# Bayesian Inference: Intro via Conjugacy

June 8, 2021

## Bayes' Rule

$$\underbrace{p(\theta|x)}_{posterior} = \frac{\underbrace{p(x|\theta)}_{likelihood}\ \underbrace{p(\theta)}_{prior}}{\underbrace{p(x)}_{evidence}} = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta)p(\theta)\,d\theta}$$

## Posterior

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 $\theta$.

## Evidence

The evidence, or *marginal likelihood*, can be used for model comparison.

# 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) $\theta$.
Prior probability over $\theta$ 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) $\theta$.

Prior probability over $\theta$ 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

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

# Bayesian inference: conjugacy

## The idea

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).

## Definition

Conjugacy can be defined as follows (*gelman2013bayesian*) . If $\mathcal{F}$ is a class of sampling distributions and $\mathcal{P}$ is a class of prior distributions for $\theta$, then the class $\mathcal{P}$ is *conjugate* for $\mathcal{F}$ if

$$p(\theta \mid y) \in \mathcal{P} \text{ for all } p(\cdot \mid \theta) \in \mathcal{F} \text{ and } p(\cdot) \in \mathcal{P}$$

Technical note: the condition trivially holds if $\mathcal{P}$ is taken to be the space of all probability distributions!

# Bayesian inference: 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 $\theta$ that we still had after seeing data $x$.

# Bayesian inference: tractability notes

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

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

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