Bayesian Analysis
A brief introduction

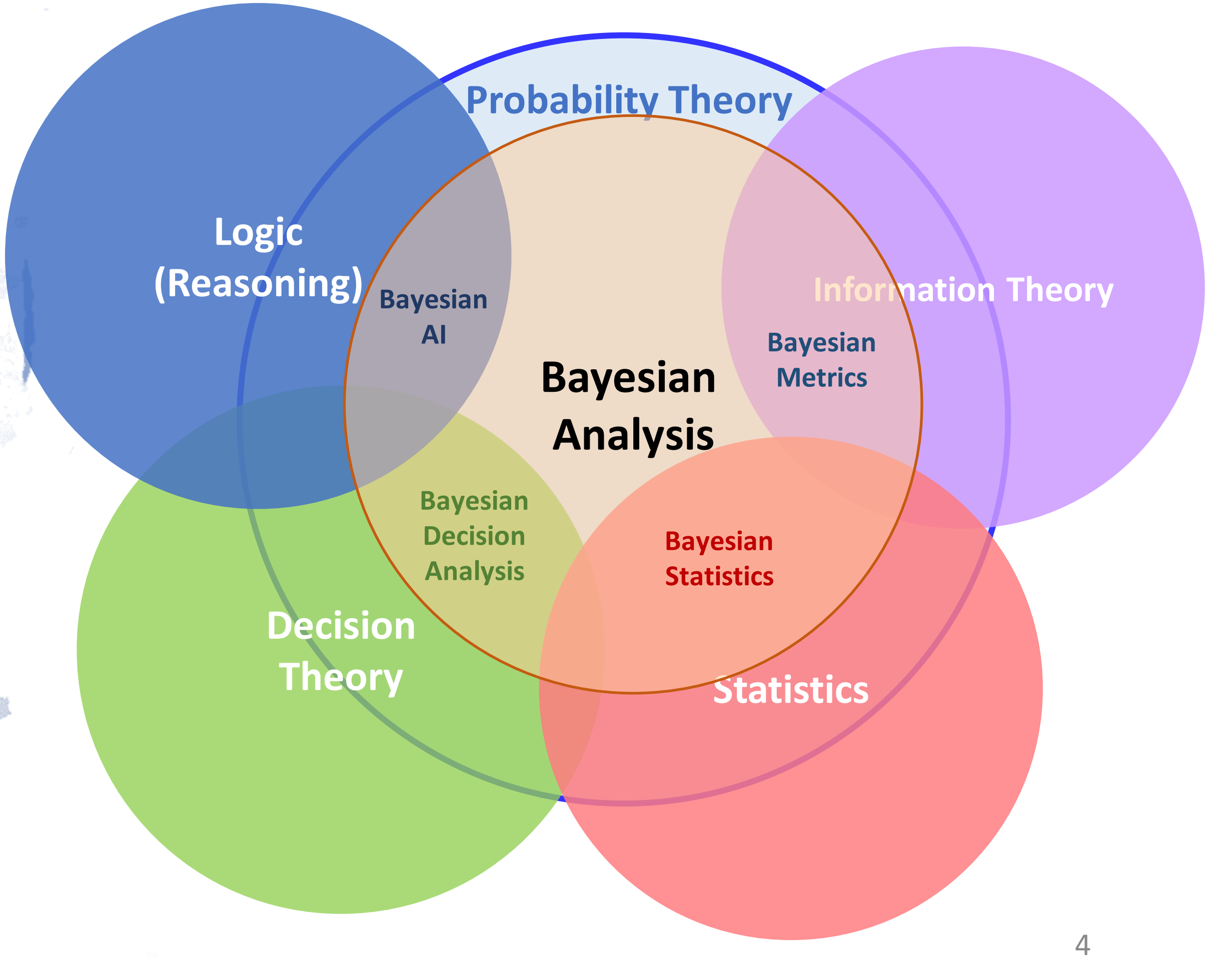
Michael L. Thompson, Ph.D.

April 26, 2021

Outline

- **What is Bayesian Analysis?**
  - Bayes Theorem: Core Concepts
- **How is Bayesian Analysis applied in AI & Data Science?**
  - Bayesian Model-based Machine Learning
- **How do you perform Bayesian Analysis?**
  - Bayesian Workflow & Tools
- **When is using Bayesian Analysis most strongly motivated?**
  - Key Traits to Look for in Your Problems
- **Where can you learn more about Bayesian Analysis?**
  - Resources & Readings

What is “Bayesian Analysis”?



Bayesian Analysis

- The application of **Bayes Theorem** …
  - To **reason** about **unknowns** by considering **evidence**, leveraging ***Probability Theory.***
  - To **infer predictions** by transforming **data**, leveraging ***Statistical Modeling.***
  - To **decide** on **actions** by accounting for **uncertainty**, leveraging ***Decision Theory.***

Bayes Theorem

Posterior

Prior

Likelihood

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)}$$

Marginal Likelihood

Hypothesis

Evidence

Bayes Theorem

(It has marvelous implications….)

Bayes Theorem

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)}$$

Bayesian:

Core Concepts

- **Probabilities** represent your **beliefs**
  - Acknowledge & quantify **uncertainty**
  - **Condition upon context** – ***all*** probabilities are conditional
- **Alternatives** are given full consideration
  - Account for **entire joint probability** of all possible combinations of evidence & hypotheses
- **Evidence** updates beliefs
  - **Prior belief** is essential – must not limit consideration to the strength of evidence

How is Bayesian Analysis applied in AI & Data Science?

Prevalence of Bayesian Applications

Product Ranking & Recommendations at Wayfair

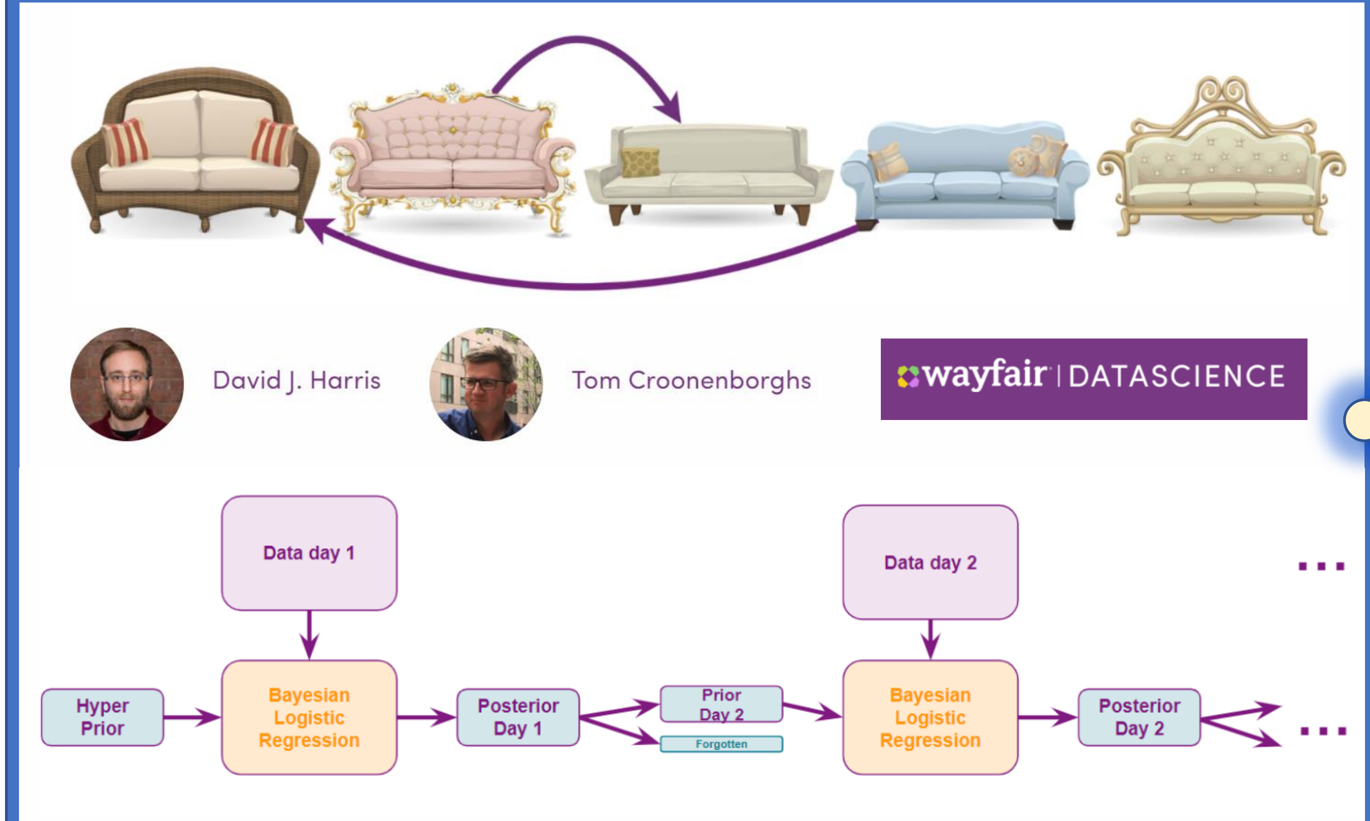


Fig. 4: A sketch of our daily update loop. Each day, our logistic regression model combines our observations of customer behavior with our prior knowledge to produce a posterior distribution. This posterior distribution then informs the next day's prior, but we intentionally "forget" a small portion of the information we have obtained, so that the model doesn't become too fixated on past performance.

"Wayfair has a huge catalog with over 14 million items. Our site features a diverse array of products for people's homes, with product categories ranging from 'appliances' to 'décor and pillows' to 'outdoor storage sheds.'

"At Wayfair, we are constantly working to improve our customers' shopping experiences…. This post features a new Bayesian system developed at Wayfair to (1) identify these products and (2) present them to our customers."

"Bayesian Product Ranking at Wayfair", Jan 20, 2020.

The Enabling Feature of Bayesian Analysis:
Multilevel Modeling & Updating → Personalization & Adaptation

Source:
"Bayesian Product Ranking at Wayfair", Harris & Croonenborghs, Wayfair DataScience blog post Jan. 20, 2020

Maximizing Return on Ad Spend at Smartly.io

At Smartly.io, we already had a model in place for the cost-per-action part that utilizes this funnel and is able to react fast to changes. It's used as part of **our budget allocation** that's based on Bayesian multi-armed bandits, and allocates the budget between multiple ad sets based on their performance. To change our Predictive Budget Allocation to optimize towards ROAS instead of CPA, we only needed to plug in stable estimates for the mean distributions of the revenue per conversion for multiple ads (or actually ad sets in the Facebook context as budget and target audience is defined on that level).

"Overall, about one million euros of advertising spend on daily level is managed with our Predictive Budget Allocation. In future, we see that Stan or some other probabilistic programming language plays a big role in the optimization features of Smartly.io."

"Tutorial: How We Productized Bayesian Revenue Estimation with Stan", Markus Ojala, Jun 21, 2017.

The Enabling Feature of Bayesian Analysis:

Uncertainty Quantification →

Risk Assessment & Mitigation

Source:
"Tutorial: How We Productized Bayesian Revenue Estimation with Stan",
Ojala, M., Smartly.io blog post 2017

How do you perform Bayesian Analysis – What is the workflow & what are the tools?

Bayesian Analysis
Fully pooled model

- **Example:** Product ranking, model the probability that a given customer will order product n:
  - $p = 1/(1 + \exp(-\alpha_{pop}))$
- **Hypothesis, *H*** – unknown parameter: α_{pop} ("pop"=population-level parameter)
- **Evidence, *E*** – given data:
  - y_n = number of customers ordering product n
  - K = number of times customers shown product n

Probabilistic Graphical Model (PGM):
Encodes our **conditional independence** (Markovian) assumptions about the joint probability distribution P(H,E).
(See **Judea Pearl** & *d-separation*.)
The Bayesian network is **a generative model** suitable for both simulation and inference.

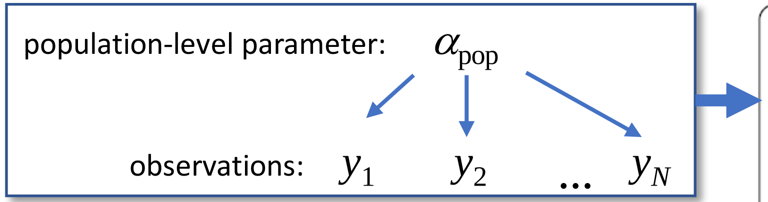
- Joint probability density function of **fully pooled** model:

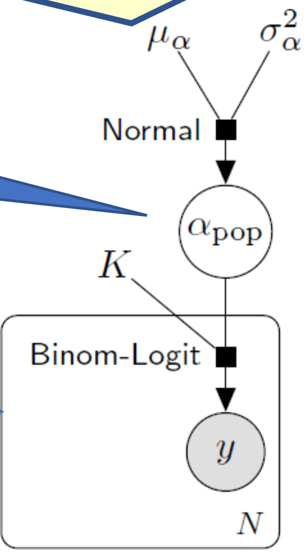
$$\underbrace{\mathrm{p}(y_1,\ldots,y_N,\alpha_{\text{pop}}|K,\mu_\alpha,\sigma_\alpha^2)}_{\text{P(H,E)}} = \underbrace{\mathrm{p}(\alpha_{\text{pop}}|\mu_\alpha,\sigma_\alpha^2)}_{\text{P(H)}} \underbrace{\prod_{n=1}^{N} \mathrm{p}(y_n|\alpha_{\text{pop}},K)}_{\text{P(E|H)}}$$

Distribution declaration:

$$\alpha_{\text{pop}} \sim \text{Normal}(\mu_\alpha,\sigma_\alpha^2)$$
$$\forall n \in \{1,\ldots,N\}:$$
$$y_n \sim \text{Binomial-Logit}(\alpha_{\text{pop}},K)$$

All N subjects share the exact same population-level parameter.





The *factor graph* of our Bayesian network.

Bayesian **Hierarchical/Multilevel** Models

Partially pooled model

- Joint probability density function of **partially pooled** model:

$$p(y_1, \ldots, y_N, \alpha_1, \ldots, \alpha_N, \sigma^2, \alpha_{\text{pop}} | K, \mu_\alpha, \sigma^2_\alpha, a_\sigma, b_\sigma) = p(\alpha_{\text{pop}} | \mu_\alpha, \sigma^2_\alpha)\, p(\sigma^2 | a_\sigma, b_\sigma) \prod_{n=1}^{N} p(y_n | \alpha_n, K) p(\alpha_n | \alpha_{\text{pop}}, \sigma^2)$$

Distribution declaration:

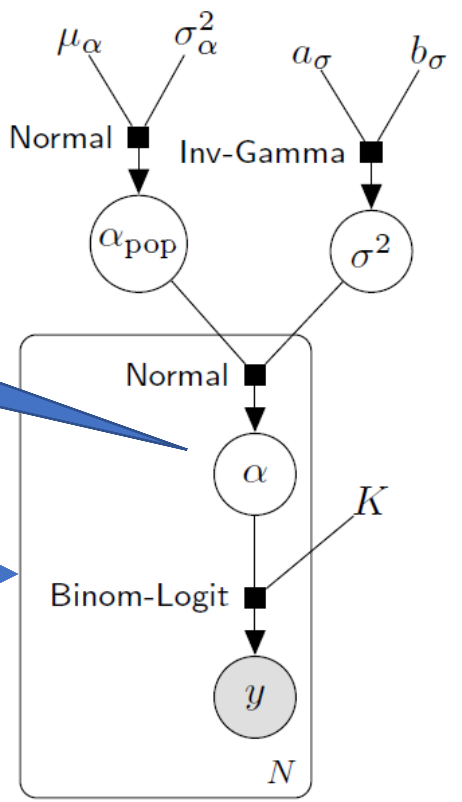
$$\begin{aligned} \alpha_{\text{pop}} &\sim \text{Normal}(\mu_\alpha, \sigma^2_\alpha) \\ \sigma^2 &\sim \text{Inv-Gamma}(a_\sigma, b_\sigma) \\ \forall n \in \{1, \ldots, N\} &: \\ \alpha_n &\sim \text{Normal}(\alpha_{\text{pop}}, \sigma^2) \\ y_n &\sim \text{Binomial-Logit}(\alpha_n, K) \end{aligned}$$

Each subject has its own parameter that depends upon the population-level parameter as the mean of its prior. So, all N subjects share information through the population-level parameter.

population-level parameters: $\alpha_{\text{pop}}, \sigma$

subject-level parameters: α_1 α_2 … α_N

observations: y_1 y_2 … y_N



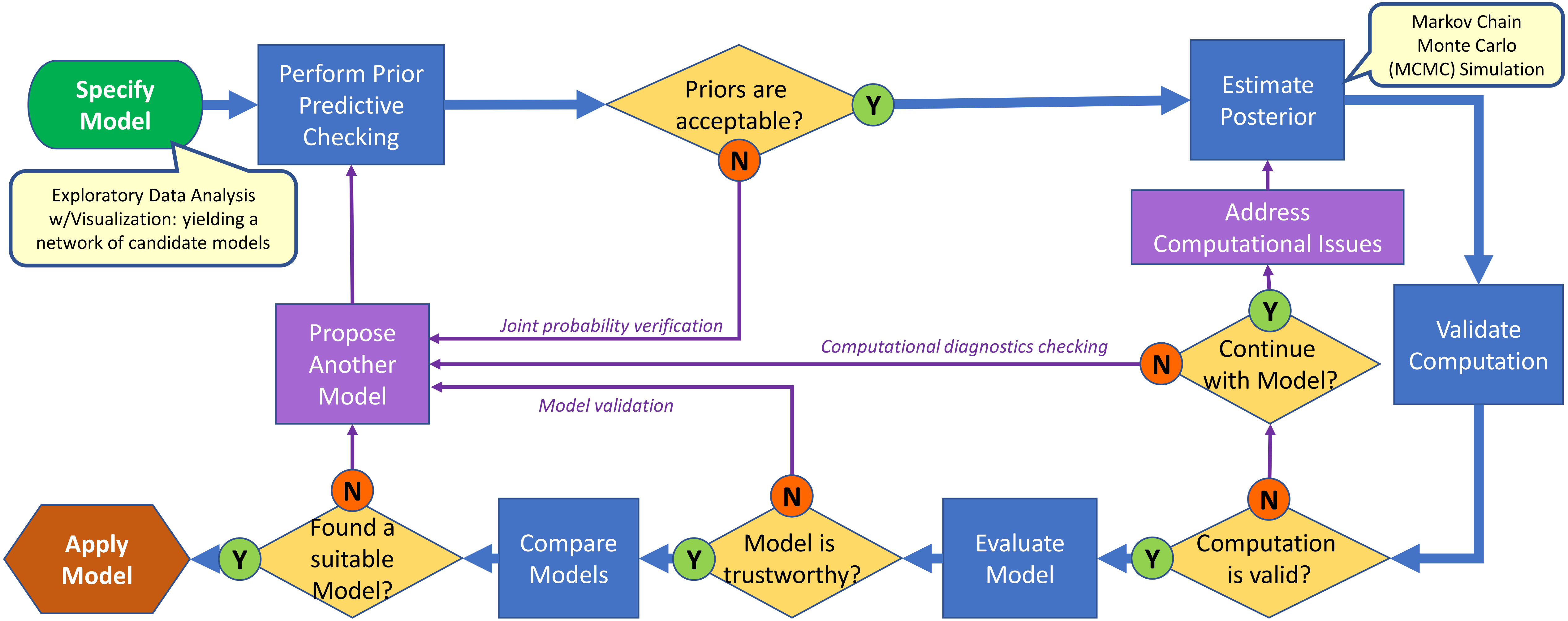
Tools for Bayesian Analysis

- **Probabilistic Programming Languages (PPL)**
  - Stan – C++ (Columbia University)
  - Pymc3 – Python
  - Pyro/NumPyro – Python/PyTorch (Uber AI)
  - TensorFlow Probability – Python/TensorFlow (Google)
  - Infer.NET – java (Microsoft)
  - Gen – Julia (MIT)
- **Bayesian Statistical Modeling & Machine Learning**
  - R Packages: **rstan** (BHM), **brms** (BHM), rstanarm (BHM), blavaan (SEM), **bnlearn** (BN), **gRain** (BN), HydeNet (ID), causal, prophet – Facebook (STSF)
  - Python Packages: pystan, prophet (STSF), **pgmpy** (PGM), bnlearn (BN)
  - Commercial: **BayesiaLab (BN)**, Hugin (BN), Netica (BN), AgenaRisk (ID)

- **PGM/BN** = **Probabilistic Graphical Models**/Bayesian Networks
- **BHM** = Bayesian Hierarchical/Multilevel Models
- **ID** = Influence Diagrams
- **SEM** = Structural Equations Models
- **STSF** = Structural Time Series & Forecasting

Bayesian Workflow

(adapted from Fig. 1, ["Bayesian Workflow", Prof. Andrew Gelman, *et al.*])



When is using Bayesian Analysis most strongly motivated?

Motivation for Bayesian Analysis:

Turn to Bayesian methods when faced with Data="Complex", Model="Complex", or Decision="Complex"

"Simple"

- **Data**
  - Single source
  - Single variable types/distribution families
  - Tabular & Ample – "Big Data"
    - Non-missing
    - Regular, exchangeable

Homogeneous

- **Model**
  - Observations linked to observations (Modeling the Data)
  - Empirical structure
    - Single-level
    - Correlative (acausal)
  - Single hypothesis
  - Component-level estimation; Low-level integration

Data-to-Data

- **Decision**
  - Deterministic assumptions
  - Modal/point estimate solutions
  - Predictive inference (What will happen?)
  - Single objective, Static

Deterministic Predictions

"Complex"

- **Data**
  - Multiple sources
  - Multiple variable types/distribution families
  - Ragged & Sparse
    - Missing
    - Multigranular aggregation

Heterogeneous

- **Model**
  - Latent spaces (Modeling the Domain)
  - Causal structure
    - Multi-level
    - Mechanisms
  - Mixture of phenomena/Multi-Hypothesis
  - System-level integration

True-to-True

- **Decision**
  - Reasoning under uncertainty (UQ)
  - Risk analysis
  - Explanatory inference (Why did it happen?)
  - Multi-Objective, Dynamic updating

Probabilistic Explanations

Where can you learn more about Bayesian Analysis?

Resources on Bayesian Analysis

- Basic Concepts: YouTube videos
  - "Bayes Theorem, and making probability intuitive" [15'45"],
    "The quick proof of Bayes' Theorem" [3'47"], and
    "The medical test paradox: Can redesigning Bayes rule help?" [21'13"]
  - "You Know I'm All About that Bayes: Crash Course Statistics #24" [12'04"] and
    "Bayes in Science and Everyday Life: Crash Course Statistics #25" [11'13"]
  - For more philosophical takes on Bayesian reasoning, check out these YouTube videos by Julia Galef:
    "Bayes: How one equation changed the way I think" [3'28"]
    "A visual guide to Bayesian thinking" [11'24"]

3Blue1Brown

CrashCourse

- How To: YouTube videos & blogs
  - Rasmus Baath's 3-part Tutorial "Bayesian Analysis" (1. What [29'29"], 2. Why [22'59"], 3. How [37'51"])
  - YouTube: "Corrie Bartelheimer: A Bayesian Workflow with PyMC and ArviZ | PyData Berlin 2019" [29'28"] (code is **here**)
  - Kurt, Will, "Count Bayesie" blog series, "A Guide to Bayesian Statistics", May 2, 2016
  - Fang & van de Schoot, "Intro to Bayesian (Multilevel) Generalised Linear Models (GLM) in R with brms" (2019)
  - Stan Development Team: Tutorials – Learn to use Stan; & Case Studies – Open-source methods & models

If you want to see how it's done, go here.

Stan

- Latest Research & Applications
  - Michael Thompson's Flipboard e-zine mashup: "Bayesian"
  - Prof. Andrew Gelman's blog:
    "Statistical Modeling, Causal Inference, and Social Science"

Thank you!