Prior Distributions I

Bayesian Modeling for Socio-Environmental Data

Mary B. Collins

August 2016

Outline

- Bayes Theorem
- Priors
- Conjugacy

Bayes Theorem

$$\underbrace{[\theta|y]}_{\text{Posterior}} = \frac{\overbrace{[y|\theta]}^{\text{Likelihood}} \; \overbrace{[\theta]}^{\text{Prior}}}{\underbrace{\int [y|\theta][\theta] d\theta}_{\text{Marginal}}}$$

- Likelihood: Links unobserved θ to observed y.

Bayes Theorem

$$\underbrace{[\theta|y]}_{\text{Posterior}} = \frac{\overbrace{[y|\theta]}^{\text{Likelihood}}\ \overbrace{[\theta]}^{\text{Prior}}}{\underbrace{\int [y|\theta][\theta]d\theta}_{\text{Marginal}}}$$

- Likelihood: Links unobserved θ to observed y.
- Prior Distribution: What is already known about θ.

Bayes Theorem

$$\underbrace{[\theta|y]}_{\text{Posterior}} = \frac{\overbrace{[y|\theta]}^{\text{Likelihood}}\ \overbrace{[\theta]}^{\text{Prior}}}{\underbrace{\int [y|\theta][\theta]d\theta}_{\text{Marginal}}}$$

- Likelihood: Links unobserved θ to observed y.
- Prior Distribution: What is already known about θ.
- Marginal Distribution: the area under the joint distribution curve. Serves to normalize the curve with respect to θ.

Bayes Theorem

$$\underbrace{[\theta|y]}_{\text{Posterior}} = \frac{\overbrace{[y|\theta]}^{\text{Likelihood}} \quad \overbrace{[\theta]}^{\text{Prior}}}{\underbrace{\int [y|\theta][\theta]d\theta}_{\text{Marginal}}}$$

- Likelihood: Links unobserved θ to observed y.
- Prior Distribution: What is already known about θ.
- Marginal Distribution: the area under the joint distribution curve. Serves to normalize the curve with respect to θ.
- Posterior Distribution: a true PDF.

Prior distributions

Prior distributions can be informative, reflecting knowledge gained in previous research *or* they can be vague, reflecting a lack of information about θ before data are collected.

Prior distributions

- In special cases the posterior, $[\theta|y]$, has the same form as the prior, $[\theta]$.

Prior distributions

- In special cases the posterior, $[\theta|y]$, has the same form as the prior, $[\theta]$.
- In these cases, the prior and the posterior are said to be *conjugate*.

Prior distributions

- In special cases the posterior, $[\theta|y]$, has the same form as the prior, $[\theta]$.
- In these cases, the prior and the posterior are said to be *conjugate*.
- Conjugacy is important for two reasons:

Prior distributions

- In special cases the posterior, $[\theta|y]$, has the same form as the prior, $[\theta]$.
- In these cases, the prior and the posterior are said to be *conjugate*.
- Conjugacy is important for two reasons:

1. Conjugacy in simple cases minimizes computational work and, in more complicated cases, allows us to break down calculations into manageable chunks.

Prior distributions

- In special cases the posterior, $[\theta|y]$, has the same form as the prior, $[\theta]$.
- In these cases, the prior and the posterior are said to be *conjugate*.
- Conjugacy is important for two reasons:

1. Conjugacy in simple cases minimizes computational work and, in more complicated cases, allows us to break down calculations into manageable chunks.

2. Conjugacy plays an important role in Markov chain Monte Carlo procedures (more on this later).

Conjugate priors

Table A.3: Table of conjugate distributions

Likelihood	Prior distribution	Posterior distribution
$y_i \sim$ binomial (n, ϕ)	$\phi \sim$ beta (α, β)	$\phi \sim$ beta $\left(\sum y_i + \alpha, n - \sum y_i + \beta\right)$
$y_i \sim$ Bernoulli(ϕ)	$\phi \sim$ beta (α, β)	$\phi \sim$ beta $\left(\sum_{i=1}^n y_i + \alpha, \sum_{i=1}^n (1 - y_i) + \beta\right)$
$y_i \sim$ Poisson (λ)	$\lambda \sim$ gamma (α, β)	$\lambda \sim$ gamma $\left(\alpha + \sum_{i=1}^n y_i,\ \beta + n\right)$
$y_i \sim$ normal(μ, σ^2) σ^2 is known.	$\mu \sim$ normal (μ_0, σ_0^2)	$\mu \sim$ normal $\left(\frac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n y_i}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)}, \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1} \right)$
$y_i \sim$ normal(μ, σ^2) μ is known.	$\sigma^2 \sim$ inverse gamma(α, β)	$\sigma^2 \sim$ inverse gamma $\left(\alpha + \frac{n}{2}, \beta + \frac{\sum_{i=1}^n (y_i - \mu)^2}{2}\right)$
$y_i \sim$ lognormal(μ, σ^2) , μ is known	$\sigma^2 \sim$ inverse gamma(α, β) ,	$\sigma^2 \sim$ inverse gamma $\left(n/2 + \alpha, \frac{(\log(y_i) - \mu)^2}{2} + \beta\right)$
$y_i \sim$ lognormal (μ, σ^2) σ^2 is known	$\mu \sim$ normal (μ_0, σ_0^2)	$\mu \sim$ normal $\left(\frac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n \log y_i}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)}, \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1} \right)$

(Hol

and Hooten, 2015)

Just a bit more on priors

- There is no such thing as a truly non-informative prior—only those that influence the posterior more (or less) than others.

Just a bit more on priors

- There is no such thing as a truly non-informative prior—only those that influence the posterior more (or less) than others.
- Informative priors can be justified and useful.

Just a bit more on priors

- There is no such thing as a truly non-informative prior—only those that influence the posterior more (or less) than others.
- Informative priors can be justified and useful.
- “Non-informative," vague, or flat priors are provisional starting points for a Bayesian analysis.

Example: childhood asthma and PM

- You are studying the relationship between childhood asthma and industrial airborne PM.
- In one school 17 of 80 students have been hospitalized for asthma-related issues.

1. What distributions would you choose for the likelihood and the prior?
2. How would you draw the DAG?

Picking distributions

Prior: $([\phi])$

- continuous quantity ranging from 0 to 1

Likelihood: $([\boldsymbol{y}|\phi])$

- count data: $y = 17$ successes, given $n = 80$ trials

Posterior: $([\phi|\boldsymbol{y}])$

- Is there a conjugate to use?

Picking distributions

Prior: $([\phi])$

- continuous quantity ranging from 0 to 1.
- Uniform beta with parameters, $\alpha_{prior} = 1, \beta_{prior} = 1$
- $\phi \sim \text{beta}(1, 1)$

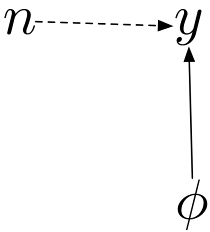
Likelihood: $([\boldsymbol{y}|\phi])$

- binomial distribution with $y = 17$ successes, given $n = 80$ trials

Posterior: $([\phi|\boldsymbol{y}])$

- Using the beta-binomial conjugate prior relationship
- $\phi \sim \text{beta}(\alpha_{post}, \beta_{post})$

Drawing the DAG



Writing out the full posterior

$$\text{beta}(\alpha_{post}, \beta_{post}) = \frac{\text{binomial}(y|\phi, n)\,\text{beta}(\alpha_{prior}, \beta_{prior})}{[y]}$$

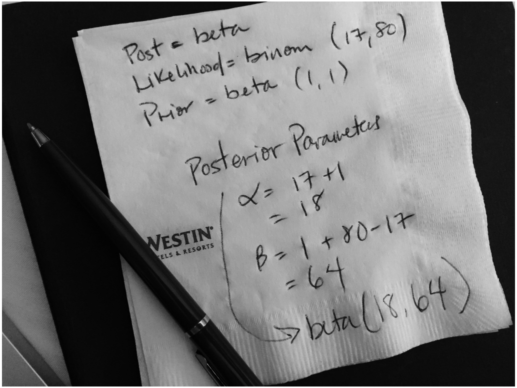
Posterior distribution parameters

$$\alpha_{post} = \Sigma_{y_i} + \alpha_{prior}$$
$$\beta_{post} = n - \Sigma_{y_i} + \beta_{prior}$$

This means you can...

This means you can...

...literaly calcluate the parameters of the posterior distribution on the back of a hotel napkin.



Posterior distribution parameters

```
aPrior <- 1
bPrior <- 1
y <- 17
n <- 80
aPost <- aPrior + y
aPost
```

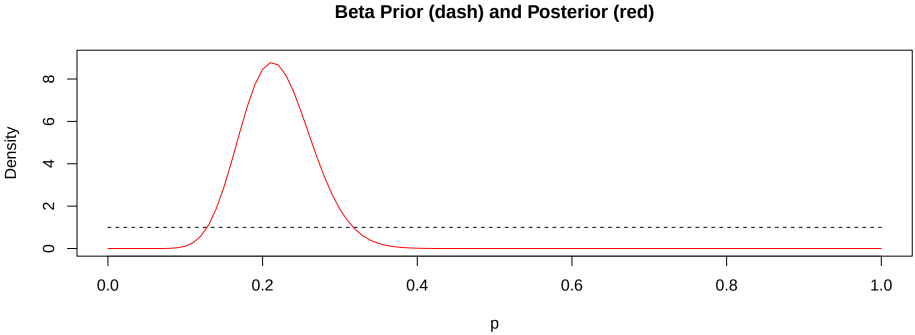
```
## [1] 18
```

```
bPost <- bPrior + n - y
bPost
```

```
## [1] 64
```

$\phi \sim \text{beta}(18, 64)$

Plotting the prior and posterior



Between which quantiles does ϕ lie with probability 0.95?

```
## [1] 0.1373419 0.3146275
```

Between which quantiles does ϕ lie with probability 0.95?

```
## [1] 0.1391926 0.2630073
```

What if, instead, the new study showed a much lower incidence of hospitalization, rather than closer finding (e.g. 3 of 75 hospitalized)?

```r
aPrior <- 18
bPrior <- 64
y <- 3
n <- 75
aPost <- aPrior + y
aPost
```

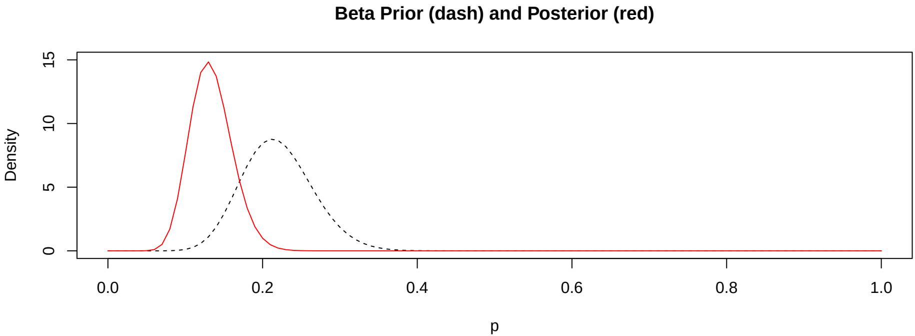
```
## [1] 21
```

```r
bPost <- bPrior + n - y
bPost
```

```
## [1] 136
```

$\phi \sim \text{beta}(21, 136)$

Plotting the prior and posterior



Between which quantiles does ϕ lie with probability 0.95?

```
## [1] 0.08529956 0.19103790
```

What if, instead, the new study showed a much higher incidence of hospitalization, rather than closer finding (e.g. 80 of 80 hospitalized)?

```r
aPrior <- 18
bPrior <- 64
y <- 80
n <- 80
aPost <- aPrior + y
aPost
```

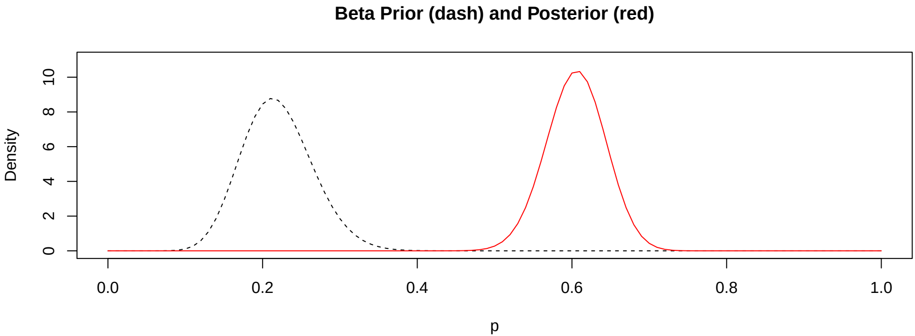
```
## [1] 98
```

```r
bPost <- bPrior + n - y
bPost
```

```
## [1] 64
```

$\phi \sim \text{beta}(98, 64)$

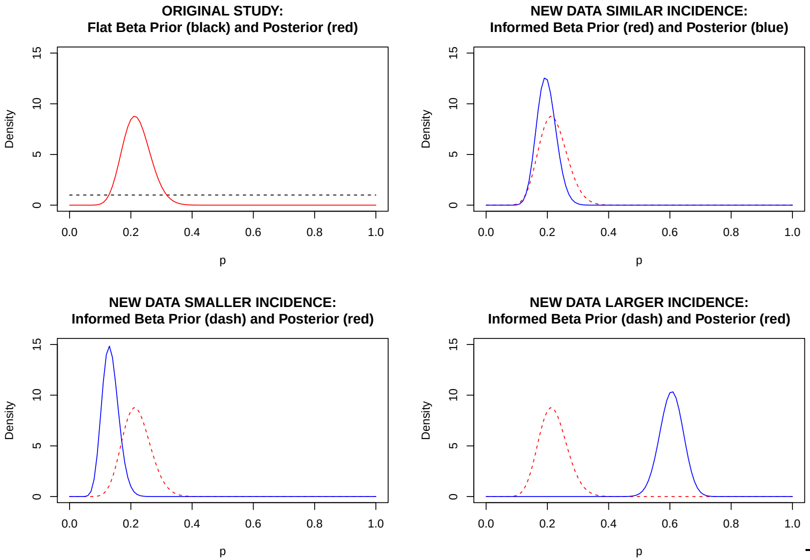
Plotting the prior and posterior



Between which quantiles does ϕ lie with probability 0.95?

```
## [1] 0.5287686 0.6786532
```

Exploring the role of new data

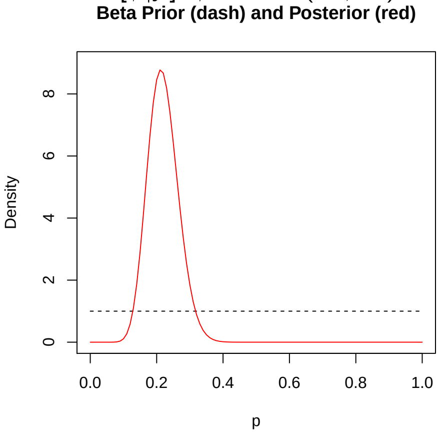


−> ##

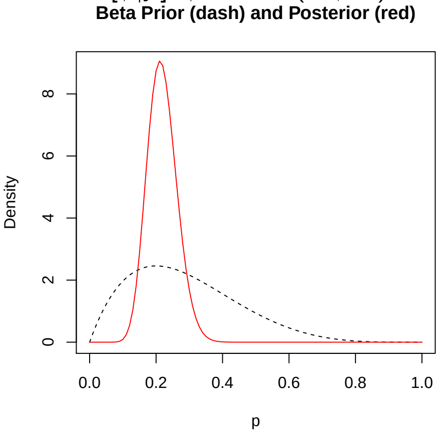
All priors and data influence on the posterior distribution

Comparing priors

Prior $[\phi]$: $\phi \sim \text{beta}(1,1)$
Post $[\phi|y]$:$\phi \sim \text{beta}(18,64)$



Prior $[\phi]$: $\phi \sim \text{beta}(2,5)$
Post $[\phi|y]$:$\phi \sim \text{beta}(19,68)$



Now consider the following data:

School: Hospitalizations/Total Students
School 1: 17/80
School 2: 17/75
School 3: 19/100
School 4: 10/55
School 5: 33/111

For a childhood asthma hospitalization model that, uses the new data across schools, what is the DAG?

For a childhood asthma hospitalization model that, uses the new data across schools, what is the DAG?

$$\begin{tikzcd} n_i \arrow[r, dashed] & y_i \\ & \phi_i \arrow[u] \\ & \alpha, \beta \arrow[u] \end{tikzcd}$$

Full posterior

$$
\begin{aligned}
y_i &\sim \text{binomial}(n_i, \phi_i) \\
\phi_i &\sim \text{beta}(\alpha, \beta) \\
\alpha &\sim \text{uniform}(0, 500) \\
\beta &\sim \text{uniform}(0, 500)
\end{aligned}
$$

$$
[\boldsymbol{\phi}, \alpha, \beta | \mathbf{y}] \propto \prod_{i=1}^{5} [y_i | \phi_i, n_i][\phi_i | \alpha, \beta][\alpha][\beta]
$$

This is a hierarchical model because parameter ϕ, is on both sides of the conditionning symbol.

–>

Things to remember

- There is no such thing as a noninformative prior, but certain priors influence the posterior distribution more than others.

Things to remember

- There is no such thing as a noninformative prior, but certain priors influence the posterior distribution more than others.
- Informative priors, when properly justified, can be useful.

Things to remember

- There is no such thing as a noninformative prior, but certain priors influence the posterior distribution more than others.
- Informative priors, when properly justified, can be useful.
- Strong data overwhelms a prior.

Role of priors in science

- Priors represent our current knowledge (or lack of current knowledge), which is updated with data.

Role of priors in science

- Priors represent our current knowledge (or lack of current knowledge), which is updated with data.
- We encourage you to think of vague priors as a provisional starting point.