450D Political Methodology IV
TA Section 8

Vincent Bauer

December 1, 2017

Goal Today

- Remaining questions about hierarchical models?
- Multilevel regression with poststratification (MRP)
- Reversible Jump Markov Chain Monte Carlo (RJMCMC)

Multilevel regression with poststratification (MRP)

Motivation: Selection Bias

- There are situations where we know that simple random sampling will give us biased answers

Motivation: Selection Bias

- There are situations where we know that simple random sampling will give us biased answers (selection bias).

Motivation: Selection Bias

- There are situations where we know that simple random sampling will give us biased answers (selection bias).
  - For example, we know that older people are less likely to respond to an internet survey

Motivation: Selection Bias

- There are situations where we know that simple random sampling will give us biased answers (selection bias).
  - For example, we know that older people are less likely to respond to an internet survey
  - Similarly, Doug's QJPS paper shows that there's selection bias on partisanship around salient political events, leading to bigger perceived swings in support.

Motivation: Selection Bias

- There are situations where we know that simple random sampling will give us biased answers (selection bias).
  - For example, we know that older people are less likely to respond to an internet survey
  - Similarly, Doug's QJPS paper shows that there's selection bias on partisanship around salient political events, leading to bigger perceived swings in support.
- The general solution is to weight respondents by their strata.

Motivation: Selection Bias

- There are situations where we know that simple random sampling will give us biased answers (selection bias).
  - For example, we know that older people are less likely to respond to an internet survey
  - Similarly, Doug's QJPS paper shows that there's selection bias on partisanship around salient political events, leading to bigger perceived swings in support.
- The general solution is to weight respondents by their strata.
  - In the survey example, we would get a nationally representative age distribution from the Census

Motivation: Selection Bias

- There are situations where we know that simple random sampling will give us biased answers (selection bias).
  - For example, we know that older people are less likely to respond to an internet survey
  - Similarly, Doug's QJPS paper shows that there's selection bias on partisanship around salient political events, leading to bigger perceived swings in support.
- The general solution is to weight respondents by their strata.
  - In the survey example, we would get a nationally representative age distribution from the Census and weight older respondents more to correct the imbalance.

Motivation: Selection Bias

- There are situations where we know that simple random sampling will give us biased answers (selection bias).
  - For example, we know that older people are less likely to respond to an internet survey
  - Similarly, Doug's QJPS paper shows that there's selection bias on partisanship around salient political events, leading to bigger perceived swings in support.
- The general solution is to weight respondents by their strata.
  - In the survey example, we would get a nationally representative age distribution from the Census and weight older respondents more to correct the imbalance.
  - When we apply these strata after having collected the data

Motivation: Selection Bias

- There are situations where we know that simple random sampling will give us biased answers (selection bias).
  - For example, we know that older people are less likely to respond to an internet survey
  - Similarly, Doug's QJPS paper shows that there's selection bias on partisanship around salient political events, leading to bigger perceived swings in support.
- The general solution is to weight respondents by their strata.
  - In the survey example, we would get a nationally representative age distribution from the Census and weight older respondents more to correct the imbalance.
  - When we apply these strata after having collected the data (instead of sampling these strata by design)

Motivation: Selection Bias

- There are situations where we know that simple random sampling will give us biased answers (selection bias).
  - For example, we know that older people are less likely to respond to an internet survey
  - Similarly, Doug's QJPS paper shows that there's selection bias on partisanship around salient political events, leading to bigger perceived swings in support.
- The general solution is to weight respondents by their strata.
  - In the survey example, we would get a nationally representative age distribution from the Census and weight older respondents more to correct the imbalance.
  - When we apply these strata after having collected the data (instead of sampling these strata by design), this is called post-stratification.

Motivation: Weighting

In theory, post-stratification is great,

Motivation: Weighting

In theory, post-stratification is great, but in practice, we can run into problems:

Motivation: Weighting

In theory, post-stratification is great, but in practice, we can run into problems:

- There are a lot of dimensions that we care about balancing:

Motivation: Weighting

In theory, post-stratification is great, but in practice, we can run into problems:

- There are a lot of dimensions that we care about balancing: age,

Motivation: Weighting

In theory, post-stratification is great, but in practice, we can run into problems:

- There are a lot of dimensions that we care about balancing: age, race,

Motivation: Weighting

In theory, post-stratification is great, but in practice, we can run into problems:

- There are a lot of dimensions that we care about balancing: age, race, gender,

Motivation: Weighting

In theory, post-stratification is great, but in practice, we can run into problems:

- There are a lot of dimensions that we care about balancing: age, race, gender, education,

Motivation: Weighting

In theory, post-stratification is great, but in practice, we can run into problems:

- There are a lot of dimensions that we care about balancing: age, race, gender, education, income.

Motivation: Weighting

In theory, post-stratification is great, but in practice, we can run into problems:

- There are a lot of dimensions that we care about balancing: age, race, gender, education, income.
- The chances are high that we might get some cells with very few (or no) respondents.

Motivation: Weighting

In theory, post-stratification is great, but in practice, we can run into problems:

- There are a lot of dimensions that we care about balancing: age, race, gender, education, income.
- The chances are high that we might get some cells with very few (or no) respondents. Our estimates of these cells will not be very good.

Motivation: Weighting

In theory, post-stratification is great, but in practice, we can run into problems:

- There are a lot of dimensions that we care about balancing: age, race, gender, education, income.
- The chances are high that we might get some cells with very few (or no) respondents. Our estimates of these cells will not be very good.
  - This will, at the very least, increase the variance in our estimates, and

Motivation: Weighting

In theory, post-stratification is great, but in practice, we can run into problems:

- There are a lot of dimensions that we care about balancing: age, race, gender, education, income.
- The chances are high that we might get some cells with very few (or no) respondents. Our estimates of these cells will not be very good.
    - This will, at the very least, increase the variance in our estimates, and
    - in some cases, could make it impossible to correctly weight our estimates.

Motivation: Weighting

In theory, post-stratification is great, but in practice, we can run into problems:

- There are a lot of dimensions that we care about balancing: age, race, gender, education, income.
- The chances are high that we might get some cells with very few (or no) respondents. Our estimates of these cells will not be very good.
  - This will, at the very least, increase the variance in our estimates, and
  - in some cases, could make it impossible to correctly weight our estimates.
- Weighting, therefore, involves a bias-variance tradeoff. We correct for bias but have to accept some additional variance.

Motivation: LA Times Example

There was a great example of these problems during the 2016 Presidential Election.

Motivation: LA Times Example

There was a great example of these problems during the 2016 Presidential Election.

- The LA Times ran a panel study of vote-choice that was consistently showing a Trump lead even when all other polls shows a double-digit deficit.

Motivation: LA Times Example

There was a great example of these problems during the 2016 Presidential Election.

- The LA Times ran a panel study of vote-choice that was consistently showing a Trump lead even when all other polls shows a double-digit deficit. What was the problem?

Motivation: LA Times Example

There was a great example of these problems during the 2016 Presidential Election.

- The LA Times ran a panel study of vote-choice that was consistently showing a Trump lead even when all other polls shows a double-digit deficit. What was the problem?
- Their estimate for the vote choice of young black men in the Midwest was based on just one respondent,

Motivation: LA Times Example

There was a great example of these problems during the 2016 Presidential Election.

- The LA Times ran a panel study of vote-choice that was consistently showing a Trump lead even when all other polls shows a double-digit deficit. What was the problem?
- Their estimate for the vote choice of young black men in the Midwest was based on just one respondent, and he was a Trump supporter.

Motivation: LA Times Example

There was a great example of these problems during the 2016 Presidential Election.

- The LA Times ran a panel study of vote-choice that was consistently showing a Trump lead even when all other polls shows a double-digit deficit. What was the problem?
- Their estimate for the vote choice of young black men in the Midwest was based on just one respondent, and he was a Trump supporter.
- This respondent got weighted 30 times higher than average, and shifted their aggregate results by several percentage points.

Motivation: Weighting

There are some solutions to low cell-size problems,

Motivation: Weighting

There are some solutions to low cell-size problems,

- You could use larger group sizes to make sure that all cells have enough respondents,

Motivation: Weighting

There are some solutions to low cell-size problems,

- You could use larger group sizes to make sure that all cells have enough respondents, but this comes at the cost of
  - not being able to ask very specific questions, and

Motivation: Weighting

There are some solutions to low cell-size problems,

- You could use larger group sizes to make sure that all cells have enough respondents, but this comes at the cost of
  - not being able to ask very specific questions, and
  - only fixing selection bias in a very coarse way.

Motivation: Weighting

There are some solutions to low cell-size problems,

- You could use larger group sizes to make sure that all cells have enough respondents, but this comes at the cost of
  - not being able to ask very specific questions, and
  - only fixing selection bias in a very coarse way.
- You could trim you weights so that no single respondent is given more weight than some arbitrary cutoff,

Motivation: Weighting

There are some solutions to low cell-size problems,

- You could use larger group sizes to make sure that all cells have enough respondents, but this comes at the cost of
  - not being able to ask very specific questions, and
  - only fixing selection bias in a very coarse way.
- You could trim you weights so that no single respondent is given more weight than some arbitrary cutoff, but this is even worse

Motivation: Weighting

There are some solutions to low cell-size problems,

- You could use larger group sizes to make sure that all cells have enough respondents, but this comes at the cost of
  - not being able to ask very specific questions, and
  - only fixing selection bias in a very coarse way.
- You could trim you weights so that no single respondent is given more weight than some arbitrary cutoff, but this is even worse
  - you're fixing the wrong problem,

Motivation: Weighting

There are some solutions to low cell-size problems,

- You could use larger group sizes to make sure that all cells have enough respondents, but this comes at the cost of
  - not being able to ask very specific questions, and
  - only fixing selection bias in a very coarse way.
- You could trim you weights so that no single respondent is given more weight than some arbitrary cutoff, but this is even worse
  - you're fixing the wrong problem, there is nothing wrong with the weights,

Motivation: Weighting

There are some solutions to low cell-size problems,

- You could use larger group sizes to make sure that all cells have enough respondents, but this comes at the cost of
  - not being able to ask very specific questions, and
  - only fixing selection bias in a very coarse way.
- You could trim you weights so that no single respondent is given more weight than some arbitrary cutoff, but this is even worse
  - you're fixing the wrong problem, there is nothing wrong with the weights, the problem is with the cell estimate.

Motivation: MRP

- To summarize, post-stratification will have problems with unreliable sample stratum mean estimates when some strata have small sample sizes

Motivation: MRP

- To summarize, post-stratification will have problems with unreliable sample stratum mean estimates when some strata have small sample sizes
- What techniques are good for solving small sample size problems?

Motivation: MRP

- To summarize, post-stratification will have problems with unreliable sample stratum mean estimates when some strata have small sample sizes
- What techniques are good for solving small sample size problems? Bayesian statistics and hierarchical models!

Motivation: MRP

- To summarize, post-stratification will have problems with unreliable sample stratum mean estimates when some strata have small sample sizes
- What techniques are good for solving small sample size problems? Bayesian statistics and hierarchical models!
- Instead of altering the knowns (the population proportions in each stratum),

Motivation: MRP

- To summarize, post-stratification will have problems with unreliable sample stratum mean estimates when some strata have small sample sizes
- What techniques are good for solving small sample size problems? Bayesian statistics and hierarchical models!
- Instead of altering the knowns (the population proportions in each stratum), we use a hierarchical model to estimate the unknowns (stratum means).

Motivation: MRP

- To summarize, post-stratification will have problems with unreliable sample stratum mean estimates when some strata have small sample sizes
- What techniques are good for solving small sample size problems? Bayesian statistics and hierarchical models!
- Instead of altering the knowns (the population proportions in each stratum), we use a hierarchical model to estimate the unknowns (stratum means).
- Partial pooling of data across strata allows estimates in undersized strata to be improved by borrowing strength from similar strata.

Execution: Overview

1. Collect individual-level response data
2. Collect a poststratification dataset
3. Collect geographic-level predictors (optional)
4. Fit a hierarchical model for individual responses given demographics and geographic unit
5. Post-stratify

Execution: Specific

1. Collect individual-level response data.

Execution: Specific

1. Collect individual-level response data.It should contain
   a. a vector $\mathbf{y} = [y_1, \ldots, y_n]$ of individual-level responses
   b. a matrix $\mathbf{X}$ of individual-level covariates (i.e. demographics)
   c. a vector $\mathbf{G}$ indicating the individual's geographic location
2. Collect a poststratification dataset.

Execution: Specific

1. Collect individual-level response data.It should contain
    a. a vector $\mathbf{y} = [y_1, \ldots, y_n]$ of individual-level responses
    b. a matrix $\mathbf{X}$ of individual-level covariates (i.e. demographics)
    c. a vector **G** indicating the individual's geographic location
2. Collect a poststratification dataset.
    a. Ideally this comes from a massive individual-level survey with the same (or more granular) demographics than $\mathbf{X}$
    b. Use this to create state-level cross-tabulations for each of these demographic variables
3. Collect geographic-level predictors (optional
    a. These will help get better geographic-unit estimates.

Execution: Specific

4. Fit a hierarchical model for individual responses given demographics and geographic unit
   a. This gives a predicted response for each combination of the individual-level variables and geographic unit
   b. You can now generate a predicted response for females with college degrees between 18 and 29 in NY
5. Post-stratify

Execution: Specific

4. Fit a hierarchical model for individual responses given demographics and geographic unit
    a. This gives a predicted response for each combination of the individual-level variables and geographic unit
    b. You can now generate a predicted response for females with college degrees between 18 and 29 in NY
5. Post-stratify
    a. Now that we have predicted outcome for every cell, weight to get estimates for every geographic area θ_j.

$$\theta_j = \frac{\sum_c n_{jc}\theta_{jc}}{\sum_c n_{jc}}$$

where n_{jc} refers to the number of individuals in state j in cell c (e.g. number of males with college degrees

Reversible Jump Markov Chain Monte Carlo (RJMCMC)

Motivation: Changepoints

- RJMCMC is an extension of the MH algorithm described by Green (1995) which lets you draw from a posterior distribution that varies in dimensions over iterations.

Motivation: Changepoints

- RJMCMC is an extension of the MH algorithm described by Green (1995) which lets you draw from a posterior distribution that varies in dimensions over iterations.
- There are lots of applications but I will use it study time-series data,

Motivation: Changepoints

- RJMCMC is an extension of the MH algorithm described by Green (1995) which lets you draw from a posterior distribution that varies in dimensions over iterations.
- There are lots of applications but I will use it study time-series data, event data are assume to arise from a Poisson process with an unknown number of steps.

Motivation: Changepoints

- RJMCMC is an extension of the MH algorithm described by Green (1995) which lets you draw from a posterior distribution that varies in dimensions over iterations.
- There are lots of applications but I will use it study time-series data, event data are assume to arise from a Poisson process with an unknown number of steps.
- The objective of the algorithm is to determine 1) the number of steps,

Motivation: Changepoints

- RJMCMC is an extension of the MH algorithm described by Green (1995) which lets you draw from a posterior distribution that varies in dimensions over iterations.
- There are lots of applications but I will use it study time-series data, event data are assume to arise from a Poisson process with an unknown number of steps.
- The objective of the algorithm is to determine 1) the number of steps, 2) the location of these steps, and

Motivation: Changepoints

- RJMCMC is an extension of the MH algorithm described by Green (1995) which lets you draw from a posterior distribution that varies in dimensions over iterations.
- There are lots of applications but I will use it study time-series data, event data are assume to arise from a Poisson process with an unknown number of steps.
- The objective of the algorithm is to determine 1) the number of steps, 2) the location of these steps, and 3) the rate parameter of these steps.

Motivation: Changepoints

- RJMCMC is an extension of the MH algorithm described by Green (1995) which lets you draw from a posterior distribution that varies in dimensions over iterations.
- There are lots of applications but I will use it study time-series data, event data are assume to arise from a Poisson process with an unknown number of steps.
- The objective of the algorithm is to determine 1) the number of steps, 2) the location of these steps, and 3) the rate parameter of these steps.
- The algorithm accomplishes this objective through four different kinds of moves: height, position, birth, and death.

Derivation: Move types

There are four possible moves,

Derivation: Move types

There are four possible moves,

1. Height: propose a random change to the height of a selected step and use the MH algorithm to accept or reject this change.

Derivation: Move types

There are four possible moves,

1. Height: propose a random change to the height of a selected step and use the MH algorithm to accept or reject this change.
2. Position: propose a random change to the location of a randomly selected break, using MH to accept or reject.

Derivation: Move types

There are four possible moves,

1. Height: propose a random change to the height of a selected step and use the MH algorithm to accept or reject this change.
2. Position: propose a random change to the location of a randomly selected break, using MH to accept or reject.
3. Birth: Propose a new break within a randomly selected step.

Derivation: Move types

There are four possible moves,

1. Height: propose a random change to the height of a selected step and use the MH algorithm to accept or reject this change.
2. Position: propose a random change to the location of a randomly selected break, using MH to accept or reject.
3. Birth: Propose a new break within a randomly selected step.
4. Death: Remove a randomly selected step, combining previous heights.

Derivation: Move types

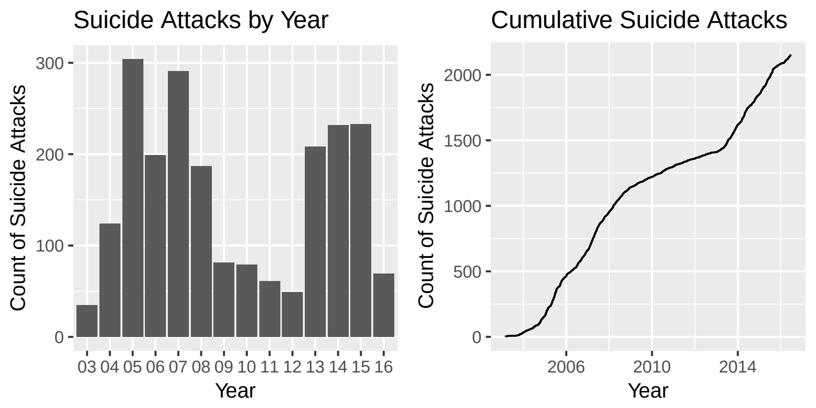
There are four possible moves,

1. Height: propose a random change to the height of a selected step and use the MH algorithm to accept or reject this change.
2. Position: propose a random change to the location of a randomly selected break, using MH to accept or reject.
3. Birth: Propose a new break within a randomly selected step.
4. Death: Remove a randomly selected step, combining previous heights.

I'll post a handout that gives the specifics of the log-likelihoods you calculate to decide whether to accept or reject these proposals.

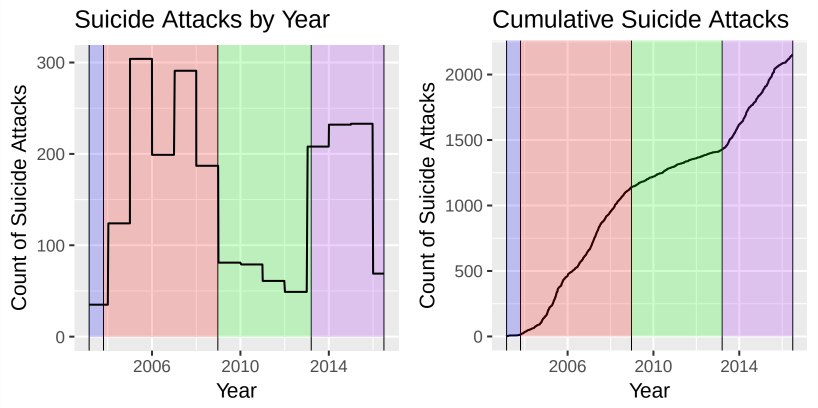
Application

I analyzed the number of suicide bombings in Iraq from the beginning of the war, March 2003, until June 2016, and identified change-points in the level of violence. The data come from CPOST and identify the date of several thousand attacks.



Application: Result

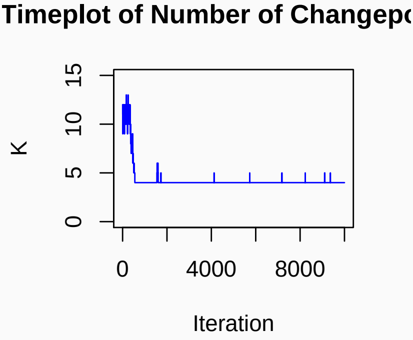
Applying the algorithm, I find the following breakpoints, which fit our intuitions.

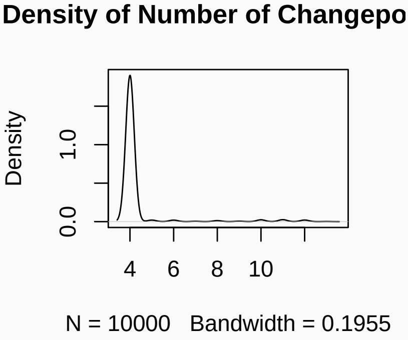


2004 is the beginning of violence, 2009 is the Surge and Anbar Awakening, 2013 is the emergence of ISIS.

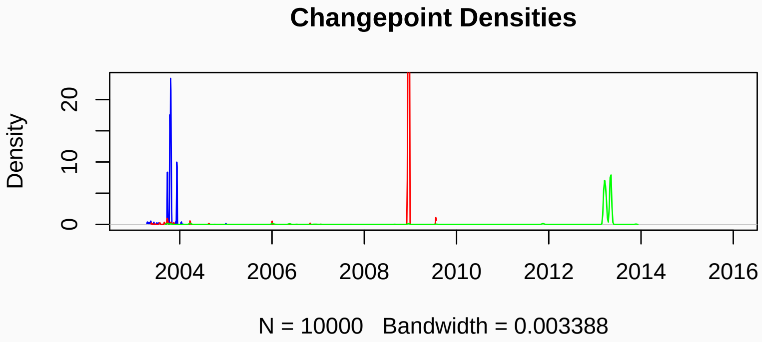
Application: Diagnostics

I start the algorithm with possible but fairly unrealistic initial values: a change point at every year (for twelve total) and a low rate parameter of 3 for each change point. It converges quickly to a much more reasonable 4 change points after only 1000 iterations.



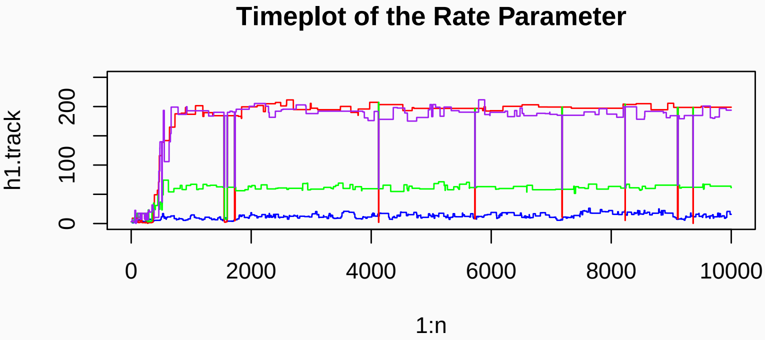


The confidence bounds around these changepoints are fairly narrow.



Application: Diagnostics

The rate parameters also converge to reasonable estimates after only a small number of iterations.



The heights sometimes change dramatically when new steps are introduced, or steps are deleted.