A few Bayesian lectures for the Uninitiated

—

Part 4: priors

Carsten F. Dormann

Biometry & Environmental System Analysis, University of Freiburg

7. März 2022

Bayes?!

Remember Bayes' Theorem in terms of fitting a model:

$$\boxed{P(\theta|\text{data}) = \frac{P(\text{data}|\theta)P(\theta)}{P(\text{data})}} \tag{1}$$

Bayes?!

Remember Bayes' Theorem in terms of fitting a model:

$$\boxed{P(\theta|\text{data}) = \frac{P(\text{data}|\theta)P(\theta)}{P(\text{data})}} \tag{1}$$

Now: how to choose the *prior*, $P(\theta)$?

Priors galore!

Name
uninformative/vague/diffuse
improper
conjugate
Jeffreys (scale invariant)
shrinkage
savvy (AIC-like)
hyperprior

See here for more:

https://jrnold.github.io/bayesian_notes/priors.html

Priors galore!

Name	Comment
uninformative/vague/diffuse	avoids imposing information

Priors galore!

Name	Comment
uninformative/vague/diffuse	avoids imposing information
Jeffreys	uninformative prior, scale invariant
	(and cumbersome: $p(\vec{\theta}) \sim \sqrt{\det \mathcal{I}(\vec{\theta})}$)

Priors galore!

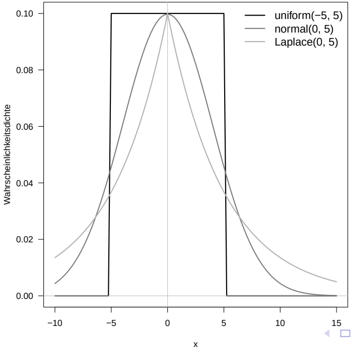
Name	Comment
uninformative/vague/diffuse	avoids imposing information
Jeffreys	uninformative prior, scale invariant (and cumbersome: $p(\vec{\theta}) \sim \sqrt{\det \mathcal{I}(\vec{\theta})}$)
improper	prior distribution not integrating to 1 (e.g. infinitive uniform)

Priors galore!

Name	Comment
uninformative/vague/diffuse	avoids imposing information
Jeffreys	uninformative prior, scale invariant (and cumbersome: $p(\vec{\theta}) \sim \sqrt{\det \mathcal{I}(\vec{\theta})}$)
improper	prior distribution not integrating to 1 (e.g. infinitive uniform)
conjugate	such that posterior is same distribution as prior

Priors galore!

Name	Comment
uninformative/vague/diffuse	avoids imposing information
Jeffreys	uninformative prior, scale invariant
	(and cumbersome: $p(\vec{\theta}) \sim \sqrt{\det \mathcal{I}(\vec{\theta})}$)
improper	prior distribution not integrating to 1
	(e.g. infinitive uniform)
conjugate	such that posterior is same
	distribution as prior
shrinkage	to implement model selection



Priors galore!

Name	Comment
uninformative/vague/diffuse	avoids imposing information
Jeffreys	uninformative prior, scale invariant
	(and cumbersome: $p(\vec{\theta}) \sim \sqrt{\det \mathcal{I}(\vec{\theta})}$)
improper	prior distribution not integrating to 1
	(e.g. infinitive uniform)
conjugate	such that posterior is same
	distribution as prior
shrinkage	to implement model selection
savvy (K-L-model)	to yield AIC-weight for model i:
	$p_i = c \cdot e^{\frac{1}{2}k_i \ln(n) - k_i}$

Priors galore!

Name	Comment
uninformative/vague/diffuse	avoids imposing information
Jeffreys	uninformative prior, scale invariant (and cumbersome: $p(\vec{\theta}) \sim \sqrt{\det \mathcal{I}(\vec{\theta})}$)
improper	prior distribution not integrating to 1 (e.g. infinitive uniform)
conjugate	such that posterior is same distribution as prior
shrinkage	to implement model selection
savvy (K-L-model)	to yield AIC-weight for model i: $p_i = c \cdot e^{\frac{1}{2}k_i \ln(n) - k_i}$
hyperprior	prior for a hyper-parameter

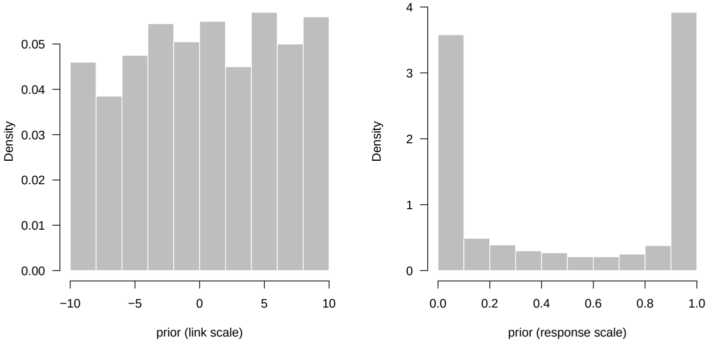
Which prior distribution to choose?

1. Informative or not? $\longrightarrow$ informative if possible![1]

[1] Lemoine, N. (2019) Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos* 128, 912.

Which prior distribution to choose?

1. Informative or not? $\longrightarrow$ informative if possible![1]



Uniform$(-10, 10)$ as prior for binomial model's intercept: it is no longer uninformative at the response scale!
(Jeffreys priors *remain* uninformative!)

[1] Lemoine, N. (2019) Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos* 128, 912.

Which prior distribution to choose?

1. Informative or not? $\longrightarrow$ informative if possible![2]
2. Is parameter a "location" (i.e. mean or slope) or a "scale" (i.e. variance-like)? $\longrightarrow$ normal/t vs. γ/half-Cauchy[1]

[2] Lemoine, N. (2019) Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos* 128, 912.

Which prior distribution to choose?

1. Informative or not? $\longrightarrow$ informative if possible![2]
2. Is parameter a "location" (i.e. mean or slope) or a "scale" (i.e. variance-like)? $\longrightarrow$ normal/t vs. γ/half-Cauchy[1]
3. Does parameter have upper/lower limits? $\longrightarrow$ think of β, γ and truncated distributions (e.g. half-normal)

[2] Lemoine, N. (2019) Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos* 128, 912.

Which prior distribution to choose?

1. Informative or not? $\longrightarrow$ informative if possible![2]
2. Is parameter a “location” (i.e. mean or slope) or a “scale” (i.e. variance-like)? $\longrightarrow$ normal/t vs. γ/half-Cauchy[1]
3. Does parameter have upper/lower limits? $\longrightarrow$ think of β, γ and truncated distributions (e.g. half-normal)
4. Are parameters correlated? $\longrightarrow$ multivariate priors (e.g. inverse (!) Wishart for MVN)

[2]Lemoine, N. (2019) Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos* 128, 912.

Prior: take-home-message

- The prior quantifies your expectation *before* seeing the data.

Prior: take-home-message

- The prior quantifies your expectation *before* seeing the data.
- It forces us to think about such expectation, which is a good thing.

Prior: take-home-message

- The prior quantifies your expectation *before* seeing the data.
- It forces us to think about such expectation, which is a good thing.
- When pretending to knowing nothing, *uninformative priors* are used.

Prior: take-home-message

- The prior quantifies your expectation *before* seeing the data.
- It forces us to think about such expectation, which is a good thing.
- When pretending to knowing nothing, *uninformative priors* are used.
- Priors are our friends: we can use them to our advantage!

Prior: take-home-message

- The prior quantifies your expectation *before* seeing the data.
- It forces us to think about such expectation, which is a good thing.
- When pretending to knowing nothing, *uninformative priors* are used.
- Priors are our friends: we can use them to our advantage!
- We may not like priors, but they are inevitable in Bayesian analysis.

Next time:
Mixed effect models from a Bayesian perspective (with R)