Bayesian Estimation

Bayes' Theorem

$$P[\theta|X] = \frac{P[X|\theta]P[\theta]}{P[X]}$$

Bayes' Theorem

Posterior

Likelihood

Prior

$$P[\theta|X] = \frac{P[X|\theta]P[\theta]}{P[X]}$$

Marginal likelihood

Bayes’ Theorem

The **posterior** gives the probability of parameter θ given the observations

The **likelihood** is the model from which we believe the data arises – choose carefully!

The **prior** includes any knowledge that we might have; may be “uninformative”

The **marginal likelihood** sums probability of the data over all possible values of θ

$$P[X] = \sum_{\theta} P[X|\theta]P[\theta]$$

We are embarking on a new study of the black-capped chickadee (*Poecile atricapillus*)/Carolina chickadee (*Poecile carolinensis*) hybrid zone. There are only very slight visual differences between these species. We would like to know the proportion of the population that are hybrids. Before we go out and collect some data, we can start to build our model.

https://www.sibleyguides.com/bird-info/black-capped-chickadee/black-capped-carolina-chickadee/

Let’s try an example!

Build the model

What can we use for the likelihood?

- Proportions → binomial

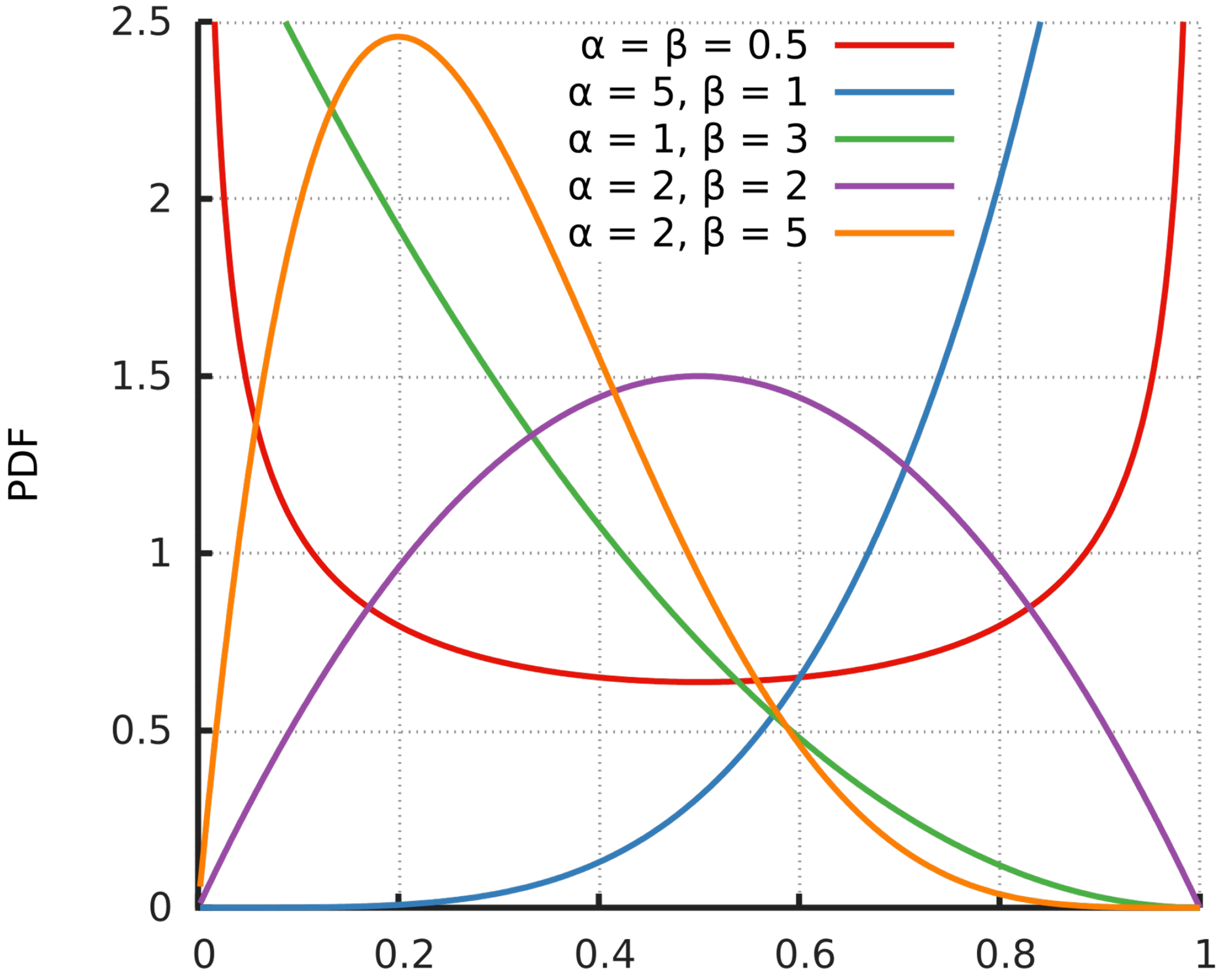
What can we use for the prior?

- Many things!
- Conjugate priors → for binomial, conjugate prior is beta
- Let's say we have some information that suggests the proportion may be 0.5 i.e. 1:1 ratio

The beta distribution

$$\text{pmf}(\alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\mathrm{B}(\alpha, \beta)}$$

Let's say $\alpha = 2$ and $\beta = 2$



Remember the formula

X: the species we observe
θ: the proportion alewives

$$P[\theta|X] = \frac{P[X|\theta]P[\theta]}{P[X]}$$

How de we actually use the formula?

$$P[X|\theta] = \text{binomial distribution} = \binom{n}{k}\theta^k(1-\theta)^{n-k}$$

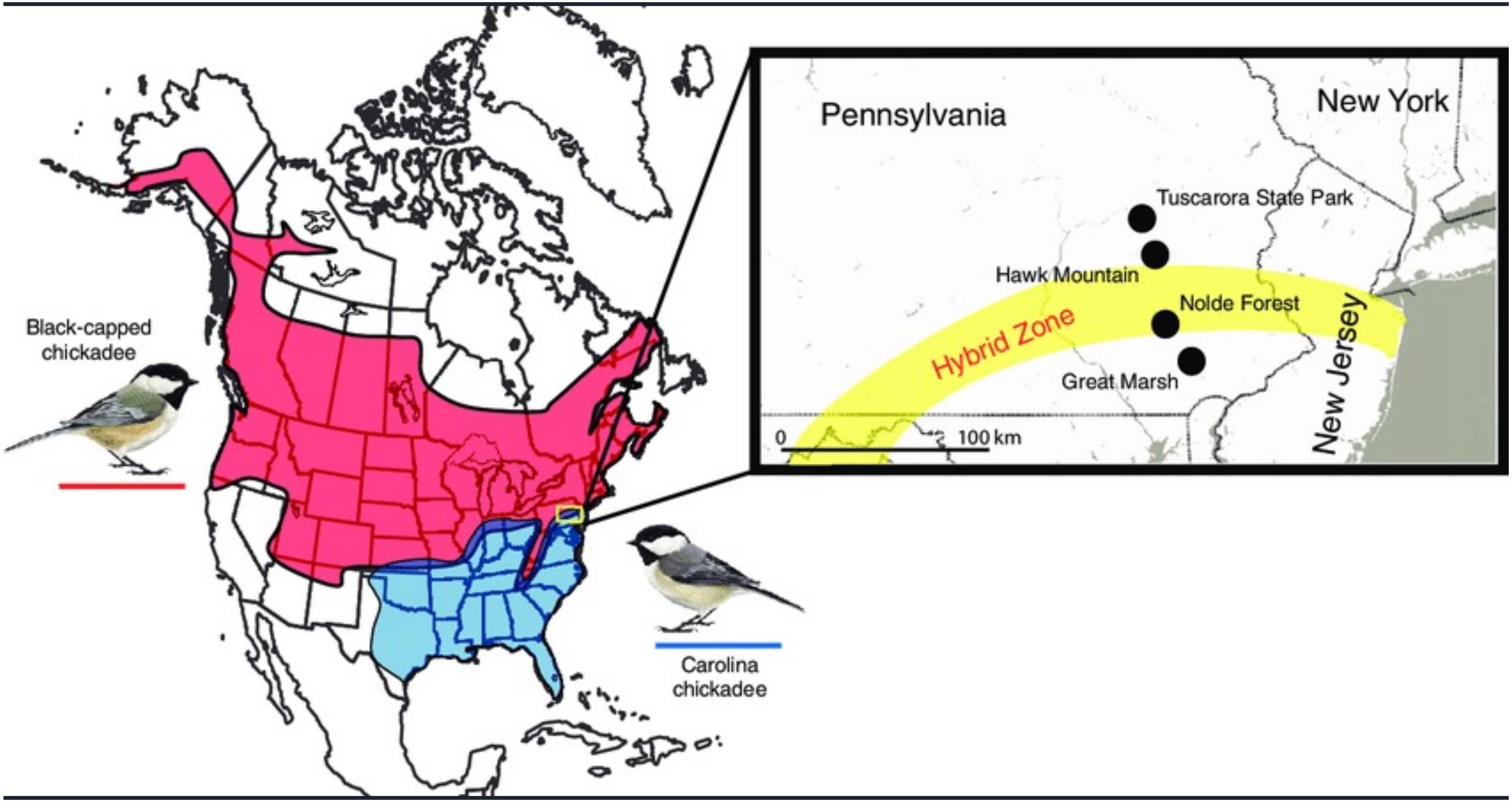
$$P[\theta] = \text{beta distribution} = \frac{\theta^{2-1}(1-\theta)^{2-1}}{\mathrm{B}(2,2)}$$

$$P[X] = \text{beta distribution} = \sum_{\theta} P[X|\theta]P[\theta] = \frac{\binom{n}{k}\theta^{k+2-1}(1-\theta)^{n-k+2-1}}{\mathrm{B}(\alpha,\beta)}$$

Marginal Likelihood

$$\mathrm{B}(2,2) = \frac{\Gamma(2)\Gamma(2)}{\Gamma(2+2)} = \frac{(2-1)!\,(2-1)!}{(2+2)!} = \frac{1}{6}$$

$$\sum_{\theta} P[X|\theta]P[\theta] = 6\binom{n}{k}\theta^{k+1}(1-\theta)^{n-k+1}$$



Dominique N. Wagner, Robert L. Curry, Nancy Chen, Irby J. Lovette, Scott A. Taylor, Genomic regions underlying metabolic and neuronal signaling pathways are temporally consistent in a moving avian hybrid zone, *Evolution*, Volume 74, Issue 7, 1 July 2020, Pages 1498–1513, https://doi.org/10.1111/evo.13970

Data! Just 3 birds to start with...



https://ebird.org/home

$$X = \{0,1,1\}$$
$$n = 3$$
$$k = 2$$

Model – marginal likelihood

$$\sum_{\theta} P[X|\theta]P[X] = 6\binom{n}{k}\theta^{k+1}(1-\theta)^{n-k+1} = 18\theta^3(1-\theta)^2\text{x}$$

$$18\int_0^1 \theta^3(1-\theta)^2 d\theta$$

$$= 18\int_0^1 (\theta^3 - 2\theta^4 + \theta^5)d\theta$$

$$= 18(\frac{1}{4}\theta^4 - \frac{2}{5}\theta^5 + \frac{1}{6}\theta^6\Big|_0^1)$$

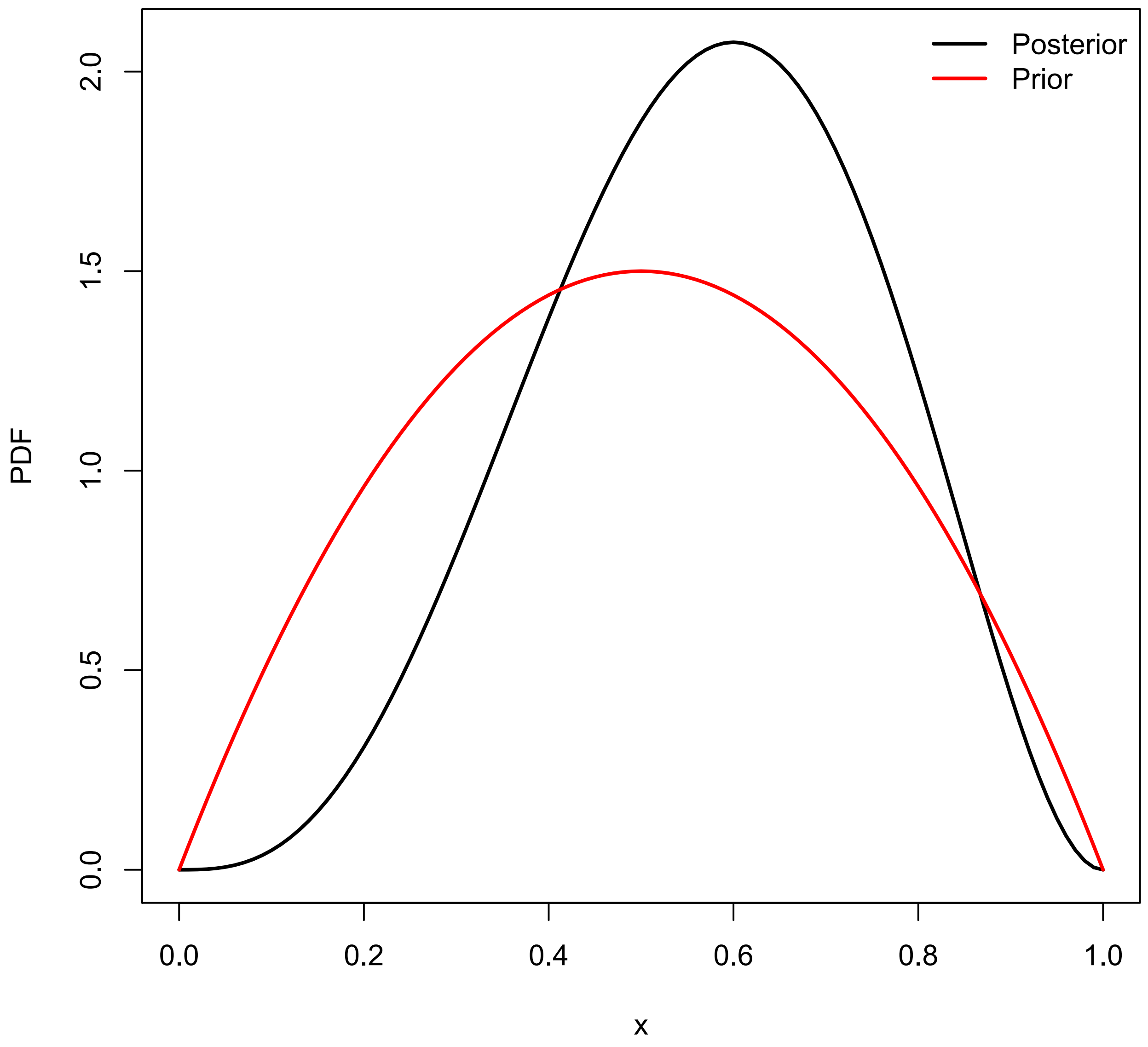
$$= \frac{18}{60} = \frac{3}{10}x$$

Model –likelihood*prior

$$P[X|\theta]P[\theta] = \binom{n}{k}\theta^k(1-\theta)^{n-k}\frac{\theta(1-\theta)}{\mathrm{B}(2,2)}$$
$$= 6\binom{n}{k}\theta^{k+1}(1-\theta)^{n-k+1}$$
$$= 18\theta^3(1-\theta)^2$$

Model –posterior

$$\frac{P[X|\theta]P[\theta]}{P[X]} = \frac{18\theta^3(1-\theta)^2}{\frac{3}{10}} = 60\ \theta^3(1-\theta)^2$$



Posterior
Prior
PDF
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
x

Summary

Bayes' Theorem can be used directly for simple models

Informative priors can have an influence on our posterior

What happens with more complex models?

- The marginal likelihood is often too difficult and too small to calculate!
- We use Markov Chain Monte Carlo to sample the **numerator** (likelihood*prior) and estimate the marginal
- We write the model in the Stan language and then rstan does the sampling and summary for us