Priors, Part 2

DS6040 Fall 2024
Teague R. Henry

UNIVERSITY of VIRGINIA | SCHOOL of DATA SCIENCE

Outline

- Priors Review
- Eliciting Priors – Placenta Previa Example
- Eliciting Priors - Kidney Cancer Example

Bayes' Theorem

Likelihood
(Probability of Data Given Model)

Posterior Distribution
(Probability of Model Given Data)

$$\boxed{P(\theta|X)} = \frac{\boxed{P(X|\theta)}\,\boxed{P(\theta)}}{\boxed{P(X)}}$$

Prior Distribution
(Probability of Model)

$$P(X) = \sum_i P(X \cap \theta_i) = \sum_i P(X|\theta_i)P(\theta_i)$$

Probability of ***your data*** occurring for any choice of θ
Note: Possible θ_i are constrained by the model type.

Marginal Distribution
(Probability of Data)

Priors Review

Priors represent our prior knowledge about the phenomena under study.

- More specifically, they are probability distributions assigned to each parameter in a model that represent our prior beliefs about the *location* and *uncertainty* of the parameters.

There are many “classifications” of priors

- Conjugate priors – priors that result in the posterior being from the same distributional family
- Informative priors – priors that have something to say
- Weakly/uninformative priors – priors without much information given

Key Point: A posterior represents a compromise between data and the priors

Placenta Previa

Placenta Previa is a material condition in which the placenta is located low in the uterus.

- This blocks normal vaginal delivery, and if not accounted for, would result in hemorrhaging.

The question here is: Are male or female births more likely to have placenta previa?

Early study from Germany found that out of 980 PP births, 437 were female.

Placenta Previa – The likelihood

437/980 female births. We are interested in the proportion of female births relative to male births. Is it higher than .485, which is the proportion of female births in the overall population.

What is a reasonable likelihood for this data? --> The Binomial Distribution

- Our parameter of interest is a proportion. --> we're dealing w/ counts
- Our data is # of female births out of # of total births.

An appropriate likelihood for this is the *binomial distribution.*

Binomial Likelihood

Binomial Distribution

$$f(k|n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

- n – Total number of births
- k – Number of female births
- p – The probability of female births.

We want to perform inference on p. So what do we need to specify?

Prior Choice

$$f(k|n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

We need to specify our prior for p. This represents our prior (to seeing the data) belief about the probability of a female birth given placenta previa.

p is a probability/proportion.

- Bound by 0/1.

What would make for a good prior distribution? The Beta Distribution

Prior Choice - Beta

The Beta distribution is the conjugate prior for a binomial likelihood.

$$g(p|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

Alpha and beta are the hyper parameters and we have to choose them. What do they mean?

Two *hyperparameters*, α and β

- What do these mean?

    - Number of successess vs number of trials

Prior Choice – Hyperparameters

Alpha here is male birds,
beta is female birds

$$g(p|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

$$E[p|\alpha,\beta] = \frac{\alpha}{\alpha+\beta}$$

$$Var[p|\alpha,\beta] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

- What are some informative priors, let say we believe that male births have more prevalence of placenta previa?
- What is an “uninformative” prior choice?

The Posterior

The Beta distribution is a conjugate prior for the binomial likelihood. This means that the posterior will be distributed as a???

- Beta distribution! Specifically this one: $Beta(\alpha + k, n - k + \beta)$

when alpha and beta get larger, then the posterior starts to look more like our prior distr. b/c the data doesn't matter as much

What do you think is going to happen as the information content in the prior increases?

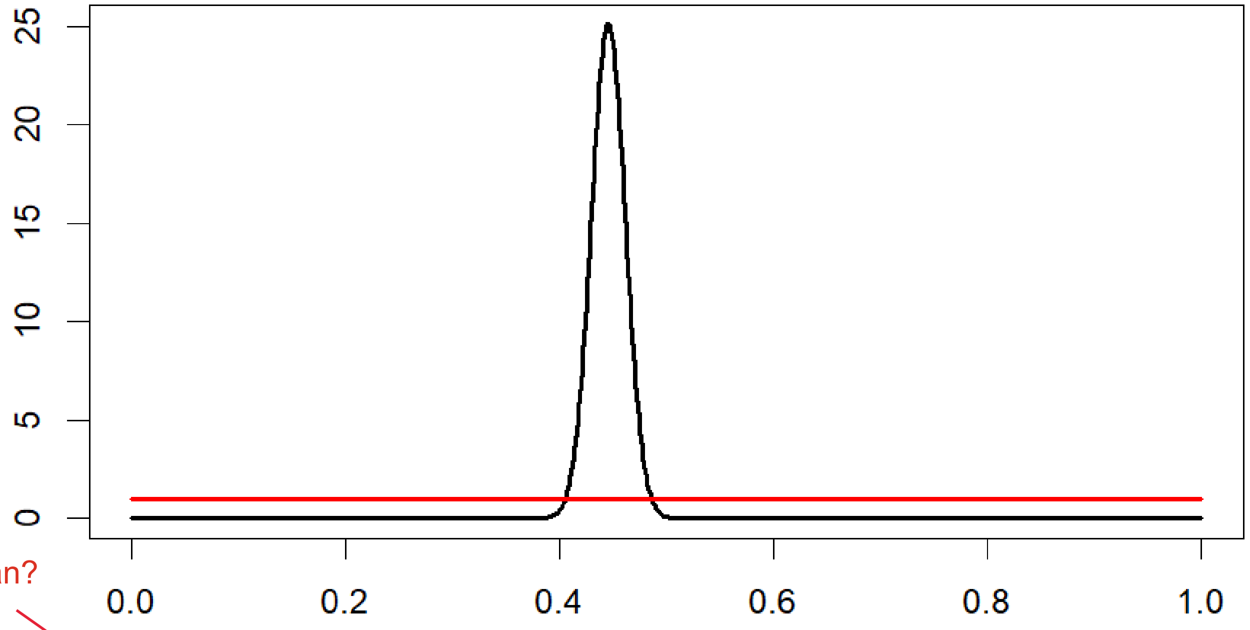
- What about if we had even more data?

This could overwhelm our prior, WHICH IS WHAT WE WANT (in most cases)

The Posterior

Using a non-informative prior $Beta(1,1)$

- Posterior - $Beta(438, 544)$
- EAP: .446 we are missing uncertainty with the EAP (we need location and uncertainty)
- 95% Credible Interval: [.415, .476] Remember the population of females is .485. What does this mean?



If we use a non-informative prior, we could conclude that the rate of female births with placenta previa is lower than the general rate (With extremely high probability)

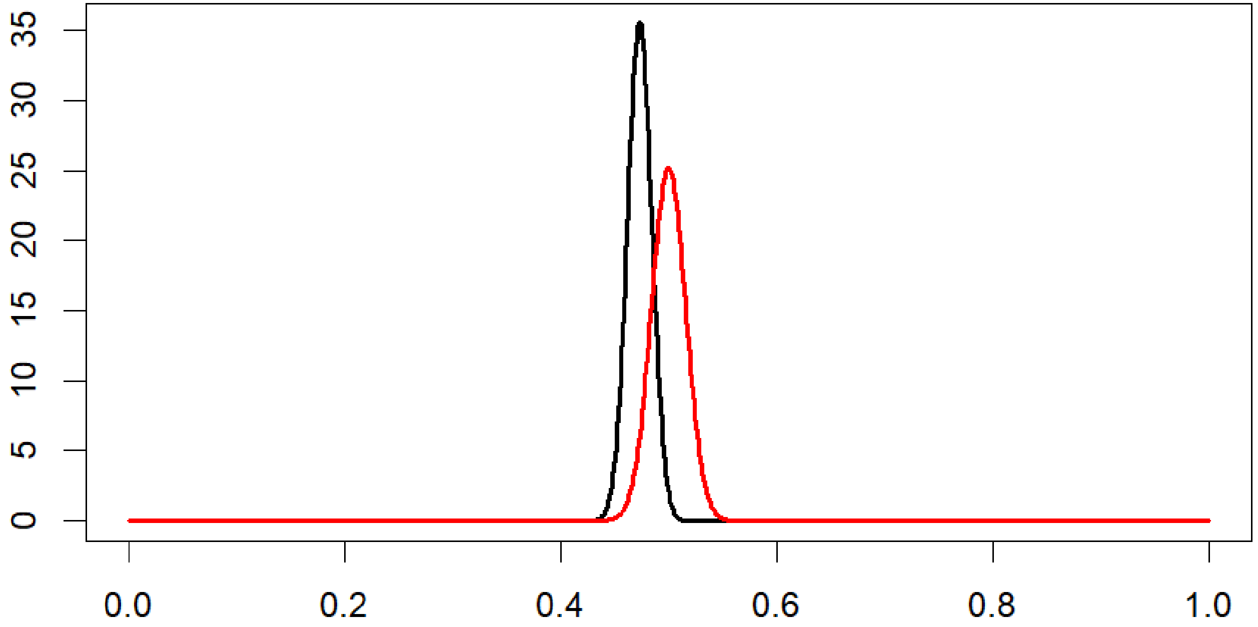
As Bayesians, we can concludet that there is a very strong likelihood that the proportion of female birds w/ placentaprivera is lower than the overall proportion of birds in the population.

What if we use an informative prior...

The Posterior

Using an informative prior $Beta(500,500)$

- Posterior - $Beta(937, 1043)$
- EAP: .473
- 95% Credible Interval: [.451, .495]



If we use an informative prior, we wouldn't conclude the same…

- What's wrong with this prior?

- It's too informative. It is WAY too certain for our data

Although this interval contains .485, we can't conclude this. So, our prior choice very does matter.

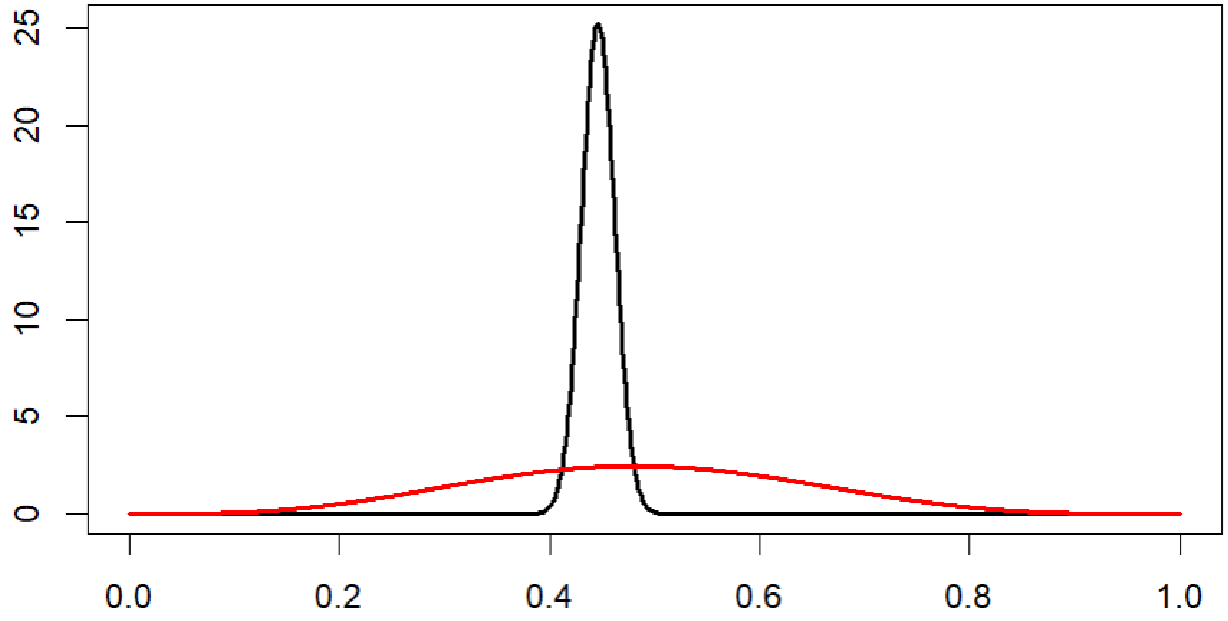
If you chose such a strong beta prior, you better have some evidence b/c of this. This strongly biases our results in a direction that has no results.

Take away: we chose a prior of 1:1 on the last page so that we had the least effect on the distribution

The Posterior

Using a weakly informative prior $Beta(4.85, 5.15)$

- Posterior - $Beta(441.85, 548.15)$
- EAP: .446 This EAP looks very similar to our uninformative prior's EAP
- 95% Credible Interval: [.415, .477]



Placenta Previia

Under a very simple model (binomial likelihood) with uninformative or weakly informative priors, we can conclude that female births are slightly less likely in placenta previa than in the general population.

- So many caveats to this point…

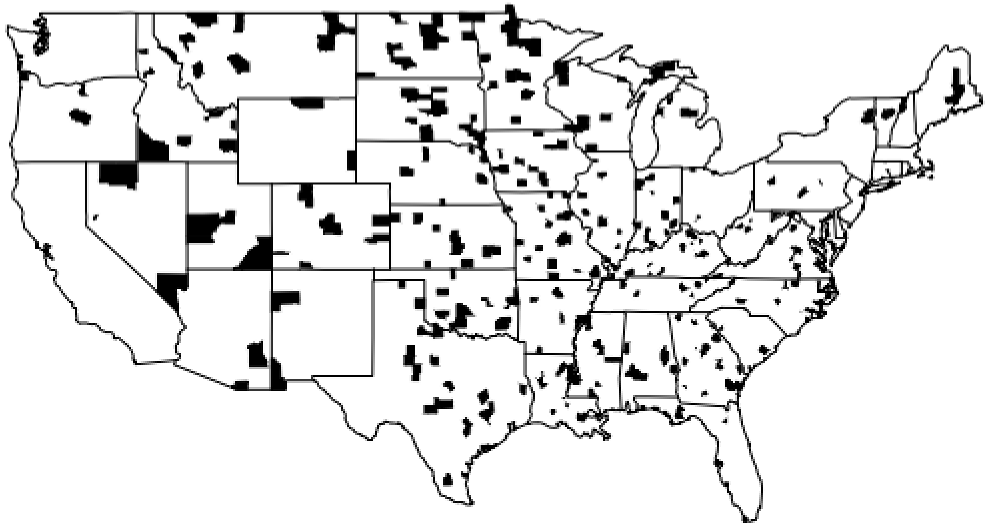
Take away points:

- **Like all statistics, you can think about a prior in terms of *location* and *uncertainty*.**
  - The informativeness of a prior is based on its uncertainty, not it's location.
- As data increases, the prior is overwhelmed.

We saw this in both cases. Now consider the last problem. If we had 10 observations, our prior would be infomrative

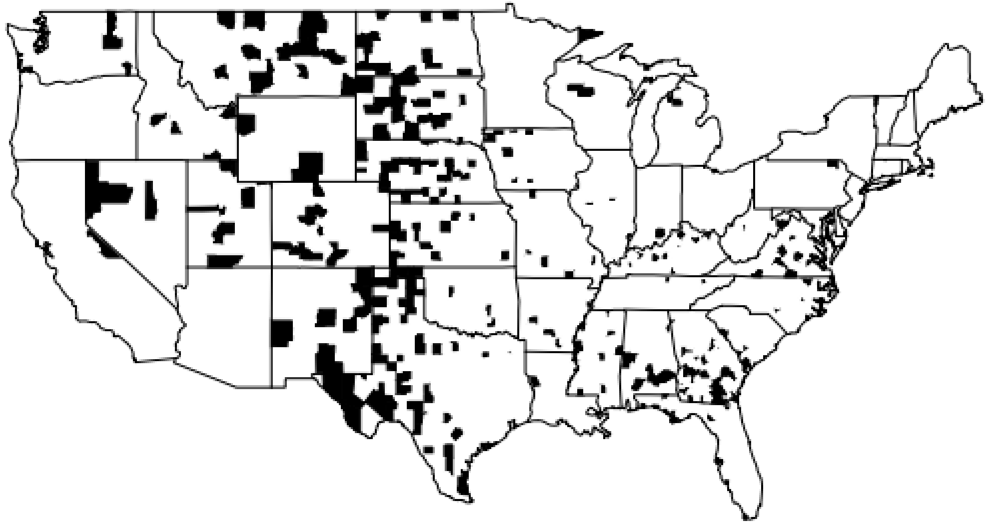
Kidney Cancer in the 1980s

Highest kidney cancer death rates



^This is also looking at the midwest b/c barely any people live there :/

Lowest kidney cancer death rates



^There's not many people in the Midwest anyways, so of course this is going to be the case

Kidney Cancer in the 1980s

As before, we start by defining our model (the likelihood):

$$y_j \sim Poisson(10 n_j \theta_j)$$

Poisson distributions are useful for modeling rates of events.

- Why not use a binomial here?

Data:

- y_j – Number of deaths in the county between 1980-1989
- n_j - Population size of the county

Parameter-

- θ_j- Rate of kidney cancer death in the county per year.

Kidney Cancer in the 1980s

$$y_j \sim Poisson(10 n_j \theta_j)$$

Now we need a prior for θ_j

- It's actually more accurate to say we need a prior for $n_j\theta_j$. Why?

B/C it's all wrapped into the likelihood

The conjugate prior for a Poisson is a Gamma:

$$Gamma(\alpha, \beta)$$

Hyperparameters:

- α – Prior count of the events
- β – Prior time interval

These are not the same alpha and betas in the data distrbutions

Kidney Cancer in the 1980s

$$y_j \sim Poisson(n_j \theta_j)$$
$$\theta_j \sim Gamma(\alpha, \beta)$$

Hyperparameters:

- α – Prior count of the events
- β – Prior population size

Due to conjugacy:

$$\theta_j | y_j, n_j \sim Gamma(\alpha + y_j, \beta + 10 n_j)$$

This is slightly different than the beta distribution

Kidney Cancer in the 1980s

$$\theta_j | y_j, n_j \sim Gamma(\alpha + y_j, \beta + 10n_j)$$

Recall that Gamma distributions have:

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

What is a non-informative prior?

- Balance variance with expected value.

How about an informative prior, when we know that on average the rates of kidney cancer deaths are .0005?

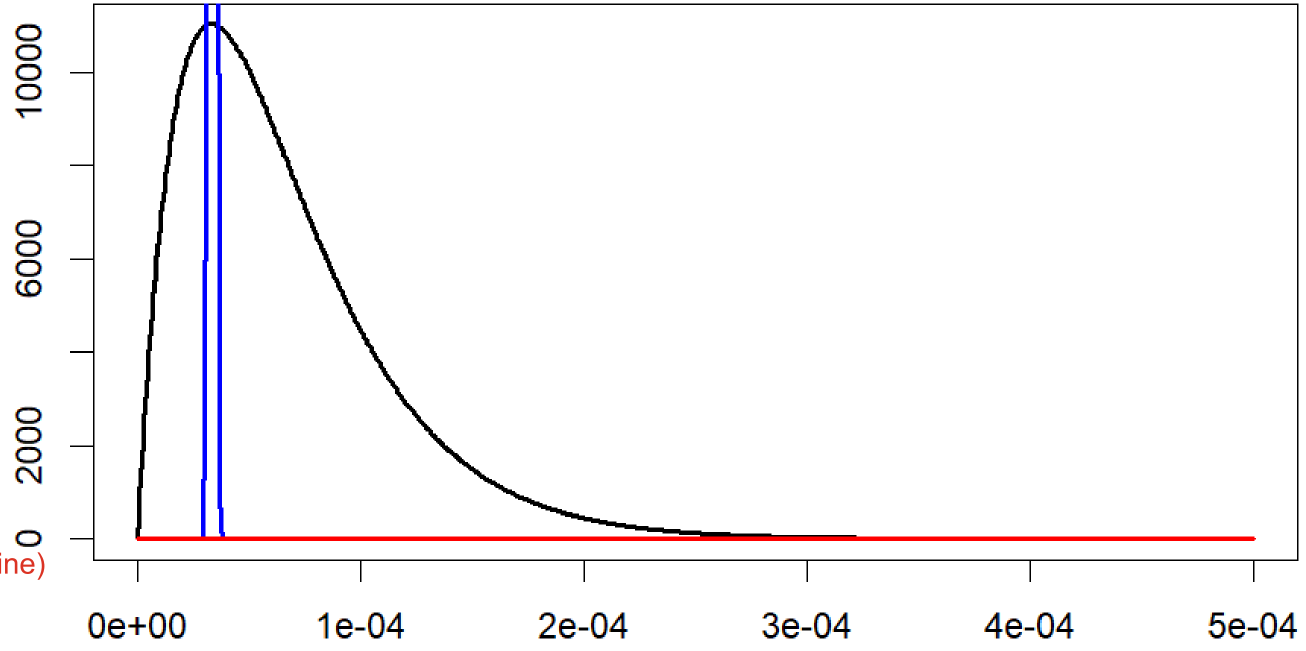
How do we make this informative prior highly informative?

-We'll have to up Beta considerably!

Non-Informative Prior

$$\theta_j | y_j, n_j \sim Gamma(\alpha + y_j, \beta + 10 n_j)$$

- How about a Gamma(1,1)?
- For a low population county
  - 1 death out of 3000
- For a high population county
  - 1000 deaths out of 3 million (blue line)



Is the prior really impacting the posterior?

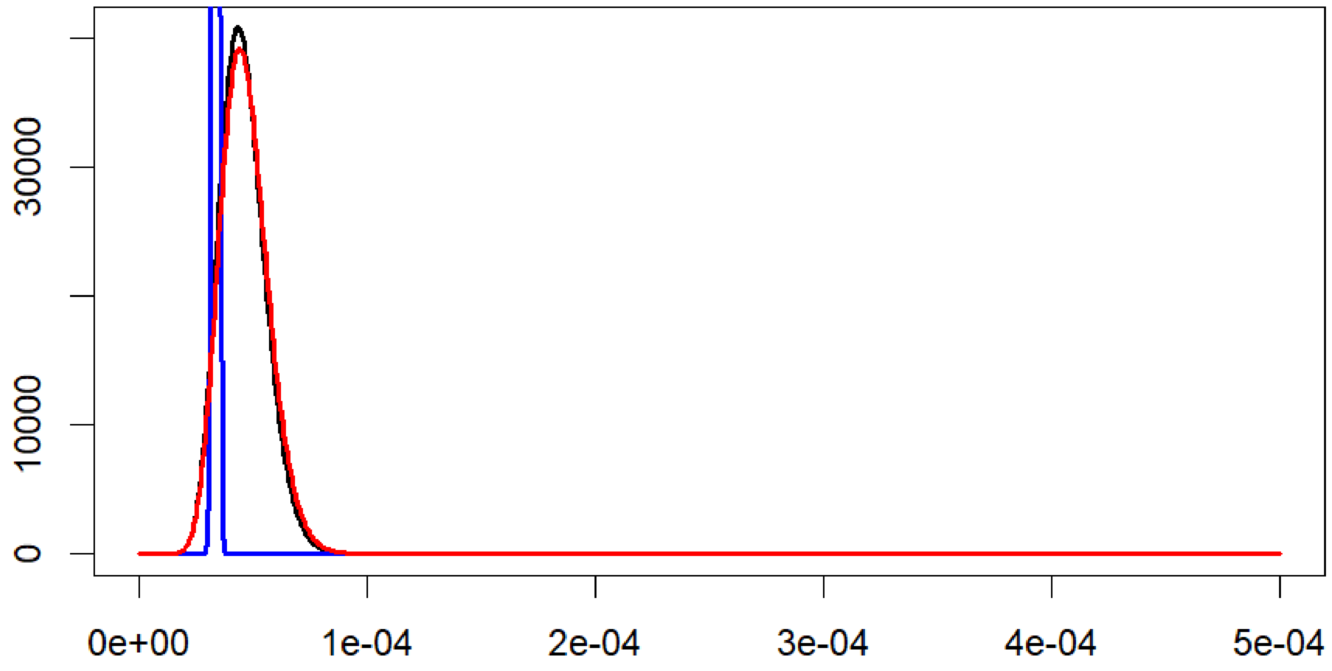
In the low population case, it is. In the high population case, adding 1 isn't really going to change anything

Non-Informative Prior

$$\theta_j | y_j, n_j \sim Gamma(\alpha + y_j, \beta + 10n_j)$$

How about a Gamma(20,430000)?

- Derived from *moment matching*



The blue line is the posterior for the high population

The black line is the posterior for the low population

New Concept: Predictive Distribution

We can calculate/provide the following:

- Posterior – $P(\theta|X)$
- Likelihood - $P(X|\theta)$
- Prior - $P(\theta)$
- Marginal - $P(X)$

But what about the distribution for new observations of X, given our data and priors?

$$P(X|X^*, \theta)$$

We can calculate the posterior, likelihood, and marginal, but what about this distribution, where X* is the data we've already observed?
This predictive posterior distribution is great for predictive models.

What do you have to do to make this happen? You have to give this specific values for our parameters.

Predictive Distribution

For any given parameter value, we can use the likelihood to predict new values of X. But, in Bayesian Statistics, parameters are not single values.

So, we need to integrate over our posterior distribution.

$$P(X|X^*) = \int P(X|\theta)P(\theta|X^*)d\theta$$

In some cases, this distribution has an easy form. But in most cases, you need to sample from this distribution.

1. Sample a value from the posterior of θ
2. Sample a value from the likelihood given the θ you sampled.

This example is taken from BDA3 by Gelman et al.

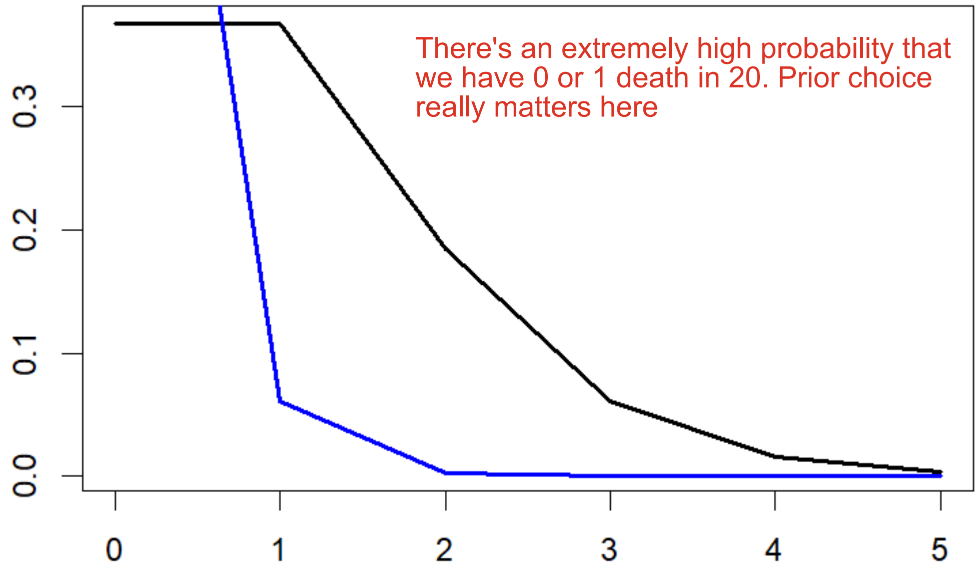
Back to Kidney Cancer

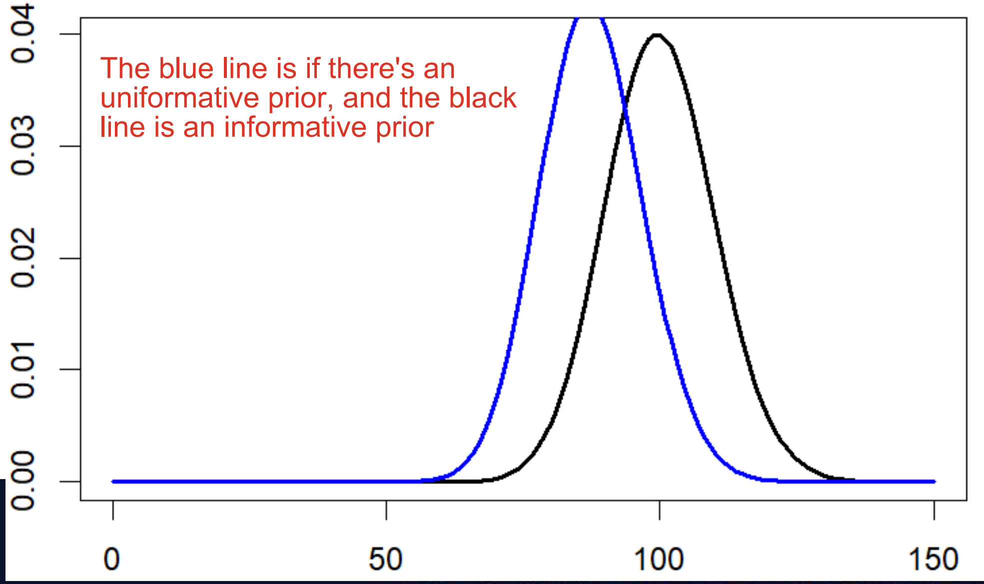
As it turns out, the predictive distribution for a Poisson Likelihood with Gamma Prior is a Negative Binomial

$$NegBin\left(\alpha + 10n_j y_j, \frac{\beta + 10n_j}{\beta + 10n_j + 1}\right)$$

Top: 1 death, 3000 n, $\alpha: 1{,}20$ $\beta: 1{,}4300000$

Bot: 100 deaths, 300000 n





Next Week on Bayes ML

Tuesday

- Bayesian Classification
- Conjugate Bayesian Regression

End notes: choosing a prior is very difficult and flat priors tend to work great in most testing scenarios, but once you start working on proportions and chance, then it becomes more trivila

the more data you have, the more informative the prior

Thursday

- Bayesian Inference and Samplers!