \epsilon$, and $f_3(x)$ for $x \geq \epsilon$, can still be obtained. This was observed by Chover and King (1981), and their derivation follows. Let F_2^ϵ be the conditional distribution of x given $x \geq \epsilon$. Defining μ_ϵ by

$$\mu_\epsilon = \int_\epsilon^R \sqrt{r^2 - \epsilon^2}\, f_3(r)dr, \qquad (1.3)$$

it follows from (1.1) that

$$1 - F_2(\epsilon) = \frac{\mu_\epsilon}{\mu}, \qquad (1.4)$$

hence

$$1 - F_2^\epsilon(x) = \frac{1 - F_2(x)}{1 - F_2(\epsilon)} = \frac{\mu}{\mu_\epsilon}(1 - F_2(x)). \qquad (1.5)$$

Substituting (1.5) into (1.1) gives

$$F_2^\epsilon(x) = 1 - \frac{1}{\mu_\epsilon}\int_x^R \sqrt{r^2 - x^2}\, f_3(r)dr. \qquad (1.6)$$

The problem now is to estimate $f_3(r)$, $r \geq \epsilon$ (or, rather, $f_3^\epsilon(r) = f_3(r)/(1 - F_3(\epsilon))$, given a sample from F_2^ϵ.

The problem of estimating the distribution of sphere sizes in a medium from the cross sections of a randomly oriented slice, given a sample cumulative distribution function from F_2, is a classical problem in stereology.

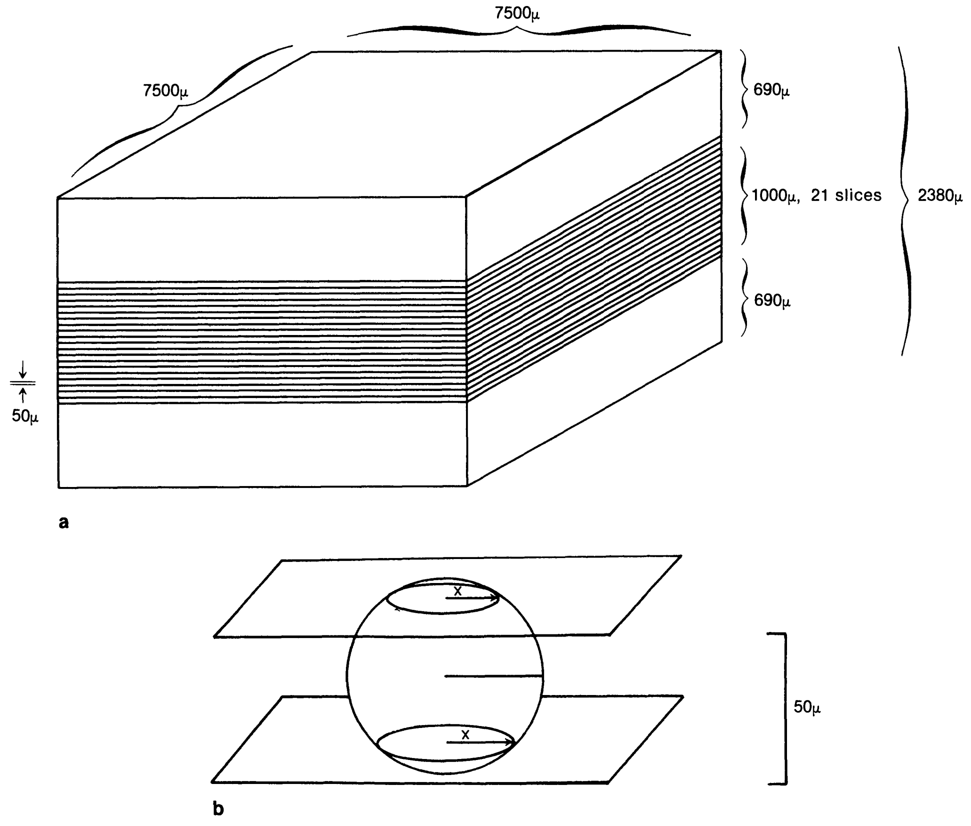

Figure 1. Schematic Diagram of the Slicing Design: (a) Slicing Design; (b) Detail of Sphere Intersected by Two Slices.

For the case $\epsilon = 0$, several approaches have been proposed, including maximum likelihood, regression, and nonparametric methods (see Keiding, Jensen, and Ranek 1972; Nicholson and Merck 1969; Nicholson 1970,1976; and Tallis 1970). Recently, Kuk (1982) has placed this problem in the context of estimating a mixing distribution. Watson (1971) discussed the estimation of moments of f_3. Anderssen and Jakeman (1975) obtained an estimate of f_3 from the inversion formula

$$f_3(r) = \frac{d}{dr}\frac{2}{\pi}\int_r^R [dF_2(x)]/\sqrt{x^2 - r^2}.$$

They use spectral differentiation and product integration to evaluate the integral. Mendelsohn and Rice (1982) studied a similar problem in which the desired density g and the density h from which observations are made are related by

$$h(r) = \int w(r, x)g(x)dx \qquad (1.7)$$

for a normal kernel w. Their work is somewhat related to the work described here and will be discussed later.

The problem of recovering estimates of f_3 from observations on f_2 is harder than might appear at first glance because it is ill posed. Here this means that large changes in the true f_3 lead to changes in the sample histogram that are imperceptible compared to the sampling error. In particular, high-frequency components in f_3 will not in general be recoverable from medium or even large samples from F_2^ϵ. For this reason parametric methods (if a parametric form is known) or nonparametric methods, which estimate a smooth solution, are most likely to be successful. If the true solution is smooth, then a good nonparametric smoothing method is a promising candidate for recovering the "truth." If the truth is not smooth, then such a method should recover the smooth part of f_3. Similar remarks have appeared in Anderssen and Jakeman (1975), Mendelsohn and Rice (1983) and elsewhere, but are worth repeating.

In Section 2.1 we show how the problem of estimating f_3^ϵ from a sample from f_2 can be converted to the problem of solving an integral equation given noisy data. We can then apply cross-validated spline methods for solving ill-posed integral equations. These methods have been shown to be successful in a variety of applications (see Crump and Seinfeld 1982, Merz 1980, and Wahba 1977,1979,1980,1982a,b).

In Section 2.2 we develop a numerical algorithm using certain carefully matched quadrature approximations, which are particularly suited to the application of cross-validated spline methods to integral equations like (1.6).

In Section 3 we apply the estimation procedure to a sample of cross-sectional mouse-liver data obtained by two of us (Goldfarb and Pugh). The mouse liver from which this data was taken was exhaustively dissected, and the true distribution of the three-dimensional tumors from the subject mouse was determined. Thus we have a unique opportunity to compare the estimated distribution with an actual distribution in circumstances that accurately reflect laboratory experiments.

The results appear to be quite successful.

Convergence properties of this estimate can be obtained by adapting known techniques for regularized solutions to ill-posed linear-operator equations (e.g., see Cox 1983, Lukas 1981, Silverman 1983, and Wahba 1977). The results appeared in Nychka (1983). More to the immediate point, the experimenter would like to know how well the method will recover size distributions with a sample size and slicing design similar to those encountered in practice. We have designed a Monte Carlo experiment to answer this question for an experiment similar to the laboratory experiment described in Section 3. This experiment is in the spirit of the recent landmark paper of Diaconis and Efron (1983). Some of the results are given in Section 4. In general, the accuracy of the estimate is quite impressive, considering the modest sample size and ill posedness of the problem. It is, however, difficult to estimate $f_3(r)$ for r near ϵ with sample sizes like those in Section 3. This is not surprising considering that f_3 is subject to length-biased sampling and that large tumors can give rise to both large and small cross sections. Thus information in the data concerning the behavior of f_3 near ϵ is scanty. The method described here extrapolates from data-rich to data-poor regions of r in a linear manner. In Section 5 we describe how a priori information concerning the behavior of f_3 near ϵ can, if available, be incorporated into the estimate.

In Section 6 we show how certain eigensequence plots can provide important insight into the precise degree of ill posedness of this problem, and we discuss the effects of "binning" the data.

In Section 7 some related methods are described, and we describe the very important interplay between statistical smoothing methods and approximation theoretic methods such as quadrature and finite-element methods.

2. CROSS-VALIDATED SPLINE METHODS FOR ILL-POSED LINEAR-OPERATOR EQUATIONS

2.1 The Cross-Validated Spline Estimate $f_{\hat\lambda}$ for f_3

Let $\mathcal{H}$ be the (Sobolev) Hilbert space of real-valued functions on $[\epsilon, R]$: $\mathcal{H} = \{h\colon h, h'$ absolutely continuous, $h'' \epsilon \mathcal{L}_2[\epsilon, R]\}$. The (usual) model behind cross-validated spline methods for integral equations is

$$z_i = L_i h + \epsilon_i, \qquad i = 1, 2, \ldots, n, \qquad (2.1)$$

where the $\{\epsilon_i\}$ are independent zero-mean random variables with common unknown variance, and $L_1, \ldots, L_n$ are bounded linear functionals on $\mathcal{H}$ (see Wahba 1977,1978,1980,1982a,b). Given data $z = (z_1, \ldots, z_n)'$, the cross-validated spline estimate $h_{\hat\lambda}$ for h is obtained as the minimizer in $\mathcal{H}$ of

$$\frac{1}{n}\sum_{i=1}^n (L_i h - z_i)^2 + \lambda\int_\epsilon^R (h''(r))^2 dr, \qquad (2.2)$$

where the smoothing (bandwidth) parameter is taken as the generalized cross validation (GCV) estimate of λ (see Craven and Wahba 1979).

In the problem under study, let $\hat{F}_2^\epsilon$ be the sample cdf of the cross-sectional radiuses, let $\{P_i\}_{i=1,n}$ be a partition of the interval $[\epsilon, R]$, $\epsilon = P_1 < P_2 < \cdots < P_n < R$, and let z_i be the fraction of all observations in the ith bin, $[P_i, P_{i+1})$.

Then

$$z_i = \hat{F}_2^\epsilon(P_{i+1}) - \hat{F}_2^\epsilon(P_i)$$
$$= F_2^\epsilon(P_{i+1}) - F_2^\epsilon(P_i) + \epsilon_i, \quad (2.3)$$

where the ϵ_i are random variables. If the observations are an independent sample from F_2^ϵ, then the $\{\epsilon_i\}$ will have zero mean and be jointly asymptotically normal and only weakly correlated. In this work we are going to ignore the fact that the variances of the ϵ_i are not necessarily the same. (Various reweighting schemes are possible; see Cox 1970 and Villalobos and Wahba 1982.) Letting $h = f_3/\mu_\epsilon$, and setting

$$L_i h = \int_{P_i}^{R} \sqrt{r^2 - P_i^2}\, h(r)dr$$
$$- \int_{P_{i+1}}^{R} \sqrt{r^2 - P_{i+1}{}^2}\, h(r)dr, \quad (2.4)$$

(2.3) becomes (with the aid of (1.6)) $z_i = L_i h + \epsilon_i$. Given z, we let h_λ be the minimizer of (2.2) in $\mathscr{H}$ and let f_λ be

$$f_\lambda(r) = h_\lambda(r) \Big/ \int_\epsilon^R h_\lambda(s)ds. \quad (2.5)$$

Our estimate $\hat{f}_3$ is then $f_{\hat\lambda}$, where $\hat\lambda$ is the GCV estimate of λ. (Note that $\int_\epsilon^R h_{\hat\lambda}(s)ds$ is an estimate for $1/\mu_\epsilon$.) The estimate obviously integrates to 1, but it is not required to be positive. Negativity was not a problem with the actual mouse-liver data. In one of the Monte Carlo examples the estimate went negative, and we have truncated the estimate in the plots. If desired, non-negativity constraints can be added to the problem of (2.2) (see Wahba 1982a and Villalobos 1983).

We remark that a penalized log-likelihood estimate for $g = \log f_3^\epsilon$ may be defined by extending the results of Silverman (1983) and by differentiating (1.6) to obtain a relationship between f_2^ϵ and g. O'Sullivan (1983) has recently shown how to use GCV to estimate the smoothing parameter in similar estimates. While the penalized log-likelihood estimate appears to be more computationally burdensome than the estimate under study here, we believe that further investigation is warranted. Note that the smoothness penalty in such a method is applied to log f_3^ϵ as opposed to f_3^ϵ.

2.2 The Numerical Method for Computing $f_{\hat\lambda}$

Using known but scattered results, we next give an efficient numerical procedure for computing (a very good approximation to) the minimizer of (2.2) and the GCV estimate $\hat\lambda$ of λ. The method is readily implemented for n less than a few hundred. In all of our calculations, n will be 80, and the bins are equally spaced in log x between ϵ and R. The log spacing is a crude variance-stabilizing spacing for our mouse-liver data. For the actual and most of the Monte Carlo data, the number of observed cross sections was between 150 and 450. The choice of $n = 80$ bins is large enough so that the binning is not doing any appreciable smoothing. Binning as smoothing will be discussed in further detail in Section 6.

As with any ill-posed problem, care must be taken in the actual calculation of the solution, or garbage may result from dividing random or roundoff errors by small eigenvalues. It will be seen here and in Sections 6 and 7 that the numerical analysis and the estimation procedure can become inextricably intertwined in ill-posed problems. Approximation-theoretic methods become smoothing procedures and vice versa. For completeness, and to allow discussion of this point, we outline the major steps of our numerical method here, pointing out the steps developed particularly for the problem at hand.

Using the results in Kimeldorf and Wahba (1971), Wahba (1978), and Wahba and Wendelberger (1980), an explicit formula for h_λ, the minimizer of (2.2) in $\mathscr{H}$, can be given as follows. Under the inner product

$$\langle f, g\rangle_{\mathscr{H}} = f(\epsilon)g(\epsilon) + f'(\epsilon)g'(\epsilon) + \int_\epsilon^R f''(r)g''(r)dr,$$

$\mathscr{H}$ is a reproducing-kernel Hilbert space. The reproducing kernel for $\mathscr{H}$ with this inner product is

$$Q(r, s) = 1 + (r - \epsilon)(s - \epsilon) + Q_1(r, s),$$
$$\epsilon \le r, s \le R, \quad (2.6)$$

where

$$Q_1(r, s) = \frac{(r - \epsilon)^2(s - \epsilon)}{2} - \frac{(r - \epsilon)^3}{6}, \qquad r \le s$$
$$= \frac{(r - \epsilon)(s - \epsilon)^2}{2} - \frac{(s - \epsilon)^3}{6}, \qquad r \ge s.$$

Let $\phi_1(r) = 1$, $\phi_2(r) = (r - \epsilon)$, and $\xi_i(r) = L_i(Q_1(\cdot, r))$, where L_i is given by (2.4) and $L_i(Q_1(\cdot, r))$ means that L_i is applied to $Q_1(s, r)$ considered as a function of s. Let T be the $n \times 2$ matrix with $i\nu$th entry $\tau_{i\nu} = L_i\phi_\nu$ ($\nu = 1, 2$), and let K be the $n \times n$ matrix with ijth entry $k_{ij} = \int^R \xi_i''(r)\xi_j''(r)dr$. If T is of rank 2, h_λ is uniquely determined and given by

$$h_\lambda(r) = \sum_{i=1}^{n} c_i\xi_i(r) + \sum_{\nu=1}^{2} d_\nu\phi_\nu(r), \quad (2.7)$$

where $c = (c_1, \ldots, c_n)'$ and $d = (d_1, d_2)'$ satisfy

$$(K + n\lambda I)c + Td = z, \quad (2.8)$$

$$T'c = 0. \quad (2.9)$$

The GCV estimate $\hat\lambda$ of λ is the minimizer of the cross-validation function $V(\lambda)$,

$$V(\lambda) = 1/n(\,\|(I - A(\lambda))z\,\|^2)/((1/n)\operatorname{tr}(I - A(\lambda)))^2\,, \quad (2.10)$$

where $A(\lambda)$ is the $n \times n$ influence matrix defined by

$$\begin{bmatrix} L_1(h_\lambda) \\ \vdots \\ L_n(h_\lambda) \end{bmatrix} = A(\lambda)z.$$

From, for example, Wahba and Wendelberger (1980) it is known that

$$I - A(\lambda) = Q(QKQ' + n\lambda I)^{-2}Q', \qquad (2.11)$$

where Q can be taken as any $n \times n - 2$ matrix whose $n - 2$ columns are linearly independent and perpendicular to the two columns of T. The numerical problem now is to compute the minimizer $\hat{\lambda}$ of $V(\lambda)$ and $h_{\hat{\lambda}}$.

In this problem closed-form expressions can be obtained for the $\{\xi_i\}$ and $\{\tau_{i\nu}\}$ and are given in Appendix A. Unfortunately we were unable to find a closed-form expression for $k_{ij} = \int_\epsilon^R \xi_i''(r)\xi_j''(r)dr$, so some form of quadrature must be used. It is not at all clear that just applying the nearest handy quadrature formula to obtain approximations to the entries k_{ij} of K is appropriate. In particular, the non-negative definiteness of K could easily be lost, leading to problems in the calculation of $\hat{\lambda}$.

The following form of matched quadrature can be used to avoid this problem. The particular form of matched quadrature chosen is motivated by (a) the fact that $\xi_i^{(\nu)}(\epsilon) = 0$ ($\nu = 0, 1, i = 1, 2, \ldots, n$) and (b) the desire to do as little quadrature approximation as possible by exploiting the known closed-form expressions for ξ_i and $\tau_{i\nu}$.

First, let η_i be the representer of L_i in $\mathscr{H}$; that is, $L_ih = \langle\eta_i, h\rangle$. It is known that $\eta_i = \xi_i + a_{i1}\phi_1 + a_{i2}\phi_2$ for some a_{i1}, a_{i2} (e.g., see Kimeldorf and Wahba 1971). Note that (2.2) may be rewritten as

$$\frac{1}{n}\sum_{i=1}^{n} (\langle\eta_i, h\rangle - z_i)^2 + \lambda\int_\epsilon^R (h''(r))^2dr.$$

Now choose a fine grid of $N + 1$ points, $\epsilon = s_0 < s_1 < s_2 < \cdots < s_N = R$, and for any h let P_Nh be that element in $\mathscr{H}$ that minimizes $J(h)$ subject to $(P_Nh)(s_l) = h(s_l)$ ($l = 0, 1, 2, \ldots, N$) and $(P_Nh)'(\epsilon) = h'(\epsilon)$. P_Nh will be a cubic interpolating spline subject to the left-boundary conditions. The matched quadrature consists of approximating L_i by $\tilde{L}_i$, where $\tilde{L}_i$ is the linear functional on $\mathscr{H}$ defined by $\tilde{L}_ih = \langle P_N\eta_i, h\rangle$. We are now in a position to solve the approximate problem. Minimize

$$\frac{1}{n}\sum_{i=1}^{n} (\tilde{L}_ih - z_i)^2 + \lambda\int_\epsilon^R (h''(r))^2dr, \qquad (2.12)$$

in $\mathscr{H}$, *exactly*. This is easily done using formula (2.7), since it can be shown that $\tilde{\tau}_{i\nu} \equiv \tilde{L}_i\phi_\nu = L_i\phi_\nu = \tau_{i\nu}$ and $\tilde{L}_i(Q_1(\cdot, r)) = P_N\xi_i = \tilde{\xi}_i$, say. (The procedure we used for computing $P_N\xi_i$ is given in Appendix B.) $\check{k}_{ij} = \int_\epsilon^R \tilde{\xi}_i''(r)\tilde{\xi}_j''(r)dr$ is readily evaluated exactly, since the $\{\tilde{\xi}_i\}$ are piecewise polynomials. The minimizer h_λ of (2.12) is given by (2.7), (2.8), and (2.9), with ξ_i replaced by $\tilde{\xi}_i$ and K replaced by $\tilde{K} = \{\check{k}_{ij}\}$. The cross-validation function $\tilde{V}(\lambda)$ for this problem is given by (2.10) and (2.11) with $A(\lambda)$ replaced by $\tilde{A}(\lambda)$, defined by replacing K by $\tilde{K}$ in (2.10). $Q\tilde{K}Q'$ will be non-negative definite.

Given $\tilde{K}$ and T, we give an efficient procedure for minimizing $\tilde{V}(\lambda)$ and computing c and d.

1. Use LINPACK (Dongarra et al. 1979) to find the QR decomposition of T, to obtain

$$T = (Q_1 \vdots Q_2)\begin{pmatrix} R_1 \\ \cdot\cdot \\ 0 \end{pmatrix},$$

where Q_2 is an $n \times n - 2$ matrix with $Q_2'Q_2 = I_{n-2\times n-2}$ ($Q_2'T = 0$) and R_1 is upper triangular. The Q appearing in (2.11) can be taken as Q_2.

2. Let $B = Q_2'\tilde{K}Q_2$ and use EISPACK (Smith et al. 1976) to find the eigenvalue eigenvector decompositon UD_BU' of B, where b_ν^2 ($\nu = 1, 2, \ldots, n - 2$) are the $n - 2$ diagonal entries of D_B (eigenvalues of B) and the $n - 2$ columns of U are the eigenvectors of B. Then

$$\operatorname{tr}(I - \tilde{A}(\lambda)) = \sum_{\nu=1}^{n-2} \frac{n\lambda}{b_\nu^2 + n\lambda},$$

$$(I - \tilde{A}(\lambda))z = n\lambda Q_2U(D_B + n\lambda)^{-1}U'Q_2'z.$$

3. Letting $w = U'Q_2'z$, then

$$\tilde{V}(\lambda) = \frac{1}{n}\sum_{\nu=1}^{n-2}\left(\frac{n\lambda w_\nu}{b_\nu^2 + n\lambda}\right)^2 \bigg/ \left(\frac{1}{n}\sum_{\nu=1}^{n-2}\frac{n\lambda}{b_\nu^2 + n\lambda}\right)^2,$$

$c = Q_2U(D_B + n\lambda I)^{-1}w$, and d is obtained by solving $R_1d = Q_1'(z - \tilde{K}c)$. $\tilde{V}(\lambda)$ is minimized by a global search in log λ. If $n\lambda$ is much smaller than the smallest b_ν^2 or much larger than the largest b_ν^2, it may be taken as 0 or ∞, respectively, so this limits the region required to be searched. $\int_\epsilon^R h_{\hat{\lambda}}(r)dr$ is easily evaluated.

It is useful to note that if $h_{\hat{\lambda}}$ is to be obtained for repeated samples, with the same bins, the cost is quite modest for runs after the first, since the expensive calculations involve the calculation of Q_1, Q_2, R, U, and D_B, and these need only be computed once because they do not depend on the data.

The preceding procedure appeared in Wendelberger (1981) and has been found to work well in similar problems for n as large as 350. In the calculations that follow we used $N = 80$, with the s_i equally spaced. Further discussion of the choice of N appears in Section 6.

3. NUMERICAL RESULTS WITH THE LABORATORY DATA

The liver being sliced fits roughly into a box about 7,500 × 7,500 microns (μ) square by 2,380μ deep (100μ = .01 cm), and for the experimental data studied, it is sliced perpendicular to the short dimension in 21 equally spaced slices (of negligible thickness) 50 microns apart, through the central 1,000 microns of the block, to be called the slicing region. Figure 1 gives a schematic diagram of the slicing design.

Thus, in practice, slices of the paraffin block containing the liver are all parallel to one another, the spacing is equal, and only the "phase" of the tumors with respect to the slicing grid is random. In this experiment, it was

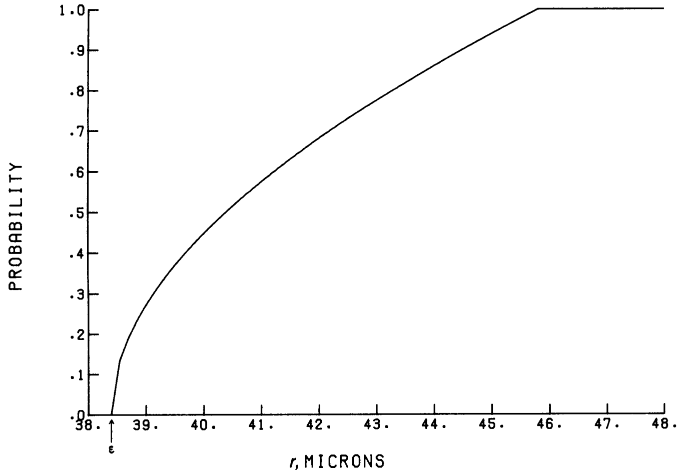

Figure 2. Probability That a Sphere of Radius r Will Have at Least One Observed Cross Section.

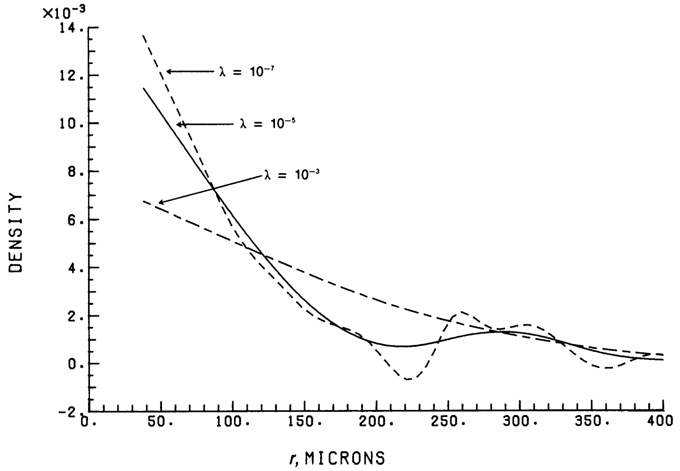

Figure 4. The Estimate f_λ (r), for Three Different Values of λ.

determined in the laboratory that $\epsilon = 38.46$ microns was the smallest cross-sectional radius reliably detected by all of the personnel identifying cross sections. This determination was made after comparing replicated slide readings by the same and different technicians. Smaller cross sections, when observed, were ignored. If ϵ is chosen too small, an erroneous estimate may result, whereas if ϵ is chosen too large, an unnecessary loss of information results. The behavior of the estimate near the left endpoint may be sensitive to the choice of ϵ. Spherical tumors of three-dimensional radius greater than 45.8 microns and lying wholly in the slicing region will be observed in at least one slice, and tumors with radiuses between ϵ and 45.8 may or may not be observed. Figure 2 gives the probability that a sphere of radius r that lies wholly in the slicing region will have at least one observed cross section. Equation (1.1) still holds, but a little reflection will show that if the spacing is uniform and spheres can be sliced more than once, the sampling variance will become smaller as the spacing becomes finer.

R was taken as 690μ. This number was chosen as a conservative upper bound for the largest possible tumor radius after examining the reconstructed tumor size distributions from a group of seven mice (which include the mouse used as an example in Section 3). The spline estimate has been found not to be sensitive to R as R increases, provided R is somewhat larger than the maximum tumor radius. It is our experience that once outside the range of the observed data, the estimate of f_3 will be, effectively, 0.

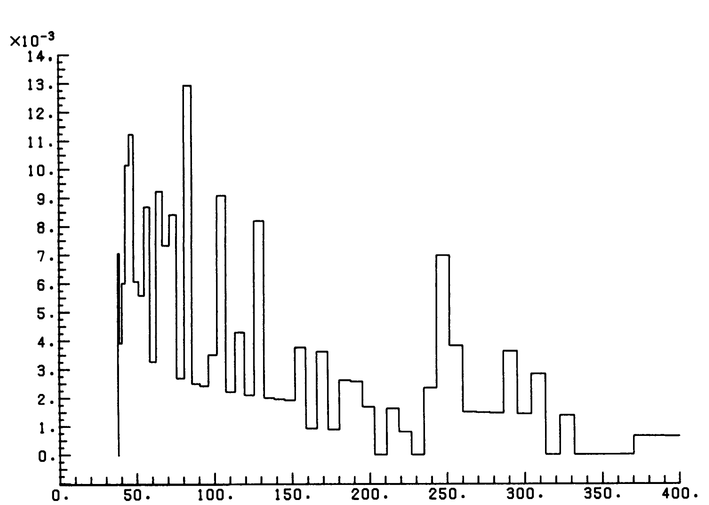

Figure 3. Histogram of Cross-Sectional Radiuses: 154 Observed Cross Sections.

With this slicing, 154 tumor cross sections were observed. Figure 3 gives a histogram of the observed cross-sectional radiuses using the bins $[P_i, P_{i+1}]$. Figure 4 gives a plot of the estimate $f_\lambda(r)$, $\epsilon \le x \le R$, for three different values of λ. This figure demonstrates the sensitivity of f_λ as λ varies. For large λ, the resulting spline is very smooth, but it may have ignored some features of the data. When λ is small, the estimate fits the cross-sectional distribution well, but it yields an oscillating estimate for the tumor size density. In these particular data, one wonders whether the mode at 280 microns is an actual component of the distributions or, rather, just an artifact from undersmoothing.

Figure 5 gives a plot of log $\tilde{V}(\lambda)$ versus log λ. $\tilde{V}(\lambda)$ is minimized for λ around 10^{-5}. This suggests that the

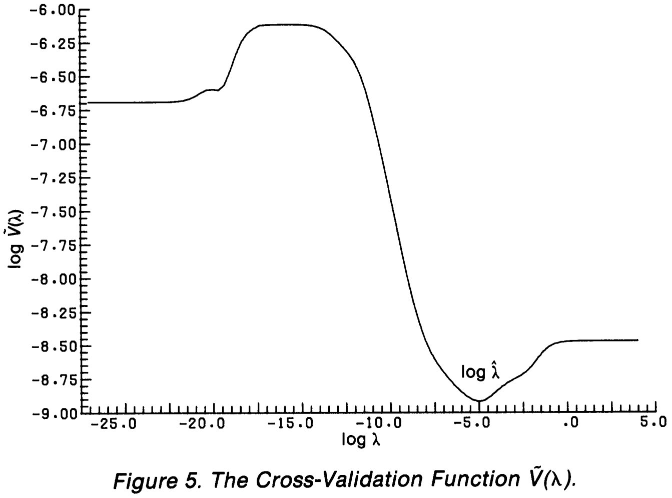

Figure 5. The Cross-Validation Function $\tilde{V}(\lambda)$.

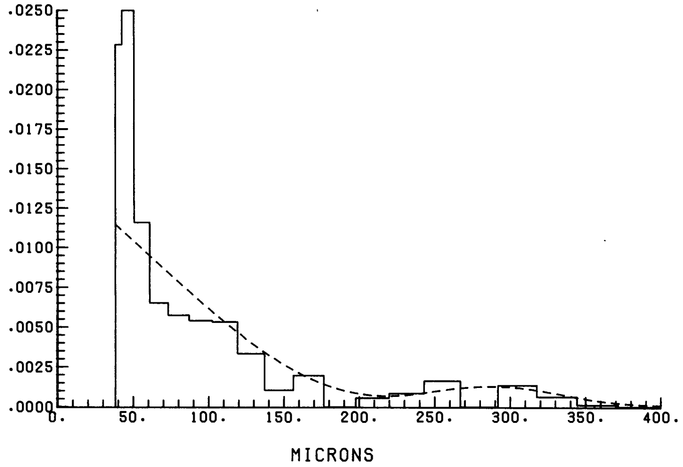

Figure 6. Histogram of True Tumor Size Distribution and $f_{\hat{\lambda}}$.

solid curve in Figure 4 is a good estimate for the size density. Note that this estimate retains a mode at around 280 microns. To compare $f_{\hat{\lambda}}$ with the true f_3^{ϵ}, the slicing region was completely dissected by fine slicing. Figure 6 gives a plot of $f_{\hat{\lambda}}$ (the same as the middle curve in Figure 2) and a histogram of the true tumor-size distribution. There were 53 tumors at least partially in the slicing region. Tumors that were only partially in the slicing region were counted as a fraction of an observation, that fraction being the ratio of the volume inside the slicing region to the whole tumor volume as estimated by the curvature of the portion in the slicing region.

Overall the agreement between a histogram of the reconstructed data and the cross-validated smoothing spline is good. These results are particularly striking because there are only 53 reconstructed tumors in the tissue sample, although of course, the systematic sampling helps. The concentration of tumors around 280 microns predicted by the spline is an actual feature of the reconstructed data. Close to the lower limit, ϵ, however, the spline underestimates the reconstructed distribution.

4. MONTE CARLO EXPERIMENTS

We studied the sampling properties of the estimate by Monte Carlo methods designed to mimic the effects of multiple sampling of large tumors, as well as edge effects, as they actually occur in the mouse experiment.

The geometry of the Monte Carlo experiment is exactly that described in Figure 1; however, R in that figure may take on other values. A pseudo-random number for the total number of spheres was generated according to a Poisson distribution with mean equal to the volume of the entire block × 900 tumors/cc. (The actual mouse had a tumor number density of about 900/cc.) If the number of spheres is n_3, then n_3 "centers" are uniformly distributed throughout the entire block. For each center, a random radius was generated according to the density f_3. Twenty-one parallel, infinitely thin slices 50 microns apart were then made through the shaded region, and the radiuses of all (two-dimensional) intersections greater than ϵ were recorded.

There are now at least two ways of defining the "true" distribution of the three-dimensional radiuses in this experiment. One is as the "theoretical" distribution determined by the density f_3, from which the pseudo-random radiuses were drawn. The second is as the "actual" distribution of the three-dimensional radiuses that were actually drawn. For comparison purposes we will display both the theoretical density and a histogram of the actual distribution as just defined. (The actual distribution is defined here a little differently than the true distribution of Section 3, since tumors in the block but outside the slicing region can be counted.) Experimenters will likely want to focus on the actual distribution if they are interested in a single mouse and on the theoretical distribution if they consider a single mouse as a member of some "superpopulation."

We present the results of four Monte Carlo studies. In each of the studies six replicates were performed. A replicate consists of drawing a sample of tumors, slicing the block, recording the observed cross-sectional radiuses, and computing the estimate $f_{\hat{\lambda}}$.

Experiment 1 was roughly designed to mimic the number density and theoretical size density f_3 of six experimental mice, one of which has been described in Section 3. The theoretical f_3 was taken as a Weibull density that approximates the data of Figure 6. R in this experiment was 690 microns. (R in the definition of $J(\cdot)$ and in Figure 1 have been taken to be the same.) The number of tumors replicated in the entire block averaged 115 with about 49% of them having recorded intersections. The number of observed cross sections averaged 204. Figure 7 shows the results of the six replications. The solid curve in the upper-left plot is the theoretical Weibull curve, the histograms represent the actual size distributions, and the broken lines are the estimates $f_{\hat{\lambda}}$. Although the overall shape of the estimate is good, a tendency to underestimate the density near the cutoff is evident in four of the six replicates. Experiment 2 (see Figure 8) studies a density with different behavior near ϵ. f_3 is a truncated β density. The average number of tumors in the block was 113, of which about 60% had recorded intersections; the average number of observed cross-sectional radiuses was 426. In this experiment most replicates overestimated f_3 near ϵ, and overall, the shape of f_3 is quite good, particularly for larger r. In Experiment 3 we wished to examine the ability of the estimate to resolve distinct peaks. It is of some interest to know to what extent this is possible. The theoretical f_3 was a mixture of two truncated normal densities. R = 450. Figure 9 shows the results of six replications for Experiment 3. The average number of tumors in the block was 80 with about 70% having recorded intersections, and the average number of observed cross-sectional radiuses was 341. Although in all of the six replicates excellent recovery of the two peaks of f_3 was obtained, the estimate of the replicate in the lower-left corner demonstrated "flaky" behavior near ϵ. Inspection of the observational data reveals that the five well-behaved estimates had no observations in the smallest observation bin whereas the flaky estimate had two. In Experiment 4 we wished to see the effect of increasing sample size in Experiment 3. The same f_3, but with a tumor number density five times as big as that of Experiment 3, was used. The average number of tumors in the block was 460 with about 70% having recorded intersections, and the average number of observed cross-sectional radiuses was 1,781. Figure 10 gives the results of this experiment. Extremely good recovery of f_3^ϵ is seen.

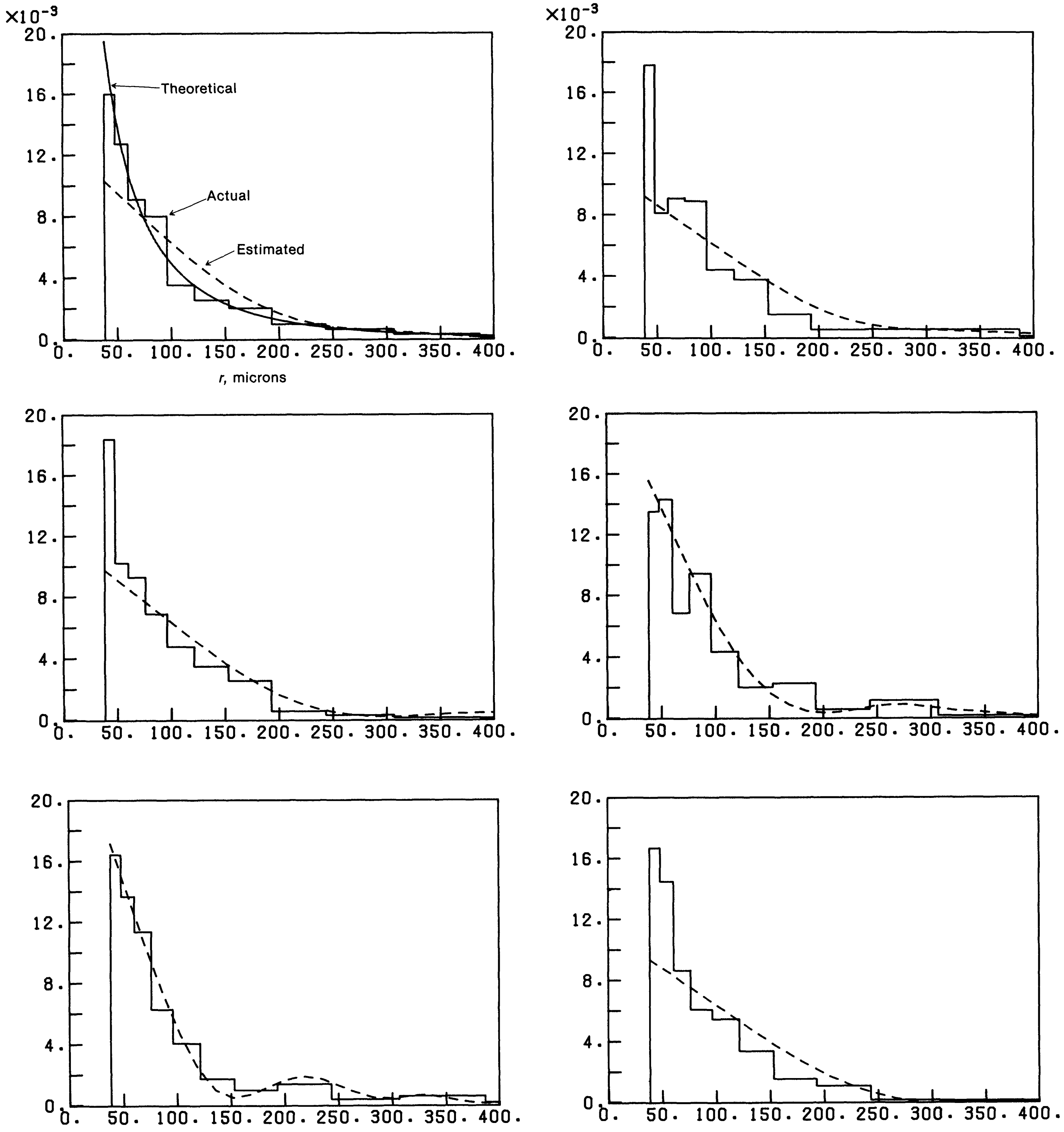

Figure 7. Experiment 1: Weibull Theoretical Density; Histogram for Actual Distribution and f_λ; Six Replicates; (——) $f_3(r)$; (– – –) $f_{\hat{\lambda}}(r)$.

We were also interested in the properties of this smooth estimate in the limiting case when the slicing region is sliced an infinite number of times. For an example, we used the size distribution of the 53 tumors from the mouse of Figure 1 and assumed that all of the tumors were actually contained within the sample. It is not difficult to show in this case that the appropriate data are the expected values of the profile histogram for this discrete distribution (Nychka 1983). A plot of $f_{\hat{\lambda}}$ appears in Figure 11, superimposed on a histogram of the theoretical distribution.

5. MODIFICATIONS FOR END EFFECTS

We conclude that this approach is quite successful on the random-spheres problem, although the behavior of the estimate is not as good as might be hoped near ϵ for small sample sizes. We believe that this is a problem primarily of the data (as opposed to the estimate), since due to the length-biased sampling, and the fact that large tumors have small cross sections as well as large ones, there is very little information in the data concerning the behavior of f_3 near ϵ. It should be recognized, however,

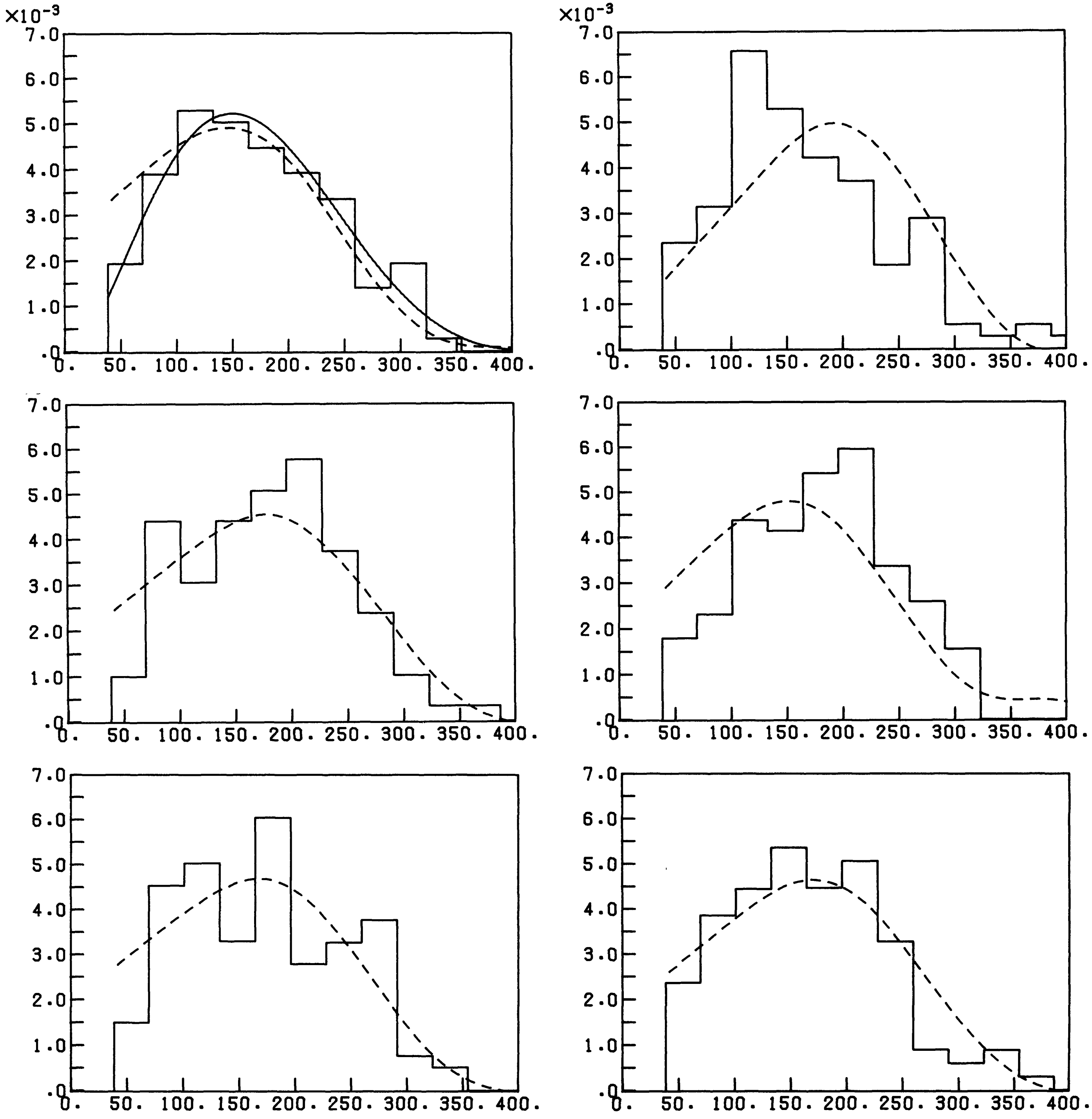

Figure 8. Experiment 2: Truncated Beta Theoretical Density; Observed Distribution and $f_{\hat{\lambda}}$; Six Replicates; (——) $f_3(r)$; (- - -) $f_{\hat{\lambda}}$ (r).

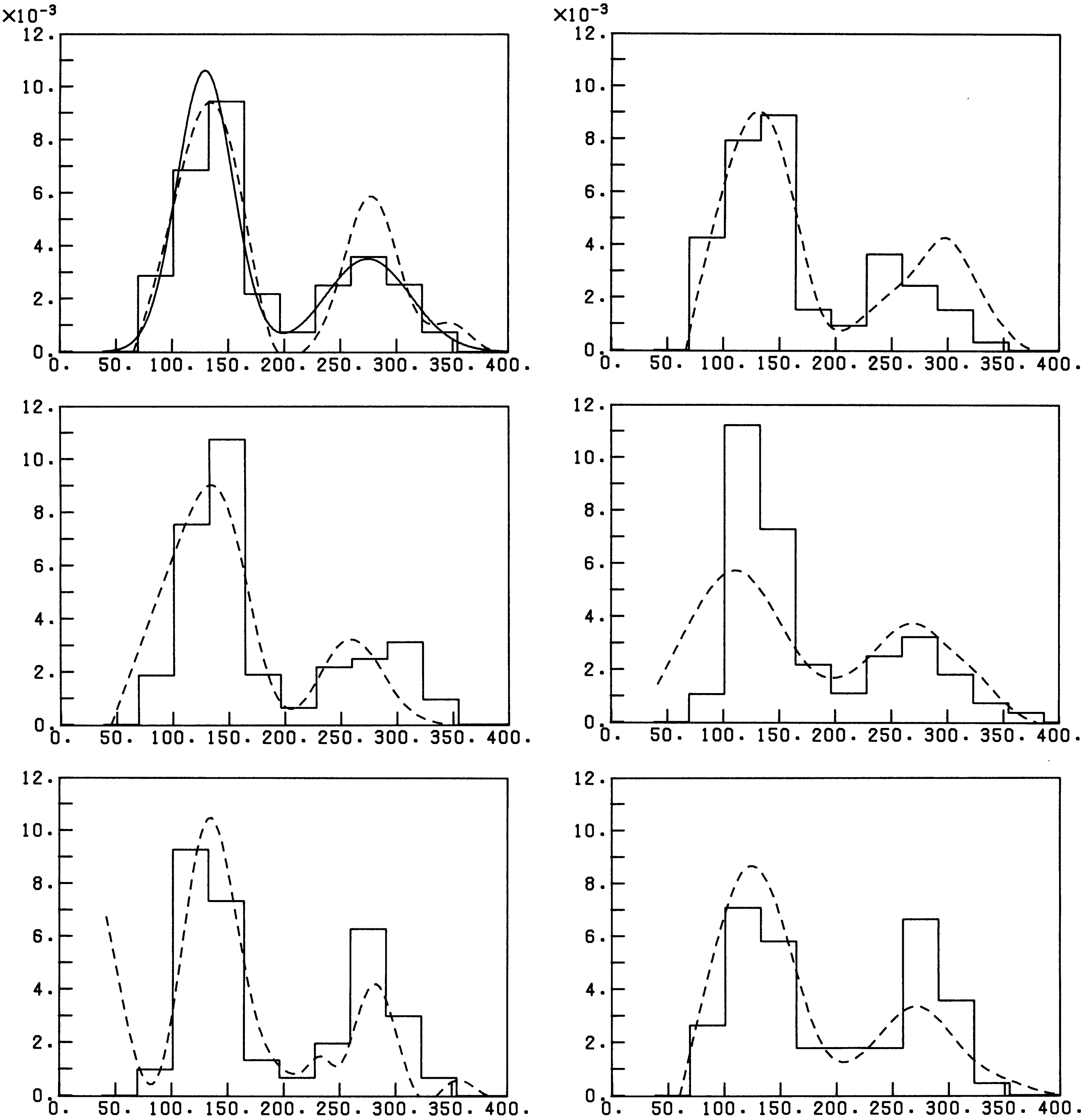

Figure 9. Experiment 3: Truncated Normal Mixture Theoretical Density; Histogram for Actual Distribution and $f_{\hat{\lambda}}$; Six Replicates; (——) $f_3(r)$; (– – –) $f_{\hat{\lambda}}$ (r).

that the estimate extrapolates smoothly from data-rich values of r to data-poor values of r, where smoothness is essentially determined by the choice of penalty functional $J(\cdot)$, which has been taken in this article as $J(f) = \int_{\epsilon}^{R} (f''(r))^2dr$. The null space of $J(\cdot)$ is the linear functions; thus, where there is insufficient information in the data, extrapolation will be linear. In this problem, the penalty functional could have been replaced by, say, $J(f) = \int_{\epsilon}^{R} (f'''(r))^2dr$, in which case the extrapolation would have been quadratic. If prior information concerning the behavior of f_3 near ϵ were available from some external source, then this information could be included in the cross-validated spline estimate by appropriately modifying $J(\cdot)$. For example, suppose it was known that f_3 behaves like a particular negative exponential density g, say. This information may be incorporated in the estimate by replacing $J(f) = \int^{R} (f''(r))^2dr$ by, for example, $J(f) = \| P_g f \|^2 Q$, where $P_g f$ is the projection of f onto the orthocomplement of span $\{g, \phi_1, \phi_2\}$. Then extrapolation from data-rich to data-poor regions (i.e., near ϵ) will proceed via Bayesian information that the true f has negative exponential behavior there. (We are compelled to report, however, that for the mouse liver problem, behavior of f_3^ϵ near ϵ was something of a surprise.) The abstract idea behind this approach may be found in Wahba (1978, section 3). For details of the application to this problem, see Nychka (1983); also see the remarks in Silverman (1983), where a normal density is in the null space of his penalty functional. For a different approach to modifying boundary behavior, see Gasser and Muller (1979). Bias near the boundaries in certain spline estimates that extrapolate linearly has recently been discussed by Rice and Rosenblatt (1983). However, despite the apparent bias near the left boundary in the real data and in Experiment 1, GCV appears to be doing its job well (see Figures 4 and 7).

6. QUANTITATIVE "ILL POSEDNESS," EIGENSEQUENCE PLOTS, AND THE CHOICE OF n

Since the cost of the numerical calculations increase rapidly with n (for the first estimate computed), it is tempting to choose n fairly small. If n is much less than

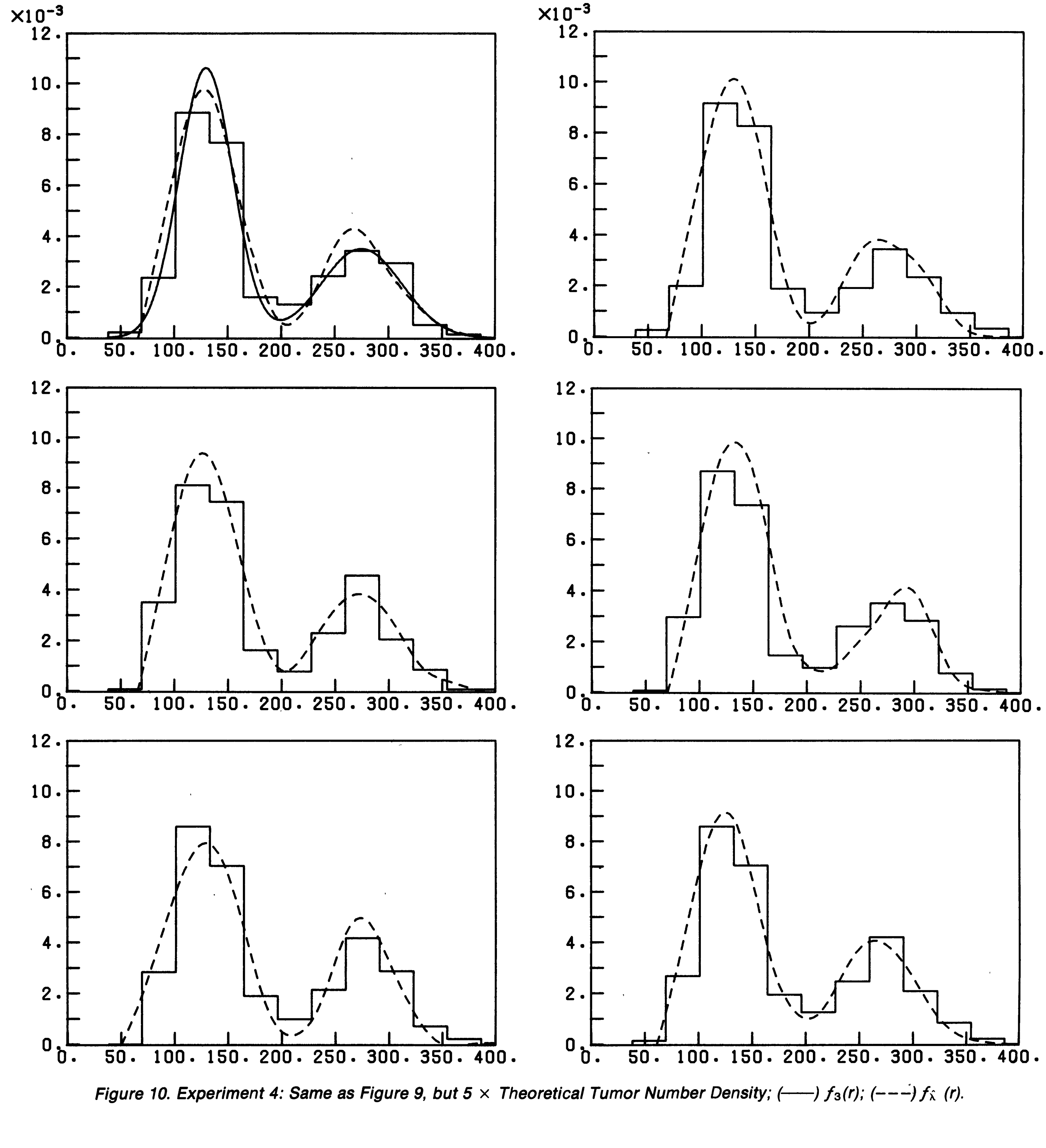

Figure 10. Experiment 4: Same as Figure 9, but 5 × Theoretical Tumor Number Density; (——) $f_3(r)$; (– – –) $f_{\hat{\lambda}}(r)$.

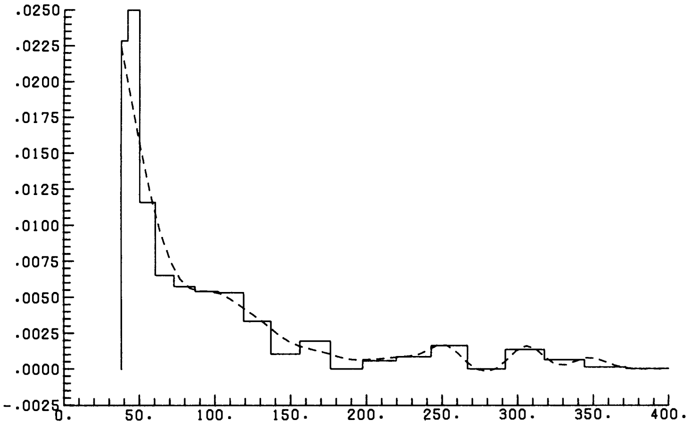

Figure 11. Smooth Estimate With an Infinite Sample Size From a Discrete Theoretical Distribution.

the number of observations, it may act as a smoothing parameter. Using n as a smoothing parameter can be justified theoretically, from an asymptotic point of view (e.g., see Wahba 1975). It is our numerical experience, however, that when there is a relatively small amount of information about the solution available in the data, then smoothing by binning can result in loss of fine structure in the estimate that would be observable if λ were allowed to do most of the smoothing. Thus we set out in this problem to choose n large enough so that little or no smoothing is done at the binning step.

Since this problem is ill posed, however, increasing n beyond some point will not retain much more information, even if the sample size were infinite.

Inspection of the computed eigenvalues b_ν^2 ($\nu = 1, 2, \ldots, n - 2$) can be a valuable procedure in studying this question, and we describe how next. First, given the bins, let $\mathcal{K}_n$ be the operator with domain $\mathcal{H}$ and range E_n, which maps f to $\mathcal{K}_n f = (L_1 f, \ldots, L_n f)$. Then $\mathcal{K}_n$ is analogous to the design matrix X in the usual regression problem $y = X\beta + \epsilon$, and the role of XX' is played by the $n \times n$ gram matrix Σ with ijth entry $\langle \eta_i, \eta_j \rangle$. Inspection of the eigenvalues of Σ thus provides important information on the effective dimension of the range of $\mathcal{K}_n$ when the domain of $\mathcal{K}_n$ is $\mathcal{H}$. In some ill-posed problems, Σ is theoretically of full rank but has fewer than n eigenvalues that are actually larger than machine double-precision 0. For an extreme example, see Wahba (1979). Now, since $\eta_i = \xi_i + a_{i1}\phi_1 + a_{i2}\phi_2$ for some a_{i1}, a_{i2}, the matrix Σ can be obtained from the matrix K with ijth entry $\langle \xi_i, \xi_j \rangle$ by the addition of some rank 2 matrix, which is not important to our problem. (The a_{ij} depend on the definition of $\langle \phi_\mu, \phi_\nu \rangle$, μ, $\nu = 1, 2$, which is irrelevant to the estimate being studied.) Furthermore, if $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_n$ are the eigenvalues of K, and $\delta_1 \geq \delta_2 \geq \cdots \geq \delta_{n-2}$ are the eigenvalues of QKQ', then $\gamma_{\nu-2} \leq \delta_\nu \leq \gamma_\nu$. Now as part of

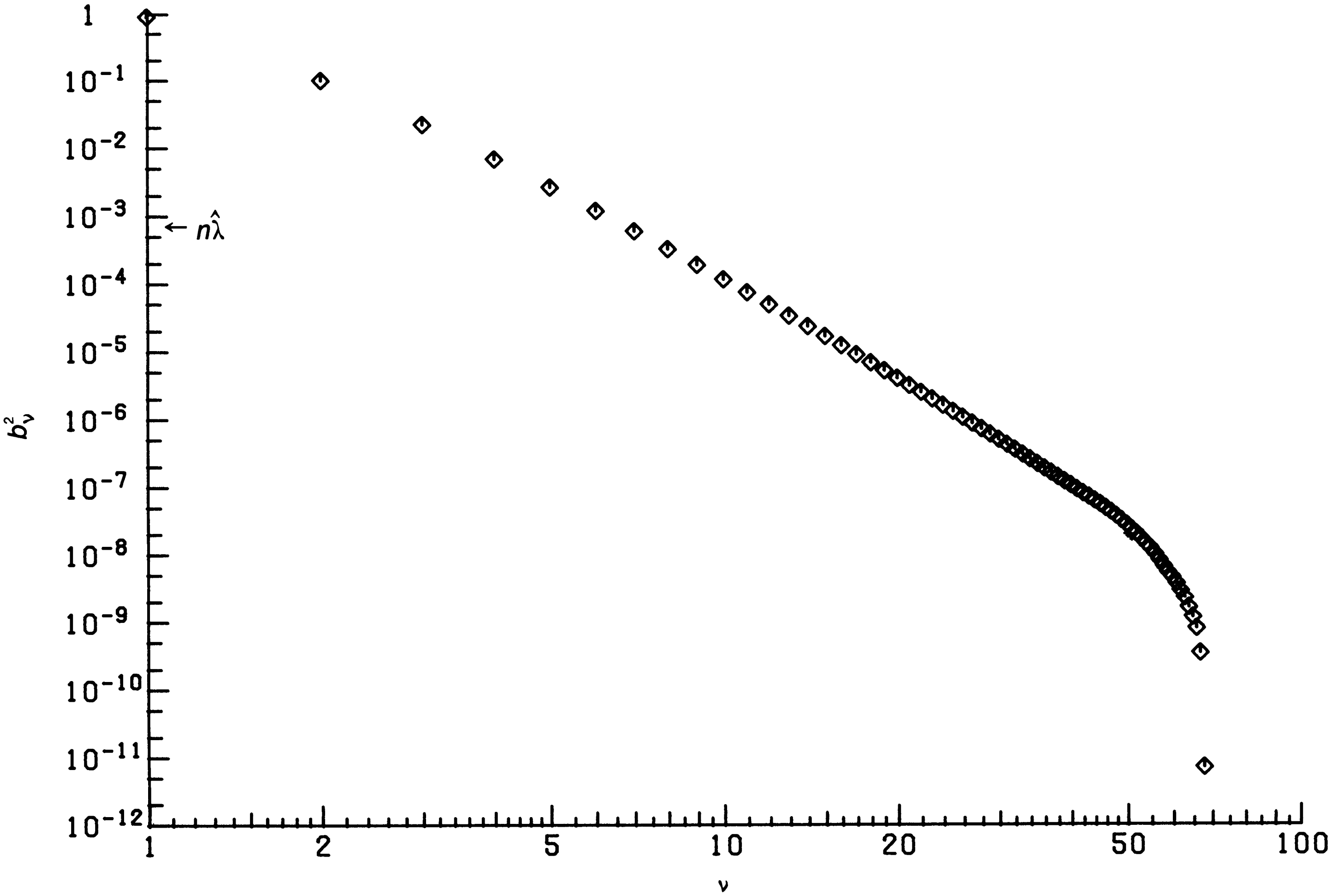

Figure 12. The Eigenvalues of QKQ'; n = 80, N = 80.

the calculations for Sections 3 and 4 we have computed $b_1^2 \geq \cdots \geq b_{n-2}^2$, which are the eigenvalues of $Q\tilde{K}Q'$, $\tilde{K}$ being the $n \times n$ matrix with ijth entry $\langle P_N\xi_i, P_N\xi_j\rangle$. The b_ν^2's satisfy $b_\nu^2 \leq \gamma_\nu$ ($i = 1, 2, \ldots, n - 2$), and the number of nonzero b_ν^2's cannot be bigger than the dimension of the range of P_N. In the limit as $N \to \infty$, $b_\nu^2 \to \gamma_\nu$. If N is too small, it, too, can act as a smoothing parameter.

Figure 12 gives a plot of the first 68 b_ν^2's on a log-log plot, with $n = 80$, $N = 80$. (The vertical unit is arbitrary and depends on the units in which r is carried in the computer. It is reasonable to choose these units so that $b_1^2 \approx 1$.) For comparison, an arrow marks $n\hat{\lambda} = 80 \times 10^{-5}$. Recall that the eigenvalues of $\tilde{A}(\lambda)$ are $(1, 1, b_1^2/(b_1^2 + n\lambda), \ldots, b_{n-2}^2/(b_{n-2} + n\lambda))$. $\tilde{A}(\lambda)$ plays the role of the influence matrix $X(X'X + n\lambda I)X'$ in the regression problem when a ridge estimate is used for β.

Based on trying several values of n and N, it is our belief that at least the first 30 or 40 b_ν^2's approximate the δ_ν's very well and that increasing N would have no appreciable effect on the resulting estimate $f_{\hat{\lambda}}$. If n is increased, our unpublished plots as well as recent analytical work suggest that the slope of (the major part of) the eigensequence log-log plot will tend to a limit (e.g., see Utreras 1981 and Wahba 1977). Note that b_{40}^2/b_1^2 is already down to 10^{-7}. We conclude that increasing n (with $N \geq n$) much past 80 would not change $f_{\hat{\lambda}}$, certainly not to plot accuracy, and that we have thus succeeded in choosing n and N so that they are not acting as smoothing parameters.

Eigensequence plots can provide insight about practical limits on the amount of information concerning f_3^ϵ in the data, and we suggest that these plots be routinely examined in problems of this sort. It is seen that with $\lambda = 10^{-5}$, the eigenvalues $b_\nu^2/(b_\nu^2 + n\lambda)$ of the influence matrix $A(\lambda)$ have decreased to .5 by about the eighth eigenvalue ($\nu = 6$).

7. RELATED ESTIMATES AND THEIR SMOOTHING PARAMETERS

Another approach to the approximate numerical calculation of the minimizer of (2.2) in $\mathcal{H}$ is to minimize (2.2), in a convenient approximating subspace $\mathcal{S}_N = \text{span}\{B_l\}$, say.

Then one finds h_λ of the form $h_\lambda = \sum_l \theta_l B_l$ to minimize (2.2). In the problems studied here, a space of cubic B splines (see deBoor 1978) would be appropriate. If the basis functions have compact support, this would be considered to be a finite-element method.

Given $\epsilon = s_0 < s_1 < \cdots < s_N = R$, let S_Nh be that function in $\mathcal{H}$ that minimizes $J(h)$ subject to $(S_Nh)(s_l) = h(s_l)$ ($l = 0, 1, \ldots, N$) and $(S_Nh)'(s_0) = h'(s_0)$, $(S_Ng)'(s_N) = h'(s_N)$. S_Nh is the cubic spline interpolating to h at $s_0, s_1, \ldots, s_N$, and to h' at s_0 and s_N. Let $\mathcal{S}_N$ be a set of $N + 3$ cubic B splines whose span is the range of S_N (e.g., see deBoor 1978, chap. 9). Then the minimizer of the exact expression

$$\frac{1}{n}\sum_{i=1}^{n} (\langle \eta_i, h\rangle - z_i)^2 + \lambda J(h), \tag{7.1}$$

in the approximating subspace $\mathcal{S}_N$, is the same as the minimizer of the approximate expression

$$\frac{1}{n}\sum_{i=1}^{n} (\langle \eta_i, S_Nh\rangle - z_i)^2 + \lambda J(h) \tag{7.2}$$

in $\mathcal{H}$. This can be shown without difficulty by writing $h = S_Nh + (I - S_N)h = \sum_l \theta_l B_l + (I - S_N)h$ for some $\{\theta_l\}$ and using the property of the cubic spline interpolant $\int_\epsilon^R (S_Nh)''((I - S_N)h)'' = 0$ to obtain $J(h) = J(S_Nh) + J((I - S_N)h)$. It can then be shown that the minimizer of (7.2) must be in $\mathcal{S}_N$. Upon observing that $S_Nh = h$ for any h in $\mathcal{S}_N$ (spline interpolation is idempotent), it follows that problems (7.1) and (7.2) are the same. For comparison with (2.12) we can write (7.2) as

$$\frac{1}{n}\sum_{i=1}^{n} (\tilde{\tilde{L}}_i h - z_i)^2 + \lambda J(h), \tag{7.3}$$

where $\tilde{\tilde{L}}_i h = \langle S_N^*\eta_i, h\rangle$, S_N^* being the adjoint operator to S_N. The cross-validation function $\tilde{\tilde{V}}(\lambda)$ for the problem (7.3) can also be readily obtained.

The minimizer of the exact problem (7.1) in some space $\mathcal{S}_N$ of B splines is the "hybrid estimate" proposed by Wahba (1980) and mentioned by Mendelsohn and Rice (1983). If the dimension of $\mathcal{S}_N$ is chosen large, then this hybrid estimate will, numerically, be a good approximation to the original cross-validated spline estimate (minimizer of (2.2) in $\mathcal{H}$). On the other hand, if $\mathcal{S}_N$ is relatively small, then N will act as a smoothing parameter. Thus there will be a pair of smoothing parameters (λ, N), which, in principle, could be chosen objectively by GCV. Mendelsohn and Rice (1983) solved the problem mentioned in (1.7) by using the minimizer of

$$\frac{1}{n}\sum_{i=1}^{n} (L_ih - z_i)^2$$

in $\mathcal{S}_N$. In their work, n was very large, and N was the only smoothing parameter.

When N is the only smoothing parameter, the optimal integrated mean-squared error (IMSE) value of N grows very slowly with n. (For certain regression problems $N = 0(n^{1/5})$; see Agarwal and Studden 1980.) When n is very large and the data is nearly exact, then one can sometimes profitably use N as the sole smoothing parameter, since the optimal N will be large enough so that recoverable structure in the solution will not be lost. (An N-only estimate is the easiest to compute.) In Mendelsohn and Rice's (1983) problem, n was several hundred and the data could be considered extremely "exact," since 10^5 observations were in the n bins. They found an N of 12 subjectively. In our problem with much "noisier" data, we conjecture that the optimal N in an N-only estimate would result in N of more like 3–6, and in general the estimate would not show the peak resolution that is evident in Figures 9 and 10 unless the true solution was actually in $\mathcal{S}_N$. Efficient numerical methods for the hybrid estimate for problems with very large N (as might occur in image processing, for example) can be found in Bates and Wahba (1982).

We see now that there are actually three possible smoothing parameters—λ, n, and N. In the matched-quadrature method, it is natural to have $N > n$, and the computing load is sensitive to n and insensitive to N. In the hybrid method it is natural to take $N < n$, and the computing load will be sensitive to N and insensitive to n. In the matched-quadrature method, one could easily use n and λ as joint smoothing parameters, and in the hybrid method one could easily use N and λ as joint smoothing parameters. (There may, however, be a region in (λ, n) or (λ, N) space in which decreasing both parameters simultaneously will have little effect on the IMSE.) In the problem at hand, where the data is very noisy (because the sample size is small) and the problem is somewhat ill posed, we believe that one can do a better job of recovering structure in the solution if one lets λ do all or most of the smoothing and one chooses n and N just large enough so that they are not doing appreciable smoothing. When there is a very large amount of information in the data, using n or N to do (some of) the smoothing, can be very cost effective.

APPENDIX A: FORMULA FOR $\xi_i(r)$ AND $\tau_{i\nu}$

$\xi_i(r) = L_i(Q_1(\cdot, r)) = \psi(P_{i-1}, r) - \psi(P_i, r)$, where

$$\psi(P, t) = \frac{(t-\epsilon)^2}{2}\mathcal{I}_1(\epsilon, P, P, R) - \frac{(t-\epsilon)^3}{2} I_0(P, P, R), \qquad \epsilon \le t \le P$$

$$= \frac{(t-\epsilon)}{2}\mathcal{I}_2(\epsilon, P, P, t) - I_3(\epsilon, P, P, t) + \frac{(t-\epsilon)}{2}\mathcal{I}_1(\epsilon, P, t, R) - \frac{(t-\epsilon)^3}{6} I_0(P, P, R), \qquad P \le t \le R,$$

where

$$I_k(x, a, b) = \int_a^b u^k \sqrt{u^2 - x^2 du}, \qquad x \le a, k = 0, 1, 2, 3,$$

$$\mathcal{I}_k(\epsilon, x, a, b) = \sum_{i=0}^{k} \binom{k}{i} \epsilon^{k-i} I_i(x, a, b), \qquad k = 0, 1, 2, 3.$$

The definite integrals I_k have closed-form analytic representation (see Selby's 1979 formulas 156, 167, 168, and 170, p. 425): $\tau_{i1} = I_0(P_{i-1}, P_{i-1}, R) - I_0(P_i, P_i, R)$ and $\tau_{i2} = \mathcal{I}_1(P_{i-1}, P_{i-1}, R) - \mathcal{I}_1(P_i, P_i, R)$.

APPENDIX B: COMPUTATION OF $P_N\xi_i$ AND $\tilde{k}_{ij}$

Since $\xi_i(0) = \xi_i'(0) = 0$, and Q_1 is the reproducing kernel for the subspace of $\mathcal{H}$ satisfying these boundary conditions, we must have

$$P_N\xi_i = \sum_{k=1}^{N} \alpha_{ik} Q(\cdot, s_k)$$

for some $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iN})$. The α_{ij} are chosen so that the interpolation conditions are satisfied; that is,

$$\tilde{Q}\alpha_i = \begin{pmatrix} \xi_i(s_1) \\ \vdots \\ \xi_i(s_N) \end{pmatrix},$$

where $\tilde{Q}$ is the $N \times N$ matrix with ijth entry $Q(s_i, s_j)$. Since $\tilde{Q}$ is positive definite, the α_i can be efficiently computed via a Cholesky factorization of $\tilde{Q}$ (see Dongarra et al. 1979, chap. 3). Now

$$\check{k}_{ij} = \langle P_N\xi_i, P_N\xi_j \rangle = \sum_k \alpha_{ik} \sum_l \alpha_{jl} \langle Q(\cdot, s_k), Q(\cdot, s_l) \rangle$$

$$= \alpha_i' \tilde{Q} \alpha_j.$$

[*Received June 1983. Revised May 1984.*]

REFERENCES

AGARWAL, G., and STUDDEN, W. (1980), "Asymptotic Integrated Mean Square Error Using Least Squares and Bias Minimizing Splines," *Annals of Statistics*, 8, 1307–1325.

ANDERSSEN, R.S., and JAKEMAN, A.S. (1975), "Abel Type Integral Equations in Stereology, II: Computational Methods of Solution and the Random Spheres Approximation," *Journal of Microscopy*, 105, 135–153.

BATES, D., and WAHBA, G. (1982), "Computational Methods for Generalized Cross-Validation With Large Data Sets," in *Treatment of Integral Equations by Numerical Methods*, eds. C. T. H. Baker and G. F. Miller, London: Academic Press, 283–296.

COX, D.R. (1970), *Analysis of Binary Data*, London: Chapman and Hall.

COX, DENNIS (1983), "Approximation of Method of Regularization Estimators," Technical Report 723, University of Wisconsin–Madison, Statistics Dept.

CHOVER, J., and KING, J. (1981), personal communication.

CRAVEN, P., and WAHBA, G. (1979), "Smoothing Noisy Data With Spline Functions: Estimating the Correct Degree of Smoothing by the Methods of Generalized Cross Validation," *Numerische Mathematik*, 31, 377–403.

CRUMP, J.G., and SEINFELD, J.H. (1982), "A New Algorithm for Inversion of Aerosol Size Distribution Data," *Aerosol Science and Technology*, 1, 15–34.

deBOOR, C. (1978), *A Practical Guide to Splines*, New York: Springer-Verlag.

DIACONIS, P., and EFRON, B. (1983), "Computer-Intensive Methods in Statistics," *Scientific American*, 248, 5, 116–130.

DONGARRA, J.J., BUNCH, J.R., MOLER, C.B., and STEWART, G.W. (1979), *LINPACK Users Guide*, Philadelphia: Society for Industrial and Applied Mathematics.

GASSER, T., and MULLER, M. (1979), "Kernel Estimation of Regression Functions," in *Smoothing Techniques for Curve Estimation* (Lecture Notes in Mathematics No. 757), eds. T. Gasser and M. Rosenblatt, New York: Springer-Verlag.

KEIDING, N., JENSEN, S.T., and RANEK, L. (1972), "Maximum Likelihood Estimation of the Size Distribution of Liver Cell Nuclei From the Observed Distribution in a Plane Section," *Biometrics*, 28, 813–829.

KIMELDORF, G., and WAHBA, G. (1971), "Some Results of Tchebycheffian Spline Functions," *Journal of Mathematical Analysis and Applications*, 33, 82–95.

KOEN, H., PUGH, T., and GOLDFARB, S. (1983), "Hepatocarcinogenesis in the Mouse Combined Morphologic-Stereologic Studies," *American Journal of Pathology*, 112, 89–100.

KUK, A.Y.C. (1982), "A Mixing Distribution Approach to Estimating

Particle Size Distributions," Technical Report 328, Stanford University, Statistics Dept.

LUKAS, M. (1981), *Regularization of Linear Operator Equations*, unpublished thesis, Australian National University.

MENDELSOHN, J., and RICE, J. (1983), "Deconvolution of Microfluorometric Histograms With B-splines," *Journal of the American Statistical Association*, 77, 748–753.

MERZ, P. (1980), "Determination of Adsorption Energy Distribution by Regularization and a Characterization of Certain Adsorption Isotherms," *Journal of Computational Physics*, 38, 64–85.

NICHOLSON, W.L. (1970), "Estimation of Linear Properties of Particle Size Distributions," *Biometrika*, 57, 273–297.

——— (1976), "Estimation of Linear Functions by Maximum Likelihood," *Journal of Microscopy*, 113, 113–239.

NICHOLSON, W.L., and MERCK, K.R. (1969), "Unfolding Particle Size Distributions," *Technometrics*, 11, 707–720.

NYCHKA, D. (1983), *The Solution of Abel-Type Integral Equations With an Application in Stereology*, unpublished Ph.D. thesis, University of Wisconson–Madison, Statistics Dept.

O'SULLIVAN, F. (1983), *The Analysis of Some Penalized Likelihood Estimation Schemes*, unpublished Ph.D. thesis, University of Wisconsin–Madison, Statistics Dept.

RICE, J., and ROSENBLATT, M. (1983), "Smoothing Splines: Regression, Derivatives and Deconvolution," *Annals of Mathematical Statistics*, 11, 141–156.

SELBY, S. (ed.) (1979), *CRC Standard Mathematical Tables* (21st ed.), Cleveland, Ohio: Central Rubber Company.

SILVERMAN, B. (1983), "On the Estimation of a Probability Density Function by the Maximum Penalized Likelihood Method," *Annals of Statistics*, 10, 795–810.

SMITH, B.T., BOYLE, J.M., DONGARRA, J.J., GARBOW, B.S., IKEBE, Y., KLEMA, V.V., and MOLER, C.B. (1976), "Matrix Eigensystem Routines—EISPACK Guide," in *Lecture Notes in Computer Science*, New York: Springer-Verlag.

TALLIS, G.M. (1970), "Estimating the Distribution of Spherical and Elliptical Bodies in Conglomerates From Plate Sections," *Biometrics*, 26, 87–103.

UTRERAS, F. (1981), "Optimal Smoothing of Noisy Data Using Spline Functions," *SIAM Journal of Scientific and Statistical Computing*, 2, 349–362.

VILLALOBOS, M. (1983), *Multivariate Spline Estimates for the Posterior Probabilities in the Classification Problem*, unpublished Ph.D. thesis, University of Wisconsin–Madison, Statistics Dept.

VILLALOBOS, M., and WAHBA, G. (1983), "Multivariate Thin Plate Spline Estimates for the Posterior Probabilities in the Classification Problem," *Communications in Statistics, Part B—Simulation and Computation*, 12, 100–120.

WAHBA, G. (1975), "Interpolating Spline Methods for Density Estimation, I: Equispaced Knots," *Annals of Statistics*, 3, 30–48.

——— (1977), "Practical Approximate Solutions to Linear Operator Equations When the Data Are Noisy," *SIAM Journal of Numerical Analysis*, 14, 4.

——— (1978), "Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression," *Journal of the Royal Statistical Society*, Ser. B, 40, 3.

——— (1979), "Smoothing and Ill Posed Problems," in *Solution Methods for Integral Equations With Applications*, ed. Michael Golberg, New York: Plenum Press, 183–194.

——— (1980), "Ill Posed Problems: Numerical and Statistical Methods for Mildly, Moderately and Severely Ill Posed Problems With Noisy Data," Technical Report 595, University of Wisconsin–Madison, Statistics Dept.

——— (1982a), "Constrained Regularization for Ill Posed Linear Operator Equations With Applications in Meteorology and Medicine," in *Statistical Design Theory and Related Topics: III* (Vol. 2), eds. S. S. Gupta and J. O. Berger, New York: Academic Press.

——— (1982b), "Cross Validated Spline Methods for Direct and Indirect Sensing Experiments," Technical Report 694, University of Wisconsin, Statistics Dept.

WAHBA, G., and WENDELBERGER, J. (1980), "Some New Mathematical Methods for Variational Objective Analysis Using Splines and Cross-Validation," *Monthly Weather Review*, 108, 1122–1143.

WATSON, G.S. (1971), "Estimating Functionals of Particle Size Distributions," *Biometrika*, 58, 483.

WENDELBERGER, J. (1981), "The Computation of Laplacian Smoothing Splines With Examples," Technical Report 648, University of Wisconsin, Statistics Dept.

WICKSELL, D.S. (1925), "The Corpuscle Problem, Part I," *Biometrika*, 17, 87–97.