Parameter Estimation

Brendon J. Brewer

Department of Statistics
The University of Auckland

https://www.stat.auckland.ac.nz/~brewer/

Parameter estimation

Bayesian inference is mostly used for *parameter estimation*. Let θ be an unknown quantity or 'parameter', and D be some data.

Parameter estimation

Bayesian inference is mostly used for *parameter estimation*. Let θ be an unknown quantity or 'parameter', and D be some data. A version of Bayes' rule applies to probability distributions:

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{p(D)} \tag{1}$$

Each term is a distribution, not a single probability.

Why?

Consider three statements *about the value of* θ: for example, $\theta = 3.5, \theta = 3.6, \theta = 3.7$. Consider one statement *about the value of* D: for example, $D = 5.9$.

Why?

$$P(\theta = 3.5|D = 5.9) = \frac{P(\theta = 3.5)P(D = 5.9|\theta = 3.5)}{P(D = 5.9)} \quad (2)$$

$$P(\theta = 3.6|D = 5.9) = \frac{P(\theta = 3.6)P(D = 5.9|\theta = 3.6)}{P(D = 5.9)} \quad (3)$$

$$P(\theta = 3.7|D = 5.9) = \frac{P(\theta = 3.7)P(D = 5.9|\theta = 3.7)}{P(D = 5.9)} \quad (4)$$

Why?

$$P(\theta = 3.5|D = 5.9) = \frac{P(\theta = 3.5)P(D = 5.9|\theta = 3.5)}{P(D = 5.9)} \quad (5)$$

$$P(\theta = 3.6|D = 5.9) = \frac{P(\theta = 3.6)P(D = 5.9|\theta = 3.6)}{P(D = 5.9)} \quad (6)$$

$$P(\theta = 3.7|D = 5.9) = \frac{P(\theta = 3.7)P(D = 5.9|\theta = 3.7)}{P(D = 5.9)} \quad (7)$$

Green = posterior distribution
Red = prior distribution
Blue = likelihood function
Black = Marginal likelihood / evidence

Why?

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{p(D)} \tag{8}$$

Green = posterior distribution
Red = prior distribution
Blue = likelihood function
Black = Marginal likelihood / evidence

Inputs and outputs

Inference:
(choice of prior, choice of likelihood, the data) $\rightarrow$ (posterior, marginal likelihood)

Example

If we knew the intensity of an X-ray source was λ, we would expect to see λt photons arrive in a time interval of length t. Let x_1, x_2, x_3 be the number of photons observed in three consecutive minutes.

$$x_i|\lambda \sim \text{Poisson}(\lambda) \tag{9}$$

i.e.,

$$p(x_1, x_2, x_3|\lambda) = \prod_{i=1}^{3} \frac{\lambda^{x_i} e^{-\lambda x_i}}{x_i!} \tag{10}$$

Suppose we observed $(x_1, x_2, x_3) = (21, 14, 22)$, and want to infer λ.

Example

For Bayes' rule:

$$p(\lambda|x_1, x_2, x_3) = \frac{p(\lambda)p(x_1, x_2, x_3|\lambda)}{p(x_1, x_2, x_3)} \quad (11)$$

we need a prior, $p(\lambda)$.

The log-uniform prior

How long is a piece of string?

The log-uniform prior

How long is a piece of string?
Twice the distance from the middle to one end.
$\implies$

$$p(\lambda) \propto \frac{1}{\lambda} \tag{12}$$

This is called a log-uniform prior. It ‘puts the error bars in the exponent’.

Numerical solution with a grid

```
# The data
x = np.array([21, 14, 22])

# Grid of possible parameter values
lamb = np.linspace(0.01, 100.0, 10001)

# Prior density
prior = 1.0 / lamb
prior = prior / np.trapz(prior, x=lamb)

# Likelihood
lik = lamb**(np.sum(x)) * np.exp(-3*lamb)\
          / np.prod(scipy.misc.factorial(x))

# Calculate and plot posterior
h = prior*lik
Z = np.trapz(h, x=lamb)
post = h/Z
plt.plot(lamb, post)
```