Why Not to Be Afraid of Priors

Paul Bürkner

2018-04-08

What's a prior?

A prior is a probability distribution over *unknown quantities*

A Likelihood is a probability distribution over *known quantities*

The posterior of a model M is the product of likelihood and prior:

$$p(\theta|D, M) \propto p(D|\theta, M)p(\theta|M)$$

Commonly heard reasons to be afraid of priors

"Priors are inherently subjective"

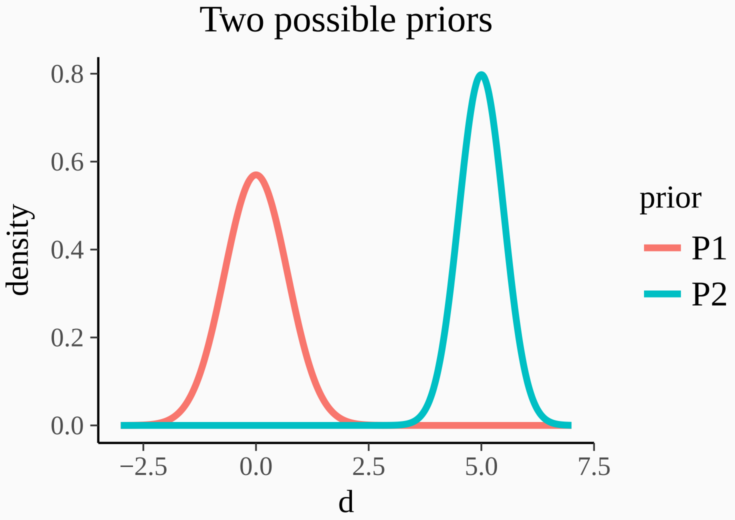
"Priors bias your analysis"

"You should let the data speak for themselves"

"Classical statistics is fine without priors"

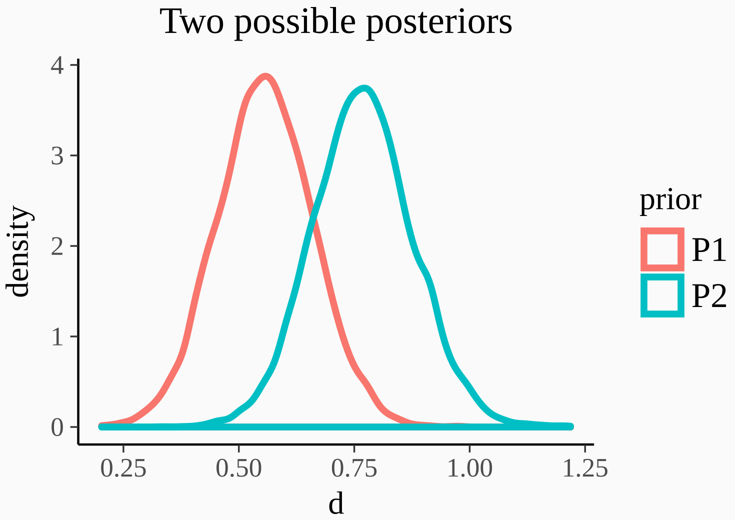
"I have no idea how to set appropriate priors"

Are priors inherently subjective?



Priors influence posteriors

Simulation of 100 observations with a true $d = 0.5$:



Do priors bias your analysis?

Definition:

An estimator $\hat{\theta}$ of the parameter θ is called *unbiased* if $\mathsf{E}(\hat{\theta}) = \theta$.

Corrollary:

If an estimator $\hat{\theta}$ is unbiased for a given prior, it will be biased for most other priors.

Corrollary:

If an estimator $\hat{\theta}$ is unbiased at a given scale, it will be biased for most non-linear transformations $f(\hat{\theta})$:

$$\mathsf{E}(f(\hat{\theta})) \neq f(\theta)$$

Unbiasedness is fragile and often only holds in the limit

Priors are unavoidable

Model 1:

```
y ~ normal(d, 1)
d ~ normal(0, 0.7)
```

Model 2:

```
y ~ normal(d, 1)
d ~ normal(5, 0.5)
```

Model 3:

```
y ~ normal(d, 1)
```

Is the third model free of priors?

Priors appear in unexpected places (1)

Basic multilevel model:

```
y ~ normal(b0[j] + b1 * x, sigma)
b0[j] ~ normal(b0, sd0)
```

Is `b0[j] ~ normal(b0, sd0)` a likelihood, prior, or something else?

Priors appear in unexpected places (2)

Model 1:

```
y ~ normal(b0 + b1 * x, sigma)
(b0, b1) ~ normal(0, 1)
sigma ~ exponential(1)
```

Model 2:

```
ymes ~ normal(y, sdy)
y ~ normal(b0 + b1 * x, sigma)
(b0, b1) ~ normal(0, 1)
sigma ~ exponential(1)
```

Is `y ~ normal(b0 + b1 * x, sigma)` a likelihood or prior?

How do I set appropriate priors?

How do I set appropriate priors?

Possible goals of priors:

- Incorporate explicit prior knowledge
- Make your estimates less excited about extremes
- Identify your model
- Allow hypothesis testing via Bayes factors

Your priors have to depend on your goals.

Setting reasonable priors is hard in any case!

Incorporate explicit prior knowledge

Suppose we want to estimate the speed of light c. A previous experiment has found c to be roughly $299,792,457 m/s$ with an uncertainty (standard deviation) of $2 m/s$. We can update this prior knowledge by means of our own data y:

```
y ~ normal(c, sigma)
c ~ normal(299792457, 2)
```

Make your estimates less exited about extremes

Basic multilevel model:

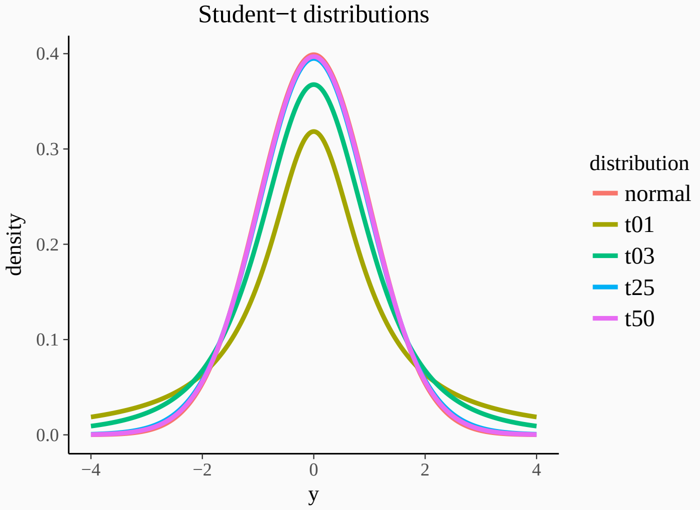
```
y ~ normal(b0[j] + b1 * x, sigma)
b0[j] ~ normal(b0, sd0)
sd0 ~ <some prior facilitating shrinkage>
```

Setting hierachical priors on parameters provides *shrinkage*

Shrinkage can be increased by setting appropriate priors on hyperparameters

This is one of the main purposes of multilevel models!

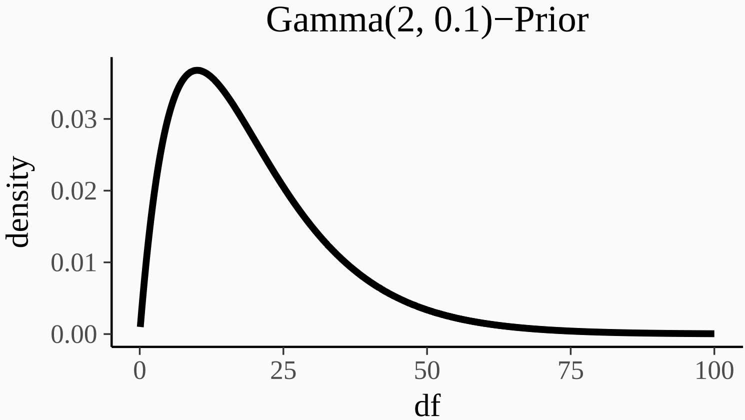
Identify your model (1)



Identify your model (2)

Robust linear regression:

```
y ~ student_t(df, b0 + b1 * x, sigma)
df ~ gamma(2, 0.1)
```



Allow hypothesis testing via Bayes factors

Bayes factor of model M_1 against model M_2:

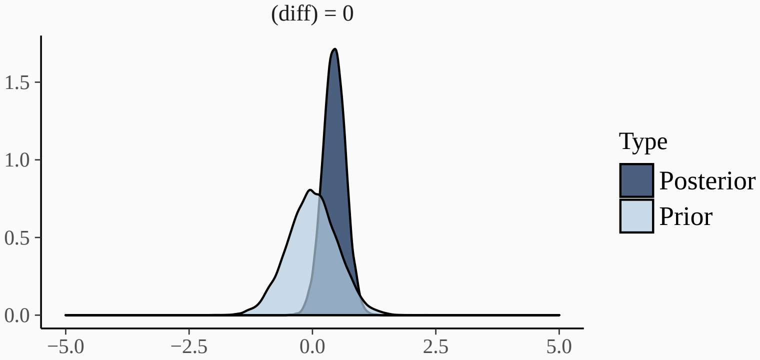
$$BF_{12} = \frac{\int p(D|\theta, M_1)p(\theta|M_1)d\theta}{\int p(D|\theta, M_2)p(\theta|M_2)d\theta}$$

When testing $M_1 : \theta = \theta_0$ against $M_2 : \theta \neq \theta_0$ we can simplify:

$$BF_{12} = \frac{p(\theta_0|D, M_2)}{p(\theta_0|M_2)}$$

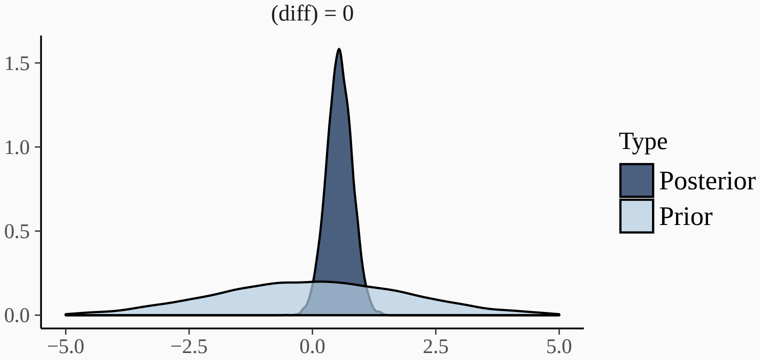
Bayes factors without reasonable priors are meaningless!

Bayes Factors: Influence of Priors (1)



$BF_{12} = 3.02$ for `diff ~ normal(0, 0.5)`.

Bayes factors: Influence of Priors (2)



$BF_{12} = 1.09$ for `diff ~ normal(0, 2)`.

Take home messages

- Priors are part of every model
- Priors can help to improve inference considerably
- Priors should always be made explicit
- Setting reasonable priors is hard...
- ... but there is no reason to be afraid of priors