Variational Bayes Markov Chain Monte Carlo Bayesian Inference Probability (statistics)

Statistics (academic discipline) Machine Learning

When should I prefer variational inference over MCMC for Bayesian analysis?

For example, if I have a really large amount of computational time and power, should I ever use variational inference? The only thing I can think of that makes variational inference better in some situations is its speed or perhaps its theoretical guarantees.

4 Answers

Jason Eisner, computer science professor at Johns Hopkins
22.3k Views • Upvoted by Jay Verkuilen and Jordan Boyd-Graber
Most Viewed Writer in Bayesian Inference

[A2A] Speed is indeed the main reason to use variational methods. David Blei told me long ago, "Variational inference is that thing you implement while waiting for your Gibbs sampler to converge." :-)

Variational inference may get better results than running MCMC for the same amount of time.

Bias: However, variational inference is irredeemably biased, whereas MCMC's bias approaches 0 as you run the Markov chain for longer and longer. (In fact, you can make the bias *exactly* 0 by using perfect sampling .) So if you have unlimited computational resources, then variational inference will lose on bias. Variational inference will even lose to trivial algorithms such as brute force enumeration or rejection sampling, which are far slower but have bias of exactly 0.

Variance: The sample variance of an MCMC estimate (or a rejection sampling estimate)

Related Questions

What is variational inference?

Bayesian Inference: Are there any text books on Baysian data analysis on the same level as "Introduction to Error Analysis" by John Taylor?

What are some nice Bayesian analysis projects that can be done in one month?

When should I use variational inference vs. expectation-maximization for fitting a Gaussian mixture model?

What is variational Bayes?

What papers should I read if I want to understand Variational Bayesian methods well?

Bayesian Inference: Which companies develop or use Bayes Net/Graphical Model tools?

Are there any tech companies that use Bayesian MCMC?

Bayesian Inference: What is a Dirichlet process in layman's terms?

Bayesian Inference: Is Edwin Thompson Jaynes right that frequentism is only applicable to a limited set of problems?

usually approaches 0 as you draw more and more samples. This means that you don't have to worry about it if you have unlimited computational resources.

A variational estimate already has sample variance of exactly 0 because it's deterministic

parameters (or the observations). This lets you tune the model parameters to get the correct estimates on training data (I have papers about that). However, that's not a convincing argument for variational estimation, since it's not hard to adapt the idea to use MCMC (details on request).

Written Jun 27, 2015 • View Upvotes • Answer requested by Palmer Lao

Related Questions

More Answers Below

What is variational inference?

Bayesian Inference: Are there any text books on Baysian data analysis on the same level as "Introduction to Error Analysis" by John Taylor?

What are some nice Bayesian analysis projects that can be done in one month?

When should I use variational inference vs. expectation-maximization for fitting a Gaussian mixture model?

What is variational Bayes?

Jordan Boyd-Graber, Assistant Prof working on Machine Learning at U Colorado

1.7k Views • Upvoted by Jay Verkuilen and Jason Eisner

Jordan has 90+ answers in Machine Learning

Jason Eisner's answer is right on the money. Some additional reasons to use variational inference:

- if you have non-conjugate distributions, it is often more straightforward to use variational distributions and much faster than falling back to MH sampling.

- variational inference is trivial to parallelize, since you can create a variational distribution that matches your computational topology. [Page on colorado.edu]
- variational inference is also trivial to turn into an online algorithm, which can also be faster if you don't have access to a cluster (SVI in the lingo).

Updated Jan 17 • View Upvotes

Wray Buntine, Professor, Bayesian computation and document analysis
1k Views

Yes, so colleagues Eisner and Boyd-Graber answer the question pretty much.

Some additional comments:

1. When working on exponential family distributions with the KL divergence, it is the (Fisher) Scoring algorithm in the approximation space, which is approximate Newton-Raphson. Hence its speed!
2. I've got two instances where variational methods perform poorly: (1) "bursty" modelling for LDA where one is required to do variational inference on really small samples (a single document), and (2) hierarchical Pitman-Yors where fast collapsed Gibbs samplers operate on a Boolean space whereas the corresponding variational algorithms (on the trees of Beta distributions from stick-breaking) operate on a far more complex and deeply nested vector space. Both in our KDD 2014 paper. Both are peculiar cases in my mind.

Seem to remember the Bayesian network crowd long ago showed variational algorithms didn't perform well for some more complex networks, and this spawned a whole "beyond variational algorithms" industry including things like Minka's Expectation Propagation (EP). All of these work brilliantly sometimes and fail other times. EP is poor when applied to LDA, whereas variational algorithms work brilliantly with LDA.

Updated Jul 25, 2015 • View Upvotes

Jay Verkuilen, PhD Psychometrics, MS Mathematical Statistics, UIUC
808 Views • Most Viewed Writer in Statistics (academic discipline) with 1050+ answers

I almost think you answered your own question. I last wrote a sampler myself several years back, but while setting up *an* MCMC sampler isn't that hard, setting up a *good* MCMC sampler is. So the fact that Variational Bayes is deterministic is a pretty big benefit. You always get the same answer (given the same starting values and an objective function without huge local optima problems). Beyond that, the speed benefits can be substantial. VB can be used to narrow down models to reasonable candidates for subsequent MCMC as well.

Updated Mar 20 • View Upvotes

Related Questions

What papers should I read if I want to understand Variational Bayesian methods well?

Bayesian Inference: Which companies develop or use Bayes Net/Graphical Model tools?

Are there any tech companies that use Bayesian MCMC?

Bayesian Inference: What is a Dirichlet process in layman's terms?

Bayesian Inference: Is Edwin Thompson Jaynes right that frequentism is only applicable to a limited set of problems?

What are variational autoencoders?

Bayesian Inference: In what way is Bayes' law unscientific?

When should I use MCMC rather than Stochastic Gradient Descent?

What is Hamiltonian MCMC?

Are stochastic variational approaches the way to do large scale Bayesian ML or do you see any hope of scaling up MCMC-based algorithms?

What are some good texts on MCMC?

Bayesian Inference: How to integrate over a parameter distributed according to a Dirichlet Process?

Does string theory = Bayesian inference?

What is Bayesian inference in statistics?

What is bayesian analysis?

Top Stories

Why aren't supporters of Bernie Sanders voting for Jill Stein?

Jacob Sommer, Citizen, student of politics (BA), longtime Sanders admirer

2.8k Views • Jacob has 870+ answers in Politics of the United States of America

Bernie Sanders supporters are not voting for Jill Stein because they don't know who she is, or if they do know then they feel that Bernie is considerably more electable with similarly good policies. I know Jill Stein. I interviewed her during her 2010 run for the governorship of Massachusetts and we had a

Read More

What would your reason be for picking a Democrat like Hillary Clinton instead of a Republican like Ben Carson?

David Stewart, Interested observer.

58.5k Views • Upvoted by Peter Flom, BA in political science. Avid follower of US politics. and Rod Young, Was a Community Organizer and activist for the Green Party of Tennessee, Former David has 570+ answers in Politics

Ignore party politics for a moment and cast aside everything the two candidates have said regarding policy and issues. If you were

Read More

Why is Greece so poor?

Alec Cawley, Physics degree followed by forty years of embedded software

2.9k Views • Most Viewed Writer in Eurozone

Because it has never got rich. It is a tiny country on the fringes of Europe, consisting mostly of mountain. It tore itself out of the Ottoman Empire in the first half of the nineteenth century, at which time it was about as poor as it could be. A nation of subsistence farmers and fishermen, with nothing that you could call infrastructure of any sort. It has few natural resources, not

Read More