STAT40380/STAT40390/STAT40850
Bayesian Analysis

Dr Niamh Russell

School of Mathematics and Statistics
University College Dublin

niamh.russell@ucd.ie

February 2016

The exponential family

- A distribution is said to be in the *one-parameter exponential family* if it can be written in the form:

$$p(x|\theta) = g(x)h(\theta)\exp[t(x)\psi(\theta)]$$

- Examples include the mean and variance of the Normal distribution, the rate parameter in the Poisson distribution, and the probability parameter in the Binomial distribution, amongst many others.
- If we have multiple observations then the likelihood is:

$$p(\mathbf{x}|\theta) = \prod_{i=1}^{n} p(x_i|\theta) \propto h(\theta)^n \exp\left(\sum_{i=1}^{n}(t(x_i)\psi(\theta))\right)$$

- The statistic $\sum t(x_i)$ is sufficient for θ.

Conjugacy and the exponential family

- Recall that if a prior distribution is conjugate to the likelihood then the prior and the posterior are of the same distributional form.
- Thus, to choose prior distributions that are conjugate we need them to be of a similar form to the likelihood.
- Following this line of thought, if we are to find a prior distribution conjugate to the likelihood we require that:

$$p(\theta) \propto h(\theta)^{\nu} \exp(\tau \psi(\theta))$$

  where τ and ν are constants.
- For example the prior for the mean of a normal distribution with fixed variance σ^2 can be of the form:

$$p(\theta) \propto \exp\left[-\frac{\nu\theta^2}{2\sigma^2}\right] \exp\left[\frac{\tau\theta}{\sigma^2}\right]$$

The two-parameter exponential family

- The exponential family can be extended to cover situations where we have two unknown parameters.
- A density is said to come from the *two-parameter exponential family* if:

$$p(x|\theta,\phi) = g(x)h(\theta,\phi)\exp[t(x)\psi(\theta,\phi) + u(x)\chi(\theta,\phi)]$$

- As before, if we have multiple observations then:

$$\prod_{i=1}^{n} p(x_i|\theta,\phi) \propto h(\theta,\phi)^n \exp\left[\sum t(x_i)\psi(\theta,\phi) + \sum u(x_i)\chi(\theta,\phi)\right]$$

- Here, the quantities $\sum t(x_i), \sum u(x_i)$ are jointly sufficient for $[\theta,\phi]$ and are sometimes represented by the vector $[\sum t(x_i), \sum u(x_i)]$.
- The family of priors for conjugate densities here now has the form:

$$p(\theta,\phi) \propto h(\theta,\phi)^\nu \exp[\tau\psi(\theta,\phi) + \omega\chi(\theta,\phi)]$$

  with constants ν, ω and τ.

Multi-parameter problems

- We can extend Bayes' theorem to deal with problems with two parameters:

$$p(\theta, \phi|\mathbf{x}) \propto p(\mathbf{x}|\theta, \phi)p(\theta, \phi)$$

- By extension, we can consider a vector $\boldsymbol{\theta}$ of any length and write:

$$p(\boldsymbol{\theta}, |\mathbf{x}) \propto p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

- We are now finding a *joint* posterior distribution for a set of parameters $\boldsymbol{\theta}$ given a likelihood and a *joint* prior.
- All the techniques we have learnt previously still apply, but it will be harder to find conjugate distributions and harder to visualise the results.

Multi-parameter problems 2

- When deciding on our joint prior we might consider three possibilities:
  1. That the parameters are *a priori jointly distributed*
     $$p(\boldsymbol{\theta}) = p(\theta_1, \theta_2, \ldots, \theta_n)$$
  2. That the parameters are *a priori independent*
     $$p(\boldsymbol{\theta}) = p(\theta_1) \times p(\theta_2) \times \cdots \times p(\theta_n)$$
  3. That the parameters are *a priori hierarchically related*, eg
     $$p(\theta, \phi) = p(\theta|\phi)p(\phi)$$
- When we have formed our joint posterior we often try and find *marginal posterior distributions* of the parameters, eg:
  $$p(\theta|\mathbf{x}) = \int p(\theta, \phi|\mathbf{x}) d\phi$$

Example 1: Multi-parametric Bayesian inference for the normal distribution

Example

Suppose that $x_i \sim N(\theta, \phi)$ with θ and ϕ unknown. Find the joint posterior distribution of (θ, ϕ) when using the improper prior distribution $p(\theta, \phi) \propto \frac{1}{\phi}$

Example 2: Marginal distribution for the mean

Example

Using your joint posterior distribution obtained in Example 1, find the marginal distribution of the mean to give $p(\theta|\mathbf{x})$

Example 3: Marginal distribution for the variance

Example

Again using your joint posterior distribution obtained in Example 1, find the marginal distribution of the variance to give $p(\phi|\mathbf{x})$

Example 4: Marginal distribution fof the mean for rats' weight

Example

From the rats' weight data, we had $n = 20, \sum x_i = 420$, $\sum x_i^2 = 9484$, and $\sum (x_i - \bar{x})^2 = 664$. Use these data to find the marginal distributions of the mean and variance.

Multi-parameter problems: some remarks

- Bayesian methods are just as applicable with thousands of parameters as they are with just one
- However, the practical problem of calculating a posterior and understanding the results gets much more complicated
- Most importantly, specifying a reasonable prior distribution in the presence of many interrelated parameters is a very hard task