Monte Carlo Methods

MATH 565

Git website and repository

Canvas

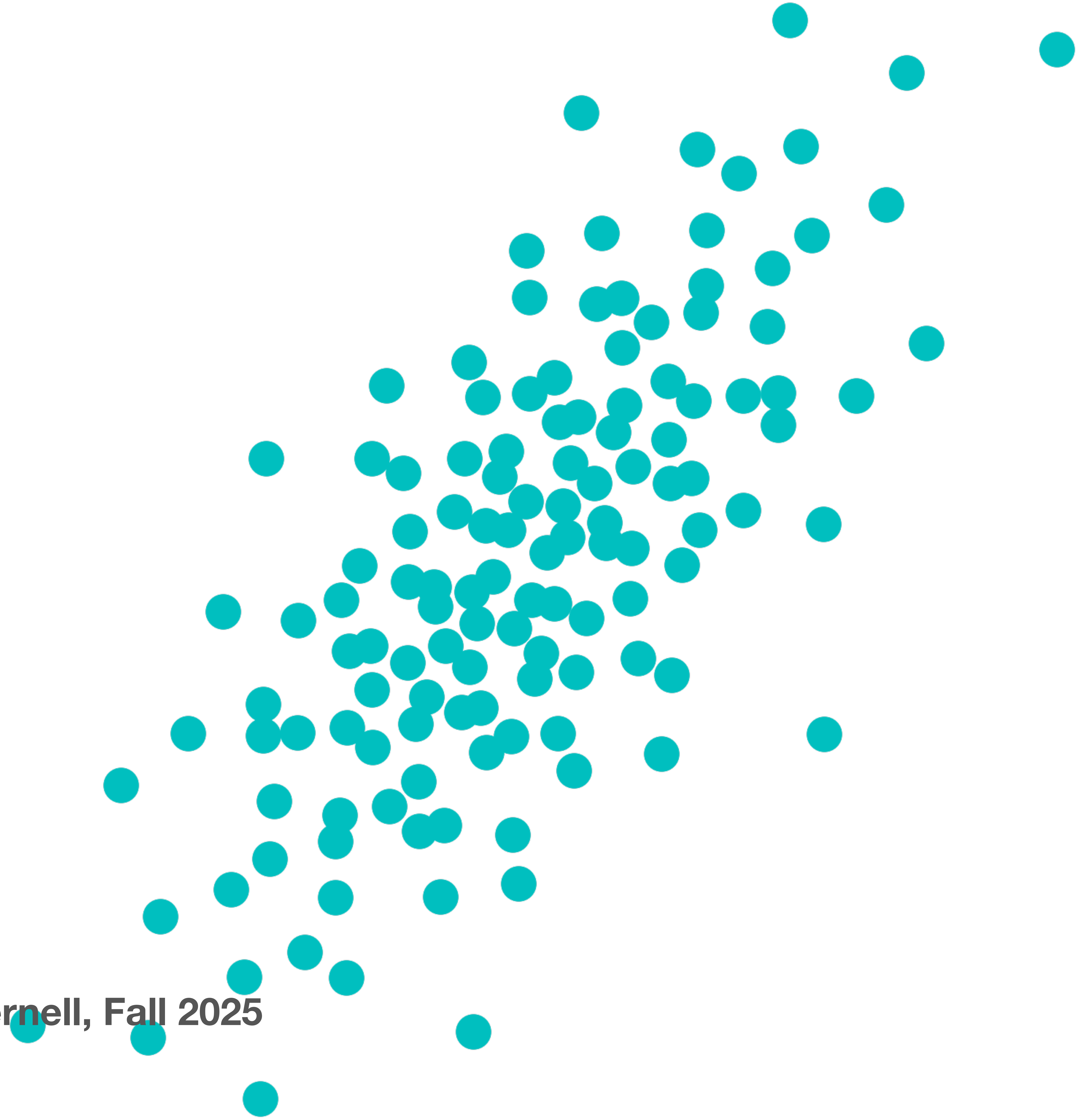
Fred Hickernell, Fall 2025

Updated 2025 August 25

Introduction

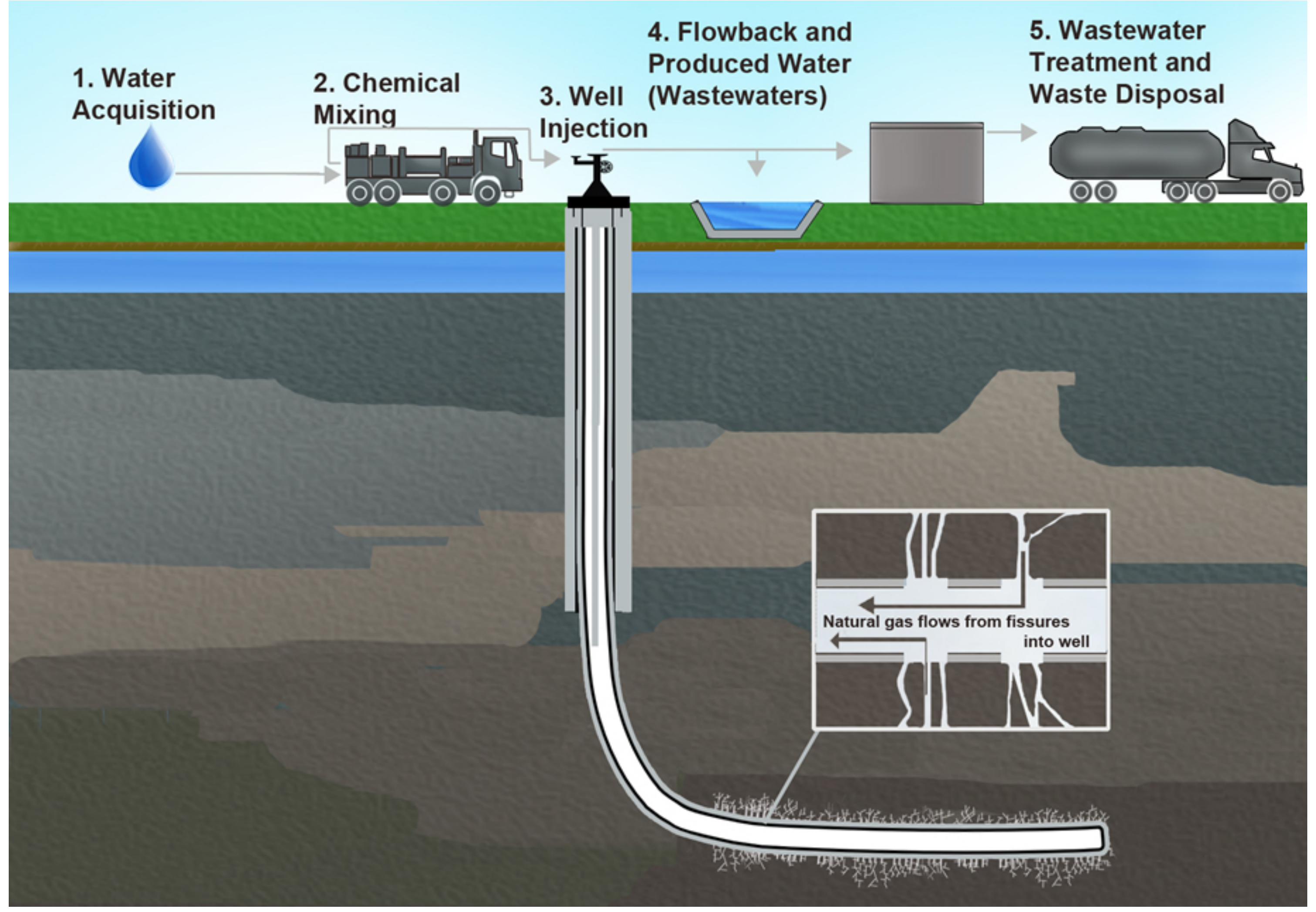
Owen, Chapters 1 & 2

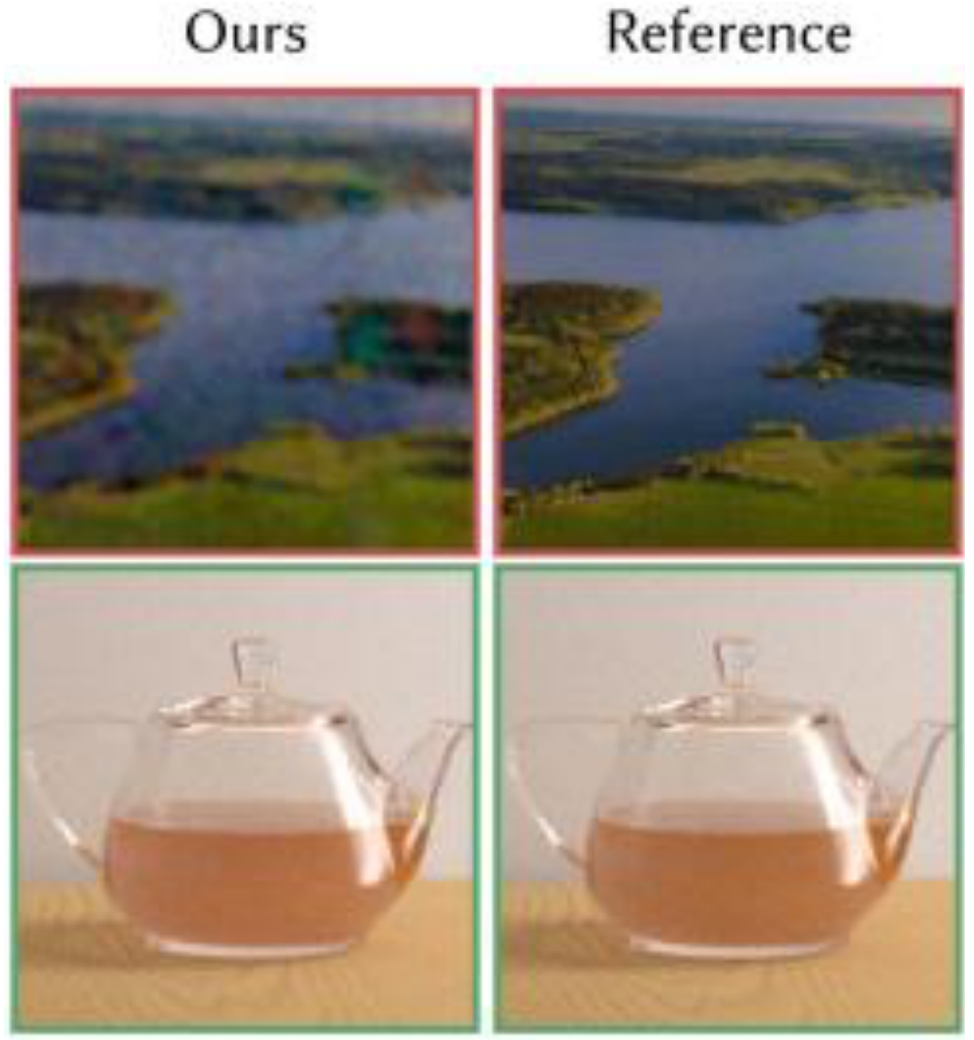
Assignment 1 Due Sep 5

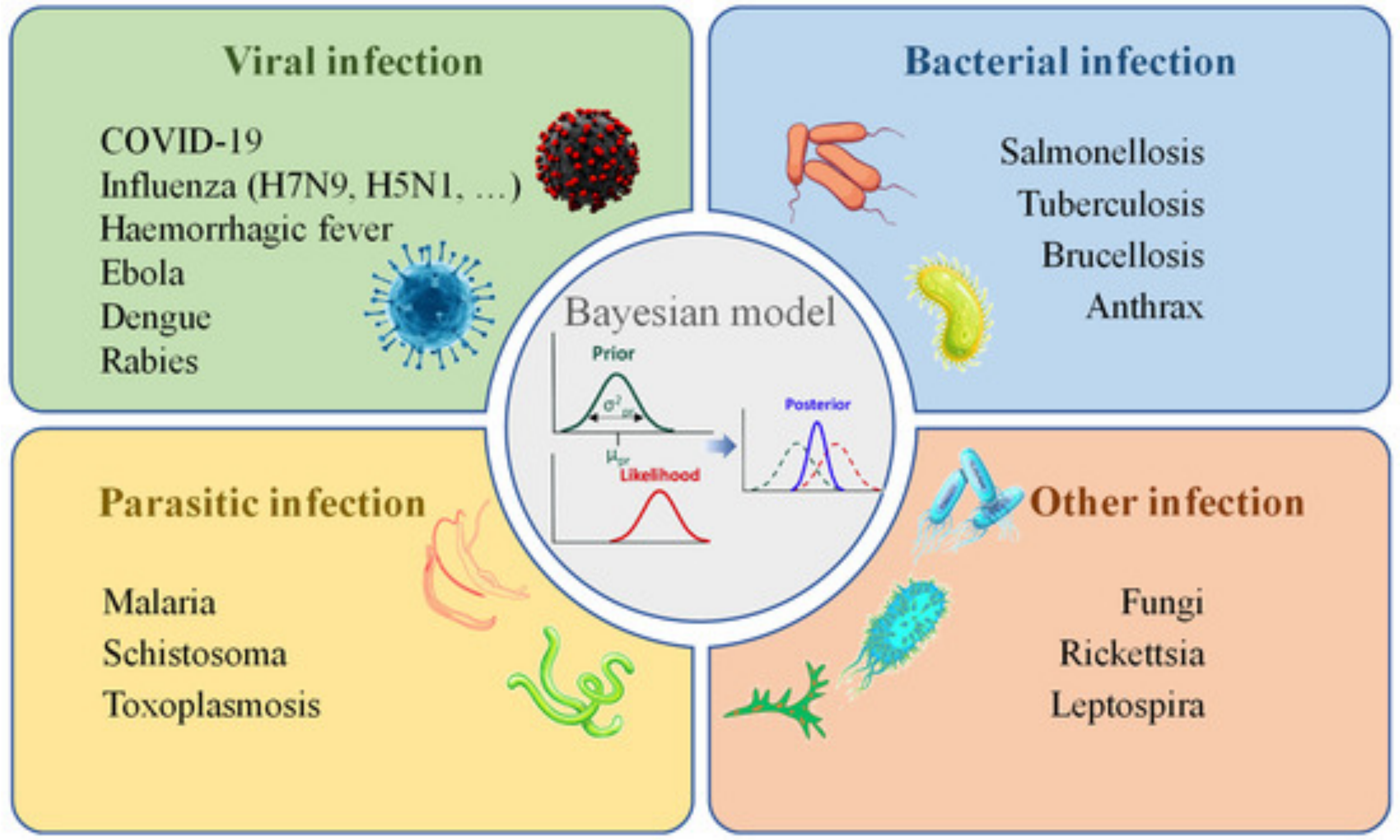


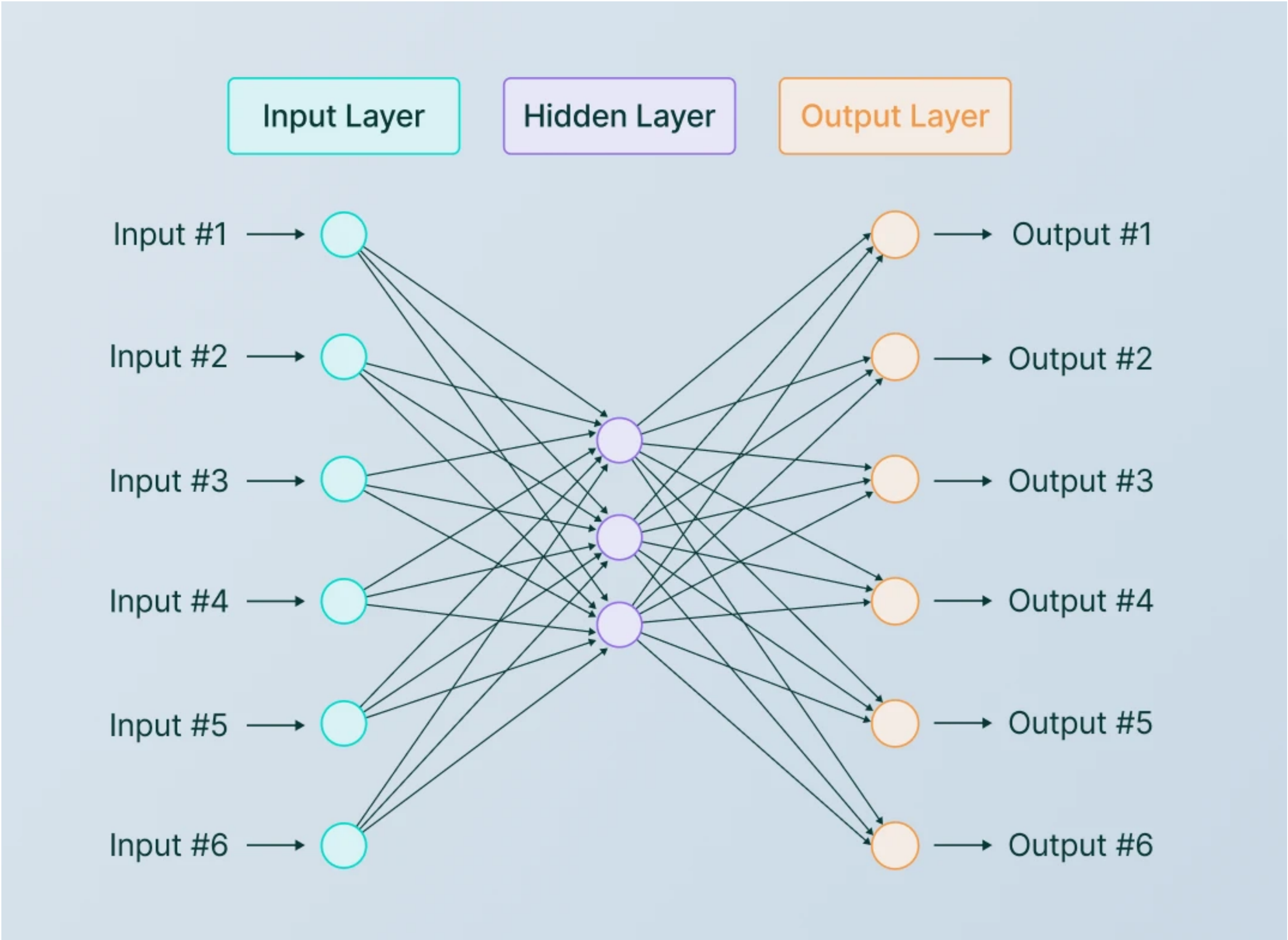
MATH 565 Monte Carlo Methods, Fred Hickernell, Fall 2025

Monte Carlo Helps With Uncertainty [WSJ 2016]









Why is there uncertainty?

Why is there uncertainty?

- Finance — market forces, often modeled by stochastic processes driven by Brownian motion

Why is there uncertainty?

- Finance — market forces, often modeled by stochastic processes driven by Brownian motion
- Engineering — variability of system parameters, sometimes modeled by Gaussian processes

Why is there uncertainty?

- Finance — market forces, often modeled by stochastic processes driven by Brownian motion

- Engineering — variability of system parameters, sometimes modeled by Gaussian processes

- Image rendering — can only trace some rays, must rely on a finite sample

Why is there uncertainty?

- Finance — market forces, often modeled by stochastic processes driven by Brownian motion

- Engineering — variability of system parameters, sometimes modeled by Gaussian processes

- Image rendering — can only trace some rays, must rely on a finite sample

- Bayesian inference — the posterior probability distribution is a combination of a prior and what is learned from data

Why is there uncertainty?

- Finance — market forces, often modeled by stochastic processes driven by Brownian motion
- Engineering — variability of system parameters, sometimes modeled by Gaussian processes
- Image rendering — can only trace some rays, must rely on a finite sample
- Bayesian inference — the posterior probability distribution is a combination of a prior and what is learned from data
- Neural networks — many parameters need to be tuned, but one cannot search in all possible directions

Why is there uncertainty?

- Finance — market forces, often modeled by stochastic processes driven by Brownian motion
- Engineering — variability of system parameters, sometimes modeled by Gaussian processes
- Image rendering — can only trace some rays, must rely on a finite sample
- Bayesian inference — the posterior probability distribution is a combination of a prior and what is learned from data
- Neural networks — many parameters need to be tuned, but one cannot search in all possible directions
- Queues — arrival times and service times of customers

How is this expressed quantitatively?

$$Y = \text{random variable denoting quantity of interest} = \begin{cases} \text{option payoff} \\ \text{fluid pressure} \\ \text{pixel intensity} \\ \text{statistical model parameter} \\ \text{neural network parameter} \\ \text{service time} \end{cases}$$

$$= f(\boldsymbol{X}), \text{ where}$$

$$\boldsymbol{X} = \text{multivariate random variable with a simpler distribution}$$

We want to estimate the mean, variance, quantile, or probability distribution of Y using sample versions

Are we there yet?

You are visiting your friend and it will require

- A 5 minute walk to the ‘L’ station
- Waiting for the train, which arrives every 20 minutes
- Traveling 35 minutes by ‘L’
- Catching a taxi at the ‘L’ destination
  - There is a 20% chance that the car is waiting for you
  - Otherwise the average wait time is 10 minutes
- A 12 minute taxi ride

How long should you plan for the trip to take?

Let’s look at this Jupyter Notebook AreWeThereYet on the class website

Let me know your degree program?

- Go to menti.com
- Use code 6222 6078

Why should you attend synchronously?

- You will be better keep pace?
- You will get real time answers to questions?
- You can influence the pace and direction of the course?
- You can help your peers learn and benefit from them—partake in a leaning community
- To help me know you, in case you want a reference, or to add me to your LinkedIn network

Review of probability

Let Y be a random variable with a sample space $\mathscr{Y} \subseteq \mathbb{R}$

$\mathbb{P}(Y \in \Omega)$ means the probability of the event Ω

F is the cumulative distribution function $F(y) := \mathbb{P}(Y \leq y), \; y \in \mathscr{Y}$

Q is the quantile function $Q(p) := \inf\{y \in \mathscr{Y} : F(y) \geq p\}, \quad 0 < p < 1$

$\mathbb{E}[g(Y)] := \int_{\mathscr{Y}} g(y)\,\mathrm{d}F(y)$ is the expectation of $g(Y)$

discrete	continuous
ϱ is the probability mass function	ϱ is the probability density function
$\varrho(y) := \mathbb{P}(Y = y)$	$\varrho(y) := F'(y)$
$\mathbb{E}[g(Y)] = \sum_{y \in \mathscr{Y}} g(y)$	$\mathbb{E}[g(Y)] = \int_{y \in \mathscr{Y}} g(y)\,\varrho(y)\,\mathrm{d}y$

Moments

$$\mu := \mathbb{E}(Y) \text{ is the mean of } Y$$

$$\sigma^2 = \mathrm{var}(Y) := \mathbb{E}[(Y - \mu)^2] \text{ is the variance of } Y$$

σ is the standard deviation

$$\mathrm{cov}(\boldsymbol{Y}) := \big(\mathbb{E}[(Y_j - \mu_j)(Y_k - \mu_k)]\big)_{j,k=1}^{d}$$

is the covariance matrix of the random vector $\boldsymbol{Y}$

Binomial Distribution

This is a coin-flipping distribution with the probability p of heads

$$Y \sim \text{Binomial}(n, p) \qquad \texttt{scipy.stats.binom(n, p)}$$

$$\varrho(y) := \mathbb{P}(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y = 0, 1, \ldots, n$$

$$F(y) = \mathbb{P}(Y \leq y) = \sum_{j=0}^{\lfloor y \rfloor} \binom{n}{j} p^j (1-p)^{n-j}$$

$$Q(p) = \inf\{y \in \{0, 1, \ldots, n\} : F(y) \geq p\}$$

$$\mu = \mathbb{E}(Y) = np \qquad \sigma^2 = \text{var}(Y) = np(1-p)$$

Multivariate Normal/Gaussian Distribution

Used in finance and uncertainty quantification

$Y \sim \mathcal{N}(\mu, \Sigma)$ is a d-dimensional random vector

`scipy.stats.multivariate_normal(mean, cov)`

$\mu =$ mean

$\Sigma =$ covariance matrix

$$\varrho(\boldsymbol{y}) := \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\tfrac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right)$$

$$F(\boldsymbol{y}) = \int_{(-\infty,\boldsymbol{y})} \varrho(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}$$

Sampling

Let Y be be a random variable, often $Y = f(\boldsymbol{X})$

$Y_1, Y_2, \ldots$ be a sample

not necessarily random or independent and identically distributed (IID)

$$\widehat{\mu}_n := \frac{1}{n} \sum_{i=1}^{n} Y_i \text{ is the sample mean}$$

$$\widehat{\sigma}_n^2 := \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \widehat{\mu}_n)^2 \text{ is the sample variance}$$

Under random sampling

Let Y be be a random variable, often $Y = f(\boldsymbol{X})$

$$Y_1, Y_2, \ldots \sim Y$$

$$\mathbb{E}(\widehat{\mu}_n) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Y_i) = \mu \text{ so } \widehat{\mu}_n \text{ is unbiased}$$

If in addition, $Y_1, Y_2, \ldots$ are uncorrelated, then

$$\mathbb{E}(\widehat{\sigma}_n^2) = \sigma^2 \text{ so } \widehat{\sigma}_n^2 \text{ is unbiased}$$

$$\mathrm{mse}(\widehat{\mu}_n) := \mathbb{E}[(\mu - \widehat{\mu}_n)^2] = \mathrm{var}(\widehat{\mu}_n) = \frac{\sigma^2}{n}, \qquad \mathrm{rmse}(\widehat{\mu}_n) = \frac{\sigma}{\sqrt{n}}$$

Central Limit Theorem

Let Y be be a random variable, often $Y = f(\boldsymbol{X})$

$$Y_1, Y_2, \ldots \overset{\text{IID}}{\sim} Y$$

The distribution of

$$\frac{\widehat{\mu}_n - \mu}{\sigma/\sqrt{n}} \text{ approaches } \mathcal{N}(0,1) \text{ as } n \to \infty$$

which means that

$$\widehat{\mu}_n \mathrel{\dot{\sim}} \mathcal{N}(\mu, \sigma^2/n) \text{ for large } n$$

Monte Carlo estimates of properties of Y

Let Y be a random variable and let $Y_1, Y_2, \ldots$ be an IID random sample

Population Quantity	Sample Approximation
mean μ	$\widehat{\mu}_n \pm 2.58\widehat{\sigma}^2_{n_0}$, `np.mean(data)`
variance σ^2	$\widehat{\sigma}^2_n$, `np.var(data, ddof=1)`
cumulative distribution function F	empirical distribution function $\widehat{F}_n$ `statsmodels.distributions.empirical_distribution.ECDF(data)`
quantile function Q	empirical quantile function $\widehat{Q}_n$ `np.quantile(data, probabilities)`
probability density or mass function ϱ	historgram, `np.histogram(data, bins = "auto")` kernel density estimator, `scipy.stats.gaussian_kde(data)`

For μ we have an interval estimator, but for the others typically point estimators

Are we there yet?

You are visiting your friend and it will require

- A 5 minute walk to the 'L' station
- Waiting for the train, which arrives every 20 minutes
- Traveling 35 minutes by 'L'
- Catching a taxi at the 'L' destination
  - There is a 20% chance that the car is waiting for you
  - Otherwise the average wait time is 10 minutes
- A 12 minute taxi ride

How long should you plan for the trip to take?

Let's look at this Jupyter Notebook AreWeThereYet on the class website

Let me know your professional aspirations?

- Go to menti.com
- Use code 6222 6078

Conditional Probability

Let (Y, Z) be be a vector of random variables

$$\mathbb{P}(Y = y \mid Z = z) = \frac{\mathbb{P}(Y = y \,\&\, Z = z)}{\mathbb{P}(Z = z)}$$

$$\mathbb{P}(Y = y \mid Z = z) = \frac{\mathbb{P}(Z = z \mid Y = y)\mathbb{P}(Y = y)}{\mathbb{P}(Z = z)}$$ Bayes' Theorem

$$\varrho_{Y|Z}(y \mid z) = \frac{\varrho_{Y,Z}(y, z)}{\varrho_Z(z)}$$

$$\mathbb{E}(Y) = \mathbb{E}_Z\big[\mathbb{E}_Y(Y \mid Z)\big]$$

$$\mathrm{var}(Y) = \mathbb{E}_Z[\mathrm{var}_Y(Y \mid Z)] + \mathrm{var}_Z[\mathbb{E}_Y(Y \mid Z)]$$

Conditional Monte Carlo

Let $(Y = f(\boldsymbol{X}_{1:d}), \boldsymbol{X}_{1:d})$ be be a vector of random variables

Let $\boldsymbol{Z} = \boldsymbol{X}_{2:d}$

If $g(\boldsymbol{X}_{2:d}) = \mathbb{E}_Y(Y \mid \boldsymbol{X}_{2:d})$ has an analytic form, then

$\mu = \mathbb{E}(Y) = \mathbb{E}_{\boldsymbol{X}_{2:d}}[g(\boldsymbol{X}_{2:d})]$ and $\operatorname{var}(Y) \geq \operatorname{var}_Z[g(\boldsymbol{X}_{2:d})]$

If $h(y, \boldsymbol{X}_{2:d}) = \varrho_{Y|\boldsymbol{X}_{2:d}}(y \mid \boldsymbol{x}_{2:d})$ has an analytic form, then

$\varrho_Y(y) = \mathbb{E}_{\boldsymbol{X}_{2:d}}[h(y, \boldsymbol{X}_{2:d})]$ probability density expressed as expectation

Summary of the Introduction

- Sample quantities, often using IID samples, used to estimate population properties of a random variable with a complex distribution, including
  - Means, variances, covariances
  - Probability density/mass functions, cumulative distribution functions, and quantile functions
- For means we have interval estimators, which indicate the uncertainty in our estimates, using the CLT
- For the other properties we have point estimators, i.e., only one approximate value for each thing to be estimated