Module 6 : Bayesian inference

Marie-Abèle Bind

Design of Experiments - Stat 140

Spring 2026

Bayesian causal model

- Rubin (1975) departs from Fisher and Neyman : Y(0) and Y(1) are now random variables, as are all causal estimands.

- This "model-based" or "predictive" approach is very flexible : can easily accommodate a wide variety of estimands (e.g., quantiles).

- Goal : Conditional distribution of the full vector of missing potential outcomes given the observed data : $f(Y^{mis}|Y^{obs}, W^{obs})$, from which any causal estimand can be calculated.

Job training data subset (Table 8.2)

- Randomized evaluation of a job training (National Supported Work) program. We focus our attention on six units.
- W_i was completely randomized (1 : job training program, 0 : otherwise)
- Y corresponds to the earnings received with and without the job training program.

Unit i	W_i	$Y_i(W_i = 0)$	$Y_i(W_i = 1)$	Y_i^{obs}
1	0	0	?	0
2	1	?	9.9	9.9
3	0	12.4	?	12.4
4	1	?	3.6	3.6
5	0	0	?	0
6	1	?	24.9	24.9

Causal estimand and estimator

Unit i	W_i	$Y_i(W_i = 0)$	$Y_i(W_i = 1)$	Y_i^{obs}
1	0	0	?	0
2	1	?	9.9	9.9
3	0	12.4	?	12.4
4	1	?	3.6	3.6
5	0	0	?	0
6	1	?	24.9	24.9

$\tau = f_1(Y(0), Y(1)) = \frac{1}{6}\sum_{i=1}^{6}[Y_i(W_i = 1) - Y_i(W_i = 0)]$

$\tau = f_2(Y^{obs}, Y^{mis}) = \frac{1}{6}\sum_{i=1}^{6}[(2W_i - 1)(Y_i^{obs} - Y_i^{mis})]$

Game here : to multiply impute the missing potential outcomes $\hat{Y}_i^{mis}$ given the observed values Y^{obs} and the treatment assignments W.

$\hat{\tau} = f_2(Y^{obs}, \hat{Y}_i^{mis}) = \frac{1}{6}\sum_{i=1}^{6}[(2W_i - 1)(Y_i^{obs} - \hat{Y}_i^{mis}]$

Naive approaches

Unit i	W_i	$Y_i(W_i = 0)$	$Y_i(W_i = 1)$	Y_i^{obs}
1	0	0	?	0
2	1	?	9.9	9.9
3	0	12.4	?	12.4
4	1	?	3.6	3.6
5	0	0	?	0
6	1	?	24.9	24.9

How to impute Y_1^{mis} ?

Naive approaches

Unit i	W_i	$Y_i(W_i = 0)$	$Y_i(W_i = 1)$	Y_i^{obs}
1	0	0	11.0	0
2	1	?	9.9	9.9
3	0	12.4	?	12.4
4	1	?	3.6	3.6
5	0	0	?	0
6	1	?	24.9	24.9

How to impute Y_1^{mis} ?

- ? <- abs(rnorm(1,mean=0,sd=27))=11.0

Naive approaches

Unit i	W_i	$Y_i(W_i = 0)$	$Y_i(W_i = 1)$	Y_i^{obs}
1	0	0	8.7	0
2	1	2.8	9.9	9.9
3	0	12.4	44.9	12.4
4	1	11.1	3.6	3.6
5	0	0	24.2	0
6	1	28.2	24.9	24.9

$$\hat{\tau} = 19.4 - 9.1 = 10.3$$

How to impute Y_1^{mis} ?

- ? <- abs(rnorm(1,mean=0,sd=27))=11.0

Naive approaches

Unit i	W_i	$Y_i(W_i = 0)$	$Y_i(W_i = 1)$	Y_i^{obs}
1	0	0	12.8	0
2	1	4.1	9.9	9.9
3	0	12.4	12.8	12.4
4	1	4.1	3.6	3.6
5	0	0	12.8	0
6	1	4.1	24.9	24.9

$$\hat{\tau} = 12.8 - 4.1 = 8.7$$

How to impute Y_1^{mis} ?

- ? <- abs(rnorm(1,mean=0,sd=27))=11.0
- ? <- $\frac{1}{N_t} \sum_{i:W_i=1} Y_i^{obs}$=12.8

Naive approaches

Unit i	W_i	$Y_i(W_i = 0)$	$Y_i(W_i = 1)$	Y_i^{obs}
1	0	0	9.9	0
2	1	12.4	9.9	9.9
3	0	12.4	24.9	12.4
4	1	12.4	3.6	3.6
5	0	0	24.9	0
6	1	12.4	24.9	24.9

$$\hat{\tau} = 16.4 - 8.3 = 8.1$$

How to impute Y_1^{mis} ?

- ? <- abs(rnorm(1,mean=0,sd=27))=11.0
- ? <- $\frac{1}{N_t} \sum_{i:W_i=1} Y_i^{obs}$=12.8
- ? <- sample(1,x=c(9.9,3.6,24.9))=9.9

Single imputation

- These naive approaches rely on single imputation. By treating a single value as known, single imputation does not account for imputation uncertainty. Thus, standard errors computed are systematically underestimated (e.g., confidence intervals are too narrow).

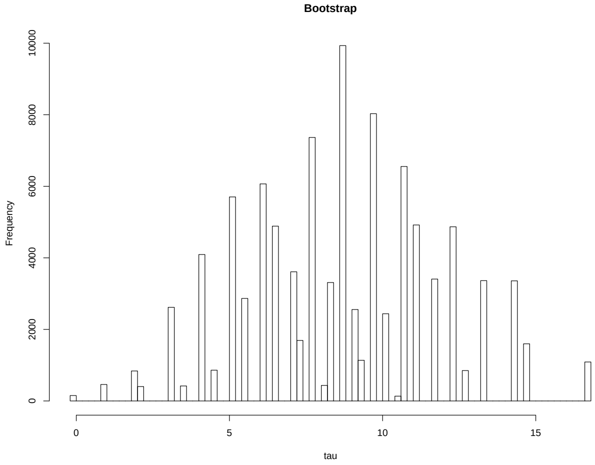
Drawing at random from the empirical distribution of Y_{obs}

Let N_{obs} and N_{mis} be the number of observed and missing potential outcomes.

1. Draw N_{mis} missing potential outcomes from the empirical distribution of the observed potential outcomes.

2. Repeat Step 1 N_{rep} times.

Drawing at random from the empirical distribution of Y_{obs}



$$\hat{\tau} = 8.7 \text{ and } \sqrt{\hat{var}(\hat{\tau})} = 3.1$$

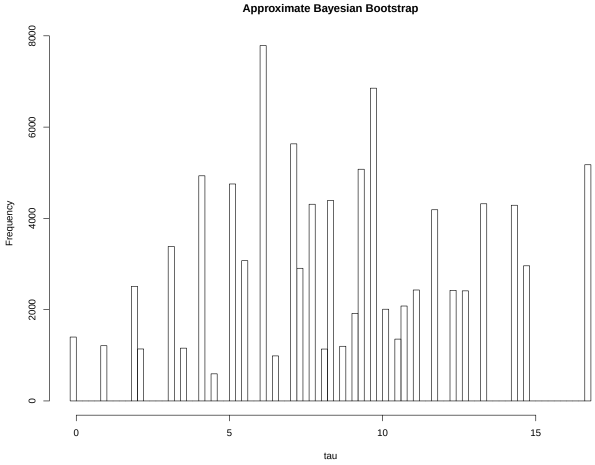
Approximate Bayesian Bootstrap (Rubin and Schenker, 1986)

1. Draw N_{mis} variables from the empirical distribution of the observed potential outcomes, call it Y_{temp}.

2. With this first draw, Y_{temp}, draw each missing potential outcome at random from Y_{temp}.

3. Repeat Steps 1 and 2 N_{rep} times.

Approximate Bayesian Bootstrap

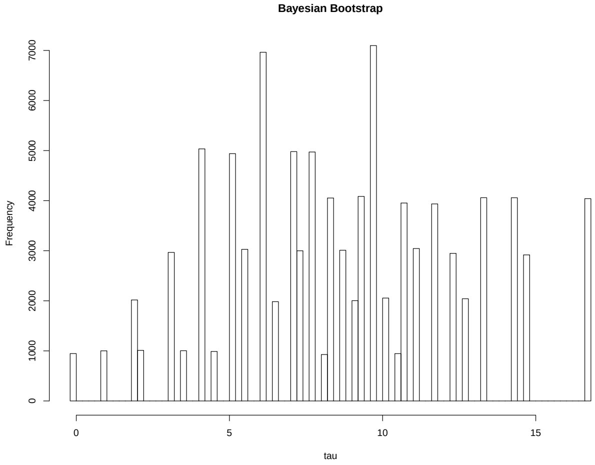


$$\hat{\tau} = 8.6 \text{ and } \sqrt{\hat{var}(\hat{\tau})} = 4.0$$

Bayesian Bootstrap (Rubin, 1981)

1. Draw N_{obs}-1 variables from uniform(0,1).
2. Order them.
3. Calculate the N_{obs} distances corresponding to the gaps on the [0,1] interval.
4. To impute the N_{mis} missing potential outcomes, draw from the empirical distribution of the observed potential outcomes (with replacement) with probabilities equal to the N_{obs} gaps calculated above.
5. Repeat Steps 1 to 4 N_{rep} times.

Bayesian Bootstrap



$$\hat{\tau} = 8.7 \text{ and } \sqrt{\hat{var}(\hat{\tau})} = 3.8$$

Draws from the empirical distribution

- Improvement : we impute by drawing from the observed outcomes for each treatment, with replacement, which provides an estimate of the entire distribution of the estimand and an estimate of the variability of $\hat{\tau}$. Resampling methods propagate imputation uncertainty.
- Limitation : drawing from the empirical distributions still does not address fully the uncertainty in estimating the probability of each observed value : we impute Y^{mis} as if we knew the exact distribution of each of the potential outcomes.
- None of these methods is model-free, in the sense that they all make assumptions about the predictive distribution of the missing values in order to generate estimates.

Multiple imputation and repeated imputation

- Multiple imputation refers to the procedure of replacing each missing value by a more than two imputed values. First proposed by Rubin in 1978. Takes more work to create the imputations and analyze the results.

- Repeated imputation : Y^{mis} should be drawn according to the following protocol. For each model being considered, the imputations of Y^{mis} are repetitions from the posterior predictive distribution of Y^{mis}, each repetition corresponding to an independent drawing of the parameters and missing values.

Bayesian causal model (Rubin, 1975)

- To take into account the uncertainty, we need a model for the potential outcomes (observed and missing), which formally addresses the uncertainty about possible values of missing potential outcomes, and a model for the assignment mechanism.
- For each replication, assuming ignorability of the assignment mechanism :
  1. Draw model parameters given the observed data.
  2. Impute all missing potential outcomes given model parameters.
  3. Calculate estimand given observed and imputed potential outcomes.
  4. Repeat Steps 1 to 4 to obtain an interval for the estimand.

Bayesian causal model

- Intermediate goal : to build a model for the missing potential outcomes, given the observed data, $f(Y^{mis}|Y^{obs}, W)$. Once we have such a model, we can derive the distribution for the estimand of interest, $\tau = \tau(Y(0), Y(1), W)$.

- Assuming ignorability, the first input for the Bayesian causal model is a model for the joint distribution of the two potential outcomes (Y(0),Y(1)) :

$$f(Y(0), Y(1)) = \int \prod_{i=1}^{N} f(Y_i(0), Y_i(1)|\theta)p(\theta)d\theta$$

(e.g., bivariate normal distribution).

- The second input is the prior distribution of θ, p(θ).

Bayesian causal model

Four steps that can be followed in a completely randomized experiment with no covariates :

1. Derive $f(Y^{mis}|Y^{obs}, W, \theta)$.
2. Derive the posterior distribution of the parameter θ, $p(\theta|Y^{obs}, W)$.
3. Derive the posterior distribution of the missing outcomes $f(Y^{mis}|Y^{obs}, W)$.
4. Derive the posterior distribution of τ, $f(\tau|Y^{obs}, W)$.

Simplified example of a Bayesian causal model

Input 1 : $(Y(0), Y(1) \mid \mu_C, \mu_T, \sigma_C^2, \sigma_T^2) \sim N\left(\begin{pmatrix} \mu_C \\ \mu_T \end{pmatrix}, \begin{pmatrix} \sigma_C^2 & 0 \\ 0 & \sigma_T^2 \end{pmatrix} \right)$

Note : Because the correlation between Y(0) and Y(1) is assumed to be zero, we can impute the missing potential outcomes among the treated separately from the missing potential outcomes among the controls.

Input 2 : $\sigma_C^2 \sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2)$ and $\mu_C \mid \sigma_C^2 \sim N(\mu_0, \frac{\sigma_C^2}{\kappa_0})$

Algorithm :

1. Draw σ_C^2 such that $\frac{1}{\sigma_C^2} \sim \frac{1}{(N-1)s_C^2} \chi^2(N-1)$
2. Draw $\mu_C \mid \sigma_C^2$, Y, W $\sim N(\bar{y}_C, \frac{\sigma_C^2}{n})$
3. Draw $Y_i^{mis} \mid Y^{obs}, W, \mu_C, \mu_T, \sigma_C^2, \sigma_T^2 \sim N(W_i\mu_C + (1-W_i)\mu_T); W_i\sigma_C^2 + (1-W_i)\sigma_T^2)$
4. We repeat this logic for imputing the missing treated potential outcomes among the controls.

Bayesian causal model

- Limitations :
  - Wider intervals, generally, but honest.
  - Have to choose a model and all models are wrong.
- Advantages :
  - Flexibility : can separate models for active treatment and control outcomes.
  - Estimand does not have to be difference in means.
  - Takes into account the uncertainty about the distribution of potential outcomes and model parameters.