Random Fields

Oswin Krause, PML, 2022

UNIVERSITY OF COPENHAGEN

Random Functions

- In ML, we search for optimal models
  - We pick a model class f_θ with parameters θ
  - Then find optimal θ given data.
  - E.g., θ maximizing a-posteriori probability of the labels given the data

Random Functions

- In ML, we search for optimal models
  - We pick a model class f_θ with parameters θ
  - Then find optimal θ given data.
  - E.g., θ maximizing a-posteriori probability of the labels given the data
- That feels indirect

Random Functions

- In ML, we search for optimal models
  - We pick a model class f_θ with parameters θ
  - Then find optimal θ given data.
  - E.g., θ maximizing a-posteriori probability of the labels given the data
- That feels indirect
- Can't we directly assign probabilities to functions in a function class?

Random Functions

- In ML, we search for optimal models
  - We pick a model class f_θ with parameters θ
  - Then find optimal θ given data.
  - E.g., θ maximizing a-posteriori probability of the labels given the data
- That feels indirect
- Can't we directly assign probabilities to functions in a function class?
- This could allow us to search in "large" classes (models with infinite parameters)

Random Functions: It is complicated.

- Consider functions $f : \mathbb{N} \to \mathbb{R}$
- These are sequences: $f(i) = f_i$, $i = 1, 2, \ldots$

Random Functions: It is complicated.

- Consider functions $f : \mathbb{N} \to \mathbb{R}$
- These are sequences: $f(i) = f_i$, $i = 1, 2, \dots$
- Let us pick a distribution

$$f_i \sim \mathcal{N}(0, 0.01)$$

- What is $p(f_1, f_2, \dots)$?

Random Functions: It is complicated.

- Consider functions $f : \mathbb{N} \to \mathbb{R}$
- These are sequences: $f(i) = f_i$, $i = 1, 2, \ldots$
- Let us pick a distribution

$$f_i \sim \mathcal{N}(0, 0.01)$$

- What is $p(f_1, f_2, \ldots)$?
- For finite sets, we have:

$$p(f_1, f_2, \ldots, f_\ell) = \prod_{i=1}^{\ell} \mathcal{N}(f_i, 0, 0.01)$$

Random Functions: It is complicated.

- We have $p(f_1, f_2, \ldots, f_\ell) = \prod_{i=1}^{\ell} \mathcal{N}(f_i; 0, 0.01)$

  What happens for $\ell \to \infty$?

Random Functions: It is complicated.

- We have $p(f_1, f_2, \ldots, f_\ell) = \prod_{i=1}^{\ell} \mathcal{N}(f_i; 0, 0.01)$

  What happens for $\ell \to \infty$?
- For $f_i = 1$ we have

$$p(1, 1, \ldots) = \prod_{i=1}^{\ell} \underbrace{\mathcal{N}(1; 0, 0.01)}_{<1} \xrightarrow{\ell \to \infty} 0$$

Random Functions: It is complicated.

- We have $p(f_1, f_2, \ldots, f_\ell) = \prod_{i=1}^{\ell} \mathcal{N}(f_i; 0, 0.01)$
  What happens for $\ell \to \infty$?
- For $f_i = 1$ we have

$$p(1, 1, \ldots) = \prod_{i=1}^{\ell} \underbrace{\mathcal{N}(1; 0, 0.01)}_{<1} \xrightarrow{\ell\to\infty} 0$$

- For $f_i = 0$, we have

$$p(0, 0, \ldots) = \prod_{i=1}^{\ell} \mathcal{N}(0; 0, 0.01) = \left(\frac{10}{\sqrt{2\pi}}\right)^{\ell} \xrightarrow{\ell\to\infty} \infty$$

Random Functions: It is complicated.

Final Example: Probability Integrals are all zero.

$$P(F_1 < u_i, F_2 < u_2, \dots) = \lim_{\ell \to \infty} \prod_{i=1}^{\ell} \underbrace{\int_{-\infty}^{u_i} \mathcal{N}(f_i; 0, 0.01)\, df_i}_{<1} = 0 \ .$$

- Integration in infinite dimensional spaces does not work.
- We need new tools!

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 $x \in \mathcal{X}$

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

- 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*
- Example:
  - $f \in \mathbb{R}^N \sim \mathcal{N}(\mu, \Sigma)$
  - $\mathcal{X} = \{1, 2, \dots, N\}$
  - Then, $f_x, x \in \mathcal{X}$ is a random field (just the indexed vector elements of f)

Example

- Remember earlier:
  - $f : \mathbb{N} \to \mathbb{R}$
  - With distribution $f_\ell \sim \mathcal{N}(0, 0.01)$
- This is a random field over index set $\mathcal{X} = \mathbb{N}$
- Real-valued, real-world example: The wave height of the ocean at any point

Random Fields and Marginals

- Random Field $F_x \in \Omega$ with observations f_x, $\forall x \in \mathcal{X}$
- Idea to save probabilities
  - We can have infinitely many random variables
  - But only observe a finite set at any time
  - Means: We can draw a random function but only observe its value at finitely many pre-selected points.

Random Fields and Marginals

- Random Field $F_x \in \Omega$ with observations f_x, $\forall x \in \mathcal{X}$
- Idea to save probabilities
  - We can have infinitely many random variables
  - But only observe a finite set at any time
  - Means: We can draw a random function but only observe its value at finitely many pre-selected points.
- Marginals: distribution of observed variables
  - Pick any finite subset $S_\ell = \{x_1, \ldots, x_\ell\} \subseteq \mathcal{X}$
  - Marginal: $p(f_{x_1}, f_{x_2}, \ldots, f_{x_\ell})$
  - Notation: $p(f_1, \ldots, f_\ell | S_\ell) = p(f_{x_1}, \ldots, f_{x_\ell})$

Defining Property of Random Fields

- Marginal $p(f_1, \ldots, f_\ell | S_\ell) = p(f_{x_1}, \ldots, f_{x_\ell})$
- It is easy to define a set of marginals.
- Open Questions:
  - Can we define marginals arbitrarily?
  - Is there more to a random field than its marginals?

Defining Property of Random Fields

- Marginal $p(f_1, \dots, f_\ell | S_\ell) = p(f_{x_1}, \dots, f_{x_\ell})$
- It is easy to define a set of marginals.
- Open Questions:
  - Can we define marginals arbitrarily?
  - Is there more to a random field than its marginals?
- Kolmogorov Consistency Theorem:
  - The marginals $p(f_1, \dots, f_\ell | S_\ell)$ for all sets $S_\ell \subset \mathcal{X}$, $\forall \ell \in \mathbb{N}$ uniquely define a random field
  - The marginals must be *consistent*

Consistency of Marginals

- Let $S = \{x_1, \ldots, x_\ell\} \subseteq \mathcal{X}$
- Consistency

Consistency of Marginals

- Let $S = \{x_1, \ldots, x_\ell\} \subseteq \mathcal{X}$
- Consistency
  - Let $T \subset S$

Consistency of Marginals

- Let $S = \{x_1, \ldots, x_\ell\} \subseteq \mathcal{X}$
- Consistency
  - Let $T \subset S$
  - Partition $(f_{x_1}, \ldots, f_{x_\ell}) = (f_T, f_C)$
  - f_T vector of variables indexed by T
  - f_C vector of variables only indexed by S and not T

Consistency of Marginals

- Let $S = \{x_1, \ldots, x_\ell\} \subseteq \mathcal{X}$
- Consistency
  - Let $T \subset S$
  - Partition $(f_{x_1}, \ldots, f_{x_\ell}) = (f_T, f_C)$
  - f_T vector of variables indexed by T
  - f_C vector of variables only indexed by S and not T
  - If for all S, T

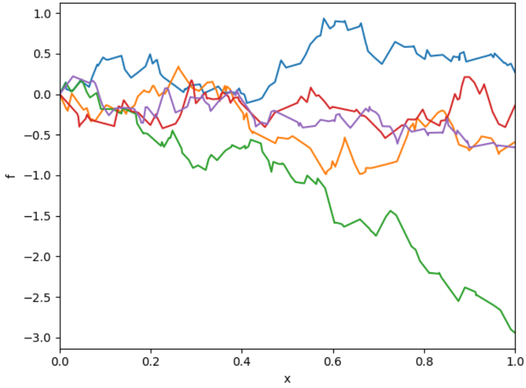
$$\underbrace{p(f_T|T)}_{\text{Marginal generated by } T} = \underbrace{\int p(f_T, f_C|S)\, df_C}_{\text{explicit integration of all variables in } S\setminus T}$$

Then $p(f_1, \ldots, f_\ell|S)$ is consistent

Example: Wiener Process

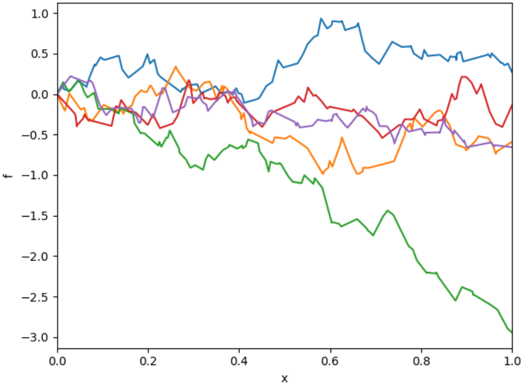
Wiener Process: Definition

- Take points
  $S = \{x_1, \ldots, x_\ell\} \subseteq [0, 1]$,
  $0 = x_0 < x_1 < x_2 < \cdots < x_\ell$
- Observation f_i at point x_i



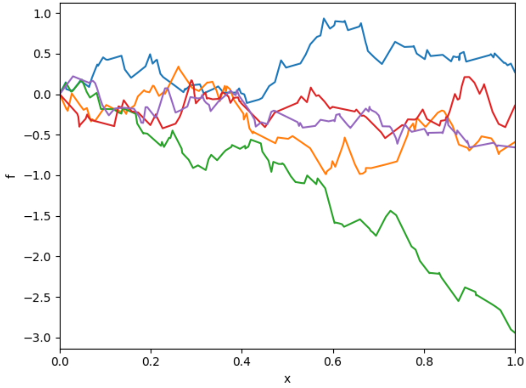
Wiener Process: Definition

- Take points
  $S = \{x_1, \dots, x_\ell\} \subseteq [0,1]$,
  $0 = x_0 < x_1 < x_2 < \cdots < x_\ell$
- Observation f_i at point x_i
- Sample f_i as
  - $f_0 = 0$
  - $f_{i+1} = f_i + W_{i+1}$,
    $W_{i+1} \sim \mathcal{N}(0, x_{i+1} - x_i)$
- Then f follows a Wiener Process

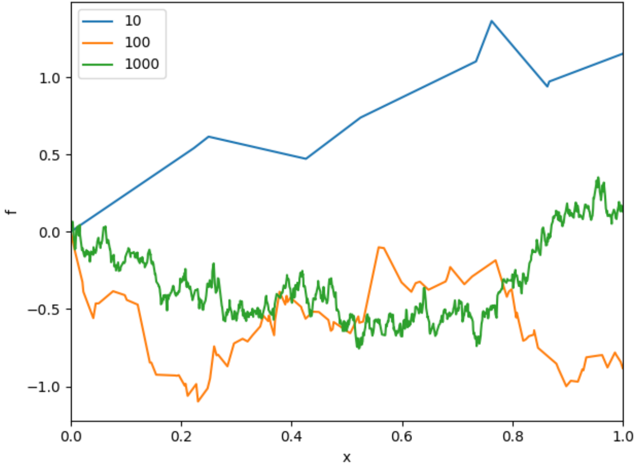


Wiener Process: Definition

- Take points
  $S = \{x_1, \ldots, x_\ell\} \subseteq [0, 1]$,
  $0 = x_0 < x_1 < x_2 < \cdots < x_\ell$
- Observation f_i at point x_i
- Sample f_i as
  - $f_0 = 0$
  - $f_{i+1} = f_i + W_{i+1}$,
    $W_{i+1} \sim \mathcal{N}(0, x_{i+1} - x_i)$
- Then f follows a Wiener Process
- Marginal pdf:
  $p(f_0, \ldots, f_\ell | S) = \prod_{i=0}^{\ell-1} p(f_{i+1} | f_i, S)$
  $p(f_{i+1} | f_i, S) = \mathcal{N}(f_{i+1}; f_i, x_{i+1} - x_i)$



Samples for different number of points ℓ



Wiener Process: Marginal distribution

What is the distribution of $p(f|S) = p(f_1, \dots f_\ell|S)$?

Wiener Process: Marginal distribution

What is the distribution of $p(f|S) = p(f_1, \ldots f_\ell|S)$?
We have

- $f_i = f_{i-1} + W_i,\ W_i \sim \mathcal{N}(0, x_i - x_{i-1})$

Wiener Process: Marginal distribution

What is the distribution of $p(f|S) = p(f_1, \dots f_\ell|S)$?
We have

- $f_i = f_{i-1} + W_i,\ W_i \sim \mathcal{N}(0, x_i - x_{i-1})$
- By repeated application

$$f_1 = \underbrace{f_0}_{0} + W_1 = W_1$$

Wiener Process: Marginal distribution

What is the distribution of $p(f|S) = p(f_1, \dots f_\ell|S)$?
We have

- $f_i = f_{i-1} + W_i,\ W_i \sim \mathcal{N}(0, x_i - x_{i-1})$
- By repeated application

$$f_2 = f_1 + W_2 = W_1 + W_2$$

Wiener Process: Marginal distribution

What is the distribution of $p(f|S) = p(f_1, \dots f_\ell|S)$?
We have

- $f_i = f_{i-1} + W_i$, $W_i \sim \mathcal{N}(0, x_i - x_{i-1})$
- By repeated application

$$f_i = \sum_{j=1}^{i} W_j$$

Wiener Process: Marginal distribution

What is the distribution of $p(f|S) = p(f_1, \ldots f_\ell|S)$?
We have

- $f_i = f_{i-1} + W_i,\ W_i \sim \mathcal{N}(0, x_i - x_{i-1})$
- By repeated application

$$f_i = \sum_{j=1}^{i} W_j$$

- In matrix form

$$\begin{bmatrix} f_1 \\ f_2 \\ \ldots \\ f_\ell \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & 0 \\ 1 & 1 & 1 & \ldots & 1 \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \ldots \\ W_\ell \end{bmatrix}$$

Marginal distribution

What is the distribution of $p(f|S)$?

- We have

$$\begin{bmatrix} f_1 \\ f_2 \\ \dots \\ f_\ell \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & 0 \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \dots \\ W_\ell \end{bmatrix}$$

Marginal distribution

What is the distribution of $p(f|S)$?

- We have

$$\underbrace{\begin{bmatrix} f_1 \\ f_2 \\ \dots \\ f_\ell \end{bmatrix}}_{f} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & 0 \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} W_1 \\ W_2 \\ \dots \\ W_\ell \end{bmatrix}}_{W}$$

Marginal distribution

What is the distribution of $p(f|S)$?

- We have

$$f = AW$$

$$W \sim \mathcal{N}\left(0, \underbrace{\begin{bmatrix} x_1 - x_0 & 0 & 0 & \dots & 0 \\ 0 & x_2 - x_1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & 0 \\ 0 & 0 & 0 & \dots & x_\ell - x_{\ell-1} \end{bmatrix}}_{D}\right)$$

Marginal distribution

What is the distribution of $p(f|S)$?

- We have

$$f = AW, \; W \sim \mathcal{N}(0, D)$$

Marginal distribution

What is the distribution of $p(f|S)$?

- We have
$$f = AW, \; W \sim \mathcal{N}(0, D)$$
- f is multivariate normal with
$$f \sim \mathcal{N}(0, ADA^T)$$

Marginal distribution

- We have

$$K = ADA^T$$

Marginal distribution

- We have

$$K = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & 0 \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} x_1 - x_0 & 0 & 0 & \dots & 0 \\ 0 & x_2 - x_1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & 0 \\ 0 & 0 & 0 & \dots & x_\ell - x_{\ell-1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \dots & 1 \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

Marginal distribution

- We have

$$K = \begin{bmatrix} x_1 - x_0 & 0 & 0 & \dots & 0 \\ x_1 - x_0 & x_2 - x_1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & 0 \\ x_1 - x_0 & x_2 - x_1 & x_3 - x_2 & \dots & x_\ell - x_{\ell-1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \dots & 1 \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

Marginal distribution

- We have

$$K = \begin{bmatrix} x_1 - x_0 & 0 & 0 & \dots & 0 \\ x_1 - x_0 & x_2 - x_1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & 0 \\ x_1 - x_0 & x_2 - x_1 & x_3 - x_2 & \dots & x_\ell - x_{\ell-1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \dots & 1 \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

- The entries are

$$K_{ij} = \sum_{l=1}^{i} (x_l - x_{l-1}) \mathbb{I}\{l \le j\}$$

Marginal distribution

- We have

$$K = \begin{bmatrix} x_1 - x_0 & 0 & 0 & \dots & 0 \\ x_1 - x_0 & x_2 - x_1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & 0 \\ x_1 - x_0 & x_2 - x_1 & x_3 - x_2 & \dots & x_\ell - x_{\ell-1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \dots & 1 \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

- The entries are

$$\begin{aligned} K_{ij} &= \sum_{l=1}^{i} (x_l - x_{l-1}) \mathbb{I}\{l \leq j\} \\ &= \sum_{l=1}^{\min\{i,j\}} (x_l - x_{l-1}) \\ &= x_{\min\{i,j\}} - x_0 = \min\{x_i, x_j\} \end{aligned}$$

Marginal distribution

- We have

$$K = \begin{bmatrix} x_1 & x_1 & x_1 & \dots & x_1 \\ x_1 & x_2 & x_2 & \dots & x_2 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ x_1 & x_2 & x_3 & \dots & x_\ell \end{bmatrix}$$

- The entries are

$$\begin{aligned} K_{ij} &= \sum_{l=1}^{i} (x_l - x_{l-1}) \mathbb{I}\{l \leq j\} \\ &= \sum_{l=1}^{\min\{i,j\}} (x_l - x_{l-1}) \\ &= x_{\min\{i,j\}} - x_0 = \min\{x_i, x_j\} \end{aligned}$$

Marginal distribution

What is the distribution of $p(f|S)$?
Our answer is

Marginal distribution

What is the distribution of $p(f|S)$?

Our answer is

- For a set $S = \{x_1, x_2, \ldots, x_\ell\}$

$$p(f|S) = \mathcal{N}(f; 0, K(S))$$

- With entries $K(S)_{ij} = \min\{x_i, x_j\}$

Marginal distribution

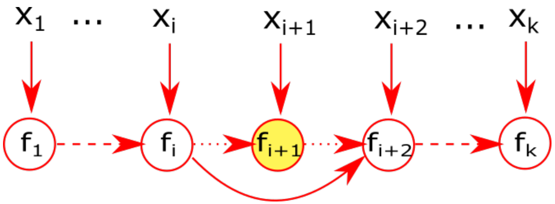
What is the distribution of $p(f|S)$?
Our answer is

- For a set $S = \{x_1, x_2, \ldots, x_\ell\}$

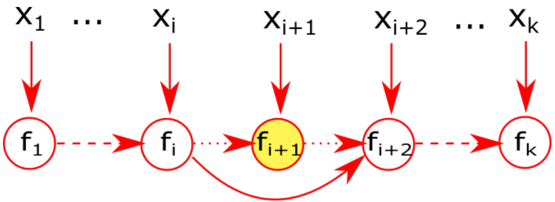
$$p(f|S) = \mathcal{N}(f; 0, K(S))$$

- With entries $K(S)_{ij} = \min\{x_i, x_j\}$
- Is this set of marginals consistent?

Kolmogorov Consistency

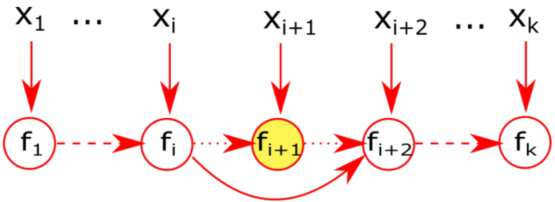


Kolmogorov Consistency



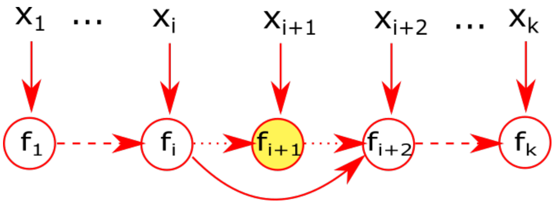
- $T = \{x_1, \ldots, x_i, x_{i+2}, \ldots, x_\ell\}$

Kolmogorov Consistency



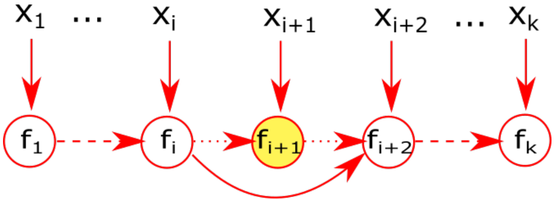
- $T = \{x_1, \ldots, x_i, x_{i+2}, \ldots, x_\ell\}$
- Is $p(f_1, \ldots, f_i, f_{i+2}, \ldots f_\ell | S) = p(f_1, \ldots, f_i, f_{i+2}, \ldots f_\ell | T)$?

Kolmogorov Consistency



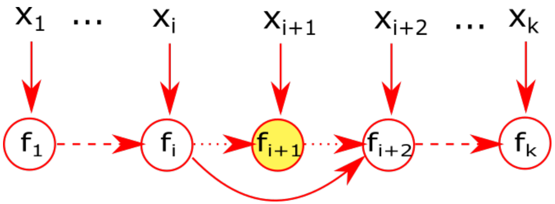
- $T = \{x_1, \ldots, x_i, x_{i+2}, \ldots, x_\ell\}$
- Is $p(f_1, \ldots, f_i, f_{i+2}, \ldots f_\ell | S) = p(f_1, \ldots, f_i, f_{i+2}, \ldots f_\ell | T)$?
- We know $p(f|S) = \mathcal{N}(f; 0, K(S))$

Kolmogorov Consistency



- $T = \{x_1, \ldots, x_i, x_{i+2}, \ldots, x_\ell\}$
- Is $p(f_1, \ldots, f_i, f_{i+2}, \ldots f_\ell | S) = p(f_1, \ldots, f_i, f_{i+2}, \ldots f_\ell | T)$?
- We know $p(f|S) = \mathcal{N}(f; 0, K(S))$
- Easy to see:
    - Marginal removes $i+1$th row/column from $K(S)$
    - Same as $K(T)$

Kolmogorov Consistency



- $T = \{x_1, \dots, x_i, x_{i+2}, \dots, x_\ell\}$
- Is $p(f_1, \dots, f_i, f_{i+2}, \dots f_\ell | S) = p(f_1, \dots, f_i, f_{i+2}, \dots f_\ell | T)$?
- We know $p(f|S) = \mathcal{N}(f; 0, K(S))$
- Easy to see:
  - Marginal removes $i+1$th row/column from $K(S)$
  - Same as $K(T)$
- Iteratively apply to generalize to any $T \subseteq S$

Bayesian Linear Regression as Random Process

Bayesian Linear Regression as Random Process

Bayesian Linear Regression with Gaussian prior

- Linear function $f_\theta(x) = \theta^T \phi(x)$
- $\phi(x) : \mathbb{R}^d \to \mathbb{R}^K$
- Prior $\theta \in \mathbb{R}^K \sim \mathcal{N}(0, \Sigma_\theta)$

This lecture: We are interested in distribution of sampled f_θ

Example: Sampling polynomials

Polynomial Features $k = 1, \dots, K$ for $x \in \mathbb{R}$

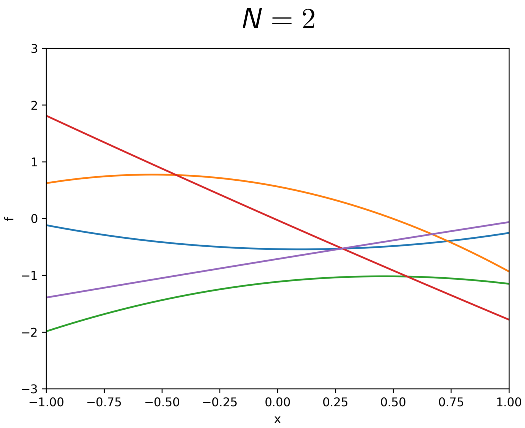
- $\phi_k(x) = x^{k-1}$
- $\phi(x) = (1, x, x^2, \dots)$

Prior

- $\theta_1 \sim \mathcal{N}(0, 1)$
- $\theta_k \sim \mathcal{N}\left(0, \frac{1}{(k-1)^2}\right), \; k > 1$

Sampled Polynomial

$$f(x) = \theta^T \phi(x) = \theta_1 + \theta_2 x + \theta_3 x^2 + \dots$$



Example: Sampling polynomials

Polynomial Features $k = 1, \ldots, K$ for $x \in \mathbb{R}$

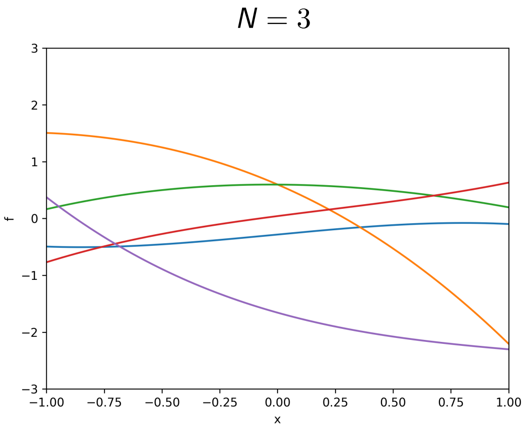
- $\phi_k(x) = x^{k-1}$
- $\phi(x) = (1, x, x^2, \ldots)$

Prior

- $\theta_1 \sim \mathcal{N}(0, 1)$
- $\theta_k \sim \mathcal{N}\left(0, \frac{1}{(k-1)^2}\right), \ k > 1$

Sampled Polynomial

$$f(x) = \theta^T \phi(x) = \theta_1 + \theta_2 x + \theta_3 x^2 + \ldots$$



Example: Sampling polynomials

Polynomial Features $k = 1, \ldots, K$ for $x \in \mathbb{R}$

- $\phi_k(x) = x^{k-1}$
- $\phi(x) = (1, x, x^2, \ldots)$

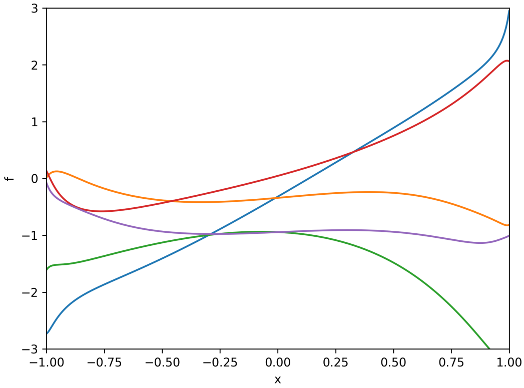
Prior

- $\theta_1 \sim \mathcal{N}(0, 1)$
- $\theta_k \sim \mathcal{N}\left(0, \frac{1}{(k-1)^2}\right),\ k > 1$

Sampled Polynomial

$$f(x) = \theta^T \phi(x) = \theta_1 + \theta_2 x + \theta_3 x^2 + \ldots$$

$N = 100$



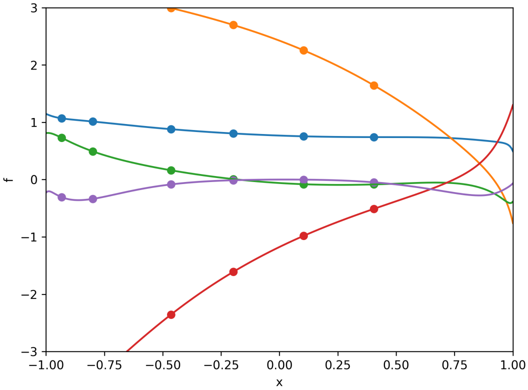
Bayesian Linear Regression as Random Process

If we take $\mathcal{X} = \mathbb{R}^d$ as index set and $x \in \mathcal{X}$, then

$$f_x = \theta^T \phi(x), \theta \sim p(\theta)$$

is a random process (random field).

Random Process view



- Functions f: created by random draws of θ
- Random Field: function values f_x at the marked positions

Marginals of the Process

We have:

- Process: $f_x = \theta^T\phi(x), \theta \sim \mathcal{N}(0, \Sigma_\theta)$
- Set $S = \{x_1, \ldots, x_\ell\}$

What is $p(f|S) = p(f_1, \ldots, f_\ell|S)$?

Marginals of the Process

We have:

- Process: $f_x = \theta^T \phi(x), \theta \sim \mathcal{N}(0, \Sigma_\theta)$
- Set $S = \{x_1, \dots, x_\ell\}$

What is $p(f|S) = p(f_1, \dots, f_\ell|S)$?

We have

$$f = \begin{bmatrix} f_1 \\ f_2 \\ \dots \\ f_\ell \end{bmatrix} = \begin{bmatrix} \phi(x_1)^T\theta \\ \phi(x_2)^T\theta \\ \dots \\ \phi(x_\ell)^T\theta \end{bmatrix} = \underbrace{\begin{bmatrix} \phi(x_1)^T \\ \phi(x_2)^T \\ \dots \\ \phi(x_\ell)^T \end{bmatrix}}_{\Phi(S)\in\mathbb{R}^{\ell\times N}} \theta = \Phi(S)\theta$$

Reminder: Linear Transformation of Multivariate normal random variables

Let $X \in \mathbb{R}^d \sim \mathcal{N}(\mu_x, \Sigma_X)$ and $Q \in \mathbb{R}^{N \times d}$ then

$$Z = QX$$

is a multivariate normal distributed variable and $Z \sim \mathcal{N}(Q\mu_X, Q\Sigma_X Q^T)$

Marginals of the Process

We have

- $f = \Phi(S)\theta$
- $\theta \sim \mathcal{N}(0, \Sigma_\theta)$

$$f \sim \mathcal{N}(0, \underbrace{\Phi(S)\Sigma_\theta\Phi(S)^T}_{K(S)})$$

with elements

$$K(S)_{ij} = \phi(x_i)^T\Sigma_\theta\phi(x_j)$$

Takeaway

- Probabilities on function spaces are difficult.
- Random Fields introduce consistent probabilities on subsets of observed function values
- Those estimates are integrals over many functions, all passing through the observations
- We can Phrase Bayesian Linear Regression as a random process