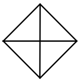


Why

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

Definition

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

Let e be a normal random vector with mean zero and covariance Σ_e. Let $a^1, \dots, a^n \in A$. We sometimes call the sequence $a^1, \dots, a^n$ the *design*. Define $y : \Omega \to \mathsf{R}^d$ 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 $b^1, \dots, b^m \in A$. Define $z : \Omega \to \mathsf{R}^d$ by $z_i = f(b^i)$ for $i = 1, \dots, n$. 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 $m_a \in \mathsf{R}^n$ by $(m(a^1), \cdots, m(a^n))^\top$ and* *define* m_b *by* $(m(b^1), \cdots, m(b^m))^\top$.[1] *Define* $\Sigma_a \in \mathsf{R}^{n \times n}$ *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 $\Sigma_{ba} \in \mathsf{R}^{m \times n}$ *by*

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

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

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

and covariance

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

Regressor

Ultimately, we want a regressor $h : A \to \mathsf{R}$. We maximize the the predictive density. The *normal random function predictor* or *gaussian process predictor* for dataset $(a^1, \gamma_1), \dots, (a^n, \gamma_n)$ in $A \times \mathsf{R}$ is $h : A \to \mathsf{R}$ defined by

$$h(x) = m(x) + \Big(k(x, a^1) \cdots k(x, a^n)\Big) \left(\Sigma_a + \Sigma_e\right)^{-1} (\gamma - m_a).$$

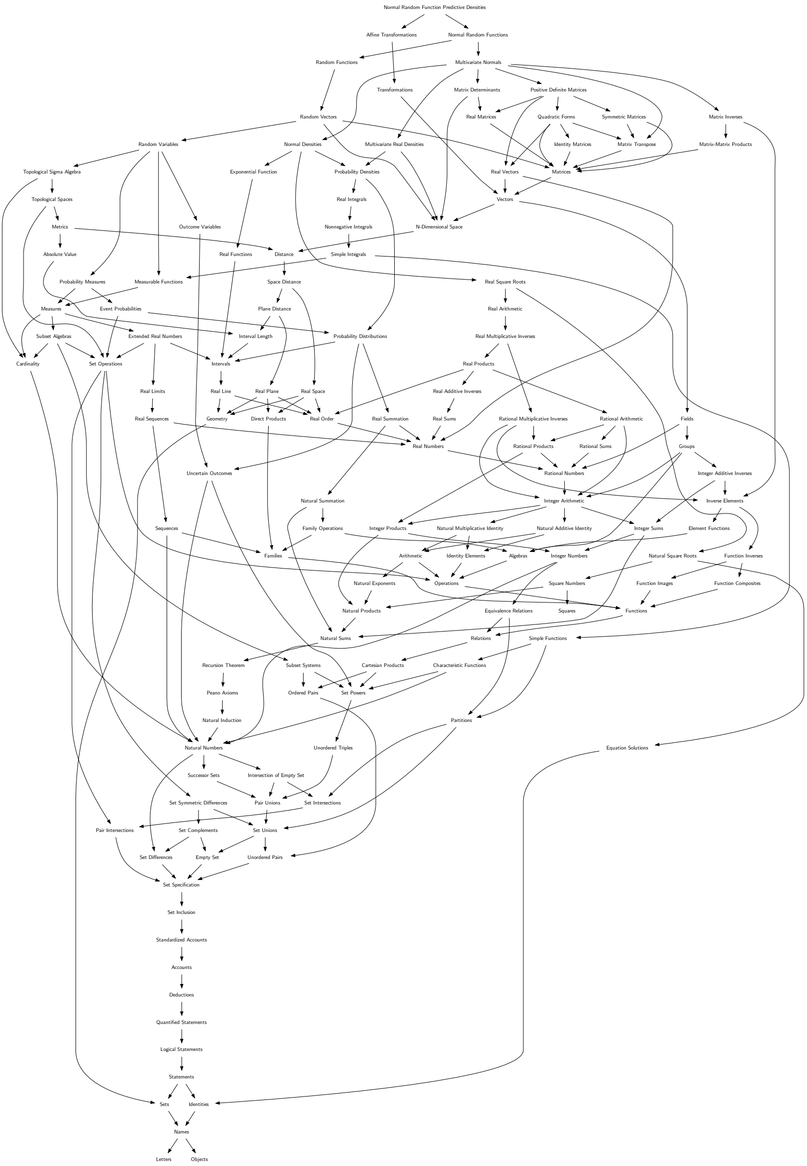
[1] Future editions will fix the re-use of the symbol m.

Notice that h is an affine function of γ. If the mean function $m \equiv 0$ then h is linear in γ. This is sometimes called a *linear estimator.*[2] Alternatively, notice (in the zero mean setting) that h is a linear combination of n kernel function $k(x, a^i)$ for $i = 1, \dots, n$. Specifically, h is a linear combination of

The process of using a normal random function predictor is often called *Gaussian process regression.* The upside is that a gaussian process predicor interpolates the data, is smooth, and the so-called variance increases with the distance from the data.[3]

[2] We avoid the other terminology we have seen used—linear predictor—because the predictor h is not linear in its input x.

[3] This last piece is true for certain covariance kernels and will be clarified in future editions.


Normal Random Function Predictive Densities
Affine Transformations
Normal Random Functions
Random Functions
Multivariate Normals
Transformations
Matrix Determinants
Positive Definite Matrices
Random Vectors
Real Matrices
Quadratic Forms
Symmetric Matrices
Matrix Inverses
Random Variables
Normal Densities
Multivariate Real Densities
Identity Matrices
Matrix Transpose
Matrix-Matrix Products
Topological Sigma Algebra
Exponential Function
Probability Densities
Real Vectors
Matrices
Topological Spaces
Real Integrals
Vectors
Metrics
Outcome Variables
Nonnegative Integrals
N-Dimensional Space
Absolute Value
Real Functions
Distance
Simple Integrals
Probability Measures
Measurable Functions
Space Distance
Real Square Roots
Measures
Event Probabilities
Plane Distance
Real Arithmetic
Subset Algebras
Extended Real Numbers
Interval Length
Probability Distributions
Real Multiplicative Inverses
Cardinality
Set Operations
Intervals
Real Products
Real Limits
Real Line
Real Plane
Real Space
Real Additive Inverses
Real Sequences
Geometry
Direct Products
Real Order
Real Summation
Real Sums
Rational Multiplicative Inverses
Rational Arithmetic
Fields
Real Numbers
Rational Products
Rational Sums
Groups
Uncertain Outcomes
Rational Numbers
Integer Additive Inverses
Natural Summation
Integer Arithmetic
Inverse Elements
Sequences
Family Operations
Integer Products
Natural Multiplicative Identity
Natural Additive Identity
Integer Sums
Element Functions
Families
Arithmetic
Identity Elements
Algebras
Integer Numbers
Natural Square Roots
Function Inverses
Natural Exponents
Operations
Square Numbers
Function Images
Function Composites
Natural Products
Equivalence Relations
Squares
Functions
Natural Sums
Relations
Simple Functions
Recursion Theorem
Subset Systems
Cartesian Products
Characteristic Functions
Peano Axioms
Ordered Pairs
Set Powers
Natural Induction
Partitions
Natural Numbers
Unordered Triples
Equation Solutions
Successor Sets
Intersection of Empty Set
Set Symmetric Differences
Pair Unions
Set Intersections
Pair Intersections
Set Complements
Set Unions
Set Differences
Empty Set
Unordered Pairs
Set Specification
Set Inclusion
Standardized Accounts
Accounts
Deductions
Quantified Statements
Logical Statements
Statements
Sets
Identities
Names
Letters
Objects