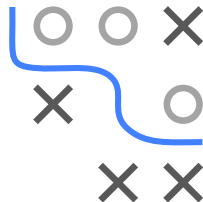
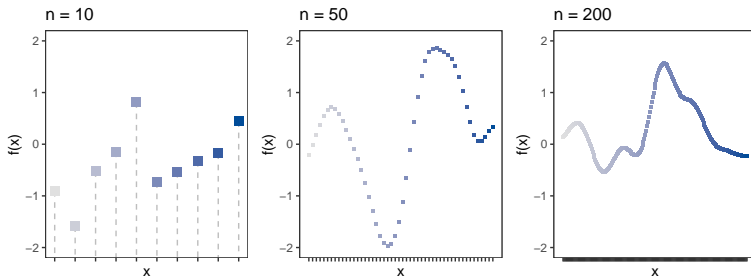


# FROM DISCRETE TO CONTINUOUS FUNCTIONS

- We defined distributions on functions with finite domain by putting a finite Gaussian on it

$$\mathbf{f} = [f(\mathbf{x}^{(1)}), f(\mathbf{x}^{(2)}), \dots, f(\mathbf{x}^{(n)})] \sim \mathcal{N}(\mathbf{m}, \mathbf{K})$$

- We can do this for  $n \rightarrow \infty$  (as “granular” as we want)



# FROM DISCRETE TO CONTINUOUS FUNCTIONS

- No matter how large  $n$  is, we are still considering a function over a discrete domain.
- How can we extend our definition to functions with **continuous domain**  $\mathcal{X} \subset \mathbb{R}$ ?
- Intuitively, a function  $f$  drawn from **Gaussian process** can be understood as an “infinite” long Gaussian random vector.
- It is unclear how to handle an “infinite” long Gaussian random vector!



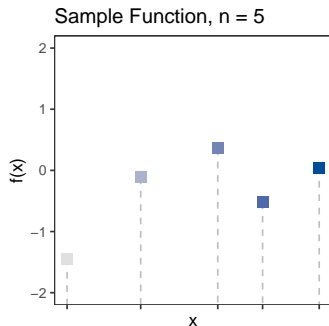
# GAUSSIAN PROCESSES: INTUITION

- Thus, it is required that for **any finite set** of inputs  $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\} \subset \mathcal{X}$ , the vector  $\mathbf{f}$  has a Gaussian distribution

$$\mathbf{f} = \left[ f(\mathbf{x}^{(1)}), \dots, f(\mathbf{x}^{(n)}) \right] \sim \mathcal{N}(\mathbf{m}, \mathbf{K}),$$

with  $\mathbf{m}$  and  $\mathbf{K}$  being calculated by a mean function  $m(\cdot)$  / covariance function  $k(\cdot, \cdot)$ .

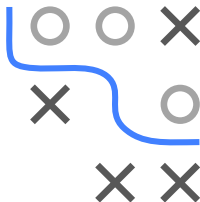
- This property is called **marginalization property**.



$f(x)$



$$\sim \mathcal{N}(\mu, \Sigma)$$



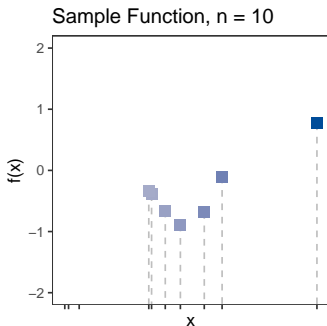
# GAUSSIAN PROCESSES: INTUITION

- Thus, it is required that for **any finite set** of inputs  $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\} \subset \mathcal{X}$ , the vector  $\mathbf{f}$  has a Gaussian distribution

$$\mathbf{f} = \left[ f(\mathbf{x}^{(1)}), \dots, f(\mathbf{x}^{(n)}) \right] \sim \mathcal{N}(\mathbf{m}, \mathbf{K}),$$

with  $\mathbf{m}$  and  $\mathbf{K}$  being calculated by a mean function  $m(\cdot)$  / covariance function  $k(\cdot, \cdot)$ .

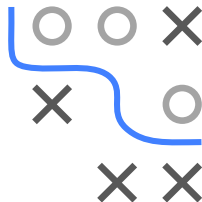
- This property is called **marginalization property**.



$f(x)$



$$\sim \mathcal{N}(\mu, \Sigma)$$



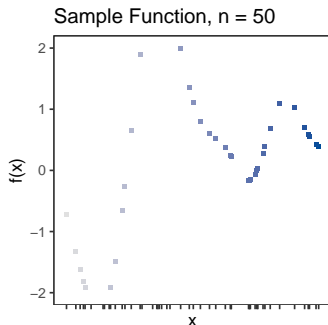
# GAUSSIAN PROCESSES: INTUITION

- Thus, it is required that for **any finite set** of inputs  $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\} \subset \mathcal{X}$ , the vector  $\mathbf{f}$  has a Gaussian distribution

$$\mathbf{f} = \left[ f(\mathbf{x}^{(1)}), \dots, f(\mathbf{x}^{(n)}) \right] \sim \mathcal{N}(\mathbf{m}, \mathbf{K}),$$

with  $\mathbf{m}$  and  $\mathbf{K}$  being calculated by a mean function  $m(\cdot)$  / covariance function  $k(\cdot, \cdot)$ .

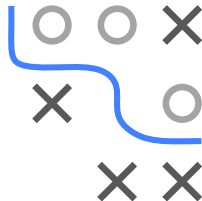
- This property is called **marginalization property**.



$f(x)$



$$\sim \mathcal{N}(\mu, \Sigma)$$



# GAUSSIAN PROCESSES

This intuitive explanation is formally defined as follows:

A function  $f(\mathbf{x})$  is generated by a GP  $\mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$  if for **any finite** set of inputs  $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\}$ , the associated vector of function values  $\mathbf{f} = (f(\mathbf{x}^{(1)}), \dots, f(\mathbf{x}^{(n)}))$  has a Gaussian distribution

$$\mathbf{f} = \left[ f(\mathbf{x}^{(1)}), \dots, f(\mathbf{x}^{(n)}) \right] \sim \mathcal{N}(\mathbf{m}, \mathbf{K}),$$

with

$$\mathbf{m} := \left( m(\mathbf{x}^{(i)}) \right)_i, \quad \mathbf{K} := \left( k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) \right)_{i,j},$$

where  $m(\mathbf{x})$  is called mean function and  $k(\mathbf{x}, \mathbf{x}')$  is called covariance function.



# GAUSSIAN PROCESSES / 2

A GP is thus **completely specified** by its mean and covariance function

$$\begin{aligned}m(\mathbf{x}) &= \mathbb{E}[f(\mathbf{x})] \\k(\mathbf{x}, \mathbf{x}') &= \mathbb{E}\left[(f(\mathbf{x}) - \mathbb{E}[f(\mathbf{x})]) (f(\mathbf{x}') - \mathbb{E}[f(\mathbf{x}')])\right]\end{aligned}$$

**Note:** For now, we assume  $m(\mathbf{x}) \equiv 0$ . This is not necessarily a drastic limitation - thus it is common to consider GPs with a zero mean function.



## SAMPLING FROM A GAUSSIAN PROCESS PRIOR

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2\ell^2}\|\mathbf{x} - \mathbf{x}'\|^2\right), \quad \ell = 1.$$

(\*) We will talk later about different choices of covariance functions.





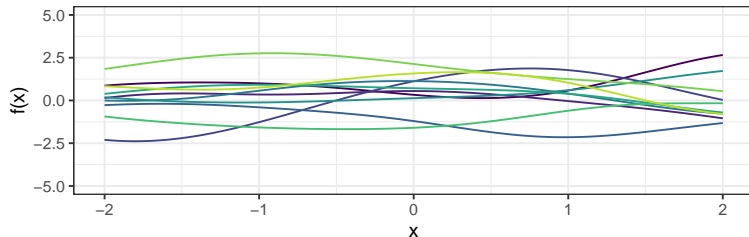
# SAMPLING FROM A GAUSSIAN PROCESS PRIOR

/ 2

To visualize a sample function, we

- choose a high number  $n$  (equidistant) points  $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\}$
- compute the corresponding covariance matrix  $\mathbf{K} = (k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}))_{i,j}$  by plugging in all pairs  $\mathbf{x}^{(i)}, \mathbf{x}^{(j)}$
- sample from a Gaussian  $\mathbf{f} \sim \mathcal{N}(\mathbf{0}, \mathbf{K})$ .

We draw 10 times from the Gaussian, to get 10 different samples.



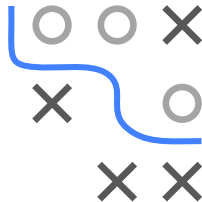
# SAMPLING FROM A GAUSSIAN PROCESS PRIOR

/ 3

Since we specified the mean function to be zero  $m(\mathbf{x}) \equiv 0$ , the drawn functions have zero mean.



# Gaussian Processes as Indexed Family



# GAUSSIAN PROCESSES AS AN INDEXED FAMILY

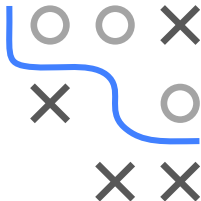
A Gaussian process is a special case of a **stochastic process** which is defined as a collection of random variables indexed by some index set (also called an **indexed family**). What does it mean?

An **indexed family** is a mathematical function (or “rule”) to map indices  $t \in T$  to objects in  $\mathcal{S}$ .

## Definition

A **family of elements in  $\mathcal{S}$  indexed by  $T$**  (indexed family) is a surjective function

$$\begin{aligned}s : T &\rightarrow \mathcal{S} \\ t &\mapsto s_t = s(t)\end{aligned}$$



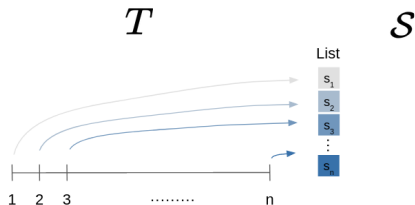
# INDEXED FAMILY

Some simple examples for indexed families are:

- finite sequences (lists):

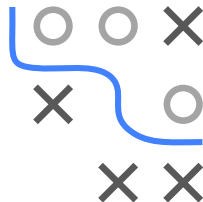
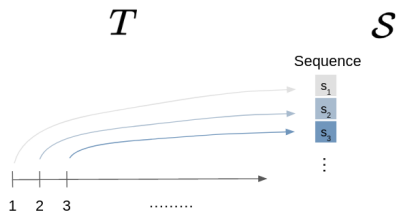
$$T = \{1, 2, \dots, n\} \text{ and}$$

$$(s_t)_{t \in T} \in \mathbb{R}$$



- infinite sequences:

$$T = \mathbb{N} \text{ and } (s_t)_{t \in T} \in \mathbb{R}$$



# INDEXED FAMILY / 2

But the indexed set  $\mathcal{S}$  can be something more complicated, for example functions or **random variables** (RV):

- $T = \{1, \dots, m\}$ ,  $Y_t$ 's are RVs: Indexed family is a random vector.
- $T = \{1, \dots, m\}$ ,  $Y_t$ 's are RVs: Indexed family is a stochastic process in discrete time
- $T = \mathbb{Z}^2$ ,  $Y_t$ 's are RVs: Indexed family is a 2D-random walk.

