Why

We use a normal random function model to make a regressor.

Definition

Let $F : \Omega \to (A \to \mathbf{R})$ be a normal random function with mean function $m : A \to \mathbf{R}$ and covariance function $k : A \times A \to \mathbf{R}$ over the probability space $(\Omega, \mathcal{A}, \mathbf{P})$. Let the family of random variables (or stochastic process) of F be $f : A \to (\Omega \to \mathbf{R})$.

Let e be a normal random vector with mean zero and covariance Σ_e. Let $a^1, \ldots, a^n \in A$. We sometimes call the sequence $a^1, \ldots, a^n$ the *design*. Define $y : \Omega \to \mathbf{R}^d$ by

$$y_i = f(a^i) + e_i$$

We call y the *observation vector* or *observation random vector*. We call e the *error vector* or *noise vector*. In this context, $f(a^i)$ is sometimes called the *signal*.

Let $b^1, \ldots, b^m \in A$. Define $z : \Omega \to \mathbf{R}^d$ by $z_i = f(b^i)$ for $i = 1, \ldots, n$. So z_i is the random variable corresponding to the family at index $b^i \in A$. Then (y, z) is normal. We call the conditional density of z given y the *predictive density* for b given a.

Proposition 1. *Define $m_a \in \mathbf{R}^n$ by $(m(a^1), \cdots, m(a^n))^\top$ and define m_b by $(m(b^1), \cdots, m(b^m))^\top$.[1] Define $\Sigma_a \in \mathbf{R}^{n \times n}$ by*

$$\begin{pmatrix} k(a^1, a^1) & \cdots & k(a^1, a^n) \\ \vdots & \ddots & \vdots \\ k(a^n, a^1) & \cdots & k(a^n, a^n) \end{pmatrix}$$

and define $\Sigma_{ba} \in \mathbf{R}^{m \times n}$ by

$$\begin{pmatrix} k(b^1, a^1) & \cdots & k(b^1, a^n) \\ \vdots & \ddots & \vdots \\ k(b^m, a^1) & \cdots & k(b^m, a^n) \end{pmatrix}.$$

---

[1]Future editions will fix the re-use of the symbol m.

The predictive density $g_{z|y}(\cdot, \gamma) : \mathbf{R}^m \to \mathbf{R}$ of $b \in A$ for design $a^1, \ldots, a^n$ is normal with mean.

$$m_b + K_{ba}(K_a + \Sigma_e)^{-1}(\gamma - m_a)$$

and covariance

$$\Sigma_b - \Sigma_{ba}(\Sigma_a + \Sigma_e)^{-1}\Sigma_{ab}.$$