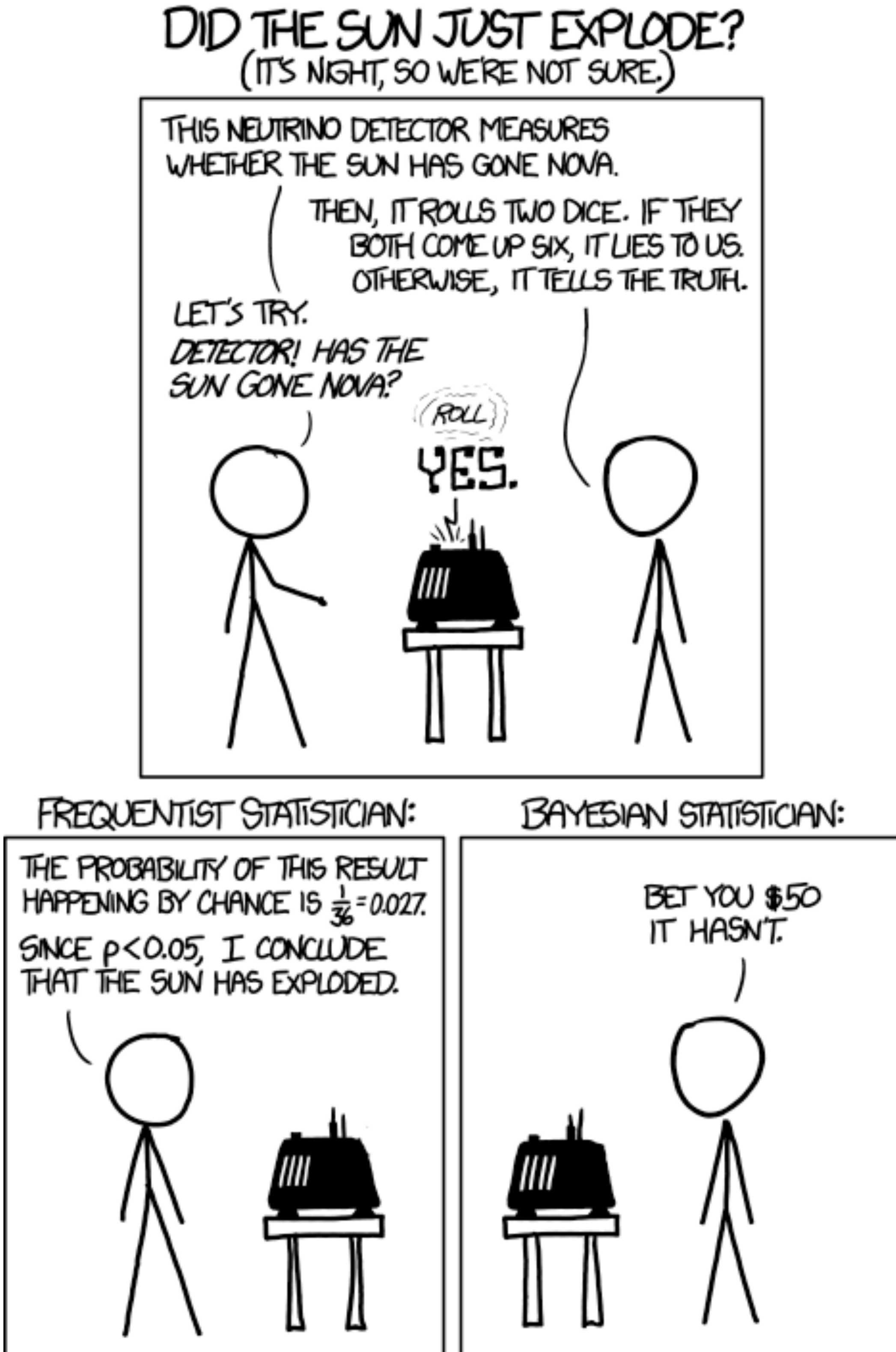


A Brief Introduction to Bayes

Lesson 3

The unifying principal is statistical estimation based on **<u>probability</u>**

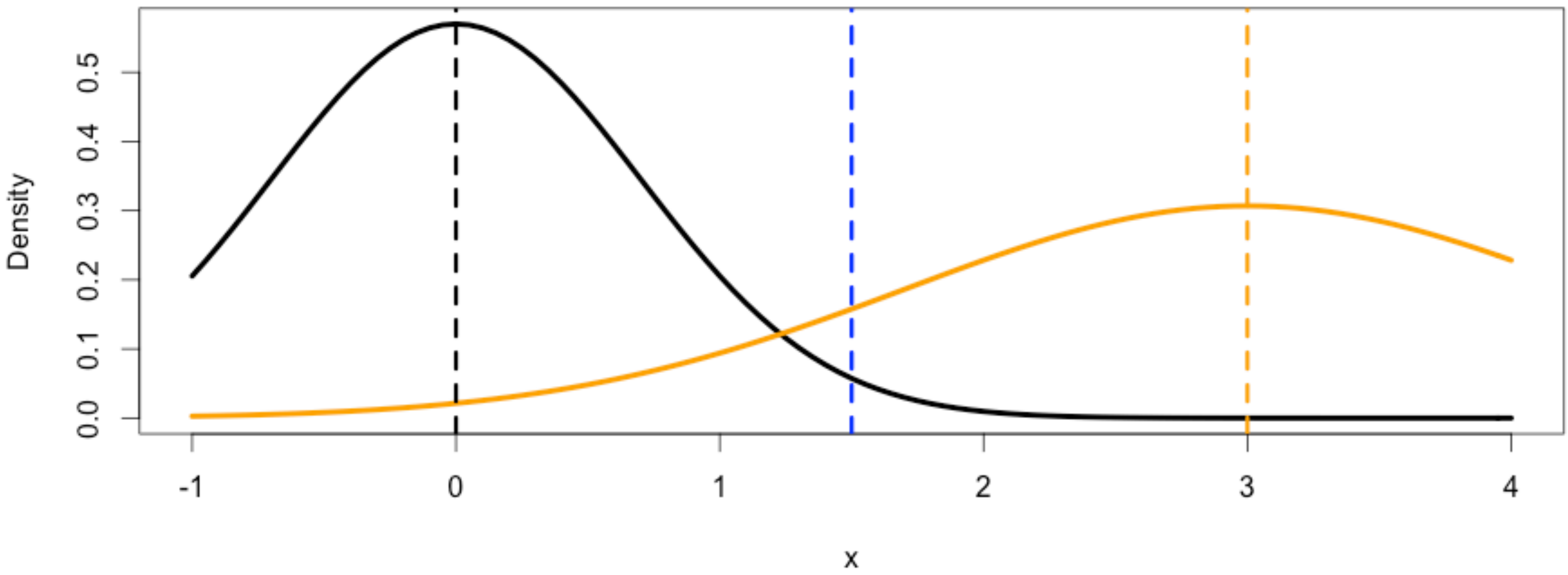
A bit on motivation....

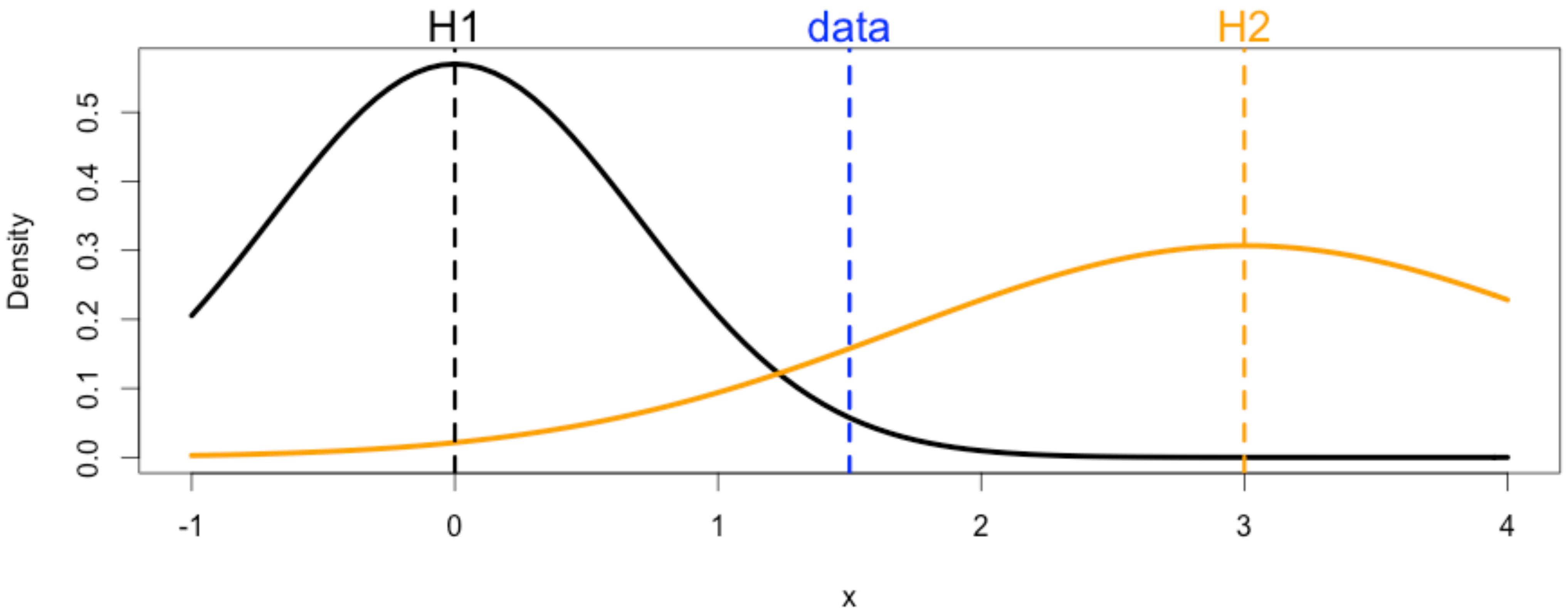
Data are usually complex &
violate the assumptions of classical tests

Forecasts need to be updated (iteratively)

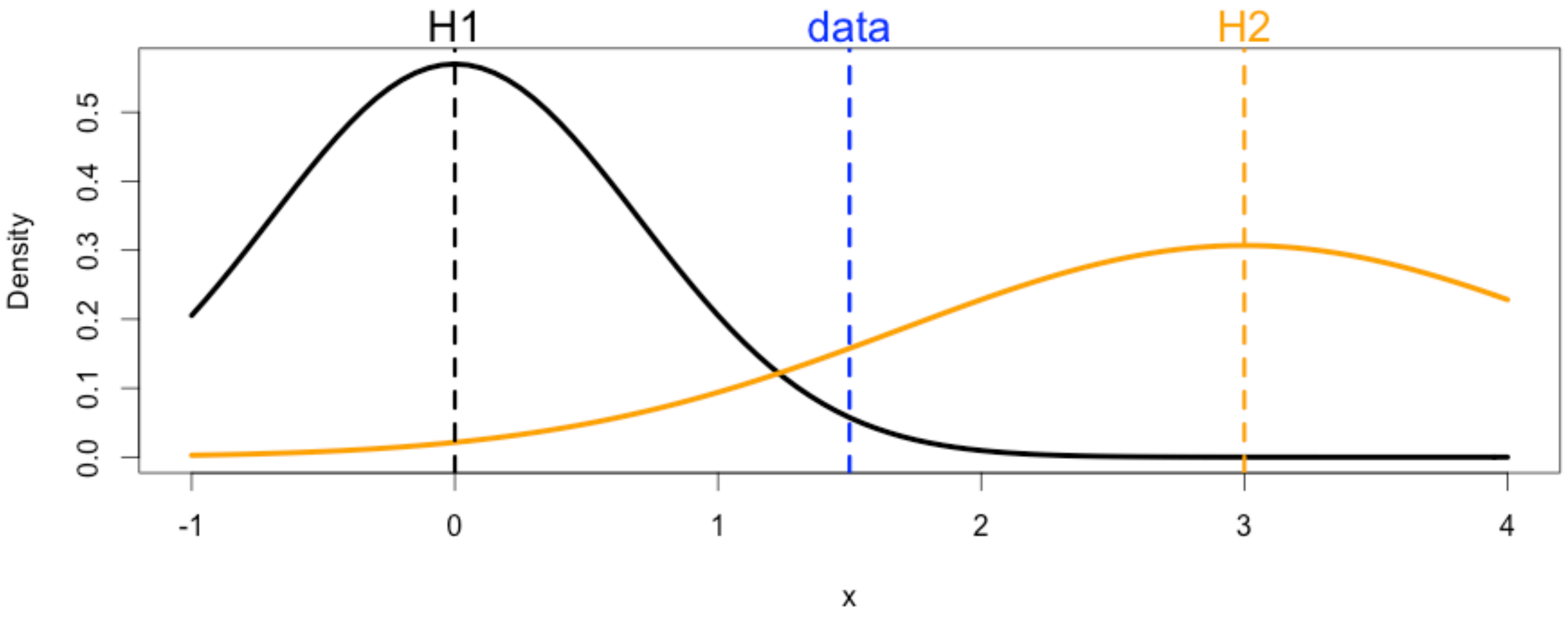
May have multiple sources of data, variability

This complexity can be addressed with modern techniques

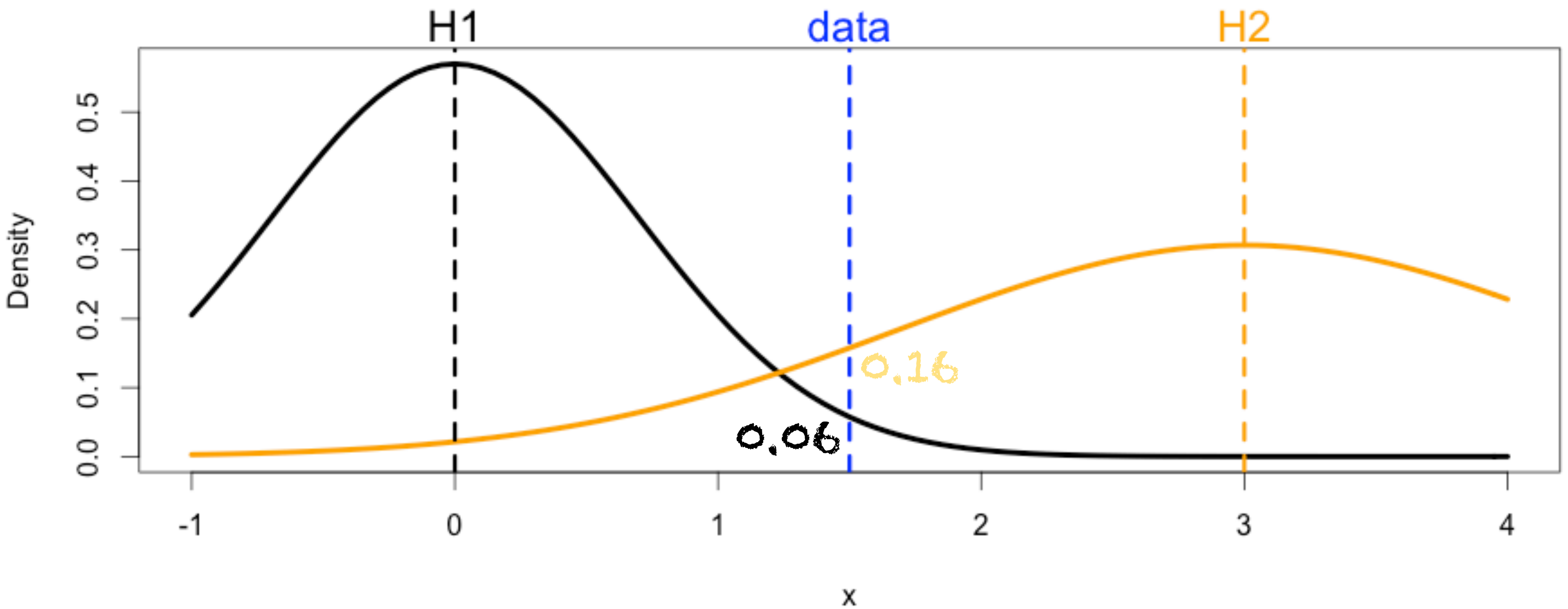

Density
0.0
0.1
0.2
0.3
0.4
0.5
-1
0
1
2
3
4
x


H1
data
H2
Density
0.0
0.1
0.2
0.3
0.4
0.5
-1
0
1
2
3
4
x

dnorm(1.5,0,0.7) dnorm(1.5,3,1.3)



dnorm(1.5,0,0.7) dnorm(1.5,3,1.3)



LIKELIHOOD

$$L = P(X = x | \theta) = P(data | model)$$

- Probability of observing a given data point x conditional on parameter value θ
- Likelihood principle: a parameter value is more likely than another if it is the one for which the data are more probable

Linear Regression

$$y_i = a_0 + a_1 x_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2)$$

Linear Regression

$$y_i = a_0 + a_1 x_i + \epsilon_i$$
$$\epsilon_i \sim N(0, \sigma^2)$$

Process Model

Linear Regression

$$y_i = a_0 + a_1 x_i + \epsilon_i$$

Process Model

$$\epsilon_i \sim N(0, \sigma^2)$$

Data Model

Linear Regression

$$y_i = a_0 + a_1 x_i + \epsilon_i$$ Process Model

$$\epsilon_i \sim N(0, \sigma^2)$$ Data Model

- Likelihood

$$L = \prod_{i=1}^{n} N(y_i | a_0 + a_1 x_i, \sigma^2)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n \exp\left[\frac{-1}{2\sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)^2 \right]$$

How is the data modeled?

How is the data modeled?

- What type of data is it?
  - Continuous
  - Integer / Count
  - Boolean (0/1)
  - Factor / categorical

How is the data modeled?

- What type of data is it?
  - Continuous
  - Integer / Count
  - Boolean (0/1)
  - Factor / categorical
- Are there range restrictions on the data?
  - Are negative values allowed?
  - Is there an upper bound?
  - Are the observed data near the bounds?

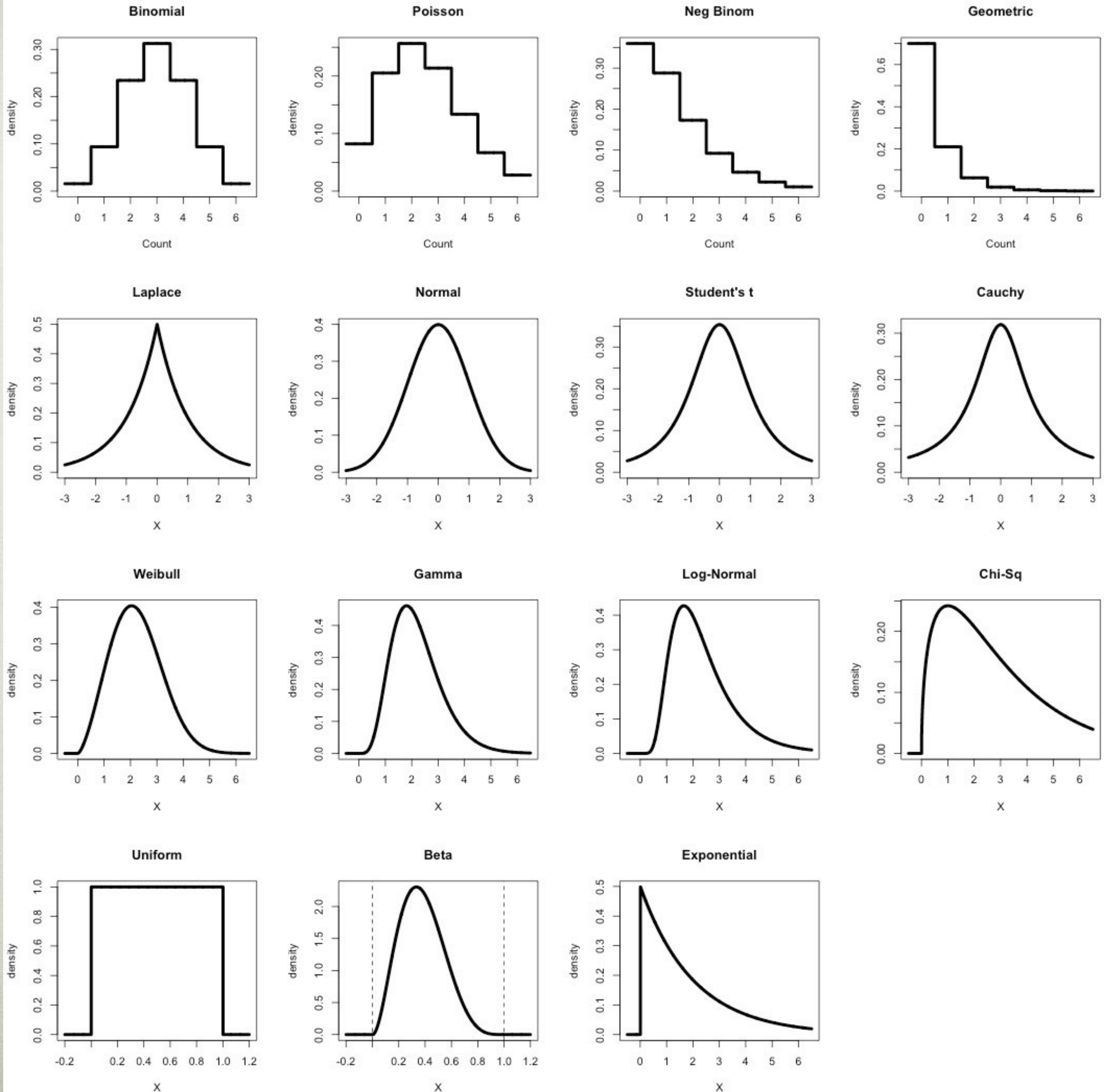
How is the data modeled?

- What type of data is it?
  - Continuous
  - Integer / Count
  - Boolean (0/1)
  - Factor / categorical
- Are there range restrictions on the data?
  - Are negative values allowed?
  - Is there an upper bound?
  - Are the observed data near the bounds?
- What process generated this data?

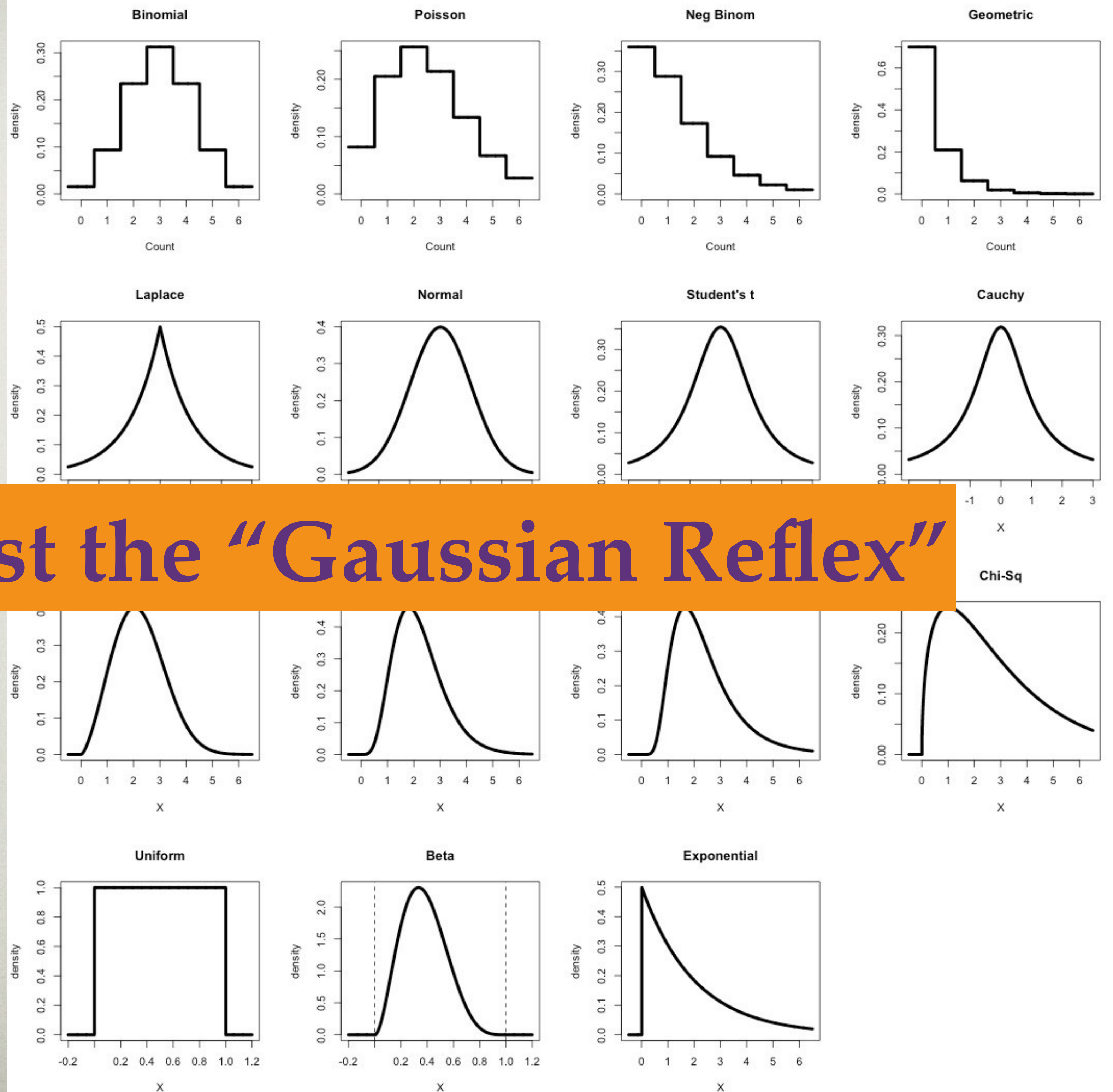
How is the data modeled?

- What type of data is it?
  - Continuous
  - Integer / Count
  - Boolean (0/1)
  - Factor / categorical
- Are there range restrictions on the data?
  - Are negative values allowed?
  - Is there an upper bound?
  - Are the observed data near the bounds?
- What process generated this data?
- What distributions are an appropriate description of the data?

Choosing a Distribution



Choosing a Distribution



Resist the "Gaussian Reflex"

How is the process modeled?

How is the process modeled?

- Constant mean

How is the process modeled?

- Constant mean
- Multiple means by factor (ANOVA)

How is the process modeled?

- Constant mean
- Multiple means by factor (ANOVA)
- As a function of covariates
  - Linear models
  - Generalized linear models
  - Generalize additive models
  - Nonlinear models

How is the process modeled?

- Constant mean
- Multiple means by factor (ANOVA)
- As a function of covariates
  - Linear models
  - Generalized linear models
  - Generalize additive models
  - Nonlinear models
- Hierarchical models (space/time variability)

How is the process modeled?

- Constant mean
- Multiple means by factor (ANOVA)
- As a function of covariates
  - Linear models
  - Generalized linear models
  - Generalize additive models
  - Nonlinear models
- Hierarchical models (space/time variability)
- Dynamic models $N_{t+1} = f(N_t)$

How is the process modeled?

- Constant mean
- Multiple means by factor (ANOVA)
- As a function of covariates
  - Linear models
  - Generalized linear models
  - Generalize additive models
  - Nonlinear models
- Hierarchical models (space/time variability)
- Dynamic models $N_{t+1} = f(N_t)$
- Mechanistic models

Likelihood limitations

Likelihood limitations

- P(data | model)? Why not P(model | data)?

Likelihood limitations

- P(data | model)? Why not P(model | data)?
- Uncertainties: Returns a point estimate (single optimum) not a distribution

Likelihood limitations

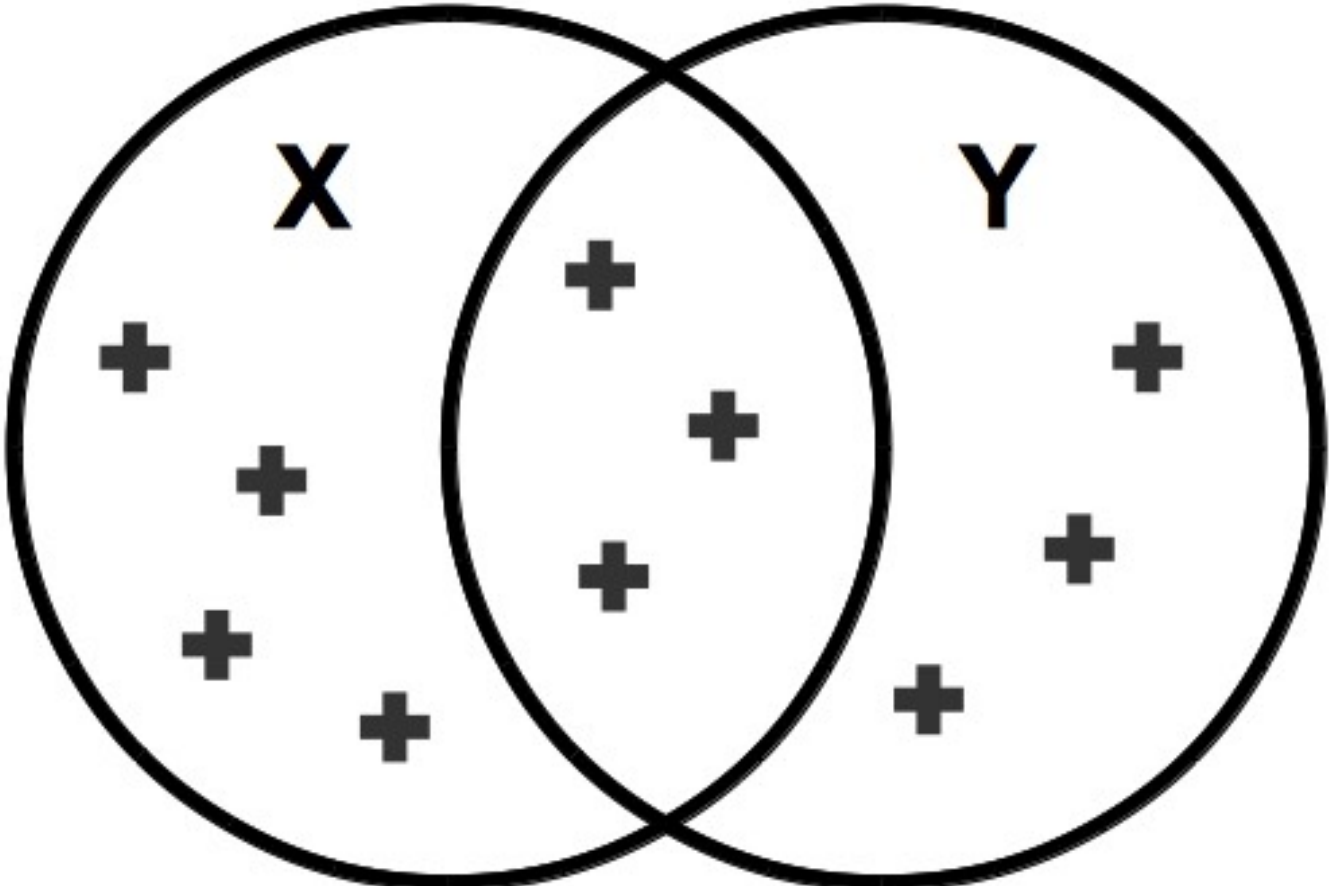
- P(data | model)? Why not P(model | data)?
- Uncertainties: Returns a point estimate (single optimum) not a distribution
  - requires additional (strained?) assumptions to calculate uncertainties

Likelihood limitations

- P(data | model)? Why not P(model | data)?
- Uncertainties: Returns a point estimate (single optimum) not a distribution
  - requires additional (strained?) assumptions to calculate uncertainties
- Pragmatic: numeric optimization often unreliable for complex models, many parameters

Likelihood limitations

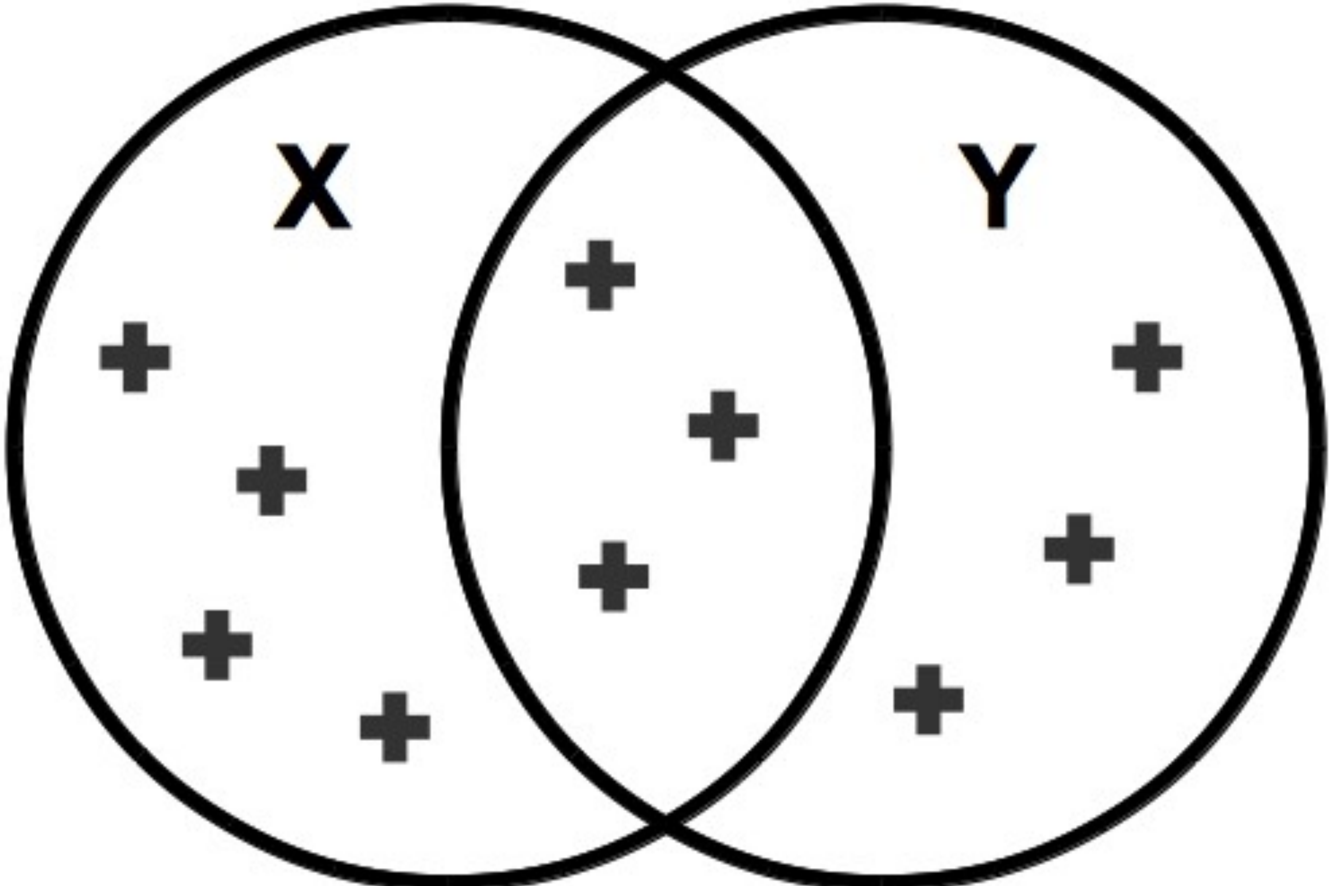
- P(data | model)? Why not P(model | data)?
- Uncertainties: Returns a point estimate (single optimum) not a distribution
  - requires additional (strained?) assumptions to calculate uncertainties
- Pragmatic: numeric optimization often unreliable for complex models, many parameters
- Inference in a vacuum: no prior knowledge, no updating, harder to combine sources of info.



JOINT = P(X,Y)

CONDITIONAL = P(X|Y)

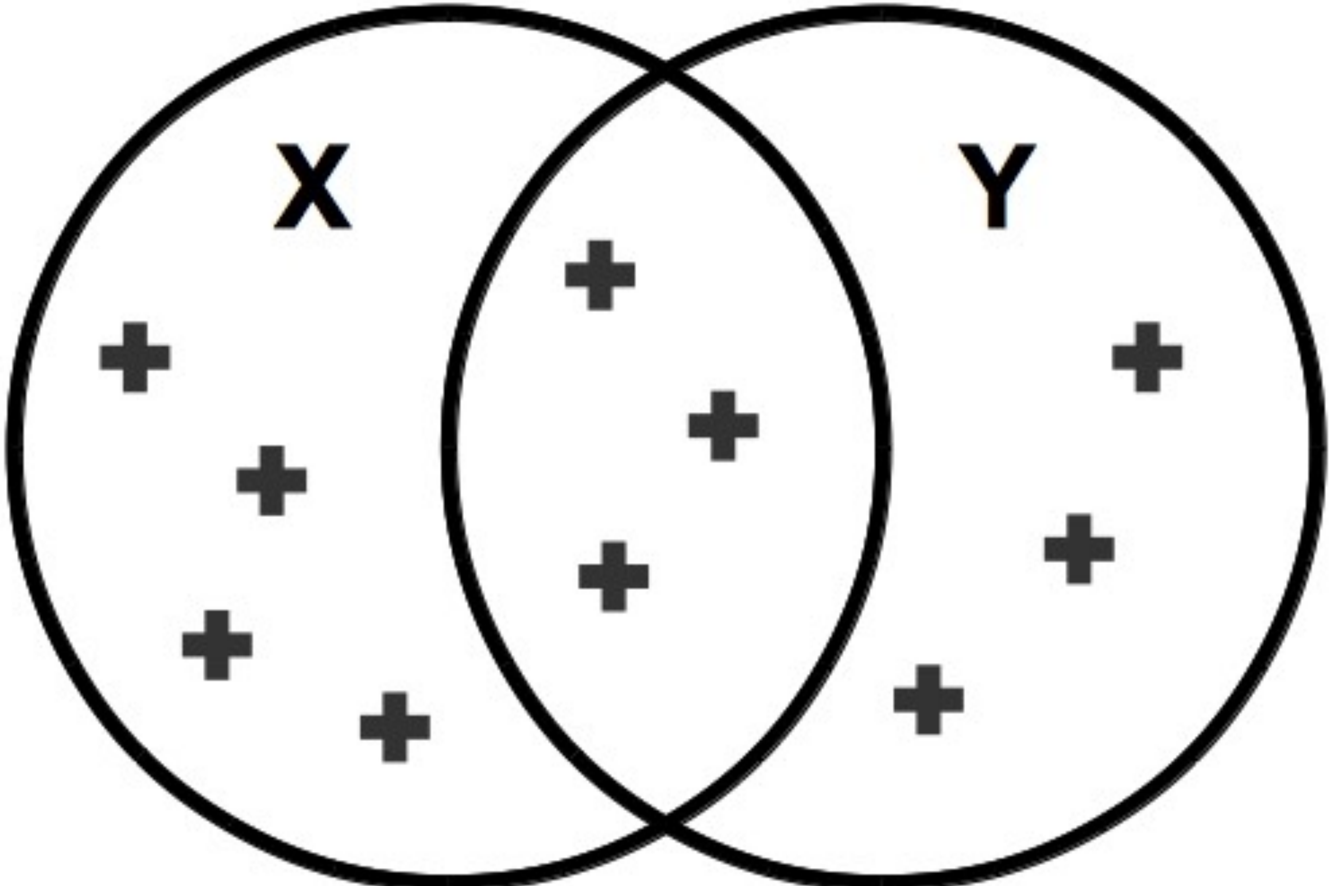
MARGINAL = P(X)



JOINT = P(X,Y)

CONDITIONAL = P(X|Y)

MARGINAL = P(X) $= \int P(x,y)dy$

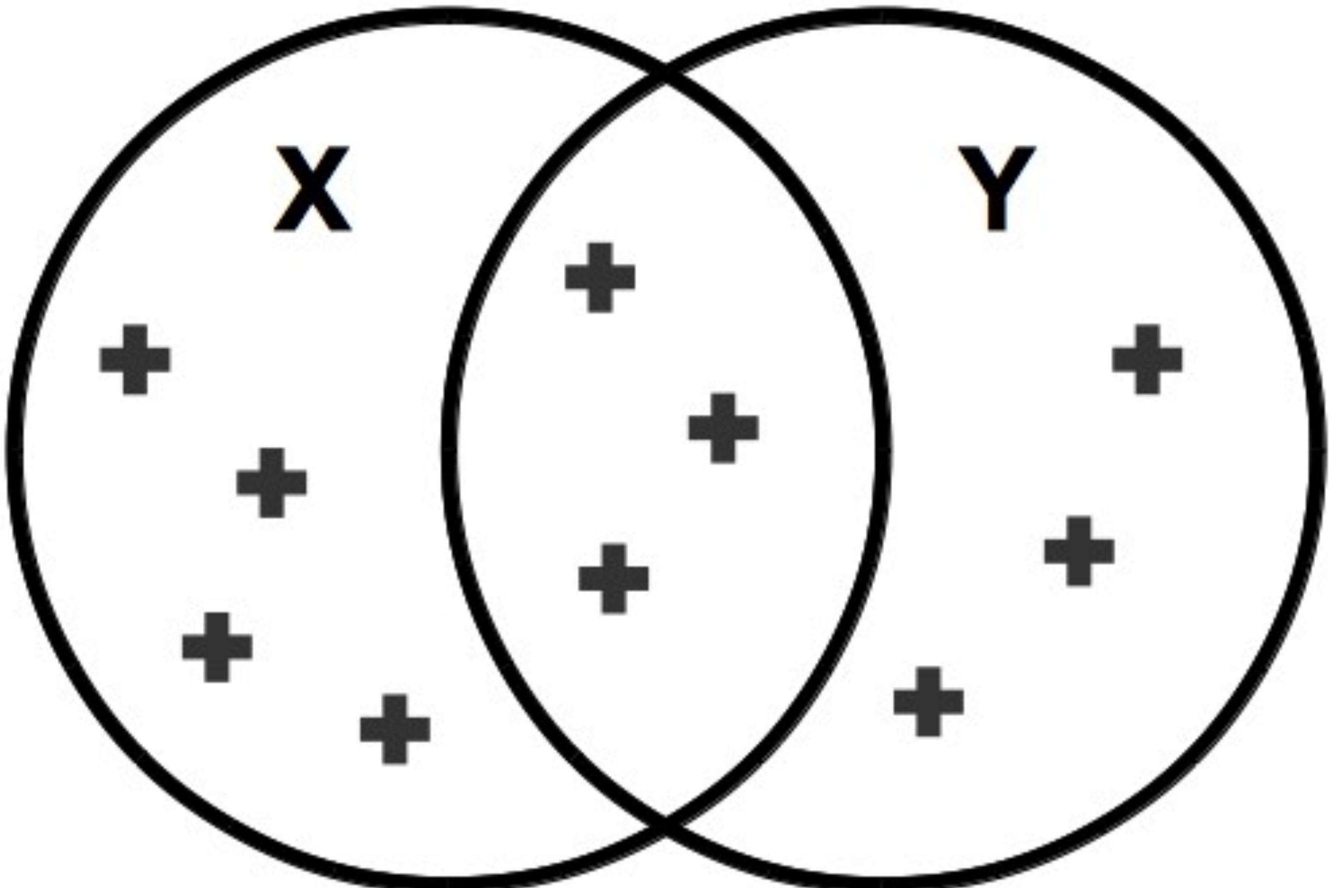


JOINT = P(X,Y)

CONDITIONAL = P(X|Y)

MARGINAL = P(X) $= \int P(x,y)dy$

	Y	!Y	
X	0.3	0.4	0.7
!X	0.3	0	0.3
	0.6	0.4	1



JOINT = P(X,Y)

CONDITIONAL = P(X|Y)

MARGINAL = P(X) $= \int P(x,y)dy$

	Y	!Y	
X	0.3	0.4	0.7
!X	0.3	0	0.3
	0.6	0.4	1

JOINT = P(X,Y)

CONDITIONAL = P(X|Y)

MARGINAL = P(X) $= \int P(x,y)dy$

	Y	!Y	
X	0.3	0.4	0.7
!X	0.3	0	0.3
	0.6	0.4	1

$$CONDITIONAL = P(X \mid Y) = \frac{P(X,Y)}{P(Y)} = \frac{JOINT}{MARGINAL}$$

JOINT = P(X,Y)

CONDITIONAL = P(X|Y)

MARGINAL = P(X) $= \int P(x,y)dy$

	Y	!Y	
X	0.3	0.4	0.7
!X	0.3	0	0.3
	0.6	0.4	1

$$CONDITIONAL = P(X \mid Y) = \frac{P(X,Y)}{P(Y)} = \frac{JOINT}{MARGINAL}$$

$$P(X,Y) = P(X \mid Y)P(Y)$$

JOINT = P(X,Y)

CONDITIONAL = P(X|Y)

MARGINAL = P(X) $= \int P(x,y)dy$

	Y	!Y	
X	0.3	0.4	0.7
!X	0.3	0	0.3
	0.6	0.4	1

$$CONDITIONAL = P(X|Y) = \frac{P(X,Y)}{P(Y)} = \frac{JOINT}{MARGINAL}$$

$$P(X,Y) = P(X|Y)P(Y)$$

$$P(X) = \int P(X,Y) = \int P(X|Y)P(Y)$$

JOINT = CONDITIONAL * MARGINAL

$$P(X, \Theta) = P(X \mid \Theta) * P(\Theta)$$

JOINT = CONDITIONAL * MARGINAL

$$P(X, \Theta) = P(X \mid \Theta) * P(\Theta)$$

$$P(\Theta, X) = P(\Theta \mid X) * P(X)$$

JOINT = CONDITIONAL * MARGINAL

$$P(X, \Theta) = P(X \mid \Theta) * P(\Theta)$$

$$P(\Theta, X) = P(\Theta \mid X) * P(X)$$

$$P(\Theta \mid X) * P(X) = P(X \mid \Theta) * P(\Theta)$$

JOINT = CONDITIONAL * MARGINAL

$$P(X, \Theta) = P(X \mid \Theta) * P(\Theta)$$

$$P(\Theta, X) = P(\Theta \mid X) * P(X)$$

$$P(\Theta \mid X) * P(X) = P(X \mid \Theta) * P(\Theta)$$

$$\overbrace{P(\theta|X)}^{posterior} = \frac{\overbrace{P(X|\theta)}^{likelihood}\ \overbrace{P(\theta)}^{prior}}{P(X)}$$

BAYES' THEOREM

JOINT = CONDITIONAL * MARGINAL

$$P(X, \Theta) = P(X \mid \Theta) * P(\Theta)$$

$$P(\Theta, X) = P(\Theta \mid X) * P(X)$$

$$P(\Theta \mid X) * P(X) = P(X \mid \Theta) * P(\Theta)$$

$$\overbrace{P(\theta|X)}^{posterior} = \frac{\overbrace{P(X|\theta)}^{likelihood}\ \overbrace{P(\theta)}^{prior}}{P(X)}$$

JOINT = CONDITIONAL * MARGINAL

$$P(X, \Theta) = P(X \mid \Theta) * P(\Theta)$$

$$P(\Theta, X) = P(\Theta \mid X) * P(X)$$

$$P(\Theta \mid X) * P(X) = P(X \mid \Theta) * P(\Theta)$$

$$\overbrace{P(\theta|X)}^{posterior} = \frac{\overbrace{P(X|\theta)}^{likelihood}\ \overbrace{P(\theta)}^{prior}}{P(X)}$$

$$= \int P(X, \theta) = \int P(X \mid \theta) P(\theta)$$

$$P(\theta|\mathrm{X}) = \frac{P(\mathrm{X}|\theta)P(\theta)}{\int P(\mathrm{X}|\theta)P(\theta)}$$

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{\int P(X|\theta)P(\theta)}$$

P(MODEL | DATA)

FALSE POSITIVES

- If a patient has a disease the test returns a positive 99% of the time P(+ | D)
- If a patient does not have the disease, the test returns positive 5% of the time P(+ | !D)
- 0.1% of the population has the disease P(D)
- What is the probability that someone who tested positive has the disease? P(D | +)

$$P(D \mid +) = \frac{P(+ \mid D)P(D)}{P(+ \mid D)P(D) + P(+ \mid !D)P(!D)}$$

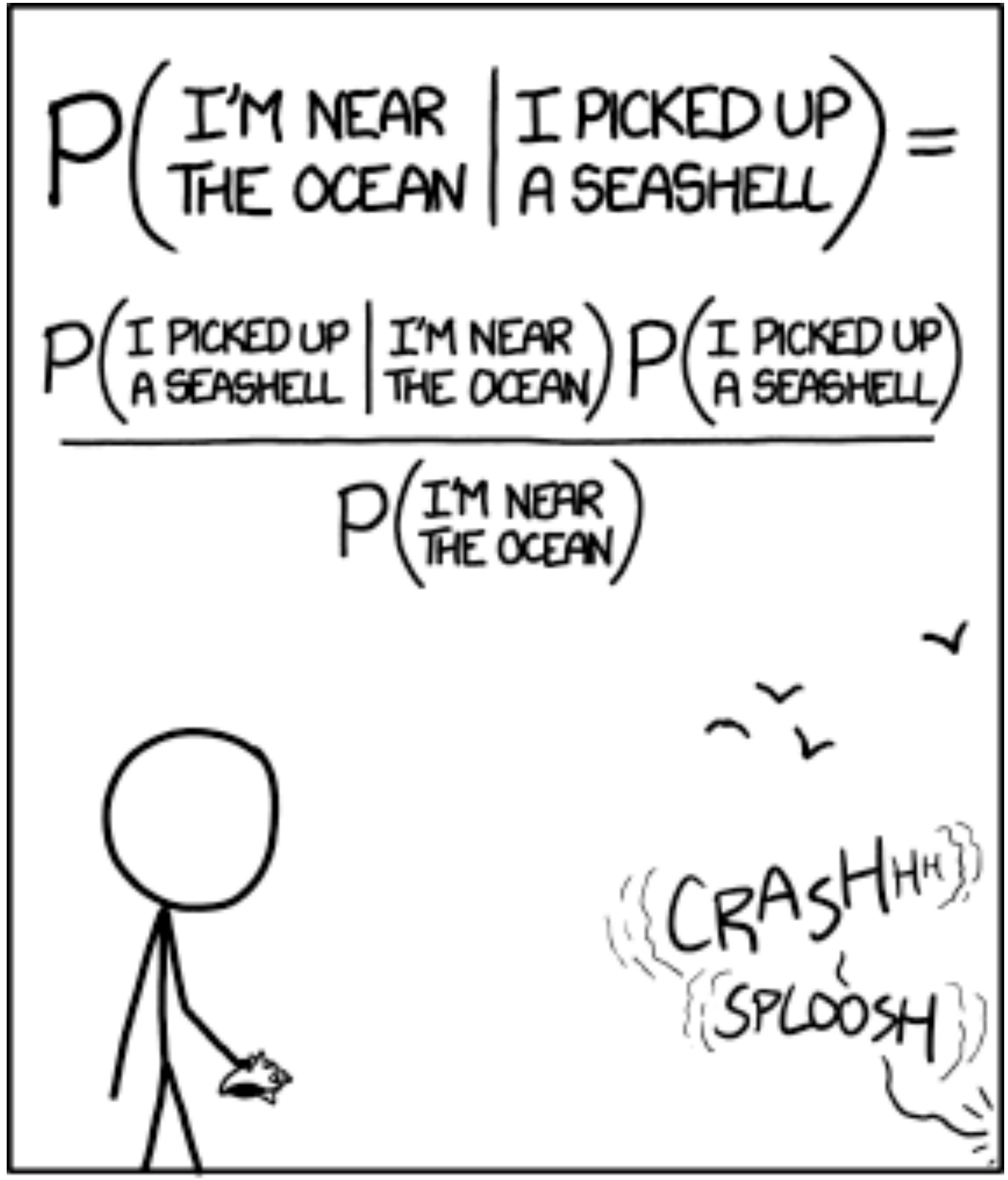
$$P(D \mid +) = \frac{P(+ \mid D)P(D)}{P(+ \mid D)P(D) + P(+ \mid !D)P(!D)}$$

$$P(D \mid +) = \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + 0.05 \cdot 0.999}$$

$$P(D \mid +) = \frac{P(+ \mid D)P(D)}{P(+ \mid D)P(D) + P(+ \mid !D)P(!D)}$$

$$P(D \mid +) = \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + 0.05 \cdot 0.999}$$

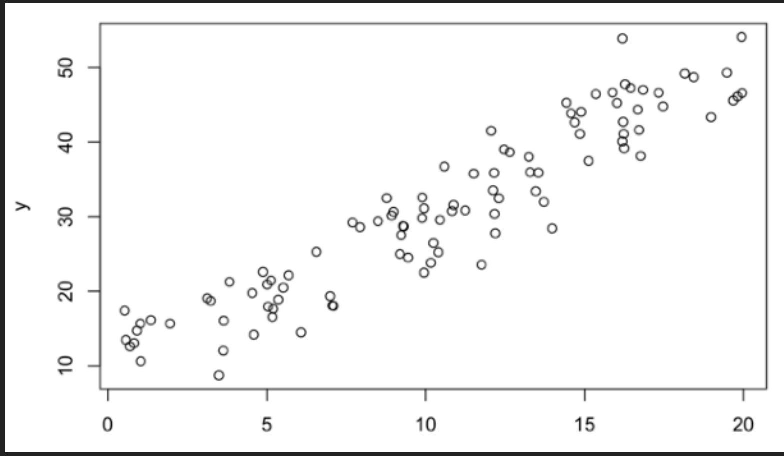
$$P(D \mid +) = \frac{0.00099}{0.00099 \cdot 0.04995} \approx 0.019$$



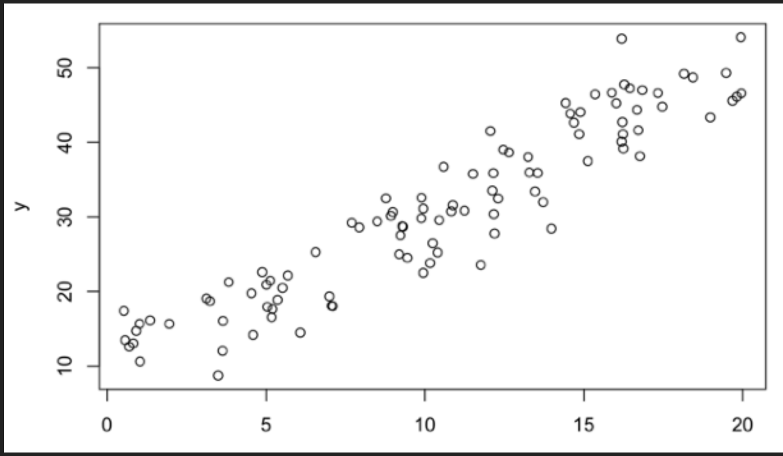
STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND *DON'T* HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.

ALSO WORKS WITH DISTRIBUTIONS AND MODELS

ALSO WORKS WITH DISTRIBUTIONS AND MODELS

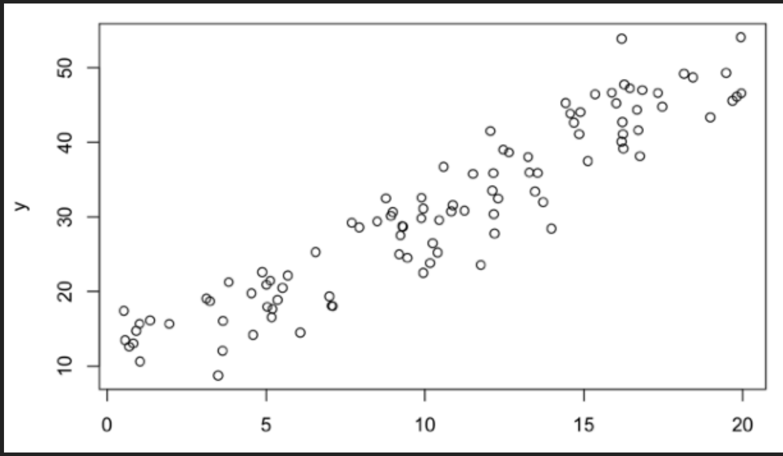


ALSO WORKS WITH DISTRIBUTIONS AND MODELS

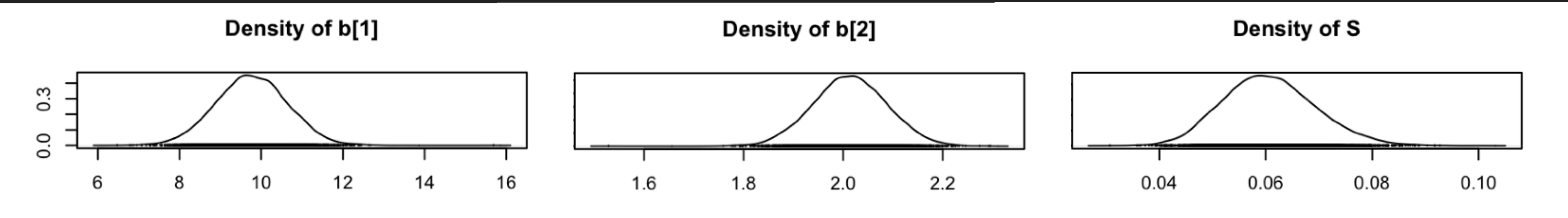


$$P(\vec{\beta}, \sigma^2 | X, Y) = \frac{N(Y | X\vec{\beta}, \sigma^2)P(\sigma^2)P(\beta)}{\int N(Y | X\vec{\beta}, \sigma^2)P(\sigma^2)P(\beta)}$$

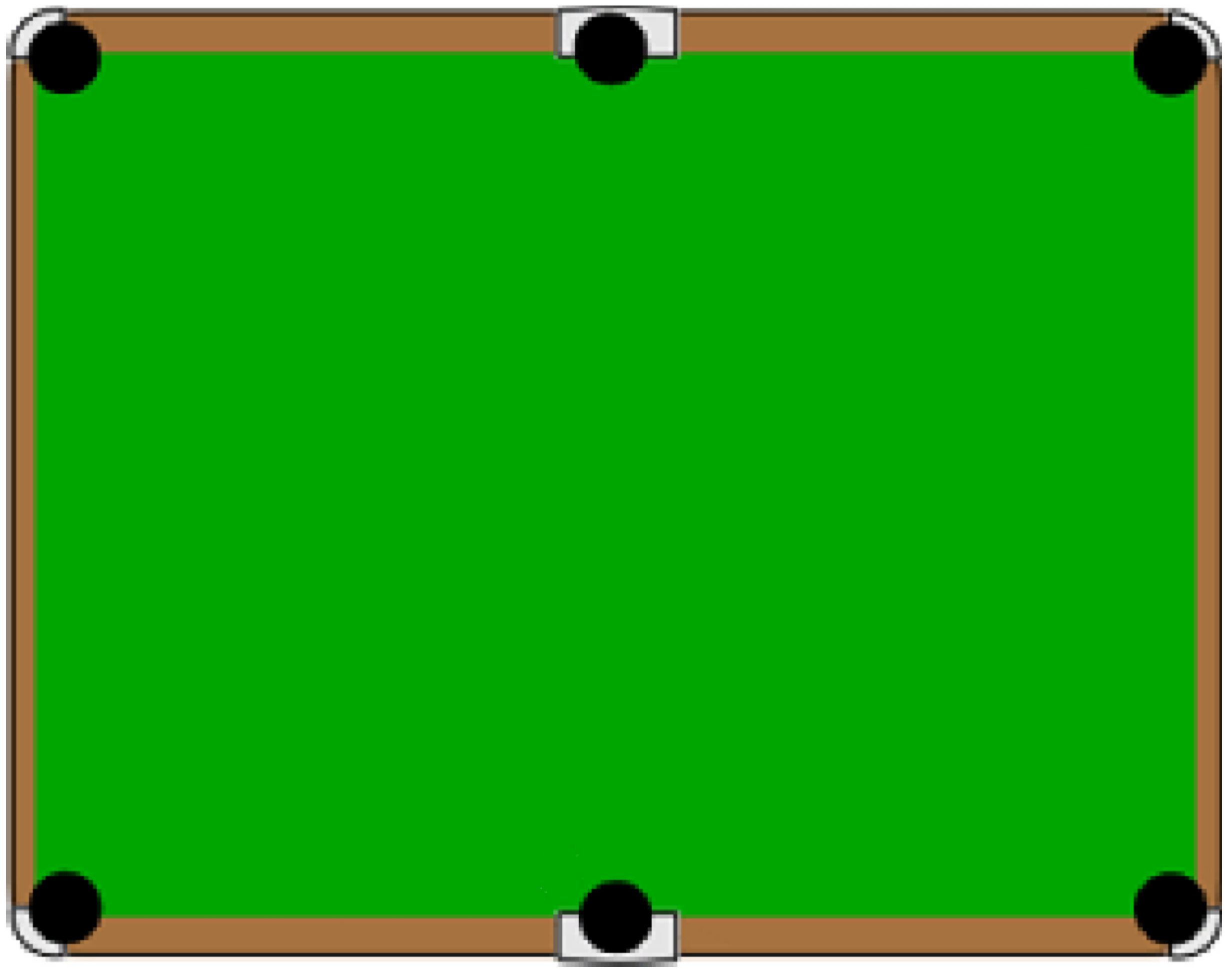
ALSO WORKS WITH DISTRIBUTIONS AND MODELS



$$P(\vec{\beta}, \sigma^2 | X, Y) = \frac{N(Y | X\vec{\beta}, \sigma^2)P(\sigma^2)P(\beta)}{\int N(Y | X\vec{\beta}, \sigma^2)P(\sigma^2)P(\beta)}$$

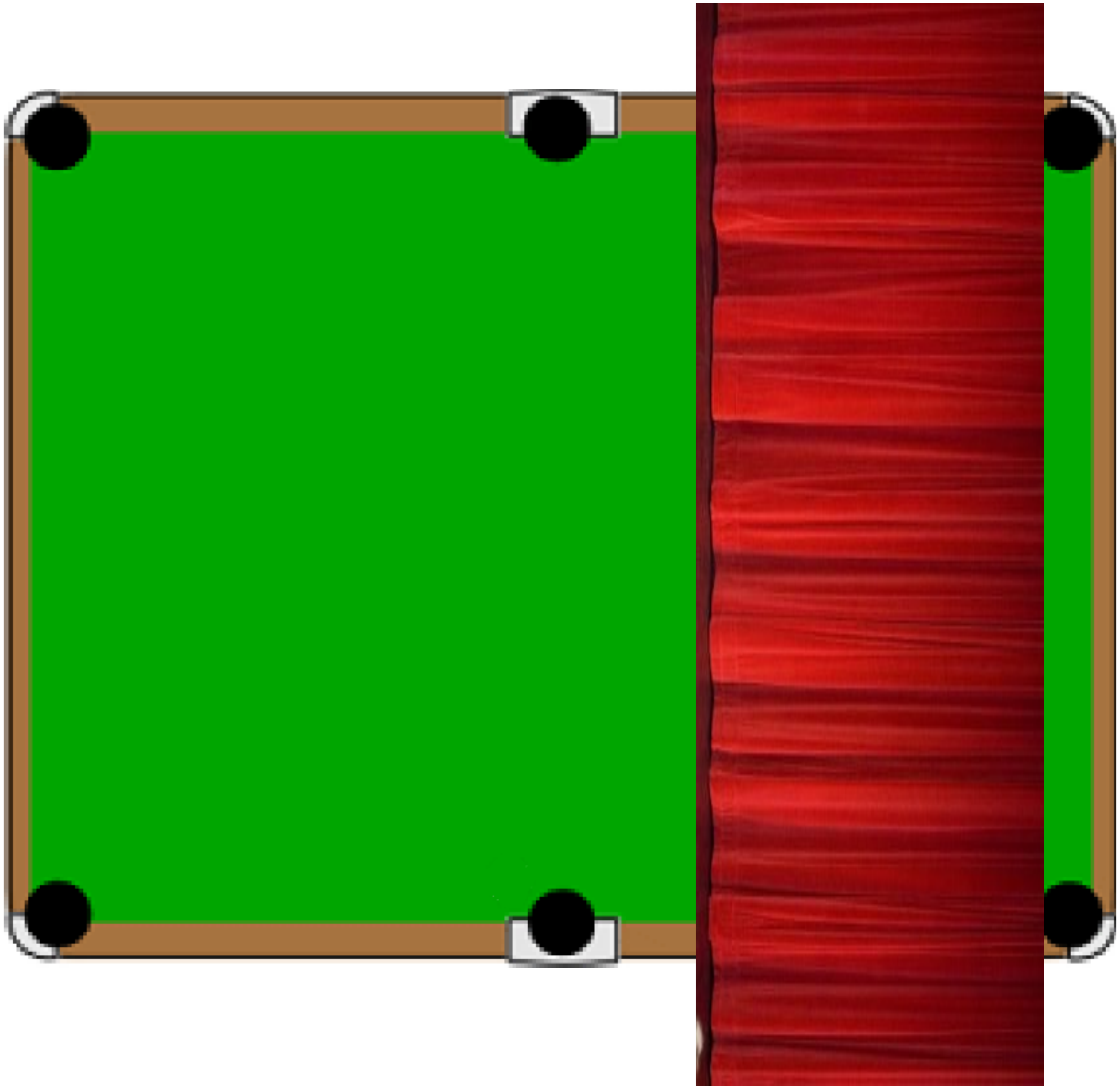


Bayes' Billiard Table



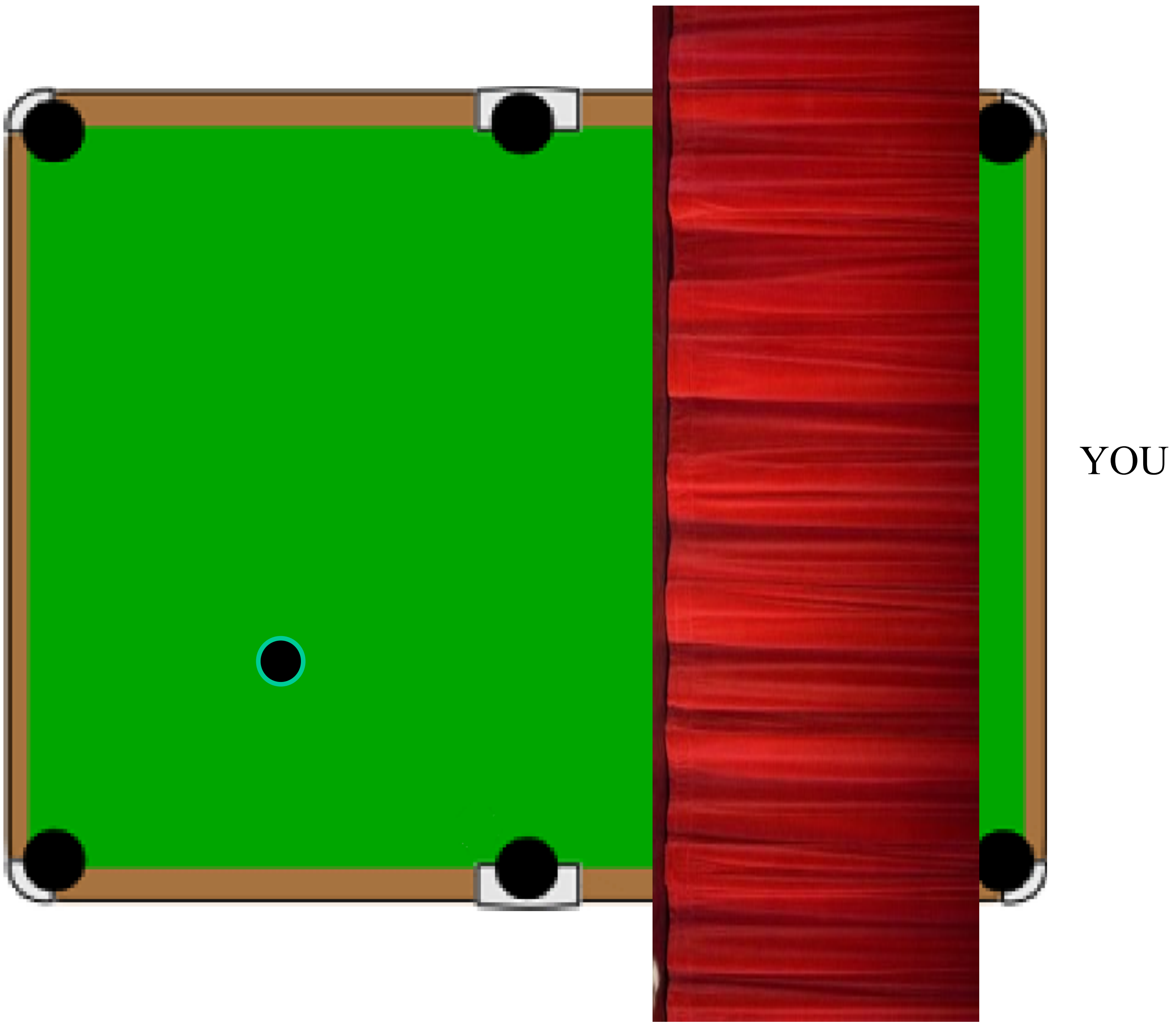
YOU

Bayes' Billiard Table

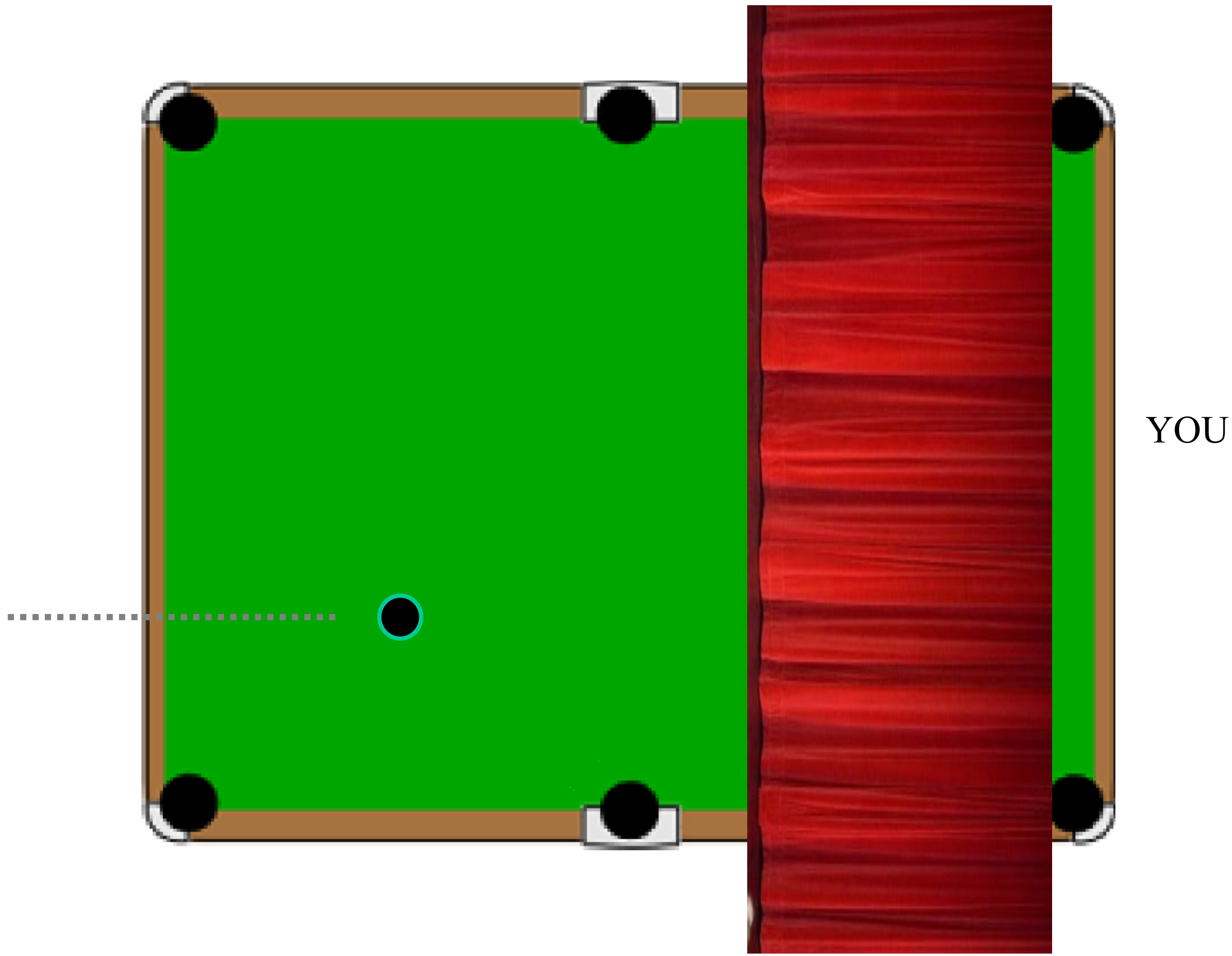


YOU

Bayes' Billiard Table

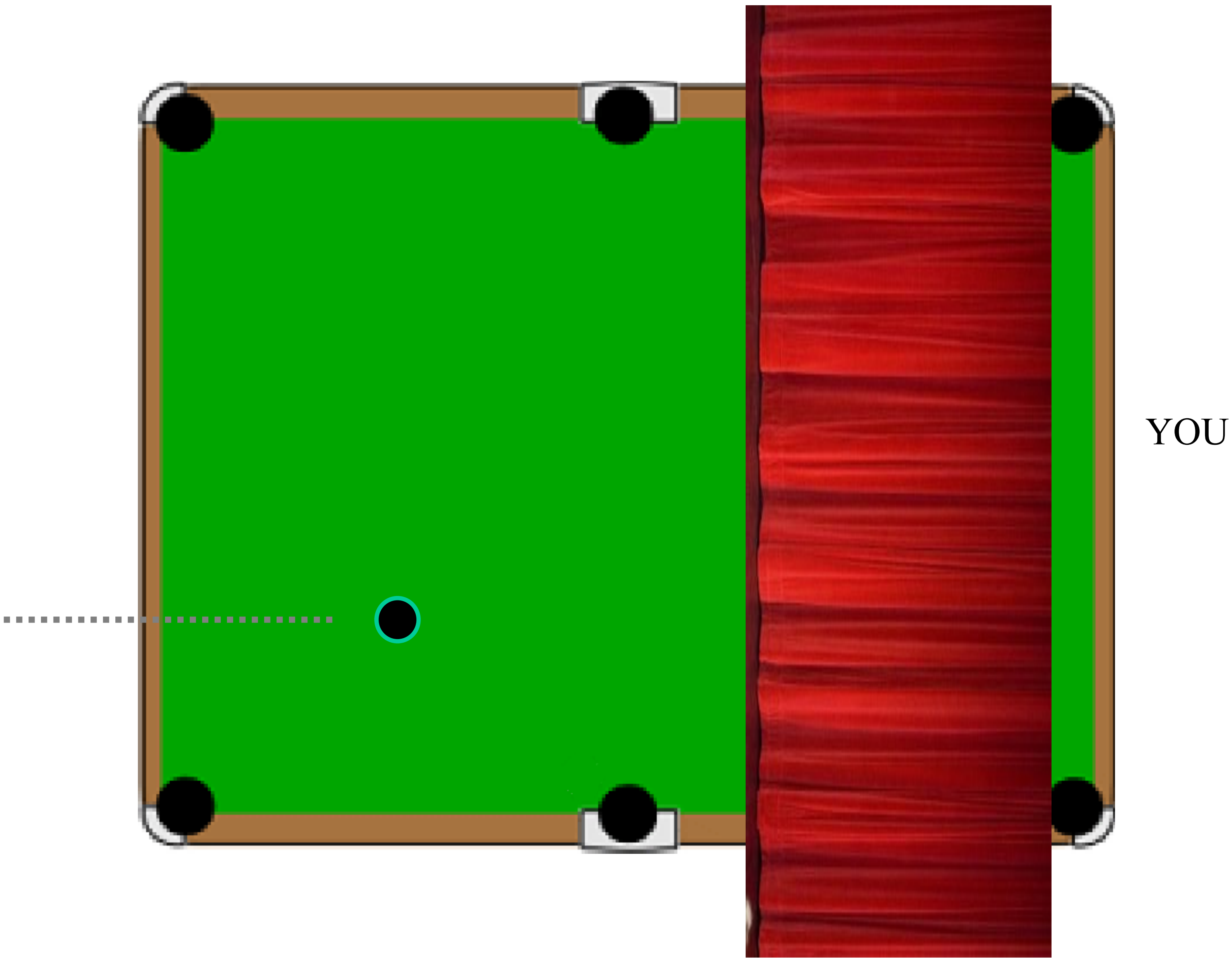


Bayes' Billiard Table



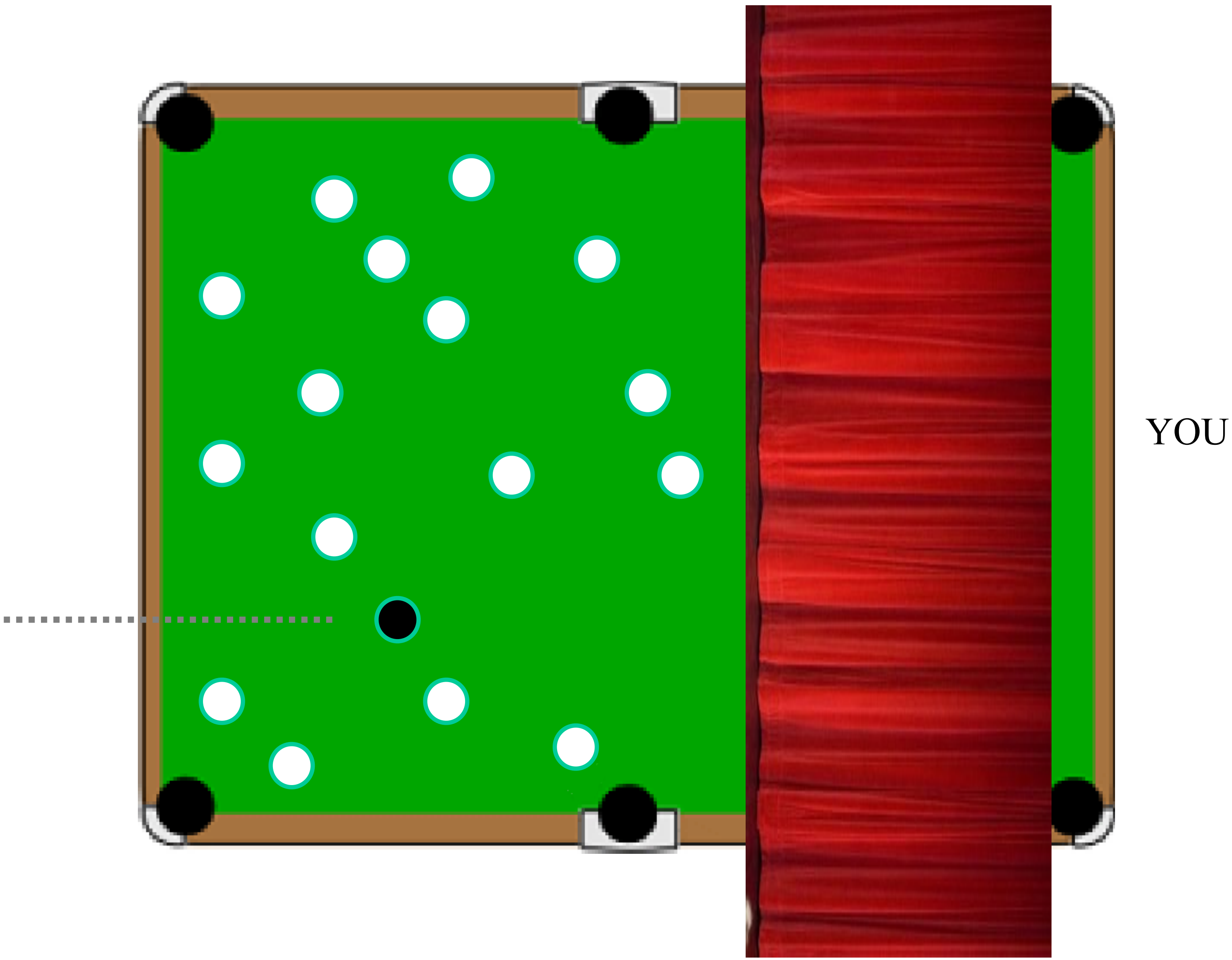
Bayes' Billiard Table

$$P(\theta) = Unif$$



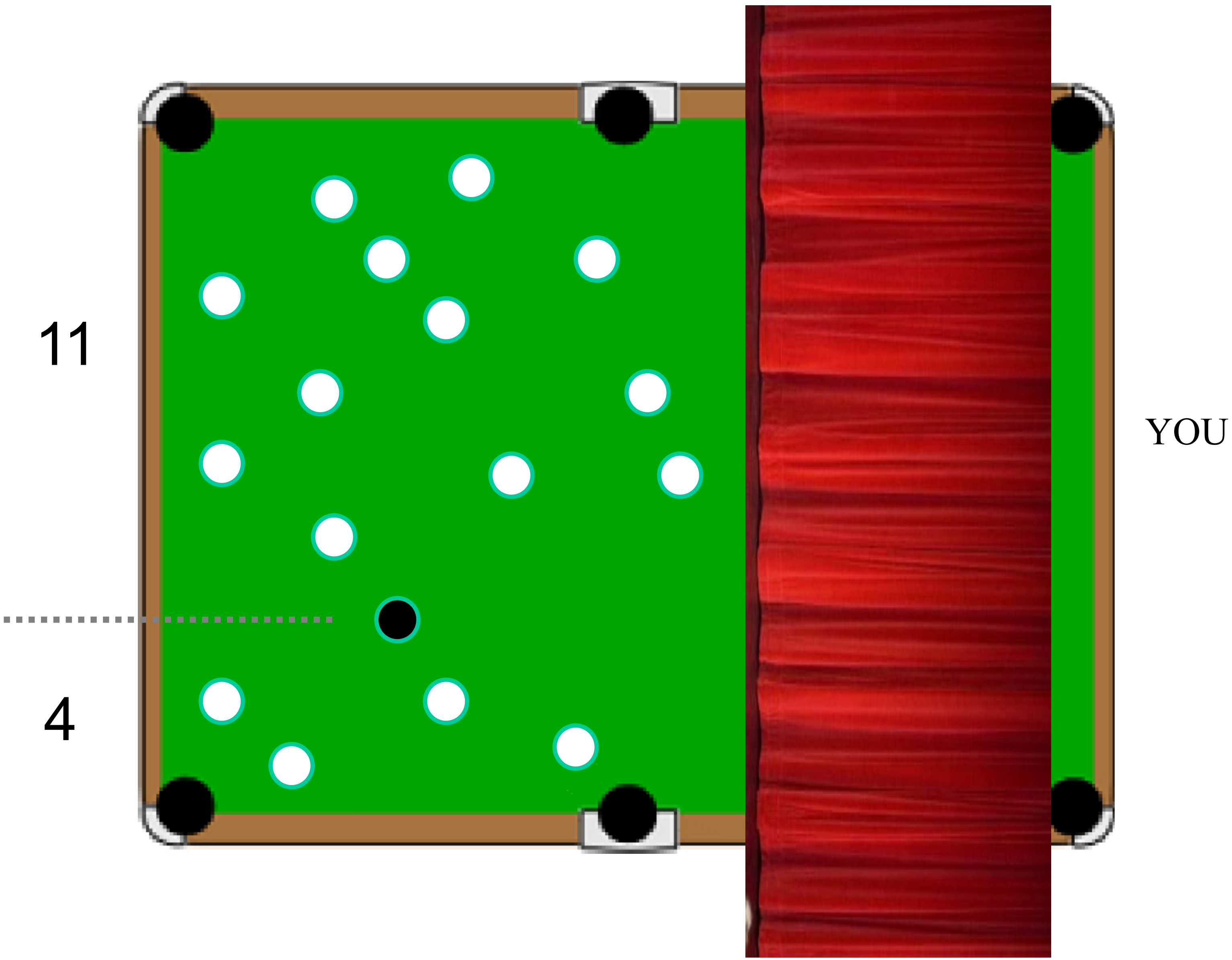
Bayes' Billiard Table

$P(\theta) = Unif$



Bayes' Billiard Table

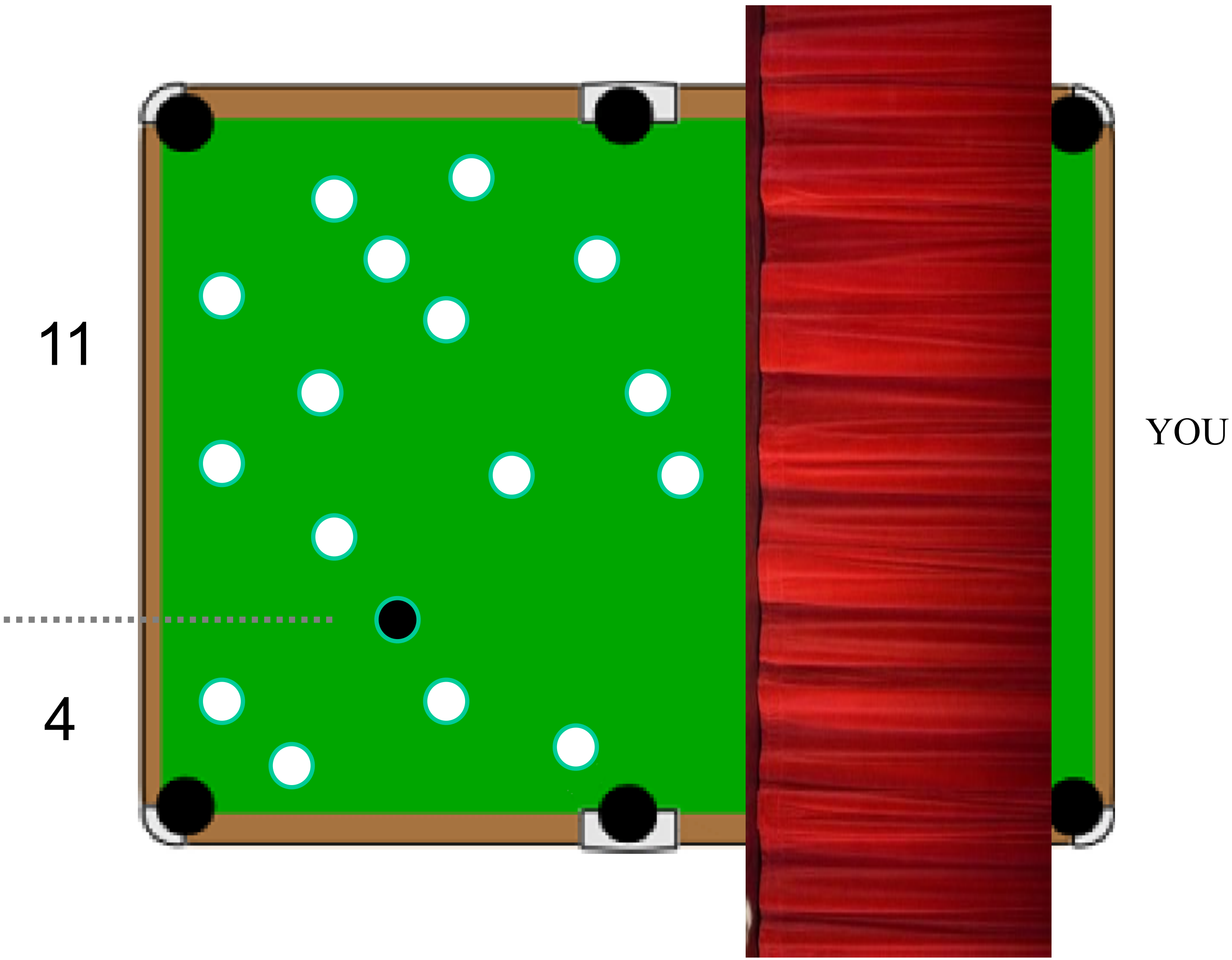
$$P(\theta) = Unif$$



Bayes' Billiard Table

$$P(\theta) = Unif$$

$$P(\theta \,|\, Y) = ??$$



$$P(\theta \mid Y) \propto P(Y \mid \theta)P(\theta)$$

$$P(\theta \mid Y) \propto P(Y \mid \theta)P(\theta)$$

Unif(0,1)

$$P(\theta \mid Y) \propto P(Y \mid \theta)P(\theta)$$

Unif(0,1)

What is $P(y \mid \theta)$?

$$P(\theta \mid Y) \propto P(Y \mid \theta)P(\theta)$$

Unif(0,1)

What is P(y | θ)?

$$L = P(Y \mid \theta) = Binom(Y \mid N, \theta)$$

$$P(\theta \mid Y) = \frac{Binom(Y \mid N, \theta) Unif(\theta \mid 0,1)}{\int_0^1 Binom(Y \mid N, \theta) Unif(\theta \mid 0,1)}$$

$$P(\theta \mid Y) = \frac{Binom(Y \mid N, \theta)Unif(\theta \mid 0,1)}{\int_0^1 Binom(Y \mid N, \theta)Unif(\theta \mid 0,1)}$$

$$P(\theta \mid Y) = \frac{\binom{N}{Y} \theta^Y (1-\theta)^{N-Y} \cdot 1}{\int_0^1 \binom{N}{Y} \theta^Y (1-\theta)^{N-Y} \cdot 1}$$

$$P(\theta \mid Y) = \frac{\theta^Y (1-\theta)^{N-Y}}{\int_0^1 \theta^Y (1-\theta)^{N-Y}}$$

$$P(\theta \mid Y) = \frac{\theta^Y(1-\theta)^{N-Y}}{\int_0^1 \theta^Y(1-\theta)^{N-Y}}$$

What do I do with this?

$$P(\theta \mid Y) = \frac{\theta^{Y}(1-\theta)^{N-Y}}{\int_0^1 \theta^{Y}(1-\theta)^{N-Y}}$$

What do I do with this?

mean? Var? CI?

$$P(\theta \mid Y) = \frac{\theta^{Y}(1-\theta)^{N-Y}}{\int_0^1 \theta^{Y}(1-\theta)^{N-Y}}$$

What do I do with this?

mean? Var? CI?

$$Beta(x \mid \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}}$$

$$P(\theta \mid Y) = \frac{\theta^{Y}(1-\theta)^{N-Y}}{\int_0^1 \theta^{Y}(1-\theta)^{N-Y}}$$

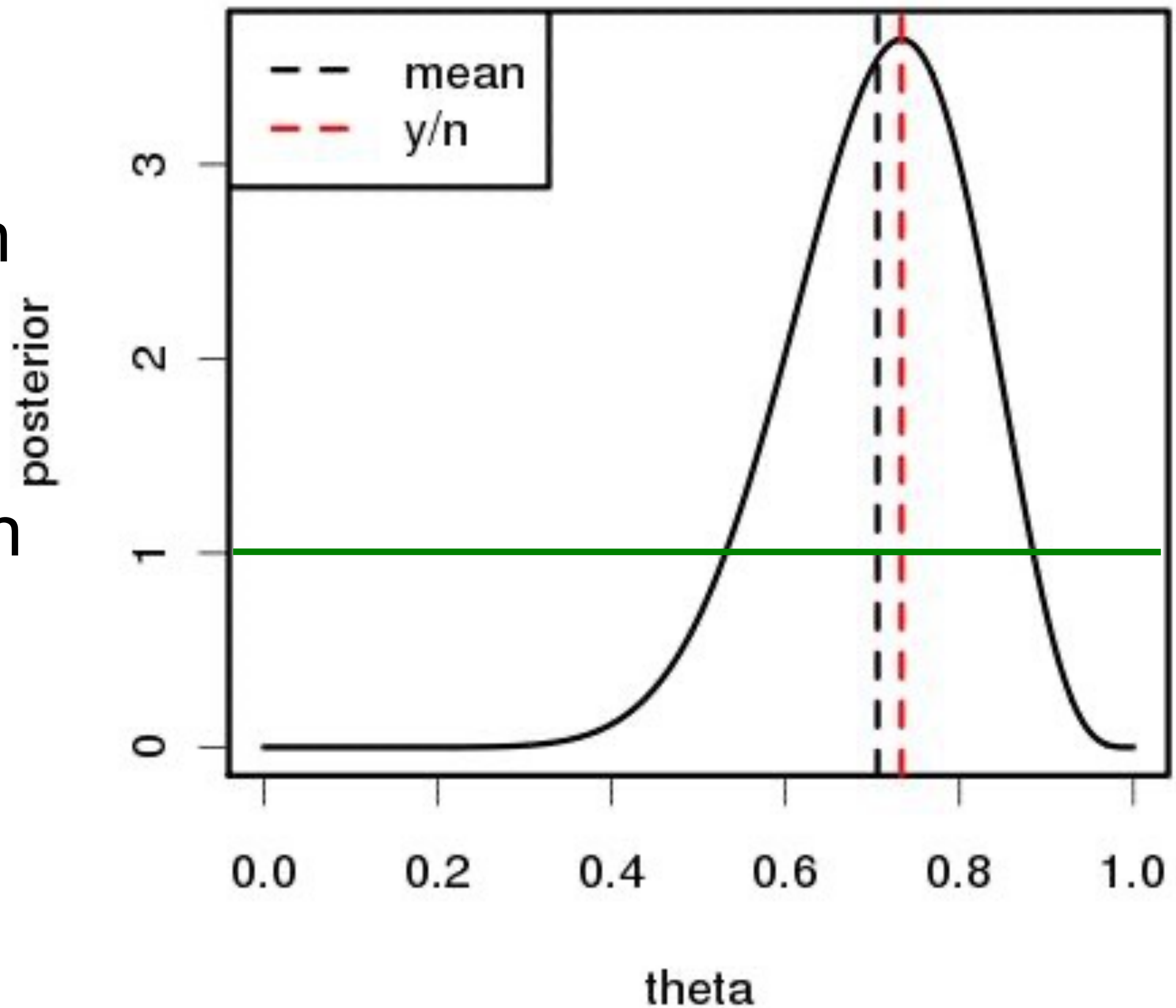
What do I do with this?

mean? Var? CI?

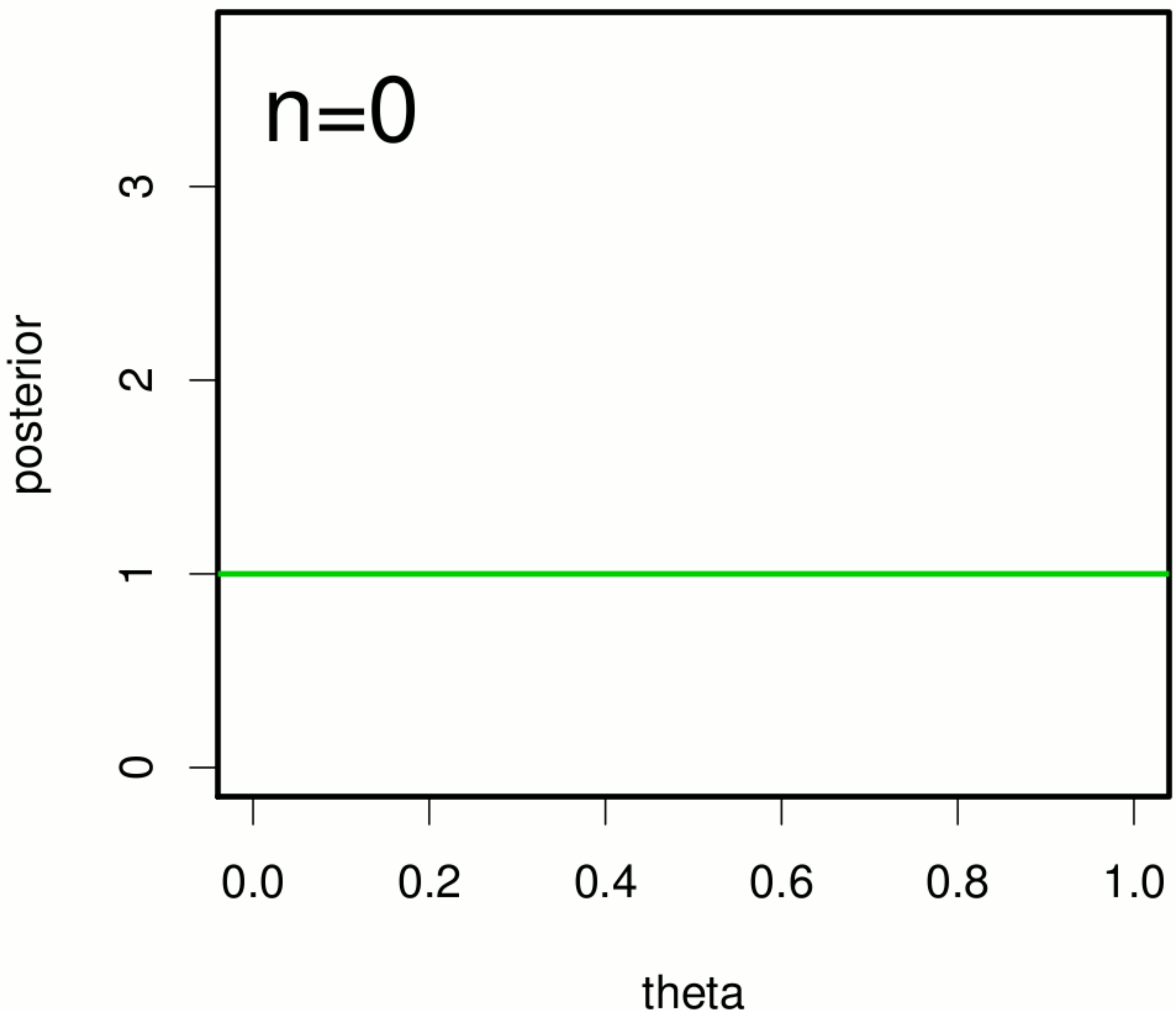
$$Beta(x \mid \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}}$$

$$P(\theta \mid Y) = Beta(\theta \mid Y+1, N-Y+1)$$

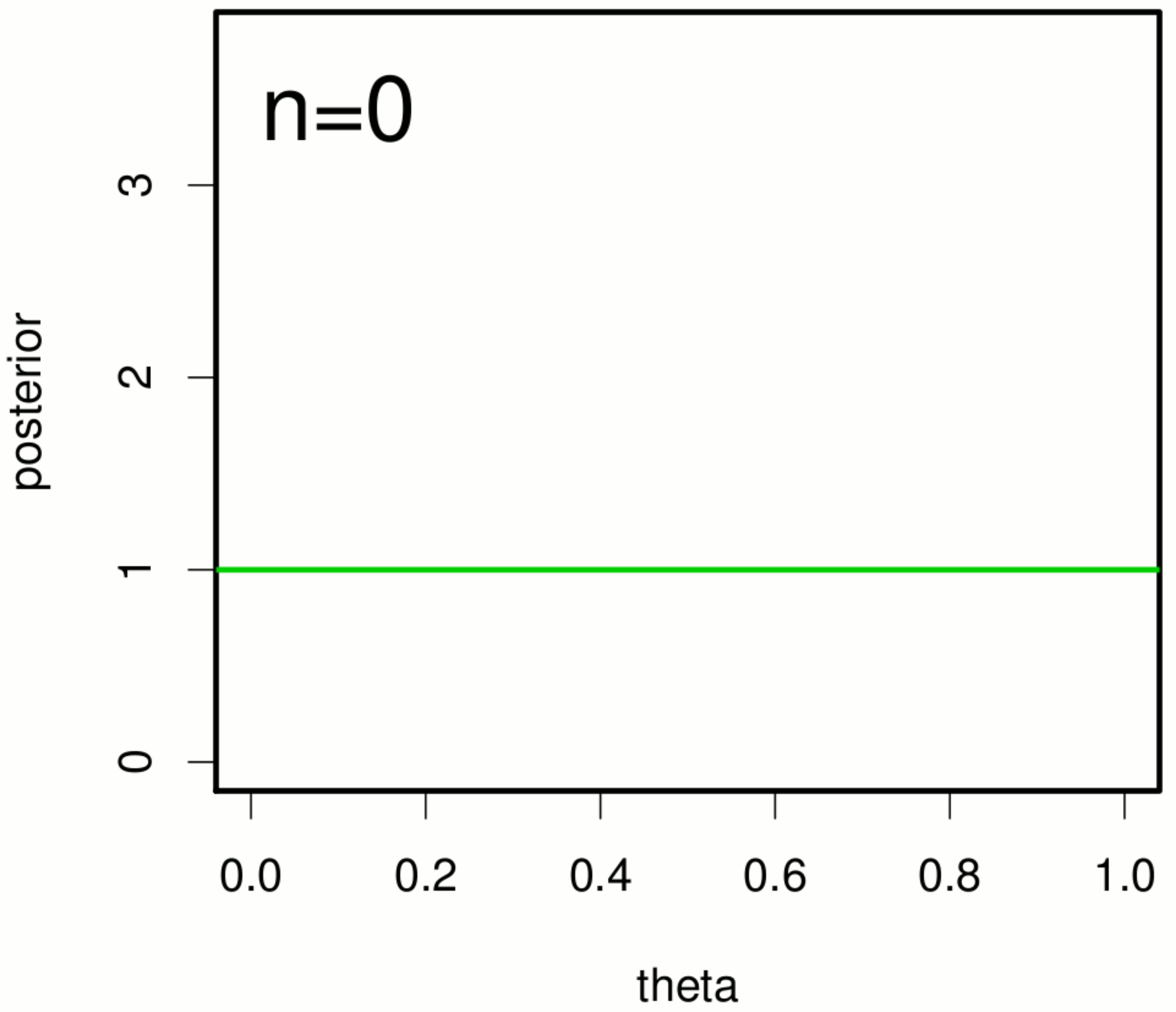
- Posterior **__is__** a PDF
- θ **__is__** a random variable
- Interested in full distribution



Data updates the prior



Data updates the prior



Normal mean and variance

Normal mean and variance

$$L = P(Y \mid \mu) = N(Y \mid \mu, \sigma^2) \propto exp \left[\frac{-(Y-\mu)^2}{2\sigma^2} \right]$$

Normal mean and variance

$$L = P(Y \mid \mu) = N(Y \mid \mu, \sigma^2) \propto exp\left[\frac{-(Y-\mu)^2}{2\sigma^2}\right]$$

$$prior = P(\mu) = N(\mu \mid \mu_0, \tau^2) \propto exp\left[\frac{-(\mu-\mu_0)^2}{2\tau^2}\right]$$

Normal mean and variance

$$L = P(Y \mid \mu) = N(Y \mid \mu, \sigma^2) \propto exp\left[\frac{-(Y-\mu)^2}{2\sigma^2}\right]$$

$$prior = P(\mu) = N(\mu \mid \mu_0, \tau^2) \propto exp\left[\frac{-(\mu-\mu_0)^2}{2\tau^2}\right]$$

Normal mean and variance

$$L = P(Y \mid \mu) = N(Y \mid \mu, \sigma^2) \propto exp\left[\frac{-(Y-\mu)^2}{2\sigma^2}\right]$$

$$prior = P(\mu) = N(\mu \mid \mu_0, \tau^2) \propto exp\left[\frac{-(\mu-\mu_0)^2}{2\tau^2}\right]$$

Prior Mean

Normal mean and variance

$$L = P(Y \mid \mu) = N(Y \mid \mu, \sigma^2) \propto exp \left[\frac{-(Y-\mu)^2}{2\sigma^2} \right]$$

$$prior = P(\mu) = N(\mu \mid \mu_0, \tau^2) \propto exp \left[\frac{-(\mu-\mu_0)^2}{2\tau^2} \right]$$

Prior Mean

Prior Variance

$$P(\mu \mid Y) = N(Y \mid \mu, \sigma^2) \cdot N(\mu \mid \mu_0, \tau^2)$$

$$P(\mu \mid Y) = N(Y \mid \mu, \sigma^2) \cdot N(\mu \mid \mu_0, \tau^2)$$

$$\propto exp\left[\frac{-(Y-\mu)^2}{2\sigma^2}\right] \cdot exp\left[\frac{-(\mu-\mu_0)^2}{2\tau^2}\right]$$

$$P(\mu \mid Y) = N(Y \mid \mu, \sigma^2) \cdot N(\mu \mid \mu_0, \tau^2)$$

$$\propto exp\left[\frac{-(Y-\mu)^2}{2\sigma^2}\right] \cdot exp\left[\frac{-(\mu-\mu_0)^2}{2\tau^2}\right]$$

$$\propto exp\left[\frac{-(Y-\mu)^2}{2\sigma^2} + \frac{-(\mu-\mu_0)^2}{2\tau^2}\right]$$

$$P(\mu \mid Y) = N\left(\mu \,\middle|\, \frac{\left(\frac{Y}{\sigma^2} + \frac{\mu_0}{\tau^2}\right)}{\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)}, \frac{1}{\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)}\right)$$

$$P(\mu \mid Y) = N\left(\mu \,\middle|\, \frac{\left(\frac{Y}{\sigma^2} + \frac{\mu_0}{\tau^2}\right)}{\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)}, \frac{1}{\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)}\right)$$

Precision = 1/variance

$$P(\mu \mid Y) = N\left(\mu \middle| \frac{\left(\frac{Y}{\sigma^2} + \frac{\mu_0}{\tau^2}\right)}{\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)}, \frac{1}{\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)}\right)$$

Precision = 1/variance $\quad S = 1/\sigma^2$

$$P(\mu \mid Y) = N\left(\mu \,\middle|\, \frac{\left(\frac{Y}{\sigma^2} + \frac{\mu_0}{\tau^2}\right)}{\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)}, \frac{1}{\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)}\right)$$

Precision = 1/variance $\quad S = 1/\sigma^2 \quad T = 1/\tau^2$

$$P(\mu \mid Y) = N\left(\mu \,\middle|\, \frac{\left(\frac{Y}{\sigma^2} + \frac{\mu_0}{\tau^2}\right)}{\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)}, \frac{1}{\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)}\right)$$

Precision = 1/variance $\quad S = 1/\sigma^2 \quad T = 1/\tau^2$

$$P(\mu \mid Y) = N\left(\mu \,\middle|\, Y \cdot \frac{S}{S+T} + \mu_0 \cdot \frac{T}{S+T}, \frac{1}{S+T}\right)$$

$$P(\mu \mid Y) = N\left(\mu \,\middle|\, \frac{\left(\frac{Y}{\sigma^2} + \frac{\mu_0}{\tau^2}\right)}{\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)}, \frac{1}{\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)}\right)$$

Precision = 1/variance $\quad S = 1/\sigma^2 \quad T = 1/\tau^2$

$$P(\mu \mid Y) = N\left(\mu \,\middle|\, Y \cdot \frac{S}{S+T} + \mu_0 \cdot \frac{T}{S+T}, \frac{1}{S+T}\right)$$

Precision weighted average of data and prior

What if we can't solve the model analytically??

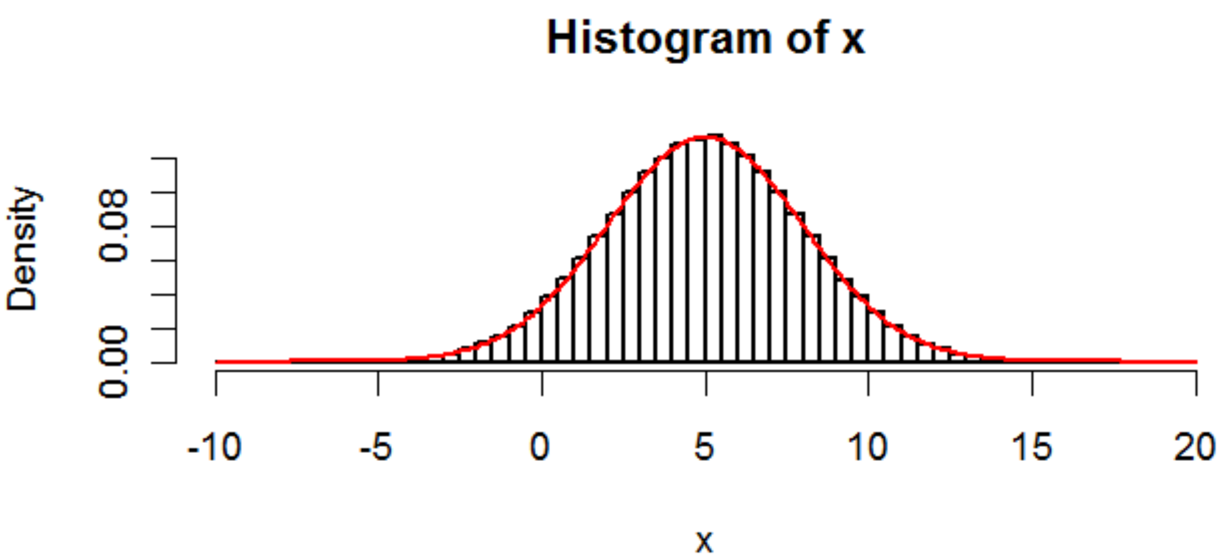
Numerical Methods for Bayes

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{\int P(X|\theta)P(\theta)}$$

- Not just optimization
- Need to integrate denominator
  - Numerical Integration
- Would also like to know the mean, median, mode, variance, quantiles, confidence intervals, etc.

Idea:
Random samples from the posterior

- Approximate PDF with the histogram
- Performs *Monte Carlo Integration*
- Allows all quantities of interest to be calculated from the sample (mean, quantiles, var, etc)



	TRUE	Sample
mean	5.000	5.000
median	5.000	5.004
var	9.000	9.006
Lower CI	-0.880	-0.881
Upper CI	10.880	10.872

Markov Chain Monte Carlo

Markov Chain Monte Carlo

1) Start from some initial parameter value

Markov Chain Monte Carlo

1) Start from some initial parameter value θ

Markov Chain Monte Carlo

1) Start from some initial parameter value θ
2) Calculate the unnormalized posterior

Markov Chain Monte Carlo

1) Start from some initial parameter value θ

2) Calculate the unnormalized posterior $P(Y \mid \theta)P(\theta)$

Markov Chain Monte Carlo

1) Start from some initial parameter value θ
2) Calculate the unnormalized posterior $P(Y \mid \theta)P(\theta)$
3) Propose a new parameter value

Markov Chain Monte Carlo

1) Start from some initial parameter value θ
2) Calculate the unnormalized posterior $P(Y|\theta)P(\theta)$
3) Propose a new parameter value $\theta*$

Markov Chain Monte Carlo

1) Start from some initial parameter value θ
2) Calculate the unnormalized posterior $P(Y|\theta)P(\theta)$
3) Propose a new parameter value $\theta*$
4) Calculate the new unnormalized posterior

Markov Chain Monte Carlo

1) Start from some initial parameter value θ
2) Calculate the unnormalized posterior $P(Y|\theta)P(\theta)$
3) Propose a new parameter value θ^*
4) Calculate the new unnormalized posterior $P(Y|\theta^*)P(\theta^*)$

Markov Chain Monte Carlo

1) Start from some initial parameter value θ
2) Calculate the unnormalized posterior $P(Y \mid \theta)P(\theta)$
3) Propose a new parameter value θ^*
4) Calculate the new unnormalized posterior $P(Y \mid \theta^*)P(\theta^*)$
5) Decide whether or not to accept the new value

Markov Chain Monte Carlo

1) Start from some initial parameter value θ

2) Calculate the unnormalized posterior $P(Y|\theta)P(\theta)$

3) Propose a new parameter value $\theta*$

4) Calculate the new unnormalized posterior $P(Y|\theta*)P(\theta*)$

5) Decide whether or not to accept the new value

$$P(accept) = \frac{P(Y|\theta*)P(\theta*)}{P(Y|\theta)P(\theta)}$$

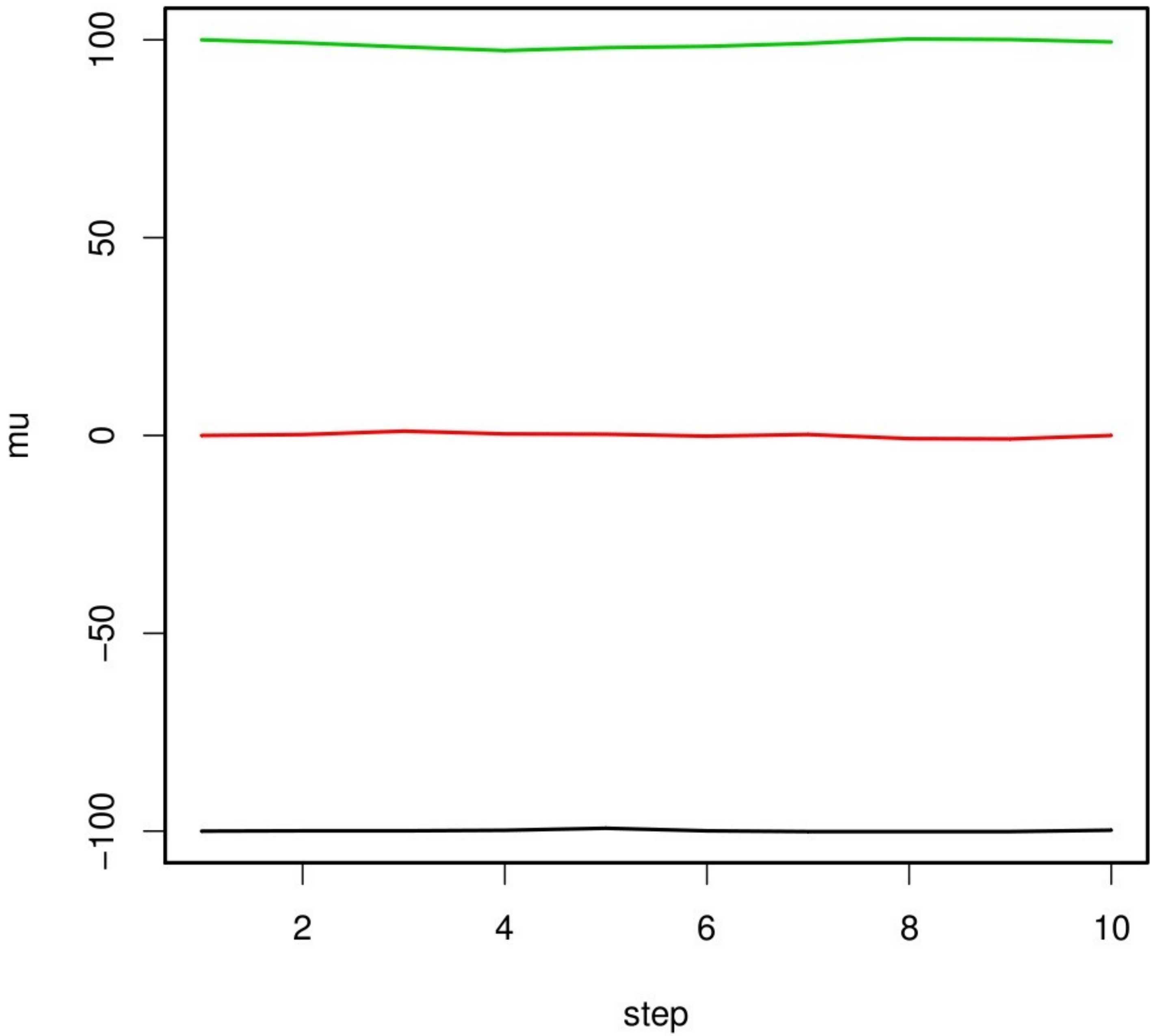
Markov Chain Monte Carlo

1) Start from some initial parameter value θ
2) Calculate the unnormalized posterior $P(Y|\theta)P(\theta)$
3) Propose a new parameter value θ^*
4) Calculate the new unnormalized posterior $P(Y|\theta^*)P(\theta^*)$
5) Decide whether or not to accept the new value
6) Repeat 3-5

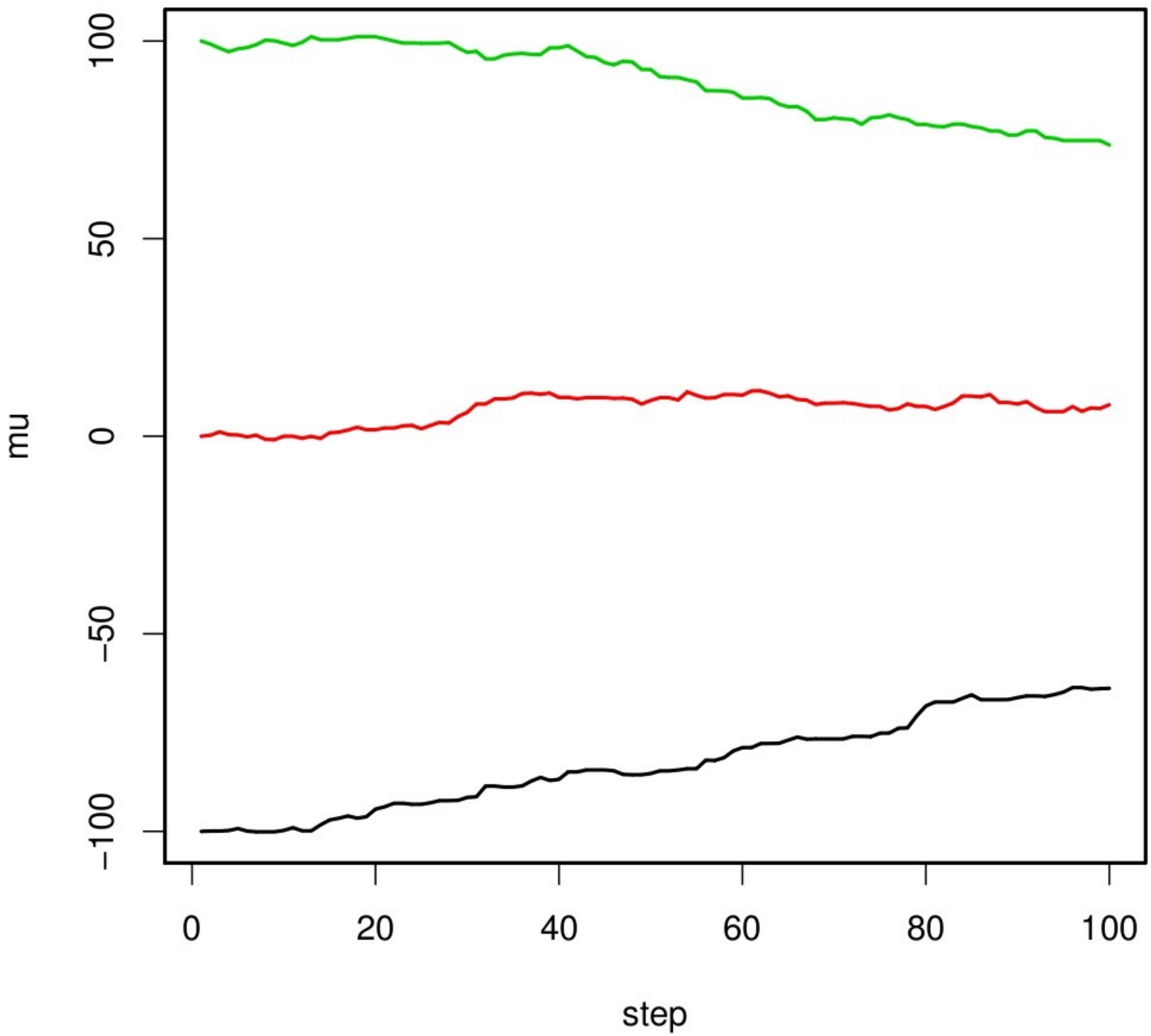
$$P(accept) = \frac{P(Y|\theta^*)P(\theta^*)}{P(Y|\theta)P(\theta)}$$

Example

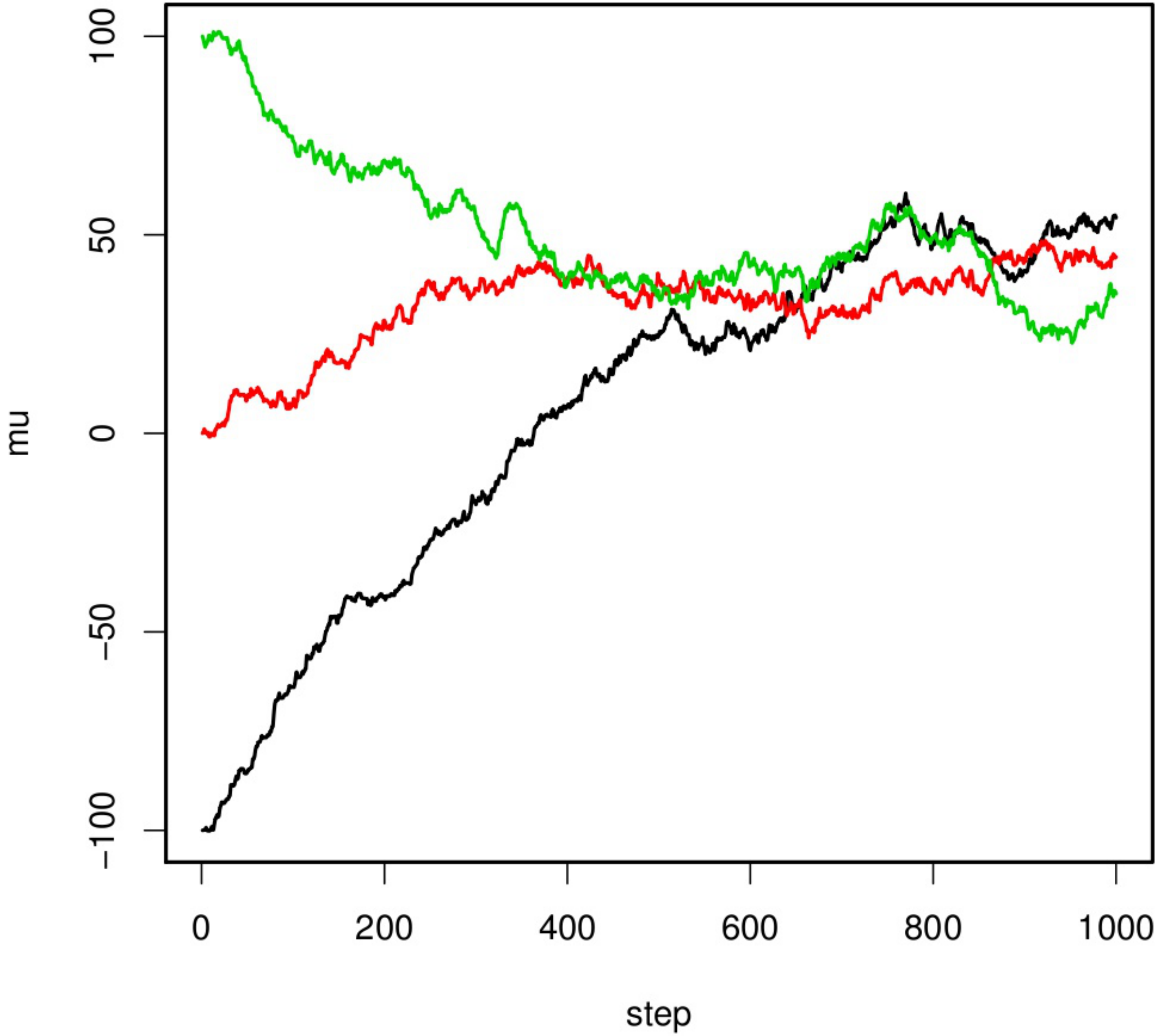
- Normal with known variance, unknown mean
  - Prior: N(53,10000)
  - Data: y = 43
  - Known variance: 100
  - Initial conditions, 3 chains starting at -100, 0, 100

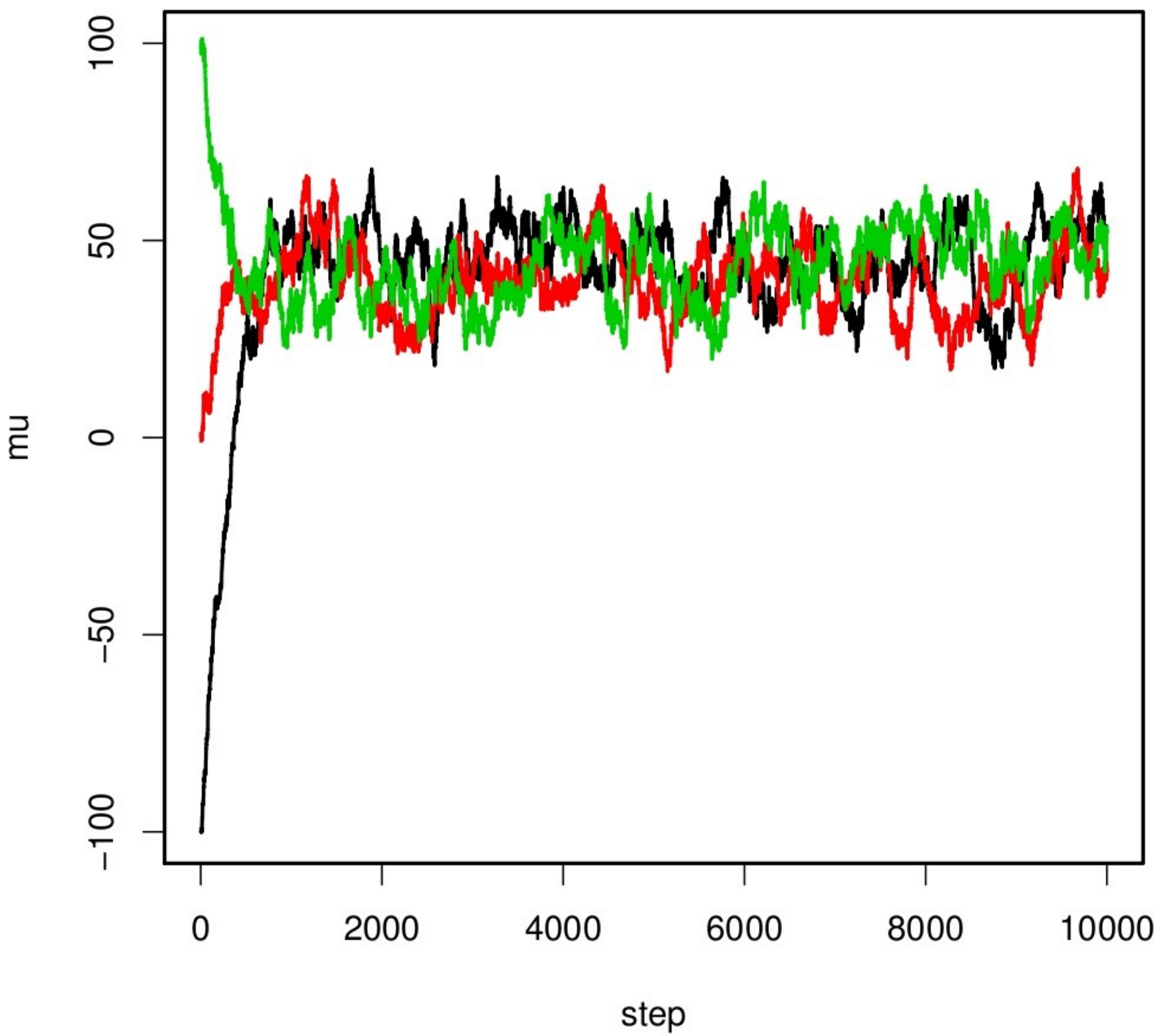


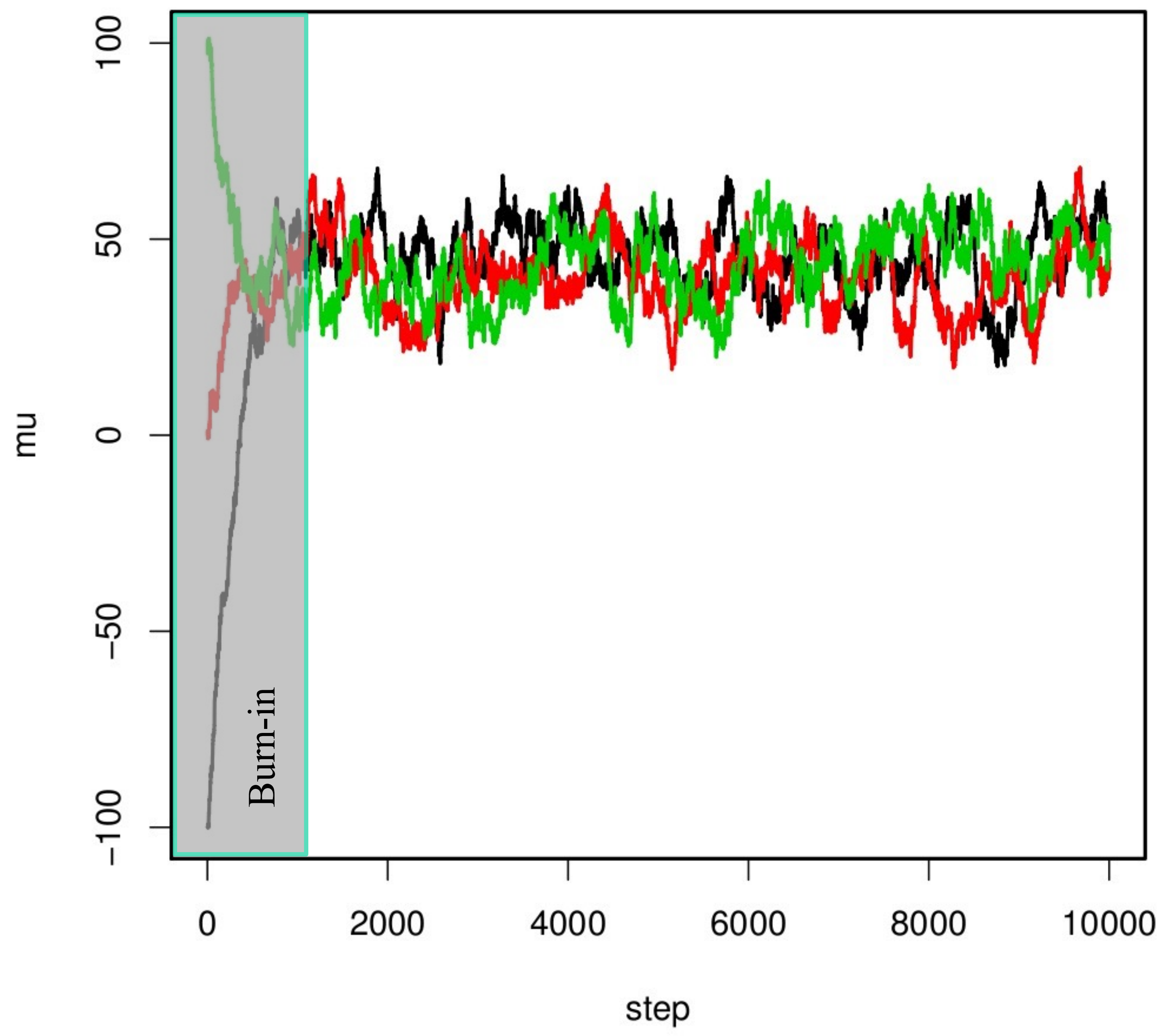
mu
100
50
0
−50
−100
step
2
4
6
8
10

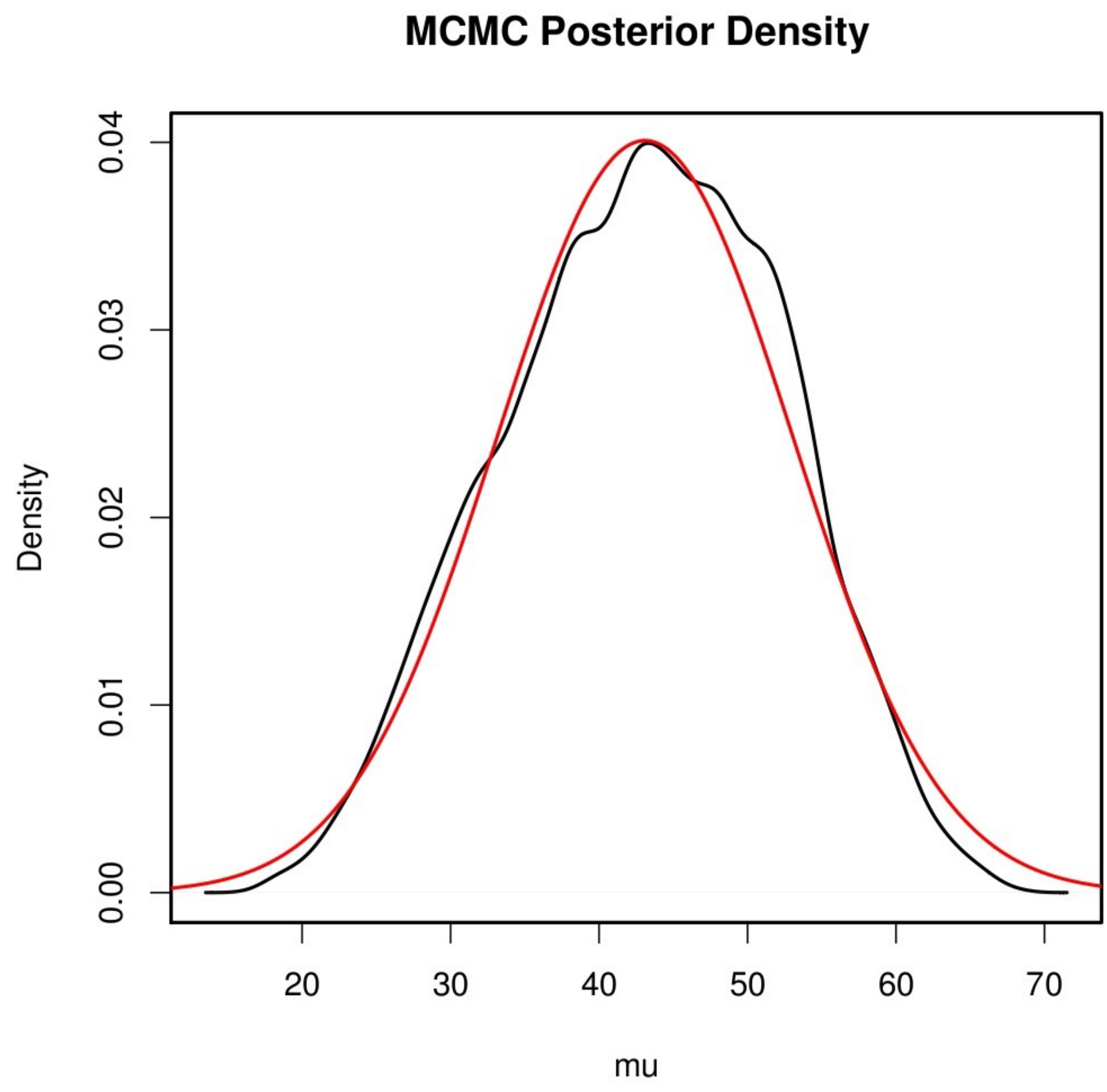


100
50
0
−50
−100
mu
0
20
40
60
80
100
step


mu
100
50
0
−50
−100
0
200
400
600
800
1000
step


100
50
0
−50
−100
mu
0
2000
4000
6000
8000
10000
step


mu
100
50
0
−50
−100
Burn-in
0
2000
4000
6000
8000
10000
step


MCMC Posterior Density
Density
mu
0.00
0.01
0.02
0.03
0.04
20
30
40
50
60
70

- Advantages
  - Multi-dimensional, joint
  - Simple
  - Robust
- Disadvantages
  - Sequential samples not independent
  - Computationally intensive
  - Discard “Burn – in” period before convergence
  - Assessing convergence

Priors

- Makes it possible to calculate a posterior <u>density of the model parameter</u> rather than the likelihood of the data
- Provides a way of incorporating information that is external to the data set(s) at hand
- Inherently sequential

Previous Posterior = New Prior

Where do Priors come from?

- Uninformative / vague
  - Chosen to have minimal information content, allows the likelihood to dominate the analysis
- Previous analyses
  - Must be equivalent
  - Variance inflation
- “The literature”
  - Meta-analysis
- Expert knowledge

Where do Priors come from?

- Uninformative / vague
  - Ch nt, allows the
- Prev
  - Mu
  - Va
- "The
  - Me
- Expert knowledge

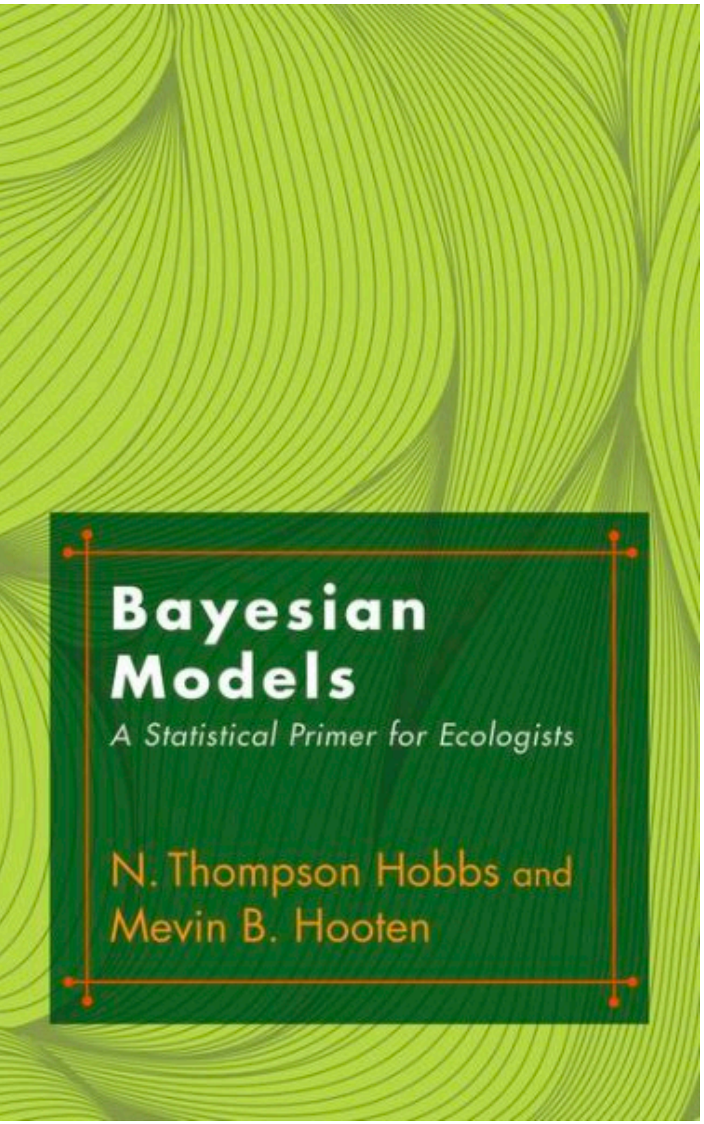
Prior specification must be "blind" to the data in the analysis!!

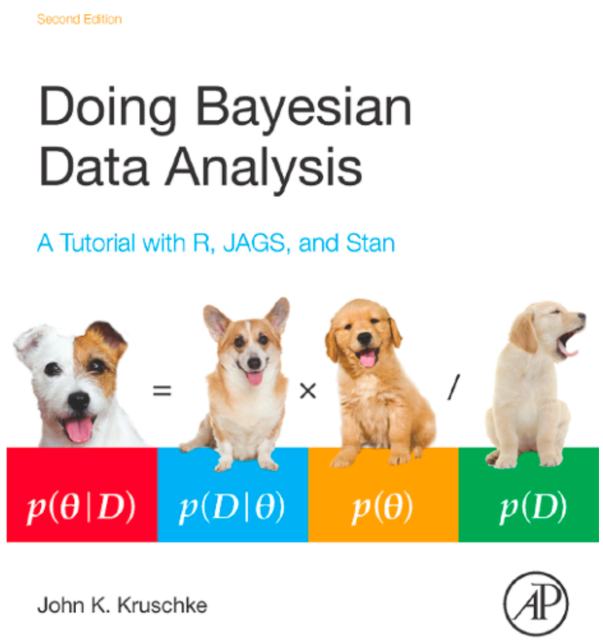
No "double dipping" -- leads to falsely overconfident results

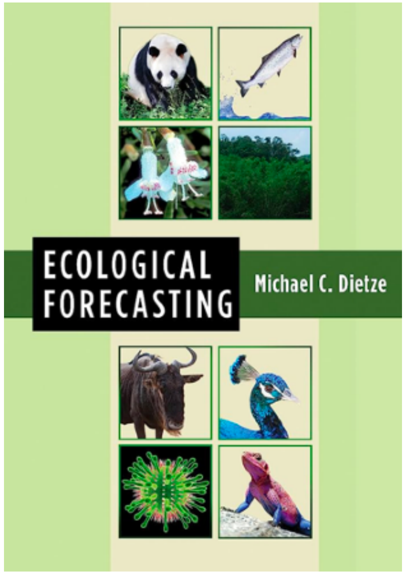
How do I choose a prior PDF?

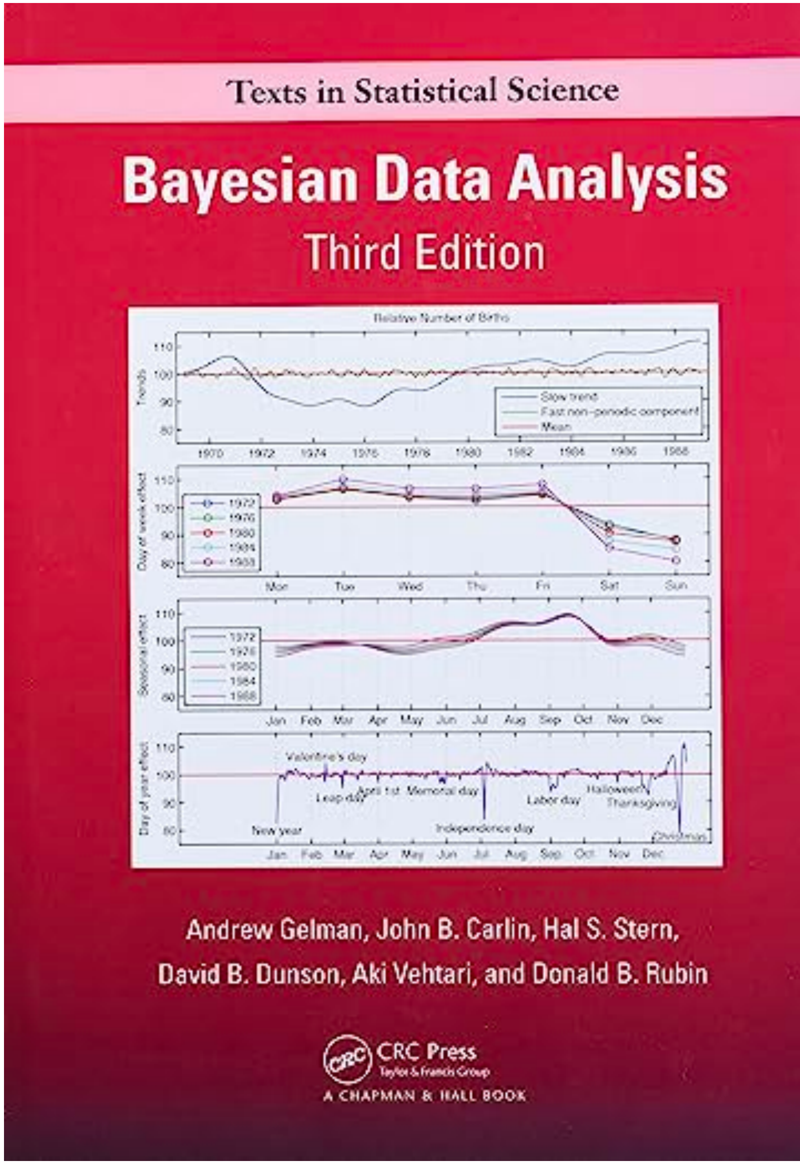
- Analogous to how we choose the data model
  - Range restrictions, shape, etc.
- Conjugacy
  - A prior is conjugate to the likelihood if the posterior PDF is in the same family as the prior
  - Allow for closed-form analytical solutions to either full posterior or (in multiparameter models) for the conditional distribution of that parameter.
  - Modern computational methods no longer require conjugacy

Resources


Bayesian
Models
A Statistical Primer for Ecologists
N. Thompson Hobbs and
Mevin B. Hooten


Second Edition
Doing Bayesian
Data Analysis
A Tutorial with R, JAGS, and Stan
=
×
/
$p(\theta|D)$
$p(D|\theta)$
$p(\theta)$
$p(D)$
John K. Kruschke


ECOLOGICAL
FORECASTING
Michael C. Dietze


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
Taylor & Francis Group
A CHAPMAN & HALL BOOK