Bayesian Inference in Psychology: A Workshop

Jeffrey N. Rouder

October, 2013
Collaborators

- Richard Morey, Groningen
- Mike Pratte, Vanderbilt
- Jory Province, Mizzou
- Paul Speckman, Mizzou, Stats
- Dongchu Sun, Mizzou, Stats
Crisis of Confidence

Belief That Evidence for Effects Has Been Overstated

Publication of Fantastic Extra-Sensory Perception Claims in Mainstream Journals

Several Cases of Outright Fraud

Crisis in How We Produce, Understand, and Evaluate Evidence.
Crisis of Confidence

- Belief That Evidence For Effects Has Been Overstated
Crisis of Confidence

- Belief That Evidence For Effects Has Been Overstated
- Publication of Fantastic Extra-Sensory Perception Claims in Mainstream Journals
Crisis of Confidence

- Belief That Evidence For Effects Has Been Overstated
- Publication of Fantastic Extra-Sensory Perception Claims in Mainstream Journals
- Several Cases of Outright Fraud
Crisis of Confidence

- Belief That Evidence For Effects Has Been Overstated
- Publication of Fantastic Extra-Sensory Perception Claims in Mainstream Journals
- Several Cases of Outright Fraud
- Crisis in How We Produce, Understand, and Evaluate Evidence.
Crisis of Confidence

Some Possible Causes:

▶ Incentives for quantity, especially for young people
▶ Relatively low-level of bona-fide theory
▶ Fundamental Difficulties in Statistical Analysis
Crisis of Confidence

Some Possible Causes:

▶ Incentives for quantity, especially for young people
Crisis of Confidence

Some Possible Causes:

▶ Incentives for quantity, especially for young people
▶ Relatively low-level of bona-fide theory
Crisis of Confidence

Some Possible Causes:

- Incentives for quantity, especially for young people
- Relatively low-level of bona-fide theory
- Fundamental Difficulties in Statistical Analysis
Crisis of Confidence

The \( p < .05 \) Rule
Crisis of Confidence

The $p < .05$ Rule

- People Know It is Not Perfect
Crisis of Confidence

The \( p < .05 \) Rule

- People Know It is Not Perfect
- Performs Admirably; Keeps Spurious Findings Out of The Literature.
Crisis of Confidence

The $p < .05$ Rule

- People Know It is Not Perfect
- Performs Admirably; Keeps Spurious Findings Out of The Literature.
- Easy to Standardize; Easy to Follow; Fun To Teach
Crisis of Confidence

The $p < .05$ Rule

- People Know It is Not Perfect
- Performs Admirably; Keeps Spurious Findings Out of The Literature.
- Easy to Standardize; Easy to Follow; Fun To Teach
- Feels About Right / Natural
Crisis of Confidence

Why the Crisis? What is to be done?

We are prone to bad practices that violate the basic assumptions of the p <.05 Rule and consequently inflate the probability of getting an effect. Bad practices include censoring data, adding more subjects, ignoring pilots, etc, and are now called p-hacking.

Solution: Be good.

- Increased focus on recording intent (preregistration)
- Increased focus on reporting all data
- Increased focus on replication

Prudent to examine the p <.05 Rule.
Crisis of Confidence

- Why the Crisis? What is to be done?
- We are prone to bad practices that violate the basic assumptions of the $p < .05$ Rule and consequently inflate the probability of getting an effect.

Solution: Be good.

- Increased focus on recording intent (preregistration)
- Increased focus on reporting all data
- Increased focus on replication
- Prudent to examine the $p < .05$ Rule.
Crisis of Confidence

- Why the Crisis? What is to be done?
- We are prone to bad practices that violate the basic assumptions of the $p < .05$ Rule and consequently inflate the probability of getting an effect.
- Bad practices include censoring data, adding more subjects, ignoring pilots, etc, and are now called *p-hacking*.
Why the Crisis? What is to be done?

We are prone to bad practices that violate the basic assumptions of the $p < .05$ Rule and consequently inflate the probability of getting an effect.

Bad practices include censoring data, adding more subjects, ignoring pilots, etc, and are now called *p-hacking*.

Solution: Be good.
Crisis of Confidence

Why the Crisis? What is to be done?
- We are prone to bad practices that violate the basic assumptions of the $p < .05$ Rule and consequently inflate the probability of getting an effect.
- Bad practices include censoring data, adding more subjects, ignoring pilots, etc, and are now called $p$-hacking.
- Solution: Be good.
  - Increased focus on recording intent (preregistration)
  - Increased focus on reporting all data
  - Increased focus on replication

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Crisis of Confidence

- Why the Crisis? What is to be done?
- We are prone to bad practices that violate the basic assumptions of the $p < .05$ Rule and consequently inflate the probability of getting an effect.
- Bad practices include censoring data, adding more subjects, ignoring pilots, etc, and are now called *p-hacking*.
- Solution: Be good.
  - Increased focus on recording intent (preregistration)
  - Increased focus on reporting all data
  - Increased focus on replication
- Prudent to examine the $p < .05$ Rule.
The Free Lunch

- The $p < .05$ Rule a “A Free Lunch” Property.
The Free Lunch

- The $p < .05$ Rule a “A Free Lunch” Property.
- Inference is based on the null hypothesis alone and the analyst need not make assumptions about the alternative.
The Free Lunch

- The $p < .05$ Rule a “A Free Lunch” Property.
- Inference is based on the null hypothesis alone and the analyst need not make assumptions about the alternative.
- Lack of assumptions about the alternative is the “Free-Lunch” part.
The Free Lunch

- The \( p < .05 \) Rule a “A Free Lunch” Property.
- Inference is based on the null hypothesis alone and the analyst need not make assumptions about the alternative.
- Lack of assumptions about the alternative is the “Free-Lunch” part.
- Contrast with power analysis. Power requires assumptions about the alternative, “Paid Lunch.”
The Free Lunch

- The $p < .05$ Rule a “A Free Lunch” Property.
- Inference is based on the null hypothesis alone and the analyst need not make assumptions about the alternative.
- Lack of assumptions about the alternative is the “Free-Lunch” part.
- Contrast with power analysis. Power requires assumptions about the alternative, “Paid Lunch.”
- Theme: You Must Pay For Lunch If Inference Is To Be Principled.
The Free Lunch

- The $p < .05$ Rule a “A Free Lunch” Property.
- Inference is based on the null hypothesis alone and the analyst need not make assumptions about the alternative.
- Lack of assumptions about the alternative is the “Free-Lunch” part.
- Contrast with power analysis. Power requires assumptions about the alternative, “Paid Lunch.”
- Theme: You Must Pay For Lunch If Inference Is To Be Principled.
- *The free lunch is rotten, and eating it contributes to the stench in the field.*
I. Frequentist And Bayesian Probability
I. Frequentist And Bayesian Probability

II. What Do We Want To Know: Invariances and Effects
I. Frequentist And Bayesian Probability
II. What Do We Want To Know: Invariances and Effects
III. Pay For Lunch: The Frequentist Perspective
I. Frequentist And Bayesian Probability
II. What Do We Want To Know: Invariances and Effects
III. Pay For Lunch: The Frequentist Perspective
IV. Pay For Lunch: The Bayesian Perspective
Part I. Frequentist and Bayesian Probability
Frequentist Interpretation of Probability

Coin Flip:

Probability is a long-run property of heads over flips, with the limit approaching the probability of the coin. Frequentists are obligated to get it right in the long run.
Frequentist Interpretation of Probability

Coin Flip:

- Probability is a long-run property
Frequentist Interpretation of Probability

Coin Flip:
- Probability is a long-run property
- \[ \frac{\text{\# heads}}{\text{\# of flips}} \], limit of lots of flips
Frequentist Interpretation of Probability

Coin Flip:

- Probability is a long-run property
- \( \frac{\text{# heads}}{\text{# of flips}} \), limit of lots of flips
- Probability is a property of the coin, much like weight and area.
Frequentist Interpretation of Probability

Coin Flip:

- Probability is a long-run property

\[
\frac{\# \text{ heads}}{\# \text{ of flips}}, \text{ limit of lots of flips}
\]

- Probability is a property of the coin, much like weight and area.

- *Frequentists Obligation: Get It Right In The Long Run.*
A Constructive Challenge
A Constructive Challenge

- Dylan Byers, Politico Blogger
A Constructive Challenge

- Dylan Byers, Politico Blogger
- Blog Entry Entitled: *Nate Silver: One-term Celebrity?*
A Constructive Challenge

- Dylan Byers, Politico Blogger
- Blog Entry Entitled: *Nate Silver: One-term Celebrity?*
- Context: Who will win the 2012 Presidential Election as seen in late October
Nate Silver’s Prediction
Byers’ Column

Silver is an ideologue, wants Obama to win.
Silver is an ideologue, wants Obama to win.

- Joe Scarborough, “Nate Silver says this is a 73.6 percent chance that the president is going to win? Nobody in that campaign thinks they have a 73 percent chance. They think they have a 50.1 percent chance of winning. And you talk to the Romney people, it’s the same thing. Both sides understand that it is close, and it could go either way. And anybody that thinks that this race is anything but a tossup right now is such an ideologue, they should be kept away from typewriters, computers, laptops and microphones for the next 10 days, because they’re jokes.”
Silver is an ideologue, wants Obama to win.

- Joe Scarborough, “Nate Silver says this is a 73.6 percent chance that the president is going to win? Nobody in that campaign thinks they have a 73 percent chance. They think they have a 50.1 percent chance of winning. And you talk to the Romney people, it’s the same thing. Both sides understand that it is close, and it could go either way. And anybody that thinks that this race is anything but a tossup right now is such an ideologue, they should be kept away from typewriters, computers, laptops and microphones for the next 10 days, because they’re jokes.”

- “And even then [if Obama wins], you won’t know if he actually had a 50.1 percent chance or a 74.6 percent chance of getting there.”
“And even then [if Obama wins], you won’t know if he actually had a 50.1 percent chance or a 74.6 percent chance of getting there.”
“And even then [if Obama wins], you won’t know if he actually had a 50.1 percent chance or a 74.6 percent chance of getting there.”

By symmetry: a 99% chance or a 1% chance
“And even then [if Obama wins], you won’t know if he actually had a 50.1 percent chance or a 74.6 percent chance of getting there.”

- By symmetry: a 99% chance or a 1% chance
- In what sense is frequentist probability a useful concept here? No large sample limit.
Probability is a statement of belief
Bayesian Interpretation of Probability

- Probability is a statement of belief
- Bayesian focus is on how beliefs should change in light of data.
Bayesian Interpretation of Probability

- Probability is a statement of belief
- Bayesian focus is on how beliefs should change in light of data.
- Long-run properties are consequences rather than primitives.
Bayesian Interpretation of Probability

- Probability is a statement of belief
- Bayesian focus is on how beliefs should *change* in light of data.
- Long-run properties are consequences rather than primitives.
- Probability is used to describe the observer or analyst, not the coin (or system more generally).
Probability is a statement of belief.
Bayesian Probability

Probability is a statement of belief.

- Silver’s $p = 0.75$ means 3:1 odds.
Bayesian Probability

Probability is a statement of belief.

- Silver’s $p = .75$ means 3:1 odds.
  - Silver should bet Byers $2.5$ to win $1$ if Obama wins
  - Silver should not bet $3.5$ to win $1$ if Obama wins.
Bayesian Probability

Probability is a statement of belief.

- Silver’s $p = .75$ means 3:1 odds.
  - Silver should bet Byers $2.5$ to win $1$ if Obama wins.
  - Silver should not bet $3.5$ to win $1$ if Obama wins.

- Probabilities may be placed on anything.
Bayesian Probability

Probability is a statement of belief.

- Silver’s $p = .75$ means 3:1 odds.
  - Silver should bet Byers $2.5$ to win $1$ if Obama wins.
  - Silver should not bet $3.5$ to win $1$ if Obama wins.

- Probabilities may be placed on anything
  - Unreplicated events.
  - Models, Theories, Conclusions
  - How much would you wager for your conclusions to be correct?
    What wager should a journal editor accept?
Bayesian focus is on how beliefs should change in light of data.
Bayesian Probability

- Bayesian focus is on how beliefs should *change* in light of data.
- Bayes Rule: How to update beliefs in light of data
Bayesian Probability

- Bayesian focus is on how beliefs should *change* in light of data.
- Bayes Rule: How to update beliefs in light of data
- *Bayesian Obligation: Use Bayes’ Rule Always.*
Prior Beliefs: How Well Can I Shoot Free Throws?
Prior Beliefs: How Well Can I Shoot Free Throws?

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Prior Beliefs: How Well Can I Shoot Free Throws?

![Graph showing the probability of success against density. The graph peaks at a probability of around 0.6 and decreases towards 1.0. The label "Looks Bad" is indicated at the peak of the curve.](image-url)
Posterior Beliefs: 8 Makes in 12 Tries

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Posterior Beliefs: 8 Makes in 12 Tries

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Posterior Beliefs: 8 Makes in 12 Tries

Density

Probability of Success

Density

Probability of Success

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Part II: What Do We Want To Know?
Is there an effect?
Do women and men differ in working-memory capacity?
Do women and men differ in working-memory capacity?

- If significant difference, then go publish.
Simplified Example

Do women and men differ in working-memory capacity?

- If significant difference, then go publish.
- If no significant difference, then better luck next time.
Simplified Example

Do women and men differ in working-memory capacity?

- If significant difference, then go publish.
- If no significant difference, then better luck next time.
- Focus is on effects, not on invariances.
Invariances are at the heart of the physical sciences
Johannes Kepler (1571-1630)

Planets varied greatly in the speed & direction of their paths through the sky. Kepler extracted the invariants of celestial motion (e.g., elliptical orbits, equal area circumscribed in equal time).
Johannes Kepler (1571-1630)

- Planets varied greatly in the speed & direction of their paths through the sky.
Johannes Kepler (1571-1630)

- Planets varied greatly in the speed & direction of their paths through the sky.

- Kepler extracted the invariants of celestial motion (e.g., elliptical orbits, equal area circumscribed in equal time).
Invariances At The Heart of Science

▶ Conservation Laws: e.g., $F = MA$ implies $F_1 M_1 = F_2 M_2$ in a common gravitational field.

▶ In genetics, adenine binds to thymine, guanine binds to cytosine, across all DNA in all species.

▶ In chemistry, mechanisms of covalent bonding are the same across all atoms.
Invariances At The Heart of Science

- Conservation Laws: e.g., $F = MA$ implies $\frac{F_1}{M_1} = \frac{F_2}{M_2}$ in a common gravitational field.

In genetics, adenine binds to thymine, guanine binds to cytosine, across all DNA in all species.

In chemistry, mechanisms of covalent bonding are the same across all atoms.
Conservation Laws: e.g., $F = MA$ implies $\frac{F_1}{M_1} = \frac{F_2}{M_2}$ in a common gravitational field.

In genetics, adenine binds to thymine, guanine binds to cytosine, across all DNA in all species.
Invariances At The Heart of Science

- Conservation Laws: e.g., $F = MA$ implies $\frac{F_1}{M_1} = \frac{F_2}{M_2}$ in a common gravitational field.
- In genetics, adenine binds to thymine, guanine binds to cytosine, across all DNA in all species.
- In chemistry, mechanisms of covalent bonding are the same across all atoms.
Hot-Hand Phenomena

- Gilovich, Vallone, and Tversky (1985)
- Hot Hand: The outcome of a shot attempt in basketball is correlated with the outcome of previous attempts.
- Conclusion: No such correlation. Probability of success is invariant to previous history.
Invariance in Models: Top-Down Selection
Invariance in Models: Top-Down Selection

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Invariance in Models: Top-Down Selection

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Invariance in Models: Top-Down Selection

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Maybe there are no invariances (Cohen)
Commentary

- Maybe there are no invariances (Cohen)
- Example: Planets don’t follow ellipses to arbitrary precision
Maybe there are no invariances (Cohen)

Example: Planets don’t follow ellipses to arbitrary precision

Perhaps there are exact invariances. For example, imagining yourself winning the lottery may not improve to any degree whatsoever your chances of winning the lottery.
Maybe there are no invariances (Cohen)

Example: Planets don’t follow ellipses to arbitrary precision

Perhaps there are exact invariances. For example, imagining yourself winning the lottery may not improve to any degree whatsoever your chances of winning the lottery.

Models are not true or false, they just vary in usefulness
Maybe there are no invariances (Cohen)

Example: Planets don’t follow ellipses to arbitrary precision

Perhaps there are exact invariances. For example, imagining yourself winning the lottery may not improve to any degree whatsoever your chances of winning the lottery.

Models are not true or false, they just vary in usefulness

Constructive Challenge: The goal is to find theoretically-useful platonic invariances.
Lawfulness, regularity, invariance, and constraint correspond to null hypotheses.
The $p < .05$ Rule Is Unhelpful for Stating Evidence for Invariances
The $p < .05$ Rule Is Unhelpful for Stating Evidence for Invariances

- Wrong side of the null
The $p < .05$ Rule Is Unhelpful for Stating Evidence for Invariances

- Wrong side of the null
- Can’t state evidence for a null.
The $p < .05$ Rule Is Unhelpful for Stating Evidence for Invariances

- Wrong side of the null
- Can’t state evidence for a null.
- State a lack of evidence for an effect.
Researchers Often Accept The Null

HOW?
Researchers Often Accept The Null

HOW?

- Inspection: “Figure x shows the lack of effect”
Researchers Often Accept The Null

HOW?

▶ Inspection: “Figure x shows the lack of effect”
▶ The $p > .05$ Rule
Researchers Often Accept The Null

HOW?

► Inspection: “Figure x shows the lack of effect”
► The $p > .05$ Rule
► $p < .05$, but the effect is inconsequentially small
Researchers Often Accept The Null

HOW?

- Inspection: “Figure x shows the lack of effect”
- The $p > .05$ Rule
- $p < .05$, but the effect is inconsequentially small
- Unprincipled
Suppose it is theoretically meaningful to show effects.
Suppose it is theoretically meaningful to show effects. Shouldn’t we use the $p < .05$ Rule then? **No!**
Consistency

Consistency refers to a method's behavior in the large-sample limit. A consistent method is one that leads to the right decision always in the large-sample limit. It is a minimal obligation for the interpretation of frequentist probability. Significance Testing is not consistent:

- If there is a true effect, then in the large-sample limit, \( Pr(\text{reject}) = 1 \)
- BUT, if there is no effect, then \( Pr(\text{reject}) = \alpha = 0.05 \), even in the large-sample limit.
A consistent method is one that leads to the right decision always in the large-sample limit.
Consistency

- A consistent method is one that leads to the right decision always in the large-sample limit.
- It is a minimal obligation for the interpretation of frequentist probability.

Significance Testing is not consistent:
- If there is a true effect, then in the large-sample limit, \( \Pr(\text{reject}) = 1 \)
- BUT, if there is no effect, then \( \Pr(\text{reject}) = \alpha \approx 0.05 \), even in the large-sample limit.
Consistency

- A consistent method is one that leads to the right decision always in the large-sample limit.
- It is a minimal obligation for the interpretation of frequentist probability.
- Significance Testing is not consistent:

  - If there is a true effect, then in the large-sample limit, \( \Pr(\text{reject}) = 1 \).
  - BUT, if there is no effect, then \( \Pr(\text{reject}) = \alpha = 0.05 \), even in the large-sample limit.
Consistency

- A consistent method is one that leads to the right decision always in the large-sample limit.
- It is a minimal obligation for the interpretation of frequentist probability.
- Significance Testing is not consistent:
  - If there is a true effect, then in the large-sample limit, $\Pr(\text{reject}) = 1$
Consistency

- A consistent method is one that leads to the right decision always in the large-sample limit.
- It is a minimal obligation for the interpretation of frequentist probability.
- Significance Testing is not consistent:
  - If there is a true effect, then in the large-sample limit, \( \Pr(\text{reject}) = 1 \)
  - BUT, if there is no effect, then \( \Pr(\text{reject}) = \alpha = .05 \), even in the large-sample limit.
Consistent Frequentist Inference

Usual: \( \alpha(N) = 0.05 \), \( N \) is sample size.

To Be Consistent: \( \alpha(N) \) must go to 0 as \( N \) increases.

Idea: \( \alpha(N) = \min(0.05, \beta(N)) \), where \( \beta(N) \) is Type II error rate, or 1-Power.

Consistent because as \( N \) increases \( \beta(N) \) goes to zero and \( \alpha(N) \) goes to zero too.

To compute \( \beta(N) \), need to specify an alternative (Pay For Lunch).

\( \delta \), effect size, is 0.4.
Consistent Frequentist Inference

- Usual: \( \alpha(N) = 0.05 \), \( N \) is sample size

- To be consistent: \( \alpha(N) \) must go to 0 as \( N \) increases

- Idea: \( \alpha(N) = \min(0.05, \beta(N)) \), where \( \beta(N) \) is Type II error rate, or 1-Power

- Consistent because as \( N \) increases \( \beta(N) \) goes to zero and \( \alpha(N) \) goes to zero too.

- To compute \( \beta(N) \), need to specify an alternative (Pay For Lunch).

- \( \delta \), effect size, is .4
Consistent Frequentist Inference

- Usual: \( \alpha(N) = .05 \), \( N \) is sample size
- To Be Consistent: \( \alpha(N) \) must go to 0 as \( N \) increases
Consistent Frequentist Inference

▷ Usual: $\alpha(N) = .05$, $N$ is sample size
▷ To Be Consistent: $\alpha(N)$ must go to 0 as $N$ increases
▷ Idea: $\alpha(N) = \min(.05, \beta(N))$, where $\beta(N)$ is Type II error rate, or 1-Power
Consistent Frequentist Inference

- **Usual:** $\alpha(N) = 0.05$, $N$ is sample size
- **To Be Consistent:** $\alpha(N)$ must go to 0 as $N$ increases
- **Idea:** $\alpha(N) = \min(0.05, \beta(N))$, where $\beta(N)$ is Type II error rate, or 1-Power
- **Consistent because** as $N$ increases $\beta(N)$ goes to zero and $\alpha(N)$ goes to zero too.

To compute $\beta(N)$, need to specify an alternative (Pay For Lunch).

$\delta$, effect size, is 0.4
Consistent Frequentist Inference

- Usual: $\alpha(N) = .05$, $N$ is sample size
- To Be Consistent: $\alpha(N)$ must go to 0 as $N$ increases
- Idea: $\alpha(N) = \min(.05, \beta(N))$, where $\beta(N)$ is Type II error rate, or 1-Power
- Consistent because as $N$ increases $\beta(N)$ goes to zero and $\alpha(N)$ goes to zero too.
- To compute $\beta(N)$, need to specify an alternative (Pay For Lunch).
Consistent Frequentist Inference

- Usual: $\alpha(N) = .05$, $N$ is sample size
- To Be Consistent: $\alpha(N)$ must go to 0 as $N$ increases
- Idea: $\alpha(N) = \min(.05, \beta(N))$, where $\beta(N)$ is Type II error rate, or 1-Power
- Consistent because as $N$ increases $\beta(N)$ goes to zero and $\alpha(N)$ goes to zero too.
- To compute $\beta(N)$, need to specify an alternative (Pay For Lunch).
- $\delta$, effect size, is .4
A Small Change To Make Significance Testing Consistent

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
A Small Change To Make Significance Testing Consistent

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
A Small Change To Make Significance Testing Consistent

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>0.0</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed Effect Size</td>
<td>5</td>
<td>10</td>
<td>20</td>
<td>50</td>
<td>100</td>
<td>200</td>
</tr>
</tbody>
</table>

\[ \alpha = 0.05 \]
\[ \alpha = \min(0.05, \beta) \]
\[ \alpha = \beta/5 \]
Tail Whispering

Does whispering "tails" to coins increase the odds of tails?

527 out of 1000 flips, \( p < 0.05 \)

Implication: Alternative is substantially more likely than the null.

Let \( q \) be true probability of a tail

\( H_0: q = 0.5 \)

\( H_1: q > 0.5 \)
Does whispering “tails” to coins increase the odds of tails?
Tail Whispering

- Does whispering “tails” to coins increase the odds of tails?
- 527 out of 1000 flips, $p < .05$
Tail Whispering

- Does whispering “tails” to coins increase the odds of tails?
- 527 out of 1000 flips, \( p < .05 \)
- Implication: Alternative is substantially more likely than the null.
Tail Whispering

▶ Does whispering “tails” to coins increase the odds of tails?
▶ 527 out of 1000 flips, $p < .05$
▶ Implication: Alternative is substantially more likely than the null.
▶ Let $q$ be true probability of a tail
  $H_0$: $q = .5$
  $H_1$: $q > .5$. 
If $q = .5$
Tail Whispering

▷ Does whispering “tails” to coins increase the odds of tails?
▷ 527 out of 1000 flips $p \approx .05$
▷ If we reject the null, then for what alternative?
Does whispering “tails” to coins increase the odds of tails?

527 out of 1000 flips $p \approx .05$

If we reject the null, then for what alternative?

Let’s look at $q = .527$, most favorable alternative
$q = .5 \text{ vs. } q = .527$
Overstating Evidence Against The Null

- Does whispering “tails” to coins increase the odds of tails?
- 527 out of 1000 flips $p \approx .05$
  1. $q = .5$ vs. $q = .527$, Evidence: 4.3 to 1 for alternative
     A researcher should bet $4 to win $1, but not $5 to win $1.

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Does whispering “tails” to coins increase the odds of tails?

527 out of 1000 flips $p \approx 0.05$

1. $q = 0.5$ vs. $q = 0.527$, Evidence: 4.3 to 1 for alternative
   A researcher should bet $4$ to win $1$, but not $5$ to win $1$.

2. $q = 0.5$ vs. $q = 0.55$, Evidence: 1.5 to 1 for alternative
Overstating Evidence Against The Null

- Does whispering “tails” to coins increase the odds of tails?
- 527 out of 1000 flips $p \approx .05$
  1. $q = .5$ vs. $q = .527$, Evidence: 4.3 to 1 for alternative
     A researcher should bet $4 to win $1, but not $5 to win $1.
  2. $q = .5$ vs. $q = .55$, Evidence: 1.5 to 1 for alternative
  3. $q = .5$ vs. $.5 < q < .6$, Evidence: 6.2 to 1 for null

No alternative is much better than the null.
Many reasonable ones are less concordant with the data than the null.
Does whispering “tails” to coins increase the odds of tails?

527 out of 1000 flips $p \approx .05$

1. $q = .5$ vs. $q = .527$, Evidence: 4.3 to 1 for alternative
   A researcher should bet $4 to win $1, but not $5 to win $1.$
2. $q = .5$ vs. $q = .55$, Evidence: 1.5 to 1 for alternative
3. $q = .5$ vs $q < .6$, Evidence: 6.2 to 1 for null

No alternative is much better than the null
Overstating Evidence Against The Null

- Does whispering “tails” to coins increase the odds of tails?
- 527 out of 1000 flips $p \approx .05$
  1. $q = .5$ vs. $q = .527$, Evidence: 4.3 to 1 for alternative
     A researcher should bet $4 to win $1, but not $5 to win $1.
  2. $q = .5$ vs. $q = .55$, Evidence: 1.5 to 1 for alternative
  3. $q = .5$ vs. $5 < q < .6$, Evidence: 6.2 to 1 for null
- No alternative is much better than the null
- Many reasonable ones are less concordant with the data than the null.
Overstating-Evidence Argument
Overstating-Evidence Argument

Eff.Size=.2, N=50
Eff.Size=.2, N=500
Null

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Overstating-Evidence Argument

Eff.Size=.2, N=50
Null

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Overstating-Evidence Argument

Eff.Size=.2, N=500

Null

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
The \( p < .05 \) Rule is unprincipled. Overstates evidence against null.
The $p < .05$ Rule is unprincipled. Overstates evidence against null.

Principled frequentist inference requires paying for lunch.
Part IV: Bayesian Inference
Bayesians place beliefs on models

Prior Beliefs:

\[ P(M_0) \quad P(M_1) \]

Posterior Beliefs:

\[ P(M_0|\text{Data}) \quad P(M_1|\text{Data}) \]

Use Bayes Theorem to Update Beliefs:

\[ P(M_0|\text{Data}) P(M_1|\text{Data}) = P(\text{Data}|M_0) P(\text{Data}|M_1) \times P(M_0) P(M_1) \]
Bayesians place beliefs on models

- Odds

Prior Beliefs:
- $P(M_0)$
- $P(M_1)$

Posterior Beliefs:
- $P(M_0|\text{Data})$
- $P(M_1|\text{Data})$

Use Bayes' Theorem to update beliefs:

$$P(M_0|\text{Data})P(M_1|\text{Data}) = P(\text{Data}|M_0)P(\text{Data}|M_1) \times P(M_0)P(M_1)$$
Bayesians place beliefs on models

- Odds
- Prior Beliefs:

\[
\frac{P(M_0)}{P(M_1)}
\]
Bayesians place beliefs on models

- Odds
- Prior Beliefs: \[
\frac{P(M_0)}{P(M_1)}
\]
- Posterior Beliefs: \[
\frac{P(M_0|\text{Data})}{P(M_1|\text{Data})}
\]
Bayesians place beliefs on models

- Odds
- Prior Beliefs:
  \[ \frac{P(M_0)}{P(M_1)} \]
- Posterior Beliefs:
  \[ \frac{P(M_0|\text{Data})}{P(M_1|\text{Data})} \]
- Use Bayes Theorem to Update Beliefs:
  \[ \frac{P(M_0|\text{Data})}{P(M_1|\text{Data})} = \frac{P(\text{Data}|M_0)}{P(\text{Data}|M_1)} \times \frac{P(M_0)}{P(M_1)} \]
Bayes Factor: A Fully Bayesian Approach

Use Bayes Theorem to Update Beliefs:

\[ \frac{P(M_0|\text{Data})}{P(M_1|\text{Data})} = \frac{P(\text{Data}|M_0)}{P(\text{Data}|M_1)} \times \frac{P(M_0)}{P(M_1)} \]
Bayes Factor: A Fully Bayesian Approach

- Use Bayes Theorem to Update Beliefs:

\[
\frac{P(M_0|\text{Data})}{P(M_1|\text{Data})} = \frac{P(\text{Data}|M_0)}{P(\text{Data}|M_1)} \times \frac{P(M_0)}{P(M_1)}
\]

- Bayes factors are the updating factor

\[
B_{01} = \frac{P(\text{Data}|M_0)}{P(\text{Data}|M_1)}.
\]
Bayes Factors Change In Beliefs

\[
\frac{P(M_0|\text{Data})}{P(M_1|\text{Data})} = B_{01} \times \frac{P(M_0)}{P(M_1)}
\]
Bayes Factors Change In Beliefs

\[
\frac{P(M_0|\text{Data})}{P(M_1|\text{Data})} = B_{01} \times \frac{P(M_0)}{P(M_1)}
\]

- Bayes factor is the change in belief due to the data
- Bayes factor is independent of prior odds
- Ratios can be interpreted in terms of wagers.
- Prior odds are a good place to add value.
- \( B_{10} = \frac{1}{B_{01}} \)
Bayes Factors Change In Beliefs

\[
\frac{P(M_0|\text{Data})}{P(M_1|\text{Data})} = B_{01} \times \frac{P(M_0)}{P(M_1)}
\]

- Bayes factor is the change in belief due to the data
- Bayes factor is independent of prior odds
Bayes Factors Change In Beliefs

\[
\frac{P(M_0|\text{Data})}{P(M_1|\text{Data})} = B_{01} \times \frac{P(M_0)}{P(M_1)}
\]

- Bayes factor is the change in belief due to the data
- Bayes factor is independent of prior odds
- Ratios can be interpreted in terms of wagers.
Bayes Factors Change In Beliefs

\[
\frac{P(M_0|\text{Data})}{P(M_1|\text{Data})} = B_{01} \times \frac{P(M_0)}{P(M_1)}
\]

- Bayes factor is the change in belief due to the data
- Bayes factor is independent of prior odds
- Ratios can be interpreted in terms of wagers.
- Prior odds are a good place to add value.
Bayes Factors Change In Beliefs

\[
\frac{P(M_0|\text{Data})}{P(M_1|\text{Data})} = B_{01} \times \frac{P(M_0)}{P(M_1)}
\]

- Bayes factor is the change in belief due to the data
- Bayes factor is independent of prior odds
- Ratios can be interpreted in terms of wagers.
- Prior odds are a good place to add value.
- \( B_{10} = 1/B_{01} \)
Bayes Factor & Priors on Parameters

What is $P(\text{Data}|\mathcal{M})$?
Bayes Factor & Priors on Parameters

What is $P(\text{Data}|M)$?

- Models have parameters: $\theta$
What is $P(\text{Data}|\mathcal{M})$?

- Models have parameters: $\theta$
- $P(\text{Data}|\theta, \mathcal{M})$ is easy. Likelihood, $L(\theta)$. 
What is $P(\text{Data}|\mathcal{M})$?

- Models have parameters: $\theta$
- $P(\text{Data}|\theta, \mathcal{M})$ is easy. Likelihood, $L(\theta)$.
- If we define the null and alternative as specific parameter values $\theta_0$ and $\theta_1$, then

$$B_{01} = \frac{P(\text{Data}|\mathcal{M}_0)}{P(\text{Data}|\mathcal{M}_1)} = \frac{L(\theta_0)}{L(\theta_1)}$$
Parameters can take on more than a point:

\[ B_{01} = \frac{P(\text{Data}|M_0)}{P(\text{Data}|M_1)} = \frac{\int \theta_0 L_0(\theta_0)f_0(\theta_0)d\theta_0}{\int \theta_1 L_1(\theta_1)f_1(\theta_1)d\theta_1} \]

- Marginal or averaged likelihood of a model with respect to the prior \( f_0(\theta_0) \) or \( f_1(\theta_1) \).
Bayes Factor & Priors on Parameters

Parameters can take on more than a point:

\[
B_{01} = \frac{P(\text{Data}|M_0)}{P(\text{Data}|M_1)} = \frac{\int_\theta L_0(\theta_0)f_0(\theta_0)d\theta_0}{\int_\theta L_1(\theta_1)f_1(\theta_1)d\theta_1}
\]

- Marginal or averaged likelihood of a model with respect to the prior \(f_0(\theta_0)\) or \(f_1(\theta_1)\).
- If priors are too broad, then the average will include many parameter values with low likelihood. Lower average likelihood, natural penalty for complexity.
Bayes Factor & Priors on Parameters

Parameters can take on more than a point:

\[
B_{01} = \frac{P(\text{Data} | M_0)}{P(\text{Data} | M_1)} = \frac{\int_{\theta} L_0(\theta_0)f_0(\theta_0)d\theta_0}{\int_{\theta} L_1(\theta_1)f_1(\theta_1)d\theta_1}
\]

- Marginal or averaged likelihood of a model with respect to the prior \(f_0(\theta_0)\) or \(f_1(\theta_1)\).
- If priors are too broad, then the average will include many parameter values with low likelihood. Lower average likelihood, natural penalty for complexity.
- Specification of priors is paying for lunch.
Simple Problem:

- Bunch of Observations From a Single Condition ($N = 50$)
Simple Problem:

- Bunch of Observations From a Single Condition \((N = 50)\)
- Define Priors on Effect Size, \(\delta\)
Bayes Factor & Priors on Parameters

Simple Problem:
- Bunch of Observations From a Single Condition \((N = 50)\)
- Define Priors on Effect Size, \(\delta\)
- Null, \(M_0: \delta = 0\)
Bayes Factor & Priors on Parameters

Simple Problem:

- Bunch of Observations From a Single Condition ($N = 50$)
- Define Priors on Effect Size, $\delta$
- Null, $M_0$: $\delta = 0$
- Alternative, $M_1$ (no free lunch)?
Bayes Factor & Priors on Parameters

Simple Problem:

- Bunch of Observations From a Single Condition \((N = 50)\)
- Define Priors on Effect Size, \(\delta\)
- Null, \(M_0\): \(\delta = 0\)
- Alternative, \(M_1\) (no free lunch)?
- You Know A Lot. Is \(\delta = 1000\) reasonable? How about \(\delta = .001\)?
Simple Problem:

- Bunch of Observations From a Single Condition ($N = 50$)
- Define Priors on Effect Size, $\delta$
- Null, $M_0$: $\delta = 0$
- Alternative, $M_1$ (no free lunch)?
- You Know A Lot. Is $\delta = 1000$ reasonable? How about $\delta = .001$?
- Show some alternatives, then show you the BF for them.
Bayes Factor & Priors on Parameters

Implausibly Wide: Scale=10.0

Effect Size

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Bayes Factor & Priors on Parameters

Effect Size

Narrow: Scale=.2

Jeffrey N. Rouder

Bayesian Inference in Psychology: A Workshop
Bayes Factor & Priors on Parameters

Implausibly Narrow: Scale=.02

Effect Size

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Some Properties of Bayes Factors
Consequence #1: Sample Size Considerations

![Graph showing critical t-values for different sample sizes and prior strengths.](image)

- **Critical t-values**
  - $B_{10} = 10$
  - $B_{10} = 3$
  - $B_{10} = 1$
  - $p$-value = .05

---

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Consequence #2: Calibration in Real Data

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Consequence #3: Respects Resolution of Data

Bayesian Inference in Psychology: A Workshop
Consequence #4: Collect data as you wish

Optional Stopping, collect data until the results are clear

Seems reasonable, save resources when results are clear and spend them when they are not.

Violates assumptions of significance tests and is to be avoided.

What about Bayes Factor? Eminent 20th century statisticians, Lindley, Jeffreys, Savage, said the interpretation of Bayes factor and posterior odds was valid irrespective of the stopping rule.

Yu, Yu, Sprenger, Thomas, and Dougherty (in press) write, "Bayesian analysis is not the magic elixir it is sometimes made out to be. One cannot simply apply Bayesian statistics to any old dataset and be confident that the outcome is free of bias..... [Bayes factors] are not interpretable if researchers used a data-dependent stopping heuristic."
Consequence #4: Collect data as you wish

Optional Stopping, collect data until the results are clear

▶ Seems reasonable, save resources when results are clear and spend them when they are not.
Consequence #4: Collect data as you wish

Optional Stopping, collect data until the results are clear

▶ Seems reasonable, save resources when results are clear and spend them when they are not.
▶ Violates assumptions of significance tests and is to be avoided.
Consequence #4: Collect data as you wish

Optional Stopping, collect data until the results are clear

▶ Seems reasonable, save resources when results are clear and spend them when they are not.
▶ Violates assumptions of significance tests and is to be avoided.
▶ What about Bayes Factor? Eminent 20th century statisticians, Lindley, Jeffreys, Savage, said the interpretation of Bayes factor and posterior odds was valid irrespective of the stopping rule.

Yu, Yu, Sprenger, Thomas, and Dougherty (in press) write, "Bayesian analysis is not the magic elixir it is sometimes made out to be. One cannot simply apply Bayesian statistics to any old dataset and be confident that the outcome is free of bias..... [Bayes factors] are not interpretable if researchers used a data-dependent stopping heuristic."
Consequence #4: Collect data as you wish

Optional Stopping, collect data until the results are clear

- Seems reasonable, save resources when results are clear and spend them when they are not.
- Violates assumptions of significance tests and is to be avoided.
- What about Bayes Factor? Eminent 20th century statisticians, Lindley, Jeffreys, Savage, said the interpretation of Bayes factor and posterior odds was valid irrespective of the stopping rule.

- Yu, Yu, Sprenger, Thomas, and Dougherty (in press) write, “Bayesian analysis is not the magic elixir it is sometimes made out to be. One cannot simply apply Bayesian statistics to any old dataset and be confident that the outcome is free of bias..... [Bayes factors] are not interpretable if researchers used a data-dependent stopping heuristic.”
Consequence #4: Collect data as you wish

Do posterior odds *mean what they say*:
Consequence #4: Collect data as you wish

Do posterior odds *mean what they say*:

- Generate 20,000 experiments of 10 samples each from null
- Generate 20,000 experiments of 10 samples each from $\delta = 0.4$
- Compute posterior odds for all 40,000 experiments.
- Show you how these simulations show that posterior odds *mean what they say*, they accurately report the probability that each replicate is from one or the other hypothesis.
Consequence #4: Collect data as you wish

Do posterior odds *mean what they say*:

- Generate 20,000 experiments of 10 samples each from null
- Generate 20,000 experiments of 10 samples each from $\delta = .4$
Consequence #4: Collect data as you wish

Do posterior odds *mean what they say*:

- Generate 20,000 experiments of 10 samples each from null
- Generate 20,000 experiments of 10 samples each from $\delta = .4$
- Compute posterior odds for all 40,000 experiments.
Consequence #4: Collect data as you wish

Do posterior odds *mean what they say*:

- Generate 20,000 experiments of 10 samples each from null
- Generate 20,000 experiments of 10 samples each from $\delta = 0.4$
- Compute posterior odds for all 40,000 experiments.
- Show you how these simulations show that posterior odds mean what they say, they accurately report the probability that each replicate is from one or the other hypothesis.
Consequence #4: Collect data as you wish

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Consequence #4: Collect data as you wish
Consequence #4: Collect data as you wish

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Consequence #4: Collect data as you wish

Nominal Posterior Odds

Frequency

0.01 0.1 1 10 100

3000 1000 1000 3000

Jeffrey N. Rouder

Bayesian Inference in Psychology: A Workshop
Consequence #4: Collect data as you wish
Consequence #4: Collect data as you wish
Consequence #4: Collect data as you wish

What about with optional stopping?

- Recomputed posterior odds after every observation
- Stop when posterior odds is 10-1 in favor of one of the hypotheses OR \( n = 22 \).
Consequence #4: Collect data as you wish
Consequence #4: Collect data as you wish

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Bayes Factor

A Most Excellent Means of Model Comparison:
A Most Excellent Means of Model Comparison:

- **Principle:** Update Beliefs Rationally in Light of Data
A Most Excellent Means of Model Comparison:

▶ **Principle:** Update Beliefs Rationally in Light of Data
▶ State evidence, avoid decisions
Bayes Factor

A Most Excellent Means of Model Comparison:

- **Principle**: Update Beliefs Rationally in Light of Data
- State evidence, avoid decisions
- Evidence is in odds, natural language in terms of wagers.
Bayes Factor

A Most Excellent Means of Model Comparison:

- **Principle: Update Beliefs Rationally in Light of Data**
- State evidence, avoid decisions
- Evidence is in odds, natural language in terms of wagers.
- State evidence for or against invariances, as dictated by data
Bayes Factor

A Most Excellent Means of Model Comparison:

- **Principle: Update Beliefs Rationally in Light of Data**
- State evidence, avoid decisions
- Evidence is in odds, natural language in terms of wagers.
- State evidence for or against invariances, as dictated by data
- Context through prior odds
A Most Excellent Means of Model Comparison:

- **Principle: Update Beliefs Rationally in Light of Data**
- State evidence, avoid decisions
- Evidence is in odds, natural language in terms of wagers.
- State evidence for or against invariances, as dictated by data
- Context through prior odds
- Natural penalty for model complexity.
Bayes Factor

A Most Excellent Means of Model Comparison:

- **Principle: Update Beliefs Rationally in Light of Data**
- State evidence, avoid decisions
- Evidence is in odds, natural language in terms of wagers.
- State evidence for or against invariances, as dictated by data
- Context through prior odds
- Natural penalty for model complexity.
- Respects resolution of data; best account of data.
Bayes Factor

A Most Excellent Means of Model Comparison:

- **Principle: Update Beliefs Rationally in Light of Data**
- State evidence, avoid decisions
- Evidence is in odds, natural language in terms of wagers.
- State evidence for or against invariances, as dictated by data
- Context through prior odds
- Natural penalty for model complexity.
- Respects resolution of data; best account of data.
- Collect/monitor data as you wish
Other “Bayesian” Competitors

- Some analysts recommend using Bayesian estimation with exceedingly large priors so they do not have to pay for lunch.
- Focus on Kruschke as his approach is now the recommended one for The Psychonomic Society.
Other “Bayesian” Competitors

Credible Interval Logic Contradicts Bayes Rule

I computed the posterior from $N = 50$ and $t = 2.5$.

Prior scale on effect size was 1000.

Posterior belief effect size is around zero: .015

Q. What were the a priori beliefs that effect size is around zero? .00004

Q. By what factor are these beliefs updated? 360-to-1 in favor of the null!

Why? The alternative is so diffuse that these data are better described by the null than by the alternative.
Other “Bayesian” Competitors

Credible Interval Logic Contradicts Bayes Rule

- I computed the posterior from \( N = 50 \) and \( t = 2.5 \)

Prior scale on effect size was 1000.

Posterior belief effect size is around zero: .015

Q. What were the \textit{a priori} beliefs that effect size is around zero? .00004

Q. By what factor are these beliefs updated? 360-to-1 in favor of the null!

Why? The alternative is so diffuse that these data are better described by the null than by the alternative.

Credible Interval Logic Contradicts Bayes Rule

- I computed the posterior from $N = 50$ and $t = 2.5$
- Prior scale on effect size was 1000.

Q. What were the a priori beliefs that effect size is around zero? .00004

Q. By what factor are these beliefs updated? 360-to-1 in favor of the null!

Why? The alternative is so diffuse that these data are better described by the null than by the alternative.
Credible Interval Logic Contradicts Bayes Rule

- I computed the posterior from $N = 50$ and $t = 2.5$
- Prior scale on effect size was 1000.
- Posterior belief effect size is around zero: .015

Q. What were the a priori beliefs that effect size is around zero? .00004
Q. By what factor are these beliefs updated? 360-to-1 in favor of the null!

Why? The alternative is so diffuse that these data are better described by the null than by the alternative.
Credible Interval Logic Contradicts Bayes Rule

- I computed the posterior from $N = 50$ and $t = 2.5$.
- Prior scale on effect size was 1000.
- Posterior belief effect size is around zero: .015
- Q. What were the *a priori* beliefs that effect size is around zero? .00004

Why? The alternative is so diffuse that these data are better described by the null than by the alternative.
Other “Bayesian” Competitors

Credible Interval Logic Contradicts Bayes Rule

- I computed the posterior from $N = 50$ and $t = 2.5$.
- Prior scale on effect size was 1000.
- Posterior belief effect size is around zero: .015
- Q. What were the *a priori* beliefs that effect size is around zero? .00004
- Q. By what factor are these beliefs updated? 360-to-1 in favor of the null!
Credible Interval Logic Contradicts Bayes Rule

- I computed the posterior from \( N = 50 \) and \( t = 2.5 \).
- Prior scale on effect size was 1000.
- Posterior belief effect size is around zero: .015
- Q. What were the \textit{a priori} beliefs that effect size is around zero? .00004
- Q. By what factor are these beliefs updated? 360-to-1 in favor of the null!
- Why? The alternative is so diffuse that these data are better described by the null than by the alternative.
Other “Bayesian” Competitors

Credible Interval Logic Contradicts Bayes Rule

- Take the argument to the limit. Prior scale could be $\infty$ (flat or non informative prior).

"Now Bayes rule is a very attractive way of reasoning, and fun to use, but using Bayes rule doesn’t make one a Bayesian. Always using Bayes rule does..."
Other “Bayesian” Competitors

Credible Interval Logic Contradicts Bayes Rule

▶ Take the argument to the limit. Prior scale could be $\infty$ (flat or non informative prior).

▶ Posterior is about the same; posterior belief in a near-zero effect size: .015

Noninformative priors are so diffuse that all data is more compatible with the infinitely simpler null.

Any method that allows for inference with non informative alternatives is in violation of Bayes Rule

"Now Bayes rule is a very attractive way of reasoning, and fun to use, but using Bayes rule doesn’t make one a Bayesian. Always using Bayes rule does..."

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Other “Bayesian” Competitors

Credible Interval Logic Contradicts Bayes Rule

▶ Take the argument to the limit. Prior scale could be $\infty$ (flat or non-informative prior).

▶ Posterior is about the same; posterior belief in a near-zero effect size: 0.015

▶ Prior belief in a near-zero effect size: 0

Updating factor: infinitely large in favor of the null.

Non-informative priors are so diffuse that all data is more compatible with the infinitely simpler null.

Any method that allows for inference with non-informative alternatives is in violation of Bayes Rule.

“Now Bayes rule is a very attractive way of reasoning, and fun to use, but using Bayes rule doesn’t make one a Bayesian. Always using Bayes rule does...”
Other “Bayesian” Competitors

Credible Interval Logic Contradicts Bayes Rule

▶ Take the argument to the limit. Prior scale could be \( \infty \) (flat or non informative prior).
▶ Posterior is about the same; posterior belief in a near-zero effect size: .015
▶ Prior belief in a near-zero effect size: 0
▶ Updating factor: infinitely large in favor of the null.
Other “Bayesian” Competitors

Credible Interval Logic Contradicts Bayes Rule

▶ Take the argument to the limit. Prior scale could be $\infty$ (flat or non informative prior).
▶ Posterior is about the same; posterior belief in a near-zero effect size: .015
▶ Prior belief in a near-zero effect size: 0
▶ Updating factor: infinitely large in favor of the null.
▶ Noninformative priors are so diffuse that all data is more compatible with the infinitely simpler null.
Other “Bayesian” Competitors

Credible Interval Logic Contradicts Bayes Rule

- Take the argument to the limit. Prior scale could be $\infty$ (flat or non-informative prior).
- Posterior is about the same; posterior belief in a near-zero effect size: 0.015
- Prior belief in a near-zero effect size: 0
- Updating factor: infinitely large in favor of the null.
- Noninformative priors are so diffuse that all data is more compatible with the infinitely simpler null.
- Any method that allows for inference with non-informative alternatives is in violation of Bayes Rule
Credible Interval Logic Contradicts Bayes Rule

- Take the argument to the limit. Prior scale could be $\infty$ (flat or non informative prior).
- Posterior is about the same; posterior belief in a near-zero effect size: 0.015
- Prior belief in a near-zero effect size: 0
- Updating factor: infinitely large in favor of the null.
- Noninformative priors are so diffuse that all data is more compatible with the infinitely simpler null.
- Any method that allows for inference with non informative alternatives is in violation of Bayes Rule
- “Now Bayes rule is a very attractive way of reasoning, and fun to use, but using Bayes rule doesn’t make one a Bayesian. *Always* using Bayes rule does...”
Pay For Lunch...

Frequentist:

Consistent Inference is Possible When the Null and Alternative Are Well Specified.

Principled Frequentist Analysts Pay For Lunch.

Bayesian:

Proper Updating Is Possible When the Null and Alternative Are Well Specified.

Principled Bayesian Analysts Pay For Lunch.

Jeffrey N. Rouder

Bayesian Inference in Psychology: A Workshop
Pay For Lunch...

- **Frequentist:**
  - Consistent Inference is Possible When the Null and Alternative Are Well Specified.
  - Principled Frequentist Analysts Pay For Lunch.
Pay For Lunch...

- **Frequentist:**
  - Consistent Inference is Possible When the Null and Alternative Are Well Specified.
  - Principled Frequentist Analysts Pay For Lunch.

- **Bayesian:**
  - Proper Updating Is Possible When the Null and Alternative Are Well Specified.
  - Principled Bayesian Analysts Pay For Lunch.
The free-lunch is mindless. Usage promotes a shallow if not cavalier attitude toward analysis.

With the free lunch, there is a mental distance between what we know and what we can show.

It is in this mental distance between what we know and what we can show that we may compromise our methodology.
The free-lunch is mindless. Usage promotes a shallow if not cavalier attitude toward analysis.
Does the Free Lunch Make Us Ill?

- The free-lunch is mindless. Usage promotes a shallow if not cavalier attitude toward analysis.
- With the free lunch, there is a mental distance between what we know and what we can show.
The free-lunch is mindless. Usage promotes a shallow if not cavalier attitude toward analysis.

With the free lunch, there is a mental distance between what we know and what we can show.

It is in this mental distance between what we know and what we can show that we may compromise our methodology.
The Psychonomic Society is a great organization where exceptional people give freely and generously of their time and skills. The guidelines committee worked hard and I appreciate their efforts.
The Psychonomic Society is a great organization where exceptional people give freely and generously of their time and skills. The guidelines committee worked hard and I appreciate their efforts.

Many of the current guidelines are reasonable when taken individually.
The Psychonomic Society is a great organization where exceptional people give freely and generously of their time and skills. The guidelines committee worked hard and I appreciate their efforts.

Many of the current guidelines are reasonable when taken individually.

Unaddressed is the free-lunch problem, which is a matter of principle.
The Psychonomic Society is a great organization where exceptional people give freely and generously of their time and skills. The guidelines committee worked hard and I appreciate their efforts.

Many of the current guidelines are reasonable when taken individually.

Unaddressed is the free-lunch problem, which is a matter of principle.

The guidelines may be summarized as, “You can eat the free lunch, just clean up your crumbs and leave the lunchroom looking nice.”
My Guidelines

▶ Good scientists do not believe that science is objective; good analysts do not believe that inference is objective.

▶ Good analysts add value by considering well-specified models that embed meaningful constraint. Good analysts state support for or against meaningful constraint by considering judiciously chosen alternatives.

▶ Good analysts are transparent in and responsible for their choices.

▶ Good readers have the responsibility for critically assessing the analyst's choice of models as well as for forming a personal opinion about the interpretation of evidence.
Good scientists do not believe that science is objective; good analysts do not believe that inference is objective.
My Guidelines

- Good scientists do not believe that science is objective; good analysts do not believe that inference is objective.
- Good analysts add value by considering well-specified models that embed meaningful constraint. Good analysts state support for or against meaningful constraint by considering judiciously chosen alternatives.
My Guidelines

▶ Good scientists do not believe that science is objective; good analysts do not believe that inference is objective.
▶ Good analysts add value by considering well-specified models that embed meaningful constraint. Good analysts state support for or against meaningful constraint by considering judiciously chosen alternatives.
▶ Good analysts are transparent in and responsible for their choices.
My Guidelines

▷ Good scientists do not believe that science is objective; good analysts do not believe that inference is objective.

▷ Good analysts add value by considering well-specified models that embed meaningful constraint. Good analysts state support for or against meaningful constraint by considering judiciously chosen alternatives.

▷ Good analysts are transparent in and responsible for their choices.

▷ Good readers have the responsibility for critically assessing the analyst’s choice of models as well as for forming a personal opinion about the interpretation of evidence.
Let's have a Q & A and then take a brief break.
Computing Bayes Factors In Common Designs

Two Methods: web applets, R package

Preliminary for web applet

Open up a browser of your choice (Safari, Explorer, Chrome, etc.)

Go to pcl.missouri.edu/bayesfactor

Minimize your browser window. Raise hand when done.

Preliminary for R Package

Get Version of R 3.0 or higher

http://cran.us.r-project.org

Start R

Let's Install BayesFactor Package:

install.packages("BayesFactor")

Let's Start Bayes Factor Package:

library("BayesFactor")

Raise hand when done
Computing Bayes Factors In Common Designs

- Two Methods: web applets, R package

  ▶ Open up a browser of your choice (Safari, Explorer, Chrome, etc.)
  ▶ Go to pcl.missouri.edu/bayesfactor
  ▶ Minimize your browser window. Raise hand when done.

  ▶ Get Version of R 3.0 or higher
  ▶ http://cran.us.r-project.org
  ▶ Start R
  ▶ Let's Install BayesFactor Package: install.packages("BayesFactor")
  ▶ Let's Start Bayes Factor Package: library("BayesFactor")
  ▶ Raise hand when done
Computing Bayes Factors In Common Designs

- Two Methods: web applets, R package
- Preliminary for web applet
Computing Bayes Factors In Common Designs

- Two Methods: web applets, R package
- Preliminary for web applet
  - Open up a browser of your choice (Safari, Explorer, Chrome, etc.)
  - Go to pcl.missouri.edu/bayesfactor
  - Minimize your browser window. Raise hand when done.
Computing Bayes Factors In Common Designs

- Two Methods: web applets, R package

- Preliminary for web applet
  - Open up a browser of your choice (Safari, Explorer, Chrome, etc.)
  - Go to pcl.missouri.edu/bayesfactor
  - Minimize your browser window. Raise hand when done.

- Preliminary for R Package

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Computing Bayes Factors In Common Designs

▶ Two Methods: web applets, R package

▶ Preliminary for web applet
  ▶ Open up a browser of your choice (Safari, Explorer, Chrome, etc.)
  ▶ Go to pcl.missouri.edu/bayesfactor
  ▶ Minimize your browser window. Raise hand when done.

▶ Preliminary for R Package
  ▶ Get Version of R 3.0 or higher
    http://cran.us.r-project.org
Computing Bayes Factors In Common Designs

- Two Methods: web applets, R package
- Preliminary for web applet
  - Open up a browser of your choice (Safari, Explorer, Chrome, etc.)
  - Go to pcl.missouri.edu/bayesfactor
  - Minimize your browser window. Raise hand when done.
- Preliminary for R Package
  - Get Version of R 3.0 or higher
    [http://cran.us.r-project.org](http://cran.us.r-project.org)
  - Start R
Computing Bayes Factors In Common Designs

▶ Two Methods: web applets, R package

▶ Preliminary for web applet
  ▶ Open up a browser of your choice (Safari, Explorer, Chrome, etc.)
  ▶ Go to pcl.missouri.edu/bayesfactor
  ▶ Minimize your browser window. Raise hand when done.

▶ Preliminary for R Package
  ▶ Get Version of R 3.0 or higher
    http://cran.us.r-project.org
  ▶ Start R
  ▶ Let’s Install BayesFactor Package:
    install.packages("BayesFactor")
Computing Bayes Factors In Common Designs

- Two Methods: web applets, R package
- Preliminary for web applet
  - Open up a browser of your choice (Safari, Explorer, Chrome, etc.)
  - Go to pcl.missouri.edu/bayesfactor
  - Minimize your browser window. Raise hand when done.

- Preliminary for R Package
  - Get Version of R 3.0 or higher
    http://cran.us.r-project.org
  - Start R
  - Let’s Install BayesFactor Package:
    install.packages("BayesFactor")
  - Let’s Start Bayes Factor Package:
    library("BayesFactor")
Computing Bayes Factors In Common Designs

- Two Methods: web applets, R package
- Preliminary for web applet
  - Open up a browser of your choice (Safari, Explorer, Chrome, etc.)
  - Go to pcl.missouri.edu/bayesfactor
  - Minimize your browser window. Raise hand when done.
- Preliminary for R Package
  - Get Version of R 3.0 or higher
    [Link](http://cran.us.r-project.org)
  - Start R
  - Let’s Install BayesFactor Package:
    ```r
    install.packages("BayesFactor")
    ```
  - Let’s Start Bayes Factor Package:
    ```r
    library("BayesFactor")
    ```
  - Raise hand when done
One Sample or Paired \( t \)-Test
One Sample or Paired $t$-Test

One Sample or Paired $t$-Test

- Maximize your browser window with the web applet.
One Sample or Paired $t$-Test

Ex.: Grider & Malberg (2010)

- Recognition memory, manipulate whether materials were emotionally negative or emotionally neutral
- Within-participant manipulation

Accuracy: 0.76 vs. 0.79 in favor of emotionally negative words.

Paired $t$-test: $t(79) = 2.03$, $p < .05$, evidence for an emotional word effect on memory.
One Sample or Paired \( t \)-Test

Ex.: Grider & Malberg (2010)

- Recognition memory, manipulate whether materials were emotionally negative or emotionally neutral
- Within-participant manipulation
- Accuracy: .76 vs. .79 in favor of emotionally negative words.

\[ t(79) = 2.03, p < .05, \] evidence for an emotional word effect on memory.
One Sample or Paired $t$-Test

Ex.: Grider & Malberg (2010)

- Recognition memory, manipulate whether materials were emotionally negative or emotionally neutral
- Within-participant manipulation
- Accuracy: .76 vs. .79 in favor of emotionally negative words.
- paired $t$-test: $t(79) = 2.03, p < .05$, evidence for an emotional word effect on memory.
One Sample or Paired $t$-Test

- Null: $\delta = 0$
One Sample or Paired $t$-Test

- Null: $\delta = 0$
- Alternative: Scale on Effect Size is .7
One Sample or Paired $t$-Test

Effect Size

Scale = .7

![Graph showing effect size distribution with a scale of 0.7.](image-url)
One Sample or Paired $t$-Test

- Enter data into one-sample applet. Don’t forget to set $r = .7$. 

$\text{Scaled JZS Bayes Factor} = 1.15187$

Evidence is equivocal for null and alternative

What if we tried to handpick a more effect-favorable scale, try $r = 2$.

$\text{Scaled JZS Bayes Factor} = 0.6050631$

$BF$ is 1.65-to-1 in favor of effect, scant evidence
One Sample or Paired $t$-Test

- Enter data into one-sample applet. Don’t forget to set $r = .7$.
- JZS Bayes Factor, (“Jeffreys, Zellner, Siow”)
One Sample or Paired $t$-Test

- Enter data into one-sample applet. Don’t forget to set $r = .7$.
- JZS Bayes Factor, (“Jeffreys, Zellner, Siow”)
- Output are odds in the form Null/Alternative:
  
  Scaled JZS Bayes Factor = 1.15187
One Sample or Paired $t$-Test

- Enter data into one-sample applet. Don’t forget to set $r = 0.7$.
- JZS Bayes Factor, ("Jeffreys, Zellner, Siow")
- Output are odds in the form Null/Alternative: Scaled JZS Bayes Factor = 1.15187
- Evidence is equivocal for null and alternative
One Sample or Paired $t$-Test

- Enter data into one-sample applet. Don’t forget to set $r = .7$.
- JZS Bayes Factor, ("Jeffreys, Zellner, Siow")
- Output are odds in the form Null/Alternative:
  Scaled JZS Bayes Factor = 1.15187
- Evidence is equivocal for null and alternative
- What if we tried to handpick a more effect-favorable scale, try $r = .2$
- Output are odds in the form Null/Alternative:
  Scaled JZS Bayes Factor = 0.6050631
- BF is 1.65-to-1 in favor of effect, scant evidence
One Sample or Paired $t$-Test

“We assessed the relative evidence for an effect vs. a null effect using the JZS Bayes factor as described by Rouder et al. (2009). The prior effect-size scale was .7, a reasonable value in this context. The resulting Bayes factor is 1.15 in favor of the null, indicating equivocal evidence.”
How big of a $t$ value would be needed for an effect in this situation?
One Sample or Paired $t$-Test

- How big of a $t$ value would be needed for an effect in this situation?
- $t = 3, B = 7.7$ for effect
One Sample or Paired $t$-Test

- How big of a $t$ value would be needed for an effect in this situation?
- $t = 3, B = 7.7$ for effect
- $t = 4, B = 206$ for effect
One Sample or Paired $t$-Test

- How big of a $t$ value would be needed for an effect in this situation?
- $t = 3$, $B = 7.7$ for effect
- $t = 4$, $B = 206$ for effect
- $t = 5$, $B = 6,885$, for effect
Two-Sample or Grouped \( t \)-Test

Use the two-sample link. Other than that, same deal.
Linear Regression


Web applet, but scale $r$ is fixed to 1.0, and one can only analyze a single model at a time.

Let's use the more flexible R package.

Keep your browser window handy, point browser to pcl.missouri.edu/jeff

Open the Tab, Multiple Regression

Make sure your R window is open as well
Linear Regression

Linear Regression

- Web applet, but scale $r$ is fixed to 1.0, and one can only analyze a single model at a time.
Linear Regression

- Web applet, but scale $r$ is fixed to 1.0, and one can only analyze a single model at a time.
- Let’s use the more flexible R package.
Linear Regression

- Web applet, but scale $r$ is fixed to 1.0, and one can only analyze a single model at a time.
- Let’s use the more flexible R package.
- Keep your browser window handy, point browser to pcl.missouri.edu/jeff

Web applet, but scale $r$ is fixed to 1.0, and one can only analyze a single model at a time.

Let’s use the more flexible R package.

Keep your browser window handy, point browser to pcl.missouri.edu/jeff

Open the Tab, Multiple Regression
Linear Regression


- Web applet, but scale $r$ is fixed to 1.0, and one can only analyze a single model at a time.

- Let’s use the more flexible R package.

- Keep your browser window handy, point browser to pcl.missouri.edu/jeff

- Open the Tab, Multiple Regression

- Make sure your R window is open as well
Hominid cranial capacity grew rapidly over last 2 million years.

What are the correlates:

1. local climate variability
2. global temperature
3. parasite load
4. population density
Example from Bailey & Geary (2011)

- Hominid cranial capacity grew rapidly over last 2 million years.
- What are the correlates:
Linear Regression

Example from Bailey & Geary (2011)

- Hominid cranial capacity grew rapidly over last 2 million years.
- What are the correlates:
  1. local climate variability, *climate*
  2. global temperature
  3. parasite load
  4. population density

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Linear Regression

Example from Bailey & Geary (2011)

- Hominid cranial capacity grew rapidly over last 2 million years.
- What are the correlates:
  1. local climate variability, *climate*
  2. global temperature, *o2*
Example from Bailey & Geary (2011)

- Hominid cranial capacity grew rapidly over last 2 million years.
- What are the correlates:
  1. local climate variability, *climate*
  2. global temperature, *o2*
  3. parasite load *para*
Example from Bailey & Geary (2011)

- Hominid cranial capacity grew rapidly over last 2 million years.
- What are the correlates:
  1. local climate variability, \( \text{climate} \)
  2. global temperature, \( o2 \)
  3. parasite load \( \text{para} \)
  4. population density \( \text{pop} \)
Highlight BLOCK #1 in the webpage

Copy & Paste Into R

Type "dat", see end of data file

Raise Your Hand When Done

Jeffrey N. Rouder
Bayesian Inference in Psychology: A Workshop
Highlight BLOCK #1 in the webpage
Getting Data into R

- Highlight BLOCK #1 in the webpage
- Copy & Paste Into R
Getting Data into R

- Highlight BLOCK #1 in the webpage
- Copy & Paste Into R
- Type “dat”, see end of data file
Getting Data into R

- Highlight BLOCK #1 in the webpage
- Copy & Paste Into R
- Type “dat”, see end of data file
- Raise Your Hand When Done
Conventional Analysis

- Copy & Paste BLOCK #2
- 4 tests, one for each covariate
- Correlation (colinearity) among covariates
Set Up For One Covariate:

\[ Y_i = \alpha + (X_i - \bar{X}) \theta + \epsilon_i \]

Units of \( \theta \) are (units of \( Y \)) / (units of \( X \))

Need to standardized with respect to units of \( Y \) (as before) and units of \( X \)

Effect size: \( \delta = \theta \left( \frac{s_X}{\sigma} \right) \)

Null: \( \delta = 0 \)

Alt: \( \delta \) is from a Cauchy with scale \( r \), set to \( r = 0.5 \).
Set Up For One Covariate:

\[ Y_i = \alpha + (X_i - \bar{X})\theta + \epsilon_i \]

- Units of \( \theta \) are \((\text{units of } Y) / (\text{units of } X)\)
- Need to standardized with respect to units of \( Y \) (as before) and units of \( X \)
- Effect size: \( \delta = \theta \left( \frac{s}{\sigma} \right) \)
- Null: \( \delta = 0 \)
- Alt: \( \delta \) is from a Cauchy with scale \( r \), set to \( r = 0.5 \).
Linear Regression: Bayes Factor Setup

Set Up For One Covariate:

- $Y_i = \alpha + (X_i - \bar{X})\theta + \epsilon_i$
- Units of $\theta$ are (units of $Y$) / (units of $X$)
Linear Regression: Bayes Factor Setup

Set Up For One Covariate:

- \( Y_i = \alpha + (X_i - \bar{X})\theta + \epsilon_i \)
- Units of \( \theta \) are (units of \( Y \)) / (units of \( X \))
- Need to standardized with respect to units of \( Y \) (as before) and units of \( X \)
Linear Regression: Bayes Factor Setup

Set Up For One Covariate:

- \( Y_i = \alpha + (X_i - \bar{X})\theta + \epsilon_i \)
- Units of \( \theta \) are \(( \text{units of } Y) / (\text{units of } X)\)
- Need to standardized with respect to units of \( Y \) (as before) and units of \( X \)
- Effect size: \( \delta = \theta \left( \frac{s_X}{\sigma} \right) \)
Set Up For One Covariate:

- $Y_i = \alpha + (X_i - \bar{X})\theta + \epsilon_i$
- Units of $\theta$ are $(\text{units of } Y) / (\text{units of } X)$
- Need to standardized with respect to units of $Y$ (as before) and units of $X$
- Effect size: $\delta = \theta \left( \frac{sX}{\sigma} \right)$
- Null: $\delta = 0$
Set Up For One Covariate:

- \( Y_i = \alpha + (X_i - \bar{X})\theta + \epsilon_i \)
- Units of \( \theta \) are \( \text{(units of } Y) / \text{(units of } X) \)
- Need to standardized with respect to units of \( Y \) (as before) and units of \( X \)
- Effect size: \( \delta = \theta \left( \frac{sx}{\sigma} \right) \)
- Null: \( \delta = 0 \)
- Alt: \( \delta \) is from a Cauchy with scale \( r \), set to \( r = .5 \)
Generalization To Multiple Covariates:

\[ Y_i = \mu \mathbf{1} + X \theta + \epsilon \]

\[ \theta | g \sim \text{Normal}(\mathbf{0}, g \sigma^2 N(X'X)^{-1}) \]

\[ g \sim \text{Inverse Gamma}(1/2, r^2/2) \]
Top-Down Linear Regression

- 4 covariates, test each one,
Top-Down Linear Regression

- 4 covariates, test each one,
- Example: Is parasite load needed?
4 covariates, test each one,
Example: Is parasite load needed?
Full Model: \( cc \sim \text{intercept} + \text{para} + \text{pos} + \text{climate} + \text{o2} \)
Top-Down Linear Regression

- 4 covariates, test each one,
- Example: Is parasite load needed?
- Full Model: \( cc \sim \text{intercept} + \text{para} + \text{pos} + \text{climate} + \text{o2} \)
- Omit para Model: \( cc \sim \text{intercept} + \text{pos} + \text{climate} + \text{o2} \)
Top-Down Linear Regression

- 4 covariates, test each one,
- Example: Is parasite load needed?
- Full Model: cc $\sim$ intercept+para+pos+climate+o2
- Omit para Model: cc $\sim$ intercept+pos+climate+o2
- Evidence for parasite load: Full Model is preferred to the model without parasite load.
Top-Down Linear Regression

- 4 covariates, test each one,
- Example: Is parasite load needed?
- Full Model: $cc \sim \text{intercept} + \text{para} + \text{pos} + \text{climate} + o2$
- Omit para Model: $cc \sim \text{intercept} + \text{pos} + \text{climate} + o2$
- Evidence for parasite load: Full Model is preferred to the model without parasite load.
- Evidence against parasite load: Model without parasite load is preferred to the Full Model.
Top-Down Linear Regression

- 4 covariates, test each one,
- Example: Is parasite load needed?
- Full Model: $cc \sim \text{intercept} + \text{para} + \text{pos} + \text{climate} + \text{o2}$
- Omit para Model: $cc \sim \text{intercept} + \text{pos} + \text{climate} + \text{o2}$
- Evidence for parasite load: Full Model is preferred to the model without parasite load.
- Evidence against parasite load: Model without parasite load is preferred to the Full Model.
- Bayes Factor between models, BLOCK #3
Better Approach, Consider All Models

- 4 covariates
Better Approach, Consider All Models

- 4 covariates
- 16 possible models with the inclusion or exclusion of all covariates
Better Approach, Consider All Models

- 4 covariates
- 16 possible models with the inclusion or exclusion of all covariates
- We can do all 16 quickly and easily, go to BLOCK #4
- Note the structure and functionality of the bf object
“We assessed the evidence for covariates using the Bayes factor approach described in Rouder & Morey (2012) in the BayesFactor package (Morey & Rouder, 2012) with default settings. The best model includes population density and global temperature and excludes local climate variation and parasite load. Adding local climate and parasite load results in less favorable models with Bayes factors of 1-to-11.9 and of 1-to-4.1, respectively. In contrast, population density and global temperature are needed. Omitting these covariates results in less favorable models with Bayes factors of $1 \times 10^{18}$ and $1 \times 10^8$, respectively.”
Linear Regression

Some Unexpected Benefits

Colinearity in Covariates is not problematic, goes into prior term ($N(X'X)^{-1}$).

No need for stepwise:

Bayes factor provides easy comparison regardless of number of covariates

Algorithm is fast, can blow through 1000 models quickly.
Some Unexpected Benefits

- Colinearity in Covariates is not problematic, goes into prior term \( N(X'X)^{-1} \)
Some Unexpected Benefits

- Colinearity in Covariates is not problematic, goes into prior term \((N(X'X)^{-1})\)
- No need for stepwise:
  - Bayes factor provides easy comparison regardless of number of covariates
  - Algorithm is fast, can blow through 1000 models quickly.
Conceptually, Bayesian analysis of ANOVA is surprisingly complicated.

- Random vs. fixed
- Interaction Models
- Linear Constraints
- What to shrink to, how much to borrow across factors?
- Which models to compare?

Rouder et al. (2012) was my midlife dissertation.
Conceptually, Bayesian analysis of ANOVA is surprisingly complicated.

- Random vs. fixed
- Interaction Models
- Linear Constraints
- What to shrink to, how much to borrow across factors?
- Which models to compare?
- Rouder et al. (2012) was my midlife dissertation.
Example: Participants Read Nouns and Verbs in Red and Green; RT is dependent measure.
ANOVA

Example: Participants Read Nouns and Verbs in Red and Green; RT is dependent measure.

- Factor POS (noun vs. verb)
Example: Participants Read Nouns and Verbs in Red and Green; RT is dependent measure.
  ▶ Factor POS (noun vs. verb)
  ▶ Factor COLOR (red vs. green)
Example: Participants Read Nouns and Verbs in Red and Green; RT is dependent measure.

- Factor POS (noun vs. verb)
- Factor COLOR (red vs. green)
- Factorial, 4 cells
Example: Participants Read Nouns and Verbs in Red and Green; RT is dependent measure.

- Factor POS (noun vs. verb)
- Factor COLOR (red vs. green)
- Factorial, 4 cells
- Between Subjects: A separate bunch of participants responded in each cell.
Conventional Analysis for 2-way ANOVA:

Q1: Is there a main effect of POS?
Q2: Is there a main effect of COLOR?
Q3: Is there a POS-BY-COLOR interaction?

These three questions comprise a top-down approach to ANOVA.
ANOVA

Conventional Analysis for 2-way ANOVA:

- Q1: Is there a main effect of POS?
- Q2: Is there a main effect of COLOR?
- Q3: Is there a POS-BY-COLOR interaction?
Conventional Analysis for 2-way ANOVA:

- Q1: Is there a main effect of POS?
- Q2: Is there a main effect of COLOR?
- Q3: Is there a POS-BY-COLOR interaction?

- These three questions comprise a top-down approach to ANOVA
Conventional Analysis for 2-way ANOVA:

- Q1: Is there a main effect of POS?
- Q2: Is there a main effect of COLOR?
- Q3: Is there a POS-BY-COLOR interaction?
- These three questions comprise a top-down approach to ANOVA
- Go to pcl.missouri.edu/jeff, AOV Ex 1, BLOCK #1
Bayesian Analysis for 2-way ANOVA

- Full Model: $RT \sim POS + COLOR + POS*COLOR$
- Omit one-at-a-time
- 3 BFs, go to BLOCK #2
Useful Alternative Approach: Look at all models
Useful Alternative Approach: Look at all models

- Full Model: $\text{RT} \sim \text{POS} + \text{COLOR} + \text{POS} \times \text{COLOR}$
Useful Alternative Approach: Look at all models

- Full Model: RT \sim POS + COLOR + POS*COLOR
- Total of 8 models
Useful Alternative Approach: Look at all models

- Full Model: RT $\sim$ POS + COLOR + POS*COLOR
- Total of 8 models
- 7 BF$s$, go to BLOCK #3
In Reconsideration....

I am not content with the treatment of interactions

- Consider the following three models:
  \[ RT \sim COL + COL*POS \]
  \[ RT \sim POS + COL*POS \]
  \[ RT \sim COL*POS \]

- Commonality: interaction without main effect.
- Worry: fortuitous choices of levels that balance out a main effect in the presence of an interaction.
- Had we chose different levels of the factors, the main effect MUST be present if indeed the interaction is there.
- So, presence of interactions implies main effects and the above three models are not considered.
- Otherwise, one is considering inappropriately constrained models and overstating evidence for interactions.
In Reconsideration....

I am not content with the treatment of interactions
▶ Consider the following three models:
  RT $\sim$ COL + COL*POS
  RT $\sim$ POS + COL*POS
  RT $\sim$ COL*POS
▶ Commonality: interaction without main effect.
In Reconsideration....

I am not content with the treatment of interactions

- Consider the following three models:
  - $RT \sim COL + COL*POS$
  - $RT \sim POS + COL*POS$
  - $RT \sim COL*POS$

- Commonality: interaction without main effect.
- Worry: fortuitous choices of levels that balance out a main effect in the presence of an interaction
In Reconsideration....

I am not content with the treatment of interactions

- Consider the following three models:
  - \( RT \sim COL + COL*POS \)
  - \( RT \sim POS + COL*POS \)
  - \( RT \sim COL*POS \)

- Commonality: interaction without main effect.
- Worry: fortuitous choices of levels that balance out a main effect in the presence of an interaction
- Had we chose different levels of the factors, the main effect MUST be present if indeed the interaction is there.
I am not content with the treatment of interactions

- Consider the following three models:
  \[
  RT \sim COL + COL*POS \\
  RT \sim POS + COL*POS \\
  RT \sim COL*POS
  \]
- Commonality: interaction without main effect.
- Worry: fortuitous choices of levels that balance out a main effect in the presence of an interaction
- Had we chose different levels of the factors, the main effect MUST be present if indeed the interaction is there.
- So, presence of interactions implies main effects and the above three models are not considered.
In Reconsideration....

I am not content with the treatment of interactions

- Consider the following three models:
  \[ RT \sim COL + COL*POS \]
  \[ RT \sim POS + COL*POS \]
  \[ RT \sim COL*POS \]

- Commonality: interaction without main effect.
- Worry: fortuitous choices of levels that balance out a main effect in the presence of an interaction.
- Had we chose different levels of the factors, the main effect MUST be present if indeed the interaction is there.
- So, presence of interactions implies main effects and the above three models are not considered.
- Otherwise, one is considering inappropriately constrained models and overstating evidence for interactions.
2-Way ANOVA, With Above Constrains

Go to BLOCK #4
"We assessed the evidence for part-of-speech and color with the Bayes factor approach described in Rouder et al. (2012) using the BayesFactor package (Morey & Rouder, 2012) in R with default settings. There is strong evidence for a part-of-speech effect (BF: 983-to-1) and evidence against any addition effect of color (BF: 1-to-2.8)."
Mixed ANOVA

Older and younger adults named nouns and verbs in red and green.

- AGE: Between participants, 30 of each, *fixed*
- POS & COLOR: Within participants, ea. person observed all four combinations, *fixed*
- SUB, large number of levels *random*
Conventional Approach

See AOV Example 2, BLOCK 1
Several complexities for Mixed ANOVA

1. Fixed and Random Factors, Fixed Interactions, Random Interactions, Mixed Interactions
2. Within vs. Between Subject Manipulations
3. Really large number of models
4. Hierarchical Structure (Gelman's approach; shrinkage within a factor but not across factors)
Mixed ANOVA

Several complexities for Mixed ANOVA

1. Fixed and Random Factors, Fixed Interactions, Random Interactions, Mixed Interactions
Mixed ANOVA

Several complexities for Mixed ANOVA

1. Fixed and Random Factors, Fixed Interactions, Random Interactions, Mixed Interactions

2. Within vs. Between Subject Manipulations
Several complexities for Mixed ANOVA

1. Fixed and Random Factors, Fixed Interactions, Random Interactions, Mixed Interactions

2. Within vs. Between Subject Manipulations

3. Really large number of models
   Total of 15 covariates: 4 main effects, 6 two-way, 4 three-way, 1 four-way Total of $2^{15} = 32,768$ Models
Mixed ANOVA

Several complexities for Mixed ANOVA

1. Fixed and Random Factors, Fixed Interactions, Random Interactions, Mixed Interactions
2. Within vs. Between Subject Manipulations
3. Really large number of models
   Total of 15 covariates: 4 main effects, 6 two-way, 4 three-way, 1 four-way Total of $2^{15} = 32,768$ Models
4. Hierarchical Structure (Gelman’s approach; shrinkage within a factor but not across factors)
Mixed ANOVA

Fixed vs. Random

- A fixed main effect has one linear constraint.
- A fixed interaction has $a + b - 1$ linear constraints where $a$ is the number of levels for the first factor and $b$ is the number of levels for the second.
- A mixed interaction has $a$ or $b$ constraints, depending on which factor is fixed and which is random.
Mixed ANOVA

Within vs. Between

- Not a problem
- Package knows which is which by the data input.
- Analyses the appropriate model automatically
Mixed ANOVA

Large Number of Models

- Cull by considering main effects when considering interactions
- Consider eliminating subject-by-stimulus-factor interactions. Why?
  - Interpretation of a true subject-by-color interaction: Some people truly read red words faster while others truly read green words faster.
  - If so, would the main effect of color be that meaningful?
  - For example would the main effect of color be meaningful if 40% of participants read green 100 ms better while 60% of read red 80 ms better?
  - I drop these complicated subject-by-stimulus-factor interactions because I would rather it go into noise than interpret these difficult interactions
Mixed ANOVA

Bayesian Analysis, go to Block 2
Mixed ANOVA

“We assessed the evidence for age, part-of-speech, and color effects with the Bayes factor approach described in Rouder et al. (2012) using the BayesFactor package (Morey & Rouder, 2012) in R with default settings. The best model included an age effect, a part-of-speech effect, and an interaction between them with young adults responding differentially better to nouns than old adults. The evidence for this interaction was equivocal ($B=1.1$-to-1), that is, there is effectively the same evidence for the model without it. Removing the age effect resulted in a dramatically worse models ($B=1$-to-29,900) as did removing the part-of-speech effect ($B=1$-to-$4.3 \times 10^{52}$). There evidence for a lack of a color effect was $B=3.1$.\"
I am at a loss here and appreciate your input and concerns. I do not think you will elevate the probability of drop-dead rejection because of use of Bayes factor. You may be critiqued, however, probably by someone who doesn't feel comfortable with statistical innovation. I take these critiques as opportunities to educate my reviewers and editors, sometimes against their will.

Do you think the field reasonable? If yes, then use Bayes factors for they are reasonable. If no, then maybe our problems transcend statistics.
Will My Reviewers Buy It?

- I am at a loss here and appreciate your input and concerns.
I am at a loss here and appreciate your input and concerns.

I do not think you will elevate the probability of drop-dead rejection because of use of Bayes factor. You may be critiqued, however, probably by someone who doesn’t feel comfortable with statistical innovation.
Will My Reviewers Buy It?

- I am at a loss here and appreciate your input and concerns.
- I do not think you will elevate the probability of drop-dead rejection because of use of Bayes factor. You may be critiqued, however, probably by someone who doesn’t feel comfortable with statistical innovation.
- I take these critiques as opportunities to educate my reviewers and editors, sometimes against their will.
Will My Reviewers Buy It?

- I am at a loss here and appreciate your input and concerns.
- I do not think you will elevate the probability of drop-dead rejection because of use of Bayes factor. You may be critiqued, however, probably by someone who doesn’t feel comfortable with statistical innovation.
- I take these critiques as opportunities to educate my reviewers and editors, sometimes against their will.
- Do you think the field reasonable?
Will My Reviewers Buy It?

- I am at a loss here and appreciate your input and concerns.
- I do not think you will elevate the probability of drop-dead rejection because of use of Bayes factor. You may be critiqued, however, probably by someone who doesn’t feel comfortable with statistical innovation.
- I take these critiques as opportunities to educate my reviewers and editors, sometimes against their will.
- Do you think the field reasonable?
  - If yes, then use Bayes factors for they are reasonable.
Will My Reviewers Buy It?

▶ I am at a loss here and appreciate your input and concerns
▶ I do not think you will elevate the probability of drop-dead rejection because of use of Bayes factor. You may be critiqued, however, probably by someone who doesn’t feel comfortable with statistical innovation.
▶ I take these critiques as opportunities to educate my reviewers and editors, sometimes against their will.
▶ Do you think the field reasonable?
  ▶ If yes, then use Bayes factors for they are reasonable.
  ▶ If no, then maybe our problems transcend statistics
Thank You