Statistics and Machine Learning 1

Lecture 7D: Conjugate Priors

Mark Muldoon
Department of Mathematics, Alan Turing Building
University of Manchester

Week 7

Bayesian polling: the ingredients

In our polling example, k out of N people polled supported Asha and we sought to infer p the proportion of voters who support her.

The Bayesian ingredients were:

$P(k \,|\, p, N)$ the likelihood,

$$P(k \,|\, p, N) = \frac{N!}{k!\,(N-k)!}\, p^k (1-p)^{N-k}.$$

$P(p)$ the *prior on the parameter*. We chose an uninformative, uniform prior which turns out to be a Beta distribution with $\alpha = \beta = 1$.

$P(D)$ the *prior over the data*. Determined by the likelihood and the prior on the parameter: we'll compute it soon.

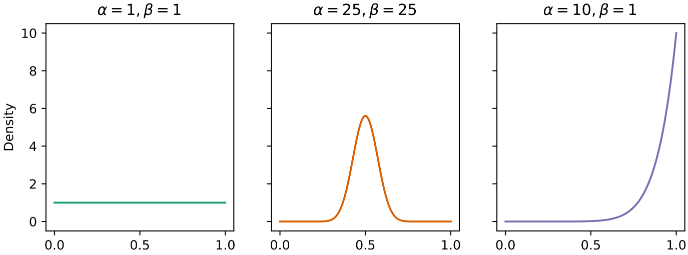
$P(p \,|\, D)$ the *posterior distribution over* p. We'll compute this too.

The Beta distribution, a refresher

The density of the Beta distribution depends on two positive *shape parameters*, α and β, and is given by

$$f_{\alpha\beta}(x) = \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right) x^{\alpha-1}(1-x)^{\beta-1}.$$

where $\Gamma(z)$ is a function that, among other properties, satisfies $\Gamma(n) = (n-1)!$ for positive integers n. The mean of the Beta distribution is $\mu = \alpha/(\alpha+\beta)$.



A useful integral

The Beta distribution is a properly-normalised probability distribution, so:

$$
\begin{aligned}
1 &= \int_0^1 f_{\alpha\beta}(x)\,dx \\
&= \int_0^1 \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right) x^{\alpha-1}(1-x)^{\beta-1}\,dx \\
&= \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right) \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx
\end{aligned}
$$

which implies the following useful result:

$$
\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx.
$$

Binomial Bayes: priors on the data

To compute the prior on the data, we need to do the following integral

$$
\begin{aligned}
P(k) &= \int_0^1 P(k \,|\, p, N) f_{\alpha\beta}(p)\, dp \\
&= \int_0^1 \left(\frac{N!}{k!\,(N-k)!}\, p^k (1-p)^{N-k} \right) \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, p^{\alpha-1} (1-p)^{\beta-1} \right) dp \\
&= \frac{N!}{k!\,(N-k)!} \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \right) \int_0^1 p^{k+\alpha-1} (1-p)^{N-k+\beta-1}\, dp
\end{aligned}
$$

Note that, except for the constants out front, this looks like an integral over a Beta distribution with shape parameters

$$
\alpha' = k + \alpha \qquad \text{and} \qquad \beta' = N - k + \beta,
$$

which implies

$$
P(k) = \frac{N!}{k!\,(N-k)!} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(k+\alpha)\Gamma(N-k+\beta)}{\Gamma(N+\alpha+\beta)}
$$

Binomial Bayes: the posterior distribution

Using Bayes' Theorem to combine all the results from the previous slides produces

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

That is, the posterior is a Beta distribution! This the main reason to use a Beta distribution for the prior on p.

For this reason, the Beta distribution is called the *conjugate prior* to the Binomial likelihood. There are a handful of commonly-used conjugate pairs of distributions:

Likelihood	Prior	Posterior
Binomial	Beta	Beta
Gaussian	Gaussian	Gaussian
Gaussian	Gamma	Gamma