Bayesian Inference

November 9, 2020

Bayesian approaches

- Typically contrasted with **frequentist** approaches
- Treat parameters as uncertain, data as fixed

Bayes' Rule

Bayes' Rule

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta)p(\theta)}$$

Notes:

$p(\theta|x)$ - posterior
$p(\theta)$ - prior
$p(x|\theta)$ - likelihood

The posterior distribution is proportional to the prior times the likelihood: $p(\theta|x) \propto p(x|\theta)p(\theta)$

The posterior distribution *is a distribution* over θ.

Posterior predictive distribution

Given

$p(\theta|x)$ - posterior
$p(\theta)$ - prior
$p(x|\theta)$ - likelihood

Posterior predictive distribution

Consider the probability of new data x'. Posterior predictive distribution is:

$$p(x'|x) = \int p(x', \theta|x)\, d\theta = \int p(x'|\theta, x) p(\theta|x)\, d\theta = \int p(x'|\theta) p(\theta|x)\, d\theta$$

Incorporates the knowledge and uncertainty about θ that we already had after seeing data x.

Bayesian inference: conjugate example

Sometimes, we can compute the posterior distribution by hand, given prior and likelihood.

Setup: flipping a coin

Probability that it lands heads is (unknown) θ.
Prior probability over θ assumed to follow a $Beta(3, 3)$ distribution:

$$p(\theta) = \frac{\theta^{3-1}(1-\theta)^{3-1}}{B(3, 3)}$$

Note: $\theta \sim Beta(a, b)$ means $p(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}$
Will collect data by flipping coin once. Likelihood of observing heads ($x = 1$) or tails ($x = 0$) is given by a Bernoulli distribution:

$$p(x|\theta) = \theta^{x}(1-\theta)^{1-x}$$

.

Bayesian inference: conjugate example

Setup: flipping a coin

Probability that it lands heads is (unknown) θ.

Prior probability over θ assumed to follow a $Beta(3,3)$ distribution:

$$p(\theta) = \frac{\theta^{3-1}(1-\theta)^{3-1}}{B(3,3)}$$

Note: $\theta \sim Beta(a,b)$ means $p(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}$

Will collect data by flipping coin once. Likelihood of observing heads ($x=1$) or tails ($x=0$) is given by a Bernoulli distribution:

$$p(x|\theta) = \theta^x(1-\theta)^{1-x}$$

.

Computing the posterior after observing x=1

$$p(\theta|x) \propto p(x|t)p(\theta) = \theta^1(1-\theta)^0\theta^2(1-\theta)^2 = \theta^3(1-\theta)^2 \implies \theta|x \sim Beta(4,3)$$

Bayesian inference: tractability notes

Conjugacy

We have conjugacy when the prior and the posterior distributions are in the same family (e.g. in the previous example, the prior and posterior are beta distributions).

Generally

Generally, computing the posterior distribution is much harder than in this example!

Consider the denominator in $p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta)p(\theta)}$ - integrals are hard

In nonconjugate examples, we need approaches to work with the posterior distribution when we cannot calculate it directly. Stay tuned!