2020 SISG Module 8: Bayesian Statistics for Genetics
Lecture 3: Binomial Sampling 2

Jon Wakefield

Departments of Statistics and Biostatistics
University of Washington

Outline

This lecture

In this lecture we continue our examination of Bayesian inference for binomial data and discuss:

- prior specification,
- predictive distributions and
- testing.

Prior Specification

Prior Choice and Prior Sensitivity

- For small datasets in particular it is a good idea to examine the sensitivity of inference to the prior choice, particularly for those parameters for which there is little information in the data.
- An obvious way to determine the latter is to compare the prior with the posterior, but experience often aids the process.
- Sometimes one may specify a prior that, in some sense, allows the data to dominate the posterior.
- In some situations, priors can be found that produce point and interval estimates that mimic a standard non-Bayesian analysis, i.e., have good frequentist properties.
- Such priors provide a baseline to compare analyses with more substantive priors.
- Other names for such priors are objective, reference and non-subjective.
- We discuss subjective priors that reflect the data analysts belief about the unknowns.

Choosing a Prior, Approach One

- Recall that to specify a beta distribution, we need to specify two quantities, a and b, which are difficult to interpret.
- The posterior mean is

$$\mathrm{E}[\theta|y] = \frac{y+a}{N+a+b}.$$

- Viewing the denominator as a sample size suggests a method for choosing a and b.
- We may specify the prior mean $m_{\text{prior}} = a/(a+b)$ and the "prior sample size" $N_{\text{prior}} = a+b$
- We then solve for a and b via

$$\begin{aligned} a &= N_{\text{prior}} \times m_{\text{prior}} \\ b &= N_{\text{prior}} \times (1 - m_{\text{prior}}). \end{aligned}$$

- Intuition: a is like a prior number of successes and b like the prior number of failures.

An Example

- Suppose we set $N_{\text{prior}} = 5$ and $m_{\text{prior}} = \frac{2}{5}$.
- It is as if we saw 2 successes out of 5.
- Suppose we obtain data with $N = 10$ and $\frac{y}{N} = \frac{7}{10}$.
- Hence $W = 10/(10+5)$ and

$$\begin{aligned} \mathrm{E}[\theta|y] &= \frac{7}{10} \times \frac{10}{10+5} + \frac{2}{5} \times \frac{5}{10+5} \\ &= \frac{9}{15} = \frac{3}{5}. \end{aligned}$$

- Solving:

$$\begin{aligned} a &= N_{\text{prior}} \times m_{\text{prior}} = 5 \times \frac{2}{5} = 2 \\ b &= N_{\text{prior}} \times (1 - m_{\text{prior}}) = 5 \times \frac{3}{5} = 3 \end{aligned}$$

- This gives a $\text{Beta}(y+a, N-y+b) = \text{Beta}(7+2, 3+3)$ posterior.

Beta Prior, Likelihood and Posterior

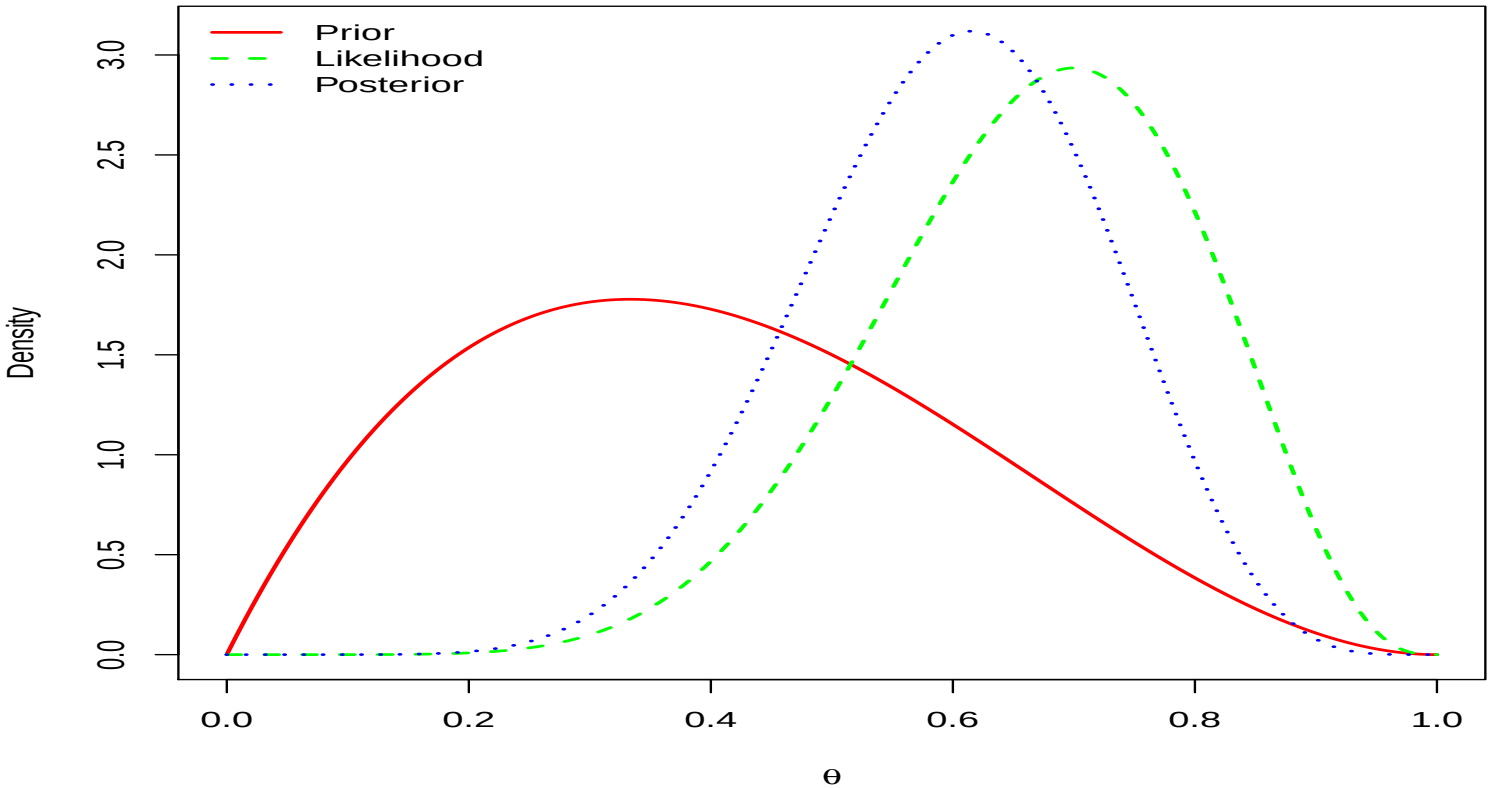

Figure 1: The prior is Beta(2,3) the likelihood is proportional to a Beta(7,3) and the posterior is Beta(7+2,3+3).

Choosing a Prior, Approach Two

- An alternative convenient way of choosing a and b is by specifying two quantiles for θ with associated (prior) probabilities.
- For example, we may wish $\Pr(\theta < 0.1) = 0.05$ and $\Pr(\theta > 0.6) = 0.05$.
- The values of a and b may be found numerically. For example, we may solve

$$[p_1 - \Pr(\theta < q_1|a,b)]^2 + [p_2 - \Pr(\theta < q_2|a,b)]^2 = 0$$

for a, b.

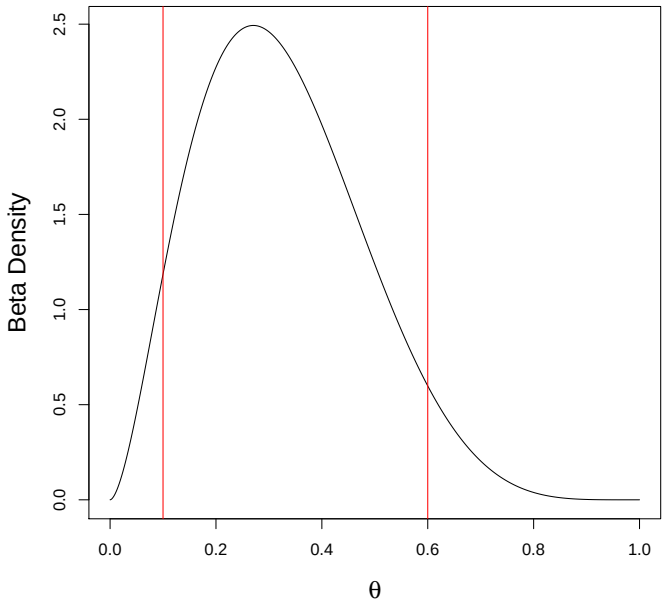

Figure 2: Beta(2.73,5.67) prior with 5% and 95% quantiles highlighted.

Beware improper priors (and especially improper posteriors!)

- The beta distribution is proper (i.e., integrates to 1 over [0,1]), if $a, b > 0$.
- If we use a proper beta prior we are guaranteed a proper posterior.
- If we choose $a = b = 0$, the prior is

$$p(\theta) \propto \frac{1}{\theta(1-\theta)},$$

  which is improper.
- If $y = 0$ or $y = N$ the posterior is also improper (one infinite weight endpoint remains).

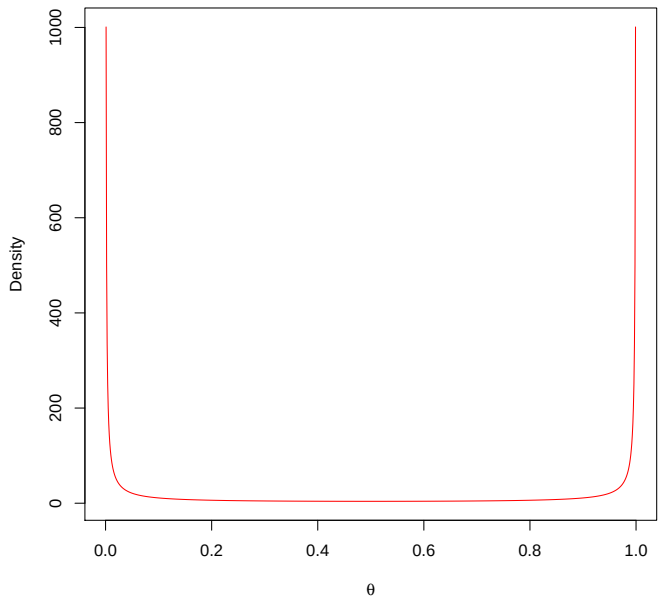

Figure 3: Beta(0,0) improper prior.

Bayesian Sequential Updating

- We show how probabilistic beliefs are updated as we receive more data.
- Suppose the data arrives sequentially via two experiments:
  1. Experiment 1: (y_1, N_1).
  2. Experiment 2: (y_2, N_2).
- Prior 1: $\theta \sim \text{Beta}(a, b)$.
- Likelihood 1: $y_1|\theta \sim \text{Binomial}(N_1, \theta)$.
- Posterior 1: $\theta|y_1 \sim \text{Beta}(a + y_1, b + N_1 - y_1)$.
- This posterior forms the prior for experiment 2.
- Prior 2: $\theta \sim \text{Beta}(a^\star, b^\star)$ where $a^\star = a + y_1$, $b^\star = b + N_1 - y_1$.
- Likelihood 2: $y_2|\theta \sim \text{Binomial}(N_2, \theta)$.
- Posterior 2: $\theta|y_1, y_2 \sim \text{Beta}(a^\star + y_2, b^\star + N_2 - y_2)$.
- Substituting for $a^\star, b^\star$:

$$\theta|y_1, y_2 \sim \text{Beta}(a + y_1 + y_2, b + N_1 - y_1 + N_2 - y_2).$$

Bayesian Sequential Updating

- Schematically:

  $$(a, b) \to (a + y_1, b + N_1 - y_1) \to (a + y_1 + y_2, b + N_1 - y_1 + N_2 - y_2)$$

- Suppose we obtain the data in one go as $y^\star = y_1 + y_2$ successes from $N^\star = N_1 + N_2$ trials.
- The posterior is

  $$\theta | y^\star \sim \text{Beta}(a + y^\star, b + N^\star - y^\star),$$

  which is the same as when we receive in two separate instances.

Prediction

Predictive Distribution

- Suppose we see y successes out of N trials, and now wish to obtain a predictive distribution for a future experiment with M trials.
- Let $Z = 0, 1, \ldots, M$ be the number of successes.
- Predictive distribution:

$$\begin{aligned}
\Pr(z|y) &= \int_0^1 p(z, \theta|y) d\theta \\
&= \int_0^1 \Pr(z|\theta, y) p(\theta|y) d\theta \\
&= \int_0^1 \underbrace{\Pr(z|\theta)}_{\text{binomial}} \times \underbrace{p(\theta|y)}_{\text{posterior}} d\theta
\end{aligned}$$

where we move between lines 2 and 3 because z is conditionally independent of y given θ.

Predictive Distribution

Continuing with the calculation:

$$
\begin{aligned}
\Pr(z|y) &= \int_0^1 \Pr(z|\theta) \times p(\theta|y) d\theta \\
&= \int_0^1 \binom{M}{z} \theta^z (1-\theta)^{M-z} \\
&\times \frac{\Gamma(N+a+b)}{\Gamma(y+a)\Gamma(N-y+b)} \theta^{y+a-1} (1-\theta)^{N-y+b-1} d\theta \\
&= \binom{M}{z} \frac{\Gamma(N+a+b)}{\Gamma(y+a)\Gamma(N-y+b)} \int_0^1 \theta^{y+a+z-1} (1-\theta)^{N-y+b+M-z-1} d\theta \\
&= \binom{M}{z} \frac{\Gamma(N+a+b)}{\Gamma(y+a)\Gamma(N-y+b)} \frac{\Gamma(a+y+z)\Gamma(b+N-y+M-z)}{\Gamma(a+b+N+M)}
\end{aligned}
$$

for $z = 0, 1, \ldots, M$.

A likelihood approach would take the predictive distribution as Binomial$(M, \widehat{\theta})$ with $\widehat{\theta} = y/N$: this does not account for estimation uncertainty.

In general, we have sampling uncertainty (which we can't get away from) and estimation uncertainty.

Predictive Distribution

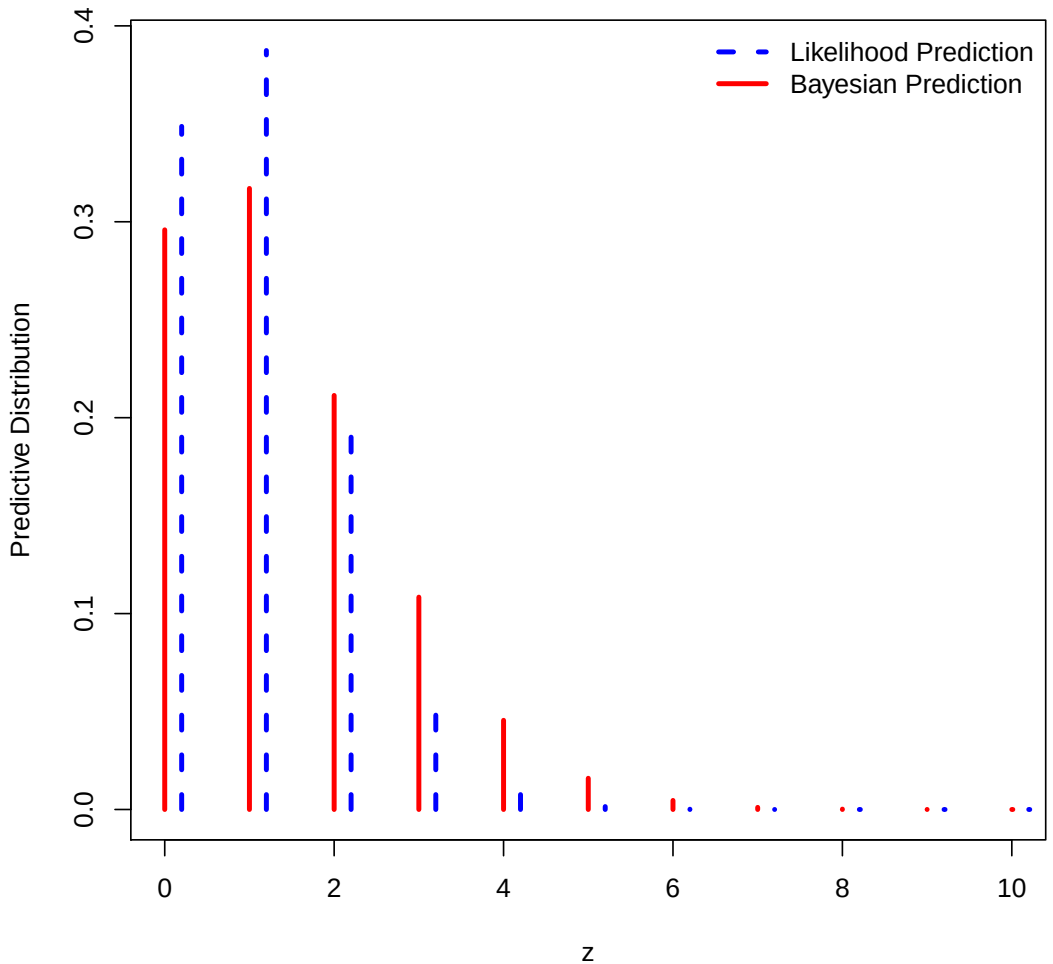

Figure 4: Likelihood and Bayesian predictive distribution of seeing $z = 0, 1, \ldots, M = 10$ successes, after observing $y = 2$ out of $N = 20$ successes (with $a = b = 1$).

Predictive Distribution: A General Approach

The posterior and sampling distributions won't usually combine so conveniently.

In general, we may form a Monte Carlo estimate of the predictive distribution:

$$\begin{aligned} p(z|y) &= \int p(z|\theta)p(\theta|y)d\theta \\ &= \mathrm{E}_{\theta|y}[p(z|\theta)] \\ &\approx \frac{1}{S}\sum_{s=1}^{S} p(z|\theta^{(s)}) \end{aligned}$$

where $\theta^{(s)} \sim p(\theta|y)$, $s = 1, \ldots, S$, is a sample from the posterior.

This provides an estimate of the predictive distribution at the point z.

Predictive Distribution: A General Approach

- Alternatively, we may sample from $p(z|\theta^{(s)})$ a large number of times to reconstruct the predictive distribution.
- First sample from the posterior:

$$\theta^{(s)}|y \sim p(\theta|y).$$

- Next sample from the likelihood:

$$z^{(s)}|\theta^{(s)} \sim p(z|\theta^{(s)}),$$

for $s = 1, \ldots, S$.

- To give a sample $z^{(s)}$ from the posterior, this is illustrated to the right.

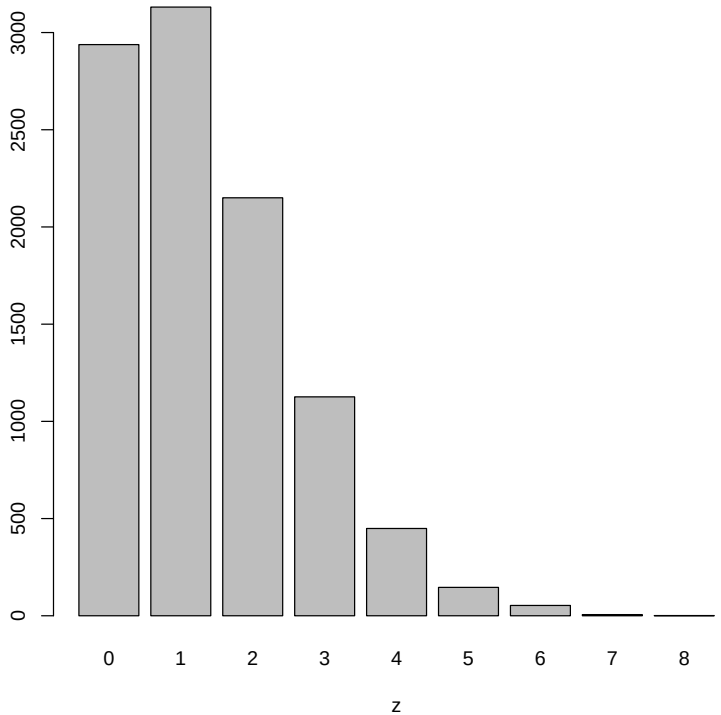

Figure 5: Sampling version of prediction in Figure 4, based on $S = 10,000$ samples.

Difference in Binomial Proportions

- It is straightforward to extend the methods presented for a single binomial sample to a pair of samples.
- Suppose we carry out two binomial experiments:

$$\begin{aligned} Y_1|\theta_1 &\sim \text{Binomial}(N_1, \theta_1) \quad \text{for sample 1} \\ Y_2|\theta_2 &\sim \text{Binomial}(N_2, \theta_2) \quad \text{for sample 2} \end{aligned}$$

- Interest focuses on $\theta_1 - \theta_2$, and often in examing the possibitlity that $\theta_1 = \theta_2$.
- With a sampling-based methodology, and independent beta priors on θ_1 and θ_2, it is straightforward to examine the posterior $p(\theta_1 - \theta_1|y_1, y_2)$.

Difference in Binomial Proportions

- Savage *et al.* (2008) give data on allele frequencies within a gene that has been linked with skin cancer.
- It is interest to examine differences in allele frequencies between populations.
- We examine one SNP and extract data on Northern European (NE) and United States (US) populations.
- Let θ_1 and θ_2 be the allele frequencies in the NE and US population from which the samples were drawn, respectively.
- The allele frequencies were 10.69% and 13.21% with sample sizes of 650 and 265, in the NE and US samples, respectively.
- We assume independent Beta(1,1) priors on each of θ_1 and θ_2.
- The posterior probability that $\theta_1 - \theta_2$ is greater than 0 is 0.12 (computed as the proportion of the samples $\theta_1^{(s)} - \theta_2^{(s)}$ that are greater than 0), so there is little evidence of a difference in allele frequencies between the NE and US samples.

Binomial Two Sample Example

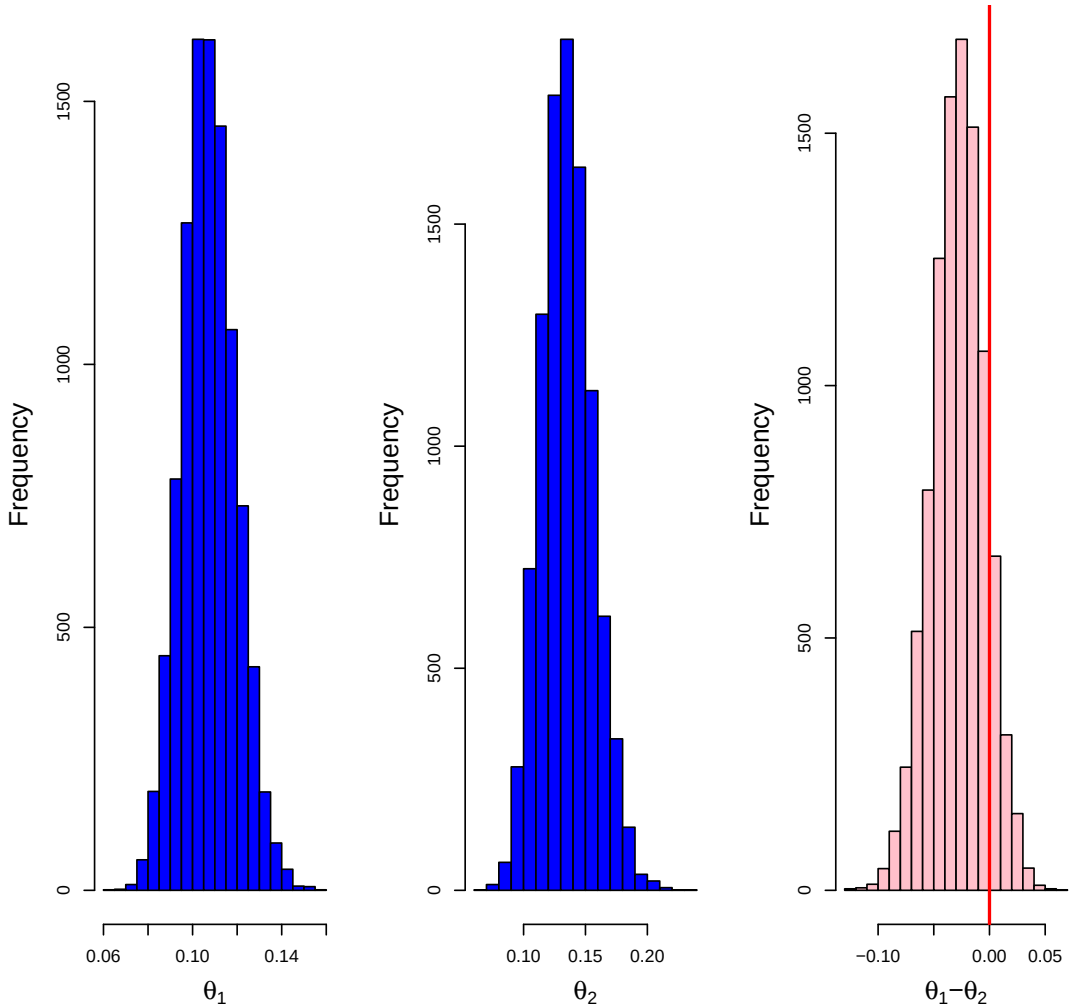

Figure 6: Histogram representations of $p(\theta_1|y_1)$, $p(\theta_2|y_2)$ and $p(\theta_1 - \theta_2|y_1, y_2)$. The red line in the right plot is at the reference point of zero.

Bayes Factors

Bayes Factors for Hypothesis Testing

- The Bayes factor provides a summary of the evidence for a particular hypothesis (model) as compared to another.
- The Bayes factor is

$$\text{BF} = \frac{\Pr(y|H_0)}{\Pr(y|H_1)}$$

  and so is simply the probability of the data under H_0 divided by the probability of the data under H_1.
- Values of BF > 1 favor H_0 while values of BF < 1 favor H_1.
- Note the similarity to the likelihood ratio

$$\text{LR} = \frac{\Pr(y|H_0)}{\Pr(y|\widehat{\theta})}$$

  where $\widehat{\theta}$ is the MLE under H_1.
- If there are no unknown parameters in H_0 and H_1 (for example, $H_0 : \theta = 0.5$ versus $H_1 : \theta = 0.3$), then the Bayes factor is identical to the likelihood ratio.

Calibration of Bayes Factors

- Kass and Raftery (1995) suggest intervals of Bayes factors for reporting:

1/Bayes Factor	Evidence Against H_0
1 to 3.2	Not worth more than a bare mention
3.2 to 20	Positive
20 to 150	Strong
$>$150	Very strong

- These provide a guideline, but should not be followed without question.

Example: Bayes Factors for Binomial Data

For each gene in the ASE dataset we may be interested in $H_0 : \theta = 0.5$ versus $H_1 : \theta \neq 0.5$.

The numerator and denominator of the Bayes factor are:

$$
\begin{aligned}
\Pr(y|H_0) &= \binom{N}{y} 0.5^y 0.5^{N-y} \\
\Pr(y|H_1) &= \int_0^1 \binom{N}{y} \theta^y (1-\theta)^{N-y} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1} d\theta \\
&= \binom{N}{y} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(y+a)\Gamma(N-y+b)}{\Gamma(N+a+b)}
\end{aligned}
$$

Values Taken by the Negative Log Bayes Factor, as a Function of y

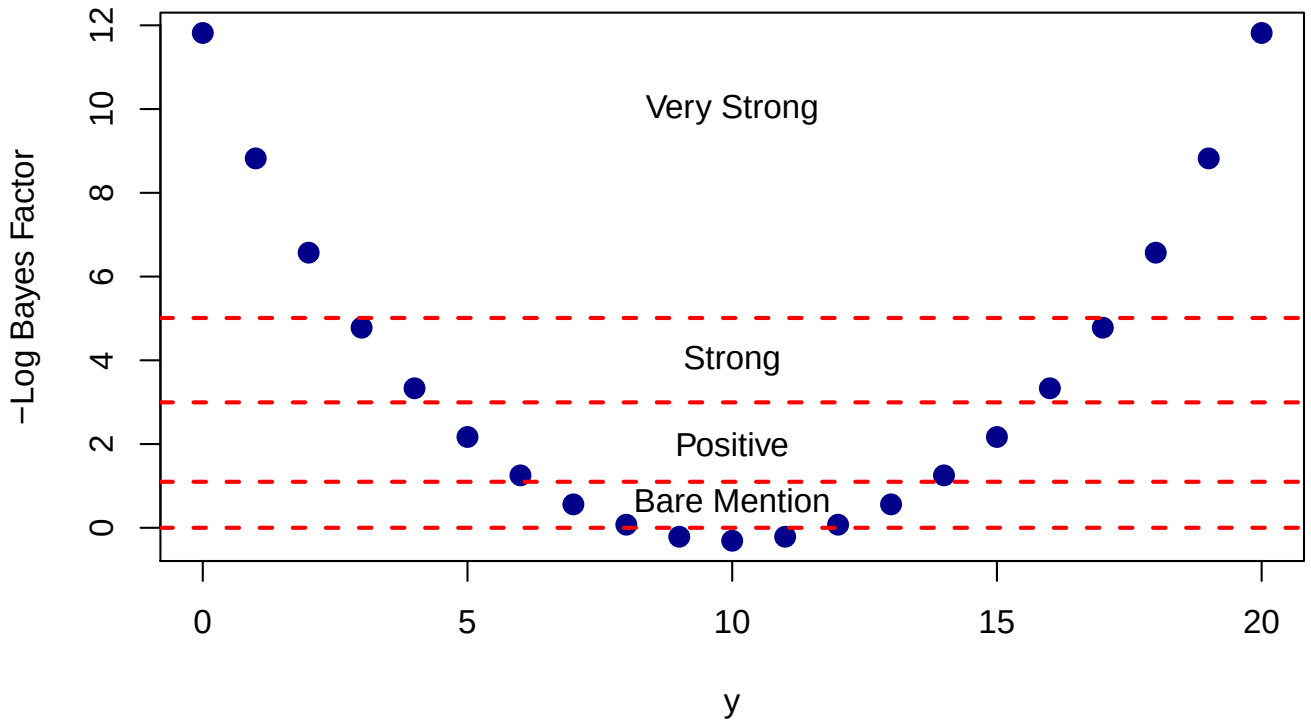

Figure 7: Negative Log Bayes factor as a function of $y|\theta \sim \text{Binomial}(20, \theta)$ for $y = 0, 1, \ldots, 20$ and $a = b = 1$. High values indicate evidence against the null.

Analysis of ASE Data

Three Approaches to Inference for the ASE Data

1. Posterior Probabilities:
   - A simple approach to testing is to calculate the posterior probability that $\theta < 0.5$.
   - We can then pick a threshold for indicating worthy of further study, e.g. if $\Pr(\theta < 0.5|y) < 0.01$ or $\Pr(\theta < 0.5|y) > 0.99$
2. Bayes Factors:
   - Calculating the Bayes factor.
   - Pick a threshold for indicating worthy of further study, e.g. if reciprocal of the Bayes factor is greater than 150.
3. Decision theory:
   - Place priors on the null and alternative hypotheses.
   - Calculate the posterior odds:

$$\begin{aligned} \frac{\Pr(H_0|y)}{\Pr(H_1|y)} &= \frac{\Pr(y|H_0)}{\Pr(y|H_1)} \times \frac{\Pr(H_0)}{\Pr(H_1)} \\ \text{Posterior Odds} &= \text{Bayes Factor} \times \text{Prior Odds} \end{aligned}$$

   - Pick a threshold R, so that if the Posterior Odds $<$ R we choose H_1.

Bayesian Analysis of the ASE Data

- Here we give a histogram of the posterior probabilities $\Pr(\theta < 0.5|y)$ and we see large numbers of genes have probabilities close to 0 and 1, indicating allele specific expression (ASE).

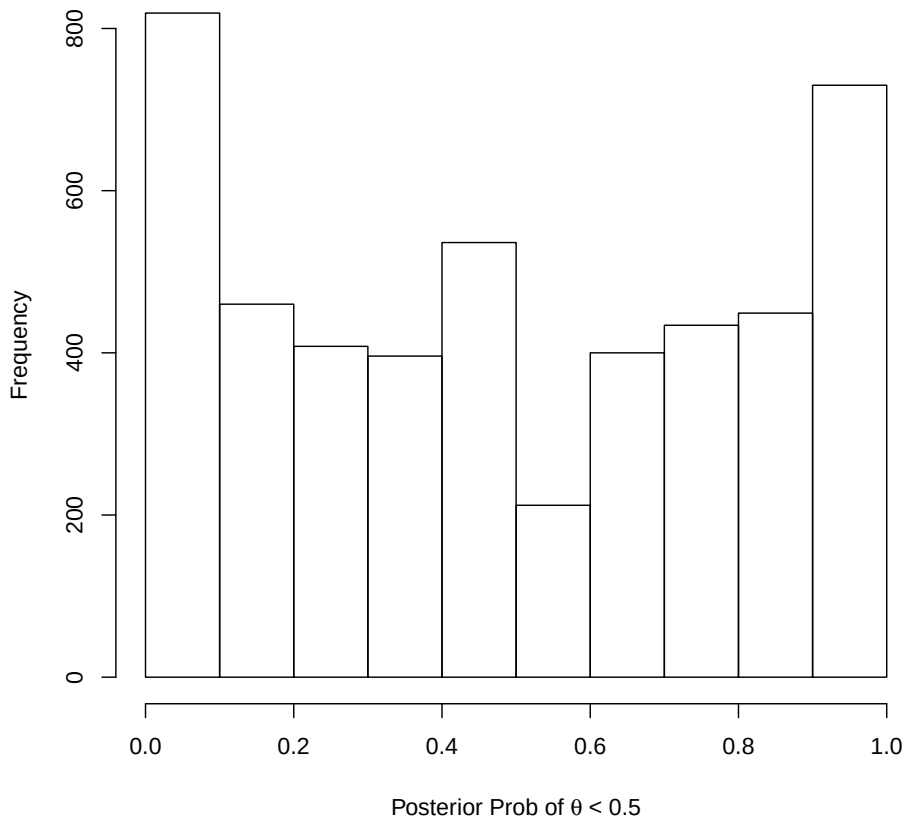

Figure 8: Histogram of 4,844 posterior probabilities of $\theta < 0.5$.

Bayesian Analysis of the ASE Data

- To the left we plot $\Pr(\theta < 0.5|y)$ versus the p-values and the general pattern is what we would expect — small p-values have posterior probabilities close to 0 and 1.
- The weird lines are due to discreteness of the data.

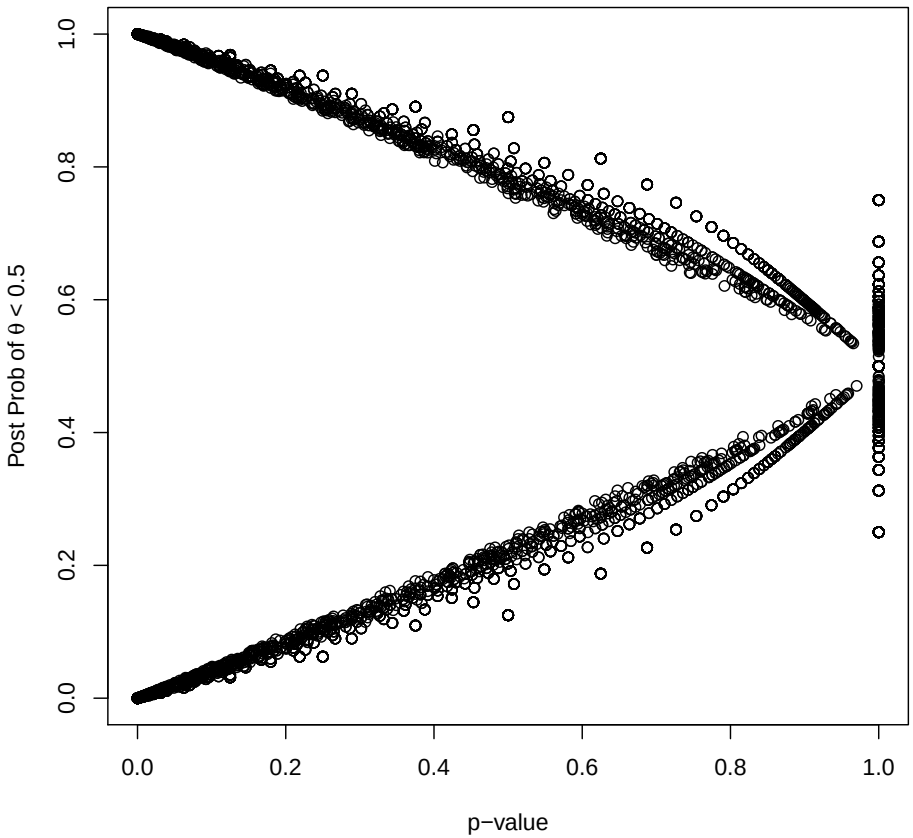

Figure 9: Posterior probabilities of $\theta < 0.5$ and *p*-values from exact tests.

Bayesian Analysis of the ASE Data

- Here we plot the -Log Bayes Factor against $\Pr(\theta < 0.5|y)$.
- Large values of the former correspond to strong evidence of ASE.
- Again we see an agreement in inference, with large values of the negative log Bayes factor corresponding with $\Pr(\theta < 0.5|y)$ close to 0 and 1.

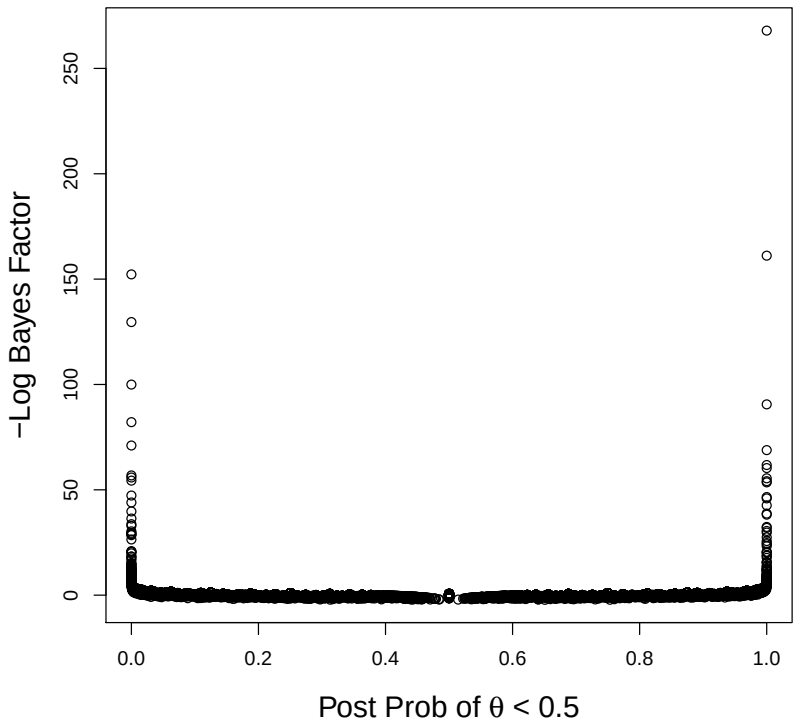

Figure 10: Negative Log Bayes factor versus posterior probabilities of $\theta < 0.5$.

ASE Example

Applying a Bonferroni correction to control the family wise error rate at 0.05, gives a p-value threshold of $0.05/4844 = 10^{-5}$ and 111 rejections. More on this later!

There were 278 genes with $\Pr(\theta < 0.5|y) < 0.01$ and 242 genes with $\Pr(\theta < 0.5|y) > 0.99$.

Following the guideline of requiring very strong evidence, there were 197 genes with the reciprocal Bayes factor greater than 150.

Requiring less stringent evidence, i.e. strong and very strong (reciprocal BF greater than 20), there were 359 genes.

We later consider a formal decision theory approach to testing.

In this example, the rankings of the different approaches are similar, but the calibration, i.e., picking a threshold, is not straightforward.

ASE Output Data

- Below are some summaries from the ASE analysis – we order with respect to the variable `logBFr`, which is the reciprocal Bayes factor (so that high numbers correspond to strong evidence against the null).
- The `postprob` variable is the posterior probability of $\theta < 0.5$.

```
allvals <- data.frame(Nsum,ysum,pvals,postprob,logBFr)
oBF <- order(-logBFr)
orderallvals <- allvals[oBF,]
head(orderallvals)
     Nsum ysum          pvals      postprob    logBFr
4751  437    6 5.340324e-119 1.000000e+00 267.9572
4041  625   97  1.112231e-72 1.000000e+00 161.1355
2370  546  468  8.994944e-69 2.621622e-69 152.2517
2770  256  245  1.127211e-58 2.943484e-59 129.6198
tail(orderallvals)
     Nsum ysum      pvals  postprob     logBFr
824   761  382 0.9422103 0.4567334 -2.086604
2163  776  390 0.9142477 0.4429539 -2.091955
3153  769  384 1.0000000 0.5143722 -2.097079
2860 1076  546 0.6474878 0.3129473 -2.146555
```

Summary

- Predictions are very natural under the Bayesian approach.
- Monte Carlo sampling provides flexibility of inference.
- All this lecture considered Binomial sampling, for which there is only a single parameter. For more parameters, prior specification and computing becomes more interesting...as we shall see.
- Multiple testing is considered in Lecture 9.
- For estimation and with middle to large sample sizes, conclusions from Bayesian and non-Bayesian approaches often coincide.
- For testing it's more complex, as discussed in Lecture 9.

Conclusions

Benefits of a Bayesian approach:

- Inference is based on probability and output is very intuitive.
- Framework is flexible, and so complex models can be built.
- Can incorporate prior knowledge!
- If the sample size is large, prior choice is less crucial.

Challenges of a Bayesian analysis:

- Require a likelihood and a prior, and inference is only as good as the appropriateness of these choices.
- Computation can be daunting, though software is becoming more user friendly and flexible; later we will describe and illustrate a number of approaches including `INLA` and `Stan`.
- One should be wary of model becoming too complex – we have the technology to contemplate complicated models, but do the data support complexity?

References

Kass, R. and Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795.

Savage, S. A., Gerstenblith, M. R., Goldstein, A. M., Mirabello, L., Fargnoli, M. C., Peris, K., and Landi, M. T. (2008). Nucleotide diversity and population differentiation of the melanocortin 1 receptor gene, mc1r. *BMC Genetics*, **9**, 31.