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 $n \to \infty$ (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!

× 0 0 × × ×

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 **f** has a Gaussian distribution

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

with **m** and **K** being calculated by a mean function m(.) / covariance function k(.,.).

• This property is called marginalization property.



× 0 0 × 0 × ×

GAUSSIAN PROCESSES: INTUITION

Thus, it is required that for any finite set of inputs
 {x⁽¹⁾,...,x⁽ⁿ⁾} ⊂ X, the vector f has a Gaussian distribution

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

with **m** and **K** being calculated by a mean function m(.) / covariance function k(.,.).

• This property is called marginalization property.



× × 0 × × ×

GAUSSIAN PROCESSES: INTUITION

Thus, it is required that for any finite set of inputs
 {x⁽¹⁾,...,x⁽ⁿ⁾} ⊂ X, the vector f has a Gaussian distribution

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

with **m** and **K** being calculated by a mean function m(.) / covariance function k(.,.).

• This property is called marginalization property.



× × ×

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

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

with

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

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

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$$

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

× 0 0 × × ×

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

We can draw functions from a Gaussian process prior. Let us consider $f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$ with the squared exponential covariance function ^(*)

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

This specifies the Gaussian process completely.

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



SAMPLING FROM A GAUSSIAN PROCESS PRIOR

To visualize a sample function, we

- choose a high number *n* (equidistant) points $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}\}$
- compute the corresponding covariance matrix $\mathbf{K} = (k (\mathbf{x}^{(i)}, \mathbf{x}^{(j)}))_{i,i}$ 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.



× 0 0 × 0 × × ×

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.

× 0 0 × × ×

× > 0 × ×

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 S.

Definition

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

$$egin{array}{rcl} m{s}: m{T} &
ightarrow & m{\mathcal{S}} \ t & \mapsto & m{s}_t = m{s}(t) \end{array}$$

× 0 0 × 0 × ×

INDEXED FAMILY

Some simple examples for indexed families are:





INDEXED FAMILY / 2

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

- *T* = {1,...,*m*}, *Y*_t's are RVs: Indexed family is a random vector.
- T = {1,...,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.



0 0 X X 0 X X

INDEXED FAMILY

- A Gaussian process is also an indexed family, where the random variables *f*(**x**) are indexed by the input values **x** ∈ *X*.
- Their special feature: Any indexed (finite) random vector has a multivariate Gaussian distribution (which comes with all the nice properties of Gaussianity!).





Visualization for a one-dimensional \mathcal{X} .

INDEXED FAMILY

- A Gaussian process is also an indexed family, where the random variables *f*(**x**) are indexed by the input values **x** ∈ *X*.
- Their special feature: Any indexed (finite) random vector has a multivariate Gaussian distribution (which comes with all the nice properties of Gaussianity!).





Visualization for a two-dimensional \mathcal{X} .