Statistics 250:

A graduate topics course on applied and computational statistics.

Ethan Anderes
Fall 2019

Central focus of the course

- Random fields, stochastic processes & Spatial stats:
- Topics centered around real data projects
- The FFT and programming language Julia.

Some persistant themes

- The power of FFT
- Deep dives into real data gives insight on statistical methodology.
- Random fields & covariance functions

$\Updownarrow$ essentially the same via analogy

Random vectors & covariance matrices

e.g. a random field $Z(x)$ is typically understood as a random function of $x \in \mathbb{R}^d$ and a random vector $X = (X_1, \ldots, X_n)$ as a tuple of random variables.

However, both are essentially the same, an indexing of random variables:

index $i \in \{1, \ldots, n\} \longmapsto$ random variable X_i

index $x \in \mathbb{R}^d \longmapsto$ random variable $Z(x)$

ndeed most concepts for X have an analogy for Z.

cov matrix eigen decomp

$\Longleftrightarrow$ Karhunen-Loève decomp

Cholesky decomp

$\Longleftrightarrow$ triangular embedding

FFT & circulant matrices

$\Longleftrightarrow$ Fourier trans & stationary fields.

$N(0, \frac{1}{\Omega_x} I)$, Ω_x = pixel volumn

$\Longleftrightarrow$ white noise

Random Fields

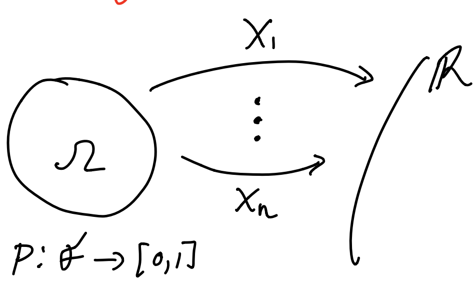
A random field (RF) is just like a random vector (RV) $(X_1, \dots, X_n)$ but where the index i, that enumerates the random variables X_i, ranges over some subset $\mathcal{T} \subset \mathbb{R}^d$.

RV $X = (X_1, \dots, X_n)$

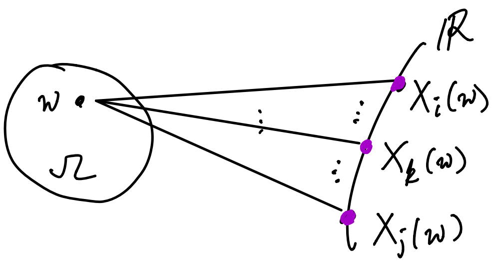
A collection of random variables

$$\{X_i : i = 1, 2, \dots, n\}$$

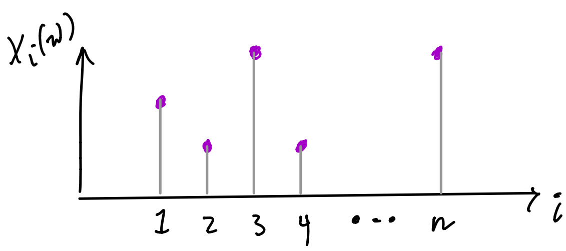
all defined on the same probability space $(\Omega, \mathcal{F}, P)$



Fixing a draw $\omega \in \Omega$ gives n real numbers



which can be plotted as a function of index i

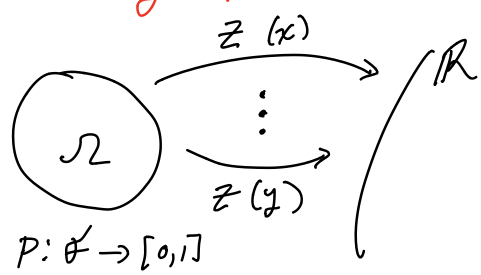


RF $(Z(x) : x \in \mathcal{T})$

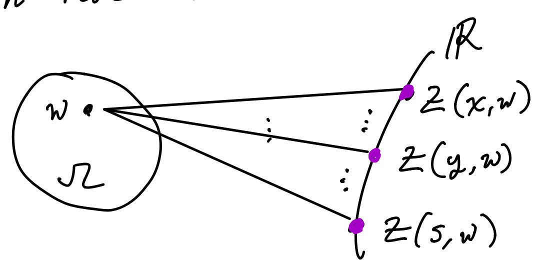
A collection of random variables

$$\{Z(x) : x \in \mathcal{T}\}$$

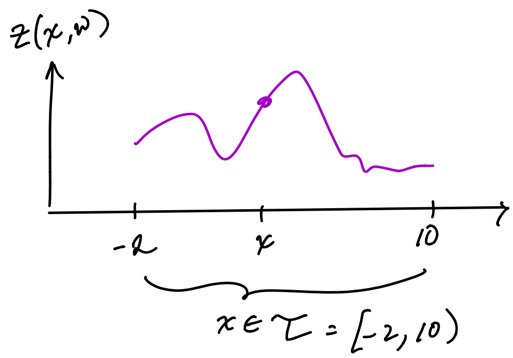
all defined on the same probability space $(\Omega, \mathcal{F}, P)$



Fixing a draw $\omega \in \Omega$ gives n real numbers



which can be plotted as a function of index $x \in \mathcal{T}$



Notice that this gives two viewpoints of a RV X and a RF Z

(i) collection of random variables:

$$X_i \text{ for } i = 1, \dots, n$$

$$Z(x) \text{ for } x \in \mathcal{X}$$

i.e. a single random variable associated to each index i or x.

i.e. $$\{ \omega \mapsto X_i(\omega) : i = 1, \dots, n \}$$
$$\& \ \{ \omega \mapsto Z(x, \omega) : x \in \mathcal{X} \}$$

(ii) A random function

$$\left. \begin{aligned} X &: \{1, 2, \dots, n\} \longrightarrow \mathbb{R} \\ Z &: \mathcal{X} \longrightarrow \mathbb{R} \end{aligned} \right\}$$

i.e. a single function associated to each draw $\omega \in \Omega$

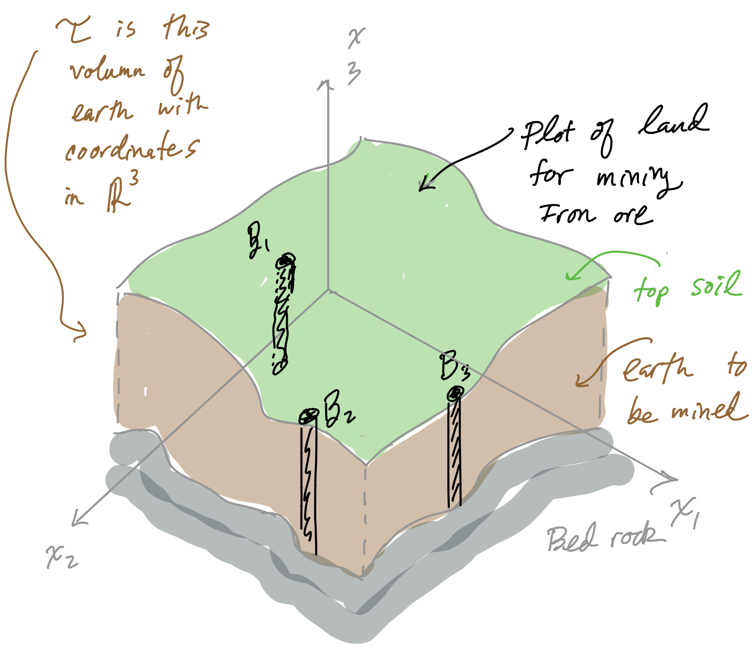
i.e. $$\{ i \mapsto X_i(\omega) : \omega \in \Omega \}$$
$$\& \ \{ x \mapsto Z(x, \omega) : \omega \in \Omega \}$$

Both (i) and (ii) are useful. (ii) is more intuitive but difficult to manage mathematically since probability measures on σ-fields over function spaces is <u>very</u> difficult.

To give a hint at the mathematical difficulty notice that we can use densities w.r.t Lebesgue measure on $\mathbb{R}^n$ to model the distribution of $(X_1, \dots, X_n)$.

However there doesn't really exist the notion of Lebesgue measure on $\mathbb{R}^\infty$ or $\mathbb{R}^{\mathbb{R}^d}$ so densities w.r.t. a canonical base measure are not available to us for modeling RFs.

I think its instructive to look at an example from mining which, in a way, started the field of spatial statistics:



For $x = (x_1, x_2, x_3)^T \in \mathcal{T}$ let

$\underbrace{Z(x)}_{\text{unknown, modeled as a RF.}}$ = Iron ore concentration in the earth at spatial location x.

Goal: Estimate

$$\int_{\mathcal{T}} Z(x)\,dx = \text{"total iron ore weight in } \mathcal{T}\text{"}$$

from bore hole samples $Y_1, \ldots, Y_n$ where $Y_i = \int_{B_i} Z(x)\,dx$.

Notice that a RF model on Z allows one to estimate with

$$E\left(\int_{\mathcal{T}} Z(x)\,dx \,\middle|\, Y_1, \ldots, Y_n\right)$$

and quantify estimation error via

$$\operatorname{var}\left(\int Z(x)\,dz \,\middle|\, Y_1, \ldots, Y_n\right)$$

Note: The RF model ensures

$$Y_1, Y_2, \ldots, Y_n, \int_{\mathcal{T}} Z(x)\,dz$$

are random variables all defined on $(\Omega, \mathcal{F}, P)$ ... which allows us to make sense of the conditional dependence necessary for

$$E\left(\int_{\mathcal{T}} Z(x)\,dx \,\middle|\, Y_1, \ldots, Y_n\right)$$

$$\operatorname{var}\left(\int Z(x)\,dz \,\middle|\, Y_1, \ldots, Y_n\right)$$

Constructing random fields

There are two common ways to construct/model a random field

(e) Explicit construnction

(basis expansions for e.g.)

(i) Implicit construction using the finite dimensional distributions (fdd)

These are the set of all joint probability distributions of $(Z(t_1), \ldots, Z(t_n))^T$ for $n \in \mathbb{N}$ and $t_1, \ldots, t_n \in \mathcal{T}$

Example of (e): Cosine process

Let $\mathcal{T} = \mathbb{R}$

$\lambda_1, \ldots, \lambda_n \in \mathbb{R}$

$X_1, \ldots, X_n, Y_1, \ldots, Y_n \overset{iid}{\sim} P$

and define the RF $Z(t)$ by

$$Z(t) := \underbrace{\sum_{i=1}^{n} X_i \sin(\lambda_i t) + Y_i \cos(\lambda_i t)}_{\text{Fixed non-random basis function with random coefficients}}$$

Since $X_1, \ldots, Y_1 \ldots$ can all be defined on a common prob space its obvious that $\{Z(t) : t \in \mathbb{R}\}$ is a RF.

Why is it called a cosine process (when there are sines too?)

The reason is that

$$A\cos(x) + B\sin(x) = \sqrt{A^2+B^2}\,\cos\left(x - \tan^{-1}\left(\frac{B}{A}\right)\right)$$

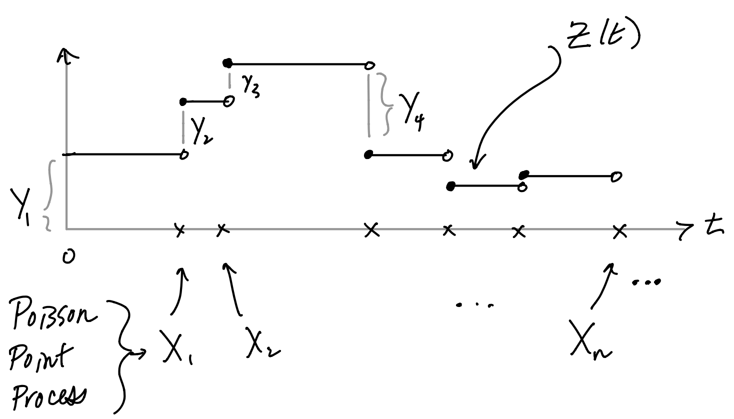
(side note: this $\tan^{-1}(B/A) \in (-\pi, \pi]$ is computed in Julia by atan(B,A))

Therefore

$$Z(t) := \sum_{i=1}^{n} R_i \cos(\lambda_i t - T_i)$$

with $R_i := \sqrt{X_i^2 + Y_i^2}$ and $T_i := \tan^{-1}\left(\frac{X_i}{Y_i}\right)$

Example of (e): Jump process



$X_1, X_2, \ldots$ random jump locations

$Y_1, Y_2, \ldots$ random function innovations

Example of (e):

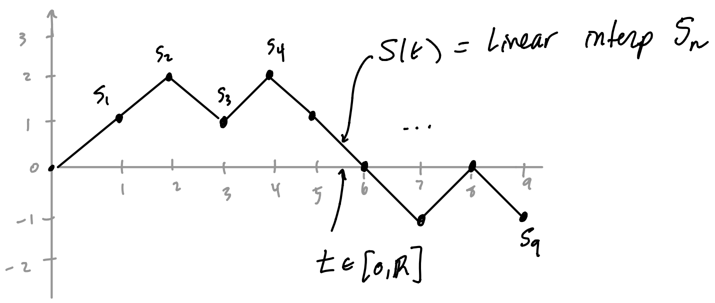
Brownian motion via limits of a R.W.
start with a random walk

$$S_0, S_1, S_2, \ldots$$

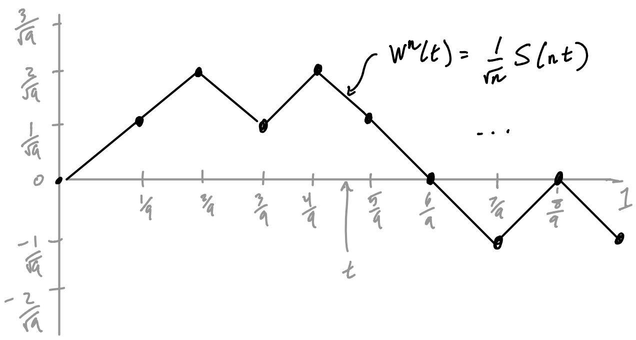
where

$$S_n = \sum_{k=1}^{n} R_k \,, \quad S_0 = 0$$

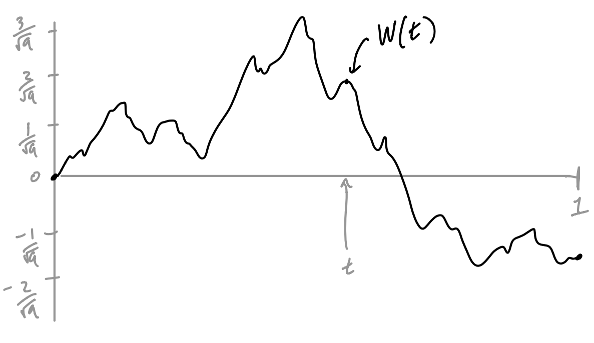
$$R_k \overset{iid}{\sim} \begin{cases} 1 & \text{w/prob } 1/2 \\ -1 & \text{w/prob } 1/2 \end{cases}$$



Now re-scale the x-axis by $\frac{1}{n}$ and the y-axis by $\frac{1}{\sqrt{n}}$



one can show W^n a functial limit (in distribution) $W^n \xrightarrow{\mathcal{D}} W$ as $n \to \infty$



This is a direct construction of a Wiener Process $W(t)$ or Brownian motion on $\mathcal{T} = [0,1]$.

Side note: There is also a construction via basis expansion similar to the cosine process construction. (see Lévy)

Example of (i)

Now lets look an implicit construction of Brownian motion via the implicit technique (i.e. modeling the fdds) for constructing a Brownian motion $W(t)$

Let $\{W(t) : t \in [0,1]\}$ be a collection of random variables, all defined on a common probability space, which satisfies:

For any $n \in \mathbb{N}$ and $t_1, \ldots, t_n \in [0,1]$

$$\begin{bmatrix} W(t_1) \\ \vdots \\ W(t_n) \end{bmatrix} \sim N(0, \Sigma)$$

$$\Sigma_{ij} = \frac{1}{2}\left(|t_i| + |t_j| - |t_i - t_j|\right)$$
$$= \min(t_i, t_j)$$

defines the f.d.d.

Note:

It is not at all clear such a collection of R.V.s exists.

First, one needs to check that the matrices $\left(\min(t_i,t_j)\right)_{i,j=1}^{n}$ are valid cov matrices (i.e. pos. def.).

Second, one needs to check that the fdd don't contradict.

e.g. these would contradict:

$$\begin{bmatrix} W(1/4) \\ W(1/2) \end{bmatrix} \sim N\left(\begin{bmatrix} 5 \\ 0 \end{bmatrix}, I\right) \quad \& \quad W(1/4) \sim N(0,1)$$

Once these are checked then we can invoke Kolmogorov's Consistency Thm to establish the existance of a RF $W(t)$ satisfying the fdd specified above.

Gaussian random fields (GRF)

A random field $(Z(x) : x \in \mathcal{X})$ is Gaussian if all the fdds of Z are multivariate Gaussian.

$$\Updownarrow$$

if for any $n \in \mathbb{N}$ and $x_1, \dots, x_n \in \mathcal{X}$ the random vector $(Z(x_1), \dots, Z(x_n))^T$ is jointly Gaussian, i.e. $\exists \mu \in \mathbb{R}^n$ and positive definite $\Sigma \in \mathbb{R}^{n \times n}$ s.t.

$$P\left((Z(x_1), \dots, Z(x_n)) \in z + dz\right) \simeq \frac{\exp\left(-\frac{1}{2}(z-\mu)^T \Sigma^{-1}(z-\mu)\right)}{\sqrt{\det(2\pi\Sigma)}}\, dz$$

In which case

$$\mu = \begin{bmatrix} E(Z(x_1)) \\ \vdots \\ E(Z(x_n)) \end{bmatrix}$$

$$\Sigma_{ij} = \operatorname{cov}\left(Z(x_i), Z(x_j)\right)$$

and usually denoted

$$(Z(x_1), \dots, Z(x_n))^T \sim N(\mu, \Sigma).$$

Instead of indexing μ & Σ with i,j as

$$\mu_i = E(Z(x_i))$$

$$\Sigma_{ij} = \operatorname{cov}(Z(x_i), Z(x_j))$$

and needing to redefine it for each $x_1, \dots, x_n$ its natural to just use functions:

$$\mu : \mathcal{T} \to \mathbb{R}$$

$$\Sigma : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$$

such that

$$\mu(x) = E(Z(x))$$

$$\Sigma(x,y) = \operatorname{cov}(Z(x), Z(y))$$

which are called the mean and covariance functions of the RF $(Z(x) : x \in \mathcal{T})$

(notice that for general RFs these may not exist)

Warning!

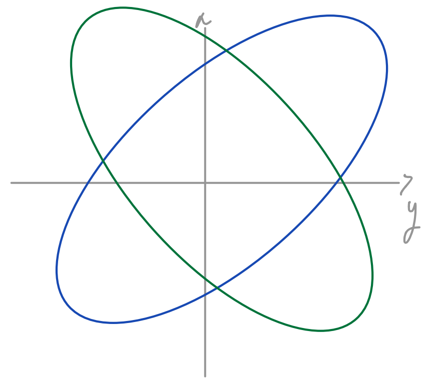
A random vector $(Z_1, \dots, Z_n)^T$ can have all marginals Z_i Gaussian but not be jointly Gaussian.

Example: Let

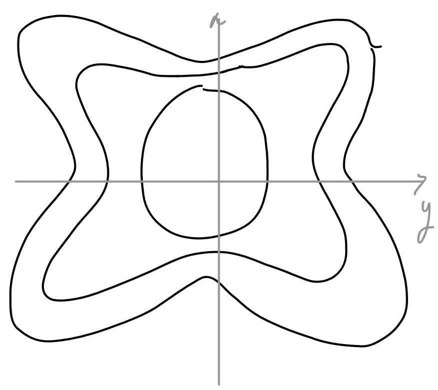
$f(x,y)$ denote the density of $N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 1/2\\ 1/2 & 1\end{pmatrix}\right)$

$g(x,y)$ denote the density of $N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & -1/2\\ -1/2 & 1\end{pmatrix}\right)$

Contours of $f(x,y)$ and $g(x,y)$



Contours of $\frac{1}{2} f(x,y) + \frac{1}{2} g(x,y)$



It is easy to see that

$$\frac{1}{2} f(x,y) + \frac{1}{2} g(x,y)$$

Is a density and its marginals are $N(0,1)$ ... but the joint is not Gaussian.

Example:

A less trivial example can be found as follows.

Let $(Y(x): x \in \mathcal{T})$ a non-Gaussian RF and try to transform it to Gaussian by a pointwise transformation

$$F_x(y) := P(Y(x) \le y) \equiv \text{CDF for } Y(x)$$

$$\Phi(w) := P(W \le w) \text{ where } W \sim N(0,1)$$

$$Z(x) := \Phi^{-1}\Big(\underbrace{F_x(Y(x))}_{\sim \mathrm{Unif}(0,1)}\Big)$$

At each $x \in \mathcal{T}$, $Z(x) \sim N(0,1)$ but is not, in general, a Gaussian random field

Here is a useful critearion for checking joint Gaussianity.

Theorem:

A RV $X \in \mathbb{R}^n$ is jointly Gaussian iff $\forall \beta \in \mathbb{R}^n$. $\beta^T X$ is univariate Gaussian

Notation:

$$\boxed{Z \sim GRF_{\mathcal{T}}(\mu, K)}$$

$$\Updownarrow$$

$(Z(x) : x \in \mathcal{T})$ is a Gaussian random field (separable version) with mean function $\mu : \mathcal{T} \to \mathbb{R}$ and covariance function $K : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$

Covariance functions ... The basics

We already know that given a GRF $(Z(x): x \in \mathcal{T})$ there exists a covariance function $K : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ s.t.

$$K(x,y) = \mathrm{cov}(Z(x), Z(y)).$$

Two key properties of this K

(I) $K(x,y) = K(y,x)$

which follows since $\mathrm{cov}(Z(x), Z(y))$ is the same as $\mathrm{cov}(Z(y), Z(x))$.

For the second property let $n \in \mathbb{N}$ and choose any $x_1, \dots, x_n \in \mathcal{T}$ and $c_1, \dots, c_n \in \mathbb{R}$. Notice

$$\begin{aligned} 0 &\leq \mathrm{var}\Big(\sum_{i=1}^{n} c_i Z(x_i)\Big) \\ &= \sum_{i,j=1}^{n} c_i c_j \,\mathrm{cov}(Z(x_i), Z(x_j)) \\ &= \sum_{i,j=1}^{n} c_i c_j K(x_i, x_j). \end{aligned}$$

In particular

(II) For any $n \in \mathbb{N}$, $x_1, \dots, x_n \in \mathcal{T}$ and $c_1, \dots, c_n \in \mathbb{R}$ one has

$$0 \leq \sum_{i,j=1}^{n} K(x_i, x_j)$$

Any $K : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ which satisfies (I) & (II) is said to be positive definite over $\mathcal{T}$.

Claim: Conditions (I) & (II) are necessary and sufficient for the existance of a GRF $(Z(x): x \in \mathcal{T})$ with $K(x,y) = \mathrm{cov}(Z(x), Z(y))$.

Due to the above claim we use positive define function on $\mathcal{T} \times \mathcal{T}$ interchangably with covariance function on $\mathcal{T}$

Note:

The mean and covariance function

$$\mu(x) = E(Z(x))$$
$$K(x,y) = \mathrm{cov}(Z(x), Z(y))$$

completely characterize $(Z(x): x \in \mathcal{T})$ if it is a (separable) Gaussian random field. In this case knowing $\mu(x)$ & $K(x,y)$ allows one to simulate, predict and do likelihood inference when working on a finite number of observation points $x_1, x_2, \dots, x_n \in \mathcal{T}$ without worrying about the rest of $\mathcal{T}$.

Example :

Simulating a GRF $(Z(x) : x \in \mathbb{R})$ with mean and cov fun $\mu(x)$, $K(x,y)$ at points $x_1, x_2, \ldots, x_n \in \mathbb{R}$.

Let $\vec{Z} = \begin{bmatrix} Z(x_1) \\ \vdots \\ Z(x_n) \end{bmatrix}$,

$$\vec{\mu} = \begin{bmatrix} \mu(x_1) \\ \vdots \\ \mu(x_n) \end{bmatrix},$$

$$\Sigma = \Big[K(x_i, x_j) \Big]_{i,j=1}^{n}$$

since the fdd of Z are Gaussian

$$\therefore \quad \vec{Z} \sim N(\underbrace{\vec{\mu}, \Sigma}_{})$$

since $\mu(\cdot)$ & $K(\cdot,\cdot)$ are the mean and cov functions of Z.

Indeed one can simulate $\vec{Z}$ by

$$\vec{Z} \overset{\mathcal{D}}{=} L \begin{pmatrix} W_1 \\ \vdots \\ W_n \end{pmatrix} + \vec{\mu}$$

where $W_1, \ldots, W_n \overset{iid}{\sim} N(0,1)$

and where L is the lower triangular cholsky decomposition of Σ, i.e.

$$\Sigma = LL^T.$$

Alternatively use a eigen decomposition of Σ as follows:

$$\Sigma = \mathcal{U} \Lambda \mathcal{U}^T$$

orthornormal Matrix

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

with $\lambda_i \geq 0$ since Σ is positive definite

With this decomposition we have

iid $N(0,1)$

$$\vec{Z} \overset{\mathcal{D}}{=} \mathcal{U} \Lambda^{1/2} \mathcal{U}^T \begin{bmatrix} W_1 \\ \vdots \\ W_n \end{bmatrix} + \vec{\mu}$$

$$\overset{\mathcal{D}}{=} \mathcal{U} \Lambda^{1/2} \begin{bmatrix} W_1 \\ \vdots \\ W_n \end{bmatrix} + \vec{\mu}$$

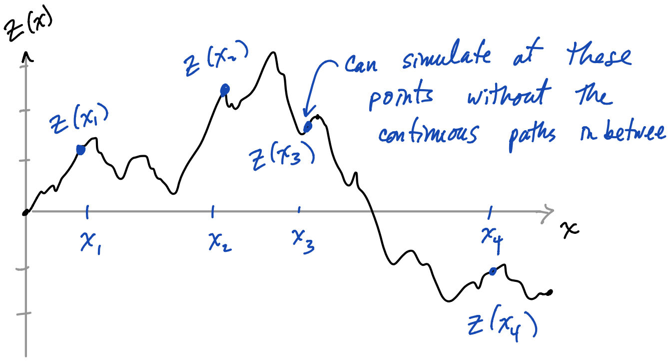
since $\mathcal{U}$ is orthonormal

$$= \sum_{i=1}^{n} u_i \underbrace{\lambda_i^{1/2} W_i}_{} + \vec{\mu}$$

columns of $\mathcal{U}$. i.e. the eigenvectors of Σ

independent, Gaussian, mean zero random variables with variance given by the eigenvalues of Σ.

So, even thought a realization of
$$(Z(x) : x \in \mathbb{R})$$
is defined on all of $\mathbb{R}$, we can simulate just the values at a discrete set of locations $x_1, \ldots, x_n \in \mathbb{R}$



We can also predict $Z(y)$ from data
$$Z(x_1), \ldots, Z(x_n)$$
where $y \notin \{x_1, \ldots, x_n\}$

since

$$\begin{bmatrix} Z(y) \\ Z(x_1) \\ \vdots \\ Z(x_n) \end{bmatrix} \sim N\left(\begin{bmatrix} \mu(y) \\ \vdots \\ \vec{\mu} \\ \vdots \end{bmatrix}, \underbrace{\begin{bmatrix} K(y,y) & K(y,x_1) & \cdots & K(y,x_n) \\ K(x_1,y) & & & \\ \vdots & & \Sigma & \\ K(x_n,y) & & & \end{bmatrix}}_{\text{write this as } \begin{bmatrix} \Sigma_{00} & \Sigma_{01} \\ \Sigma_{10} & \Sigma_{11} \end{bmatrix}} \right)$$

with $\Sigma_{00} = K(y,y)$, $\Sigma_{11} = \Sigma$

by conditional Gaussian distributions we have

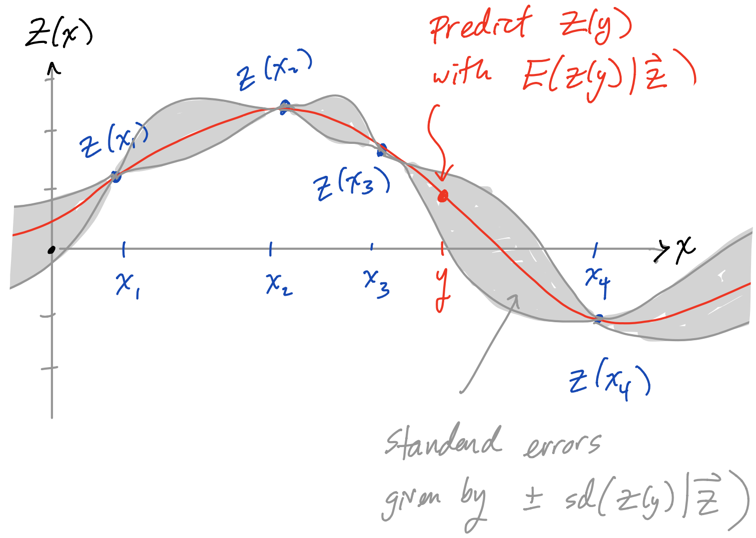
$$Z(y) \mid \vec{Z} \sim N\left(E(Z(y) \mid \vec{Z}), \operatorname{var}(Z(y) \mid \vec{Z}) \right)$$

$\vec{Z} = (Z(x_1), \ldots, Z(x_n))^T$

where

$$E(Z(y) \mid \vec{Z}) = \Sigma_{01} \Sigma_{11}^{-1} (\vec{Z} - \vec{\mu}) + \mu(y)$$

$$\operatorname{var}(Z(y) \mid \vec{Z}) = \Sigma_{00} - \Sigma_{01} \Sigma_{11}^{-1} \Sigma_{10}$$



Finally consider the case that $K(x,y)$ and $\mu(x)$ depend on unknown parameters such as $K_\theta(x,y)$, $\mu_\theta(x)$.

Given observations
$$\vec{Z} = (Z(x_1), \ldots, Z(x_n))^T$$
one can infer θ based on the log likelihood since $\vec{Z}$ is Gaussian:

$$\vec{Z} \sim N(\vec{\mu}_\theta, \Sigma_\theta)$$

$(\mu_\theta(x_i))_{i=1}^n$ $\qquad$ $(K_\theta(x_i, x_j))_{i,j=1}^n$

$\therefore \log \Pr(\vec{z}|\theta)$

$$= -\frac{1}{2}(\vec{z}-\vec{\mu}_\theta)^T \Sigma_\theta^{-1}(\vec{z}-\vec{\mu}_\theta) - \frac{1}{2}\log|\Sigma_\theta| + \text{const}$$

$$= -\frac{1}{2}\| L_\theta^{-1}(\vec{z}-\vec{\mu}_\theta)\|^2 - \sum_{i=1}^{n} \log L_\theta[i,i] + \text{const}$$

where L_θ is the lower cholesky of Σ_θ

$$= -\frac{1}{2}\| \Lambda_\theta^{-1/2} U_\theta^T(\vec{z}-\vec{\mu}_\theta)\|^2 - \frac{1}{2}\sum_{i=1}^{n} \log \Lambda_\theta[i,i] + \text{const}$$

with SVD $\Sigma_\theta = U_\theta \Lambda_\theta U_\theta^T$

diag.

the argument domain

It is important to realize that the domain $\mathcal{T}$ is an integral part of checking if $K: \mathcal{T}\times\mathcal{T} \to \mathbb{R}$ is positive define.

To illustrate suppose $\mathcal{X} \subset \mathcal{T}$ and $K: \mathcal{T}\times\mathcal{T} \to \mathbb{R}$.

Then clearly

K is pos. def. on $\mathcal{T}\times\mathcal{T}$

$\Longrightarrow K$ is pos. def. on $\mathcal{X}\times\mathcal{X}$

However in general

K is pos. def. on $\mathcal{X}\times\mathcal{X}$

$\not\Longrightarrow K$ is pos. def. on $\mathcal{T}\times\mathcal{T}$

Example: It can be shown that

$$K(x,y) = (1-\|x-y\|)^+$$

is positive definite on $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$ only for $d=1$ ...

i.e. there exists no RF on $\mathbb{R}^d$, $d>1$, with covariance function given by this K.

Even the existance of a positive definite extension to a larger space ... not much can be said.

Example:

sphere $\subseteq \mathbb{R}^3$

Suppose $K(x,y): S^2\times S^2 \to \mathbb{R}$ is positive define.

Not much is known about how to find a positive definite $\tilde{K}(x,y): \mathbb{R}^3\times\mathbb{R}^3 \to \mathbb{R}$ s.t.

$\tilde{K}(x,y) = K(x,y)$ on $x,y \in S^2$

New pos. def. fns from old

Now we introduce some simple properties of positive definite functions which we will use to construct parametric covariance models

Claim:

Suppose $K(x,y)$ is positive definite on $\mathcal{T}\times\mathcal{T}$

(i) $K(x,x)\geq 0$

(ii) $|K(x,y)| \leq \sqrt{K(x,x)}\sqrt{K(y,y)}$

(iii) If $C(x,y)$ is also positive definite on $\mathcal{T}$ and $a_1, a_2 \in \mathbb{R}^+$ then

$$a_1 C(x,y) + a_2 K(x,y)$$

and

$$C(x,y)K(x,y)$$

are both positive definite on $\mathcal{T}$.

(iv) If $\lim_{n\to\infty} \underbrace{C_n}_{\substack{\text{all pos.}\\ \text{def. on } \mathcal{T}}}(x,y) = C(x,y) < \infty$ for all $x,y\in\mathcal{T}$

then $C(x,y)$ is pos. def. on $\mathcal{T}$

Note:

Conditions (iii) and (iv) allow us to build new covariance models by limits of variance and range re-scaling

In particular let $\sigma^2_{n,i}, \alpha_{n,i} > 0$ and suppose K is positive definite on $\mathbb{R}^d$.

Now if

$$\sum_{i=1}^{n} \underbrace{\sigma^2_{n,i} K(\alpha_{n,i}x, \alpha_{n,i}y)}_{\substack{\text{changes of spatial scale}\\ \text{and variance ... these}\\ \text{are positive definite on } \mathbb{R}^d}} \xrightarrow{\text{pointwise}} C(x,y) < \infty$$

Then $C(x,y)$ is positive definite on $\mathbb{R}^d$

Notice that well behaved integrals w.r.t. densities are also limits of positive linear combinations.

So in general

$$C(x,y) = \underbrace{\int_0^\infty K\left(\frac{x}{\tau}, \frac{y}{\tau}\right)\mu(\tau)\,d\tau}_{\substack{\text{positive mixture}\\ \text{of } K\left(\frac{x}{\tau}, \frac{y}{\tau}\right)}}$$

is positive definite on $\mathbb{R}^d$ when $\mu(\tau)\geq 0$, $\int_0^\infty \mu(\tau)\,d\tau < \infty$ and $K(x,y)$ is positive definite on $\mathbb{R}^d$.

Stationary & isotropic cov. funs

Suppose $Z(x)$ is a random field on $\mathbb{R}^d$ with covariance function $K(x,y)$ and mean function $m(x)$

Z is said to be stationary if its f.d.d.s are shift invariant and isotropic if additionally its f.d.d.s are rotation invariant,

i.e. if $\forall h \in \mathbb{R}^d$ & $\forall U \in SO(d)$

(i) $\big(Z(x+h) : x \in \mathbb{R}^d\big) \overset{f.d.d.}{=} \big(Z(x) : x \in \mathbb{R}^d\big)$

(ii) $\big(Z(Ux) : x \in \mathbb{R}^d\big) \overset{f.d.d.}{=} \big(Z(x) : x \in \mathbb{R}^d\big)$

Definition: (i) $\Leftrightarrow$ Z is stationary
(i) & (ii) $\Leftrightarrow$ Z is isotropic

Note that

Z is stationary

$$\Rightarrow E\big(Z(x+h)\big) = E\big(Z(x)\big)$$
$$\mathrm{cov}\big(Z(x+h), Z(y+h)\big) = \mathrm{cov}\big(Z(x), Z(y)\big), \quad \forall h \in \mathbb{R}^d$$

$$\Rightarrow m(x+h) = m(x)$$
$$K(x+h, y+h) = K(x,y), \quad \forall h \in \mathbb{R}^d$$

$$\Rightarrow m(x) = m(0) \leftarrow \text{mean is constant}$$
$$\underbrace{K(x,y) = K(x-y, 0)}_{K(x,y) \text{ only depends on } x-y}$$

Similarly

Z is isotropic

$$\Rightarrow m(x) = m(0) \leftarrow \text{again, constant}$$
$$K(x,y) = \underbrace{K(\|x-y\|, 0)}_{\text{only depends on the distance between } x \text{ \& } y.}$$

Definition: Let $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be positive definite on $\mathbb{R}^d \times \mathbb{R}^d$

K is stationary

$\Updownarrow$

$K(x,y)$ is a function of $(x-y)$ for all $x, y \in \mathbb{R}^d$.

K is isotropic

$\Updownarrow$

$K(x,y)$ is a function of $\|x-y\|$. for all $x, y \in \mathbb{R}^d$.

Notice that if K is a stationary covariance function then $Z \sim GRF_{\mathbb{R}^d}(0, K)$ is a stationary random field (similarly for isotropic case).

However Z can be non-stationary but with stationary covariance function (sometimes called weakly stationary).

4 Fundamental covariance functions

... which are positive definite on $\mathbb{R}^d \times \mathbb{R}^d$ for any $d = 1, 2, 3, \ldots$

Let $x, y \in \mathbb{R}^d$ and $\|x\|$ denote Euclidean norm.

Linear or inner product cov. fun.

$$K_{Ln}(x,y) = \langle x, y \rangle = \sum_{i=1}^{d} x_i y_i$$

Brownian or Wiener cov. fun.

$$K_{Br}(x,y) = \frac{1}{2}\left(\|x\| + \|y\| - \|x-y\|\right)$$

Exponential cov. fun.

$$K_{Ex}(x,y) = \exp(-\|x-y\|)$$

Gaussian cov. fun.

$$K_{Ga}(x,y) = \exp(-\|x-y\|^2)$$

K_{Ex} & K_{Ga} are isotropic.

K_{Br} is not isotropic which is clear from

$$Z_{Br} \sim GRF_{\mathbb{R}^d}(0, K_{Br})$$

$$\Longrightarrow \mathrm{var}(Z_{Br}(x)) = \|x\|$$

$$\not\Rightarrow Z_{Br}(x+h) \overset{\mathcal{D}}{=} Z_{Br}(x) \text{ for } \forall h$$

K_{Ln} simply corresponds to a random hyperplane ...

If $W_i \overset{iid}{\sim} N(0,1)$ then

$$Z_{Ln}(x) = \underbrace{x_1 W_1 + \cdots + x_d W_d}_{\text{linear maps with random indep coeffs}}$$

Since

$$\mathrm{cov}(Z_{Ln}(x), Z_{Ln}(y)) = \sum_{i,j=1}^{d} x_i y_j \underbrace{E(W_i W_j)}_{\delta_{ij}}$$

$$= \langle x, y \rangle$$

$$= K_{Ln}(x,y)$$

The fact that K_{Ln} is an inner product shouldn't be a surprise ...

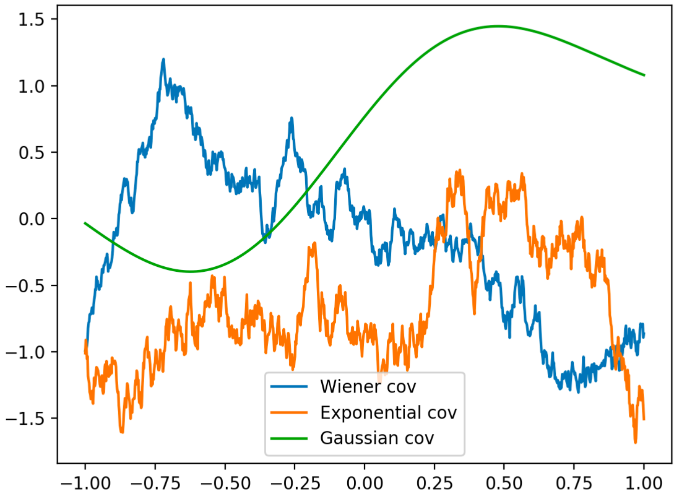
The positive definite requirement for covariance functions effectively says that covariance <u>are</u> inner products on a higher dimensional space.

Simulation illustration

$$Z_{Ga} \sim GRF_{\mathbb{R}}(0, K_{Ga})$$

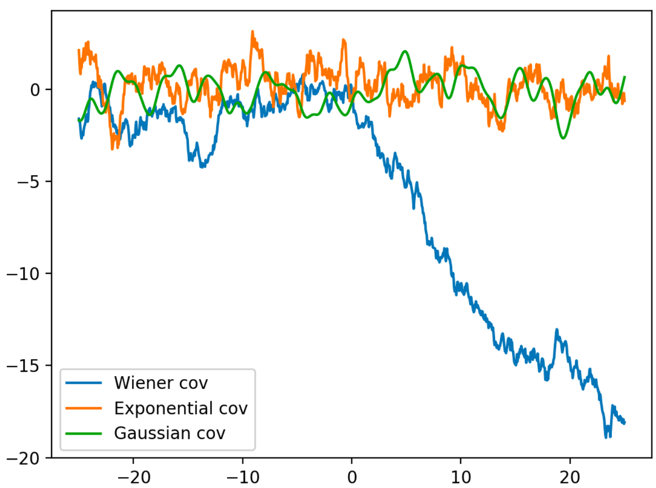
$$Z_{Br} \sim GRF_{\mathbb{R}}(0, K_{Br})$$

$$Z_{Ex} \sim GRF_{\mathbb{R}}(0, K_{Ex})$$



$x \in [-1,1]$

Z_{Ga} is very smooth | Z_{Ex} is not and looks similar to Z_{Br} in smoothness | Notice how $Z_{Br}(x)$ has non-stationary fluctuations



$x \in [-25,25]$

Auto Covariance functions

Notation:

$K : \mathbb{R}^d \to \mathbb{R}$ is an auto-covariance function on $\mathbb{R}^d$

$\Updownarrow$

$K(x-y)$ is a covariance function on $\mathbb{R}^d$.

$K : \mathbb{R}^+ \to \mathbb{R}$ is an isotropic auto-covariance function on $\mathbb{R}^d$

$\Updownarrow$

$K(\|x-y\|)$ is a covariance function on $\mathbb{R}^d$.

The first thing to notice about auto-covariance functions is that the behavior of $K(x)$ at $x=\vec{0}$ is what determines the smoothness of the corresponding

$$Z \sim GRF_{\mathbb{R}^d}(0, K).$$

Here is an illustration. Fix $x, v \in \mathbb{R}^d$ and let $\varepsilon \in \mathbb{R}$ vary in a neighborhood of 0. Two cases:

Z is smooth

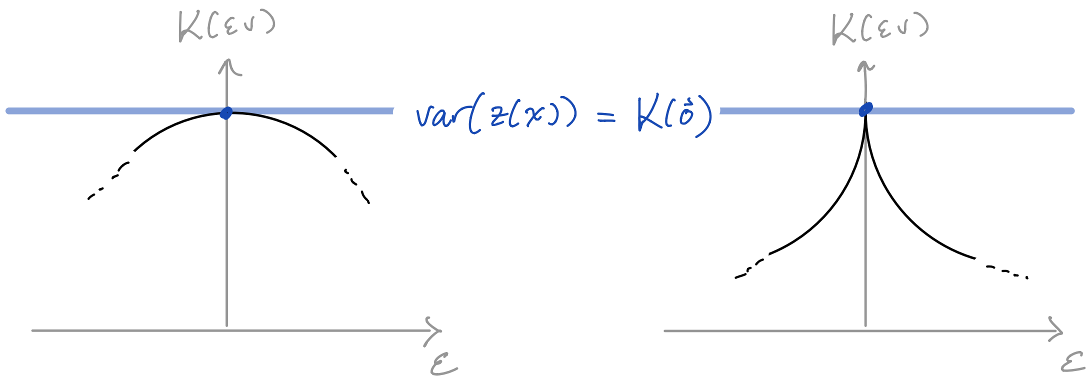
$Z(x+\varepsilon v)$ has small variability as $\varepsilon \to 0$

$\mathrm{Cov}(Z(x+\varepsilon v), Z(x))$ is large relative to $\mathrm{var}(Z(x))$ as $\varepsilon \to 0$

Z is not smooth

$Z(x+\varepsilon v)$ has large variability as $\varepsilon \to 0$

$\mathrm{Cov}(Z(x+\varepsilon v), Z(x))$ is small relative to $\mathrm{var}(Z(x))$ as $\varepsilon \to 0$

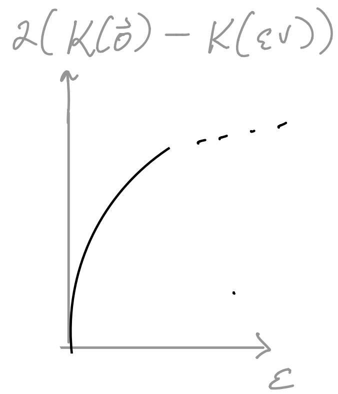


Z is smooth

$\mathrm{var}(Z(x+\varepsilon v) - Z(x)) \to 0$ fast as $\varepsilon \to 0$

Z is not smooth

$\mathrm{var}(Z(x+\varepsilon v) - Z(x)) \to 0$ slowly as $\varepsilon \to 0$



Notation: $\gamma(x,y) = \mathrm{var}(Z(x) - Z(y))$ is called the variogram

when Z is stationary $\gamma(x+\varepsilon v, x) = 2(K(\vec{0}) - K(\varepsilon v))$.

In a bit more detail a back-of-the-envelope calculations suggests:

$v \cdot \nabla Z(x)$ exists (in mean square) if

$$\infty > \operatorname{var}\left(\lim_{\varepsilon \to 0} \frac{Z(x+\varepsilon v) - Z(x)}{\varepsilon}\right)$$

$$= \lim_{\varepsilon \to 0} \operatorname{var}\left(\frac{Z(x+\varepsilon v) - Z(x)}{\varepsilon}\right)$$

$$= \lim_{\varepsilon \to 0} \frac{2\left(K(0) - K(\varepsilon v)\right)}{\varepsilon^2}$$

so that

$|K(0) - K(\varepsilon v)| \lesssim \varepsilon^2$ as $\varepsilon \to 0$ $\Longrightarrow$ Z is mean square differentiable in direction v.

Same loose arguments for α-Hölder smoothness (with $\alpha \in (0,1)$).

$|K(0) - K(\varepsilon v)| \leq c|\varepsilon|^\alpha$
for suff small ε

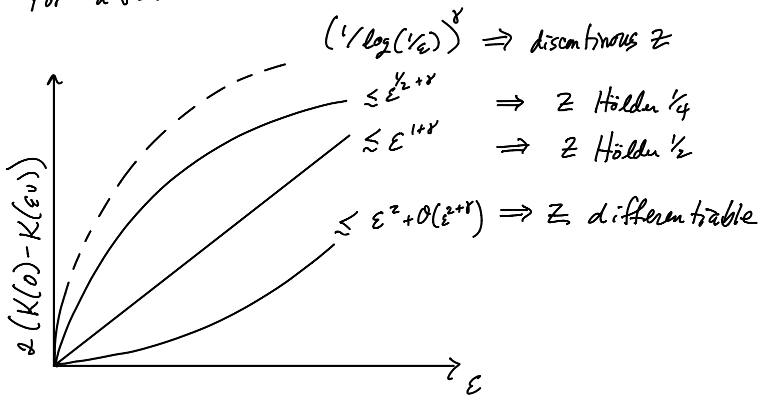
$\Rightarrow$ $\operatorname{var}\left(Z(x+\varepsilon v) - Z(x)\right) \leq 2c|\varepsilon|^\alpha$
suff small ε

$\Rightarrow$ $E\left(|Z(x+\varepsilon v) - Z(x)|^2\right) \leq 2c|\varepsilon|^\alpha$
suff small ε ... when $E(Z(x))$ is 0

$\overset{\text{not rig}}{\Longrightarrow}$ $|Z(x+\varepsilon v) - Z(x)| \leq 2c|\varepsilon|^{\alpha/2}$
suff small ε

These arguments can be made rigourous.

For a $\gamma > 0$



$(v \cdot \nabla)^n Z(x)$ exists in mean square $\iff$ $|(v\cdot\nabla)^{2n} K(0)| < \infty$

(Needs a refence)

Keep in mind that mean square smoothness & diffentiability and sample path properties are different.

You should loosly have in your mind that for a random field Z

(*) $\left\{\begin{array}{l}\text{sample path} \\ \text{smoothness of } Z\end{array}\right.$ $\not\Rightarrow$ mean square smoothness of Z

(**) $\left\{\begin{array}{c}\text{mean square} \\ \text{smoothness of } Z \\ + \\ \text{some extra} \\ \text{conditions}\end{array}\right.$ $\Longrightarrow$ sample path smoothness of Z

Here is an example (*):

For $t \in \mathbb{R}$ let

$$Z(t) = \cos(Xt + Y)$$

where

$$\left.\begin{array}{l} X \sim \text{Cauchy}\left(\text{w/density } \frac{1}{\pi(1+x^2)}\right) \\ Y \sim \text{Unif}([0, 2\pi]) \end{array}\right\} \text{indep.}$$

Now

$$\operatorname{cov}(z(t), z(s)) = E\left[\cos(Xt+Y)\cos(Xs+Y)\right]$$

since $2\cos\theta\cos\varphi = \cos(\theta-\varphi) + \cos(\theta+\varphi)$

$$= \frac{1}{2} E\left[\cos(X(t-s))\right] + \frac{1}{2} E\left[\cos(X(t+s)+2Y)\right]$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} \frac{\cos(x(t-s))}{\pi(1+x^2)}\,dx + \frac{1}{2}\int_{-\infty}^{\infty}\int_0^{2\pi} \frac{\cos(x(t+s)+2y)}{\pi(1+x^2)\,2\pi}\,dy\,dx$$

$$= \frac{1}{2}\exp(-|t-s|)$$

So $Z(t)$ has sample path derivatives of <u>all</u> order but is not even mean square differentiable ... hence an illustration of (*).

Covariance smoothness away from the origin

We have seen that smoothness of an autocovariance $K(x)$ at $x = 0 \in \mathbb{R}^d$ charaterizes the mean square smoothness of $Z \sim GRF_{\mathbb{R}^d}(0, K)$.

Intuitively it is therefore natural that smoothess of $K(x)$ at $x=0$ imposes restrictions on the smoothness of $K(x)$ at $x \neq 0$.

Example:

Let $Z \sim GRF_{\mathbb{R}}(0, K)$ with $|K''(0)| < \infty$ then $Z'(x)$ exists in mean sq and

$$\operatorname{cov}(Z'(x), Z'(0)) = \frac{d}{ds}\frac{d}{dt} K(s-t)\Big|_{s=x,\, t=0}$$

$$= -K''(x)$$

So in this case

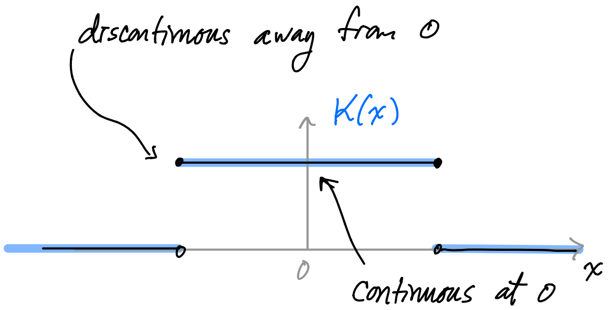
$$K \text{ has 2 finite derivatives at } x=0$$
$$\Downarrow$$
$$K \text{ has 2 finite derivatives at all } x$$

<u>Vague claim</u>: Let K be an auto-cov on $\mathbb{R}^d$.

$$\begin{array}{c}\text{Smoothness of} \\ K(x) \text{ at } x \neq 0\end{array} \geq \begin{array}{c}\text{Smoothness of} \\ K(x) \text{ at } x = 0\end{array}$$

Example:

This $K(x): \mathbb{R} \to \mathbb{R}$ is not an auto-covariance



In general you want to use auto-covariances s.t.

$$\text{smoothness of } K(x) \text{ for } x \neq 0 \gg \text{smoothness of } K(x)$$

Example:

good

$$K_{Ex}(x) = \exp(-\|x\|)$$

- at $x=0$ it is continuous but not differentiable
- at $x \neq 0$ this is infinitely differentiable

trouble

$$K_{tri}(x) = (1 - \|x\|)^+$$

- at $x=0$ it has the same smoothness as $\exp(-\|x\|)$ but has a non-differentiability at $\|x\| = 1$

Here is an illustration of the weirdness one gets when using K_{tri} for prediction.

Let

$$Z_{Ex} \sim GRF_{\mathbb{R}}(0, K_{Ex})$$
$$Z_{tri} \sim GRF_{\mathbb{R}}(0, K_{tri})$$

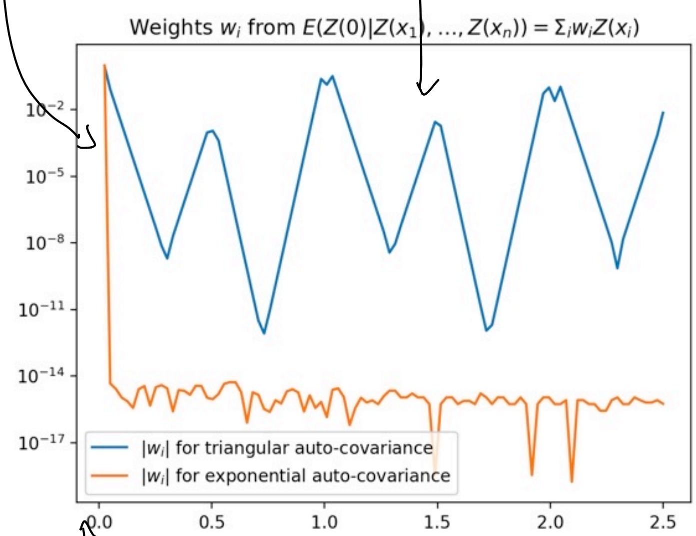
For both compute

$$E(Z(0) \mid Z(x_1), \dots, Z(x_n)) = \sum_{i=1}^{n} w_i Z(x_i)$$

and plot the weights $|w_i|$.

The weights for K_{Ex} are natural

The weights have huge spikes. for K_{tri}



These are the x_i's

The Matérn auto covariance

Recall that one can build new covariance functions by positive mixtures of re-scalings of a fixed covariance function.

A popular class of auto covariance functions developed this way is called the Matérn class. It is based on mixtures of the Gaussian auto-cov via an identity of the form

$$\|x-y\|^{\nu} \mathcal{K}_\nu(\|x-y\|) = \int_0^\infty \underbrace{\exp(-\|x-y\|^2/\tau)}_{K_{Ga}(\frac{x}{\tau}, \frac{y}{\tau})} \underset{\substack{\uparrow \\ \text{positive density}}}{\mu(\tau)}\, d\tau$$

Where $\mathcal{K}_\nu$ is the modified Bessel function of the second kind of order $\nu > 0$.

By another re-scaling to stabilize the width & height one can define the auto-covariance function

$$\mathcal{M}_\nu(\underset{\substack{\uparrow \\ t \in \mathbb{R} \text{ st. } t>0}}{t}) = \frac{2^{1-\nu}}{\Gamma(\nu)} (\sqrt{2\nu}\, t)^{\nu} \mathcal{K}_\nu(\sqrt{2\nu}\, t).$$

The parameter $\nu > 0$ controls the smoothness of $Z \sim GRF_{\mathbb{R}^d}(0, \mathcal{M}_\nu)$.

Note: $\mathcal{K}_\nu(t) \to \infty$ as $t \to 0$

but $t^\nu \mathcal{K}_\nu(t) \to \frac{\Gamma(\nu)}{2^{1-\nu}}$ as $t \to 0$

so $\mathcal{M}_\nu(0) := 1$, i.e. $\mathcal{M}_\nu$ is an auto-correlation function.

Finally we add an overall variance parameter $\sigma^2 > 0$ and a correlation range parameter $\rho > 0$.

Matérn class of covariance functions

$$\mathcal{M}_{\nu,\rho,\sigma^2}(\|x-y\|) := \sigma^2 \mathcal{M}_\nu\left(\frac{\|x-y\|}{\rho}\right)$$

Parameters: $\nu > 0$, $\rho > 0$, $\sigma^2 > 0$.

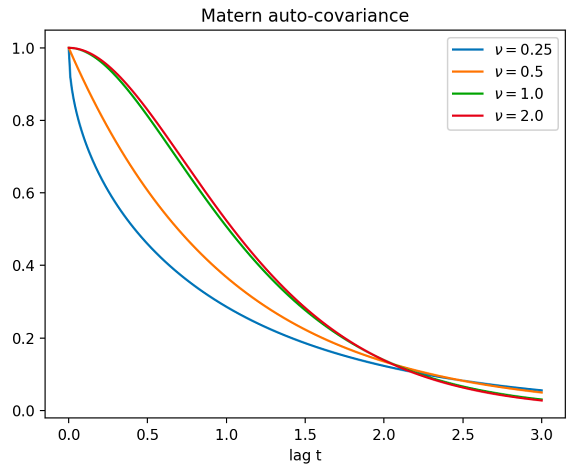
positive definite over $x, y \in \mathbb{R}^d$ for any $d \geq 1$.

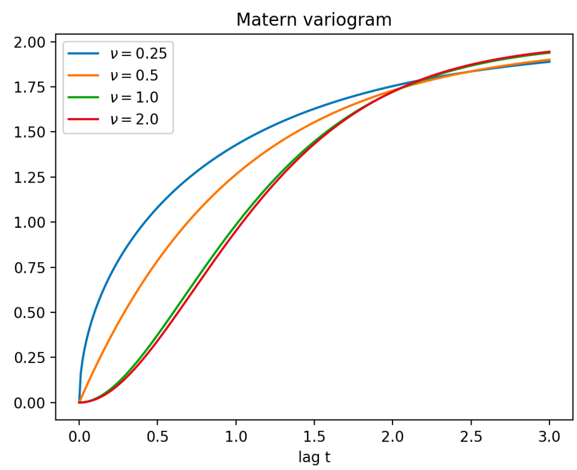
Note: If $Y \sim GRF_{\mathbb{R}^d}(0, \mathcal{M}_{\nu,\rho,\sigma^2})$ and $Z_\nu \sim GRF_{\mathbb{R}^d}(0, \mathcal{M}_\nu)$ then

$$Y(x) \overset{f.d.d}{=} \underbrace{\sigma Z_\nu\left(\frac{x}{\rho}\right)}_{\text{change of units in both argument and range}}$$

so we can focus on understanding Z_ν.

Plots of $\mathcal{M}_\nu(t)$ and $2(\mathcal{M}_\nu(0) - \mathcal{M}_\nu(t))$ for $\nu < 1$

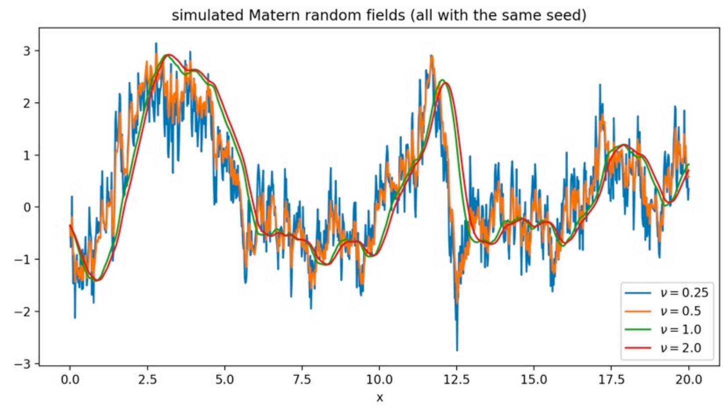




- Note that for $\nu = 1, 2$ the behavior near $t=0$ looks close to quadratic & doesn't visably change as much as with $\nu = 0.25, 0.5$

- Also note that for $\nu = 1, 2$ the inflection points are close to $t = 1$.

Corresponding simulations for $d=1$:



- Note that higher order differentiability isn't visible to the eye without taking increments ... This is analogous to higher order terms beyond quadratic in the auto cov for $\nu > 1$:

$$\mathcal{M}_\nu(t) \approx c_0 + c_1 t^2 + \underbrace{o(t^2)}$$

as $t \to 0$

This term drives the post-derivative smoothness

Some properties of $\mathcal{M}_\nu$

Many of the properties of $\mathcal{M}_\nu$ can be found in

- Gradshteyn & Ryzhik [GR]
- Abromowitz & Stegun [AS]
- Arfken & Weber 6th ed [AW]
- Rasmussen & Williams [RW]

Here are some examples:

(I) In [AW, p.717] they give the following recursive relations:

$$K_\nu(t) = -\frac{x}{2\nu}\left(K_{\nu-1}(t) - K_{\nu+1}(t)\right)$$

$$\frac{d}{dt} K_\nu(t) = -\frac{1}{2}\left(K_{\nu-1}(t) + K_{\nu+1}(t)\right)$$

$$K_{-\nu}(t) = K_\nu(t)$$

(II) when $n = 0, 1, 2, \ldots$

$$\mathcal{M}_{n+1/2}(t) = \exp(-\sqrt{2\nu}|t|)\cdot(\text{polynomial in } |t| \text{ of order } n)$$

More explicitly [RW, p.85] gives

$$\mathcal{M}_{n+1/2}(t) = \exp(-\sqrt{2\nu}|t|)\left[\frac{\Gamma(n+1)}{\Gamma(2n+1)}\sum_{k=0}^{n}\frac{(n+k)!}{k!(n-k)!}\left(\frac{\sqrt{8\nu}|t|}{\rho}\right)^{n-k}\right]$$

Special cases:

$$\mathcal{M}_{1/2}(t) = \exp(-|t|).$$

$$\mathcal{M}_{3/2}(t) = \left(1 + \frac{\sqrt{3}|t|}{\rho}\right)\exp\left(-\frac{\sqrt{3}|t|}{\rho}\right)$$

$$\mathcal{M}_{5/2}(t) = \left(1 + \frac{\sqrt{5}|t|}{\rho} + \frac{5|t|^2}{3\rho^2}\right)\exp\left(-\frac{\sqrt{5}|t|}{\rho}\right)$$

(III) $$\mathcal{M}_\nu(t) \xrightarrow{\nu\to\infty} \exp\left(-\frac{t^2}{2}\right)$$

Gaussian auto-cov

(IV) If $Z \sim GRF_{\mathbb{R}}(0, \mathcal{M}_\nu)$ then

Z is k-times mean square differentiable

$$\Updownarrow$$

$$\nu > k$$

(V) For $0 < \nu < 1$

$$\underbrace{2\left(\mathcal{M}_\nu(0) - \mathcal{M}_\nu(t)\right)}_{\text{variogram}} \sim \underbrace{(\text{constant})}_{\text{positive}}\, t^{2\nu}$$

as $t \to 0$

(VI) $$\mathcal{M}_\nu(t) = \underbrace{\sum_{k=0}^{\lfloor\nu\rfloor} b_k t^{2k}}_{\text{order } 2\lfloor\nu\rfloor} + c_\nu \mathcal{G}_\nu(t) + \underbrace{\mathcal{O}(t^{2\lfloor\nu\rfloor+2})}_{\text{order } 2\lfloor\nu\rfloor+2}$$

as $t \to 0$ where $c_\nu > 0$ and

$$\mathcal{G}_\nu(t) := (-1)^{\lfloor\nu\rfloor+1} \times \begin{cases} t^{2\nu}\log t & \text{if } \nu = 1, 2, \ldots \\ t^{2\nu} & \text{otherwise} \end{cases}$$

(VII) $$t^\nu K_\nu(t) \sim \sqrt{\frac{\pi}{2}}\, t^{\nu - 1/2} e^{-t} \quad \text{as } t \to \infty$$

[AW, p.721]

(VIII) $$K_\nu(t) = \int_0^\infty \exp(-t\cosh(k))\cosh(\nu k)\,dk$$

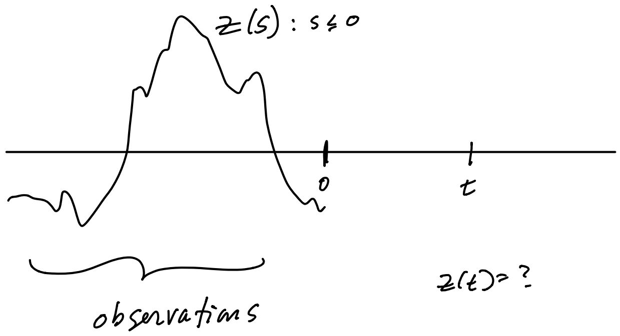
[AW, p.718]

Projection application to Gaussian random field prediction

Let $\{z(t) : t \in \mathbb{R}\}$ be a Gaussian random field (GRF) s.t. $\forall\, t, s \in \mathbb{R}$

$$\left.\begin{aligned} &E(z(t)) = 0 \\ &\mathrm{cov}(z(t), z(s)) = e^{-|t-s|} \end{aligned}\right\} \text{specifies the f.d.d.s.}$$

Suppose we observe $z(s)$, $\forall s \le 0$ & want to predict $z(t)$ for some $t > 0$.



Using Hilbert space theory it is possible to show that

$$E(z(t) \mid z(s), s \le 0) = \mathcal{P}_0\, z(t)$$

where $\mathcal{P}_0$ is the projection operator onto the closed linear span of

$$\mathcal{H}_0 = \overline{\mathrm{span}}\{z(s) : s \ge 0\}$$

with inner product given by

$$\langle X, Y \rangle := \mathrm{cov}(X, Y)$$

To find $\mathcal{P}_0\, z(t)$ lets guess that $\mathcal{P}_0\, z(t) = a_t\, z(0)$ and prove the residuals are orthognal to $\mathcal{H}_0$.

Note that by linearity & continuity of $\langle \cdot, \cdot \rangle$ it is sufficient to show

$$\langle z(t) - a_t z(0), z(s) \rangle = 0 \quad \forall s \le 0.$$

For $t > 0$ & $s \le 0$ we have

$$\begin{aligned} &\langle z(t) - a_t z(0), z(s) \rangle \\ &= e^{-|t-s|} - a_t e^{-|s|} \\ &= e^{-(t-s)} - a_t e^{s} \quad \text{since } s \le 0 < t \\ &= 0 \quad \text{iff } a_t := e^{-t} \end{aligned}$$

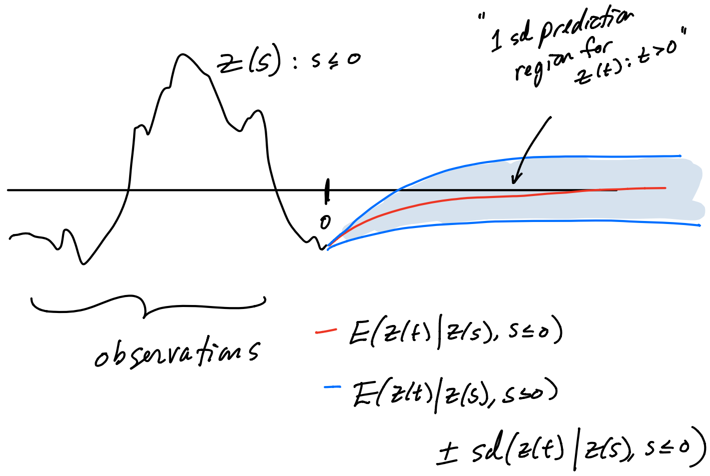
$$\therefore\; E(z(t) \mid z(s), s \le 0) = e^{-t} z(0)$$

and

$$\begin{aligned} &\mathrm{var}(z(t) \mid z(s), s \le 0) \\ &= E\Big(\big[\underbrace{z(t) - e^{-t} z(0)}\big]^2 \,\Big|\, \underbrace{z(s), s \le 0}\Big) \end{aligned}$$

Since these are orthognal in L_2, they are uncorrelated. Since they are Gaussian they are then independent

$$\begin{aligned} &= \mathrm{var}(\underbrace{z(t) - e^{t} z(0)}_{\text{marginal residual}}) \\ &= e^{0} - 2e^{-t} e^{-|t|} + e^{2t} e^{0} \\ &= 1 - e^{-2t} \end{aligned}$$



Variograms

If $K(x,y) = \mathrm{cov}(Z(x), Z(y))$ then the variogram is given by

$$\gamma(x,y) = \mathrm{var}(Z(x) - Z(y))$$
$$= K(x,x) + K(y,y) - 2K(x,y)$$

Notice that to write $K(x,y)$ in terms of $\gamma(x,y)$ one needs to know $\mathrm{var}(Z(x))$ and $\mathrm{var}(Z(y))$.

$$K(x,y) = \frac{1}{2}\Big(K(x,x) + K(y,y) - \gamma(x,y)\Big).$$

One can think of $\gamma(x,y)$ as a lossy compression of $K(x,y)$ that models a "Projection of $Z(x)$ that removes an overall random constant".

In particular define

$$\tilde{Z}(x) = Z(x) - Z(a_0)$$

some arbitrary $a_0 \in \mathcal{X}$.

with covariance $\tilde{K}(x,y) = \mathrm{cov}(\tilde{Z}(x), \tilde{Z}(y))$

Now

$$\left.\begin{aligned}\gamma(x,y) &= \mathrm{var}(Z(x) - Z(y)) \\ &= \mathrm{var}(\tilde{Z}(x) - \tilde{Z}(y))\end{aligned}\right\} \text{$\tilde{Z}$ \& Z have the same variogram}$$

$$= \mathrm{var}(\tilde{Z}(x)) + \mathrm{var}(\tilde{Z}(y)) - 2\tilde{K}(x,y)$$

$$= \mathrm{var}(Z(x) - Z(a_0)) + \mathrm{var}(Z(y) - Z(a_0)) - 2\tilde{K}(x,y)$$

$$= \gamma(x,a_0) + \gamma(y,a_0) - 2\tilde{K}(x,y)$$

which means

$$\tilde{K}(x,y) = \frac{1}{2}\Big(\gamma(x,a_0) + \gamma(y,a_0) - \gamma(x,y)\Big)$$

$Z, \tilde{Z}$ have the same variogram γ but only $\tilde{K}$ can be recovered by γ.

The variogram is sometimes easier to model ...

- Being a lossy compression there is less to model
- A general heuristic is that a cov function is analogous to a inner product and a variogram is analogous to a squared distance... intuition being that are easier to think about than inner products.

Example:

For $K_{Ln}(x,y) = \langle x,y \rangle$

$$\mathrm{cov}(Z_{Ln}(x), Z_{Ln}(y)) = \langle x,y \rangle$$

$$\mathrm{var}(Z_{Ln}(x) - Z_{Ln}(y)) = \|x-y\|^2$$

$$\langle x,x \rangle + \langle y,y \rangle - 2\langle x,y \rangle$$

Z_{Ln} is non-stationary but $\mathrm{var}(Z_{Ln}(x) - Z_{Ln}(y))$ only depends on $\|x-y\|^2$. In this example $K_{Ln}(x,y) = 0$ when x or y is 0... so it already has a overall constant removed and $\gamma(x,y)$ is sufficient to recover K_{Ln} as we see:

$$K_{Ln}(x,y) = \langle x,y \rangle$$

$$= \frac{1}{2}\Big(\underbrace{\|x\|^2}_{\gamma(x,0)} + \underbrace{\|y\|^2}_{\gamma(y,0)} - \underbrace{\|x-y\|^2}_{\gamma(x,y)}\Big)$$

Here is a fundamental result by Schoenberg which characterizes variograms and shows how to construct covariance functions from them.

Theorem: Let $\gamma(x,y): \mathcal{T}\times\mathcal{T} \to \mathbb{R}$ where $\mathcal{T} \in \mathbb{R}^d$ and $a_0 \in \mathcal{T}$ be arbitrary. Then the following are equivalent:

(i) $\gamma(x,y) = \operatorname{var}(Z(x) - Z(y))$ for some $Z \sim GRF_{\mathcal{T}}$

(ii) $K(x,y) = \frac{1}{2}\left(\gamma(x,a_0) + \gamma(y,a_0) - \gamma(x,y)\right)$ is pos. def. on $\mathcal{T}\times\mathcal{T}$

(iii) $\exp(-\beta\gamma(x,y))$ is pos. def. on $\mathcal{T}\times\mathcal{T}$ $\forall \beta > 0$,

(iv) $$\left.\begin{array}{l} x_1,\dots,x_n \in \mathcal{T} \\ c_1,\dots,c_n \in \mathbb{R} \\ c_1+\dots+c_n = 0 \end{array}\right\} \Longrightarrow \sum_{i,j=1}^{n} c_i c_j \gamma(x_i,x_j) \le 0$$

Condition (iv) shows that variograms behave similar to (the negative of) a positive definite function ... but with a slightly weaker condition to show

i.e. can focus on $c_1,\dots,c_n$ ...

$c_1+\dots+c_n = 0 \leftarrow \sum_{i=1}^{n} f(x_i)\, c_i = 0$ if $f(x) \equiv$ constant.

Lets see how (i) $\Longrightarrow$ (ii) & (iii) allow us to construct covariance functions

$$\boxed{\gamma(x,y) = \|x-y\|^2}$$

via $\frac{1}{2}\left(\gamma(x,0)+\gamma(y,0) - \gamma(x,y)\right)$

$$K_{Ln}(x,y) = \langle x,y\rangle$$

via $\exp(-\gamma(x,y))$

$$K_{Ga}(x,y) = \exp(-\|x-y\|^2)$$

$$\boxed{\gamma(x,y) = \|x-y\|}$$

via $\frac{1}{2}\left(\gamma(x,0)+\gamma(y,0) - \gamma(x,y)\right)$

$$K_{Br}(x,y) = \frac{1}{2}\left(\|x\| + \|y\| - \|x-y\|\right)$$

via $\exp(-\gamma(x,y))$

$$K_{Ex}(x,y) = \exp(-\|x-y\|)$$

Conversly lets see how to get new variograms from covariance functions

Recall the Matérn auto covariance M_ν, for $0<\nu<1$, satisfies

$$\underbrace{2(M_\nu(0) - M_\nu(t))}_{\text{variogram}} \sim (\text{constant})\, t^{2\nu} \quad \text{as } t\to 0$$

The constant term here is positive so that

$$\lim_{\varepsilon\to 0} \frac{M_\nu(0) - M_\nu(\varepsilon\|x-y\|)}{\varepsilon^{2\nu}} \propto \|x-y\|^{2\nu}$$

we can verify (iv) in Schoenberg's thm as follows. Let $x_1, \dots, x_n \in \mathbb{R}^d$ and

$$c_1 + \cdots + c_n = 0$$

$$\sum_{i,j=1}^{n} c_i c_j \|x_i - x_j\|^{2\nu}$$

$$\propto \lim_{\varepsilon \to 0} \sum_{i,j=1}^{n} c_i c_j \frac{\mathcal{M}_\nu(0) - \mathcal{M}_\nu(\varepsilon \|x_i - x_j\|)}{\varepsilon^{2\nu}}$$

$\sum_{i=1}^{n} c_i \mathcal{M}(0) = 0$

$$= \underbrace{- \lim_{\varepsilon \to 0} \underbrace{\sum_{i,j=1}^{n} c_i c_j \frac{\mathcal{M}_\nu(\varepsilon \|x_i - x_j\|)}{\varepsilon^{2\nu}}}_{\geq 0 \ \forall \varepsilon > 0}}_{\text{so this will be negative}}$$

This implies (iv) holds for

$$\gamma(x,y) = \|x-y\|^{2\nu}, \qquad \nu \in (0,1)$$

and therefore, via (i), there exists a

$$Z \sim GRF_{\mathbb{R}^d}(0, K) \quad (\text{not Matérn btw}) \text{ s.t.}$$

$$\mathrm{var}(Z(x) - Z(y)) = \|x-y\|^{2\nu}$$

for $\nu \in (0,1)$.

Applying (ii) again (with $\nu \in (0,1)$)

$$\boxed{\gamma(x,y) = \|x-y\|^{2\nu}}$$

via $\exp(-\gamma(x,y))$

$$\exp\left(-\|x-y\|^{2\nu}\right)$$

Power exponential auto-covariance.

via $\frac{1}{2}\left(\gamma(x,0) + \gamma(y,0) - \gamma(x,y)\right)$

$$\frac{1}{2}\left(\|x\|^{2\nu} + \|y\|^{2\nu} - \|x-y\|^{2\nu}\right)$$

recognized as Fractional Brownian Field / Process