STSCI 4780/5780
The Stan platform
for probabilistic programming

Tom Loredo, CCAPS & SDS, Cornell University

Stan: mc-stan.org

About Stan

Stan is a state-of-the-art platform for statistical modeling and high-performance statistical computation. Thousands of users rely on Stan for statistical modeling, data analysis, and prediction in the social, biological, and physical sciences, engineering, and business.

Users specify log density functions in Stan's probabilistic programming language and get:

- full Bayesian statistical inference with MCMC sampling (NUTS, HMC)
- approximate Bayesian inference with variational inference (ADVI)
- penalized maximum likelihood estimation with optimization (L-BFGS)

`http://mc-stan.org/`

Stan forum: `https://discourse.mc-stan.org/`

Stan has over 50 active core developers: statisticians, computer scientists, and users in various disciplines

Core Developers (in order of joining)

Andrew Gelman
Columbia University

Bob Carpenter
Flatiron Institute

Daniel Lee

Ben Goodrich
Columbia University

Michael Betancourt
Independent Consultant

Dan Simpson
University of Toronto

Yi Zhang
Metrum Research Group

Paul Bürkner
University of Stuttgart

Steve Bronder
Capital One, Columbia University

Rok Cesnovar
University of Ljubljana

. . .

"Stan"?

Stanislaw Ulam, namesake of Stan and co-inventor of Monte Carlo methods (Metropolis and Ulam, 1949), shown here holding the Fermiac, Enrico Fermi's physical Monte Carlo simulator for neutron diffusion.

Image from (Giesler, 2000).

(Stan developers report that the name also alludes to the Eminem song, "Stan")

The Stan platform

- A probabilistic programming language for specifying and exploring complex probability distributions
- A C++ library implementing probability distributions, mathematical functions, optimizers, MCMC samplers...
- A source-to-source (Stan to C++) compiler that generates C++ code from a model specified in the Stan language, making use of the Stan library

Stan relies on platform-specific C++ compilers to generate the code the Stan compiler produces

Stan's foundations

Stan v1 was released in Aug 2012; v2 in Oct 2013 (current: 2.29)

Stan brings together three key technologies that weren't well-integrated in 2012

- Probabilistic programming languages — BUGS, JAGS...
- Hamiltonian Monte Carlo (HMC) gradient-based MCMC algorithm
- Automatic differentiation (Autodiff)

A key innovation is Stan's No U-Turn Sampler (NUTS) for tuning the parameters of the HMC algorithm (Hoffman & Gelman 2014)

The Stan ecosystem

- The Stan language, library, and compiler
- CmdStan — Using compiled Stan models via the command line
- Language interfaces: R, Python, Julia, MATLAB, Scala, Stat, Mathematica
- General-purpose tools built on Stan: RStanArm, brms...

Python access

- `CmdStanPy`: Python functions that run command-line Stan programs, communicating data between those programs and Python via files
- `PyStan`:
  - v2: Classes directly accessing compiled Stan libraries, sharing data via C++ data structures (Linux, macOS, Windows)
  - v3: Classess accessing libraries via an intermediate HTTP REST API (currently for a subset of Stan capabilities; no Windows support except via WSL)

Stan "Pumps" example (number counts!)

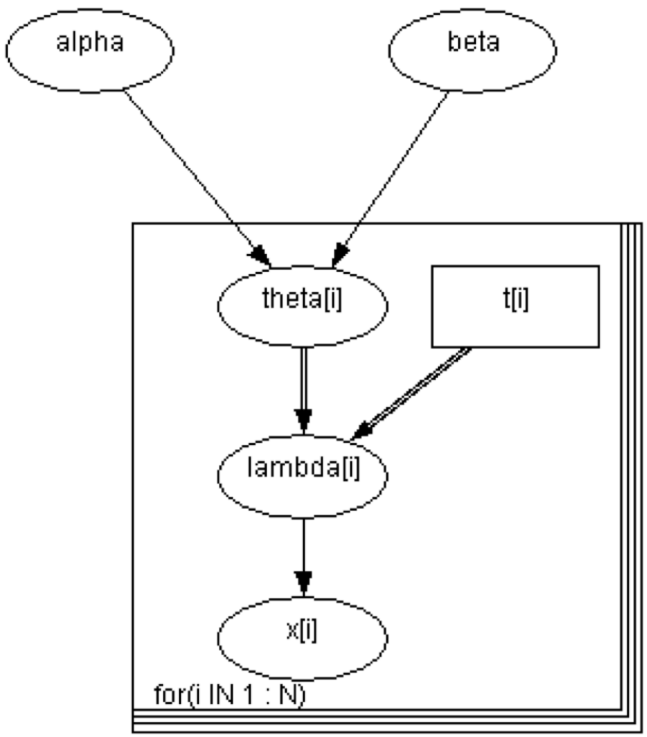
Power law slope

Flux $\theta_i \sim \text{Gamma}(\alpha, \beta)$

Exponential cutoff

Expected counts $\lambda_i = \theta_i t_i$

Observed counts $x_i \sim \text{Poisson}(\lambda_i)$



22 lines (18 sloc) | 0.313 kb

```
1
2  data {
3    int<lower=0> N;
4    int<lower=0> x[N];
5    real  t[N];
6  }
7
8  parameters {
9    real<lower=0> alpha;
10   real<lower=0> beta;
11   real<lower=0> theta[N];
12 }
13
14 model {
15   alpha ~ exponential(1.0);
16   beta ~ gamma(0.1, 1.0);
17   for (i in 1:N){
18     theta[i] ~ gamma(alpha, beta);
19     x[i] ~ poisson(theta[i] * t[i]);
20   }
21 }
```

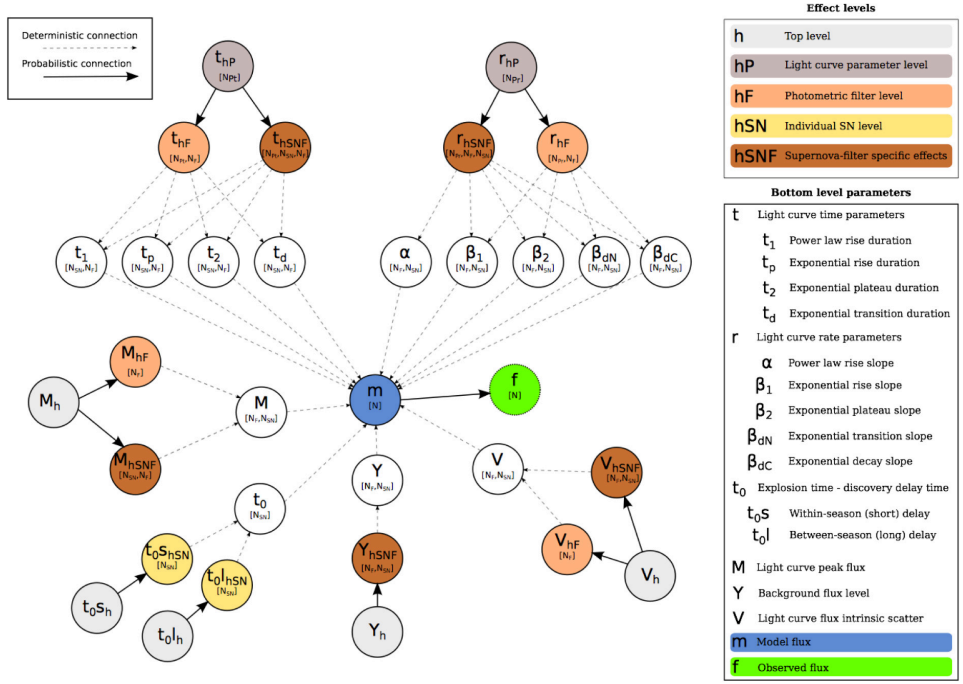
Inaugural "Stan model of the week"

arXiv.org > astro-ph > arXiv:1404.3619 Search

Astrophysics > Instrumentation and Methods for Astrophysics

Unsupervised Transient Light Curve Analysis Via Hierarchical Bayesian Inference

Nathan Sanders (CfA), Michael Betancourt (University of Warwick), Alicia Soderberg (CfA)



Models light curves of 20,000 Pan-STARRS1 observations of 80 SN IIP

Stan Store

Rainbow Stan logo - Coffee/Tea Mug
$14.99

Rainbow Stan logo - Women's T-Shirt
$18.99

Rainbow Stan logo - Men's T-Shirt
$18.99

New Logo Website - Men's Premium T-Shirt
$21.49

New Logo Website - Coffee/Tea Mug
$14.99

New Logo Website - Women's T-Shirt
$18.99

New Logo Website - Men's T-Shirt
$18.99

New Logo Website - Unisex Tri-Blend T-Shirt
$23.99

New Logo Website - Face Mask
$16.99

hamiltonian express tshirt png - Men's Ringer T-Shirt
$23.49

Support Stan

Support

Contribute to the Stan Project

Stan is now linked to NumFOCUS, a U.S. 501(c)(3) nonprofit organization that serves several other open-source software projects, including NumPy, Julia, Jupyter, and others.

Stan can now accept tax-deductible contributions through this foundation. If you're interested in contributing to Stan's ongoing development, you may contact us directly or just follow the link:

- Contribute to Stan (Salsa Labs payment processing)

It is also possible to donate to Stan via the GitHub sponsors program:

- Sponsor Stan via GitHub sponsors

Why Support Stan?

Stan is a worthy cause: our software is free and, as users can attest, we also give a lot of free and open help through the Stan Forums. By supporting Stan, you or your organization is helping to make this state-of-the-art Bayesian data analysis program even better. You're pushing forward the frontiers of science and helping your own work as well as others'.

Stan goods for good

- Stan t-shirts and mugs