## **ITERATING OVER STATISTICAL MODELS** NCAA TOURNAMENT EDITION



## WHY ARE YOU LISTENING TO ME?

Stan developer
<u>http://mc-stan.org</u>

Researcher at Columbia

 Co-founder of Stan Group training / statistical support / consulting



email: bearlee@alum.mit.edu
 twitter: @djsyclik / @mcmc\_stan
 web: http://syclik.com



#### WHAT IS MARCH MADNESS?

First Round March 17-18	Second Pound		16 Holv Cross	- 11	Vanderbilt
1 Kancas	March 19-20		16 Southern		Wichita St
		March 24-25			
	1	Louisiville, KY	-		
8 Colorado		1. I	Regional Finals		
9 Connecticut	1				
5 Maryland Spokane, WA	1	E.,			
12 S. Dakota St.	1				
4 California	1	1			1.5
13 Hawaii	1			Fina April	l Four 2
6 Arizona	1	500.	r H		
11 Vanderbilt/Wichita St.		<u></u>			
3 Miami	1	Louisville, KY			
14 Buffalo	1	P .			
Brooklyn, NY	7				
/ Iowa	h.	h			
	1				
2 Villanova					
15 UNC Asheville	1				
Spokane, WA	7				
10 Usly Gross (Southown		h			
		Anaheim, CA	-		
8 Saint Joseph's	<i>b</i>	pi			
9 Cincinnati	1				
5 Baylor	۲	L.			
12 Yale	1				
4 Duke	T				
13 UNC Wilmington	1	P			
6 Texas		WE:	<b>5</b> T		
11 Northern Iowa		h			
Oklahoma City, OK	1	Anaheim, CA			
14 Green Bay					
Oklahoma City, OK					
/ Oregon State		6			
2 Oklahoma	1				
15 Cal St Bakersfield	1				



#### WHAT IS MARCH MADNESS?

				First
F <b>irst Round</b> March 17-18	Second Round		16 Holy Cross	11 Vanderbil
Des Moines, IA	March 19-20	Regional Semis	16 Southern	11 Wichita St
16 Austin Peay	Kansas	March 24-25	KY	
B Colorado	1	Kansas	Regional Finals	
9 Connecticut	Connecticut		March 26-27	
F Maxuland Spokane, WA			Kansas	
	Maryland			
2 S. DdKOld Sl.		Maryland		
l California	Hawaii			<b>Final Four</b>
3 Hawaii	3	SOU	TH	April 2
B Arizona	Wichita St			Villanova
l Wichita St.	Wichita St.	Louisville, I	(Y	
Providence, RI	1	iviiami	1	
4 Buffalo	Miami			
Brooklyn, NY	1		Villanova	P:
	lowa		5	
Brooklyn, NY		Villanova		
2 Villanova	Villanova			
15 UNC Asheville				
1 Oregon				
6 Holy Cross	Oregon			
Spokane, WA		Anaheim, C	.́А	
B Saint Joseph's	St. Joseph's			
			Oregon	
5 Baylor	Vale			
12 Yale		Duke		
Providence, RI =	1	Duke		
13 UNC Wilmington	Duke			
Oklahoma City, OK		WE	<b>S.I.</b>	Oklahoma
1 Northern Iowa	Northern Iowa			
Oklahoma City, OK	7	Texas A&M		
M Groop Bay	Texas A&M			
Oklahama City, OK			Oklahoma	
7 Oregon State	VCU	-		-
10 VCU		Oklahoma		
2 Oklahoma City, OK	Oklahoma			
15 Cal St Bakersfield	ORIGIUIIId			



#### CHAMPIONSHIP GAME: VILLANOVA VS NORTH CAROLINA

## CHAMPIONSHIP GAME: VILLANOVA VS NORTH CAROLINA

# P(VILLANOVA WIN)?





- 1:00. 70 69
- ▶ 0:35. 72 69
- ▶ 0:23. 72 71



- ▶ 1:00. 70 69
- ▶ 0:35. 72 69
- ▶ 0:23. 72 71
- ▶ 0:13. 74 71



- ▶ 1:00. 70 69
- ▶ 0:35. 72 69
- ▶ 0:23. 72 71
- ▶ 0:13. 74 71
- ▶ 0:06. 74 74



- ▶ 1:00. 70 69
- ▶ 0:35. 72 69
- ▶ 0:23. 72 71
- ▶ 0:13. 74 71
- ▶ 0:06. 74 74
- 0:00. 77 74. Villanova wins.



NHL CFB RECRUITING CBB UFC SOCCER SHOP MORE

ALL 311 BLOGS 🛉 🕑

MARCH MADNESS 2016: NCAA TOURNAMENT COVERAG

#### The national championship game had the greatest ending in NCAA **Tournament history**

By Rodger Sherman 🕑 @rodger\_sherman on Apr 5, 2016, 12:09a 14



*The national championship game ended on a buzzer-beater — but it had more* than one shining moment.

#### Armour: Villanova win vs. UNC will be remembered as best title game ever

Nancy Armour 18 Hours Ago







Villanova-UNC was the best NCAA championship game ever

# TREAT AS CODE





#### WRITING CODE IS A DISCIPLINE

#### Pragmatic Programmer



from journeyman to master

#### Andrew Hunt David Thomas

Copyrighted Material



Robert C. Martin Series

#### MARTIN FOWLER

With Contributions by Kent Beck, John Brant William Opdyke, and Don Roberts

Foreword by Erich Gamma Object Technology International Inc.

#### **Clean Code**

**Copyrighted Material** 

A Handbook of Agile Software Craftsmanship



Foreword by James O. Coplie@opyrighted Internal



#### CAN BE WRITING CODE <del>IS</del> A DISCIPLINE

- design patterns
- testing
- code review
- maintenance
- modularity
- collaboration

#### Pragmatic Programmer



from journeyman to master

#### Andrew Hunt David Thomas

Copyrighted Material



Robert C. Martin Series

#### MARTIN FOWLER

With Contributions by Kent Beck, John Brant William Opdyke, and Don Roberts

Foreword by Erich Gamma Object Technology International Inc.

#### **Clean Code**

**Copyrighted Materia** 

A Handbook of Agile Software Craftsmanship



Foreword by James O. Coplietopyrighted Material



#### **IS STATISTICAL MODELING A DISCIPLINE?**

Art or science?

Models have names

Statistical model vs implementation

Collaboration on the statistical model?

#### **Texts in Statistical Science**

#### **Bayesian Data Analysis**

#### **Third Edition**



Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin

> CRC Press Tyler & Functs Group CHAPMAN & HALL BOOK Copyrighted Material

#### ANALYTICAL METHODS FOR SOCIAL RESEARCH



#### Data Analysis Using Regression and Multilevel/Hierarchical Models

ANDREW GELMAN JENNIFER HILL

Matorial

#### Doing Bayesian Data Analysis

A Tutorial with R, JAGS, and Stan



John K. Kruschke



## WHAT DO WE NEED TO DO

Elevate statistical models: first class entity

Modularize

Language

Discuss subtle details

Collaborate

## **STAN GETS US CLOSE**

- statistical modeling language
  - domain-specific language has its own grammar; not R or BUGS!
  - rstan

shinystan, RStudio integration, rstanarm

open-source core libraries are new BSD



#### Stan program

- plain text
- plays nicely with source repositories
- imperative language

Note: Stan isn't the only thing you can do this with



## MODELING BASKETBALL



## **BASKETBALL HISTORY**

- 1891
- Dr. James Naismith. Springfield, MA
- Non-contact conditioning
- 13 simple rules
- Peach basket

## **BASKETBALL HISTORY**

- 1891
- Dr. James Naismith. Springfield, MA
- Non-contact conditioning
- 13 simple rules
- Peach basket

10. The umpire shall be judge of the men and shall note the fouls and notify the referee when three consecutive fouls have been made. He shall have power to disqualify men according to Rule 5.

## **BASKETBALL NOW**

- 2 x 20 min half
- Increasing score
- Points increment by 2, 3, and 1
- 5 players, unlimited substitutions
- player DQ: 5th foul
- Bonus: 7 team fouls
   Double bonus: 10 team fouls

#### Shot Chart



#### DATA

- ► 351 NCAA Division 1 Men's basketball teams
- ► 33 conferences
- 5421 games
- > 24 35 games per team
- Max 3 observations





## TALL DATA VS WIDE DATA

Tall data
 lots of replications

Wide datalots of fields

day, home, score, ot, fgm, fga, 3pm, 3pa, 3a, fta, ftm, or, dr, ast, to, stl, blk, pf



#### 1. Set up full probability model

#### 2. Condition on observed data

#### 3. Evaluate the fit of the model

Set up full probability model
 Write a Stan program

2. Condition on observed data

3. Evaluate the fit of the model

Set up full probability model
 Write a Stan program

2. Condition on observed data Run RStan

3. Evaluate the fit of the model

1. Set up full probability model Write a Stan program

2. Condition on observed data Run RStan

3. Evaluate the fit of the model R, ShinyStan, posterior predictive checks

## **TERATING OVER** MODELS



## **STATISTICAL MODEL #1**

- Only 2015-2016 matters
- Teams have a latent ability
- "logistic regression"
- "Bradley-Terry model"





```
data {
  int N_games;
  int W_id[N_games];
  int L_id[N_games];
transformed data {
  int<lower=1> win[N_games];
  win <- rep_array(1, N_games);</pre>
parameters {
  vector[351] team_ability;
model {
  team_ability \sim normal(0, 1);
```

win ~ bernoulli\_logit(team\_ability[W\_id] - team\_ability[L\_id]);

source("dat\_2015.dat.R") fit <- stan("ncaa.stan")</pre> print(fit, pars=c("team\_ability[326]", "team\_ability[205]"))

Inference for Stan model: model1. 4 chains, each with iter=2000; warmup=1000; thin=1; post-warmup draws per chain=1000, total post-warmup draws=4000.

mean se\_mean sd 2.5% 25% 50% 75% 97.5% n\_eff Rhat team\_ability[326] 2.82 0.01 0.48 1.93 2.50 2.80 3.13 3.80 4000 team\_ability[205] 1.82 0.01 0.39 1.05 1.55 1.82 2.07 2.59 4000 1

Samples were drawn using NUTS(diag\_e) at Sat Apr 9 01:12:53 2016. For each parameter, n\_eff is a crude measure of effective sample size, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat=1).

## P(VILLANOVA > UNC | DATA) = 0.73







Save & Close	SHINYSTAN	🗐 DIAGNO	SE 📲 E	STIMATE	EXPLORE	MORE -
Select parameter					Rhat	n_ef
team_ability[326]		•			1	4000
Multiview			y-axis			
Bivariate			team_	ability[20	)5]	
Trivariate			□ Shov	v/Hide O	ptions	
Density			3 -			
Histogram						
			team_ability[205]			
			1-			

terminated its evolution normally.

2





For Stan models using the NUTS algorithm, red points indicate iterations that encountered a divergent transition. Yellow points indicate a transition that hit the maximum treedepth rather than



## TREAT STATISTICAL MODELS AS CODE

- Statistical model in a separate file
- **Git**
- Testing
  - Inspection of fit
  - Backtest on historical data
- Priors
- "Model #1"

## IN WRITING, YOU MUST KILL ALL YOUR DARLINGS.







## **STATISTICAL MODEL #2**

Home court advantage!

- Teams have a latent ability
- "logistic regression"
- "Bradley-Terry model"

 $y \sim \text{bernoulli}(\text{logit}^{-1}(\alpha + \theta_1 - \theta_2))$ 

```
data {
  int N_games;
  int W_id[N_games];
  int L_id[N_games];
  int W_h_ind[N_games];
transformed data {
  int<lower=1> win[N_games];
 win <- rep_array(1, N_games);</pre>
parameters {
  vector[351] team_ability;
  vector[3] court_advantage;
model {
  team_ability \sim normal(0, 1);
  court_advantage \sim normal(0, 1);
  win ~ bernoulli_logit(court_advantage[W_h_ind]
```

+ team\_ability[W\_id] - team\_ability[L\_id]);



Inference for Stan model: model2. 4 chains, each with iter=2000; warmup=1000; thin=1; post-warmup draws per chain=1000, total post-warmup draws=4000.

mean se\_mean sd 2.5% 25% 50% 75% 97.5% n\_eff Rhat team\_ability[205] 0.06 0.01 0.94 -1.77 -0.58 0.07 0.69 1.93 4000

Samples were drawn using NUTS(diag\_e) at Sat Apr 9 01:33:33 2016. For each parameter, n\_eff is a crude measure of effective sample size, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat=1).

## P(VILLANOVA > UNC | DATA) = 0.52



Inference for Stan model: model2. 4 chains, each with iter=2000; warmup=1000; thin=1; post-warmup draws per chain=1000, total post-warmup draws=4000.

	mean	se_mean	S
<pre>court_advantage[1]</pre>	6.46	0.01	0.40
court_advantage[2]	7.01	0.01	0.38
court_advantage[3]	5.48	0.01	0.44
team_ability[326]	0.13	0.02	0.97
team_ability[205]	0.06	0.01	0.94

Samples were drawn using NUTS(diag\_e) at Sat Apr 9 01:33:33 2016. For each parameter, n\_eff is a crude measure of effective sample size, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat=1).

d 2.5% 25% 50% 75% 97.5% n\_eff Rhat 0 5.74 6.18 6.44 6.72 7.30 4000 8 6.31 6.74 7.00 7.27 7.81 4000 4 4.69 5.18 5.45 5.77 6.41 4000 7 -1.68 -0.52 0.10 0.76 2.05 4000 4 -1.77 -0.58 0.07 0.69 1.93 4000



## **STATISTICAL MODEL #3**

- Assumptions:
  - Only 2015-2016 matters
  - Teams have a latent ability
  - Model points
  - Add a home court advantage

```
data {
  int N_games;
  vector<lower=0>[N_games] WL_spread;
  int W_id[N_games];
  int L_id[N_games];
  int W_h_ind[N_games];
parameters {
 vector[351] team_ability;
  real<lower=0> sigma;
  vector[3] court_advantage;
model {
  team_ability \sim normal(0, 1);
  court_advantage \sim normal(0, 1);
  sigma ~ normal(0, 1);
  WL_spread ~ normal(court_advantage[W_h_ind]
                      + team_ability[W_id] - team_ability[L_id],
                      sigma);
generated quantities {
  real championship_game;
  real villanova_wins;
  championship_game <- normal_rng(team_ability[326] - team_ability[205],</pre>
                                   sigma);
 villanova_wins <- if_else(championship_game > 0, 1, 0);
```

Inference for Stan model: model3. 4 chains, each with iter=2000; warmup=1000; thin=1; post-warmup draws per chain=1000, total post-warmup draws=4000.

mean se\_mean sd 2.5% 25% 50% 75% 97.5% n\_eff Rhat 8.33 0.00 0.08 8.16 8.27 8.33 8.38 8.49 sigma 4000 team\_ability[326] 2.21 0.01 0.80 0.61 1.67 2.20 2.75 3.77 4000 team\_ability[205] 2.22 0.01 0.79 0.70 1.68 2.21 2.76 3.78 4000 championship\_game -0.24 0.13 8.35 -17.13 -5.87 -0.15 5.37 16.42 4000 villanova\_wins 0.49 0.01 0.50 0.00 0.00 0.00 1.00 3616 1.00

Samples were drawn using NUTS(diag\_e) at Sat Apr 9 02:07:34 2016. For each parameter, n\_eff is a crude measure of effective sample size, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat=1).





## HOW DID WE DO?

#### Kaggle: 87 / 608

#### What went wrong?

#### What's next?

# TREAT AS CODE





#### THANKS

#### Collaborative effort

NCAA modeling: **Rob Trangucci** 

#### Stan team

Andrew Gelman, Bob Carpenter, Matt Hoffman, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, Allen Riddell, Marco Ignacio, Jeff Arnold, Mitzi Morris, Rob Goedman, Brian Lau, Jonah Gabry, Alp Kucukelbir, Robert Grant, Dustin Tran, Krzysztof Sakrejda, Aki Vehtari, Rayleigh Lei, Sebastian Weber

#### HELP

- http://mc-stan.org
- stan-users mailing list



A Fiscally Sponsored Project of **OPEN CODE = BETTER SCIENCE** 

Stan Group Inc. http://stan.fit training / statistical support / consulting

bearlee@alum.mit.edu / @djsyclik / @mcmc\_stan





