Special Topic: Bayesian Finite Population Survey Sampling

Sudipto Banerjee

Division of Biostatistics
School of Public Health
University of Minnesota

April 22, 2010

- Scientific survey sampling (Neyman, 1934) represents one of statistics' greatest contribution to science

- Scientific survey sampling (Neyman, 1934) represents one of statistics' greatest contribution to science
- It provides the remarkable ability to obtain useful inferences about large populations from modest samples, with measurable uncertainty

- Scientific survey sampling (Neyman, 1934) represents one of statistics' greatest contribution to science
- It provides the remarkable ability to obtain useful inferences about large populations from modest samples, with measurable uncertainty
- It forms the backbone of data collection in science and government

- Scientific survey sampling (Neyman, 1934) represents one of statistics' greatest contribution to science
- It provides the remarkable ability to obtain useful inferences about large populations from modest samples, with measurable uncertainty
- It forms the backbone of data collection in science and government
- Traditional (and widely favoured) approach: Randomization-based

- Scientific survey sampling (Neyman, 1934) represents one of statistics' greatest contribution to science
- It provides the remarkable ability to obtain useful inferences about large populations from modest samples, with measurable uncertainty
- It forms the backbone of data collection in science and government
- Traditional (and widely favoured) approach: Randomization-based
- Advocating Bayes for survey sampling is like "swimming upstream": Modelling assumptions of any kind are anathema here, let alone priors and further subjectivity that Bayes brings along!

- For a population with N units, let $\mathbf{Y} = (Y_1, \ldots, Y_N)$ denote the population variable

- For a population with N units, let $\mathbf{Y} = (Y_1, \ldots, Y_N)$ denote the population variable
- Y_i is the value of unit i in the population

- For a population with N units, let $\mathbf{Y} = (Y_1, \ldots, Y_N)$ denote the population variable
- Y_i is the value of unit i in the population
- Let $\mathbf{I} = (I_1, \ldots, I_N)$ be the set of *inclusion indicator variables*

- For a population with N units, let $\mathbf{Y} = (Y_1, \ldots, Y_N)$ denote the population variable
- Y_i is the value of unit i in the population
- Let $\mathbf{I} = (I_1, \ldots, I_N)$ be the set of *inclusion indicator variables*
- $I_i = 1$ if unit i is selected in the sample; $I_i = 0$ otherwise

- For a population with N units, let $\mathbf{Y} = (Y_1, \ldots, Y_N)$ denote the population variable
- Y_i is the value of unit i in the population
- Let $\mathbf{I} = (I_1, \ldots, I_N)$ be the set of *inclusion indicator variables*
- $I_i = 1$ if unit i is selected in the sample; $I_i = 0$ otherwise
- Let $\mathbf{I}'$ denote the complement of $\mathbf{I}$, obtained by "switching" the 1's to 0's and 0's to 1's in $\mathbf{I}$

- For a population with N units, let $\mathbf{Y} = (Y_1, \ldots, Y_N)$ denote the population variable
- Y_i is the value of unit i in the population
- Let $\mathbf{I} = (I_1, \ldots, I_N)$ be the set of *inclusion indicator variables*
- $I_i = 1$ if unit i is selected in the sample; $I_i = 0$ otherwise
- Let $\mathbf{I}'$ denote the complement of $\mathbf{I}$, obtained by "switching" the 1's to 0's and 0's to 1's in $\mathbf{I}$
- Thus, $\mathbf{I}'$ represents the *exclusion indicator variables*, indexing units *not* selected in the sample.

- Simple Random Sampling With Replacement (SRSWR): We draw a unit from the population, put it back and draw again.

- Simple Random Sampling With Replacement (SRSWR): We draw a unit from the population, put it back and draw again.
- Simple Random Sampling Without Replacement (SRSWOR): We draw a unit from the population, keep it aside and draw again

- Simple Random Sampling With Replacement (SRSWR): We draw a unit from the population, put it back and draw again.
- Simple Random Sampling Without Replacement (SRSWOR): We draw a unit from the population, keep it aside and draw again
- SRSWR leads to binomial distribution; chances of units being drawn remain the same between draws

- Simple Random Sampling With Replacement (SRSWR): We draw a unit from the population, put it back and draw again.
- Simple Random Sampling Without Replacement (SRSWOR): We draw a unit from the population, keep it aside and draw again
- SRSWR leads to binomial distribution; chances of units being drawn remain the same between draws
- SRSWOR leads to hypergeometric distribution: chances of units being drawn change from draw to draw

- Simple Random Sampling With Replacement (SRSWR): We draw a unit from the population, put it back and draw again.
- Simple Random Sampling Without Replacement (SRSWOR): We draw a unit from the population, keep it aside and draw again
- SRSWR leads to binomial distribution; chances of units being drawn remain the same between draws
- SRSWOR leads to hypergeometric distribution: chances of units being drawn change from draw to draw
- Key questions:
  - What are the random variables?

- Simple Random Sampling With Replacement (SRSWR): We draw a unit from the population, put it back and draw again.
- Simple Random Sampling Without Replacement (SRSWOR): We draw a unit from the population, keep it aside and draw again
- SRSWR leads to binomial distribution; chances of units being drawn remain the same between draws
- SRSWOR leads to hypergeometric distribution: chances of units being drawn change from draw to draw
- Key questions:
    - What are the random variables?
    - What are the parameters?

- Let $\mathbf{Y}_I$ be the collection of units selected in the sample; we often write $\mathbf{Y}_I$ as $\mathbf{y} = (y_1, \ldots, y_n)$.

- Let $\mathbf{Y}_I$ be the collection of units selected in the sample; we often write $\mathbf{Y}_I$ as $\mathbf{y} = (y_1, \ldots, y_n)$.
- IMPORTANT: The lower-case y_i's simply index the sample; thus y_i *does not* necessarily represent Y_i, but the i-th member of the sample.

- Let $\mathbf{Y}_I$ be the collection of units selected in the sample; we often write $\mathbf{Y}_I$ as $\mathbf{y} = (y_1, \ldots, y_n)$.
- IMPORTANT: The lower-case y_i's simply index the sample; thus y_i *does not* necessarily represent Y_i, but the i-th member of the sample.
- Let $\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i$ be the population mean

- Let $\mathbf{Y}_I$ be the collection of units selected in the sample; we often write $\mathbf{Y}_I$ as $\mathbf{y} = (y_1, \ldots, y_n)$.

- IMPORTANT: The lower-case y_i's simply index the sample; thus y_i *does not* necessarily represent Y_i, but the i-th member of the sample.

- Let $\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i$ be the population mean

- Let $\bar{Y}_I = \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ be the sample mean; i.e. the mean of $\mathbf{Y}_I$

- Let $\mathbf{Y}_I$ be the collection of units selected in the sample; we often write $\mathbf{Y}_I$ as $\mathbf{y} = (y_1, \ldots, y_n)$.

- IMPORTANT: The lower-case y_i's simply index the sample; thus y_i *does not* necessarily represent Y_i, but the i-th member of the sample.

- Let $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ be the population mean

- Let $\bar{Y}_I = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ be the sample mean; i.e. the mean of $\mathbf{Y}_I$

- Let $\mathbf{Y}_{I'}$ be the collection of units *excluded* from the sample; let $\bar{Y}_{I'}$ be the mean of $\mathbf{Y}_{I'}$.

- Let $\mathbf{Y}_I$ be the collection of units selected in the sample; we often write $\mathbf{Y}_I$ as $\mathbf{y} = (y_1, \ldots, y_n)$.

- IMPORTANT: The lower-case y_i's simply index the sample; thus y_i *does not* necessarily represent Y_i, but the i-th member of the sample.

- Let $\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i$ be the population mean

- Let $\bar{Y}_I = \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ be the sample mean; i.e. the mean of $\mathbf{Y}_I$

- Let $\mathbf{Y}_{I'}$ be the collection of units *excluded* from the sample; let $\bar{Y}_{I'}$ be the mean of $\mathbf{Y}_{I'}$.

- Then $\bar{Y} = f\bar{y} + (1 - f)\bar{Y}_{I'}$; $f = \frac{n}{N}$ is the *sampling fraction*

- Key questions answered:
  - What are the random variables? Answer: The I_i's
  - What are the parameters? Answer: The Y_i's.

- Key questions answered:
  - What are the random variables? Answer: The I_i's
  - What are the parameters? Answer: The Y_i's.
- For SRSWOR: $E[I_i] = P(I_i = 1) = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$ for each i.

- Key questions answered:
  - What are the random variables? Answer: The I_i's
  - What are the parameters? Answer: The Y_i's.
- For SRSWOR: $E[I_i] = P(I_i = 1) = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$ for each i.
- Unbiasedness of the sample mean $\bar{y}$: Write $\bar{y} = \frac{1}{n}\sum_{i=1}^{N} I_i Y_i$ and see:

$$E[\bar{y}] = \frac{1}{n}\sum_{i=1}^{N} E[I_i]Y_i = \frac{1}{n}\sum_{i=1}^{N} \frac{n}{N}Y_i = \bar{Y}.$$

- Key questions answered:
  - What are the random variables? Answer: The I_i's
  - What are the parameters? Answer: The Y_i's.
- For SRSWOR: $E[I_i] = P(I_i = 1) = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$ for each i.
- Unbiasedness of the sample mean $\bar{y}$: Write $\bar{y} = \frac{1}{n}\sum_{i=1}^{N} I_i Y_i$ and see:

$$E[\bar{y}] = \frac{1}{n}\sum_{i=1}^{N} E[I_i]Y_i = \frac{1}{n}\sum_{i=1}^{N} \frac{n}{N} Y_i = \bar{Y}.$$

- For SRSWR: $\bar{y} = \sum_{i=1}^{N} W_i Y_i$, where W_i is the number of times unit i appears in the sample. Then, $W_i \sim Bin(n, \frac{1}{N})$, so $E[W_i] = \frac{n}{N}$, hence $E[\bar{y}] = \frac{1}{n}\sum_{i=1}^{N} E[W_i]Y_i = \frac{1}{n}\sum_{i=1}^{N} \frac{n}{N} Y_i = \bar{Y}$

- We need a prior for each population unit: Y_i

- We need a prior for each population unit: Y_i
- Suppose $Y_i \overset{iid}{\sim} N(\mu, \sigma^2)$

- We need a prior for each population unit: Y_i
- Suppose $Y_i \overset{iid}{\sim} N(\mu, \sigma^2)$
- Let us first assume σ^2 is known and $P(\mu) \propto 1$.

- We need a prior for each population unit: Y_i
- Suppose $Y_i \overset{iid}{\sim} N(\mu, \sigma^2)$
- Let us first assume σ^2 is known and $P(\mu) \propto 1$.
- What is “seen”? $D = (\mathbf{Y}_I, \mathbf{I})$

- We need a prior for each population unit: Y_i
- Suppose $Y_i \overset{iid}{\sim} N(\mu, \sigma^2)$
- Let us first assume σ^2 is known and $P(\mu) \propto 1$.
- What is "seen"? $D = (\mathbf{Y}_I, \mathbf{I})$
- Note that $P(D \mid \mu, \sigma^2) = P(\mathbf{I})P(\mathbf{Y}_I \mid \mu, \sigma^2) \propto P(\mathbf{Y}_I \mid \mu, \sigma^2)$; so likelihood is product of n independent $N(\mu, \sigma^2)$.

- We need a prior for each population unit: Y_i
- Suppose $Y_i \overset{iid}{\sim} N(\mu, \sigma^2)$
- Let us first assume σ^2 is known and $P(\mu) \propto 1$.
- What is “seen”? $D = (\mathbf{Y}_I, \mathbf{I})$
- Note that $P(D \mid \mu, \sigma^2) = P(\mathbf{I})P(\mathbf{Y}_I \mid \mu, \sigma^2) \propto P(\mathbf{Y}_I \mid \mu, \sigma^2)$; so likelihood is product of n independent $N(\mu, \sigma^2)$.
- We need the posterior distribution

$$P(\bar{Y} \mid D, \sigma^2) = \int P(\bar{Y} \mid D, \mu, \sigma^2) P(\mu \mid D, \sigma^2) d\mu.$$

- Let us take a closer look at $P(\bar{Y} \mid D, \sigma^2)$.

- Let us take a closer look at $P(\bar{Y} \mid D, \sigma^2)$.
- Observe that $\bar{Y} = f\bar{y} + (1-f)\bar{Y}_{I'}$.

- Let us take a closer look at $P(\bar{Y} \mid D, \sigma^2)$.
- Observe that $\bar{Y} = f\bar{y} + (1-f)\bar{Y}_{I'}$.
- Given D, $\bar{y}$ is fixed; so the posterior distribution of $\bar{Y}_{I'}$ will determine the posterior distribution of $\bar{Y}$.

- Let us take a closer look at $P(\bar{Y} \mid D, \sigma^2)$.
- Observe that $\bar{Y} = f\bar{y} + (1-f)\bar{Y}_{I'}$.
- Given D, $\bar{y}$ is fixed; so the posterior distribution of $\bar{Y}_{I'}$ will determine the posterior distribution of $\bar{Y}$.
- So, what is the posterior distribution $P(\bar{Y}_{I'} \mid D, \sigma^2)$?

- Let us take a closer look at $P(\bar{Y} \mid D, \sigma^2)$.
- Observe that $\bar{Y} = f\bar{y} + (1-f)\bar{Y}_{I'}$.
- Given D, $\bar{y}$ is fixed; so the posterior distribution of $\bar{Y}_{I'}$ will determine the posterior distribution of $\bar{Y}$.
- So, what is the posterior distribution $P(\bar{Y}_{I'} \mid D, \sigma^2)$?

$$P(\bar{Y}_{I'} \mid D, \sigma^2) = \int P(\bar{Y}_{I'} \mid \mu, \sigma^2) P(\mu \mid D, \sigma^2) d\mu,$$

where we use $P(\bar{Y}_{I'} \mid D, \mu, \sigma^2) = P(\bar{Y}_{I'} \mid \mu, \sigma^2)$. Note $P(\bar{Y}_{I'} \mid \mu, \sigma^2) = N\left(\mu, \frac{\sigma^2}{N(1-f)}\right)$

- Let us take a closer look at $P(\bar{Y} \mid D, \sigma^2)$.
- Observe that $\bar{Y} = f\bar{y} + (1-f)\bar{Y}_{I'}$.
- Given D, $\bar{y}$ is fixed; so the posterior distribution of $\bar{Y}_{I'}$ will determine the posterior distribution of $\bar{Y}$.
- So, what is the posterior distribution $P(\bar{Y}_{I'} \mid D, \sigma^2)$?

$$P(\bar{Y}_{I'} \mid D, \sigma^2) = \int P(\bar{Y}_{I'} \mid \mu, \sigma^2) P(\mu \mid D, \sigma^2) d\mu,$$

  where we use $P(\bar{Y}_{I'} \mid D, \mu, \sigma^2) = P(\bar{Y}_{I'} \mid \mu, \sigma^2)$. Note $P(\bar{Y}_{I'} \mid \mu, \sigma^2) = N\left(\mu, \frac{\sigma^2}{N(1-f)}\right)$
- Can we simplify this posterior without integrating?

- What is the posterior distribution $P(\mu \mid D, \sigma^2)$?

- What is the posterior distribution $P(\mu \mid D, \sigma^2)$?
- Note that $P(\mu \mid D, \sigma^2) \propto P(D \mid \mu, \sigma^2)$

- What is the posterior distribution $P(\mu \mid D, \sigma^2)$?
- Note that $P(\mu \mid D, \sigma^2) \propto P(D \mid \mu, \sigma^2)$
- $\implies P(\mu \mid D, \sigma^2) = N\left(\bar{y}, \frac{\sigma^2}{n}\right)$.

- What is the posterior distribution $P(\mu \,|\, D, \sigma^2)$?
- Note that $P(\mu \,|\, D, \sigma^2) \propto P(D \,|\, \mu, \sigma^2)$
- $\implies P(\mu \,|\, D, \sigma^2) = N\left(\bar{y}, \frac{\sigma^2}{n}\right)$.
- Conditional on D and σ^2, we can write:

$$\begin{aligned} \bar{Y} &= f\bar{y} + (1-f)\bar{Y}_{I'}; \\ \bar{Y}_{I'} &= \mu + u_1;\ u_1 \sim N\left(0, \frac{\sigma^2}{N(1-f)}\right); \\ \Rightarrow \bar{Y} &= f\bar{y} + (1-f)\mu + (1-f)u_1 \end{aligned}$$

- Also, conditional on D and σ^2, $\mu = \bar{y} + u_2;\ u_2 \sim N\left(0, \frac{\sigma^2}{n}\right)$.

- Then:

$$\bar{Y} = \bar{y} + (1 - f)u_2 + (1 - f)u_1$$

- Then:

$$\bar{Y} = \bar{y} + (1 - f)u_2 + (1 - f)u_1$$

- The posterior is $P(\bar{Y} \mid D, \sigma^2)$ is:

$$N\left(\bar{y}, \frac{\sigma^2}{n}(1 - f)\right)$$

- Then:

$$\bar{Y} = \bar{y} + (1-f)u_2 + (1-f)u_1$$

- The posterior is $P(\bar{Y} \mid D, \sigma^2)$ is:

$$N\left(\bar{y}, \frac{\sigma^2}{n}(1-f)\right)$$

- Let σ^2 be unknown. Prior: $P(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$

- Then:

$$\bar{Y} = \bar{y} + (1-f)u_2 + (1-f)u_1$$

- The posterior is $P(\bar{Y} \mid D, \sigma^2)$ is:

$$N\left(\bar{y}, \frac{\sigma^2}{n}(1-f)\right)$$

- Let σ^2 be unknown. Prior: $P(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$
- Then we reproduce classical result:

$$\frac{\bar{Y} - \bar{y}}{\sqrt{\frac{s^2}{n}(1-f)}} \sim t_{n-1}$$

where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ is the sample variance.

- Note that the marginal posterior distribution $P(\sigma^2 \mid D) = IG\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$, where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ is the sample variance.

- Note that the marginal posterior distribution $P(\sigma^2 \mid D) = IG\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$, where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ is the sample variance.

- We can perform exact posterior sampling as follows. For $l = 1, \ldots, L$ do the following:

- Note that the marginal posterior distribution $P(\sigma^2 \mid D) = IG\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$, where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ is the sample variance.

- We can perform exact posterior sampling as follows. For $l = 1, \ldots, L$ do the following:
    - Draw $\sigma^2_{(l)} \sim IG\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$

- Note that the marginal posterior distribution $P(\sigma^2 \mid D) = IG\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$, where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ is the sample variance.

- We can perform exact posterior sampling as follows. For $l = 1, \ldots, L$ do the following:
  - Draw $\sigma^2_{(l)} \sim IG\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$
  - Draw $\mu_{(l)} \sim N\left(\bar{y}, \frac{\sigma^2_{(l)}}{n}(1-f)\right)$

- Note that the marginal posterior distribution $P(\sigma^2 \mid D) = IG\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$, where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ is the sample variance.

- We can perform exact posterior sampling as follows. For $l = 1, \ldots, L$ do the following:
  - Draw $\sigma^2_{(l)} \sim IG\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$
  - Draw $\mu_{(l)} \sim N\left(\bar{y}, \frac{\sigma^2_{(l)}}{n}(1-f)\right)$
  - Draw $\bar{Y}^{(l)}_{I'} \sim N\left(\mu_{(l)}, \frac{\sigma^2_{(l)}}{N(1-f)}\right)$

- Note that the marginal posterior distribution $P(\sigma^2 \mid D) = IG\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$, where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ is the sample variance.

- We can perform exact posterior sampling as follows. For $l = 1, \ldots, L$ do the following:
    - Draw $\sigma^2_{(l)} \sim IG\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$
    - Draw $\mu_{(l)} \sim N\left(\bar{y}, \frac{\sigma^2_{(l)}}{n}(1-f)\right)$
    - Draw $\bar{Y}^{(l)}_{I'} \sim N\left(\mu_{(l)}, \frac{\sigma^2_{(l)}}{N(1-f)}\right)$
    - Set $\bar{Y}^{(l)} = f\bar{y} + (1-f)\bar{Y}^{(l)}_{I'}$.

- Note that the marginal posterior distribution $P(\sigma^2 \mid D) = IG\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$, where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ is the sample variance.

- We can perform exact posterior sampling as follows. For $l = 1, \ldots, L$ do the following:
  - Draw $\sigma^2_{(l)} \sim IG\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$
  - Draw $\mu_{(l)} \sim N\left(\bar{y}, \frac{\sigma^2_{(l)}}{n}(1-f)\right)$
  - Draw $\bar{Y}^{(l)}_{I'} \sim N\left(\mu_{(l)}, \frac{\sigma^2_{(l)}}{N(1-f)}\right)$
  - Set $\bar{Y}^{(l)} = f\bar{y} + (1-f)\bar{Y}^{(l)}_{I'}$.

- The resulting collection $\{\bar{Y}^{(l)}\}_{l=1}^{L}$ is a sample from the posterior distribution of the population mean.

HOMEWORK

- A simple random sample of 30 households was drawn from a township containing 576 households. The numbers of persons per household in the sample were as follows:

$$5, 6, 3, 3, 2, 3, 3, 3, 4, 4, 3, 2, 7, 4, 3, 5, 4, 4, 3, 3, 4, 3, 3, 1, 2, 4, 3, 4, 2, 4$$

  Use a non-informative Bayesian analysis to estimate the total number of people in the area and compute the posterior probability that the population total lies within 10% of the sample estimate.

- From a list of 468 small 2-year colleges in the North-Eastern United States, a simple sample of 100 colleges was drawn. Data for the number of students (y) and the number of teachers (x) for these colleges were summarized as follows:
  - The total number of students in the sample was: $44,987$
  - The total number of teachers in the sample was: $3,079$
  - Also given are the sample sums of squares: $\sum_{i=1}^{n} y_i^2 = 29,881,219$ and $\sum_{i=1}^{n} x_i^2 = 111,090.$

  Assuming a non-informative Bayesian setting and where the population of students and teachers are independent, find the posterior mean and 95% credible interval for the student-teacher ratio in the population (i.e. all the 468 colleges combined).

- A general theory for Bayesian missing data analysis

- A general theory for Bayesian missing data analysis
- Let $y = (y_1, \ldots, y_N)$ be the units in the population. We assume a likelihood for this population, say $p(y \mid \theta)$, where θ are the parameters in the model.

- A general theory for Bayesian missing data analysis
- Let $y = (y_1, \ldots, y_N)$ be the units in the population. We assume a likelihood for this population, say $p(y \mid \theta)$, where θ are the parameters in the model.
- Let $I = (I_1, \ldots, I_N)$ be the set of “inclusion indicators”.

- A general theory for Bayesian missing data analysis
- Let $y = (y_1, \ldots, y_N)$ be the units in the population. We assume a likelihood for this population, say $p(y \mid \theta)$, where θ are the parameters in the model.
- Let $I = (I_1, \ldots, I_N)$ be the set of "inclusion indicators".
- Let us denote by "obs" the index set $\{i : I_i = 1\}$ and by "mis" the index set $\{i; I_i = 0\}$

- A general theory for Bayesian missing data analysis
- Let $y = (y_1, \ldots, y_N)$ be the units in the population. We assume a likelihood for this population, say $p(y \,|\, \theta)$, where θ are the parameters in the model.
- Let $I = (I_1, \ldots, I_N)$ be the set of "inclusion indicators".
- Let us denote by "obs" the index set $\{i : I_i = 1\}$ and by "mis" the index set $\{i; I_i = 0\}$
- So, y_{obs} is the set of observed data points and y_{mis} is for the unobserved data. We will often write $y = (y_{obs}, y_{mis})$.

- We will break the joint probability model into two parts

- We will break the joint probability model into two parts
  - The model for the "complete data" $p(y \mid \theta)$ – including observed and missing data

- We will break the joint probability model into two parts
  - The model for the “complete data” $p(y \mid \theta)$ – including observed and missing data
  - The model for the inclusion vector I, say $p(I \mid y, \phi)$

- We will break the joint probability model into two parts
  - The model for the "complete data" $p(y \mid \theta)$ – including observed and missing data
  - The model for the inclusion vector I, say $p(I \mid y, \phi)$
- We define the *complete-likelihood* as:

$$p(y, I \mid \theta, \phi) = p(y \mid \theta) p(I \mid y, \phi)$$

- We will break the joint probability model into two parts
  - The model for the "complete data" $p(y \mid \theta)$ – including observed and missing data
  - The model for the inclusion vector I, say $p(I \mid y, \phi)$
- We define the *complete-likelihood* as:

$$p(y, I \mid \theta, \phi) = p(y \mid \theta) p(I \mid y, \phi)$$

- Parameters: Scientific interest usually revolves around estimation of θ (sometimes called the "super-population" parameters)

- We will break the joint probability model into two parts
  - The model for the "complete data" $p(y \mid \theta)$ – including observed and missing data
  - The model for the inclusion vector I, say $p(I \mid y, \phi)$
- We define the *complete-likelihood* as:

$$p(y, I \mid \theta, \phi) = p(y \mid \theta) p(I \mid y, \phi)$$

- Parameters: Scientific interest usually revolves around estimation of θ (sometimes called the "super-population" parameters)
- The parameters ϕ index the missingness model (often called "design" parameters).

- The “complete-likelihood” is *not* the data likelihood as it involves the missing data component as well.

- The "complete-likelihood" is *not* the data likelihood as it involves the missing data component as well.

- The actual information available is (y_{obs}, I). So the data likelihood is:

$$p(y_{obs}, I \mid \theta, \phi) = \int p(y, I \mid \theta, \phi) dy_{mis}$$

- The "complete-likelihood" is *not* the data likelihood as it involves the missing data component as well.

- The actual information available is (y_{obs}, I). So the data likelihood is:

$$p(y_{obs}, I \mid \theta, \phi) = \int p(y, I \mid \theta, \phi) dy_{mis}$$

- The joint posterior distribution is

$$\begin{aligned} p(\theta, \phi \mid y_{obs}, I) &\propto p(\theta, \phi) p(y_{obs}, I \mid \theta, \phi) \\ &\propto p(\theta, \phi) \int p(y \mid \theta) p(I \mid y, \phi) dy_{mis} \end{aligned}$$

- The “complete-likelihood” is *not* the data likelihood as it involves the missing data component as well.
- The actual information available is (y_{obs}, I). So the data likelihood is:

$$p(y_{obs}, I \mid \theta, \phi) = \int p(y, I \mid \theta, \phi) dy_{mis}$$

- The joint posterior distribution is

$$\begin{aligned} p(\theta, \phi \mid y_{obs}, I) &\propto p(\theta, \phi) p(y_{obs}, I \mid \theta, \phi) \\ &\propto p(\theta, \phi) \int p(y \mid \theta) p(I \mid y, \phi) dy_{mis} \end{aligned}$$

- These integrals are evaluated by drawing samples (y_{mis}, θ, ϕ) from $p(y_{mis}, \theta, \phi \,|\, y_{obs}, I)$

- These integrals are evaluated by drawing samples (y_{mis}, θ, ϕ) from $p(y_{mis}, \theta, \phi \,|\, y_{obs}, I)$

- Typically, we first compute samples $(\theta^{(l)}, \phi^{(l)})$ from $p(\theta, \phi \,|\, y_{obs}, I)$

- These integrals are evaluated by drawing samples (y_{mis}, θ, ϕ) from $p(y_{mis}, \theta, \phi \,|\, y_{obs}, I)$
- Typically, we first compute samples $(\theta^{(l)}, \phi^{(l)})$ from $p(\theta, \phi \,|\, y_{obs}, I)$
- Then we draw $y_{mis}^{(l)} \sim p(y_{mis} \,|\, y_{obs}, I, \theta^{(l)}, \phi^{(l)})$

- These integrals are evaluated by drawing samples (y_{mis}, θ, ϕ) from $p(y_{mis}, \theta, \phi \,|\, y_{obs}, I)$
- Typically, we first compute samples $(\theta^{(l)}, \phi^{(l)})$ from $p(\theta, \phi \,|\, y_{obs}, I)$
- Then we draw $y_{mis}^{(l)} \sim p(y_{mis} \,|\, y_{obs}, I, \theta^{(l)}, \phi^{(l)})$
- Computations often simplify from assumptions:
  - $p(I \,|\, y, \phi) = p(I \,|\, y_{obs}, \phi)$

- These integrals are evaluated by drawing samples (y_{mis}, θ, ϕ) from $p(y_{mis}, \theta, \phi \mid y_{obs}, I)$
- Typically, we first compute samples $(\theta^{(l)}, \phi^{(l)})$ from $p(\theta, \phi \mid y_{obs}, I)$
- Then we draw $y_{mis}^{(l)} \sim p(y_{mis} \mid y_{obs}, I, \theta^{(l)}, \phi^{(l)})$
- Computations often simplify from assumptions:
  - $p(I \mid y, \phi) = p(I \mid y_{obs}, \phi)$
  - $p(I \mid y, \phi) = p(I \mid \phi) = p(I)$

- These integrals are evaluated by drawing samples (y_{mis}, θ, ϕ) from $p(y_{mis}, \theta, \phi \mid y_{obs}, I)$
- Typically, we first compute samples $(\theta^{(l)}, \phi^{(l)})$ from $p(\theta, \phi \mid y_{obs}, I)$
- Then we draw $y_{mis}^{(l)} \sim p(y_{mis} \mid y_{obs}, I, \theta^{(l)}, \phi^{(l)})$
- Computations often simplify from assumptions:
  - $p(I \mid y, \phi) = p(I \mid y_{obs}, \phi)$
  - $p(I \mid y, \phi) = p(I \mid \phi) = p(I)$
  - $p(\phi \mid \theta) = p(\phi)$
  - $p(y_{mis} \mid \theta^{(l)}, y_{obs}) = p(y_{mis} \mid \theta^{(l)})$.