An Introduction to Theoretical Statistics Research

Blair Bilodeau

July 13, 2021

Independent Summer Stats Community

Theoretical Guarantees

Data in theory.

Data in practice.

Theoretical Guarantees

Data in theory.

Data in practice.

To make guarantees, we must assume something about the world.
These assumptions and guarantees may not be related to utility in practice.

Theoretical Guarantees

Data in theory.

Data in practice.

To make guarantees, we must assume something about the world.
These assumptions and guarantees may not be related to utility in practice.

How to trade-off assumptions needed for theory with the reality of applications?

Theoretical Guarantees

Data in theory.

Data in practice.

To make guarantees, we must assume something about the world.
These assumptions and guarantees may not be related to utility in practice.

How to trade-off assumptions needed for theory with the reality of applications?

The Theoretician’s Toolkit

The Theoretician's Toolkit

Common assumptions

The Theoretician's Toolkit

Common assumptions

- Normality of data (sample size is "large enough").

The Theoretician's Toolkit

Common assumptions

- Normality of data (sample size is "large enough").
- Data generated by some analytically tractable parametric family.

The Theoretician's Toolkit

Common assumptions

- Normality of data (sample size is "large enough").
- Data generated by some analytically tractable parametric family.
- Observations are independent and identically distributed (I.I.D.)...

  ...or they satisfy some other dependence structure (e.g., Markov).

The Theoretician's Toolkit

Common assumptions

- Normality of data (sample size is "large enough").
- Data generated by some analytically tractable parametric family.
- Observations are independent and identically distributed (I.I.D.)...

  ...or they satisfy some other dependence structure (e.g., Markov).

Ways to sanitize reality

The Theoretician's Toolkit

Common assumptions

- Normality of data (sample size is "large enough").
- Data generated by some analytically tractable parametric family.
- Observations are independent and identically distributed (I.I.D.)...
  ...or they satisfy some other dependence structure (e.g., Markov).

Ways to sanitize reality

- Study asymptotic limits (number of data points tends to infinity).

The Theoretician's Toolkit

Common assumptions

- Normality of data (sample size is "large enough").
- Data generated by some analytically tractable parametric family.
- Observations are independent and identically distributed (I.I.D.)...

  ...or they satisfy some other dependence structure (e.g., Markov).

Ways to sanitize reality

- Study asymptotic limits (number of data points tends to infinity).
- Ignore constant, logarithmic, or even lower-order polynomial factors.

The Theoretician's Toolkit

Common assumptions

- Normality of data (sample size is "large enough").
- Data generated by some analytically tractable parametric family.
- Observations are independent and identically distributed (I.I.D.)...

  ...or they satisfy some other dependence structure (e.g., Markov).

Ways to sanitize reality

- Study asymptotic limits (number of data points tends to infinity).
- Ignore constant, logarithmic, or even lower-order polynomial factors.
- Avoid commenting on any specific instance of using a method, only providing guarantees for performance on average or in the worst case.

The Theoretician's Toolkit

Common assumptions

- Normality of data (sample size is "large enough").
- Data generated by some analytically tractable parametric family.
- Observations are independent and identically distributed (I.I.D.)...

  ...or they satisfy some other dependence structure (e.g., Markov).

Ways to sanitize reality

- Study asymptotic limits (number of data points tends to infinity).
- Ignore constant, logarithmic, or even lower-order polynomial factors.
- Avoid commenting on any specific instance of using a method, only providing guarantees for performance on average or in the worst case.

Impacts of Theoretical Research

Impacts of Theoretical Research

Questions theory tries to answer

Impacts of Theoretical Research

Questions theory tries to answer

- How well can we guarantee this method used in practice will perform?

Impacts of Theoretical Research

Questions theory tries to answer

- How well can we guarantee this method used in practice will perform?
- Is one statistical task inherently more or less difficult than another?

Impacts of Theoretical Research

Questions theory tries to answer

- How well can we guarantee this method used in practice will perform?
- Is one statistical task inherently more or less difficult than another?
- Can novel methods be motivated and designed using mathematical tools?

Impacts of Theoretical Research

Questions theory tries to answer

- How well can we guarantee this method used in practice will perform?
- Is one statistical task inherently more or less difficult than another?
- Can novel methods be motivated and designed using mathematical tools?

What might answers to these questions look like?

Impacts of Theoretical Research

Questions theory tries to answer

- How well can we guarantee this method used in practice will perform?
- Is one statistical task inherently more or less difficult than another?
- Can novel methods be motivated and designed using mathematical tools?

What might answers to these questions look like?

- New guarantees for an existing method under "realistic" assumptions.

Impacts of Theoretical Research

Questions theory tries to answer

- How well can we guarantee this method used in practice will perform?
- Is one statistical task inherently more or less difficult than another?
- Can novel methods be motivated and designed using mathematical tools?

What might answers to these questions look like?

- New guarantees for an existing method under "realistic" assumptions.
- Quantifications of the key statistical elements of a task...

  ...or a new paradigm describing a family of statistical tasks.

Impacts of Theoretical Research

Questions theory tries to answer

- How well can we guarantee this method used in practice will perform?
- Is one statistical task inherently more or less difficult than another?
- Can novel methods be motivated and designed using mathematical tools?

What might answers to these questions look like?

- New guarantees for an existing method under "realistic" assumptions.
- Quantifications of the key statistical elements of a task...
  ...or a new paradigm describing a family of statistical tasks.
- Introduction of new methods for a novel or existing statistical task, along with associated performance guarantees.

Impacts of Theoretical Research

Questions theory tries to answer

- How well can we guarantee this method used in practice will perform?
- Is one statistical task inherently more or less difficult than another?
- Can novel methods be motivated and designed using mathematical tools?

What might answers to these questions look like?

- New guarantees for an existing method under "realistic" assumptions.
- Quantifications of the key statistical elements of a task...
  ...or a new paradigm describing a family of statistical tasks.
- Introduction of new methods for a novel or existing statistical task, along with associated performance guarantees.
- Can also be negative results:

Impacts of Theoretical Research

Questions theory tries to answer

- How well can we guarantee this method used in practice will perform?
- Is one statistical task inherently more or less difficult than another?
- Can novel methods be motivated and designed using mathematical tools?

What might answers to these questions look like?

- New guarantees for an existing method under "realistic" assumptions.
- Quantifications of the key statistical elements of a task...

  ...or a new paradigm describing a family of statistical tasks.
- Introduction of new methods for a novel or existing statistical task, along with associated performance guarantees.
- Can also be negative results:
  - showing a method *does* not perform well under certain assumptions,
  - or that a statistical task is inherently more difficult than anticipated.

Impacts of Theoretical Research

Questions theory tries to answer

- How well can we guarantee this method used in practice will perform?
- Is one statistical task inherently more or less difficult than another?
- Can novel methods be motivated and designed using mathematical tools?

What might answers to these questions look like?

- New guarantees for an existing method under "realistic" assumptions.
- Quantifications of the key statistical elements of a task...
  ...or a new paradigm describing a family of statistical tasks.
- Introduction of new methods for a novel or existing statistical task, along with associated performance guarantees.
- Can also be negative results:
  - showing a method *does* not perform well under certain assumptions,
  - or that a statistical task is inherently more difficult than anticipated.

When Theory and Practice Collide

When Theory and Practice Collide

Machine Learning

When Theory and Practice Collide

Machine Learning

- Classical statistics says that one should tradeoff complexity (size) of a model:
  - Larger models have less bias; they can learn very complicated relationships.
  - Smaller models have smaller variance; they avoid overfitting.

When Theory and Practice Collide

Machine Learning

- Classical statistics says that one should tradeoff complexity (size) of a model:
  - Larger models have less bias; they can learn very complicated relationships.
  - Smaller models have smaller variance; they avoid overfitting.
- Practitioners (many from Toronto!) observed excellent empirical performance of massive neural networks, and they keep getting bigger.

When Theory and Practice Collide

Machine Learning

- Classical statistics says that one should tradeoff complexity (size) of a model:
  - Larger models have less bias; they can learn very complicated relationships.
  - Smaller models have smaller variance; they avoid overfitting.
- Practitioners (many from Toronto!) observed excellent empirical performance of massive neural networks, and they keep getting bigger.
- Theoreticians were forced to revisit what “complexity” means. (Bartlett, 1996; Bartlett et al., 2017; Arora et al., 2018)

When Theory and Practice Collide

Machine Learning

- Classical statistics says that one should tradeoff complexity (size) of a model:
  - Larger models have less bias; they can learn very complicated relationships.
  - Smaller models have smaller variance; they avoid overfitting.
- Practitioners (many from Toronto!) observed excellent empirical performance of massive neural networks, and they keep getting bigger.
- Theoreticians were forced to revisit what "complexity" means. (Bartlett, 1996; Bartlett et al., 2017; Arora et al., 2018)

Adam Optimizer

When Theory and Practice Collide

Machine Learning

- Classical statistics says that one should tradeoff complexity (size) of a model:
  - Larger models have less bias; they can learn very complicated relationships.
  - Smaller models have smaller variance; they avoid overfitting.
- Practitioners (many from Toronto!) observed excellent empirical performance of massive neural networks, and they keep getting bigger.
- Theoreticians were forced to revisit what "complexity" means. (Bartlett, 1996; Bartlett et al., 2017; Arora et al., 2018)

Adam Optimizer

- An efficient, empirically strong optimization method (Kingma and Ba, 2015).

When Theory and Practice Collide

Machine Learning

- Classical statistics says that one should tradeoff complexity (size) of a model:
  - Larger models have less bias; they can learn very complicated relationships.
  - Smaller models have smaller variance; they avoid overfitting.
- Practitioners (many from Toronto!) observed excellent empirical performance of massive neural networks, and they keep getting bigger.
- Theoreticians were forced to revisit what "complexity" means. (Bartlett, 1996; Bartlett et al., 2017; Arora et al., 2018)

Adam Optimizer

- An efficient, empirically strong optimization method (Kingma and Ba, 2015).
- Provably does worse than theory initially claimed (Reddi et al., 2018).

When Theory and Practice Collide

Machine Learning

- Classical statistics says that one should tradeoff complexity (size) of a model:
  - Larger models have less bias; they can learn very complicated relationships.
  - Smaller models have smaller variance; they avoid overfitting.
- Practitioners (many from Toronto!) observed excellent empirical performance of massive neural networks, and they keep getting bigger.
- Theoreticians were forced to revisit what "complexity" means. (Bartlett, 1996; Bartlett et al., 2017; Arora et al., 2018)

Adam Optimizer

- An efficient, empirically strong optimization method (Kingma and Ba, 2015).
- Provably does worse than theory initially claimed (Reddi et al., 2018).
- Still used *extensively* and *successfully* in modern machine learning.

When Theory and Practice Collide

Machine Learning

- Classical statistics says that one should tradeoff complexity (size) of a model:
  - Larger models have less bias; they can learn very complicated relationships.
  - Smaller models have smaller variance; they avoid overfitting.
- Practitioners (many from Toronto!) observed excellent empirical performance of massive neural networks, and they keep getting bigger.
- Theoreticians were forced to revisit what "complexity" means. (Bartlett, 1996; Bartlett et al., 2017; Arora et al., 2018)

Adam Optimizer

- An efficient, empirically strong optimization method (Kingma and Ba, 2015).
- Provably does worse than theory initially claimed (Reddi et al., 2018).
- Still used *extensively* and *successfully* in modern machine learning.

Bayesian Statistics

When Theory and Practice Collide

Machine Learning

- Classical statistics says that one should tradeoff complexity (size) of a model:
  - Larger models have less bias; they can learn very complicated relationships.
  - Smaller models have smaller variance; they avoid overfitting.
- Practitioners (many from Toronto!) observed excellent empirical performance of massive neural networks, and they keep getting bigger.
- Theoreticians were forced to revisit what "complexity" means. (Bartlett, 1996; Bartlett et al., 2017; Arora et al., 2018)

Adam Optimizer

- An efficient, empirically strong optimization method (Kingma and Ba, 2015).
- Provably does worse than theory initially claimed (Reddi et al., 2018).
- Still used *extensively* and *successfully* in modern machine learning.

Bayesian Statistics

- Theoretically shown to be optimal while still practically intractable.

When Theory and Practice Collide

Machine Learning

- Classical statistics says that one should tradeoff complexity (size) of a model:
  - Larger models have less bias; they can learn very complicated relationships.
  - Smaller models have smaller variance; they avoid overfitting.
- Practitioners (many from Toronto!) observed excellent empirical performance of massive neural networks, and they keep getting bigger.
- Theoreticians were forced to revisit what "complexity" means. (Bartlett, 1996; Bartlett et al., 2017; Arora et al., 2018)

Adam Optimizer

- An efficient, empirically strong optimization method (Kingma and Ba, 2015).
- Provably does worse than theory initially claimed (Reddi et al., 2018).
- Still used *extensively* and *successfully* in modern machine learning.

Bayesian Statistics

- Theoretically shown to be optimal while still practically intractable.
- Modern computational advancements have led to wide application in science.

When Theory and Practice Collide

Machine Learning

- Classical statistics says that one should tradeoff complexity (size) of a model:
  - Larger models have less bias; they can learn very complicated relationships.
  - Smaller models have smaller variance; they avoid overfitting.
- Practitioners (many from Toronto!) observed excellent empirical performance of massive neural networks, and they keep getting bigger.
- Theoreticians were forced to revisit what "complexity" means. (Bartlett, 1996; Bartlett et al., 2017; Arora et al., 2018)

Adam Optimizer

- An efficient, empirically strong optimization method (Kingma and Ba, 2015).
- Provably does worse than theory initially claimed (Reddi et al., 2018).
- Still used *extensively* and *successfully* in modern machine learning.

Bayesian Statistics

- Theoretically shown to be optimal while still practically intractable.
- Modern computational advancements have led to wide application in science.

Quantifying the Difficulty of Statistical Tasks

Quantifying the Difficulty of Statistical Tasks

Task: Observe data pairs $(X_{1:n}, Y_{1:n})$

Receive a new covariate X and predict the response Y.

Quantifying the Difficulty of Statistical Tasks

Task: Observe data pairs $(X_{1:n}, Y_{1:n})$

Receive a new covariate X and predict the response Y.

Regression: Approximate $\mathbb{E}[Y \mid X]$ with $\hat{f} : \mathcal{X} \to \mathcal{Y}$; (e.g., $\hat{f}$ is linear).

Quantifying the Difficulty of Statistical Tasks

Task: Observe data pairs $(X_{1:n}, Y_{1:n})$

Receive a new covariate X and predict the response Y.

Regression: Approximate $\mathbb{E}[Y \mid X]$ with $\hat{f} : \mathcal{X} \to \mathcal{Y}$; (e.g., $\hat{f}$ is linear).

Problem: Cannot capture uncertainty in the data without strong assumptions.

Quantifying the Difficulty of Statistical Tasks

Task: Observe data pairs $(X_{1:n}, Y_{1:n})$
Receive a new covariate X and predict the response Y.

Regression: Approximate $\mathbb{E}[Y \mid X]$ with $\hat{f} : \mathcal{X} \to \mathcal{Y}$; (e.g., $\hat{f}$ is linear).

Problem: Cannot capture uncertainty in the data without strong assumptions.

Our work: Predict full conditional densities rather than means.

Quantifying the Difficulty of Statistical Tasks

Task: Observe data pairs $(X_{1:n}, Y_{1:n})$

Receive a new covariate X and predict the response Y.

Regression: Approximate $\mathbb{E}[Y \mid X]$ with $\hat{f} : \mathcal{X} \to \mathcal{Y}$; (e.g., $\hat{f}$ is linear).

Problem: Cannot capture uncertainty in the data without strong assumptions.

Our work: Predict full conditional densities rather than means.

We answer the question...

What is the best I can do in the worst case situation?

Quantifying the Difficulty of Statistical Tasks

Task: Observe data pairs $(X_{1:n}, Y_{1:n})$
Receive a new covariate X and predict the response Y.
Regression: Approximate $\mathbb{E}[Y \mid X]$ with $\hat{f} : \mathcal{X} \to \mathcal{Y}$; (e.g., $\hat{f}$ is linear).
Problem: Cannot capture uncertainty in the data without strong assumptions.
Our work: Predict full conditional densities rather than means.

We answer the question...
What is the best I can do in the worst case situation?

- For discrete data, with no assumptions on the dependence structure. (Bilodeau, Foster, and Roy, 2020a)

Quantifying the Difficulty of Statistical Tasks

Task: Observe data pairs $(X_{1:n}, Y_{1:n})$
Receive a new covariate X and predict the response Y.
Regression: Approximate $\mathbb{E}[Y \mid X]$ with $\hat{f} : \mathcal{X} \to \mathcal{Y}$; (e.g., $\hat{f}$ is linear).
Problem: Cannot capture uncertainty in the data without strong assumptions.
Our work: Predict full conditional densities rather than means.

We answer the question...
What is the best I can do in the worst case situation?

- For discrete data, with no assumptions on the dependence structure. (Bilodeau, Foster, and Roy, 2020a)
- For arbitrary data, under the assumption that observations are independent and identically distributed. (Bilodeau, Foster, and Roy, 2021a)

Quantifying the Difficulty of Statistical Tasks

Task: Observe data pairs $(X_{1:n}, Y_{1:n})$
Receive a new covariate X and predict the response Y.
Regression: Approximate $\mathbb{E}[Y \mid X]$ with $\hat{f} : \mathcal{X} \to \mathcal{Y}$; (e.g., $\hat{f}$ is linear).
Problem: Cannot capture uncertainty in the data without strong assumptions.
Our work: Predict full conditional densities rather than means.

We answer the question...
What is the best I can do in the worst case situation?

- For discrete data, with no assumptions on the dependence structure. (Bilodeau, Foster, and Roy, 2020a)
- For arbitrary data, under the assumption that observations are independent and identically distributed. (Bilodeau, Foster, and Roy, 2021a)

This reveals the inherent difficulty of identifying the best conditional density depending on the complexity of the assumed model.

Quantifying the Difficulty of Statistical Tasks

Task: Observe data pairs $(X_{1:n}, Y_{1:n})$
Receive a new covariate X and predict the response Y.
Regression: Approximate $\mathbb{E}[Y \mid X]$ with $\hat{f} : \mathcal{X} \to \mathcal{Y}$; (e.g., $\hat{f}$ is linear).
Problem: Cannot capture uncertainty in the data without strong assumptions.
Our work: Predict full conditional densities rather than means.

We answer the question...
What is the best I can do in the worst case situation?

- For discrete data, with no assumptions on the dependence structure. (Bilodeau, Foster, and Roy, 2020a)
- For arbitrary data, under the assumption that observations are independent and identically distributed. (Bilodeau, Foster, and Roy, 2021a)

This reveals the inherent difficulty of identifying the best conditional density depending on the complexity of the assumed model.

Introducing Novel Statistical Paradigms

Introducing Novel Statistical Paradigms

Main Idea: How is prediction affected by the i.i.d. assumption failing to hold?

Introducing Novel Statistical Paradigms

Main Idea: How is prediction affected by the i.i.d. assumption failing to hold?

Existing results: Well-understood at the endpoints of i.i.d. and “adversarial”.

Introducing Novel Statistical Paradigms

Main Idea: How is prediction affected by the i.i.d. assumption failing to hold?

Existing results: Well-understood at the endpoints of i.i.d. and "adversarial".

Our Solution (Bilodeau, Negrea, and Roy, 2020b)

Introducing Novel Statistical Paradigms

Main Idea: How is prediction affected by the i.i.d. assumption failing to hold?

Existing results: Well-understood at the endpoints of i.i.d. and “adversarial”.

Our Solution (Bilodeau, Negrea, and Roy, 2020b)

- Define a spectrum of settings inbetween i.i.d. and adversarial.

Introducing Novel Statistical Paradigms

Main Idea: How is prediction affected by the i.i.d. assumption failing to hold?
Existing results: Well-understood at the endpoints of i.i.d. and "adversarial".

Our Solution (Bilodeau, Negrea, and Roy, 2020b)

- Define a spectrum of settings inbetween i.i.d. and adversarial.
- Define what it means to do *as well as possible*, without knowing the true dependence structure.

Introducing Novel Statistical Paradigms

Main Idea: How is prediction affected by the i.i.d. assumption failing to hold?
Existing results: Well-understood at the endpoints of i.i.d. and "adversarial".

Our Solution (Bilodeau, Negrea, and Roy, 2020b)

- Define a spectrum of settings inbetween i.i.d. and adversarial.
- Define what it means to do *as well as possible*, without knowing the true dependence structure.
- Design a prediction algorithm that achieves this notion of optimality.

Introducing Novel Statistical Paradigms

Main Idea: How is prediction affected by the i.i.d. assumption failing to hold?
Existing results: Well-understood at the endpoints of i.i.d. and "adversarial".

Our Solution (Bilodeau, Negrea, and Roy, 2020b)

- Define a spectrum of settings inbetween i.i.d. and adversarial.
- Define what it means to do *as well as possible*, without knowing the true dependence structure.
- Design a prediction algorithm that achieves this notion of optimality.

Broader Implications

Introducing Novel Statistical Paradigms

Main Idea: How is prediction affected by the i.i.d. assumption failing to hold?
Existing results: Well-understood at the endpoints of i.i.d. and "adversarial".

Our Solution (Bilodeau, Negrea, and Roy, 2020b)

- Define a spectrum of settings inbetween i.i.d. and adversarial.
- Define what it means to do *as well as possible*, without knowing the true dependence structure.
- Design a prediction algorithm that achieves this notion of optimality.

Broader Implications

- How to define spectrums along assumptions in more general settings?

Introducing Novel Statistical Paradigms

Main Idea: How is prediction affected by the i.i.d. assumption failing to hold?
Existing results: Well-understood at the endpoints of i.i.d. and "adversarial".

Our Solution (Bilodeau, Negrea, and Roy, 2020b)

- Define a spectrum of settings inbetween i.i.d. and adversarial.
- Define what it means to do *as well as possible*, without knowing the true dependence structure.
- Design a prediction algorithm that achieves this notion of optimality.

Broader Implications

- How to define spectrums along assumptions in more general settings?
- Propose that statistical methods should be designed to do well *regardless* of the true setting, rather than only under a specific assumption.

Introducing Novel Statistical Paradigms

Main Idea: How is prediction affected by the i.i.d. assumption failing to hold?
Existing results: Well-understood at the endpoints of i.i.d. and "adversarial".

Our Solution (Bilodeau, Negrea, and Roy, 2020b)

- Define a spectrum of settings inbetween i.i.d. and adversarial.
- Define what it means to do *as well as possible*, without knowing the true dependence structure.
- Design a prediction algorithm that achieves this notion of optimality.

Broader Implications

- How to define spectrums along assumptions in more general settings?
- Propose that statistical methods should be designed to do well *regardless* of the true setting, rather than only under a specific assumption.

Performance Guarantees for Existing Methods

Performance Guarantees for Existing Methods

The Problem: Bayesian inference requires an intractable integral. Most solutions use computationally-heavy sampling techniques to approximate.

Performance Guarantees for Existing Methods

The Problem: Bayesian inference requires an intractable integral. Most solutions use computationally-heavy sampling techniques to approximate.

Existing Solutions: Numerical approximations to integrals have been proposed for a long time, but theory not complete.

Performance Guarantees for Existing Methods

The Problem: Bayesian inference requires an intractable integral. Most solutions use computationally-heavy sampling techniques to approximate.

Existing Solutions: Numerical approximations to integrals have been proposed for a long time, but theory not complete.

Our Contributions (Bilodeau, Stringer, and Tang, 2021b)

Performance Guarantees for Existing Methods

The Problem: Bayesian inference requires an intractable integral. Most solutions use computationally-heavy sampling techniques to approximate.

Existing Solutions: Numerical approximations to integrals have been proposed for a long time, but theory not complete.

Our Contributions (Bilodeau, Stringer, and Tang, 2021b)

- First theoretical guarantees for a family of sampling-free approximations in the context of Bayesian inference. These require "classical" assumptions.

Performance Guarantees for Existing Methods

The Problem: Bayesian inference requires an intractable integral. Most solutions use computationally-heavy sampling techniques to approximate.

Existing Solutions: Numerical approximations to integrals have been proposed for a long time, but theory not complete.

Our Contributions (Bilodeau, Stringer, and Tang, 2021b)

- First theoretical guarantees for a family of sampling-free approximations in the context of Bayesian inference. These require "classical" assumptions.
- Computational method that uses such approximations for a full Bayesian inference estimation procedure.

Performance Guarantees for Existing Methods

The Problem: Bayesian inference requires an intractable integral. Most solutions use computationally-heavy sampling techniques to approximate.

Existing Solutions: Numerical approximations to integrals have been proposed for a long time, but theory not complete.

Our Contributions (Bilodeau, Stringer, and Tang, 2021b)

- First theoretical guarantees for a family of sampling-free approximations in the context of Bayesian inference. These require "classical" assumptions.
- Computational method that uses such approximations for a full Bayesian inference estimation procedure.
- Efficient implementation of the method in an R package.

Performance Guarantees for Existing Methods

The Problem: Bayesian inference requires an intractable integral. Most solutions use computationally-heavy sampling techniques to approximate.

Existing Solutions: Numerical approximations to integrals have been proposed for a long time, but theory not complete.

Our Contributions (Bilodeau, Stringer, and Tang, 2021b)

- First theoretical guarantees for a family of sampling-free approximations in the context of Bayesian inference. These require “classical” assumptions.
- Computational method that uses such approximations for a full Bayesian inference estimation procedure.
- Efficient implementation of the method in an R package.

Next Steps for Theory

Performance Guarantees for Existing Methods

The Problem: Bayesian inference requires an intractable integral. Most solutions use computationally-heavy sampling techniques to approximate.

Existing Solutions: Numerical approximations to integrals have been proposed for a long time, but theory not complete.

Our Contributions (Bilodeau, Stringer, and Tang, 2021b)

- First theoretical guarantees for a family of sampling-free approximations in the context of Bayesian inference. These require "classical" assumptions.
- Computational method that uses such approximations for a full Bayesian inference estimation procedure.
- Efficient implementation of the method in an R package.

Next Steps for Theory

- Develop an improved method with theory for the more general procedure.

Performance Guarantees for Existing Methods

The Problem: Bayesian inference requires an intractable integral. Most solutions use computationally-heavy sampling techniques to approximate.

Existing Solutions: Numerical approximations to integrals have been proposed for a long time, but theory not complete.

Our Contributions (Bilodeau, Stringer, and Tang, 2021b)

- First theoretical guarantees for a family of sampling-free approximations in the context of Bayesian inference. These require "classical" assumptions.
- Computational method that uses such approximations for a full Bayesian inference estimation procedure.
- Efficient implementation of the method in an R package.

Next Steps for Theory

- Develop an improved method with theory for the more general procedure.
- Define a novel paradigm to better understand the role of assumptions.

Performance Guarantees for Existing Methods

The Problem: Bayesian inference requires an intractable integral. Most solutions use computationally-heavy sampling techniques to approximate.

Existing Solutions: Numerical approximations to integrals have been proposed for a long time, but theory not complete.

Our Contributions (Bilodeau, Stringer, and Tang, 2021b)

- First theoretical guarantees for a family of sampling-free approximations in the context of Bayesian inference. These require "classical" assumptions.
- Computational method that uses such approximations for a full Bayesian inference estimation procedure.
- Efficient implementation of the method in an R package.

Next Steps for Theory

- Develop an improved method with theory for the more general procedure.
- Define a novel paradigm to better understand the role of assumptions.

Summary

Summary

- Theoreticians must make assumptions to obtain guarantees, but these assumptions should be made with practice in mind.

Summary

- Theoreticians must make assumptions to obtain guarantees, but these assumptions should be made with practice in mind.
- Theory can impact practice by introducing new methods, understanding existing methods, or precisely quantifying the difficulty of statistical tasks.

Summary

- Theoreticians must make assumptions to obtain guarantees, but these assumptions should be made with practice in mind.
- Theory can impact practice by introducing new methods, understanding existing methods, or precisely quantifying the difficulty of statistical tasks.
- Theoretical guarantees often disagree with empirical observations, suggesting that the assumptions do not reflect reality.

Summary

- Theoreticians must make assumptions to obtain guarantees, but these assumptions should be made with practice in mind.
- Theory can impact practice by introducing new methods, understanding existing methods, or precisely quantifying the difficulty of statistical tasks.
- Theoretical guarantees often disagree with empirical observations, suggesting that the assumptions do not reflect reality. **In my opinion, these are some of the most interesting open problems.**

Summary

- Theoreticians must make assumptions to obtain guarantees, but these assumptions should be made with practice in mind.
- Theory can impact practice by introducing new methods, understanding existing methods, or precisely quantifying the difficulty of statistical tasks.
- Theoretical guarantees often disagree with empirical observations, suggesting that the assumptions do not reflect reality. **In my opinion, these are some of the most interesting open problems.**
- So...how do you get started doing theoretical research?

Summary

- Theoreticians must make assumptions to obtain guarantees, but these assumptions should be made with practice in mind.
- Theory can impact practice by introducing new methods, understanding existing methods, or precisely quantifying the difficulty of statistical tasks.
- Theoretical guarantees often disagree with empirical observations, suggesting that the assumptions do not reflect reality. **In my opinion, these are some of the most interesting open problems.**
- So...how do you get started doing theoretical research?
  Pick a task people try and do in practice, and then start understanding the assumptions surrounding it that are both
    a) mathematically convenient and
    b) reflect reality to some extent.

Summary

- Theoreticians must make assumptions to obtain guarantees, but these assumptions should be made with practice in mind.
- Theory can impact practice by introducing new methods, understanding existing methods, or precisely quantifying the difficulty of statistical tasks.
- Theoretical guarantees often disagree with empirical observations, suggesting that the assumptions do not reflect reality. **In my opinion, these are some of the most interesting open problems.**
- So...how do you get started doing theoretical research?
  Pick a task people try and do in practice, and then start understanding the assumptions surrounding it that are both
    a) mathematically convenient and
    b) reflect reality to some extent.

S. Arora, R. Ge, B. Neyshabur, and Y. Zhang. Stronger generalization bounds for deep nets via a compression approach. In *Proceedings of the 35th International Conference on Machine Learning*, 2018.

P. L. Bartlett. For valid generalization, the size of the weights is more important than the size of the network. In *Advances in Neural Information Processing Systems 9*, 1996.

P. L. Bartlett, D. J. Foster, and M. Telgarsky. Spectrally-normalized margin bounds for neural networks. In *Advances in Neural Information Processing Systems 31*, 2017.

B. Bilodeau, D. J. Foster, and D. M. Roy. Tight bounds on minimax regret under logarithmic loss via self-concordance. In *Proceedings of the 37th International Conference on Machine Learning*, 2020a.

B. Bilodeau, J. Negrea, and D. M. Roy. Relaxing the i.i.d. assumption: Adaptively minimax optimal regret via root-entropic regularization, 2020b. arXiv:2007.06552.

B. Bilodeau, D. J. Foster, and D. M. Roy. Minimax rates for conditional density estimation via empirical entropy, 2021a. In preparation.

B. Bilodeau, A. Stringer, and Y. Tang. Stochastic convergence rates and applications of adaptive quadrature in Bayesian inference, 2021b. arXiv:2102.06801.

D. P. Kingma and J. L. Ba. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representation*, 2015.

S. J. Reddi, S. Kale, and S. Kumar. On the convergence of Adam and beyond. In *Proceedings of the 6th International Conference on Learning Representation*, 2018.