Bayesian Data Analysis, class 2a

Andrew Gelman

Chapter 2. Single-parameter models (part 1)

2. Single-parameter models (part 1)

- The basics:
  - Data model (thus, likelihood)
  - Prior density
  - Posterior density
- Work with analytic conjugate forms
- Likelihood comes before the prior; why?

2. Single-parameter models (part 1)

- The basics:
  - Data model (thus, likelihood)
  - Prior density
  - Posterior density
- Work with analytic conjugate forms
- Likelihood comes before the prior; why?

2. Single-parameter models (part 1)

- The basics:
  - Data model (thus, likelihood)
  - Prior density
  - Posterior density
- Work with analytic conjugate forms
- Likelihood comes before the prior; why?

2. Single-parameter models (part 1)

- The basics:
  - Data model (thus, likelihood)
  - Prior density
  - Posterior density
- Work with analytic conjugate forms
- Likelihood comes before the prior; why?

2. Single-parameter models (part 1)

- The basics:
  - Data model (thus, likelihood)
  - Prior density
  - Posterior density
- Work with analytic conjugate forms
- Likelihood comes before the prior; why?

2. Single-parameter models (part 1)

- The basics:
  - Data model (thus, likelihood)
  - Prior density
  - Posterior density
- Work with analytic conjugate forms
- Likelihood comes before the prior; why?

2. Single-parameter models (part 1)

- The basics:
  - Data model (thus, likelihood)
  - Prior density
  - Posterior density
- Work with analytic conjugate forms
- Likelihood comes before the prior; why?

2.1. Estimating a probability from binomial data

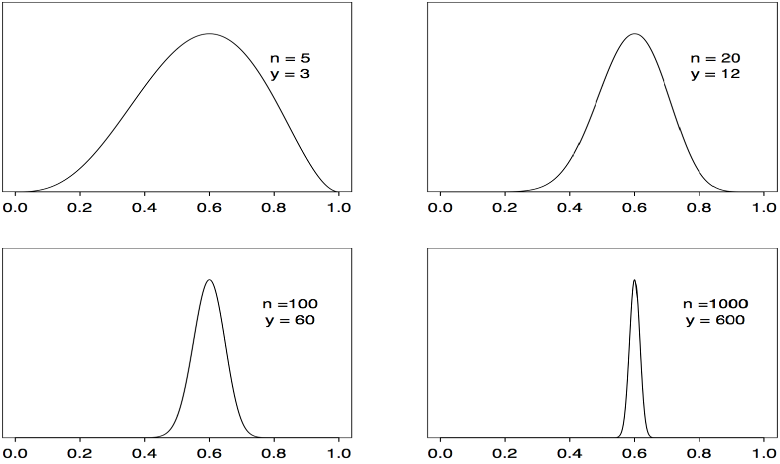


Figure 2.1 *Unnormalized posterior density for binomial parameter θ, based on uniform prior distribution and y successes out of n trials. Curves displayed for several values of n and y.*

2.1. Estimating a probability from binomial data

- Bayes' original example
- $\frac{y}{n}$, $\frac{y+1}{n+2}$, and other alternatives
- Consider some examples

- Dependence of θ and n

2.1. Estimating a probability from binomial data

- Bayes' original example
- $\frac{y}{n}$, $\frac{y+1}{n+2}$, and other alternatives
- Consider some examples
  - Opinion polls
  - Sampling of financial records
  - Rare diseases
- Dependence of θ and n

2.1. Estimating a probability from binomial data

- Bayes' original example
- $\frac{y}{n}$, $\frac{y+1}{n+2}$, and other alternatives
- Consider some examples
  - Opinion polls
  - Sampling of financial records
  - Rare diseases
- Dependence of θ and n

2.1. Estimating a probability from binomial data

- Bayes' original example
- $\frac{y}{n}$, $\frac{y+1}{n+2}$, and other alternatives
- Consider some examples
  - Opinion polls
  - Sampling of financial records
  - Rare diseases
- Dependence of θ and n

2.1. Estimating a probability from binomial data

- Bayes' original example
- $\frac{y}{n}$, $\frac{y+1}{n+2}$, and other alternatives
- Consider some examples
  - Opinion polls
  - Sampling of financial records
  - Rare diseases
- Dependence of θ and n

2.1. Estimating a probability from binomial data

- Bayes' original example
- $\frac{y}{n}$, $\frac{y+1}{n+2}$, and other alternatives
- Consider some examples
  - Opinion polls
  - Sampling of financial records
  - Rare diseases
- Dependence of θ and n

2.1. Estimating a probability from binomial data

- Bayes' original example
- $\frac{y}{n}$, $\frac{y+1}{n+2}$, and other alternatives
- Consider some examples
  - Opinion polls
  - Sampling of financial records
  - Rare diseases
- Dependence of θ and n

2.1. Estimating a probability from binomial data

- Bayes' original example
- $\frac{y}{n}$, $\frac{y+1}{n+2}$, and other alternatives
- Consider some examples
  - Opinion polls
  - Sampling of financial records
  - Rare diseases
- Dependence of θ and n

2.2. Posterior as compromise between data and prior information

- The posterior variance is on average smaller than the prior variance
- When does it happen that posterior variance is *larger* than the prior variance?
  - Bad luck
  - Bad model
- Research problem: when do models have "warning lights"?

2.2. Posterior as compromise between data and prior information

- The posterior variance is on average smaller than the prior variance
- When does it happen that posterior variance is *larger* than the prior variance?
  - Bad luck
  - Bad model
- Research problem: when do models have "warning lights"?

2.2. Posterior as compromise between data and prior information

- The posterior variance is on average smaller than the prior variance
- When does it happen that posterior variance is *larger* than the prior variance?
  - Bad luck
  - Bad model
- Research problem: when do models have "warning lights"?

2.2. Posterior as compromise between data and prior information

- The posterior variance is on average smaller than the prior variance
- When does it happen that posterior variance is *larger* than the prior variance?
  - Bad luck
  - Bad model
- Research problem: when do models have "warning lights"?

2.2. Posterior as compromise between data and prior information

- The posterior variance is on average smaller than the prior variance
- When does it happen that posterior variance is *larger* than the prior variance?
  - Bad luck
  - Bad model
- Research problem: when do models have "warning lights"?

2.2. Posterior as compromise between data and prior information

- The posterior variance is on average smaller than the prior variance
- When does it happen that posterior variance is *larger* than the prior variance?
  - Bad luck
  - Bad model
- Research problem: when do models have "warning lights"?

2.3. Summarizing posterior inference

- Simulation draws
- Simulations + analytics
- Posterior intervals

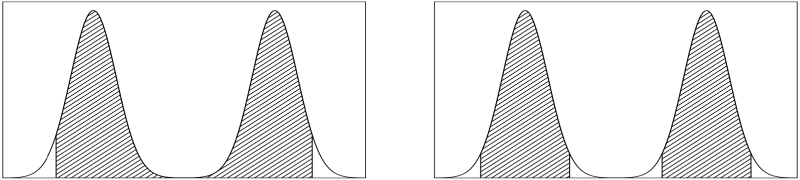


Figure 2.2 *Hypothetical posterior density for which the 95% central interval and 95% highest posterior density region dramatically differ: (a) central posterior interval, (b) highest posterior density region.*

2.3. Summarizing posterior inference

- Simulation draws
- Simulations + analytics
- Posterior intervals

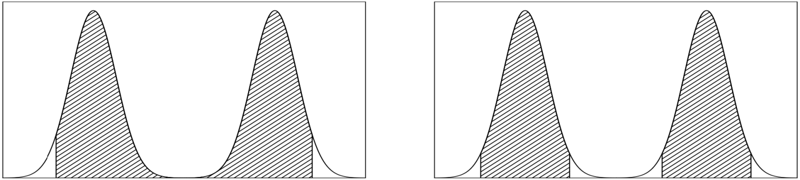


Figure 2.2 *Hypothetical posterior density for which the 95% central interval and 95% highest posterior density region dramatically differ: (a) central posterior interval, (b) highest posterior density region.*

2.3. Summarizing posterior inference

- Simulation draws
- Simulations + analytics
- Posterior intervals

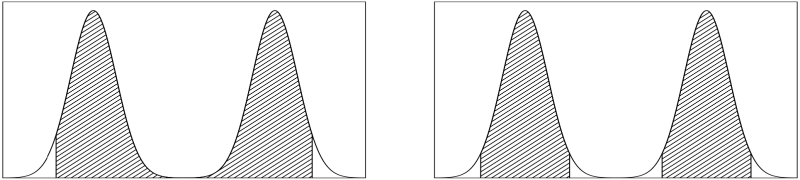


Figure 2.2 *Hypothetical posterior density for which the 95% central interval and 95% highest posterior density region dramatically differ: (a) central posterior interval, (b) highest posterior density region.*

2.3. Summarizing posterior inference

- Simulation draws
- Simulations + analytics
- Posterior intervals

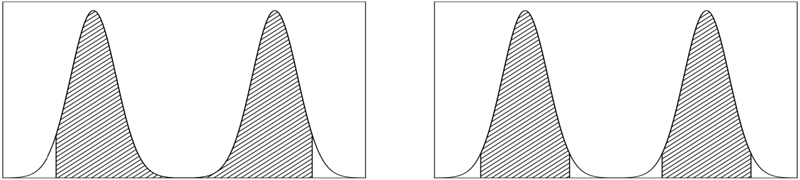
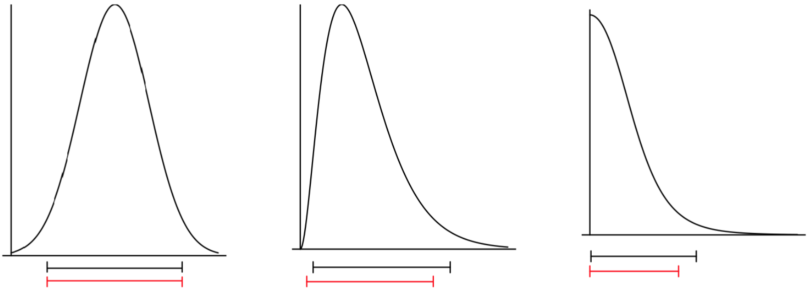
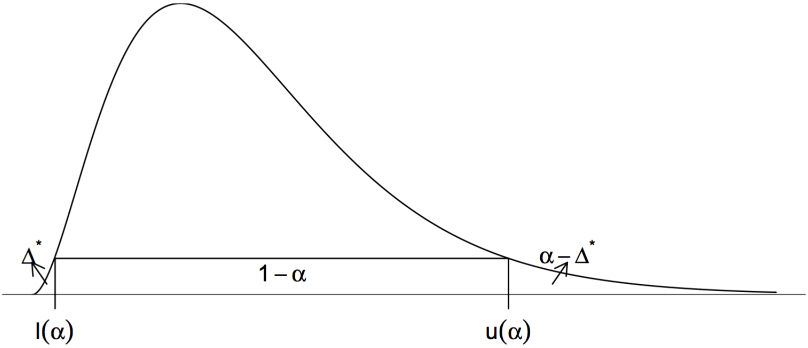


Figure 2.2 *Hypothetical posterior density for which the 95% central interval and 95% highest posterior density region dramatically differ: (a) central posterior interval, (b) highest posterior density region.*

Central and shortest intervals



Shortest posterior intervals



Posterior intervals: choices and goals

- Central intervals
- Highest posterior density intervals
- Nested or non-nested?
- Asymmetry and multiple modes
- Multidimensional contours
- The purpose of interval estimation

Posterior intervals: choices and goals

- Central intervals
- Highest posterior density intervals
- Nested or non-nested?
- Asymmetry and multiple modes
- Multidimensional contours
- The purpose of interval estimation

Posterior intervals: choices and goals

- Central intervals
- Highest posterior density intervals
- Nested or non-nested?
- Asymmetry and multiple modes
- Multidimensional contours
- The purpose of interval estimation

Posterior intervals: choices and goals

- Central intervals
- Highest posterior density intervals
- Nested or non-nested?
- Asymmetry and multiple modes
- Multidimensional contours
- The purpose of interval estimation

Posterior intervals: choices and goals

- Central intervals
- Highest posterior density intervals
- Nested or non-nested?
- Asymmetry and multiple modes
- Multidimensional contours
- The purpose of interval estimation
  - Accepting what's in the interval
  - Rejecting what's outside the interval

Posterior intervals: choices and goals

- Central intervals
- Highest posterior density intervals
- Nested or non-nested?
- Asymmetry and multiple modes
- Multidimensional contours
- The purpose of interval estimation
  - Accepting what's in the interval
  - Rejecting what's outside the interval

Posterior intervals: choices and goals

- Central intervals
- Highest posterior density intervals
- Nested or non-nested?
- Asymmetry and multiple modes
- Multidimensional contours
- The purpose of interval estimation
  - Accepting what's in the interval
  - Rejecting what's outside the interval

Posterior intervals: choices and goals

- Central intervals
- Highest posterior density intervals
- Nested or non-nested?
- Asymmetry and multiple modes
- Multidimensional contours
- The purpose of interval estimation
  - Accepting what's in the interval
  - Rejecting what's outside the interval

Posterior intervals: choices and goals

- Central intervals
- Highest posterior density intervals
- Nested or non-nested?
- Asymmetry and multiple modes
- Multidimensional contours
- The purpose of interval estimation
  - Accepting what's in the interval
  - Rejecting what's outside the interval

Constructing shortest posterior intervals from simulations

- Empirical Spin
- Smoothing
- Boostrapping
- Computational tradeoffs

Constructing shortest posterior intervals from simulations

- Empirical Spin
- Smoothing
- Boostrapping
- Computational tradeoffs

Constructing shortest posterior intervals from simulations

- Empirical Spin
- Smoothing
- Boostrapping
- Computational tradeoffs

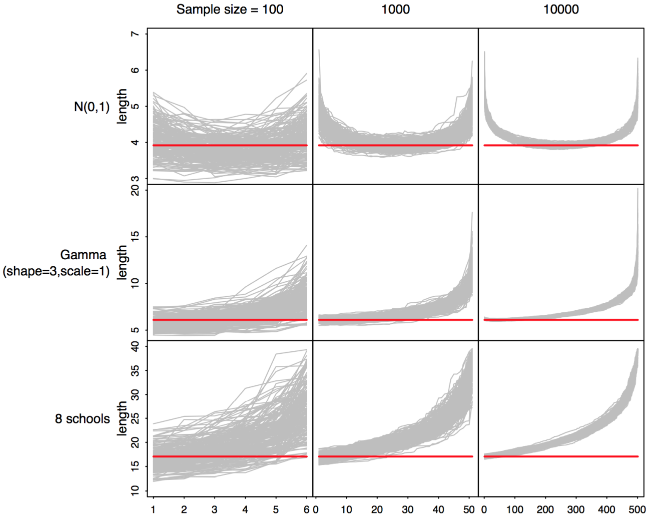
Constructing shortest posterior intervals from simulations

- Empirical Spin
- Smoothing
- Boostrapping
- Computational tradeoffs

Constructing shortest posterior intervals from simulations

- Empirical Spin
- Smoothing
- Boostrapping
- Computational tradeoffs

Sampling variability of the empirical shortest posterior interval



2.4. Informative prior distributions

- Interpretations
  - Population
  - State of knowledge
  - Software defaults (statistician in a box)
- Binomial model:
- Conjugate and non-conjugate priors

2.4. Informative prior distributions

- Interpretations
  - Population
  - State of knowledge
  - Software defaults (statistician in a box)
- Binomial model:
- Conjugate and non-conjugate priors

2.4. Informative prior distributions

- Interpretations
  - Population
  - State of knowledge
  - Software defaults (statistician in a box)
- Binomial model:
- Conjugate and non-conjugate priors

2.4. Informative prior distributions

- Interpretations
  - Population
  - State of knowledge
  - Software defaults (statistician in a box)
- Binomial model:
  - Beta(α, β) prior as eqivalent to $\alpha + \beta$ data points
    Or $\alpha + \beta - 2$ data points?
- Conjugate and non-conjugate priors

2.4. Informative prior distributions

- Interpretations
  - Population
  - State of knowledge
  - Software defaults (statistician in a box)
- Binomial model:
  - Beta(α, β) prior as eqivalent to $\alpha + \beta$ data points
  - Or $\alpha + \beta - 2$ data points?
- Conjugate and non-conjugate priors

2.4. Informative prior distributions

- Interpretations
  - Population
  - State of knowledge
  - Software defaults (statistician in a box)
- Binomial model:
  - Beta(α, β) prior as eqivalent to $\alpha + \beta$ data points
  - Or $\alpha + \beta - 2$ data points?
- Conjugate and non-conjugate priors

2.4. Informative prior distributions

- Interpretations
  - Population
  - State of knowledge
  - Software defaults (statistician in a box)
- Binomial model:
  - Beta(α, β) prior as eqivalent to $\alpha + \beta$ data points
  - Or $\alpha + \beta - 2$ data points?
- Conjugate and non-conjugate priors

2.4. Informative prior distributions

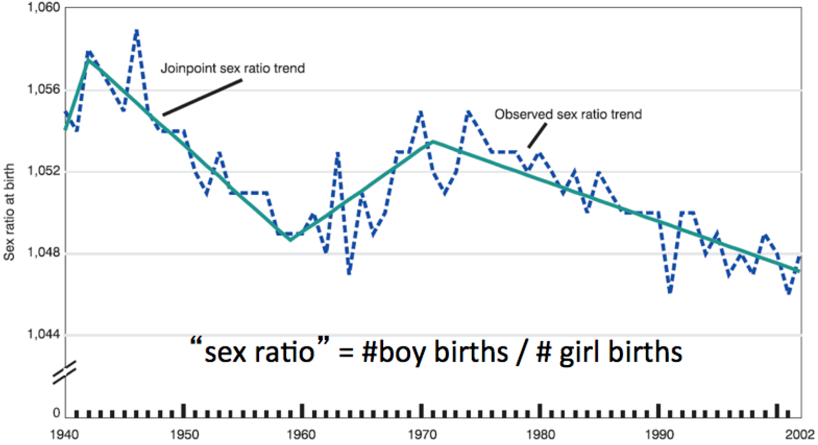
- Interpretations
  - Population
  - State of knowledge
  - Software defaults (statistician in a box)
- Binomial model:
  - Beta(α, β) prior as eqivalent to $\alpha + \beta$ data points
  - Or $\alpha + \beta - 2$ data points?
- Conjugate and non-conjugate priors

2.4. Informative prior distributions

- Interpretations
  - Population
  - State of knowledge
  - Software defaults (statistician in a box)
- Binomial model:
  - Beta(α, β) prior as eqivalent to $\alpha + \beta$ data points
  - Or $\alpha + \beta - 2$ data points?
- Conjugate and non-conjugate priors

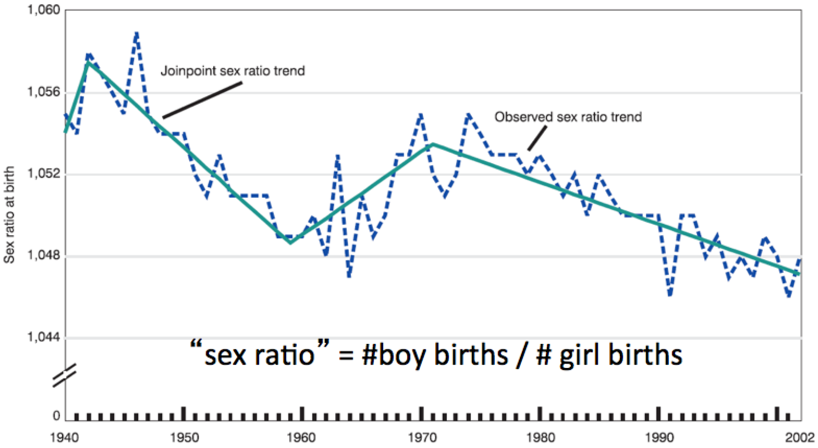
Probability of a girl birth

- In general population, Pr(girl birth) = 0.485

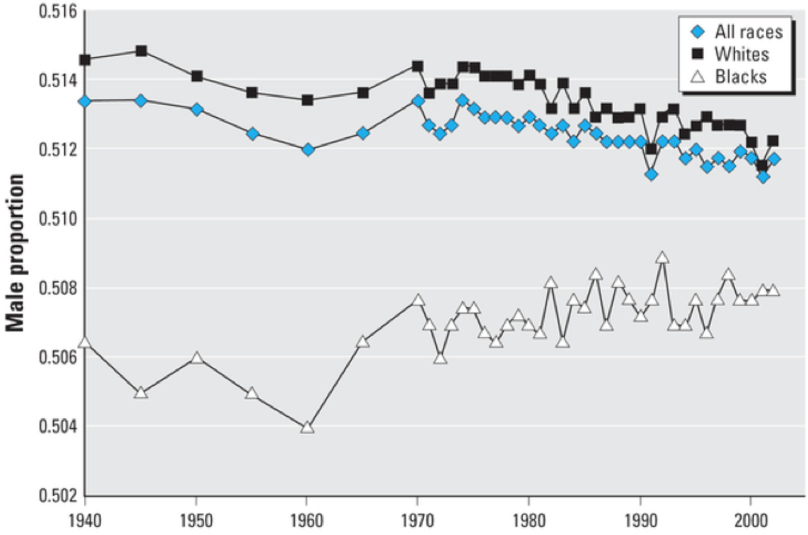


Probability of a girl birth

- In general population, Pr(girl birth) $= 0.485$



Variation!



All races
Whites
Blacks
Male proportion
0.516
0.514
0.512
0.510
0.508
0.506
0.504
0.502
1940
1950
1960
1970
1980
1990
2000

Example: probability of a girl birth given placenta previia

- Data: 437 girls out of 980 births, $y/n = 437/980 = 0.446$
- Consider different Beta(α, β) priors

$\frac{\alpha}{\alpha+\beta}$	$\alpha + \beta$	Posterior median of θ	95% posterior interval for θ
0.500	2	0.446	$[0.415, 0.477]$
0.485	2	0.446	$[0.415, 0.477]$
0.485	5	0.446	$[0.415, 0.477]$
0.485	10	0.446	$[0.415, 0.477]$
0.485	20	0.447	$[0.416, 0.478]$
0.485	100	0.450	$[0.420, 0.479]$
0.485	200	0.453	$[0.424, 0.481]$

Example: probability of a girl birth given placenta previia

- Data: 437 girls out of 980 births, $y/n = 437/980 = 0.446$
- Consider different Beta(α, β) priors

$\frac{\alpha}{\alpha+\beta}$	$\alpha + \beta$	Posterior median of θ	95% posterior interval for θ
0.500	2	0.446	[0.415, 0.477]
0.485	2	0.446	[0.415, 0.477]
0.485	5	0.446	[0.415, 0.477]
0.485	10	0.446	[0.415, 0.477]
0.485	20	0.447	[0.416, 0.478]
0.485	100	0.450	[0.420, 0.479]
0.485	200	0.453	[0.424, 0.481]

Example: probability of a girl birth given placenta previia

- Data: 437 girls out of 980 births, $y/n = 437/980 = 0.446$
- Consider different Beta(α, β) priors

$\frac{\alpha}{\alpha+\beta}$	$\alpha+\beta$	Posterior median of θ	95% posterior interval for θ
0.500	2	0.446	$[0.415, 0.477]$
0.485	2	0.446	$[0.415, 0.477]$
0.485	5	0.446	$[0.415, 0.477]$
0.485	10	0.446	$[0.415, 0.477]$
0.485	20	0.447	$[0.416, 0.478]$
0.485	100	0.450	$[0.420, 0.479]$
0.485	200	0.453	$[0.424, 0.481]$

Constructing a prior distribution

- $p =$ probability of a girl birth given placenta previa
- Suppose we believe p should be between 0.47 and 0.50
  - Beta(α, β) prior with mean 0.485 and sd 0.01
    Need $\frac{\alpha}{\alpha+\beta} = 0.485$ and $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0.01^2$
    Solve to get $\alpha + \beta = 2500$ (actually, 2496.75) and $\alpha = 1212.5$, $\beta = 1287.5$
- Combine with $n = 980$ data points

Constructing a prior distribution

- $p =$ probability of a girl birth given placenta previa
- Suppose we believe p should be between 0.47 and 0.50
  - Beta(α, β) prior with mean 0.485 and sd 0.01
  - Need $\frac{\alpha}{\alpha+\beta} = 0.485$ and $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0.01^2$
  - Solve to get $\alpha + \beta = 2500$ (actually, 2496.75) and $\alpha = 1212.5$, $\beta = 1287.5$
- Combine with $n = 980$ data points

Constructing a prior distribution

- $p =$ probability of a girl birth given placenta previa
- Suppose we believe p should be between 0.47 and 0.50
  - Beta(α, β) prior with mean 0.485 and sd 0.01
  - Need $\frac{\alpha}{\alpha+\beta} = 0.485$ and $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0.01^2$
  - Solve to get $\alpha+\beta = 2500$ (actually, 2496.75) and $\alpha = 1212.5$, $\beta = 1287.5$
- Combine with $n = 980$ data points

Constructing a prior distribution

- p = probability of a girl birth given placenta previa
- Suppose we believe p should be between 0.47 and 0.50
  - Beta(α, β) prior with mean 0.485 and sd 0.01
  - Need $\frac{\alpha}{\alpha+\beta} = 0.485$ and $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0.01^2$
  - Solve to get $\alpha + \beta = 2500$ (actually, 2496.75) and $\alpha = 1212.5$, $\beta = 1287.5$
- Combine with $n = 980$ data points

Constructing a prior distribution

- p = probability of a girl birth given placenta previa
- Suppose we believe p should be between 0.47 and 0.50
  - Beta(α, β) prior with mean 0.485 and sd 0.01
  - Need $\frac{\alpha}{\alpha+\beta} = 0.485$ and $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0.01^2$
  - Solve to get $\alpha + \beta = 2500$ (actually, 2496.75) and $\alpha = 1212.5$, $\beta = 1287.5$
- Combine with $n = 980$ data points

Constructing a prior distribution

- p = probability of a girl birth given placenta previa
- Suppose we believe p should be between 0.47 and 0.50
  - Beta(α, β) prior with mean 0.485 and sd 0.01
  - Need $\frac{\alpha}{\alpha+\beta} = 0.485$ and $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0.01^2$
  - Solve to get $\alpha + \beta = 2500$ (actually, 2496.75) and $\alpha = 1212.5$, $\beta = 1287.5$
- Combine with $n = 980$ data points

Constructing a prior distribution

- $p =$ probability of a girl birth given placenta previa
- Suppose we believe p should be between 0.47 and 0.50
  - Beta(α, β) prior with mean 0.485 and sd 0.01
  - Need $\frac{\alpha}{\alpha+\beta} = 0.485$ and $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0.01^2$
  - Solve to get $\alpha + \beta = 2500$ (actually, 2496.75) and $\alpha = 1212.5$, $\beta = 1287.5$
- Combine with $n = 980$ data points