450D Political Methodology IV
TA Section 2

Vincent Bauer

October 6, 2017

Estimated time: 1hr 30min

Goal Today

- Derive posterior distributions for binomial and poisson models with conjugate priors.

Road map to those goals:

1. Functions
   1.1 Binomial coefficient
   1.2 Gamma function
   1.3 Beta function
2. Distributions
   2.1 Beta distribution
   2.2 Binomial distribution
   2.3 Gamma distribution
   2.4 Poisson distribution
3. Posterior Derivations
   3.1 Binomial distribution
   3.2 Poisson distribution
4. Posterior Predictive
   4.1 Binomial distribution

Starting with some basic functions...

Functions

The *binomial coefficient* is the number of ways of picking k unordered outcomes from n possibilities.

Functions

The *binomial coefficient* is the number of ways of picking k unordered outcomes from n possibilities.

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

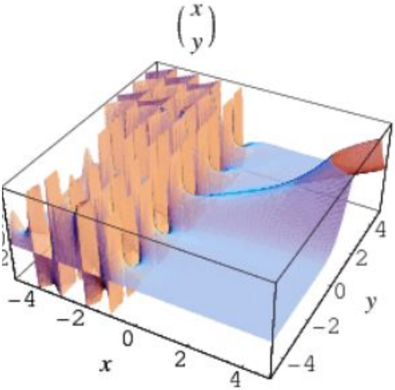
Functions

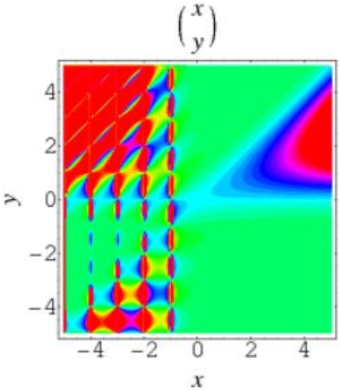
The *binomial coefficient* is the number of ways of picking k unordered outcomes from n possibilities.

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

There are no assumptions here, this is just true.

Functions





Functions

The *gamma function* is a generalization of the factorial operator that works for non-integers as well.

Functions

The *gamma function* is a generalization of the factorial operator that works for non-integers as well.

$$\Gamma(x) = \int_0^\infty t^{x-1}\, e^{-t}\, dt \qquad (x > 0)$$

Functions

The *gamma function* is a generalization of the factorial operator that works for non-integers as well.

$$\Gamma(x) = \int_0^\infty t^{x-1}\, e^{-t}\, dt \qquad (x > 0)$$

Denote the gamma function as $\Gamma(x)$, and the the gamma distribution as $Gamma(x)$.

Functions

The *gamma function* is a generalization of the factorial operator that works for non-integers as well.

$$\Gamma(x) = \int_0^{\infty} t^{x-1}\, e^{-t}\, dt \qquad (x > 0)$$

Denote the gamma function as $\Gamma(x)$, and the the gamma distribution as $Gamma(x)$.

There's no closed form solution to this integral but R knows how to do it.

Functions

The *gamma function* is a generalization of the factorial operator that works for non-integers as well.

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \, dt \qquad (x > 0)$$

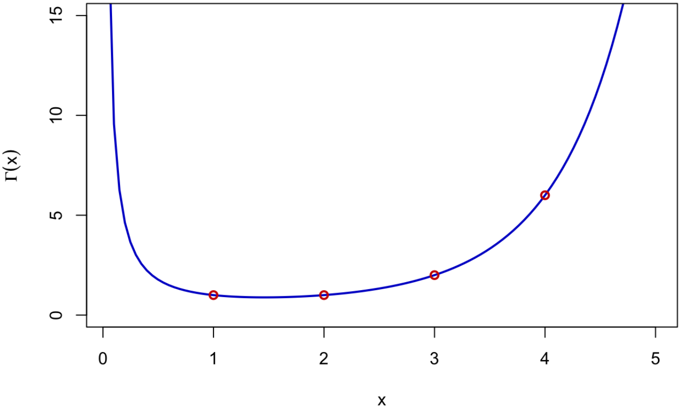
Denote the gamma function as $\Gamma(x)$, and the the gamma distribution as $Gamma(x)$.

There's no closed form solution to this integral but R knows how to do it.

Here are some properties:

1. $\Gamma(x+1) = x \cdot \Gamma(x)$, i.e. its a factorial
2. $\Gamma(x) = (n-1)!$

Functions



Functions

The *beta function* is related to the binomial coefficient and factorials.

Functions

The *beta function* is related to the binomial coefficient and factorials. One of its substantive applications is describing how wealth accumilates disproportionately to those who already have more.

Functions

The *beta function* is related to the binomial coefficient and factorials. One of its substantive applications is describing how wealth accumilates disproportionately to those who already have more.

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x + y)}$$

Functions

The *beta function* is related to the binomial coefficient and factorials. One of its substantive applications is describing how wealth accumilates disproportionately to those who already have more.

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

Recall that the binomial coefficient is already a ratio of factorials, so it makes sense that there's a relationship between the binomial coefficient and the Beta function.

$$\binom{n}{k} = \frac{1}{(n+1)\, B(k+1, n-k+1)}$$

Functions

The *beta function* is related to the binomial coefficient and factorials. One of its substantive applications is describing how wealth accumilates disproportionately to those who already have more.

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x + y)}$$

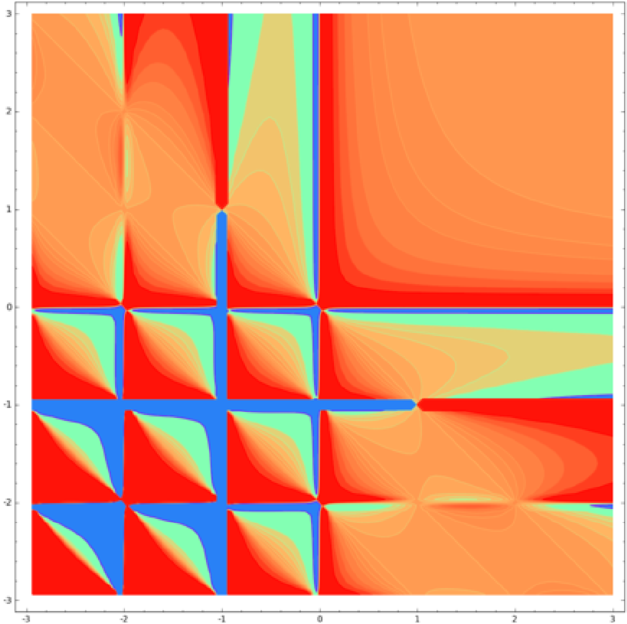
Recall that the binomial coefficient is already a ratio of factorials, so it makes sense that there's a relationship between the binomial coefficient and the Beta function.

$$\binom{n}{k} = \frac{1}{(n + 1)\, B(k + 1, n - k + 1)}$$

It can also be written in integral form, and as it relies on Gammas.

$$B(a, b) = \int_0^1 t^{a-1}(1 - t)^{b-1}\, dx \qquad (a, b > 0)$$

Functions



Moving on to distributions ...

Distributions

The *beta distribution* is bounded between [0,1] so we could use it to model proportions.

Distributions

The *beta distribution* is bounded between [0,1] so we could use it to model proportions. In practice, we care about the beta distribution as a conjugate prior for binomially distributed data.

Distributions

The *beta distribution* is bounded between [0,1] so we could use it to model proportions. In practice, we care about the beta distribution as a conjugate prior for binomially distributed data.

$$Beta(\alpha, \beta) = \frac{1}{B(\alpha, \beta)} \cdot \theta^{\alpha-1}(1-\theta)^{(\beta-1)}$$

Notice that the Beta distribution references the Beta function.

Distributions

The *beta distribution* is bounded between [0,1] so we could use it to model proportions. In practice, we care about the beta distribution as a conjugate prior for binomially distributed data.

$$Beta(\alpha, \beta) = \frac{1}{B(\alpha, \beta)} \cdot \theta^{\alpha-1}(1-\theta)^{(\beta-1)}$$

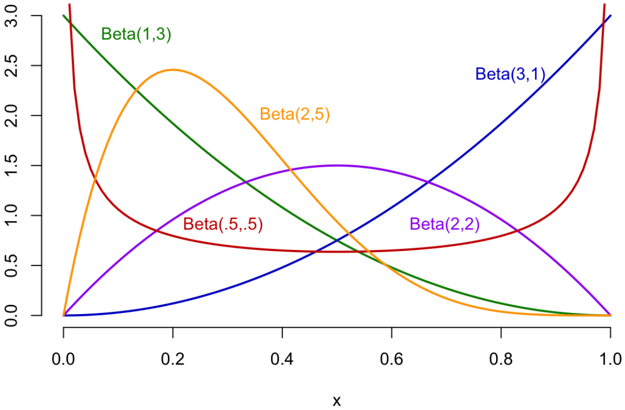
Notice that the Beta distribution references the Beta function.

If $X \sim Beta(\alpha, \beta)$ then

$$E[X] = \frac{\alpha}{\alpha + \beta}$$

$$Var[X] = \frac{\alpha \cdot \beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Distributions



Distributions

The *binomial distribution* is used for items belonging to one of two possible categories. We're modeling the count of "successes".

Distributions

The *binomial distribution* is used for items belonging to one of two possible categories. We're modeling the count of "successes".

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

Notice that the binomial distribution involves the binomial coefficient.

Distributions

The *binomial distribution* is used for items belonging to one of two possible categories. We're modeling the count of "successes".

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

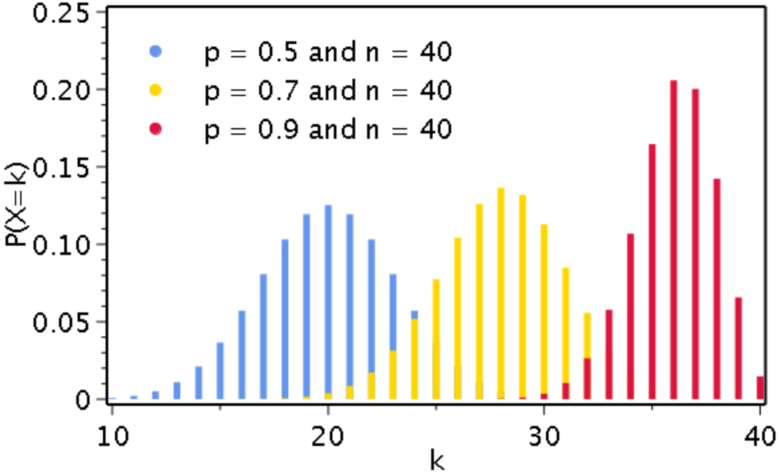
Notice that the binomial distribution involves the binomial coefficient.

If $X \sim Binomial(n, \theta)$ then

$$E[X] = n\theta$$

$$Var[X] = n\theta(1 - \theta)$$

Distributions



Distributions

The *gamma distribution* is defined for positive real numbers. If a is a positive integer, then the gamma distribution is the sum of a independent exponentially distributed random variables with mean $1/b$.

Distributions

The *gamma distribution* is defined for positive real numbers. If a is a positive integer, then the gamma distribution is the sum of a independent exponentially distributed random variables with mean $1/b$. In practice, we care about the gamma distribution as a conjugate prior for the poisson distribution.

Distributions

The *gamma distribution* is defined for positive real numbers. If a is a positive integer, then the gamma distribution is the sum of a independent exponentially distributed random variables with mean $1/b$. In practice, we care about the gamma distribution as a conjugate prior for the poisson distribution.

$$Gamma(a, b) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \qquad (\theta > 0)$$

Notice that the Gamma distribution references the Gamma function.

Distributions

The *gamma distribution* is defined for positive real numbers. If a is a positive integer, then the gamma distribution is the sum of a independent exponentially distributed random variables with mean $1/b$. In practice, we care about the gamma distribution as a conjugate prior for the poisson distribution.

$$Gamma(a, b) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \qquad (\theta > 0)$$

Notice that the Gamma distribution references the Gamma function.

If $X \sim Gamma(\alpha, \beta)$ then

$$E[X] = \frac{\alpha}{\beta}$$

$$Var[X] = \frac{\alpha}{\beta^2}$$

Distributions

The *gamma distribution* is defined for positive real numbers. If a is a positive integer, then the gamma distribution is the sum of a independent exponentially distributed random variables with mean $1/b$. In practice, we care about the gamma distribution as a conjugate prior for the poisson distribution.

$$Gamma(a, b) = \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta} \qquad (\theta > 0)$$

Notice that the Gamma distribution references the Gamma function.

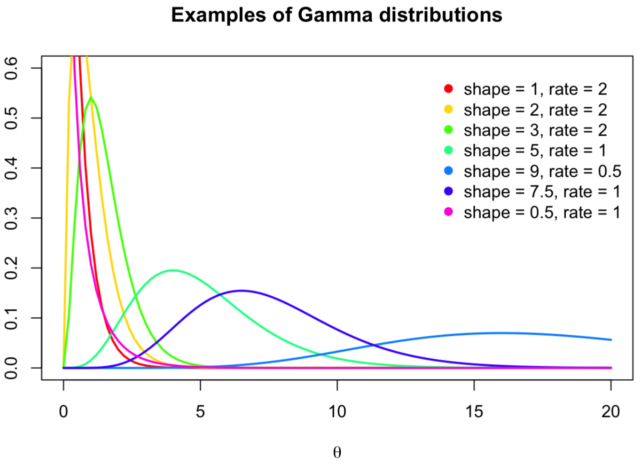
If $X \sim Gamma(\alpha, \beta)$ then

$$E[X] = \frac{\alpha}{\beta}$$

$$Var[X] = \frac{\alpha}{\beta^2}$$

Surprisingly friendly!

Distributions



Distributions

The *poisson distribution* is useful for modeling counts. It is used to model the number of events occurring in a fixed time interval.

Distributions

The *poisson distribution* is useful for modeling counts. It is used to model the number of events occurring in a fixed time interval.

$$Poisson(\theta) = \frac{1}{y_i!}\theta^{y_i}\exp(-\theta) \qquad (\theta > 0)$$

Distributions

The *poisson distribution* is useful for modeling counts. It is used to model the number of events occurring in a fixed time interval.

$$Poisson(\theta) = \frac{1}{y_i!}\theta^{y_i}\exp(-\theta) \qquad (\theta > 0)$$

Searching Google Scholar for "poison distribution" typos is always fun,

Distributions

The *poisson distribution* is useful for modeling counts. It is used to model the number of events occurring in a fixed time interval.

$$Poisson(\theta) = \frac{1}{y_i!}\theta^{y_i}\exp(-\theta) \qquad (\theta > 0)$$

Searching Google Scholar for "poison distribution" typos is always fun, if you're a nerd.

Distributions

The *poisson distribution* is useful for modeling counts. It is used to model the number of events occurring in a fixed time interval.

$$Poisson(\theta) = \frac{1}{y_i!}\theta^{y_i}\exp(-\theta) \qquad (\theta > 0)$$

Searching Google Scholar for "poison distribution" typos is always fun, if you're a nerd.

If $X \sim Poisson(\alpha, \beta)$ then

$$E[X] = \theta$$

$$Var[X] = \theta$$

Distributions

The *poisson distribution* is useful for modeling counts. It is used to model the number of events occurring in a fixed time interval.

$$Poisson(\theta) = \frac{1}{y_i!}\theta^{y_i}\exp(-\theta) \qquad (\theta > 0)$$

Searching Google Scholar for "poison distribution" typos is always fun, if you're a nerd.

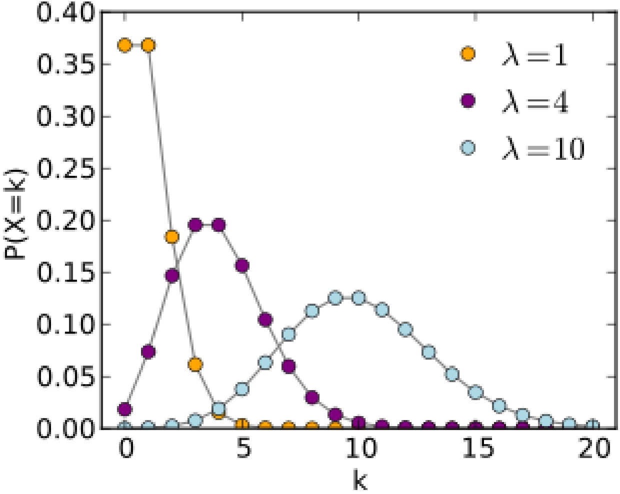
If $X \sim Poisson(\alpha, \beta)$ then

$$E[X] = \theta$$

$$Var[X] = \theta$$

Having the same mean and variance is convenient for now but is usually a hinderence in political science data so we will use a *negative binomial* distribution.

Distributions



Moving on to the derivations ...

Deriving the Binomial Posterior

How do we actually do Bayesian statistics with these distributions?

Deriving the Binomial Posterior

How do we actually do Bayesian statistics with these distributions?

1. Write out the likelihood of the data:

Deriving the Binomial Posterior

How do we actually do Bayesian statistics with these distributions?

1. Write out the likelihood of the data:

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

Drop the multiplicative constant, doesn't depend on θ.

$$\propto \theta^y (1-\theta)^{n-y}$$

Deriving the Binomial Posterior

How do we actually do Bayesian statistics with these distributions?

1. Write out the likelihood of the data:

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

   Drop the multiplicative constant, doesn't depend on θ.

$$\propto \theta^y(1-\theta)^{n-y}$$

2. We need a prior for θ that is bound on [0,1] so the beta distribution is a good choice.

Deriving the Binomial Posterior

How do we actually do Bayesian statistics with these distributions?

1. Write out the likelihood of the data:

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

   Drop the multiplicative constant, doesn't depend on θ.

$$\propto \theta^y(1-\theta)^{n-y}$$

2. We need a prior for θ that is bound on [0,1] so the beta distribution is a good choice.

$$p(\theta) = Beta(a, b) = \frac{1}{B(a, b)} \cdot \theta^{a-1} \cdot (1-\theta)^{b-1}$$

   Drop the multiplicative constant, doesn't depend on θ.

$$\propto \theta^{a-1} \cdot (1-\theta)^{b-1}$$

Deriving the Binomial Posterior

3. Write out the posterior

$$p(\theta|y,n) \propto \overbrace{\theta^{a-1} \cdot (1-\theta)^{b-1}}^{p(\theta)} \times \overbrace{\theta^{y}(1-\theta)^{n-y}}^{p(y|\theta,n)}$$

Group exponents together

$$\propto \theta^{a+y-1} \cdot (1-\theta)^{b+n-y-1}$$

Recognize this as

$$\propto Beta(a+y, b+n-y)$$

Deriving the Binomial Posterior

3. Write out the posterior

$$p(\theta|y,n) \propto \overbrace{\theta^{a-1} \cdot (1-\theta)^{b-1}}^{p(\theta)} \times \overbrace{\theta^{y}(1-\theta)^{n-y}}^{p(y|\theta,n)}$$

Group exponents together

$$\propto \theta^{a+y-1} \cdot (1-\theta)^{b+n-y-1}$$

Recognize this as

$$\propto Beta(a+y, b+n-y)$$

Therefore, the beta distribution and the binomial distribution are *conjugate*, our posterior has the same distribution as the prior, we have just added additional data.

Deriving the Poisson Posterior

Exactly the same process for the Poisson distribution

1. Write out the likelihood of the data:

Deriving the Poisson Posterior

Exactly the same process for the Poisson distribution

1. Write out the likelihood of the data:

$$
\begin{aligned}
p(y|\theta) = Poisson(\theta) &= \prod_{i=1}^{n} \frac{1}{y_i!} \cdot \theta^{y_i} \exp(-\theta) \\
&= \frac{1}{\prod_{i=1}^{n} y_i!} \cdot \theta^{\sum_{i=1}^{n} y_i} \exp(-n\theta) \\
&\propto \theta^{n\bar{y}} \exp(-n\theta)
\end{aligned}
$$

Deriving the Poisson Posterior

Exactly the same process for the Poisson distribution

1. Write out the likelihood of the data:

$$
\begin{aligned}
p(y|\theta) = Poisson(\theta) &= \prod_{i=1}^{n} \frac{1}{y_i!} \cdot \theta^{y_i} \exp(-\theta) \\
&= \frac{1}{\prod_{i=1}^{n} y_i!} \cdot \theta^{\sum_{i=1}^{n} y_i} \exp(-n\theta) \\
&\propto \theta^{n\bar{y}} \exp(-n\theta)
\end{aligned}
$$

2. We need a prior for θ that is bound to positive real numbers, so the gamma distribution is a good choice.

Deriving the Poisson Posterior

Exactly the same process for the Poisson distribution

1. Write out the likelihood of the data:

$$
\begin{aligned}
p(y|\theta) = Poisson(\theta) &= \prod_{i=1}^{n} \frac{1}{y_i!} \cdot \theta^{y_i} \exp(-\theta) \\
&= \frac{1}{\prod_{i=1}^{n} y_i!} \cdot \theta^{\sum_{i=1}^{n} y_i} \exp(-n\theta) \\
&\propto \theta^{n\bar{y}} \exp(-n\theta)
\end{aligned}
$$

2. We need a prior for θ that is bound to positive real numbers, so the gamma distribution is a good choice.

$$
\begin{aligned}
p(\theta) = Gamma(a, b) &= \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta) \\
&\propto \theta^{a-1} \exp(-b\theta)
\end{aligned}
$$

Deriving the Poisson Posterior

3. Write out the posterior

$$p(\theta|y,n) \propto \overbrace{\theta^{a-1}\exp(-b\theta)}^{p(\theta)} \times \overbrace{\theta^{n\bar{y}}\exp(-n\theta)}^{p(y|\theta,n)}$$

Group exponents together

$$\propto \theta^{a+n\bar{y}-1}\exp(-(b+n)\theta)$$

Recognize this as

$$\propto Gamma(a+n\bar{y}, b+n)$$

Deriving the Poisson Posterior

3. Write out the posterior

$$p(\theta|y,n) \propto \overbrace{\theta^{a-1}\exp(-b\theta)}^{p(\theta)} \times \overbrace{\theta^{n\bar{y}}\exp(-n\theta)}^{p(y|\theta,n)}$$

Group exponents together

$$\propto \theta^{a+n\bar{y}-1}\exp(-(b+n)\theta)$$

Recognize this as

$$\propto Gamma(a+n\bar{y}, b+n)$$

Therefore, the gamma distribution and the poisson distribution are *conjugate*, our posterior has the same distribution as the prior, we have just added additional data.

Posterior Predictive distributions

Knowing the posterior distribution of θ is great for academics, but not usually what your audience cares about.

Posterior Predictive distributions

Knowing the posterior distribution of θ is great for academics, but not usually what your audience cares about. How have our predictions about outcome y changed after collecting this data in comparison to our prior predictions?

Posterior Predictive distributions

Knowing the posterior distribution of θ is great for academics, but not usually what your audience cares about. How have our predictions about outcome y changed after collecting this data in comparison to our prior predictions?

Let's create a posterior predictive distribution

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

Posterior Predictive distributions

Knowing the posterior distribution of θ is great for academics, but not usually what your audience cares about. How have our predictions about outcome y changed after collecting this data in comparison to our prior predictions?

Let's create a posterior predictive distribution

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

Unfortunately, we don't <u>know</u> θ, instead we have a distribution of possible values of θ.

Posterior Predictive distributions

Knowing the posterior distribution of θ is great for academics, but not usually what your audience cares about. How have our predictions about outcome y changed after collecting this data in comparison to our prior predictions?

Let's create a posterior predictive distribution

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

Unfortunately, we don't know θ, instead we have a distribution of possible values of θ. We need to integrate out θ using its posterior distribution $p(\theta|y)$ to obtain the predictive distribution.

Posterior Predictive distributions

Knowing the posterior distribution of θ is great for academics, but not usually what your audience cares about. How have our predictions about outcome y changed after collecting this data in comparison to our prior predictions?

Let's create a posterior predictive distribution

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

Unfortunately, we don't know θ, instead we have a distribution of possible values of θ. We need to integrate out θ using its posterior distribution $p(\theta|y)$ to obtain the predictive distribution.

$$p(\tilde{y}|y) = \int_0^1 p(\tilde{y}, \theta|y)\, d\theta = \int_0^1 p(\tilde{y}|\theta, y)\, p(\theta|y)\, d\theta$$

Posterior Predictive distributions

$$p(\tilde{y}|y) = \int_0^1 p(\tilde{y}, \theta|y)\, d\theta = \int_0^1 p(\tilde{y}|\theta, y)\, p(\theta|y)\, d\theta$$

Posterior Predictive distributions

$$p(\tilde{y}|y) = \int_0^1 p(\tilde{y}, \theta|y)\, d\theta = \int_0^1 p(\tilde{y}|\theta, y)\, p(\theta|y)\, d\theta$$

Plug in the Beta-binomial case

$$= \int_0^1 \overbrace{\binom{\tilde{n}}{\tilde{y}} \theta^{\tilde{y}} (1-\theta)^{\tilde{n}-\tilde{y}}}^{p(\tilde{y}|\theta,y)} \cdot \overbrace{\frac{1}{B(a', b')} \theta^{a'-1} (1-\theta)^{b'-1} d\theta}^{p(\theta|y)}$$

Posterior Predictive distributions

$$p(\tilde{y}|y) = \int_0^1 p(\tilde{y}, \theta|y)\, d\theta = \int_0^1 p(\tilde{y}|\theta, y)\, p(\theta|y)\, d\theta$$

Plug in the Beta-binomial case

$$= \int_0^1 \overbrace{\binom{\tilde{n}}{\tilde{y}} \theta^{\tilde{y}} (1-\theta)^{\tilde{n}-\tilde{y}}}^{p(\tilde{y}|\theta,y)} \cdot \overbrace{\frac{1}{B(a', b')} \theta^{a'-1} (1-\theta)^{b'-1} d\theta}^{p(\theta|y)}$$

Move and group terms

$$= \binom{\tilde{n}}{\tilde{y}} \frac{1}{B(a', b')} \int \theta^{a'+\tilde{y}-1} (1-\theta)^{b'+\tilde{n}-\tilde{y}-1} d\theta$$

Posterior Predictive distributions

$$p(\tilde{y}|y) = \binom{\tilde{n}}{\tilde{y}} \frac{1}{B(a', b')} \int \theta^{a'+\tilde{y}-1}(1-\theta)^{b'+\tilde{n}-\tilde{y}-1} d\theta$$

Posterior Predictive distributions

$$p(\tilde{y}|y) = \binom{\tilde{n}}{\tilde{y}} \frac{1}{B(a', b')} \int \theta^{a'+\tilde{y}-1}(1-\theta)^{b'+\tilde{n}-\tilde{y}-1} d\theta$$

Remember the definition of the beta function

$$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1}\, dx$$

Posterior Predictive distributions

$$p(\tilde{y}|y) = \binom{\tilde{n}}{\tilde{y}} \frac{1}{B(a', b')} \int \theta^{a'+\tilde{y}-1}(1-\theta)^{b'+\tilde{n}-\tilde{y}-1} d\theta$$

Remember the definition of the beta function

$$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1}\, dx$$

Plug this into our equation

$$p(\tilde{y}|y) = \binom{\tilde{n}}{\tilde{y}} \frac{B(a'+\tilde{y}, b'+\tilde{n}-\tilde{y})}{B(a', b')}$$

Posterior Predictive distributions

$$p(\tilde{y}|y) = \binom{\tilde{n}}{\tilde{y}} \frac{1}{B(a', b')} \int \theta^{a'+\tilde{y}-1}(1-\theta)^{b'+\tilde{n}-\tilde{y}-1} d\theta$$

Remember the definition of the beta function

$$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1}\, dx$$

Plug this into our equation

$$p(\tilde{y}|y) = \binom{\tilde{n}}{\tilde{y}} \frac{B(a'+\tilde{y}, b'+\tilde{n}-\tilde{y})}{B(a', b')}$$

This is called the Beta-Binomial Distribution

Posterior Predictive distributions

Table: Summary of Distribution

Data	Prior	Posterior	Posterior Predictive
Binomial	Beta	Beta	Beta-Binomial
Poisson	Gamma	Gamma	Negative-Binomial

(Wikipedia has a good page on *conjugate priors*)

Posterior Predictive distributions

Table: Summary of Distribution

Data	Prior	Posterior	Posterior Predictive
Binomial	Beta	Beta	Beta-Binomial
Poisson	Gamma	Gamma	Negative-Binomial

(Wikipedia has a good page on *conjugate priors*)

The overarching idea is that we need to incorporate our uncertainty about θ into our predictions. This creates greater variance than we would see given a set θ.

Feedback on section...

Feedback on section...too easy? ...

Feedback on section...too easy? ... too hard? ...

Feedback on section...too easy? ... too hard? ... just right?