Priors, Part 2

$y \sim N(\theta, 1)$ $p(\theta) \sim c \Leftarrow$ flat prior

$y_1 = 1 \rightarrow P(\theta \mid y_1 > 0) = 0.84$

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)

what data is modeled by

Posterior Distribution
(Probability of Model Given Data)

what want

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

Prior Distribution
(Probability of Model)

Prob of param estimates that reflect what you believe about phenomena

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

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. Doesn't matter location if have huge uncertainty

There are many "classifications" of priors (more data = less prior matters)

- Conjugate priors – priors that result in the posterior being from the same distributional family (Only apply in simple testing cases)
- Informative priors – priors that have something to say (lookup prev. results)
- Weakly/uninformative priors – priors without much information given

<u>Key Point: A posterior represents a compromise between data (observed) and the priors</u>

Placenta Previa

Placenta Previa is a ~~material~~ maternal 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. (yikes...)

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. $F = 437/980$ (make assumption M/F births = 50%)

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.

F = 0.485, PP|F = 437/980

What is a reasonable likelihood for this data?

- Our parameter of interest is a proportion.
- Our data is # of female births out of # of total births.

An appropriate likelihood for this is the *binomial distribution.* (proportion) (counts)

Binomial Likelihood

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

- 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?

1. Parameter: p
2. Prior for p:

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?

Prior Choice - Beta

→ posterior is same family

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

→ prob of prop being observed given α, β

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

→ for binom likelihood, when Beta is applied, is a beta with params

Two *hyperparameters*, α and β

- What do these mean? (are params for beta dist = prior)

Are #successes vs total trials

Prior Choice – Hyperparameters

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

expected val → $$E[p|\alpha,\beta] = \frac{\alpha}{\alpha+\beta}$$ ← $\frac{\text{successes (female births)}}{\text{total \# (F+M births)}}$

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

- What are some informative priors, let say we believe that male births have more prevalence of placenta previa? $\beta > \alpha$ (how large #'s are infl variance)
- What is an "uninformative" prior choice? $\alpha = \beta$ (if big, var=high) so if both 1, will go to flat prior, high var

Beta(1,1)

The Posterior

The Beta distribution is a conjugate prior for the binomial likelihood. This means that the posterior will be distributed as a??? Beta! (same since conj prior)

1 437

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

prior knowledge of female births

prior male births

What do you think is going to happen as the information content in the prior increases? (uncertainty decreases) ($\alpha, \beta \to \uparrow$) posterior looks like prior distribution

- What about if we had even more data?

Could overwhelm the prior, what we want

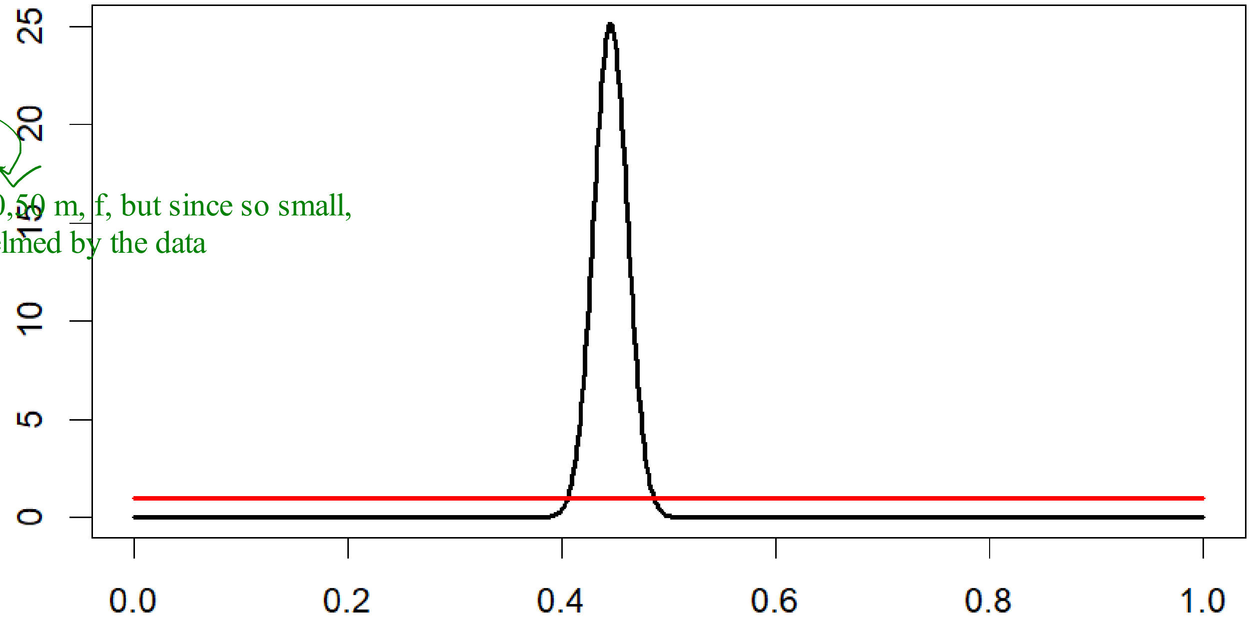
The Posterior

Using a non-informative prior $Beta(1,1)$ (*)for now want non-inf. b/c have no previous info, does slightly matter

it reflects a 50,50 m, f, but since so small, it is overwhelmed by the data

- Posterior - $Beta(438, 544)$ #F+1 #M+1
- EAP: .446 - location (expected val of posterior)
- 95% Credible Interval: [.415, .476] –

0.485 if not in this range (true prop of female births)



Care about prop of female births with PP compared to overall female births

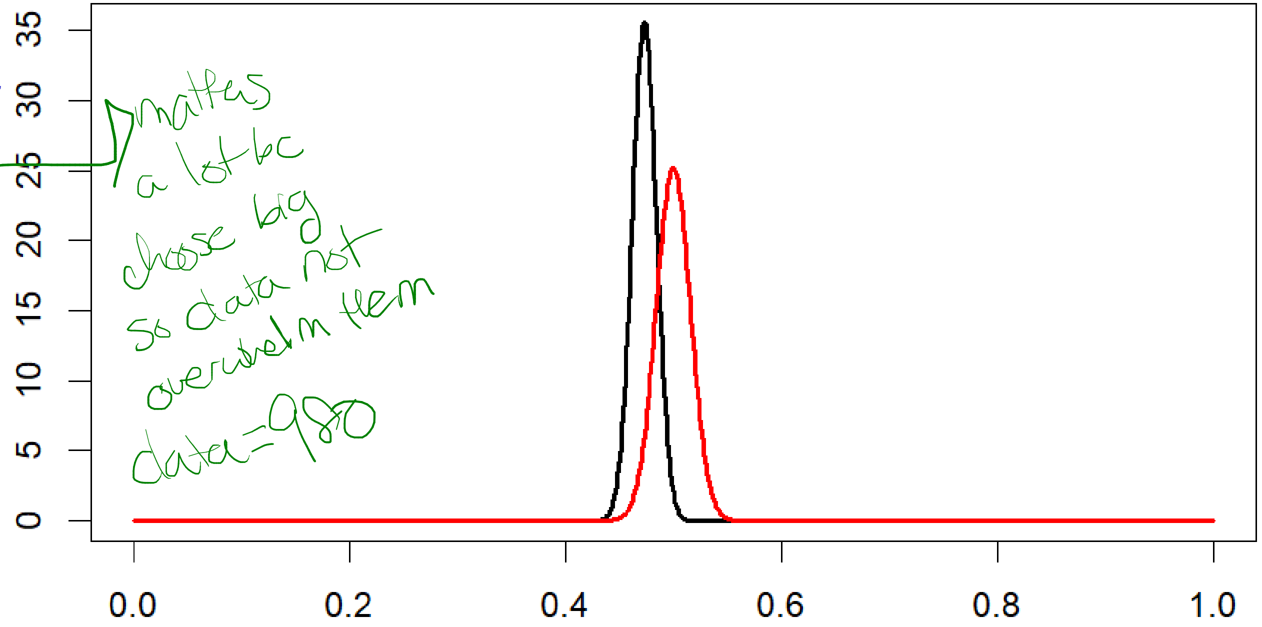
Males more likley to be born under PP than females=conclusion

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

The Posterior

Using an informative prior (bad) - better have justified) not even pop reflective
$Beta(500{,}500)$ $E(\theta|\alpha,\beta) = \frac{1}{2}$
(saying we expect out of 1000 births, 500 F 500 M

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



matters a lot bc choose big so data not overwhelm them
data = 980

If we use an informative prior, we wouldn't conclude the same… Contains 0.485, so can't conclude

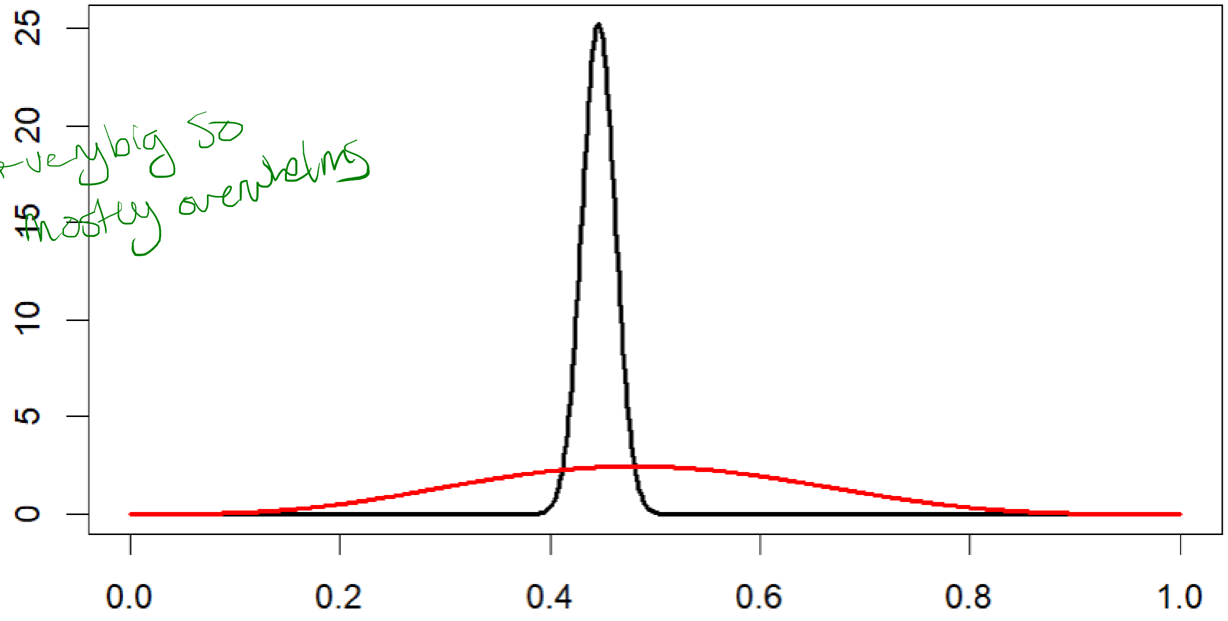
- What's wrong with this prior?

The Posterior

Using a weakly informative prior $Beta(4.85, 5.15)$ E( ) =0.485

Out of 10 births, 4.85 = F 5.15 = M, reflective of pop., but not very big so data mostly overwhelms

- Posterior - $Beta(441.85, 548.15)$
- EAP: .446
- 95% Credible Interval: [.415, .477] → 0.485 not in the interval, so we could say that rate of F w/ PP is lower than prop of female births overall



IF did BETA(485,515) it would be highly informative bc big and so data would not overwhelm the prior

* If data = 10 observations, this example would be informative

Placenta Previa

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.

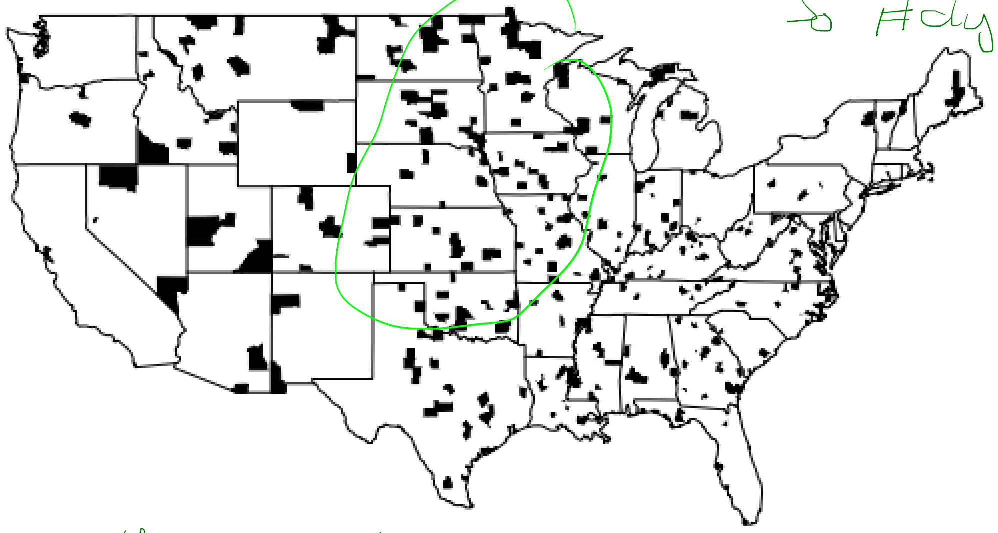
- So many caveats to this point… only from germany

Take away points:

- **Like all statistics, you can think about a prior in terms of *location* and *uncertainty*.** matters more
  - The informativeness of a prior is based on its uncertainty, not it's location.
- As data increases, the prior is overwhelmed. prior matter less as $n \to \infty$

Kidney Cancer in the 1980s

Highest kidney cancer death rates (not very many pple, so #dying/#ppl)

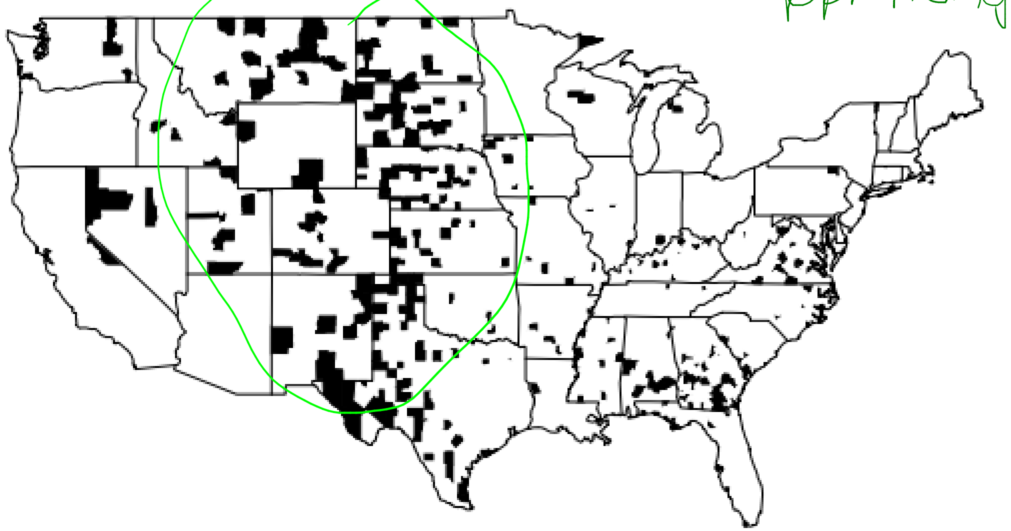


define the model:

#pple w/out/total

(bc not many ppl living there)

Lowest kidney cancer death rates



Kidney Cancer in the 1980s

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

$\lambda = 10 n_j \theta_j$

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

j = county

10 yrs, per yr — pop — rate

Poisson distributions are useful for modeling rates of events.

- Why not use a binomial here? Have # in county, but we care rate over time controlling for overall pop

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 (need prior for $n_j \theta_j$ = more accurate)

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

The conjugate prior for a Poisson is a Gamma: (looked it up)

easier to walk thru math

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

Hyperparameters:

- α – Prior count of the events
- β – Prior time interval ← not as nice interpretation

Kidney Cancer in the 1980s

$$y_j \sim Poisson(n_j \theta_j)$$ ← likelihood

$$\theta_j \sim Gamma(\alpha, \beta)$$ ← prior

Hyperparameters:

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

Due to conjugacy:

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

rate

posterior

pop size

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.

informative

(α, β)

$(1, 1)$

rel to # deaths

being strong on # deaths =1

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

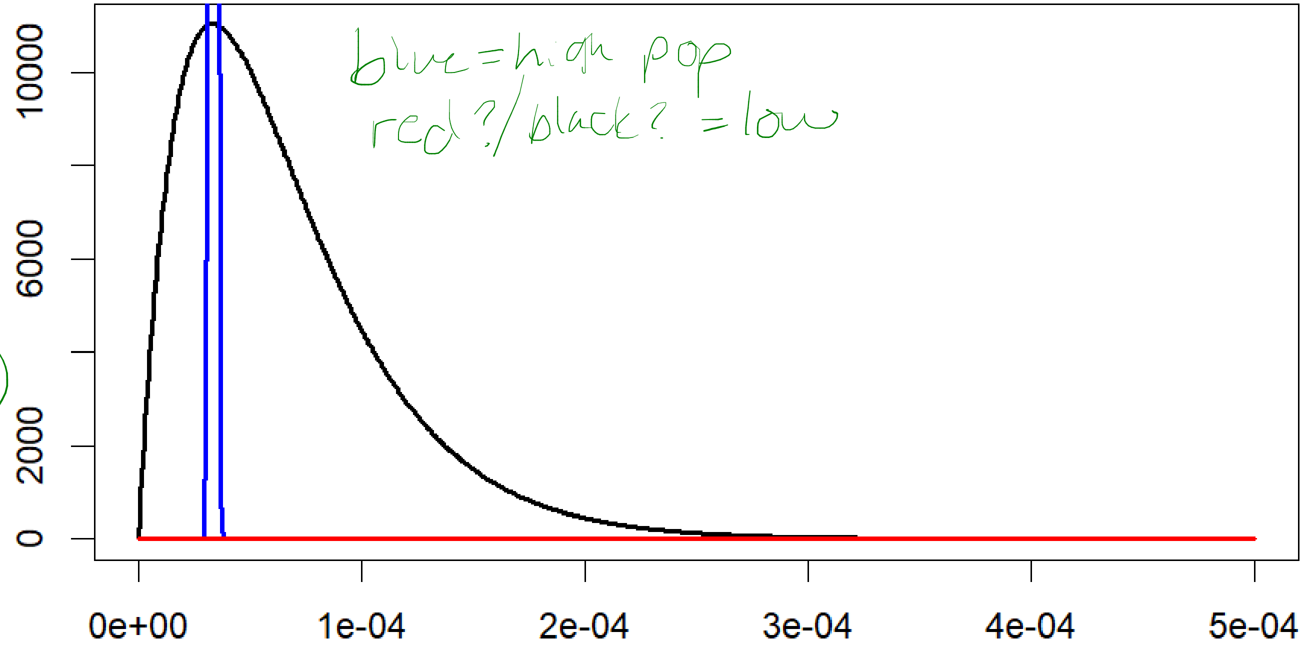
How do we make this informative prior highly informative?

choose α, β so get E we want, but high variance

Non-Informative Prior

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

- How about a Gamma(1,1)?
  (uniform)
- For a low population county
  - 1 death out of 3000 (1 person dies over 10 yrs)
- For a high population county
  - 1000 deaths out of 3 million
    (1000 deaths)



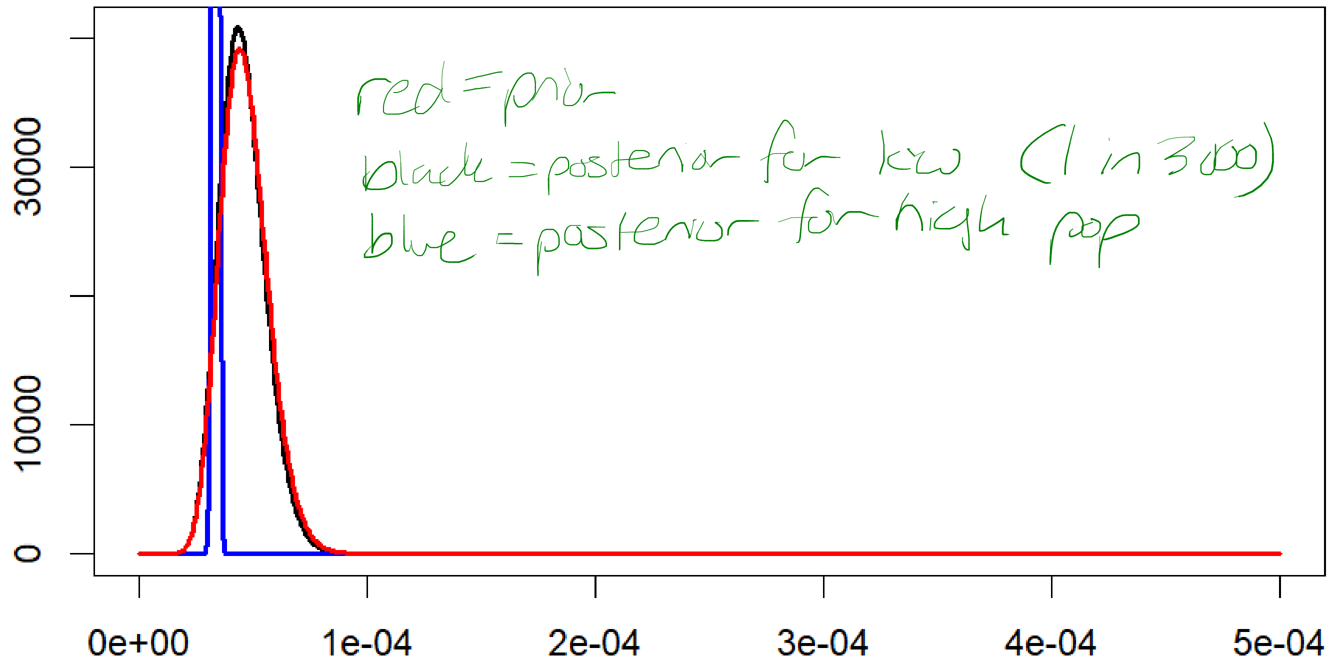
Is the prior really impacting the posterior? In low pop, it is
In high pop, not as much

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*



This example is taken from BDA3 by Gelman et al.

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?

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

X^{*} = data we observed already

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.

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

likelihood integrated over post?

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

plug into likelihood, get dist

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

Back to Kidney Cancer

As it turns out, the predictive distribution for a Poisson Likelihood with Gamma Prior is a Negative Binomial ← # of failures going to see before get a success or certain # successes

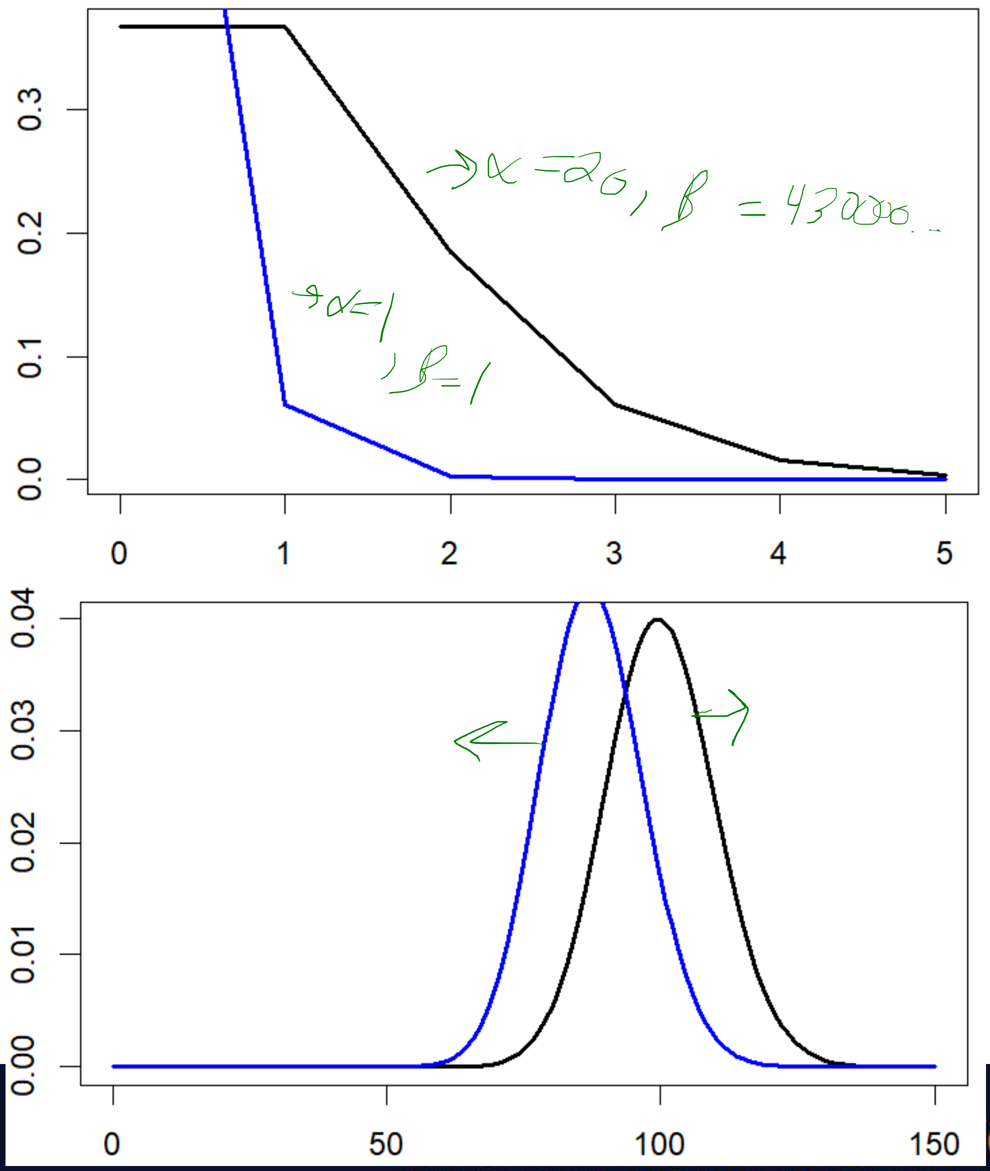
$$NegBin\left(\alpha + 10n_j y_j, \frac{\beta + 10n_j}{\beta + 10n_j + 1}\right)$$

Top: 1 death, 3000 n, α: 1,20

β: 1,4300000

↓ 1, then 20

Bot: 100 deaths, 300000 n



This example is taken from BDA3 by Gelman et al.

Next Week on Bayes ML

Tuesday

- Bayesian Classification
- Conjugate Bayesian Regression

Thursday

- Bayesian Inference and Samplers!

Diff priors = might be informative based on data

Think what model, what know

General rule, flat prior works great

$n \to \infty$, more overwhelms priors