Gaussian Random Variables

Definition

A Gaussian random variable X is completely specified by its mean μ and standard deviation σ. Its density function is:

$$\mathbf{P}[X = x] = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

Definition

A multivariate Gaussian random variable $\boldsymbol{X}$ is completely specified by its mean $\boldsymbol{\mu}$ and covariance matrix Σ (positive definite and symmetric). Its density function is:

$$\mathbf{P}[\boldsymbol{X} = \boldsymbol{x}] = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

Definition

A Gaussian process $f(x)$ is a collection of random variables, any finite number of which have a joint Gaussian distribution. A Gaussian process is completely specified by its mean function $\mu(x)$ and its covariance function $k(x, y)$. For $n \in \mathbb{N}$ and $x_1, \ldots, x_n$:

$$(f(x_1), \ldots, f(x_n)) \sim \mathcal{N}((\mu(x_1), \ldots, \mu(x_n)), \boldsymbol{K})$$

$$\boldsymbol{K} := \begin{pmatrix} k(x_1, x_1) & k(x_1, x_2) & \ldots \\ k(x_2, x_1) & k(x_2, x_2) & \ldots \\ \ldots & & \end{pmatrix}$$

Sampling from a GP

- Goal: Generate a draw from a GP with mean μ and covariance $\boldsymbol{K}$.
- Compute Cholesky decomposition of $\boldsymbol{K}$, i.e.
$$\boldsymbol{K} = \boldsymbol{L}\boldsymbol{L}^\top,$$
and $\boldsymbol{L}$ is lower triangular.
- Generate
$$\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}).$$
- Compute
$$\boldsymbol{x} = \mu + \boldsymbol{L}\boldsymbol{u}.$$
- $\boldsymbol{x}$ has the right distribution, i.e.
$$\mathbf{E}(\boldsymbol{x} - \mu)(\boldsymbol{x} - \mu)^\top = \boldsymbol{L}\mathbf{E}[\boldsymbol{u}\boldsymbol{u}^\top]\boldsymbol{L}^\top = \boldsymbol{K}.$$
- Often numerical unstable: Add $\epsilon \boldsymbol{I}$ to the covariance.

Sampling from a GP

- Goal: Generate a draw from a GP with mean μ and covariance $\boldsymbol{K}$.
- Compute Cholesky decomposition of $\boldsymbol{K}$, i.e.

$$\boldsymbol{K} = \boldsymbol{L}\boldsymbol{L}^\top,$$

  and $\boldsymbol{L}$ is lower triangular.
- Generate

$$\boldsymbol{u} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}).$$

- Compute

$$\boldsymbol{x} = \mu + \boldsymbol{L}\boldsymbol{u}.$$

- $\boldsymbol{x}$ has the right distribution, i.e.

$$\mathbf{E}(\boldsymbol{x} - \mu)(\boldsymbol{x} - \mu)^\top = \boldsymbol{L}\mathbf{E}[\boldsymbol{u}\boldsymbol{u}^\top]\boldsymbol{L}^\top = \boldsymbol{K}.$$

- Often numerical unstable: Add $\epsilon\boldsymbol{I}$ to the covariance.

Sampling from a GP

- Goal: Generate a draw from a GP with mean μ and covariance $\boldsymbol{K}$.
- Compute Cholesky decomposition of $\boldsymbol{K}$, i.e.

$$\boldsymbol{K} = \boldsymbol{L}\boldsymbol{L}^\top,$$

  and $\boldsymbol{L}$ is lower triangular.
- Generate

$$\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}).$$

- Compute

$$\boldsymbol{x} = \mu + \boldsymbol{L}\boldsymbol{u}.$$

- $\boldsymbol{x}$ has the right distribution, i.e.

$$\mathbf{E}(\boldsymbol{x} - \mu)(\boldsymbol{x} - \mu)^\top = \boldsymbol{L}\mathbf{E}[\boldsymbol{u}\boldsymbol{u}^\top]\boldsymbol{L}^\top = \boldsymbol{K}.$$

- Often numerical unstable: Add $\epsilon \boldsymbol{I}$ to the covariance.

Sampling from a GP

- Goal: Generate a draw from a GP with mean μ and covariance $\boldsymbol{K}$.
- Compute Cholesky decomposition of $\boldsymbol{K}$, i.e.

$$\boldsymbol{K} = \boldsymbol{L}\boldsymbol{L}^\top,$$

  and $\boldsymbol{L}$ is lower triangular.
- Generate

$$\boldsymbol{u} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}).$$

- Compute

$$\boldsymbol{x} = \mu + \boldsymbol{L}\boldsymbol{u}.$$

- $\boldsymbol{x}$ has the right distribution, i.e.

$$\mathbf{E}(\boldsymbol{x} - \mu)(\boldsymbol{x} - \mu)^\top = \boldsymbol{L}\mathbf{E}[\boldsymbol{u}\boldsymbol{u}^\top]\boldsymbol{L}^\top = \boldsymbol{K}.$$

- Often numerical unstable: Add $\epsilon \boldsymbol{I}$ to the covariance.

Sampling from a GP

- Goal: Generate a draw from a GP with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{K}$.
- Compute Cholesky decomposition of $\boldsymbol{K}$, i.e.

$$\boldsymbol{K} = \boldsymbol{L}\boldsymbol{L}^\top,$$

  and $\boldsymbol{L}$ is lower triangular.
- Generate

$$\boldsymbol{u} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}).$$

- Compute

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{u}.$$

- $\boldsymbol{x}$ has the right distribution, i.e.

$$\mathbf{E}(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^\top = \boldsymbol{L}\mathbf{E}[\boldsymbol{u}\boldsymbol{u}^\top]\boldsymbol{L}^\top = \boldsymbol{K}.$$

- Often numerical unstable: Add $\epsilon \boldsymbol{I}$ to the covariance.

Sampling from a GP

- Goal: Generate a draw from a GP with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{K}$.
- Compute Cholesky decomposition of $\boldsymbol{K}$, i.e.

$$\boldsymbol{K} = \boldsymbol{L}\boldsymbol{L}^\top,$$

  and $\boldsymbol{L}$ is lower triangular.
- Generate

$$\boldsymbol{u} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}).$$

- Compute

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{u}.$$

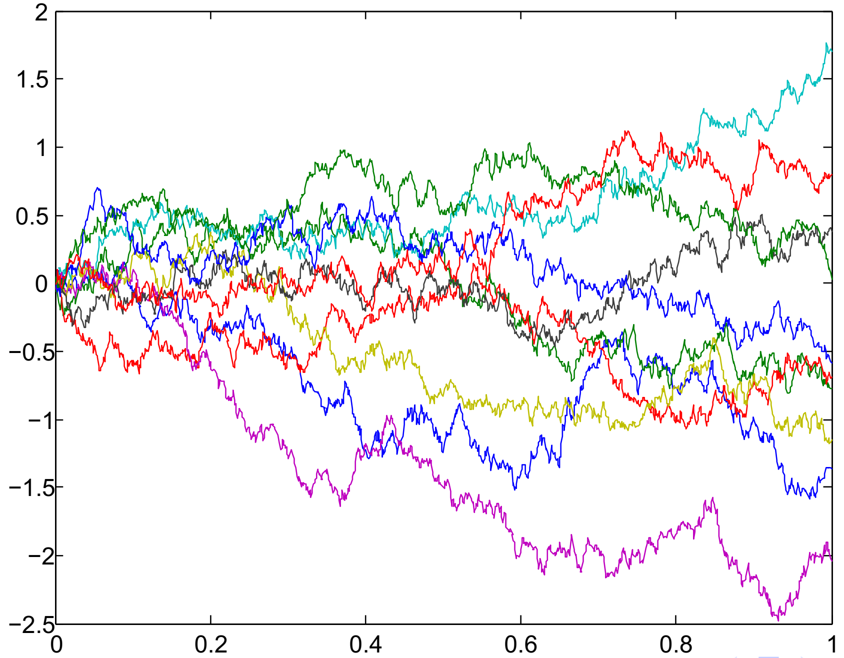
- $\boldsymbol{x}$ has the right distribution, i.e.

$$\mathbf{E}(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^\top = \boldsymbol{L}\mathbf{E}[\boldsymbol{u}\boldsymbol{u}^\top]\boldsymbol{L}^\top = \boldsymbol{K}.$$

- Often numerical unstable: Add $\epsilon\boldsymbol{I}$ to the covariance.

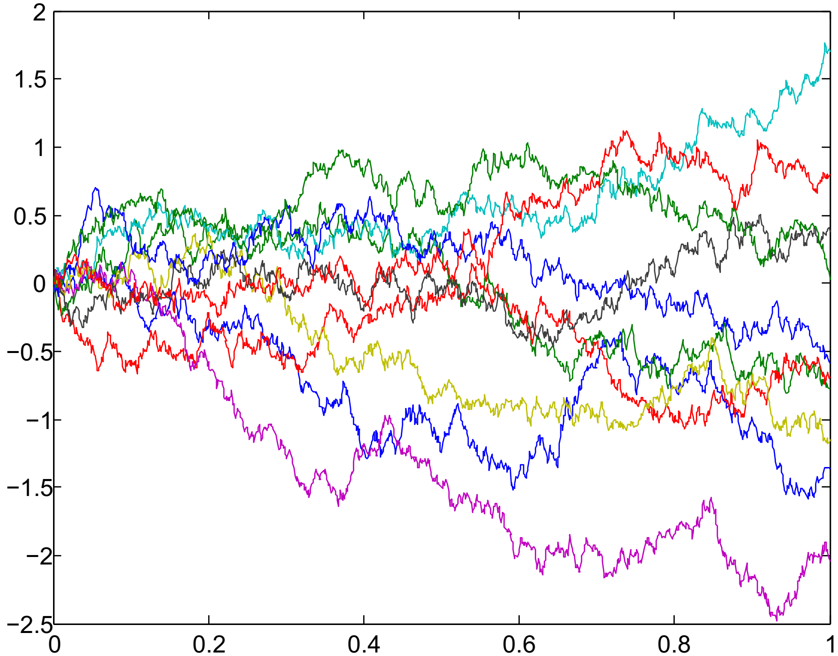
Example 1: Brownian Motion

- Most famous GP: Brownian Motion.
- Process on the real line starting at time 0 with value $f(0) = 0$.
- Covariance: $k(s, t) = \min\{s, t\}$.
- Brownian Motion is a Markov process. Means intuitively that for times $t_1 < t_2 < t_3$ the value of $f(t_3)$ conditional on $f(t_2)$ is independent of $f(t_1)$.



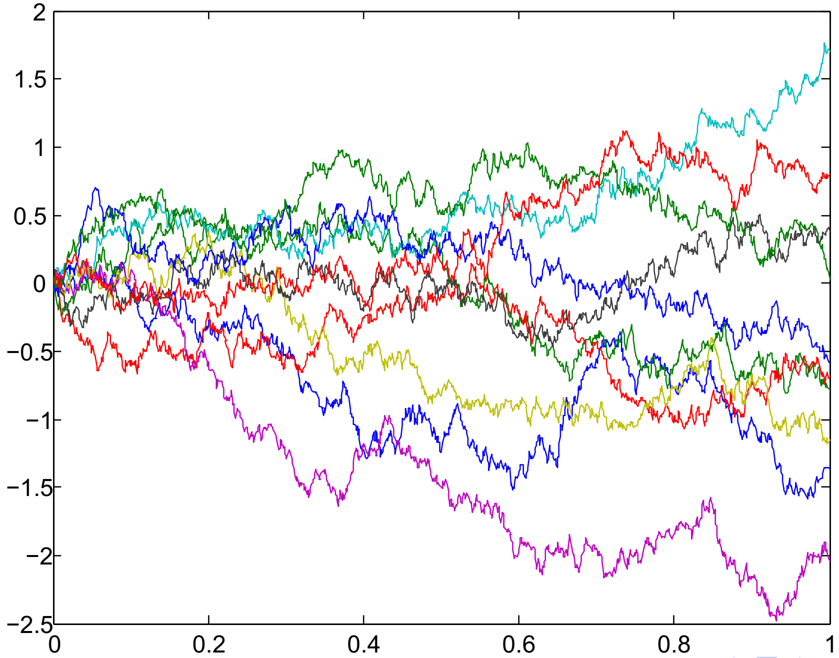
Example 1: Brownian Motion

- Most famous GP: Brownian Motion.
- Process on the real line starting at time 0 with value $f(0) = 0$.
- Covariance: $k(s, t) = \min\{s, t\}$.
- Brownian Motion is a Markov process. Means intuitively that for times $t_1 < t_2 < t_3$ the value of $f(t_3)$ conditional on $f(t_2)$ is independent of $f(t_1)$.



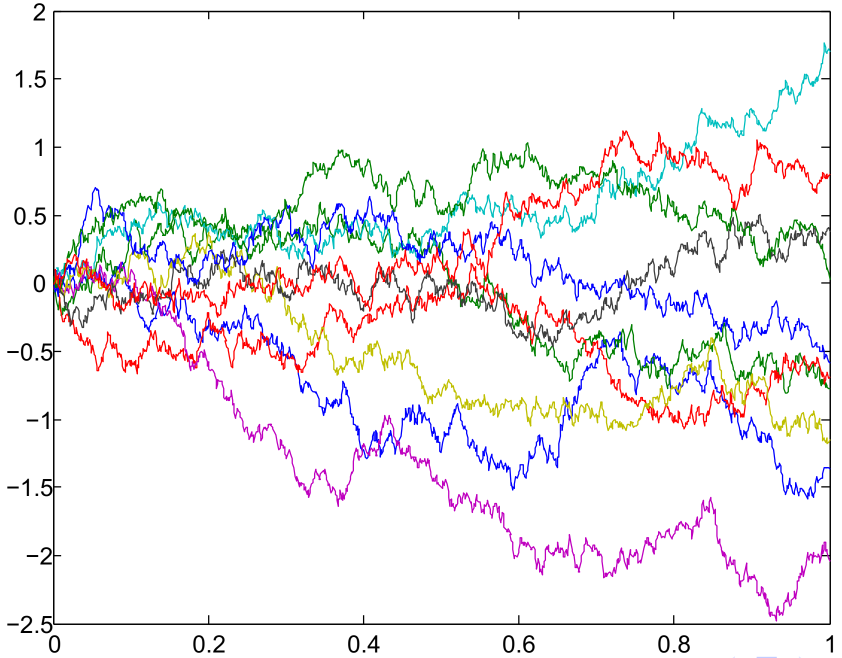
Example 1: Brownian Motion

- Most famous GP: Brownian Motion.
- Process on the real line starting at time 0 with value $f(0) = 0$.
- Covariance: $k(s, t) = \min\{s, t\}$.
- Brownian Motion is a Markov process. Means intuitively that for times $t_1 < t_2 < t_3$ the value of $f(t_3)$ conditional on $f(t_2)$ is independent of $f(t_1)$.



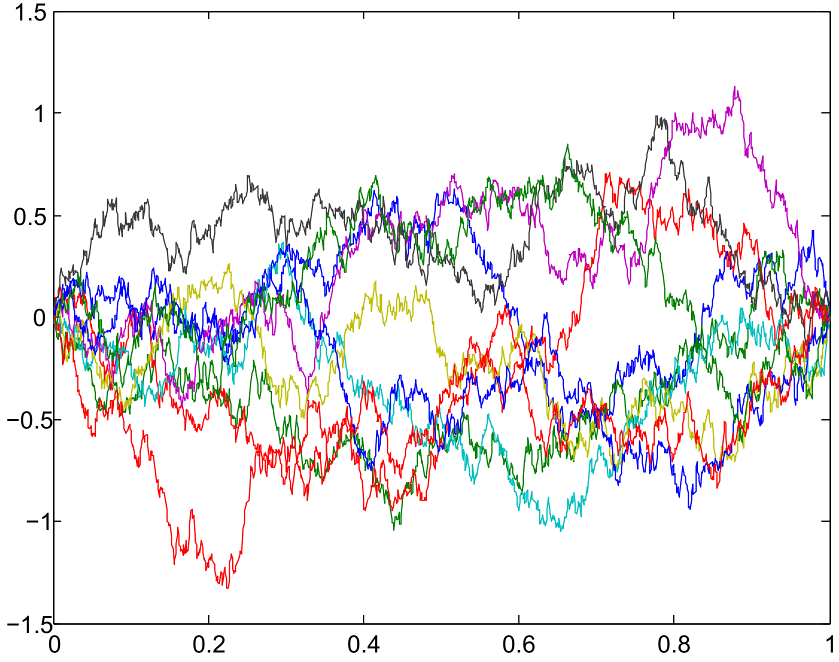
Example 1: Brownian Motion

- Most famous GP: Brownian Motion.
- Process on the real line starting at time 0 with value $f(0) = 0$.
- Covariance: $k(s, t) = \min\{s, t\}$.
- Brownian Motion is a Markov process. Means intuitively that for times $t_1 < t_2 < t_3$ the value of $f(t_3)$ conditional on $f(t_2)$ is independent of $f(t_1)$.



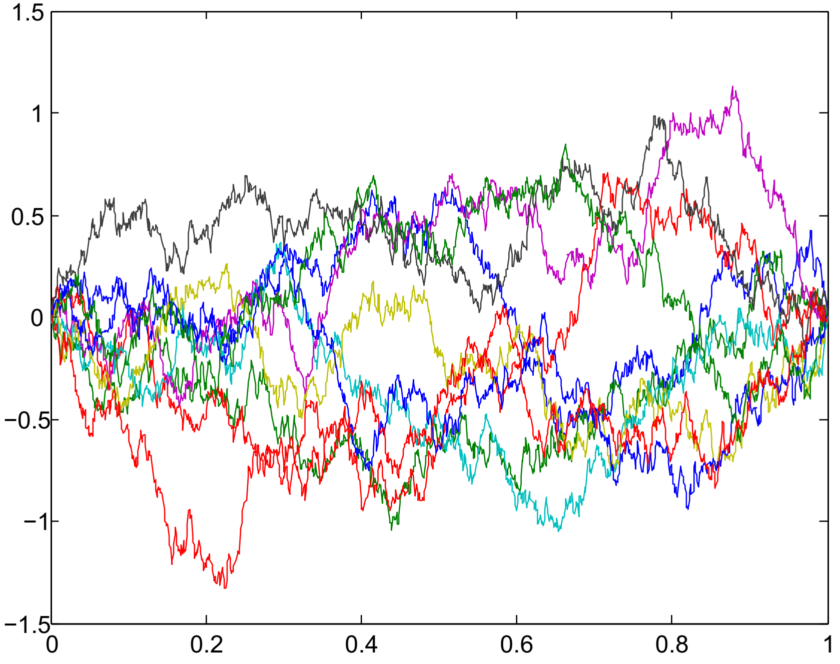
Example 2: Brownian Bridge

- A bridge is a stochastic process that is "clamped" at two points, i.e. each path goes (w.p. 1) through two specified points.
- Example: Brownian Bridge on $[0, 1]$ with $f(0) = f(1) = 0$.
- Covariance: $k(s, t) = \min\{s, t\} - st$



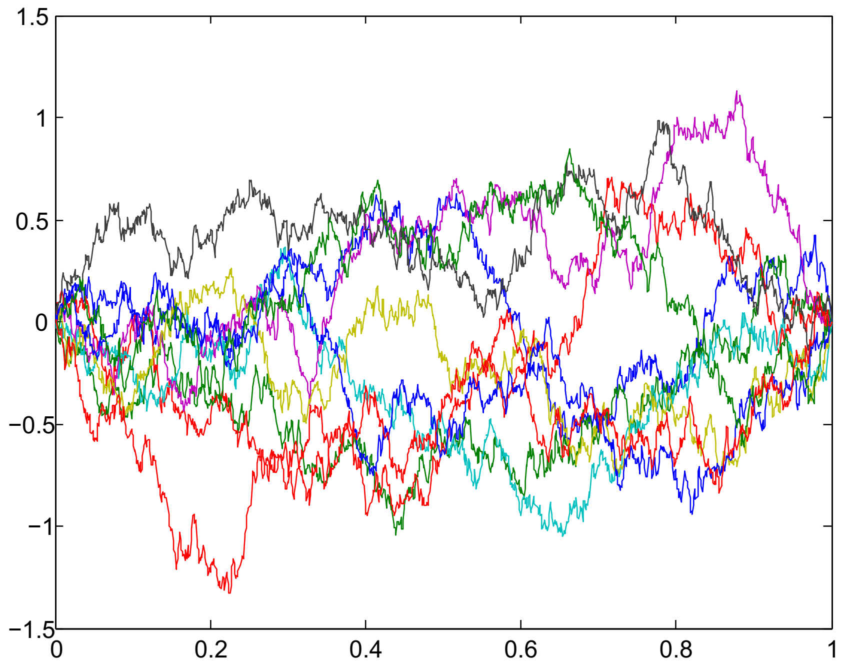
Example 2: Brownian Bridge

- A bridge is a stochastic process that is "clamped" at two points, i.e. each path goes (w.p. 1) through two specified points.
- Example: Brownian Bridge on $[0, 1]$ with $f(0) = f(1) = 0$.
- Covariance: $k(s, t) = \min\{s, t\} - st$



Example 2: Brownian Bridge

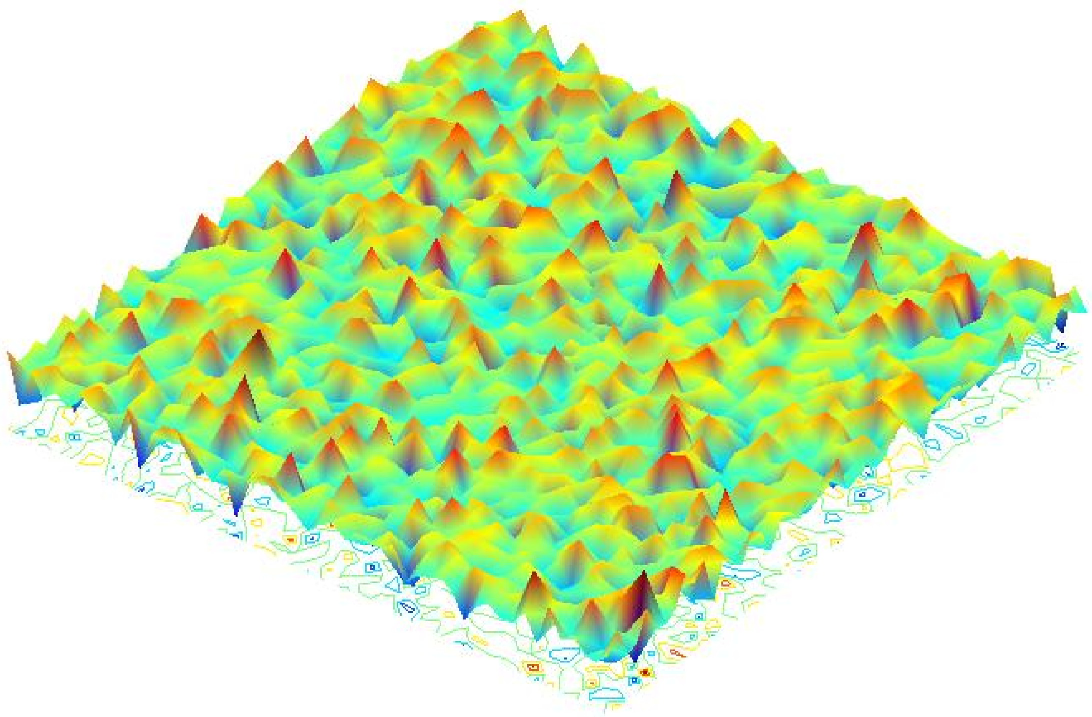
- A bridge is a stochastic process that is "clamped" at two points, i.e. each path goes (w.p. 1) through two specified points.
- Example: Brownian Bridge on $[0, 1]$ with $f(0) = f(1) = 0$.
- Covariance: $k(s, t) = \min\{s, t\} - st$



Example 3: Gauss Covariance

- Gauss covariance function:

$$k(x, y) = \exp\left(-\frac{1}{2\sigma}||x-y||_2^2\right).$$



Continuity and Differentiability of Sample Paths

- These three processes have continuous sample paths (w.p. 1).
- The process with the Gauss covariance has furthermore sample paths that are infinitely often differentiable (w.p. 1).
- Sample paths of Markov processes are very “rough” with a lot of fluctuations. The sample paths of Brownian motion are, for example, nowhere differentiable (w.p. 1).
- It is useful for modelling purposes to be able to specify the smoothnes of a process in terms of how often the sample paths are differentiable. The Matérn class of covariance functions allows to do that.

Continuity and Differentiability of Sample Paths

- These three processes have continuous sample paths (w.p. 1).
- The process with the Gauss covariance has furthermore sample paths that are infinitely often differentiable (w.p. 1).
- Sample paths of Markov processes are very “rough” with a lot of fluctuations. The sample paths of Brownian motion are, for example, nowhere differentiable (w.p. 1).
- It is useful for modelling purposes to be able to specify the smoothnes of a process in terms of how often the sample paths are differentiable. The Matérn class of covariance functions allows to do that.

Continuity and Differentiability of Sample Paths

- These three processes have continuous sample paths (w.p. 1).
- The process with the Gauss covariance has furthermore sample paths that are infinitely often differentiable (w.p. 1).
- Sample paths of Markov processes are very “rough” with a lot of fluctuations. The sample paths of Brownian motion are, for example, nowhere differentiable (w.p. 1).
- It is useful for modelling purposes to be able to specify the smoothnes of a process in terms of how often the sample paths are differentiable. The Matérn class of covariance functions allows to do that.

Continuity and Differentiability of Sample Paths

- These three processes have continuous sample paths (w.p. 1).
- The process with the Gauss covariance has furthermore sample paths that are infinitely often differentiable (w.p. 1).
- Sample paths of Markov processes are very “rough” with a lot of fluctuations. The sample paths of Brownian motion are, for example, nowhere differentiable (w.p. 1).
- It is useful for modelling purposes to be able to specify the smoothnes of a process in terms of how often the sample paths are differentiable. The Matérn class of covariance functions allows to do that.