Quantitative Genomics and Genetics
BTRY 4830/6830; PBSB.5201.01

Lecture 21: (Brief) Introduction to Bayesian Inference

Jason Mezey
jgm45@cornell.edu
April 28, 2016 (Th) 8:40-9:55
Announcements

- I will have truncated office hours today: 3-4PM (!!)

Summary of lecture 21

- We will provide a (brief) introduction to Bayesian inference and how to we could apply such inference using the quantitative genetic model.
Introduction to Bayesian analysis

• Up to this point, we have considered statistical analysis (and inference) using a Frequentist formalism.

• There is an alternative formalism called Bayesian that we will now introduce in a very brief manner.

• Note that there is an important conceptual split between statisticians who consider themselves Frequentist of Bayesian but for GWAS analysis (and for most applications where we are concerned with analyzing data) we do not have a preference, i.e. we only care about getting the “right” biological answer so any (or both) frameworks that get us to this goal are useful.

• In GWAS (and mapping) analysis, you will see both frequentist (i.e. the framework we have built up to this point!) and Bayesian approaches applied.
• In both frequentist and Bayesian analyses, we have the same probabilistic framework (sample spaces, random variables, probability models, etc.) and when assuming our probability model falls in a family of parameterized distributions, we assume that a single fixed parameter value(s) describes the true model that produced our sample.

• However, in a Bayesian framework, we now allow the parameter to have its own probability distribution (we DO NOT do this in a frequentist analysis), such that we treat it as a random variable.

• This may seem strange - how can we consider a parameter to have a probability distribution if it is fixed?

• However, we can if we have some prior assumptions about what values the parameter value will take for our system compared to others and we can make this prior assumption rigorous by assuming there is a probability distribution associated with the parameter.

• It turns out, this assumption produces major differences between the two analysis procedures (in how they consider probability, how they perform inference, etc.).
Introduction to Bayesian analysis III

- To introduce Bayesian statistics, we need to begin by introducing Bayes' theorem.

- Consider a set of events (remember events!?) $A = A_1 \ldots A_k$ of a sample space $\Omega$ (where $k$ may be infinite), which form a partition of the sample space, i.e. $\bigcup_i^k A_i = \Omega$ and $A_i \cap A_j = \emptyset$ for all $i \neq j$.

- For another event $B \subset \Omega$ (which may be $\Omega$ itself) define the Law of total probability:

$$Pr(B) = \sum_{i=1}^{k} Pr(B \cap A_i) = \sum_{i=1}^{k} Pr(B|A_i)Pr(A_i)$$

- Now we can state Bayes' theorem:

$$Pr(A_i|B) = \frac{Pr(A_i \cap B)}{Pr(B)} = \frac{Pr(B|A_i)Pr(A_i)}{Pr(B)} = \frac{Pr(B|A_i)Pr(A)}{\sum_{i=1}^{k} Pr(B|A_i)Pr(A_i)}$$
Introduction to Bayesian analysis IV

- Remember that in a Bayesian (not frequentist!) framework, our parameter(s) have a probability distribution associated with them that reflects our belief in the values that might be the true value of the parameter.

- Since we are treating the parameter as a random variable, we can consider the joint distribution of the parameter AND a sample $\mathbf{y}$ produced under a probability model:

$$Pr(\theta \cap \mathbf{y})$$

- For inference, we are interested in the probability the parameter takes a certain value given a sample:

$$Pr(\theta | \mathbf{y})$$

- Using Bayes theorem, we can write:

$$Pr(\theta | \mathbf{y}) = \frac{Pr(\mathbf{y} | \theta)Pr(\theta)}{Pr(\mathbf{y})}$$

- Also note that since the sample is fixed (i.e. we are considering a single sample) $Pr(\mathbf{y}) = c$, we can rewrite this as follows:

$$Pr(\theta | \mathbf{y}) \propto Pr(\mathbf{y} | \theta)Pr(\theta)$$
Introduction to Bayesian analysis V

- Let's consider the structure of our main equation in Bayesian statistics:

\[ Pr(\theta|y) \propto Pr(y|\theta)Pr(\theta) \]

- Note that the left hand side is called the posterior probability:

\[ Pr(\theta|y) \]

- The first term of the right hand side is something we have seen before, i.e. the likelihood (!!):

\[ Pr(y|\theta) = L(\theta|y) \]

- The second term of the right hand side is new and is called the prior:

\[ Pr(\theta) \]

- Note that the prior is how we incorporate our assumptions concerning the values the true parameter value may take

- In a Bayesian framework, we are making two assumptions (unlike a frequentist where we make one assumption: 1. the probability distribution that generated the sample, 2. the probability distribution of the parameter
Probability in a Bayesian framework

- By allowing for the parameter to have an prior probability distribution, we produce a change in how we consider probability in a Bayesian versus Frequentist perspective.

- For example, consider a coin flip, with $\text{Bern}(p)$.

  - In a Frequentist framework, we consider a conception of probability that we use for inference to reflect the outcomes as if we flipped the coin an infinite number of times, i.e. if we flipped the coin 100 times and it was “heads” each time, we do not use this information to change how we consider a new experiment with this same coin if we flipped it again.

  - In a Bayesian framework, we consider a conception of probability can incorporate previous observations, i.e. if we flipped a coin 100 times and it was “heads” each time, we might want to incorporate this information in to our inferences from a new experiment with this same coin if we flipped it again.

- Note that this philosophic distinction is very deep (=we have only scratched the surface with this one example).
Debating the Frequentist versus Bayesian frameworks

- Frequentists often argue that because they “do not” take previous experience into account when performing their inference concerning the value of a parameter, such that they do not introduce biases into their inference framework.

- In response, Bayesians often argue:
  - Previous experience is used to specify the probability model in the first place.
  - By not incorporating previous experience in the inference procedure, prior assumptions are still being used (which can introduce logical inconsistencies!)
  - The idea of considering an infinite number of observations is not particularly realistic (and can be a non-sensical abstraction for the real world)
  - The impact of prior assumptions in Bayesian inference disappear as the sample size goes to infinite

- Again, note that we have only scratched the surface of this debate!
Types of priors in Bayesian analysis

- Up to this point, we have discussed priors in an abstract manner.
- To start making this concept more clear, let's consider one of our original examples where we are interested in the knowing the mean human height in the US (what are the components of the statistical framework for this example!? Note the basic components are the same in Frequentist / Bayesian!)
- If we assume a normal probability model of human height (what parameter are we interested in inferring in this case and why?) in a Bayesian framework, we will at least need to define a prior:
  \[ Pr(\mu) \]
- One possible approach is to make the probability of each possible value of the parameter the same (what distribution are we assuming and what is a problem with this approach), which defines an improper prior:
  \[ Pr(\mu) = c \]
- Another possible approach is to incorporate our previous observations that heights are seldom infinite, etc. where one choice for incorporating this observations is my defining a prior that has the same distribution as our probability model, which defines a conjugate prior (which is also a proper prior):
  \[ Pr(\mu) \sim N(\kappa, \phi^2) \]
Constructing the posterior probability

- Let’s put this all together for our “heights in the US” example
- First recall that our assumption is the probability model is normal (so what is the form of the likelihood?):

\[ Y \sim N(\mu, \sigma^2) \]

- Second, assume a normal prior for the parameter we are interested in:

\[ Pr(\mu) \sim N(\kappa, \phi^2) \]

- From the Bayesian equation, we can now put this together as follows:

\[ Pr(\theta|y) \propto Pr(y|\theta)Pr(\theta) \]

\[ Pr(\mu|y) \propto \left( \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \mu)^2}{2\sigma^2}} \right) \frac{1}{\sqrt{2\pi\phi^2}} e^{-\frac{(\mu - \kappa)^2}{2\phi^2}} \]

- Note that with a little rearrangement, this can be written in the following form:

\[ Pr(\mu|y) \sim N\left( \frac{\frac{\kappa}{\sigma^2} + \frac{\sum_{i} y_i}{\sigma^2}}{\frac{1}{\phi^2} + \frac{n}{\sigma^2}}, \left( \frac{1}{\phi^2} + \frac{n}{\sigma^2} \right)^{-1} \right) \]
Bayesian inference: estimation

- Inference in a Bayesian framework differs from a frequentist framework in both estimation and hypothesis testing.

- For example, for estimation in a Bayesian framework, we always construct estimators using the posterior probability distribution, for example:

  \[ \hat{\theta} = \text{mean}(\theta|y) = \int \theta \Pr(\theta|y) d\theta \quad \text{or} \quad \hat{\theta} = \text{median}(\theta|y) \]

- For example, in our “heights in the US” example our estimator is:

  \[ \hat{\mu} = \text{median}(\mu|y) = \text{mean}(\mu|y) = \frac{\left( \frac{\kappa}{\sigma^2} + \frac{n\bar{y}}{\sigma^2} \right)}{\left( \frac{1}{\phi^2} + \frac{n}{\sigma^2} \right)} \]

- Note 1: again notice that the impact of the prior disappears as the sample size goes to infinite (=same as MLE under this condition):

  \[ \frac{\left( \frac{\kappa}{\sigma^2} + \frac{n\bar{y}}{\sigma^2} \right)}{\left( \frac{1}{\phi^2} + \frac{n}{\sigma^2} \right)} \approx \frac{\left( \frac{n\bar{y}}{\sigma^2} \right)}{\left( \frac{n}{\sigma^2} \right)} \approx \bar{y} \]

- Note 2: estimates in a Bayesian framework can be different than in a likelihood (Frequentist) framework since estimator construction is fundamentally different (!!)
Bayesian inference: hypothesis testing

- For hypothesis testing in a Bayesian analysis, we use the same null and alternative hypothesis framework:
  \[ H_0 : \theta \in \Theta_0 \]
  \[ H_A : \theta \in \Theta_A \]

- However, the approach to hypothesis testing is completely different than in a frequentist framework, where we use a Bayes factor to indicate the relative support for one hypothesis versus the other:
  \[ Bayes = \frac{\int_{\theta \in \Theta_0} Pr(y|\theta)Pr(\theta)d\theta}{\int_{\theta \in \Theta_A} Pr(y|\theta)Pr(\theta)d\theta} \]

- Note that a downside to using a Bayes factor to assess hypotheses is that it can be difficult to assign priors for hypotheses that have completely different ranges of support (e.g. the null is a point and alternative is a range of values)

- As a consequence, people often use an alternative “psuedo-Bayesian” approach to hypothesis testing that makes use of credible intervals (which is what we will use in this course)
Bayesian credible intervals (versus frequentist confidence intervals)

- Recall that in a Frequentist framework that we can estimate a confidence interval at some level (say 0.95), which is an interval that will include the value of the parameter 0.95 of the times we performed the experiment an infinite number of times, calculating the confidence interval each time (note: a strange definition...)

- In a Bayesian interval, the parallel concept is a credible interval that has a completely different interpretation: this interval has a given probability of including the parameter value (!!!)

- The definition of a credible interval is as follows:

\[
c.i.(\theta) = \int_{-c_\alpha}^{c_\alpha} Pr(\theta|y)d\theta = 1 - \alpha
\]

- Note that we can assess a null hypothesis using a credible interval by determining if this interval includes the value of the parameter under the null hypothesis (!!!)
Bayesian inference: genetic model 1

- We are now ready to tackle Bayesian inference for our genetic model (note that we will focus on the linear regression model but we can perform Bayesian inference for any GLM!):

\[ Y = \beta_\mu + X_a \beta_a + X_d \beta_d + \epsilon \]

\( \epsilon \sim N(0, \sigma_\epsilon^2) \)

- Recall for a sample generated under this model, we can write:

\[ y = x\beta + \epsilon \]

\( \epsilon \sim multiN(0, I\sigma_\epsilon^2) \)

- In this case, we are interested in the following hypotheses:

\[ H_0 : \beta_a = 0 \cap \beta_d = 0 \]

\( H_A : \beta_a \neq 0 \cup \beta_d \neq 0 \)

- We are therefore interested in the marginal posterior probability of these two parameters
Bayesian inference: genetic model II

• To calculate these probabilities, we need to assign a joint probability distribution for the prior

\[ Pr(\beta_\mu, \beta_a, \beta_d, \sigma^2_\epsilon) \]

• One possible choice is as follows (are these proper or improper!?):

\[ Pr(\beta_\mu, \beta_a, \beta_d, \sigma^2_\epsilon) = Pr(\beta_\mu)Pr(\beta_a)Pr(\beta_d)Pr(\sigma^2_\epsilon) \]

\[ Pr(\beta_\mu) = Pr(\beta_a) = Pr(\beta_d) = c \]

\[ Pr(\sigma^2_\epsilon) = c \]

• Under this prior the complete posterior distribution is multivariate normal (!!):

\[ Pr(\beta_\mu, \beta_a, \beta_d, \sigma^2_\epsilon | \mathbf{y}) \propto Pr(\mathbf{y}|\beta_\mu, \beta_a, \beta_d, \sigma^2_\epsilon) \]

\[ Pr(\theta|\mathbf{y}) \propto (\sigma^2_\epsilon)^{-\frac{n}{2}} e^{-\frac{(\mathbf{y}-\mathbf{x}\beta)^T(\mathbf{y}-\mathbf{x}\beta)}{2\sigma^2_\epsilon}} \]
Bayesian inference: genetic model III

- For the linear model with sample:

\[ y = x\beta + \epsilon \]
\[ \epsilon \sim multiN(0, I\sigma^2_\epsilon) \]

- The complete posterior probability for the genetic model is:

\[ Pr(\beta_\mu, \beta_a, \beta_d, \sigma^2_\epsilon | y) \propto Pr(y | \beta_\mu, \beta_a, \beta_d, \sigma^2_\epsilon) Pr(\beta_\mu, \beta_a, \beta_d, \sigma^2_\epsilon) \]

- With a uniform prior is:

\[ Pr(\beta_\mu, \beta_a, \beta_d, \sigma^2_\epsilon | y) \propto Pr(y | \beta_\mu, \beta_a, \beta_d, \sigma^2_\epsilon) \]

- The marginal posterior probability of the parameters we are interested in is:

\[ Pr(\beta_a, \beta_d | y) = \int_0^\infty \int_{-\infty}^\infty Pr(\beta_\mu, \beta_a, \beta_d, \sigma^2_\epsilon | y) d\beta_\mu d\sigma^2_\epsilon \]
Bayesian inference: genetic model IV

- Assuming uniform (improper!) priors, the marginal distribution is:

\[ Pr(\beta_a, \beta_d|y) = \int_{-\infty}^{\infty} \int_{0}^{\infty} Pr(\beta_\mu, \beta_a, \beta_d, \sigma_\epsilon^2|y) d\beta_\mu d\sigma_\epsilon^2 \sim \text{multi-t-distribution} \]

- With the following parameter values:

\[
\text{mean}(Pr(\beta_a, \beta_d|y)) = \left[ \hat{\beta}_a, \hat{\beta}_d \right]^T = C^{-1} \left[ X_a, X_d \right]^T y \\
C = \begin{bmatrix} X_a^T X_a & X_a^T X_d \\ X_d^T X_a & X_d^T X_d \end{bmatrix} \\
\text{cov} = \frac{(y - \left[ X_a, X_d \right] \left[ \hat{\beta}_a, \hat{\beta}_d \right]^T)^T(y - \left[ X_a, X_d \right] \left[ \hat{\beta}_a, \hat{\beta}_d \right]^T)}{n - 6} C^{-1} \\
df(\text{multi-t}) = n - 4
\]

- With these estimates (equations) we can now construct a credible interval for our genetic null hypothesis and test a marker for a phenotype association and we can perform a GWAS by doing this for each marker (!!)
Bayesian inference: genetic model V

\[ \Pr(\beta_a, \beta_d|y) \]

0.95 credible interval

Cannot reject H0!

\[ \Pr(\beta_a, \beta_d|y) \]

0.95 credible interval

Reject H0!
That’s it for today

• Next lecture: we will continue our brief introduction to Bayesian statistics