CHAPTER 5

"Confidence Intervals"

5.1 Bayesian "confidence intervals."

Continuing with the Bayesian model

$$\begin{aligned} F(t) &= \sum_{\nu=1}^{M} \theta_\nu \phi_\nu(t) + b^{1/2} X(t),\ t \in \mathcal{T}, \\ Y_i &= L_i F + \epsilon_i \end{aligned}$$

as in (1.5.8), we know that

$$\lim_{a \to \infty} E(F(t) | Y_i = y_i, i = 1, \ldots, n) = f_\lambda(t)$$

with $\lambda = \sigma^2/nb$. The covariance of $f_\lambda(s)$ and $f_\lambda(t)$, call it $c_\lambda(s,t)$, can be obtained by standard multivariate techniques. A formula is given in Wahba (1983), which we do not reproduce here. (This formula also involves b.)

By the arguments in Section 1.5,

$$E(L_0 F | y_1, \ldots, y_n) = L_0 f_\lambda,$$

and it is not hard to show that the covariance of $L_0 f_\lambda$ and $L_{00} f_\lambda$ is $L_{0(s)} L_{00(t)} c_\lambda(s,t)$.

An important special case that will be used to construct "confidence intervals" is: The covariance matrix of $(L_1 f_\lambda, \ldots, L_n f_\lambda)$ is

$$\text{cov}\ (L_1 f_\lambda, \ldots, L_n f_\lambda) = \sigma^2 A(\lambda). \tag{5.1.1}$$

One way to derive (5.1.1) is to consider the Bayes model of (1.5.8) before letting $a \to \infty$. Then the joint covariance matrix of $(L_1 F, \ldots, L_n F, Y_1, \ldots, Y_n)$ is

$$\begin{pmatrix} aTT' + b\Sigma & aTT' + b\Sigma & \\ & & \\ aTT' + b\Sigma & aTT' + b\Sigma & +\sigma^2 I \end{pmatrix}.$$

Then we have

$$\begin{pmatrix} L_1 f_\lambda \\ \vdots \\ L_n f_\lambda \end{pmatrix} = \lim_{a \to \infty} A^a(\lambda) y$$

where $A^a(\lambda) = (aTT' + b\Sigma)(aTT' + b\Sigma + \sigma^2 I)^{-1}$, with $\lambda = \sigma^2/nb$. It can be verified using the limit formulae (1.5.11) and (1.5.12) that $\lim_{a\to\infty} A^a(\lambda) = A(\lambda)$. Then

$$\begin{pmatrix} L_1 f_\lambda - L_1 F \\ \vdots \\ L_n f_\lambda - L_n F \end{pmatrix} = -\lim_{a\to\infty} (I - A^a(\lambda)) \begin{pmatrix} L_1 F \\ \vdots \\ L_n F \end{pmatrix} + A^a(\lambda)\epsilon,$$

with covariance

$$\lim_{a\to\infty} [(I - A^a(\lambda))(aTT' + b\Sigma)(I - A^a(\lambda)) + \sigma^2 A^a(\lambda) \cdot A^a(\lambda)]. \tag{5.1.2}$$

The collection of terms in (5.1.2) shows that the quantity in brackets in (5.1.2) is equal to $\sigma^2 A^a(\lambda)$, giving the result.

Considering the case $L_i F = F(t_i)$, we have that (5.1.1) suggests using as a confidence interval

$$f_{\hat\lambda}(t_i) \pm z_{\alpha/2}\sqrt{\hat\sigma^2 a_{ii}(\hat\lambda)},$$

where $\hat\lambda$ and $\hat\sigma^2$ are appropriate estimates of λ and σ^2 and $z_{\alpha/2}$ is the $\alpha/2$ point of the normal distribution.

The estimate

$$\hat\sigma^2 = \frac{\mathrm{RSS}(\hat\lambda)}{\mathrm{Trace}\,(I - A(\hat\lambda))} \tag{5.1.3}$$

where $\mathrm{RSS}(\hat\lambda)$ is the residual sum of squares, was used in the example below. Although to the author's knowledge theoretical properties of this estimate have not been published, good numerical results in simulation studies have been found by several authors (see, e.g., O'Sullivan and Wong (1988), Nychka (1986, 1988, 1990), Hall and Titterington (1987)). The argument is that trace $A(\hat\lambda)$ should be considered the degrees of freedom (d.f.) for signal, by analogy with the regression case, and trace $(I - A(\hat\lambda))$ is the d.f. for noise. On a hunch, it was decided to study these "confidence intervals" numerically with smooth functions and the GCV estimates $\hat\lambda$ of λ.

Figure 5.1 gives a test function from Wahba (1983). Data were generated according to the model

$$y_{ij} = f\left(\frac{2i+1}{2N}, \frac{2j+1}{2N}\right) + \epsilon_{ij}, i, j = 1, \ldots, N$$

with $N = 13$, giving $n = N^2 = 169$ data points. The peak height of f was approximately 1.2 and σ was taken as .03. $f_{\hat\lambda}$ was the thin-plate spline of Section 2.4 with $d = 2, m = 2$. Figure 5.2 gives four selected cross sections for four fixed values of $x_1, x_1 = (2i+1)/N$, for $i = 7, 9, 11, 13$. In each cross section is plotted $f((2i+1)/N, x_2)$, $0 \le x_2 \le 1$ (solid line), $f_{\hat\lambda}((2i+1)/N, x_2)$, $0 \le x_2 \le 1$, where $f_{\hat\lambda}$ is the thin plate smoothing line (dashed line), the data y_{ij}, $j = 1, \ldots, 13$, for i fixed, and confidence bars, which extend between

$$f_{n,\hat\lambda}((2i+1)/N, x_2(j)) \pm 1.96\hat\sigma(\hat\lambda)\sqrt{a_{ij,ij}(\hat\lambda)}.$$

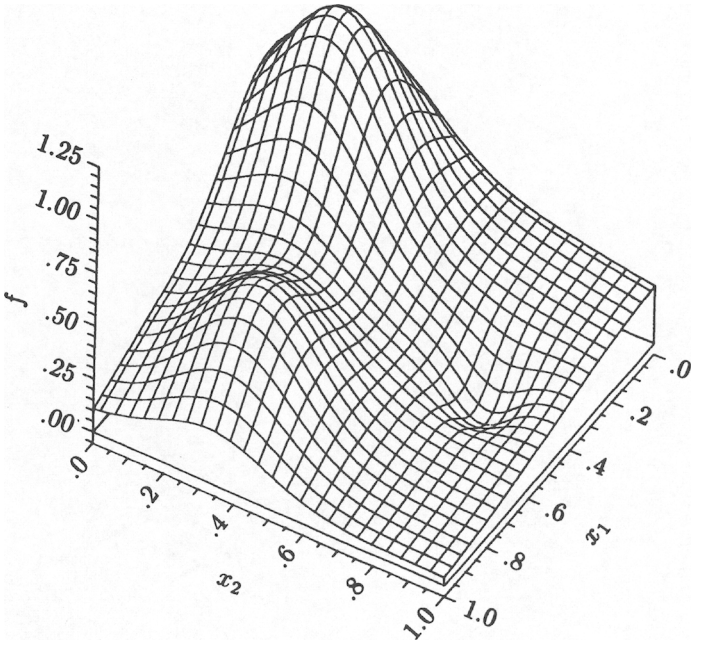

FIG. 5.1. *Test function for confidence intervals.*

Of the 169 confidence intervals, 162 or 95.85 percent covered the true value of $f(x_1(i), x_2(j))$.

We take pains to note that these "confidence intervals" must be interpreted "across the function," as opposed to pointwise. If this experiment were repeated with the same f and new ϵ_{ij}'s then it would be likely that about 95 percent of the confidence intervals would cover the corresponding true values, but it may be that the value at the same (x_1, x_2) is covered each time. This effect is more pronounced if the true curve or surface has small regions of particularly rapid change. In an attempt to understand why these Bayesian confidence intervals have the frequentist properties that they apparently do, it was shown that

$$ET(\lambda^*) = \alpha \frac{\sigma^2}{n} \sum_{i=1}^{n} a_{ii}(\lambda^*)(1 + o(1)), \tag{5.1.4}$$

where λ^* is the minimizer of $ET(\lambda^*)$ and, in the case of the univariate polynomial spline of degree $2m-1$ with equally spaced data, $\alpha \in [(1+1/4m)(1-1/2m), 1]$, that is, $(\sigma^2/n)\,\mathrm{Tr}\, A(\lambda^*)$ is actually quite close to $ET(\lambda^*) = b^2(\lambda^*) + \sigma^2\mu_2(\lambda^*)$. Nychka (1986, 1988, 1990) and Hall and Titterington (1987) later showed that the lower bound obtained, and Nychka gave a nice argument rigorizing the interpretation of intervals as confidence intervals "across the function" by

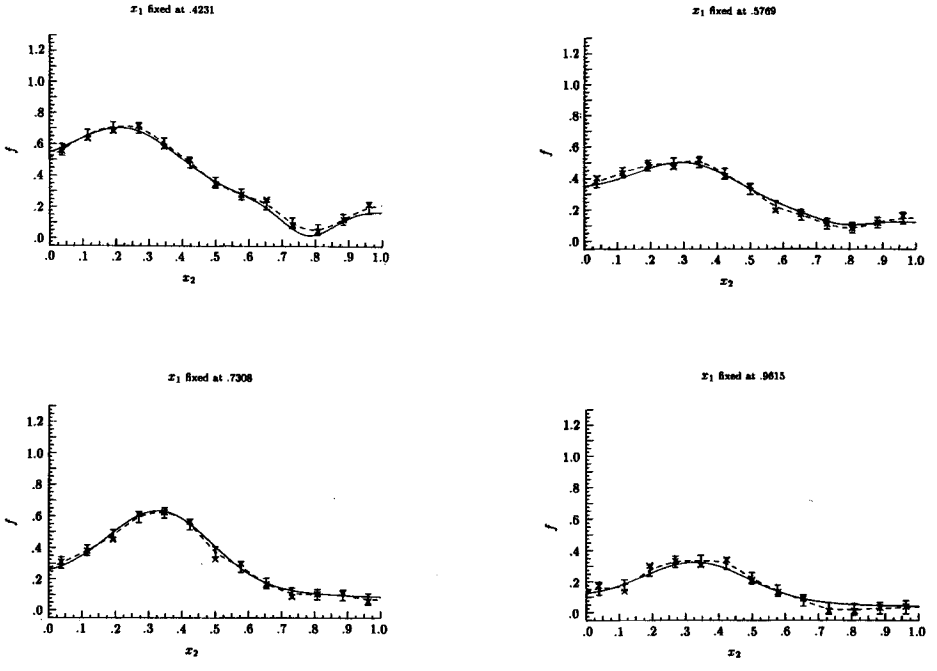

FIG. 5.2. *Cross sections of $f, f_{\hat{\lambda}}$, and "confidence intervals."*

working with the "average coverage probability" (ACP), defined by

$$\frac{1}{n}\sum_{i=1}^{n} P\{f(t_i) \in C_{.95}(t_i)\}$$

where $C_{.95}(t_i)$ is the ith confidence interval. These and similar confidence intervals have also been discussed by Silverman (1984) and O'Sullivan (1986a).

Nychka (see also Shiau (1985)) argued that a confidence interval based on the distribution $\mathcal{N}(0, \mu_i^2 + \sigma_i^2)$ should not be, in a practical sense, too far from correct when the true distribution is $N(\mu_i, \sigma_i^2)$, provided that μ_i^2 is not large compared to σ_i^2. Here let $\mu_i = Ef_{\hat{\lambda}}(t_i) - f(t_i)$ and $\sigma_i^2 = E(f_{\hat{\lambda}}(t_i) - Ef_{\hat{\lambda}}(t_i))^2$. We are, "on the average," replacing $\mathcal{N}(\mu_i, \sigma_i^2)$ with $\mathcal{N}(0, \mu_i^2 + \sigma_i^2)$, since $1/n \sum_{i=1}^n (\mu_i^2 + \sigma_i^2) = b^2(\hat{\lambda}) + \sigma^2 \mu_2(\hat{\lambda})$ $\simeq b^2(\lambda^*) + \sigma^2 \mu_2(\lambda^*) = \alpha\sigma^2/n \sum_{i=1}^n a_{ii}(\lambda^*)(1 + o(1))$ by (5.1.4). The minimization of $T(\lambda)$ with respect to λ entails that the square bias $b^2(\lambda^*)$ be of the same order as the variance $\sigma^2\mu_2(\lambda^*)$. In examples it tends to be of moderate size with respect to the variance.

Considering the case of univariate spline smoothing in W_m, we remark that if f is going to be in $\mathcal{H}_{R^p}$ of (4.5.4) for some $p > 1$, then f must satisfy some boundary

conditions. For example, let $p = 1 + k/m$ for some $k \leq m$; then $f \in \mathcal{H}_{R^{1+k/m}}$ if $f^{(k+m)} \epsilon \mathcal{L}_2$ and $f^{(\nu)}(0) = f^{(\nu)}(1) = 0$ for $\nu = m, m+1, \ldots, m+k-1$. Thus f can be "very smooth" in the interior of $[0,1]$ in the sense that $f^{(k+m)} \epsilon \mathcal{L}_2$, but if f does not satisfy the additional boundary conditions, then the higher-order convergence rates will not hold (see Rice and Rosenblatt (1983)).

In the case of confidence intervals, if f is "very smooth" in the interior, but fails to satisfy the higher-order boundary conditions, this would tend to cause the 5 percent of coverage failures for 95 percent confidence intervals to repeatedly fall near the boundary. This is similar to the way that the failed confidence intervals tend to repeat over a break in the first derivative of the true f if it occurs in the interior of $[0,1]$. (See Wahba (1983) for examples of this.) Nychka (1988) has proposed procedures for excluding the boundaries.

5.2 Estimate-based bootstrapping.

Another approach to confidence interval estimation may be called "estimate-based bootstrapping" (see also Efron (1982), Efron and Tibshirani (1986)). It goes as follows. From the data obtain $f_{\hat{\lambda}}$ and $\hat{\sigma}^2(\hat{\lambda})$, then, pretending that $f_{\hat{\lambda}}$ is the "true" f, generate data

$$\tilde{y}_i = f_{\hat{\lambda}}(t_i) + \tilde{\epsilon}_i,$$

where $\tilde{\epsilon}_i \sim \mathcal{N}(0, \hat{\sigma}^2(\hat{\lambda}))$, from a random number generator. (Here we are supposing that $L_i f = f(t_i)$.) Then find $\tilde{f}_{\hat{\lambda}}$, based on the data $\tilde{y}$. Upon repeating this calculation l times (with l different $\tilde{\epsilon}$), one has a distribution of l values of $\tilde{f}_{\hat{\lambda}}(t_i)$ at each t, and the $\alpha/2$ lth and $(1-\alpha/2)$ lth values can be used for a "confidence interval" (see, for example O'Sullivan (1988a)). The properties of these "confidence intervals" are not known. Plausible results have been obtained in simulation experiments, however. It is possible that the results will be too "rosy," since $f_{\hat{\lambda}}$ can be expected to display less "fine structure" than f. It would be a mistake to take the raw residuals $\hat{\epsilon}_i(\hat{\lambda}) = y_i - f_{\hat{\lambda}}(t_i)$, $i = 1, \ldots, n$ and generate data by

$$\tilde{y}_i = f_{\hat{\lambda}}(t_i) + \tilde{\epsilon}_i$$

where $\tilde{\epsilon}_i$ is drawn from the population $\{\hat{\epsilon}_i(\hat{\lambda}), \ldots, \hat{\epsilon}_n(\hat{\lambda})\}$, since $1/n \sum \hat{\epsilon}_i^2(\hat{\lambda}) = \mathrm{RSS}(\hat{\lambda}) \simeq \sigma^2/\mathrm{Tr}\,(I - A(\hat{\lambda}))$, the $\tilde{\epsilon}_i$ should be corrected for d.f. noise first if this approach were to be used.

Other important diagnostic tools are discussed in Eubank (1984, 1985).