Time Series 2: Imputation and MCMC

-Experimental Statistics II-

Lecturer: Darren Homrighausen, PhD

An example

A clinical trial

We are running a clinical trial for a depression medication

We gather 50 participants and (blindly) randomly assign half to a treatment and half to placebo

We interview all 50 participants and score them on a depression index

As well, we record some general health information about them, including risk factors for heart attacks

After 6 weeks, we re-interview them and score them on the same depression index

There are 49 participants that ultimately showed up for the exit interview

A clinical trial

Let's consider some possibilities that happened to that 50th person:

1. The person died after being struck by a car
2. The person died of a heart attack
3. The person died of suicide

In each case, the mechanism that caused the missing data is...

1. ... independent of depression status
2. ... confounded with depression status. However, we can account for the confounding using the auxiliary information (in particular, heart attack risk)
3. ... confounded with depression status and we cannot account for the confounding of missing status and the depression status

Types of missing data

Missingness

MCAR

- Missingness mechanism is independent of the data

$$\mathbb{P}(\text{missing}|\text{complete data}) = \mathbb{P}(\text{missing})$$

- Only statistical implication is a loss of power
  (E.g. The person died after being struck by a car)

MAR

- Missingness mechanism is conditionally independent of the missing data given the observed data

$$\mathbb{P}(\text{missing}|\text{complete data}) = \mathbb{P}(\text{missing}|\text{observed data})$$

- Statistical implications are a loss of power and bias. The missing data can be recovered consistently from the observed data
  (E.g. The person died of a heart attack)

MNAR

- Everything else
- Statistical implications are a loss of power and bias. Needs outside information to recover missing data
  (E.g. The person died of suicide)

Single imputation

Any method that estimates values that missing data would have taken is known as imputation

There are some basic imputation approaches:

- List-wise deletion: This is the simplest approach in which any observation or explanatory variable that has missing values is deleted
- Mean/Median/Mode: Any missing values in an explanatory variable is imputed with the same mean/median/mode from non-missing observations
- Regression: We can run a regression in which the explanatory variable we want to impute is treated as a response and we predict the missing observations.

These are all examples of single imputation schemes as we are inserting/estimating/imputing a single value for a missing measurement

Multiple imputation

Having a missing observation and then using single imputation tends to underestimate the variability in a problem

This can cause any uncertainty quantification to make erroneously small confidence intervals/p-values

To account for this extra uncertainty, multiple imputation can be used to generate many random values for each missing observation

SAS PROCEDURES

By default, `SAS` (really, all software) uses list-wise deletion

But, in many cases, we do a little better using a more sophisticated imputation scheme

We can use the `SAS` procedure `PROC MI` to do multiple imputation

The key component of `PROC MI` is the "MCMC method"

To understand what MCMC means, we need to revisit "Bayesian Statistics"

What is Bayesian Statistics?

There are two many philosophies in statistics:

- Treat parameters as fixed, but unknown quantities

  (This is known as "frequentist statistics")
- Treat parameters as random to represent out uncertainty about them

  (This is known as "Bayesian statistics")

Example: Suppose we want to estimate the mean μ from a sample $Y_1, \ldots, Y_n$

- The frequentist perspective is that μ is a fixed, but unknown, quantity (So, statements like $\overline{Y} \to \mu$ as n goes to infinity has meaning)
- The Bayesian perspective is that μ is random. Hence, it has a distribution...

What is Bayesian Statistics?

We need to specify

- The likelihood $p(y|\mu)$
- The prior $p(\mu)$

Hence, via Bayes' theorem

$$p(\mu|y) = \frac{p(y|\mu)p(\mu)}{p(y)}$$

In words, we want to update the distribution of the parameter μ after observing the data y

Note that we know $p(y|\mu)$ and $p(\mu)$ (as we specified them)

The whole Bayesian bottleneck: We don't know $p(y)$!

What is MCMC?

MCMC stands for "Markov Chain Monte Carlo" and it is a method for sampling from $p(\mu|y)$ even if it is unknown

A Markov Chain is a process in which the past is conditionally independent given the present

- Suppose I flip a coin and take one step to the right if heads and one step to the left if tails. Then, to take the next step, I only need to know where I am, not how I got there

  (This is a Markov Chain)

- Conversely, I flip a coin and take one step to the right if I have cumulatively flipped more heads than tails and left otherwise. Then my movement depends on the entire sequence of coin flips

  (This is not a Markov Chain)

How do Markov Chains and Bayesian statistics relate?

Reminder: The whole Bayesian goal is to compute $p(\mu|y)$

Often, this is impossible

But, if we can sample from it, that's almost as good as knowing what it is

It turns out, we can set up a Markov Chain so that the "long term" behavior of the chain looks like the posterior

Back to that example: Suppose I flip a coin and take one step to the right if heads and one step to the left if tails.

After flipping the coin a large number of times, I can look at the fraction of times I'm at each location. This is known as the stationary distribution

How do Markov Chains and Bayesian statistics relate?

General idea: Sample from the Markov Chain a bunch of times. Eventually, we will be getting draws from the posterior. However, these draws will not be independent. So, we need to...

- Throw away the first few (hundred?) observations (as they were not part of the "long term" behavior)
  (This is known as a burn in)
- Subsample from the draws so that they look more uncorrelated
  (Connection to time series: how to we examine whether something looks uncorrelated?)

(Let's look at the script `MCMCstationaryDist`)

How does the "MCMC method" work?

We iterate between the following two steps

1. Imputation: [Input μ, Σ]

$$\begin{bmatrix} Y_{obs} \\ Y_{mis} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11}, \Sigma_{12} \\ \Sigma_{21}, \Sigma_{22} \end{bmatrix} \right)$$

   Using the BLUP we can form the conditional mean:

$$\mathbb{E} Y_{mis} | Y_{obs} = \mu_2 + \Sigma_{12}^{\top} \Sigma_{11}^{-1} (Y_{obs} - \mu_1)$$

   and conditional variance $\mathbb{V} Y_{mis} | Y_{obs}$ (equation omitted)

   $\rightarrow$ Output: $\hat{Y}_{mis} \sim N(\mathbb{E} Y_{mis} | Y_{obs}, \mathbb{V} Y_{mis} | Y_{obs})$

2. Estimation: [Input: $Y = (Y_{obs}, \hat{Y}_{mis})$]

   2.1 Assume multivariate normality for Y
   (This is the likelihood)
   2.2 Specify a prior

   $\rightarrow$ Output: Estimates of μ, Σ $(\hat{\mu}, \hat{\Sigma})$

How does MCMC work? Prior

Bayes' Theorem:

$$p(\mu|y) = \frac{p(y|\mu)p(\mu)}{p(y)}$$

For the prior, some choices are

- Jeffrey's: A noninformative prior
  (This does not mean uniform, but it does make it invariate to reparameterization)
- Ridge: If the sample covariance $\hat{\Sigma} \propto \sum_{i=1} Y_i Y_i^\top$ is nearly singular, then adding a bit to its diagonals can stabilize the procedure

How does MCMC work?

The output of these iterations will look like

$$(\hat{Y}^1_{mis}, (\mu^1, \Sigma^1)), (\hat{Y}^2_{mis}, (\mu^2, \Sigma^2)), \ldots$$

(This is an example of a Markov chain, which explains one of the "MCs")

It has stationary distribution $p(Y_{mis}, (\mu, \Sigma) | Y_{obs})$

Assuming we

- Perform enough iterations
- Subselect from the iterations to break correlations

  (You can have SAS produce an autocorrelation plot)

Then we have independent draws of the missing observations from this distribution

These draws will constitute the multiple imputation

(Go to `imputation.sas`)

Missingness in practice

Possible considerations:

- Data size/complexity
  (Does it fit in RAM?)
- Business purpose
  (Is data precious? Development time?)
- Are any observations/explanatory variables missing a large fraction of values?
- Type of explanatory variables
  (Any sparsity? Is multivariate normality appropriate?)
- Any atypical missingness indicators?
  (e.g. using -1000 for income to indicate a missing value)

Some resources

https://support.sas.com/documentation/onlinedoc/stat/141/mi.pdf

https://uisug.org.uiowa.edu/sites/uisug.org.uiowa.edu/files/wysiwyg_uploads/Handling%20Missing%20Values%20with%20SAS.pdf