More About Priors I

Bayesian Models for Ecologists

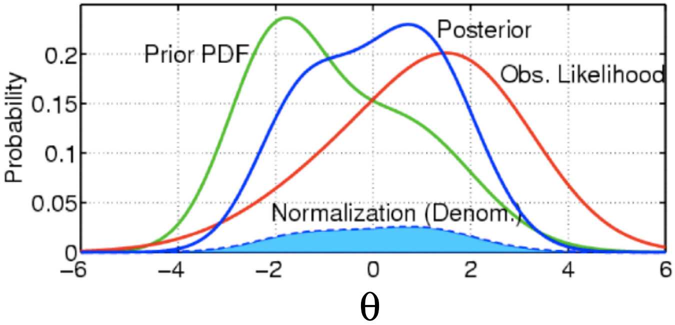
Becky Tang

June 05, 2024

DEB 2042028

Acknowledgment: slides largely adapted from Baily Fosdick.

The posterior distribution uses information from both the likelihood and prior distributions. *An informative prior is more influential than a vague prior*.



Outline

- Informative priors
- Vague/Flat priors
- Conjugate priors

Choosing priors example

- Suppose I'm interested in learning about the proportion of ecologists who are left-handed. Call this θ.
- It's impossible to survey all ecologists, but luckily we can obtain a sample here today to perform inference for θ.
- We need to select a prior for θ. This means specifying a distribution and the values for the parameters of that distribution.
  - Which one(s)? A good place to start is to consider what values θ can take on.

Informative priors

Informative priors are distributions that are not diffuse relative to the posterior. These distributions may be based on

- statistics reported in the literature
- posterior distributions from previous studies
- meta-analyses
- “plausible” assumptions

Why use informative priors?

- They speed up convergence
- They reduce problems with identifiability and allow you to estimate difficult/impossible quantities
  - Can help "pin down" a parameter by restricting the range of values it can take

Why don't we find people using informative priors more often?

- *Cultural*: "All studies stand alone" argument
- Texts often use vague priors (including H&H)
- Hard work!
- Concerns about "excessive subjectivity"

If you wanted to use an informative prior, how would you do it?

- Strong scholarship is the basis for strong priors
- Moment match, converting means and standard deviations (or other moments) to usable parameters
- Pilot studies
- Allometric relationships
- Deterministic models with parameters that have specific meaning

Example where parameter has specific meaning

Consider the predator-prey Lotka-Volterra model:

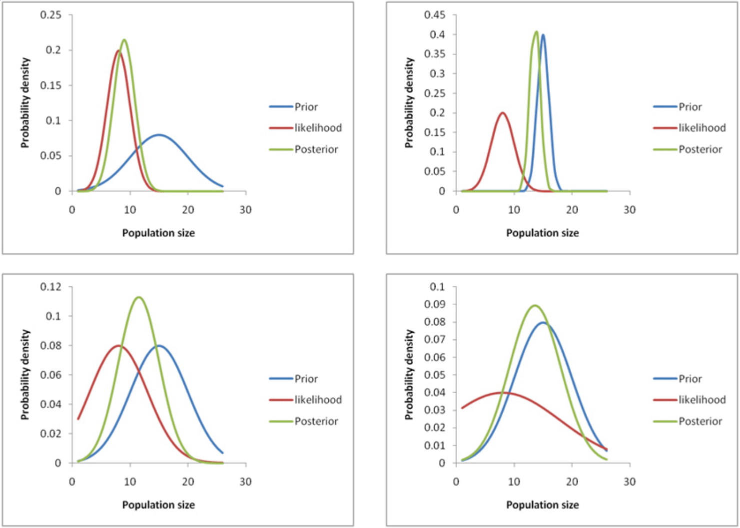
$$\frac{dx}{dt} = \alpha x + \beta xy \qquad \frac{dy}{dt} = \delta xy + \gamma y$$

- x and y are population densities of of prey and predator, respectively, and equations are the the instantaneous growth rates
- α represents maximum per-capita growth rate for prey and β the effect of predator on prey's death rate
- δ represents effect of the presence of prey on the predator's growth rate and γ the predator's per-capita death rate
- In order to be *ecologically* sensible, $\alpha, \beta, \delta, \gamma$ cannot take on any arbitrary value
  - e.g. β and γ need to restricted to be negative, and α and δ positive. Magnitudes must also be appropriate for populations (otherwise populations would explode or die out immediately). Choose priors that encode these constraints, e.g. $\gamma \sim \text{Uniform}(-0.20, 0)$

Choosing priors example cont.

- Let's use moment matching to formulate some informative priors for θ.
- Visualize

How much does an informative prior influence the posterior?



Gündoğdu, Sedat and Akar, Mustafa. "Bayesian Update For Descriptive Statistics In Fisheries Science" Transylvanian Review

Communicating your use of informative priors

Ecological Monographs, 85(4), 2015, pp. 525–556
© 2015 by the Ecological Society of America

State-space modeling to support management of brucellosis in the Yellowstone bison population

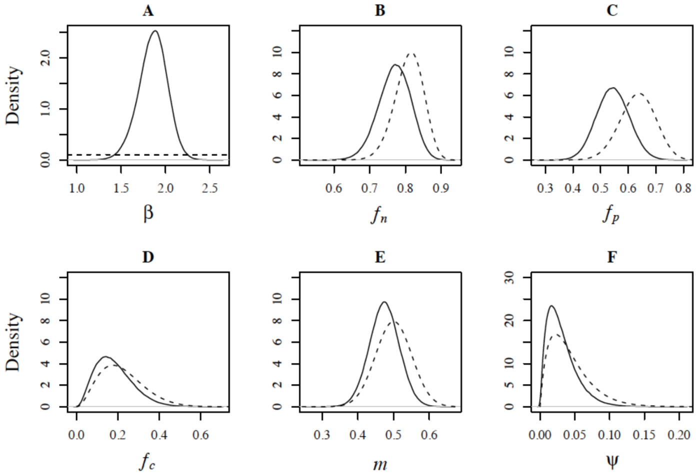
N. Thompson Hobbs,[1,5] Chris Geremia,[2] John Treanor,[2] Rick Wallen,[2] P. J. White,[2] Mevin B. Hooten,[3] and Jack C. Rhyan[4]

Communicating your use of informative priors

Table 3: Prior distributions for parameters in model of brucellosis in the Yellowstone bison population. Sources are given for informative priors.

Parameter	Definition	Distribution	Mean	SD	Source
β	Rate of transmission (yr^{-1})	uniform(0,50)	25	14.3	vague
f_n	Number of offspring recruited per seronegative (susceptible) female	beta(77,18)	.81	.04	Fuller et al., 2007
f_p	Number of offspring recruited per seropositive (recovered) female	beta(37,20)	.64	.06	Fuller et al., 2007
f_c	Number of offspring recruited per seroconverting	beta(3.2,11)	.22	.10	Fuller et al., 2007

Communicating your use of informative priors



Vague/Flat Priors

A vague prior is a distribution with a range of uncertainty that is clearly wider than the range of reasonable values for the parameter (Gelman and Hill 2007:347).

Vague/Flat Priors

- Avoid calling a prior "uninformative" or "non-informative" rather:
  - diffuse
  - flat
  - automatic
  - nonsubjective
  - locally uniform
  - objective

Choosing priors example cont.

- Let's use moment matching to formulate some weakly informative priors for θ.
- Then we'll collect data and obtain the posteriors for θ based on our choices of priors!
  - We could then use the posterior distribution $[\theta|\mathbf{y}]$ to answer questions about the proportion of ecologists who are left-handed, in light of our observed data

Commonly Used Vague/flat Priors

- For strictly non-negative quantities: Gamma(.001, .001)
- For variances: Inverse Gamma(.001, .001) or Uniform(0, some large number)
- For regression coefficients: Normal(0, a variance much greater than the mean)

Important Note: The Uniform and Normal must be scaled properly! For example $\beta_0 \sim \text{normal}(0, 1000)$ is extremely informative if $\beta_0 = 10000$.

Issues With Vague/Flat Priors

- Computational: failure to converge, slicer errors, failure to calculate log density, etc.
- Sensitivity: changes in parameters of "vague" priors can lead to meaningful changes in the posterior when data sets are small or when they have high variance
  - E.g. $\sigma^2 \sim \text{InverseGamma}(\varepsilon, \varepsilon)$ where ε is small can be problematic if σ^2 is estimated to be near zero; this may come up in the multilevel modeling lab

Conjugacy

- In special cases the posterior, $[\theta|y]$, has the same distributional form (i.e. family) as the prior, $[\theta]$.
    - For example: if you have $\theta \sim \text{Gamma}(\alpha, \beta)$ and $\theta|y \sim \text{Gamma}(\alpha_{new}, \beta_{new})$
- In these cases, the prior and the posterior are said to be **conjugate** for the *particular kind of data/sampling model*

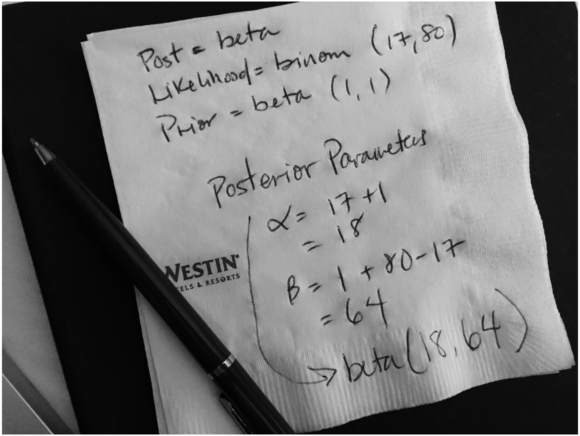
Conjugacy is important for two reasons:

1. Conjugacy minimizes computational work and, in more complicated cases, allows us to break down calculations into manageable chunks.
    1. There is often a nice interpretation of how the prior and data come together to influence shape of posterior
2. Conjugacy plays an important role in Markov chain Monte Carlo (more on this later).

Conjugate priors

Table A.3: Table of conjugate distributions

Likelihood	Prior distribution	Posterior distribution
$y_i \sim \text{binomial}\,(n, \phi)$	$\phi \sim \text{beta}\,(\alpha, \beta)$	$\phi \sim \text{beta}\,(\sum y_i + \alpha, n - \sum y_i + \beta)$
$y_i \sim \text{Bernoulli}(\phi)$	$\phi \sim \text{beta}\,(\alpha, \beta)$	$\phi \sim \text{beta}\,(\sum_{i=1}^n y_i + \alpha, \sum_{i=1}^n (1 - y_i) + \beta)$
$y_i \sim \text{Poisson}\,(\lambda)$	$\lambda \sim \text{gamma}\,(\alpha, \beta)$	$\lambda \sim \text{gamma}\,(\alpha + \sum_{i=1}^n y_i,\ \beta + n)$
$y_i \sim \text{normal}(\mu, \sigma^2)$ σ^2 is known.	$\mu \sim \text{normal}\,(\mu_0, \sigma_0^2)$	$\mu \sim \text{normal}\left(\frac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n y_i}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)}, \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}\right)$
$y_i \sim \text{normal}(\mu, \sigma^2)$ μ is known.	$\sigma^2 \sim$ inverse gamma(α, β)	$\sigma^2 \sim$ inverse gamma $\left(\alpha + \frac{n}{2}, \beta + \frac{\sum_{i=1}^n (y_i - \mu)^2}{2}\right)$
$y_i \sim \text{lognormal}(\mu, \sigma^2)$, μ is known	$\sigma^2 \sim$ inverse gamma(α, β) ,	$\sigma^2 \sim$ inverse gamma $\left(n/2 + \alpha, \frac{(\log(y_i) - \mu)^2}{2} + \beta\right)$
$y_i \sim \text{lognormal}\,(\mu, \sigma^2)$ σ^2 is known	$\mu \sim \text{normal}\,(\mu_0, \sigma_0^2)$	$\mu \sim \text{normal}\left(\frac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n \log y_i}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)}, \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}\right)$


Post = beta
Likelihood = binom (17, 80)
Prior = beta (1, 1)
Posterior Parameters
α = 17 + 1
= 18
β = 1 + 80 - 17
= 64
WESTIN
ELS & RESORTS
→ beta(18, 64)

Why Use Conjugacy

- It is not necessary, conjugate priors will accelerate MCMC.
- For simple models, you can use conjugate priors to obtain the posterior distribution in closed form, without any simulation.

Things to remember

- There is no such thing as an uninformative prior, but certain priors influence the posterior distribution more than others.
- Informative priors, when properly justified, can be useful.
- Strong data overwhelm a prior.
- Priors represent current knowledge (or lack of), which is updated with data.
- We encourage you to think of vague priors as a provisional starting point.
- You don't have to use conjugate priors, especially if they do not actually represent your prior beliefs!
    - Related: should I place a prior on the variance or the standard deviation? Conjugacy might exist for variance parameter, but a prior for the standard deviation might be more interpretable.

Lab exercises.