UCLA Health | Jonsson Comprehensive Cancer Center

# Bayesian Statistics

Jaron Arbet

# Outline

- Bayes Theorem
- 3 Key steps of Bayesian Analysis
- Bayes vs Frequentist stats

# Bayes' Theorem

Thomas Bayes invented "Bayes' Theorem" (1763)

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

- $p(A|B)$: probability of A given that B is true
- $p(B|A)$: probability of B given that A is true
- $p(A), p(B)$: probability that A (B) occurs



https://en.wikipedia.org/wiki/Thomas_Bayes

# Example: Disease testing


- A = disease/healthy, B = +/- test
- Prevalence: $\Pr(A = disease)$ = 0.02
- Sensitivity = $\Pr(B = +\mid A = disease)$ = 0.99
- False positive rate = $\Pr(B = +\mid A = healthy)$ = 0.10

$$\boldsymbol{Pr(A = disease \mid B = +)} = \frac{\Pr(A = disease) * \Pr(B = + | A = disease)}{\sum_i \Pr(A = a_i) * \Pr(B = + | A = a_i)}$$

Summing over all $a_i$ gives $p(B)$

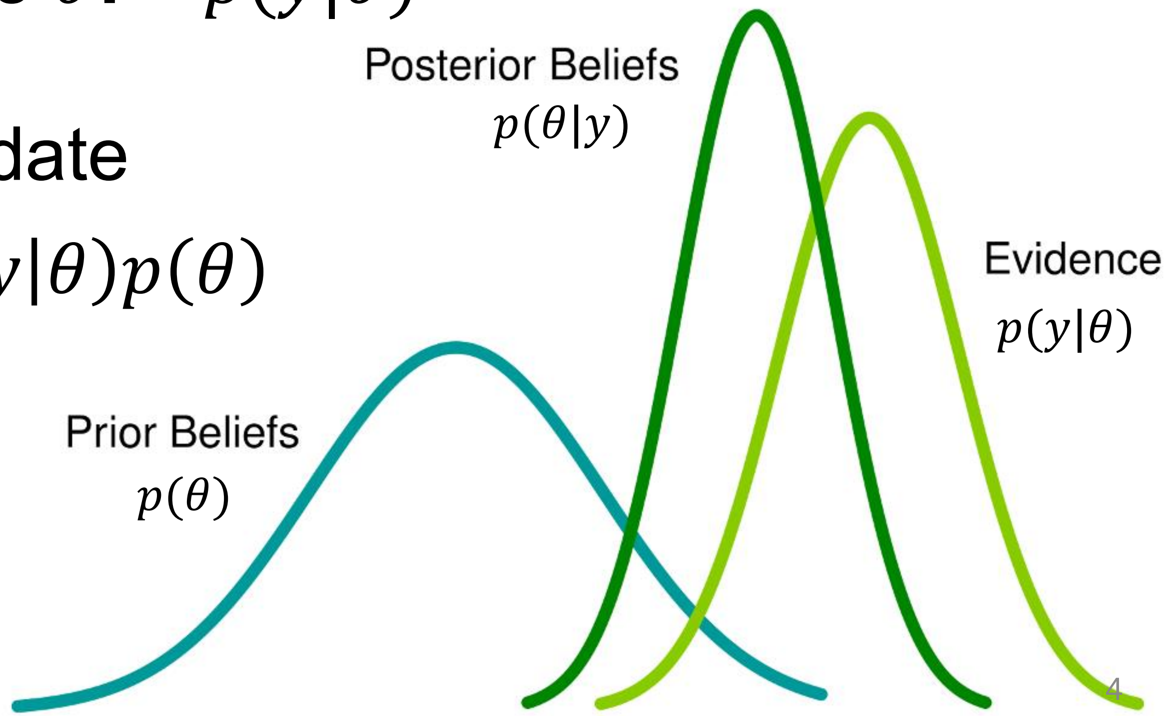
$$= \frac{\Pr(A = disease) * \Pr(B = + | A = disease)}{\Pr(A = disease) * \Pr(B = + | A = disease) + \Pr(A = healthy) * \Pr(B = + | A = healthy)}$$

$$= \frac{0.02 * 0.99}{0.02 * 0.99 + 0.98 * 0.1} = \mathbf{16.8\%}$$

# 3 Key steps of Bayesian Analysis

Unknown population parameter “$\theta$” (*e.g.* mean, proportion, correlation)

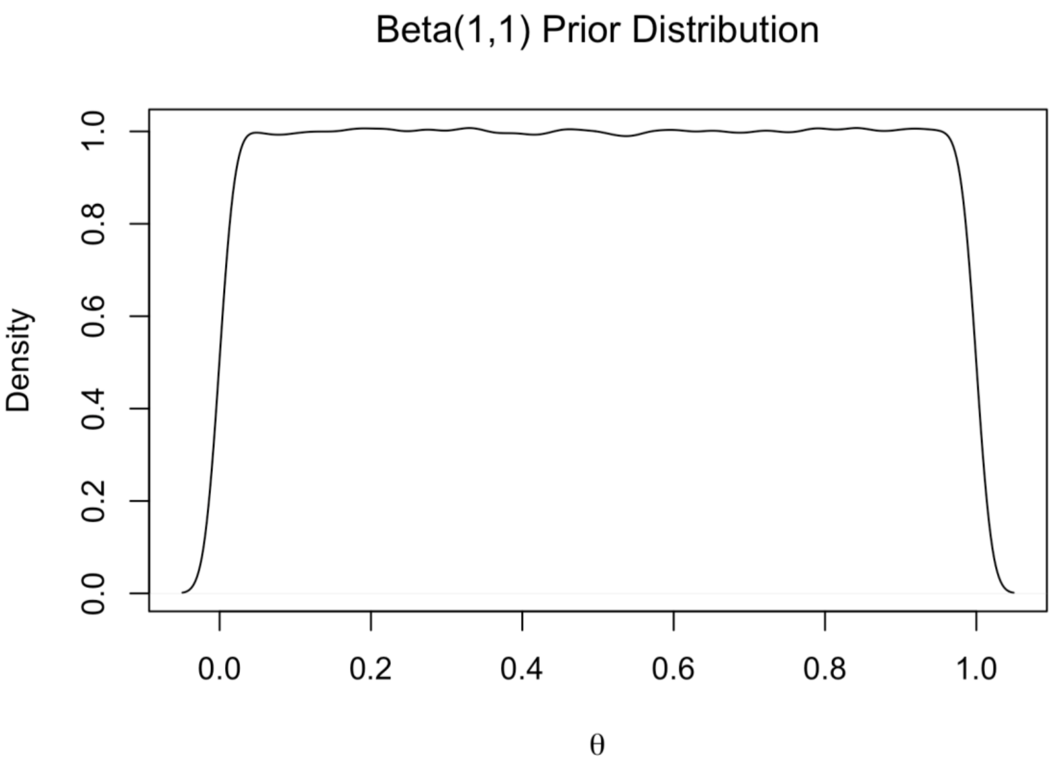
1. **Prior** beliefs about $\theta$: “$p(\theta)$”
2. **Model** (“likelihood”) links new data $y$ to $\theta$: “$p(y|\theta)$”
3. **Posterior**: use Bayes’ Theorem to update knowledge: $p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \propto p(y|\theta)p(\theta)$
   - Weighted avg. of prior knowledge and new data



# Proportion Example

$\theta$ = Kobe Bryant's field goal percentage 1999-2000

**Step 1**: $p(\theta) = Beta(a,b) = \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}$





**"Flat, non-informative prior"**

- a=1, b=1
- Treat all values equally *apriori:* use when highly uncertain or want to let the data speak for itself
- Similar results as traditional non-Bayes methods



Modified from: Ch1 of Ntzoufras, Ioannis. *Bayesian modeling using WinBUGS*. John Wiley & Sons, 2011.

## Step 2: model links new data $y$ to $\theta$

- $Y$ = 554 successful field goals
- $N$ = 1183 attempts
- Model:

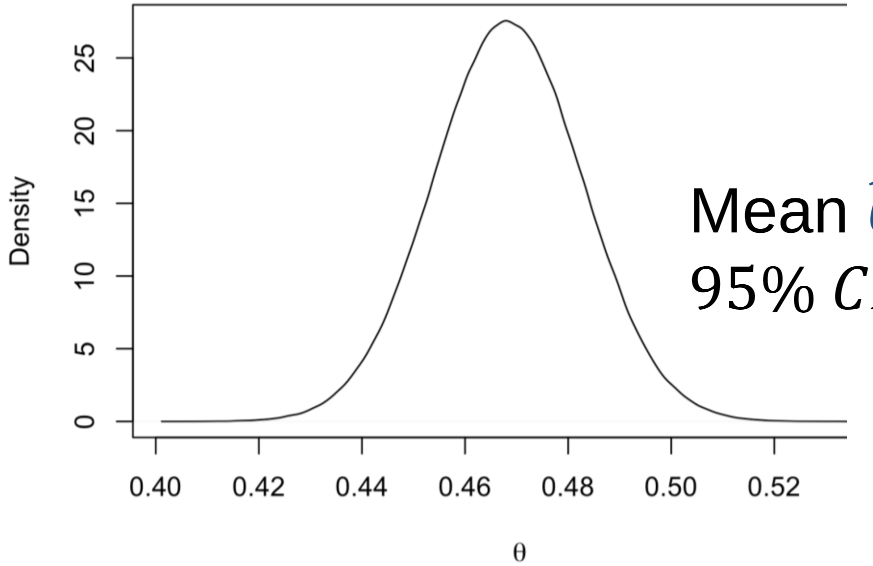
$$p(Y\,|\theta) = \; Binomial(N, \theta)$$
$$= \binom{N}{Y} \theta^{Y} (1-\theta)^{N-1}$$

## Step 3: Bayes' Thm update knowledge

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}$$

$=$  ..calculus..

$$= Beta(a^* = a + Y, \ldots b^* = b + N - Y)$$
$$= Beta(a^* = 555, \quad b^* = 630)$$




Mean $\boldsymbol{\hat{\theta} = 46.8\%}$
95% $CI$: 44.0 to 49.7%

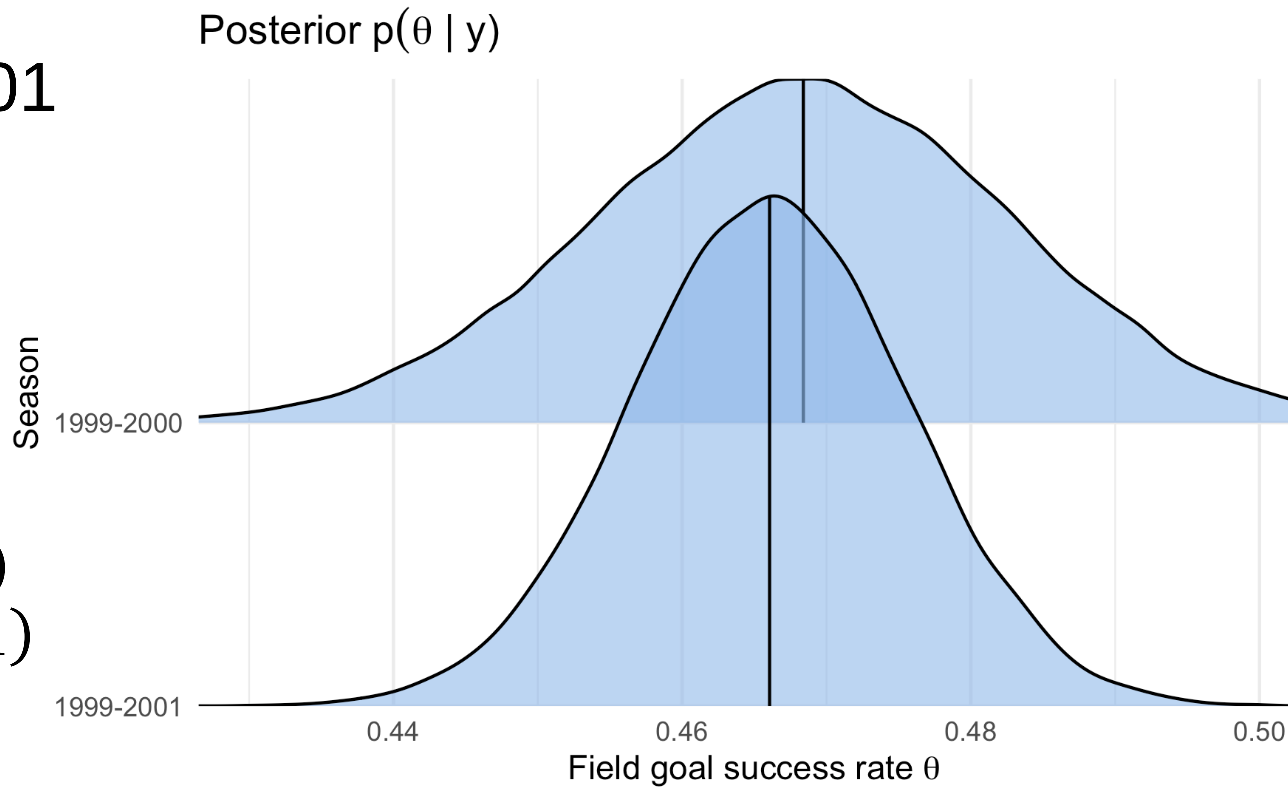
UCLA Health | Jonsson Comprehensive Cancer Center

# Continuous updating of knowledge

- 1999-2000: Kobe's FG% ~ Beta(555, 630)
- What if we collect more data? 2000-2001
- $p(\theta)$ = Beta(555, 630)

  $Y$ = 701 successes; $N$ = 1510 attempts

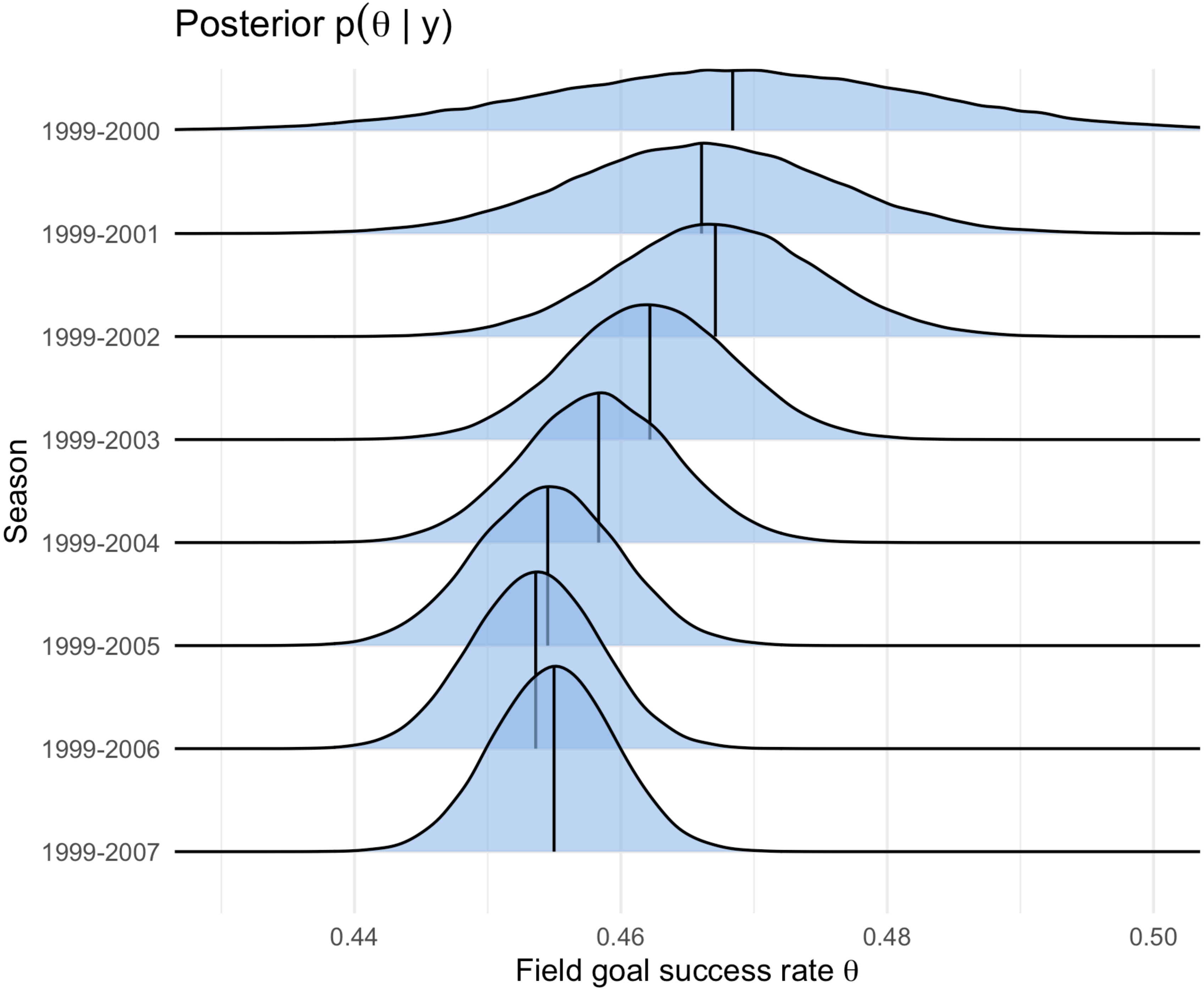
$$p(\theta|y) = Beta(a^* = a + Y, .. b^* = b + N - Y)$$
$$= Beta(555 + 701, 630 + 1510 - 701)$$
$$= Beta(1256, 1439)$$

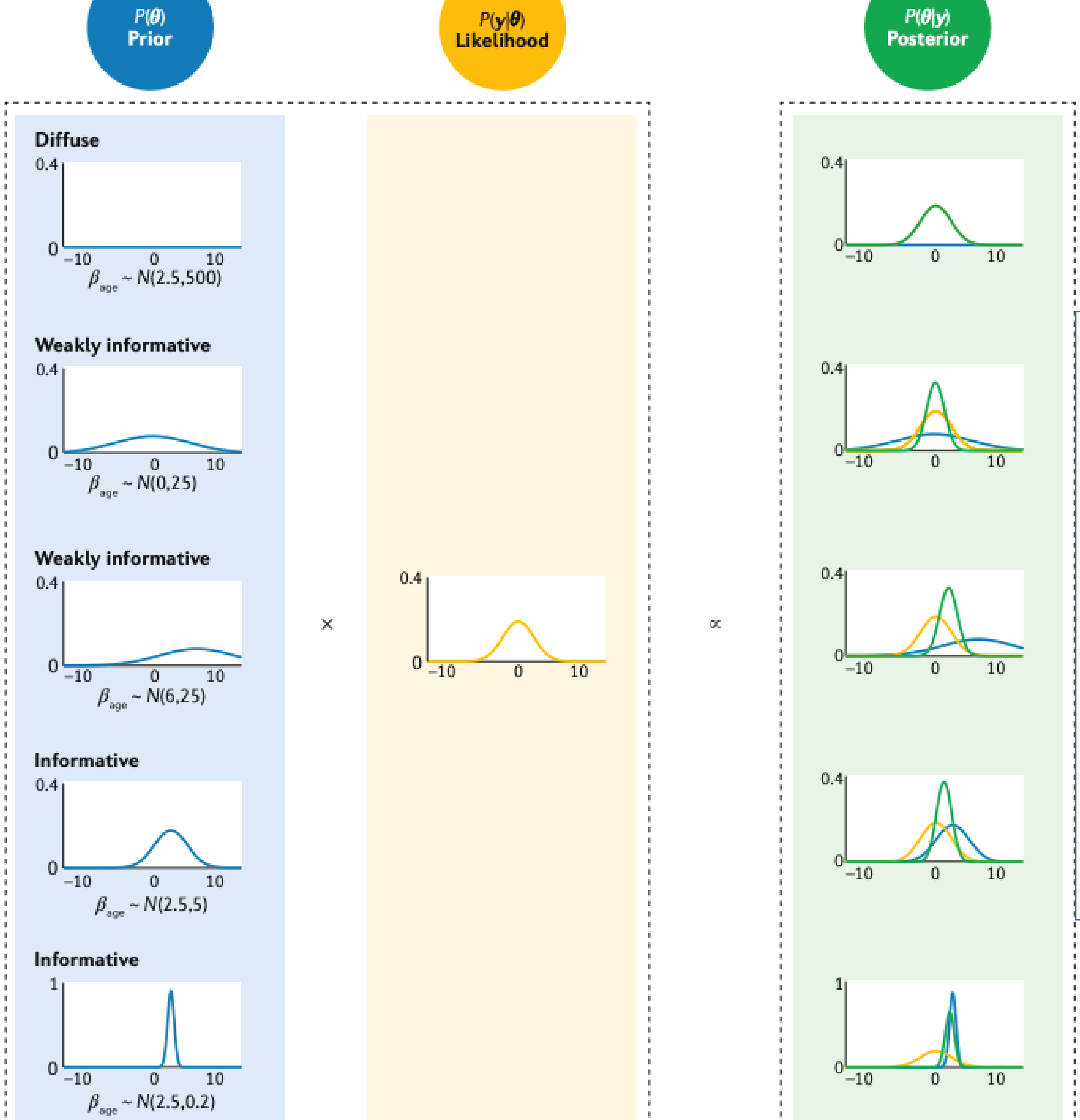



"Today's Posterior is tomorrow's Prior" (Lindley 2000)



Lindley, Dennis V. "The philosophy of statistics." *Journal of the Royal Statistical Society: Series D (The Statistician)* 49.3 (2000): 293-337.


Posterior p(θ | y)
Season
1999-2000
1999-2001
1999-2002
1999-2003
1999-2004
1999-2005
1999-2006
1999-2007
0.44
0.46
0.48
0.50
Field goal success rate θ




# How Prior affects Posterior

- **"Flat/diffuse prior"** → more weight to Likelihood (new data)
- **"Informative prior"** (concentrated on smaller range) → more weight to prior
  - Although as N gets larger, the Likelihood gets more weight

van de Schoot, Rens, et al. "Bayesian statistics and modelling." *Nature Reviews Methods Primers* 1.1 (2021): 1-26.

# MCMC for general Bayesian models

- Bayes' Thm:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}$$

- In general, no closed form solution
- Use MCMC (Markov chain Monte Carlo) to simulate from $p(\theta|y)$

Table 2 | **A non-exhaustive summary of commonly used and open Bayesian software programs**

| Software package | Summary |
|---|---|
| ***General-purpose Bayesian inference software*** | |
| BUGS[231,232] | The original general-purpose Bayesian inference engine, in different incarnations. These use Gibbs and Metropolis sampling. Windows-based software (WinBUGS[233]) with a user-specified model and a black-box MCMC algorithm. Developments include an open-source version (OpenBUGS[234]) also available on Linux and Mac |
| JAGS[235] | An open-source variation of BUGS that can run cross-platform and can run from R via rjags[236] |
| PyMC3[237] | An open-source framework for Bayesian modelling and inference entirely within Python; includes Gibbs sampling and Hamiltonian Monte Carlo |
| Stan[98] | An open-source, general-purpose Bayesian inference engine using Hamiltonian Monte Carlo; can be run from R, Python, Julia, MATLAB and Stata |
| NIMBLE[238] | Generalization of the BUGS language in R; includes sequential Monte Carlo as well as MCMC. Open-source R package using BUGS/JAGS-model language to develop a model; different algorithms for model fitting including MCMC and sequential Monte Carlo approaches. Includes the ability to write novel algorithms |
| ***Programming languages that can be used for Bayesian inference*** | |
| TensorFlow Probability[239,240] | A Python library for probabilistic modelling built on Tensorflow[203] from Google |
| Pyro[241] | A probabilistic programming language built on Python and PyTorch[204] |
| Julia[242] | A general-purpose language for mathematical computation. In addition to Stan, numerous other probabilistic programming libraries are available for the Julia programming language, including Turing.jl[243] and Mamba.jl[244] |
| ***Specialized software doing Bayesian inference for particular classes of models*** | |
| JASP[245] | A user-friendly. higher-level interface offering Bayesian analysis. Open source and relies on a collection of open-source R packages |
| R-INLA[230] | An open-source R package for implementing INLA[246]. Fast inference in R for a certain set of hierarchical models using nested Laplace approximations |
| GPstuff[247] | Fast approximate Bayesian inference for Gaussian processes using expectation propagation; runs in MATLAB, Octave and R |



van de Schoot, Rens, et al. "Bayesian statistics and modelling." *Nature Reviews Methods Primers* 1.1 (2021): 1-26.

# Frequentist *vs.* Bayesian

- **Frequentist:** if repeated same study many times, how *frequently* would these ***unobserved*** studies be similar to our observed study?

- **Bayesian**: conditional on prior knowledge + new data; no unobserved hypothetical studies

- **P-value**: assuming $H_0$ is true and repeat study many times, $p =$ probability these ***unobserved*** studies generate more extreme results than the current study…
  - Does NOT tell you probability that $H_0$, $H_A$ are true/false
  - Bayes' **posterior probabilities**: probability that $H_0$, $H_A$ are true/false

- **95% Confidence Interval**: if repeated same study many times, expect 95% of CI's from these **unobserved** studies to contain the true $\theta$…?
  - Bayesian **Credible Interval**: 95% probability the interval contains $\theta$

- **What about questions that are fundamentally not repeatable?**
  - Will Biden be re-elected in 2024?
  - Will COVID be eradicated by 2025?
  - Unlike Bayes, it's unclear how Frequentist applies here

# Summary

- 3 key steps: prior knowledge $p(\theta)$ → model new data $p(y|\theta)$ → updated knowledge $p(\theta|y)$
- If uncomfortable using prior knowledge, then can use a flat non-informative prior → similar results as traditional frequentist methods
- **Everyone is Bayesian at "design stage"**: we use prior info to make "educated guess" for effect sizes, variances, etc.. prior info is not bad!

# Learn more


- van de Schoot, Rens, et al. "Bayesian statistics and modelling." *Nature Reviews Methods Primers* 1.1 (2021): 1-26.
- Stephens, Matthew, and David J. Balding. "Bayesian statistical methods for genetic association studies." *Nature Reviews Genetics* 10.10 (2009): 681-690.
- Kruschke, John. "Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan."

# Questions