

# Monte Carlo methods

- ▶ Assume of interest

$$\mu = E^p[h(\mathbf{X})] = \int_{\mathbf{x}} h(\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

Can be **approximated** by

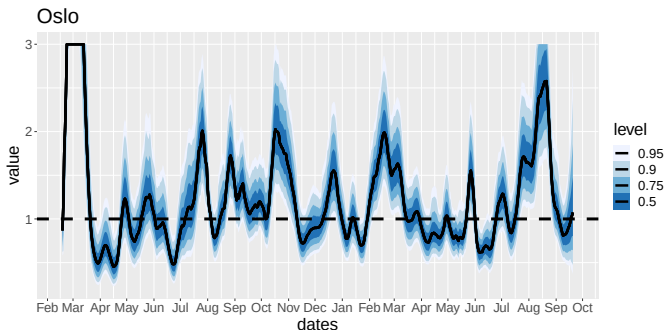
$$\hat{\mu} = \frac{1}{B} \sum_{b=1}^b h(\mathbf{X}^b), \quad \mathbf{X}^b \sim p(\mathbf{x})$$

- ▶ If  $X^1, \dots, X^B$  are iid,

$$\text{Var}[\hat{\mu}] = \frac{1}{B} \text{Var}[h(\mathbf{X}^b)]$$

- ▶ Easy if  $\mathbf{X}$  is low-dimensional

# Example Covid-19



- ▶ Aim: Estimate  $R_{1:T}$  based on data  $\mathbf{y}_{1:T} = (y_{1:T}^{hosp}, y_{1:T}^{test})$
- ▶ 550 days, 11 counties, 6050 variables
- ▶ Also how many susceptible, infected, asymptomatic, symptomatic, recovered,  $550 \times 11 \times 5 = 30250$  variables
- ▶ Some extra as well

# Bayesian neural networks

- ▶ **Neural networks**: Highly flexible, but many parameters to estimate  
⇒ high uncertainty
- ▶ **Bayesian approaches**: Take uncertainty into account

$$p(\mathbf{y}^*|\mathbf{x}^*, \mathcal{D}) = \int_{\mathbf{w}} p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{w})p(\mathbf{w}|\mathcal{D})d\mathbf{w}$$
$$\approx \frac{1}{B} \sum_{b=1}^B p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{w}^b), \quad \mathbf{w}^b \sim p(\mathbf{w}|\mathcal{D})$$

- ▶ Dimension of  $\mathbf{w}$  typically **huge**

# Monte Carlo in high dimension

- ▶ Aim: Simulate from  $p(\mathbf{x}|\mathbf{y})$
- ▶ **Makov chain Monte Carlo:**
  - ▶ Sample sequentially  $\mathbf{x}^b, b = 1, \dots, B$  such that  $\mathbf{x}^B \approx p(\mathbf{x}|\mathbf{y})$
  - ▶ May perform bad with large amounts of data: **Subsampling**
- ▶ **Sequential Monte Carlo:**
  - ▶ Utilise that  $\mathbf{x} = (x_1, \dots, x_T)$
  - ▶ Sequentially simulate  $x_t \sim p(x_t|y_{1:t})$
  - ▶ Also struggles in high dimension
  - ▶ Especially problematic if **static parameters** are to be estimated simultaneously

# Possible master projects

- ▶ Many new approaches for **more efficient Monte Carlo sampling**
- ▶ Starting point: Try out these approaches
- ▶ Further work:
  - ▶ Modify/improve existing approaches
  - ▶ Study theoretical properties of algorithms
    - ▶ Probability calculations
- ▶ Requirements
  - ▶ **STK4051 - Computational statistics (spring course)**
  - ▶ Will be helpful with **STK4021 - Applied Bayesian Analysis**