Ornstein–Uhlenbeck Processes

Oswin Krause, PML, 2022

UNIVERSITY OF COPENHAGEN

Reminder Random Fields

- Let Ω be an event space (e.g., $\mathbb{R}^N$)
- Let $\mathcal{X}$ be an index set (e.g. $\mathbb{N}$ or $\mathbb{R}^d$)
- A random field is a collection of random variables
  - $F_x \in \Omega, \forall x \in \mathcal{X}$ with realizations f_x
  - Intuitively: A function that assigns a random variable to each point $x \in \mathcal{X}$
- If $\mathcal{X} = \mathbb{R}^d$ it is also called a *random process*

Reminder Random Fields

- Let Ω be an event space (e.g., $\mathbb{R}^N$)
- Let $\mathcal{X}$ be an index set (e.g. $\mathbb{N}$ or $\mathbb{R}^d$)
- A random field is a collection of random variables
    - $F_x \in \Omega, \forall x \in \mathcal{X}$ with realizations f_x
    - Intuitively: A function that assigns a random variable to each point $x \in \mathcal{X}$
- If $\mathcal{X} = \mathbb{R}^d$ it is also called a *random process*
- Random Fields are defined by their Marginals:
    - Pick any finite subset $S_\ell = \{x_1, \dots, x_\ell\} \subseteq \mathcal{X}$
    - Marginal: $p(f_1, \dots, f_\ell | S_\ell) = p(f_{x_1}, \dots, f_{x_\ell})$

Reminder: Gaussian Processes

Definition

Let $\mathcal{X}$ be an index set

A random field $F_x \in \mathbb{R}$ whose marginals $p(f|S)$ are Multivariate Normal distributions, is called a Gaussian Process.

Moreover, there exists a kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and a function $m : \mathcal{X} \to \mathbb{R}$ such, that

$$p(f|S) = \mathcal{N}(m(S), K(S)), \forall S = \{x_1, \ldots, x_\ell\} \subset \mathcal{X}, \forall \ell \in \mathbb{N}$$

with $m(S) = (m(x_1), \ldots, m(x_\ell))$ and $K(S)_{ij} = k(x_i, x_j)$. If m and k are known, we write

$$f \sim \mathcal{GP}(m(\cdot), k(\cdot, \cdot))$$

Ornstein–Uhlenbeck Processes

Special case of GP:

- index set $\mathbb{R}_+$
- Today t is the index variable (for time) and X_t the random variable.
- Mean $m(t) = 0$
- Kernel

$$k(t, t') = \frac{\sigma_k^2}{2\theta} e^{-\theta|t-t'|} \quad .$$

Ornstein–Uhlenbeck Processes

Special case of GP:

- index set $\mathbb{R}_+$
- Today t is the index variable (for time) and X_t the random variable.
- Mean $m(t) = 0$
- Kernel

$$k(t, t') = e^{-\frac{1}{2}|t-t'|} \quad .$$

- Today:
  - $\theta = \frac{1}{2}$
  - $\sigma_k^2 = 1$

Why the OU Process?

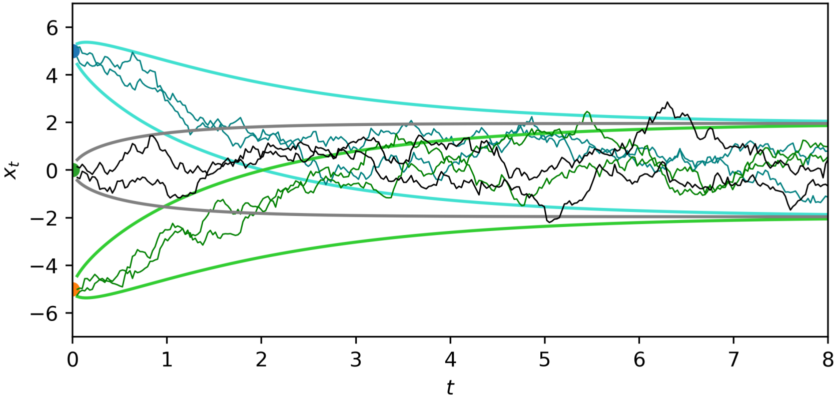
- Stationary process:

$$p(X_t) = p(X_{t'})$$

- It is a "forgetting" process
  - for large $t' > t$: $X_{t'}$ almost independent of X_t
- Markov process:

$$p(x_1, x_2, \ldots, x_T | S) = p(x_1)p(x_2|x_1)\ldots p(x_T|x_{T-1})$$

OU Process Samples



How does the OU process "forget"?

- for large $t' > t$: $X_{t'}$ almost independent of X_t
- That means

$$p(x_{t'}|x_t) \rightarrow p(x_{t'}), \text{ as } |t - t'| \rightarrow \infty$$

Deriving $p(x_{t'}|x_t)$

The Marginal of the OU process for the two variables $X_t, X_{t'}$ is

$$\begin{bmatrix} X_t \\ X_{t'} \end{bmatrix} \sim \mathcal{N}\left(0, \underbrace{\begin{bmatrix} 1 & e^{-\frac{1}{2}|t-t'|} \\ e^{-\frac{1}{2}|t-t'|} & 1 \end{bmatrix}}_{K(S)}\right)$$

Deriving $p(x_{t'}|x_t)$

Conditioning leads to

$$X_{t'}|X_t \sim \mathcal{N}(e^{-\frac{1}{2}|t-t'|}X_t, 1-e^{-|t-t'|})$$

Deriving $p(x_{t'}|x_t)$

Conditioning leads to

$$X_{t'}|X_t \sim \mathcal{N}(e^{-\frac{1}{2}|t-t'|}X_t, 1 - e^{-|t-t'|})$$

- Let $|t - t'| \to \infty$

Deriving $p(x_{t'}|x_t)$

Conditioning leads to

$$X_{t'}|X_t \sim \mathcal{N}(e^{-\frac{1}{2}|t-t'|}X_t, 1 - e^{-|t-t'|})$$

- Let $|t - t'| \to \infty$
- Then $e^{-\frac{1}{2}|t-t'|} \to 0$
- Approaches marginal $\mathcal{N}(0, 1)$.

OU is a Markov Process

For a set $S = \{t_1, \ldots, t_T\}$, ordered such that $t_1 < \ldots, t_T$ we can write the marginal as

$$p(x_1, x_2, \ldots, x_T | S) = p(x_1)p(x_2|x_1)\ldots p(x_T|x_{T-1})$$

With

$$p(x_j|x_{j-1}) = \mathcal{N}\left(x_j; e^{-\frac{1}{2}|t_j - t_{j-1}|}X_{j-1}, 1 - e^{-|t_j - t_{j-1}|}\right)$$

(This is assignment work!)

A different Parameterization

For a set $S = \{t_1, \dots, t_T\}$, ordered such that $t_1 < \dots, t_T$ we can write the marginal as

$$p(x_1, x_2, \dots, x_T | S) = p(x_1)p(x_2|x_1)\dots p(x_T|x_{T-1})$$

With

$$p(x_j|x_{j-1}) = \mathcal{N}\left(x_j; e^{-\frac{1}{2}|t_j - t_{j-1}|}X_{j-1}, 1 - e^{-|t_j - t_{j-1}|}\right)$$

Reparameterize with $\beta_t = 1 - e^{-|t_j - t_{j-1}|}$

A different Parameterization

For a set $S = \{t_1, \ldots, t_T\}$, ordered such that $t_1 < \ldots, t_T$ we can write the marginal as

$$p(x_1, x_2, \ldots, x_T | S) = p(x_1)p(x_2|x_1) \ldots p(x_T|x_{T-1})$$

With

$$p(x_j|x_{j-1}) = \mathcal{N}\left(x_j; \sqrt{1-\beta_j}X_{j-1}, \beta_t\right)$$

Reparameterize with $\beta_j = 1 - e^{-|t_j - t_{j-1}|}$

The marginal can be written without t_j

A different Parameterization

- Instead of choosing $S = \{t_0, \ldots, t_T\}$
- We can choose $0 < \beta_j < 1$, $j = 1 \ldots, T$
- Assuming t_0 is fixed and $t_{j-1} < t_j$, both formulations are equivalent

A different Parameterization

- Instead of choosing $S = \{t_0, \ldots, t_T\}$
- We can choose $0 < \beta_j < 1$, $j = 1 \ldots, T$
- Assuming t_0 is fixed and $t_{j-1} < t_j$, both formulations are equivalent
- Proof:

$$
\begin{aligned}
\beta_t &= 1 - e^{-|t_j - t_{j-1}|} \\
\Leftrightarrow t_j &= t_{j-1} - \log(1 - \beta_j) \\
\Leftrightarrow t_j &= \ldots \\
\Leftrightarrow t_j &= t_0 - \sum_{l=1}^{j} \log(1 - \beta_l)
\end{aligned}
$$

Marginal in β parameterisation

- What is $X_j|X_0$ with β parameters?

Marginal in β parameterisation

- What is $X_j | X_0$ with β parameters?
- We have $t_j = t_0 - \sum_{l=1}^{j} \log(1 - \beta_l)$

Marginal in β parameterisation

- What is $X_j|X_0$ with β parameters?
- We have $t_j = t_0 - \sum_{l=1}^{j} \log(1-\beta_l)$
- The marginal is

$$X_j|X_0 \sim \mathcal{N}\left(\sqrt{\bar{\alpha}_j}X_0, 1-\bar{\alpha}_j\right)$$
$$\bar{\alpha}_j = e^{-|t_j-t_0|} = \prod_{i=1}^{j}(1-\beta_i) \quad .$$

The OU process as a diffusion Process

- Diffusion: the slow mixing of atoms/molecules in a fluid over time
- For example: ink in water diffuses until the water is equally colored
- Diffusion in statistics: an initial complex distribution becomes similar to a simple distribution over time

The OU process as a diffusion Process

- OU process properties:
  - As time goes on, $X_j|X_0 \to X_j$
  - Due to Markov property $p(X_j|X_{j-1})$ each trajectory follows the same rules
    - The mean is shrunk a bit
    - pure noise is added to make up for the lost scale
    - Models random particle movement with drift towards zero
  - This means we forget the starting points after a while as noise gradually covers all information
- Next up: using OU processes for generative models

The OU process as a diffusion Process