University of Nevada, Reno

Lecture 10: Bayes' Theorem I

Perry Williams, PhD

NRES 779

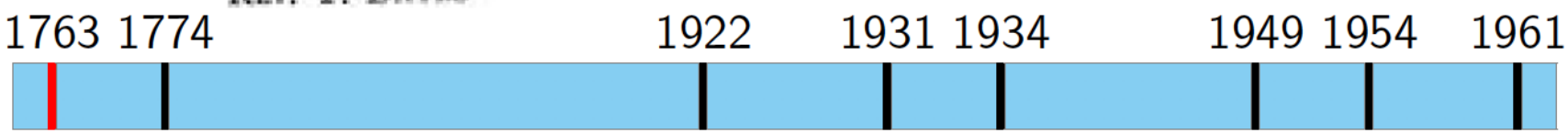
Bayesian Hierarchical Modeling in Natural Resources

History

Rev. T. Bayes

Bayes' Theorem appeared in "*An Essay Towards Solving a Problem in the Doctrine of Chances*"

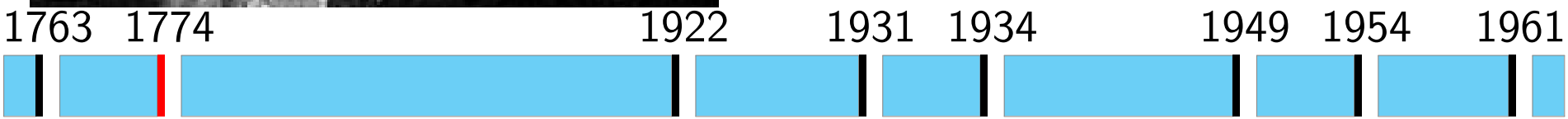
"Aldrich suggests that we interpret [Bayes' definition of probability] in terms of expected utility, and thus that Bayes' result would make sense only to the extent to which one can bet on its observable consequences."
-Stephen Fienberg, 2006.



History

Laplace published "Memoire sur la Probabilité des Causes par les Évènements"

- Elaborate example of inverse probability
- Uniform prior distributions
- Methods for choosing estimators that minimize posterior loss



The Theory of Inverse Probability

$$P(C|E) = \frac{P(E|C)P_{\text{prior}}(C)}{\sum(E|C'')P_{\text{prior}}(C')}$$

"I propose to determine the probability of the causes of events, a question which has not been given due consideration before, but which deserves even more to be studied, for it is principally from this point of view that the science of chances can be useful in civil life."

Originally publishes as "Mémoire sur la probabilité des causes par les évènemens"

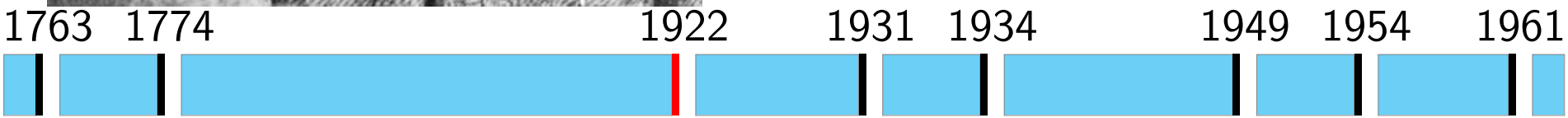
"Laplace's principle being dead, it should be decently buried out of sight, and not embalmed in text-books and examination papers...The indiscretion of great men should be quietly allowed to be forgotten."

George Chriystal 1891

History

Fisher published "On the Mathematical Foundations of Theoretical Statistics"

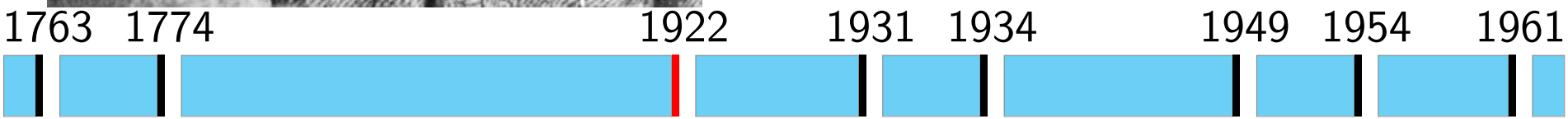
- Rejected inverse probability
- Grounded his theory on frequency interpretation of probability
- Obviated the need for prior distributions
- Introduced likelihood
- Tests of significance



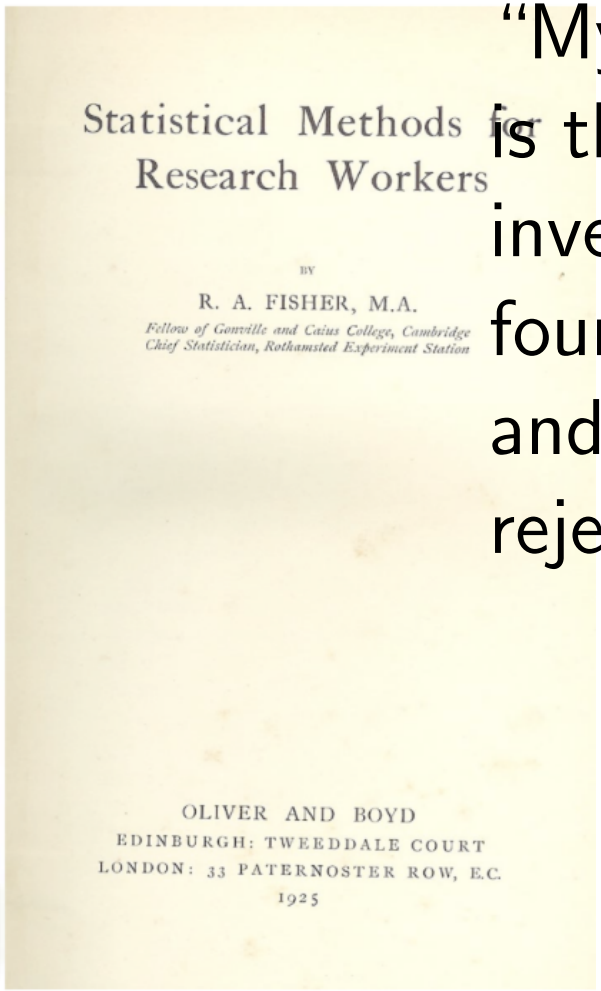
History

Fisher on Probability and Decisions

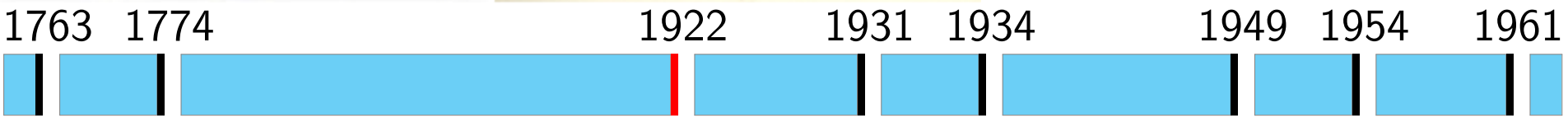
> "We aim, in fact, at methods of inference which should be equally convincing to all rational minds, irrespective of any intentions they may have in utilizing the knowledge inferred."



History



"My personal conviction is that the theory of inverse probability is founded upon an error and must be wholly rejected"



History

1763 1774 1922 1931 1934 1949 1954 1961

"There has not been a single date in the history of the law of gravitation when a modern significance test would not have rejected the law outright"

Harold Jeffreys

"The p-value is almost nothing sensible you can think of. I tell students to give up trying"

Stephen Goodman

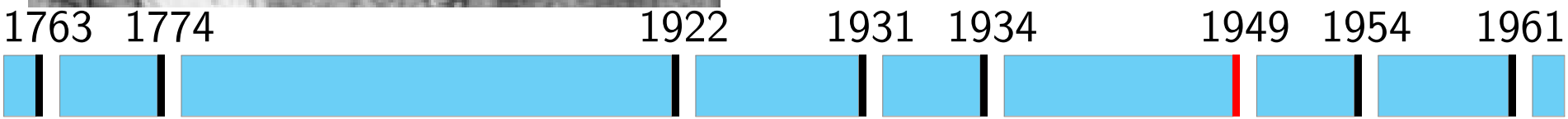
What is the collective noun for a group of statisticians?


ENIGMA

History

"It is well recognized that the statistical estimation theory should and can be organized within the framework of the theory of statistical decision functions (Wald 1950)"

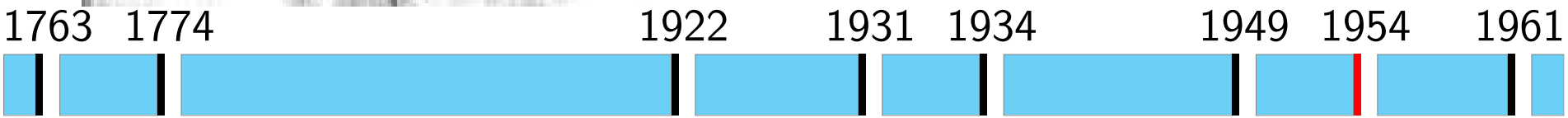
Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle.



History

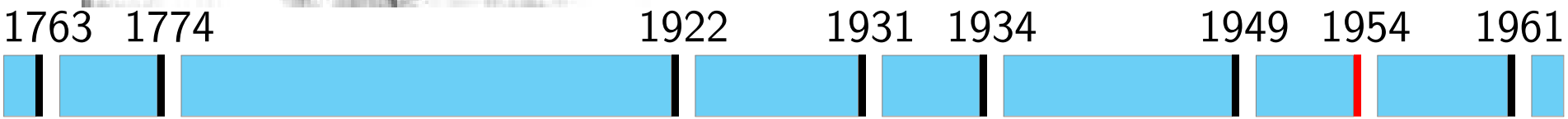
Savage published "The Foundations of Statistics"

- Set the stage for Bayesian revival

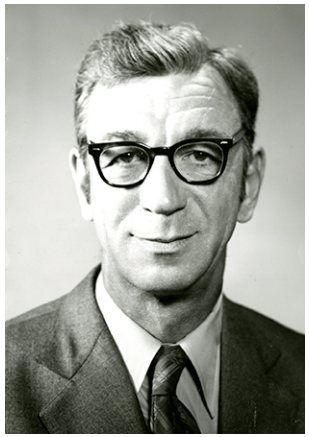


History

"Decision theory is the best and most stimulating, if not the only, systematic model of statistics."

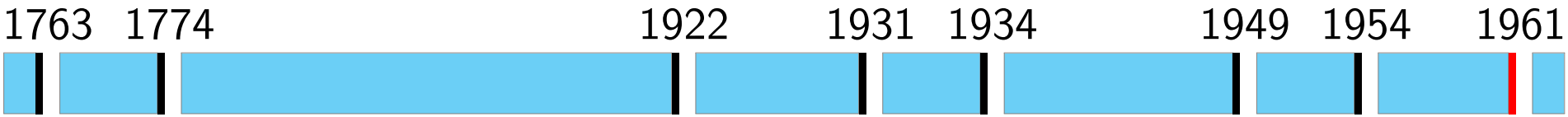


History



Raiffa and Schlaifer published "Applied Statistical Decision Theory"

- "Methods of Fisher, Neyman, and Pearson did not address the main problem of a business[person]: how to make decisions under uncertainty"
- Developed Bayesian decision theory



- F.P. Ramsey
- B. De Finetti
- J.M. Keynes
- H. Jeffreys
- D.V. Lindley
- D.R. Cox
- J.W. Tukey
- A. Birnbaum
- M. Kendall

Second Bayesian Revival

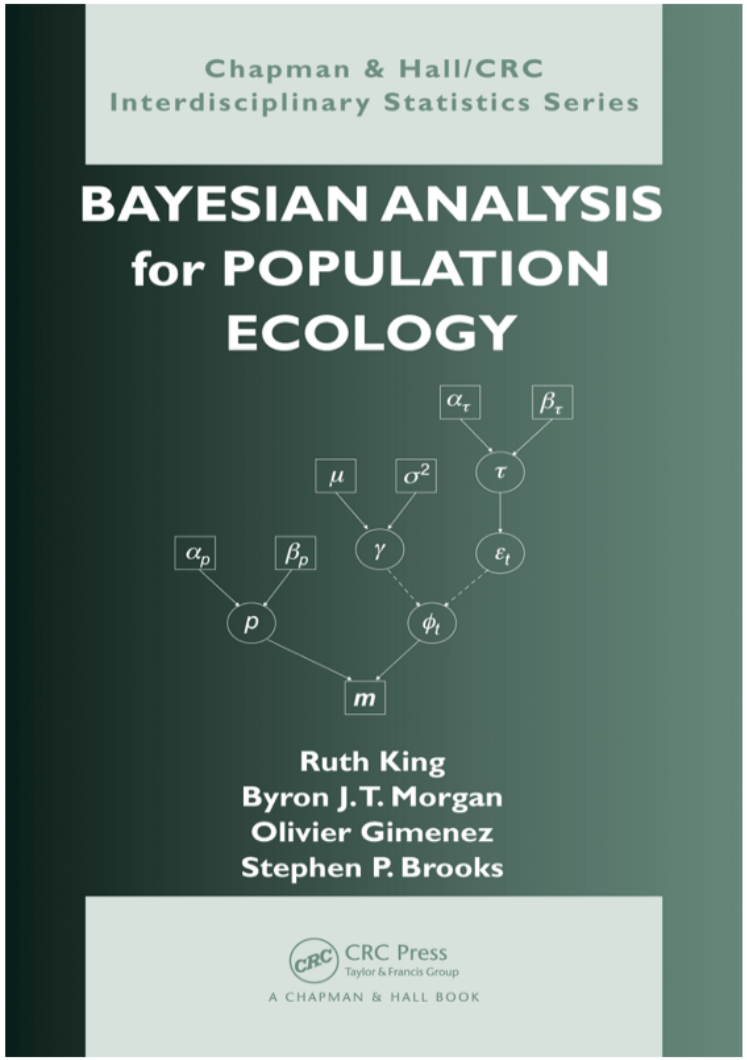
Gelfand, A. E., and A. F. M. Smith. 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 85:398-409.

Second Bayesian Revival

Dr. Ruth King, Thomas Bayes' Chair of Statistics, University of Edinburgh

Second Bayesian Revival

Dr. Ruth King, Thomas Bayes' Chair of Statistics, University of Edinburgh



Second Bayesian Revival

Dr. Jennifer Hoeting, Bayesian Model Averaging, Professor of Statistics, Colorado State University

Second Bayesian Revival

Dr. Monserrat Fuentes,

- President, St. Edwards University
- Provost and Professor of Statistics, University of Iowa
- Editor, JABES
- Director, STATMOS

Second Bayesian Revival

Dr. Staci Hepler, Bayesian modeling, spatial-temporal statistics, epidemiology, Associate Professor of Statistics, Wake Forest University

Second Bayesian Revival

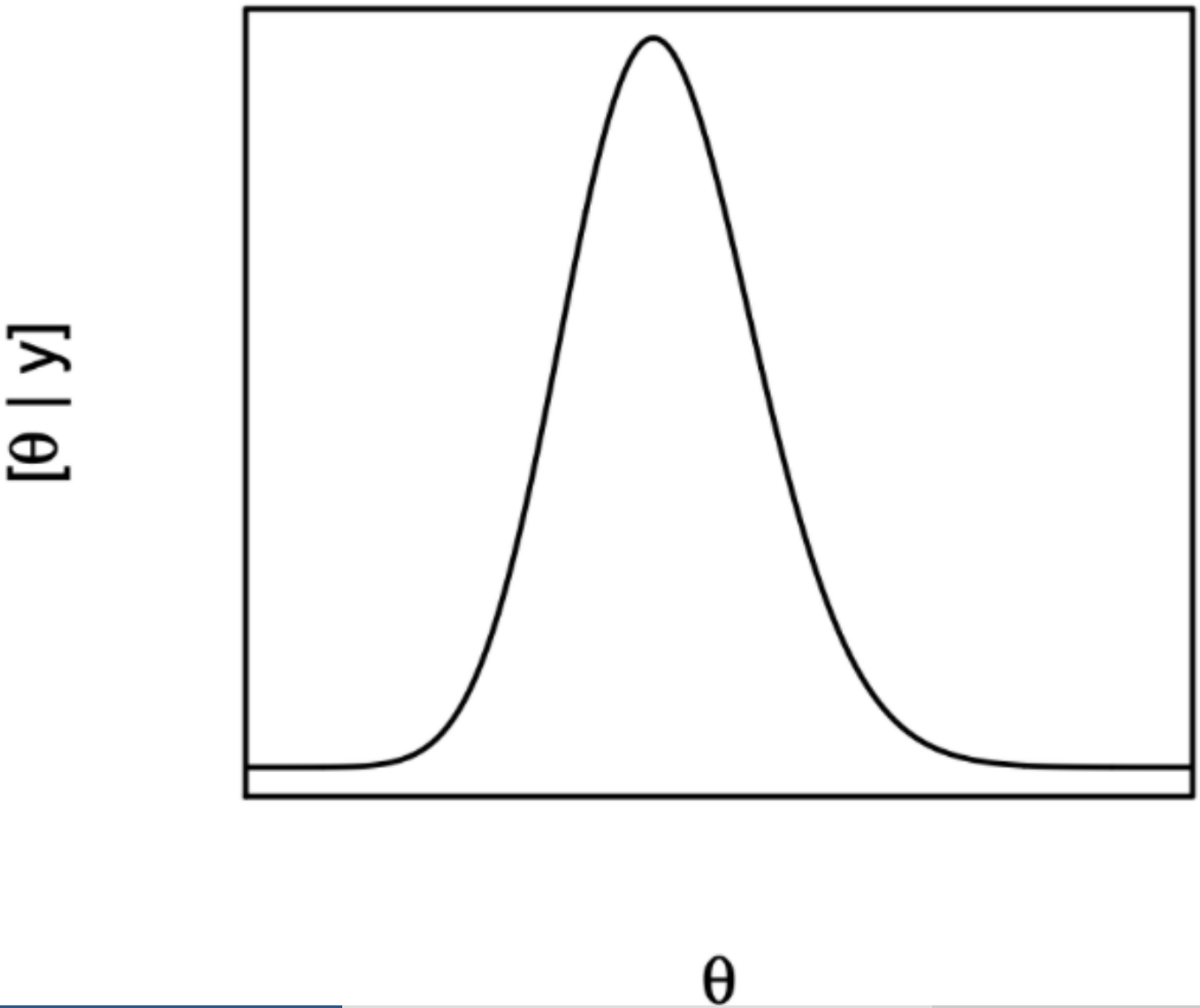
Dr. Erin Schliep, Bayesian statistics, spatial statistics, environmental statistics, Associate Professor of Statistics, North Carolina University

Second Bayesian Revival

Dr. Elizabeth Mannshardt-Hawk, Bayesian statistics, government executive, environmental statistics, Associate Director of the information Access and Analytic Services Division at U.S. EPA.

Bayesian Inference

All unobserved quantities are treated as random variables



Random Variables

All unobserved quantities are treated in exactly the same way

- Model parameters
- Latent states
- Missing data
- Predictions and forecasts
- Observations (before they are observed)

Exercise

- Assume we have two, jointly distributed random variables, θ and y. The random variable θ represents unobserved quantities of interest. The random variable y represents observations, which become fixed *after* they are observed.
- Derive Bayes' Theorem

$$[\boldsymbol{\theta}|y] = \frac{[y|\boldsymbol{\theta}][\boldsymbol{\theta}]}{[y]}$$

using your knowledge of the laws of probability, particularly the definition of conditional probability.

Derivation

Recall the definition of conditional probability

$$[\theta|y] = \frac{[\theta, y]}{[y]} \tag{1}$$

$$[y|\theta] = \frac{[\theta, y]}{[\theta]} \tag{2}$$

solving (2) for $[\theta, y]$

$$[\theta, y] = [y|\theta][\theta] \tag{3}$$

Substituting the right hand side of (3) for $[\theta, y]$ in (1) we obtain Bayes' Theorem

$$[\theta|y] = \frac{[y|\theta][\theta]}{[y]}$$

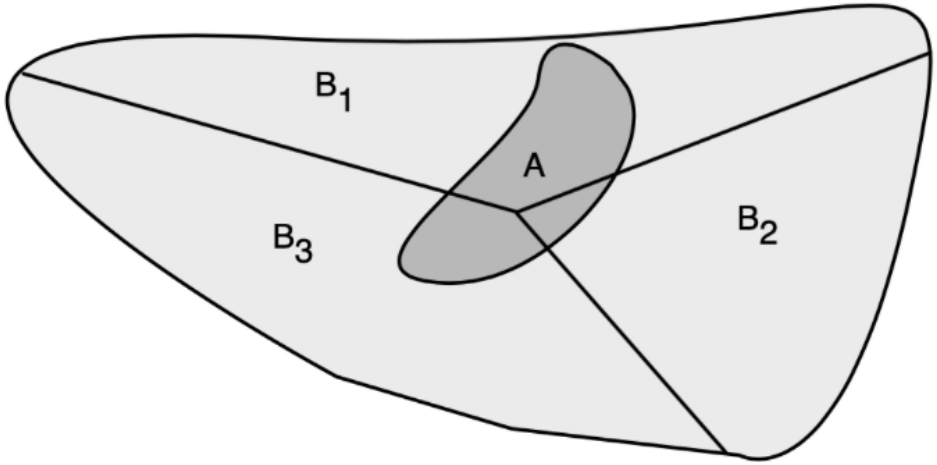
We will often make use of the equivalent equation

$$[\boldsymbol{\theta}|y] = \frac{[y, \boldsymbol{\theta}]}{[y]}$$

as a starting point for developing hierarchical models by factoring $[y, \boldsymbol{\theta}]$ into ecologically sensible components (*sensu* Berliner 1996) that can be treated in MCMC as univariate distributions. More about that soon.

What is $[y]$?

Recall the law of total probability for discrete random variables



$$[A] = \sum_i [A \mid B_i]\,[B_i]\,.$$

and for continuous random variables

$$[A] = \int_B [A|B]\,[B]\,dB.$$

What is $[y]$?

It follows that

$$[y] = \sum_{\theta_i \in \{\Theta\}} [y|\theta_i]\,[\theta_i] \text{ for discrete parameters}$$

$$[y] = \int_{\theta} [y|\theta]\,[\theta]\,d\theta \text{ for continuous parameters.}$$

Thus, Bayes theorem for discrete valued parameters is

$$[\theta|y] = \frac{[y|\theta_i]\,[\theta_i]}{\sum_{\theta_i \in \{\Theta\}} [y|\theta_i]\,[\theta_i]}$$

and for parameters that are continuous,

$$[\theta|y] = \frac{[y|\theta]\,[\theta]}{\int_{\theta} [y|\theta]\,[\theta]\,d\theta}.$$

More about $[y]$

- $[y]$ is the marginal distribution of the data, a *distribution* before the data are observed and a *normalizing constant* after the data are observed.
- It is also called the *prior predictive distribution*.
- Because $[y]$ is a constant after the data are observed,

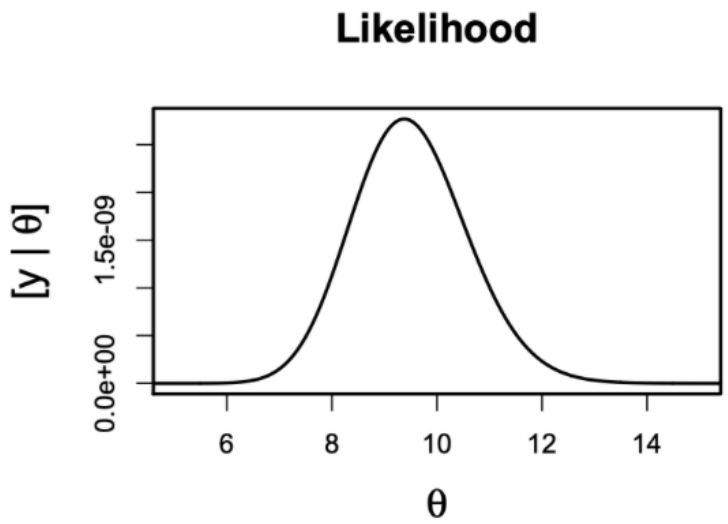
$$\begin{aligned}[\boldsymbol{\theta}|y] &\propto [y, \boldsymbol{\theta}] \\ &\propto [y|\boldsymbol{\theta}][\boldsymbol{\theta}]\end{aligned}$$

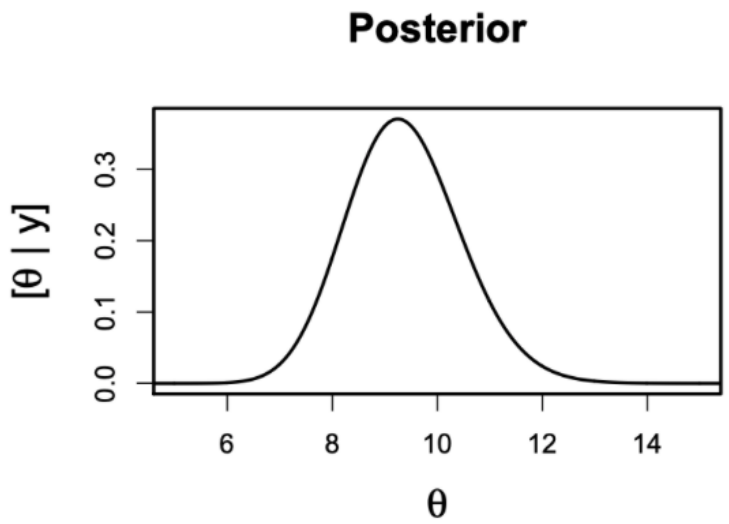
$[y]$ is critical to Bayes

$$\mathbf{y} = (5,\ 10,\ 11,\ 12,\ 14,\ 9,\ 8,\ 6)'$$

$$\text{likelihood} = \prod_{i=1}^{8} \text{Poisson}(y_i|\theta)$$

$$\text{posterior} = \frac{\prod_{i=1}^{8} \text{Poisson}(y_i|\theta)\text{gamma}(\theta|.0001,.0001)}{[y]}$$





Probability Mass Function $[y|\theta]$

θ is *known* to be $\frac{1}{2}$. Probability of number of whites conditional on three draws and $\theta = \frac{1}{2}$:

$y =$ Number of whites	$[y\|\theta]$
0	.125
1	.375
2	.375
3	.125
$\sum_{i=1}^{4} [y\|\theta_i] =$	1

Likelihood $[y|\boldsymbol{\theta}]$

Probability of two whites on three draws conditional on θ_i

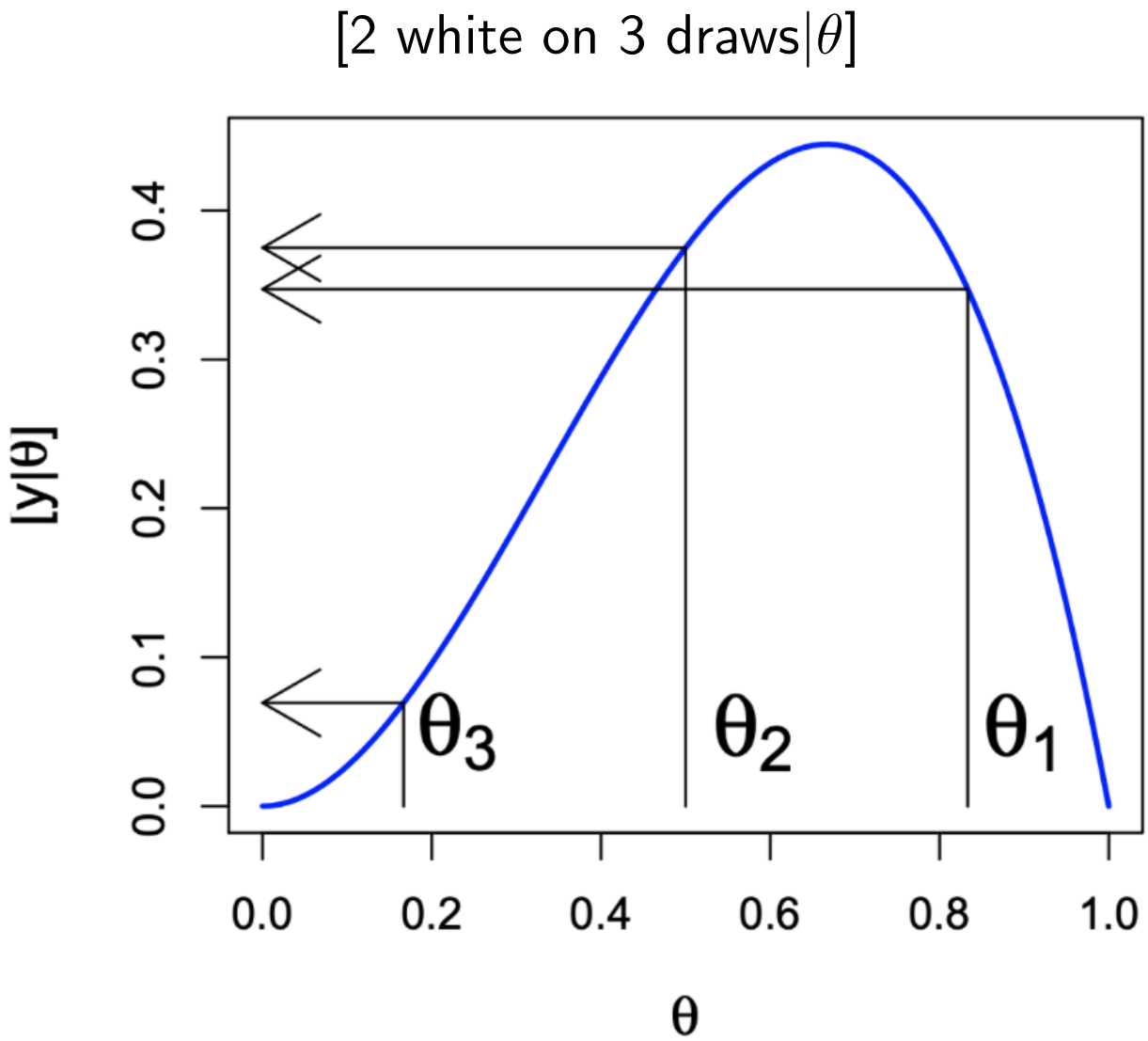
Parameter	Likelihood $[y\|\boldsymbol{\theta}_i]$
$\boldsymbol{\theta}_1 = 5/6$	.347
$\boldsymbol{\theta}_2 = 1/2$	.375
$\boldsymbol{\theta}_3 = 1/6$	.069
$\sum_{i=1}^{3} [y\|\boldsymbol{\theta}_i] =$	.791

Posterior Distribution $[\theta|y]$

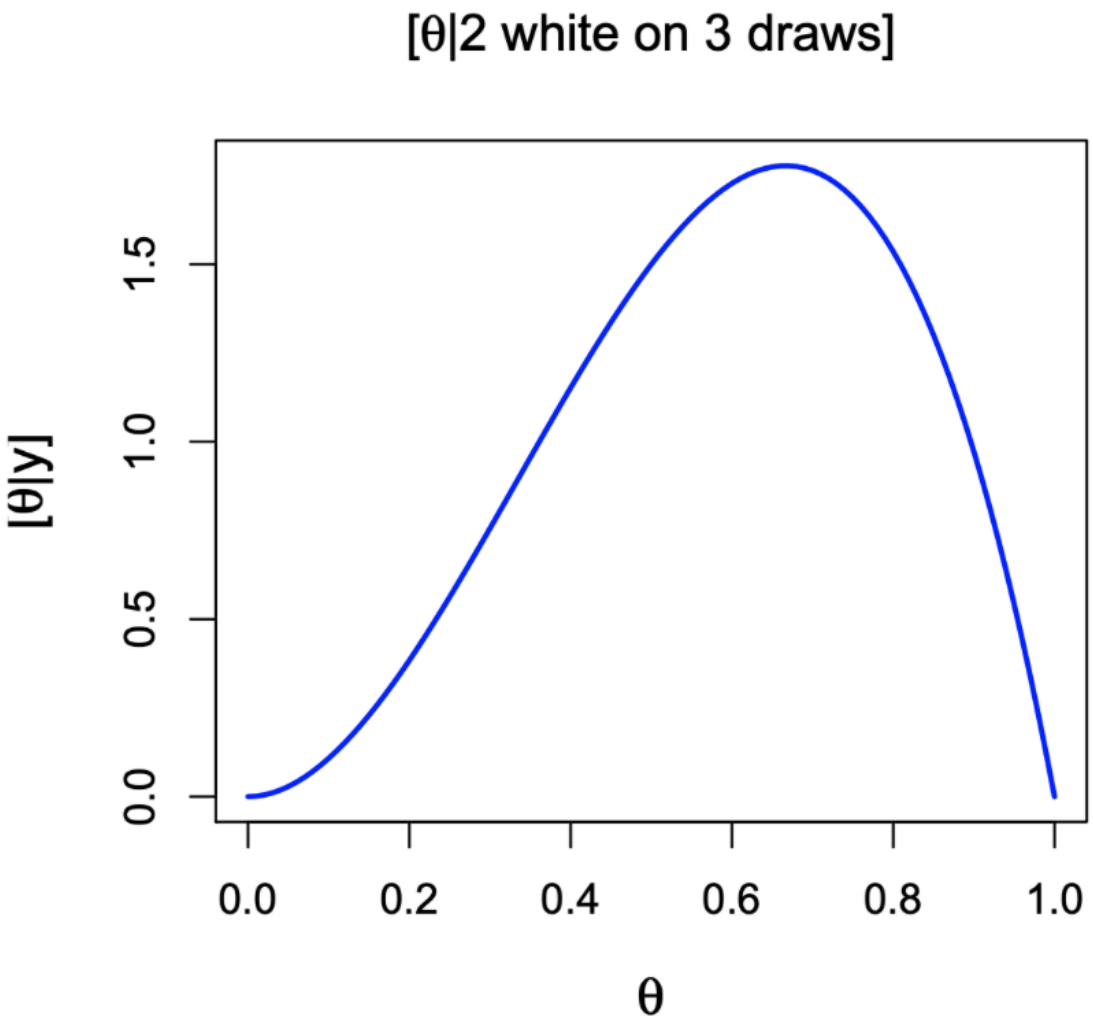
Probability of θ_i conditional on two whites on three draws

Parameter	Prior $[\theta_i]$	Likelihood $[y\|\theta_i]$	Joint $[y\|\theta_i]\,[\theta_i]$	Posterior $\frac{[y\|\theta_i][\theta_i]}{[y]} = [\theta_i\|y]$
θ_1	0.333	0.347	0.115	0.439
θ_2	0.333	0.375	0.125	0.474
θ_3	0.333	0.069	0.023	0.087
		$[y] = \sum_{i=1}^{3} [y\|\theta_i]\,[\theta_i] =$	0.261	$\sum_{i=1}^{3} [\theta_i\|y] = 1$

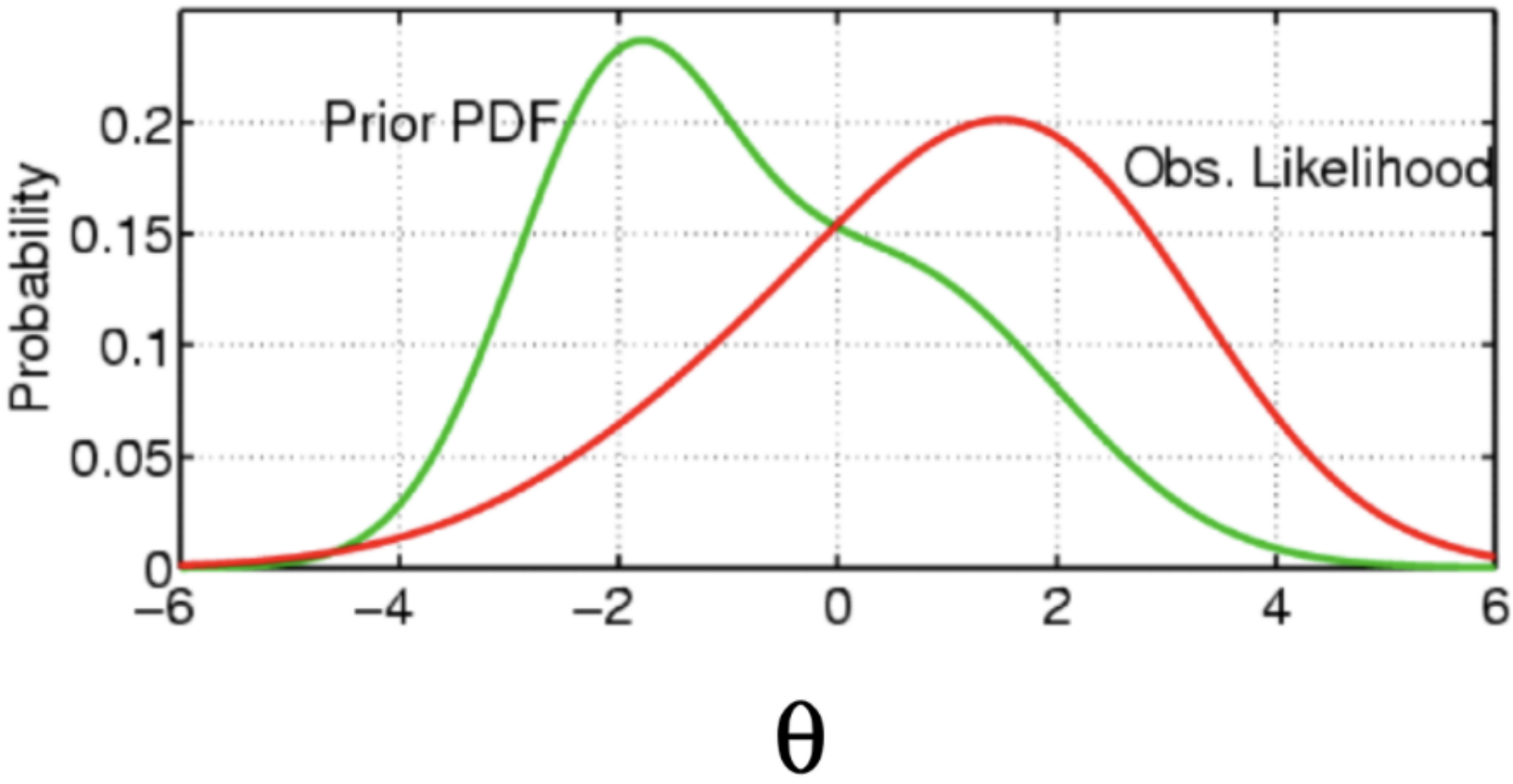
Likelihood Profile $[y|\theta]$



Posterior Distribution $[\theta|y]$



The Components of Bayes' Theorem

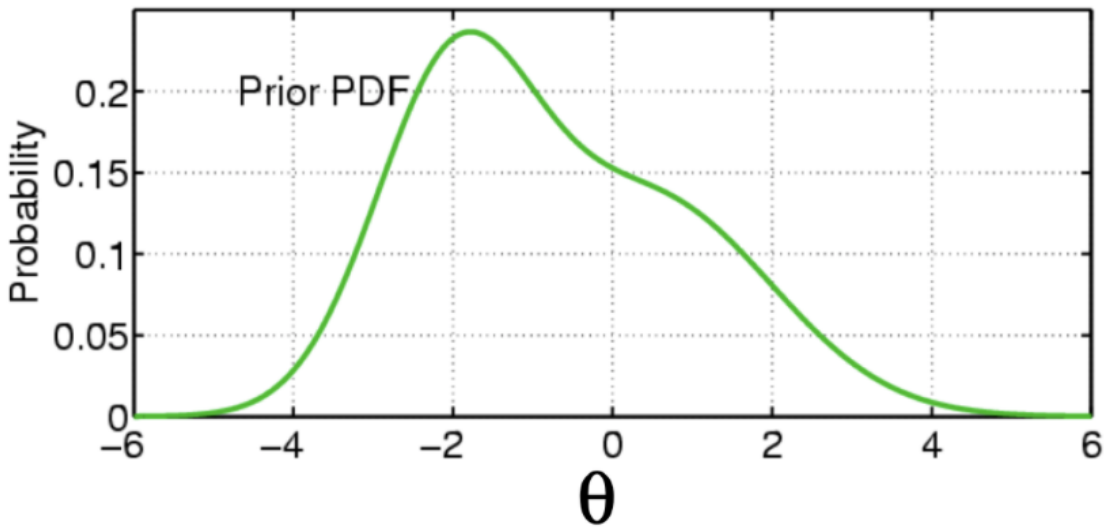


Courtesy of Chris Wikle, University of Missouri

The Prior

$$[\theta|y] = \frac{[y|\theta]\,[\theta]}{\int_\theta [y|\theta]\,[\theta]\,d\theta}$$

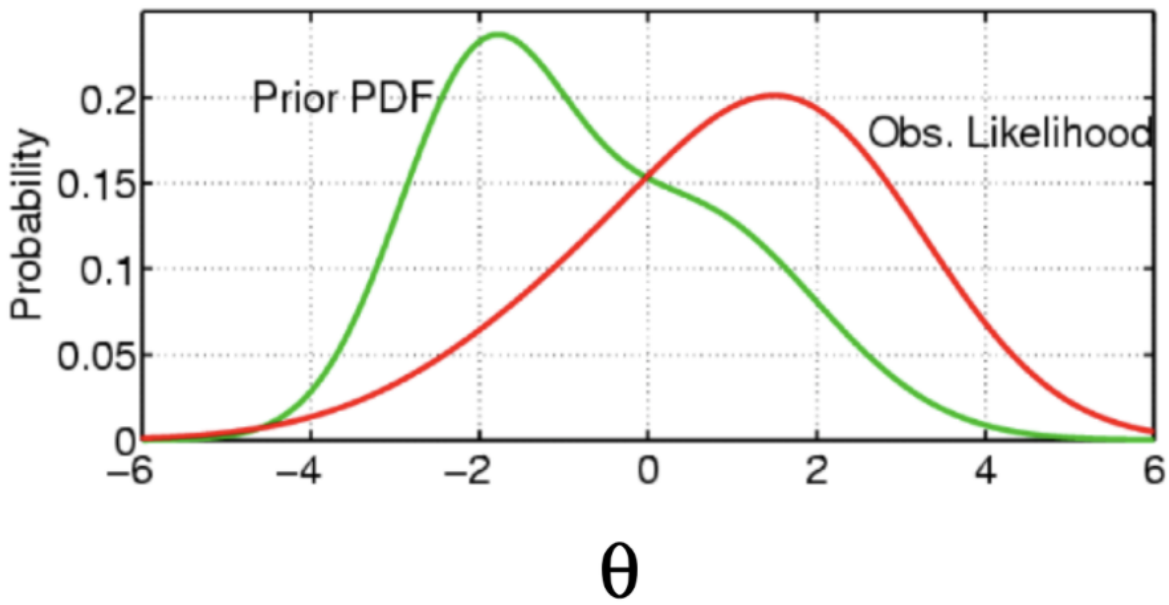
The prior, $[\theta]$, can be informative or vague.



The Likelihood

$$[\theta|y] = \frac{[y|\theta][\theta]}{\int_\theta [y|\theta][\theta]\, d\theta}$$

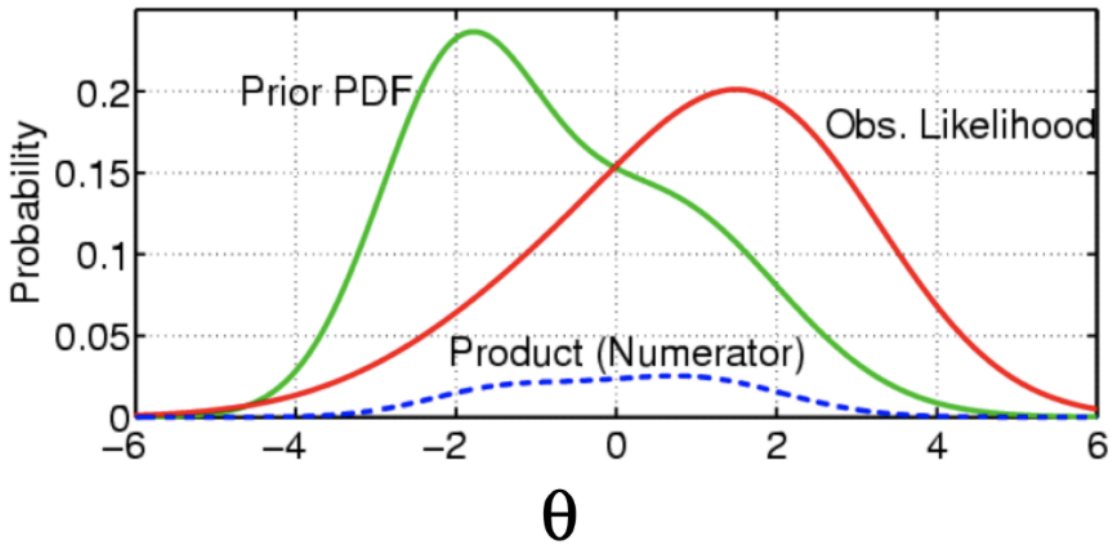
The likelihood (a.k.a. data distribution, $[y|\theta]$)



The Joint Distribution

$$[\theta|y] = \frac{[y,\theta]}{[y]} = \frac{[y|\theta][\theta]}{\int_\theta [y|\theta][\theta]\, d\theta}$$

The product of the prior and the likelihood, $[y|\theta][\theta]$, the joint distribution of the parameters and the data, $[y,\theta]$.

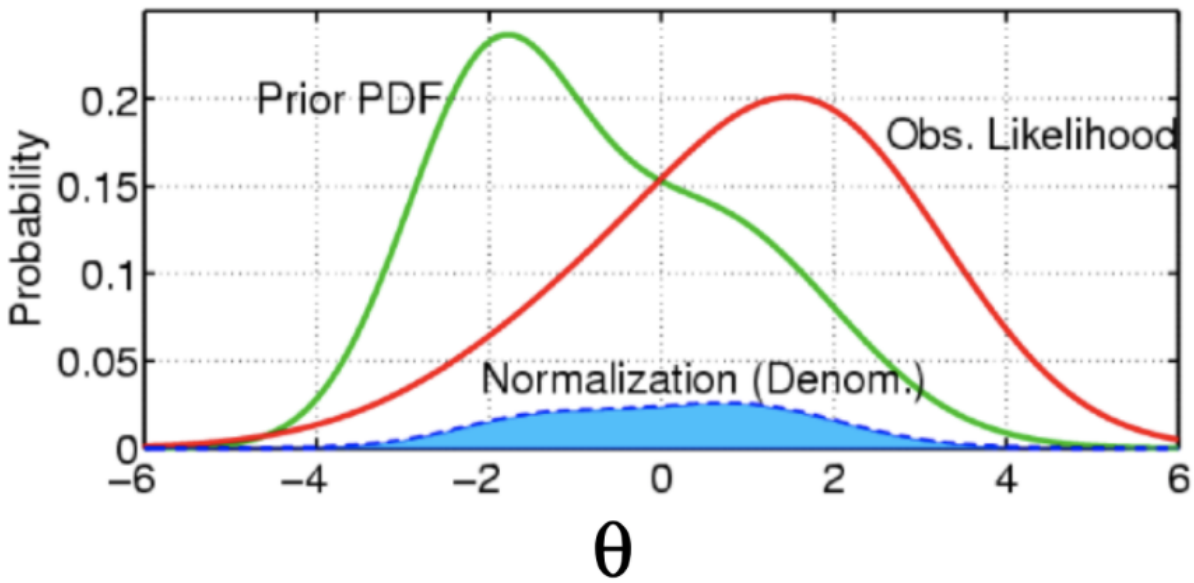


What is the maximum likelihood estimate of θ?

The Marginal Distribution of the Data

$$[\theta|y] = \frac{[y|\theta][\theta]}{\int_\theta [y|\theta][\theta]\, d\theta}$$

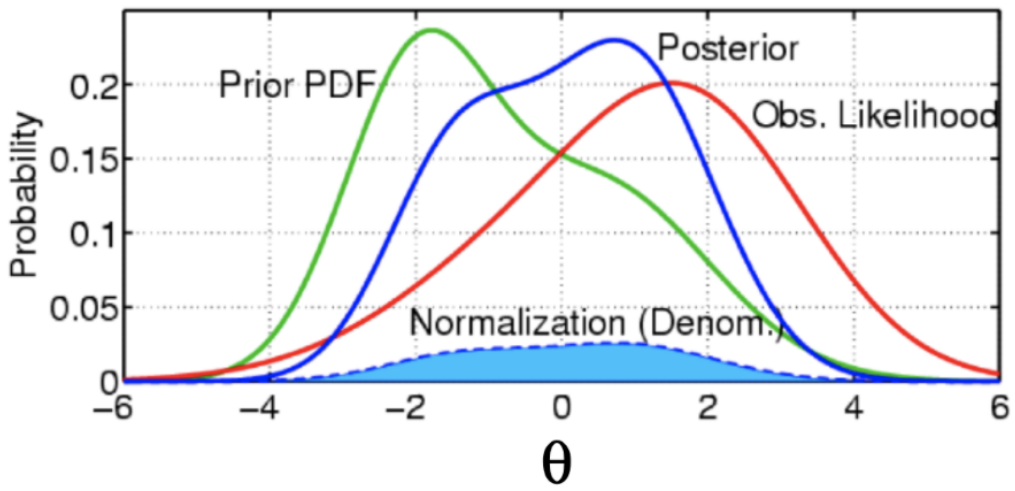
The marginal distribution of the data (the denominator) is the area under the joint distribution



The Posterior Distribution

What we are seeking: The posterior distribution, $[\theta|y]$.

$$[\theta|y] = \frac{[y|\theta][\theta]}{\int_\theta [y|\theta][\theta]\, d\theta}$$



Note that we are dividing each point on the dashed line by the area under the dashed line to obtain a probability density function reflecting our prior and current knowledge about θ

So what?

What does this enable you to do? Review factoring joint distributions: Remember from the basic laws of probability that:

$$p(z_1, z_2) = p(z_1|z_2)p(z_2) = p(z_2|z_1)p(z_1)$$

This generalizes to:

$$\mathbf{z} = (z_1, z_2, \ldots, z_n)$$

$$p(z_1, z_2, \ldots, z_n) = p(z_n|z_{n-1}, \ldots, z_1) \times \ldots \times p(z_3|z_2, z_1)p(z_1)$$

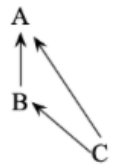
where the components z_i may be scalars or subvectors of $\mathbf{z}$ and the sequence of their conditioning is arbitrary. This equation can be simplified using knowledge of independence.

So What?

I

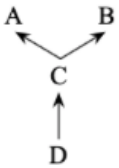
$$\Pr(A, B) = \Pr(A|B)\Pr(B)$$

II



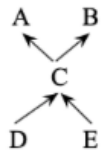
$$\Pr(A, B, C) = \Pr(A|B, C) \times \Pr(B|C)\Pr(C)$$

III



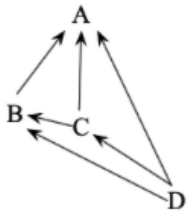
$$\Pr(A, B, C, D) = \Pr(A|C) \times \Pr(B|C)\Pr(C|D)\Pr(D)$$

IV



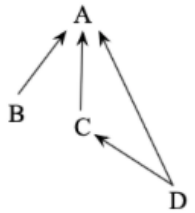
$$\Pr(A, B, C, D, E) = \Pr(A|C) \times \Pr(B|C)\Pr(C|D, E) \times \Pr(D)\Pr(E)$$

V



$$\Pr(A, B, C, D) = \Pr(A|B, C, D) \times \Pr(B|C, D) \times \Pr(C|D)\Pr(D)$$

VI



$$\Pr(A, B, C, D) = \Pr(A|B, C, D) \times \Pr(C|D) \times \Pr(B)Pr(D)$$

So What?

$$\overbrace{[\theta|y]}^{\text{Posterior}} = \frac{[y,\theta]}{[y]} = \frac{\overbrace{[y|\theta]}^{\text{likelihood}}\ \overbrace{[\theta]}^{\text{prior}}}{\underbrace{\int_\theta [y|\theta]\,[\theta]\,d\theta}_{\text{marginal}}} \tag{22}$$

Useful models will be more complex:

$$\underbrace{[\theta_1, \theta_2, \theta_3, \ldots, \theta_n, \mathbf{z}_1, \mathbf{z}_2 .. \mathbf{z}_n | \mathbf{y}_1, \mathbf{y}_2]}_{\text{multiple parameters, latent states, data sets}} \propto \underbrace{[\theta_1, \theta_2, \theta_3, \ldots, \theta_n, \mathbf{z}_1, \mathbf{z}_2 .. \mathbf{z}_n, \mathbf{y}_1, \mathbf{y}_2]}_{\text{factor into conidtional distributions}}$$

We use the rules of probability to factor complex joint distributions into a series of conditional distributions. We can then use the Markov chain Monte Carlo algorithm to escape the need for integrating the marginal data distribution, allowing us to find the marginal posterior distributions of all of the unobserved quantities. Which, of course, is where we started out. And where we are headed.