Monte Carlo Methods

Introduction

Generating Samples

Markov Chain Monte Carlo

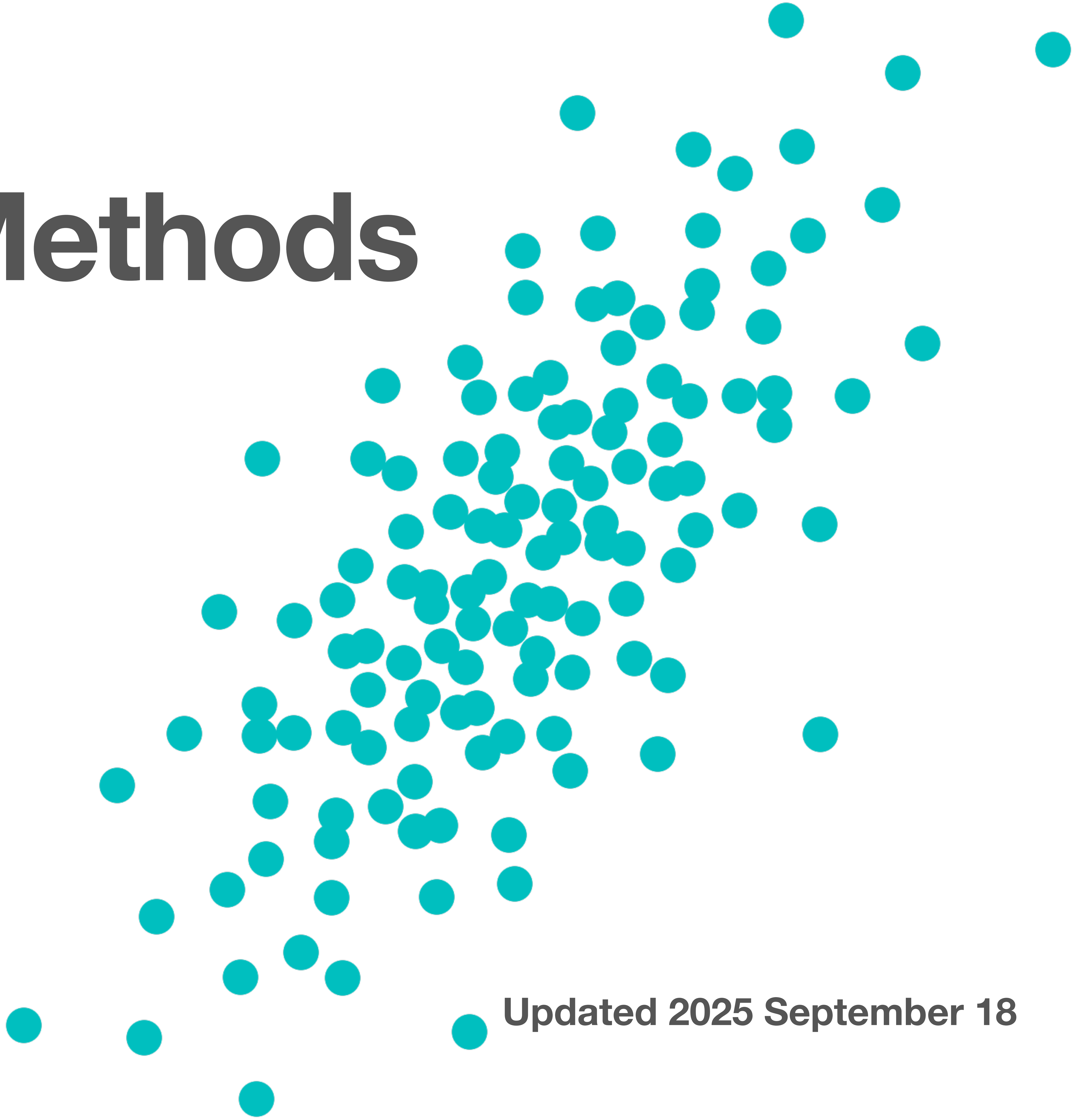
Improving Efficiency

Selected Topics

Git website and repository

Canvas

Fred Hickernell, Fall 2025



Updated 2025 September 18

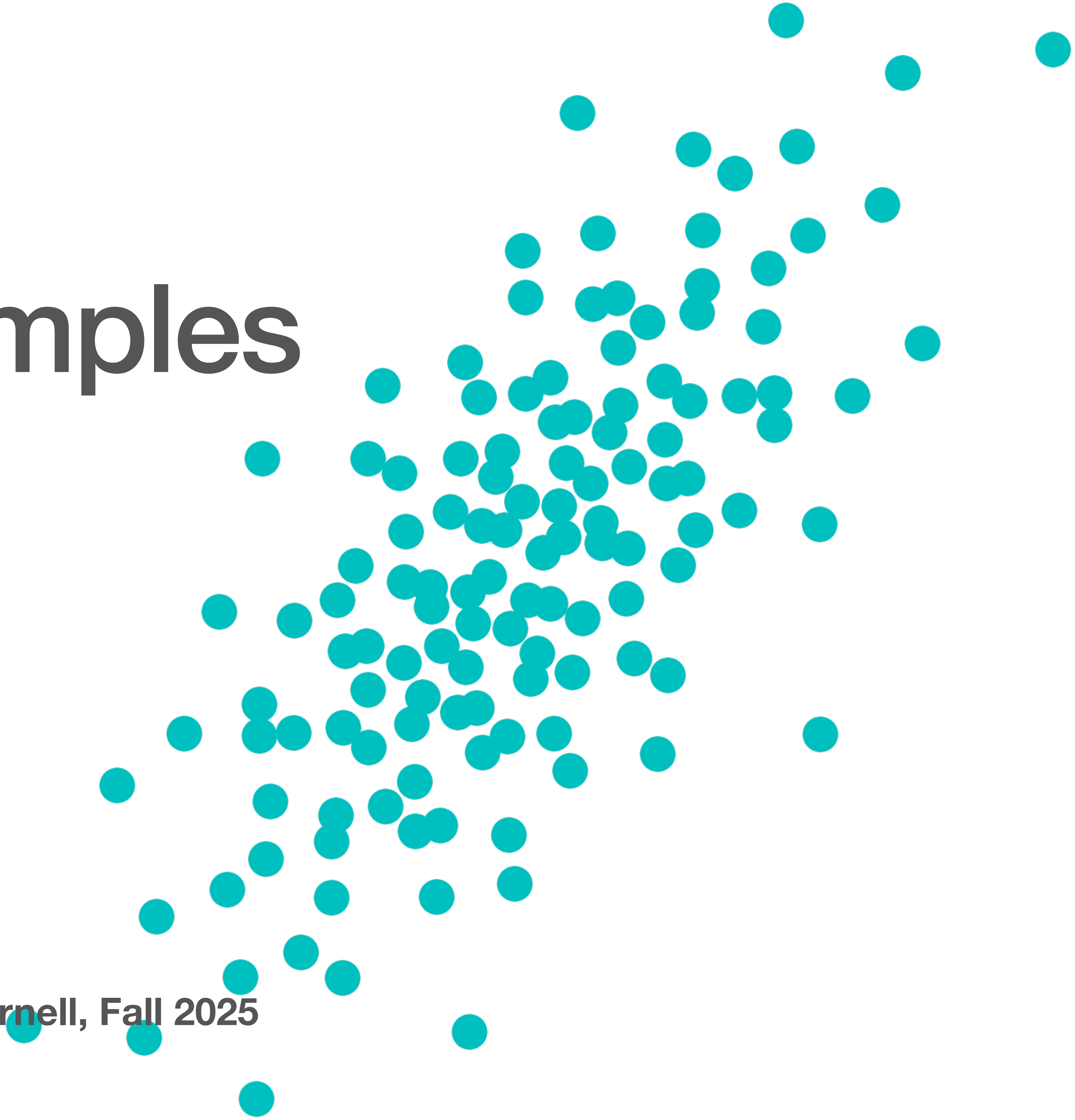
Generating Samples

Owen, Chapters 3 – 6

Assignment 2 due Sep 19

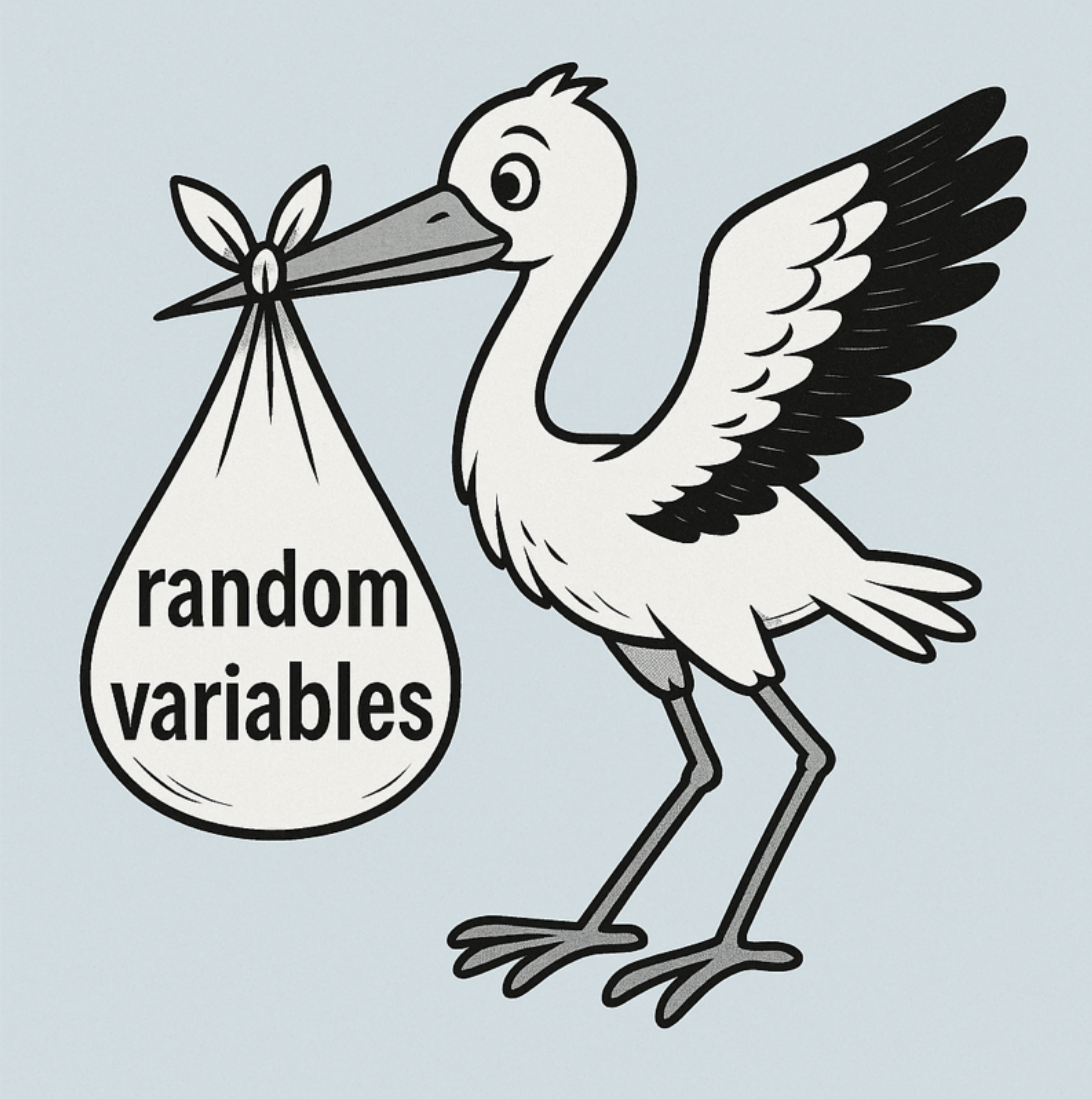
Test 1 Sep 24

Project Selection due Oct 1



MATH 565 Monte Carlo Methods, Fred Hickernell, Fall 2025

Where do random variables come from?



Where do random variables come from?

- Algorithms generate (typically uniform) pseudo-random numbers

Where do random variables come from?

- Algorithms generate (typically uniform) pseudo-random numbers
- Pseudo-random number generators are tested to ensure that their output looks IID random in every way possible

Where do random variables come from?

- Algorithms generate (typically uniform) pseudo-random numbers
- Pseudo-random number generators are tested to ensure that their output looks IID random in every way possible
- Care must be taken when working on multiple processors

Where do random variables come from?

- Algorithms generate (typically uniform) pseudo-random numbers
- Pseudo-random number generators are tested to ensure that their output looks IID random in every way possible
- Care must be taken when working on multiple processors
- For a drastic failure see RANDU

Where do random variables come from?

- Algorithms generate (typically uniform) pseudo-random numbers
- Pseudo-random number generators are tested to ensure that their output looks IID random in every way possible
- Care must be taken when working on multiple processors
- For a drastic failure see RANDU
- Physical random number generators may be random, but cannot be guaranteed to be IID random of the desired distribution

Quantile function gives non-uniform random numbers

- Most random number generators output $U_1, U_2, \ldots$ that mimic IID $\mathcal{U}[0,1]$.
- To get $X_1, X_2, \ldots$ IID with cumulative distribution function F, we use the quantile function, Q, where $Q(u) := \inf\{x \in \mathcal{X} : F(x) \geq u\}$. Note that

$$Q(u) \leq x \iff u \leq F(x)$$

- Letting $X := Q(U)$, it follows that

$$\mathbb{P}(X \leq x) = \mathbb{P}(Q(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)$$

and so $X \sim F$, and

$$X_1 := Q(U_1), X_2 := Q(U_2), \ldots \overset{\text{IID}}{\sim} F$$

Binomial random numbers

Binomial random numbers

Suppose $X \sim \text{Binomial}(3, 0.6)$

Binomial random numbers

Suppose $X \sim$ Binomial$(3,0.6)$

x	0	1	2	3
PMF $\varrho(x)$	0.4^3	$3(0.6)(0.4^2)$	$3(0.6^2)(0.4)$	0.6^3
	0.064	0.288	0.432	0.216
$x \in$	$[0,1)$	$[1,2)$	$[2,3)$	$[3,\infty)$
CDF $F(x)$	0.064	0.352	0.784	1
$u \in$	$(0,0.064]$	$(0.064,0.352]$	$(0.352,0.784]$	$(0.784,1]$
quantile $Q(u)$	0	1	2	3

Binomial random numbers

Suppose $X \sim$ Binomial$(3,0.6)$

x	0	1	2	3
PMF $\varrho(x)$	0.4^3	$3(0.6)(0.4^2)$	$3(0.6^2)(0.4)$	0.6^3
	0.064	0.288	0.432	0.216
$x \in$	$[0,1)$	$[1,2)$	$[2,3)$	$[3,\infty)$
CDF $F(x)$	0.064	0.352	0.784	1
$u \in$	$(0,0.064]$	$(0.064,0.352]$	$(0.352,0.784]$	$(0.784,1]$
quantile $Q(u)$	0	1	2	3

So $U_1 = 0.63,\ U_2 = 0.02,\ U_3 = 0.47$ produces $X_1 = 2,\ X_2 = 0,\ X_3 = 2$

Binomial random numbers

Suppose $X \sim$ Binomial$(3{,}0.6)$

x	0	1	2	3
PMF $\varrho(x)$	0.4^3	$3(0.6)(0.4^2)$	$3(0.6^2)(0.4)$	0.6^3
	0.064	0.288	0.432	0.216
$x \in$	$[0, 1)$	$[1, 2)$	$[2, 3)$	$[3, \infty)$
CDF $F(x)$	0.064	0.352	0.784	1
$u \in$	$(0, 0.064]$	$(0.064, 0.352]$	$(0.352, 0.784]$	$(0.784, 1]$
quantile $Q(u)$	0	1	2	3

So $U_1 = 0.63,\ U_2 = 0.02,\ U_3 = 0.47$ produces $X_1 = 2,\ X_2 = 0,\ X_3 = 2$

Note that F is right-continuous and Q is left-continuous

Explorations in generating random vectors

Generating Samples

Zero-inflated exponential

Suppose that X is a non-negative random variable with probability p_0 of being zero and otherwise an exponential distribution

$$F(x) = \begin{cases} 0 & -\infty < x < 0 \\ p_0 + (1 - p_0)[1 - \exp(-\lambda x)] & 0 \le x < \infty \end{cases}$$

To generate $X_1, X_2, \ldots \overset{\text{IID}}{\sim} F$ from $U_1, U_2, \ldots \overset{\text{IID}}{\sim} \mathcal{U}[0,1]$, we need the quantile function

Quantile function for zero-inflated exponential

$$F(x) = \begin{cases} 0 & -\infty < x < 0 \\ p_0 + (1-p_0)[1-\exp(-\lambda x)] & 0 \le x < \infty \end{cases}$$

$$\left.\begin{matrix} u \le F(x) \\ u = F(x) \end{matrix}\right\} \implies \begin{cases} x = 0, & 0 < u \le p_0 \\ u = p_0 + (1-p_0)[1-\exp(-\lambda x)], & p_0 \le u < 1 \end{cases}$$

$$\iff \begin{cases} x = 0, & 0 < u \le p_0 \\ \dfrac{u-p_0}{1-p_0} = 1-\exp(-\lambda x), & p_0 \le u < 1 \end{cases}$$

$$\iff x = Q(u) = \begin{cases} 0, & 0 < u \le p_0 \\ -\dfrac{1}{\lambda}\log\left(1-\dfrac{u-p_0}{1-p_0}\right), & p_0 \le u < 1 \end{cases}$$

Random vectors with independent marginals

Suppose that $\boldsymbol{X} = (X_1, \ldots, X_d)$ is a random vector with independent marginals. Then

$$\boldsymbol{X} = \big(Q_1(\boldsymbol{U}_1), \ldots, Q_d(\boldsymbol{U}_d)\big), \quad \boldsymbol{U} \sim \mathcal{U}[0,1]^d$$

has the desired distribution, provided that $Q_1, \ldots, Q_d$ are the corresponding quantile functions.

We express n samples of $\boldsymbol{X}$ as an $n \times d$ matrix or array,

$$\mathsf{X} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1d} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nd} \end{pmatrix},$$ where X_{ij} is the j^{th} component of the i^{th} sample

What is your expertise in computing and probability?

- Go to menti.com
- Use code 4845 0474

Multivariate normal distribution

If $Z \sim \mathcal{N}(\mathbf{0}_d, \mathrm{I}_d)$, i.e., a d-dimensional standard normal random variable, then $X = \mathrm{A}Z + \boldsymbol{b}$ (thinking of Z as a column vector) has

$$\boldsymbol{\mu} = \mathbb{E}(X) = \boldsymbol{b}$$

$$\Sigma = \operatorname{cov}(X) = \mathbb{E}[(X - \boldsymbol{\mu})(X - \boldsymbol{\mu})^\top]$$

$$= \mathbb{E}[\mathrm{A}ZZ^T\mathrm{A}^T] = \mathrm{A}\mathrm{A}^\top$$

So $X \sim \mathcal{N}(\boldsymbol{b}, \Sigma)$, where Σ is symmetric and positive-definite. So, give a desired $\boldsymbol{\mu}$ and Σ, one needs only to find A with $\Sigma = \mathrm{A}\mathrm{A}^\top$, and then

$$X = \mathrm{A}Z + \boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

Note that there are multiple ways to decompose $\Sigma = \mathrm{A}\mathrm{A}^\top$.

Cholesky decomposition of Σ

We want to find a lower triangular matrix A, such that $\Sigma = AA^\top$.

$$\begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix} = \Sigma = AA^\top = \begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} \\ 0 & a_{22} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}^2 & a_{11}a_{21} \\ a_{11}a_{21} & a_{21}^2 + a_{22}^2 \end{pmatrix}$$

$$\implies a_{11} = \sqrt{3}$$

$$a_{21} = -1/a_{11} = -1/\sqrt{3}$$

$$a_{22} = \sqrt{1 - a_{21}^2} = \sqrt{2/3}$$

$$A = \begin{pmatrix} \sqrt{3} & 0 \\ -1/\sqrt{3} & \sqrt{2/3} \end{pmatrix}$$

Gaussian processes

- Random functions such that evaluating them at a finite number of points gives a random vector with a multivariate normal distribution
- The distribution is denoted $\mathcal{GP}(\mu, K)$, where μ is the mean function and K is the covariance kernel
- $g \sim \mathcal{GP}(\mu, K) \implies$
  - $\mathbb{E}[g(t)] = \mu(t)$
  - $\mathbb{E}[\{g(t) - \mu(t)\}\{g(x) - \mu(x)\}] = K(t, x)$

Brownian motion

Special case of a Gaussian process

$$\mu(t) = 0 \quad \text{and} \quad K(t, x) = \min(t, x), \qquad 0 \le t$$

For $0 \le t_1 < t_2 < \cdots \le t_d$,

$$\boldsymbol{X} = (B(t_1), \ldots, B(t_d))^\top \sim \mathcal{N}(\boldsymbol{0}, \Sigma)$$

where

$$\Sigma = \begin{pmatrix} t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & \cdots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & \cdots & t_d \end{pmatrix} = \begin{pmatrix} \sqrt{t_1} & 0 & \cdots & 0 \\ \sqrt{t_1} & \sqrt{t_2 - t_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{t_1} & \sqrt{t_2 - t_1} & \cdots & \sqrt{t_d - t_{d-1}} \end{pmatrix} \begin{pmatrix} \sqrt{t_1} & \sqrt{t_1} & \cdots & \sqrt{t_1} \\ 0 & \sqrt{t_2 - t_1} & \cdots & \sqrt{t_2 - t_1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{t_d - t_{d-1}} \end{pmatrix}$$

Brownian motion

For this Cholesky decomposition of the covariance matrix

$$\Sigma = \begin{pmatrix} t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & \cdots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & \cdots & t_d \end{pmatrix} = \begin{pmatrix} \sqrt{t_1} & 0 & \cdots & 0 \\ \sqrt{t_1} & \sqrt{t_2 - t_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{t_1} & \sqrt{t_2 - t_1} & \cdots & \sqrt{t_d - t_{d-1}} \end{pmatrix} \begin{pmatrix} \sqrt{t_1} & \sqrt{t_1} & \cdots & \sqrt{t_1} \\ 0 & \sqrt{t_2 - t_1} & \cdots & \sqrt{t_2 - t_1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{t_d - t_{d-1}} \end{pmatrix}$$

One may construct, IID random samples $\boldsymbol{X}_i = (B(t_1), \ldots, B(t_d))^\top \overset{\text{IID}}{\sim} \mathcal{N}(\boldsymbol{0}, \Sigma)$ by

$$X_{i1} = \sqrt{t_1} Z_{i1}, \quad X_{i2} = X_{i1} + \sqrt{t_2 - t_1} Z_{i2}, \quad \ldots, \quad X_{id} = X_{i,d-1} + \sqrt{t_d - t_{d-1}} Z_{id}$$

where $\boldsymbol{Z}_i \overset{\text{IID}}{\sim} \mathcal{N}(\boldsymbol{0}, \mathrm{I}_d)$

Stock/asset prices

Brownian motions are used in financial modeling. If

- S_0 is the initial price of a stock/asset
- r is the interest rate
- σ is the volatility

Then

$$S(t) = S_0 \exp\Big((r - \sigma^2/2)t + \sigma B(t)\Big)$$

is a geometric Brownian motion model of the random stock price

Option payoffs and prices

Give some model of asset prices, we may model the payoff of an option.
Let K be the strike price and T be the time to maturity. Examples of discounted payoffs are

European call	$\max\big(S(T) - K, 0\big)\exp(-rT)$
European put	$\max\big(K - S(T), 0\big)\exp(-rT)$
Asian Arithmetic mean call	$\max\left(\frac{1}{T}\int_0^T S(t)\mathrm{d}t - K, 0\right)\exp(-rT)$

The fair price of an option is

$$\mu = \mathbb{E}[\text{payoff}]$$

Low discrepancy sampling

If we are willing to accept samples that are not IID, then we may get substantial gains in the convergence rate. We say that

$$U_1, \ldots, U_n \overset{\text{LD}}{\sim} \mathcal{U}[0,1]^d$$

if the empirical distribution of $\{U_1, \ldots, U_n\}$ is close to the uniform distribution. The measure of how far away these distributions are is the discrepancy, and it will be defined precisely later. Some examples are

- Digital sequences, especially Sobol’ sequences
- Lattice sequences
- Halton sequences
- Kronecker sequences

These come in deterministic and random forms.

Low discrepancy sampling

- LD sequences are available through `scipy` and `qmcpy`, which has been developed in our research group
- Rows correspond to samples and columns to coordinates
- See Generating Samples for empirical comparisons
- The theory of LD sampling will be touched on later in the semester

Bias–variance decomposition

- Let μ be any population quantity, e.g., the mean
- Let $\widehat{\mu}$ be any estimator for μ, e.g., the sample mean
- Then the mean squared error of $\widehat{\mu}$ is

$$\begin{aligned}\operatorname{mse}(\widehat{\mu}) = \mathbb{E}[(\mu - \widehat{\mu})^2] &= \mathbb{E}\big[\{\mu - \mathbb{E}(\widehat{\mu})\} + \{\mathbb{E}(\widehat{\mu}) - \widehat{\mu}\}\big]^2 \\ &= \mathbb{E}\big[\{\mu - \mathbb{E}(\widehat{\mu})\}\big]^2 + \mathbb{E}\big[\{\mathbb{E}(\widehat{\mu}) - \widehat{\mu}\}\big]^2 \\ &= [\operatorname{bias}(\widehat{\mu})]^2 + \operatorname{var}(\widehat{\mu}), \qquad \text{where} \quad \operatorname{bias}(\widehat{\mu}) = \mathbb{E}(\widehat{\mu} - \mu\end{aligned}$$

IID sample means have zero bias and positive variance

Deterministic LD sample means have positive bias but zero variance

Shrinkage estimators

Suppose that $\widehat{\mu}_n$ is the sample mean of $Y_1, \ldots, Y_n \overset{\text{IID}}{\sim} (\mu, \sigma^2)$.

For what value of α is $\alpha\widehat{\mu}_n$ the estimator of μ with smallest mean squared error?

Acceptance–rejection sampling

Suppose that

- You can sample a random variable Z with known proposal density ϱ_Z
- You really want to sample X with known (unnormalized) target density ϱ, i.e., the true density is $c\varrho$ for some positive constant c
- And we know a M for which $\varrho(x) \leq M\varrho_Z(x)$ for all x

The acceptance-rejection sampling proceeds as follows

```
i←0
repeat
  repeat
    generate Z~ϱ_Z and U~𝒰[0,1]
  until U ≤ ϱ(Z)/[Mϱ_Z(Z)]
X[i]←Z, i←i+1
until i=n
```

Why does acceptance–rejection sampling work?

$$W := \begin{cases} 1, & U \le \varrho(Z)/[M\varrho_Z(Z)] \qquad \text{accept} \\ 0, & \text{otherwise} \qquad \text{reject} \end{cases}$$

Our sampling method generates X that has a PDF of

$$\begin{aligned}
\widetilde{\varrho}(z) &= \varrho_{Z|W}(z|1) = \text{PDF of } Z \text{ conditioned on accepting } Z \\
&= \frac{\varrho_{W|Z}(1|z)\varrho_Z(z)}{\varrho_W(1)} \qquad \text{by Bayes Theorem} \\
&= \frac{\{\varrho(z)/[M\varrho_Z(z)]\}\varrho_Z(z)}{\varrho_W(1)} \qquad \text{by our acceptance rule} \\
&= \frac{\varrho(z)}{M\varrho_W(1)} = c\varrho(z) \text{ since } c\varrho \text{ is a density; } \mathbb{P}(\text{acceptance}) = \varrho_W(1) = 1/(Mc)
\end{aligned}$$

To get n samples of X, we need on average $n(Mc)$ samples of Z

Explorations in acceptance-rejection

See Acceptance Rejection Sampling for some examples and figures

Comparing methods for non-uniform sampling

Method	Requirements	LD friendly
$X = Q(U)$	quantile function	Yes
$\boldsymbol{X} = \mathsf{A}\boldsymbol{Z} + \boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$	$\Sigma = \mathsf{A}\mathsf{A}^T$	Yes
acceptance-rejection	$\varrho(z) \leq M \varrho_Z(z)$	Somewhat

For acceptance-rejection with LD see [Zhu & Dick 2024]
This is a good project topic