ORIE 6217/CS6384: Applied Bayesian Data Analysis for Research

Lecture 2: Bayesian Intro

Nikhil Garg

Announcements

Lecture overview

- Uncertainty and Bayes Rule
  - Art of probabilistic modeling
- Walking through Bayesian learning in a single parameter (binomial) model
- Stan code for single parameter model

Uncertainty and Bayes rule

Uncertainty and probabilistic modeling

- Two types of uncertainty: aleatoric and epistemic
- Representing uncertainty with probabilities
- Updating uncertainty

Two types of uncertainty

- Aleatoric uncertainty due to randomness
- Epistemic uncertainty due to lack of knowledge

Two types of uncertainty

- Aleatoric uncertainty due to randomness
  - we are not able to obtain observations which could reduce this uncertainty
- Epistemic uncertainty due to lack of knowledge

Two types of uncertainty

- Aleatoric uncertainty due to randomness
  - we are not able to obtain observations which could reduce this uncertainty
- Epistemic uncertainty due to lack of knowledge
  - we are able to obtain observations which can reduce this uncertainty
  - two observers may have different epistemic uncertainty

Updating uncertainty

- Probability of red $\frac{\#\text{red}}{\#\text{red}+\#\text{yellow}} = \theta$

Updating uncertainty

- Probability of red $\frac{\#\text{red}}{\#\text{red}+\#\text{yellow}} = \theta$

- $p(y = \#\text{red}|\theta) = \theta$ aleatoric uncertainty

Updating uncertainty

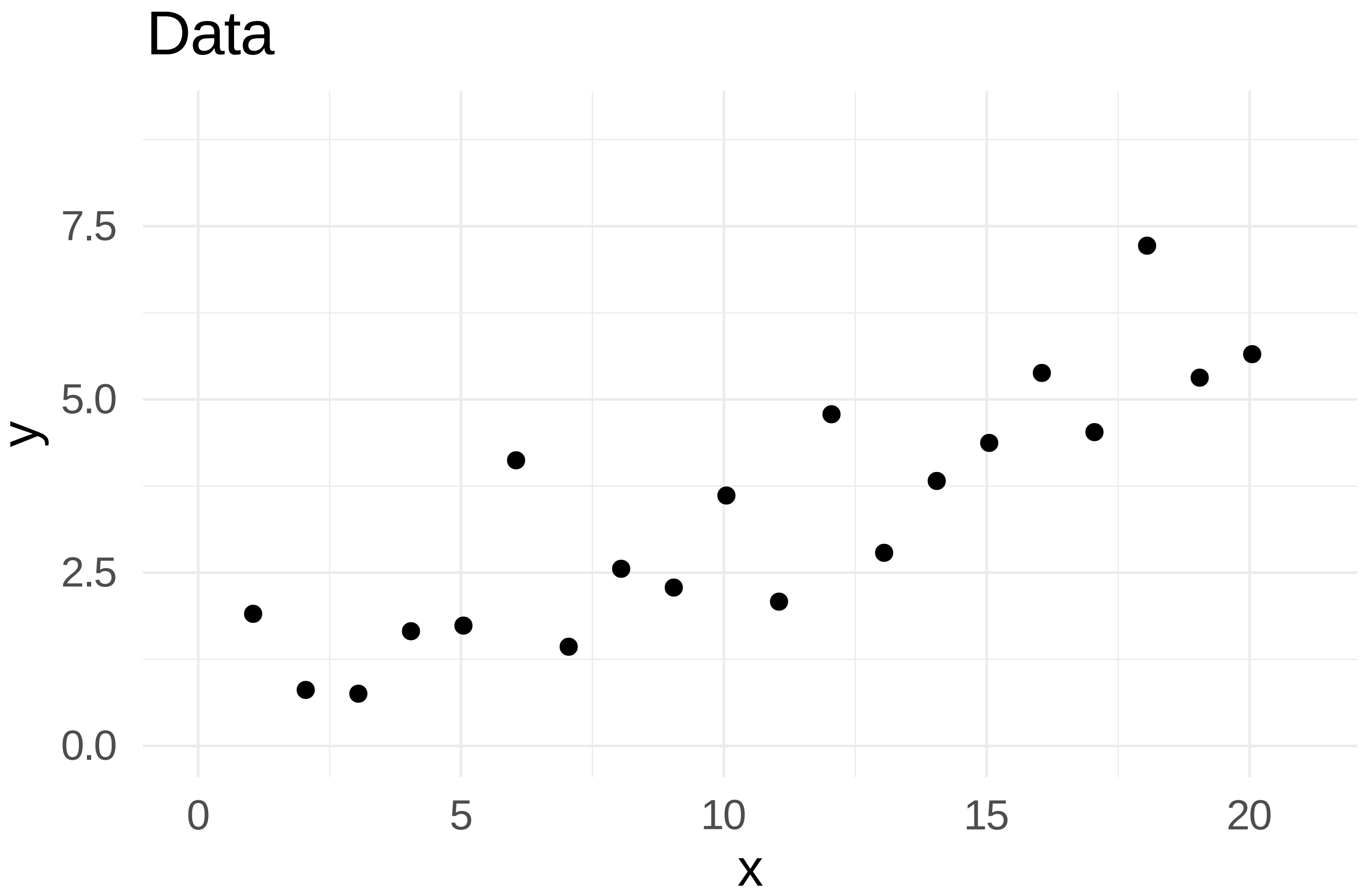
- Probability of red $\frac{\#\text{red}}{\#\text{red}+\#\text{yellow}} = \theta$
- $p(y = \#\text{red} | \theta) = \theta$ aleatoric uncertainty
- $p(\theta)$ epistemic uncertainty

Updating uncertainty

- Probability of red $\frac{\#\text{red}}{\#\text{red}+\#\text{yellow}} = \theta$
- $p(y = \#\text{red}|\theta) = \theta$ aleatoric uncertainty
- $p(\theta)$ epistemic uncertainty
- Data reduces epistemic uncertainty: Picking many chips updates our uncertainty about the proportion
- $p(\theta|\text{y} = \#\text{red}, \#\text{yellow}, \#\text{red}, \#\text{red}, \ldots) = ?$
  - Bayes rule $p(\theta|y) =$

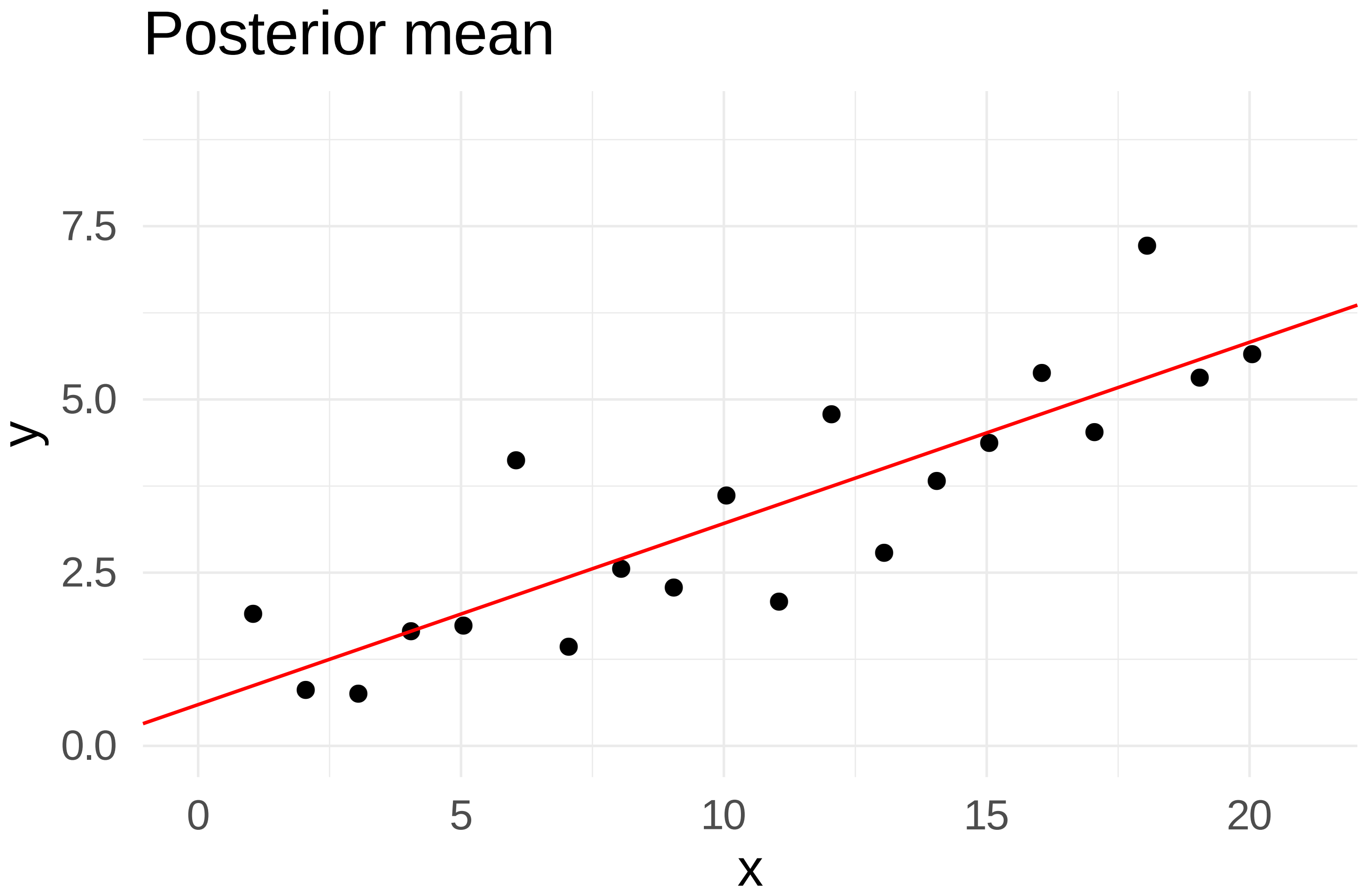
Uncertainty in modeling

$Y = \theta * x + noise$



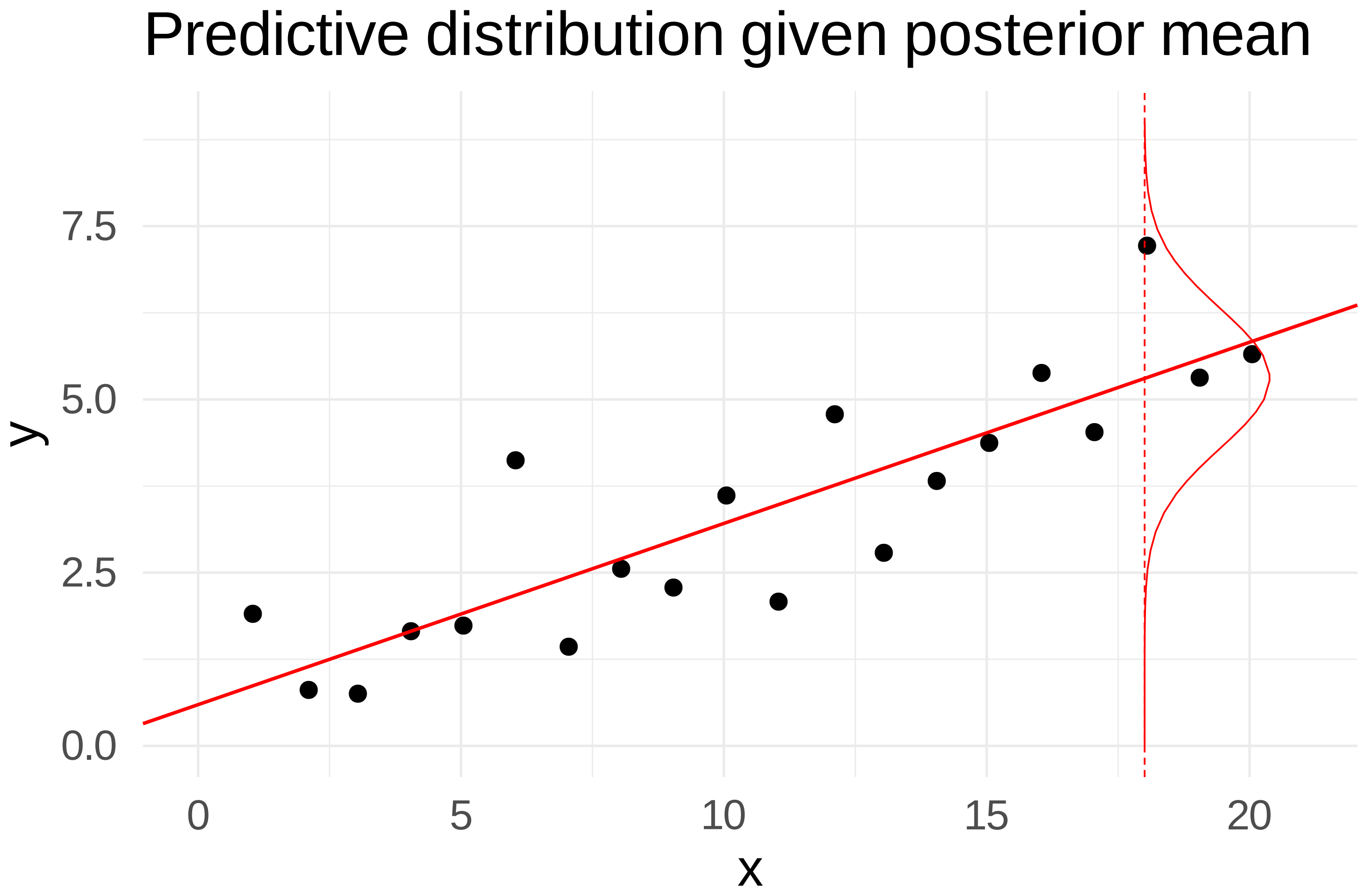
Uncertainty in modeling

$Y = \theta * x + noise$



Uncertainty in modeling

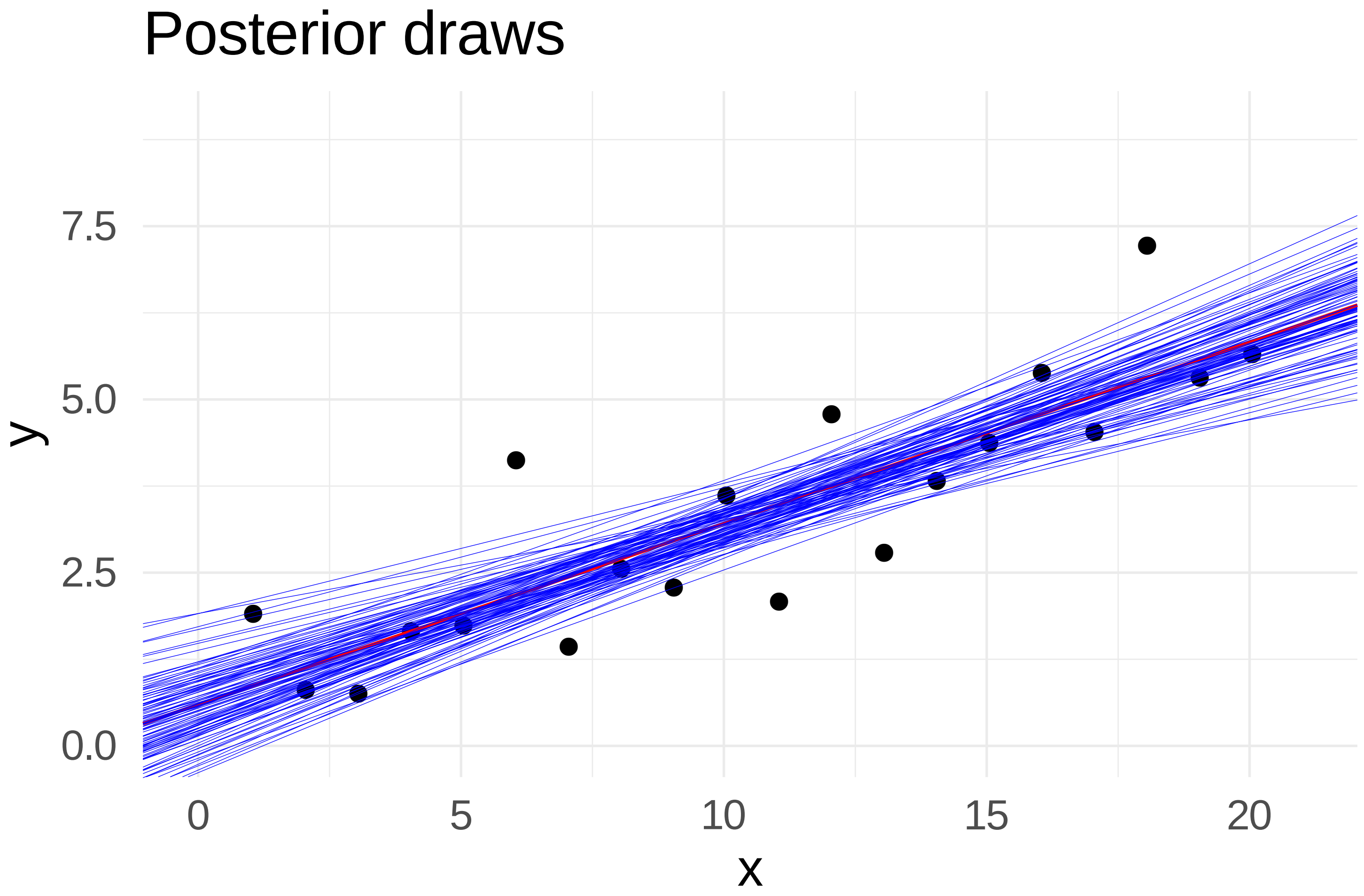
$Y = \theta * x + noise$



- Model: $p(\mathbf{y}|\theta)$ as a function of **y** given fixed θ describes the aleatoric uncertainty

Uncertainty in modeling

$Y = \theta * x + noise$



- ***θ*: our knowledge of *θ* also varies (epistemic uncertainty)**

Model vs. likelihood

- Bayes rule $p(\theta|y) \propto p(y|\theta)p(\theta)$

- Model: $p(\mathbf{y}|\theta)$ as a function of **y** given fixed θ describes the aleatoric uncertainty

- Likelihood: $p(y|\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$ given fixed y provides information about epistemic uncertainty, but is not a probability distribution

- Bayes rule combines the likelihood with prior uncertainty $p(\theta)$ and transforms them to updated posterior uncertainty

The art of probabilistic modeling

- The art of probabilistic modeling is to describe in a mathematical form (model and prior distributions) what we already know and what we don't know

The art of probabilistic modeling

- The art of probabilistic modeling is to describe in a mathematical form (model and prior distributions) what we already know and what we don't know

- "Easy" part is to use Bayes rule to update the uncertainties
  - computational challenges

The art of probabilistic modeling

- The art of probabilistic modeling is to describe in a mathematical form (model and prior distributions) what we already know and what we don’t know
- “Easy” part is to use Bayes rule to update the uncertainties
  - computational challenges
- Other parts of the art of probabilistic modeling are, for example,
  - Finding the right model that says something useful
  - model checking: is data in conflict with our prior knowledge?
  - presentation: presenting the model and the results to the application experts

Single parameter example again

Binomial model for binary data

- Binomial model is the simplest model
  - useful to discuss likelihood, posterior, prior, integration, posterior summaries
  - very commonly used as a building block
  - examples:
    - coin tossing
    - chips from bag
    - covid tests and vaccines
    - classification / logistic regression

Binomial: known θ

- Probability of event 1 in trial is θ

Binomial: known θ

- Probability of event 1 in trial is θ
- Probability of event 2 in trial is $1 - \theta$

Binomial: known θ

- Probability of event 1 in trial is θ
- Probability of event 2 in trial is $1 - \theta$
- Probability of several events in independent trials is e.g. $\theta\theta(1-\theta)\theta(1-\theta)(1-\theta)\ldots$

Binomial: known θ

- Probability of event 1 in trial is θ
- Probability of event 2 in trial is $1 - \theta$
- Probability of several events in independent trials is e.g. $\theta\theta(1-\theta)\theta(1-\theta)(1-\theta)\ldots$
- If there are n trials and we don't care about the order of the events, then the probability that event 1 happens y times is

$$p(y|\theta, n) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

Binomial: known θ

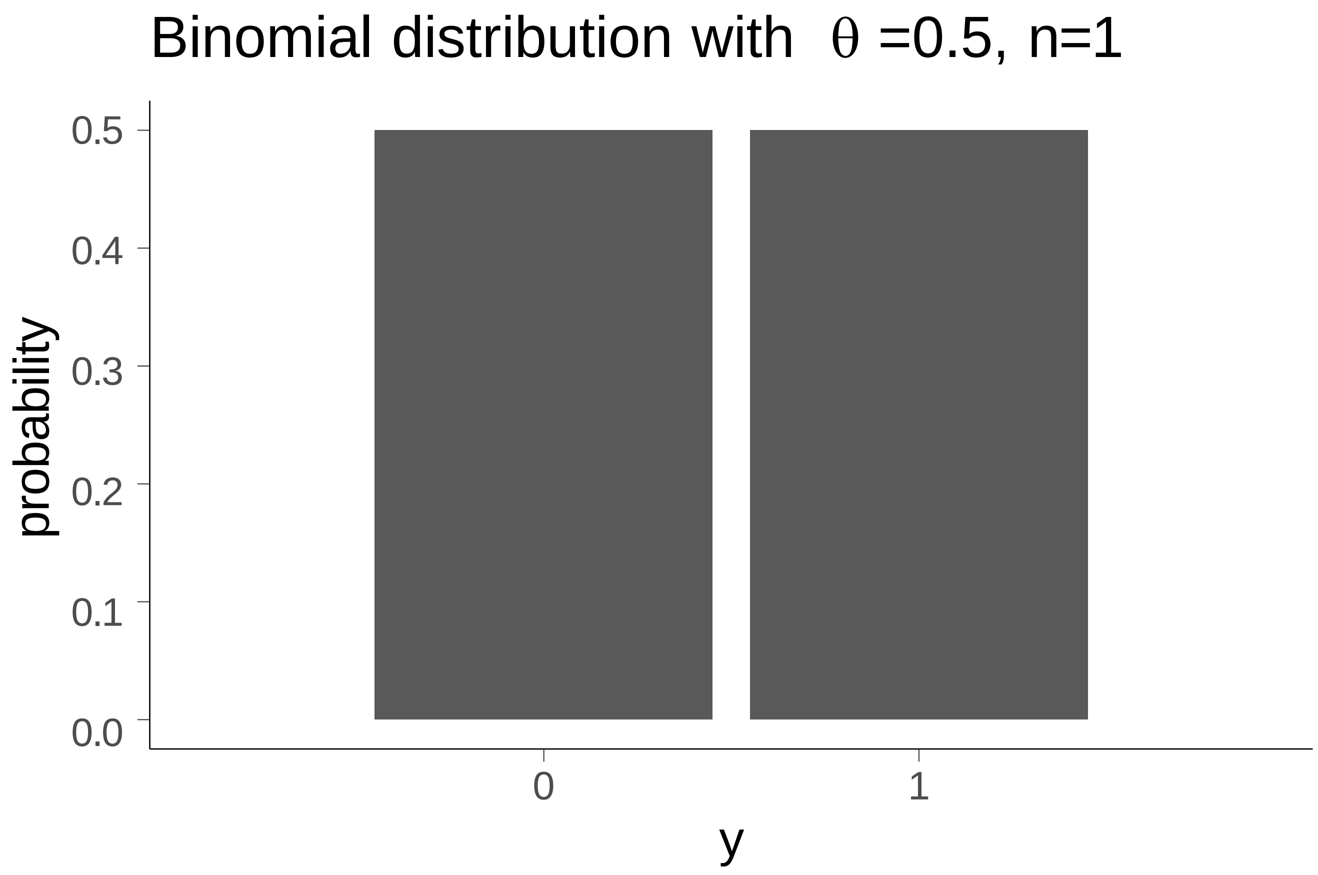
- Observation model (function of y, discrete)

$$p(y|\theta, n) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

Binomial: known θ

- Observation model (function of y, discrete)

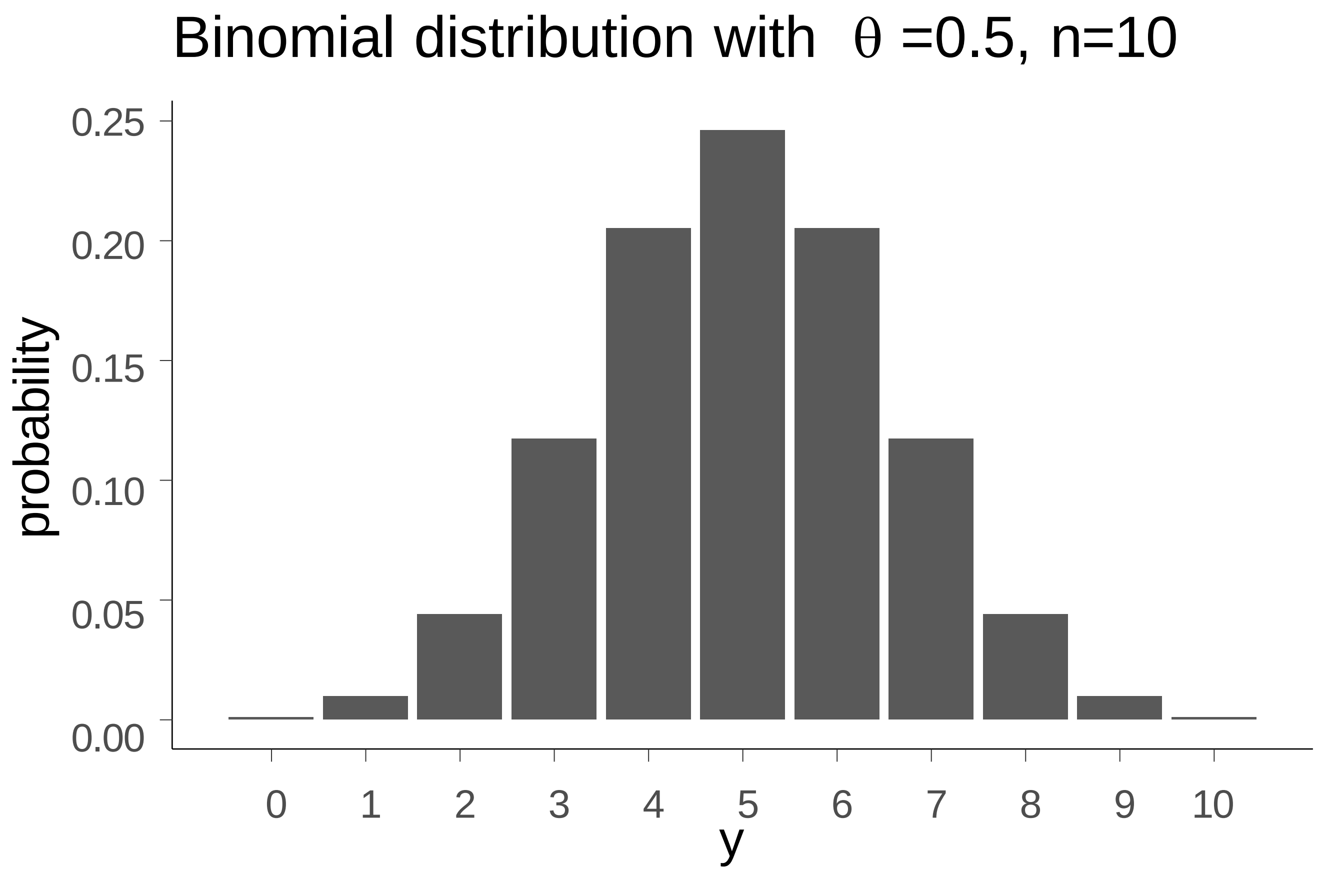
$$p(y|\theta, n) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$



Binomial: known θ

- Observation model (function of y, discrete)

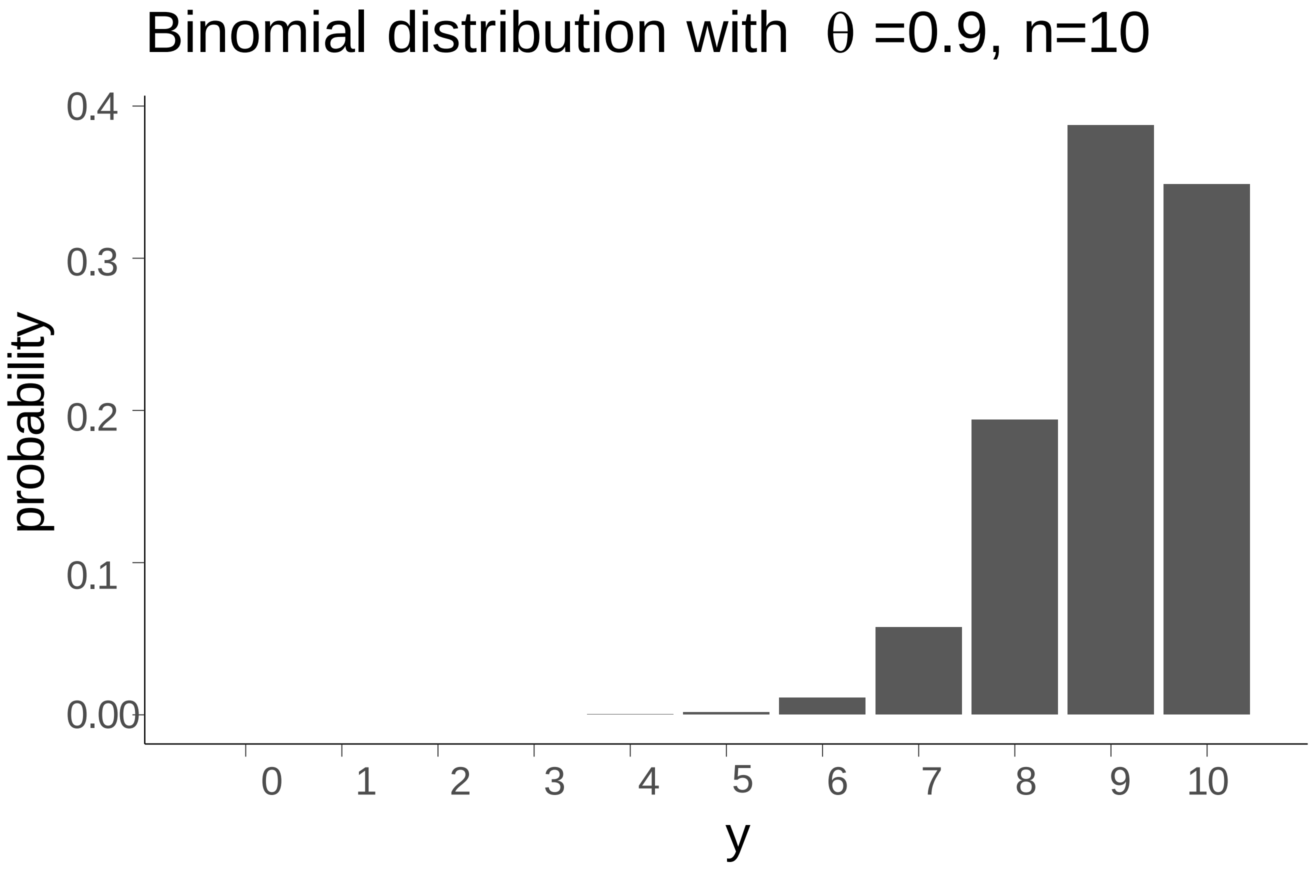
$$p(y|\theta, n) = \binom{n}{y} \theta^{y} (1-\theta)^{n-y}$$



Binomial: known θ

- Observation model (function of y, discrete)

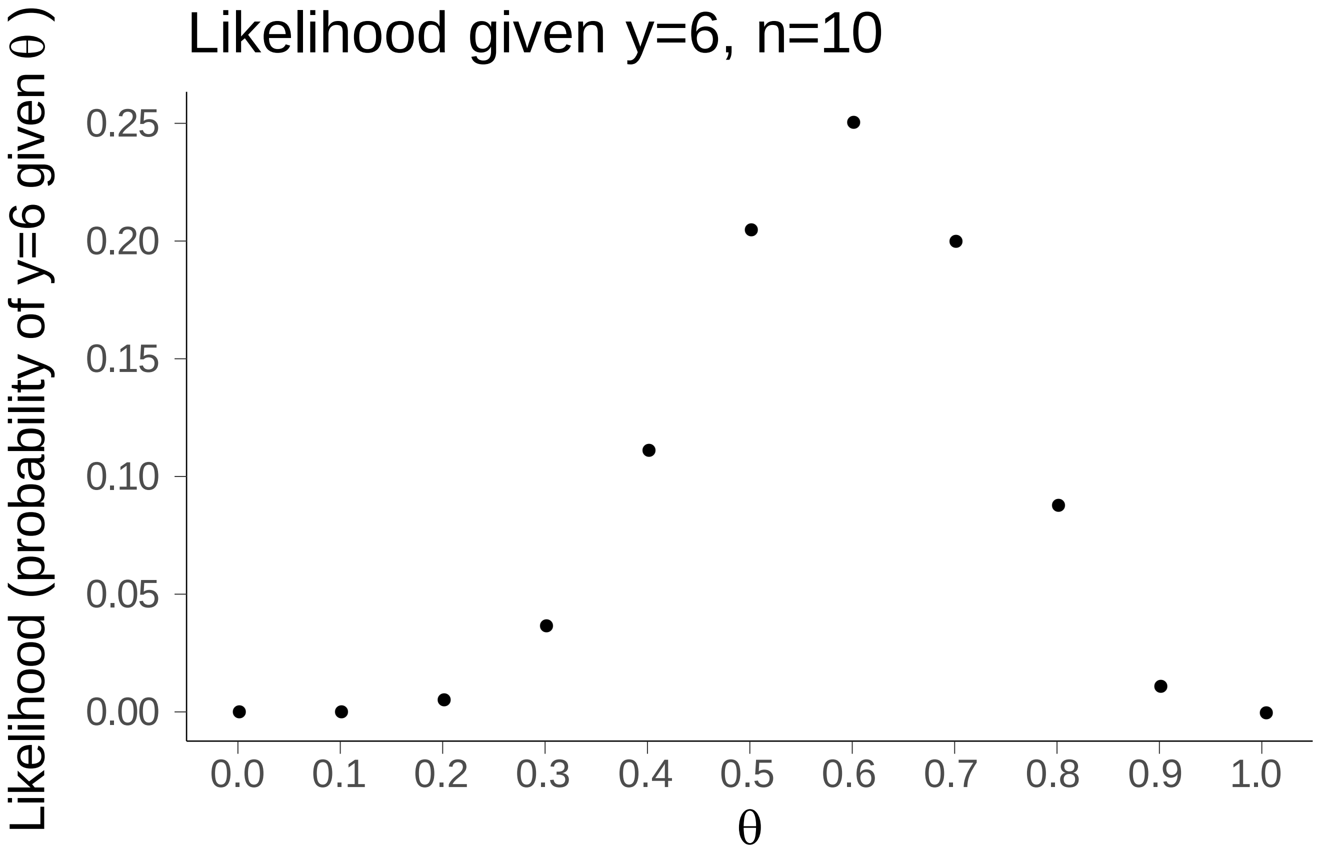
$$p(y|\theta, n) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$



Binomial: unknown θ

- Likelihood (function of θ, continuous)

$$p(y|\theta, n) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$$

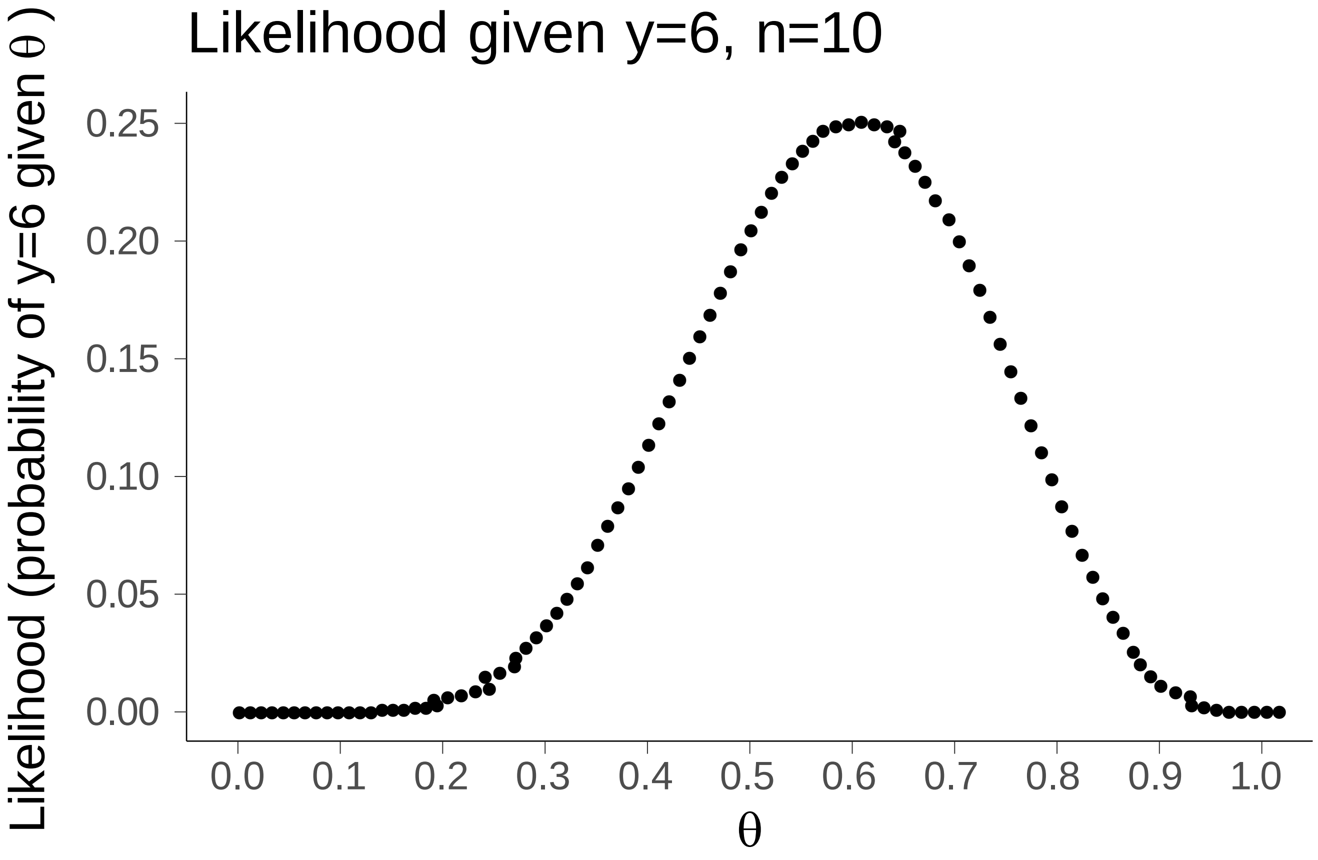


$p(y = 6|n = 10, \theta)$: 0.00 0.00 0.01 0.04 0.11 **0.21** 0.25 0.20 0.09 **0.01** 0.00

Binomial: unknown θ

- Likelihood (function of θ, continuous)

$$p(y|\theta, n) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$



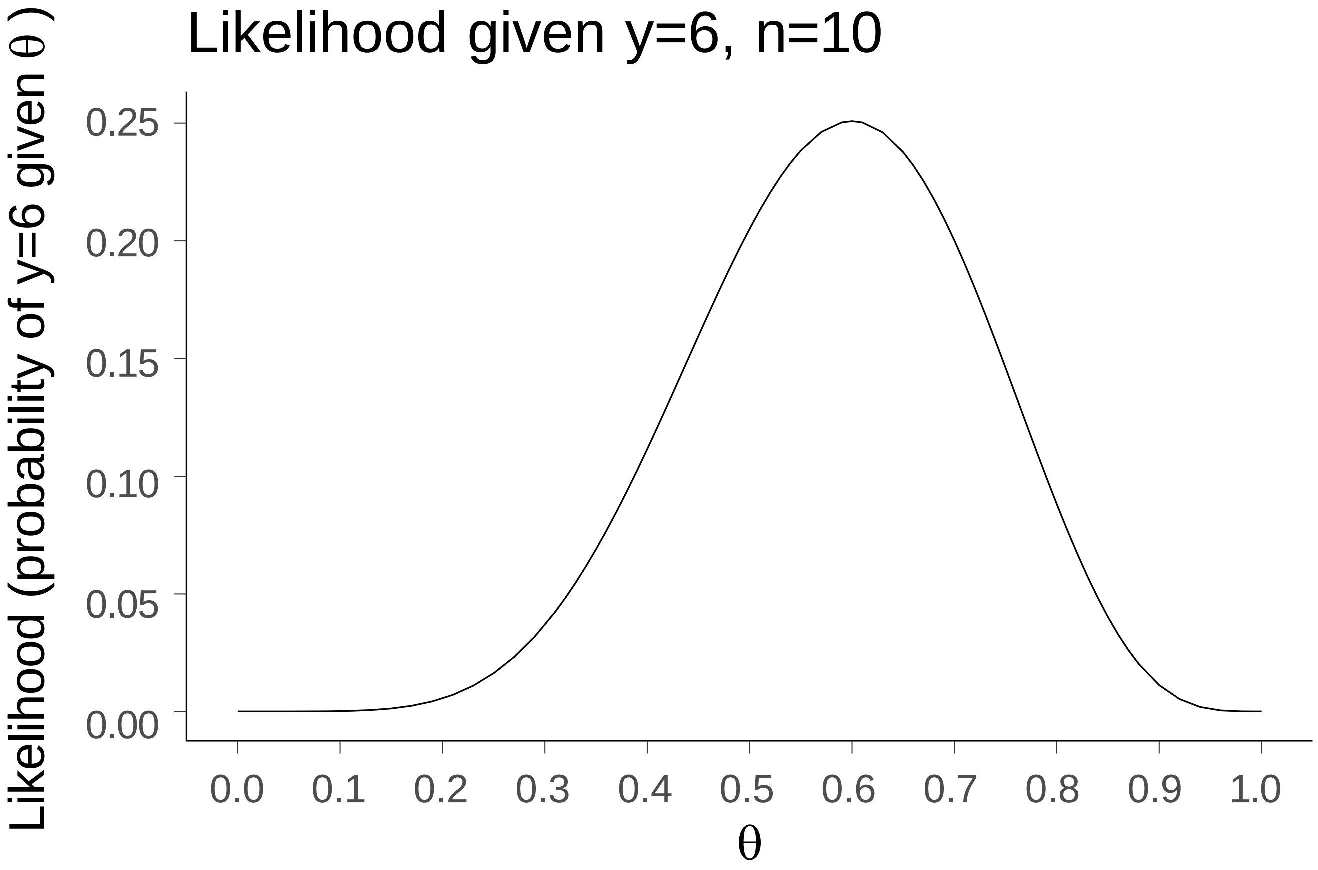
we can compute the value for any θ, but in practice can evaluate only finite times

Binomial: unknown θ

- Likelihood (function of θ, continuous)

$$p(y|\theta, n) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

With linear interpolation, looks smooth, and we'll get back to later to computational cost issues



Binomial: unknown θ

- Posterior with Bayes rule (function of θ, continuous)

$$p(\theta|y,n) = \frac{p(y|\theta,n)p(\theta|n)}{p(y|n)}$$

Binomial: unknown θ

- Posterior with Bayes rule (function of θ, continuous)

$$p(\theta|y,n) = \frac{p(y|\theta,n)p(\theta|n)}{p(y|n)}$$

where $p(y|n) = \int p(y|\theta,n)p(\theta|n)d\theta$

Binomial: unknown θ

- Posterior with Bayes rule (function of θ, continuous)

$$p(\theta|y,n) = \frac{p(y|\theta,n)p(\theta|n)}{p(y|n)}$$

where $p(y|n) = \int p(y|\theta,n)p(\theta|n)d\theta$

- Start with uniform prior

$$p(\theta|n) = p(\theta|M) = 1, \text{ when } 0 \leq \theta \leq 1$$

The textbook/lectures I'm borrowing from sometimes uses M to remind us that this is an assumption, and so some quantities are due to our assumptions

Binomial: unknown θ

- Posterior with Bayes rule (function of θ, continuous)

$$p(\theta|y,n) = \frac{p(y|\theta,n)p(\theta|n)}{p(y|n)}$$

where $p(y|n) = \int p(y|\theta,n)p(\theta|n)d\theta$

- Start with uniform prior

$$p(\theta|n) = p(\theta|M) = 1, \text{ when } 0 \le \theta \le 1$$

- Then

$$p(\theta|y,n) = \frac{p(y|\theta,n)}{p(y|n)} = \frac{\binom{n}{y}\theta^y(1-\theta)^{n-y}}{\int_0^1 \binom{n}{y}\theta^y(1-\theta)^{n-y}d\theta}$$

$$= \frac{1}{Z}\theta^y(1-\theta)^{n-y}$$

Binomial: unknown θ

- Normalization term Z (constant given y)

$$Z = p(y|n) = \int_0^1 \theta^y (1-\theta)^{n-y} d\theta = \frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)}$$

Binomial: unknown θ

- Normalization term Z (constant given y)

$$Z = p(y|n) = \int_0^1 \theta^y (1-\theta)^{n-y} d\theta = \frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)}$$

- Evaluate with $y = 6, n = 10$

```
y<-6;n<-10;
integrate(function(theta) theta^y*(1-theta)^(n-y), 0, 1)
```

$\approx$ 0.0004329

`gamma(6+1)*gamma(10-6+1)/gamma(10+2)` $\approx$ 0.0004329

Binomial: unknown θ

- Normalization term Z (constant given y)

$$Z = p(y|n) = \int_0^1 \theta^y (1-\theta)^{n-y} d\theta = \frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)}$$

- Evaluate with $y = 6, n = 10$

```
y<-6;n<-10;
integrate(function(theta) theta^y*(1-theta)^(n-y), 0, 1)
```

≈ 0.0004329

`gamma(6+1)*gamma(10-6+1)/gamma(10+2)` ≈ 0.0004329

usually computed via $\log \Gamma(\cdot)$ due to the limitations of floating point presentation

Binomial: unknown θ

- Posterior is

$$p(\theta|y,n) = \frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}\theta^y(1-\theta)^{n-y},$$

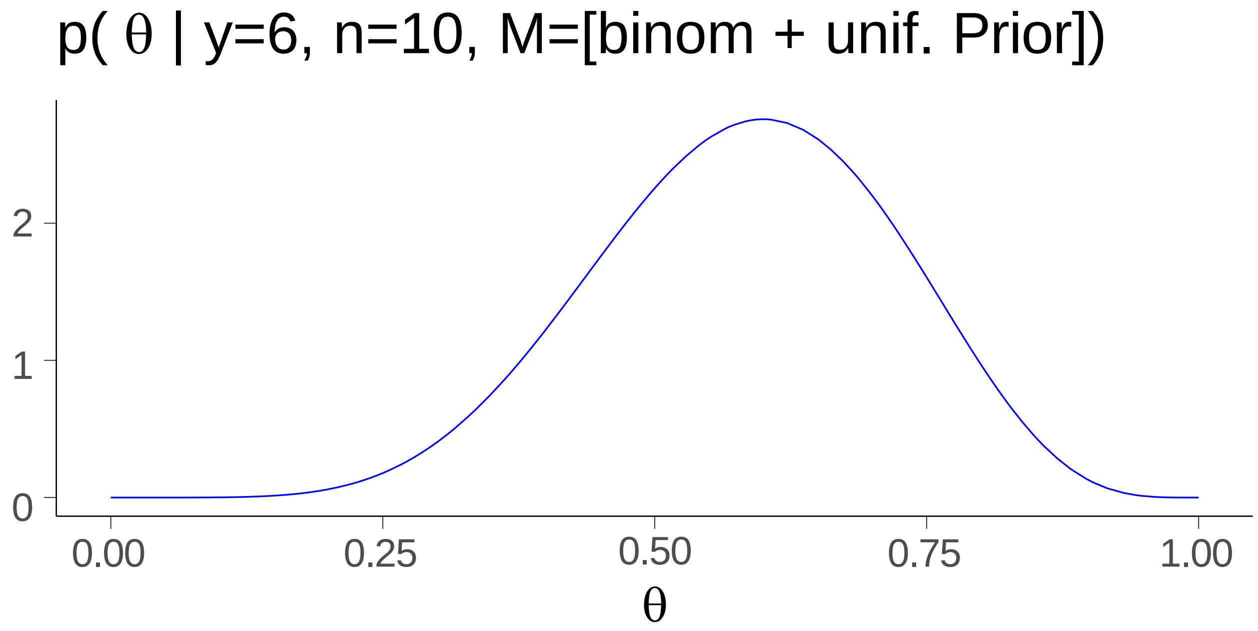
Binomial: unknown θ

- Posterior is

$$p(\theta|y,n) = \frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}\theta^y(1-\theta)^{n-y},$$

which is called Beta distribution

$$\theta|y,n \sim \text{Beta}(y+1, n-y+1)$$



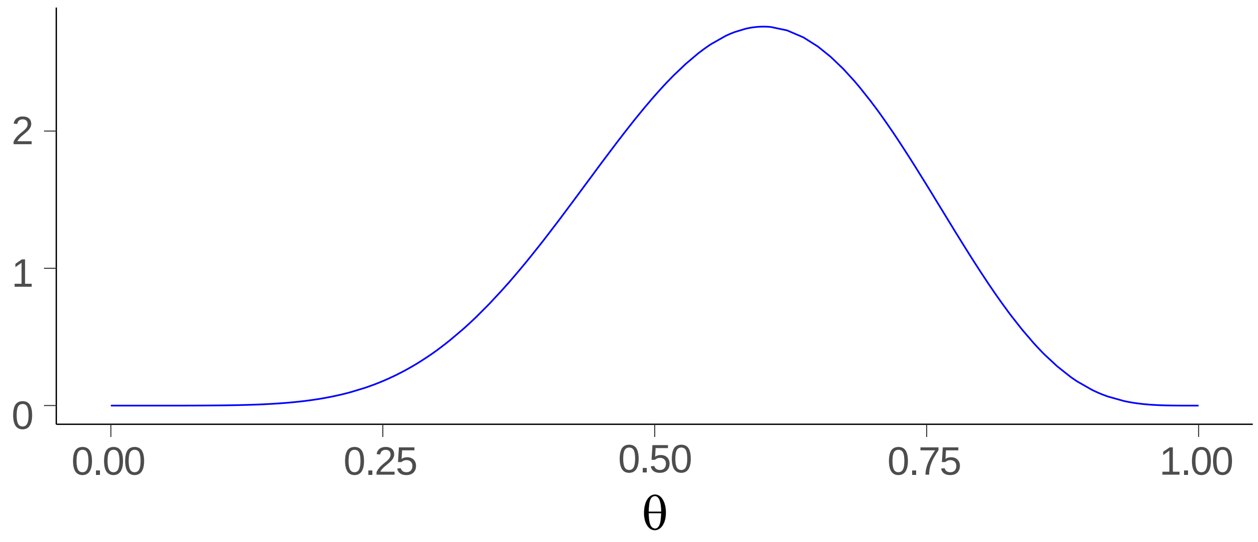
[Code demo with beta prior]

Binomial: computation

- R
  - density `dbeta`
  - CDF `pbeta`
  - quantile `qbeta`
  - random number `rbeta`
- Python
  - `from scipy.stats import beta`
  - density `beta.pdf`
  - CDF `beta.cdf`
  - prctile `beta.ppf`
  - random number `beta.rvs`

Binomial: computation

- Beta CDF not trivial to compute
- For example, `pbeta` in R uses a continued fraction with weighting factors and asymptotic expansion
- Laplace developed normal approximation (Laplace approximation), because he didn't know how to compute Beta CDF

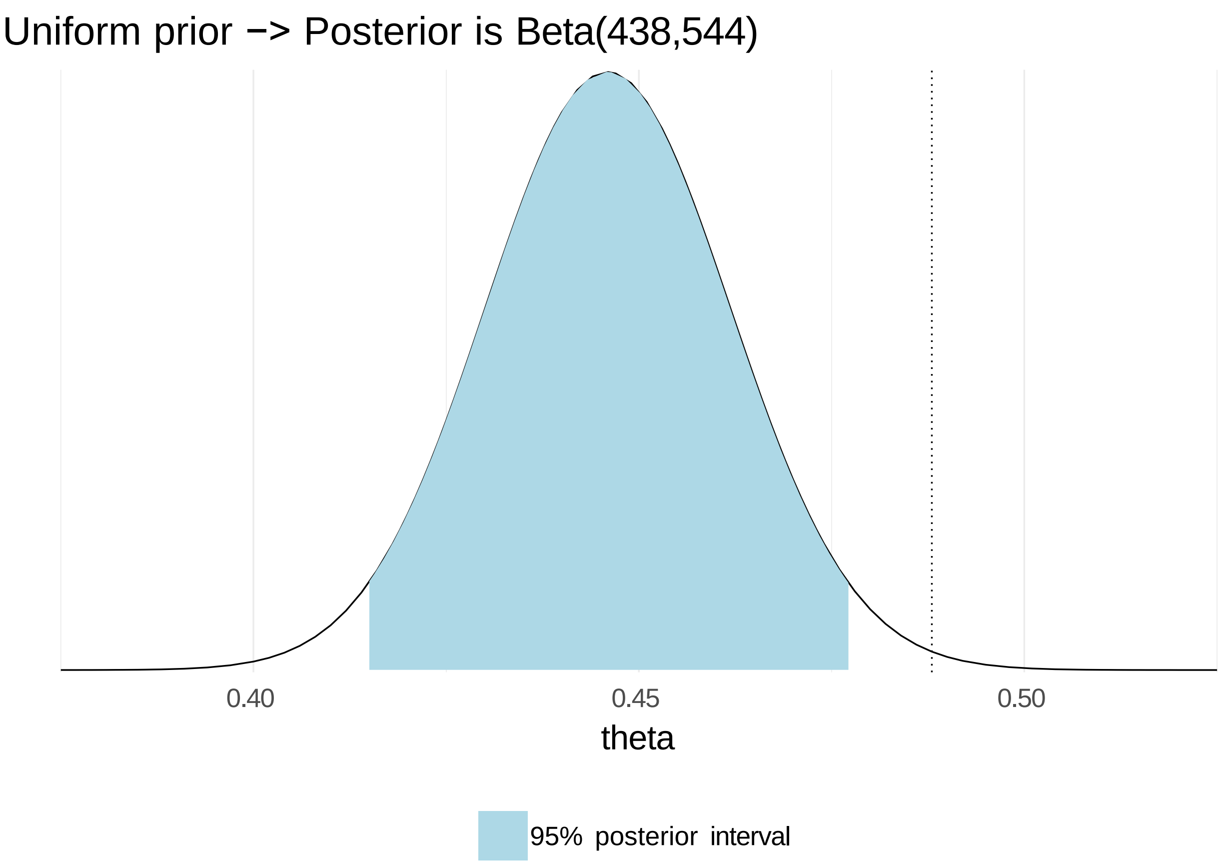


Placenta previa

- Probability of a girl birth given placenta previa (BDA3 p. 37)
  - 437 girls and 543 boys have been observed
  - is the ratio 0.445 different from the population average 0.485?

Placenta previa

- Probability of a girl birth given placenta previa (BDA3 p. 37)
  - 437 girls and 543 boys have been observed
  - is the ratio 0.445 different from the population average 0.485?



Some other one parameter models

- Poisson, useful for count data (e.g. in epidemiology)
- Exponential, useful for time to an event (e.g. particle decay)

Questions?