2022 SISG Module 13: Bayesian Statistics for Genetics
Lecture 4: Multinomial and Poisson Models

Jon Wakefield

Departments of Statistics and Biostatistics
University of Washington

Outline

Introduction and Motivating Examples
- Inference for Parameters of Interest

Bayesian Analysis of Multinomial Data
- Derivation of the Posterior and Prior Specfication

Bayes Factors

Poisson Modeling of Count Data

Ranking Countries on Excess Covid Mortality

Appendix
- Bayes Factor Details
- Non-Conjugate Analysis

Introduction

Introduction

- In this lecture we will consider the Bayesian modeling of count data, in particular multinomial and Poisson data, with an extension to negative binomial.
- The examination of Hardy-Weinberg equilibrium will be used to motivate a multinomial model.
- Again, conjugate priors will be used.
- 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.
- We will fit simple Poisson and negative binomial models to an AIDS example dataset.

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:

$$\begin{aligned} p_1 &= q_1 + \frac{q_2}{2} \\ p_2 &= \frac{q_2}{2} + q_3 \end{aligned}$$

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 Real Example

Lidicker *et al.* (1997) examined genetic variation in sea otter populations (Enhydra lutris) in the eastern Pacific.

- Locus EST gave the data $n_1 = 37, n_2 = 20, n_3 = 7$, with $n = 64$.
- Are these frequencies consistent with HWE?
- The MLEs are:

$$\begin{aligned} \widehat{q}_1 &= \frac{37}{64} = 0.58 \quad \widehat{q}_2 = \frac{20}{64} = 0.31 \quad \widehat{q}_3 = \frac{7}{64} = 0.11 \\ \widehat{p}_1 &= \frac{37 \times 2 + 20}{128} = 0.73 \quad \widehat{p}_2 = \frac{20 + 7 \times 2}{128} = 0.27. \end{aligned}$$

- 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 0.11.

A Toy Example

In this made up example we have $n = 100$ so calculations are simpler.

Example:

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

$$\begin{array}{rcccc} \widehat{q}_1 & = & 0.88 & \widehat{q}_2 = 0.10 & \widehat{q}_3 = 0.02 \\ \widehat{p}_1 & = & 0.93 & \widehat{p}_2 = 0.07 & \end{array}$$

- 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) χ^2 tests or exact tests.
- χ^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).

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 the A_1 allele for the two possibilities at a locus; so $X_1 = X_2 = 1$ corresponds to the genotype A_1A_1.
- 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

$$\begin{aligned} q_1 &= p_1^2 + p_1(1-p_1)f \\ q_2 &= 2p_1(1-p_1)(1-f) \\ q_3 &= (1-p_1)^2 + p_1(1-p_1)f \end{aligned}$$

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

Bayesian Analysis of Multinomial Data

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 $\boldsymbol{q} = (q_1, q_2, q_3)$ to represent the vector of probabilities, and $\boldsymbol{n} = (n_1, n_2, n_3)$ for the data vector.
- Via Bayes Theorem:

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

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

Elements of Bayes Theorem: The Likelihood

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

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

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

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

- 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 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:

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

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

The Dirichlet Distribution as a Prior Choice for a Multinomial q

- The Dirichlet(v_1, v_2, v_3) prior:

$$\begin{aligned} p(q_1, q_2, q_3) &= \frac{\Gamma(v_1 + v_2 + v_3)}{\Gamma(v_1)\Gamma(v_2)\Gamma(v_3)} \times q_1^{v_1 - 1} q_2^{v_2 - 1} q_3^{v_3 - 1} \\ &\propto q_1^{v_1 - 1} q_2^{v_2 - 1} q_3^{v_3 - 1}. \end{aligned}$$

- $v_1, v_2, v_3 > 0$ 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 $k = 2, 3, ...$ categories).
- The beta distribution is a special case of the dirichlet when there are two categories only.

Dirichlet Prior

- The mean and variance are

$$\begin{aligned} \mathrm{E}[q_i] &= \frac{v_i}{v_1 + v_2 + v_3} = \frac{v_i}{v} \\ \mathrm{var}(q_i) &= \frac{\mathrm{E}[q_i](1 - \mathrm{E}[q_i])}{v_1 + v_2 + v_3 + 1} = \frac{\mathrm{E}[q_i](1 - \mathrm{E}[q_i])}{v + 1} \end{aligned}$$

  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.

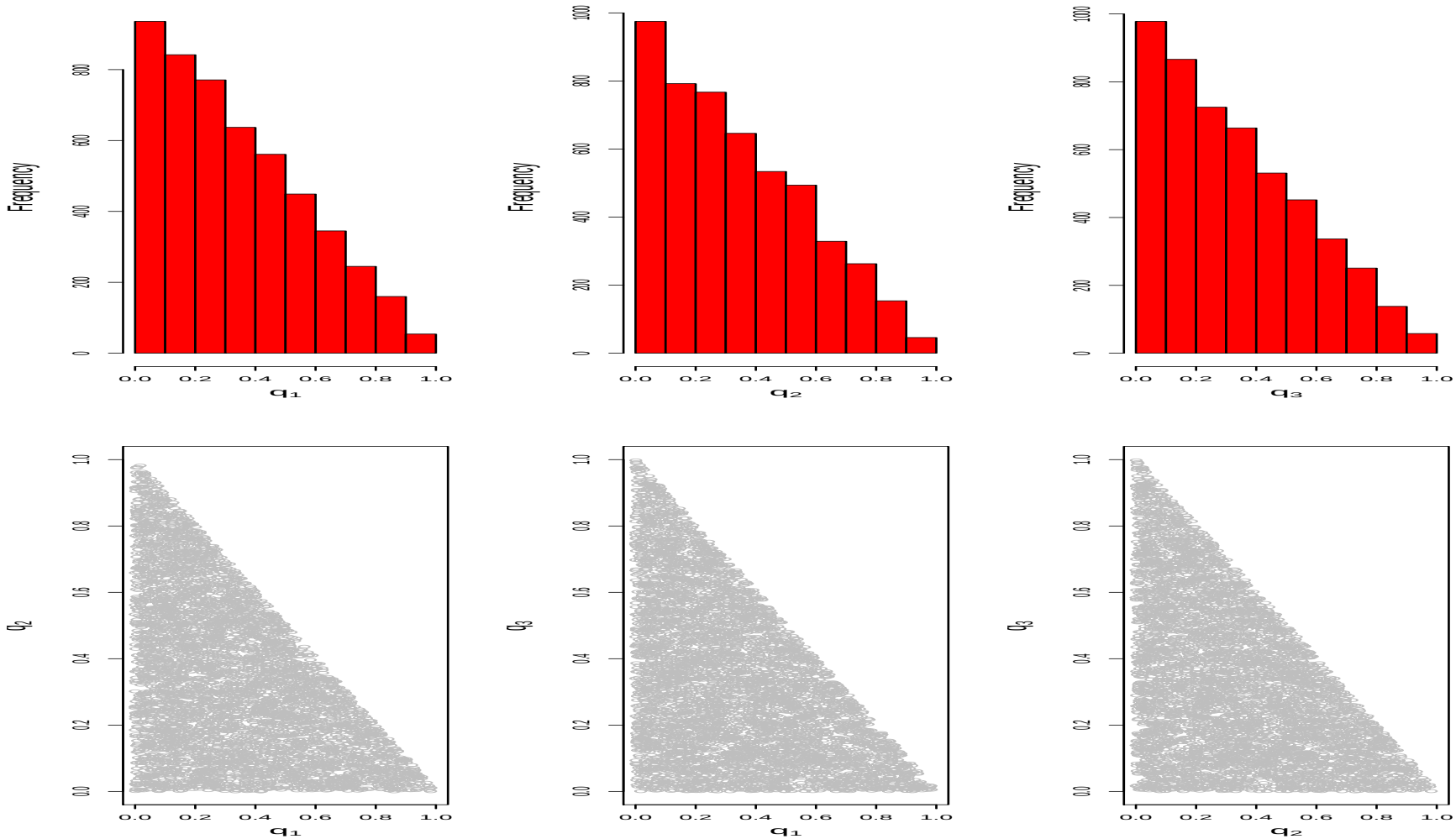


Figure 1: Samples from a Dirichlet$(1, 1, 1)$ distribution. The mean is $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.

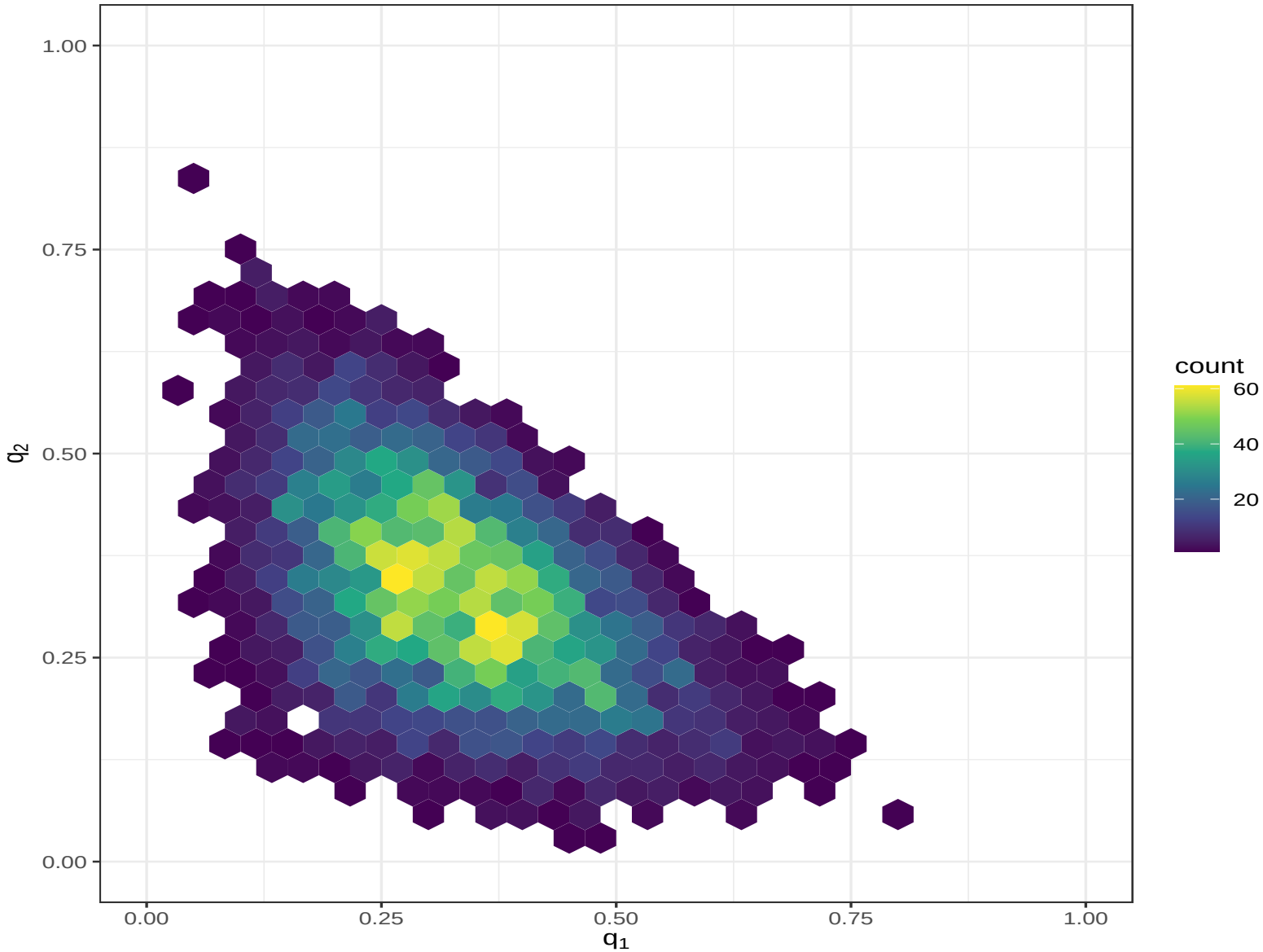


Figure 2: q_1, q_2 samples from a Dirichlet$(5, 5, 5)$. The mean is $\left(\frac{1}{3}, \frac{1}{3}\right)$.

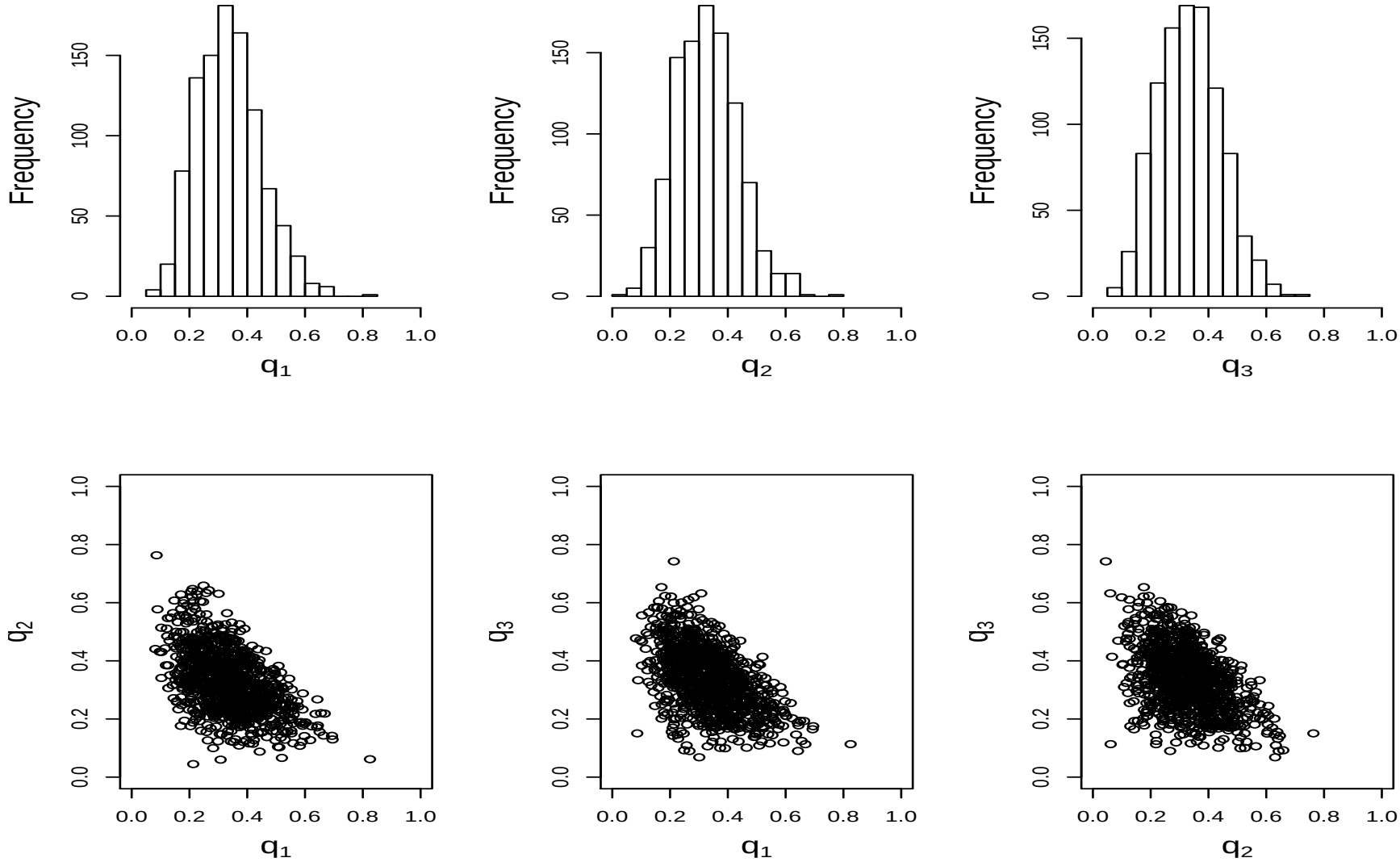


Figure 3: Samples from a Dirichlet(6, 6, 6) distribution. The mean is $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.

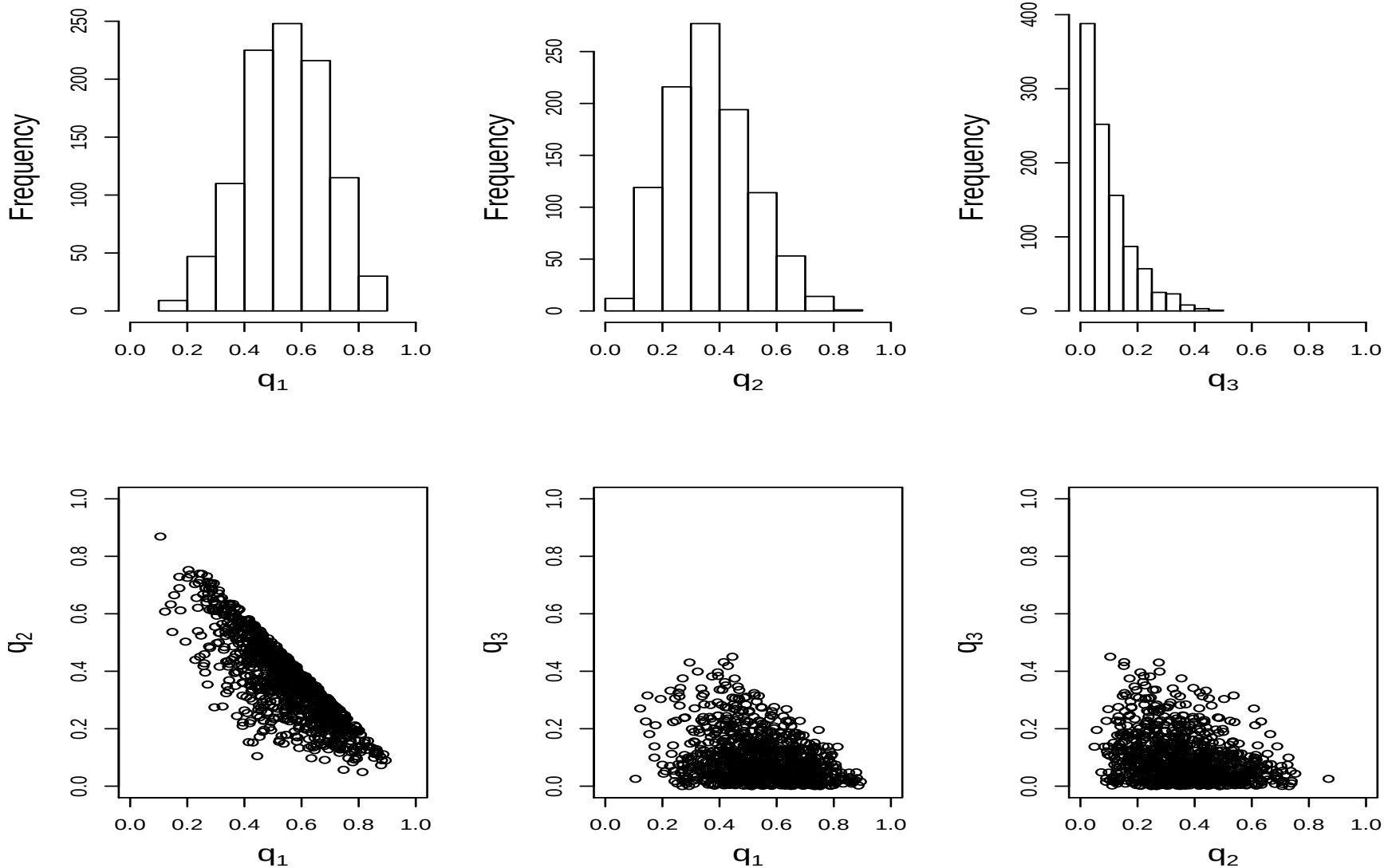


Figure 4: Samples from a Dirichlet(6, 4, 1) distribution. The mean is $\left(\frac{6}{11}, \frac{4}{11}, \frac{1}{11}\right) = (0.55, 0.36, 0.09)$.

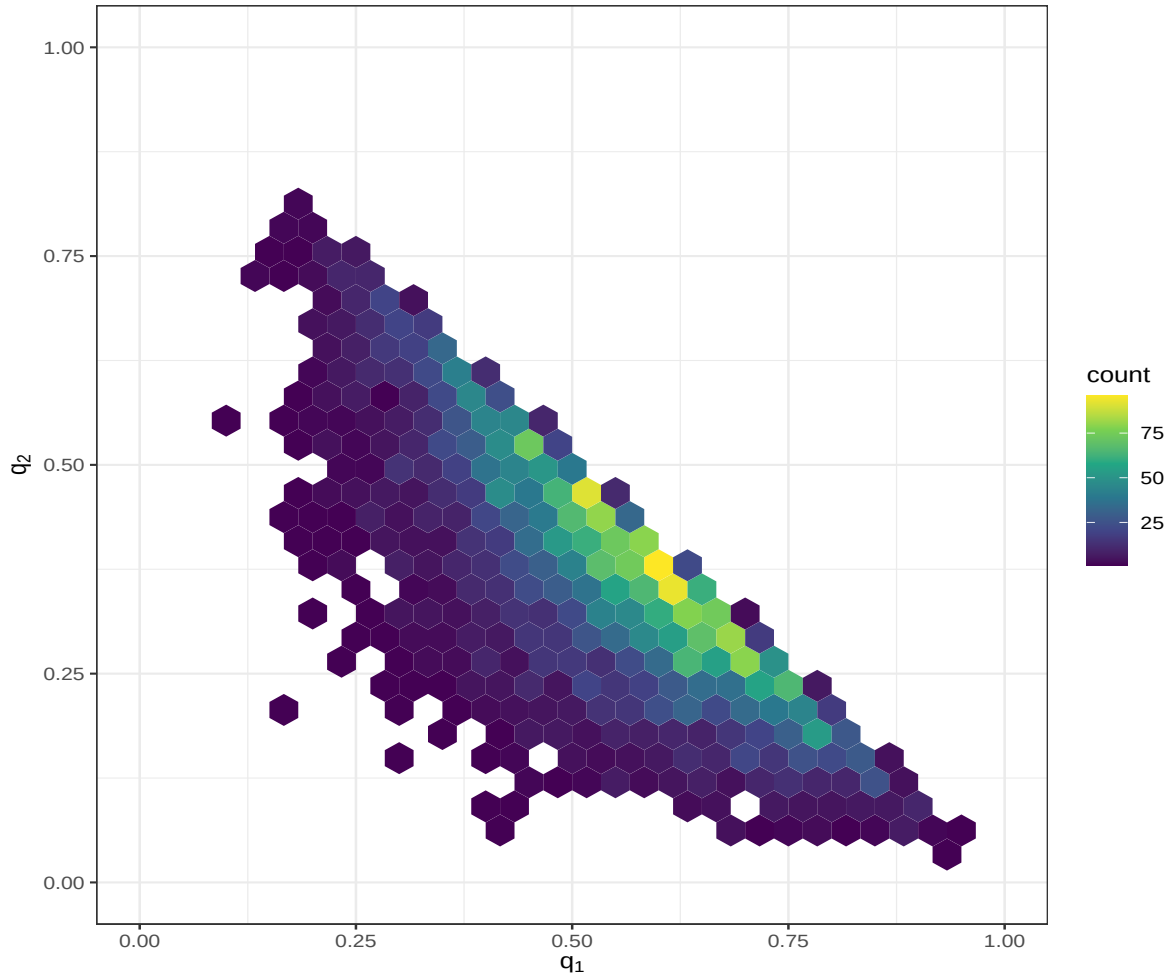


Figure 5: Hexbin plot of q_1, q_2 samples from a Dirichlet$(6, 4, 1)$ distribution.

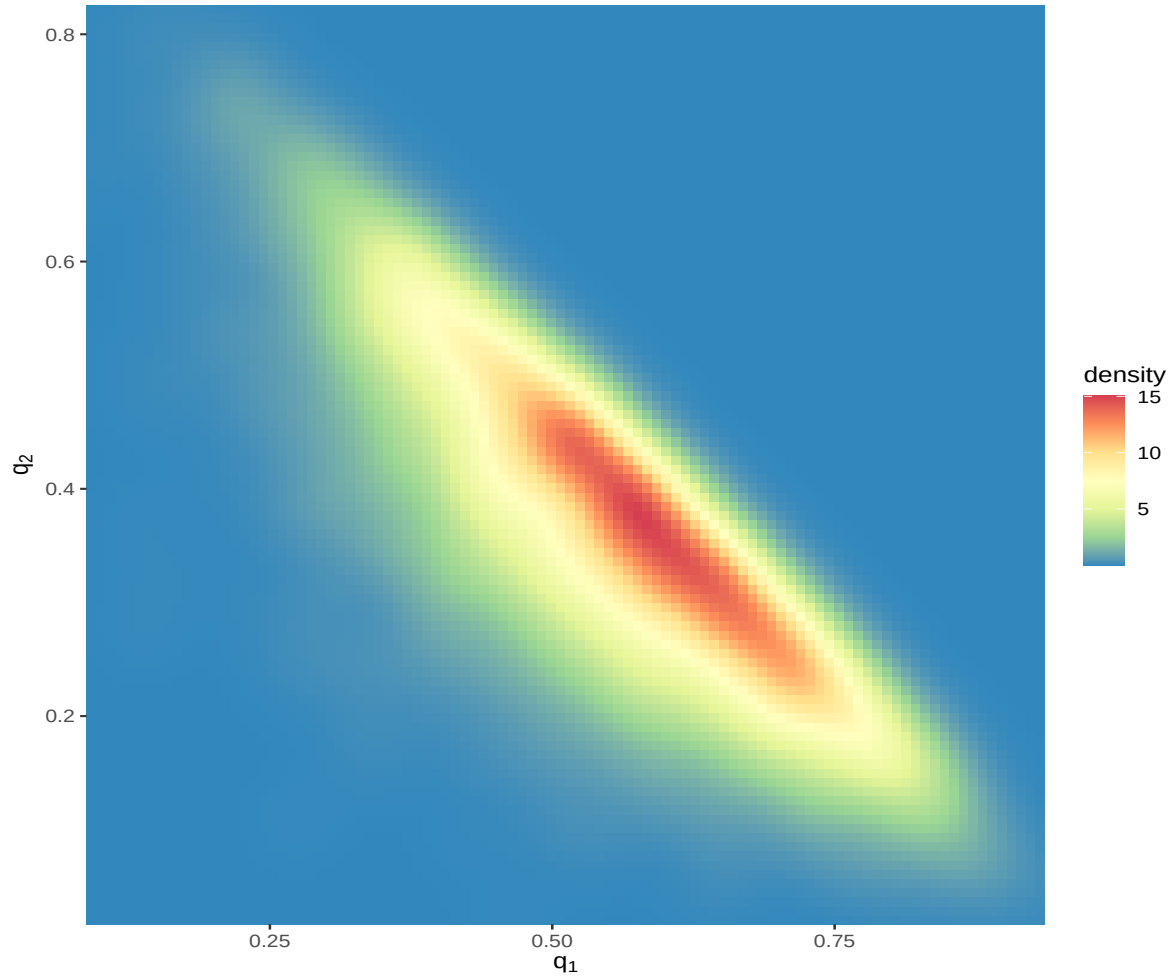


Figure 6: Image plot of q_1, q_2 from a Dirichlet$(6, 4, 1)$ distribution.

Parameters of Interest

- Each of D, ψ 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 $\boldsymbol{q}$, Dirichlet$(1, 1, 1)$, does not correspond to a flat prior for D, f, ψ, as Figure 7 shows.
- With a "flat" Dirichlet prior Dirichlet(1,1,1) the prior probability that $f > 0$ is 0.67.

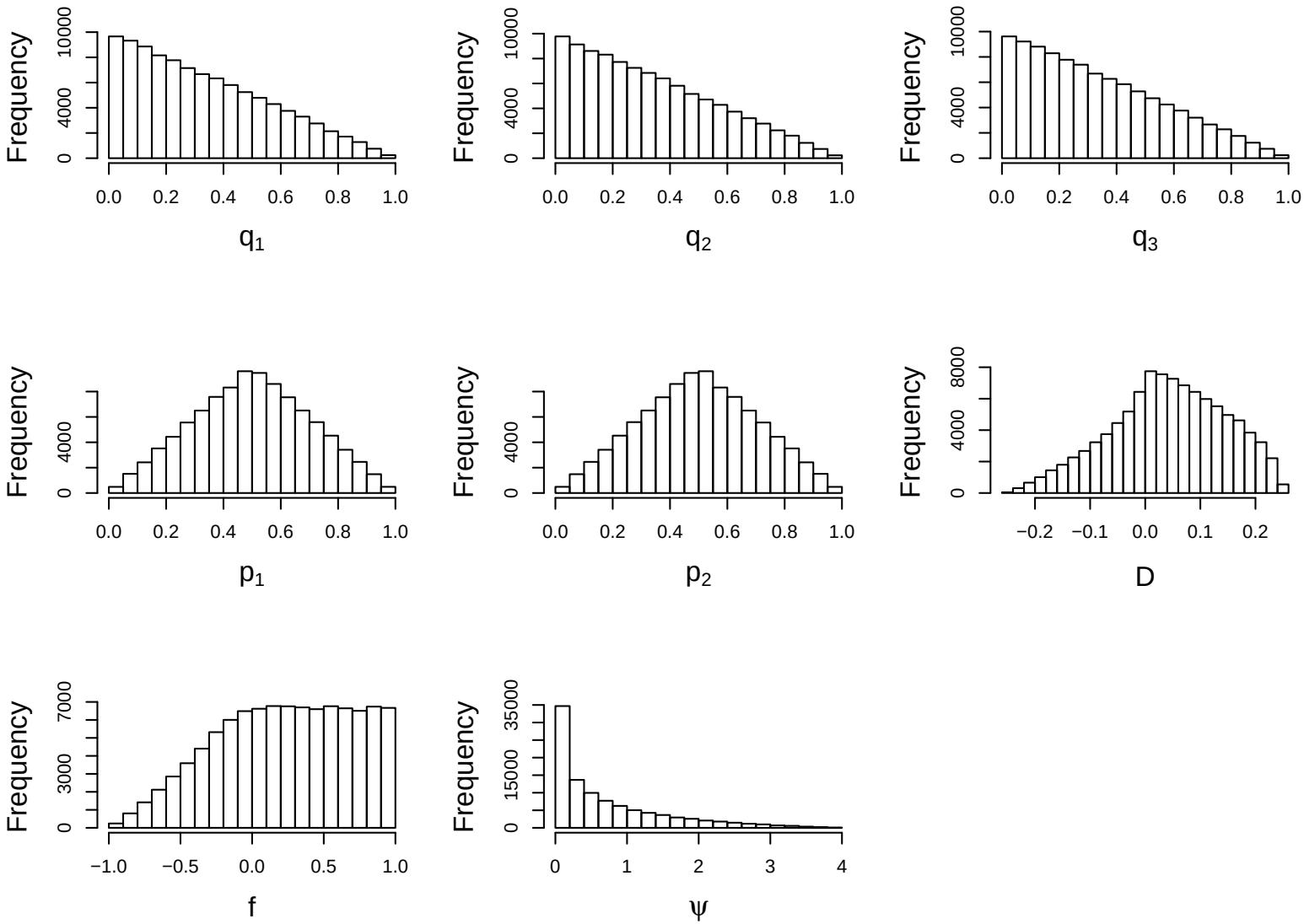


Figure 7: Samples from a Dirichlet(1,1,1) for various functions.

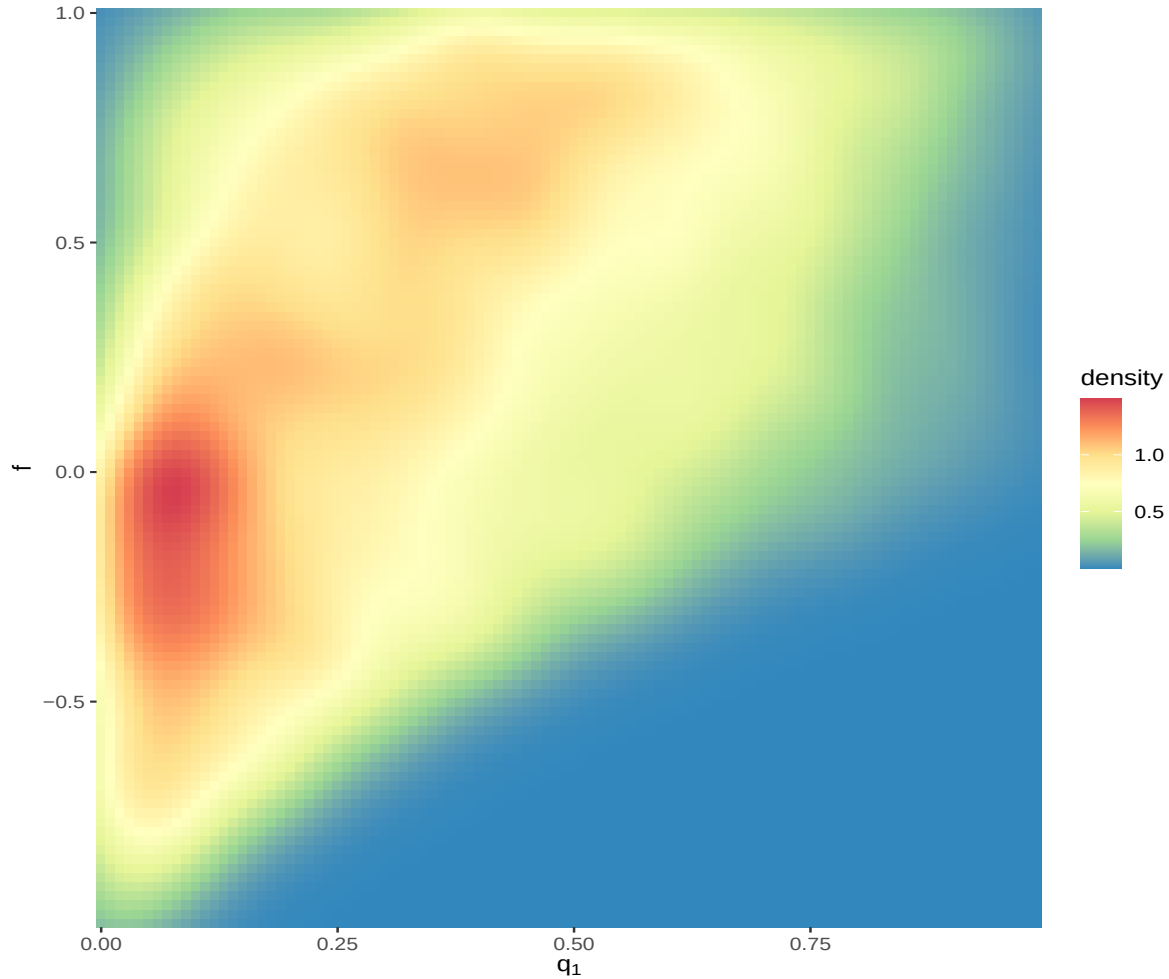


Figure 8: Image plot of q_1, f from a Dirichlet$(1, 1, 1)$ distribution.

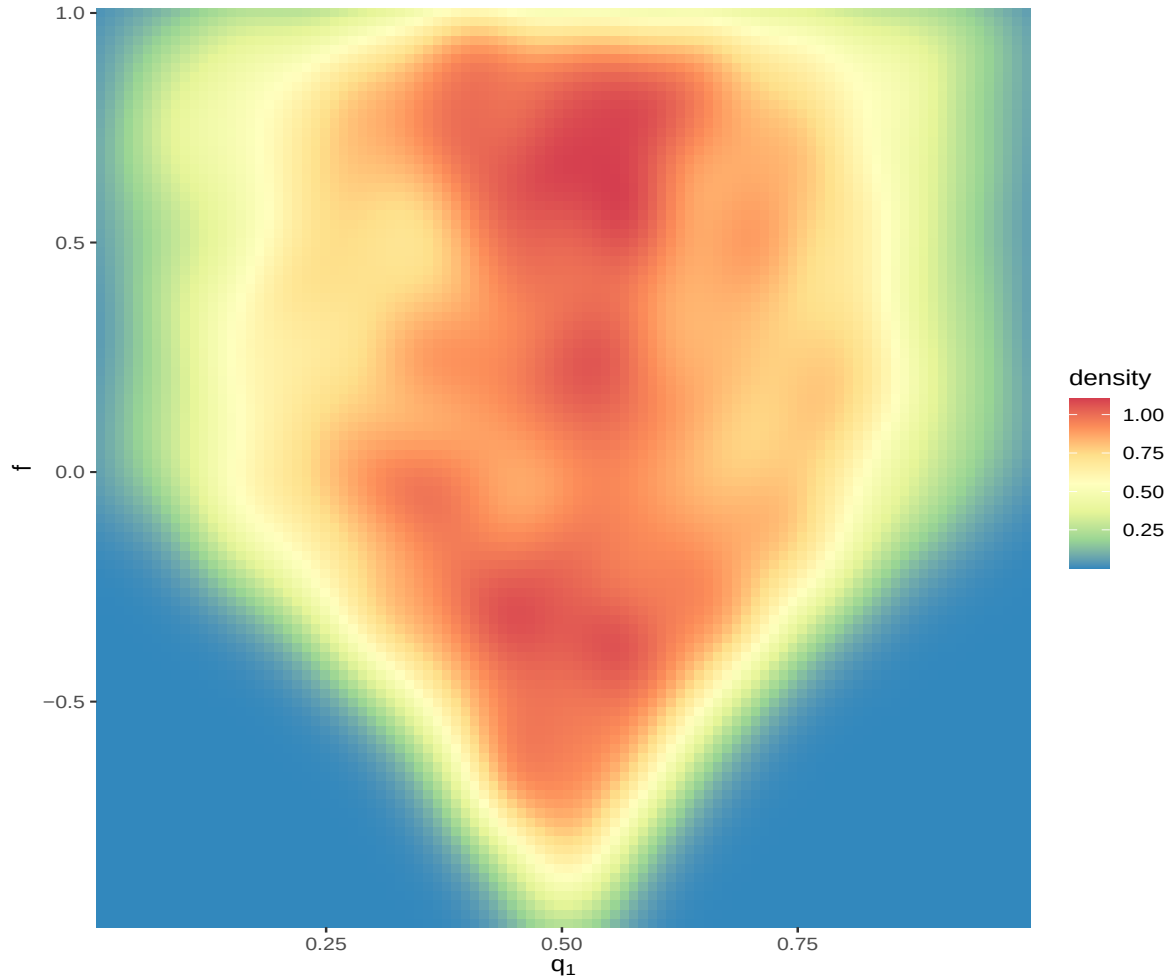


Figure 9: Image plot of p_1, f from a Dirichlet$(1, 1, 1)$ distribution.

Posterior Distribution

- Combining the Dirichlet prior, Dirichlet(v_1, v_2, v_3), with the multinomial likelihood gives the posterior:

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

- 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 $i = 1, 2, 3$:

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

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

- The weight W is

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

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

Choosing a Prior

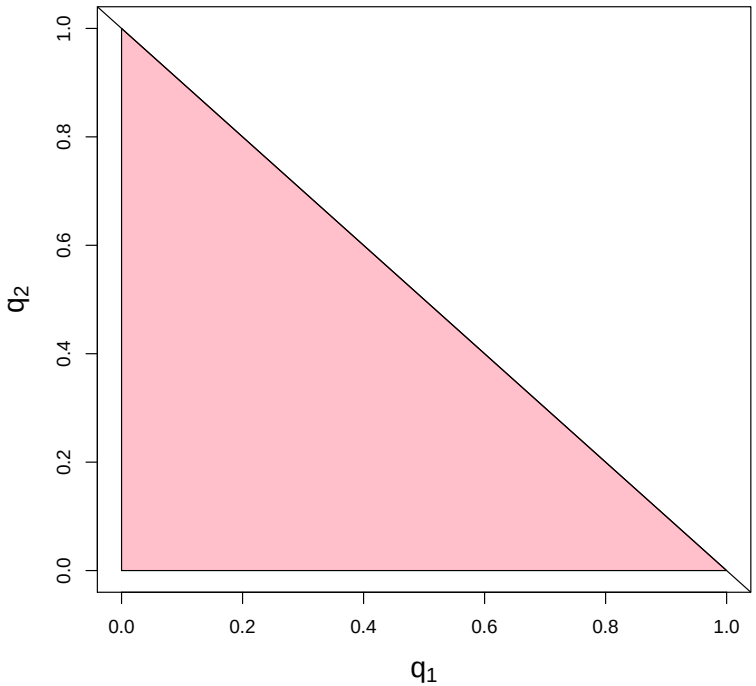
- Recall the prior mean is

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

- 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 have seen.

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 $\boldsymbol{q}$:
$$v_1 = v_2 = v_3 = 1.$$
- This gives the posterior as Dirichlet$(88+1, 10+1, 2+1)$ with posterior means:
$$\begin{aligned} \mathrm{E}[q_1|\boldsymbol{n}] &= \frac{1+88}{3+100} = \frac{89}{103} \\ \mathrm{E}[q_2|\boldsymbol{n}] &= \frac{1+10}{3+100} = \frac{11}{103} \\ \mathrm{E}[q_3|\boldsymbol{n}] &= \frac{1+2}{3+100} = \frac{3}{103}. \end{aligned}$$
- 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

$$\text{Dirichlet}(88+1, 10+1, 2+1).$$

- As posterior summaries we display, in Figure 13:
  - Histograms of the 3 univariate marginal distributions $p(q_1|\boldsymbol{y})$, $p(q_2|\boldsymbol{y})$, $p(q_3|\boldsymbol{y})$.
  - Scatterplots of the 3 bivariate marginal distributions $p(q_1, q_2|\boldsymbol{y})$, $p(q_1, q_3|\boldsymbol{y})$, $p(q_2, q_3|\boldsymbol{y})$.
- 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

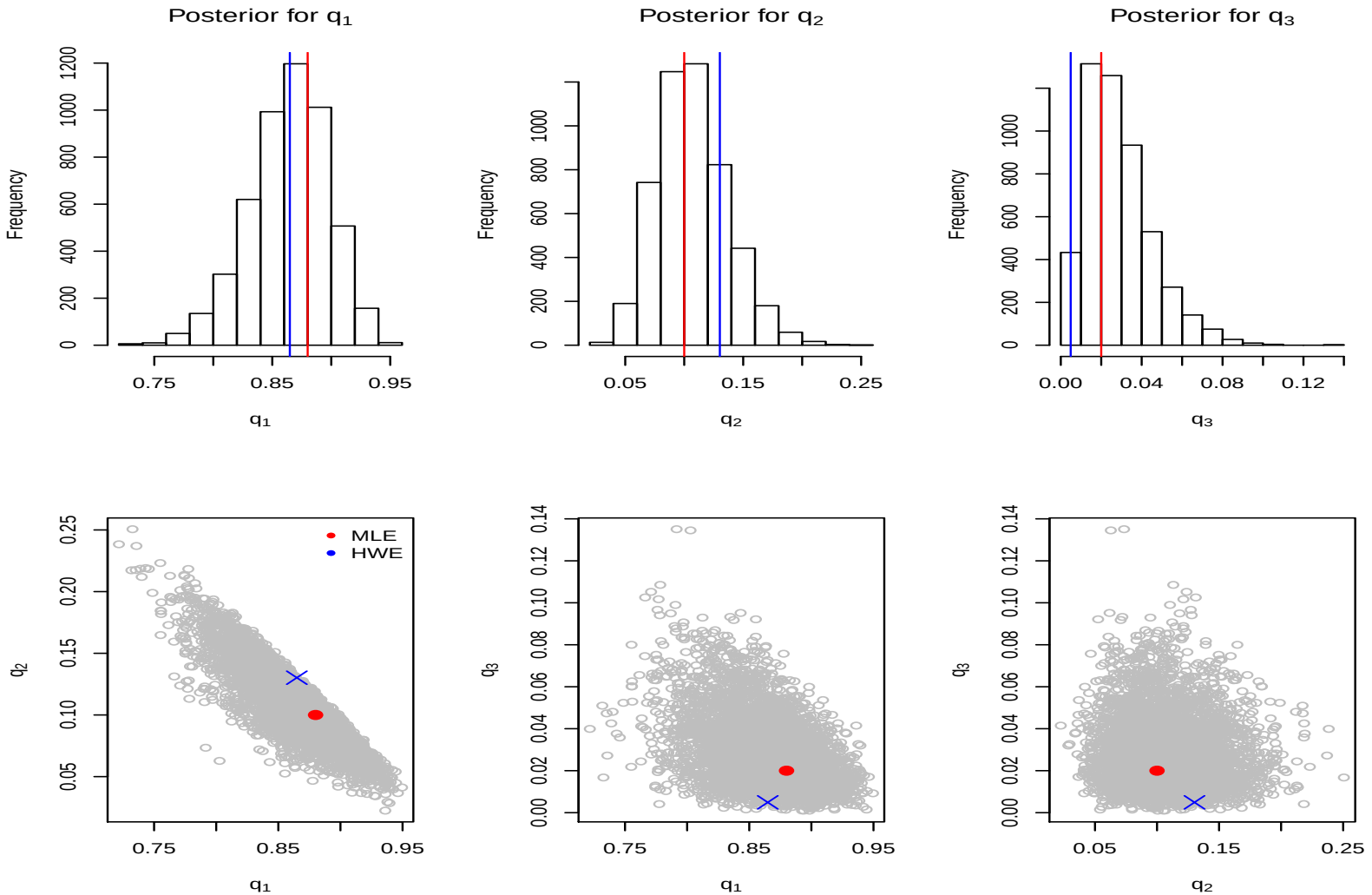


Figure 10: Univariate and bivariate posterior distributions for $\boldsymbol{n} = (88, 10, 2)$. MLEs in red for the general model and in blue for the HWE model.

Bayes analysis of (88,10,2) data

- As expected with a sample size of $n = 100$ and a flat prior, the MLEs (in red) lie close to the center of the posteriors.
- Note the asymmetry of the posteriors.
- Asymptotic confidence intervals of the form $\widehat{q}_i \pm 1.96 \times \text{se}(\widehat{q}_i)$ 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 θ.
- In a multinomial context the situation is more complex; shortly we will examine Bayes factors to carry out hypothesis testing.

Bayes Factors

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

$$H_0 : \text{HWE} \quad \text{versus} \quad H_1 : \text{Not HWE}.$$

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

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

  where p_1 and (q_1, q_2) are the parameters under the null and alternative, respectively.
- Under the null we have $(p_1, p_2) \sim \text{Beta}(w_1, w_2)$ and under the alternative $(q_1, q_2, q_3) \sim \text{Dirichlet}(v_1, v_2, v_3)$.

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[1].
- If 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.

[1] When we work out a χ^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 π_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.

A Simple Example

The posterior probability of the null is

$$\Pr(H_0|\boldsymbol{n}) = \frac{1.54}{1+1.54} = 0.61.$$

This probability is very sensitive to the prior on the null, π_0.

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

$$\Pr(H_0|\boldsymbol{n}) = \frac{3.08}{1+3.08} = 0.75.$$

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).

Poisson Modeling of Count Data

AIDS Data

Whyte et al. (1987) reported deaths due to AIDS in Australian 3-month periods from January 1983 to June 1986.

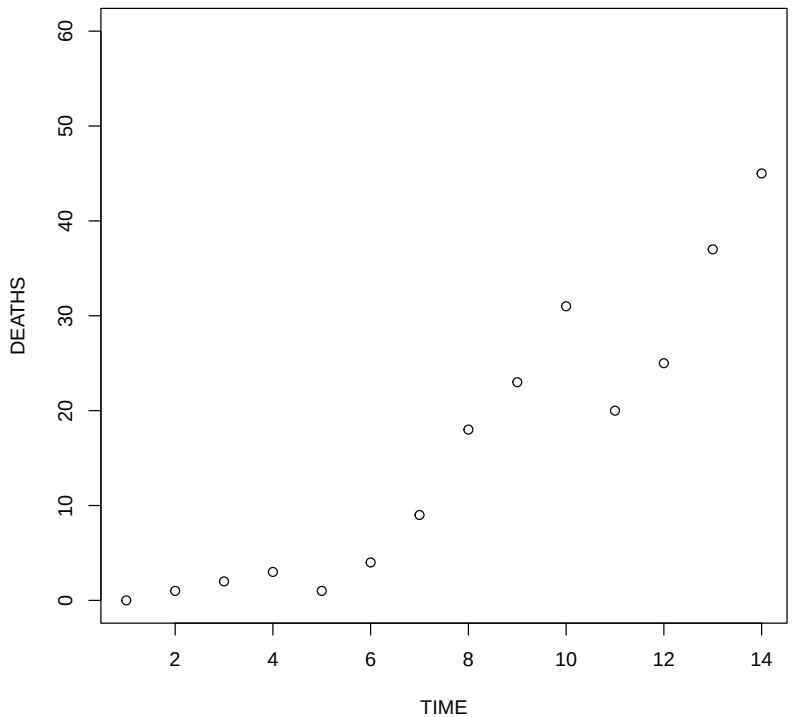


Figure 11: AIDS death in Australia as a function of time.

AIDS Data

We illustrate Bayesian modeling of these count data using a very simple Poisson loglinear model:

$$\begin{aligned} Y_i|\mu_i &\sim \text{Poisson}(\mu_i) \\ \log \mu_i &= \beta_0 + \beta_1 \log(\text{time}_i) \end{aligned}$$

For this model, we require priors on β_0 and β_1, and as with logistic regression, conjugate priors don't exist to provide an analytically tractable analysis.

But it is straightforward to fit such models in INLA, with independent normal priors on β_0, β_1: with the default priors:

```
AIDS.inla1 <- inla(DEATHS ~ log(TIME),
                   data = AIDS, family = "poisson")
round(AIDS.inla1$summary.fixed[,1:5], 4)
               mean     sd 0.025quant 0.5quant 0.975quant
(Intercept) -1.9429 0.5112    -2.9902  -1.9275    -0.9829
log(TIME)    2.1749 0.2149     1.7687   2.1693     2.6132
```

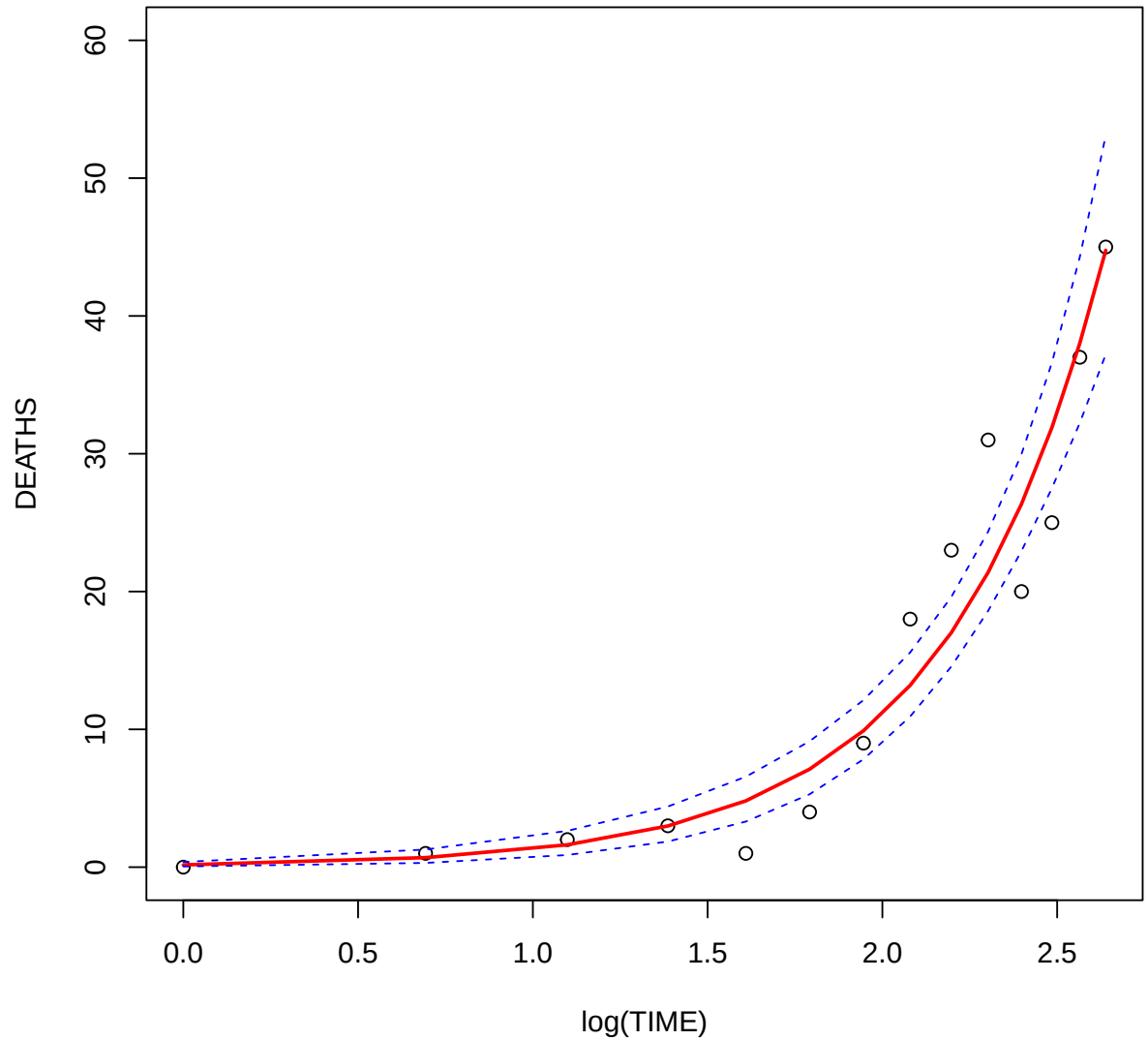


Figure 12: AIDS death in Australia as a function of time, with posterior mean (and 95% credible interval) of the expected value, Poisson model.

Negative Binomial Model

The Poisson model is often inadequate, as the variance is constrained to equal the mean.

The negative binomial model adds an overdispersion parameter ϕ, such that

$$\text{var}(Y_i) = \mu_i(1 + \mu_i/\phi),$$

to increase flexibility.

This model is also straightforward to fit in INLA:

```
AIDS.inla2 <- inla(DEATHS ~ log(TIME),
                   data = AIDS, family = "nbinomial")
round(AIDS.inla2$summary.fixed[,1:5], 4)
                mean     sd 0.025quant 0.5quant 0.975quant
(Intercept) -2.0210 0.5784    -3.2318  -1.9965    -0.9527
log(TIME)    2.2101 0.2491     1.7485   2.1998     2.7310
```

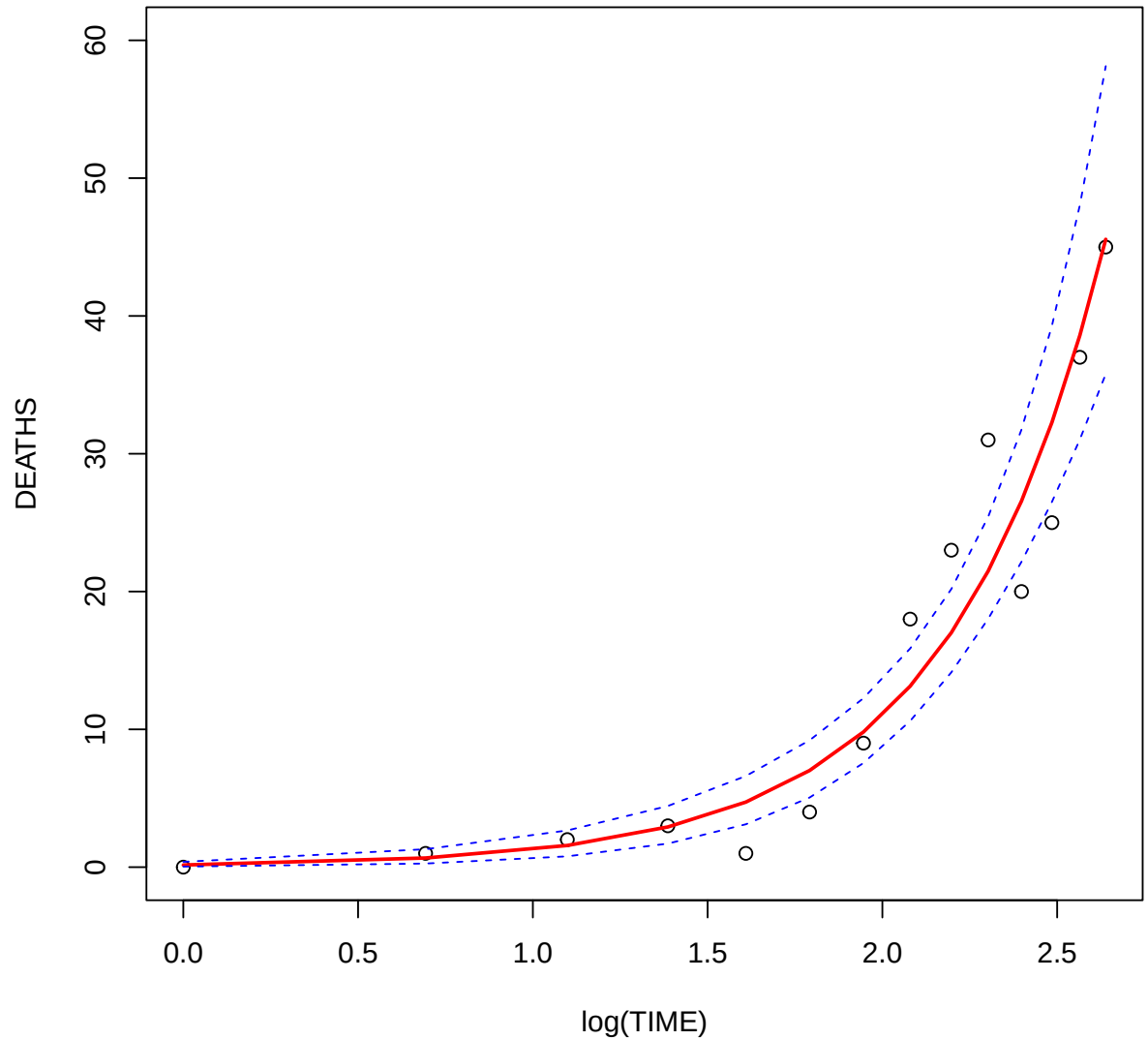


Figure 13: AIDS death in Australia as a function of time, with posterior mean (and 95% credible interval) of the expected value, negative binomial model.

Ranking Countries on Excess Covid Mortality

Excess Covid Mortality in Europe

?) describe a model to estimate:

$$\text{Excess Mortality in Country } c \text{ in Month } t = Y_{c,t} - E_{ct},$$

where

- $Y_{c,t}$ is the mortality count for country c in month t.
- $E_{c,t}$ is the expected mortality (based on historic data) count for country c in month t – these is uncertainty in this count.

Whether we like it or not, people will rank countries, on the basis of the excess rate, i.e., the count divided

Excess Covid Mortality in Europe

We obtain samples from the posterior the excess, and at any month we can take the set of sampled rates across countries and rank them.

We illustrate using data on a small set of European countries, before moving to a larger group.

The probabilities sum to one for each country within each month, i.e., a country has to be in one of the ranking positions.

Also for a fixed month and ranking position across all countries the probabilities sum to one because one of the countries has to be in rank 1, 2, etc, in each month.

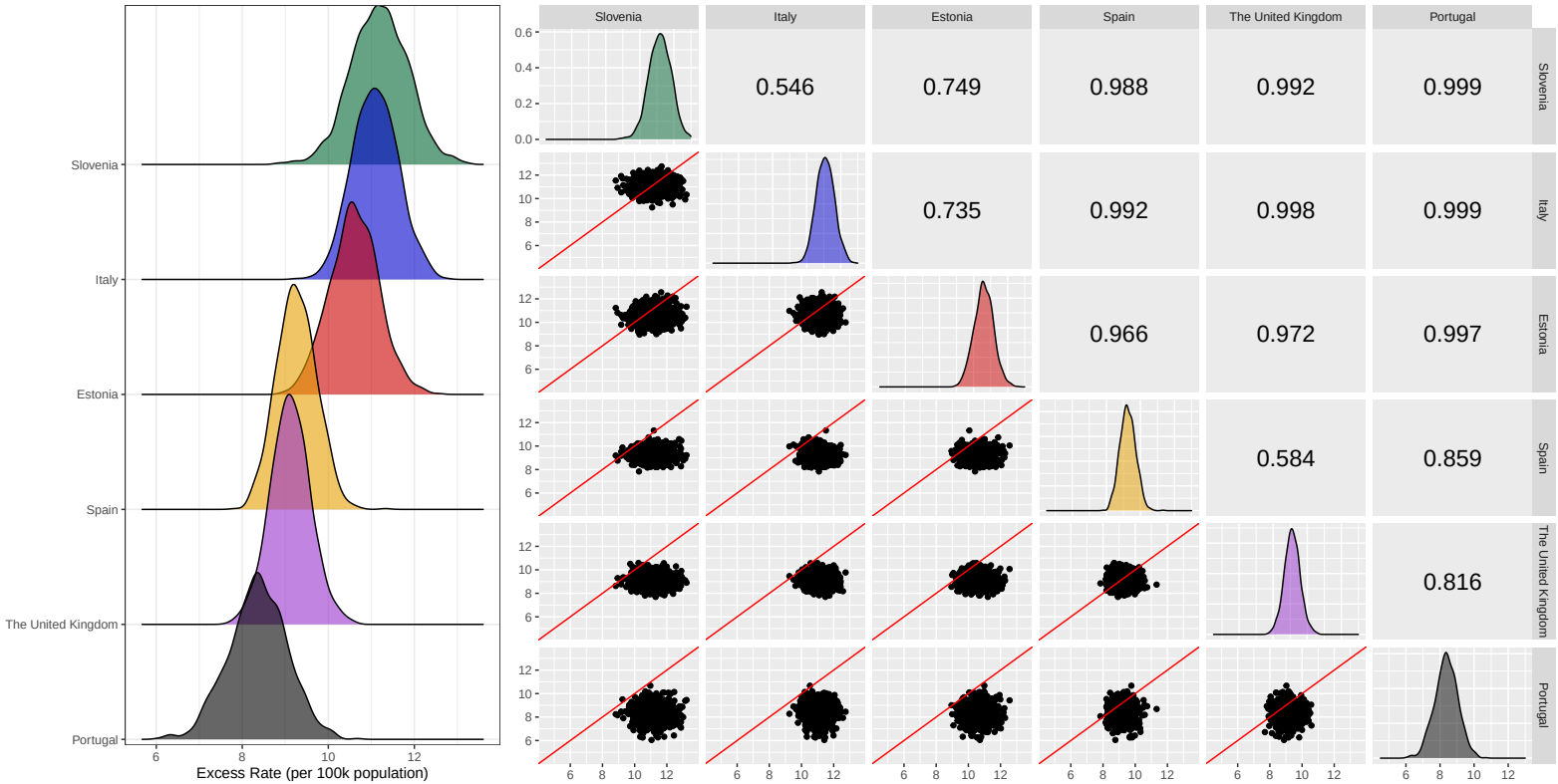


Figure 14: Left: Ridgeplots representing the uncertainty in the cumulative excess monthly mortality rate over January 2020–December 2021 for six European countries. Right: Bivariate plots of pairs of excess rates (lower triangular), 1-dimensional summaries for individual countries (diagonal), and probabilities that the excess for the country labeled on the left exceeds the rate for the country labeled at the top. These probabilities are the fraction of points that lie above the red line in the corresponding bivariate plot.

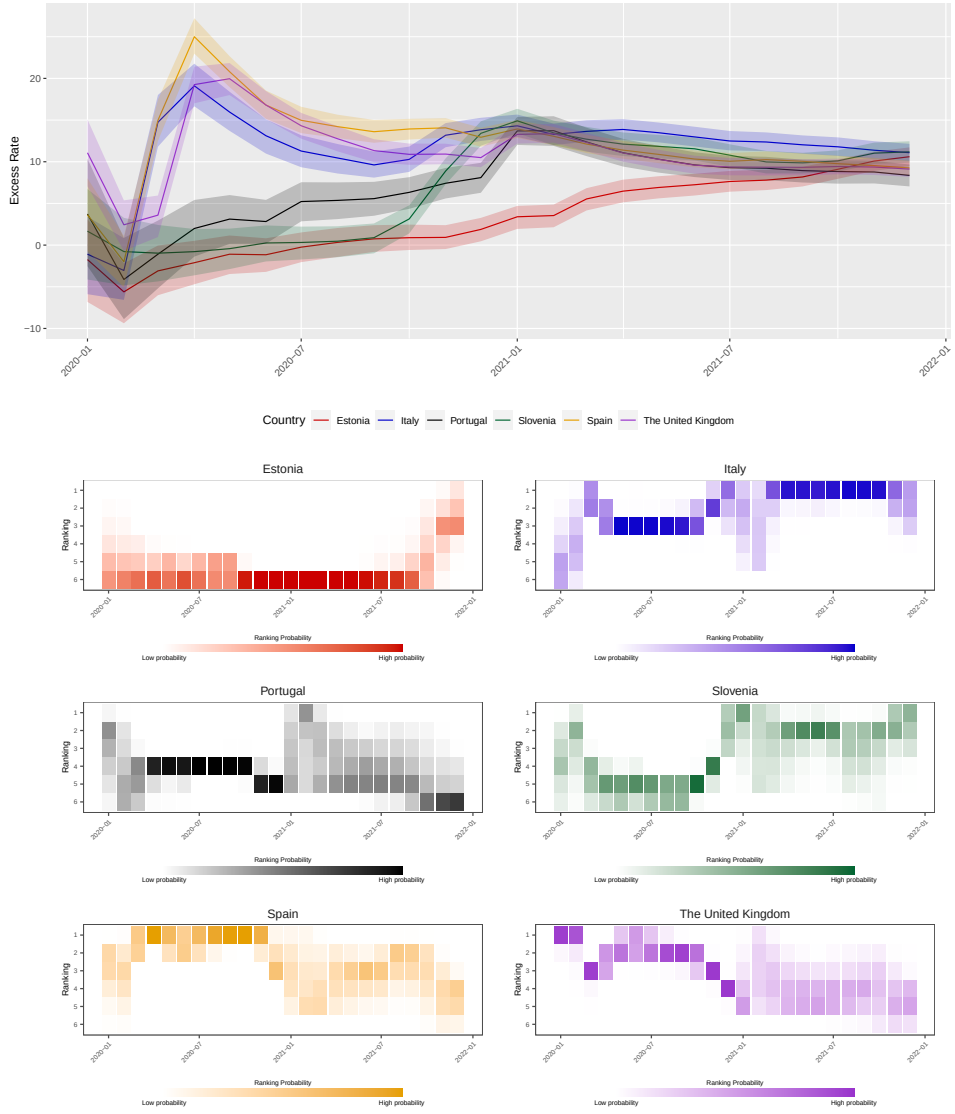


Figure 15: Top: excess monthly mortality rate (per 100K of population), with 95% uncertainty intervals, by month, for six European countries. The bottom five panels display the ranking probabilities for each of the individual countries.

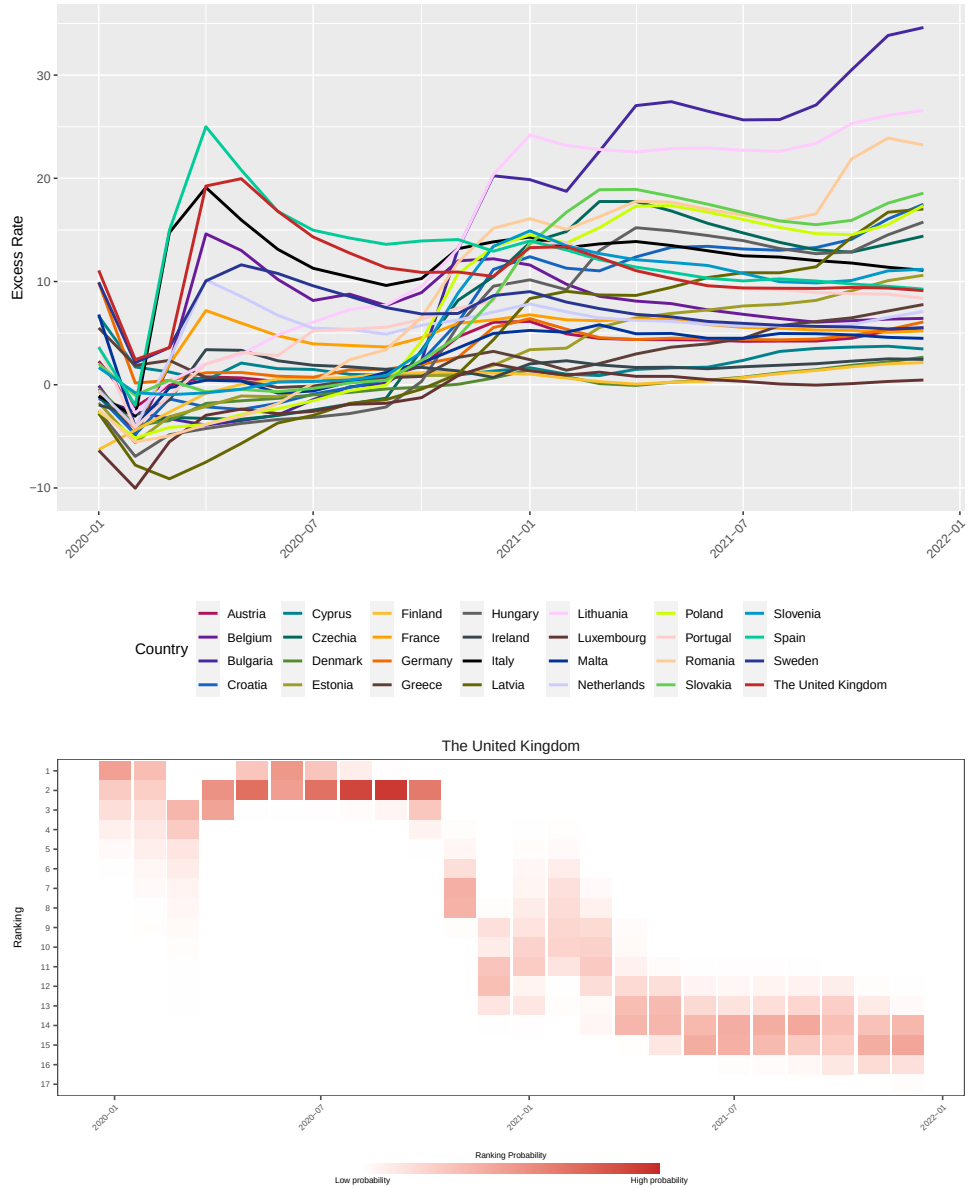


Figure 16: The UKs rank (bottom plot) as compared to 27 other European countries. The rates by month are also shown (top plot).

Conclusions

HWE Example:

- 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) – see Appendix.
- For multiple alleles computation is slow whether the approach is frequentist or Bayesian.
- On the course website there is `Stan` code to analyze multinomial data, and this allows flexibility in prior specification.

Conclusions

Poisson/negative binomial example:

- Poisson and negative binomial models are straightforward to fit using INLA.
- In fact, any generalized linear models (GLMs) are easy.
- In Lecture 7, we extend these models, to allow for random effects.

Conclusions

General Conclusions:

- In multiparameter situations, integration is required.
- INLA can perform the necessary integrations, and is fast and relatively easy to use, though can't be used for all models.
- Bayes factors are sensitive to the prior.
- Monte Carlo sampling is a powerful tool for inference.

References

Lidicker, W., William, Z., and McCollum, F. (1997). Allozymic variation in California sea otters. *Journal of Mammalogy*, **78**, 417–425.

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.

Wakefield, J. (2013). *Bayesian and Frequentist Regression Methods*. Springer, New York.

Appendix

Derivation of Bayes Factor for Assessing HWE

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

$$\frac{\Pr(\boldsymbol{n}|H_0)}{\Pr(\boldsymbol{n}|H_1)} = \frac{\int \Pr(\boldsymbol{n}|p_1)p(p_1)dp_1}{\int \Pr(\boldsymbol{n}|q_1,q_2)p(q_1,q_2)dq_1\,dq_2}$$

  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,

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

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

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

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

HWE Bayes Factor

- The Bayes factor is

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

  and we have just given the form of the numerator.
- We now turn to the denominator.
- Under the alternative we assume $\boldsymbol{q} \sim \text{Dirichlet}(v_1, v_2, v_3)$.
- The probability of the data under the alternative is:

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

- 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 BF $>$ 1($<$ 1) the data are more (less) likely to have come from the null.
- Can be readily extended to $k > 2$ alleles.

A Non-Conjugate Test of HWE

The above prior specifications are convenient analytically, but in some situations we would like to perform Bayesian inference using priors that are based on contextual information.

If we are really interested in the deviations from HWE of a sample from a particular population, then we may have strong prior information which perhaps can be represented through a prior on the inbreeding coefficient f.

A Different Prior for the Alternative

Under the null we have a single probability p_1, the probability of an A_1 allele.

Under the alternative we may specify the prior

$$\pi(p_1, f) = \pi(p_1) \times \pi(f|p_1)$$

where the conditioning allows the constraints on f:

$$f_{\min} = \max\left(-\frac{p_1}{1-p_1}, -\frac{1-p_1}{p_1}\right) < f < 1$$

Unfortunately there is no closed form calculations for finding posterior distributions and Bayes factors, instead we describe a simulation-based technique —the rejection algorithm.

A Rejection Algorithm

Let θ denote the parameters with prior distribution $\pi(\theta)$, and let $\widehat{\theta}$ be the MLE and $p(\boldsymbol{y}|\widehat{\theta})$ the maximized likelihood.

Then the rejection algorithm (e.g., Wakefield, 2013, Chapter 3) proceeds as follows:

1. Generate $U \sim U(0,1)$ and $\theta \sim \pi(\theta)$, independently.
2. Accept θ if
$$U < \frac{p(\boldsymbol{y}|\theta)}{p(\boldsymbol{y}|\widehat{\theta})},$$
otherwise reject θ.
3. Return to 1.

The resultant $\theta^{(s)}$, $s = 1, \ldots, S$, are an independent sample from the posterior $p(\theta|\boldsymbol{y})$.

A Rejection Algorithm

The rejection algorithm may be very inefficient if the prior and likelihood differ substantially (e.g., prior is dispersed and/or likelihood is peaked).

An estimate of the normalizing constant (required for Bayes factor calculation) is given by

$$\widehat{p}(\boldsymbol{y}) = \frac{1}{S}\sum_{s=1}^{S} p(\boldsymbol{y}|\boldsymbol{\theta}^{(s)})$$

where $\boldsymbol{\theta}^{(s)} \sim \pi(\cdot)$.

Note that this only requires samples from the prior — the rejection algorithm is not needed.

In the HW context the maximized likelihood is available in closed form.

Specific Non-Conjugate Priors

Recall the prior is

$$\pi(p_1, f) = \pi(p_1) \times \pi(f|p_1)$$

Two components:

- For $\pi(p_1)$ we take a Be(w_1, w_2) prior.
- For $\pi(f|p_1)$ we transform to

$$\phi = \log((f - f_{\min})/(1 - f))$$

and assume $\phi|p_1$ is normal.

HWE Example Revisited

We again consider the data $n_{11} = 88, n_{12} = 10, n_{22} = 2$.

These data give a p-value of 0.0654. The MLE for f is 0.23 with asymptotic standard error 0.17. MLE of HWE proportions: $(0.865, 0.130, 0.05)$.

With flat conjugate Dirichlet priors we obtained a Bayes factor of 1.54 so that the data are 50 % more likely under the null, but the evidence is low.

HWE Example Revisited

We assume that the 50% point of the prior on f is 0, and the 95% point is 0.5

We obtain a Bayes factor of 0.29 so that the data are 3.4 times as likely under the alternative, but the evidence is again weak.

The posterior probability that $f > 0$ is 0.98.

The difference between the priors is that the non-conjugate version gives more weight close to where the data are located.

Graphical Summaries

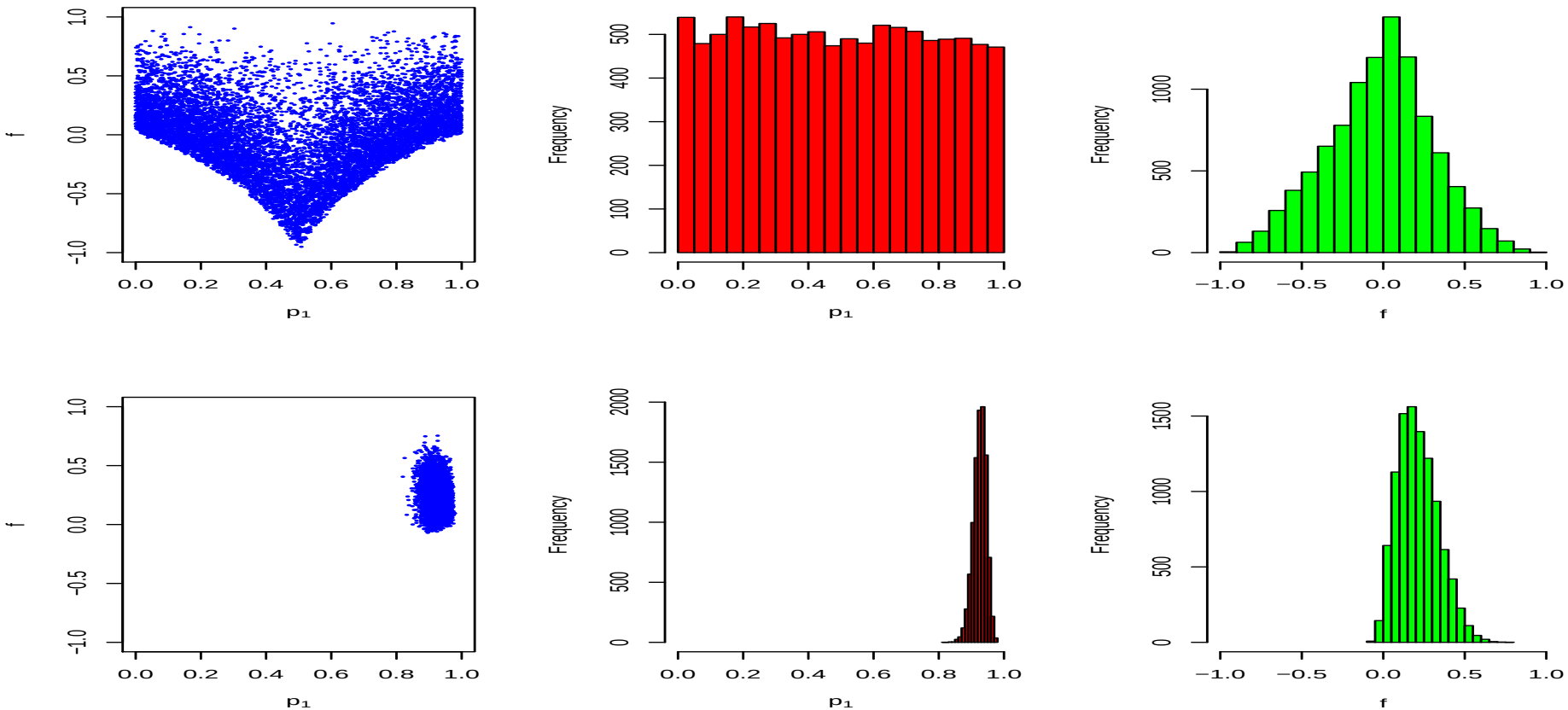


Figure 17: Prior (top) and Posterior (bottom). Notice the clear constraint in the top left plot.

Influence of Prior

In contrast to estimation, in which the prior influence generally disappears with increasing sample size, the Bayes factor remains influenced by the prior.

To illustrate we multiply the data of the previous example by different factors.

Factor	Conj BF	Non-conj BF	Post prob $f > 0$	p-value
1	1.54	0.29	0.984	0.0654
2	0.40	0.070	0.997	0.0089
5	0.0039	0.000639	1	3.6×10^{-5}
10	1.2×10^{-6}	1.8×10^{-7}	1	5.3×10^{-9}

The conjugate and non-conjugate Bayes factors remain quite different (though the substantive conclusions are the same).