Introduction to Applied Bayesian Modeling

ICPSR

Day 4

Simple Priors

- Remember Bayes' Law:

$$P(A|B) \propto P(A) \times P(A|B)$$

- Where P(A) is the prior probability of A

Simple prior

- Recall the ‘test for disease’ example where we specified the prior probability of a disease as a point estimate.
- In that example we said the probability of having the disease was 0.01 and then computed the probability of having the disease given a positive test result.

Too simple?

- Considering the prior probability of having a disease (or being pregnant, or winning an election, or winning a war, etc…) to be a point estimate is probably unreasonable.
- Conditional on age (for disease), money raised (for elections) or economic strength (for conflicts), wouldn’t we expect the prior probabilities to vary?

Distributions of priors

- Even without conditioning on covariates, we would likely find different sources of information providing slightly different unconditional prior probabilities.
- That is, our prior beliefs would be derived from different samples and are not known and fixed population quantities.

More realistic priors…

- So…rather than assigning a single value for P(A), it makes sense, from a Bayesian perspective, to assign a probability distribution to P(A) that captures prior uncertainty about it's true value.
- This results in the posterior probability that is also a probability distribution, rather than a point estimate.

How do we do this?

- As we’ve seen from earlier lectures, we must specify the form of the prior probability distribution
  - Must be an ‘appropriate’ distribution
  - That is, the characteristics of the parameter help dictate which distribution(s) is/are acceptable

Appropriate priors

- For example, if we want to specify a prior for a probability (i.e. p in a binomial), we need a distribution that varies between 0 and 1.
- The Beta distribution is perfect for this:

$$p(p|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha) \quad \Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

More priors…

- If we are modeling counts with a Poisson, then we want our prior distribution to only allow positive values (negative counts make no sense).
- The Gamma distribution fits the bill nicely:

$$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}$$

And more priors…

- What about normally distributed data?

- What do think would be a good prior for μ in a normal model?

Hmmm?

- How about Normal?

$$p(\mu) = \frac{1}{\sqrt{2\pi\tau^2}} exp[-\frac{(\mu - M)^2}{2\tau^2}]$$

- How about a prior for the variance?

Prior for variance

- The standard prior for the variance in the normal model is the inverse gamma distribution.

$$p(\sigma) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \sigma^{-(\alpha-1)} e^{-\frac{\beta}{\sigma}}$$

So, what does this all mean

- Well, it means that we can specify uncertainty about our prior beliefs by using appropriate probability distributions.

- It also means that our ‘answers’ are distributions rather than point estimates

Do I have to use these priors?

- NO.

- The examples on the previous slides are all what we call *conjugate* priors.

- Conjugacy has some very nice properties

How nice?

- If a prior is conjugate to a given likelihood function, then we are ***guaranteed*** to have a closed-form solution for the moments of the posterior.
- HUH?
- This means, that we know the distribution of the posterior and can easily solve for means, medians, variance, quantiles etc.

We know the form of the posterior?

- YES.

- If a prior is conjugate to the likelihood, then the posterior has the same distribution as the prior.
  - It just has slightly different parameters
  - Which are weighted averages of the prior and the likelihood.
  - We saw this yesterday.

Conjugacy

Table 1: Some Exponential Family Forms and Conjugate Priors

Form	Conjugate Prior Distribution	Hyperparameters
Bernoulli	Beta	$\alpha > 0,\ \beta > 0$
Binomial	Beta	$\alpha > 0,\ \beta > 0$
Multinomial	Dirichlet	$\theta_j > 0,\ \Sigma\theta_j = \theta_0$
Negative Binomial	Beta	$\alpha > 0,\ \beta > 0$
Poisson	Gamma	$\alpha > 0,\ \beta > 0$
Exponential	Gamma	$\alpha > 0,\ \beta > 0$
Gamma (incl. χ^2)	Gamma	$\alpha > 0,\ \beta > 0$
Normal for μ	Normal	$\mu \in \mathbb{R},\ \sigma^2 > 0$
Normal for σ^2	Inverse Gamma	$\alpha > 0,\ \beta > 0$
Pareto for α	Gamma	$\alpha > 0,\ \beta > 0$
Pareto for β	Pareto	$\alpha > 0,\ \beta > 0$
Uniform	Pareto	$\alpha > 0,\ \beta > 0$

Non-informative Priors

★ A noninformative prior is one in which little new explanatory power about the unknown parameter is provided by intention.

★ Noninformative priors are very useful from the perspective of traditional Bayesianism that sought to mitigate frequentist criticisms of intentional subjectivity.

★ Fisher was characteristically negative on the subject: "…how are we to avoid the staggering falsity of saying that however extensive our knowledge of the values of x may be, yet we know nothing and can know nothing about the values of θ?"

Uniform Priors

★ An obvious choice for the noninformative prior is the uniform distribution:

- Uniform priors are particularly easy to specify in the case of a parameter with bounded support.

- *Proper* uniform priors can be specified for parameters defined over unbounded space if we are willing to impose prior restrictions.

- Thus if it reasonable to restrict the range of values for a variance parameter in a normal model, instead of specifying it over $[0:\infty]$, we restrict it to $[0:\nu]$ and can now articulate it as $p(\sigma) = 1/\nu,\ 0 \leq \theta \leq \nu$.

- *Improper* uniform priors that do not possess bounded integrals and surprisingly, these can result in fully proper posteriors under some circumstances (although this is far from guaranteed).

- Consider the common case of a noninformative uniform prior for the mean of a normal distribution. It would necessarily have uniform mass over the interval: $p(\theta) = c,\ [-\infty \geq \theta \geq \infty]$. Therefore to give *any* nonzero probability to values on this support, $p(\theta) = \epsilon > 0$, would lead to a prior with infinite density: $\int_{-\infty}^{\infty} p(\theta)d\theta = \infty$.

Elicited Priors

- **Clinical Priors:** elicited from substantive experts who are taking part in the research project.

- **Skeptical Priors:** built with the assumption that the hypothesized effect does not actually exist and are typically operationalized through a probability function (PMF or PDF) with mean zero.

- **Enthusiastic Priors:** the opposite of the sceptical prior, built around the positions of partisan experts or advocates and generally assuming the existence of the hypothesized effect.

- **Reference Priors:** produced from expert opinion as a way to express informational uncertainty, but they are somewhat misguided in this context since the purpose of elicitation is to glean information that can be described formally.

So, can I use any prior

- Well, pretty much, but there are some things to consider…
  - Some priors could lead to nonsensical results
  - Some priors make life easier
  - Some priors make publication easier—unfortunately.
  - Some priors lead to intractable results…

what does that mean?

- Imagine the following prior:

$$p(\theta) = \arctan \theta^{\sin \pi^2 + 2 \prod_{i=1}^{N} (x_i)^{\cos \sqrt{-e^{\pi}}}}$$

- You could do this is you REALLY wanted to, but this leads to the following result

Weird prior result

- 1
- + Math
- = HARD

So, then what

- If we have a posterior for which there is no analytical solution—or, if the analytical solution is REALLY hard,
- Don't worry—there is an answer. It's called **MCMC**.

Next week…

- Summarizing posterior distributions

- Introducing covariates

- MCMC/Gibbs Sampling

- See you at the picnic!!