Introduction to Causal Inference

# Lecture 4

# Model-Based Inference for Completely Randomized Experiments

邓婉璐

WANLUDENG@TSINGHUA.EDU.CN



清华大学统计与数据科学系

# The National Supported Work Data



**Unit:** $N$ = 445 men substantially disadvantaged in the labor market

**Treatment:** a training program

**Control:** no active treatment

**Outcome:**

earnings in 1978 (earn’78)

**Seven covariates:**

age, education, married, no degree, black, earn’75, earn’74

**Completely randomized assignment:**

$N_t = 185, N_c = 260$

**Table 8.1.** ***Summary Statistics: National Supported Work (NSW) Program Data***

| Covariate | Mean | (S.D.) | Average Controls ($N_c = 260$) | Average Treated ($N_t = 185$) |
|---|---|---|---|---|
| age | 25.37 | (7.10) | 25.05 | 25.82 |
| education | 10.20 | (1.79) | 10.09 | 10.35 |
| married | 0.17 | (0.37) | 0.15 | 0.19 |
| nodegree | 0.78 | (0.41) | 0.83 | 0.71 |
| black | 0.83 | (0.37) | 0.83 | 0.84 |
| earn’74 | 2.10 | (5.36) | 2.11 | 2.10 |
| earn’74=0 | 0.73 | (0.44) | 0.75 | 0.71 |
| earn’75 | 1.38 | (3.15) | 1.27 | 1.53 |
| earn’75=0 | 0.65 | (0.48) | 0.68 | 0.60 |
| earn’78 | 5.30 | (6.63) | 4.56 | 6.35 |
| earn’78=0 | 0.31 | (0.46) | 0.35 | 0.24 |

# How to estimate the Average Causal Effect?



- Finite Sample
  - ✓ $H_0: Y_i(0) = Y_i(1)$
  - ✓ $\hat{\tau}^{dif} = \bar{Y}_t^{obs} - \bar{Y}_c^{obs}$; $\hat{V} \rightarrow Confidence\ Interval$
- Super Population
  - ✓ $\hat{\tau}^{dif} = \bar{Y}_t^{obs} - \bar{Y}_c^{obs}$, $\hat{V}^{Neyman}$
  - ✓ $\hat{\tau}^{ols}$: $Y_i^{obs} = \alpha + W_i\tau + \vec{X}_i\beta + \varepsilon$

**Treatment (W=1):** a training program
**Control (W=0):** no active treatment

**Outcome (Y):**
earnings in 1978 (earn’78)

**Seven covariates ($\vec{X}$):**
age, education, married, no degree, black, earn’75, earn’74

# The National Supported Work Data



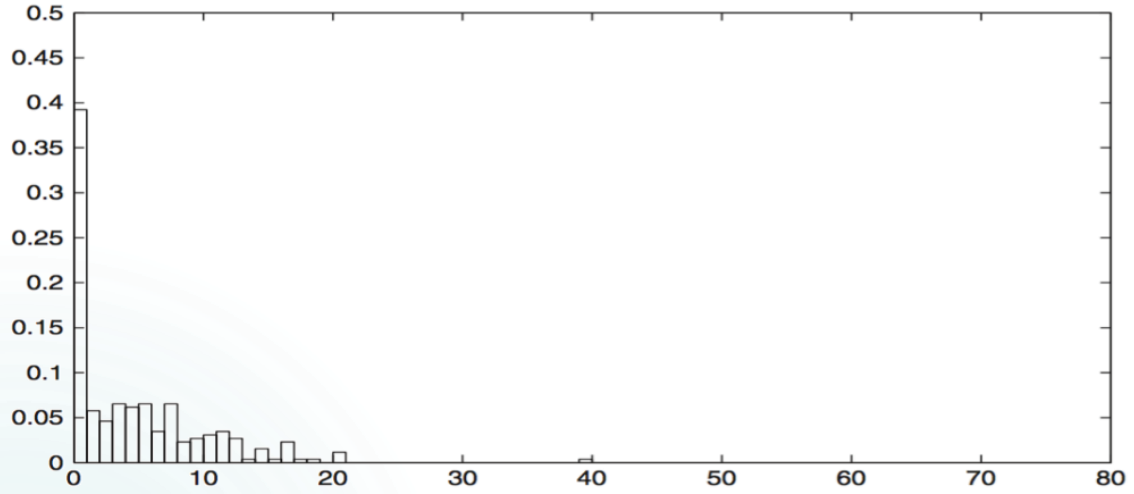



**Figure 8.1.** Histogram of earnings for control group – NSW job-training data

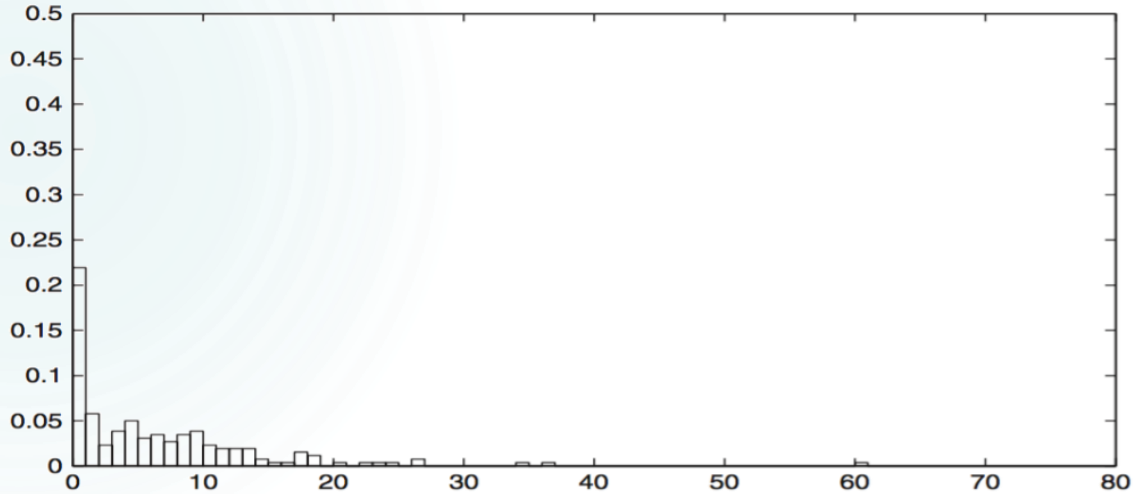



**Figure 8.2.** Histogram of earnings for trainee group – NSW job-training data

# Possible Model



- Model:

$$\begin{pmatrix} \ln(Y_i(0)) \\ \ln(Y_i(1)) \end{pmatrix} \Big| \theta \sim \mathcal{N}\left( \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} \sigma_c^2 & 0 \\ 0 & \sigma_t^2 \end{pmatrix} \right).$$

- Parameter:

$$\theta = (\mu_t, \mu_c, \sigma_t^2, \sigma_c^2)'$$

- Goal:
  - Finite sample ACE: $\tau_{fs} = \frac{1}{N}\sum_{i=1}^{N}\big(Y_i(1) - Y_i(0)\big) = \tau(\mathbf{Y}(0), \mathbf{Y}(1))$
  - Super Population ACE: $\tau_{sp} = \mathbb{E}_{\text{sp}}[Y_i(1) - Y_i(0)] = \tau(\theta)$

$$\tau_{\text{sp}} = \tau(\theta) = \exp\left(\mu_t + \frac{1}{2}\cdot\sigma_t^2\right) - \exp\left(\mu_c + \frac{1}{2}\cdot\sigma_c^2\right)$$



# Completely Randomized Experiment



- A leading case of a classical randomized experiment
- The number of treated units $N_t$ is fixed
- $N_t$ units are randomly selected, from a population of $N$ units, to receive the active treatment, with the remaining $N_c$ assigned to the control group.
- Each unit has unit assignment probability $q = N_t / N$
- The assignment mechanism:

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} 1 \Big/ \binom{N}{N_t} & \text{if } \sum_{i=1}^{N} W_i = N_t, \\ 0 & \text{otherwise,} \end{cases}$$

$$\binom{N}{N_t} = \frac{N!}{N_t!(N - N_t)!}$$

- subset of the set of possible values for $\mathbf{W}$ with positive probability

$$\mathbb{W}^+ = \left\{ \mathbf{W} \in \mathbb{W} \,\middle|\, \sum_{i=1}^{N} W_i = N_t \right\}$$



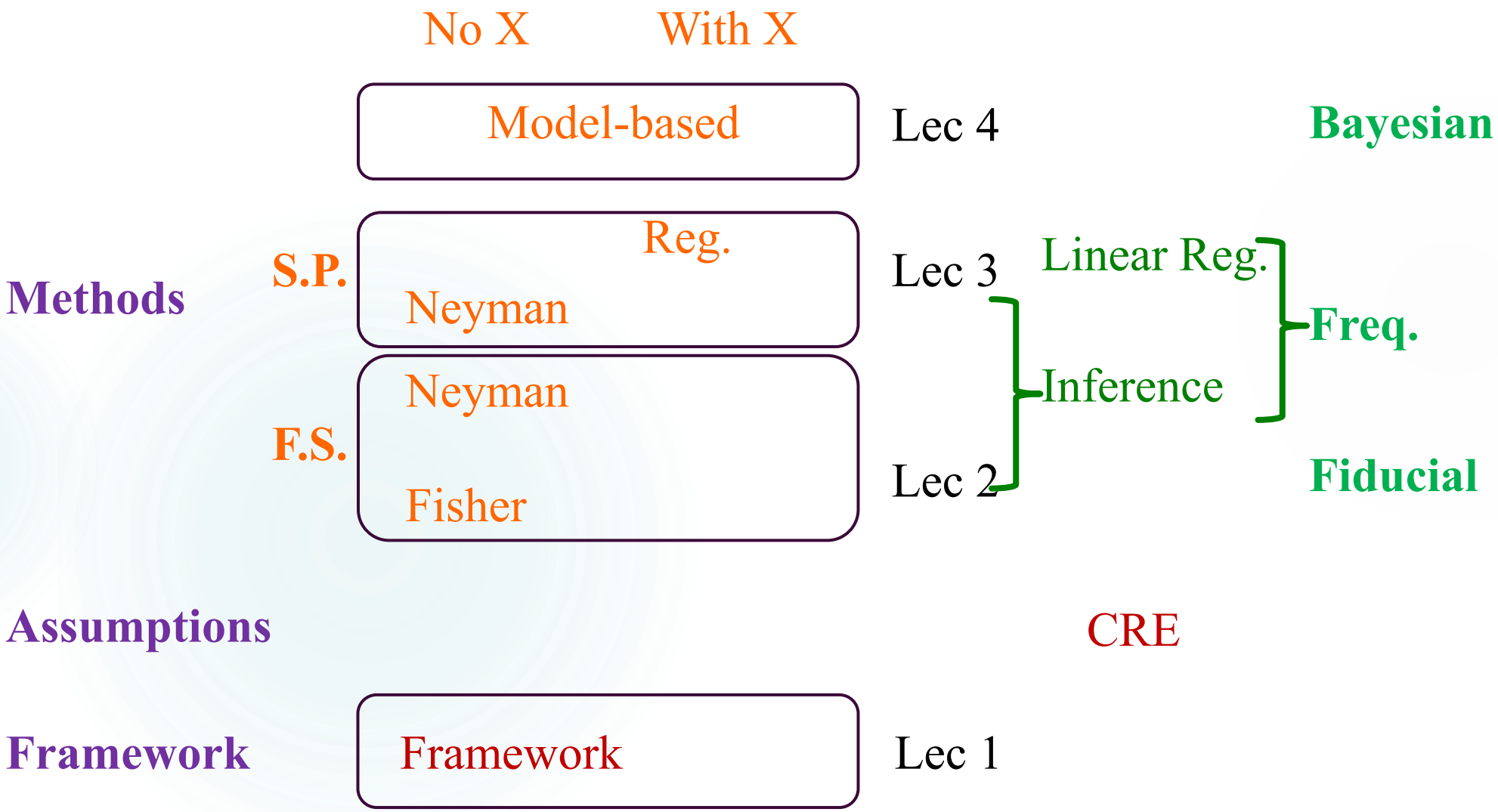
# Inference of CREs



- Fisher’s Exact P-Values
  - whether there exists a non-zero unit-level causal effect?
- Neyman’s Repeated Sampling Approach
  - how to estimate the average causal effect?
- Regression Methods
  - how covariates help estimate the average causal effect?
- Model-Based Inference
  - How concrete parametric models on potential outcomes help infer the average causal effect?



# Review


No X
With X
Model-based
Lec 4
Bayesian
Methods
S.P.
Reg.
Neyman
Lec 3
Linear Reg.
Freq.
Inference
F.S.
Neyman
Fisher
Lec 2
Fiducial
Assumptions
CRE
Framework
Framework
Lec 1

# Objectives for today



- 理解Neyman-Rubin Framework是一套描述因果的语言。其基础上进行统计推断，可以用各类统计方法，有一定的兼容性，这也是Rubin的主要贡献之一!
- 掌握最基本的Model-based inference，例如binomial, Normal。思路+简单推导
- 进一步体会Finite sample与Super population的区别



# Outline



- Motivation
  - Moving towards to Bayesian model
- Basic Bayesian theories
- Bayesian model-based imputation for finite sample
  - Theory (Required)
  - Simulation (Not required)
- Bayesian model-based imputation for Super Population
- Model comparison in application

Illustrated by a real data example



# Motivation

ILLUSTRATED WITH A REAL DATA EXAMPLE



# Background/Assumptions and Goal



- Background / Assumptions
  - ✓ A random sample from a super population
  - ✓ CRE

- Goal:
  - Finite sample ACE: $\tau_{fs} = \frac{1}{N}\sum_{i=1}^{N}\big(Y_i(1) - Y_i(0)\big) = \tau(\mathbf{Y}(0), \mathbf{Y}(1))$
  - Super Population ACE: $\tau_{sp} = \mathbb{E}_{\text{sp}}[Y_i(1) - Y_i(0)] = \tau(\theta)$

In the following, we consider finite sample ACE first!

# First Six Observations from NSW Data



$$\tau_{\text{fs}} = \tau(\mathbf{Y}(0), \mathbf{Y}(1)) = \frac{1}{6} \cdot \sum_{i=1}^{6} \big(Y_i(1) - Y_i(0)\big)$$

$$Y_i(0) = \begin{cases} Y_i^{\text{mis}} & \text{if } W_i = 1, \\ Y_i^{\text{obs}} & \text{if } W_i = 0, \end{cases} \quad \text{and} \quad Y_i(1) = \begin{cases} Y_i^{\text{mis}} & \text{if } W_i = 0, \\ Y_i^{\text{obs}} & \text{if } W_i = 1. \end{cases}$$

**Table 8.2.** *First Six Observations from NSW Program Data*

| Unit | Potential Outcomes | | Treatment | Observed Outcome |
|---|---|---|---|---|
| | $Y_i(0)$ | $Y_i(1)$ | $W_i$ | $Y_i^{\text{obs}}$ |
| 1 | 0 | ? | 0 | 0 |
| 2 | ? | 9.9 | 1 | 9.9 |
| 3 | 12.4 | ? | 0 | 12.4 |
| 4 | ? | 3.6 | 1 | 3.6 |
| 5 | 0 | ? | 0 | 0 |
| 6 | ? | 24.9 | 1 | 24.9 |

*Note*: Question marks represent missing potential outcomes.

$$\begin{aligned} \tau_{\text{fs}} &= \tilde{\tau}(\mathbf{Y}^{\text{obs}}, \mathbf{Y}^{\text{mis}}, \mathbf{W}) \\ &= \frac{1}{6} \cdot \sum_{i}^{N} \Big( (W_i \cdot Y_i^{\text{obs}} + (1 - W_i) \cdot Y_i^{\text{mis}}) - ((1 - W_i) \cdot Y_i^{\text{obs}} + W_i \cdot Y_i^{\text{mis}}) \Big) \\ &= \frac{1}{6} \cdot \sum_{i=1}^{N} \Big( (2 \cdot W_i - 1) \cdot \big( Y_i^{\text{obs}} - Y_i^{\text{mis}} \big) \Big) \cdot \end{aligned}$$

Key: how to **impute** $Y^{\text{mis}}$ given $Y^{\text{obs}}$ and W?

$$\hat{\tau} = \tilde{\tau}(\mathbf{Y}^{\text{obs}}, \hat{\mathbf{Y}}^{\text{mis}}, \mathbf{W}) = \frac{1}{6} \cdot \sum_{i=1}^{N} \Big( (2 \cdot W_i - 1) \cdot (Y_i^{\text{obs}} - \hat{Y}_i^{\text{mis}}) \Big)$$

# A Naïve Approach to Imputation



**Table 8.3.** ***The Average Treatment Effect Using Imputation of Average Observed Outcome Values within Treatment and Control Groups for the NSW Program Data***

| Unit | Potential Outcomes | | Treatment | Observed Outcome |
|---|---|---|---|---|
| | $Y_i(0)$ | $Y_i(1)$ | $W_i$ | $Y_i^{obs}$ |
| 1 | 0 | (12.8) | 0 | 0 |
| 2 | (4.13) | 9.9 | 1 | 9.9 |
| 3 | 12.4 | (12.8) | 0 | 12.4 |
| 4 | (4.13) | 3.6 | 1 | 3.6 |
| 5 | 0 | (12.8) | 0 | 0 |
| 6 | (4.13) | 24.9 | 1 | 24.9 |
| Average | 4.13 | 12.8 | | |
| Diff (ATE): | | 8.67 | | |

$\hat{\tau} = 12.8 - 4.1 = 8.7 \Leftrightarrow \hat{\tau}^{dif} = \overline{Y}_t^{obs} - \overline{Y}_c^{obs}$

$\overline{Y}_c^{obs} = (Y_1(0) + Y_3(0) + Y_5(0))/3 = (0 + 12.4 + 0)/3 = 4.1$

$\overline{Y}_t^{obs} = (Y_2(1) + Y_4(1) + Y_6(1))/3 = (9.9 + 3.6 + 24.9)/3 = 12.8$

Imputation mechanism:

$$\Pr\left(Y_i^{mis} = y \middle| \mathbf{Y}^{obs}, \mathbf{W}\right) = \begin{cases} 1 & \text{if } y = 12.8, \text{ and } W_i = 0, \\ 1 & \text{if } y = 4.1, \text{ and } W_i = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Impute the missing potential outcomes with constants as if there were no uncertainty about their values

# Motivation



The target

Distribution of $\tau = \tau(\mathbf{Y}(0), \mathbf{Y}(1))$

$(\mathbf{Y}(0), \mathbf{Y}(1)) = h(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}, \mathbf{W})$

→ $\tau(Y^{mis}, Y^{obs}, W)$

↓

$f(\tau | \mathbf{Y}^{\text{obs}}, \mathbf{W})$

↑ ?

The model

We need
(1) a more flexible model;
(2) Ways to solve the model.

# Bayesian Model-Based Imputation



The output

Distribution of $\tau = \tau(\mathbf{Y}(0), \mathbf{Y}(1))$

$(\mathbf{Y}(0), \mathbf{Y}(1)) = h(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}, \mathbf{W})$

→ $\tau(Y^{mis}, Y^{obs}, W)$

↓ $f(\tau|\mathbf{Y}^{\text{obs}}, \mathbf{W})$

↑ ?

The 3 inputs

$$\Pr(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$f(\mathbf{Y}(0), \mathbf{Y}(1)|\theta)$

$p(\theta)$

# Example: Possible Model



- Model:

$$\begin{pmatrix} \ln(Y_i(0)) \\ \ln(Y_i(1)) \end{pmatrix} \Big| \theta \sim \mathcal{N}\left( \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} \sigma_c^2 & 0 \\ 0 & \sigma_t^2 \end{pmatrix} \right).$$

- Parameter:

$$\theta = (\mu_t, \mu_c, \sigma_t^2, \sigma_c^2)'$$

**Treatment (W=1):** a training program
**Control (W=0):** no active treatment

**Outcome (Y):**
earnings in 1978 (earn’78)

The 3 inputs

$$\Pr(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$$f(\mathbf{Y}(0), \mathbf{Y}(1)|\theta)$$

$$p(\theta)$$



# Bayesian Framework

ILLUSTRATED WITH A SIMPLE EXAMPLE



# The Bayesian Paradigm



- ***Key opinion***
  - Treat unknowns (model parameters) as random variables
- ***Main elements***
  - Prior distribution: $p(\theta)$
  - Likelihood function: $p(data \mid \theta)$
  - Joint distribution: $p(data, \theta)$
  - Posterior distribution: $p(\theta \mid data)$
- ***Bayesian inference***
  - Describe quantities of scientific interest with probability
  - Let the ***law of probability*** work its way out
  - Summarize all uncertainty into posterior distribution



# Example: Binomial Model



- Laplace's work on boy birth rate in Paris. (1786)
  - Want to test if the rate is over 0.5.
  - 1745-1770, 251527 boys, 241945 girls.

- Assume that each baby is independent and the outcome each time is a Bernoulli trial with success probability $\theta$.

- This implies that $y|\theta \sim Bin(n, \theta)$

- The sample size $n$ is fixed and the success probability $\theta$ can be any number in $[0, 1]$.



# Example: Binomial Model



- The measurement model is

$$p(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}, \qquad \theta \in [0,1]$$

Since $n$ is fixed in this analysis, we'll drop it as a conditioning argument for ease of notation.

- Goal: make inference on $\theta$ given $y$ and $n$.
- Need: a prior distribution $p(\theta)$ for $\theta$.
- Bayes' choice: $\theta \sim U(0,1)$, i.e.

$$p(\theta) = \begin{cases} 1 & 0 \le \theta \le 1 \\ 0 & \text{Otherwise} \end{cases}$$



# Example: Binomial Model



- So the Bayes' rule gives

$$p(y,\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y},$$

$$p(y) = \int_0^1 \binom{n}{y}\theta^y(1-\theta)^{n-y}d\theta = \frac{1}{n+1},$$

$$p(\theta|y) = (n+1)\binom{n}{y}\theta^y(1-\theta)^{n-y}.$$

- Given data, Laplace calculated that
  $Prob(boy\ birth\ rate \leq 0.5) < 10^{-42}$.
  So he concluded that the rate is over 0.5.



# Example: Binomial Model



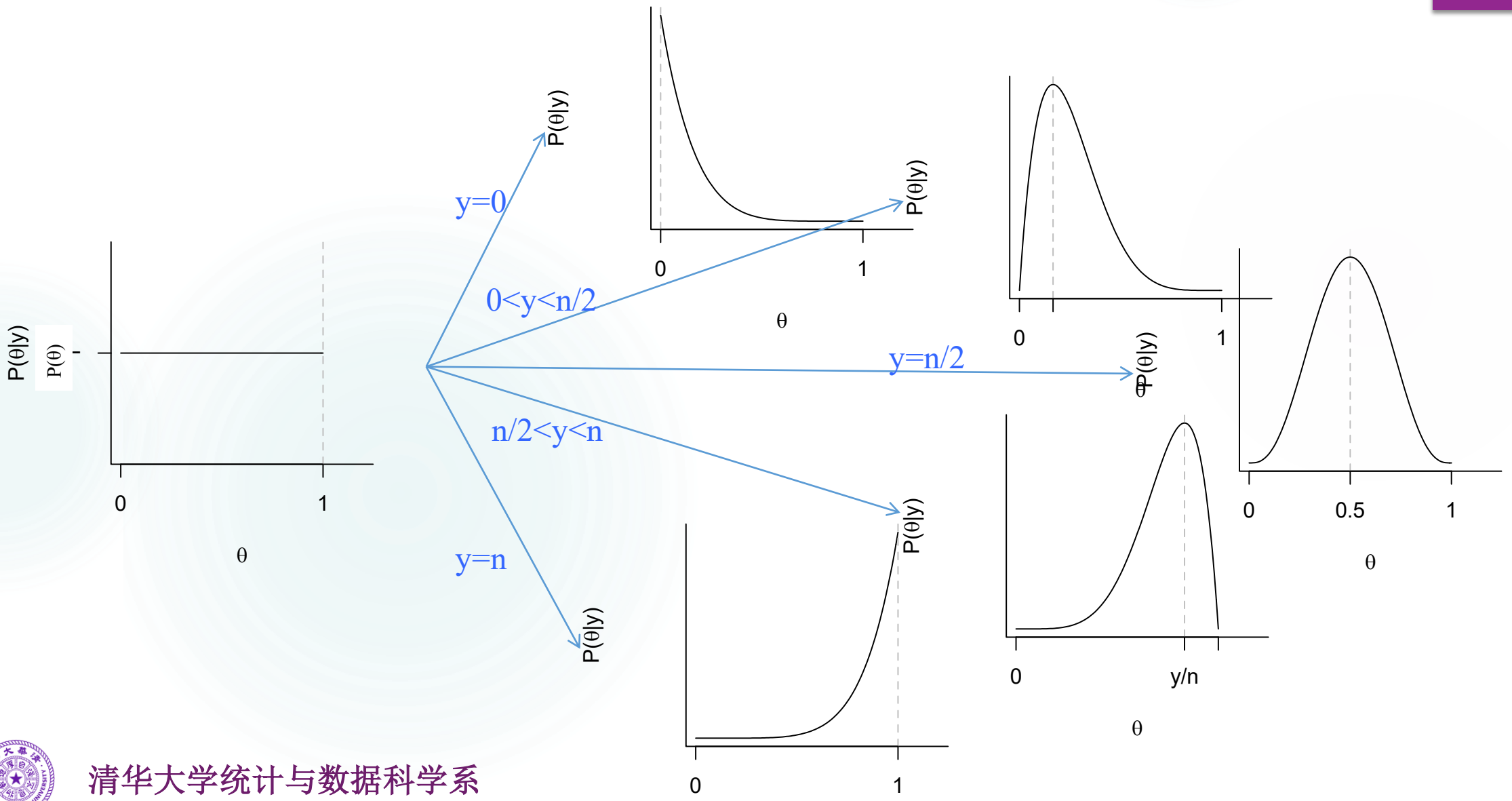
- $p(\theta|y) = (n+1)\binom{n}{y}\theta^y(1-\theta)^{n-y}$. $\theta|y \sim ?$
- The PDF for the for the beta distribution ($Beta(\alpha, \beta)$) is $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$.
- It is often easier to deal with the proportional form of Bayes' Rule

$$p(\theta|y) \propto p(\theta)p(y|\theta)$$

- So in this example,
  - $p(\theta|y) \propto 1 \times \theta^y(1-\theta)^{n-y}$.
  - This is proportional to the density of a $Beta(y+1, n-y+1)$ distribution.

# Example: Binomial Model





# Bayesian Framework of Inference



- ***Generic Notations***

  $p(y \mid \theta)$ --- Sampling distribution or data likelihood

  $p(\theta)$ --- Prior distribution

- ***The Bayes Theorem***

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)} \propto p(\theta)p(y|\theta)$$

Prior

Posterior

Data likelihood

- ***Forecast***

$$\begin{aligned} p(\tilde{y}|y) &= \int p(\tilde{y}, \theta|y)d\theta \\ &= \int p(\tilde{y}|\theta, y)p(\theta|y)d\theta \\ &= \int p(\tilde{y}|\theta)p(\theta|y)d\theta. \end{aligned}$$

Post predictive distribution



# Bayesian versus Frequentist



- ***Advantages***
  - Answers the questions that researchers are usually interested in, "What is the probability that …"
  - Almost all frequentist methods have Bayesian counterparts
  - Easy and consistent to combine information: Formal method for combining prior beliefs with observed (quantitative) information. Natural way of combining information from multiple studies.
  - Easy to deal with real problems with complicate structures
  - Many powerful sampling techniques can be used to support the analysis
- ***Critical issue***
  - Selection of the prior (subjectivity of Bayesian approach)

7 reasons to use Bayesian Inference:
https://statmodeling.stat.columbia.edu/2025/10/11/7-reasons-to-use-bayesian-inference/#comment-2404677

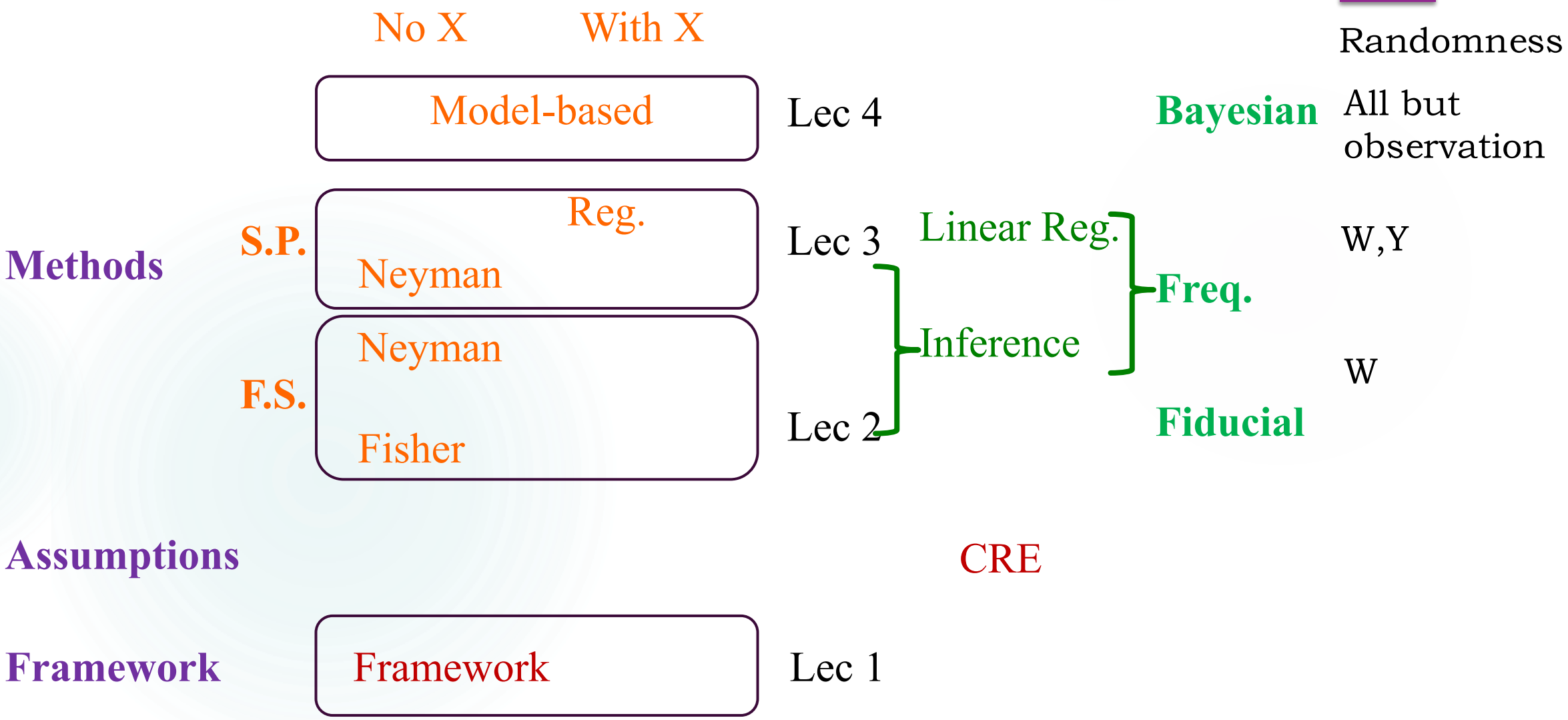


# Bayesian Model-based Imputation - Theory

- NO CORRELATION, NO COVARIATES
- CORRELATION, NO COVARIATES (IN APPENDIX)
- NO CORRELATION, COVARIATES (IN APPENDIX)



# Review


No X
With X
Randomness
Model-based
Lec 4
Bayesian
All but observation
Methods
S.P.
Reg.
Neyman
Lec 3
Linear Reg.
W,Y
Freq.
Inference
F.S.
Neyman
Fisher
Lec 2
W
Fiducial
Assumptions
CRE
Framework
Framework
Lec 1

# Example: First Six Observations from NSW Data



**Table 8.2.** ***First Six Observations from NSW Program Data***

| Unit | Potential Outcomes | | Treatment | Observed Outcome |
|---|---|---|---|---|
| | $Y_i(0)$ | $Y_i(1)$ | $W_i$ | $Y_i^{\text{obs}}$ |
| 1 | 0 | ? | 0 | 0 |
| 2 | ? | 9.9 | 1 | 9.9 |
| 3 | 12.4 | ? | 0 | 12.4 |
| 4 | ? | 3.6 | 1 | 3.6 |
| 5 | 0 | ? | 0 | 0 |
| 6 | ? | 24.9 | 1 | 24.9 |

*Note*: Question marks represent missing potential outcomes.

The target Distribution of $\tau_{\text{fs}} = \frac{1}{6} \cdot \sum_{i=1}^{N} \left( (2 \cdot W_i - 1) \cdot \left( Y_i^{\text{obs}} - Y_i^{\text{mis}} \right) \right).$

$$= \frac{1}{N} \sum_{i=1}^{N} (1 - 2 \cdot W_i) \cdot Y_i^{\text{mis}} + \frac{1}{N} \sum_{i=1}^{N} (2 \cdot W_i - 1) \cdot Y_i^{\text{obs}}$$



# Bayesian Model-Based Imputation



The output

Distribution of $\tau = \tau(\mathbf{Y}(0), \mathbf{Y}(1))$

$(\mathbf{Y}(0), \mathbf{Y}(1)) = h(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}, \mathbf{W})$

$\tau(Y^{mis}, Y^{obs}, W)$

$$= \frac{1}{N}\sum_{i=1}^{N}(1-2\cdot W_i)\cdot Y_i^{\text{mis}} + \frac{1}{N}\sum_{i=1}^{N}(2\cdot W_i - 1)\cdot Y_i^{\text{obs}}$$

$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}}, \mathbf{W})$ ? $f(\tau|\mathbf{Y}^{\text{obs}}, \mathbf{W})$

$= f(\tau(Y^{mis}, Y^{obs}, W)|Y^{obs}, W)$

The 3 inputs

$$\Pr(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$f(\mathbf{Y}(0), \mathbf{Y}(1)|\theta)$

$p(\theta)$



# Theory (Part 1)



The inputs

$$\Pr(\mathbf{W}|\mathbf{Y}(0),\mathbf{Y}(1)) = \binom{N}{N_t}^{-1} \quad f(\mathbf{Y}(0),\mathbf{Y}(1)|\theta) \quad p(\theta)$$

$$f(\mathbf{Y}^{\text{mis}},\theta|\mathbf{Y}^{\text{obs}},\mathbf{W}) = f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W},\theta)\cdot p(\theta|\mathbf{Y}^{\text{obs}},\mathbf{W}).$$

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W}) = \int_{\theta} f(\mathbf{Y}^{\text{mis}},\theta|\mathbf{Y}^{\text{obs}},\mathbf{W})\,\mathrm{d}\theta$$

The target

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W})$$



# Theory (Part 2.1)



The inputs

$$\Pr(\mathbf{W}|\mathbf{Y}(0),\mathbf{Y}(1)) = \binom{N}{N_t}^{-1} \quad f(\mathbf{Y}(0),\mathbf{Y}(1)|\theta) \quad p(\theta)$$

$$f(\mathbf{Y}^{\text{obs}},\mathbf{W}|\theta) = \int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}},\mathbf{Y}^{\text{obs}},\mathbf{W}|\theta)\,\mathrm{d}\mathbf{y}^{\text{mis}}$$

$$p(\theta|\mathbf{Y}^{\text{obs}},\mathbf{W}) = \frac{p(\theta)\cdot f(\mathbf{Y}^{\text{obs}},\mathbf{W}|\theta)}{f(\mathbf{Y}^{\text{obs}},\mathbf{W})}$$

Normalizing constant:

$$f(\mathbf{Y}^{\text{obs}},\mathbf{W}) = \int_{\theta} p(\theta)\cdot f(\mathbf{Y}^{\text{obs}},\mathbf{W}|\theta)\mathrm{d}\theta$$

$$f(\mathbf{Y}^{\text{mis}},\theta|\mathbf{Y}^{\text{obs}},\mathbf{W}) = f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W},\theta)\cdot p(\theta|\mathbf{Y}^{\text{obs}},\mathbf{W}).$$

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W}) = \int_{\theta} f(\mathbf{Y}^{\text{mis}},\theta|\mathbf{Y}^{\text{obs}},\mathbf{W})\,\mathrm{d}\theta$$

The target

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W})$$



# Theory (Part 2.2)



The inputs

$$\Pr(\mathbf{W}|\mathbf{Y}(0),\mathbf{Y}(1)) = \binom{N}{N_t}^{-1}$$

$$f(\mathbf{Y}(0),\mathbf{Y}(1)|\theta)$$

$$p(\theta)$$

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W},\theta) = \frac{f(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)}{f(\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)} = \frac{f(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)}{\int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)\mathrm{d}\mathbf{y}^{\text{mis}}}$$

$$f(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)$$

$$f(\mathbf{Y}^{\text{mis}},\theta|\mathbf{Y}^{\text{obs}},\mathbf{W}) = f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W},\theta)\cdot p(\theta|\mathbf{Y}^{\text{obs}},\mathbf{W}).$$

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W}) = \int_{\theta} f(\mathbf{Y}^{\text{mis}},\theta|\mathbf{Y}^{\text{obs}},\mathbf{W})\,\mathrm{d}\theta$$

The target

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W})$$

# Theory (Interim Summary)



The inputs

$P(\mathbf{W})$ $\quad f(\mathbf{Y}^{mis}, \mathbf{Y}^{obs}, \mathbf{W}|\theta)$

$P(\boldsymbol{\theta})$ $\quad f(\mathbf{Y}^{mis}, \mathbf{Y}^{obs}|\mathbf{W}, \theta)$

Step 1

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}}, \mathbf{W}, \theta) = \frac{f(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}|\mathbf{W}, \theta)}{f(\mathbf{Y}^{\text{obs}}|\mathbf{W}, \theta)} = \frac{f(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}|\mathbf{W}, \theta)}{\int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}|\mathbf{W}, \theta)\mathrm{d}\mathbf{y}^{\text{mis}}}$$

$$f(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}|\mathbf{W}, \theta)$$

Step 2

$$f(\mathbf{Y}^{\text{obs}}, \mathbf{W}|\theta) = \int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}, \mathbf{W}|\theta)\, \mathrm{d}\mathbf{y}^{\text{mis}}$$

$$p(\theta|\mathbf{Y}^{\text{obs}}, \mathbf{W}) = \frac{p(\theta) \cdot f(\mathbf{Y}^{\text{obs}}, \mathbf{W}|\theta)}{f(\mathbf{Y}^{\text{obs}}, \mathbf{W})}$$

Normalizing constant:

$$f(\mathbf{Y}^{\text{obs}}, \mathbf{W}) = \int_{\theta} p(\theta) \cdot f(\mathbf{Y}^{\text{obs}}, \mathbf{W}|\theta)\mathrm{d}\theta$$

Step 3

$$f(\mathbf{Y}^{\text{mis}}, \theta|\mathbf{Y}^{\text{obs}}, \mathbf{W}) = f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}}, \mathbf{W}, \theta) \cdot p(\theta|\mathbf{Y}^{\text{obs}}, \mathbf{W}).$$

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}}, \mathbf{W}) = \int_{\theta} f(\mathbf{Y}^{\text{mis}}, \theta|\mathbf{Y}^{\text{obs}}, \mathbf{W})\, \mathrm{d}\theta$$

The target

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}}, \mathbf{W})$$

# Theory (part 3)



$$(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}) = g(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W})$$

$$f(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W}|\theta) = \Pr(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1), \theta) \cdot f(\mathbf{Y}(0), \mathbf{Y}(1)|\theta). \Rightarrow f(\mathbf{Y}^{mis}, \mathbf{Y}^{obs}, \mathbf{W}|\theta)$$

$$f(\mathbf{Y}(0), \mathbf{Y}(1)|\mathbf{W}, \theta) = \frac{f(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W}|\theta)}{\Pr(\mathbf{W}|\theta)} = \frac{f(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W}|\theta)}{\int f(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W}|\theta) d\mathbf{Y}(0) d\mathbf{Y}(1)}.$$

$$(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}) = g(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W})$$

$$f(\mathbf{Y}^{mis}, \mathbf{Y}^{obs}|\mathbf{W}, \theta)$$

The 3 inputs

$$\Pr(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$$f(\mathbf{Y}(0), \mathbf{Y}(1)|\theta)$$

$$p(\theta)$$

# Example for Theory (part 3)



$$f(\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{W}|\theta) = \Pr(\mathbf{W}|\mathbf{Y}(0),\mathbf{Y}(1),\theta)\cdot f(\mathbf{Y}(0),\mathbf{Y}(1)|\theta).$$

$$f(\mathbf{Y}(0),\mathbf{Y}(1)|\mathbf{W},\theta) = \frac{f(\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{W}|\theta)}{\Pr(\mathbf{W}|\theta)} = \frac{f(\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{W}|\theta)}{\int f(\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{W}|\theta)\mathrm{d}\mathbf{Y}(0)\mathrm{d}\mathbf{Y}(1)}.$$

$$(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}) = g(\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{W})$$

$$f(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)$$

$$\begin{pmatrix} Y_i^{\text{mis}} \\ Y_i^{\text{obs}} \end{pmatrix} \Bigg| \mu_c,\mu_t,\mathbf{W} \sim \mathcal{N}\left(\begin{pmatrix} W_i\cdot\mu_c+(1-W_i)\cdot\mu_t \\ (1-W_i)\cdot\mu_c+W_i\cdot\mu_t \end{pmatrix}, \begin{pmatrix} W_i\cdot 100+(1-W_i)\cdot 64 & 0 \\ 0 & (1-W_i)\cdot 100+W_i\cdot 64 \end{pmatrix}\right)$$

The 3 inputs

$$\Pr(\mathbf{W}|\mathbf{Y}(0),\mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \Bigg| \theta \sim \mathcal{N}\left(\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} 100 & 0 \\ 0 & 64 \end{pmatrix}\right)$$

$$\theta = \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 10{,}000 & 0 \\ 0 & 10{,}000 \end{pmatrix}\right)$$

$$f(\mathbf{y}^{\text{mis}},\mathbf{Y}^{\text{obs}},\mathbf{W}|\theta)$$

# Theory (Summary)



The output

Distribution of $\tau = \tau(\mathbf{Y}(0), \mathbf{Y}(1))$

$(\mathbf{Y}(0), \mathbf{Y}(1)) = h(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}, \mathbf{W})$

$\tau(Y^{mis}, Y^{obs}, W)$

$$= \frac{1}{N}\sum_{i=1}^{N}(1 - 2 \cdot W_i) \cdot Y_i^{\text{mis}} + \frac{1}{N}\sum_{i=1}^{N}(2 \cdot W_i - 1) \cdot Y_i^{\text{obs}}$$

$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}}, \mathbf{W})$ $\Longleftrightarrow$ $f(\tau|\mathbf{Y}^{\text{obs}}, \mathbf{W})$

$= f(\tau(Y^{mis}, Y^{obs}, W)|Y^{obs}, W)$

The 3 inputs

$$\Pr(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$f(\mathbf{Y}(0), \mathbf{Y}(1)|\theta)$

$p(\theta)$

$f(\mathbf{y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}, \mathbf{W}|\theta)$

$f(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}|\mathbf{W}, \theta)$

# Example: First Six Observations from NSW Data



**Table 8.2.** ***First Six Observations from NSW Program Data***

| Unit | Potential Outcomes | | Treatment | Observed Outcome |
|---|---|---|---|---|
| | $Y_i(0)$ | $Y_i(1)$ | $W_i$ | $Y_i^{\text{obs}}$ |
| 1 | 0 | ? | 0 | 0 |
| 2 | ? | 9.9 | 1 | 9.9 |
| 3 | 12.4 | ? | 0 | 12.4 |
| 4 | ? | 3.6 | 1 | 3.6 |
| 5 | 0 | ? | 0 | 0 |
| 6 | ? | 24.9 | 1 | 24.9 |

*Note*: Question marks represent missing potential outcomes.

The target Distribution of $\tau_{\text{fs}} = \frac{1}{6} \cdot \sum_{i=1}^{N} \left( (2 \cdot W_i - 1) \cdot \left( Y_i^{\text{obs}} - Y_i^{\text{mis}} \right) \right)$.



# An Analytic Example with Six Units



$$\Pr(\mathbf{W}|\mathbf{Y}(0),\mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \Big| \theta \sim \mathcal{N}\left( \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} 100 & 0 \\ 0 & 64 \end{pmatrix} \right)$$

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 10{,}000 & 0 \\ 0 & 10{,}000 \end{pmatrix} \right)$$

Step 1

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W},\theta) = \frac{f(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)}{f(\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)} = \frac{f(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)}{\int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)\mathrm{d}\mathbf{y}^{\text{mis}}}$$

$$\downarrow$$

$$f(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)$$

$$f(\mathbf{Y}(0),\mathbf{Y}(1)|\mu_c,\mu_t) = \prod_{i=1}^{N} f(Y_i(0),Y_i(1)|\mu_c,\mu_t)$$

$$f(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta) \downarrow$$

$$\begin{pmatrix} Y_i^{\text{mis}} \\ Y_i^{\text{obs}} \end{pmatrix} \Big| \mu_c,\mu_t,\mathbf{W} \sim \mathcal{N}\left( \begin{pmatrix} W_i\cdot\mu_c + (1-W_i)\cdot\mu_t \\ (1-W_i)\cdot\mu_c + W_i\cdot\mu_t \end{pmatrix}, \begin{pmatrix} W_i\cdot 100 + (1-W_i)\cdot 64 & 0 \\ 0 & (1-W_i)\cdot 100 + W_i\cdot 64 \end{pmatrix} \right)$$

$$\downarrow$$

$$f(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\mu_c,\mu_t) = \prod_{i=1}^{N} f(Y_i^{\text{mis}},Y_i^{\text{obs}}|\mathbf{W},\mu_c,\mu_t)$$

$$\downarrow$$

$$Y_i^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W},\mu_c,\mu_t \sim \mathcal{N}\left(W_i\cdot\mu_c + (1-W_i)\cdot\mu_t, W_i\cdot 100 + (1-W_i)\cdot 64\right)$$

# An Analytic Example with Six Units



$$\Pr(\mathbf{W}|\mathbf{Y}(0),\mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \Big| \theta \sim \mathcal{N}\left( \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} 100 & 0 \\ 0 & 64 \end{pmatrix} \right)$$

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 10{,}000 & 0 \\ 0 & 10{,}000 \end{pmatrix} \right)$$

Step 1

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W},\theta) = \frac{f(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)}{f(\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)} = \frac{f(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)}{\int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)\mathrm{d}\mathbf{y}^{\text{mis}}}$$

$$\downarrow$$

$$f(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)$$

$$\begin{pmatrix} Y_1^{\text{mis}} \\ Y_2^{\text{mis}} \\ Y_3^{\text{mis}} \\ Y_4^{\text{mis}} \\ Y_5^{\text{mis}} \\ Y_6^{\text{mis}} \end{pmatrix} \Bigg| \mathbf{Y}^{\text{obs}}, \mathbf{W}, \mu_c, \mu_t \sim \mathcal{N}\left( \begin{pmatrix} \mu_t \\ \mu_c \\ \mu_t \\ \mu_c \\ \mu_t \\ \mu_c \end{pmatrix}, \begin{pmatrix} 64 & 0 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 & 0 \\ 0 & 0 & 64 & 0 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100 \end{pmatrix} \right)$$

$$\uparrow$$

$$Y_i^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W},\mu_c,\mu_t \sim \mathcal{N}\left(W_i \cdot \mu_c + (1-W_i)\cdot \mu_t, W_i \cdot 100 + (1-W_i)\cdot 64\right)$$

# An Analytic Example with Six Units



$$\Pr(\mathbf{W}|\mathbf{Y}(0),\mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \Big| \theta \sim \mathcal{N}\left( \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} 100 & 0 \\ 0 & 64 \end{pmatrix} \right)$$

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 10{,}000 & 0 \\ 0 & 10{,}000 \end{pmatrix} \right)$$

Step 2

$$f(\mathbf{Y}^{\text{obs}}, \mathbf{W}|\theta) = \int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}, \mathbf{W}|\theta)\, \mathrm{d}\mathbf{y}^{\text{mis}}$$

$$p(\theta|\mathbf{Y}^{\text{obs}}, \mathbf{W}) = \frac{p(\theta) \cdot f(\mathbf{Y}^{\text{obs}}, \mathbf{W}|\theta)}{f(\mathbf{Y}^{\text{obs}}, \mathbf{W})}$$

Normalizing constant:

$$f(\mathbf{Y}^{\text{obs}}, \mathbf{W}) = \int_{\theta} p(\theta) \cdot f(\mathbf{Y}^{\text{obs}}, \mathbf{W}|\theta) \mathrm{d}\theta$$

$$\begin{pmatrix} Y_i^{\text{mis}} \\ Y_i^{\text{obs}} \end{pmatrix} \Big| \mu_c, \mu_t, \mathbf{W} \sim \mathcal{N}\left( \begin{pmatrix} W_i \cdot \mu_c + (1 - W_i) \cdot \mu_t \\ (1 - W_i) \cdot \mu_c + W_i \cdot \mu_t \end{pmatrix}, \begin{pmatrix} W_i \cdot 100 + (1 - W_i) \cdot 64 & 0 \\ 0 & (1 - W_i) \cdot 100 + W_i \cdot 64 \end{pmatrix} \right)$$

$$Y_i^{\text{obs}} | \mathbf{W}, \mu_c, \mu_t \sim \mathcal{N}\left( (1 - W_i) \cdot \mu_c + W_i \cdot \mu_t, (1 - W_i) \cdot 100 + W_i \cdot 64 \right)$$

$f(\mathbf{Y}^{\text{obs}}, \mathbf{W}|\theta)$

$$= \mathcal{L}(\mu_c, \mu_t | \mathbf{Y}^{\text{obs}}, \mathbf{W}) \propto \prod_{i=1}^{6} \left\{ \frac{1}{\sqrt{2\pi \cdot ((1 - W_i) \cdot 100 + W_i \cdot 64)}} \times \exp\left[ -\frac{1}{2} \left( \frac{1}{(1 - W_i) \cdot 100 + W_i \cdot 64} (Y_i^{\text{obs}} - (1 - W_i) \cdot \mu_c - W_i \cdot \mu_t)^2 \right) \right] \right\}$$

$$\propto \prod_{i:W_i=0} \frac{1}{\sqrt{2\pi \cdot 100}} \exp\left[ -\frac{1}{2} \left( \frac{1}{100} (Y_i^{\text{obs}} - \mu_c)^2 \right) \right] \times \prod_{i:W_i=1} \frac{1}{\sqrt{2\pi \cdot 64}} \exp\left[ -\frac{1}{2} \left( \frac{1}{64} (Y_i^{\text{obs}} - \mu_t)^2 \right) \right]$$

# An Analytic Example with Six Units



$$\Pr(\mathbf{W}|\mathbf{Y}(0),\mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \Big| \theta \sim \mathcal{N}\left( \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} 100 & 0 \\ 0 & 64 \end{pmatrix} \right)$$

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 10{,}000 & 0 \\ 0 & 10{,}000 \end{pmatrix} \right)$$

Step 2

$$f(\mathbf{Y}^{\text{obs}}, \mathbf{W}|\theta) = \int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}, \mathbf{W}|\theta)\, d\mathbf{y}^{\text{mis}}$$

$$p(\theta|\mathbf{Y}^{\text{obs}}, \mathbf{W}) = \frac{p(\theta) \cdot f(\mathbf{Y}^{\text{obs}}, \mathbf{W}|\theta)}{f(\mathbf{Y}^{\text{obs}}, \mathbf{W})}$$

Normalizing constant:

$$f(\mathbf{Y}^{\text{obs}}, \mathbf{W}) = \int_{\theta} p(\theta) \cdot f(\mathbf{Y}^{\text{obs}}, \mathbf{W}|\theta) d\theta$$

$$f(\mathbf{Y}^{\text{obs}}, \mathbf{W}|\theta) \propto \prod_{i:W_i=0} \frac{1}{\sqrt{2\pi \cdot 100}} \exp\left[ -\frac{1}{2} \left( \frac{1}{100} (Y_i^{\text{obs}} - \mu_c)^2 \right) \right] \times \prod_{i:W_i=1} \frac{1}{\sqrt{2\pi \cdot 64}} \exp\left[ -\frac{1}{2} \left( \frac{1}{64} (Y_i^{\text{obs}} - \mu_t)^2 \right) \right]$$

$$p(\mu_c, \mu_t|\mathbf{Y}^{\text{obs}}, \mathbf{W}) \propto \exp\left[ -\frac{1}{2} \left( \frac{\mu_c^2}{10{,}000} \right) \right] \cdot \prod_{i:W_i=0} \frac{1}{\sqrt{2\pi \cdot 100}} \exp\left[ -\frac{1}{2} \left( \frac{(Y_i^{\text{obs}} - \mu_c)^2}{100} \right) \right] \times \exp\left[ -\frac{1}{2} \left( \frac{\mu_t^2}{10{,}000} \right) \right] \cdot \prod_{i:W_i=1} \frac{1}{\sqrt{2\pi \cdot 64}} \exp\left[ -\frac{1}{2} \left( \frac{(Y_i^{\text{obs}} - \mu_t)^2}{64} \right) \right]$$

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \Big| \mathbf{Y}^{\text{obs}}, \mathbf{W} \sim \mathcal{N}\left( \begin{pmatrix} \overline{Y}_{\text{c}}^{\text{obs}} \cdot \dfrac{N_{\text{c}} \cdot 10{,}000}{N_{\text{c}} \cdot 10{,}000 + 100} \\ \overline{Y}_{\text{t}}^{\text{obs}} \cdot \dfrac{N_{\text{t}} \cdot 10{,}000}{N_{\text{t}} \cdot 10{,}000 + 64} \end{pmatrix}, \begin{pmatrix} \dfrac{1}{N_{\text{c}}/100 + 1/10{,}000} & 0 \\ 0 & \dfrac{1}{N_{\text{t}}/64 + 1/10{,}000} \end{pmatrix} \right)$$

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \Big| \mathbf{Y}^{\text{obs}}, \mathbf{W} \sim \mathcal{N}\left( \begin{pmatrix} 4.1 \\ 12.8 \end{pmatrix}, \begin{pmatrix} 5.8^2 & 0 \\ 0 & 4.6^2 \end{pmatrix} \right)$$

# An Analytic Example with Six Units



Step 1

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W},\theta)=\frac{f(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)}{f(\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)}=\frac{f(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)}{\int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)\mathrm{d}\mathbf{y}^{\text{mis}}}$$

$$\begin{pmatrix} Y_1^{\text{mis}} \\ Y_2^{\text{mis}} \\ Y_3^{\text{mis}} \\ Y_4^{\text{mis}} \\ Y_5^{\text{mis}} \\ Y_6^{\text{mis}} \end{pmatrix} \Bigg| \mathbf{Y}^{\text{obs}},\mathbf{W},\mu_c,\mu_t \sim \mathcal{N}\left( \begin{pmatrix} \mu_t \\ \mu_c \\ \mu_t \\ \mu_c \\ \mu_t \\ \mu_c \end{pmatrix}, \begin{pmatrix} 64 & 0 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 & 0 \\ 0 & 0 & 64 & 0 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100 \end{pmatrix} \right)$$

Step 2

$$p(\theta|\mathbf{Y}^{\text{obs}},\mathbf{W})=\frac{p(\theta)\cdot\mathcal{L}(\theta|\mathbf{Y}^{\text{obs}},\mathbf{W})}{f(\mathbf{Y}^{\text{obs}},\mathbf{W})}$$

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \Bigg| \mathbf{Y}^{\text{obs}},\mathbf{W} \sim \mathcal{N}\left( \begin{pmatrix} 4.1 \\ 12.8 \end{pmatrix}, \begin{pmatrix} 5.8^2 & 0 \\ 0 & 4.6^2 \end{pmatrix} \right)$$

Step 3

$$f(\mathbf{Y}^{\text{mis}},\theta|\mathbf{Y}^{\text{obs}},\mathbf{W})=f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W},\theta)\cdot p(\theta|\mathbf{Y}^{\text{obs}},\mathbf{W})$$

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W})=\int_\theta f(\mathbf{Y}^{\text{mis}},\theta|\mathbf{Y}^{\text{obs}},\mathbf{W})\,\mathrm{d}\theta$$

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W}) \sim \mathcal{N}(*,*)$$

# An Analytic Example with Six Units



Step 1

$$\begin{pmatrix} Y_1^{\text{mis}} \\ Y_2^{\text{mis}} \\ Y_3^{\text{mis}} \\ Y_4^{\text{mis}} \\ Y_5^{\text{mis}} \\ Y_6^{\text{mis}} \end{pmatrix} \Bigg| \mathbf{Y}^{\text{obs}}, \mathbf{W}, \mu_c, \mu_t \sim \mathcal{N}\left( \begin{pmatrix} \mu_t \\ \mu_c \\ \mu_t \\ \mu_c \\ \mu_t \\ \mu_c \end{pmatrix}, \begin{pmatrix} 64 & 0 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 & 0 \\ 0 & 0 & 64 & 0 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100 \end{pmatrix} \right)$$

Step 2

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \Bigg| \mathbf{Y}^{\text{obs}}, \mathbf{W} \sim \mathcal{N}\left( \begin{pmatrix} 4.1 \\ 12.8 \end{pmatrix}, \begin{pmatrix} 5.8^2 & 0 \\ 0 & 4.6^2 \end{pmatrix} \right)$$

Step 3

$$f(\mathbf{Y}^{\text{mis}}, \theta | \mathbf{Y}^{\text{obs}}, \mathbf{W}) = f(\mathbf{Y}^{\text{mis}} | \mathbf{Y}^{\text{obs}}, \mathbf{W}, \theta) \cdot p(\theta | \mathbf{Y}^{\text{obs}}, \mathbf{W}).$$

$$f(\mathbf{Y}^{\text{mis}} | \mathbf{Y}^{\text{obs}}, \mathbf{W}) \sim \mathcal{N}(*,*)$$

$$\mathbb{E}\left[ Y_i^{\text{mis}} \middle| \mathbf{Y}^{\text{obs}}, \mathbf{W}, \mu_c, \mu_t \right] = W_i \cdot \mu_c + (1 - W_i) \cdot \mu_t$$

$$\mathbb{E}\left[ \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \middle| \mathbf{Y}^{\text{obs}}, \mathbf{W} \right] = \begin{pmatrix} \overline{Y}_{\text{c}}^{\text{obs}} \cdot \dfrac{N_{\text{c}} \cdot 10{,}000}{N_{\text{c}} \cdot 10{,}000 + 100} \\ \overline{Y}_{\text{t}}^{\text{obs}} \cdot \dfrac{N_{\text{t}} \cdot 10{,}000}{N_{\text{t}} \cdot 10{,}000 + 64} \end{pmatrix}$$

$$\mathbb{E}\left[ Y_i^{\text{mis}} | \mathbf{Y}^{\text{obs}}, \mathbf{W} \right] = W_i \cdot \left( \overline{Y}_{\text{c}}^{\text{obs}} \cdot \frac{N_{\text{c}} \cdot 10{,}000}{N_{\text{c}} \cdot 10{,}000 + 100} \right) + (1 - W_i) \cdot \left( \overline{Y}_{\text{t}}^{\text{obs}} \cdot \frac{N_{\text{t}} \cdot 10{,}000}{N_{\text{t}} \cdot 10{,}000 + 64} \right)$$

$$= 4.1\, W_i + 12.8\, (1 - W_i)$$

$$\mathbb{V}\left( Y_i^{\text{mis}} \middle| \mathbf{Y}^{\text{obs}}, \mathbf{W} \right) = \mathbb{E}\left[ \mathbb{V}\left( Y_i^{\text{mis}} \middle| \mathbf{Y}^{\text{obs}}, \mathbf{W}, \mu_c, \mu_t \right) \middle| \mathbf{Y}^{\text{obs}}, \mathbf{W} \right] + \mathbb{V}\left( \mathbb{E}\left[ Y_i^{\text{mis}} \middle| \mathbf{Y}^{\text{obs}}, \mathbf{W}, \mu_c, \mu_t \right] \middle| \mathbf{Y}^{\text{obs}}, \mathbf{W} \right)$$

$$\mathbb{C}\left( Y_i^{\text{mis}}, Y_{i'}^{\text{mis}} \middle| \mathbf{Y}^{\text{obs}}, \mathbf{W} \right) = \mathbb{E}\left[ \mathbb{C}\left( Y_i^{\text{mis}}, Y_{i'}^{\text{mis}} \middle| \mathbf{Y}^{\text{obs}}, \mathbf{W}, \mu_c, \mu_t \right) \middle| \mathbf{Y}^{\text{obs}}, \mathbf{W} \right] + \mathbb{C}\left( \mathbb{E}\left[ Y_i^{\text{mis}} \middle| \mathbf{Y}^{\text{obs}}, \mathbf{W}, \mu_c, \mu_t \right], \mathbb{E}\left[ Y_{i'}^{\text{mis}} \middle| \mathbf{Y}^{\text{obs}}, \mathbf{W}, \mu_c, \mu_t \right] \middle| \mathbf{Y}^{\text{obs}}, \mathbf{W} \right)$$

# An Analytic Example with Six Units



Step 1

$$\begin{pmatrix} Y_1^{\text{mis}} \\ Y_2^{\text{mis}} \\ Y_3^{\text{mis}} \\ Y_4^{\text{mis}} \\ Y_5^{\text{mis}} \\ Y_6^{\text{mis}} \end{pmatrix} \Bigg| \mathbf{Y}^{\text{obs}}, \mathbf{W}, \mu_c, \mu_t \sim \mathcal{N}\left( \begin{pmatrix} \mu_t \\ \mu_c \\ \mu_t \\ \mu_c \\ \mu_t \\ \mu_c \end{pmatrix}, \begin{pmatrix} 64 & 0 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 & 0 \\ 0 & 0 & 64 & 0 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100 \end{pmatrix} \right)$$

Step 2

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \Bigg| \mathbf{Y}^{\text{obs}}, \mathbf{W} \sim \mathcal{N}\left( \begin{pmatrix} 4.1 \\ 12.8 \end{pmatrix}, \begin{pmatrix} 5.8^2 & 0 \\ 0 & 4.6^2 \end{pmatrix} \right)$$

Step 3

$$f(\mathbf{Y}^{\text{mis}}, \theta | \mathbf{Y}^{\text{obs}}, \mathbf{W}) = f(\mathbf{Y}^{\text{mis}} | \mathbf{Y}^{\text{obs}}, \mathbf{W}, \theta) \cdot p(\theta | \mathbf{Y}^{\text{obs}}, \mathbf{W}).$$

$$f(\mathbf{Y}^{\text{mis}} | \mathbf{Y}^{\text{obs}}, \mathbf{W}) \sim \mathcal{N}(*,*)$$

$$\begin{pmatrix} Y_1^{\text{mis}} \\ Y_2^{\text{mis}} \\ Y_3^{\text{mis}} \\ Y_4^{\text{mis}} \\ Y_5^{\text{mis}} \\ Y_6^{\text{mis}} \end{pmatrix} \Bigg| \mathbf{Y}^{\text{obs}}, \mathbf{W} \sim \mathcal{N}\left( \begin{pmatrix} 12.8 \\ 4.1 \\ 12.8 \\ 4.1 \\ 12.8 \\ 4.1 \end{pmatrix}, \begin{pmatrix} 85.3 & 0 & 21.3 & 0 & 21.3 & 0 \\ 0 & 133.2 & 0 & 33.2 & 0 & 33.2 \\ 21.3 & 0 & 85.3 & 0 & 21.3 & 0 \\ 0 & 0 & 0 & 133.2 & 0 & 33.2 \\ 21.3 & 0 & 21.3 & 0 & 85.3 & 0 \\ 0 & 33.2 & 0 & 33.2 & 0 & 133.2 \end{pmatrix} \right)$$

# An Analytic Example with Six Units



The output

Distribution of $\tau = \tau(\mathbf{Y}(0), \mathbf{Y}(1))$

$$(\mathbf{Y}(0), \mathbf{Y}(1)) = h(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}, \mathbf{W})$$

$$\tau(Y^{mis}, Y^{obs}, W)$$

$$= \frac{1}{N}\sum_{i=1}^{N}(1 - 2 \cdot W_i) \cdot Y_i^{\text{mis}} + \frac{1}{N}\sum_{i=1}^{N}(2 \cdot W_i - 1) \cdot Y_i^{\text{obs}}$$

$f(\mathbf{Y}^{\text{mis}} | \mathbf{Y}^{\text{obs}}, \mathbf{W})$ $\Longleftrightarrow$ $f(\tau | \mathbf{Y}^{\text{obs}}, \mathbf{W})$ Normal dist.

$$\mathbb{E}\left[\tau_{\text{fs}} \middle| \mathbf{Y}^{\text{obs}}, \mathbf{W}\right] \qquad \mathbb{V}\left(\tau_{\text{fs}} \middle| \mathbf{Y}^{\text{obs}}, \mathbf{W}\right)$$

$$\begin{pmatrix} Y_1^{\text{mis}} \\ Y_2^{\text{mis}} \\ Y_3^{\text{mis}} \\ Y_4^{\text{mis}} \\ Y_5^{\text{mis}} \\ Y_6^{\text{mis}} \end{pmatrix} \middle| \mathbf{Y}^{\text{obs}}, \mathbf{W} \sim \mathcal{N}\left( \begin{pmatrix} 12.8 \\ 4.1 \\ 12.8 \\ 4.1 \\ 12.8 \\ 4.1 \end{pmatrix}, \begin{pmatrix} 85.3 & 0 & 21.3 & 0 & 21.3 & 0 \\ 0 & 133.2 & 0 & 33.2 & 0 & 33.2 \\ 21.3 & 0 & 85.3 & 0 & 21.3 & 0 \\ 0 & 0 & 0 & 133.2 & 0 & 33.2 \\ 21.3 & 0 & 21.3 & 0 & 85.3 & 0 \\ 0 & 33.2 & 0 & 33.2 & 0 & 133.2 \end{pmatrix} \right)$$



# An Analytic Example with Six Units - Summary



$$\tau = \tau(\mathbf{Y}(0), \mathbf{Y}(1))$$
$$(\mathbf{Y}(0), \mathbf{Y}(1)) = h(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}, \mathbf{W})$$

$$\tau(Y^{mis}, Y^{obs}, W)$$

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}}, \mathbf{W}) \quad \Longleftrightarrow \quad f(\tau|\mathbf{Y}^{\text{obs}}, \mathbf{W})$$

$$\tau_{\text{fs}}|\mathbf{Y}^{\text{obs}}, \mathbf{W} \sim \mathcal{N}\left(8.7, 5.2^2\right)$$

Step 4

Step 3

$$\begin{pmatrix} Y_1^{\text{mis}} \\ Y_2^{\text{mis}} \\ Y_3^{\text{mis}} \\ Y_4^{\text{mis}} \\ Y_5^{\text{mis}} \\ Y_6^{\text{mis}} \end{pmatrix} \Bigg| \mathbf{Y}^{\text{obs}}, \mathbf{W} \sim \mathcal{N}\left( \begin{pmatrix} 12.8 \\ 4.1 \\ 12.8 \\ 4.1 \\ 12.8 \\ 4.1 \end{pmatrix}, \begin{pmatrix} 85.3 & 0 & 21.3 & 0 & 21.3 & 0 \\ 0 & 133.2 & 0 & 33.2 & 0 & 33.2 \\ 21.3 & 0 & 85.3 & 0 & 21.3 & 0 \\ 0 & 0 & 0 & 133.2 & 0 & 33.2 \\ 21.3 & 0 & 21.3 & 0 & 85.3 & 0 \\ 0 & 33.2 & 0 & 33.2 & 0 & 133.2 \end{pmatrix} \right)$$

$$p(\theta|\mathbf{Y}^{\text{obs}}, \mathbf{W}) = \frac{p(\theta) \cdot \mathcal{L}(\theta|\mathbf{Y}^{\text{obs}}, \mathbf{W})}{f(\mathbf{Y}^{\text{obs}}, \mathbf{W})}$$

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \Bigg| \mathbf{Y}^{\text{obs}}, \mathbf{W} \sim \mathcal{N}\left( \begin{pmatrix} 4.1 \\ 12.8 \end{pmatrix}, \begin{pmatrix} 5.8^2 & 0 \\ 0 & 4.6^2 \end{pmatrix} \right)$$

Step 2

Step 1

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}}, \mathbf{W}, \theta) = \frac{f(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}|\mathbf{W}, \theta)}{f(\mathbf{Y}^{\text{obs}}|\mathbf{W}, \theta)} = \frac{f(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}|\mathbf{W}, \theta)}{\int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}|\mathbf{W}, \theta) \mathrm{d}\mathbf{y}^{\text{mis}}}$$

$$\begin{pmatrix} Y_1^{\text{mis}} \\ Y_2^{\text{mis}} \\ Y_3^{\text{mis}} \\ Y_4^{\text{mis}} \\ Y_5^{\text{mis}} \\ Y_6^{\text{mis}} \end{pmatrix} \Bigg| \mathbf{Y}^{\text{obs}}, \mathbf{W}, \mu_c, \mu_t \sim \mathcal{N}\left( \begin{pmatrix} \mu_t \\ \mu_c \\ \mu_t \\ \mu_c \\ \mu_t \\ \mu_c \end{pmatrix}, \begin{pmatrix} 64 & 0 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 & 0 \\ 0 & 0 & 64 & 0 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100 \end{pmatrix} \right)$$

$$\Pr(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \Big| \theta \sim \mathcal{N}\left( \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} 100 & 0 \\ 0 & 64 \end{pmatrix} \right)$$

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 10{,}000 & 0 \\ 0 & 10{,}000 \end{pmatrix} \right)$$

# Bayesian Model-based Imputation - Simulation

NOT REQUIRED



# Challenges in Bayesian Inference



Usually impossible to compute analytically in close form

- Two key computational challenges:
  - computation of the posterior distribution $p(\theta|y)$
  - computation of the posterior predictive distribution $p(\tilde{y}|y)$

- Normalized and unnormalized densities

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)} \propto p(\theta)p(y|\theta)$$

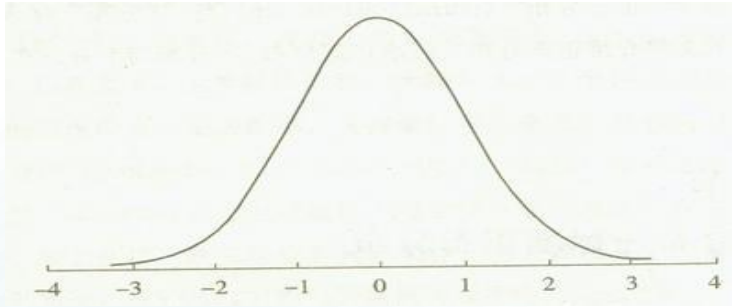
target distribution

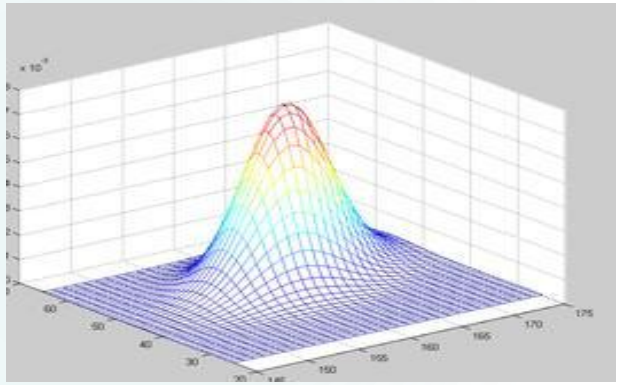
unnormalized density

# Inference via Monte Carlo Simulation

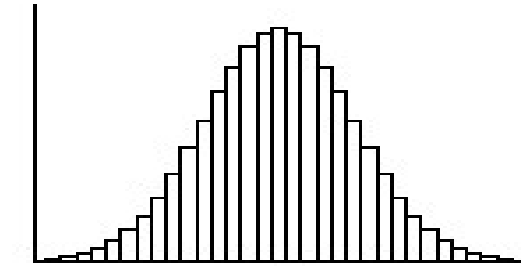


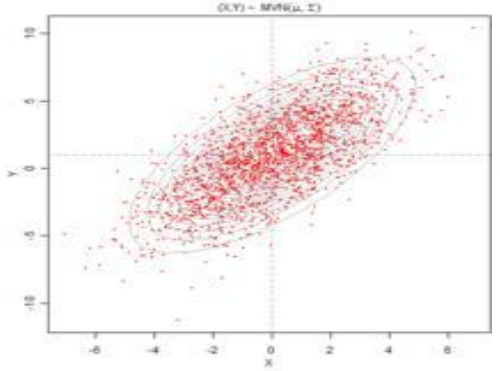
- A distribution can be well represented by its samples



Approximation based on samples





Remark: with samples from the posterior distribution, we can do any posterior inference we want in a numerical way!

# Significance of Sampling



- Numerical integration

$i.i.d.$ samples from $p(\theta|y)$

$$E(h(\theta)|y) = \int h(\theta)p(\theta|y)d\theta \approx \frac{1}{S}\sum_{s=1}^{S} h(\theta^s).$$

LLN. Convergence rate = $O(\frac{1}{\sqrt{S}})$ based on CLT.

- Very many quantities of interest may be cast as expectations:
  - Probabilities: $P(Y \in A) = E\left(I_{\{A\}}(Y)\right)$
  - Integrals: $(b-a)\int_a^b q(x)\frac{1}{b-a}dx = (b-a)\int_a^b q(x)f_U(x)dx = (b-a)E\big(q(U)\big)$
  - Discrete Sums: $\sum_{x\in A} q(x) = \frac{1}{p}\sum_{x\in A} q(x)p = \frac{1}{p}E\big(q(W)\big)$
- Monte Carlo - approximating an expectation by the sample mean of a function of simulated random variables. (unbiased)



# Approximation based on Simulation



The inputs

$f(\mathbf{y}^{mis}, \mathbf{Y}^{obs}, \mathbf{W}|\theta)$

$p(\theta)$

$\Rightarrow f(\mathbf{y}^{mis}, \mathbf{Y}^{obs}, \mathbf{W}, \theta)$

Step 1

$$f(\mathbf{Y}^{obs}, \mathbf{W}|\theta) = \int_{\mathbf{y}^{mis}} f(\mathbf{y}^{mis}, \mathbf{Y}^{obs}, \mathbf{W}|\theta)\, d\mathbf{y}^{mis}$$

$$p(\theta|\mathbf{Y}^{obs}, \mathbf{W}) = \frac{p(\theta) \cdot f(\mathbf{Y}^{obs}, \mathbf{W}|\theta)}{f(\mathbf{Y}^{obs}, \mathbf{W})}$$

Step 2

$$f(\mathbf{Y}^{mis}|\mathbf{Y}^{obs}, \mathbf{W}, \theta) = \frac{f(\mathbf{Y}^{mis}, \mathbf{Y}^{obs}|\mathbf{W}, \theta)}{f(\mathbf{Y}^{obs}|\mathbf{W}, \theta)} = \frac{f(\mathbf{Y}^{mis}, \mathbf{Y}^{obs}|\mathbf{W}, \theta)}{\int_{\mathbf{y}^{mis}} f(\mathbf{y}^{mis}, \mathbf{Y}^{obs}|\mathbf{W}, \theta) d\mathbf{y}^{mis}}$$

Step 3

$$f(\mathbf{Y}^{mis}, \theta|\mathbf{Y}^{obs}, \mathbf{W}) = f(\mathbf{Y}^{mis}|\mathbf{Y}^{obs}, \mathbf{W}, \theta) \cdot p(\theta|\mathbf{Y}^{obs}, \mathbf{W})$$

$$f(\mathbf{Y}^{mis}|\mathbf{Y}^{obs}, \mathbf{W}) = \int_{\theta} f(\mathbf{Y}^{mis}, \theta|\mathbf{Y}^{obs}, \mathbf{W})\, d\theta$$

$f(\mathbf{Y}^{mis}|\mathbf{Y}^{obs}, \mathbf{W})$

Step 4

The target

$f(\tau|\mathbf{Y}^{obs}, \mathbf{W})$

Repeat:

1. Impute $\theta$
2. Impute $Y^{mis}$
3. Calculate $\tau$

If analytic solutions are hard to get, we can sample by MCMC



# Review: An Analytic Example with Six Units



$$\tau = \tau(\mathbf{Y}(0), \mathbf{Y}(1))$$
$$(\mathbf{Y}(0), \mathbf{Y}(1)) = h(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}, \mathbf{W}).$$

$$\tau\left(Y^{mis}, Y^{obs}, W\right)$$

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}}, \mathbf{W}) \quad \Longleftrightarrow \quad f(\tau|\mathbf{Y}^{\text{obs}}, \mathbf{W})$$

$$\tau_{\text{fs}}|\mathbf{Y}^{\text{obs}}, \mathbf{W} \sim \mathcal{N}\left(8.7, 5.2^2\right)$$

Step 4

Step 3

$$\begin{pmatrix} Y_1^{\text{mis}} \\ Y_2^{\text{mis}} \\ Y_3^{\text{mis}} \\ Y_4^{\text{mis}} \\ Y_5^{\text{mis}} \\ Y_6^{\text{mis}} \end{pmatrix} \Bigg| \mathbf{Y}^{\text{obs}}, \mathbf{W} \sim \mathcal{N}\left( \begin{pmatrix} 12.8 \\ 4.1 \\ 12.8 \\ 4.1 \\ 12.8 \\ 4.1 \end{pmatrix}, \begin{pmatrix} 85.3 & 0 & 21.3 & 0 & 21.3 & 0 \\ 0 & 133.2 & 0 & 33.2 & 0 & 33.2 \\ 21.3 & 0 & 85.3 & 0 & 21.3 & 0 \\ 0 & 0 & 0 & 133.2 & 0 & 33.2 \\ 21.3 & 0 & 21.3 & 0 & 85.3 & 0 \\ 0 & 33.2 & 0 & 33.2 & 0 & 133.2 \end{pmatrix} \right)$$

$$p(\theta|\mathbf{Y}^{\text{obs}}, \mathbf{W}) = \frac{p(\theta) \cdot \mathcal{L}(\theta|\mathbf{Y}^{\text{obs}}, \mathbf{W})}{f(\mathbf{Y}^{\text{obs}}, \mathbf{W})}$$

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \Bigg| \mathbf{Y}^{\text{obs}}, \mathbf{W} \sim \mathcal{N}\left( \begin{pmatrix} 4.1 \\ 12.8 \end{pmatrix}, \begin{pmatrix} 5.8^2 & 0 \\ 0 & 4.6^2 \end{pmatrix} \right)$$

Step 2

Step 1

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}}, \mathbf{W}, \theta) = \frac{f(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}|\mathbf{W}, \theta)}{f(\mathbf{Y}^{\text{obs}}|\mathbf{W}, \theta)} = \frac{f(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}|\mathbf{W}, \theta)}{\int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}|\mathbf{W}, \theta) d\mathbf{y}^{\text{mis}}}$$

$$\begin{pmatrix} Y_1^{\text{mis}} \\ Y_2^{\text{mis}} \\ Y_3^{\text{mis}} \\ Y_4^{\text{mis}} \\ Y_5^{\text{mis}} \\ Y_6^{\text{mis}} \end{pmatrix} \Bigg| \mathbf{Y}^{\text{obs}}, \mathbf{W}, \mu_c, \mu_t \sim \mathcal{N}\left( \begin{pmatrix} \mu_t \\ \mu_c \\ \mu_t \\ \mu_c \\ \mu_t \\ \mu_c \end{pmatrix}, \begin{pmatrix} 64 & 0 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 & 0 \\ 0 & 0 & 64 & 0 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100 \end{pmatrix} \right)$$

$$\Pr(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \Bigg| \theta \sim \mathcal{N}\left( \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} 100 & 0 \\ 0 & 64 \end{pmatrix} \right)$$

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 10{,}000 & 0 \\ 0 & 10{,}000 \end{pmatrix} \right)$$

# A Simulation Example with Six Units



Repeat:

1. Impute $\theta$
2. Impute $Y^{mis}$
3. Calculate $\tau$

$$(\mu_c^{(1)}, \mu_t^{(1)}) = (1.63, 5.09).$$

$$\begin{pmatrix} Y_1^{mis} \\ Y_2^{mis} \\ Y_3^{mis} \\ Y_4^{mis} \\ Y_5^{mis} \\ Y_6^{mis} \end{pmatrix} \Bigg| \mathbf{Y}^{obs}, \mathbf{W}, \theta \sim \mathcal{N}\left( \begin{pmatrix} 5.09 \\ 1.63 \\ 5.09 \\ 1.63 \\ 5.09 \\ 1.63 \end{pmatrix}, \begin{pmatrix} 64 & 0 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 & 0 \\ 0 & 0 & 64 & 0 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100 \end{pmatrix} \right),$$

| $Y_i(0)$ | $Y_i(1)$ | $W_i$ |
|---|---|---|
| 0 | (6.1) | 0 |
| (13.5) | 9.9 | 1 |
| 12.4 | (7.4) | 0 |
| (13.5) | 3.6 | 1 |
| 0 | (−4.1) | 0 |
| (1.3) | 24.9 | 1 |

$$\hat{\tau}^{(1)} = 1.2$$

Step 2

$$f(\mathbf{Y}^{mis}|\mathbf{Y}^{obs}, \mathbf{W}, \theta) = \frac{f(\mathbf{Y}^{mis}, \mathbf{Y}^{obs}|\mathbf{W}, \theta)}{f(\mathbf{Y}^{obs}|\mathbf{W}, \theta)} = \frac{f(\mathbf{Y}^{mis}, \mathbf{Y}^{obs}|\mathbf{W}, \theta)}{\int_{\mathbf{y}^{mis}} f(\mathbf{y}^{mis}, \mathbf{Y}^{obs}|\mathbf{W}, \theta) d\mathbf{y}^{mis}}$$

$$\begin{pmatrix} Y_1^{mis} \\ Y_2^{mis} \\ Y_3^{mis} \\ Y_4^{mis} \\ Y_5^{mis} \\ Y_6^{mis} \end{pmatrix} \Bigg| \mathbf{Y}^{obs}, \mathbf{W}, \mu_c, \mu_t \sim \mathcal{N}\left( \begin{pmatrix} \mu_t \\ \mu_c \\ \mu_t \\ \mu_c \\ \mu_t \\ \mu_c \end{pmatrix}, \begin{pmatrix} 64 & 0 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 & 0 \\ 0 & 0 & 64 & 0 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100 \end{pmatrix} \right)$$

$$p(\theta|\mathbf{Y}^{obs}, \mathbf{W}) = \frac{p(\theta) \cdot \mathcal{L}(\theta|\mathbf{Y}^{obs}, \mathbf{W})}{f(\mathbf{Y}^{obs}, \mathbf{W})}$$

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \Bigg| \mathbf{Y}^{obs}, \mathbf{W} \sim \mathcal{N}\left( \begin{pmatrix} 4.1 \\ 12.8 \end{pmatrix}, \begin{pmatrix} 5.8^2 & 0 \\ 0 & 4.6^2 \end{pmatrix} \right)$$

Step 1

$$\Pr(\mathbf{W}|\mathbf{Y}(0), \mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \Big| \theta \sim \mathcal{N}\left( \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} 100 & 0 \\ 0 & 64 \end{pmatrix} \right)$$

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 10{,}000 & 0 \\ 0 & 10{,}000 \end{pmatrix} \right)$$

# A Simulation Example with Six Units



**Table 8.5.** ***The Average Treatment Effect Using Full Model-Based Imputations for the NSW Program Data***

| Unit | Potential Outcomes | | Treatment | Observed Outcome |
|---|---|---|---|---|
| | $Y_i(0)$ | $Y_i(1)$ | $W_i$ | $Y_i^{\text{obs}}$ |
| Panel A: First Parameter Draw $(\mu_c^{(1)}, \mu_t^{(1)}) = (1.63, 5.09)$ | | | | |
| 1 | 0 | (6.1) | 0 | 0 |
| 2 | (13.5) | 9.9 | 1 | 9.9 |
| 3 | 12.4 | (7.4) | 0 | 12.4 |
| 4 | (13.5) | 3.6 | 1 | 3.6 |
| 5 | 0 | (−4.1) | 0 | 0 |
| 6 | (1.3) | 24.9 | 1 | 24.9 |
| Average | 6.8 | 8.0 | | |
| $\tau_{\text{fs}}^{(1)}$ | | 1.2 | | |
| Panel B: Second Parameter Draw $(\mu_c^{(2)}, \mu_t^{(2)}) = (6.01, 13.58)$ | | | | |
| 1 | 0 | (12.1) | 0 | 0 |
| 2 | (27.8) | 9.9 | 1 | 9.9 |
| 3 | 12.4 | (19.4) | 0 | 12.4 |
| 4 | (4.6) | 3.6 | 1 | 3.6 |
| 5 | 0 | (8.9) | 0 | 0 |
| 6 | (7.1) | 24.9 | 1 | 24.9 |
| Average | 8.7 | 13.1 | | |
| $\tau_{\text{fs}}^{(2)}$ | | 4.5 | | |

$$N_R = 10{,}000$$

$$\frac{1}{N_R}\sum_{r=1}^{N_R}\tau_{\text{fs}}^{(r)} = \bar{\tau} = 8.6, \qquad \frac{1}{N_R-1}\sum_{r=1}^{N_R}\left(\tau_{\text{fs}}^{(r)} - \bar{\tau}\right)^2 = 5.3^2.$$

For comparison, the analytic results:

$$\tau_{\text{fs}} | \mathbf{Y}^{\text{obs}}, \mathbf{W} \sim \mathcal{N}\left(8.7, 5.2^2\right)$$

# In view of Super-population



# Bayesian Model-Based Imputation without Covariates – super population

The 3 inputs

$$\Pr(\mathbf{W}|\mathbf{Y}(0),\mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$f(\mathbf{Y}(0),\mathbf{Y}(1)|\theta)$

$p(\theta)$

Step 1

$$f(\mathbf{Y}^{\text{obs}},\mathbf{W}|\theta) = \int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}},\mathbf{Y}^{\text{obs}},\mathbf{W}|\theta)\, d\mathbf{y}^{\text{mis}}$$

$$p(\theta|\mathbf{Y}^{\text{obs}},\mathbf{W}) = \frac{p(\theta)\cdot f(\mathbf{Y}^{\text{obs}},\mathbf{W}|\theta)}{f(\mathbf{Y}^{\text{obs}},\mathbf{W})}$$

Simulation:

$$\tau^{(1)} = \tau(\theta^{(1)})$$

$$\{\hat{\tau}_{\text{sp}}^{(r)}, r = 1,\ldots,N_R\}$$

$\approx$

Step 2 $\quad \tau_{\text{sp}} = \tau(\theta)$

$$\tau_{\text{sp}} = \mathbb{E}_{\text{sp}}[Y_i(1) - Y_i(0)]$$

$$\tau_{\text{sp}} = \tau(\theta)$$

$$= \int\int (y(1) - y(0)) f(y(1), y(0)|\theta)\, dy(1)\, dy(0)$$

The target

$f(\tau|\mathbf{Y}^{\text{obs}},\mathbf{W})$



# An Analytic Example with Six Units



The 3 inputs

$$\Pr(\mathbf{W}|\mathbf{Y}(0),\mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \Big| \theta \sim \mathcal{N}\left( \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} 100 & 0 \\ 0 & 64 \end{pmatrix} \right)$$

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 10{,}000 & 0 \\ 0 & 10{,}000 \end{pmatrix} \right)$$

No need to to do simulation in this case.

Step 1

$$p(\theta|\mathbf{Y}^{\text{obs}},\mathbf{W}) = \frac{p(\theta)\cdot\mathcal{L}(\theta|\mathbf{Y}^{\text{obs}},\mathbf{W})}{f(\mathbf{Y}^{\text{obs}},\mathbf{W})}$$

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \Big| \mathbf{Y}^{\text{obs}},\mathbf{W} \sim \mathcal{N}\left( \begin{pmatrix} 4.1 \\ 12.8 \end{pmatrix}, \begin{pmatrix} 5.8^2 & 0 \\ 0 & 4.6^2 \end{pmatrix} \right)$$

Step 2

$$\tau_{\text{sp}} = \tau(\theta)$$

$$\tau_{\text{sp}} = \mu_t - \mu_c$$

$$\tau_{\text{fs}}|\mathbf{Y}^{\text{obs}},\mathbf{W} \sim \mathcal{N}\left(8.7, 7.4^2\right)$$

Larger variance than finite sample (Recall the results under Neyman's approach)

# Model Comparison in Application

NOT REQUIRED



# The National Supported Work Data



**Unit:** $N$ = 445 men substantially disadvantaged in the labor market

**Treatment:** a training program

**Control:** no active treatment

**Outcome:**

earnings in 1978 (earn’78)

**Seven covariates:**

age, education, married, no degree, black, earn’75, earn’74

**Completely randomized assignment:**

$N_t$ = 185, $N_c$=260

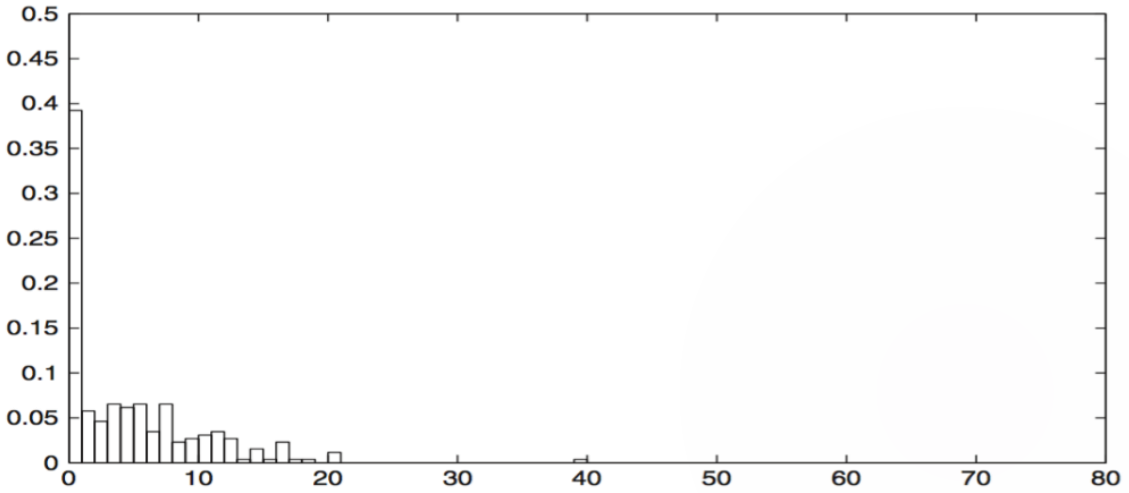



**Figure 8.1.** Histogram of earnings for control group – NSW job-training data

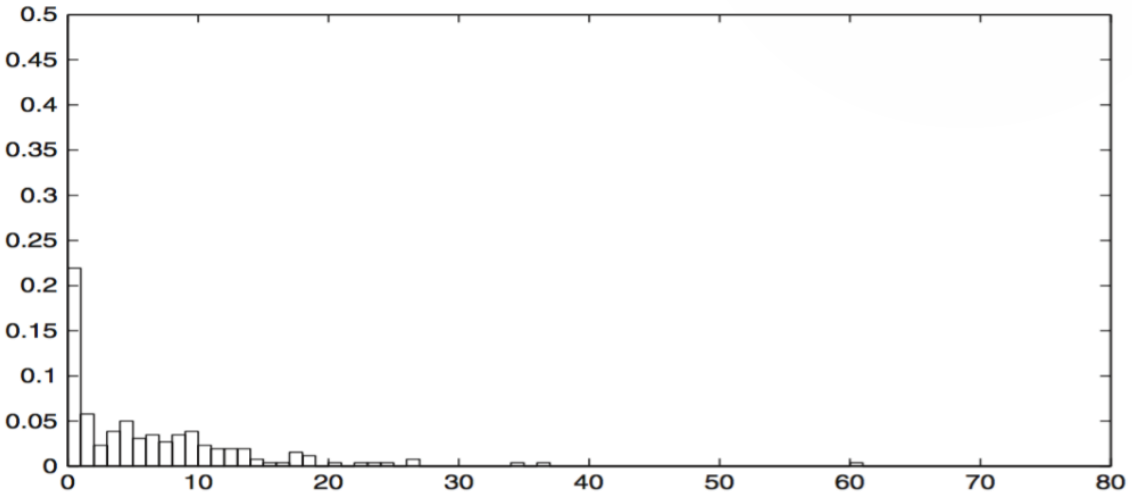



**Figure 8.2.** Histogram of earnings for trainee group – NSW job-training data

# Four Models for the NSW Data



Model 1: dependent Gaussian model without covariates (same variance in two groups)

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \Big| X_i, \theta \sim \mathcal{N}\left( \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 \end{pmatrix} \right)$$

Model 2: independent Gaussian model without covariates

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \Big| X_i, \theta \sim \mathcal{N}\left( \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} \sigma_c^2 & 0 \\ 0 & \sigma_t^2 \end{pmatrix} \right)$$

Model 3: independent Gaussian model with covariates

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \Big| X_i, \theta \sim \mathcal{N}\left( \begin{pmatrix} X_i\beta_c \\ X_i\beta_t \end{pmatrix}, \begin{pmatrix} \sigma_c^2 & 0 \\ 0 & \sigma_t^2 \end{pmatrix} \right)$$

Model 4: independent zero-inflated Gaussian model with covariates

$$\Pr(Y_i(0) > 0 | X_i, W_i, \theta) = \frac{\exp(X_i\gamma_c)}{1 + \exp(X_i\gamma_c)}$$

$$\Pr(Y_i(1) > 0 | X_i, W_i, \theta) = \frac{\exp(X_i\gamma_t)}{1 + \exp(X_i\gamma_t)}$$

$$\ln(Y_i(0)) | Y_i(0) > 0, X_i, W_i, \theta \sim \mathcal{N}\left(X_i\beta_c, \sigma_c^2\right)$$

$$\ln(Y_i(1)) | Y_i(1) > 0, X_i, W_i, \theta \sim \mathcal{N}\left(X_i\beta_t, \sigma_t^2\right)$$

# Result Summary of the Four Models



Neyman : 1.79 (0.67)

Regression : 1.67 (0.64) (incl. covariate Hispanics)

In the finite sample, the 0.25 quant of Y(1) – the 0.25 quant of Y(0).

**Table 8.6.** ***Posterior Means and Standard Deviations for Treatment Effects under Four Models for NSW Program Data***

| | Mean Covariate Dependent | Variance Treatment Specific | Potential Outcome Independent | Two-Part Model | Mean Effect | | Effect on Quantiles: 0.25 quant | | 0.50 quant | | 0.75 quant | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Mean | (S.D.) | Mean | (S.D.) | Mean | (S.D.) | Mean | (S.D.) |
| $M_1$ | No | No | No | No | 1.79 | (0.63) | 1.79 | (0.63) | 1.79 | (0.63) | 1.79 | (0.63) |
| $M_2$ | No | Yes | Yes | No | 1.78 | (0.49) | 0.63 | (0.35) | 1.63 | (0.55) | 3.07 | (0.64) |
| $M_3$ | Yes | Yes | Yes | No | 1.57 | (0.50) | 0.42 | (0.34) | 1.40 | (0.55) | 2.89 | (0.63) |
| $M_4$ | Yes | Yes | Yes | Yes | 1.79 | (0.85) | 0.25 | (0.30) | 1.03 | (0.53) | 1.69 | (0.72) |

M4 is more realistic in this example.

# Details for Simulation



- Algorithms for sampling:
  - M 1-3: Gibbs sampling;
  - M 4:
    - Zero-Inflated part: Independence H-M,
    - Log-Normal part : Gibbs sampling;
- Implementation:
  - Initials: MLE
  - burn-in: the first 1/3 samples

# M1

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \mid X_i, \theta \sim \mathcal{N}\left( \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 \end{pmatrix} \right).$$



- Prior: $\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} \nu_c \\ \nu_t \end{pmatrix}, \begin{pmatrix} \omega_c^2 & 0 \\ 0 & \omega_t^2 \end{pmatrix} \right)$,
- $\sigma^2 \sim \text{Scale}-\text{Inv}-\chi^2(M, S_0^2)$, i.e.g $\sigma^{-2} M S_0^2 \sim \chi_M^2$
- Posterior: $\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} | \sigma^2, \boldsymbol{Y}^{\text{obs}}, \boldsymbol{W}$

$$\sim \mathcal{N}\left( \begin{pmatrix} \frac{\bar{Y}_c^{\text{obs}} \cdot N_c/\sigma^2 + \nu_c/\omega_c^2}{N_c/\sigma^2 + 1/\omega_c^2} \\ \frac{\bar{Y}_t^{\text{obs}} \cdot N_t/\sigma^2 + \nu_t/\omega_t^2}{N_t/\sigma^2 + 1/\omega_t^2} \end{pmatrix}, \begin{pmatrix} \frac{1}{N_c/\sigma^2 + 1/\omega_c^2} & 0 \\ 0 & \frac{1}{N_t/\sigma^2 + 1/\omega_t^2} \end{pmatrix} \right),$$

- $\sigma^{-2}\left( M S_0^2 + \sum_{W_i=1} \left(Y_i^{\text{obs}} - \mu_t\right)^2 + \sum_{W_i=0} \left(Y_i^{\text{obs}} - \mu_c\right)^2 \right) | \, \mu_c, \mu_t, \boldsymbol{Y}^{\text{obs}}, \boldsymbol{W} \sim \chi_{M+N}^2$
- Fixed hyper-parameters: $\nu_c = \nu_t = 0,\ \omega_c = \omega_t = 100,\ M = 2,\ S_0 = 0.1$

# M2

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \Big| X_i, \theta \sim \mathcal{N}\left( \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} \sigma_c^2 & 0 \\ 0 & \sigma_t^2 \end{pmatrix} \right)$$



- Prior: $\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} \nu_c \\ \nu_t \end{pmatrix}, \begin{pmatrix} \omega_c^2 & 0 \\ 0 & \omega_t^2 \end{pmatrix} \right)$,
- $\sigma_c^2 \perp \sigma_t^2 \sim \text{Scale}-\text{Inv}-\chi^2(M, S_0^2)$, i.e. $\sigma_c^{-2} M S_0^2 \perp \sigma_t^{-2} M S_0^2 \sim \chi_M^2$
- Posterior: $\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} | \sigma_c^2, \sigma_t^2, \boldsymbol{Y}^{\text{obs}}, \boldsymbol{W} \sim \mathcal{N}\left( \begin{pmatrix} \frac{\bar{Y}_c^{\text{obs}} \cdot N_c/\sigma_c^2 + \nu_c/\omega_c^2}{N_c/\sigma_c^2 + 1/\omega_c^2} \\ \frac{\bar{Y}_t^{\text{obs}} \cdot N_t/\sigma_t^2 + \nu_t/\omega_t^2}{N_t/\sigma_t^2 + 1/\omega_t^2} \end{pmatrix}, \begin{pmatrix} \frac{1}{N_c/\sigma_c^2 + 1/\omega_c^2} & 0 \\ 0 & \frac{1}{N_t/\sigma_t^2 + 1/\omega_t^2} \end{pmatrix} \right)$,
- $\sigma_c^{-2}\left( M S_0^2 + \sum_{W_i=0} \left( Y_i^{\text{obs}} - \mu_c \right)^2 \right) | \, \mu_c, \mu_t, \boldsymbol{Y}^{\text{obs}}, \boldsymbol{W} \sim \chi_{M+N_c}^2$
- $\sigma_t^{-2}\left( M S_0^2 + \sum_{W_i=1} \left( Y_i^{\text{obs}} - \mu_t \right)^2 \right) | \, \mu_c, \mu_t, \boldsymbol{Y}^{\text{obs}}, \boldsymbol{W} \sim \chi_{M+N_t}^2$

(条件独立)

- Fixed hyper-parameters： $\nu_c = \nu_t = 0,\ \omega_c = \omega_t = 100,\ M = 2,\ S_0 = 0.1$

# M3

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \Big| X_i, \theta \sim \mathcal{N}\left( \begin{pmatrix} X_i\beta_c \\ X_i\beta_t \end{pmatrix}, \begin{pmatrix} \sigma_c^2 & 0 \\ 0 & \sigma_t^2 \end{pmatrix} \right).$$



- Prior: $\boldsymbol{\beta}_c \perp \boldsymbol{\beta}_t \sim \mathcal{N}(\mathbf{0}, \omega^2 \boldsymbol{I}_p)$,
- $\sigma_c^2 \perp \sigma_t^2 \sim \text{Scale}-\text{Inv}-\chi^2(M, S_0^2)$, i.e. $\sigma_c^{-2} M S_0^2 \perp \sigma_t^{-2} M S_0^2 \sim \chi_M^2$
- Conditional posterior:
- $\boldsymbol{\beta}_c | \sigma_c^2, \sigma_t^2, \boldsymbol{Y}^{\text{obs}}, \boldsymbol{W}, \boldsymbol{X} \sim \mathcal{N}\left( \left( \sigma_c^{-2} \boldsymbol{X}_c^{\mathrm{T}} \boldsymbol{X}_c + \omega^{-2} \boldsymbol{I}_p \right)^{-1} \sigma_c^{-2} \boldsymbol{X}_c^{\mathrm{T}} \boldsymbol{Y}_c,\ \left( \sigma_c^{-2} \boldsymbol{X}_c^{\mathrm{T}} \boldsymbol{X}_c + \omega^{-2} \boldsymbol{I}_p \right)^{-1} \right)$
- $\boldsymbol{\beta}_t | \sigma_c^2, \sigma_t^2, \boldsymbol{Y}^{\text{obs}}, \boldsymbol{W}, \boldsymbol{X} \sim \mathcal{N}\left( \left( \sigma_t^{-2} \boldsymbol{X}_t^{\mathrm{T}} \boldsymbol{X}_t + \omega^{-2} \boldsymbol{I}_p \right)^{-1} \sigma_t^{-2} \boldsymbol{X}_t^{\mathrm{T}} \boldsymbol{Y}_t,\ \left( \sigma_t^{-2} \boldsymbol{X}_t^{\mathrm{T}} \boldsymbol{X}_t + \omega^{-2} \boldsymbol{I}_p \right)^{-1} \right)$

(条件独立)

- $\sigma_c^{-2} \left( M S_0^2 + \sum_{W_i=0} \left( Y_i^{\text{obs}} - \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\beta}_c \right)^2 \right) | \boldsymbol{\beta}_c, \boldsymbol{\beta}_t, \boldsymbol{Y}^{\text{obs}}, \boldsymbol{W}, \boldsymbol{X} \sim \chi_{M+N_c}^2$
- $\sigma_t^{-2} \left( M S_0^2 + \sum_{W_i=1} \left( Y_i^{\text{obs}} - \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\beta}_t \right)^2 \right) | \boldsymbol{\beta}_c, \boldsymbol{\beta}_t, \boldsymbol{Y}^{\text{obs}}, \boldsymbol{W}, \boldsymbol{X} \sim \chi_{M+N_t}^2$

(条件独立)

- Fixed hyper-parameters: $\omega = 100,\ M = 2,\ S_0 = 0.1$

# M3

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \Big| X_i, \theta \sim \mathcal{N}\left( \begin{pmatrix} X_i\beta_c \\ X_i\beta_t \end{pmatrix}, \begin{pmatrix} \sigma_c^2 & 0 \\ 0 & \sigma_t^2 \end{pmatrix} \right).$$



- Additional tips：Matrix Calculation
- Full and reduced SVD $(n > p)$: $\boldsymbol{X}_{n\times p} = \boldsymbol{U}_{n\times n}\boldsymbol{\Sigma}_{n\times p}(\boldsymbol{V}_{p\times p})^{\mathrm{T}} = \boldsymbol{U}_{1_{n\times p}}\boldsymbol{D}_{p\times p}(\boldsymbol{V}_{p\times p})^{\mathrm{T}}$
- Singular values： $\boldsymbol{D} = \mathrm{diag}(d_1, d_2, \cdots, d_p)$
- Posterior variance： $\left(\sigma^{-2}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + \omega^{-2}\boldsymbol{I}_p\right)^{-1}$

$$= \boldsymbol{V}\begin{pmatrix} (\sigma^{-2}d_1^2 + \omega^{-2})^{-1} & & \\ & \ddots & \\ & & \left(\sigma^{-2}d_p^2 + \omega^{-2}\right)^{-1} \end{pmatrix}\boldsymbol{V}^{\mathrm{T}}$$

- Posterior Mean： $\left(\sigma^{-2}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + \omega^{-2}\boldsymbol{I}_p\right)^{-1}\sigma^{-2}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y} = \boldsymbol{V}\begin{pmatrix} \frac{d_1}{d_1^2+\sigma^2\omega^{-2}} & & \\ & \ddots & \\ & & \frac{d_p}{d_p^2+\sigma^2\omega^{-2}} \end{pmatrix}\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{Y}$

- Faster and more stable

# M4 Zero-Inflated Log-Normal (ZILN)

$$\Pr(Y_i(0) > 0 | X_i, W_i, \theta) = \frac{\exp(X_i\gamma_c)}{1+\exp(X_i\gamma_c)},$$

$$\Pr(Y_i(1) > 0 | X_i, W_i, \theta) = \frac{\exp(X_i\gamma_t)}{1+\exp(X_i\gamma_t)}.$$

$$\ln(Y_i(0)) | Y_i(0) > 0, X_i, W_i, \theta \sim \mathcal{N}\left(X_i\beta_c, \sigma_c^2\right),$$

$$\ln(Y_i(1)) | Y_i(1) > 0, X_i, W_i, \theta \sim \mathcal{N}\left(X_i\beta_t, \sigma_t^2\right).$$

- **Zero-Inflated Part (逻辑回归)**
- Prior： $\boldsymbol{\gamma}_c \perp \boldsymbol{\gamma}_t \sim \mathcal{N}(\mathbf{0}, \omega^2 \boldsymbol{I}_p)$
- Likelihood： $L_c(\boldsymbol{\gamma}_c) = \prod_{W_i=0} \frac{\exp(Y_i^{\text{obs}} \cdot \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\gamma}_c)}{1+\exp(\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\gamma}_c)}$, $L_t(\boldsymbol{\gamma}_t) = \prod_{W_i=1} \frac{\exp(Y_i^{\text{obs}} \cdot \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\gamma}_t)}{1+\exp(\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\gamma}_t)}$
- Proposal distribution： $\boldsymbol{\gamma}_c^* \sim g_c = \mathcal{N}(\hat{\boldsymbol{\gamma}}_{c,\text{MLE}}, 2\text{Cov}(\hat{\boldsymbol{\gamma}}_{c,\text{MLE}}))$

  $\boldsymbol{\gamma}_t^* \sim g_t = \mathcal{N}(\hat{\boldsymbol{\gamma}}_{t,\text{MLE}}, 2\text{Cov}(\hat{\boldsymbol{\gamma}}_{c,\text{MLE}}))$
- Accept probability: $\mathbb{P}\left(\boldsymbol{\gamma}_c^{(r+1)} = \boldsymbol{\gamma}_c^*\right) = \min\left\{1, \frac{L_c(\boldsymbol{\gamma}_c^*) \cdot \mathcal{N}(\boldsymbol{\gamma}_c^* \mid \mathbf{0}, \omega^2 \boldsymbol{I}_p) \cdot g_c(\boldsymbol{\gamma}_c^{(r)})}{L_c(\boldsymbol{\gamma}_c^{(r)}) \cdot \mathcal{N}(\boldsymbol{\gamma}_c^{(r)} \mid \mathbf{0}, \omega^2 \boldsymbol{I}_p) \cdot g_c(\boldsymbol{\gamma}_c^*)}\right\}$

  $\mathbb{P}\left(\boldsymbol{\gamma}_t^{(r+1)} = \boldsymbol{\gamma}_t^*\right) = \min\left\{1, \frac{L_t(\boldsymbol{\gamma}_t^*) \cdot \mathcal{N}(\boldsymbol{\gamma}_t^* \mid \mathbf{0}, \omega^2 \boldsymbol{I}_p) \cdot g_t(\boldsymbol{\gamma}_t^{(r)})}{L_t(\boldsymbol{\gamma}_t^{(r)}) \cdot \mathcal{N}(\boldsymbol{\gamma}_t^{(r)} \mid \mathbf{0}, \omega^2 \boldsymbol{I}_p) \cdot g_t(\boldsymbol{\gamma}_t^*)}\right\}$
- **Lor-Normal Part (线性回归)**
- Same as M3

- Fixed hyper-parameters： $\omega = 100,\ M = 2,\ S_0 = 0.1$

# Summary



# Summary



- Motivation of more flexible models and plausibility leads to Bayesian model-based imputation approach.
- Key of the theory (required):
  - Parameters are random variables/vectors
  - Specification of priors
  - Inference on posteriors
- Key of implementation (not required):
  - Simulation, e.g. MCMC
  - valid because of LLN



# Reference



- Reference: [T1] Chapter 8

# Bayesian Model-based Imputation - Theory

- NO CORRELATION, NO COVARIATES
- CORRELATION, NO COVARIATES (NOT REQUIRED)
- NO CORRELATION, COVARIATES



# Dependence between Potential Outcomes



- Given this prior, the posterior distribution of the correlation coefficient will be identical to its prior distribution.
- The correlation coefficient is somewhat different from other parameters.

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W}) \Longleftrightarrow f(\tau|\mathbf{Y}^{\text{obs}},\mathbf{W})$$

Step 4

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W}) = \int_{\theta} f(\mathbf{Y}^{\text{mis}},\theta|\mathbf{Y}^{\text{obs}},\mathbf{W})\,d\theta$$

Step 3

$$f(\mathbf{Y}^{\text{mis}},\theta|\mathbf{Y}^{\text{obs}},\mathbf{W}) = f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W},\theta)\cdot p(\theta|\mathbf{Y}^{\text{obs}},\mathbf{W})$$

$$f(\mathbf{Y}^{\text{obs}},\mathbf{W}|\theta) = \int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}},\mathbf{Y}^{\text{obs}},\mathbf{W}|\theta)\,d\mathbf{y}^{\text{mis}}$$

$$p(\theta|\mathbf{Y}^{\text{obs}},\mathbf{W}) = \frac{p(\theta)\cdot f(\mathbf{Y}^{\text{obs}},\mathbf{W}|\theta)}{f(\mathbf{Y}^{\text{obs}},\mathbf{W})}$$

Step 2

$$f(\mathbf{Y}^{\text{obs}},\mathbf{W}) = \int_{\theta} p(\theta)\cdot f(\mathbf{Y}^{\text{obs}},\mathbf{W}|\theta)d\theta$$

Step 1

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W},\theta) = \frac{f(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)}{f(\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)} = \frac{f(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)}{\int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)d\mathbf{y}^{\text{mis}}}$$

$$(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}) = g(\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{W})$$

$$f(\mathbf{Y}(0),\mathbf{Y}(1)|\mathbf{W},\theta) = \frac{f(\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{W}|\theta)}{\Pr(\mathbf{W}|\theta)} = \frac{f(\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{W}|\theta)}{\int f(\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{W}|\theta)d\mathbf{Y}(0)d\mathbf{Y}(1)}$$

$$f(\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{W}|\theta) = \Pr(\mathbf{W}|\mathbf{Y}(0),\mathbf{Y}(1),\theta)\cdot f(\mathbf{Y}(0),\mathbf{Y}(1)|\theta).$$

$$\Pr(\mathbf{W}|\mathbf{Y}(0),\mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$$f(Y_i(0),Y_i(1)|\theta) \sim \mathcal{N}\left(\begin{pmatrix}\mu_c\\ \mu_t\end{pmatrix}, \begin{pmatrix}\sigma_c^2 & \rho\sigma_c\sigma_t\\ \rho\sigma_c\sigma_t & \sigma_t^2\end{pmatrix}\right)$$

$$p(\theta) = p(\rho)\cdot p(\mu_c,\mu_t,\sigma_c^2,\sigma_t^2)$$

Parameters for dependence

# An Example for comparison



For example, $\rho = 1$

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \Big| \theta \sim \mathcal{N}\left( \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} 100 & 80 \\ 80 & 64 \end{pmatrix} \right) \qquad \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 10{,}000 & 0 \\ 0 & 10{,}000 \end{pmatrix} \right)$$

$$\begin{pmatrix} Y_i^{\text{mis}} \\ Y_i^{\text{obs}} \end{pmatrix} \Big| \mathbf{W}, \mu_c, \mu_t \sim \mathcal{N}\left( \begin{pmatrix} W_i \cdot \mu_c + (1 - W_i) \cdot \mu_t \\ (1 - W_i) \cdot \mu_c + W_i \cdot \mu_t \end{pmatrix}, \begin{pmatrix} W_i \cdot 100 + (1 - W_i) \cdot 64 & 80 \\ 80 & (1 - W_i) \cdot 100 + W_i \cdot 64 \end{pmatrix} \right)$$

$$Y_i^{\text{mis}} | \mathbf{Y}^{\text{obs}}, \mathbf{W}, \mu_c, \mu_t \sim \mathcal{N}\left( W_i \cdot \left( \mu_c + \frac{80}{64} \cdot (Y_i^{\text{obs}} - \mu_t) \right) + (1 - W_i) \cdot \left( \mu_t + \frac{80}{100} \cdot (Y_i^{\text{obs}} - \mu_c) \right), 0 \right)$$

Step 1

$$f(\mathbf{Y}^{\text{mis}} | \mathbf{Y}^{\text{obs}}, \mathbf{W}, \theta) = \frac{f(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}} | \mathbf{W}, \theta)}{f(\mathbf{Y}^{\text{obs}} | \mathbf{W}, \theta)} = \frac{f(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}} | \mathbf{W}, \theta)}{\int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}}, \mathbf{Y}^{\text{obs}} | \mathbf{W}, \theta) d\mathbf{y}^{\text{mis}}}$$

$$\Pr(\mathbf{W} | \mathbf{Y}(0), \mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$$f(Y_i(0), Y_i(1) | \theta) \sim \mathcal{N}\left( \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} \sigma_c^2 & \rho\sigma_c\sigma_t \\ \rho\sigma_c\sigma_t & \sigma_t^2 \end{pmatrix} \right)$$

$$p(\theta) = p(\rho) \cdot p(\mu_c, \mu_t, \sigma_c^2, \sigma_t^2)$$

# An Example for comparison



$$\tau_{\text{fs}} | \mathbf{Y}^{\text{obs}}, \mathbf{W} \sim \mathcal{N}\left(8.7,\ 7.4^2\right)$$

$$f(\mathbf{Y}^{\text{mis}} | \mathbf{Y}^{\text{obs}}, \mathbf{W}) \longleftrightarrow f(\tau | \mathbf{Y}^{\text{obs}}, \mathbf{W})$$ Step 4

Step 3

$$f(\mathbf{Y}^{\text{mis}} | \mathbf{Y}^{\text{obs}}, \mathbf{W}) = \int_{\theta} f(\mathbf{Y}^{\text{mis}}, \theta | \mathbf{Y}^{\text{obs}}, \mathbf{W})\, d\theta$$

$$f(\mathbf{Y}^{\text{mis}}, \theta | \mathbf{Y}^{\text{obs}}, \mathbf{W}) = f(\mathbf{Y}^{\text{mis}} | \mathbf{Y}^{\text{obs}}, \mathbf{W}, \theta) \cdot p(\theta | \mathbf{Y}^{\text{obs}}, \mathbf{W})$$

Step 2

$$f(\mathbf{Y}^{\text{obs}}, \mathbf{W} | \theta) = \int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}, \mathbf{W} | \theta)\, d\mathbf{y}^{\text{mis}}$$

$$p(\theta | \mathbf{Y}^{\text{obs}}, \mathbf{W}) = \frac{p(\theta) \cdot f(\mathbf{Y}^{\text{obs}}, \mathbf{W} | \theta)}{f(\mathbf{Y}^{\text{obs}}, \mathbf{W})}$$

Step 1

$$(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}}) = g(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W})$$

$$f(\mathbf{Y}^{\text{mis}} | \mathbf{Y}^{\text{obs}}, \mathbf{W}, \theta) = \frac{f(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}} | \mathbf{W}, \theta)}{f(\mathbf{Y}^{\text{obs}} | \mathbf{W}, \theta)} = \frac{f(\mathbf{Y}^{\text{mis}}, \mathbf{Y}^{\text{obs}} | \mathbf{W}, \theta)}{\int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}}, \mathbf{Y}^{\text{obs}} | \mathbf{W}, \theta) d\mathbf{y}^{\text{mis}}}$$

$$f(\mathbf{Y}(0), \mathbf{Y}(1) | \mathbf{W}, \theta) = \frac{f(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W} | \theta)}{\Pr(\mathbf{W} | \theta)} = \frac{f(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W} | \theta)}{\int f(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W} | \theta) d\mathbf{Y}(0) d\mathbf{Y}(1)}.$$

$$f(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W} | \theta) = \Pr(\mathbf{W} | \mathbf{Y}(0), \mathbf{Y}(1), \theta) \cdot f(\mathbf{Y}(0), \mathbf{Y}(1) | \theta).$$

$$\Pr(\mathbf{W} | \mathbf{Y}(0), \mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \Big| \theta \sim \mathcal{N}\left( \begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix}, \begin{pmatrix} 100 & 80 \\ 80 & 64 \end{pmatrix} \right)$$

$$\begin{pmatrix} \mu_c \\ \mu_t \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 10{,}000 & 0 \\ 0 & 10{,}000 \end{pmatrix} \right)$$

For example, $\rho = 1$

# Bayesian Model-based Imputation - Theory

- NO CORRELATION, NO COVARIATES
- CORRELATION, NO COVARIATES
- NO CORRELATION, COVARIATES (NOT REQUIRED)



# Model-Based Imputation with Covariates



The inputs

$$\Pr(\mathbf{W}|\mathbf{Y}(0),\mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$$f(Y_i(0), Y_i(1), X|\theta_{Y|X}, \theta_X) = f(Y_i(0), Y_i(1)|X, \theta_{Y|X}) \cdot f(X|\theta_X)$$

$$p(\theta_{Y|X}, \theta_X) = p(\theta_{Y|X}) \cdot p(\theta_X)$$

With covariates

The target $f(\tau|\mathbf{Y}^{\text{obs}}, \mathbf{W})$

# An Example



Similar procedure as before, but more difficult to get analytic solutions

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W}) \longleftrightarrow f(\tau|\mathbf{Y}^{\text{obs}},\mathbf{W})$$ Step 4

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W}) = \int_{\theta} f(\mathbf{Y}^{\text{mis}},\theta|\mathbf{Y}^{\text{obs}},\mathbf{W})\,d\theta$$ Step 3

$$f(\mathbf{Y}^{\text{mis}},\theta|\mathbf{Y}^{\text{obs}},\mathbf{W}) = f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W},\theta)\cdot p(\theta|\mathbf{Y}^{\text{obs}},\mathbf{W}).$$

$$f(\mathbf{Y}^{\text{obs}},\mathbf{W}|\theta) = \int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}},\mathbf{Y}^{\text{obs}},\mathbf{W}|\theta)\,d\mathbf{y}^{\text{mis}}$$

$$p(\theta|\mathbf{Y}^{\text{obs}},\mathbf{W}) = \frac{p(\theta)\cdot f(\mathbf{Y}^{\text{obs}},\mathbf{W}|\theta)}{f(\mathbf{Y}^{\text{obs}},\mathbf{W})}$$ Step 2

Step 1

$$(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}) = g(\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{W})$$

$$f(\mathbf{Y}^{\text{mis}}|\mathbf{Y}^{\text{obs}},\mathbf{W},\theta) = \frac{f(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)}{f(\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)} = \frac{f(\mathbf{Y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)}{\int_{\mathbf{y}^{\text{mis}}} f(\mathbf{y}^{\text{mis}},\mathbf{Y}^{\text{obs}}|\mathbf{W},\theta)d\mathbf{y}^{\text{mis}}}$$

$$f(\mathbf{Y}(0),\mathbf{Y}(1)|\mathbf{W},\theta) = \frac{f(\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{W}|\theta)}{\Pr(\mathbf{W}|\theta)} = \frac{f(\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{W}|\theta)}{\int f(\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{W}|\theta)d\mathbf{Y}(0)d\mathbf{Y}(1)}.$$

$$f(\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{W}|\theta) = \Pr(\mathbf{W}|\mathbf{Y}(0),\mathbf{Y}(1),\theta)\cdot f(\mathbf{Y}(0),\mathbf{Y}(1)|\theta).$$

$$\Pr(\mathbf{W}|\mathbf{Y}(0),\mathbf{Y}(1)) = \binom{N}{N_t}^{-1}, \quad \text{for all } \mathbf{W} \text{ such that } \sum_{i=1}^{N} W_i = N_t,$$

$$f(Y_i(0),Y_i(1),X|\theta_{Y|X},\theta_X) = f(Y_i(0),Y_i(1)|X,\theta_{Y|X})\cdot f(X|\theta_X)$$

$$p(\theta_{Y|X},\theta_X) = p(\theta_{Y|X})\cdot p(\theta_X)$$