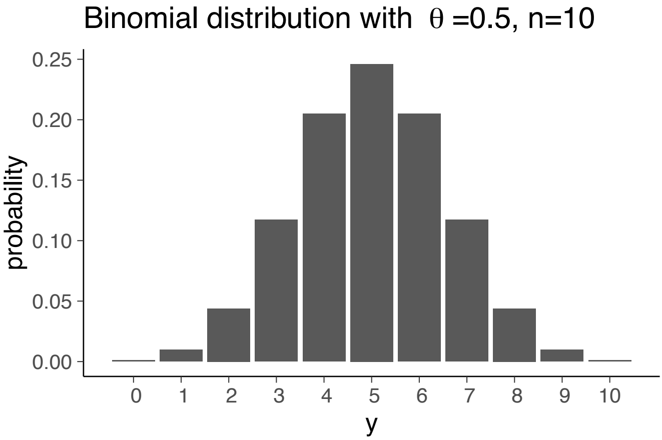
Extras for Chapter 2

- A bit more what is likelihood
- Why do we need the normalization term
- Plotting a continuous function
- Why probability density can be larger than 1
- Explicit conditioning on model M

Binomial: known θ

- Observation model (function of y, discrete)

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

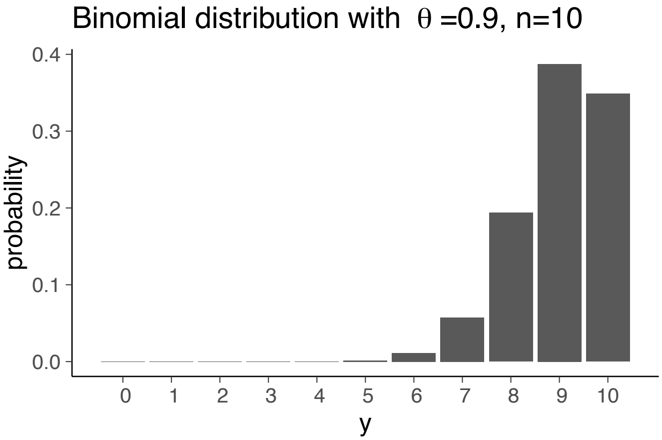


$p(y|n=10, \theta=0.5)$: 0.00 0.01 0.04 0.12 0.21 0.25 0.21 0.12 0.04 0.01 0.00

Binomial: known θ

- Observation model (function of y, discrete)

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

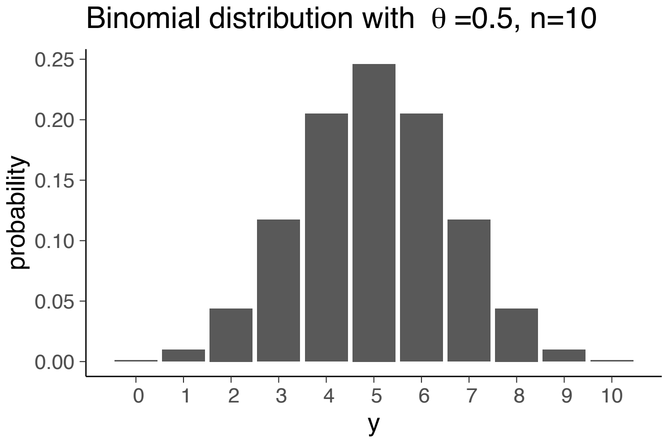


$p(y|n = 10, \theta = 0.9)$: 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.06 0.19 0.39 0.35

Binomial: known θ

- Observation model (function of y, discrete)

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

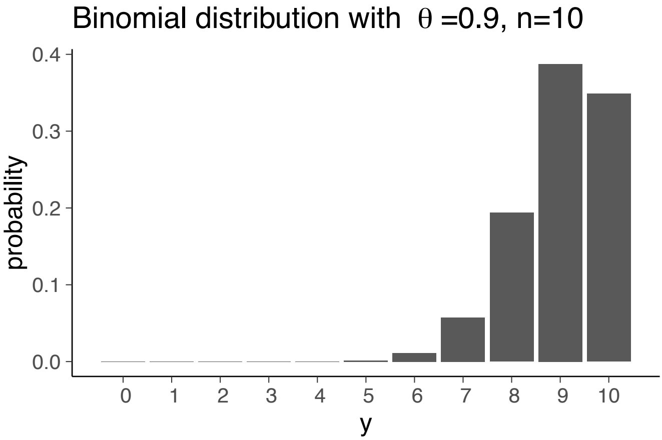


$p(y = 6|n = 10, \theta = 0.5)$: 0.00 0.01 0.04 0.12 0.21 0.25 **0.21** 0.12 0.04 0.01 0.00

Binomial: known θ

- Observation model (function of y, discrete)

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

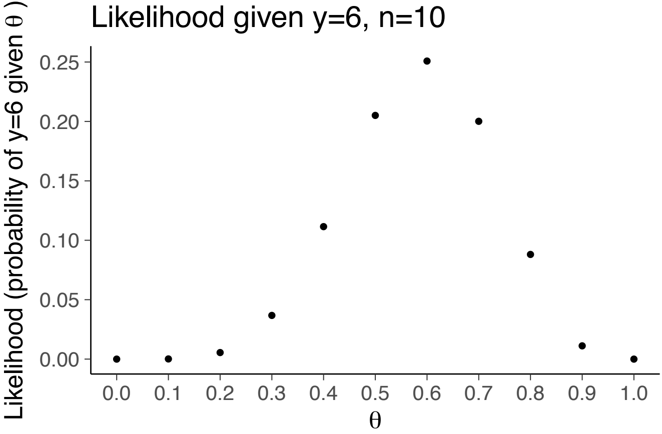


$p(y = 6|n = 10, \theta = 0.9)$: 0.00 0.00 0.00 0.00 0.00 0.00 **0.01** 0.06 0.19 0.39 0.35

Binomial: unknown θ

- Likelihood (function of θ, continuous)

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

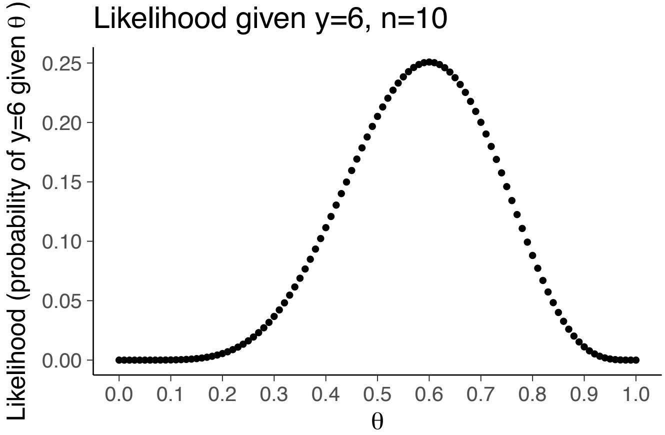


$p(y=6|n=10,\theta)$: 0.00 0.00 0.01 0.04 0.11 **0.21** 0.25 0.20 0.09 **0.01** 0.00

Binomial: unknown θ

- Likelihood (function of θ, continuous)

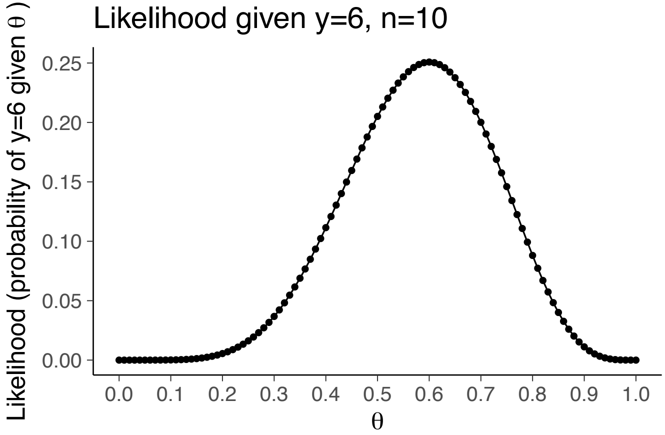
$$p(y|\theta, n) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$



Binomial: unknown θ

- Likelihood (function of θ, continuous)

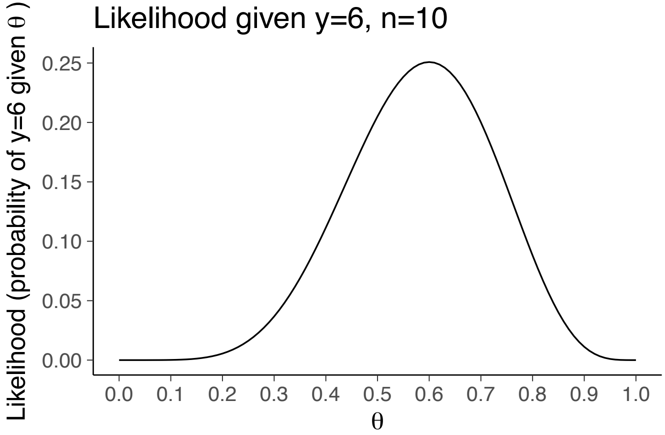
$$p(y|\theta, n) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$



Binomial: unknown θ

- Likelihood (function of θ, continuous)

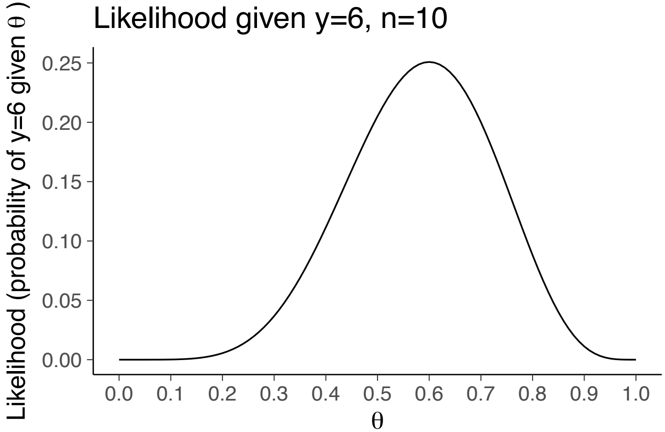
$$p(y|\theta, n) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$



Binomial: unknown θ

- Likelihood (function of θ, continuous)

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



`integrate(function(theta) dbinom(6, 10, theta), 0, 1)` ≈ 0.09

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

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

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

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

- 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)

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

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

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

$\approx$ 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)
```

$\approx$ 0.0004329

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

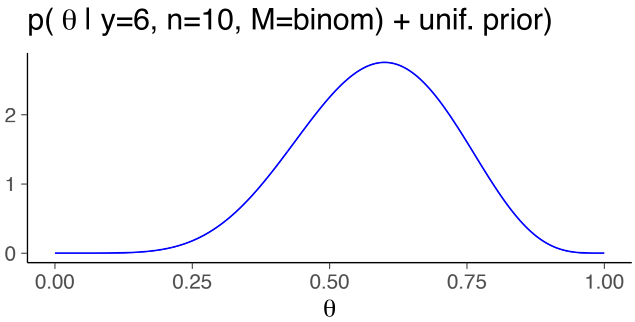
Binomial: unknown θ

- 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

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



Binomial: unknown θ

Sometimes conditioning on the model M is explicitly shown

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

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

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

Binomial: unknown θ

Sometimes conditioning on the model M is explicitly shown

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

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

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

- makes it more clear that likelihood and prior are both part of the model
- makes it more clear that there is no absolute probability for $p(y|n)$, but it depends on the model M
- in case of two models, we can evaluate marginal likelihoods $p(y|n, M_1)$ and $p(y|n, M_2)$ (more in Ch 7)

Binomial: unknown θ

Sometimes conditioning on the model M is explicitly shown

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

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

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

  - makes it more clear that likelihood and prior are both part of the model
  - makes it more clear that there is no absolute probability for $p(y|n)$, but it depends on the model M
  - in case of two models, we can evaluate marginal likelihoods $p(y|n, M_1)$ and $p(y|n, M_2)$ (more in Ch 7)
  - usually dropped to make the notation more concise