Bayesian perspectives for epidemiological research:
1. Foundations and basic methods.

Sander Greenland, International Journal of Epidemiology, 2006. 35:765-775 (with discussion)

Cancer Prevention and Control Journal Club

September 13, 2007

Presenter: Liz Garrett-Mayer

Introduction to Bayesian Statistics

- Who knows what Bayesian statistics is?
- Who remembers what Bayes' theorem is?
- (We'll get back to Bayes' theorem later)
- "**Bayesian inference** is statistical inference in which evidence or observations are used to update or to newly infer the probability that a hypothesis may be true" (wikipedia)
- Critical idea: **UPDATE**

Key Ideas of Bayesian Approach

- **Prior**: What we believe about unknown parameters BEFORE observing data

- **Posterior**: What we believe about unknown parameters AFTER observing data

Update?

- The data UPDATES our prior beliefs
- How is the data represented?
  - The likelihood function
  - A model-based representation of the data
- Prior x likelihood = posterior
  - (technically: Prior x likelihood $\propto$ posterior)
- What the heck does that mean?
- Let's use the first example...

Example from Greenland

Table 1 Case–control data on residential magnetic fields and childhood leukaemia (Savitz *et al*., 1988) and frequentist results

	$X = 1$		$X = 0$
Cases	3	(8.3%)	33
Controls	5	(2.5%)	193

Let's start with what we know

- Likelihood
- This is the probability model that we assume
- Setting:
  - Large N
  - OR (or RR)
- Common approach:
  - Use a normal approximation
- Assume log(RR) ~ Normal
  - Mean = log(ad/bc)
  - Variance = 1/a + 1/b + 1/c + 1/d
- Familiar? I HOPE SO!

Greenland's calculations

- Mean = log(3*193/(5*33)) = 3.51
- Variance = 1/3 + 1/33 + 1/5 + 1/193 = 0.57
- Then what?
- 95% CI: $\log(\hat{\psi}) \pm 1.96\sqrt{\text{var}_{\psi}}$

$$\log(3.51) \pm 1.96\sqrt{0.57} = (-0.22, 2.74)$$

- Exponentiate:

$$\exp(-0.22, 2.74) = (0.80, 15.4)$$

"What the data says"

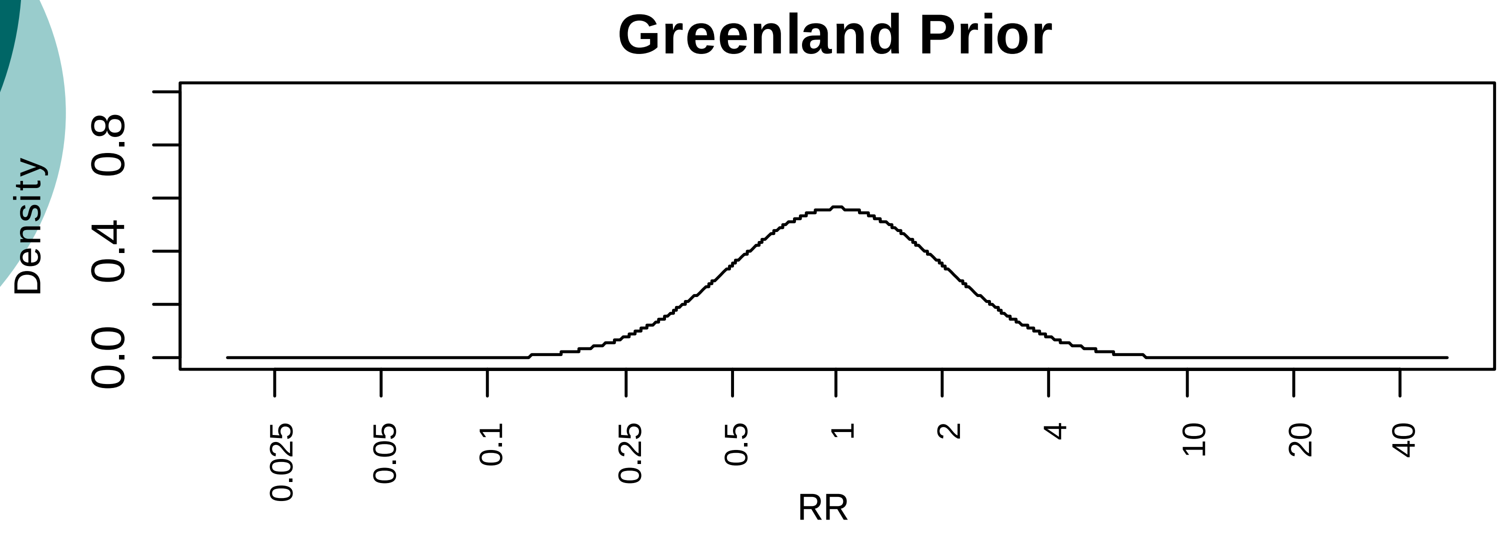
- OR = 3.51
- 95% CI: (0.80, 15.4)

- FREQUENTIST SO FAR……

What about this “Prior” business?

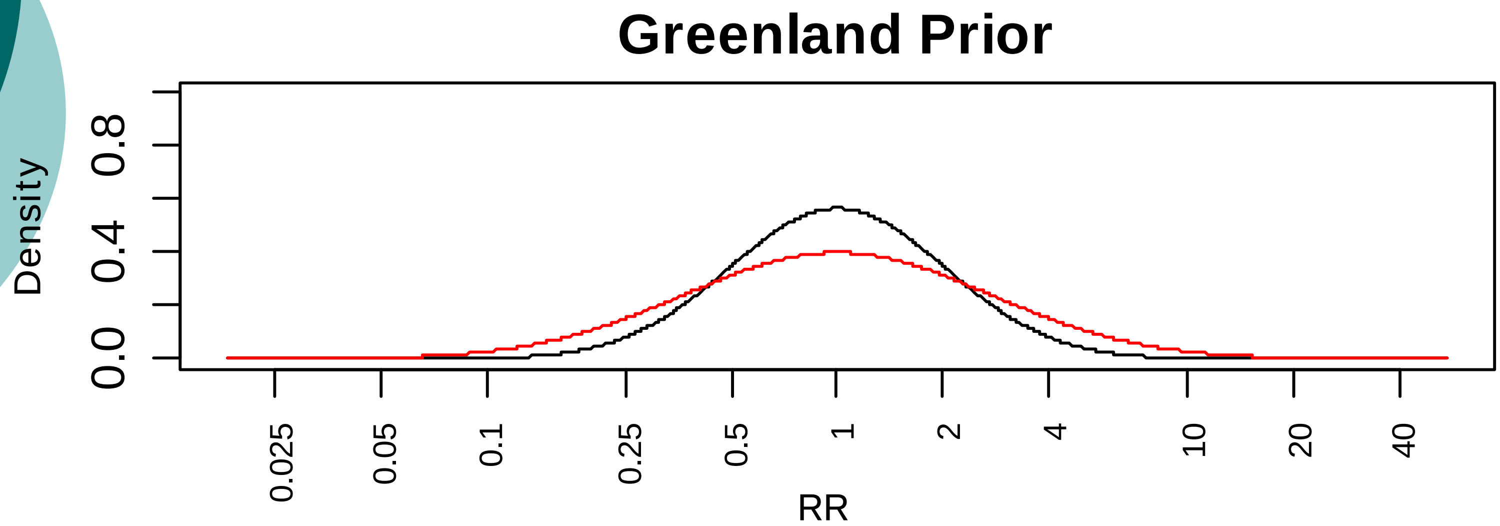
- The prior QUANTIFIES what we <u>**know**</u> or <u>**expect**</u> before doing the study
- Frequentist: “I know nothing”
  - Regardless of prior research
  - Regardless of simple intuition
- Bayesian: “I know something”
  - Might be quite a lot
  - Might be not much, but still quantifiable

Priors

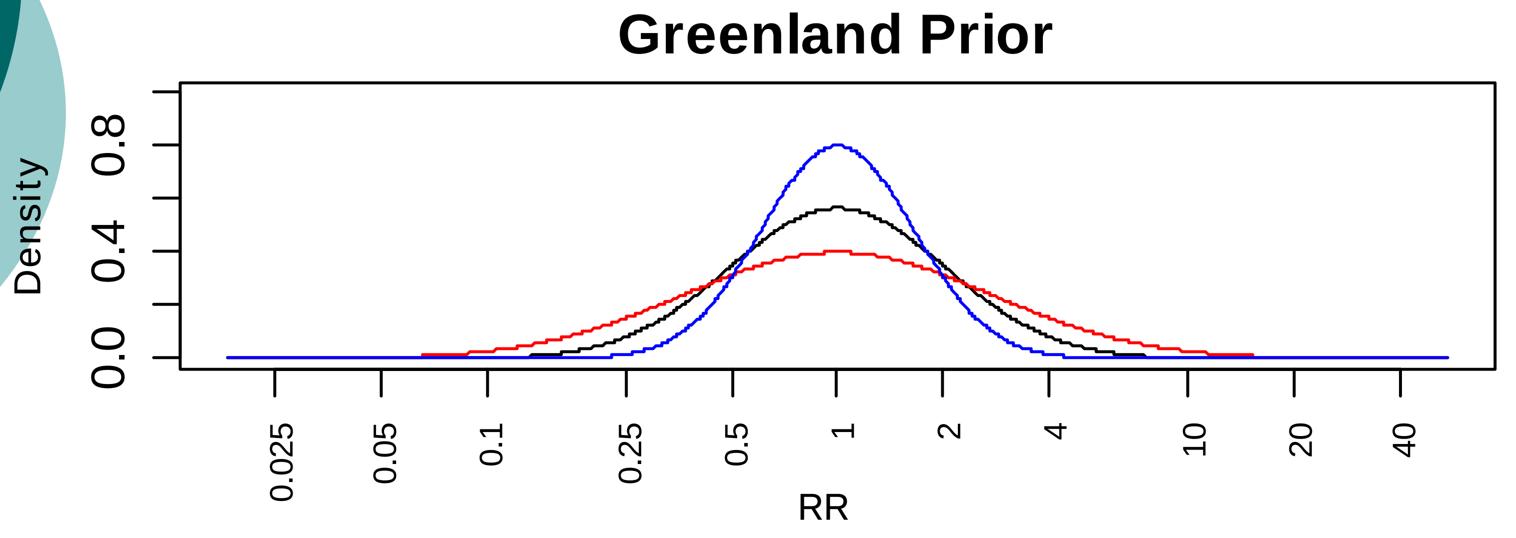


Greenland Prior
Density
0.0
0.4
0.8
0.025
0.05
0.1
0.25
0.5
1
2
4
10
20
40
RR

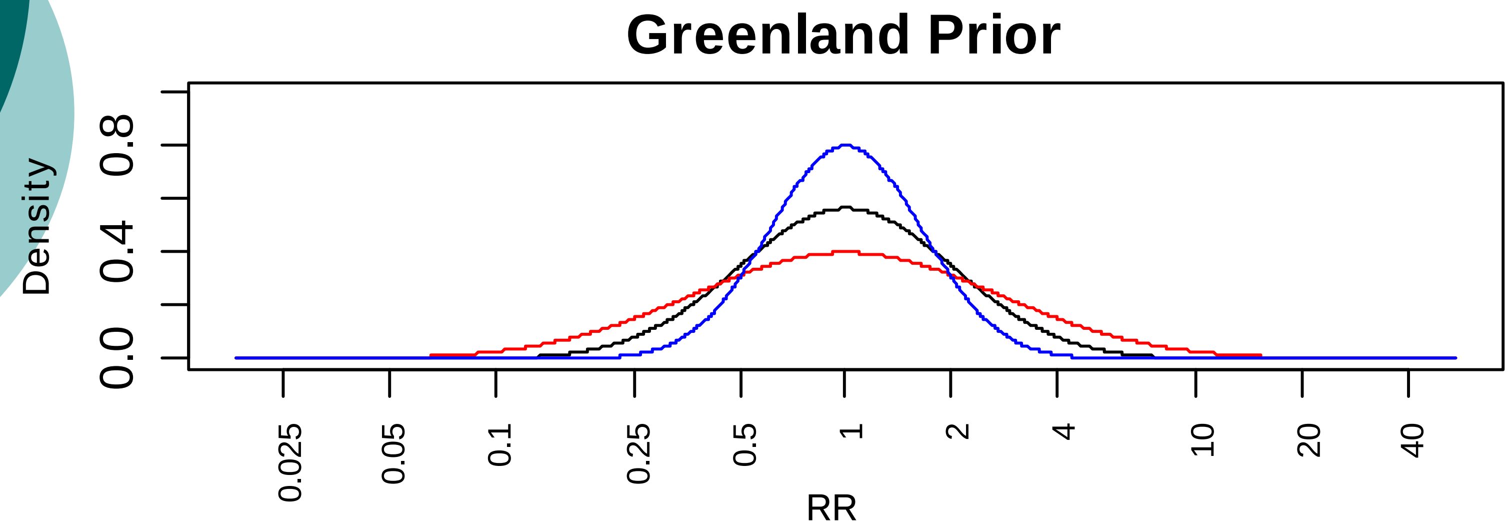
Priors

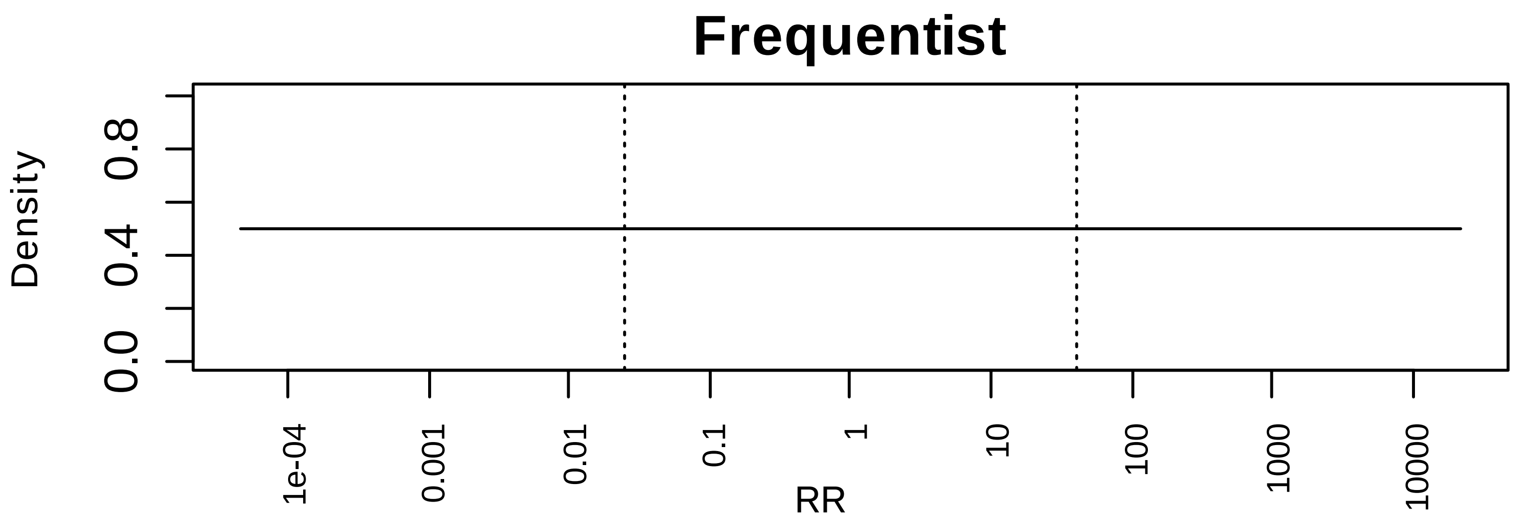

Greenland Prior
Density
0.0
0.4
0.8
0.025
0.05
0.1
0.25
0.5
1
2
4
10
20
40
RR

Priors


Greenland Prior
Density
0.0
0.4
0.8
0.025
0.05
0.1
0.25
0.5
1
2
4
10
20
40
RR

Priors


Greenland Prior
Density
0.0
0.4
0.8
0.025
0.05
0.1
0.25
0.5
1
2
4
10
20
40
RR


Frequentist
Density
0.0
0.4
0.8
1e-04
0.001
0.01
0.1
1
10
100
1000
10000
RR

What do you think?

- Is it reasonable to exclude
  - RR's>10?
  - RR's>100?
  - RR's>1000?
- Frequentists exclude nothing

Getting the posterior

- Combine the prior and the data (likelihood)
- How?
  - Multiply the prior x likelihood OR
  - Add the log-prior + log-likelihood
- Resulting distribution
  - posterior OR
  - log-posterior
- Sound hard?
  - For some prior-likelihood combinations, very simple
  - "Conjugate" prior
- Example: normal is easy

Posterior Distribution

- Define the following:

$$\log(\hat{\psi}) \sim N(\mu, \sigma^2)$$

$$and$$

$$\mu \sim N(m, \tau^2)$$

Note that σ is the standard error!

- Then, the distribution of the posterior is:

$$N\left(\frac{\frac{m}{\tau^2} + \frac{\log(\hat{\psi})}{\sigma^2}}{\frac{1}{\tau^2} + \frac{1}{\sigma^2}}, \left(\frac{1}{\tau^2} + \frac{1}{\sigma^2} \right)^{-1} \right)$$

OUCH!

Not as bad as it seems….

- m = 0, $\tau^2 = 0.5$
- Posterior becomes:

$$N\left(\frac{\dfrac{\log(\hat{\psi})}{\sigma^2}}{2+\dfrac{1}{\sigma^2}},\left(2+\frac{1}{\sigma^2}\right)^{-1}\right)=N\left(\frac{\log(\hat{\psi})}{2\sigma^2+1},\frac{\sigma^2}{2\sigma^2+1}\right)$$

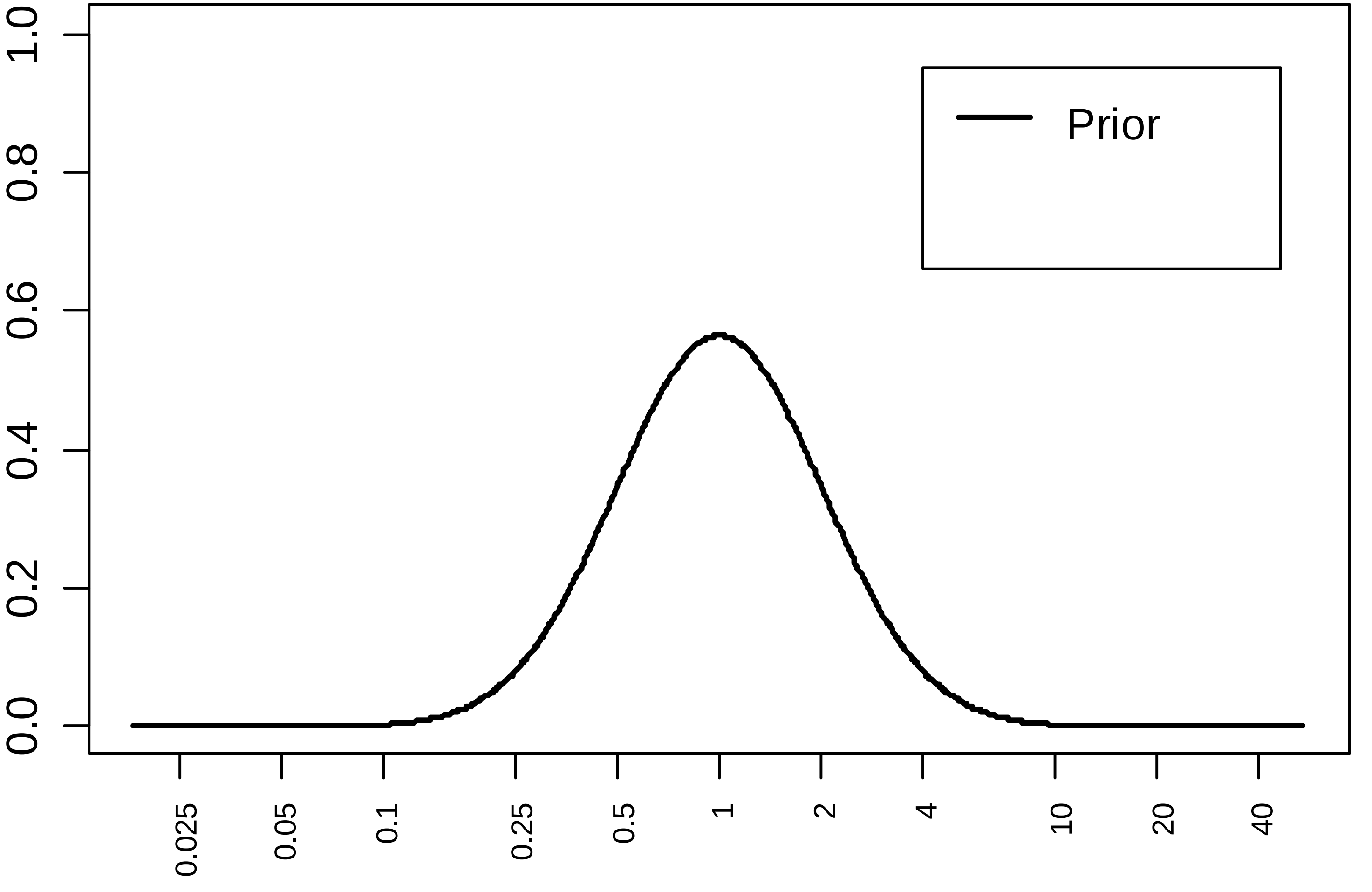
For large n, what happens?

Can you take just one more mathy slide?

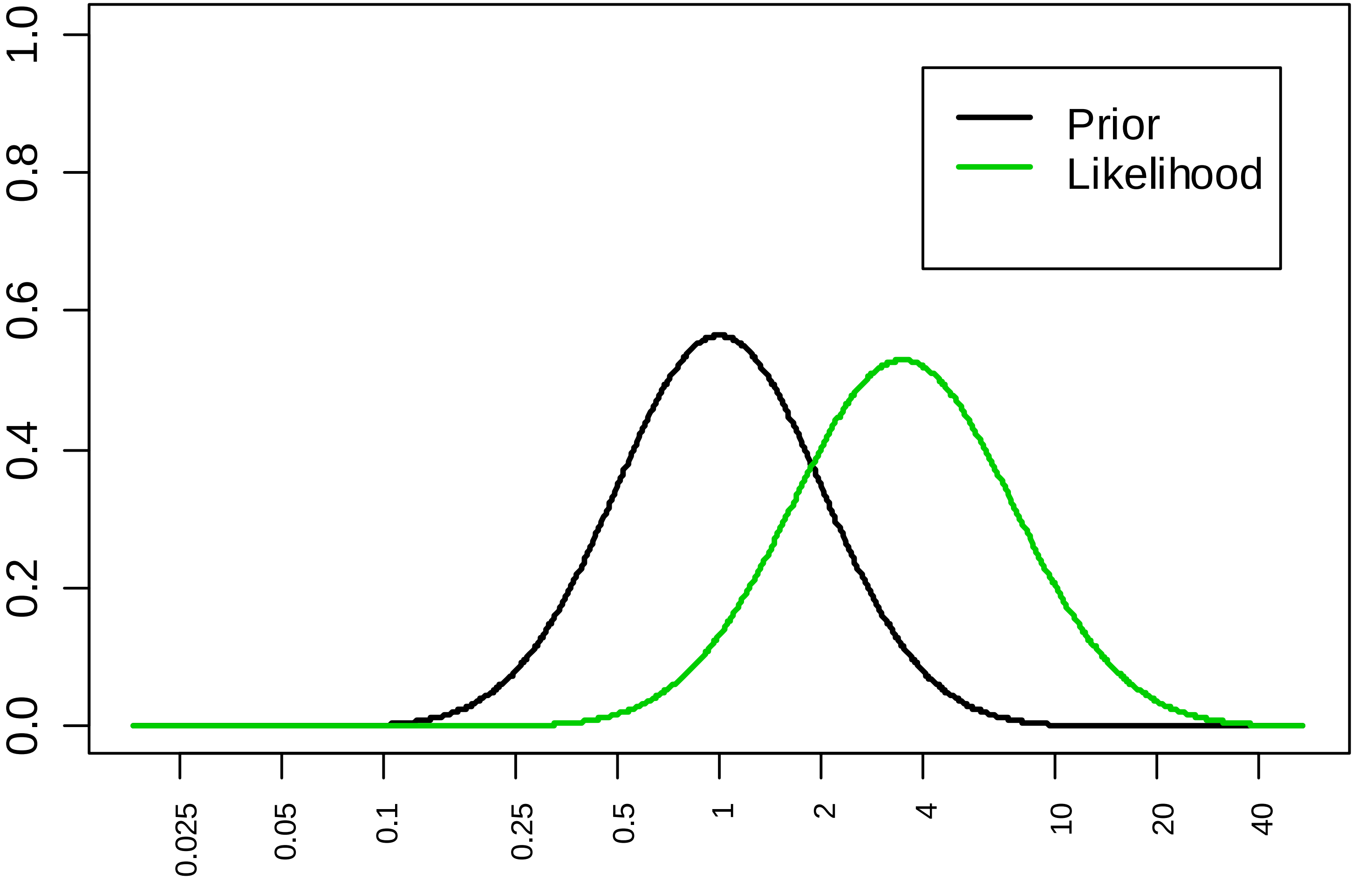
- Note our estimates:
  - $\psi = \log(3.51)$
  - $\sigma^2 = 0.57$

$$N\left(\frac{\frac{\log(\hat{\psi})}{\sigma^2}}{2+\frac{1}{\sigma^2}}, \left(2+\frac{1}{\sigma^2}\right)^{-1} \right) = N(0.59, 0.27)$$

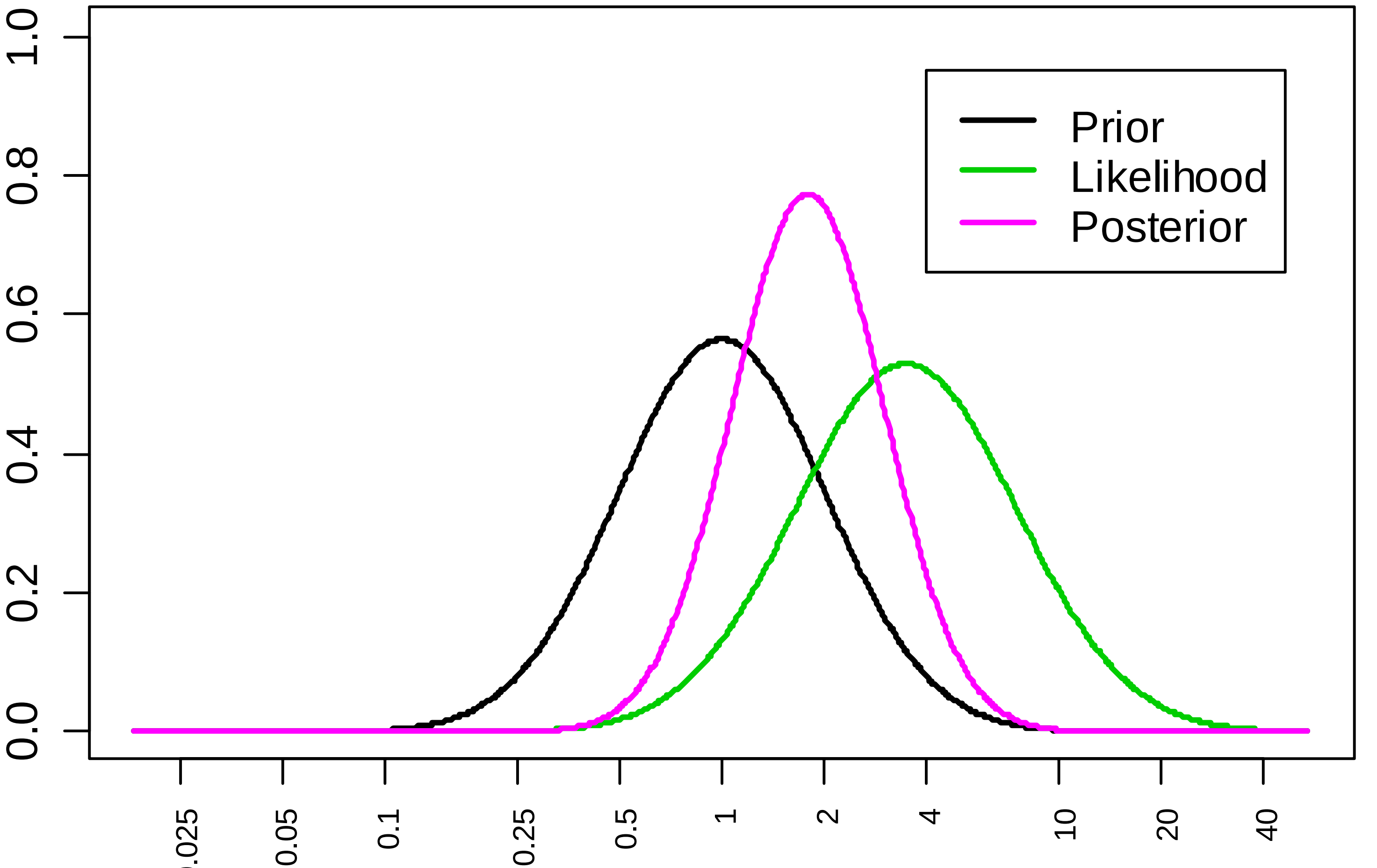
Maybe a picture would bring us back to earth…..



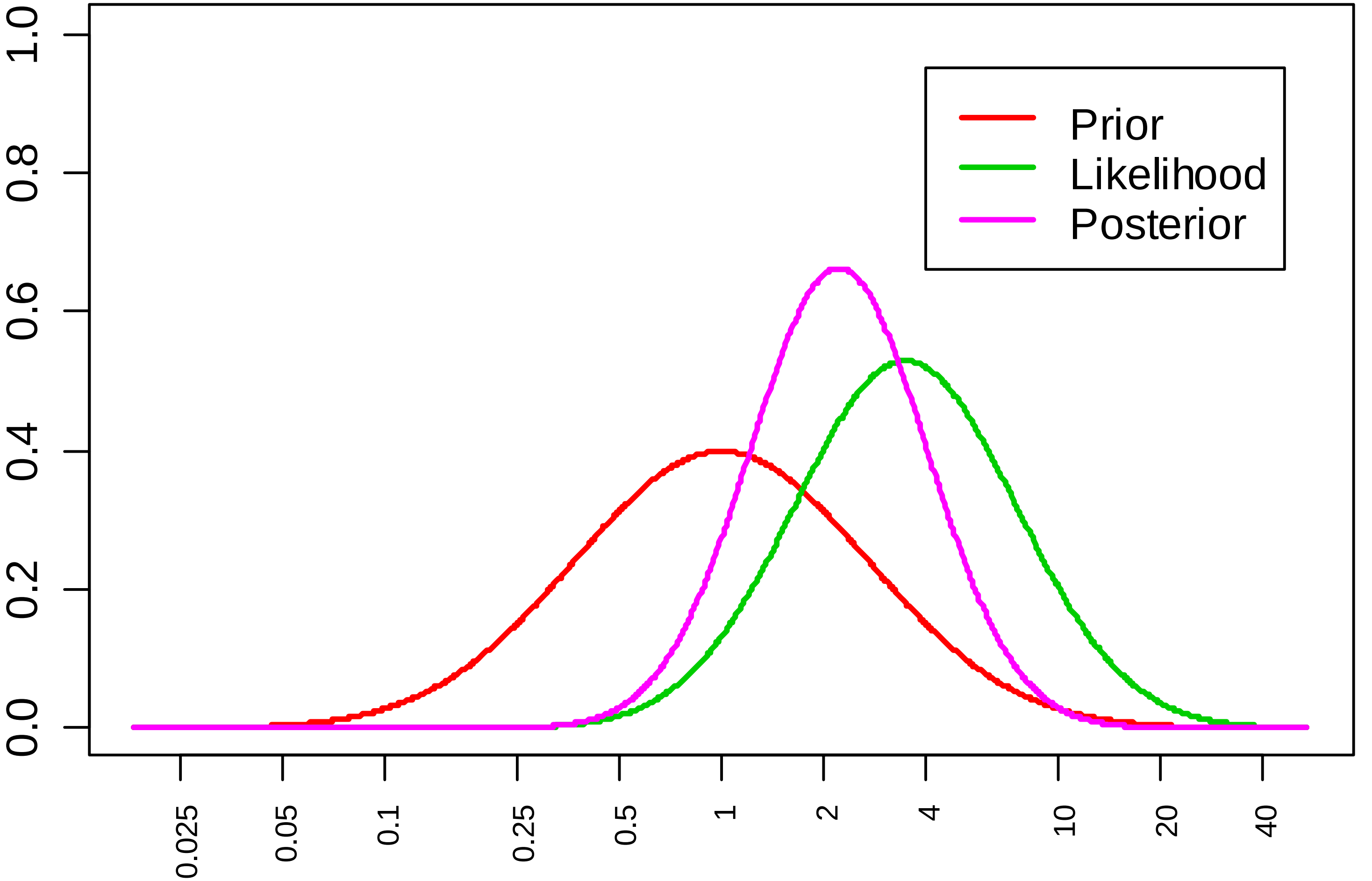
Maybe a picture would bring us back to earth…..



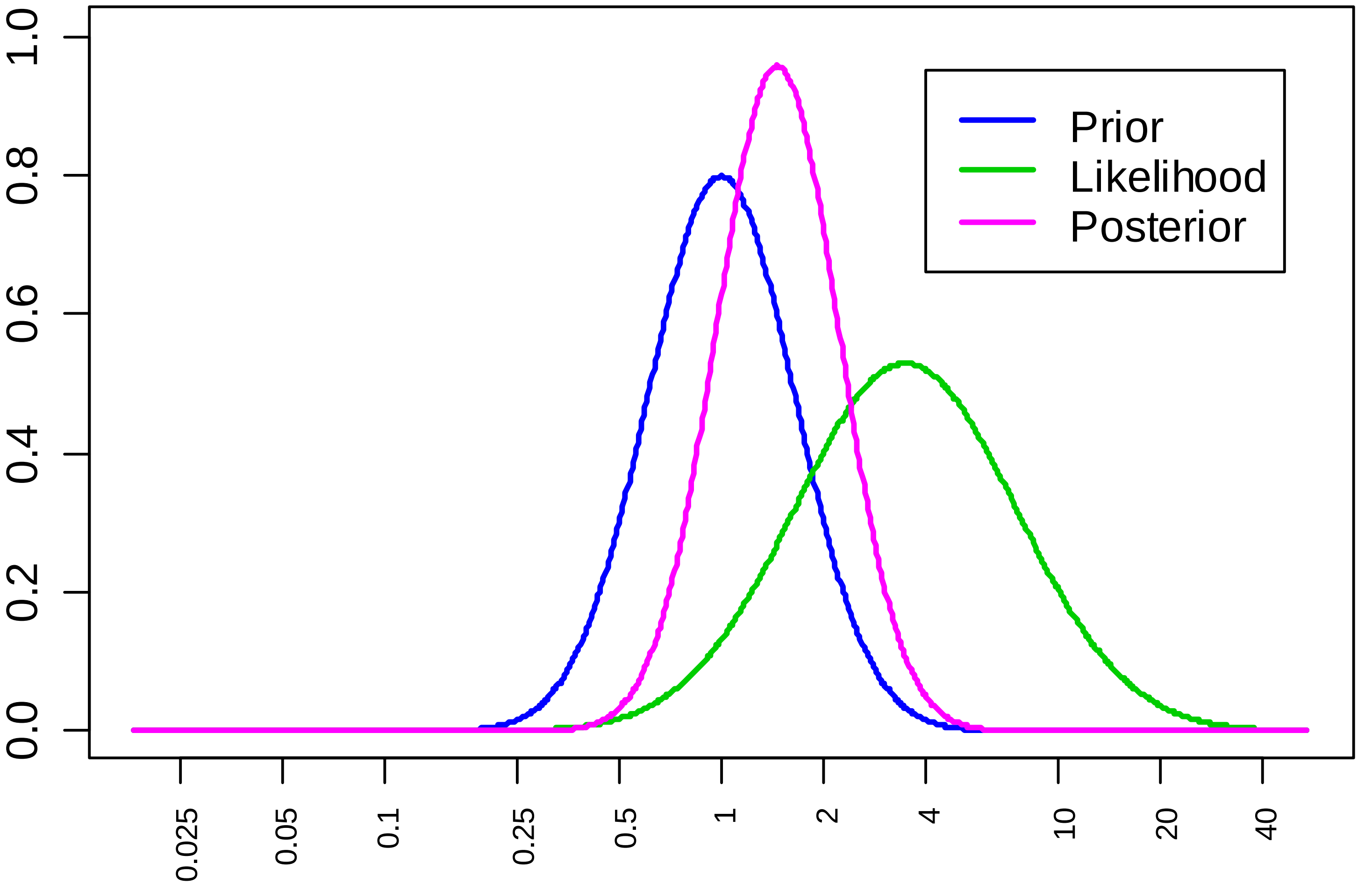
Maybe a picture would bring us back to earth…..



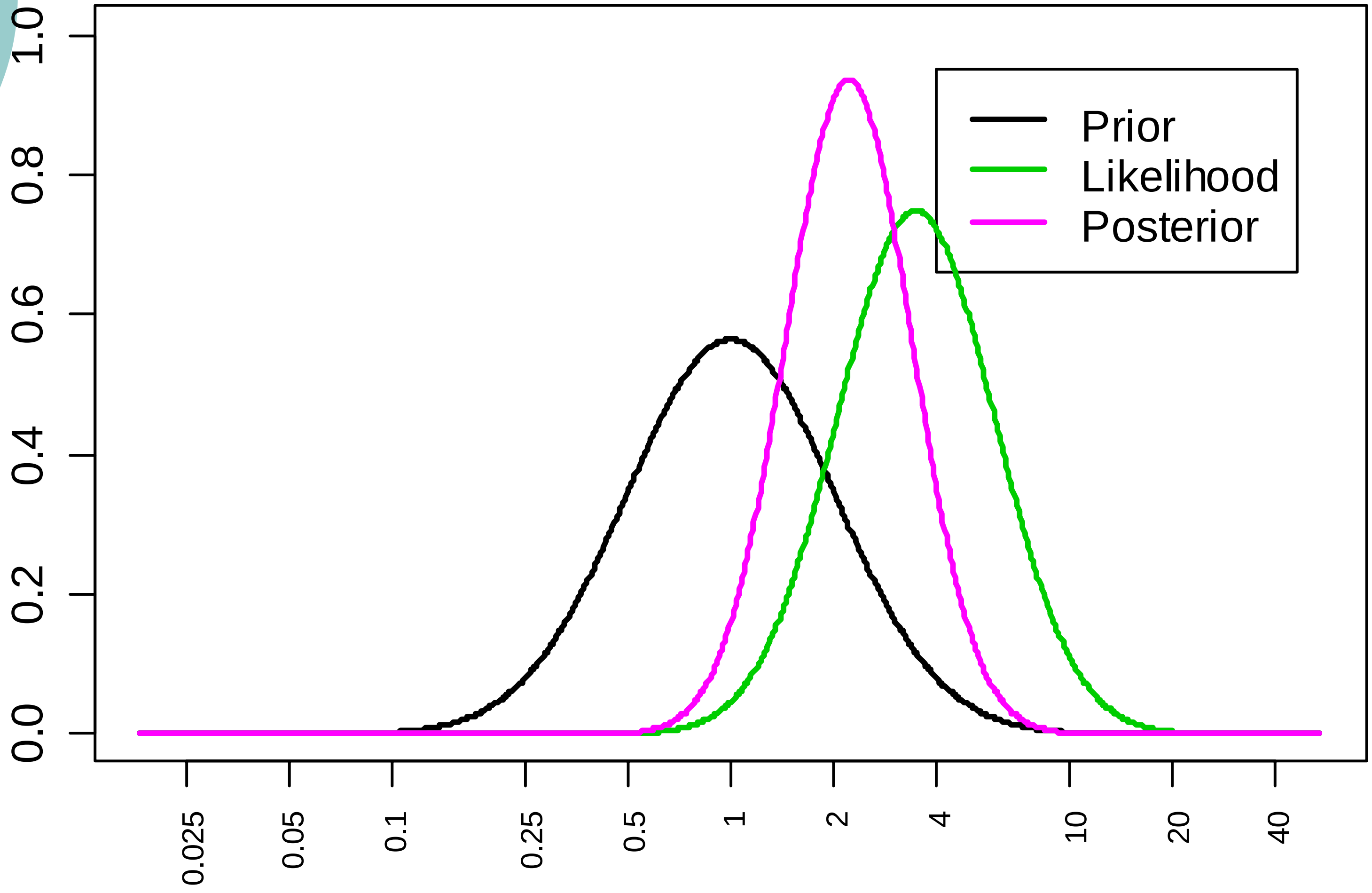
What about other priors: flatter



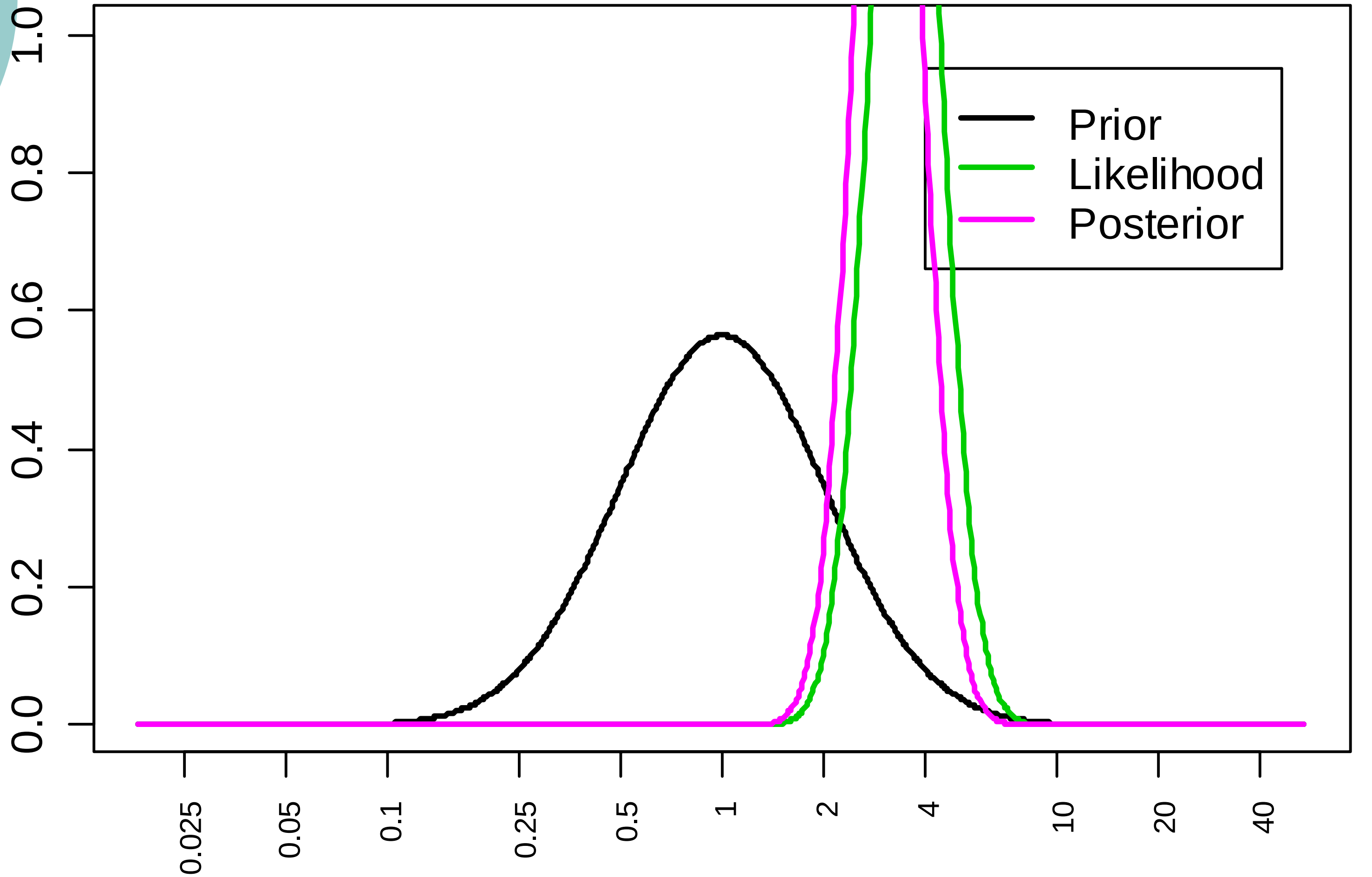
What about other priors: steeper



What if we had more data?
What if we had twice as much data?



What if we had more data?
What if we had 10 times as much data?



So….easy, right?

- There are even simpler approximations
- Combining the prior and likelihood:
  - Computationally easy
  - Statisticians have already done all the hard part!
  - Approximations work pretty well!

OK…so what does Sander Greenland have to say about all of this?

- A LOT!
- Bayesian statistical methods have been around for a long time
- They were “pushed out” awhile back in favor of frequentist methods
  - ‘randomized’ trials and surveys
  - computation
- However, epidemiologic studies are generally not randomized
- Bayesian approaches allow you to ‘adjust’ for some of the subjectivity involved
  - control for bias
  - adjust for knowledge outside the trial

Education in Bayesian approaches is lacking

- Some think that the Bayesian approach is too complicated
- Not really….needs to be put into context
- Balance of objective versus subjective
- Consideration of implications of the prior is a useful exercise in many situations
- Focus has been on the computational differences
- But, the focus should be on the philosophical differences

Why Epi?

- ‘Huge uncertainties’
  - measurement error
  - process generating the data
  - lack of information about confounders
  - lack of knowledge of what potential confounders may be
- Precise computation ‘not necessary’
- Informative priors can help!

Frequentism vs. Subjective Bayes

- Arguments against Bayes:
  - assumptions and models are subjective
  - propaganda (via the prior)
- Actually: these can also be considered arguments against Frequentism
  - frequentists pick models too
  - recall the ‘frequentist’ prior (or lack thereof
- Bayesians are often just more explicit which ends up encouraging criticism.

Priors

- Not arbitrary
- Can arise from different sources
  - data from other studies
  - mine versus yours
  - should allow for possible biases
  - should allow for lack of generalizability across studies

The other part: the likelihood

- another assumption!
- both Bayes and Frequentists requires a model
- THIS IS NOT NECESSARILY AN OBVIOUS CHOICE
- Can try more than one model
- (and can try more than one prior)
- **Generally: results are more sensitive to likelihood choice than prior choice**

“Objective Bayes”

- Use vague priors
- To Greenland, probably even more terrible
- make the presumption of complete ignorance very obvious!

- *“They make these unrealistic ‘non-informative’ priors explicit, and so produce posterior intervals that represent the inference of no one (except perhaps someone who understood nothing of the subject under study or even the meaning of the variable names).”*

Frequentist vs Bayesian Divergence

- “Frequentist methods pretend that the models are laws of chance in the real world”
- “The subjective Bayesian interpretation is much less ambitious (and less confident)…insofar as it treats the models and the analysis results as systems of personal judgements, possibly poor ones, rather than as some sort of objective reality”

Frequentist ‘fantasy’

- 95% CIs are presented as if they account for random error” (e.g., measurement error)
- No mention of random error is generally mentioned
- Frequentist approaches do not account for biases
- The approach treats observational data as if it arose from a randomized experiment or random sample
- “Letting the data speak for themselves” assumes lack of biases, which is often false assumption

Some "EASY" Bayesian approaches: Information Weighting

- Recall the example
- Observed data:
  - $\log(\psi) = \log(3.51)$
  - $var(\log(\psi)) = 0.57$
- Prior assumption:
  - $\log(\psi) = 0$
  - $var(\log(\psi)) = 0.5$

Information Weighting

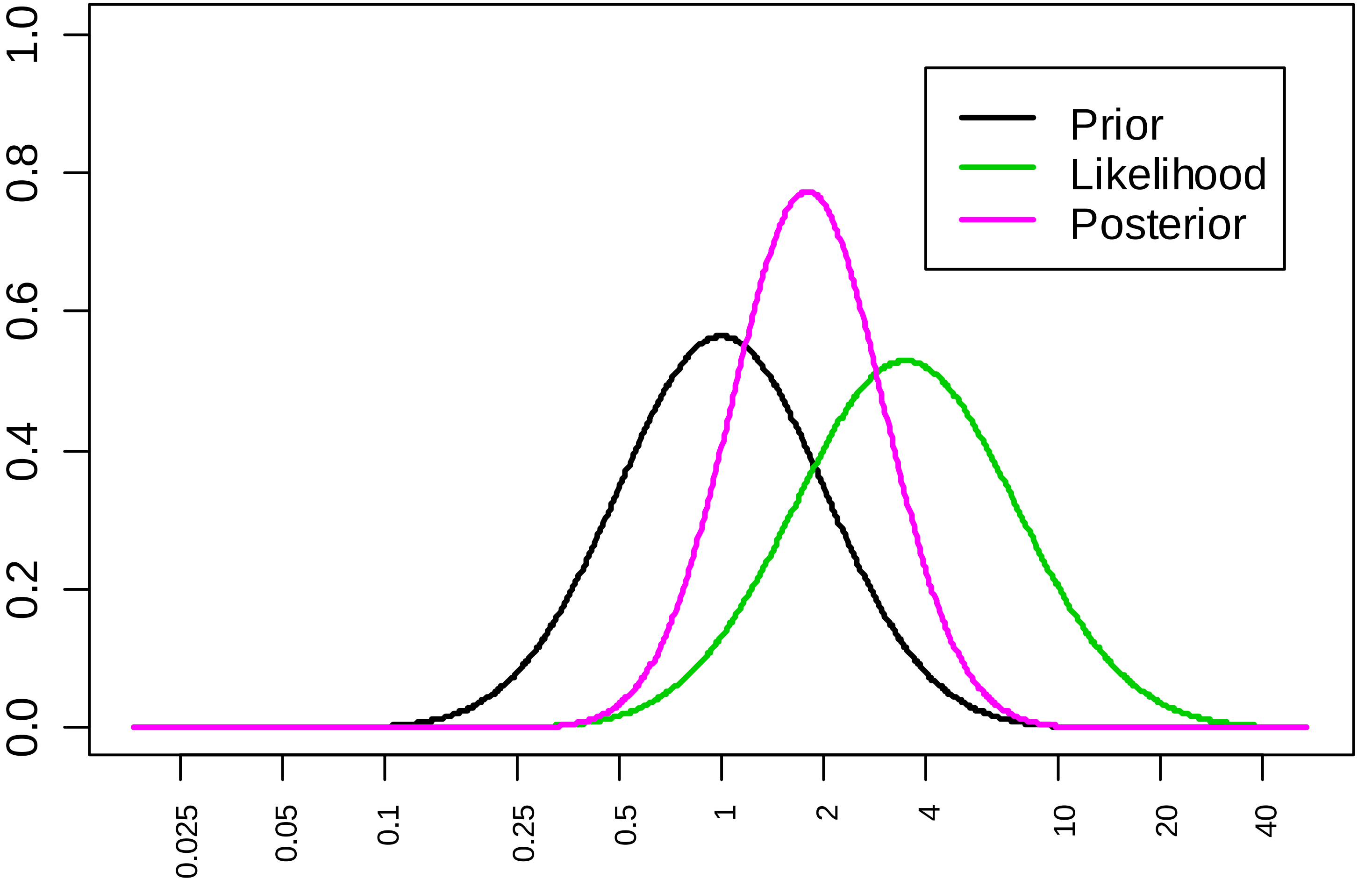
- Generate a weighted average of the means, weighting by the variances

$$\text{posterior}\log(\psi) = \frac{\frac{1}{0.5}(0) + \frac{1}{0.57}(\log(3.51))}{\frac{1}{0.5} + \frac{1}{0.57}}$$

$$= 0.587$$

- Note: exp(0.587) = 1.80

Remember this? Same result!



How can we interpret this?

- Both the DATA and the PRIOR are weak
- Both have similar variance
- But, posterior mean is quite different than 'simple' mean
  - frequentist RR = 3.51
  - Bayesian RR = 1.80 (=exp(0.266))
  - (prior RR = 1)
- Interestingly, average RR in other studies is about 1.70

Seems like a big difference!

- Yes and no
- Depends on the sample size
- Here, the data were quite weak
- N = 234, but very few cases

Extendable?

- So, sort of easy
- But, not practical....when is the last time we didn't adjust for age, race and gender!
- Can still do this with adjustment
- All you need is
  - the adjusted estimates of RR and its variance
  - your prior mean and variance

Another approach: Including "prior data"

- Frequentist approach to Bayes!
- Include "pseudo-data" that is consistent with your prior beliefs
- Can do some simple calculations to figure out how much
- Process:
  - construct data equivalent to the prior
  - add those data to dataset
  - resulting estimate and its variance are actually posterior estimates
- Aside: This is a popular approach in Bayesian phase I trials!

Discussion

- Data alone say nothing
  - need at least a model!
  - both frequentists and Bayesian must choose a model
- If frequentist results arise from 'perfect randomized trial'
  - prior not very necessary
  - vagueness appropriate
- observational studies not perfect!
- Bayesian statistics become worthwhile exactly when frequentist approaches break down