

# 2016 SISG Module 16: Bayesian Statistics for Genetics

## Lecture 5: Multinomial Sampling

Jon Wakefield

Departments of Statistics and Biostatistics  
University of Washington

# Outline

Introduction and Motivating Examples

Bayesian Analysis of Multinomial Data

Derivation of the Posterior and Prior Specification

Inference for Parameters of Interest

Analysis of HWE Data

Appendix: BF Detail

## Introduction

- In this lecture we will consider the Bayesian modeling of multinomial data.
- The examination of **Hardy-Weinberg equilibrium** will be used to motivate the multinomial model.
- Again, **conjugate priors** will be used, though sampling from the posterior will be emphasized as a method for flexible inference.
- **Bayes factors** will be used as a measure of evidence for hypothesis testing.

## Motivating Example: Testing for HWE

- For simplicity we consider a diallelic marker, and suppose we obtain a random sample of genotypes for  $n$  individuals.
- The form of the data is

	Genotype			Total
	$A_1A_1$	$A_1A_2$	$A_2A_2$	
Count	$n_1$	$n_2$	$n_3$	$n$
Population Frequency	$q_1$	$q_2$	$q_3$	1

- So the model contains 3 probabilities (which sum to 1)  $q_1, q_2, q_3$ ; hence, there are 2 **free** parameters.
- Suppose the proportions of alleles  $A_1$  and  $A_2$  in a given generation are  $p_1$  and  $p_2 = 1 - p_1$ .
- In terms of  $q_1, q_2, q_3$ :

$$p_1 = q_1 + \frac{q_2}{2}$$

$$p_2 = \frac{q_2}{2} + q_3$$

## Motivating Example: Testing for HWE

- HWE is the **statistical independence** of an individual's alleles at a locus.
- Under HWE, the probability distribution for the genotype of an individual in the next generation is:

	Genotype			
	$A_1A_1$	$A_1A_2$	$A_2A_2$	
Proportion	$p_1^2$	$2p_1p_2$	$p_2^2$	1

- Reasons for **deviation** from HWE include: small population size, selection, inbreeding and population structure.

## A Toy Example

### Example:

- Consider the data  $n_1 = 88$ ,  $n_2 = 10$ ,  $n_3 = 2$ .
- Are these frequencies consistent with HWE?
- The MLEs are:

$$\begin{aligned}\hat{q}_1 &= 0.88 & \hat{q}_2 &= 0.10 & \hat{q}_3 &= 0.02 \\ \hat{p}_1 &= 0.93 & \hat{p}_2 &= 0.07\end{aligned}$$

- For these data the exact  $p$ -value for

$$H_0 : q_1 = p_1^2, \quad q_2 = 2p_1p_2, \quad q_3 = p_2^2$$

is 0.0654.

## Critique of Non-Bayesian Approach

- Testing for HWE is carried out via (**asymptotic**, i.e., large sample)  $\chi^2$  tests or **exact** tests.
- $\chi^2$  tests require very large sample sizes for accurate  $p$ -values.
- The exact test can be **computationally expensive** to perform, when there are many alleles/samples.
- Under the null of HWE, the **discreteness** of the test statistic causes difficulties.
- In general, how to decide on a **significance level**? The level should be a function of sample size (and in particular should decrease as sample size increases), but how should it be chosen?
- Estimation depends on **asymptotic approximations** (i.e., large sample sizes).
- Estimation also difficult due to **awkward constraints** on parameters (particularly with many alleles).

## Bayes Theorem

	Genotype			Total
	$A_1A_1$	$A_1A_2$	$A_2A_2$	
Count	$n_1$	$n_2$	$n_3$	$n$
Population Frequency	$q_1$	$q_2$	$q_3$	1

- The multinomial with three counts is known as the trinomial distribution.
- We have three parameters,  $q_1, q_2, q_3$ , but they sum to 1, so that effectively we have two parameters.
- We write  $\mathbf{q} = (q_1, q_2, q_3)$  to represent the vector of probabilities, and  $\mathbf{n} = (n_1, n_2, n_3)$  for the data vector.
- Via **Bayes Theorem**:

$$p(\mathbf{q}|\mathbf{n}) = \frac{\Pr(\mathbf{n}|\mathbf{q}) \times p(\mathbf{q})}{\Pr(\mathbf{n})}$$

Posterior     $\propto$     Likelihood  $\times$  Prior



## Elements of Bayes Theorem: The Likelihood

- We assume  $n$  independent draws with common probabilities  $\mathbf{q} = (q_1, q_2, q_3)$ .
- In this case, the distribution of  $n_1, n_2, n_3$  is **multinomial**:

$$\Pr(n_1, n_2, n_3 | q_1, q_2, q_3) = \frac{n!}{n_1! n_2! n_3!} q_1^{n_1} q_2^{n_2} q_3^{n_3}. \quad (1)$$

- For fixed  $\mathbf{n}$ , we may view (1) as a function of  $\mathbf{q}$  – this is the **likelihood function**.
- The **maximum likelihood estimate** (MLE) is

$$\hat{\mathbf{q}} = \left( \frac{n_1}{n}, \frac{n_2}{n}, \frac{n_3}{n} \right).$$

- The MLE gives the highest probability to the observed data, i.e. maximizes the likelihood function.

## The Dirichlet Distribution as a Prior Choice for a Multinomial $\mathbf{q}$

- Once the likelihood is specified we need to think about the **prior** distribution.
- We require a prior distribution over  $(q_1, q_2, q_3)$  — not straightforward since the three probabilities all lie in  $[0,1]$ , and must sum to 1.
- A distribution that satisfies these requirements is the **dirichlet** distribution, denoted **dirichlet** $(v_1, v_2, v_3)$  and has density:

$$p(q_1, q_2, q_3) = \frac{\Gamma(v_1 + v_2 + v_3)}{\Gamma(v_1)\Gamma(v_2)\Gamma(v_3)} \times q_1^{v_1-1} q_2^{v_2-1} q_3^{v_3-1}$$

$$\propto q_1^{v_1-1} q_2^{v_2-1} q_3^{v_3-1}$$

where  $\Gamma(\cdot)$  denotes the gamma function.

## The Dirichlet Distribution as a Prior Choice for a Multinomial $\mathbf{q}$

- The **dirichlet**( $v_1, v_2, v_3$ ) prior:

$$p(q_1, q_2, q_3) = \frac{\Gamma(v_1 + v_2 + v_3)}{\Gamma(v_1)\Gamma(v_2)\Gamma(v_3)} \times q_1^{v_1-1} q_2^{v_2-1} q_3^{v_3-1}$$

$$\propto q_1^{v_1-1} q_2^{v_2-1} q_3^{v_3-1}.$$

- $v_1, v_2, v_3 > 0$  are specified to reflect prior beliefs about  $(q_1, q_2, q_3)$ .
- The dirichlet distribution can be used with general multinomial distributions (i.e. for  $k = 2, 3, \dots$  categories).
- The beta distribution is a special case of the dirichlet when there are two categories only.

## Dirichlet Prior

- The **mean** and **variance** are

$$E[q_i] = \frac{v_i}{v_1 + v_2 + v_3} = \frac{v_i}{v}$$

$$\text{var}(q_i) = \frac{E[q_i](1 - E[q_i])}{v_1 + v_2 + v_3 + 1} = \frac{E[q_i](1 - E[q_i])}{v + 1}$$

for  $i = 1, 2, 3$ , where  $v = v_1 + v_2 + v_3$ .

- Large values of  $v$  increase the influence of the prior.
- The dirichlet has a single parameter only ( $v$ ) to control the **spread** for all of the dimensions, which is a deficiency.
- The quartiles may be empirically calculated from samples.

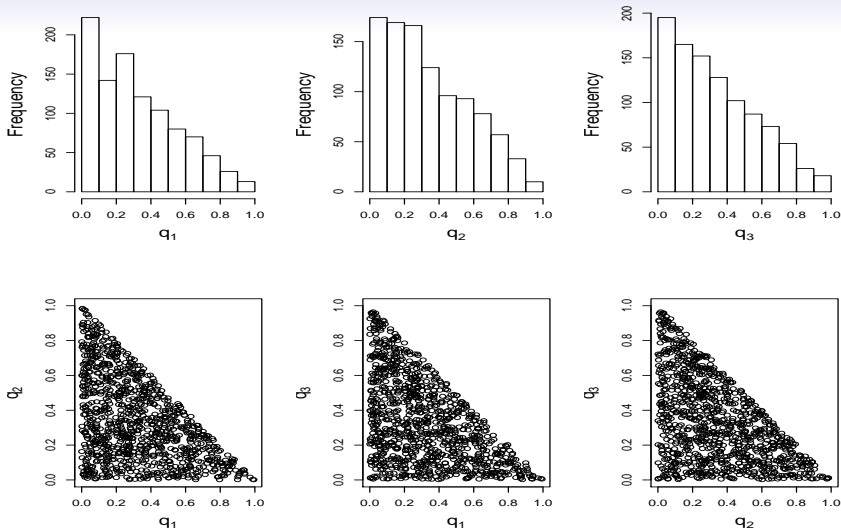


Figure 1 : Samples from a  $\text{dirichlet}(1, 1, 1)$  distribution. The mean is  $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ .

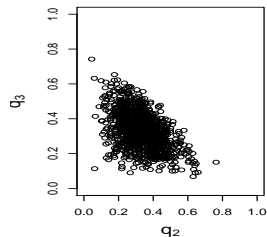
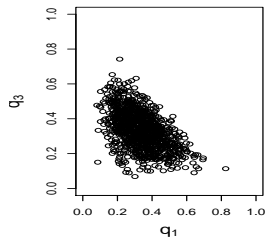
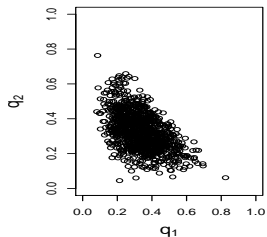
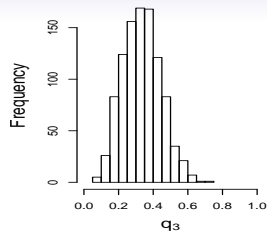
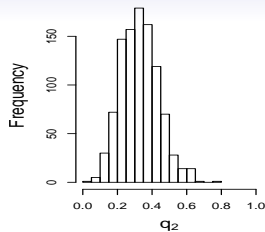
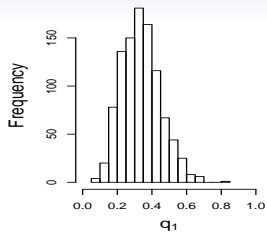


Figure 2 : Samples from a  $\text{dirichlet}(6, 6, 6)$  distribution. The mean is  $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ .

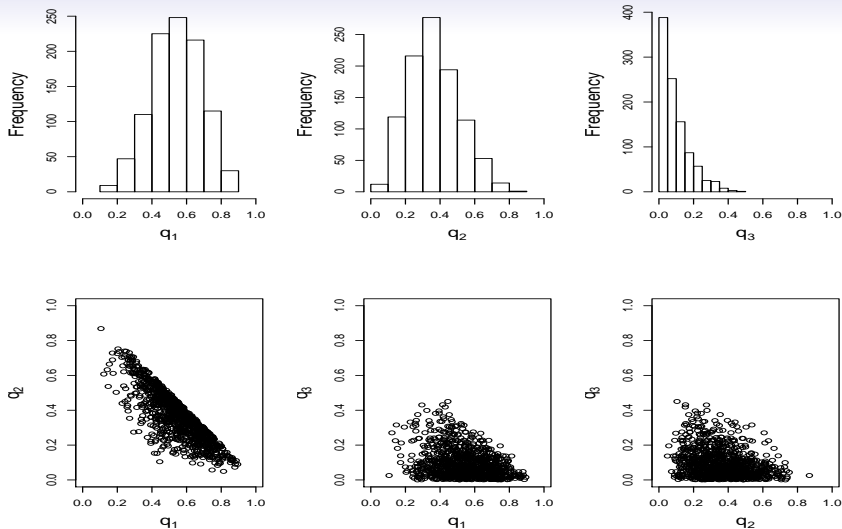


Figure 3 : Samples from a **dirichlet(6, 4, 1)** distribution. The mean is  $(\frac{6}{11}, \frac{4}{11}, \frac{1}{11}) = (0.55, 0.36, 0.09)$ .

## Posterior Distribution

- Combining the Dirichlet prior,  $\text{dirichlet}(v_1, v_2, v_3)$ , with the multinomial likelihood gives the posterior:

$$\begin{aligned} p(q_1, q_2, q_3 | \mathbf{n}) &\propto \text{Pr}(\mathbf{n} | \mathbf{q}) \times p(\mathbf{q}) \\ &\propto q_1^{n_1} q_2^{n_2} q_3^{n_3} \times q_1^{v_1-1} q_2^{v_2-1} q_3^{v_3-1} \\ &= q_1^{n_1+v_1-1} q_2^{n_2+v_2-1} q_3^{n_3+v_3-1}. \end{aligned}$$

- This distribution is another Dirichlet:

$$\text{dirichlet}(n_1 + v_1, n_2 + v_2, n_3 + v_3).$$

- Notice: “as if” we had observed counts  $(n_1 + v_1, n_2 + v_2, n_3 + v_3)$ .



## Choosing a Prior

- The **posterior mean** for the expected proportion of counts in cell  $i$  is, for  $i = 1, 2, 3$ :

$$\begin{aligned} E[q_i | \mathbf{n}] &= \frac{n_i + v_i}{n + v} \\ &= \frac{n_i}{n} \frac{n}{n + v} + \frac{v_i}{v} \frac{v}{n + v} \\ &= \text{MLE} \times W + \text{Prior Mean} \times (1 - W) \end{aligned}$$

where  $n = n_1 + n_2 + n_3$ ,  $v = v_1 + v_2 + v_3$ .

- The **weight**  $W$  is

$$W = \frac{n}{n + v}$$

which is the proportion of the **total information** ( $n + v$ ) that is contributed by the data ( $n$ ).

## Choosing a Prior

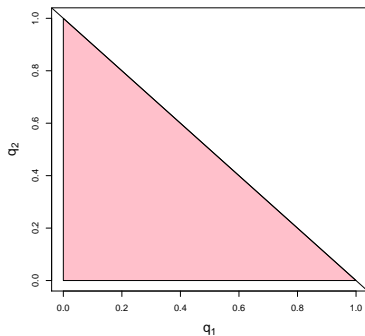
- Recall the prior mean is

$$\left( \frac{v_1}{v}, \frac{v_2}{v}, \frac{v_3}{v} \right)$$

- These forms help to choose  $v_1, v_2, v_3$ .
- As with the beta distribution we may specify the prior means, and the relative weight that the prior and data contribute:  $n$  and  $v$  are on a comparable scale.
- For example, suppose we believe that event 1 is four times as likely as each of event 2 or event 3.
- Then we may specify the means in the ratios 4:1:1.
- Suppose  $n = 24$  and we wish to allow the prior contribution to be a half of this total (and therefore a third of the complete information). Then the **prior sample size** is  $v = 12$  and the prior mean requirement gives

$$v_1 = 8, v_2 = 2, v_3 = 2.$$

## A Uniform Prior



An obvious choice of parameters is  $v_1 = v_2 = v_3 = 1$  to give a prior that is uniform over the simplex:

$$\pi(q_1, q_2, q_3) = 2$$

for

$$0 < q_1, q_2, q_3 < 1, \quad q_1 + q_2 + q_3 = 1$$

Note: not uniform over all parameter of interests, as we see shortly

## Simple HWE Example

- The data is

$$n_1 = 88, n_2 = 10, n_3 = 2.$$

- We assume a flat dirichlet prior on the allowable values of  $\mathbf{q}$ :

$$v_1 = v_2 = v_3 = 1.$$

- This gives the posterior as **dirichlet(88 + 1, 10 + 1, 2 + 1)** with posterior means:

$$\begin{aligned} E[q_1|\mathbf{n}] &= \frac{1 + 88}{3 + 100} = \frac{89}{103} \\ E[q_2|\mathbf{n}] &= \frac{1 + 10}{3 + 100} = \frac{11}{103} \\ E[q_3|\mathbf{n}] &= \frac{1 + 2}{3 + 100} = \frac{3}{103}. \end{aligned}$$

- Note the similarity to the MLEs of

$$\left( \frac{88}{100}, \frac{10}{100}, \frac{2}{100} \right).$$

## Simple HWE Example

- We continue with this example and now examine posterior distributions.
- We generate samples from

`dirichlet(88 + 1, 10 + 1, 2 + 1).`

- As posterior summaries we display, in Figure 4:
  - Histograms of the 3 univariate marginal distributions  $p(q_1|\mathbf{y})$ ,  $p(q_2|\mathbf{y})$ ,  $p(q_3|\mathbf{y})$ .
  - Scatterplots of the 3 bivariate marginal distributions  $p(q_1, q_2|\mathbf{y})$ ,  $p(q_1, q_3|\mathbf{y})$ ,  $p(q_2, q_3|\mathbf{y})$ .
- On each plot we indicate the MLEs for the general model, i.e. the non-HWE model (in red) and under the assumption of HWE (in blue).

## Samples from the Posterior

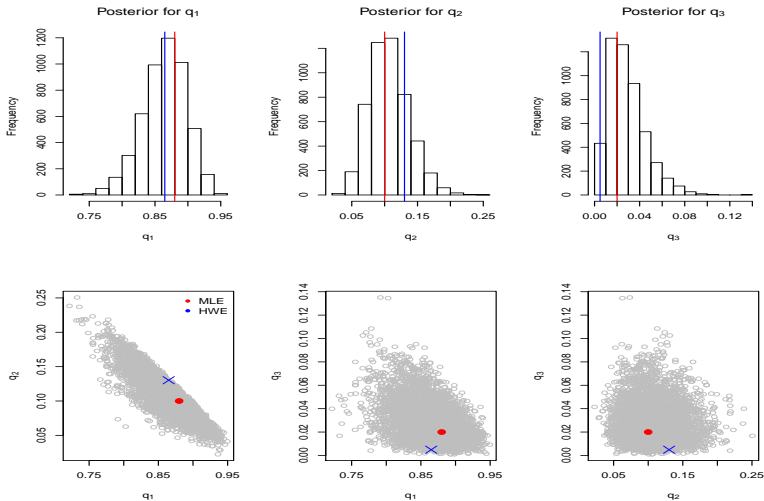


Figure 4 : Univariate and bivariate posterior distributions for  $\mathbf{n} = (88, 10, 2)$ . MLEs in red for the general model and in blue for the HWE model.

## Bayes analysis of (88,10,2) data

- As expected with a sample size of  $n = 100$  and a flat prior, the MLEs lie close to the center of the posteriors.
- Note the asymmetry of the posteriors.
- Asymptotic confidence intervals of the form  $\hat{q}_i \pm 1.96 \times \text{se}(\hat{q}_i)$  would be symmetric.

## Bayes analysis of (88,10,2) data

- In the context of a binomial sampling model and interest in a particular point (for example,  $\theta = 0.5$ ) we could examine intervals for  $\theta$ .
- In a multinomial context the situation is more complex; shortly we will examine **Bayes factors** to carry out hypothesis testing.



## Parameters of Interest

	Genotype			Total
	$A_1A_1$	$A_1A_2$	$A_2A_2$	
Population Frequency	$q_1$	$q_2$	$q_3$	1

- Rather than  $q_1, q_2, q_3$ , we may be interested in other parameters of interest.
- In the HWE context: Let  $X_1$  and  $X_2$  be 0/1 indicators of  $A_1$  for the two alleles at a locus.
- The covariance between  $X_1$  and  $X_2$  is the **disequilibrium coefficient**:

$$D = q_1 - p_1^2$$

Under HWE  $q_1 = p_1^2$ , and the covariance is zero.

- Another quantity of interest (Shoemaker *et al.*, 1998) is

$$\psi = \frac{q_2^2}{q_1 q_3}.$$

Under HWE,  $\psi = 4$ .

## Parameters of Interest

- The **inbreeding coefficient** is

$$f = \frac{q_1 - p_1^2}{p_1 p_2}$$

- The variance of  $X_1$  and  $X_2$  is  $p_1(1 - p_1) = p_1 p_2$  and so  $f$  is the correlation.
- We may express  $q_1, q_2, q_3$  as

$$q_1 = p_1^2 + p_1(1 - p_1)f$$

$$q_2 = 2p_1(1 - p_1)(1 - f)$$

$$q_3 = (1 - p_1)^2 + p_1(1 - p_1)f$$

- Positive** values of  $f$  indicate an excess of homozygotes (and may indicate inbreeding), while **negative** values indicate an excess of heterozygotes.

## Parameters of Interest

- Each of  $D$ ,  $\psi$  and  $f$  are complex functions of  $q_1, q_2, q_3$  and given a Dirichlet prior for the latter do not have known posterior forms.
- The “flat” prior for  $\mathbf{q}$ ,  $\text{dirichlet}(1, 1, 1)$ , does not correspond to a flat prior for  $D, f, \psi$ , as Figure 5 shows.

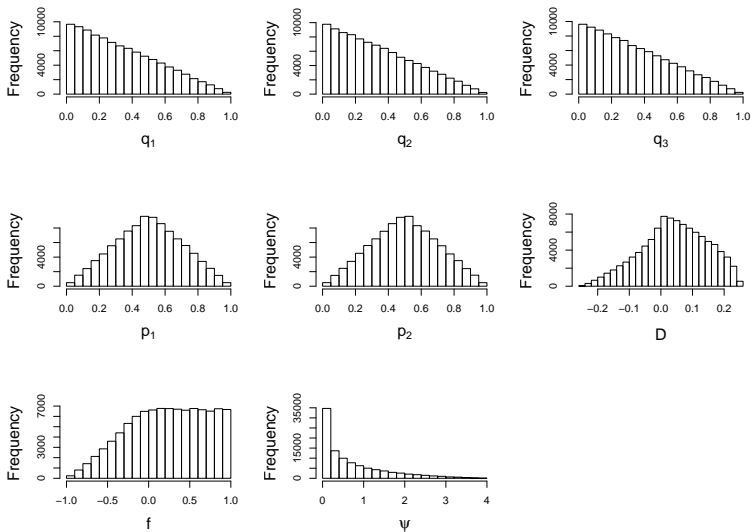


Figure 5 : Samples from a  $\text{dirichlet}(1,1,1)$  for various functions.

## Implied Prior on Functions of Interest

- **Very important point:** As we saw in the binomial development, you can't be "flat" on every scale, but the implications of particular priors can be assessed via simulation.
- We emphasize that we are not uniform on the marginal distributions for  $q_i$ , since these follow

$$\text{beta}(v_i, v - v_i)$$

distributions, i.e.  $\text{beta}(1,2)$  if  $v_1 = v_2 = v_3 = 1$ .

- The priors on the measures of distance from HWE are far from uniform.
- For example, with a "flat" Dirichlet prior  $\text{dirichlet}(1,1,1)$  the prior probability that  $f > 0$  is 0.67.

## Bayesian Approaches to HWE

Previous approaches include:

- Altham (1971).
- Pereira and Rogatko (1984).
- Lindley (1988).
- Shoemaker *et al.* (1998).
- Montoya-Delgado *et al.* (2001).
- Consonni *et al.* (2008).
- Wakefield (2010).

## Bayes factors for HWE

- Recall that Bayes factors measure the evidence in a sample for one hypothesis, as compared to an alternative.
- We derive the Bayes factor for multinomial data in the context of testing for HWE.
- We wish to test

$H_0$  : HWE versus  $H_1$  : Not HWE.

- We need to specify priors on the null and alternatives, and then calculate the **Bayes factor**:

$$\frac{\Pr(\mathbf{n}|H_0)}{\Pr(\mathbf{n}|H_1)}$$

where  $p_1$  and  $(q_1, q_2)$  are the parameters under the null and alternative, respectively.

## The HWE Bayes Factor

- The **Bayes factor**, measuring the evidence in the data for the null, as compared to the alternative is:

$$\text{BF} = \frac{2^{n_2} \Gamma(w) \Gamma(2n_1 + n_2 + w_1) \Gamma(v_1) \Gamma(v_2) \Gamma(v_3) \Gamma(n_2 + 2n_3 + w_2) \Gamma(n + v)}{\Gamma(w_1) \Gamma(w_2) \Gamma(2n + w) \Gamma(v) \Gamma(n_1 + v_1) \Gamma(n_2 + v_2) \Gamma(n_3 + v_3)}$$

- This appears complex, but is just a function of the **observed data**, and the **prior inputs**, and can be easily evaluated<sup>1</sup>.
- If  $\text{BF} > 1 (< 1)$  the data are **more (less)** likely to have come from the null.
- Can be readily extended to  $k > 2$  alleles.
- We next consider a formal decision rule.

---

<sup>1</sup>When we work out a  $\chi^2$  tail area we don't worry about the form of the distribution we just use the relevant function in our favorite software



## Bayesian Decision Theory

- Decision as to reject  $H_0$  in favor of  $H_1$  depends on the **costs** of making the two types of error:

		Decision	
		Report $H_0$	Report $H_1$
Truth	$H_0$	0	$C_I$
	$H_1$	$C_{II}$	0

- Costs of making the two types of error  $C_I$  is the cost of a type I error and  $C_{II}$  the cost of a type II error.
- The decision theory solution is to **report  $H_1$**  if:

$$\text{Posterior Odds of } H_0 = \text{BF} \times \text{Prior Odds} < \frac{C_{II}}{C_I} = R$$

so that we only need to consider the ratio of costs  $R$ .

- If  $\frac{C_{II}}{C_I} = 4$  (type II errors four times as bad as type I errors) then report  $H_1$  if

$$\text{Posterior Odds of } H_0 < 4,$$

i.e. if

$$\Pr(H_1 | \text{data}) > 0.2.$$

## A Simple Example

- We again consider the data  $n_1 = 88, n_2 = 10, n_3 = 2$ .
- These data give a **p-value** of 0.0654.
- With “flat” conjugate Dirichlet priors ( $w_1 = w_2 = v_1 = v_2 = v_3 = 1$ ) we obtain a **Bayes factor** of 1.54 so that the data are 50% more likely under the null than the alternative, so the evidence in favor of  $H_0$  is not strong.
- With a prior probability of the null  $\pi_0$ , to give a prior odds of  $\pi_0/(1 - \pi_0)$ , we have

$$\text{Posterior Odds of } H_0 = \text{BF} \times \frac{\pi_0}{1 - \pi_0}$$

- Hence, with  $\pi_0 = 0.5$  the posterior odds equal the Bayes factor, i.e. 1.54.
- The posterior probability of the null is

$$\frac{1.54}{1 + 1.54} = 0.61.$$

- This probability is very sensitive to the prior on the null. For example, with  $\pi_0 = 2/3$  we obtain a posterior odds of  $1.54 \times 2 = 3.08$  to give a posterior probability on the null of

$$\frac{3.08}{1 + 3.08} = 0.75.$$

## Posterior Inference for Functions of Interest

- Figure 6 shows functions of interest for the  $(88, 10, 2)$  example.
- Note that the asymptotic confidence interval for  $f$  is  $(-0.1032, 0.5632)$ .

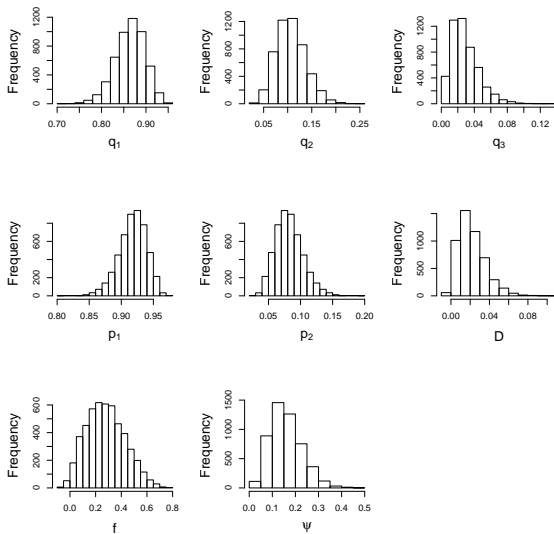
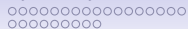


Figure 6 : Posterior inference for simple HWE example.

## The HWEBayes Package

- The R package **HWEBayes** implements the rejection algorithm and importance sampling (a numerical integration technique), for **testing and estimation** in the HWE context:

<http://cran.r-project.org/web/packages/HWEBayes/index.html>

- The *vignette* contains a worked example.
- Code for a four-allele example is here:

<http://faculty.washington.edu/jonno/HWEBayesFourAllele.R>

- More details of the methodology: Wakefield (2010).

## Conclusions

- The dirichlet distribution is convenient but quite inflexible as a prior distribution.
- Alternative priors are more difficult to specify since they are on scales that are more difficult to interpret (e.g. the logistic-normal distribution).
- Bayes factors are **sensitive** to the prior.
- **Monte Carlo sampling** is a powerful tool for inference.
- For **multiple alleles** computation is slow whether the approach is frequentist or Bayesian.

## References

- Altham, P. (1971). Exact Bayesian analysis of an intraclass  $2 \times 2$  table. *Biometrika*, **58**, 679–680.
- Consonni, G., Gutierrez-Pena, E., and Veronese, P. (2008). Compatible priors for Bayesian model comparison with an application to the Hardy-Weinberg equilibrium model. *Test*, **17**, 585–605.
- Lindley, D. (1988). Statistical inference concerning Hardy-Weinberg equilibrium. In J. Bernardo, M. DeGroot, D. Lindley, and A. Smith, editors, *Bayesian Statistics 3*, pages 307–326, Oxford University Press, Oxford.
- Montoya-Delgado, L., Irony, T., de B. Pereira, C., and Whittle, M. (2001). An unconditional exact test for the Hardy-Weinberg equilibrium law: sample-space ordering using the Bayes factor. *Genetics*, **158**, 875–883.
- Pereira, C. and Rogatko, A. (1984). The Hardy-Weinberg equilibrium under a Bayesian perspective. *Revista Brasileira de Genetica*, **4**, 689–707.
- Shoemaker, J., Painter, I., and Weir, B. (1998). A Bayesian characterization of Hardy-Weinberg disequilibrium. *Genetics*, **149**, 2079–2088.
- Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. *Biometrics*, **66**, 257–265.

## Derivation of Bayes Factor for Assessing HWE

- We need to specify priors on the null and alternatives, and then calculate the **Bayes factor**:

$$\frac{\Pr(\mathbf{n}|H_0)}{\Pr(\mathbf{n}|H_1)} = \frac{\int \Pr(\mathbf{n}|p_1)p(p_1)dp_1}{\int \Pr(\mathbf{n}|q_1, q_2)p(q_1, q_2)dq_1dq_2}$$

where  $p_1$  and  $(q_1, q_2)$  are the parameters under the null and alternative, respectively.

- Under the null we have a single parameter, and under the alternative two.
- **Important point**: When Bayes factors are evaluated we need to include the normalizing constants.



## HWE Bayes Factor

- Under  $H_0$  and  $H_1$  we must take care to evaluate the probability of the same data,  $n_1, n_2, n_3$ .
- Under the **null**,

$$\Pr(\mathbf{n}|p_1) = \Pr(n_1, n_2, n_3|p_1) = \frac{n!2^{n_1}}{n_2!n_{12}!n_3!} p_1^{2n_1+n_2} (1-p_1)^{n_2+2n_3}.$$

- With a  $\text{Be}(w_1, w_2)$  prior on  $p_1$ :

$$\begin{aligned} \Pr(n_1, n_2, n_3|H_0) &= \int \Pr(\mathbf{n}|p_1) \times p(p_1) dp_1 \\ &= \frac{n!2^{n_2} \Gamma(w) \Gamma(2n_1 + n_2 + w_1) \Gamma(n_2 + 2n_3 + w_2)}{n_1!n_2!n_3! \Gamma(w_1) \Gamma(w_2) \Gamma(2n + w)} \end{aligned} \quad (2)$$

- This is the probability of the observed data under the null.

## HWE Bayes Factor

- The Bayes factor is

$$\frac{\Pr(\mathbf{n}|H_0)}{\Pr(\mathbf{n}|H_1)}$$

and we have just given the form of the numerator.

- We now turn to the denominator.
- Under the **alternative** we assume  $\mathbf{q} \sim \text{dirichlet}(v_1, v_2, v_3)$ .
- The probability of the data under the alternative is:

$$\begin{aligned}\Pr(n_1, n_2, n_3|H_1) &= \int \Pr(\mathbf{n}|q_1, q_2) \times p(q_1, q_2) dq_1 dq_2 \\ &= \frac{n! \Gamma(v) \Gamma(n_1 + v_1) \Gamma(n_2 + v_2) \Gamma(n_3 + v_3)}{n_1! n_2! n_3! \Gamma(v_1) \Gamma(v_2) \Gamma(v_3) \Gamma(n + v)}.\end{aligned}\tag{3}$$

- Again, just a probability distribution, which we may evaluate for any realization of  $(n_1, n_2, n_3)$ .

## The HWE Bayes Factor

- Hence, the **Bayes factor**, measuring the evidence in the data for the null, as compared to the alternative is:

$$\begin{aligned} \text{BF} &= \frac{\Pr(n_1, n_2, n_3 | H_0)}{\Pr(n_1, n_2, n_3 | H_1)} \\ &= \frac{2^{n_2} \Gamma(w) \Gamma(2n_1 + n_2 + w_1) \Gamma(v_1) \Gamma(v_2) \Gamma(v_3) \Gamma(n_2 + 2n_3 + w_2) \Gamma(n + v)}{\Gamma(w_1) \Gamma(w_2) \Gamma(2n + w) \Gamma(v) \Gamma(n_1 + v_1) \Gamma(n_2 + v_2) \Gamma(n_3 + v_3)} \end{aligned}$$

which is (2) divided by (3).

- This appears complex, but is just a function of the **observed data**, and the **prior inputs**, and can be easily evaluated.
- If  $\text{BF} > 1$  ( $< 1$ ) the data are **more** (**less**) likely to have come from the null.
- Can be readily extended to  $k > 2$  alleles.
- We next consider a formal decision rule.