Suppose that \( \{X_1, \ldots, X_n\} \) are independent and identically distributed random variables with probability density/mass function \( p(x) \) and \( x \) is the vector of the unknown parameters, then the likelihood function is defined as \( L(\theta; x) = \prod_i p(x_i | \theta) \). In Bayesian inference it is assumed that these unknown distribution parameters are random variables which are given prior distributions to indicate our initial beliefs and knowledge about them, where the joint prior of these parameters is denoted by \( p(\theta) \). After observing the data we can update our beliefs by finding the posterior distribution function \( p(\theta | x) \) using Bayes theorem

\[
p(\theta | x) \propto p(x | \theta) p(\theta)
\]

where \( p(x) \) is a function of the data that makes the posterior density function integrate to one over the whole range of \( \theta \).

Suppose that we leave the observation \( x \) out of the sample and want to predict it based on the rest of the data then the Conditional Predictive Ordinate which is the predictive distribution for \( x \) conditional on all of the other observations, is given by the following expression

\[
T(x) = \int p(x | \theta) p(\theta | x) \mathrm{d} \theta
\]

where \( x \) is a vector of the whole data set excluding \( x \).

The Bayes Factor can be used to compare two different models \( M_1 \) and \( M_2 \). It is defined as

\[
B_{12} = \frac{\int p(x | \theta) p(\theta | x) \mathrm{d} \theta}{\int p(x | \theta) p(\theta | x) \mathrm{d} \theta}
\]

where \( p(x | \theta) \), \( p(\theta | x) \) are the likelihood and prior under the model \( M_1 \). In the problems that we have considered, the Bayes Factor should be less than 0.015 for us to be able to select the model \( M_1 \) in favour of \( M_2 \), noting that the smaller \( B_{12} \) is, the more evidence we have to select the model \( M_1 \).

An outlier is an observation that is extreme compared to the rest of the sample. We would like to model outliers as observations that are generated by different probability distributions to the rest of the sample.

### Background

The detection of outliers in samples from univariate distributions has received a good deal of attention. Barnard and Lewis (1993) review methods for many distributions. They suggest that for the Uniform distribution, the result that successive differences between the order statistics have an Exponential distribution can be utilised to construct tests. For multivariate observations Barnett (1982) described a series of classical tests that were based on simulations to test whether extreme observations in a sample were outliers. He did this for a variety of different probability distributions, including the Uniform, Normal, Exponential and Pareto distributions. For the Uniform he assumes the end points of the intervals are known and we do not see how any observation which lies in a known rectangle can be described as an outlier. From a Bayesian viewpoint a number of papers have considered using the conditional predictive ordinate to test if there is sufficient evidence to declare them as outliers. Pettit and Smith (2011) used a similar approach to detect outliers and Bayes Factors to test if there is sufficient evidence to declare them as outliers. Pettit and Smith (2011) use this approach for samples from the Normal distribution and linear models. Pettit (1988) considers this problem for Exponential distributions. Pettit (1990) looks at the multivariate Normal distribution in some detail and approaches this problem by deriving various results using the conditional predictive ordinate. Since then this problem has been considered using these Bayesian methods for both the Poisson and Binomial cases in Pettit (1996) and Sofroniou and Pettit (2005). We have studied this problem for both the Uniform and Pareto distributions using the ideas from these papers. In February 2011 we submitted a paper containing our results for the Uniform case and hope to submit a similar paper for the Pareto distribution by the end of June 2011.

### Modelling Outliers From One Sided Uniform Samples

If a random variable \( X \sim \text{Uniform}(\theta, \theta) \), it is said to have a one sided uniform distribution with probability density function \( p(x) = \frac{1}{2\theta} \), for \( \theta > x \). Bayesian theory told us that a sensible prior to take for \( \theta \) is one that is conjugate, meaning that the prior makes the posterior have the same form, which in this case is a Pareto(\( \alpha, \theta \)) prior with probability density function \( p(\theta) = \frac{\alpha}{\theta^{\alpha+1}} \), for \( \theta > 0 \), and where \( \alpha \) and \( \theta \) are both assumed to be known. For this problem, we defined an outlier as something from a Uniform(0, \( \theta \)) distribution, where \( \theta > 1 \).

For the one sided Uniform sample we have shown that the largest observation in the sample has the smallest conditional predictive ordinate and have derived the Bayes Factor for testing whether or not it is an outlier when both the amount of contamination is known and unknown. When \( \theta \) was unknown it was given a Pareto(\( \beta, 1 \)) prior distribution, as the probability that an observation is more and more extreme gets smaller rather quickly and symmetrically, where it is assumed that \( \beta \) is known and is less than one. We then derived the Bayes Factors for the case of testing whether a set of \( q \) observations are outliers generated by the same probability distribution and for the case of testing whether a set of \( q \) observations are upper outliers generated by the same probability distribution. Finally we saw that all of our results for both the one sided and two sided Uniform distributions could quite easily be extended to the multivariate case.

As an example to illustrate the methods in this section, we consider the following sample of 18 observations and assume that \( \alpha = \frac{1}{\theta} \). Note that the expressions for the following Bayes Factors do not depend on the observations that are not given.

\[
B_{12} = 1.1117 \times 10^{-35}
\]

It can be shown that when \( \lambda = \frac{1}{\theta} \), the Bayes Factor for comparing the models with no outliers and one upper outlier is equal to 0.00056. By comparing this to our answer from a previous example with a smaller sample size, we know that masking has definitely occurred.

We can clearly see that 0.97 cannot possibly be an upper outlier and therefore have to compare the models \( M_1 \) and \( M_{1,1} \), (one upper outlier and one lower outlier) using \( \lambda_1 = \lambda_2 = \frac{1}{\theta} \). Hence

\[
B_{11,2} = 7.7134 \times 10^{-2}
\]

and so we should certainly select \( M_{1,1} \) over \( M_1 \). We do not have to perform any more iterations because we can clearly see that 0.98 cannot possibly be a lower outlier, therefore we select \( M_{1,1} \) as our final model and conclude that 5.29 is an upper outlier and 5.22 is a lower outlier.

### Modelling Outliers From Pareto Samples

If a random variable \( X \sim \text{Pareto}(\theta, k) \), it has probability density function \( p(x, \theta, k) = \frac{\theta^k x^{-k-1}}{\Gamma(k)} \), for \( x > \theta \). In this problem we assumed that both \( \theta \) and \( k \) were unknown. We were given a conjugate gamma(\( \alpha, \beta \)) prior and \( k \) was given an improper prior such that \( k \) has a Uniform(0, \( \infty \)) distribution, so that \( p(k) = \frac{1}{\infty} \), when \( \theta \) was unknown. Then the uniform constant whose exact value did not matter. Note that if \( k \) was assumed to be known, if we use the transformation \( Y = \log (\frac{1}{k}) \) to transform our data to an Exponential(\( \theta \)) sample, then the methods in Petit (1985) can be used to model outliers. For this problem, we defined an outlier as something from a Pareto(\( \beta, \delta k \)) distribution, as this increases the expected value of \( X \) which is \( \frac{\theta}{\delta} \). Also \( \delta \) was given the improper prior \( p(\delta) = \frac{1}{\delta^2} \), where the constant \( \delta \) was determined by using the method of imaginary observations. For this problem it was sensible to use a noninformative prior on \( \delta \) as the Pareto distribution is so heavy tailed.

We came to exactly the same conclusions for this problem, as before, noting that here like for the two sided Uniform distribution it is not always reasonable to assume that our outliers are generated by the same probability distribution if they differ a lot.

### Future Work

We have still to get finish off the Pareto problem by extending these ideas to the bi-variate and multivariate cases, then obviously we submit the second paper. I would also like to find ways of detecting, modelling and measuring the influence of outliers for more general multivariate distributions.