

UCLA

UCLA Electronic Theses and Dissertations

Title

Learning from Social Data: Non-exchangeable Priors in the Social Sciences

Permalink

<https://escholarship.org/uc/item/4t1024ns>

Author

Csapo, Marika Danielle

Publication Date

2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Learning from Social Data: Non-exchangeable Priors in the Social Sciences

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Statistics

by

Marika Danielle Csapo

2016

© Copyright by
Marika Danielle Csapo
2016

ABSTRACT OF THE THESIS

Learning from Social Data: Non-exchangeable Priors in the Social Sciences

by

Marika Danielle Csapo

Master of Science in Statistics

University of California, Los Angeles, 2016

Professor Qing Zhou, Chair

Bayesian analyses are often critiqued on the basis of dubious exchangeability claims regarding the data. Not only must observed data be exchangeable, but prior “data” must be as well, and the observed data must also be exchangeable with the prior data—an assumption not typically justified by the practitioner. Yet social scientists often utilize social data—observed human behaviors that rely on human judgment—to make inferences. Social priors shared by the researcher are, therefore, non-exchangeable with social data. One common defensive argument offered by Bayesian practitioners is that as long as there is some component of new information in the observed data, repeated observation-updating cycles will still eventually produce a highly informative posterior distribution. In frequentist statistics we have power analyses—a way of estimating how much data we need to get desirable properties from our estimator. Here I develop a model that parameterizes the degree of non-exchangeability between the observed data and the prior data and offers a standard way to calculate how many observations are needed to achieve a parameterized definition of an “informative” Bayes’ estimate in a single iteration of updating, or the number of updating iterations needed given a fixed observation size n at each iteration. I illustrate the phenomenon with a combination of real and model-synthesized data showing how New York police officers who make stops learn from social data—convictions generated by jury trials in the U.S. justice system.

The thesis of Marika Danielle Csapo is approved.

Rick Schoenberg

Chad Hazlett

Qing Zhou, Committee Chair

University of California, Los Angeles

2016

To Mako and Joe

TABLE OF CONTENTS

1	Introduction	1
2	A Non-exchangeable Model of Social Data	4
2.1	New Data Non-exchangeable with Social Prior	4
2.1.1	First Period Social Prior	4
2.1.2	Unobserved Data-Generating Process	4
2.1.3	Naive Likelihood	6
2.1.4	Posterior Belief about μ After Observing the Data	6
2.1.5	Quantifying the Bias in the Bayes' Estimate	7
3	“Power Analysis” Analog for Non-exchangeable Data	9
3.1	Single Iteration Inference	9
3.2	Simulation: Necessary “n” for a single iteration	12
4	Iterative Updating	14
4.1	Find number of iterations, t , given n , to get within ϵ of truth	15
4.1.1	Calculating conservative “t” conditional on “n”	17
4.2	True Bayesian updating: estimating “t” conditional on “n”	18
4.2.1	Calculating “t” conditional on “n” using Bayesian updating	20
4.3	Can a Larger n Speed Up Convergence?	20
5	Comparing Patterns of Social Learning	24
5.1	Data Application	24
5.2	Results	29

6 Conclusion	32
-------------------------------	-----------

LIST OF FIGURES

3.1	“Power Analysis” Analog	13
4.1	Comparing Analytical Predictions with Averages from Simulated Data . . .	17
4.2	Simulation Using True Bayesian Updating Rule	21
4.3	Testing the R Function	22
4.4	Iterative Bayesian Updating with Different Sample Sizes	23
5.1	Police Stop Rates	27
5.2	Comparing Evolution of Police “Priors” with Model-Based Counterfactuals ($\delta = 0.9$)	30
5.3	Comparing Evolution of Police “Priors” with Model-Based Counterfactuals ($\delta = 0.1$)	31

LIST OF TABLES

5.1	Stop Rates per 1,000 in Population	25
-----	--	----

ACKNOWLEDGMENTS

I owe many thanks to my committee chair, Qing Zhou, who sat with me so patiently on many occasions as I tried to articulate my ideas for this project—often less than coherently. His suggestions were indispensable. I also thank my other committee members, Chad Hazlett and Rick Schoenberg for their participation in this project and for very helpful comments. Thanks also go to Jeffrey Lewis from the Department of Political Science for teaching the class that inspired the project and for reading drafts and sharing his comments.

Finally, I thank my mentor Barbara Geddes for her unending guidance and support through the academic process. I also likely would not have finished the project if it were not for the emotional support provided by my siblings, Mako and Joe, and by Ronald M. Solórzano and Shauna McManus.

CHAPTER 1

Introduction

Observed human behaviors that rely on human judgment, including consumption and investment decisions, cooperative behavior, jury trial results, voting behavior, psychological diagnoses, and behavioral responses to risk, among others, are often the source of social data. Kahneman explains that the inputs to human behavior combine essentially autonomic prior beliefs, which may include psychological biases, and which have the convenience of leading to speedy decisions since they do not require much cognitive processing time, with information processing and logic ([Kahneman, 2011](#)). Social priors—widely held beliefs that may be subject to shared narrative biases ([Bar-Tal, 2000](#))—are, therefore, non-exchangeable with social data by definition. The social priors factor into the data-generating process itself. In the standard Bayesian model, however, the researcher assumes the data are exchangeable.

We can think of individuals’ prior beliefs as being drawn from a social distribution of prior beliefs, with some mean, and many small, additive sources of individual variation. It is important to note that the researcher is, too, a sociological product of the same society as the data-generating members of society. That is, the same source of prior information informs both the data and the researcher’s priors.

We commonly note that our observations ought to be exchangeable with each other ([Diaconis and Freedman, 1980](#); [De Finetti and Machí, 2002](#); [Kerns and Székely, 2006](#)), but there is no mathematical reason that our prior observations should be exempt from this standard. Prior observations are treated mathematically as data (whether or not that data is quantitative in its raw form), just as are new observations ([Jackman, 2009](#); [Johnson et al., 2010](#); [Gelman et al., 2014](#); [Gill, 2014](#)). From a mathematical perspective, it is not clear which distribution is the prior and which is the “data.” Thus, permuting the order

in which we observe the prior observations and the new observations should not effect our results. This implies the new data should not be influenced by our prior beliefs if we are to update accurately (or, at least, the updating process should take into account any partial non-exchangeability (Bernardo, 1996)). However, we rarely attempt to explicitly justify the implicit assumption that our prior and new observations are exchangeable with each other. The problem of ignoring non-exchangeability between prior and new data is especially worrisome when prior data are both biased and non-exchangeable with the new data.

Typically Bayesians assuage fears over biased priors and non-exchangeability by noting that if we allow our priors to update with each iteration of newly observed data and if the newly observed data contain at least some component of independent information about the world, eventually (after repeated iterations of updating) the posterior distribution will still become highly informative. However, two practical problems exist with this logic. First, we often must make an inference after a single round of updating. For practical purposes, we still need to know, given the non-exchangeability and potential for bias in our priors, what size n is needed in this single iteration to produce an informative estimate? Furthermore, we may wish to know, given a certain number of new observations on each iteration, how many iterations of new data and updating would be required before the estimate becomes highly informative? Both have clearer practical implications than the claim that if we gathered information infinitely, in the limit we would definitely have a good answer.

This thesis proceeds as follows. In Chapter 2, I elaborate a model in which a researcher attempts to learn from social data that has been influenced by social prior beliefs. I provide a method for calculating which portion of the prior bias remains in the posterior Bayes' estimate. In Chapter 3, I provide an analytical solution to calculate the n that is necessary to get a posterior Bayes' estimate that falls within an arbitrary ϵ from the truth. I show that a bound exists in ϵ , after which no amount of data is sufficient to reduce bias below that range in a single updating iteration. I provide a formula to calculate this bound. Chapter 4 considers iterative observation-updating cycles and asks, given a fixed observation size n at each round, how many updating rounds are required to produce a posterior Bayes' estimate that falls within ϵ of the truth. I provide two alternative ways to calculate this quantity. The

first is an analytical solution, using a conservative updating rule (where the researcher does not increase the weight afforded to the social prior after the first or subsequent updating iterations, even though more and more data support this distribution). The second is a computational solution in the form of R code that simulates “true” Bayesian updating. I compare the rate of bias reduction across the two forms. Chapter 5 provides a practical social science application. It takes New York City police officers as the “researcher,” their stop rates of New York City residents as the social prior regarding guilt, and model-simulated conviction data as the non-exchangeable social data. I compare the observed evolution of the police officers’ priors with model-simulated counterfactuals using the social data to show which updating rule police officers are likely using. Chapter 6 concludes.

CHAPTER 2

A Non-exchangeable Model of Social Data

In this chapter, I provide a model of social data that has been influenced by social priors. For those that prefer a concrete example, consider jury convictions which are generated by jurors who consider some combination of real trial evidence and their own prior beliefs about the defendant's guilt in making a conviction decision. A researcher that uses observed conviction rates as data is using social data. The data is a function of both objective information and (possibly biased) prior beliefs.

2.1 New Data Non-exchangeable with Social Prior

2.1.1 First Period Social Prior

The central limit theorem implies that if the distribution of social beliefs is made up of a group mean and the sums of many small individual factors the distribution may be adequately characterized by $\mathcal{N}(\alpha, \tau)$. I use this to represent the social prior distribution.

2.1.2 Unobserved Data-Generating Process

Because the data is “social data,” which is data generated from human behavior reflecting human judgment, I model the socially-influenced data-generating process as a linear combination of an individual's prior beliefs, $\alpha + \nu_i$ (drawn from the social prior distribution), and some independent information, μ plus noise, z_i . That is,

$$x_i = \delta(\alpha + \nu_i) + (1 - \delta)(\mu + z_i),$$

where

$$z_i \sim \mathcal{N}(0, \sigma),$$

and where,

$$\nu_i \sim \mathcal{N}(0, \tau),$$

and where $\delta \in (0, 1)$ and both ν and Z are independently and identically distributed. Note that by design there is no information about the world contained in the first period prior beliefs, α . The researcher is interested in uncovering μ , the independent systematic contribution of the world, but the observed social data are influenced by both this independent component and the social prior beliefs (shared by the researcher and the data-generating social actors). Based on the data-generating process above, one can think of $X = (x_1, \dots, x_n)$, given a certain value of the parameter, μ , as the weighted sum of two independent normal distributions, with

$$X|\mu \stackrel{iid}{\sim} \mathcal{N}\left(\delta\alpha + (1-\delta)\mu, \sqrt{\delta^2\tau^2 + (1-\delta)^2\sigma^2}\right). \quad (2.1)$$

If the researcher did not share in the social prior (and had an idea of the true data-generating process) she could model the likelihood function in this way to account for the fact that the observed data, X , are in part informed by the social prior. This would allow her to back out the unique and systematic contribution of the world (μ) to the observed data. That is, she could manufacture a model in which the exchangeability assumption would hold. However, here I model a situation in which the researcher's beliefs are not independent of society's (nor is my researcher sufficiently self-aware to recognize her biases).

2.1.3 Naive Likelihood

When one overlooks the question of exchangeability between the new data and prior beliefs and does not use the backout method, this allows the bias in the prior beliefs to carry over into the posterior via both the prior and the data (the degree to which it carries over through the data in one period depends on δ). Unaware of the exchangeability violation, the researcher constructs the naive likelihood of the data as a normal distribution with mean, μ —assuming the data is made up of a combination of useful information and unbiased noise. For mathematical ease, I assume that the standard deviation is known to be the quantity given by $\sqrt{\delta^2\tau^2 + (1 - \delta)^2\sigma^2}$. Therefore, the likelihood function is proportional to

$$P(X|\mu) = \prod_{i=1}^n \phi(x_i; \mu, \sqrt{\delta^2\tau^2 + (1 - \delta)^2\sigma^2}),$$

where ϕ is the normal probability distribution of x_i , with mean, μ , and standard deviation, $\sqrt{\delta^2\tau^2 + (1 - \delta)^2\sigma^2}$.

2.1.4 Posterior Belief about μ After Observing the Data

Using this naive likelihood function, it is possible to construct the posterior exploiting normal-normal conjugacy. The posterior belief regarding the updated probability distribution of μ is proportional to the joint likelihood of all the observed data, conditional on μ , multiplied by the social prior belief regarding the probability of μ . Recall that the social prior belief is distributed $\mathcal{N}(\alpha, \tau)$. Therefore,

$$P(\mu|X) \propto P(X|\mu)P(\mu)$$

with,

$$\mu|X \sim \mathcal{N}(\alpha_1, \tau_1),$$

where

$$\alpha_1 = \frac{\left(\frac{\alpha}{\tau^2} + \frac{n\bar{X}}{\delta^2\tau^2 + (1-\delta)^2\sigma^2}\right)}{\left(\frac{1}{\tau^2} + \frac{n}{\delta^2\tau^2 + (1-\delta)^2\sigma^2}\right)}, \quad (2.2)$$

$$\tau_1 = \frac{1}{\sqrt{\frac{1}{\tau^2} + \frac{n}{\delta^2\tau^2 + (1-\delta)^2\sigma^2}}}. \quad (2.3)$$

Note that $E[\bar{X}] = \delta\alpha + (1-\delta)\mu$ and that $\bar{X}$ is also the (naive) frequentist maximum likelihood estimate.

2.1.5 Quantifying the Bias in the Bayes' Estimate

Both the ML estimate of the parameter value and the Bayes' estimate will be biased due to the fact that the data are influenced by human psychological biases but the researcher is not modeling them as such. It is possible to quantify the bias produced by these two estimates in one period as the absolute distance between the expectation of the Bayes' estimator and μ , the part of the data that is neither informed by noise nor by systematic psychological bias.

For comparison, the size of the bias in the MLE may be expressed as, $|E[\bar{X}] - \mu|$. The expectation of the frequentist maximum likelihood estimator is $E[\bar{X}] = \delta\alpha + (1-\delta)\mu$. Expanding and re-arranging terms, this gives

$$\begin{aligned}
E[\bar{X}] - \mu &= [\delta\alpha + (1 - \delta)\mu] - \mu \\
&= \delta\alpha + \mu - \delta\mu - \mu \\
&= \delta(\alpha - \mu).
\end{aligned} \tag{2.4}$$

Similarly, it is possible to characterize the bias in the naive Bayes' estimate as $\left| \alpha k + E[\bar{X}](1 - k) - \mu \right|$, where $k = \left(\frac{1}{\tau^2} \right) / \left(\frac{1}{\tau^2} + \frac{n}{\delta^2 \tau^2 + (1 - \delta)^2 \sigma^2} \right)$. Expanding and re-arranging terms gives

$$\begin{aligned}
\alpha k + E[\bar{X}](1 - k) - \mu &= \alpha k + [\delta\alpha + (1 - \delta)\mu](1 - k) - \mu \\
&= \alpha k + \delta\alpha - \delta\alpha k + (1 - \delta)\mu - (1 - \delta)\mu k - \mu \\
&= \alpha(\delta + k - \delta k) + \mu - \mu(\delta + k - \delta k) - \mu \\
&= [\delta + k(1 - \delta)](\alpha - \mu).
\end{aligned} \tag{2.5}$$

Note that since δ lies in the interval $(0, 1)$ by construction and k is a proportion which must also lie in the interval $(0, 1)$ one can conclude that $[\delta + k(1 - \delta)] > \delta$. That is, the bias remaining in the posterior Bayes' estimate (see Equation 2.5) is larger than the bias remaining in the MLE estimator (see Equation 2.4).

CHAPTER 3

“Power Analysis” Analog for Non-exchangeable Data

3.1 Single Iteration Inference

Social research often must lead to an inference after a single round of updating unless the inquiry is purely theoretical or the data simulated. If biased social priors partly inform the data, but the researcher uses the naive likelihood function, is there an n sufficiently large to come up with a reasonably good Bayes' estimate in a single iteration of updating? If so, what size n is needed in this single iteration to produce an informative estimate? As shown in the previous section, it is possible to characterize the bias in our Bayes' estimate as $[\delta + k(1 - \delta)](\alpha - \mu)$. In a single iteration, what size n is needed to make the absolute bias drop below an arbitrary acceptable cutoff point, ϵ ?

Referring back to Equation 2.5, the bias remaining in the posterior Bayes' estimate after one iteration of updating, it would be ideal to have an equation to calculate the size n that brings that bias below a certain threshold, thus satisfying the following condition,

$$[\delta + k(1 - \delta)] |\alpha - \mu| \leq \epsilon.$$

Substituting the definition of k from the previous chapter and solving for n gives the following:

$$\begin{aligned}
\left[\frac{\delta \left(\frac{1}{\tau^2} + \frac{n}{\delta^2 \tau^2 + (1-\delta)^2 \sigma^2} \right) + \frac{1-\delta}{\tau^2}}{\frac{1}{\tau^2} + \frac{n}{\delta^2 \tau^2 + (1-\delta)^2 \sigma^2}} \right] |\alpha - \mu| &\leq \epsilon \\
\frac{\frac{\delta}{\tau^2} + \frac{\delta n}{\delta^2 \tau^2 + (1-\delta)^2 \sigma^2} + \frac{1-\delta}{\tau^2}}{\frac{1}{\tau^2} + \frac{n}{\delta^2 \tau^2 + (1-\delta)^2 \sigma^2}} &\leq \frac{\epsilon}{|\alpha - \mu|} \\
\frac{\delta^2 \tau^2 + (1-\delta)^2 \sigma^2 + \tau^2 \delta n}{\tau^2 (\delta^2 \tau^2 + (1-\delta)^2 \sigma^2)} &\leq \frac{\epsilon}{|\alpha - \mu|} \left(\frac{\delta^2 \tau^2 + (1-\delta)^2 \sigma^2 + \tau^2 n}{\tau^2 (\delta^2 \tau^2 + (1-\delta)^2 \sigma^2)} \right) \\
\tau^2 \delta n - \left(\frac{\epsilon}{|\alpha - \mu|} \right) \tau^2 n &\leq \left(\frac{\epsilon}{|\alpha - \mu|} \right) (\delta^2 \tau^2 + (1-\delta)^2 \sigma^2) - (\delta^2 \tau^2 + (1-\delta)^2 \sigma^2) \\
\tau^2 \delta n |\alpha - \mu| - \epsilon \tau^2 n &\leq \epsilon [\delta^2 \tau^2 + (1-\delta)^2 \sigma^2] - |\alpha - \mu| [\delta^2 \tau^2 + (1-\delta)^2 \sigma^2] \\
n \tau^2 (\delta |\alpha - \mu| - \epsilon) &\leq [\delta^2 \tau^2 + (1-\delta)^2 \sigma^2] (\epsilon - |\alpha - \mu|).
\end{aligned}$$

It would now be ideal to isolate n by dividing both sides of the inequality by $\tau^2 (\delta |\alpha - \mu| - \epsilon)$. However, we do not know whether the quantity $\delta |\alpha - \mu| - \epsilon$ is greater than or less than 0 and this will determine the direction of the inequality. Therefore, I separate the results into the two possible scenarios.

First, suppose $\delta |\alpha - \mu| < \epsilon$. Then,

$$n \geq \frac{[\delta^2 \tau^2 + (1-\delta)^2 \sigma^2] (\epsilon - |\alpha - \mu|)}{\tau^2 (\delta |\alpha - \mu| - \epsilon)}.$$

It is clear from the condition that $\delta |\alpha - \mu| - \epsilon < 0$, so the denominator must be negative. This scenario further subdivides into two cases, depending on whether $\epsilon - |\alpha - \mu|$ is less than zero.

1. If $\epsilon - |\alpha - \mu| > 0$, then the entire expression to the right of the inequality is negative (and is zero if $\epsilon - |\alpha - \mu| = 0$). The conclusion, then, is that if $\epsilon \geq |\alpha - \mu|$, n needs only to be larger or equal to some negative number in order to get an estimate with bias within the tolerable range. This makes sense, as long as one recognizes that n cannot

be smaller than zero, so for this range, $n = 0$. No data is needed to update the social prior because the prior Bayes' estimate (even without updating using data) is α , and the bias in α by definition falls within the tolerable range.

2. In the more interesting scenario, $\epsilon - |\alpha - \mu| < 0$. In this case, $\epsilon - |\alpha - \mu| < 0$. Then, both the numerator and the denominator of the expression to the right of the inequality are negative, so the whole expression is a positive number and n must be at least that great to get the bias in the (posterior) Bayes' estimate to be within the tolerable range after one iteration.

Now, suppose $\delta |\alpha - \mu| \geq \epsilon$. This results in the following inequality for n ,

$$n \leq \frac{[\delta^2 \tau^2 + (1 - \delta)^2 \sigma^2](\epsilon - |\alpha - \mu|)}{\tau^2(\delta |\alpha - \mu| - \epsilon)}.$$

Since $\delta |\alpha - \mu| \geq \epsilon$, the denominator is positive. Note that since $\epsilon - \delta |\alpha - \mu| \leq 0$ and δ is in the interval $(0, 1)$, it is possible to conclude that $\epsilon - |\alpha - \mu| \leq 0$, as well. Therefore, the numerator is negative and the whole expression to the right of the inequality evaluates to a negative number. The conclusion is that n must be less than or equal to some negative number, which is nonsensical. This results because, in this range, there is no size n that is sufficient to get the (posterior) Bayes' estimate within the tolerable range of bias.

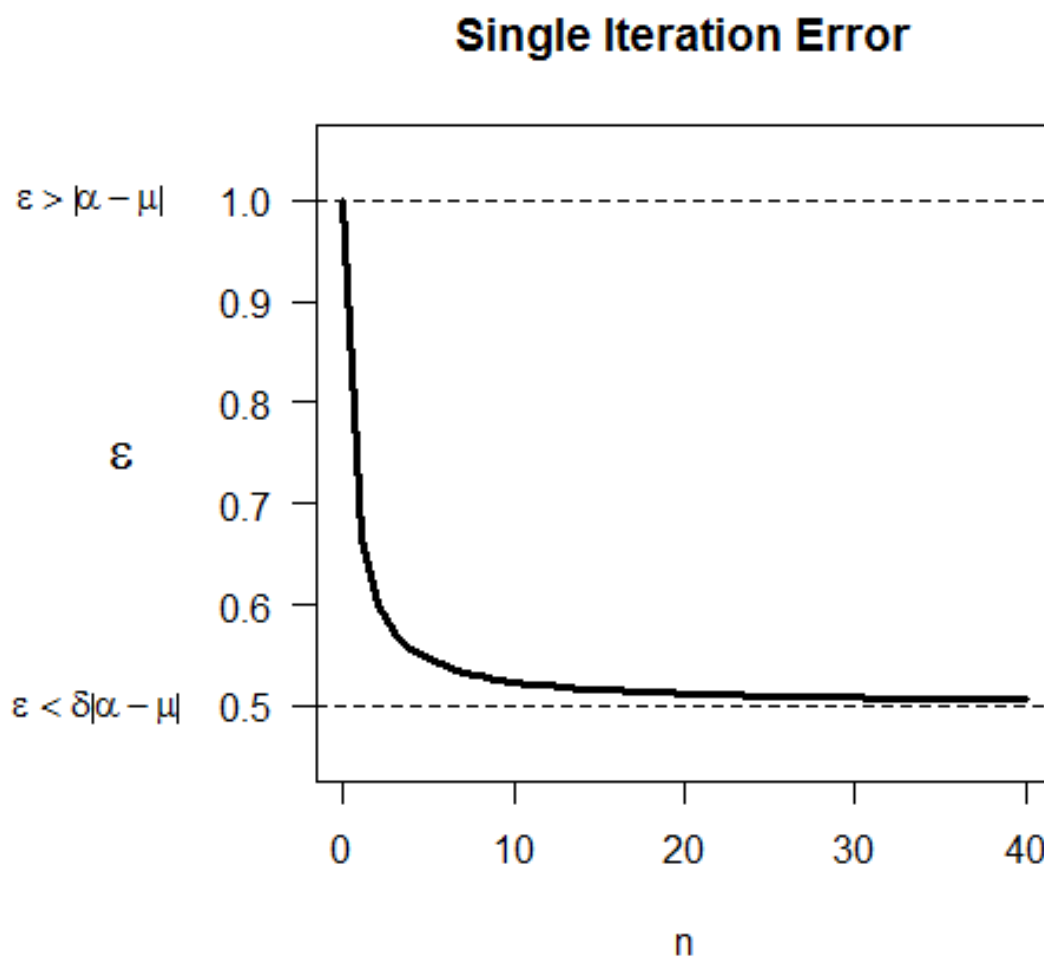
Therefore, I have identified three plausible ranges for ϵ . When $\delta |\alpha - \mu| \geq \epsilon$, no size n is sufficient to produce a Bayes' estimate that falls within the tolerable range in a single iteration, given that the researcher has overlooked the exchangeability violation. While it is still true that in the limit, infinite iterations of data observation and updating would deliver a good estimate, in this case it is simply not possible to collect enough data to overwhelm the nonexchangeability problem in a single iteration. When $\epsilon \geq |\alpha - \mu|$, any size n is sufficient. And when $\delta |\alpha - \mu| < \epsilon < |\alpha - \mu|$, it is possible to calculate the n that will deliver a posterior Bayes' estimate within the tolerable range of bias, using the inequality, $n \geq \frac{[\delta^2 \tau^2 + (1 - \delta)^2 \sigma^2](\epsilon - |\alpha - \mu|)}{\tau^2(\delta |\alpha - \mu| - \epsilon)}$. Below, I use the parameters defined above for the posterior distribution to illustrate these three ranges for a simulated case. See Appendix A

for the associated R code.

3.2 Simulation: Necessary “n” for a single iteration

The plot below shows the three ranges discussed in the text above. I’ve scaled the y-axis to be on a $[0, 1]$ interval expressing the proportion of the original prior bias remaining in the posterior Bayes’ estimate. For one range, the tolerance rate, ϵ , is large enough that no updating iterations are necessary to get a Bayes’ estimate within the acceptable range. For some range, the formula provided for “n” gives the minimum sample size necessary to get within the tolerable range. Finally, for some range of ϵ (in the plot, any $\epsilon < 0.5$), no sample size is sufficiently large to get within the tolerable range.

Figure 3.1: “Power Analysis” Analog



CHAPTER 4

Iterative Updating

Suppose, now, that the researcher has an opportunity to collect another round of social data after estimating and disseminating publicly the results from the first round of updating. I return to the notation introduced in Chapter 2 when thinking about the second, or any subsequent, iteration by making one change and indexing α and α_1 by their iteration number, $\alpha^{\{t\}}$ and $\alpha_1^{\{t\}}$, where t denotes the time period. Since the researcher and social actors are updating the social prior beliefs with each iteration (an idealistic assumption with respect to social learning), in $t = 2$ we will have $\alpha^{\{2\}} = \alpha_1^{\{1\}}$. That is, the posterior mean from $t = 1$ becomes the new prior mean for $t = 2$, and so on.

For now I assume that the researcher is an imperfect Bayesian. She updates her prior beliefs with each round of data observed, but she does not increase the data-weight on her prior belief with each updating round, despite the mounting evidence favoring the prior. In subsequent chapters I will show that this is a conservative way of calculating the posterior (and is similar to the naive maximum likelihood estimation strategy). Later on, I contrast this conservative approach with the true Bayesian approach, which allows the data-weight allocated to the prior to increase with each updating iteration. The conservative approach is mathematically simpler because the parameterization of the posterior distribution defined through Equations 2.2 and 2.3 also holds for the posterior in periods $t > 1$ (though the quantity, α , evolves with each iteration).

Note that in Equation 2.2 α_1 is the inverse-variance-weighted mean of α and the data. This implies that the size of the posterior bias, $|\alpha_1^{\{t\}} - \mu|$ is smaller than the size of the prior bias, $|\alpha^{\{t\}} - \mu|$, as one would expect with Bayesian updating since there is some independent informational content in the data. In fact, it is possible to get a sense for the rate at which

the bias decays over one time period (for a given n and δ). The proportion of the prior bias that remains in the (conservative) posterior is given by

$$\frac{\alpha_1^{\{t\}} - \mu}{\alpha^{\{t\}} - \mu}.$$

Thus, the rate of decay in the bias over one period is

$$1 - \frac{\alpha_1^{\{t\}} - \mu}{\alpha^{\{t\}} - \mu}.$$

From Chapter 2, we know that $\alpha_1^{\{t\}} - \mu = [\delta + k(1 - \delta)] (\alpha^{\{t\}} - \mu)$. Therefore, the bias decay rate for one iteration is

$$1 - [\delta + k(1 - \delta)].$$

This bias decay rate is a function (either implicitly, explicitly or both) of δ , n , τ and σ .

4.1 Find number of iterations, t , given n , to get within ϵ of truth

Using the general bias equation found above, it is possible to say that the (absolute) size of the bias after a single iteration of updating occurs is

$$[\delta + k(1 - \delta)] |\alpha^{\{1\}} - \mu|.$$

Using the “conservative” definition of Bayesian updating described above (where the researcher’s prior remains diffuse in each iteration, producing a posterior much like the ML estimator), any additional iteration, t , further reduces the bias in the previous iteration’s posterior mean by the following bias discount factor

$$[\delta + k(1 - \delta)].$$

This “conservative” version does not update prior weight. It always assumes there is one prior observation, which is drawn from an updated distribution. (Contrast this with the later version where one iteration’s posterior becomes the next iteration’s prior, so the researcher becomes increasingly sure about her biased prior, but continues to learn from new social data.) Conveniently, in the “conservative” version presented here, the bias discount factor is not a function of α and, therefore, is fixed across iterations. This makes it mathematically easy to compound over time. The remaining bias in the Bayes’ estimate at t iterations is, thus, given by the following equation:

$$[\delta + k(1 - \delta)]^t \left| \alpha^{\{1\}} - \mu \right|.$$

It is now possible to use this equation to calculate the number of iterations, given a fixed sample size, n , across iterations, that are necessary to reduce the bias in the Bayes’ estimate below some arbitrary quantity, ϵ . This occurs when the following condition is satisfied:

$$[\delta + k(1 - \delta)]^t \left| \alpha^{\{1\}} - \mu \right| \leq \epsilon.$$

Since it is true by construction that $0 < [\delta + k(1 - \delta)] < 1$, the following inequality exists for t :¹

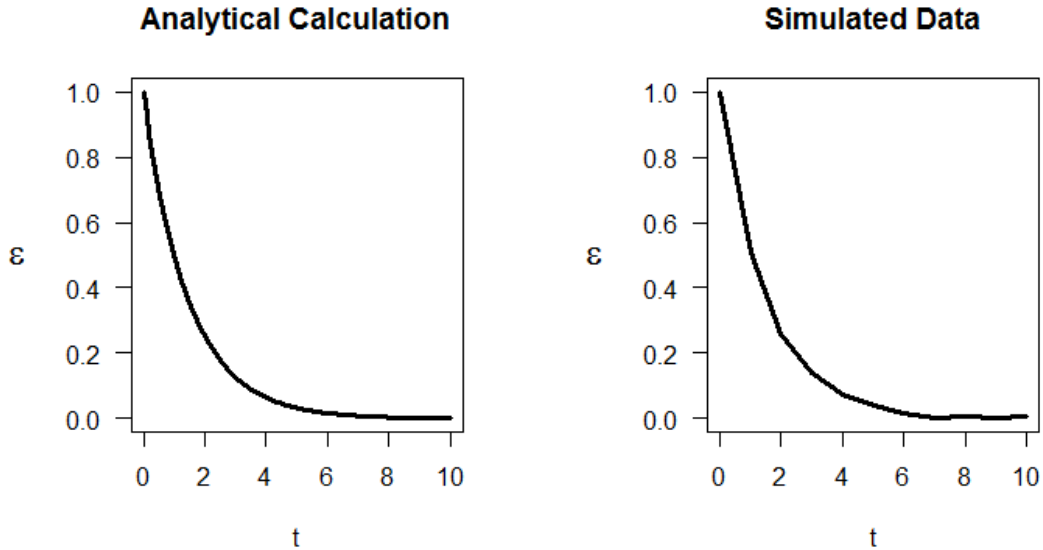
$$t \geq \frac{\log \epsilon - \log \left| \alpha^{\{1\}} - \mu \right|}{\log [\delta + k(1 - \delta)]}.$$

This implies that it is always possible to find some $t < \infty$ sufficient to produce an informative posterior, as would suggest conventional Bayesian intuition. When $\epsilon > \left| \alpha^{\{1\}} - \mu \right|$, then the equation returns some $t \leq 0$, which implies no updating is necessary to obtain a point estimate within the tolerable range. Below I compare the bias (ϵ) remaining after t iterations based on the analytical solution for t above with calculations based on data simulated in R. See Appendix B for the associated R code.

¹Note that the direction of the inequality switches because the log of a positive fraction is a negative number.

4.1.1 Calculating conservative “t” conditional on “n”

Figure 4.1: Comparing Analytical Predictions with Averages from Simulated Data



When the researcher does not update the weight placed on the prior with each iteration, the bias decay rate is similar to that of an iterative MLE process (because the researcher’s prior is always diffuse in each iteration), with social updating. The plot above shows that in both the analytical solution and the average over simulated data, the bias decays quickly in the conservative version—within 10 iterations the bias is essentially gone. Below I contrast this with the “true” Bayesian updating rule, where the researcher gets increasingly certain of the updated prior with each iteration. The analytical solution is more difficult to compound over time, but I provide R code that uses simulation to calculate the necessary number of iterations to get within ϵ of the truth when the researcher uses true Bayesian updating. The result below shows it takes far more than 10 iterations for the bias to decay to the same degree as in the conservative version.

4.2 True Bayesian updating: estimating “t” conditional on “n”

In the method above, the researcher updates the prior distribution with each iteration, but does not update the weight placed on the prior. Putting less weight on the prior makes it possible to learn faster from the data (in particular, from the part of the data that contains true information about the world). A strict interpretation of Bayesian updating, however, implies that as data accumulates across updating iterations, the researcher should begin each iteration, t , with increased weight on her prior beliefs, which are now supported by all the previous iterations’ rounds of data. If the researcher updates according to the strict interpretation, she will learn more slowly from the data. Take, for example, the illustration below, which defines $\alpha_1^{\{2\}}$ and $\tau_1^{\{2\}}$ (the posterior parameters from the second round of updating) in terms of the original α and τ .

Previously, I defined α_1 and τ_1 in the following generic sense:

$$\alpha_1 = \frac{\left(\frac{\alpha}{\tau^2} + \frac{n\bar{X}}{\delta^2\tau^2 + (1-\delta)^2\sigma^2} \right)}{\left(\frac{1}{\tau^2} + \frac{n}{\delta^2\tau^2 + (1-\delta)^2\sigma^2} \right)}$$

$$\tau_1 = \frac{1}{\sqrt{\frac{1}{\tau^2} + \frac{n}{\delta^2\tau^2 + (1-\delta)^2\sigma^2}}}.$$

This implies that

$$\alpha_1^{\{2\}} = \frac{\left(\frac{\alpha_1^{\{1\}}}{\tau_1^{2\{1\}}} + \frac{n\bar{X}^{\{2\}}}{\delta^2\tau_1^{2\{1\}} + (1-\delta)^2\sigma^2} \right)}{\left(\frac{1}{\tau_1^{2\{1\}}} + \frac{n}{\delta^2\tau_1^{2\{1\}} + (1-\delta)^2\sigma^2} \right)}$$

and

$$\tau_1^{\{2\}} = \frac{1}{\sqrt{\frac{1}{\tau_1^{2\{1\}}} + \frac{n}{\delta^2 \tau_1^{2\{1\}} + (1-\delta)^2 \sigma^2}}}.$$

It is then possible to substitute the definition of $\alpha_1^{\{1\}}$ and $\tau_1^{\{1\}}$ into these equations and simplify to express $\alpha_1^{\{2\}}$ and $\tau_1^{\{2\}}$ in terms of the original $\alpha^{\{1\}}$ and $\tau^{\{1\}}$. For simplicity, define $\alpha^{\{1\}} = \alpha$ and $\tau^{\{1\}} = \tau$. Then,

$$\alpha_1^{\{2\}} = \frac{\frac{\alpha}{\tau^2} + \frac{n\bar{X}^{\{1\}}}{\delta^2 \tau^2 + (1-\delta)^2 \sigma^2} + \frac{n\bar{X}^{\{2\}} \left(\frac{1}{\tau^2} + \frac{n}{\delta^2 \tau^2 + (1-\delta)^2 \sigma^2} \right)}{\delta^2 + (1-\delta)^2 \sigma^2}}{\frac{1}{\tau^2} + \frac{n}{\delta^2 \tau^2 + (1-\delta)^2 \sigma^2} + \frac{n \left(\frac{1}{\tau^2} + \frac{n}{\delta^2 \tau^2 + (1-\delta)^2 \sigma^2} \right)}{\delta^2 + (1-\delta)^2 \sigma^2}}.$$

Note there are three (weighted) terms in the numerator. The first term places weight on the prior belief from $t = 1$. The second term places weight on the social data from the first iteration, $X^{\{1\}}$. The third term places weight on the new social data observed in the second period, $X^{\{2\}}$. It is easy to see this by defining the weights as follows,

$$\begin{aligned} A &= \frac{1}{\tau^2} \\ B &= \frac{n}{\delta^2 \tau^2 + (1-\delta)^2 \sigma^2} \\ C &= \frac{n \left(\frac{1}{\tau^2} + \frac{n}{\delta^2 \tau^2 + (1-\delta)^2 \sigma^2} \right)}{\delta^2 + (1-\delta)^2 \sigma^2} \end{aligned}$$

and so,

$$\frac{\alpha A + \bar{X}^{\{1\}} B + \bar{X}^{\{2\}} C}{A + B + C}.$$

It can be shown that as long as $\tau^2 > 1$, then the weight put on the new data, $\bar{X}^{\{1\}}$ at

$t = 1$, $\frac{B}{A+B}$, is greater than the weight put on the new data $\bar{X}^{\{2\}}$ at $t = 2$, $\frac{C}{A+B+C}$ (see Appendix C for proof). That is, with each iteration, the contribution of the new data to the posterior decreases as the researcher becomes more certain of her non-exchangeable prior with each iteration. This suggests that the solution for t provided above is conservative, holding only when the researcher places a weight of one data point on prior beliefs at each iteration. When the researcher updates using a strict Bayesian updating rule, taking prior observations as support for the prior belief, it will take more than t iterations, as defined above, to reach a posterior estimate within the tolerable range. I further explore this using simulations, below, showing the number of iterations necessary to get within ϵ of the truth, using the true Bayesian updating rule. See Appendix D for associated R code.

4.2.1 Calculating “t” conditional on “n” using Bayesian updating

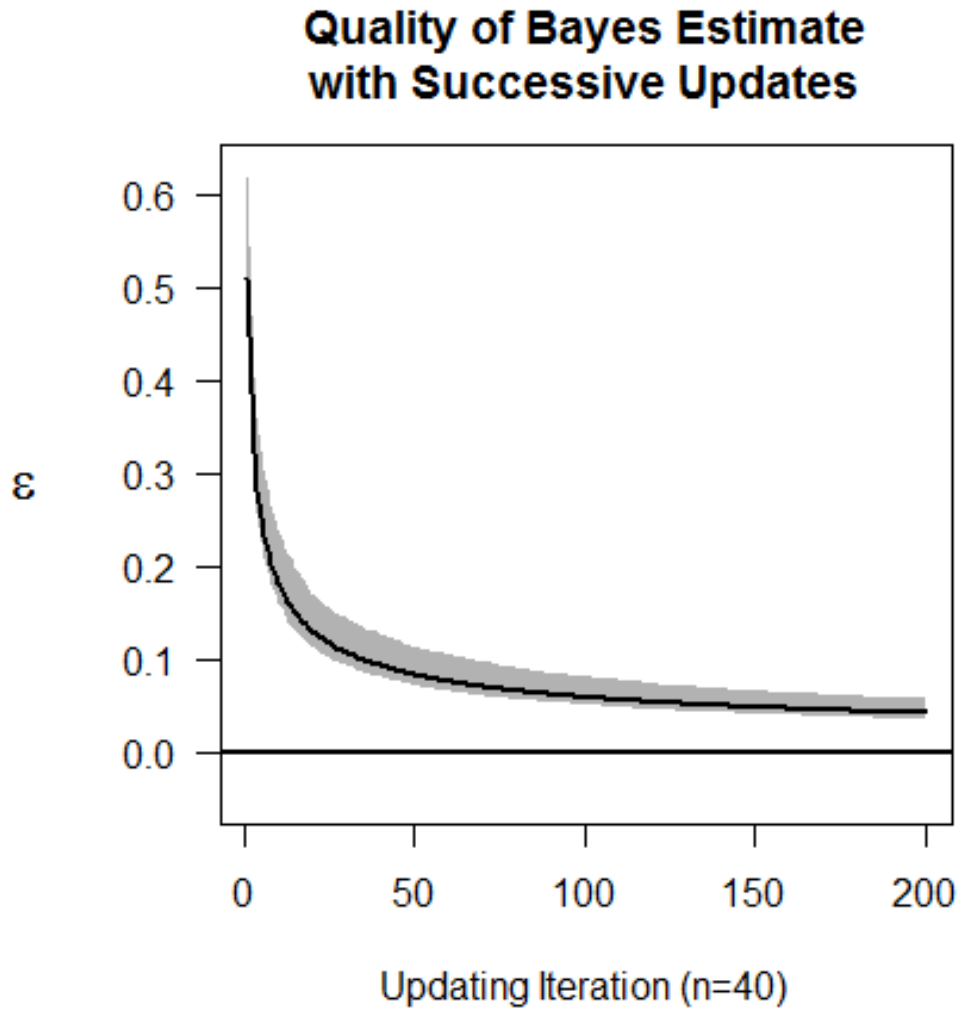
Figure 4.2 shows that even after 200 iterations, the updating process has not yet led to as good of an estimate as that produced in less than 10 iterations by the “conservative” version (with diffuse priors in each round approximating MLE learning from social data). In Appendix E, I provide an R function that can be used more generally to calculate the necessary number of iterations, t , given n , α , τ and δ , using the true Bayesian updating procedure.

Figure 4.3 uses this code to calculate the number of iterations necessary to get a Bayes’ estimate within $\epsilon = 0.0625$ of the truth using true Bayesian updating. If the researcher were to use the conservative, analytical updating rule, the answer is 5 iterations; the answer for the true Bayesian updating version turns out to be 88 iterations.

4.3 Can a Larger n Speed Up Convergence?

Finally, one might wonder whether a larger number of observations, n , of the social data on each iteration leads to quicker convergence to the truth. Figure 4.4 below shows the number of iterations needed for a given ϵ when the n at each iteration is 5, 40 and 500, respectively. Interestingly, there is very little difference in average convergence rate for different sizes of n .

Figure 4.2: Simulation Using True Bayesian Updating Rule



However, there is greater variation across simulations when n is smaller. This implies that when the data are non-exchangeable, the most efficient sample size is $n = 1$! That is, given this set up, it is worth updating the posterior (and, importantly, next period's prior beliefs which factor into next period's new data) after each new observation. Given the option to update once with $n = 40$ and the option to update 40 times with $n = 1$, the latter would be the more efficient strategy.

Figure 4.3: Testing the R Function

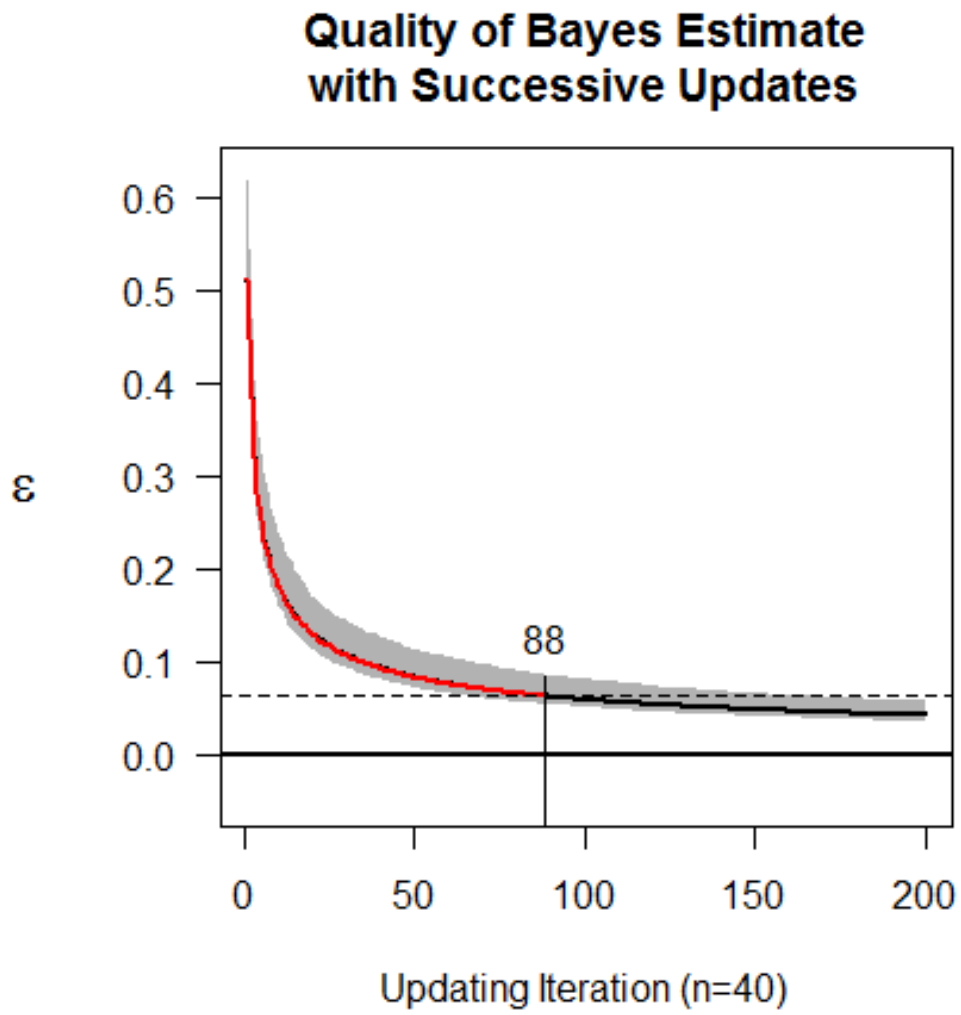
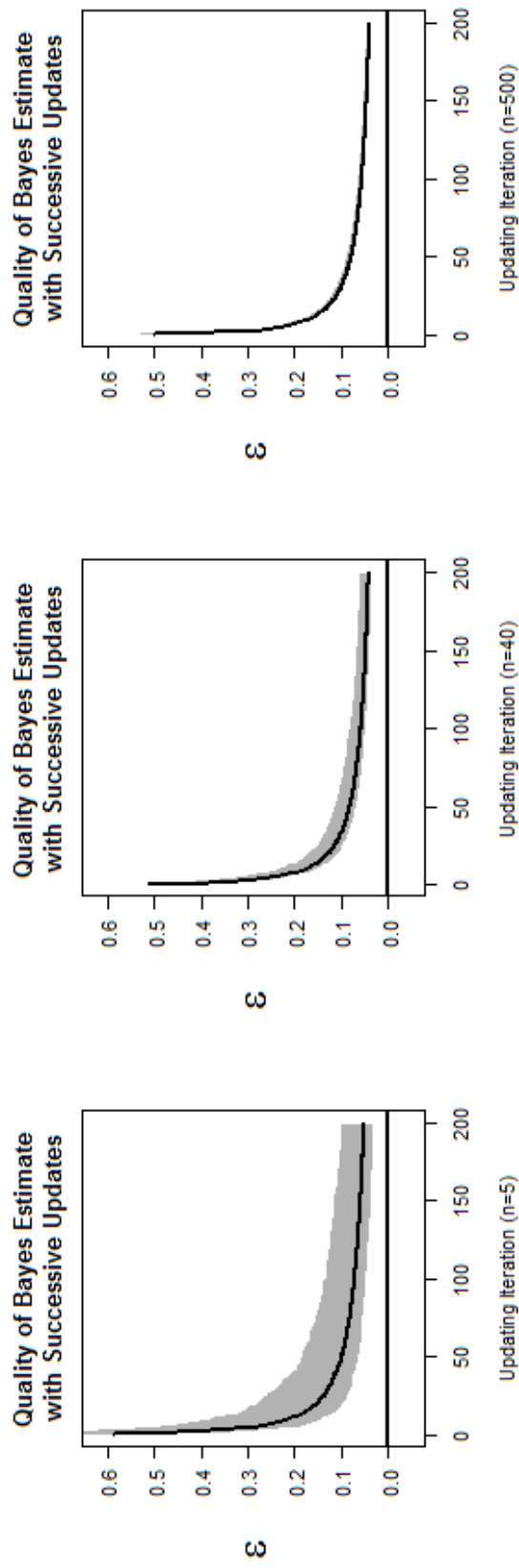


Figure 4.4: Iterative Bayesian Updating with Different Sample Sizes



CHAPTER 5

Comparing Patterns of Social Learning

Is my characterization of the social data-generating process descriptively consistent with any real-life scenarios? In this section I engage in a stylized application of the model elaborated in the previous chapters to real New York City “Stop and Frisk” data and model-based synthesized conviction data. I illustrate the use of the model to make inferences about social learning by assessing patterns of annual updating in New York City police stops in comparison to patterns predicted by the different models of learning presented in this thesis. This analysis is meant to illustrate the type of applied work that is possible with the model I present and should not be taken as an exhaustive or conclusive study of the motivations for pedestrian police stops in New York City. I examine the degree to which annual changes in police stop rates in New York City can plausibly be described by different theoretical models of social learning proposed in the previous chapters. I consider a model of learning from objective evidence uncovered in the field and contrast this with a model of learning from social data using (simulated) juror-influenced conviction data. In the latter case, I compare the observed movement in annual stop rates with counterfactuals produced by the conservative social updating model (in which the researcher puts little weight on her priors) and the Bayesian social updating model introduced in Chapter 4.

5.1 Data Application

New York City’s “Stop and Frisk” policy permits police officers to stop, question and frisk individuals based on “reasonable suspicion” (which holds a lower burden of proof than “probable cause”). Aggressive implementation of the policy began in the mid-90s under mayor

Table 5.1: Stop Rates per 1,000 in Population

Year	Stops
2003	19.8
2004	38.3
2005	48.5
2007	56.2
2008	64.7
2009	69.3
2010	73.4
2011	82.9
2012	63.9
2013	22.8
2014	5.4

Rudolph Giuliani. The New York Police Department collects data on these stops: why individuals were stopped, whether a weapon or contraband was subsequently found on the individual, and whether the individual is arrested are among the variables recorded.

How do police officers decide how many people to stop? The tradeoff between the personal infringement of experiencing an unwarranted police stop and the potential social gain that (debatably) comes with a higher stop rate is central to this negotiation. Given the option to make increased stops, how do police determine where to draw the line? Because “Stop and Frisk” allows for increased discretionality in police stops, this makes for an interesting opportunity to learn about police officers’ prior beliefs about a suspect’s guilt before stopping—and to study how they update these beliefs in light of new information. Table 5.1 shows stop rates per 1,000 people in the population of New York City (from U.S. Census data) over time. There exists considerable variation from year to year. The stop rate increases annually until 2011 and then suddenly drops off. What explains the annual variation?

The data application presented in this chapter attempts to address this question. Specifically, it asks, do police update stop rates based off the rate of evidence recovery in the field? Alternatively, do they rely principally on their own prior beliefs of guilt? Or do they update stop rates based on social data—data influenced by the judgment of human actors—such

as witness reports, victim complaints, or jury trial conviction rates? If police “learn” from social data, to what degree do stop rates reflect psychological biases in the assessment of guilt shared by social actors including witnesses, victims, jurors and police relative to objective data? I further explore this topic using a model-driven and, admittedly, simplistic observationalization of the “Stop and Frisk” data.

I treat the annual stop rate (out of the NYC resident population¹) as a stylized measure of the police officer’s prior beliefs of guilt (before questioning, frisking, or gathering more socially-processed sources of information such as a determination of “guilt” by the justice system). Figure 5.1 shows the evolution of the police officers’ prior beliefs using this stylized measure—the proportion of NYC residents stopped over time. The plot suggests that police officers engage in some form of annual updating.² That is, if stop rates were driven by the stylized measure of “prior beliefs” alone and no data was incorporated, the evolution would be a flat horizontal line, since the priors would never be updated.

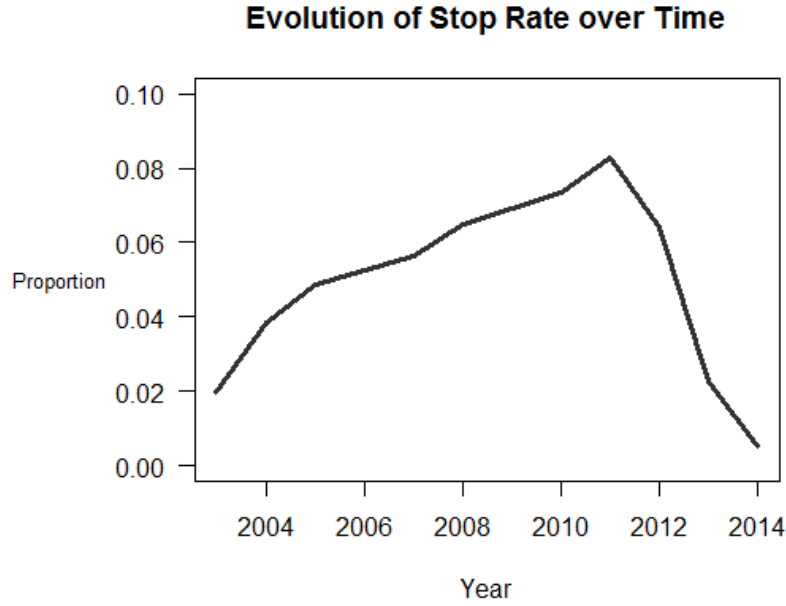
In my stylized operationalization, officers who stop individuals based on “priors” then get to observe some evidence—that is, whether the individual is carrying a weapon or contraband. I will treat the observation of this evidence, given a stop, as an (again, simplistic) measure of law-breaking intent. I treat this as a first opportunity for officers to “learn” from data (i.e., update their beliefs). I compare the evolution of the prior beliefs (the annual proportion of individuals stopped in New York City) with the evolution of the “evidence” (the annual proportion of stoppees found to be carrying weapons or contraband). The first hypothesis I explore is that police officers update their stop rates based on this measure of objective “evidence.” If the observed evidence rate drives the evolution of police officers’ future stop rates, then next year’s prior stop rate (this year’s posterior) will be a positive function of this year’s evidence recovery rate.

As an alternative hypothesis, I suppose the officer uses more socially-processed informa-

¹Note that this denominator assumes that the police have a chance to stop every resident of NYC, or at least a random sample of them.

²An alternative interpretation might exist if the number of police officers *assigned* to participate in “Stop and Frisk” were to have changed on an annual basis. This would imply the annual learning occurs at a higher level of the police hierarchy.

Figure 5.1: Police Stop Rates



tion to inform next year’s prior. In this alternative, but still stylized, operationalization of learning, the officer considers guilt to be determined by the justice system. Then “conviction” rather than “evidence” undergirds my predictions of police learning. The NYPD does not collect data on the conviction results for stop and frisk suspects. Therefore, I use my social data model as a structural equation to probabilistically generate simulated conviction data. I assume juries use my social data-generating function, placing some weight, δ on noisy social priors shared by the jury and police officers and some weight, $1 - \delta$, on noisy “evidence,” to generate a latent variable, x^* , which when above an arbitrary cutoff point, results in conviction (I will later illustrate the results for two values of $\delta \in \{0.1, 0.9\}$). Then, x^* is generated as

$$x_i^* = \delta(\text{Prior Stop Rate} + \nu_i) + (1 - \delta)(\text{Evidence Present}_i + z_i),$$

for a given year. I take the standard deviation of ν_i as equal to the standard deviation of Bernoulli trials for whether an individual was stopped or not in the data, and the standard

deviation of z_i as equal to the standard deviation of Bernoulli trials for whether a stopped individual was found to possess weapons or contraband or not. For simplicity, I take normal draws from these distributions, since the binomial distribution can be well-approximated by the normal distribution with the type of high sample size present in this data. I then use a cutoff of 0 to assign convictions. When $x^* > 0$, the suspect is convicted, if not the suspect is acquitted. I then use the aggregate conviction rate as the “social data,” X , observed by police officers.³

I then compare the evolution of the stylized measure of police officers’ prior beliefs with the annual stop rates predicted by my theoretical model of social learning. This model assumes annual learning from conviction data generated by the structural model of social data defined above. I provide two different counterfactuals for the evolution of police “priors” based on my model. In one, I assume the police learn from social data in a true Bayesian sense—they consider last year’s posterior as this year’s prior and they place the full weight of last year’s data on this year’s prior. I call this “learning from social data with Bayesian updating.” Alternatively, I refer to the counterfactual generated by the more conservative model of social learning (in which the police officer learns from social data, but does not increase the weight on his prior beliefs as he collects more data) as “learning from social data with conservative updating.” I will assume that—among the alternatives presented—the model that generates the counterfactual that most resembles the observed data is the most plausible model of social learning describing this scenario. I acknowledge, however, that the counterfactuals described by my model of social learning are non-exhaustive and it remains highly likely that more nuanced models could do a similarly good job or even a better job at describing police stop data.

³Note that while I refer to my simulated data as “conviction” data to use a concrete example, it actually represents any social data generated by the social data-generating function. For example, it could also be “police tips” where the latent variable x^* is informed by noisy prior beliefs and evidence and where the cutoff of zero determines whether the witness decides to “report” a suspected crime. So, in fact, the “conviction” results may be more generally interpreted as “social data” results.

5.2 Results

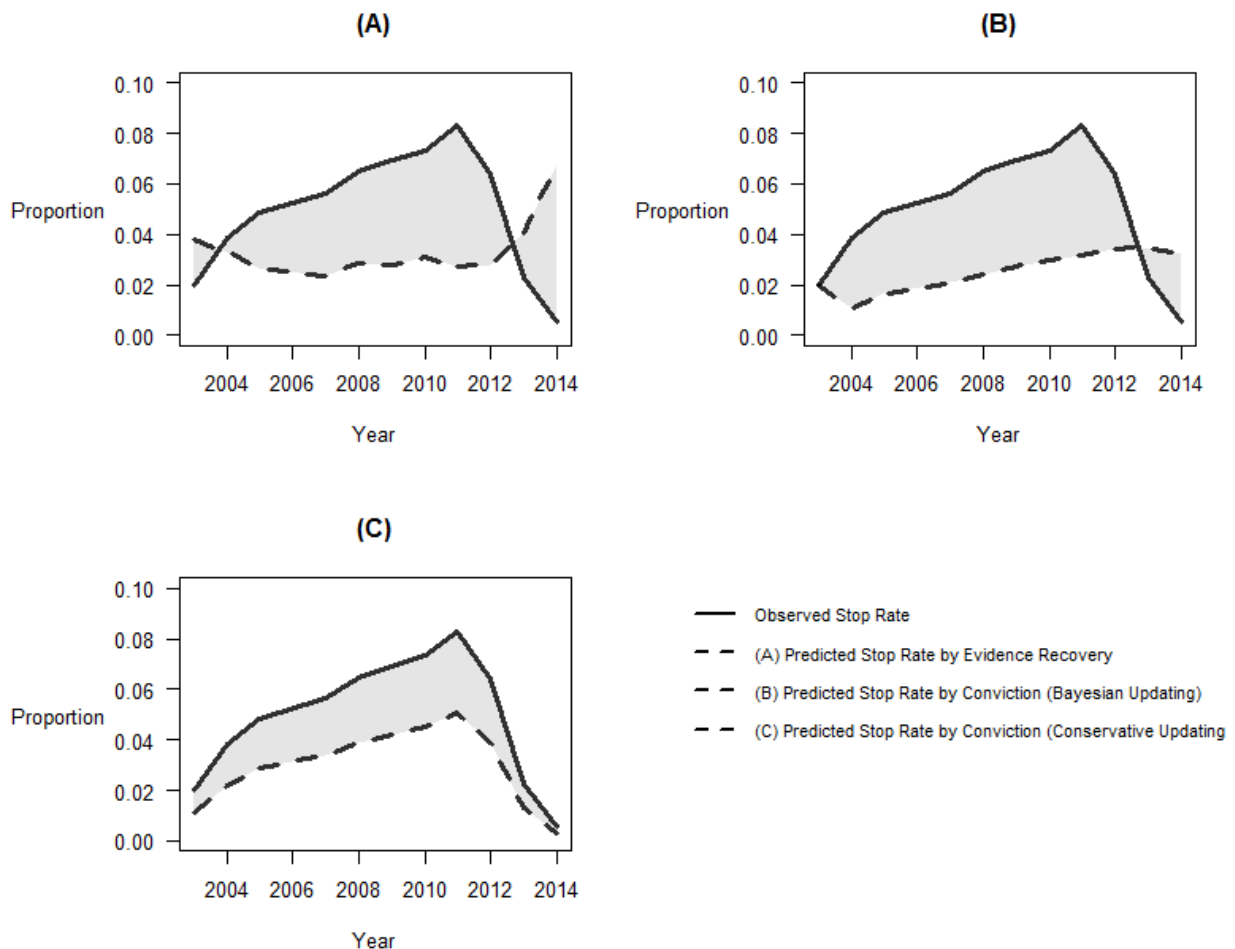
Below I compare three counterfactuals with my stylized measure of officers’ annual prior beliefs. In Figure 5.2, plot (A) compares the prior stop rates with the rate of “objective evidence” collected conditional on a stop. The large disparity between this counterfactual and the observed measure implies that the annual police stop rate is not governed principally by the contraband- and weapon-recovery rates observed among stoppees. Interestingly, this measure of “evidence” seems almost inversely correlated with the observed annual stop rate.

Plot (B) compares the observed “prior data” with a counterfactual generated by my social data-generating (conviction) function and assumes increasing weight on prior beliefs in each year as the data compiles. This counterfactual is consistent with a stylized model of police learning from social data in a true Bayesian fashion. This updating rule implies that police will be less sensitive to new data as time goes on, once they already have considerable data-weight supporting their “prior beliefs.” This counterfactual also falls short of a good fit—the officers’ stop rates vary more greatly than would be predicted by Bayesian updating from social data. Though this counterfactual is more highly (positively) correlated with the officers’ annual behavior than the evidence alone, the correlation remains weak.

Finally, plot (C) provides a counterfactual which models police learning from conviction data generated by my social data-generating function, where police do not increase the data-weight on their own priors annually, making the function more sensitive to annual fluctuations in social data as time passes. This makes sense, since the priors factor into the social data. That is, this year’s counterfactual is essentially an AR1 lag of this year’s police prior belief with some independent information from the evidence. Perhaps not surprisingly, this plot provides the best fit, as measured by correlation between the counterfactual and the observed police behavior, among the provided alternatives. Again, this does not preclude the possibility that a counterfactual model not considered here is, in fact, the best fit for the already-stylized measure of “priors.”

Plots (B) and (C) in Figure 5.2 are generated assuming that $\delta = 0.9$ in the social data-generating conviction function—that is, simulated jurors rely very highly on noisy “prior

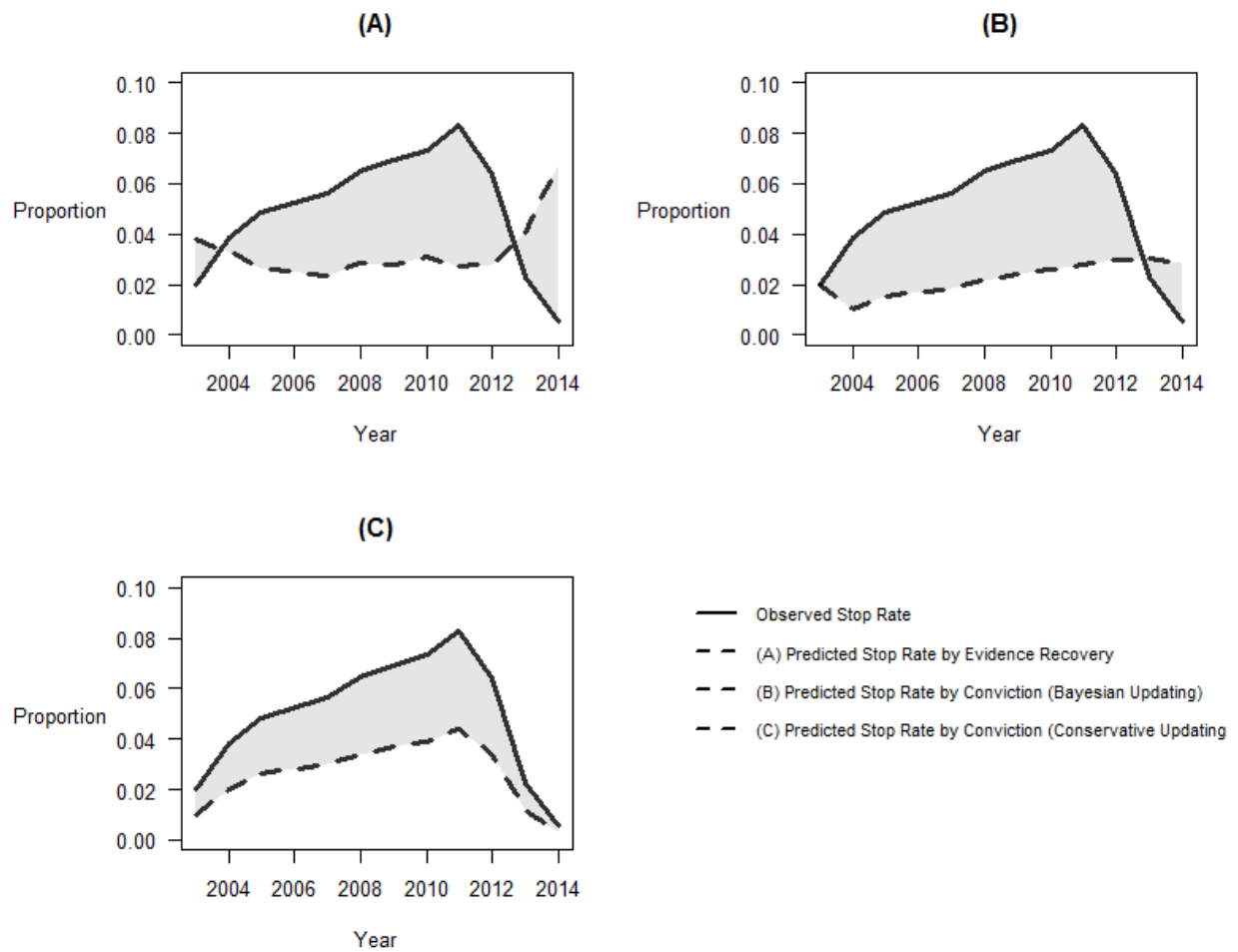
Figure 5.2: Comparing Evolution of Police “Priors” with Model-Based Counterfactuals ($\delta = 0.9$)



beliefs” about a suspect’s guilt, but do still put a weight of 10% on a noisy measure of “evidence.” Suppose one were to instead simulate a justice system with jurors more willing to be influenced by trial evidence—say, for example, one in which jurors base only 10% of their conviction decision on noisy “prior beliefs” and 90% on the measure of “evidence.” The counterfactual results for this justice system are shown in Figure 5.3. Under this assumption, plot (C) (social learning with conservative updating) still provides the best model fit (taken as the counterfactual that minimizes the sum of squared deviations from the data). Thus, changing this modeling assumption does not drastically affect the conclusion of this data application: that annual variation in police stop rates in New York City is more consistent

with conservative updating from a social data-generating function that is non-exchangeable with their own prior beliefs than it is with a model which relies purely on evidence uncovered in the field or which is heavily influenced by the officer's own prior beliefs.

Figure 5.3: Comparing Evolution of Police “Priors” with Model-Based Counterfactuals ($\delta = 0.1$)



CHAPTER 6

Conclusion

This exploration has generated practical guidelines, analogous to power analyses, for situations in which it is plausible that socio-psychological biases inform both the behaviors of the humans generating behavioral data and the prior beliefs of the researcher analyzing those data. This experiment helps to make salient several relevant methodological and philosophical points regarding Bayesian social learning. This final discussion will summarize and flesh out several key findings of this research.

First, it is important to consider the plausibility of the exchangeability assumption with respect to prior data—even if that data is non-quantitative in its raw form. Though most scholars are in the practice of considering questions of exchangeability with respect to newly observed data, the question of whether the prior data observations are exchangeable with the newly observed data is not regularly considered, at least not in practice. The tendency to forgo an explicit justification of the exchangeability assumption with respect to priors is especially common when the prior distribution is informed by non-quantitative data or beliefs.

Unfortunately, such oversights are likely to be fairly common in the social sciences where many of our data are human responses to socialized perceptions. It is therefore worth identifying contexts in which this situation is especially likely to arise. Certain contexts, in particular, informational contexts, may make our data-generating actors more or less prone to relying on prior beliefs to inform their behaviors or decisions. When a behavior requires a judgment to be made with relatively little information about the subject (such as in a jury trial or a psychological diagnosis), individuals will automatically tend to use the little bit of information they (feel they) have to fill in the blanks about the subject

in a way that is consistent with the details they know, as they understand them (Zaller, 1992). Low levels of information are important not just because they can lead to reliance on psychological biases to inform the “blanks,” but also because they can result in unreasonable levels of certainty about these conclusions. Kahneman says “you cannot help dealing with the limited information you have as if it were all there is to know...Paradoxically, it is easier to construct a coherent story when you know little” (Kahneman, 2011). When we receive little information from the world about a subject, we may find our prior beliefs constitute most of the story (since we have used them to fill in the blanks). Therefore, δ will likely be larger when information to supplement prior beliefs is limited. On the other end of the spectrum, it also seems that information overload can lead a person to give up on sorting out the facts, instead reverting to autonomic shortcuts including psychological biases—as may be true, for example, of voting behavior in elections with a high number of parties or competitors (Aguilar et al., 2015).

Second, rather than rely on limiting assumptions, we can calculate how many observations or iterations would be needed to make a reasonable inference, even assuming unfavorable prior exchangeability conditions. The justification we normally use to feel comfortable with potentially-biased Bayesian priors is that as we observe more and more new (exchangeable) data, the weight of our biased prior diminishes. In the limit, the original prior from the first time period has no influence. Thus, even if our prior were completely misinformed, there is no aggregate or long-run bias in our methodology. However, when the data themselves are informed in part by our biased priors (and are therefore non-exchangeable), the medium-run implications of our methodology become quite important. Even if in the limit we can expect reasonable posterior estimates, this gives us little notion of how close our immediate reality comes to the limit and, for practical purposes, tells us little about the quality of our inferences. We should have some notion of what number of observations or updating iterations are needed to arrive at a reasonably informative estimate. This paper provides a method to assist the researcher in making this judgment.

Appendix

Appendix A

```
n <- 0:40
alpha <- 4
tau <- 2
mu <- 0
sigma <- 2
delta <- .5

alpha_1 <- ((alpha/(tau^2)) +
            (n*(delta*alpha + (1-delta)*mu))/
            ((delta^2)*(tau^2)+((1-delta)^2)*(sigma^2)) )/
            ( (1/(tau^2))+n/((delta^2)*(tau^2)+((1-delta)^2)*(sigma^2))) )

tau_1 <- 1/(sqrt((1/(tau^2))+
                (n/((delta^2)*(tau^2)+((1-delta)^2)*(sigma^2)))))

plot(n,alpha_1-mu,type="n",ylim=c(1.8,4.2)/4,ylab="",
     main="Single Iteration Error")
mtext(side=2,line=4,expression(epsilon),cex=1.5,font=2)
lines(n,(alpha_1-mu)/4,lwd=3)
abline(h=delta*abs((alpha-mu)/4),lty=2)
abline(h=abs((alpha-mu)/4),lty=2)
mtext(side=2,line=3,at=2/4,expression(epsilon<delta*abs(alpha-mu)),
      font=2)
mtext(side=2,line=3.4,at=4/4,expression(epsilon>abs(alpha-mu)),
      font=2)
```

Appendix B

```
conservative.prior.iterator <- function(iters=20,n=40,alpha=4,tau=2,
                                       mu=0,sigma=2,delta=.5,n_r=1){

  prior_i <- rnorm(n=n,alpha,tau)
  true_info_i <- rnorm(n=n,mu,sigma)
  x_i <- delta*prior_i + (1-delta)*true_info_i
  researcher_prior <- rnorm(n=n_r,alpha,tau)
  naive_posterior <- c(researcher_prior,x_i)
  post_mean <- mean(naive_posterior)
  post_sd <- sd(naive_posterior)
  my.t <- 1
  test <- c(alpha,post_mean)

  # iterate
  repeat{
    alpha_1 <- post_mean
    tau_1 <- post_sd
    prior_i <- rnorm(n=n,alpha_1,tau_1)
    true_info_i <- rnorm(n=n,mu,sigma)
    x_i <- delta*prior_i + (1-delta)*true_info_i
    researcher_prior <- rnorm(n=n_r,alpha_1,tau_1)
    naive_posterior <- c(researcher_prior,x_i)
    post_mean <- mean(naive_posterior)
    post_sd <- sd(naive_posterior)
    my.t <- my.t+1
    test <- c(test,post_mean)
    if(my.t>=iters){
      break
    }
  }
}
```

```

    }
  }
}

iters=10
p <- 300
n <- 40

par(mar=c(5.1, 6.1, 4.1, 2.1), las = 1,mfrow=c(1,2))
k <- (1/(tau^2)) /
  (((1/(tau^2)))+(n/((delta^2)*(tau^2) + ((1-delta)^2)*sigma^2)))
bias <- delta+k*(1-delta)
my.t <- seq(0,10,length=100)
my.eps <- (bias^my.t)*abs(alpha-mu)
plot(my.t,my.eps/4,type="n",
      xlab="t",ylab="",
      main="Analytical Calculation")
mtext(side=2,expression(epsilon),line=3.5,cex=1.5)
lines(my.t,my.eps/4,lwd=3)

my.ests <- matrix(NA,nrow=iters+1,ncol=p)
for(i in 1:p){
  conservative.prior.iterator(iters=iters,n=40)
  my.ests[,i] <- test
}

res <- apply(my.ests,1,mean)
plot(0:my.t,res/4,type="n",
      xlab="t",ylab="",

```

```

    main="Simulated Data")
mtext(side=2,expression(epsilon),line=3.5,cex=1.5)
lines(0:iters,res/4,lwd=3)

```

Appendix C

The weight on the new data $\bar{X}^{\{1\}}$ at time, $t = 1$ is greater than the weight of the new data $\bar{X}^{\{2\}}$ at time, $t = 2$ whenever the following is true.

For

$$A = \frac{1}{\tau^2}$$

$$B = \frac{n}{\delta^2 \tau^2 + (1 - \delta)^2 \sigma^2}$$

$$C = \frac{n \left(\frac{1}{\tau^2} + \frac{n}{\delta^2 \tau^2 + (1 - \delta)^2 \sigma^2} \right)}{\delta^2 + (1 - \delta)^2 \sigma^2},$$

$$\frac{B}{A + B} > \frac{C}{A + B + C}.$$

Rearranging terms, the weight on the newly observed data decreases with each iteration when

$$\frac{B}{C} > \frac{A + B}{A + B + C},$$

or,

$$\frac{\frac{n}{\delta^2\tau^2+(1-\delta)^2\sigma^2}}{\frac{n\left(\frac{1}{\tau^2}+\frac{n}{\delta^2\tau^2+(1-\delta)^2\sigma^2}\right)}{\delta^2+(1-\delta)^2\sigma^2}} > \frac{\frac{1}{\tau^2} + \frac{n}{\delta^2\tau^2+(1-\delta)^2\sigma^2}}{\frac{1}{\tau^2} + \frac{n}{\delta^2\tau^2+(1-\delta)^2\sigma^2} + \frac{n\left(\frac{1}{\tau^2}+\frac{n}{\delta^2\tau^2+(1-\delta)^2\sigma^2}\right)}{\delta^2+(1-\delta)^2\sigma^2}}.$$

Since A , B and C are all greater than zero by construction, we can manipulate this inequality freely without worrying about changing the direction of the inequality. Rearranging terms, this simplifies to

$$\frac{\tau^2(\delta^2 + (1-\delta)^2\tau^2)}{\delta^2\tau^2 + (1-\delta)^2\sigma^2 + n\tau^2} > \frac{\delta^2 + (1-\delta)^2\sigma^2}{\delta^2 + (1-\delta)^2\sigma^2 + n}.$$

This further simplifies to

$$(1-\delta)^2\sigma^2\tau^2 > (1-\delta)^2\sigma^2,$$

or

$$\tau^2 > 1.$$

This suggests that whenever τ^2 is greater than one, the weight on the contribution of the new data to the posterior decreases with each iteration.

Appendix D

```
my.naive.iterator <- function(iters=20,n=40,alpha=4,tau=2,
                               mu=0,sigma=2,delta=.5,n_r=1){
```

```

prior_i <- rnorm(n=n,alpha,tau)
true_info_i <- rnorm(n=n,mu,sigma)
x_i <- delta*prior_i + (1-delta)*true_info_i
researcher_prior <- rnorm(n=n_r,alpha,tau)
naive_posterior <- c(researcher_prior,x_i)
post_mean <- mean(naive_posterior)
my.t <- 1
test <- post_mean

# iterate
repeat{
  alpha_1 <- post_mean
  tau_1 <- sd(naive_posterior)
  prior_i <- rnorm(n=n,alpha_1,tau_1)
  true_info_i <- rnorm(n=n,mu,sigma)
  x_i <- delta*prior_i + (1-delta)*true_info_i
  naive_posterior <- c(naive_posterior,x_i)
  post_mean <- mean(naive_posterior)
  my.t <- my.t+1
  test <- c(test,post_mean)
  if(my.t>=iters){
    break
  }
}

iters=200
p <- 300
lower.b <- numeric(p)

```

```

upper.b <- numeric(p)

my.ests <- matrix(NA,nrow=iters,ncol=p)
for(i in 1:p){
  my.naive.iterator(iters=iters,n=40)
  my.ests[,i] <- test
}

res <- apply(my.ests,1,mean)
upper.b <- apply(my.ests,1,function(x){
  quantile(x,.975)
})
lower.b <- apply(my.ests,1,function(x){
  quantile(x,.225)
})

par(mar=c(5.1, 7.1, 4.1, 2.1), las = 1)
plot(1:iters,seq(-.2,2.5,length=iters)/4,type="n",
     xlab="Updating Iteration (n=40)",ylab="",
     main="Quality of Bayes' Estimate\nwith Successive Updates")
mtext(side=2,expression(epsilon),line=4,cex=1.5)
abline(h=mu,lwd=2)
lines(1:iters,res/4,col="black",lwd=2)
polygon(x=c(1:iters,rev(1:iters)),y=c(lower.b,rev(upper.b))/4,
        col=rgb(0,0,0,.3),border=NA)

```

Appendix E

```

naive_post <- function(max.iters=1000,eps=.25,n=40,alpha=4,tau=2,
                       mu=0,sigma=2,delta=.5,n_r=1,p=1000){

```

```

prior_i <- matrix(rnorm(n=n*p,alpha,tau),nrow=p,ncol=n)
true_info_i <- matrix(rnorm(n=n*p,mu,sigma),nrow=p,ncol=n)
x_i <- delta*prior_i + (1-delta)*true_info_i
researcher_prior <- matrix(rnorm(n=n_r*p,alpha,tau),nrow=p,ncol=1)
naive_posterior <- cbind(researcher_prior,x_i)
my.ests <- apply(naive_posterior,1,mean)
my.sds <- apply(naive_posterior,1,sd)
alpha_1 <- mean(my.ests)
tau_1 <- mean(my.sds)
my.t <- 1
test <- alpha_1

# iterate
repeat{
  if(my.t==max.iters){
    print("max iters reached")
    break
  }
  if((test[length(test)]-mu) < eps){
    print(paste("tolerance rate obtained, t=",my.t))
    break
  }
  alpha_0 <- alpha_1
  tau_0 <- tau_1
  prior_i <- matrix(rnorm(n=n*p,alpha_0,tau_0),nrow=p,ncol=n)
  true_info_i <- matrix(rnorm(n=n*p,mu,sigma),nrow=p,ncol=n)
  x_i <- delta*prior_i + (1-delta)*true_info_i
  naive_posterior <- cbind(naive_posterior,x_i)

```

```

    my.ests <- apply(naive_posterior,1,mean)
    my.sds <- apply(naive_posterior,1,sd)
    alpha_1 <- mean(my.ests)
    tau_1 <- mean(my.sds)
    my.t <- my.t+1
    test <- c(test,alpha_1)
  }
}

naive_post(max.iters=1000,eps=.25,n=40,alpha=4,tau=2,
           mu=0,sigma=2,delta=.5,n_r=1,p=1000)
lines(1:my.t,test/4,col="red",lwd=2)
abline(h=.25/4,lty=2)
text(88,.5/4,"88")
segments(x0=88,y0=-.5/4,x1=88,y1=.33/4)

```

BIBLIOGRAPHY

- Aguilar, Rosario, Saul Cunow, Scott Desposato and Leonardo Sangali Barone. 2015. “Ballot Structure, Candidate Race, and Vote Choice in Brazil.” *Latin American Research Review* 50(3):175–202.
- Bar-Tal, Daniel. 2000. *Shared beliefs in a society: Social psychological analysis*. Sage Publications.
- Bernardo, José M. 1996. “The concept of exchangeability and its applications.” *Far East Journal of Mathematical Sciences* 4:111–122.
- De Finetti, Bruno and Antonio Machí. 2002. *Theory of probability: A critical introductory treatment*. Wiley.
- Diaconis, Persi and David Freedman. 1980. “Finite exchangeable sequences.” *The Annals of Probability* pp. 745–764.
- Gelman, Andrew, John B Carlin, Hal S Stern and Donald B Rubin. 2014. *Bayesian data analysis*. Vol. 2 Taylor & Francis.
- Gill, Jeff. 2014. *Bayesian methods: A social and behavioral sciences approach*. Vol. 20 CRC press.
- Jackman, Simon. 2009. *Bayesian analysis for the social sciences*. Vol. 846 John Wiley & Sons.
- Johnson, Sindhu R, George A Tomlinson, Gillian A Hawker, John T Granton and Brian M Feldman. 2010. “Methods to elicit beliefs for Bayesian priors: a systematic review.” *Journal of clinical epidemiology* 63(4):355–369.
- Kahneman, Daniel. 2011. *Thinking, fast and slow*. Macmillan.
- Kerns, G Jay and Gábor J Székely. 2006. “Definetti’s Theorem for Abstract Finite Exchangeable Sequences.” *Journal of Theoretical Probability* 19(3):589–608.

Zaller, John. 1992. *The nature and origins of mass opinion*. Cambridge university press.