Statistics 601 – Modern Bayesian Theory

Alexander Volfovsky

Lecture 1 - August 27, 2019

Course information

Instructor: Alexander Volfovsky, Assistant Professor, Dept of Statistical Science, alexander.volfovsky@duke.edu
Course Time: T/Th: 10:05am - 11:20 pm
Course webpage: Sakai
Office Hours: ??

Course information: TAs

Office hours held in Old Chem ??

Lead TA: Jordan Bryan

360 Lab 1 Time: F 8:30am (LINK 88)

Office Hours: (W 11:30) OR **(W 3pm)** OR (W 5pm) OR (M 5pm) OR (T 5pm)

TA: Heather Mathews

360 Lab 2 Time: F 10:05am (LINK 87)

Office Hours: **(W 11:30)** OR (M 1pm)

TA: Fan Bu

Office hours: **(W 1pm)** OR (W 11:30am)

TA: Rihui Ou

Office hours: **(Tu 3pm)** OR (M 11:30am)

TA: Andrew Cooper

Office hours (M 2pm) OR **(Tu 5pm)**

TA: Qiufeng Zhang

Office hours: **M 5pm**

A word about enrollment

- First time (in a while) that we have divorced 360 from 601!
  (By popular demand)

A word about enrollment

- First time (in a while) that we have divorced 360 from 601!
  (By popular demand)
- The awesome: This is a very popular class.

A word about enrollment

- First time (in a while) that we have divorced 360 from 601! (By popular demand)
- The awesome: This is a very popular class.
- Fall 360 enrollment is as expected; Fall 601 enrollment is at capacity.

A word about enrollment

- First time (in a while) that we have divorced 360 from 601! (By popular demand)
- The awesome: This is a very popular class.
- Fall 360 enrollment is as expected; Fall 601 enrollment is at capacity.
- The complicated: There are many of you who emailed me asking for permission to get in (I'm sorry if I did not get back to you in a timely fashion)!

A word about enrollment

- First time (in a while) that we have divorced 360 from 601! (By popular demand)
- The awesome: This is a very popular class.
- Fall 360 enrollment is as expected; Fall 601 enrollment is at capacity.
- The complicated: There are many of you who emailed me asking for permission to get in (I'm sorry if I did not get back to you in a timely fashion)!
- The solution (for graduate students): I am happy to have you in the class but it might not happen this semester. Stick around on the waitlist as there might be some shuffling/reshuffling after the first week. If you don't get in, don't despair! This class is offered every semester.

A word about enrollment

- First time (in a while) that we have divorced 360 from 601! (By popular demand)
- The awesome: This is a very popular class.
- Fall 360 enrollment is as expected; Fall 601 enrollment is at capacity.
- The complicated: There are many of you who emailed me asking for permission to get in (I'm sorry if I did not get back to you in a timely fashion)!
- The solution (for graduate students): I am happy to have you in the class but it might not happen this semester. Stick around on the waitlist as there might be some shuffling/reshuffling after the first week. If you don't get in, don't despair! This class is offered every semester.
- To all: thank you for being patient!

Course information

- What is “Bayes”?

Course information

- What is “Bayes”?
- General overview of some math.

Course information

- What is “Bayes”?
- General overview of some math.
- We will follow Hoff’s book – a chapter a week (give or take).

Course information

- What is “Bayes”?
- General overview of some math.
- We will follow Hoff’s book – a chapter a week (give or take).
- Sprinkles of additional theory...

Course information

- What is “Bayes”?
- General overview of some math.
- We will follow Hoff’s book – a chapter a week (give or take).
- Sprinkles of additional theory...
- Grading:

Course information

- What is “Bayes”?
- General overview of some math.
- We will follow Hoff’s book – a chapter a week (give or take).
- Sprinkles of additional theory...
- Grading:
  - HWs, labs and discussions (more on this in a bit) are worth 20% (no late homeworks...)
  - Quiz 1 (September 17)
  - Midterm (October 22)
  - Quiz 2 (approximately November 19)
  - Final Exam (Saturday, December 14, 5pm)
  - Each quiz is worth 12.5%, Midterm is worth 25%, Final is worth 30%. Miss a quiz or the midterm? That percentage gets pushed to the final exam.

Course information

- What is “Bayes”?
- General overview of some math.
- We will follow Hoff’s book – a chapter a week (give or take).
- Sprinkles of additional theory...
- Grading:
  - HWs, labs and discussions (more on this in a bit) are worth 20% (no late homeworks...)
  - Quiz 1 (September 17)
  - Midterm (October 22)
  - Quiz 2 (approximately November 19)
  - Final Exam (Saturday, December 14, 5pm)
  - Each quiz is worth 12.5%, Midterm is worth 25%, Final is worth 30%. Miss a quiz or the midterm? That percentage gets pushed to the final exam.
- Labs are there to help you! Come learn a concept or two and some R from the TAs and then work on the computational part of the homeworks.

Course information

- What is “Bayes”?
- General overview of some math.
- We will follow Hoff’s book – a chapter a week (give or take).
- Sprinkles of additional theory...
- Grading:
  - HWs, labs and discussions (more on this in a bit) are worth 20% (no late homeworks...)
  - Quiz 1 (September 17)
  - Midterm (October 22)
  - Quiz 2 (approximately November 19)
  - Final Exam (Saturday, December 14, 5pm)
  - Each quiz is worth 12.5%, Midterm is worth 25%, Final is worth 30%. Miss a quiz or the midterm? That percentage gets pushed to the final exam.
- Labs are there to help you! Come learn a concept or two and some R from the TAs and then work on the computational part of the homeworks.
- Post all questions to Piazza...

Course information – continued

Required things

- R – we won't grade code (though do turn it in!), but "pretty" code is generally good practice.
- Come to class! – sometimes we will go beyond the book...
- Come to lab! — will definitely go beyond the book and lecture...
- Write in complete sentences... Show all of your steps...

Course information – continued

Required things

- R – we won't grade code (though do turn it in!), but "pretty" code is generally good practice.
- Come to class! – sometimes we will go beyond the book...
- Come to lab! — will definitely go beyond the book and lecture...
- Write in complete sentences... Show all of your steps...

Desired things

- $\LaTeX$– makes our life better if we can read everything.
- A sample Markdown file will be provided for those who want to submit homework in this format (please print to pdf).
- Do all of the readings before class and ask questions!

“What discussions? I thought this was math...”

- Efron, B., 1986. Why isn’t everyone a Bayesian?. The American Statistician, 40(1), pp.1-5.
- Gelman, A., 2008. Objections to Bayesian statistics. Bayesian Analysis, 3(3), pp.445-449.
- Diaconis, P., 1977. Finite forms of de Finetti’s theorem on exchangeability. Synthese, 36(2), pp.271-281.
- Gelman, A., Meng, X.L. and Stern, H., 1996. Posterior predictive assessment of model fitness via realized discrepancies. Statistica sinica, pp.733-760.
- Dunson, D.B., 2018. Statistics in the big data era: Failures of the machine. Statistics & Probability Letters, 136, pp.4-9.
- ...

What's new this semester?

- Completely revamped labs!
- Statistics PhD students Becky Tang and Jordan Bryan (our lead TA) put together all new labs for this semester.
- Very new: the labs will now introduce STAN (via R) for fast Bayesian computation.
- What is stan?
- (similar to JAGS, BUGS, etc)
- The idea: STAN is somewhat plug-and-play and unlike JAGS and BUGS uses a much more complicated backend to produce posterior samples.
- When the time comes: you will use some of these computational tools from lab to compare to more classical tools that we will be developing in class.
- (Jumping too far ahead: we will *not* be learning how to derive Hamiltonian Monte Carlo samplers, but the goal is for you to feel comfortable with these tools from an applied perspective and to see some of the details.)

Course outline, a preview

Aug 27: intro, math quiz, how does statistics work?

Aug 29: Chapter 2 – probability and exchangeability (hw1 assigned)

Aug 30: Lab 1, review of R

Sep 3: “Pre-”Bayes review (MLE, etc)

Sep 5: Chapter 3 – Binomial and Poisson models (hw1 due, hw2 assigned)

Sep 6: Lab 2, binomial models

Sep 10: Chapter 3 – Exponential families

Sep 12: Chapter 3+ – more priors (hw2 due, hw3 assigned)

Sep 13: Lab 3

Sep 17: Chapter 4 – Monte Carlo

Sep 19: Quiz 1

Today's outline

- Data analysis and good statistical form.
- Some examples of where Bayes is used and why we might want to use it.
- math/stats quiz (15 minutes) – mainly to gage where everyone is.

A little bit about modern data analysis

1. Exploratory Data Analysis (EDA)
2. Formal model building:
   - Likelihood based methods
   - Bayesian approaches (built on top of likelihood based approaches)
3. Model checking and validation
4. Model refinement
5. Quantification of uncertainty when stating conclusions.

Operationalizing data analysis

Step 1. State the question.

????

Step k. Answer the question. (Make profit?)

Operationalizing data analysis

Step 1. State the question.

Operationalizing data analysis

Step 1. State the question.

Step 2. Collect the data.

Operationalizing data analysis

Step 1. State the question.

Step 2. Collect the data.

Step 3. Explore the data.

Operationalizing data analysis

Step 1. State the question.

Step 2. Collect the data.

Step 3. Explore the data.

Step 4. Formulate and state a modeling framework.

Operationalizing data analysis

Step 1. State the question.

Step 2. Collect the data.

Step 3. Explore the data.

Step 4. Formulate and state a modeling framework.

Step 5. Check your models.

Operationalizing data analysis

Step 1. State the question.

Step 2. Collect the data.

Step 3. Explore the data.

Step 4. Formulate and state a modeling framework.

Step 5. Check your models.

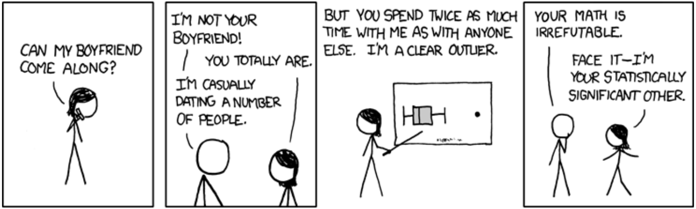
Step 6. Answer the question.

KNOW YOUR DATA

- Are there missing data? (is it random?)
- Is there non-response in your survey? (structural vs not)
- Do your variables have scales? ($\mathrm{BA} > \mathrm{HS}$ but is $2 \times \mathrm{HS} = \mathrm{BA}$?)
- Are there generic coding errors? (is person #9 really 162 years old?)

Exploring the data

- Plots, plots, plots...



Meta: ...okay, but because you said that, we're breaking up.

Modeling and model checking

...essentially what this course is about...

Conclusions

- This is much harder than it looks.

Conclusions

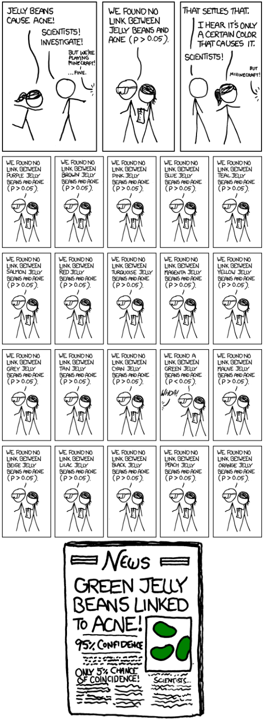
- This is much harder than it looks.
- Did you answer the question you started with?

Conclusions

- This is much harder than it looks.
- Did you answer the question you started with?
- Did you answer a different question?

Conclusions

- This is much harder than it looks.
- Did you answer the question you started with?
- Did you answer a different question?
- Did you go through a bunch of questions until you found one with a “significant” answer?



JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
...FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MIIIINECRAFT!
WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).
NEWS
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time*

Andrew Gelman[†] and Eric Loken[‡]

14 Nov 2013

"I thought of a labyrinth of labyrinths, of one sinuous spreading labyrinth that would encompass the past and the future . . . I felt myself to be, for an unknown period of time, an abstract perceiver of the world." — Borges (1941)

Abstract

Researcher degrees of freedom can lead to a multiple comparisons problem, even in settings where researchers perform only a single analysis on their data. The problem is there can be a large number of *potential* comparisons when the details of data analysis are highly contingent on data, without the researcher having to perform any conscious procedure of fishing or examining multiple p-values. We discuss in the context of several examples of published papers where data-analysis decisions were theoretically-motivated based on previous literature, but where the details of data selection and analysis were not pre-specified and, as a result, were contingent on data.

nature International weekly journal of science

NATURE | RESEARCH HIGHLIGHTS: SOCIAL SELECTION

Psychology journal bans *P* values

Test for reliability of results 'too easy to pass', say editors.

Chris Woolston

26 February 2015 | Clarified: 09 March 2015

A controversial statistical test has finally met its end, at least in one journal. Earlier this month, the editors of *Basic and Applied Social Psychology* (*BASP*) announced that the journal would no longer publish papers containing *P* values because the statistics were too often used to support lower-quality research[1].

SCIENCE THE STATE OF THE UNIVERSE. AUG. 2 2017 4:44 PM

Slate

Will Lowering P-Value Thresholds Help Fix Science?

P-values are already all over the map, and they're also not exactly the problem.

454 176 39

By *Nick Thieme*

The Problems With P-Values are not Just With P-Values

Andrew Gelman

The ASA's statement on p-values says, "Valid scientific conclusions based on p-values and related statistics cannot be drawn without at least knowing how many and which analyses were conducted." I agree, but knowledge of how many analyses were conducted etc. is not enough. The whole point of the "garden of forking paths" (Gelman and Loken 2014) is that to compute a valid p-value you need to know what analyses *would have been done* had the data been different. Even if the researchers only did a single analysis of the data at hand, they well could've done other analyses had the data been different. Remember that "analysis" here also includes rules for data coding, data exclusion, etc.

When I was sent an earlier version of the ASA's statement, I suggested changing the sentence to, "Valid p-values cannot be drawn without knowing, not just what was done with the credible intervals, Bayes factors, cross-validation: you name the method, it can and will be twisted, even if inadvertently, to create the appearance of strong evidence where none exists.

What, then, can and should be done? I agree with the ASA statement's final paragraph, which emphasizes the importance of design, understanding, and context—and I would also add measurement to that list.

What went wrong? How is it that we know that design, data collection, and interpretation of results in context are so important—and yet the practice of statistics is so associated with p-values, a typically misused and misunderstood data summary that is problematic even in the rare cases where it can be mathematically interpreted?

Conclusions

- This is much harder than it looks.
- Did you answer the question you started with?
- Did you answer a different question?
- Did you go through a bunch of questions until you found one with a “significant” answer?

Conclusions

- This is much harder than it looks.
- Did you answer the question you started with?
- Did you answer a different question?
- Did you go through a bunch of questions until you found one with a “significant” answer?
- If someone else tries to run your code will they get the same answer?

Conclusions

- This is much harder than it looks.
- Did you answer the question you started with?
- Did you answer a different question?
- Did you go through a bunch of questions until you found one with a “significant” answer?
- If someone else tries to run your code will they get the same answer?
- Are your pictures seed-dependent?

Lets take a quiz

- 10 minutes, no books, no internet, no "phone-a-friend"...
- (Not) graded – I want to see where everyone is on their math and stats...
- Don't worry about not knowing the answers – and if you know all the answers there's still lots to learn!

What is “Bayes”?

A collection of methods that provide

- parameter estimates with good statistical properties;
- parsimonious descriptions of observed data;
- predictions for missing data and forecasts of future;
- a computational framework for model estimation, selection and validation.

What is "Bayes"?

A collection of methods that provide

- parameter estimates with good statistical properties;
- parsimonious descriptions of observed data;
- predictions for missing data and forecasts of future;
- a computational framework for model estimation, selection and validation.

Mathematically: $\mathcal{Y}$ is the sample space, y is a dataset, Θ is a parameter space and θ is a parameter of interest.

What is "Bayes"?

A collection of methods that provide

- parameter estimates with good statistical properties;
- parsimonious descriptions of observed data;
- predictions for missing data and forecasts of future;
- a computational framework for model estimation, selection and validation.

Mathematically: $\mathcal{Y}$ is the sample space, y is a dataset, Θ is a parameter space and θ is a parameter of interest.

Idealized Bayesian learning:

What is “Bayes”?

A collection of methods that provide

- parameter estimates with good statistical properties;
- parsimonious descriptions of observed data;
- predictions for missing data and forecasts of future;
- a computational framework for model estimation, selection and validation.

Mathematically: $\mathcal{Y}$ is the sample space, y is a dataset, Θ is a parameter space and θ is a parameter of interest.

Idealized Bayesian learning:

1. For each $\theta \in \Theta$ specify a prior distribution $p(\theta)$ describing our beliefs about θ being the true population parameter.

What is "Bayes"?

A collection of methods that provide

- parameter estimates with good statistical properties;
- parsimonious descriptions of observed data;
- predictions for missing data and forecasts of future;
- a computational framework for model estimation, selection and validation.

Mathematically: $\mathcal{Y}$ is the sample space, y is a dataset, Θ is a parameter space and θ is a parameter of interest.

Idealized Bayesian learning:

1. For each $\theta \in \Theta$ specify a prior distribution $p(\theta)$ describing our beliefs about θ being the true population parameter.
2. For each $\theta \in \Theta, y \in \mathcal{Y}$ specify a sampling distribution $p(y|\theta)$ that describes our belief that y is the outcome of a study with true parameter θ.

What is “Bayes”?

A collection of methods that provide

- parameter estimates with good statistical properties;
- parsimonious descriptions of observed data;
- predictions for missing data and forecasts of future;
- a computational framework for model estimation, selection and validation.

Mathematically: $\mathcal{Y}$ is the sample space, y is a dataset, Θ is a parameter space and θ is a parameter of interest.

Idealized Bayesian learning:

1. For each $\theta \in \Theta$ specify a prior distribution $p(\theta)$ describing our beliefs about θ being the true population parameter.
2. For each $\theta \in \Theta, y \in \mathcal{Y}$ specify a sampling distribution $p(y|\theta)$ that describes our belief that y is the outcome of a study with true parameter θ.
3. After observing y, for each $\theta \in \Theta$ our posterior distribution $p(\theta|y)$ describes the belief about θ having observed y.

Bayes is induction

Idealized Bayesian learning:

1. For each $\theta \in \Theta$ specify a prior distribution $p(\theta)$ describing our beliefs about θ being the true population parameter.
2. For each $\theta \in \Theta, y \in \mathcal{Y}$ specify a sampling distribution $p(y|\theta)$ that describes our belief that y is the outcome of a study with true parameter θ.
3. After observing y, for each $\theta \in \Theta$ our posterior distribution $p(\theta|y)$ describes the belief about θ having observed y.

Bayes is induction

Idealized Bayesian learning:

1. For each $\theta \in \Theta$ specify a prior distribution $p(\theta)$ describing our beliefs about θ being the true population parameter.
2. For each $\theta \in \Theta, y \in \mathcal{Y}$ specify a sampling distribution $p(y|\theta)$ that describes our belief that y is the outcome of a study with true parameter θ.
3. After observing y, for each $\theta \in \Theta$ our posterior distribution $p(\theta|y)$ describes the belief about θ having observed y.

To go from Step 1 to Step 3 we need to update our beliefs. We do this using Bayes' rule

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int_\Theta p(y|\tilde{\theta})p(\tilde{\theta})d\tilde{\theta}}$$

Example: Rare events

Example: Rare events

- Number of heads for a biased coin.

Example: Rare events

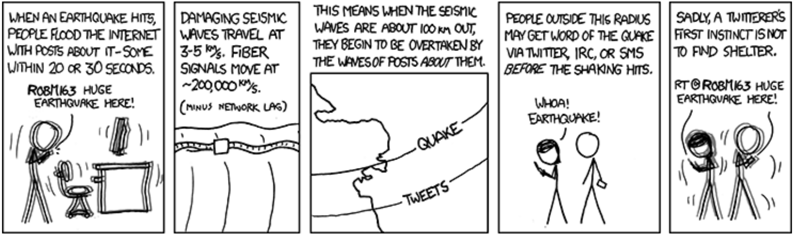
- Number of heads for a biased coin.
- Number of infected individuals in a city.

Example: Rare events

- Number of heads for a biased coin.
- Number of infected individuals in a city.
- Number of worldwise terrorism incidents.

Example: Rare events

- Number of heads for a biased coin.
- Number of infected individuals in a city.
- Number of worldwise terrorism incidents.
- Number of earthquakes of magnitude over 7.



Example: Rare events – Data analysis

Step 1. State the question:

Example: Rare events – Data analysis

Step 1. State the question:

What is the prevalence of an infectious disease in a small city?

Example: Rare events – Data analysis

Step 1. State the question:

What is the prevalence of an infectious disease in a small city?

Why? High prevalence means more public health precautions are recommended.

Example: Rare events – Data analysis

Step 1. State the question:

What is the prevalence of an infectious disease in a small city?

Why? High prevalence means more public health precautions are recommended.

Step 2. Collect the data.

Example: Rare events – Data analysis

Step 1. State the question:

What is the prevalence of an infectious disease in a small city?

Why? High prevalence means more public health precautions are recommended.

Step 2. Collect the data.

A small random sample of 20 individuals are checked.

Example: Rare events – Data analysis

Step 1. State the question:

What is the prevalence of an infectious disease in a small city?

Why? High prevalence means more public health precautions are recommended.

Step 2. Collect the data.

A small random sample of 20 individuals are checked.

Step 3. Explore the data.

Example: Rare events – Data analysis

Step 1. State the question:

What is the prevalence of an infectious disease in a small city?

Why? High prevalence means more public health precautions are recommended.

Step 2. Collect the data.

A small random sample of 20 individuals are checked.

Step 3. Explore the data.

y records the total number of infections in the sample. Sample space is any whole number from 0 to 20.

Example: Rare events – Data analysis

Step 4. Formulate and state a modeling framework.

Example: Rare events – Data analysis

Step 4. Formulate and state a modeling framework.

- Parameter of interest: θ is the fraction of infected individuals in the city.

Example: Rare events – Data analysis

Step 4. Formulate and state a modeling framework.

- Parameter of interest: θ is the fraction of infected individuals in the city.
  Parameter space is numbers between 0 and 1.

Example: Rare events – Data analysis

Step 4. Formulate and state a modeling framework.

- Parameter of interest: θ is the fraction of infected individuals in the city.
  Parameter space is numbers between 0 and 1.
  Mathematically: $\Theta = [0, 1]$, $\mathcal{Y} = \{0, 1, \ldots, 20\}$.

Example: Rare events – Data analysis

Step 4. Formulate and state a modeling framework.

- Parameter of interest: θ is the fraction of infected individuals in the city.
  Parameter space is numbers between 0 and 1.
  Mathematically: $\Theta = [0, 1]$, $\mathcal{Y} = \{0, 1, \ldots, 20\}$.
- Sampling model: Prior to sampling let Y denote the unknown number of infected individuals in the sample.

Example: Rare events – Data analysis

Step 4. Formulate and state a modeling framework.

- Parameter of interest: θ is the fraction of infected individuals in the city.
  Parameter space is numbers between 0 and 1.
  Mathematically: $\Theta = [0, 1]$, $\mathcal{Y} = \{0, 1, \ldots, 20\}$.
- Sampling model: Prior to sampling let Y denote the unknown number of infected individuals in the sample.
  If we know θ a reasonable model for Y would be $\mathrm{binomial}(20, \theta)$.

Example: Rare events – Data analysis

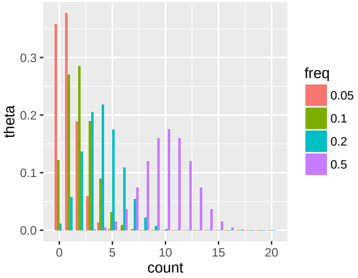
Step 4. Formulate and state a modeling framework.

- Parameter of interest: θ is the fraction of infected individuals in the city.
  Parameter space is numbers between 0 and 1.
  Mathematically: $\Theta = [0, 1]$, $\mathcal{Y} = \{0, 1, \ldots, 20\}$.
- Sampling model: Prior to sampling let Y denote the unknown number of infected individuals in the sample.
  If we know θ a reasonable model for Y would be $\mathrm{binomial}(20, \theta)$.
- We write $Y|\theta \sim \mathrm{binomial}(20, \theta)$

Example: Rare events – Data analysis

Step 4. Formulate and state a modeling framework.

- Parameter of interest: θ is the fraction of infected individuals in the city.
  Parameter space is numbers between 0 and 1.
  Mathematically: $\Theta = [0, 1]$, $\mathcal{Y} = \{0, 1, \ldots, 20\}$.
- Sampling model: Prior to sampling let Y denote the unknown number of infected individuals in the sample.
  If we know θ a reasonable model for Y would be $\mathrm{binomial}(20, \theta)$.
- We write $Y|\theta \sim \mathrm{binomial}(20, \theta)$



Example: Rare events – Data analysis

Step 4. Formulate and state a modeling framework (continued).

Example: Rare events – Data analysis

Step 4. Formulate and state a modeling framework (continued).

- Prior specification: information from previous studies — infection rate in "comparable cities" ranges from 0.05 to 0.20 with an average of 0.10.

Example: Rare events – Data analysis

Step 4. Formulate and state a modeling framework (continued).

- Prior specification: information from previous studies — infection rate in "comparable cities" ranges from 0.05 to 0.20 with an average of 0.10.
- What is a good prior? Most of the weight of $p(\theta)$ is in the interval $(0.05, 0.20)$ and the expected value of θ is close to 0.10.

Example: Rare events – Data analysis

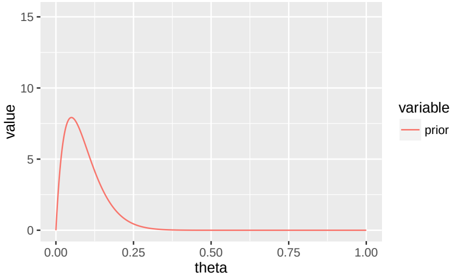
Step 4. Formulate and state a modeling framework (continued).

- Prior specification: information from previous studies — infection rate in "comparable cities" ranges from 0.05 to 0.20 with an average of 0.10.
- What is a good prior? Most of the weight of $p(\theta)$ is in the interval $(0.05, 0.20)$ and the expected value of θ is close to 0.10.
- Possible priors include a beta prior which has expectation $a/(a+b)$ and "most probable value" of $(a-1)/(a-1+b-1)$.

Example: Rare events – Data analysis

Step 4. Formulate and state a modeling framework (continued).

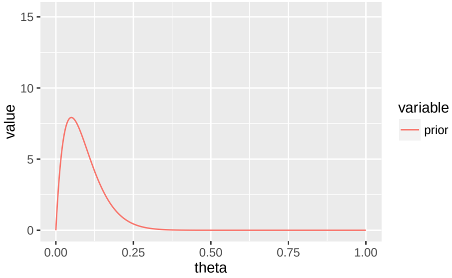
- Prior specification: information from previous studies — infection rate in “comparable cities” ranges from 0.05 to 0.20 with an average of 0.10.
- What is a good prior? Most of the weight of $p(\theta)$ is in the interval $(0.05, 0.20)$ and the expected value of θ is close to 0.10.
- Possible priors include a beta prior which has expectation $a/(a+b)$ and “most probable value” of $(a-1)/(a-1+b-1)$.



Example: Rare events – Data analysis

Step 4. Formulate and state a modeling framework (continued).

- Prior specification: information from previous studies — infection rate in "comparable cities" ranges from 0.05 to 0.20 with an average of 0.10.
- What is a good prior? Most of the weight of $p(\theta)$ is in the interval $(0.05, 0.20)$ and the expected value of θ is close to 0.10.
- Possible priors include a beta prior which has expectation $a/(a+b)$ and "most probable value" of $(a-1)/(a-1+b-1)$.



- Expectation: $E[\theta] = 2/(2+20)$
- $\mathrm{mode}[\theta] = (2-1)/(2-1+20-1)$
- $\Pr(\theta < 0.1) \approx 0.6$
- $\Pr(0.05 < \theta < 0.20) \approx 0.66$

Example: Rare events – Data analysis

Step 4. Formulate and state a modeling framework (continued).

Example: Rare events – Data analysis

Step 4. Formulate and state a modeling framework (continued).

- Posterior distribution for the model:

$$Y|\theta \sim \mathrm{binom}(n, \theta), \quad \theta \sim \mathrm{beta}(a, b)$$

is given by

$$\theta|Y = y \sim \mathrm{beta}(a + y, b + n - y)$$

Example: Rare events – Data analysis

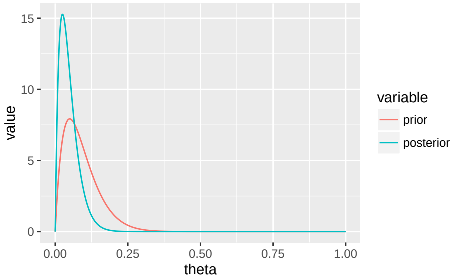
Step 4. Formulate and state a modeling framework (continued).

- Posterior distribution for the model:

$$Y|\theta \sim \mathrm{binom}(n, \theta), \quad \theta \sim \mathrm{beta}(a, b)$$

is given by

$$\theta|Y = y \sim \mathrm{beta}(a + y, b + n - y)$$



Example: Rare events – Data analysis

Step 5. Check your models.

Example: Rare events – Data analysis

Step 5. Check your models.

- Sensitivity analysis! A little more in depth...

$$
\begin{aligned}
E[\theta|Y=y] &= \frac{a+y}{a+b+n} \\
&= \frac{n}{a+b+n}\frac{y}{n} + \frac{a+b}{a+b+n}\frac{a}{a+b} \\
&= \frac{n}{w+n}\bar{y} + \frac{w}{w+n}\theta_0
\end{aligned}
$$

Example: Rare events – Data analysis

Step 5. Check your models.

- Sensitivity analysis! A little more in depth...

$$
\begin{aligned}
E[\theta|Y=y] &= \frac{a+y}{a+b+n} \\
&= \frac{n}{a+b+n}\frac{y}{n} + \frac{a+b}{a+b+n}\frac{a}{a+b} \\
&= \frac{n}{w+n}\bar{y} + \frac{w}{w+n}\theta_0
\end{aligned}
$$

- Interpretation: $\theta_0 = a/(a+b)$ is the prior expectation. $w = a+b$ is some notion of prior confidence.

Example: Rare events – Data analysis

Step 5. Check your models.

- Sensitivity analysis! A little more in depth...

$$
\begin{aligned}
E[\theta|Y=y] &= \frac{a+y}{a+b+n} \\
&= \frac{n}{a+b+n}\frac{y}{n} + \frac{a+b}{a+b+n}\frac{a}{a+b} \\
&= \frac{n}{w+n}\bar{y} + \frac{w}{w+n}\theta_0
\end{aligned}
$$

- Interpretation: $\theta_0 = a/(a+b)$ is the prior expectation. $w = a+b$ is some notion of prior confidence.
- Posterior expectation is a weighted average of the prior expectation and the observed sample mean $\bar{y}$.

Example: Rare events – Data analysis

Step 5. Check your models continued.

Example: Rare events – Data analysis

Step 5. Check your models continued.

- Compare performance of posterior mean and posterior probability that $\theta < 0.1$.

Example: Rare events – Data analysis

Step 5. Check your models continued.

- Compare performance of posterior mean and posterior probability that $\theta < 0.1$.
- We can translate back to a and b via:

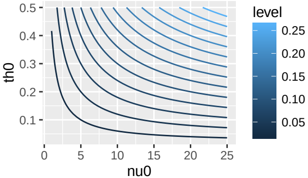
$$a = w\theta_0, \quad b = w(1 - \theta_0).$$

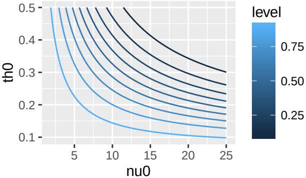
Example: Rare events – Data analysis

Step 5. Check your models continued.

- Compare performance of posterior mean and posterior probability that $\theta < 0.1$.
- We can translate back to *a* and *b* via:

$$a = w\theta_0, \quad b = w(1 - \theta_0).$$





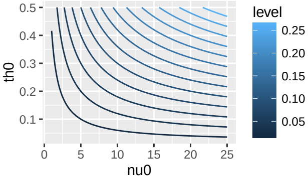
$E[\theta|Y = 0]$ $\qquad$ $P(\theta < 0.1|Y = 0)$

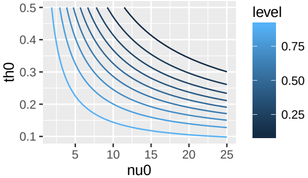
Example: Rare events – Data analysis

Step 5. Check your models continued.

- Compare performance of posterior mean and posterior probability that $\theta < 0.1$.
- We can translate back to *a* and *b* via:

$$a = w\theta_0, \quad b = w(1 - \theta_0).$$





$E[\theta|Y = 0]$ $\qquad$ $P(\theta < 0.1|Y = 0)$

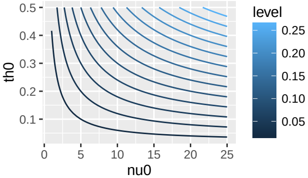
- What can we learn: People with weak prior beliefs (low *w*) or low prior expectations (small θ_0) are generally at least 90% certain that the infection rate is below 0.10.

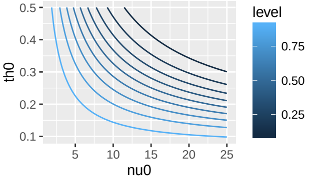
Example: Rare events – Data analysis

Step 5. Check your models continued.

- Compare performance of posterior mean and posterior probability that $\theta < 0.1$.
- We can translate back to a and b via:

$$a = w\theta_0, \quad b = w(1 - \theta_0).$$





$E[\theta|Y = 0]$ $\qquad$ $P(\theta < 0.1|Y = 0)$

- What can we learn: People with weak prior beliefs (low w) or low prior expectations (small θ_0) are generally at least 90% certain that the infection rate is below 0.10.
- High degrees of certainty require high certainty in the prior.

Example: Rare events – Data analysis

Step 6. Answer the question.

Example: Rare events – Data analysis

Step 6. Answer the question.

- Lets say we are comfortable with our prior then we have that the posterior for θ given the event $Y = 0$ is $\mathrm{beta}(2, 40)$.

Example: Rare events – Data analysis

Step 6. Answer the question.

- Lets say we are comfortable with our prior then we have that the posterior for θ given the event $Y = 0$ is $\mathrm{beta}(2, 40)$.
- It is to the left of $p(\theta)$ because the observation $Y = 0$ provides evidence of a low value of θ.

Example: Rare events – Data analysis

Step 6. Answer the question.

- Lets say we are comfortable with our prior then we have that the posterior for θ given the event $Y = 0$ is $\mathrm{beta}(2, 40)$.
- It is to the left of $p(\theta)$ because the observation $Y = 0$ provides evidence of a low value of θ.
- It is more peaked than $p(\theta)$ because it combines information and so contains more information that $p(\theta)$ alone.

Example: Rare events – Data analysis

Step 6. Answer the question.

- Lets say we are comfortable with our prior then we have that the posterior for θ given the event $Y = 0$ is $\mathrm{beta}(2, 40)$.
- It is to the left of $p(\theta)$ because the observation $Y = 0$ provides evidence of a low value of θ.
- It is more peaked than $p(\theta)$ because it combines information and so contains more information that $p(\theta)$ alone.
- The posterior expectation is 0.048

Example: Rare events – Data analysis

Step 6. Answer the question.

- Lets say we are comfortable with our prior then we have that the posterior for θ given the event $Y = 0$ is $\mathrm{beta}(2, 40)$.
- It is to the left of $p(\theta)$ because the observation $Y = 0$ provides evidence of a low value of θ.
- It is more peaked than $p(\theta)$ because it combines information and so contains more information that $p(\theta)$ alone.
- The posterior expectation is 0.048
- The posterior mode is 0.025.

Example: Rare events – Data analysis

Step 6. Answer the question.

- Lets say we are comfortable with our prior then we have that the posterior for θ given the event $Y = 0$ is $\mathrm{beta}(2, 40)$.
- It is to the left of $p(\theta)$ because the observation $Y = 0$ provides evidence of a low value of θ.
- It is more peaked than $p(\theta)$ because it combines information and so contains more information that $p(\theta)$ alone.
- The posterior expectation is 0.048
- The posterior mode is 0.025.
- The posterior probability of $\theta < 0.10$ is 0.93.