CHAPTER 3

Equivalence and Perpendicularity, or, What's So Special About Splines?

3.1 Equivalence and perpendicularity of probability measures.

In Section 1.2 we considered a penalty functional that is a seminorm in W_m and a penalty functional that is a seminorm in $\tilde{W}_m$, a topologically equivalent space. Note that it took several additional parameters (the a_i's) to specify the seminorm in $\tilde{W}_m$. Aside from some computational advantages (considerable in the case of polynomial splines) why, in practical work, should we choose one penalty functional over another?

Continuing with this inquiry, one may ask if there is a particular reason for using a kriging estimator with $K(\tau)$ of (2.5.2) given by

$$K(\tau) = \sum_{p=0}^{k} (-1)^{p+1} a_p \tau^{2p+1},$$

where the a_p must be estimated from the data, rather than the thin-plate spline estimate, that corresponds to the simpler

$$K(\tau) = (-1)^{p+1} a_p \tau^{2p+1}.$$

How far should one go in estimating parameters in the penalty functional when valid prior information is not otherwise available? The theory of equivalence and perpendicularity gives an answer to this question. We will now describe the results we need.

A probability measure P_1 is said to dominate another measure P_2 ($P_1 \succ P_2$) if $P_1(A) = 0 \Rightarrow P_2(A) = 0$. P_1 is said to be equivalent to P_2 ($P_1 \equiv P_2$) if $P_2 \prec P_1$ and $P_2 \succ P_1$. P_1 is said to be perpendicular to P_2 if there exists an event A such that $P_1(A) = 0$ and $P_2(A) = 1$. It is known that Gaussian measures are either equivalent or perpendicular. Any two nontrivial Gaussian measures on E^1 are equivalent, and two Gaussian measures on E^d are equivalent if the null spaces of their covariance matrices coincide, otherwise they are perpendicular.

Considering Gaussian measures on infinite sequences of zero-mean independent random variables $X_1, X_2, \ldots$, Hajek (1962a,b) has given necessary and sufficient conditions for equivalence. Let $EX_\nu^2 = \sigma_\nu^2(1)$ under P_1 and $\sigma_\nu^2(2)$ under P_2. If $\sigma_\nu^2(1) > 0$ and $\sigma_\nu^2(2) = 0$ or $\sigma_\nu^2(1) = 0$ and $\sigma_\nu^2(2) > 0$ for some ν, then the two processes are perpendicular. Suppose $\sigma_\nu^2(1)$ and $\sigma_\nu^2(2)$ are positive or zero together. Then $P_1 \equiv P_2$ if and only if $\sum_{\nu=1}^\infty (1-\sigma_\nu^2(1)/\sigma_\nu^2(2))^2 < \infty$.

Let us now consider a stochastic process $X(t),\ t \in \mathcal{T}$ with the Karhunen–Loeve expansion (see Section 1.1)

$$X(t) = \sum_{\nu=1}^{\infty} X_\nu \ \Psi_\nu(t), \tag{3.1.1}$$

where $\{\Psi_\nu\}$ is an orthonormal sequence in $\mathcal{L}_2(\mathcal{T})$ and the $\{X_\nu\}$ are independent, zero-mean Gaussian random variables with $EX_\nu^2 = \sigma_\nu^2(i) > 0$ under $P_i,\ i = 1,2$. Then it follows from Hajek's result that P_1 and P_2 will be equivalent or perpendicular accordingly as $\sum_{\nu=1}^\infty (1-\sigma_\nu^2(1)/\sigma_\nu^2(2))^2$ is finite or infinite.

Now consider the following example of Section 2.1:

$$\sigma_\nu^2(1) = b_1(2\pi\nu)^{-2m_1}, \qquad \sigma_\nu^2(2) = b_2(2\pi\nu)^{-2m_2},$$

P_1 and P_2 will be equivalent if $b_1 = b_2$ and $m_1 = m_2$, and perpendicular otherwise. (Here m_1 and m_2 need not be integers.)

Suppose we have a prior distribution with $\sigma_\nu^2 = b_*(2\pi\nu)^{-2m_*}$, where b_* and m_* are *unknown.* The perpendicularity fact above means that we can expect to find a consistent estimator for (b,m) given $X(t_1),\dots,X(t_n)$ as $t_1,\dots,t_n$ become dense in $\mathcal{T}$. To see this, let (b_*,m_*) be any fixed value of (b,m). Then there exists a set $A(b_*,m_*)$ in the sigma field for $\{X(t),\ t\in\mathcal{T}\}$, equivalently in the sigma field for $\{X_1,X_2,\dots\}$ such that $P(\{X_1,X_2,\dots\} \in A(b_*,m_*)) = 1$ if (b_*,m_*) is true and zero otherwise. Thus the estimate is formed by determining in which $A(b,m)$ $\{X_1,X_2,\dots\}$ lies. Under some mild regularity conditions (for example, $X(t),\ t\in\mathcal{T}$ continuous in quadratic mean), it is sufficient to observe $X(t)$ only for t in a dense subset of $\mathcal{T}$.

Now consider the case with $\sigma_\nu^2(1) = (2\pi\nu)^{-2m}$ and $\sigma_\nu^2(2) = [(2\pi\nu)^{2m} + \theta^2(2\pi\nu)^{2(m-1)}]^{-1}$. By considering the problem in its usual complex form (we omit the details), the $\{\sigma_\nu^2(2)\}$ can be shown to be the eigenvalues associated with the penalty $\int_0^1 [f^{(m)}(t) + \theta f^{(m-1)}(t)]^2\,dt$ for the periodic spline case of Section 2.1.

Then

$$\frac{\sigma_\nu^2(1)}{\sigma_\nu^2(2)} = 1 + \frac{\theta^2}{(2\pi\nu)^2}$$

and

$$\sum_{\nu=1}^{\infty}\left(1-\frac{\sigma_\nu^2(1)}{\sigma_\nu^2(2)}\right)^2 = \sum_{\nu=1}^{\infty}\left[\frac{\theta^2}{(2\pi\nu)^2}\right]^2 < \infty$$

so that P_1 and P_2 are equivalent. This means that there *cannot* be a consistent estimate of θ, since if there were we would know θ "perfectly" (w.pr.1) given $X(t),\ t\in\mathcal{T}$, and then we could tell w.pr.1 which of P_1 or P_2 is true, which contradicts the fact that they are equivalent.

Hajek (1962a) considers the nonperiodic case on a finite interval of the real line. The result, loosely stated, is that if $X(t)$ formally satisfies

$$\sum_{j=0}^{m} a_{m-j}^{(i)} X^{(j)} = dW \tag{3.1.2}$$

under P_i, $i = 1, 2$ and the boundary random variables are equivalent, then $P_1 \equiv P_2$ if $a_0^{(1)} = a_0^{(2)}$ and $P_1 \perp P_2$ if $a_0^{(1)} \neq a_0^{(2)}$. Thus $a_1, \ldots, a_{m-1}$ cannot be estimated consistently from data on a finite interval. More generally, if X is the restriction to a finite interval of a stationary Gaussian process on the real line with spectral density

$$f(w) = \left| \frac{\sum_{k=0}^{q(i)} b_{q-k}^{(i)} (iw)^k}{\sum_{k=0}^{p(i)} a_{p-k}^{(i)} (iw)^k} \right|^2, \quad i = 1, 2$$

then $P_1 \equiv P_2$ if $q(1) - p(1) = q(2) - p(2)$ and $a_0(1)/b_0(1) = a_0(2)/b_0(2)$ and $P_1 \perp P_2$ otherwise (see Hajek (1962b, Thm. 2.3) for further details). Parzen (1963) discusses conditions for the equivalence and perpendicularity in terms of the properties of reproducing kernels.

3.2 Implications for kriging.

Now we will examine the periodic version of the prior related to kriging, from the d-dimensional version of (2.5.4), to see which coefficients in the variogram should be consistently estimable from data on a bounded region. To see what happens most easily in the d-dimensional periodic case, it is convenient to think of the eigenfunctions of the reproducing kernel in complex form. Letting $\boldsymbol{\nu} = (\nu_1, \nu_2, \ldots, \nu_d)$, then the eigenfunctions $\Phi_{\boldsymbol{\nu}}$ are

$$\Phi_{\boldsymbol{\nu}}(x_1, \ldots, x_d) = e^{2\pi i(\nu_1 x_1 + \ldots + \nu_d x_d)}, \quad \begin{array}{l} \nu_j = \ldots -1, 0, 1, \ldots \\ j = 1, 2, \ldots, d. \end{array}$$

Observing that

$$\begin{aligned} \Delta \Phi_{\boldsymbol{\nu}} &= [(2\pi\nu_1)^2 + \ldots + (2\pi\nu_d)^2] \quad \Phi_{\boldsymbol{\nu}} \\ &= \|2\pi\boldsymbol{\nu}\|^2 \quad \Phi_{\boldsymbol{\nu}} \end{aligned}$$

it is not hard to see that the $\sigma_{\boldsymbol{\nu}}^2$'s for the d-dimensional periodic stochastic process corresponding to the thin-plate spline penalty functional are

$$\sigma_{\boldsymbol{\nu}}^2(1) = \|2\pi\boldsymbol{\nu}\|^{-2m}$$

and the eigenvalues corresponding to the d-dimensional periodic version of kriging, given by the d-dimensional version of the covariance of (2.5.4) are, from (2.5.5),

$$\sigma_{\boldsymbol{\nu}}^2(2) = \sum_{l=m}^{m+k} \frac{\alpha_{l-m}}{\|2\pi\boldsymbol{\nu}\|^{2l}} = \frac{1}{\|2\pi\boldsymbol{\nu}\|^{2m}} \left(1 + \sum_{j=1}^{k} \frac{\alpha_j}{\|\boldsymbol{\nu}\|^{2j}}\right), \tag{3.2.1}$$

where we have set $\alpha_0 = 1$. Then

$$\sum_{\nu_1 \ldots \nu_d=-\infty}^{\infty} \left(1 - \frac{\sigma_{\nu}^2(2)}{\sigma_{\nu}^2(1)}\right)^2 = \sum_{\nu_1, \ldots \nu_d=-\infty}^{\infty} \left(\sum_{j=1}^{k} \frac{\alpha_j}{\|\nu\|^{2j}}\right)^2. \tag{3.2.2}$$

Looking at the lowest order term in (3.2.2)

$$\sum_{\nu_1, \ldots, \nu_d=-\infty}^{\infty} \left(\frac{\alpha_1}{\|\nu\|^2}\right)^2 = \alpha_1^2 \sum_{\nu_1, \ldots, \nu_d=-\infty}^{\infty} \frac{1}{\|\nu\|^4}, \tag{3.2.3}$$

we estimate the sum by

$$\int \cdots \int_{\|\mathbf{x}\|>1} \frac{1}{\|\mathbf{x}\|^4} dx_1 \ldots dx_d = \int_{r>1} \cdots \int \frac{1}{r^4} r^{d-1}\, dr$$

where $\|\mathbf{x}\|^2 = x_1^2 + \ldots + x_d^2$ and the expression on the right is obtained by transforming to polar coordinates. The expression on the right will be finite if $4 - (d-1) = 5 - d > 1$. In particular, it will be finite for $d = 1, 2$, and 3. Thus the right-hand side of (3.2.3) will be finite for $d = 1, 2$, and 3, and the conclusion to be drawn is that P_1 and P_2 here are equivalent, and $\alpha_1, \ldots, \alpha_k$ cannot be estimated consistently. One can argue that the situation is analogous for the stochastic process with variogram

$$\tilde{E}(\tau) = |\tau|^{2m-1} + \alpha_1 |\tau|^{2m+1} + \ldots + \alpha_k |\tau|^{2m+2k-1}, \; (2m-1) > d \tag{3.2.4}$$

and if this is so, then $\alpha_1, \ldots, \alpha_k$ cannot be estimated consistently from data in a bounded region for $d = 1, 2$, and 3. Thus, in practice, if prior information is not available concerning $\alpha_1, \ldots, \alpha_k$, one might as well set $\alpha_1, \ldots, \alpha_k$ to zero, that is, use the thin-plate spline. I made observations to this effect in Wahba (1981b). In an elegant series of papers Stein (1987, 1988, 1990) has obtained results that imply the same thing.