

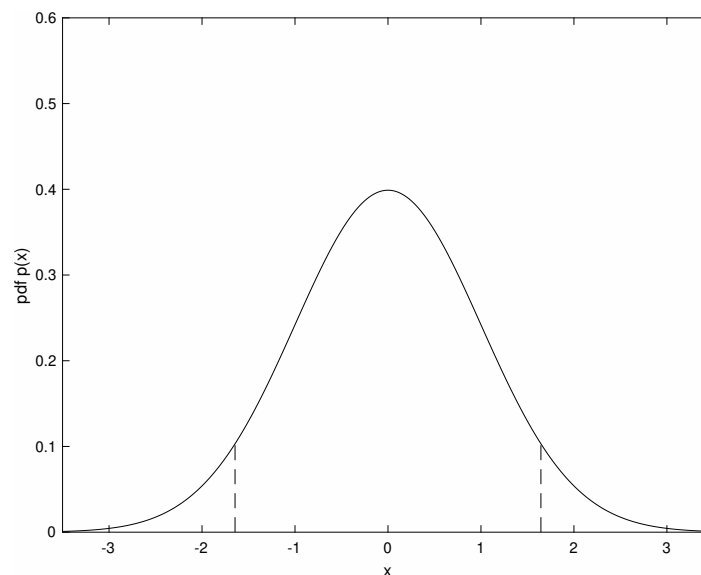
Gaussian processes and applications

Uncertainty quantification in a diverse range of applications:

- Genetic data (dependence via 'distance' in cells or in pedigrees)
- Functional data (dependence via 'distance' in covariates, use this to borrow information (smooth surface))
- Response surfaces modeling and optimization

Extremely common as building block in several machine learning applications.

Univariate Gaussian distribution



$$Y \sim N(\mu, \sigma^2).$$

$$p(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{(y - \mu)^2}{\sigma^2}\right), \quad y \in \mathbb{R}.$$

$$Z = \frac{Y - \mu}{\sigma}, \quad Y = \mu + \sigma Z.$$

Z is standard normal, mean 0 and variance 1.

Multivariate Gaussian distribution

Size $n \times 1$ vector $\mathbf{Y} = (Y_1, \dots, Y_n)$

$$p(\mathbf{y}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right), \quad \mathbf{Y} \in \mathbb{R}^n.$$

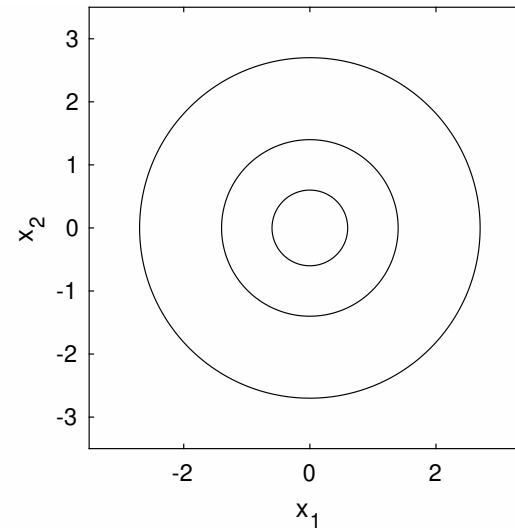
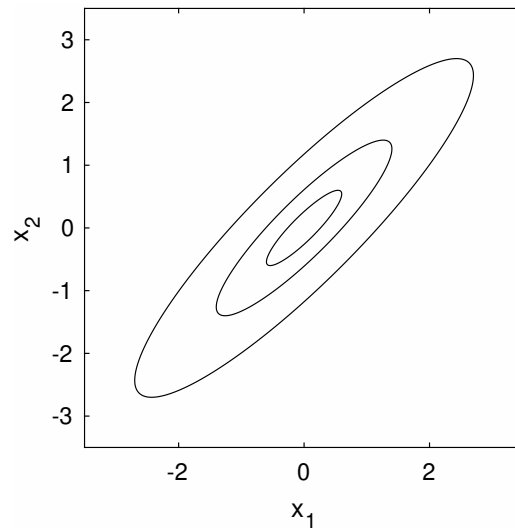
Mean is $E(\mathbf{Y}) = \boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$. Symmetric, positive-definite variance-covariance matrix is

$$\Sigma = \begin{bmatrix} \Sigma_{1,1} & \dots & \Sigma_{1,n} \\ \dots & \dots & \dots \\ \Sigma_{n,1} & \dots & \Sigma_{n,n} \end{bmatrix},$$

$$\Sigma_{i,i} = \sigma_i^2 = \text{Var}(Y_i), \quad \Sigma_{i,j} = \text{Cov}(Y_i, Y_j), \quad \text{Corr}(Y_i, Y_j) = \Sigma_{i,j} / (\sigma_i \sigma_j).$$

Illustrations $n = 2$

- correlation 0.9 (left), independent (right).

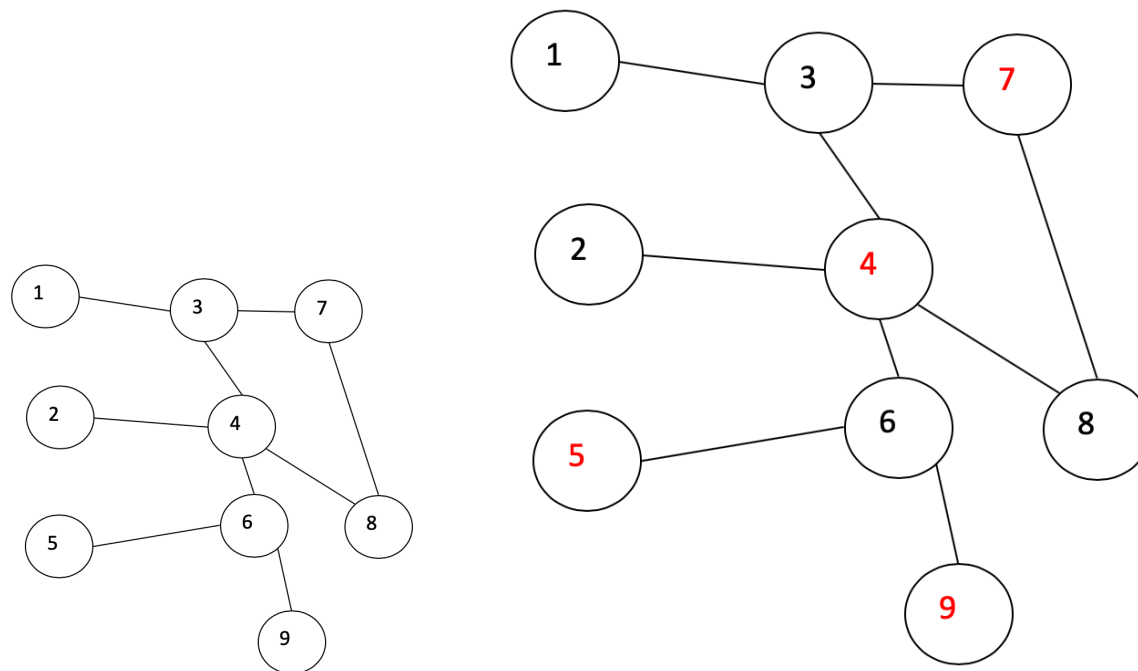


Joint for blocks

$\mathbf{Y}_A = (Y_{A,1}, \dots, Y_{A,n_A})^\top$, $\mathbf{Y}_B = (Y_{B,1}, \dots, Y_{B,n_B})^\top$, joint Gaussian with mean $(\boldsymbol{\mu}_A, \boldsymbol{\mu}_B)^\top$, covariance:

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_A, \boldsymbol{\mu}_B)^\top, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_A & \boldsymbol{\Sigma}_{A,B} \\ \boldsymbol{\Sigma}_{B,A} & \boldsymbol{\Sigma}_B \end{bmatrix},$$

Observed set - non-observed set



$$n = 9, n_A = 5, n_B = 4.$$

A set is non-observed (black). B set is the observation set (red).

Because of the dependence (here illustrated by edges), the information on variables in set B will propagate to set A variables.

Conditioning

$$\begin{aligned}E(\mathbf{Y}_A | \mathbf{Y}_B) &= \boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{A,B} \boldsymbol{\Sigma}_B^{-1} (\mathbf{Y}_B - \boldsymbol{\mu}_B), \\ \text{Var}(\mathbf{Y}_A | \mathbf{Y}_B) &= \boldsymbol{\Sigma}_A - \boldsymbol{\Sigma}_{A,B} \boldsymbol{\Sigma}_B^{-1} \boldsymbol{\Sigma}_{B,A}.\end{aligned}$$

Mean is linear in conditioning variable (data).

Variance is not dependent on data.

Transformation

$$\mathbf{Z} = \mathbf{L}^{-1}(\mathbf{Y} - \boldsymbol{\mu}), \quad \mathbf{Y} = \boldsymbol{\mu} + \mathbf{L}\mathbf{Z}, \quad \boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}'.$$

$\mathbf{Z} = (Z_1, \dots, Z_n)$ are independent standard normal, mean 0 and variance 1.

Cholesky factorization

$$\Sigma = \begin{bmatrix} \Sigma_{1,1} & \dots & \Sigma_{1,n} \\ \dots & \dots & \dots \\ \Sigma_{n,1} & \dots & \Sigma_{n,n} \end{bmatrix} = \mathbf{L}\mathbf{L}',$$

Lower triangular matrix

$$\mathbf{L} = \begin{bmatrix} L_{1,1} & 0 & \dots & 0 \\ L_{2,1} & L_{2,2} & \dots & 0 \\ \dots & \dots & \dots & 0 \\ L_{n,1} & L_{n,2} & \dots & L_{n,n} \end{bmatrix},$$

Cholesky - example

$$\Sigma = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}, \quad \mathbf{L} = \begin{bmatrix} 1 & 0 \\ 0.9 & 0.44 \end{bmatrix}.$$

Consider sampling from joint $p(y_1, y_2)$:

- Draw independent standard Gaussian variables Z_1 and Z_2 (mean 0, variance 1).
- $Y_1 = \mu_1 + L_{1,1}Z_1$.
- $Y_2 = \mu_2 + L_{2,1}Z_1 + L_{2,2}Z_2 = \mu_2 + L_{2,1} \frac{Y_1 - \mu_1}{L_{1,1}} + L_{2,2}Z_2$

Gaussian process (GP)

For any set of time locations $t_1, \dots, t_n$ we have that $Y(t_1), \dots, Y(t_n)$ is jointly multivariate Gaussian with mean $\mu(t_i)$, $i = 1, \dots, n$. and

$$\Sigma = \begin{bmatrix} \Sigma_{1,1} & \dots & \Sigma_{1,n} \\ \dots & \dots & \dots \\ \Sigma_{n,1} & \dots & \Sigma_{n,n} \end{bmatrix},$$

A GP is a distribution over functions. For every t , we can determine the mean and the variance.

Covariance function

The covariance tends to decay with 'distance':

$$\Sigma_{i,j} = \gamma(|t_i - t_j|),$$

for some covariance function γ .

Examples (distance $h = |t - t'|$):

$$\gamma_{\text{exp}}(h) = \sigma^2 \exp(-\phi_E h)$$

$$\gamma_{\text{mat}}(h) = \sigma^2 (1 + \phi_M h) \exp(-\phi_M h)$$

$$\gamma_{\text{gauss}}(h) = \sigma^2 \exp(-\phi_G h^2)$$

Properties

- Marginal variance σ^2 : how much can the process deviate from the mean.
- Range (ϕ_E, ϕ_m, ϕ_G) how far away are things dependent.
- Smoothness: Realizations are 0 times differentiable for Exponential, 1 times differentiable for Matérn-type, and infinitely many times differentiable for Gaussian.

Illustration of covariance function

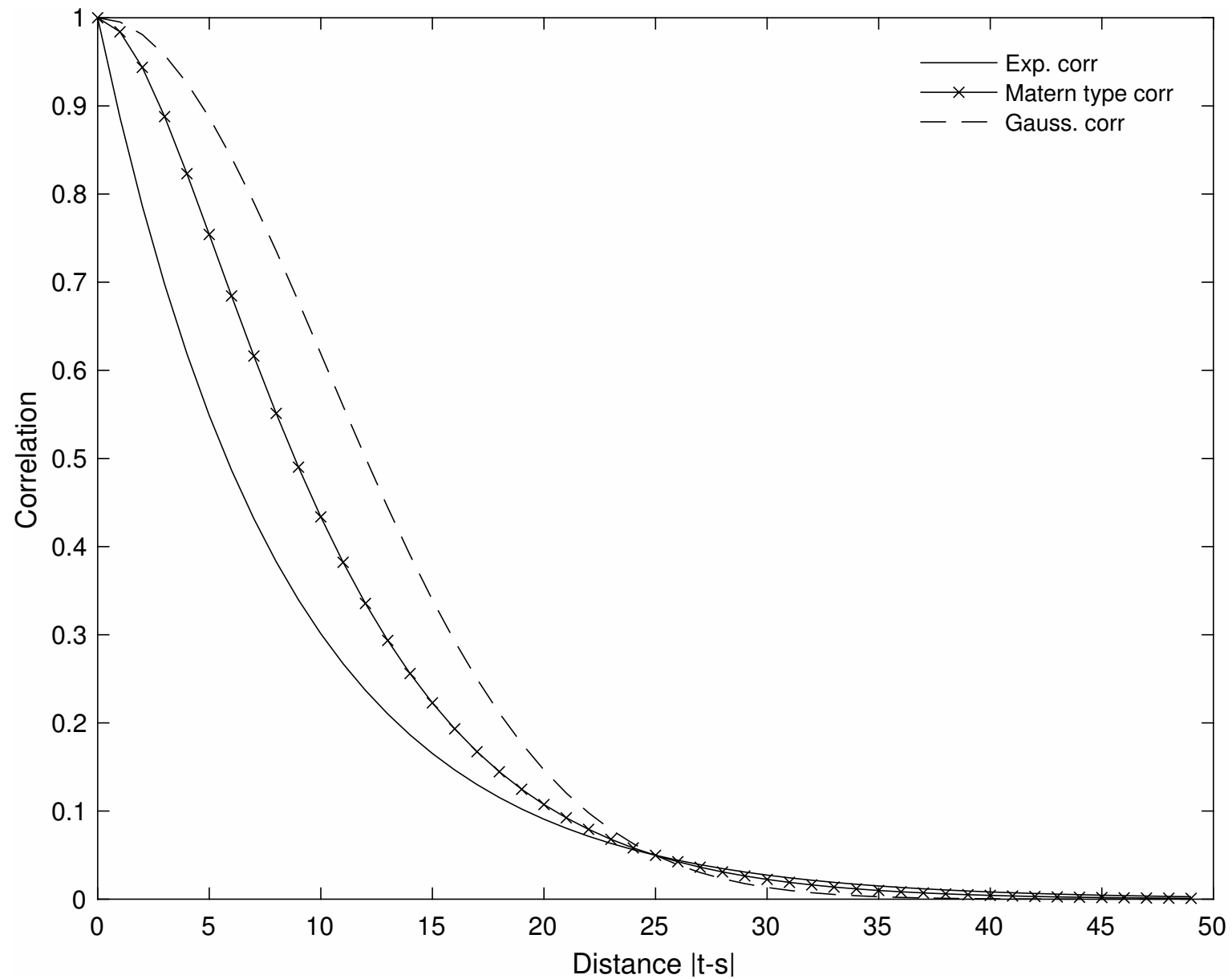
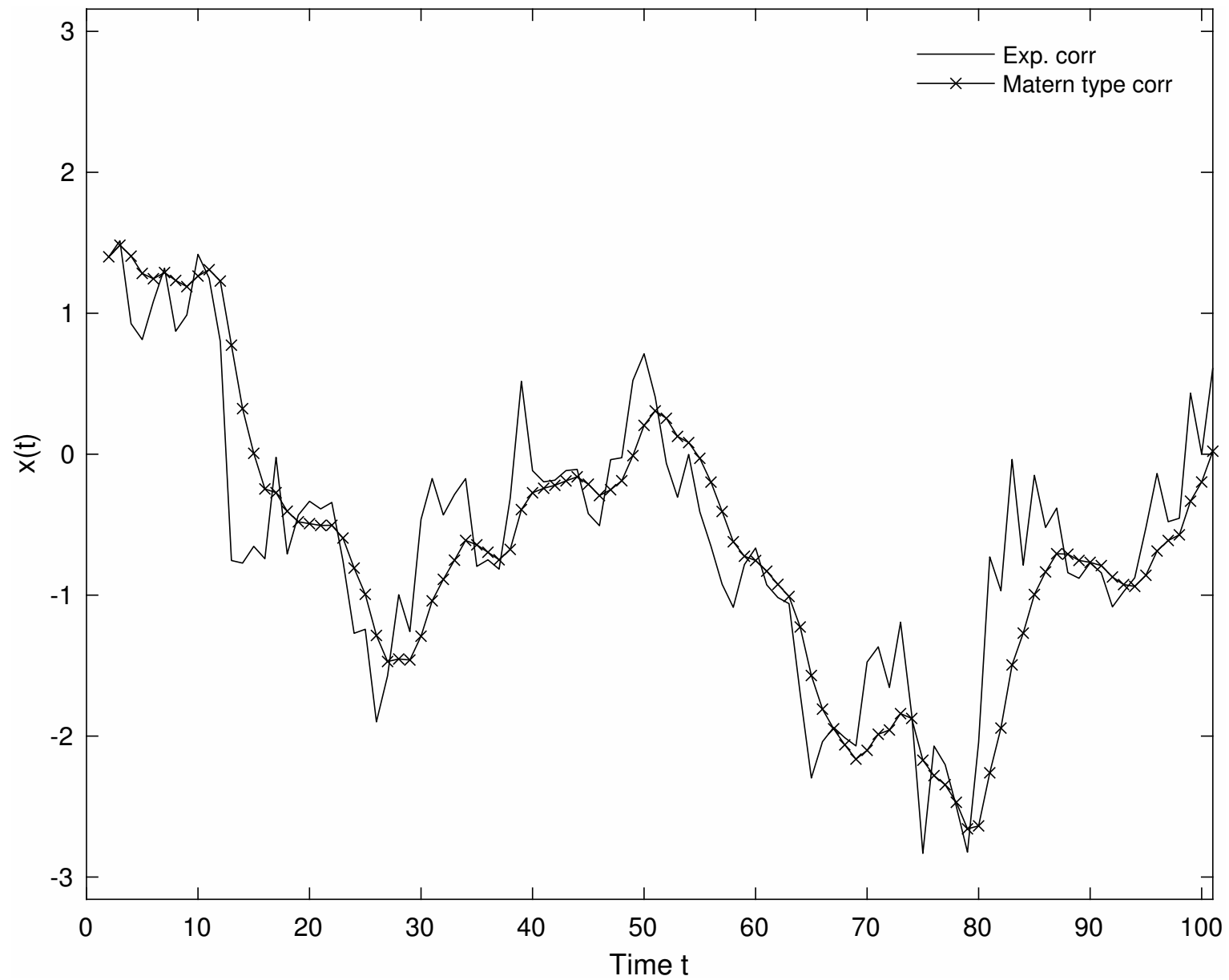


Illustration of samples of Gaussian processes



Conditional formula

$$\begin{aligned}E(\mathbf{Y}_A|\mathbf{Y}_B) &= \boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{A,B}\boldsymbol{\Sigma}_B^{-1}(\mathbf{Y}_B - \boldsymbol{\mu}_B), \\ \text{Var}(\mathbf{Y}_A|\mathbf{Y}_B) &= \boldsymbol{\Sigma}_A - \boldsymbol{\Sigma}_{A,B}\boldsymbol{\Sigma}_B^{-1}\boldsymbol{\Sigma}_{B,A}.\end{aligned}$$

- Expectation linear in data.
- Variance only dependent on data locations, not data.
- Expectation close to conditioning variables near data locations, goes to $\boldsymbol{\mu}_A$ far from data.
- Variance small near data locations, goes to $\boldsymbol{\Sigma}_A$ far from data.

Illustration of conditioning in Gaussian processes

