notes from the week of Feb. 18, 2019

Contents

We worked through Bayes' rule and then talked about the difference between confidence intervals and Bayesian posterior probability statements (see `http://phylo.bio.ku.edu/biostats/bad_bean_counter.pdf`)

1 Posterior distribution for ν

. We'll use $\nu = ut$, as the branch length in terms of the expected number of changes over a branch in a genealogy.

Back to our original example $X = 5$ and the branch length from JKK to the MRCA and back to you is 2ν.

We assume that the Poisson describes the probability of any number of mutations given the expected number In an ML context, you'd get $\hat{\nu} = 2.5$ and you could define a 95% confidence interval based on what values of ν give you log-likelihoods that are 1.92 lower than the log-likelihood at $\nu = 2.5$.

For a Bayesian inference of ν, we start with Bayes' rule applied to this problem:

$$f(\nu \mid X = 5) = \frac{g(\nu)\mathbb{P}(X = 5 \mid \nu)}{\mathbb{P}(X = 5)} \tag{1}$$

where $f(\nu \mid X = 5)$ is the "posterior probability density of ν" and $g(\nu)$ is the "prior probability density of ν." Both are probability distributions, meaning that the integral of the density over the feasible range for ν (which is $0 \leq \nu < \infty$) will be 1.

1.1 the prior

In a purely subjective Bayesian perspective we'd have to figure out how to describe our prior beliefs. In practice, we often pick an analytically tractable distribution that is close to what we want to express as our prior beliefs, while still being easy to work with mathematically.

Last week we saw that population genetics tells us the $t \sim$ Exponential$(\lambda = 1/N)$ where N is the effective population size. If we pretended that we knew the mutation rate for our locus (which is not terribly implausible as there is a lot of data about the rate of molecular evolution in mammalian mitochondrial genes), then perhaps our uncertainty about ν can be thought of as being basically some exponential distribution.

An exponentail distribution is just a special case of a Gamma distribution, where $\alpha = 1$ and $\beta = \lambda$ when we are using the "shape+rate" parameterization of the the Gamma distribution. Wikipedia tells us that if $x \sim \text{Gamma}(\alpha, \beta)$ then:

$$\begin{aligned}
g(x \mid \alpha, \beta) &= \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} && (2)\\
\mathbb{E}[x] &= \frac{\alpha}{\beta} && (3)\\
\text{Var}(x) &= \frac{\alpha}{\beta^2} && (4)
\end{aligned}$$

So if we wanted to say our prior on ν is a fairly broad exponentail distribution with a mean of 5 ($= \frac{1}{5} = 0.2$ in the typical Exponential notation or $\alpha = 1, \beta = .2$ in our preferred notation for the Gamma). So, to be concrete:

$$\begin{aligned}
g(\nu \mid \alpha, \beta) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \nu^{\alpha-1} e^{-\beta\nu} && (5)\\
g(\nu \mid \alpha = 1, \beta = 0.2) &= \frac{0.2}{\Gamma(1)} \nu^0 e^{-0.2\nu} && (6)
\end{aligned}$$

1.2 the likelihood

Recall that our expectated number of mutations in ν and our likelihood is Poisson:

$$\mathbb{P}(X = 5 \mid \nu) = \frac{(2\nu)^5 e^{-2\nu}}{5!} \qquad (7)$$

1.3 the marginal likelihood

The denominator of Bayes' rule is tough it is:

$$\mathbb{P}(X = 5) = \int_0^\infty \mathbb{P}(X = 5 \mid \nu) g(\nu) d\nu \qquad (8)$$

Crucially it is is a constant – it is the same for every value of ν. So we'll just call it K

1.4 the posterior

So what is our posterior density?

$$\begin{aligned}
f(\nu \mid X = 5) &= \frac{g(\nu)\mathbb{P}(X = 5 \mid \nu)}{\mathbb{P}(X = 5)} && (9)\\
&= \frac{\frac{0.2}{\Gamma(1)} \nu^0 e^{-0.2\nu} \frac{(2\nu)^5 e^{-2\nu}}{5!}}{K} && (10)\\
&= K_2 \nu^0 e^{-0.2\nu} \nu^5 e^{-2\nu} && (11)\\
&= K_2 \nu^5 e^{-2.2\nu} && (12)
\end{aligned}$$

where K_2 is a constant that is just a product of all of the parts of the equation that don't depend on ν.

The key thing to recognize is that the shape of the density is just the relative height as ν changes. That shape is determined by $\nu^5 e^{-2.2\nu}$ K_2 is the "normalization constant" that we'd have to multiple to every density so that the whole thing integrates to 1. Conveniently, the shape of the posterior that we derived is exactly the shape of a Gamma($\alpha = 6, \beta = 2.2$) distribution. That means that $K_2 = 2.2^6/\Gamma(6)$ (because that is the coefficient of the Gamma density, and the Gamma integrates to 1, too). Figure 1 shows the prior, likelihood, and posterior for ν

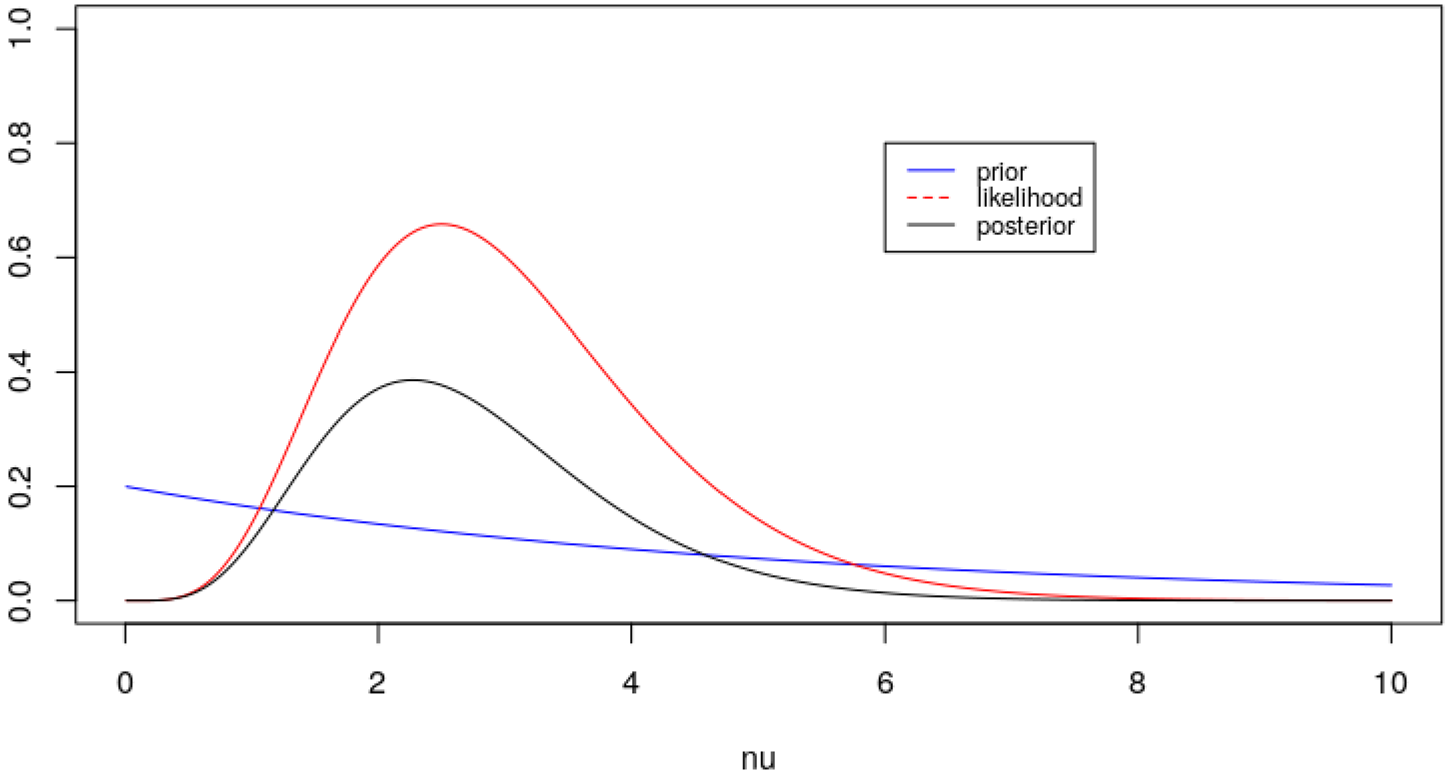

Figure 1: Poisson likelihood (red - not a probability distribution over ν), the Gamma($\alpha = 1, \beta = 0.2$) prior and Gamma($\alpha = 6, \beta = 2.2$)

1.5 a 95% credible interval

Because the posterior is a nice analytically well-characterized distribution, we can use `qgamma(0.025, shape=6, rate=2.2)` and `qgamma(0.975, shape=6, rate=2.2)` in R to find the values of ν that chop off the lower and upper 2.5% tails to get a credible interval. Most Bayesians use a slightly different Highest Posterior Density (HPD) interval, which is the smallest interval (on the ν axis) that integrates to 0.95. But these intervals would be quite similar for this problem.

Figure 2 shows the likelihood with the 1.92 log drop (the maximized likelihood divided by $e^{1.92}$). The ML-based 95% confidence interval is $0.87 < \nu < 5.37$, while the Bayesian 95% credible interval produces by cropping off 2.5% from each tail is the statement that $\mathbb{P}(1.00 < \nu < 5.30 \mid X = 5) = 0.95$ using our prior.

It is common for the 95% credible interval to be a tiny bit smaller (more precise) than the 95% confidence interval. Recall that the confidence interval is based on a recipe guaranteeing 95% coverage that was designed before we look at the data. By fully exploiting prior information and conditioning our inference fully on the data, the Bayesian approach has a bit more power (at the cost of having more inputs that it could be sensitive to).

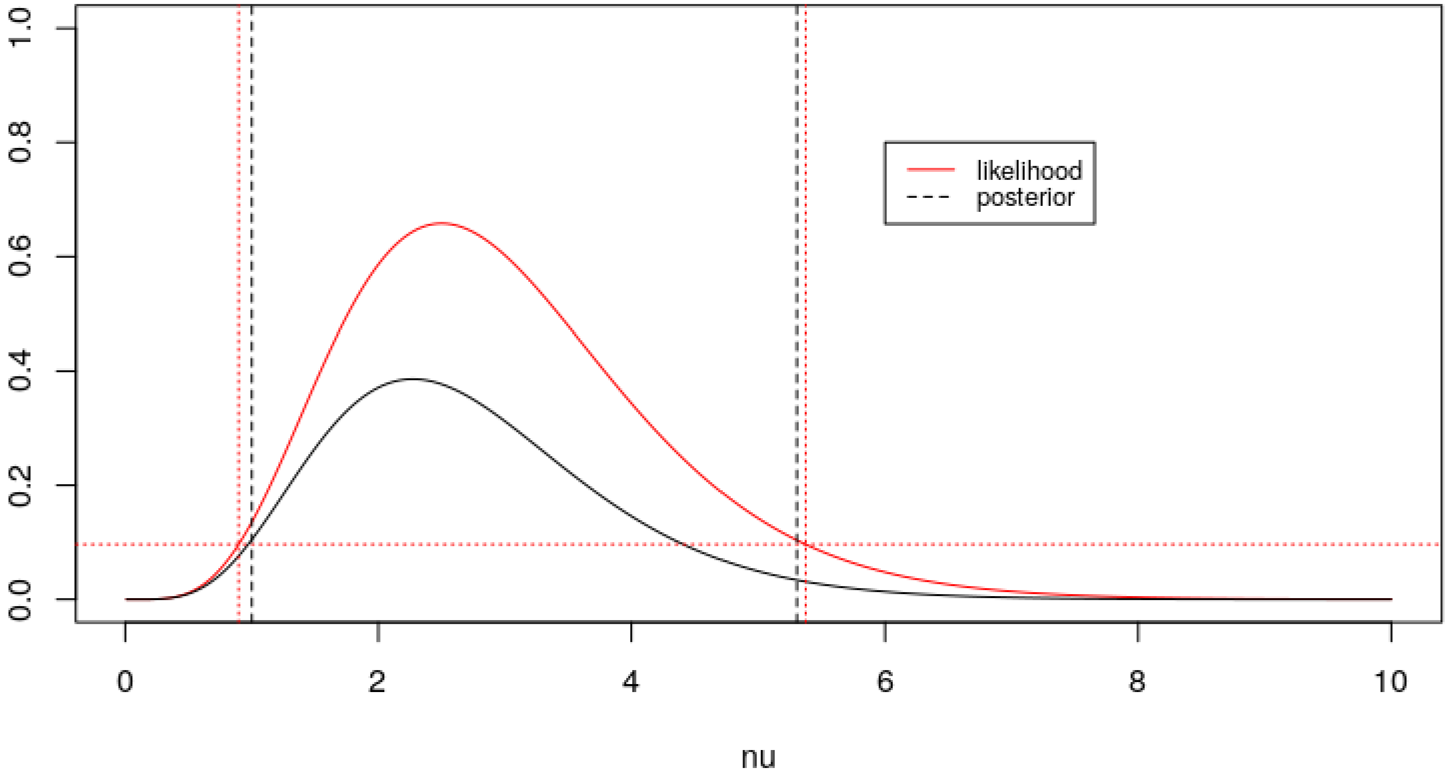

Figure 2: Likelihood and posterior. Confidence interval boundaries computed by the 1.92 ln-likelihood drop method. Credible interval boundaries computed by the `qgamma` method in R.

https://en.wikipedia.org/wiki/Conjugate_prior#Table_of_conjugate_distributions has a nice table of conjugate priors that can be used when your likelihood has a nice form to guarantee that the posterior will have an analytically tractable form.

Later in the course we'll use Markov chain Monte Carlo methods to do Bayesian estimation even when we don't have a likelihood that has a conjugate prior.