Beta-Binomial Inference, Part 2

STAT 251, Unit 4B

Overview

Review

Beta-Binomial Inference

Analytically deriving the posterior distribution of θ

Prior Elicitation for the Beta-Binomial setting

Review

Big picture outline for process in remainder of course

- Select likelihood (model)
- Select prior
- Determine posterior distribution
- Make inference with posterior distribution.

Notation

Notation: We will use

- $\mathbf{y}$ to generically represent the *data*,
- $\boldsymbol{\theta}$ to generically represent the *parameters*,
- $f(\mathbf{y}|\boldsymbol{\theta})$ to represent the *likelihood* of the data given the parameters
- $\pi(\boldsymbol{\theta})$ to represent the *prior distribution* of the parameters
- $\pi(\boldsymbol{\theta}|\mathbf{y})$ to represent the *posterior distribution* of the parameters given the observed data.
- $f(\mathbf{y})$ to represent the *marginal likelihood* of the data—note the marginal likelihood is NOT a function of θ

Key Points from Unit 4A

Familiarity with Binomial distribution

Introduction to Beta distribution

Beta-Binomial Inference

Binomial Data

Suppose we have data that are reasonable to assume follow a binomial distribution.

That is, we must have

- the observed data consisting of the total number of successes in a fixed number, n, of trials; and
- each trial must have a constant probability of success, θ (where $\theta \in [0, 1]$); and
- each trial outcome is independent of the other trials' outcomes.

Examples of Binomial Data

- The number of free throws made by a basketball player in three attempts.
- The number of defective axles from a random sample of 20 axles produced at a manufacturing plant
- The number of students opposing capital punishment in a random sample of 100 BYU students.

Making Inference with the Binomial Likelihood

Binomial Likelihood: $y|\theta$ is $f(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$ for $y \in \{0, 1, \ldots, n\}$

What do we know about θ? It represents the probability of any given trial being a "success." Because of this interpretation, it is clear why $0 \leq \theta \leq 1$.

Technically the binomial distributions depends on two parameters, n and p, but we will always assume n is known when working with the binomial distribution. So we focus entirely on prior and posterior beliefs for p. To fit the general framework we use to describe Bayesian inference—using y to describe the data and θ to describe the parameter(s)—I adopt the convention of using θ to refer to p whenever I work with the binomial distribution.

Example Interpretations of θ

Some examples:

- If $Y =$ the number of free throws made by a basketball player in three attempts, then θ would represent

Example Interpretations of θ

Some examples:

- If $Y =$ the number of free throws made by a basketball player in three attempts, then θ would represent the probability of making a free throw.

Example Interpretations of θ

Some examples:

- If $Y =$ the number of free throws made by a basketball player in three attempts, then θ would represent the probability of making a free throw.
- If $Y =$ the number of defective axles from a random sample of 20 axles produced at a manufacturing plant, then θ would represent

Example Interpretations of θ

Some examples:

- If $Y =$ the number of free throws made by a basketball player in three attempts, then θ would represent the probability of making a free throw.
- If $Y =$ the number of defective axles from a random sample of 20 axles produced at a manufacturing plant, then θ would represent the proportion of all axles produced at the plant that are defective.

Example Interpretations of θ

Some examples:

- If $Y =$ the number of free throws made by a basketball player in three attempts, then θ would represent the probability of making a free throw.
- If $Y =$ the number of defective axles from a random sample of 20 axles produced at a manufacturing plant, then θ would represent the proportion of all axles produced at the plant that are defective.
- If $Y =$ the number of students opposing capital punishment in a random sample of 100 BYU students, then θ would represent

Example Interpretations of θ

Some examples:

- If $Y =$ the number of free throws made by a basketball player in three attempts, then θ would represent the probability of making a free throw.
- If $Y =$ the number of defective axles from a random sample of 20 axles produced at a manufacturing plant, then θ would represent the proportion of all axles produced at the plant that are defective.
- If $Y =$ the number of students opposing capital punishment in a random sample of 100 BYU students, then θ would represent the proportion of all BYU students that oppose capital punishment.

Example Interpretations of θ

Some examples:

- If $Y =$ the number of free throws made by a basketball player in three attempts, then θ would represent the probability of making a free throw.
- If $Y =$ the number of defective axles from a random sample of 20 axles produced at a manufacturing plant, then θ would represent the proportion of all axles produced at the plant that are defective.
- If $Y =$ the number of students opposing capital punishment in a random sample of 100 BYU students, then θ would represent the proportion of all BYU students that oppose capital punishment.

Example Interpretations of θ

Some examples:

- If $Y =$ the number of free throws made by a basketball player in three attempts, then θ would represent the probability of making a free throw.
- If $Y =$ the number of defective axles from a random sample of 20 axles produced at a manufacturing plant, then θ would represent the proportion of all axles produced at the plant that are defective.
- If $Y =$ the number of students opposing capital punishment in a random sample of 100 BYU students, then θ would represent the proportion of all BYU students that oppose capital punishment.

In any of the above examples, do we know the numerical value of θ?

Prior Distribution for θ

If θ were known, life would be simple. But because it is not known, we reflect our prior belief about its possible values with a statistical distribution. Then, we will be able to use collected data to update our beliefs.

Prior Distribution for θ

If θ were known, life would be simple. But because it is not known, we reflect our prior belief about its possible values with a statistical distribution. Then, we will be able to use collected data to update our beliefs.

We'll use the beta distribution to describe our prior beliefs about θ when working with the binomial likelihood.

$$\pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{(0\leq\theta\leq 1)}$$

Technical note: The gamma function, $\Gamma(u)$, has an interesting and useful property. For any integer value of u such that $u \geq 1$, then $\Gamma(u) = (u-1)!$. For example, $\Gamma(6) = 5! = 120$, and $\Gamma(3) = 2! = 2$. In general, the gamma function exhibits the recursive property that for $u \in (1, \infty)$, then $\Gamma(u) = (u-1)\Gamma(u-1)$.

Why use the Beta distribution?

- It has the right support for a parameter that represents the probability of a success.
    - That is, *Binomial*(n, θ) requires that $\theta \in [0, 1]$, and the support of the Beta distribution is also $[0, 1]$.
- It is a **conjugate** prior distribution for θ when the binomial likelihood is assumed. That is, a binomial likelihood for $y|\theta$ coupled with a beta prior distribution for θ produces a (different) beta posterior distribution for $\theta|y$.

Analytically deriving the posterior distribution of θ

Detailed Derivation

$$\pi(\theta|y) = \frac{\pi(\theta)f(y|\theta)}{f(y)}$$

Detailed Derivation

$$\pi(\theta|y) = \frac{\pi(\theta)f(y|\theta)}{f(y)}$$

$$= \frac{\overbrace{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{(0\leq\theta\leq 1)}}^{\pi(\theta),\text{ because }\theta\sim Beta(a,b)}\;\overbrace{\binom{n}{y}\theta^{y}(1-\theta)^{n-y}\mathbb{1}_{(y\in\{0,1,\ldots,n\})}}^{f(y|\theta),\text{ because }y|\theta\sim Binomial(n,\theta)}}{\underbrace{\int\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}t^{a-1}(1-t)^{b-1}\mathbb{1}_{(0\leq t\leq 1)}\binom{n}{y}t^{y}(1-t)^{n-y}\mathbb{1}_{(y\in\{0,1,\ldots,n\})}dt}_{f(y),\text{ because }f(y)=\int\pi(\theta)f(y|\theta)d\theta}}$$

Detailed Derivation

$$\pi(\theta|y) = \frac{\pi(\theta)f(y|\theta)}{f(y)}$$

$$= \frac{\overbrace{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{(0\le\theta\le1)}}^{\pi(\theta),\text{ because }\theta\sim Beta(a,b)} \overbrace{\binom{n}{y}\theta^{y}(1-\theta)^{n-y}\mathbb{1}_{(y\in\{0,1,\ldots,n\})}}^{f(y|\theta),\text{ because }y|\theta\sim Binomial(n,\theta)}}{\underbrace{\int \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}t^{a-1}(1-t)^{b-1}\mathbb{1}_{(0\le t\le1)}\binom{n}{y}t^{y}(1-t)^{n-y}\mathbb{1}_{(y\in\{0,1,\ldots,n\})}dt}_{f(y),\text{ because }f(y)=\int\pi(\theta)f(y|\theta)d\theta}}$$

$$= \frac{\cancel{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}}\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{(0\le\theta\le1)}\cancel{\binom{n}{y}}\theta^{y}(1-\theta)^{n-y}\cancel{\mathbb{1}_{(y\in\{0,1,\ldots,n\})}}}{\int \cancel{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}}t^{a-1}(1-t)^{b-1}\mathbb{1}_{(0\le t\le1)}\cancel{\binom{n}{y}}t^{y}(1-t)^{n-y}\cancel{\mathbb{1}_{(y\in\{0,1,\ldots,n\})}}dt}$$

Detailed Derivation (cont. from previous slide)

$$\pi(\theta|y) = \frac{\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{(0\leq\theta\leq1)}\theta^{y}(1-\theta)^{n-y}}{\int t^{a-1}(1-t)^{b-1}\mathbb{1}_{(0\leq t\leq1)}t^{y}(1-t)^{n-y}dt}$$

Detailed Derivation (cont. from previous slide)

$$\pi(\theta|y) = \frac{\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{(0\leq\theta\leq 1)}\theta^{y}(1-\theta)^{n-y}}{\int t^{a-1}(1-t)^{b-1}\mathbb{1}_{(0\leq t\leq 1)}t^{y}(1-t)^{n-y}dt}$$

$$= \frac{\theta^{a+y-1}(1-\theta)^{b+n-y-1}\mathbb{1}_{(0\leq\theta\leq 1)}}{\int_0^1 t^{a+y-1}(1-t)^{b+n-y-1}dt}$$

Detailed Derivation (cont. from previous slide)

$$\pi(\theta|y) = \frac{\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{(0\le\theta\le1)}\theta^{y}(1-\theta)^{n-y}}{\int t^{a-1}(1-t)^{b-1}\mathbb{1}_{(0\le t\le1)}t^{y}(1-t)^{n-y}dt}$$

$$= \frac{\theta^{a+y-1}(1-\theta)^{b+n-y-1}\mathbb{1}_{(0\le\theta\le1)}}{\int_0^1 t^{a+y-1}(1-t)^{b+n-y-1}dt}$$

$$= \frac{\theta^{a+y-1}(1-\theta)^{b+n-y-1}\mathbb{1}_{(0\le\theta\le1)}}{\frac{\Gamma(a+y)\Gamma(b+n-y)}{\Gamma(a+b+n)}\underbrace{\int_0^1 \frac{\Gamma(a+b+n)}{\Gamma(a+y)\Gamma(b+n-y)}t^{a+y-1}(1-t)^{b+n-y-1}dt}_{=1}}$$

Detailed Derivation (cont. from previous slide)

$$
\begin{aligned}
\pi(\theta|y) &= \frac{\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{(0\le\theta\le1)}\theta^{y}(1-\theta)^{n-y}}{\int t^{a-1}(1-t)^{b-1}\mathbb{1}_{(0\le t\le1)}t^{y}(1-t)^{n-y}dt} \\
&= \frac{\theta^{a+y-1}(1-\theta)^{b+n-y-1}\mathbb{1}_{(0\le\theta\le1)}}{\int_0^1 t^{a+y-1}(1-t)^{b+n-y-1}dt} \\
&= \frac{\theta^{a+y-1}(1-\theta)^{b+n-y-1}\mathbb{1}_{(0\le\theta\le1)}}{\frac{\Gamma(a+y)\Gamma(b+n-y)}{\Gamma(a+b+n)}\underbrace{\int_0^1 \frac{\Gamma(a+b+n)}{\Gamma(a+y)\Gamma(b+n-y)}t^{a+y-1}(1-t)^{b+n-y-1}dt}_{=1}} \\
&= \frac{\Gamma(a+b+n)}{\Gamma(a+y)\Gamma(b+n-y)}\theta^{a+y-1}(1-\theta)^{b+n-y-1}\mathbb{1}_{(0\le\theta\le1)}
\end{aligned}
$$

Detailed Derivation (cont. from previous slide)

$$\begin{aligned}
\pi(\theta|y) &= \frac{\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{(0\le\theta\le 1)}\theta^{y}(1-\theta)^{n-y}}{\int t^{a-1}(1-t)^{b-1}\mathbb{1}_{(0\le t\le 1)}t^{y}(1-t)^{n-y}dt} \\
&= \frac{\theta^{a+y-1}(1-\theta)^{b+n-y-1}\mathbb{1}_{(0\le\theta\le 1)}}{\int_0^1 t^{a+y-1}(1-t)^{b+n-y-1}dt} \\
&= \frac{\theta^{a+y-1}(1-\theta)^{b+n-y-1}\mathbb{1}_{(0\le\theta\le 1)}}{\frac{\Gamma(a+y)\Gamma(b+n-y)}{\Gamma(a+b+n)}\underbrace{\int_0^1 \frac{\Gamma(a+b+n)}{\Gamma(a+y)\Gamma(b+n-y)}t^{a+y-1}(1-t)^{b+n-y-1}dt}_{=1}} \\
&= \frac{\Gamma(a+b+n)}{\Gamma(a+y)\Gamma(b+n-y)}\theta^{a+y-1}(1-\theta)^{b+n-y-1}\mathbb{1}_{(0\le\theta\le 1)}
\end{aligned}$$

Note, this is the pdf of the Beta$(a^\star = a+y, b^\star = b+n-y)$ distribution. That is, when $\theta \sim$ Beta(a,b) and $y|\theta \sim$ Binomial(n,θ), then $\theta|y \sim$ Beta$(a+y, b+n-y)$.

A math trick

To derive the posterior distribution, we also could have used a helpful property: **whenever two statistical distributions are proportional to each other, they must be equal to each other.**

> If we recognize that $\pi(\theta|y)$ is PROPORTIONAL to some particular distribution, we know the posterior must BE that distribution.

- Why? Let $D_1(\theta)$ be the posterior distribution, and let $D_2(\theta)$ be some known distribution. Because the total probability for a distribution must be 1, then the only possibility for $D_1(\theta) \propto D_2(\theta)$, $D_1(\theta)$ a density, and $D_2(\theta)$ a density is for $D_1 = D_2$.

Beta-Binomial Posterior

As a function of θ,

$$\begin{aligned} \pi(\theta|y) = & \quad \frac{\pi(\theta)f(y|\theta)}{f(y)} \\ \propto & \quad \pi(\theta)f(y|\theta) \end{aligned}$$

Beta-Binomial Posterior

As a function of θ,

$$\begin{aligned}\pi(\theta|y) = & \quad \frac{\pi(\theta)f(y|\theta)}{f(y)} \\ \propto & \quad \pi(\theta)f(y|\theta) \\ = & \underbrace{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{(0\leq\theta\leq1)}}_{\theta\sim Beta(a,b)}\underbrace{\binom{n}{y}\theta^{y}(1-\theta)^{n-y}\mathbb{1}_{(y\in\{0,1,\ldots,n\})}}_{y|\theta\sim Binomial(n,\theta)}\end{aligned}$$

Beta-Binomial Posterior

As a function of θ,

$$\begin{aligned}
\pi(\theta|y) = & \quad \frac{\pi(\theta)f(y|\theta)}{f(y)} \\
\propto & \quad \pi(\theta)f(y|\theta) \\
= & \underbrace{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{(0\leq\theta\leq1)}}_{\theta\sim Beta(a,b)}\underbrace{\binom{n}{y}\theta^{y}(1-\theta)^{n-y}\mathbb{1}_{(y\in\{0,1,\ldots,n\})}}_{y|\theta\sim Binomial(n,\theta)} \\
\propto & \quad \theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{(0\leq\theta\leq1)}\theta^{y}(1-\theta)^{n-y} \\
= & \quad \theta^{(a+y)-1}(1-\theta)^{(b+n-y)-1}\mathbb{1}_{(0\leq\theta\leq1)}
\end{aligned}$$

Beta-Binomial Posterior

As a function of θ,

$$
\begin{aligned}
\pi(\theta|y) = & \quad \frac{\pi(\theta) f(y|\theta)}{f(y)} \\
\propto & \quad \pi(\theta) f(y|\theta) \\
= & \underbrace{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1} \mathbb{1}_{(0 \le \theta \le 1)}}_{\theta \sim Beta(a,b)} \underbrace{\binom{n}{y} \theta^{y} (1-\theta)^{n-y} \mathbb{1}_{(y \in \{0,1,\ldots,n\})}}_{y|\theta \sim Binomial(n,\theta)} \\
\propto & \quad \theta^{a-1} (1-\theta)^{b-1} \mathbb{1}_{(0 \le \theta \le 1)} \theta^{y} (1-\theta)^{n-y} \\
= & \quad \theta^{(a+y)-1} (1-\theta)^{(b+n-y)-1} \mathbb{1}_{(0 \le \theta \le 1)} \\
\propto & \quad \frac{\Gamma((a+y)+(b+n-y))}{\Gamma(a+y)\Gamma(b+n-y)} \theta^{(a+y)-1} (1-\theta)^{(b+n-y)-1} \mathbb{1}_{(0 \le \theta \le 1)} \\
= & \quad \frac{\Gamma(a^{\star}+b^{\star})}{\Gamma(a^{\star})\Gamma(b^{\star})} \theta^{a^{\star}-1} (1-\theta)^{b^{\star}-1} \mathbb{1}_{(0 \le \theta \le 1)}
\end{aligned}
$$

The posterior dist. is $\theta|y \sim Beta(a^{\star} = a + y, b^{\star} = b + n - y)$.

Beta-Binomial Result–Very Important!

- IF WE HAVE
  - a binomial likelihood (that is, the model for $y|\theta$ is $f(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$ for $y \in \{0, 1, \dots, n\}$)

  AND
  - a beta distribution for the prior distribution of θ (i.e., $\pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{(0\leq\theta\leq 1)}$),

  THEN
  - the posterior distribution, $\pi(\theta|y)$, will be another beta distribution (which can be written as $Beta(a^\star, b^\star)$, where $a^\star = a + y$ and $b^\star = b + (n - y)$).

Definition of Conjugacy

We say a prior distribution is **conjugate** for a given likelihood if the posterior distribution must be in the same family of distributions as the prior distribution if the given likelihood is assumed.

Beta prior and binomial likelihood $\Rightarrow$ beta posterior—The beta prior is conjugate for the binomial likelihood!

Example of Posterior

Having a Beta(4,10) prior on $\theta \equiv$ probability of success, and then observing 11 successes in 25 trials, produces a Beta(15, 24) posterior for $\theta|\mathbf{y}$.

That is,

- if I (with justification) select the *Binomial*$(25, \theta)$ likelihood for the observed data (y number of successes in 25 trials); **and**

Example of Posterior

Having a Beta(4,10) prior on $\theta \equiv$ probability of success, and then observing 11 successes in 25 trials, produces a Beta(15, 24) posterior for $\theta|\mathbf{y}$.

That is,

- if I (with justification) select the *Binomial*$(25, \theta)$ likelihood for the observed data (y number of successes in 25 trials); **and**
- if I (with justfication) select the *Beta*$(4, 10)$ prior distribution for the model parameter (θ is the probability of success for each trial); **and**

Example of Posterior

Having a Beta(4,10) prior on $\theta \equiv$ probability of success, and then observing 11 successes in 25 trials, produces a Beta(15, 24) posterior for $\theta|\mathbf{y}$.

That is,

- if I (with justification) select the $\mathit{Binomial}(25, \theta)$ likelihood for the observed data (y number of successes in 25 trials); **and**
- if I (with justfication) select the $\mathit{Beta}(4, 10)$ prior distribution for the model parameter (θ is the probability of success for each trial); **and**
- if the observed data are $y = 11$ successes in the 25 total trials; **then**

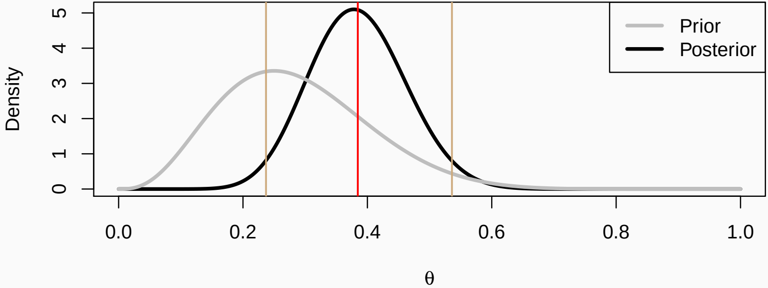
Example of Posterior

Having a Beta(4,10) prior on $\theta \equiv$ probability of success, and then observing 11 successes in 25 trials, produces a Beta(15, 24) posterior for $\theta|\mathbf{y}$.

That is,

- if I (with justification) select the $\textit{Binomial}(25, \theta)$ likelihood for the observed data (y number of successes in 25 trials); **and**
- if I (with justfication) select the $\textit{Beta}(4, 10)$ prior distribution for the model parameter (θ is the probability of success for each trial); **and**
- if the observed data are $y = 11$ successes in the 25 total trials; **then**
- the posterior distribution is the $\textit{Beta}(4 + 11 = 15, 10 + 25 - 11 = 24)$ distribution.

Posterior Summaries



posterior mean $= E(\theta|\mathbf{y}) = \int \theta\pi(\theta|\mathbf{y})d\theta = \frac{a^\star}{a^\star+b^\star} = \frac{15}{39} =$ 0.3846

posterior variance $=$

$$\frac{a^\star b^\star}{(a^\star+b^\star)^2(a^\star+b^\star+1)} = \frac{15(24)}{(15+24)^2(15+24+1)} = \frac{360}{60840} = 0.00592$$

95% posterior **credible interval**: (0.237, 0.536)

A (posterior) **credible interval** is an interval with 95% posterior probability of containing θ.

Interpretation of Beta-binomial Prior, Posterior

Common interpretations of a and b in the beta prior distribution are as the number of "prior successess" and "prior failures," respectively.

That is, **prior beliefs are expressed as though one had already observed a observations' worth of successes and b observations' worth of failures.**

But note that (1) it might be more accurate to consider having $a - 1$ and $b - 1$ observations' worth of "prior successes" and "prior failures," and that (2) a and b need not be integer valued in the prior distribution!

The motivation for this interpretation

- y observed data successes; $(n - y)$ observed data failures.
- Every increase of 1 in observed successes leads to an increase of 1 in $a^\star$
- Likewise, whenever a increases by 1, $a^\star$ increases by 1.
- Posterior has $a^\star = a + y =$ "prior successes" + observed data successes;

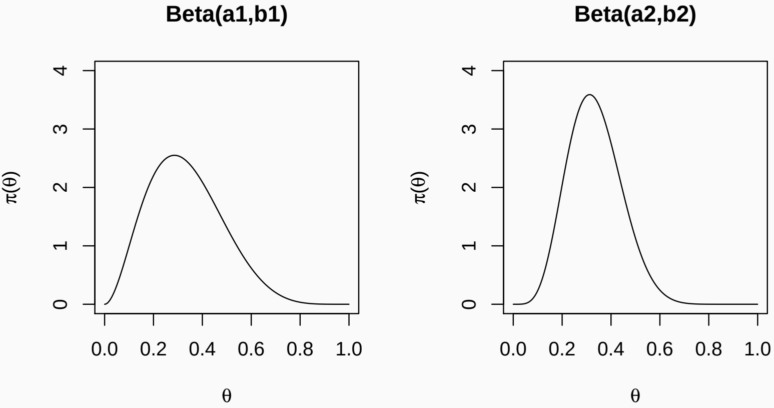
The motivation for this interpretation

- y observed data successes; $(n - y)$ observed data failures.
- Every increase of 1 in observed successes leads to an increase of 1 in $a^\star$
- Likewise, whenever a increases by 1, $a^\star$ increases by 1.
- Posterior has $a^\star = a + y =$ "prior successes" + observed data successes;
- Likewise, if $(n - y)$ increases by 1, effect on $b^\star$ is same as if b had increased by 1.
- Posterior has $b^\star = b + (n - y) =$ "prior failures" + observed data failures.

This yields the *data augmentation* interpretation (as described in Christensen et al. 2011).

Comprehension Questions

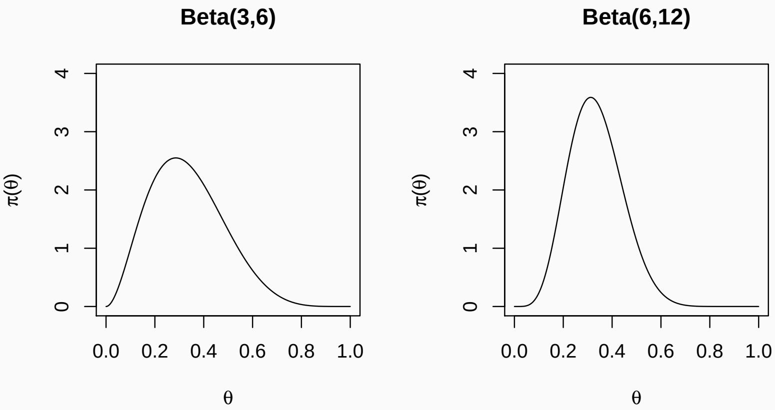
Which distribution represents a more "informative" prior?
Similarly, which distribution has a larger value for $a + b$?



Comprehension Questions

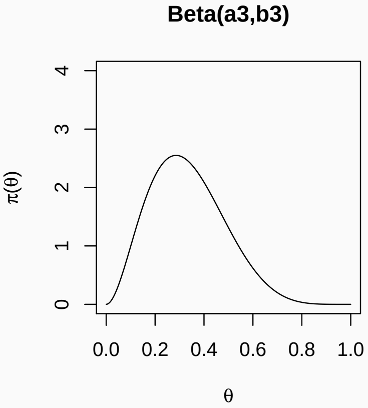
Which distribution represents a more "informative" prior?

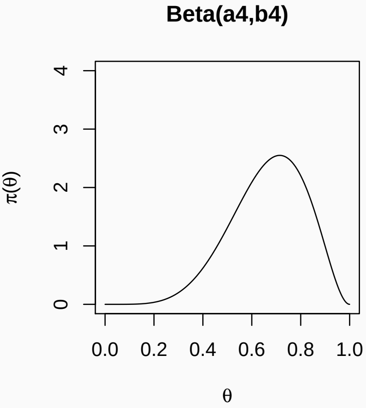
Similarly, which distribution has a larger value for $a + b$?



Comprehension Questions (cont)

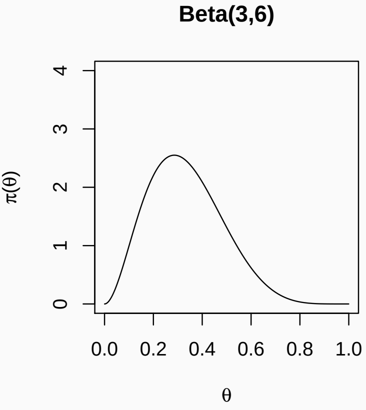
Which distribution has a larger mean? (Similarly, which distribution has $a > b$?)

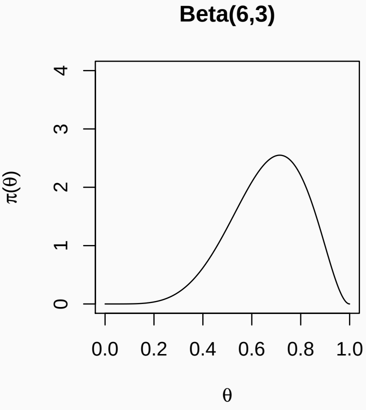




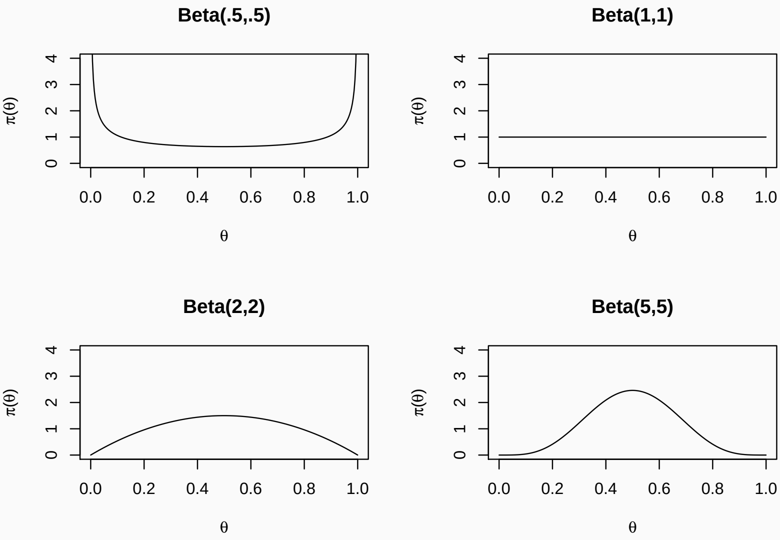
Comprehension Questions (cont)

Which distribution has a larger mean? (Similarly, which distribution has $a > b$?)





What happens if $a = b$?



Beta-binomial Example

Christensen et al. (2011, pp. 22–27) gave a hypothetical example for inference on $\theta =$ proportion of U.S. transportation workers who are on the influence of drugs while working. I am amending this example (though I heartily recommend the original to you). Suppose that my prior is $\theta \sim \mathit{Beta}(1.4, 23.6)$.

1. Graph this prior distribution in R. (use *dbeta*)
2. What is the (prior) mean of θ according to this prior?
3. What is the maximum a priori estimate of θ? (prior mode)
4. What is a 95% credible interval for the prior value of θ? (use *qbeta* twice to get the 2.5th and 97.5th percentiles)

Beta-binomial Example–Solutions

1. Graph this prior ($\theta \sim Beta(1.4, 23.6)$)

```
plot(tt<-seq(0,1,length.out=1001), dbeta(tt,1.4,23.6),
     main="Beta(1.4, 23.6) Prior Distribution",
     xlab=expression(theta), type="l", ylab=
     expression(paste(pi, "(", theta, ")", sep="")))
```

Beta-binomial Example–Solutions

1. Graph this prior ($\theta \sim Beta(1.4, 23.6)$)

```
plot(tt<-seq(0,1,length.out=1001), dbeta(tt,1.4,23.6),
     main="Beta(1.4, 23.6) Prior Distribution",
     xlab=expression(theta), type="l", ylab=
     expression(paste(pi, "(", theta, ")", sep="")))
```

2. What is the (prior) mean of θ according to this prior?
   $a/(a+b) = 1.4/25 = 0.056$

Beta-binomial Example–Solutions

1. Graph this prior ($\theta \sim Beta(1.4, 23.6)$)

```
plot(tt<-seq(0,1,length.out=1001), dbeta(tt,1.4,23.6),
     main="Beta(1.4, 23.6) Prior Distribution",
     xlab=expression(theta), type="l", ylab=
     expression(paste(pi, "(", theta, ")", sep="")))
```

2. What is the (prior) mean of θ according to this prior?
   $a/(a+b) = 1.4/25 = 0.056$
3. What is the maximum a priori estimate of θ? (prior mode)
   (a-1)/(a+b-2) = 0.4/23 = 0.0174

Beta-binomial Example–Solutions

1. Graph this prior ($\theta \sim Beta(1.4, 23.6)$)

```
plot(tt<-seq(0,1,length.out=1001), dbeta(tt,1.4,23.6),
     main="Beta(1.4, 23.6) Prior Distribution",
     xlab=expression(theta), type="l", ylab=
     expression(paste(pi, "(", theta, ")", sep="")))
```

2. What is the (prior) mean of θ according to this prior?
   $a/(a+b) = 1.4/25 = 0.056$
3. What is the maximum a priori estimate of θ? (prior mode)
   $(a\text{-}1)/(a\text{+}b\text{-}2) = 0.4/23 = 0.0174$
4. What is a 95% credible interval for the prior value of θ? (use *qbeta* twice to get the 2.5th and 97.5th percentiles)

```
c(qbeta(.025, 1.4, 23.6), qbeta(.975, 1.4, 23.6))

## [1] 0.00364109 0.17181657
```

Beta-binomial Example (cont.)

Recall that my prior distribution for θ is $Beta(1.4, 23.6)$. Suppose that a sample of workers revealed the following:

Under Influence?											
No	No	No	Yes	No	Yes	No	No	No	No	Yes	No

Table 1: My adaptation of a hypothetical example from Christensen et al. (2011, pp. 22-27)

- What is the posterior distribution for $\theta|y$?
- Plot the posterior distribution. (dbeta)
- What is the posterior mean for $\theta|y$?
- What is the maximum *a posteriori* estimate of $\theta|y$? (posterior mode)
- What is a 95% credible interval for $\theta|y$? (use qbeta twice)
- How well does the posterior agree with the prior?

Prior Elicitation for the Beta-Binomial setting

Prior Elicitation

How to select a and b for the Beta distribution? Some options:

- If you don't have much information at all about θ, can use $\mathbf{a} = \mathbf{b} = 1$. This corresponds to a uniform prior distribution: all values between 0 and 1 are considered equally likely.

Prior Elicitation

How to select a and b for the Beta distribution? Some options:

- If you don't have much information at all about θ, can use $\mathbf{a} = \mathbf{b} = 1$. This corresponds to a uniform prior distribution: all values between 0 and 1 are considered equally likely.
- Translate your prior belief into **a** observations' worth of successes and **b** observations' worth of failures.

Prior Elicitation

How to select a and b for the Beta distribution? Some options:

- If you don't have much information at all about θ, can use $\mathbf{a} = \mathbf{b} = 1$. This corresponds to a uniform prior distribution: all values between 0 and 1 are considered equally likely.
- Translate your prior belief into **a** observations' worth of successes and **b** observations' worth of failures.
- Give an expected prior mean and expected prior standard deviation, then solve for **a** and **b**.
  $E(\theta) = \frac{a}{a+b}$ and $S.D.(\theta) = \sqrt{\frac{ab}{(a+b)^2(a+b+1)}}$

Keep in mind: the ratio of a to b will affect the mean, and the size of $a + b$ will affect how informative your prior is.

Reference

Christensen, R., Johnson, W.O., Branscum, A.J., and Hanson, T.E. (2011). *Bayesian Ideas and Data Analysis*. Boca Raton: CRC Press.