Bayesian Analysis
A brief introduction

Michael L. Thompson, Ph.D.

April 26, 2021

About the Presenter

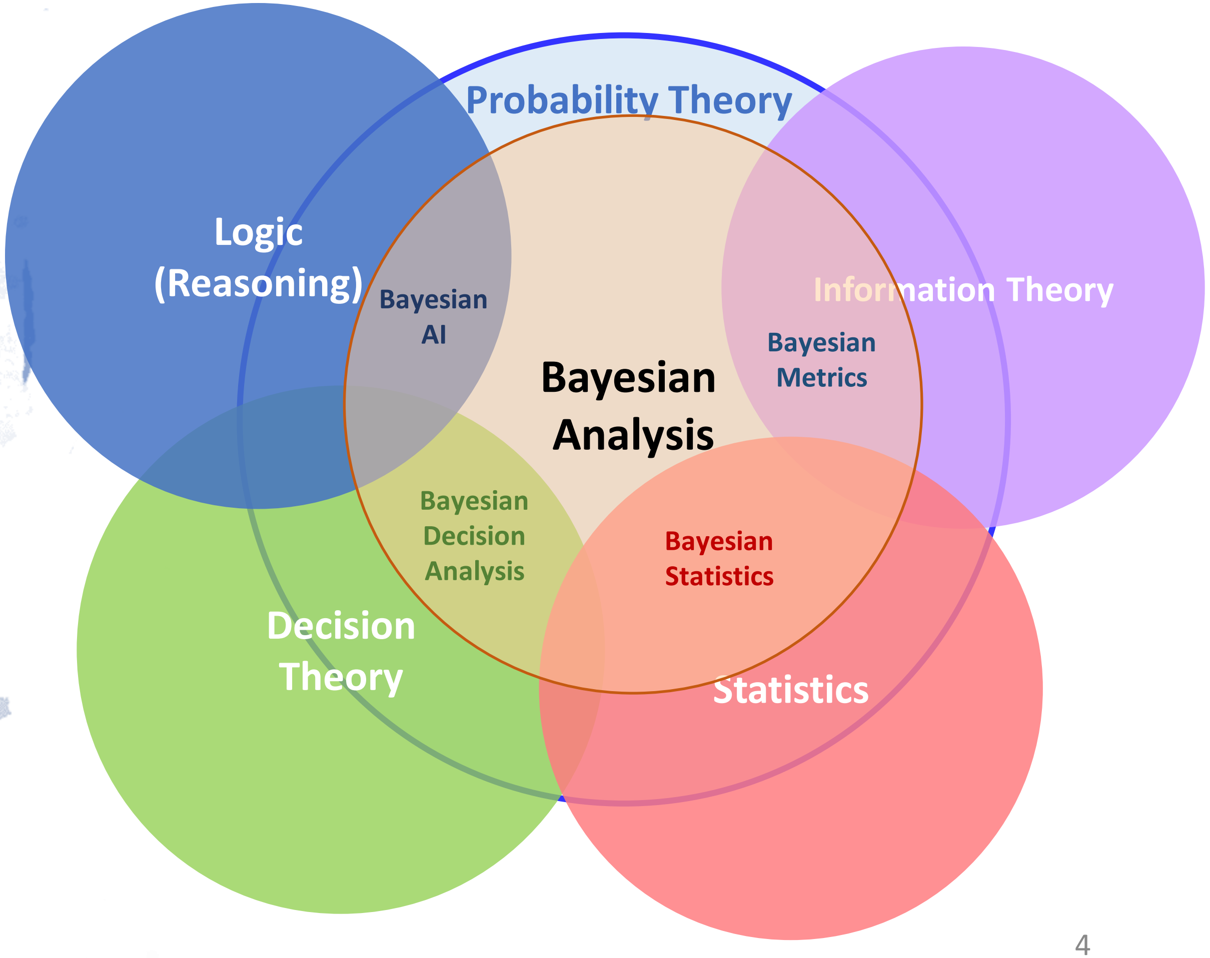
Hidden

Dr. Michael L. Thompson is retired from the Procter & Gamble Company, where he led Bayesian Analysis R&D in consumer & market modeling. His degrees are in Chemical Engineering: B.S., Northwestern University, '82; M.S., MIT, '84; and Ph.D., MIT, '96, with minor in Statistics and Artificial Intelligence. Michael has extensive experience in the process industry, having worked for Dow, Alcoa, Amoco, and Mitsubishi Chemical (Japan). At P&G for 21 years, Michael applied his expertise in Bayesian Analysis, especially Bayesian belief networks (BBN), to deliver results in the consumer-packaged goods (CPG) industry. His contributions spanned business functions including R&D, Engineering, Manufacturing, Marketing, and Business Analytics. He has authored journal articles ranging from fluidized bed reactors to hybrid probabilistic and first-principles biochemical models to optimal consumer product design. Currently, Michael is a Term Adjunct in the Lindner College of Business at the University of Cincinnati, where he teaches Bayesian Analysis to candidates for the Master of Science in Business Analytics. He also serves on the Advisory Board for the Retail AI Lab of the Northwestern University Retail Analytics Council.

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

$$\overbrace{P(\underbrace{H}_{\textit{Hypothesis}}|\underbrace{E}_{\textit{Evidence}})}^{\text{Posterior}} = \frac{\overbrace{P(H)}^{\text{Prior}}\overbrace{P(E|H)}^{\text{Likelihood}}}{\underbrace{P(E)}_{\text{Marginal Likelihood}}}$$

Hidden

Bayes Theorem

(It has marvelous implications….)

Bayes Theorem

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

Given your **Universe of Discourse**,
$U \rightarrow \{C_k\}$, context-specific traits & assumptions

For example: $\mathrm{P}(H)=\mathrm{P}(H|U)=\sum_k \mathrm{P}(H|C_k,U)\mathrm{P}(C_k,U)$.

Though rarely done, we could show this explicitly:

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

Bayesian:
Core Concepts

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

Hidden

Bayes Theorem

(It has marvelous implications….)

Bayes Theorem

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

All alternative hypotheses are captured:

$$P(E) = \Sigma_i P(E|H{=}h_i)P(H{=}h_i)$$

$$= \Sigma_i P(E,H{=}h_i) = P(E,H{=}h^*) + P(E,H{\neq}h^*).$$

So, for your specific hypothesis of interest, h^*,
upon observing specific evidence e^*:

$$P(H{=}h^*|E{=}e^*) = \frac{P(E{=}e^*,H{=}h^*)}{P(E{=}e^*,H{=}h^*) + P(E{=}e^*,H \neq h^*)}$$

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 evidences & hypotheses
- **Evidence** updates beliefs
  - **Prior belief** is essential, not merely strength of evidence

Hidden

Bayes Theorem

(It has marvelous implications….)

Bayes Theorem

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

Prior beliefs are updated by strength of evidence to yield posterior beliefs:

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

Displayed as odds, this is even more obvious:

$$\frac{P(H{=}h^*|E)}{P(H \neq h^*|E)} = \left[\frac{P(H{=}h^*)}{P(H \neq h^*)}\right]\left[\frac{P(E|H{=}h^*)}{P(E)} \frac{P(H \neq h^*)}{P(H \neq h^*|E)}\right]$$

$$O(H{=}h^*|E) = O(H{=}h^*) \times BF(H{:}E)$$

Bayes Factor: $BF(H{:}E) \equiv \dfrac{P(E|H{=}h^*)}{P(E|H \neq h^*)}$

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

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

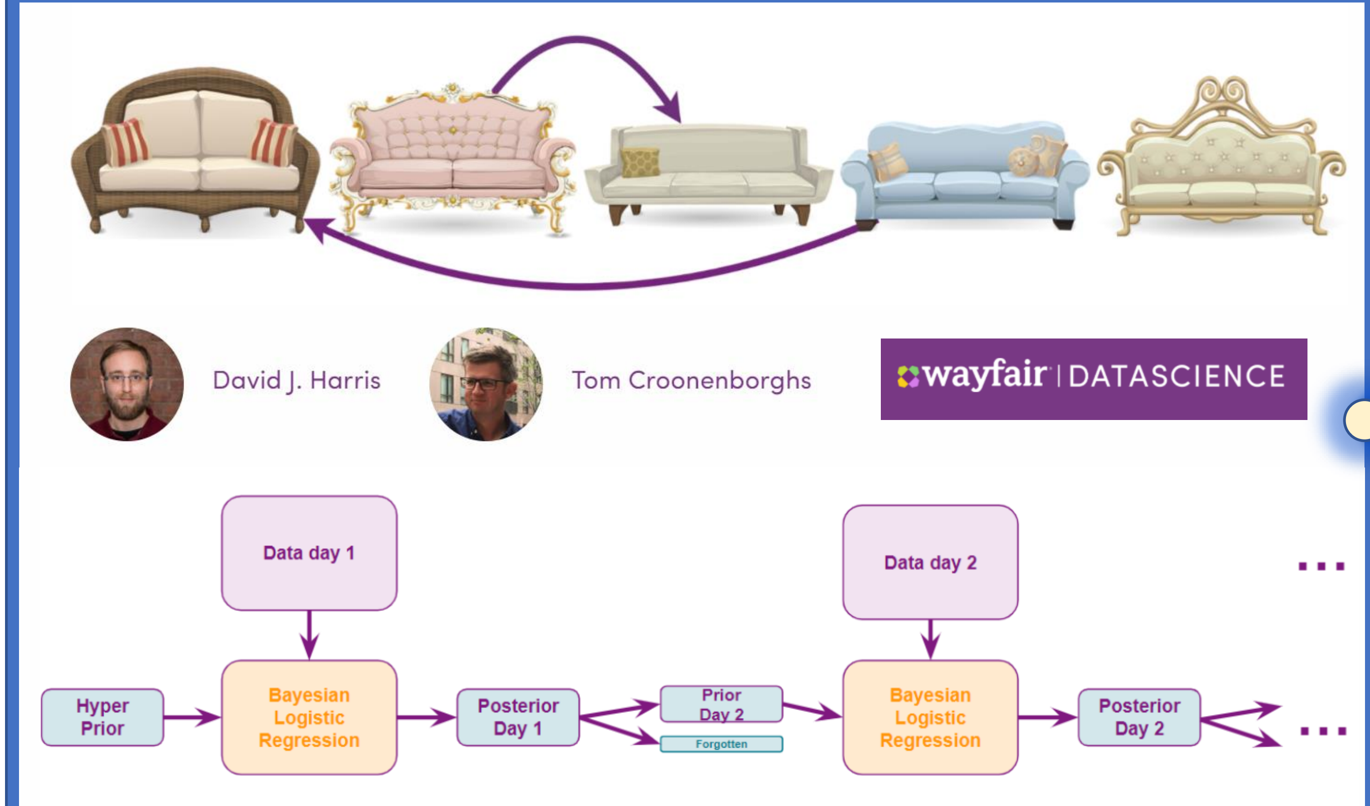


Fig. 4: A sketch of our daily update loop. Each day, our logistic regression model combines our observations of customer behavior with our prior knowledge to produce a posterior distribution. This posterior distribution then informs the next day's prior, but we intentionally "forget" a small portion 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

Prevalence of Bayesian Applications

Maximizing Return on Ad Spend at Smartly.io

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

Hidden

Prevalence of Bayesian Applications

Probabilistic Match-Making at Betterhalf.ai

Pawan Gupta
CEO

Rahul Namdev
CTO

"'We use recursive probabilistic Bayes update algorithm to estimate and learn about the correct personality of users over time,'
says **Betterhalf**'s co-founder and CTO Rahul Namdev."

"Marriages made in heaven with a little help from Bayes", Sriram Sharma, FactorDaily.com, May 14, 2018

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

Source: About Us, Betterhalf.ai

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.

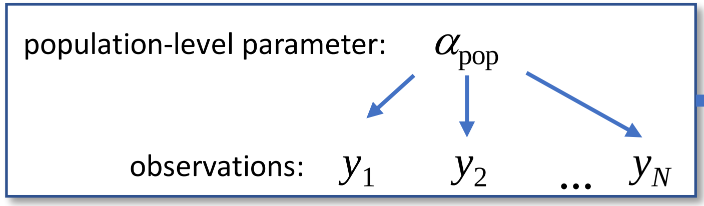
- Joint probability density function of **fully pooled** model:

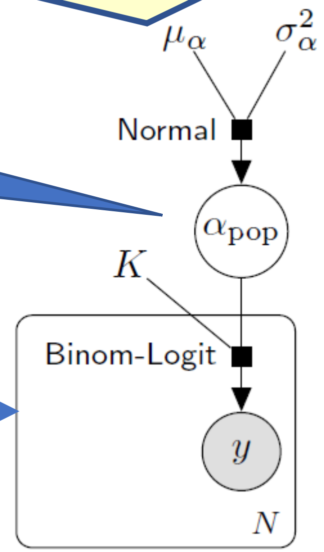
$$\underbrace{\mathrm{p}(y_1,\ldots,y_N,\alpha_{\mathrm{pop}}|K,\mu_\alpha,\sigma_\alpha^2)}_{\text{P(H,E)}} = \underbrace{\mathrm{p}(\alpha_{\mathrm{pop}}|\mu_\alpha,\sigma_\alpha^2)}_{\text{P(H)}} \underbrace{\prod_{n=1}^{N}\mathrm{p}(y_n|\alpha_{\mathrm{pop}},K)}_{\text{P(E|H)}}$$

Distribution declaration:

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

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





The *factor graph* of our Bayesian network.

Hidden

Bayesian Analysis

Logistic Regression: Frequentist vs. Bayesian

- **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

Frequentist Modeling

- Likelihood function of **fully pooled** model:

$$\mathrm{f}(\alpha_{\text{pop}}; y_1, \ldots, y_N) = \prod_{n=1}^{N} \mathrm{p}(y_n|\alpha_{\text{pop}})$$

Distribution declaration:

$$\forall n \in \{1, \ldots, N\}:$$

$$y_n \sim \mathsf{Binomial}(p(\alpha_{\text{pop}}), K)$$

$$\text{where } p(\alpha_{\text{pop}}) = \mathsf{Inv\text{-}Logit}(\alpha_{\text{pop}}) = \frac{1}{1 + \exp(-\alpha_{\text{pop}})}$$

So, $y_n \sim \mathsf{Binomial\text{-}Logit}(\alpha_{\text{pop}}, K)$

Bayesian Modeling

- 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:

P(H): $\alpha_{\text{pop}} \sim \mathsf{Normal}(\mu_\alpha, \sigma_\alpha^2)$

$\forall n \in \{1, \ldots, N\}:$

P(E|H): $y_n \sim \mathsf{Binomial\text{-}Logit}(\alpha_{\text{pop}}, K)$

Hidden

Bayesian Analysis

Logistic Regression: Frequentist vs. Bayesian

- **Example:** Product ranking, model the probability that a given customer will order product n:
  - $p = 1/(1 + \exp(-\alpha_{pop}))$

- Likelihood is treated as a deterministic function of the parameters at fixed values of data – ***so not as a probability density function on the data.***
- Finds ***a single point estimate for parameters*** by applying an optimizer to maximize likelihood.
- Assesses result (e.g., confidence intervals) under ***asymptotic assumptions*** (i.e., in the limit of infinite data) about the sampling process. Be wary of sparse data.
- Result is a point estimate and assessment of **the sampling process**, not of the true parameters. So, we cannot make any direct probability statements **about the true parameter values.**

Frequentist Modeling

- Likelihood function of **fully pooled** model:

$$f(\alpha_{pop}; y_1, \ldots, y_N) = \prod_{n=1}^{N} p(y_n|\alpha_{pop})$$

Distribution declaration:

$$\forall n \in \{1, \ldots, N\}:$$

$$y_n \sim \text{Binomial}(p(\alpha_{pop}), K)$$

$$\text{where } p(\alpha_{pop}) = \text{Inv-Logit}(\alpha_{pop}) = \frac{1}{1 + \exp(-\alpha_{pop})}$$

$$\text{So, } y_n \sim \text{Binomial-Logit}(\alpha_{pop}, K)$$

- **Probabilistic:** The full joint probability distribution is specified.
- **Alternatives considered:** Our priors on hypotheses capture plausibility of alternatives, providing domain knowledge in spite of sparse data.
- **Update beliefs w/evidence:** Data updates the prior to the posterior distribution, formally accounting for strength of evidence, which concentrates our beliefs on **most** plausible values. But computation of posterior **typically requires advanced algorithms.**
- Result is a **full posterior distribution** on our parameters. So, we **quantify uncertainty** and can make direct probability statements **about the true parameter values,** and **all derived quantities**, like predicted responses.

Bayesian Modeling

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

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

Distribution declaration:

P(H): $\alpha_{pop} \sim \text{Normal}(\mu_\alpha, \sigma_\alpha^2)$

$$\forall n \in \{1, \ldots, N\}:$$

P(E|H): $y_n \sim \text{Binomial-Logit}(\alpha_{pop}, K)$

Hidden

Bayesian Analysis
Fully pooled and unpooled models

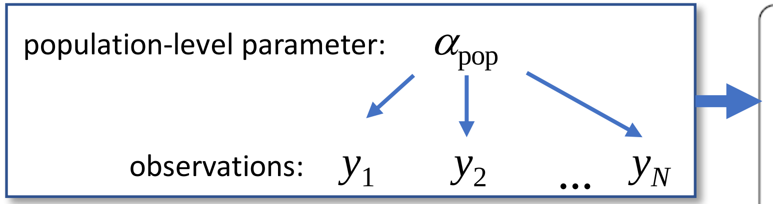
- Joint probability density function of **fully pooled** model:

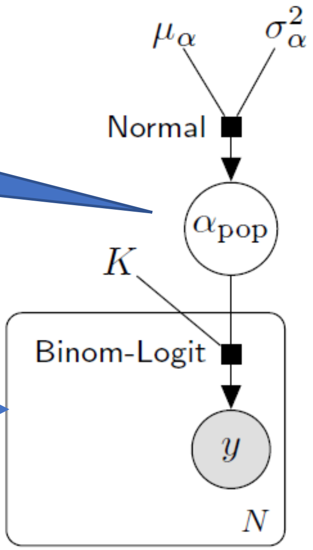
$$\overbrace{\mathrm{p}(y_1,\ldots,y_N,\alpha_{\text{pop}}|K,\mu_\alpha,\sigma_\alpha^2)}^{\text{P(H,E)}} = \overbrace{\mathrm{p}(\alpha_{\text{pop}}|\mu_\alpha,\sigma_\alpha^2)}^{\text{P(H)}} \prod_{n=1}^{N} \overbrace{\mathrm{p}(y_n|\alpha_{\text{pop}},K)}^{\text{P(E|H)}}$$

Distribution declaration:

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

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





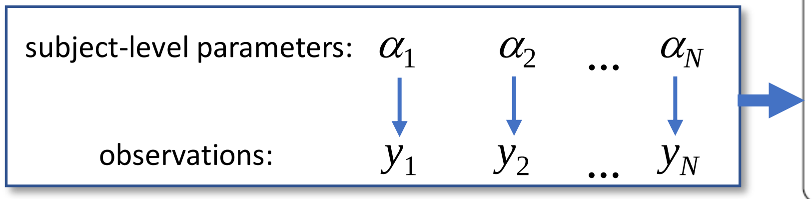
- Joint probability density function of **unpooled** model:

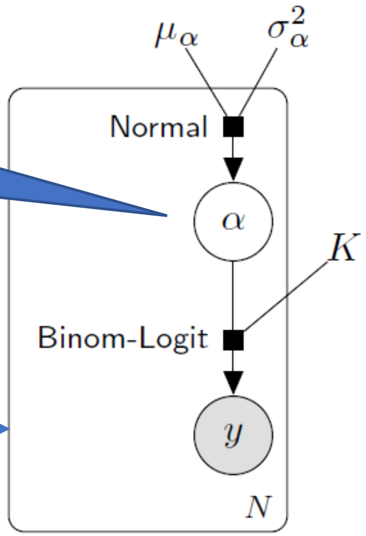
$$\overbrace{\mathrm{p}(y_1,\ldots,y_N,\alpha_1,\ldots,\alpha_N,|K,\mu_\alpha,\sigma_\alpha^2)}^{\text{P(H,E)}} = \prod_{n=1}^{N} \overbrace{\mathrm{p}(y_n|\alpha_n,K)}^{\text{P(E|H)}}\, \overbrace{\mathrm{p}(\alpha_n|\mu_\alpha,\sigma_\alpha^2)}^{\text{P(H)}}$$

Distribution declaration:

$$\begin{aligned} \forall n \in \{1,\ldots,N\}: & \\ \alpha_n &\sim \mathsf{Normal}(\mu_\alpha,\sigma_\alpha^2) \\ y_n &\sim \mathsf{Binomial\text{-}Logit}(\alpha_n,K) \end{aligned}$$

Each subject has its own subject-level parameter, independent of all other subjects.





Bayesian **Hierarchical/Multilevel** Models
Partially pooled model

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

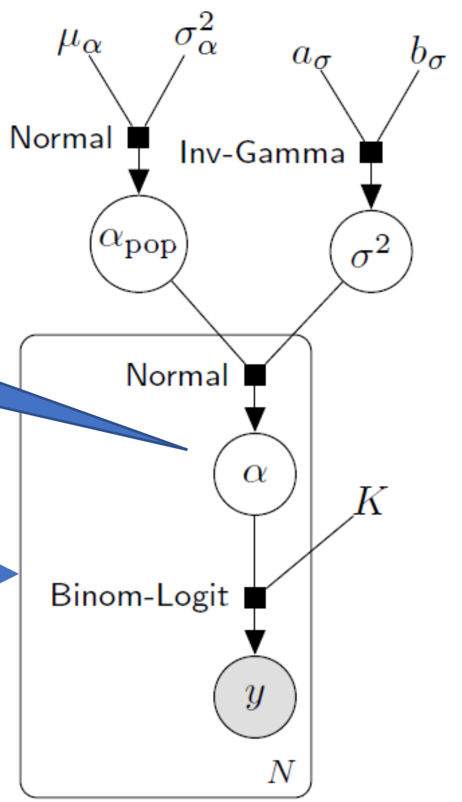
$$p(y_1,\ldots,y_N,\alpha_1,\ldots,\alpha_N,\sigma^2,\alpha_{\text{pop}}|K,\mu_\alpha,\sigma_\alpha^2,a_\sigma,b_\sigma) = p(\alpha_{\text{pop}}|\mu_\alpha,\sigma_\alpha^2)\,p(\sigma^2|a_\sigma,b_\sigma)\prod_{n=1}^{N} p(y_n|\alpha_n,K)\,p(\alpha_n|\alpha_{\text{pop}},\sigma^2)$$

Distribution declaration:

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

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:	α_{pop}, σ			
subject-level parameters:	α_1	α_2	…	α_N
observations:	y_1	y_2	…	y_N



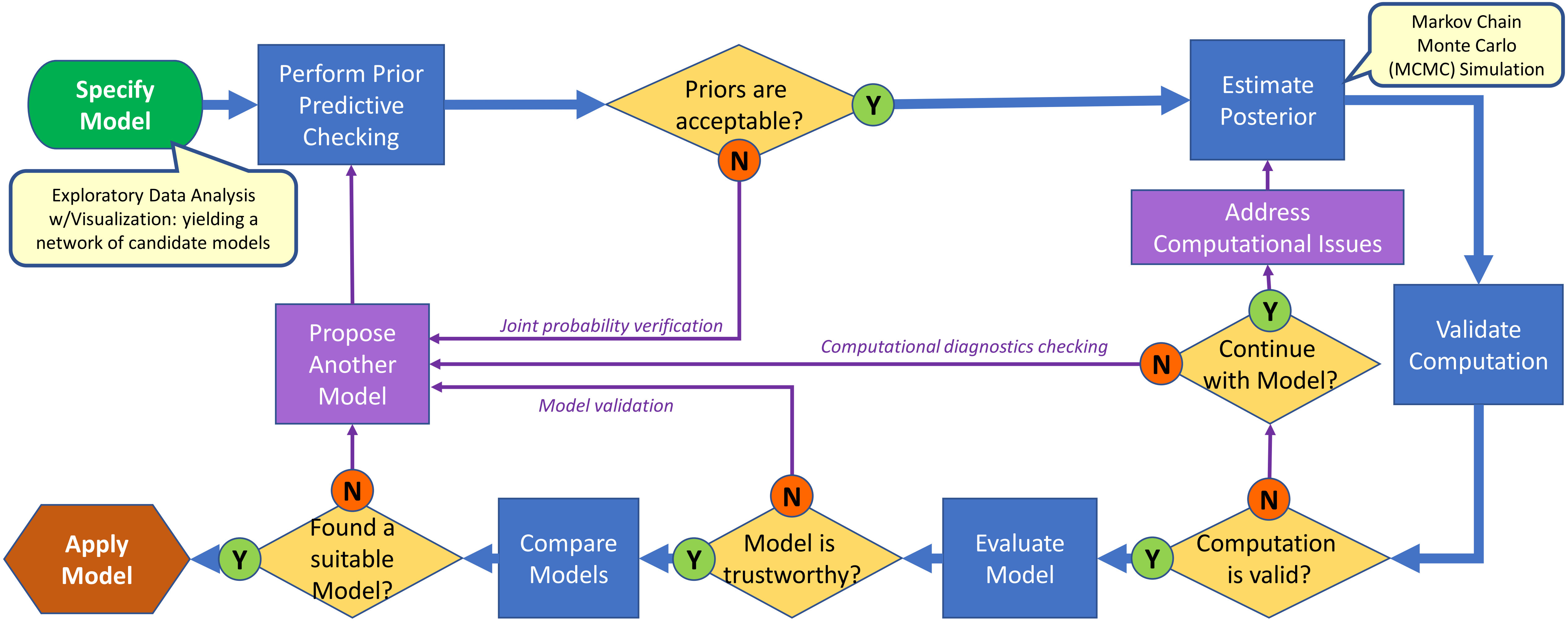
Tools for Bayesian Analysis

- **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.*)



Hidden

How to do Bayesian Analysis?

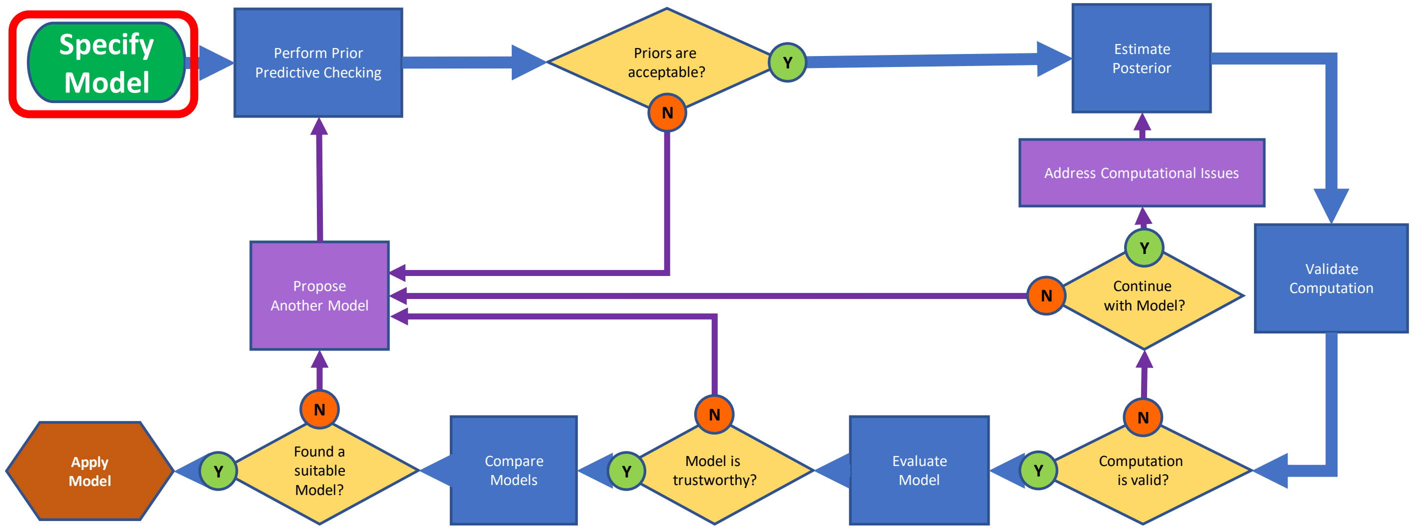
Example: Generalized linear multilevel model

- "Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists",
  T. Sorensen, S. Hohenstein & S. Vasishth, *arXiv preprint,* arXiv:1506.06201 (2015).
  - Excerpt:
    "This paper was written using a literate programming tool, `knitr` (Xie, 2015); this integrates documentation for the accompanying code with the paper. The `knitr` file that generated this paper, as well as all the code and data used in this tutorial, can be downloaded from our website:
    https://www.ling.uni-potsdam.de/~vasishth/statistics/BayesLMMs.html
    In addition, the source code for the paper, all R code, and data are available on github at: https://github.com/vasishth/BayesLMMTutorial".

How to do Bayesian Analysis?

Example: Generalized linear multilevel model

- “Finally! Bayesian Hierarchical Modelling at Scale"
  Florian Wilhelm, *Inovex* blog post, Mar. 3, 2021
    - Excerpt:
      “In this blog post, I want to draw your attention to the somewhat dusty *Bayesian Hierarchical Modelling*. Modern techniques and frameworks allow you to finally apply this cool method on datasets with sizes much bigger than what was possible before and thus letting it really shine.”



Hidden

Bayesian Workflow

1. Specify Model

- Problem Statement*
  - We wish to estimate the success rate π of our promotional ad campaign amongst our target population. We have data from a small test amongst N=16 customers, in which we succeeded with n=6.
- Generative Model
  - Base Model: $n \sim \text{Binomial}(N,\pi)$; prior $\text{p}(\pi)$
- Weakly Informative Priors
  - Previous experience tells us that typically π~0.30 and almost always our campaigns have π<0.80; we'll use $\pi \sim \text{Beta}(\alpha,\beta)$ and set the hyperparams assuming μ=0.30 and let 97.5%-ile = 0.80; so use α=0.96, β=2.24

Just as with the simple Mosaic plot examples, this gives us the joint distribution, but as a simulator rather than a 2D plot. And applying Bayes Rule is, again, as simple as selecting the generated hypotheses (parameter) values that coincide with the same generated success rates as the evidence, in this case n.

*Adapted from Rasmus Baath: "Bayesian Analysis Tutorial, Part 1: What"; YouTube (2017).

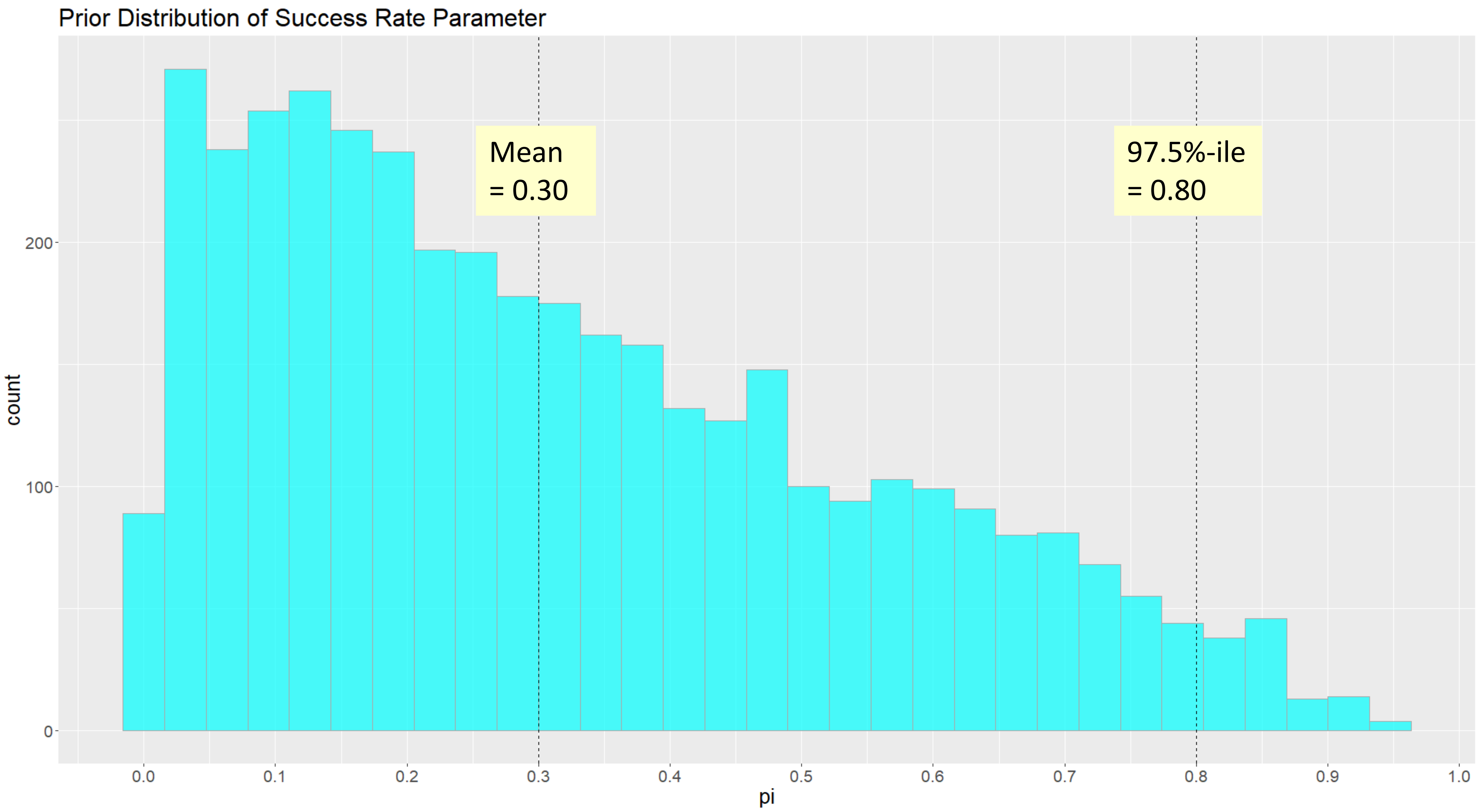
Hidden

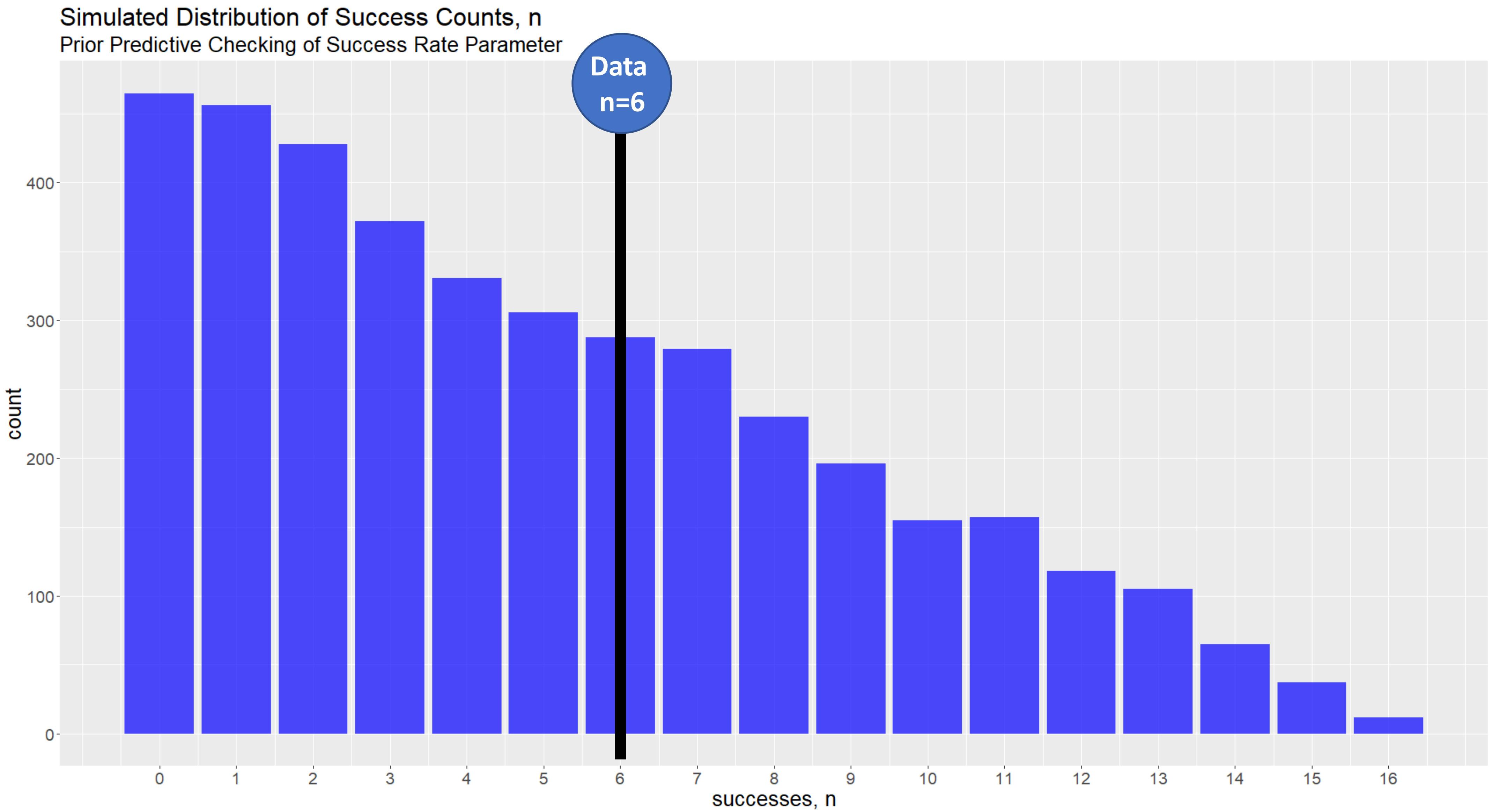
Prior Predictive Checking

1. Simulation of hypotheses given evidence using joint probability distribution

```
> library(magrittr)
> library(tidyverse)
> library(brms)
> df <- tibble(n=6,N=16)
> sim_beta <- brm(
          formula = n | trials(N) ~ 1,
          family  = binomial(link="identity"),
          data    = df,
          prior   = set_prior(prior="beta(0.96,2.24)",class="Intercept"),
          sample_prior = "only"
)
> sim_beta
 Family: binomial
  Links: mu = identity
Formula: n | trials(N) ~ 1
   Data: df (Number of observations: 1)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     0.32      0.23     0.02     0.82 1.01      454      871

Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

Tool:
R package
`brms`


Prior Distribution of Success Rate Parameter
Mean
= 0.30
97.5%-ile
= 0.80
count
0
100
200
pi
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0


Simulated Distribution of Success Counts, n
Prior Predictive Checking of Success Rate Parameter
Data
n=6
count
400
300
200
100
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
successes, n

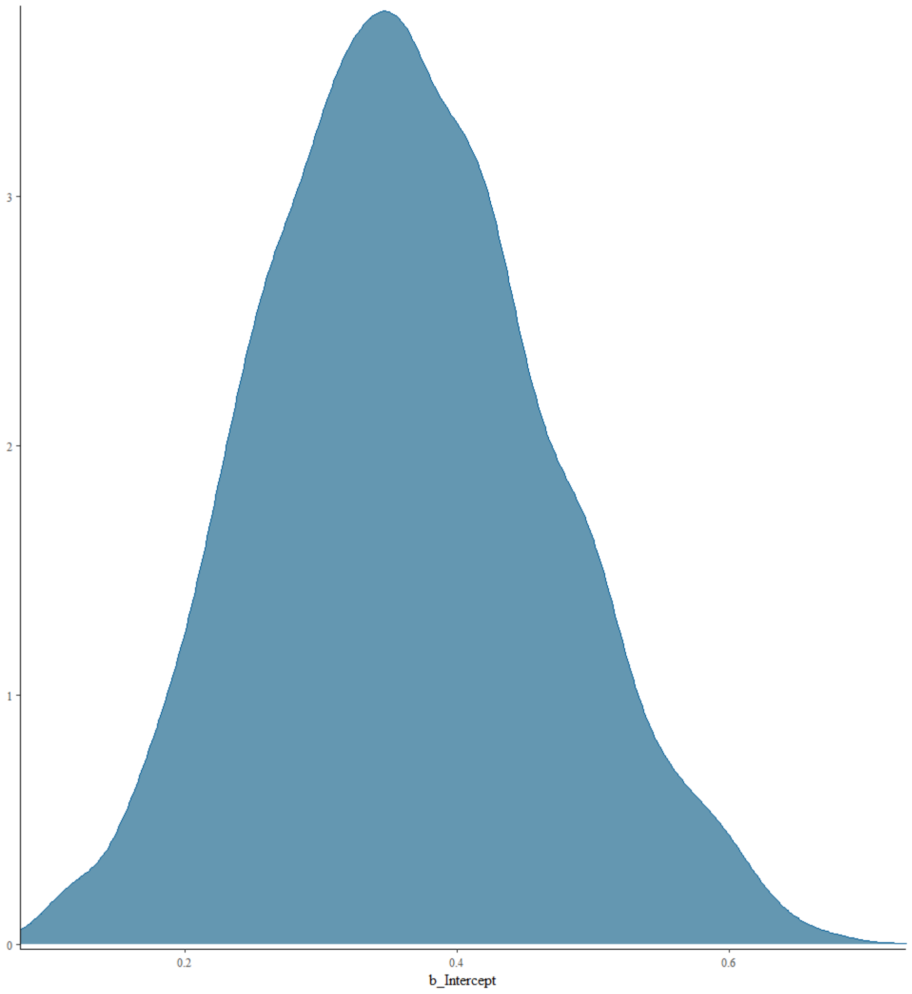
Hidden

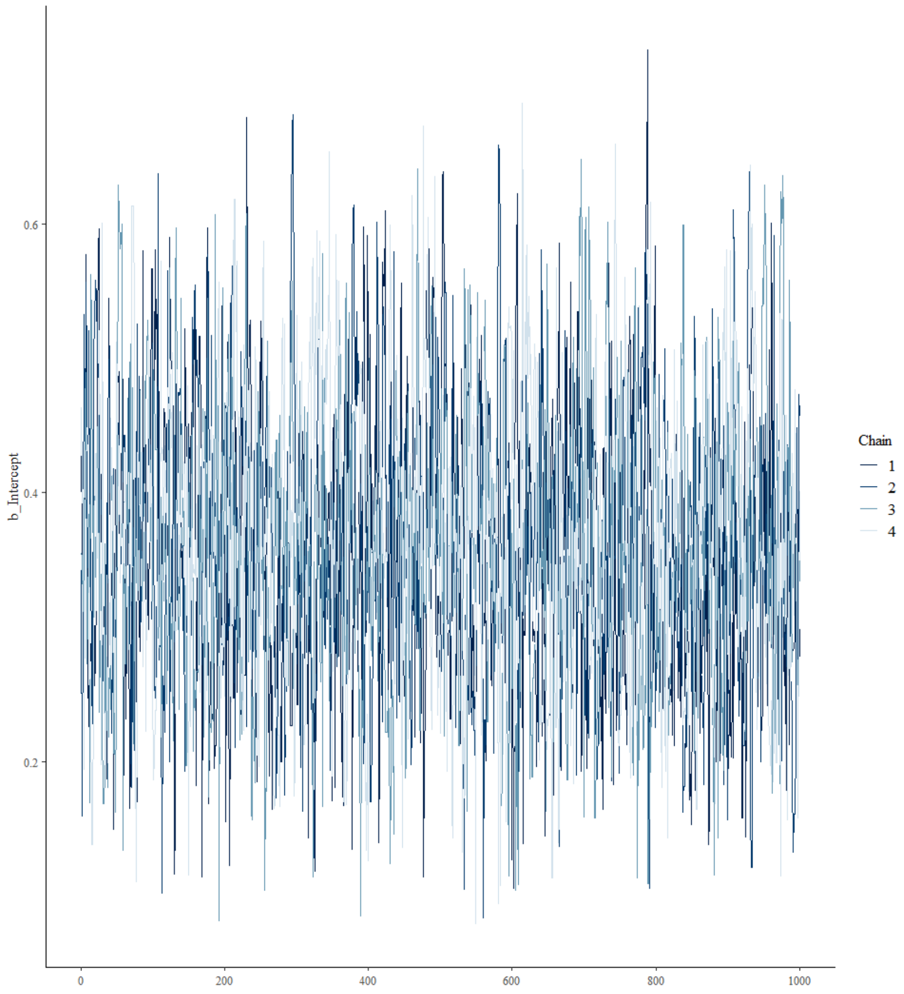
Hidden

Estimating Posterior

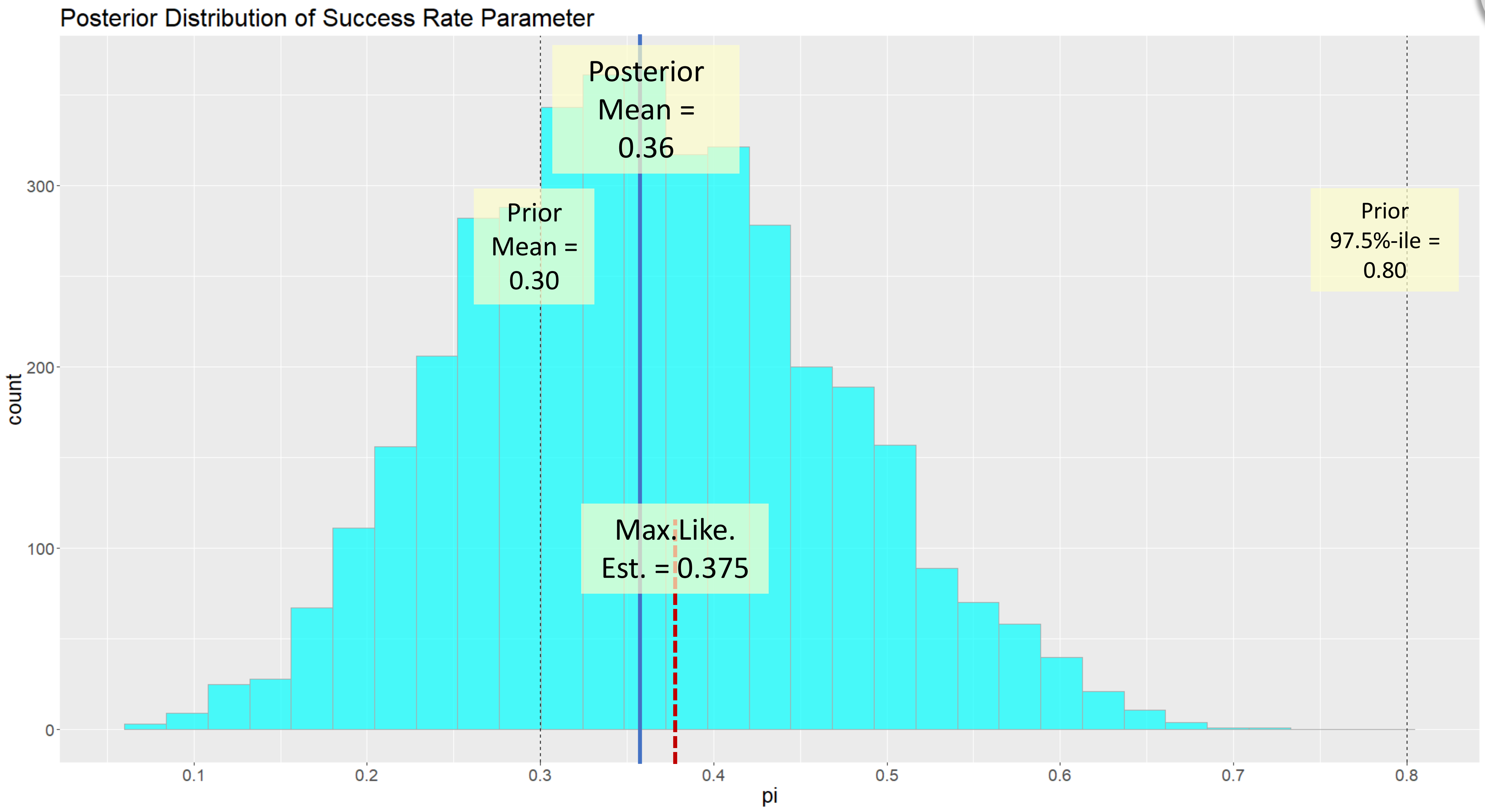
```
> mod_beta <- update(sim_beta,sample_prior="no")
> mod_beta
 Family: binomial
  Links: mu = identity
Formula: n | trials(N) ~ 1
   Data: df (Number of observations: 1)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     0.36      0.11     0.17     0.58 1.00     1537     1770

Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
Warning message:
There were 6 divergent transitions after warmup. Increasing adapt_delta above 0.8 may help.
See http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

> plot(mod_beta)
```

Posterior Density & Monte Carlo Trace Plot for success rate parameter π

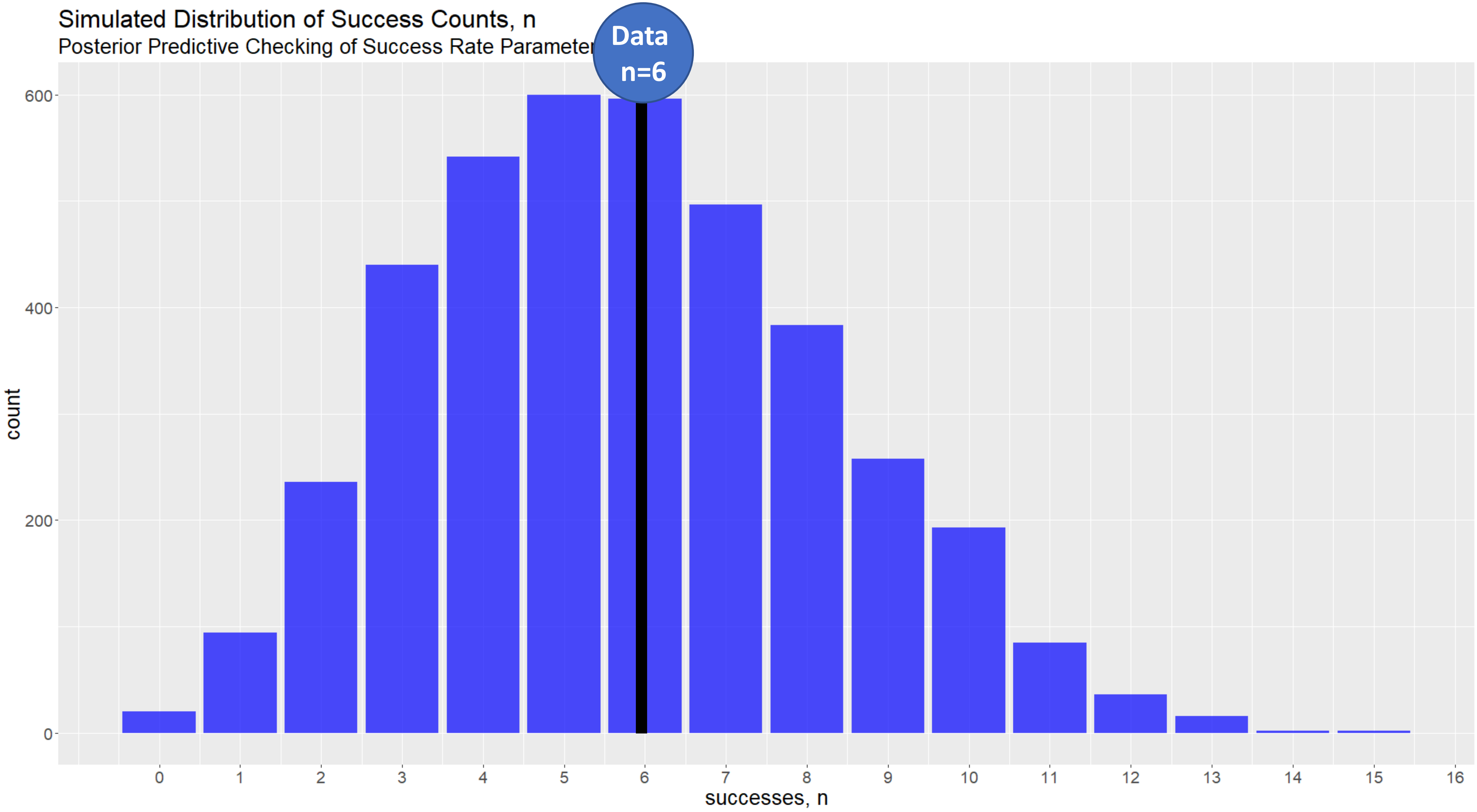




Hidden


Posterior Distribution of Success Rate Parameter
Posterior
Mean =
0.36
Prior
Mean =
0.30
Prior
97.5%-ile =
0.80
Max.Like.
Est. = 0.375
300
200
100
0
count
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
pi

Hidden


Simulated Distribution of Success Counts, n
Posterior Predictive Checking of Success Rate Parameter
Data
n=6
count
600
400
200
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
successes, n

Hidden

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"	"Complex"
Data	**Data**
• Single source	• Multiple sources
• Single variable types/distribution families	• Multiple variable types/distribution families
• Tabular & Ample – "Big Data"	• Ragged & Sparse
◦ Non-missing	◦ Missing
◦ Regular, exchangeable	◦ Multigranular aggregation
Homogeneous	Heterogeneous
Model	**Model**
• Observations linked to observations (Modeling the Data)	• Latent spaces (Modeling the Domain)
• Empirical structure	• Causal structure
◦ Single-level	◦ Multi-level
◦ Correlative (acausal)	◦ Mechanisms
Data-to-Data	True-to-True
• Single hypothesis	• Mixture of phenomena/Multi-Hypothesis
• Component-level estimation; Low-level integration	• System-level integration
Decision	**Decision**
Deterministic Predictions	Probabilistic Explanations
• Deterministic assumptions	• Reasoning under uncertainty (UQ)
• Modal/point estimate solutions	• Risk analysis
• Predictive inference (What will happen?)	• Explanatory inference (Why did it happen?)
• Single objective, Static	• Multi-Objective, Dynamic updating

Hidden

Use of Bayesian Analysis is Compelled by…

… These Issues

- **Uncertainty Quantification (UQ)**
  - Sparsity & Missing Data
  - Risk & Decision Analysis; Post-Inference App: Explanations; Confirmations
- **Domain Knowledge (DK) Inclusion**
  - Expertise; Causality/Structural/State Space/Mechanism Representation
- **Heterogeneity**
  - Multi-level Behavior (context-specific; individual-specific)
- **Latent Structure**
  - Factor Analysis; Conceptual Embeddings; Underlying Cognitive Constructs
- **Multiple Responses**
  - Information Fusion; Multimodal Data Streams
- **Adaptation (Bayesian Updating)**
  - Autonomous & Active Learning Systems; Adaptive Questionnaires; Adaptive Experiments

Example: Predicting Success Rates

Base Model: $n \sim \text{Binomial}(N,\pi)$; prior $p(\pi)$

- **UQ:** posterior captures uncertainty $p(\pi \mid \{N,n\}_{\text{Data}})$; prior $p(x)$, e.g. "error-in-variables"
- **DK Inclusion**: leverage expertise; interventions on x; $u(x) = f(x;\theta)$; captures mechanism (possibly nonlinear) $\pi = g^{-1}(u(x))$ (logit link function, $g(.)$)
- **Heterogeneity:** hyper-personalization context- and individual-specific parameters $u(x_{ij}) = f(x_{ij};\theta_{ij})$; $\theta_{ij}\sim\text{Normal}(\theta_j)$; $\theta_j\sim\text{Normal}(\bar{\theta})$
- **Latent Structure:** cognitive constructs manifest as observed behavior; e.g. customer preferences $u(x,z) = f(x,z;\theta)$; $z = h(x)$ captures theory of behavior
- **Multiple Responses:** universal latent constructs z driving behaviors measured by different modalities – e.g. text reviews, videos, photos, survey responses.Augment trial-success data $y=\{N,n\}_{\text{Data}}$ with other data & add measurement models (likelihood modules)
- **Adaptation:** leverage UQ to compute learning objectives to generate new x; e.g. optimal product design; adaptive recommenders

Hidden

Bayesian Analysis Beyond Predictive Models

Decision Support & Automated Reasoning **Systems**

Confirmatory, Explanatory & Causal Inference

Generalized Bayes Factors & Weight of Evidence

Most Compelling Confirmation (MCC)

Most Relevant Explanation (MRE)

Causal Inference

Risk Assessment & Decision Analysis

Uncertainty Quantification

Utility Functions

Costs/Rewards

Learning, Adaptation & Content Generation

Bayesian Machine Learning (Model-Based ML)

Bayesian Optimization

Reinforcement/Active Learning

Adaptive Experiments

Computational Creativity

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

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

- 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

Stan

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

Hidden

Readings: Articles & Book Excerpts

- Statistics
  - Kruschke, J. K. "Chapter 2, Introduction: Credibility, Models, and Parameters", Doing Bayesian Data Analysis, 2nd Edition: A Tutorial with R, JAGS, and Stan. Waltham, MA: Academic Press (2015).
  - Richard McElreath "Statistical Rethinking"
  - Gabry et al. "Visualization in Bayesian Workflow"
  - Gelman et al. "Bayesian Workflow"
  - Gelman et al. "The Prior Can Often Only Be Understood in the Context of the Likelihood"
  - Michael Betancourt, "A Conceptual Introduction to Hamiltonian Monte Carlo Simulation"

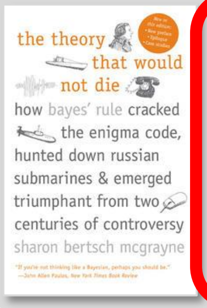
I draw heavily upon these in the graduate course I teach on Bayesian Analysis at the Univ. of Cincinnati.

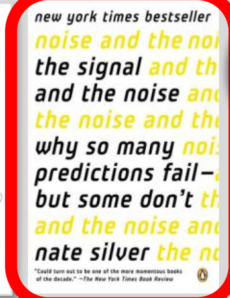
- AI/ML
  - Chris Bishop (Microsoft Research) "Model-Based Machine Learning"
  - Tenenbaum, *et al*., How to Grow a Mind: Statistics, Structure, and Abstraction (PDF), *Science*, 2011; lecture video
  - Daphne Koller & Nir Friedman "Probabilistic Graphical Models" (excerpt)
  - Norman Fenton & Martin Neil "Risk Assessment & Decision Analysis with Bayesian Networks" (sample chapters)

Bayesian Theory & Practice

(Each image is hyperlinked to the book's website, most of which include excerpts & example code)

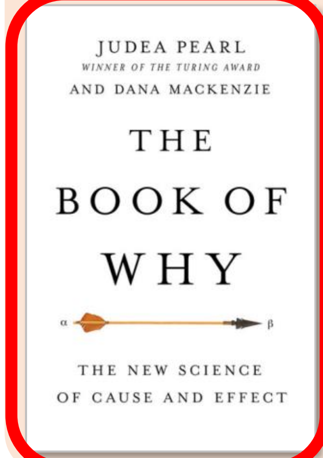
Hidden

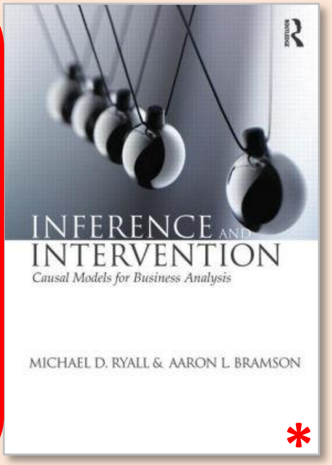




*

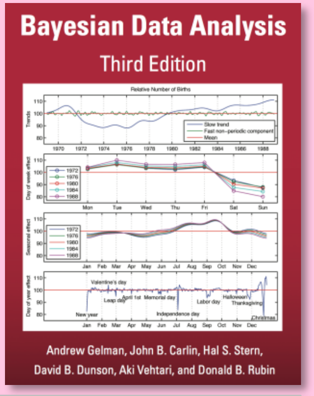
Causal Inference

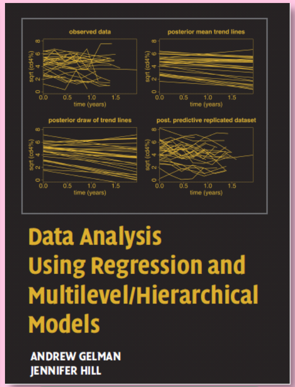


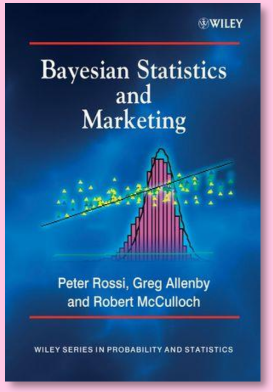


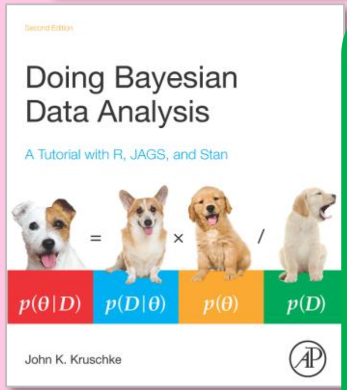
*

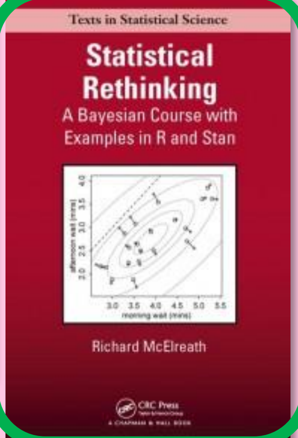
Bayesian Statistics

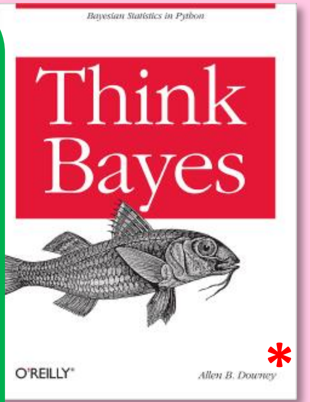








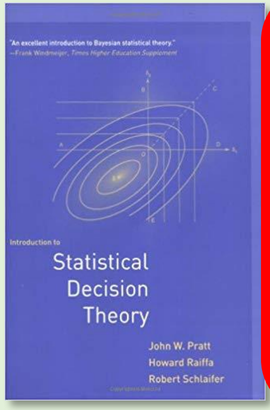


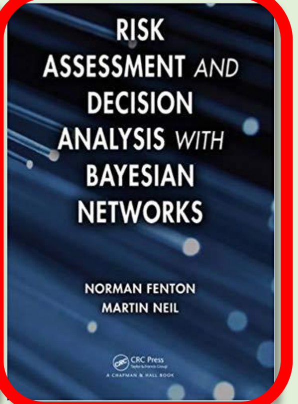


*

Bayesian Statistics

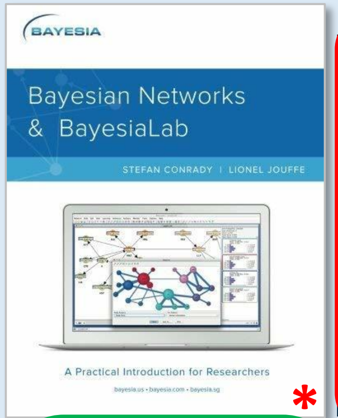
Bayesian Decision Analysis



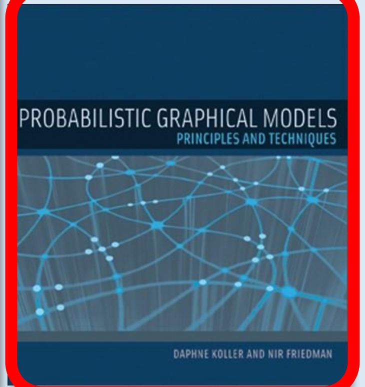


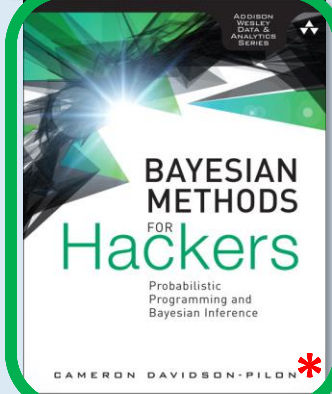
Bayesian Decision Analysis

Bayesian Machine Learning

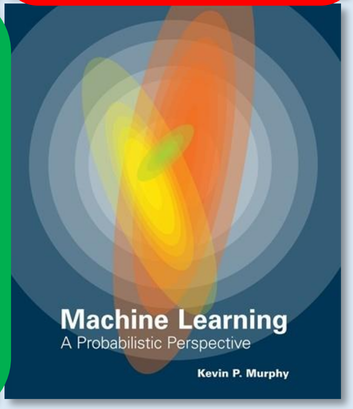


*





*



Bayesian Machine Learning

Especially good for hands-on learning.

* *Little if any theory & math; many applied examples*

Thank you!