

Stan is the future, and it's here now

Sean van der Merwe

University of the Free State, South Africa

23 August 2019

DON'T PANIC

Overview

1 What is Stan?

2 Examples

3 Conclusion

About Stan

- <https://mc-stan.org/>
- “Stan is a state-of-the-art platform for statistical modeling and high-performance statistical computation.”
- You type out your model in Stan language
- Then Stan does all the work, including:
 - Simulating the whole posterior distribution,
 - Simulating the approximate posterior, and/or
 - Maximum likelihood

Advantages of Stan

- You specify the model once and then fit it as much as you want, however you want
- It compiles your model into C++ code, so it's *very* fast
- Latest simulation methods: NUTS and HMC, better than old Gibbs and MH
- If your model is super complicated or simulation is too slow, then you can do variational Bayes or Maximum Likelihood
 - You don't change your model, just one line of code
- Most importantly, if you want to change your model then you change only the part of the model that needs changing
 - No changing software, no installing new packages, no changing notation for a different 'PROC', etc.

Main Advantage of Stan: Bayes

- Stan automatically picks practical priors, so you don't have to
- But you can easily add your own priors if you like
- Gives all the advantages of Bayes without effort:
 - Prediction is easy on any scale
 - Intervals are better because they incorporate more sources of variation and don't have to be symmetric
 - Probabilities can be estimated directly

Why not use Stan for everything?

- Basic tools are better for basic problems
- If you really only want to do a standard regression, or fit a standard distribution, then Stan is overkill
 - You don't need to build a model if all you need is descriptive statistics
 - Stan is fast, but a standard regression is 'instant'
 - Standard models have standard interpretations
- Stan is new so it needs explanation
 - Most of my clients haven't heard of SAS or R — often only hear about a p-value when an external assessor asks for one
- Stan is for modelling, so don't use it for exploratory data analysis
- Don't use Stan for data mining tasks

Overview

1 What is Stan?

2 Examples

3 Conclusion

Meta analysis

- What if you know the variance of each observation, but not the true mean?
- What if that mean depends on explanatory factors?
- What if different levels of an explanatory factor have different variances in their means?
- Standard software does not allow us to answer these questions properly, so I was forced to learn Stan
- With Stan it doesn't matter how weird the model is, as long as you can write it down in standard notations
- This model is rather complicated so we won't go into it
 - I recommend you learn Stan with easy models, not the hard way like I did

Fitting a distribution

- What if you want to fit a distribution to data, but it's not in any statistics package?
- Or maybe you want to play with the priors?
- Let's look at two examples:
 - 1 <https://seanvdm.co.za/post/tfit1/>
 - 2 <https://seanvdm.co.za/post/simulationsversterv1/>

Regression

- What if you want to fit a logistic regression model and get prediction intervals?
- Let's look at the famous Challenger O-ring failure question:
- <https://bookdown.org/egarpor/PM-UC3M/glm-challenger.html>
- <https://seanvdm.co.za/post/challenger1/>

Overview

1 What is Stan?

2 Examples

3 Conclusion

The point

- *Stan = freedom + power*
- It does the standard models, but
 - you can use distributions different from what you are used to
 - you can use new distributions that don't even have names yet
 - you can violate assumptions and model those violations
 - you can add random effects, changing variances, weights, *etc.* without stress
 - you can add priors and expert knowledge as if it is the most natural thing in the world
- So grab statistics modelling freedom and power for yourself, then teach it to your students!