Generative Models: What do they know? Do they know things? Let's find out!

J. Moreno (javier@interaxon.ca / @infrahumano)

DevTO Meetup - February 27, 2017

Probability

In the usual approach, the probability of an event is the frequency of manifestations after a large number of trials.

Confusing: we seem to have an intuition of probability for events that only happen once. (e.g., the probability of Donald Trump winning the election.)

Question: Does probability make sense in that context? (Does *anything* make sense ever since?)

Probability

In the usual approach, the probability of an event is the frequency of manifestations after a large number of trials.

Confusing: we seem to have an intuition of probability for events that only happen once. (e.g., the probability of Donald Trump winning the election.)

Question: Does probability make sense in that context? (Does *anything* make sense ever since?)

Probability

In the usual approach, the probability of an event is the frequency of manifestations after a large number of trials.

Confusing: we seem to have an intuition of probability for events that only happen once. (e.g., the probability of Donald Trump winning the election.)

Question: Does probability make sense in that context? (Does *anything* make sense ever since?)

Probability

In the usual approach, the probability of an event is the frequency of manifestations after a large number of trials.

Confusing: we seem to have an intuition of probability for events that only happen once. (e.g., the probability of Donald Trump winning the election.)

Question: Does probability make sense in that context? (Does *anything* make sense ever since?)

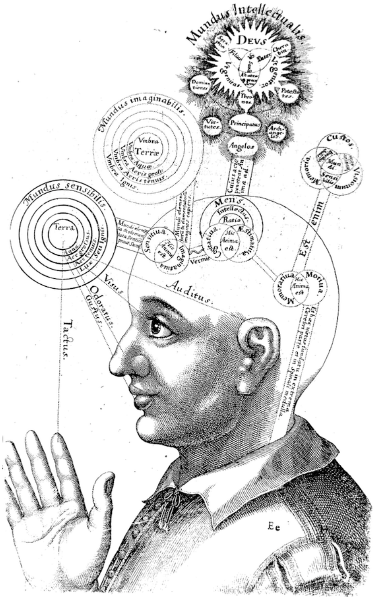
Probability

In the usual approach, the probability of an event is the frequency of manifestations after a large number of trials.

Confusing: we seem to have an intuition of probability for events that only happen once. (e.g., the probability of Donald Trump winning the election.)

Question: Does probability make sense in that context? (Does *anything* make sense ever since?)

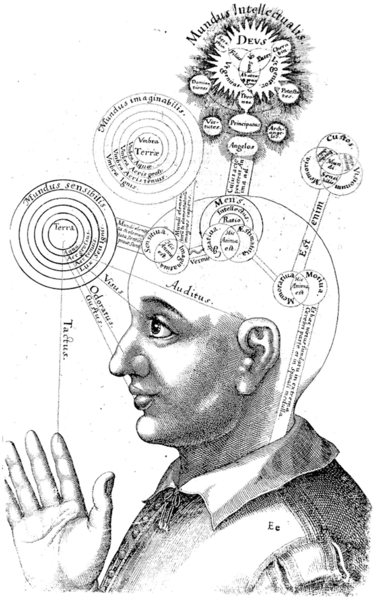
Knowledge



Instead of thinking of probability as a count of frequencies you may think of it as a quantification of (un)certainty given a state of knowledge.

This can be fully formalized. (It is what is called *The Bayesian Approach*.)

Knowledge

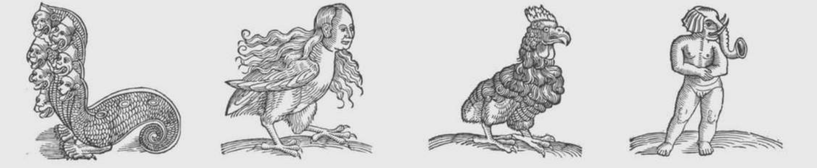


Instead of thinking of probability as a count of frequencies you may think of it as a quantification of (un)certainty given a state of knowledge.

This can be fully formalized. (It is what is called *The Bayesian Approach*.)

Got news

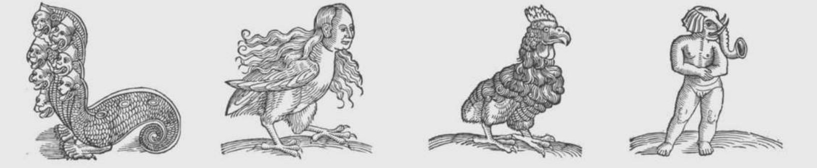
Good news: this totally formalizable version of probability comes by default with our operative system. We feel we have a sense for it.



Bad news: the implementation is quite poor. People have won Nobel prices for carefully proving this: our analytic heuristics do not understand randomness (among other things).

Got news

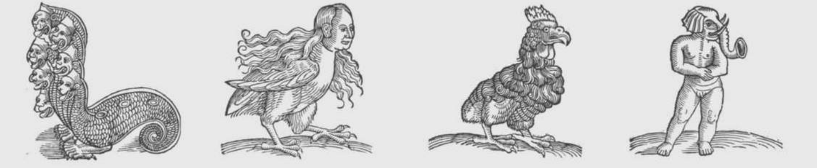
Good news: this totally formalizable version of probability comes by default with our operative system. We feel we have a sense for it.



Bad news: the implementation is quite poor. People have won Nobel prices for carefully proving this: our analytic heuristics do not understand randomness (among other things).

Got news

Good news: this totally formalizable version of probability comes by default with our operative system. We feel we have a sense for it.



Bad news: the implementation is quite poor. People have won Nobel prices for carefully proving this: our analytic heuristics do not understand randomness (among other things).

It can't be THAT bad

It is!

HHHTTT versus HTHTTH

It can't be THAT bad

It is!

HHHTTT versus HTHTTH

But who cares?

We should! We rely on data to take decisions. We collect data and build products on it. We ingest data, clean it, organize it, and even try to predict future outcomes.

"Every company is a data company." — @athomasq (Quandl)

Data is randomness digitalized. For data to be truly useful, we need to keep in mind its probabilistic nature.

But who cares?

We should! We rely on data to take decisions. We collect data and build products on it. We ingest data, clean it, organize it, and even try to predict future outcomes.

"Every company is a data company." — @athomasq (Quandl)

Data is randomness digitalized. For data to be truly useful, we need to keep in mind its probabilistic nature.

But who cares?

We should! We rely on data to take decisions. We collect data and build products on it. We ingest data, clean it, organize it, and even try to predict future outcomes.

"Every company is a data company." — @athomasq (Quandl)

Data is randomness digitalized. For data to be truly useful, we need to keep in mind its probabilistic nature.

But who cares?

We should! We rely on data to take decisions. We collect data and build products on it. We ingest data, clean it, organize it, and even try to predict future outcomes.

“Every company is a data company.” — @athomasq (Quandl)

Data is randomness digitalized. For data to be truly useful, we need to keep in mind its probabilistic nature.

What's the risk?

When extracting information from data, let yourself be guided by naïve intuitions (or half-understood recipes) regarding probability and you may end up in really dark places.

There is a way!

Be methodic (and brave.)

One method — one of many

- It is creative. Not based on blindly following recipes. Encourages productive thinking.
- The mathematics, statistics and computer science that make it possible are beautiful.
- It puts uncertainty and doubt at the center of the game. Perfect for the insecure.
- It provides you with a general framework to ask probabilistic questions about your data.
- It is based on building stories. And I like stories.

One method — one of many

- It is creative. Not based on blindly following recipes. Encourages productive thinking.
- The mathematics, statistics and computer science that make it possible are beautiful.
- It puts uncertainty and doubt at the center of the game. Perfect for the insecure.
- It provides you with a general framework to ask probabilistic questions about your data.
- It is based on building stories. And I like stories.

One method — one of many

- It is creative. Not based on blindly following recipes. Encourages productive thinking.
- The mathematics, statistics and computer science that make it possible are beautiful.
- It puts uncertainty and doubt at the center of the game. Perfect for the insecure.
- It provides you with a general framework to ask probabilistic questions about your data.
- It is based on building stories. And I like stories.

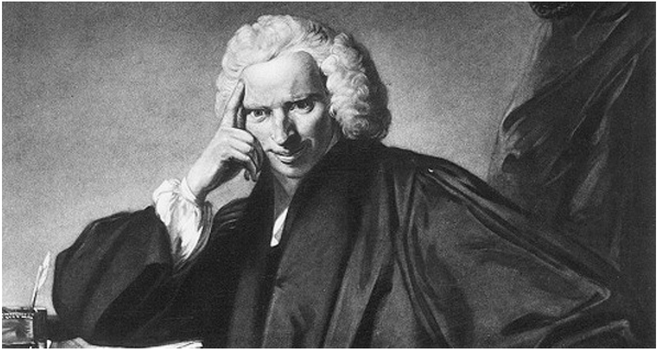
One method — one of many

- It is creative. Not based on blindly following recipes. Encourages productive thinking.
- The mathematics, statistics and computer science that make it possible are beautiful.
- It puts uncertainty and doubt at the center of the game. Perfect for the insecure.
- It provides you with a general framework to ask probabilistic questions about your data.
- It is based on building stories. And I like stories.

One method — one of many

- It is creative. Not based on blindly following recipes. Encourages productive thinking.
- The mathematics, statistics and computer science that make it possible are beautiful.
- It puts uncertainty and doubt at the center of the game. Perfect for the insecure.
- It provides you with a general framework to ask probabilistic questions about your data.
- It is based on building stories. And I like stories.

The approach



Making explicit all sources of uncertainty you may have, come up with a plausible story that explains how data came into being.

A common problem

Repeated measurements of similar events for different individuals:

Individual	Type	Location	Extra info	Y (food intake in Kg)
1	Therapod	Great Valley	$\cdots$	90.5
1	Therapod	Drought Land	$\cdots$	40.7
2	Therapod	Great Valley	$\cdots$	102.9
3	Sauropod	Great Valley.	$\cdots$	712.4
3	Sauropod	Drought Land	$\cdots$	345.2
3	Sauropod	Drought Land	$\cdots$	210.3
4	Sauropod	Great Valley	$\cdots$	476.7
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

We would like to model

$$y_{i,g} = f(d_{i,g}, p_g) + \epsilon_{i,g}$$

With ϵ normally distributed (maybe unknown variance.)

A common problem

Repeated measurements of similar events for different individuals:

Individual	Type	Location	Extra info	Y (food intake in Kg)
1	Therapod	Great Valley	$\cdots$	90.5
1	Therapod	Drought Land	$\cdots$	40.7
2	Therapod	Great Valley	$\cdots$	102.9
3	Sauropod	Great Valley.	$\cdots$	712.4
3	Sauropod	Drought Land	$\cdots$	345.2
3	Sauropod	Drought Land	$\cdots$	210.3
4	Sauropod	Great Valley	$\cdots$	476.7
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

We would like to model

$$y_{i,g} = \alpha_g + d_{i,g}\beta_g + \epsilon_{i,g}$$

With ϵ normally distributed (maybe unknown variance.)

A common problem

Repeated measurements of similar events for different individuals:

Individual	Type	Location	Extra info	Y (food intake in Kg)
1	Therapod	Great Valley	$\cdots$	90.5
1	Therapod	Drought Land	$\cdots$	40.7
2	Therapod	Great Valley	$\cdots$	102.9
3	Sauropod	Great Valley.	$\cdots$	712.4
3	Sauropod	Drought Land	$\cdots$	345.2
3	Sauropod	Drought Land	$\cdots$	210.3
4	Sauropod	Great Valley	$\cdots$	476.7
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

We would like to model

$$y_{i,g} \sim \text{Normal}(\alpha_g + d_{i,g}\beta_g, \sigma)$$

With priors for α_g and β_g reflecting (shared) current doubts.

All sources of uncertainty, I said

Perhaps some of the measurements in $d_{i,g}$ deserve explicit uncertainty! In that case, new relations may pop up in addition to the basic model.

Everything is connected

Three possible stories (could be many more):

- Happy families: groups do not matter:

$$y_{i,g} \sim \text{Normal}(\alpha + d_{i,g}\beta, \sigma).$$

- Unhappy families: groups are completely different.

$$y_{i,g} \sim \text{Normal}(\alpha_g + d_{i,g}\beta_g, \sigma).$$

- Real families: groups have varying degrees of similarity.

$$y_{i,g} \sim \text{Normal}(\alpha_g + d_{i,g}\beta_g, \sigma)$$

plus something like

$$(\alpha_g, \beta_g')' \sim \text{MultiNormal}(\mu, \Sigma).$$

And priors on μ and Σ.

Everything is connected

Three possible stories (could be many more):

- Happy families: groups do not matter:

$$y_{i,g} \sim \text{Normal}(\alpha + d_{i,g}\beta, \sigma).$$

- Unhappy families: groups are completely different.

$$y_{i,g} \sim \text{Normal}(\alpha_g + d_{i,g}\beta_g, \sigma).$$

- Real families: groups have varying degrees of similarity.

$$y_{i,g} \sim \text{Normal}(\alpha_g + d_{i,g}\beta_g, \sigma)$$

plus something like

$$(\alpha_g, \beta_g')' \sim \text{MultiNormal}(\mu, \Sigma).$$

And priors on μ and Σ.

Everything is connected

Three possible stories (could be many more):

- Happy families: groups do not matter:

$$y_{i,g} \sim \text{Normal}(\alpha + d_{i,g}\beta, \sigma).$$

- Unhappy families: groups are completely different.

$$y_{i,g} \sim \text{Normal}(\alpha_g + d_{i,g}\beta_g, \sigma).$$

- Real families: groups have varying degrees of similarity.

$$y_{i,g} \sim \text{Normal}(\alpha_g + d_{i,g}\beta_g, \sigma)$$

  plus something like

$$(\alpha_g, \beta_g')' \sim \text{MultiNormal}(\mu, \Sigma).$$

  And priors on μ and Σ.

Acts

- The Prior: *prob*(Params)
- The Likelihood: *prob*(Data|Params)
- The Posterior: *prob*(Params|Data)

Problem : given prior and likelihood we would like to get the posterior.

General probabilistic nonsense tells us that

$$\mathit{prob}(\text{Params}|\text{Data}) = \frac{\mathit{prob}(\text{Data}|\text{Params})\mathit{prob}(\text{Params})}{\int_{\text{Params}} \mathit{prob}(\text{Data}|\text{Params})\mathit{prob}(\text{Params})}.$$

However, calculating this explicitly is usually impossible.

Acts

- The Prior: $\textit{prob}(\text{Params})$
- The Likelihood: $\textit{prob}(\text{Data}|\text{Params})$
- The Posterior: $\textit{prob}(\text{Params}|\text{Data})$

Problem : given prior and likelihood we would like to get the posterior.

General probabilistic nonsense tells us that

$$\textit{prob}(\text{Params}|\text{Data}) = \frac{\textit{prob}(\text{Data}|\text{Params})\textit{prob}(\text{Params})}{\int_{\text{Params}} \textit{prob}(\text{Data}|\text{Params})\textit{prob}(\text{Params})}.$$

However, calculating this explicitly is usually impossible.

Acts

- The Prior: $\textit{prob}(\text{Params})$
- The Likelihood: $\textit{prob}(\text{Data}|\text{Params})$
- The Posterior: $\textit{prob}(\text{Params}|\text{Data})$

Problem : given prior and likelihood we would like to get the posterior.

General probabilistic nonsense tells us that

$$\textit{prob}(\text{Params}|\text{Data}) = \frac{\textit{prob}(\text{Data}|\text{Params})\textit{prob}(\text{Params})}{\int_{\text{Params}} \textit{prob}(\text{Data}|\text{Params})\textit{prob}(\text{Params})}.$$

However, calculating this explicitly is usually impossible.

Acts

- The Prior: $\textit{prob}(\text{Params})$
- The Likelihood: $\textit{prob}(\text{Data}|\text{Params})$
- The Posterior: $\textit{prob}(\text{Params}|\text{Data})$

Problem : given prior and likelihood we would like to get the posterior.

General probabilistic nonsense tells us that

$$\textit{prob}(\text{Params}|\text{Data}) = \frac{\textit{prob}(\text{Data}|\text{Params})\textit{prob}(\text{Params})}{\int_{\text{Params}} \textit{prob}(\text{Data}|\text{Params})\textit{prob}(\text{Params})}.$$

However, calculating this explicitly is usually impossible.

In randomness we trust

Monte Carlo methods allow us to estimate the posterior from likelihood and prior by randomly (and systematically) travel around the space of parameters collecting samples.

Initial algorithms were theoretically great but not very practical. Too slow!

New algorithms (e.g., Hamiltonian Monte Carlo) rely on differential geometry and automatic differentiation to make the journey faster and efficient.

In randomness we trust

Monte Carlo methods allow us to estimate the posterior from likelihood and prior by randomly (and systematically) travel around the space of parameters collecting samples.

Initial algorithms were theoretically great but not very practical. Too slow!

New algorithms (e.g., Hamiltonian Monte Carlo) rely on differential geometry and automatic differentiation to make the journey faster and efficient.

In randomness we trust

Monte Carlo methods allow us to estimate the posterior from likelihood and prior by randomly (and systematically) travel around the space of parameters collecting samples.

Initial algorithms were theoretically great but not very practical. Too slow!

New algorithms (e.g., Hamiltonian Monte Carlo) rely on differential geometry and automatic differentiation to make the journey faster and efficient.

Weapon of choice

Stan: a probabilistic programming language

http://mc-stan.org/

Interfaces for Python, R, Matlab, Julia, &c.

Weapon of choice (Code Example)

Something like

$$y_i \sim \text{Normal}(\alpha + x_i\beta, \sigma)$$

with α and β with prior Normal$(0, 0.5)$ and σ uniform becomes:

```
data {
int<lower=0> N; // observations
vector[N] x; // predictors
vector[N] y; // outcomes
}
parameters {
real alpha;
real beta;
real<lower=0> sigma;
}
model {
beta ~ normal(0, 0.5); // prior on beta
alpha ~ normal(0, 0.5); // prior on alpha
y ~ normal(alpha + beta * x, sigma); // likelihood
}
```

Modern statistical workflow (Abridged)

1. Explore data.
2. Set up your model.
3. Try it out with synthetic data until you can recover chosen parameters.
4. Run it on actual data.
5. Check simulation for convergence.
6. Evaluate fit by generating new data and compare to given data.
7. Iterate (or enjoy.)

Modern statistical workflow (Abridged)

1. Explore data.
2. Set up your model.
3. Try it out with synthetic data until you can recover chosen parameters.
4. Run it on actual data.
5. Check simulation for convergence.
6. Evaluate fit by generating new data and compare to given data.
7. Iterate (or enjoy.)

Modern statistical workflow (Abridged)

1. Explore data.
2. Set up your model.
3. Try it out with synthetic data until you can recover chosen parameters.
4. Run it on actual data.
5. Check simulation for convergence.
6. Evaluate fit by generating new data and compare to given data.
7. Iterate (or enjoy.)

Modern statistical workflow (Abridged)

1. Explore data.
2. Set up your model.
3. Try it out with synthetic data until you can recover chosen parameters.
4. Run it on actual data.
5. Check simulation for convergence.
6. Evaluate fit by generating new data and compare to given data.
7. Iterate (or enjoy.)

Modern statistical workflow (Abridged)

1. Explore data.
2. Set up your model.
3. Try it out with synthetic data until you can recover chosen parameters.
4. Run it on actual data.
5. Check simulation for convergence.
6. Evaluate fit by generating new data and compare to given data.
7. Iterate (or enjoy.)

Modern statistical workflow (Abridged)

1. Explore data.
2. Set up your model.
3. Try it out with synthetic data until you can recover chosen parameters.
4. Run it on actual data.
5. Check simulation for convergence.
6. Evaluate fit by generating new data and compare to given data.
7. Iterate (or enjoy.)

Modern statistical workflow (Abridged)

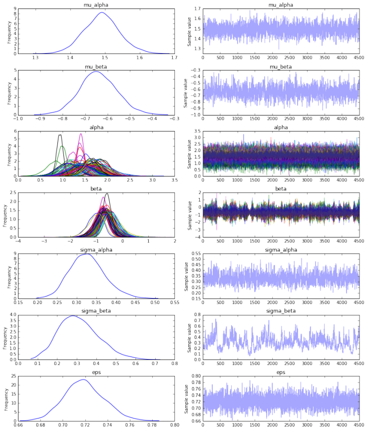
1. Explore data.
2. Set up your model.
3. Try it out with synthetic data until you can recover chosen parameters.
4. Run it on actual data.
5. Check simulation for convergence.
6. Evaluate fit by generating new data and compare to given data.
7. Iterate (or enjoy.)

Simulation is the real life

If your model survives the process you can:

- Run general inferences on the parameters.
- Forecast new observations taking into account the uncertainty of the parameters.
- Evaluate risk (for decision making).
- Update it with new data!

Crucial: every piece of information comes accompanied with explicit estimations of uncertainty.



Posterior plots from: http://twiecki.github.io/

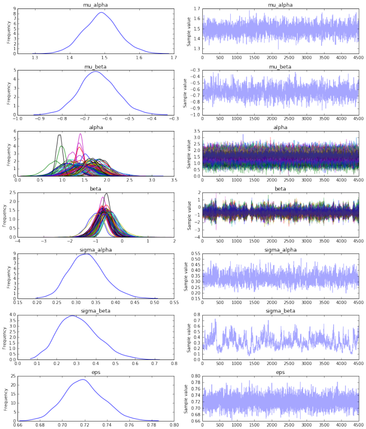
(Who actually uses PyMC3)

Simulation is the real life

If your model survives the process you can:

- Run general inferences on the parameters.
- Forecast new observations taking into account the uncertainty of the parameters.
- Evaluate risk (for decision making).
- Update it with new data!

Crucial: every piece of information comes accompanied with explicit estimations of uncertainty.



Posterior plots from: http://twiecki.github.io/

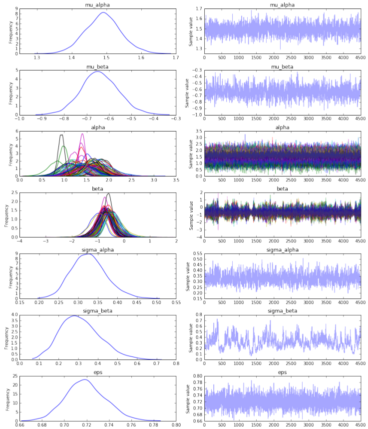
(Who actually uses PyMC3)

Simulation is the real life

If your model survives the process you can:

- Run general inferences on the parameters.
- Forecast new observations taking into account the uncertainty of the parameters.
- Evaluate risk (for decision making).
- Update it with new data!

Crucial: every piece of information comes accompanied with explicit estimations of uncertainty.



Posterior plots from: http://twiecki.github.io/

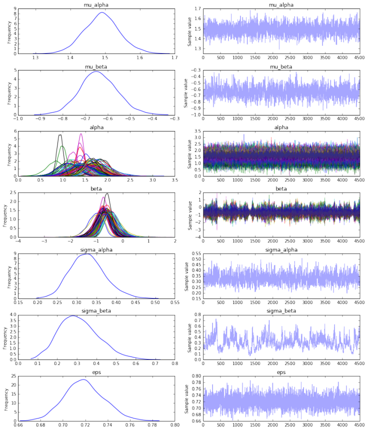
(Who actually uses PyMC3)

Simulation is the real life

If your model survives the process you can:

- Run general inferences on the parameters.
- Forecast new observations taking into account the uncertainty of the parameters.
- Evaluate risk (for decision making).
- Update it with new data!

Crucial: every piece of information comes accompanied with explicit estimations of uncertainty.



Posterior plots from: http://twiecki.github.io/

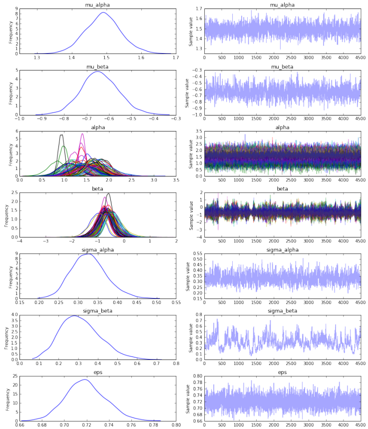
(Who actually uses PyMC3)

Simulation is the real life

If your model survives the process you can:

- Run general inferences on the parameters.
- Forecast new observations taking into account the uncertainty of the parameters.
- Evaluate risk (for decision making).
- Update it with new data!

Crucial: every piece of information comes accompanied with explicit estimations of uncertainty.



Posterior plots from: http://twiecki.github.io/

(Who actually uses PyMC3)

Some manifestations

- Lendable Marketplace: agilizing microloans in the developing world.
- Geometric Intelligence (Uber): small data smart systems.
- Whetlab (Twitter): neural network design.
- Quantopian (PyMC3 crowd): competitive algorithm trading platform.
- Academic and medical research: as a response to "the crisis of reproducibility".

Some manifestations

- Lendable Marketplace: agilizing microloans in the developing world.
- Geometric Intelligence (Uber): small data smart systems.
- Whetlab (Twitter): neural network design.
- Quantopian (PyMC3 crowd): competitive algorithm trading platform.
- Academic and medical research: as a response to "the crisis of reproducibility".

Some manifestations

- Lendable Marketplace: agilizing microloans in the developing world.
- Geometric Intelligence (Uber): small data smart systems.
- Whetlab (Twitter): neural network design.
- Quantopian (PyMC3 crowd): competitive algorithm trading platform.
- Academic and medical research: as a response to "the crisis of reproducibility".

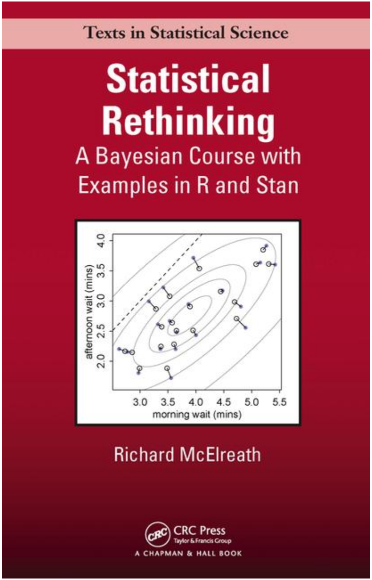
Some manifestations

- Lendable Marketplace: agilizing microloans in the developing world.
- Geometric Intelligence (Uber): small data smart systems.
- Whetlab (Twitter): neural network design.
- Quantopian (PyMC3 crowd): competitive algorithm trading platform.
- Academic and medical research: as a response to "the crisis of reproducibility".

Some manifestations

- Lendable Marketplace: agilizing microloans in the developing world.
- Geometric Intelligence (Uber): small data smart systems.
- Whetlab (Twitter): neural network design.
- Quantopian (PyMC3 crowd): competitive algorithm trading platform.
- Academic and medical research: as a response to "the crisis of reproducibility".

The truth is out there



Books

- Rethinking Statistics (R. McElreath)
- Doing Bayesian data analysis (J.K. Kruschke)
- Probabilistic Programming and Bayesian Methods for Hackers (C. Davidson Pilon)
- BDA3 (Gelman et al.)
- Stan manual

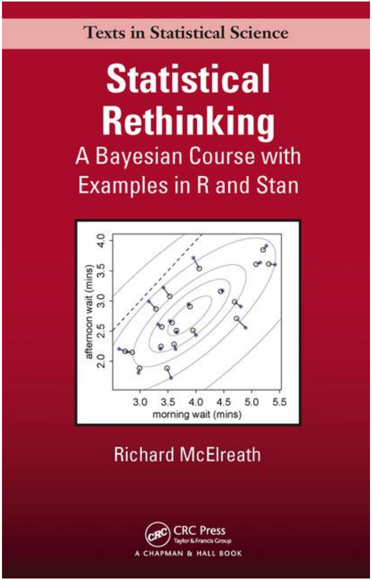
Blogs/Twitter

- Andrew Gelman and friends (@statsmodeling)
- Jim Savage (@jim_savage_)
- Thomas Wiecki (@twiecki)

Tools

- Stan
- PyMC3
- Edward

The truth is out there



Books

- Rethinking Statistics (R. McElreath)
- Doing Bayesian data analysis (J.K. Kruschke)
- Probabilistic Programming and Bayesian Methods for Hackers (C. Davidson Pilon)
- BDA3 (Gelman et al.)
- Stan manual

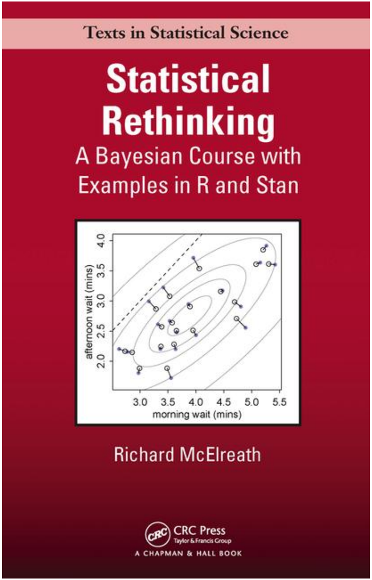
Blogs/Twitter

- Andrew Gelman and friends (@statsmodeling)
- Jim Savage (@jim_savage_)
- Thomas Wiecki (@twiecki)

Tools

- Stan
- PyMC3
- Edward

The truth is out there



Books

- Rethinking Statistics (R. McElreath)
- Doing Bayesian data analysis (J.K. Kruschke)
- Probabilistic Programming and Bayesian Methods for Hackers (C. Davidson Pilon)
- BDA3 (Gelman et al.)
- Stan manual

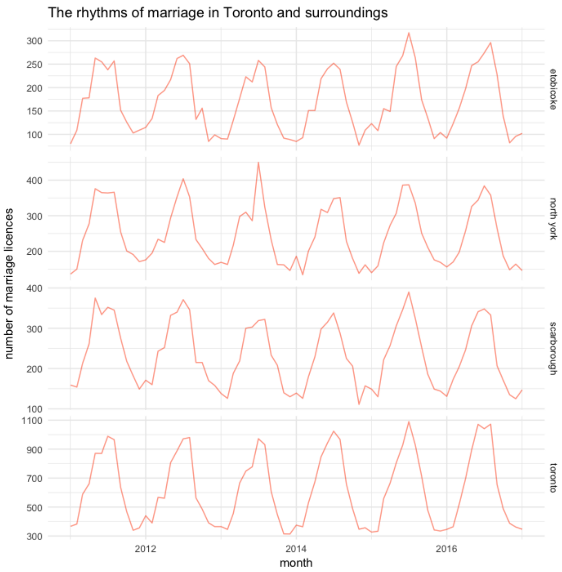
Blogs/Twitter

- Andrew Gelman and friends (@statsmodeling)
- Jim Savage (@jim_savage_)
- Thomas Wiecki (@twiecki)

Tools

- Stan
- PyMC3
- Edward

Extra: The 2015 Toronto Marriage Peak Mystery



Fin