Bayesian Statistics for Genetics

Lecture 4: Multinomial Samples & GLMs/INLA

July, 2023

Outline

- Conclude our beta/binomial discussion with its extension to *multinomial* data – and conjugate priors
- Hardy-Weinberg equilibrium examples
- INLA, for posterior calculations in more general models

Motivating Example: HWE

- We measure a diallelic marker on n unrelated individuals
- The data, and the model's notation:

	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

- There is a fixed unknown probability q_1, q_2, q_3 for each of the genotypes – and $q_1 + q_2 + q_3 = 1$ so there are two free parameters
- Define the proportions of alleles A_1 and A_2 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: HWE

- Formally, HWE is the statistical independence of an individual's alleles at a locus
- Under HWE, the genotype probabilities are

	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: genotyping error, but also small population size, selection, inbreeding and population structure

Motivating Example: HWE

Lidicker *et al* (1997) examined genetic variation in sea otters.

With n=64, they got n_1=37,n_2=20,n_3=7. Are these frequencies consistent with HWE?

- The MLEs are:

$$\widehat{q}_1 = \frac{37}{64} = 0.58 \quad \widehat{q}_2 = \frac{20}{64} = 0.31 \quad \widehat{q}_3 = \frac{7}{64} = 0.11$$

$$\widehat{p}_1 = \frac{37 \times 2 + 20}{128} = 0.73 \quad \widehat{p}_2 = \frac{20 + 7 \times 2}{128} = 0.27$$

- An 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.

Motivating Example: HWE

- Testing for HWE is carried out via χ^2 tests – that use *asymptotic* i.e. large-sample approximations – or or *exact* tests, that don't
- The accuracy of the χ^2 test's approximation depends on sample size (smallest cell, broadly) and α, the level of Type I error rate control
- Computing the exact test can be a burden, particularly when there are many alleles/samples
- The *discreteness* of the test statistic is a problem – e.g. the exact test has to be conservative to control Type I error rates
- In general, choosing α is tricky; the null of **exact** HWE isn't plausible, so how often we'd reject it when it holds (i.e. T1ER) isn't obviously relevant
- Doing estimation, the *parameter space constraints* are a further challenge, particularly when expressing uncertainty. (This gets worse with more 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 indicators of the A_1 allele for the two possibilities at a locus; so $X_1 = X_2 = 1$ corresponds to 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, Painter & Weir, 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

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

  So **positive** values of f indicate an excess of homozygotes (and may indicate inbreeding), while **negative** values indicate an excess of heterozygotes.

Derivation of the Posterior and Prior Specification

	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

- With three counts, the multinomial is known as a *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:

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

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

Derivation of the Posterior and Prior Specification

- We assume n independent draws each with common probabilities $\mathbf{q} = (q_1, q_2, q_3)$ of being in each category. The distribution of n_1, n_2, n_3 is called a *multinomial*:

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

  Viewing this as a function of $\mathbf{q}$ gives the *likelihood function*.
- The maximum likelihood estimate (MLE) is

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

  i.e. the values which give the highest probability to the observed data

The Dirichlet distribution, as a prior for $\mathbf{q}$

With the parameters specified we can think about their prior.

- We need a prior distribution over (q_1, q_2, q_3) — that respects all three probabilities lying in [0,1], and adding to 1
- The *Dirichlet* distribution satisfies these requirements. Denoted Dirichlet(v_1, v_2, v_3) it 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 for $\mathbf{q}$

- Viewed as a prior, $v_1, v_2, v_3 > 0$ are specified to reflect what we know about (q_1, q_2, q_3)
- Note that the Dirichlet generalizes the Beta, and in particular we can view v_1, v_2, v_3 as acting like having those number of observations in each category.
- 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, with only two categories

The Dirichlet distribution, as a prior for $\mathbf{q}$

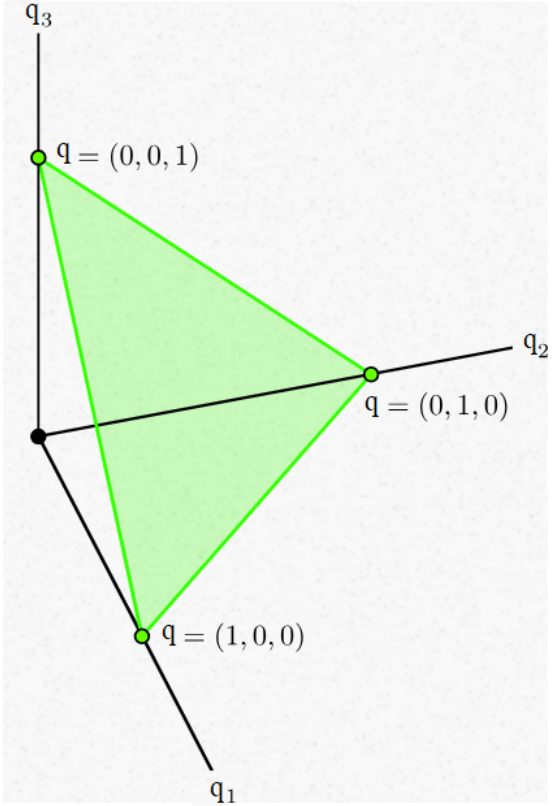
- The mean and variance are

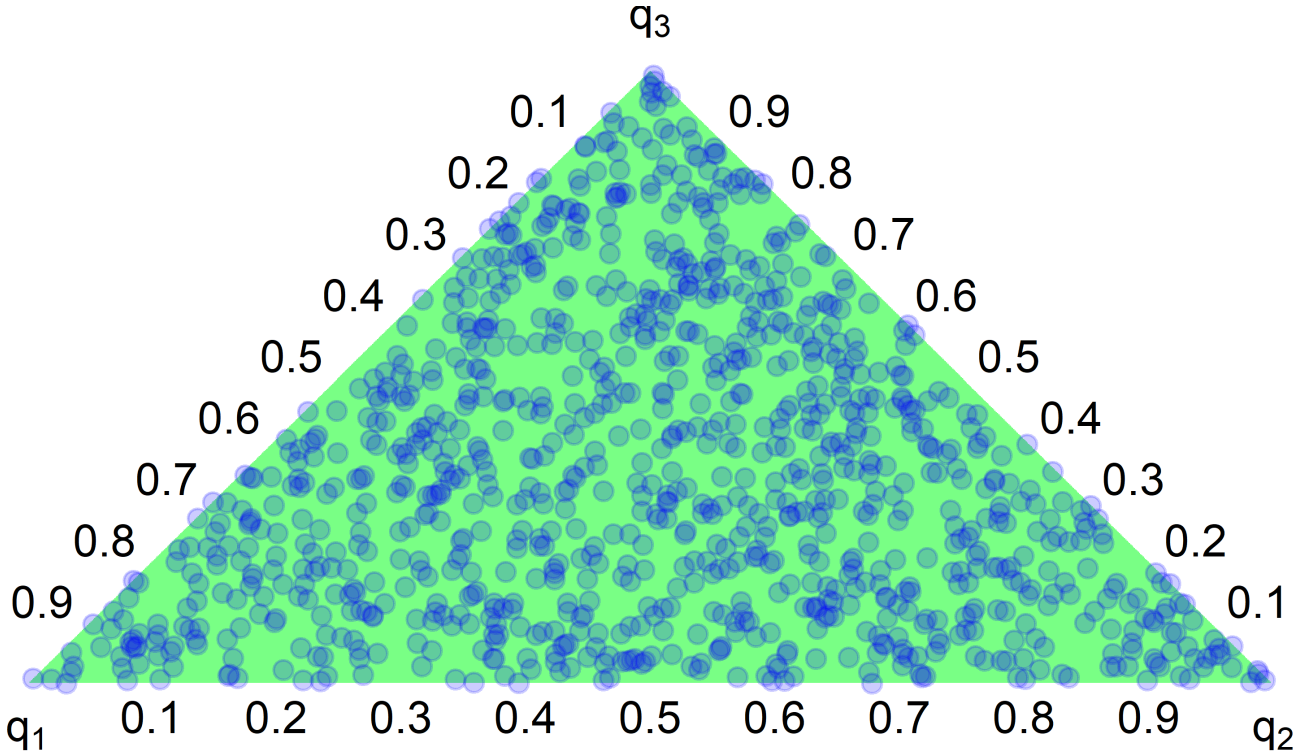
$$\begin{aligned}\mathbb{E}[q_i] &= \frac{v_i}{v_1+v_2+v_3} = \frac{v_i}{v} \\ \mathsf{Var}(q_i) &= \frac{\mathbb{E}[q_i](1-\mathbb{E}[q_i])}{v_1+v_2+v_3+1} = \frac{\mathbb{E}[q_i](1-\mathbb{E}[q_i])}{v+1}\end{aligned}$$

  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 uses it single parameter ($\mathbf{v}$) to control both location and spread, which is a deficiency.
- Quartiles can be calculated empirically, i.e. from samples.

The Dirichlet distribution, as a prior for $\mathbf{q}$

We use *ternary* plots (see below left) to illustrate Dirichlet samples (below right, from Dirichlet(1,1,1)) and densities (next slides).

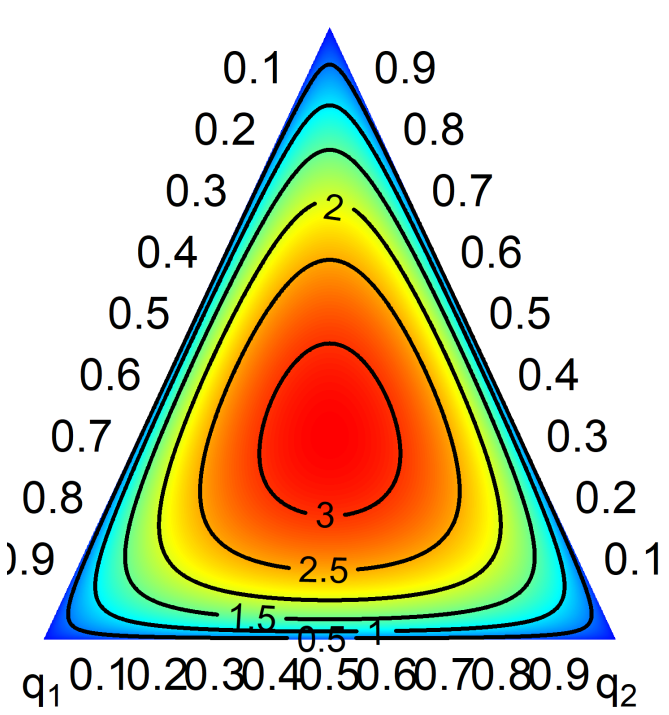




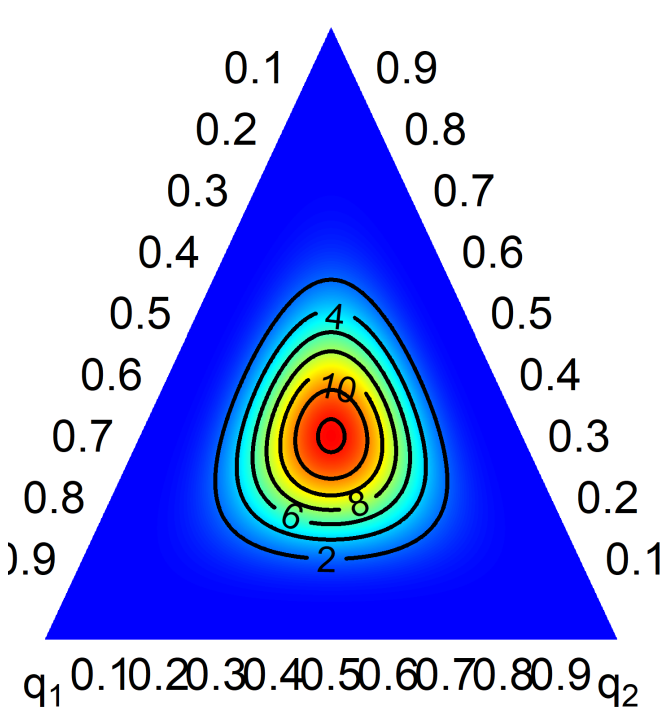
The Dirichlet distribution, as a prior for $\mathbf{q}$

Densities, shown on ternary plots:

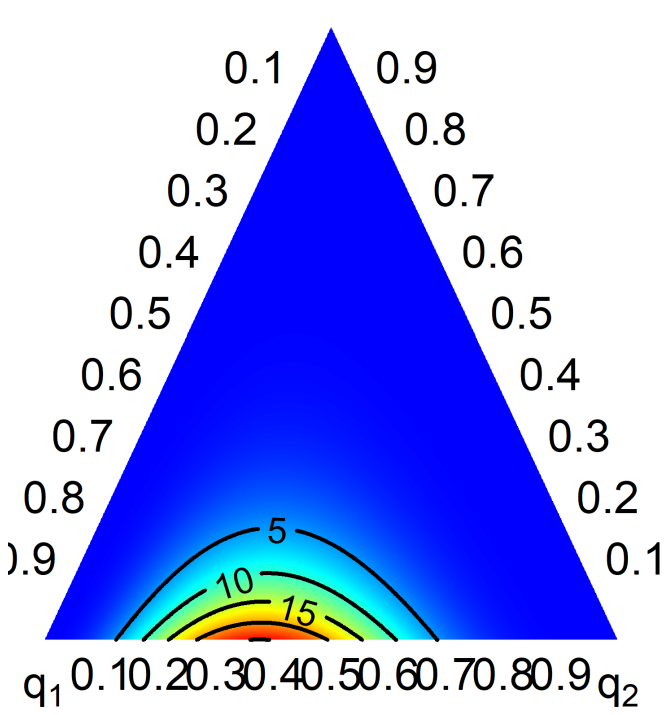
Dirichlet(1.5,1.5,1.5)



Dirichlet(6,6,6)

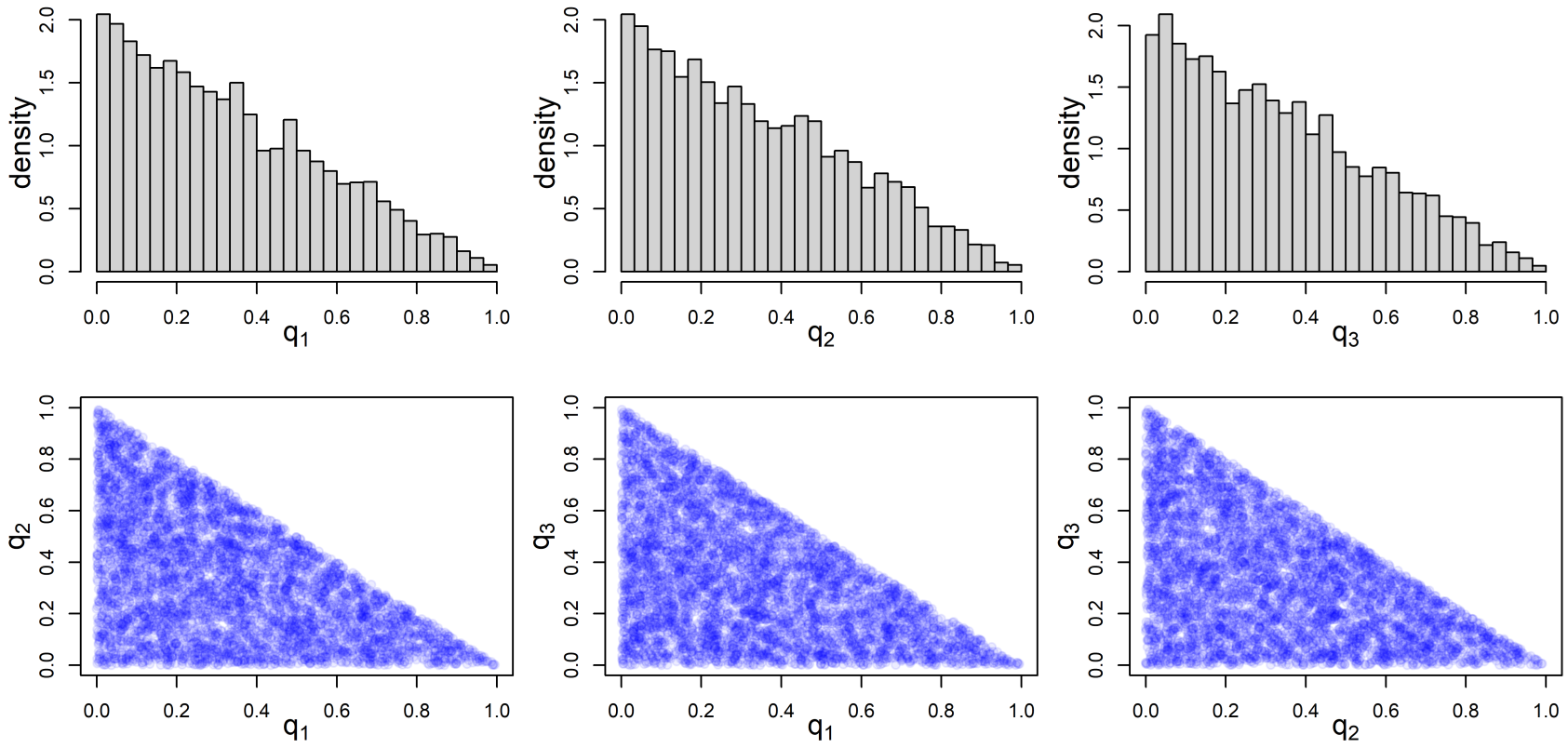


Dirichlet(6,4,1)



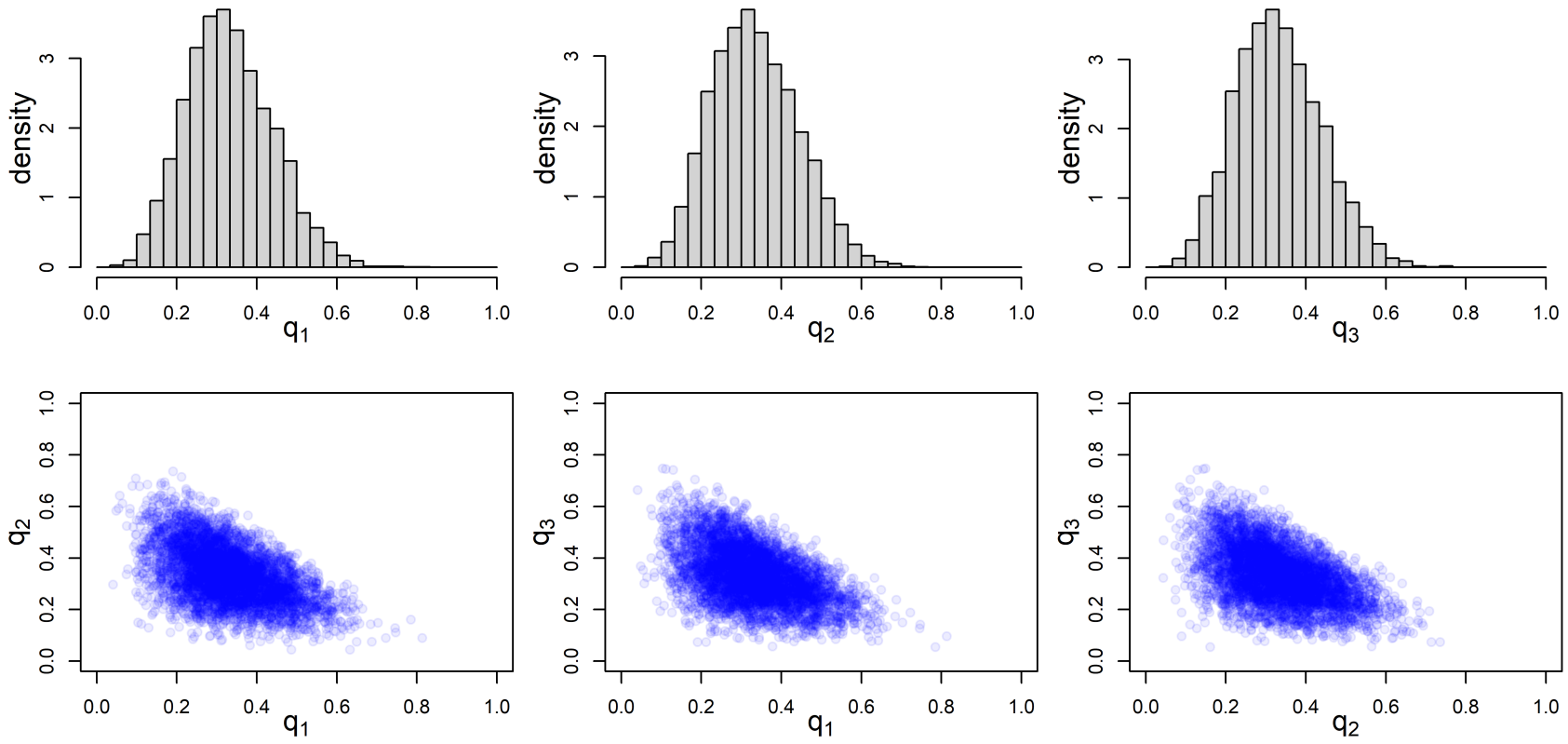
The Dirichlet distribution, as a prior for $\mathbf{q}$

Plotting one or two elements of $\mathbf{q}$ from Dirichlet(1,1,1), with mean (1/3,1/3,1/3):



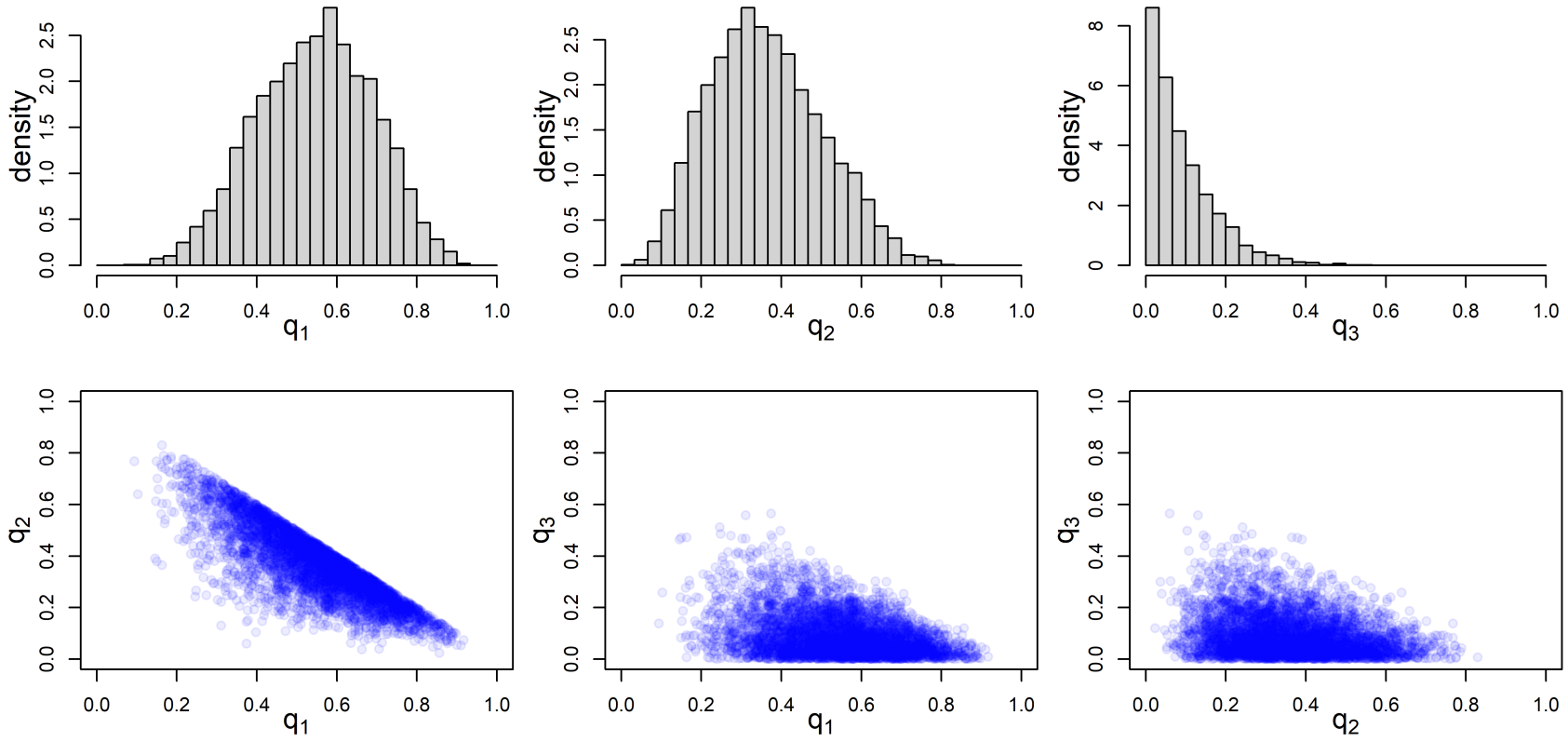
The Dirichlet distribution, as a prior for $\mathbf{q}$

Plotting one or two elements of $\mathbf{q}$ from Dirichlet(6,6,6), with mean (1/3,1/3,1/3):



The Dirichlet distribution, as a prior for $\mathbf{q}$

And from Dirichlet(6,4,1), with mean (6/11,4/11,1/11) $\approx$ (0.55, 0.36, 0.09):



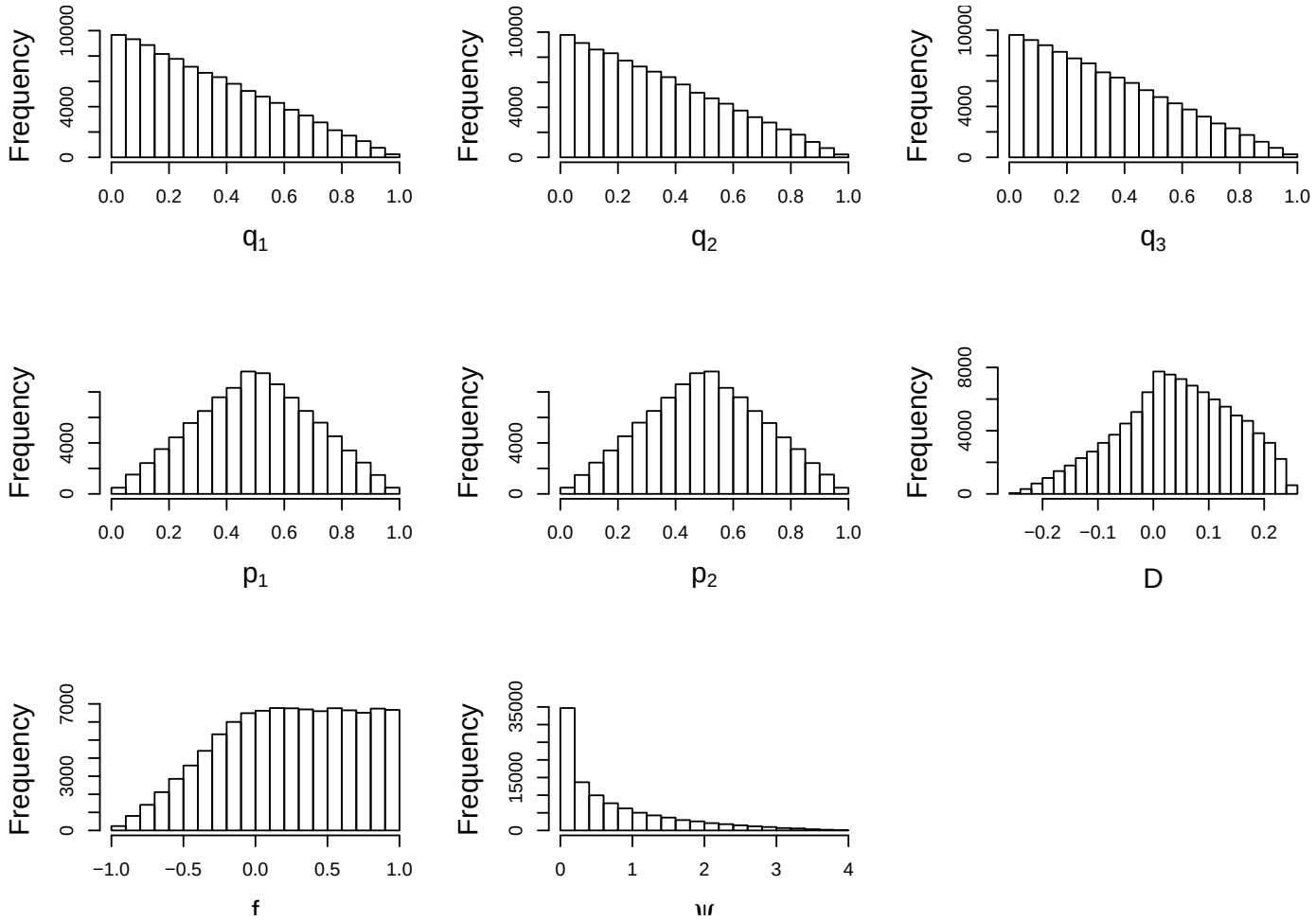
The Dirichlet distribution, as a prior for $\mathbf{q}$

While helpful, 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 $\mathbf{q}$, Dirichlet$(1, 1, 1)$, does not correspond to a flat prior for D, f, ψ, as the next slide shows
- With a 'flat' Dirichlet(1,1,1) prior the prior probability that $f > 0$ is 2/3.

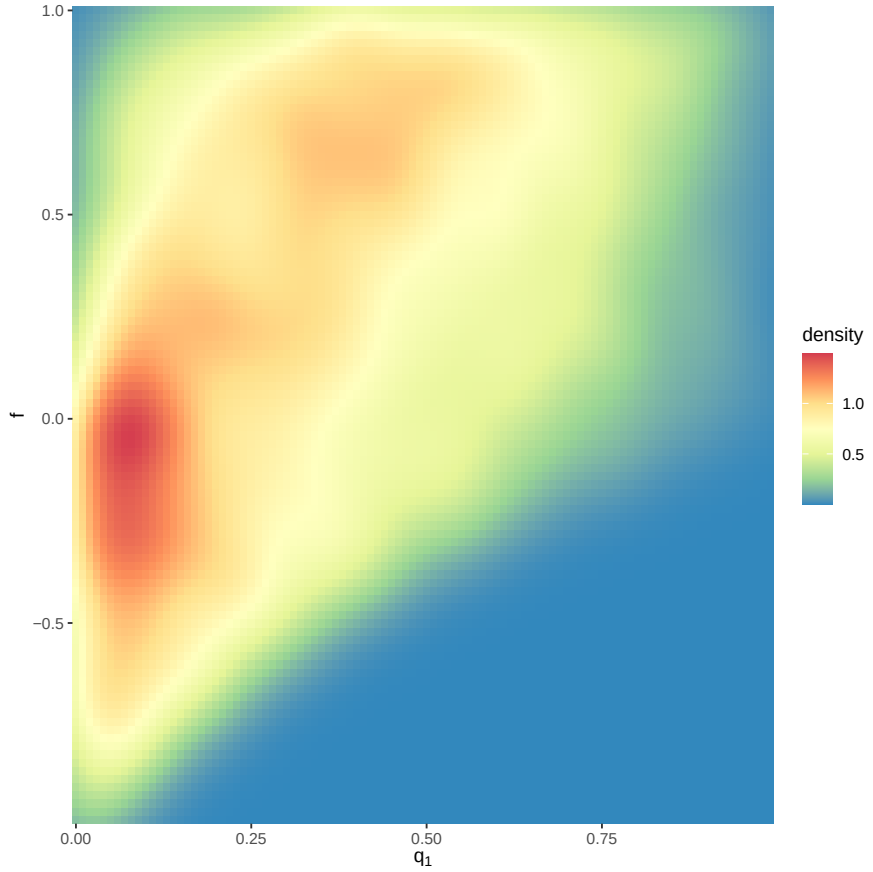
The Dirichlet distribution, as a prior for $\mathbf{q}$

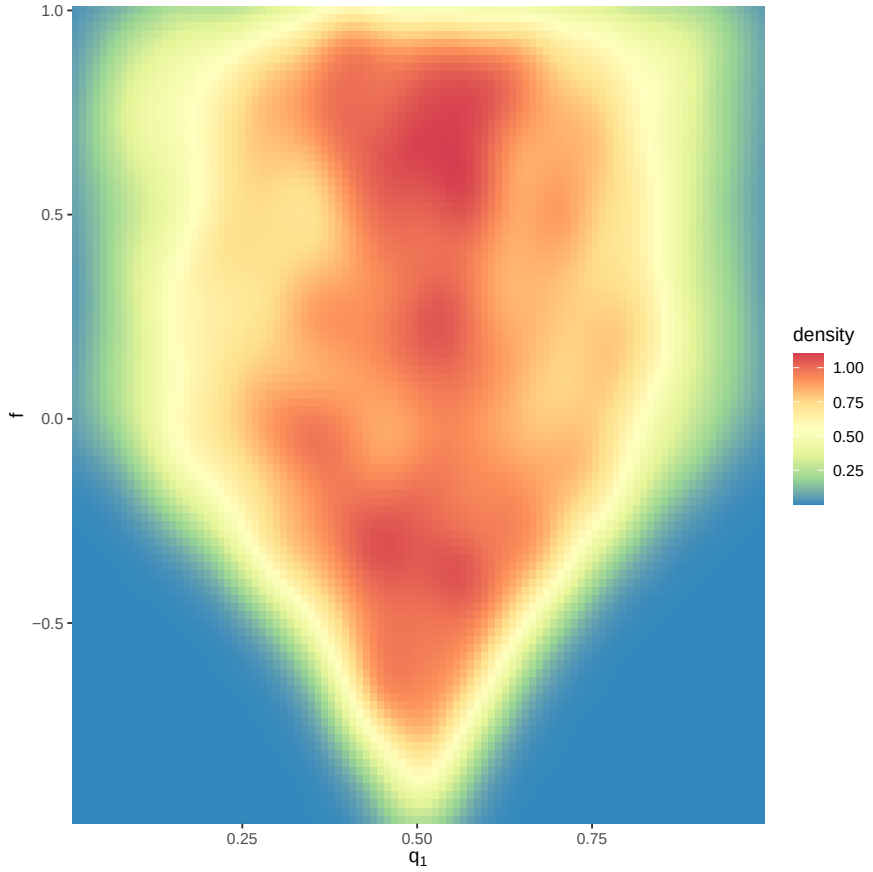
Samples from a Dirichlet(1,1,1), for various functions of $\mathbf{q}$:



The Dirichlet distribution, as a prior for $\mathbf{q}$

Image plots of $\{q_1, f\}$ and $\{p_1, f\}$ from a Dirichlet(1,1,1)





Posterior Distribution

Combining the Dirichlet (v_1, v_2, v_3), with the multinomial likelihood, conjugacy gives us the posterior:

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

which we recognize as another Dirichlet:

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

Just like the beta prior/binomal likelihood, this behaves *as if* we had observed counts $(n_1 + v_1, n_2 + v_2, n_3 + v_3)$.

Choosing a Prior

- Recall the prior mean is

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

- The posterior mean for the expected proportion of counts in cell i is

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

  where $n = n_1 + n_2 + n_3$, $v = v_1 + v_2 + v_3$ and $i = 1, 2, 3$.
- The weight $\text{W} = \frac{n}{n+v}$ is the proportion of the total information $(n + v)$ that is contributed by the data (n), versus that from the prior
- These forms help to choose v_1, v_2, v_3.

Choosing a Prior

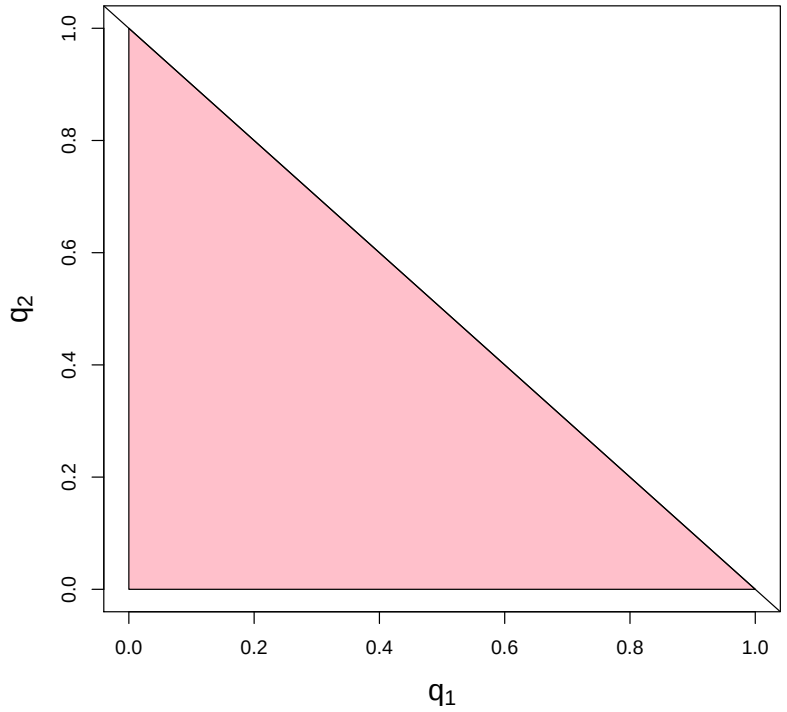
- As with the beta distribution we can specify the prior mean, and the relative weight that the prior and data contribute: n and v are on a comparable scale
- For example, suppose we believe that event 1 is four times as likely as each of event 2 or event 3. Then we specify the means in the ratios 4:1:1.
- 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

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

Choosing a Prior

An obvious choice of parameters is $v_1 = v_2 = v_3 = 1$ to give a prior that is uniform over the *simplex*: (but not over all parameters, as we've seen)

$$\pi(q_1, q_2, q_3) = 2, \quad \text{for } 0 < q_1, q_2, q_3 < 1, \text{ and } q_1 + q_2 + q_3 = 1.$$



Otters again

- The data is $n_1 = 37, n_2 = 20, n_3 = 7$.
- 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$(37+1, 20+1, 7+1)$ with posterior means:

$$\begin{aligned}
\mathbb{E}[q_1|\mathbf{n}] &= \frac{1+37}{3+64} = \frac{38}{67} \\
\mathbb{E}[q_2|\mathbf{n}] &= \frac{1+20}{3+64} = \frac{21}{67} \\
\mathbb{E}[q_3|\mathbf{n}] &= \frac{1+7}{3+64} = \frac{8}{67}.
\end{aligned}$$

- Note the similarity to the MLE

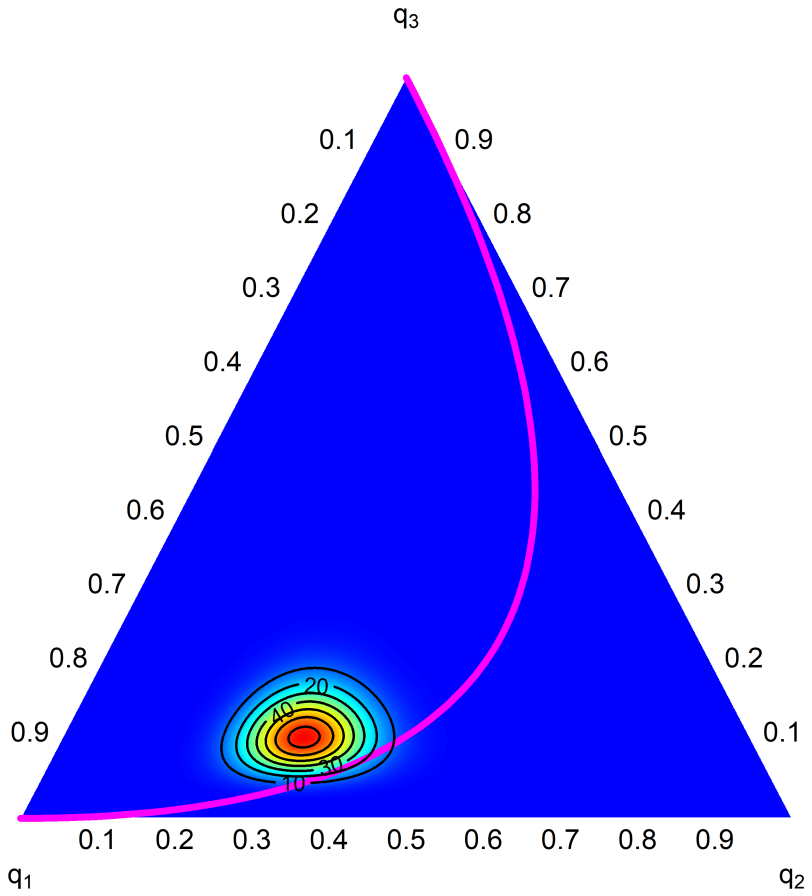
$$\hat{\mathbf{q}} = \left(\frac{37}{64}, \frac{20}{64}, \frac{7}{64}\right).$$

Otters again

The joint posterior, on a ternary plot:

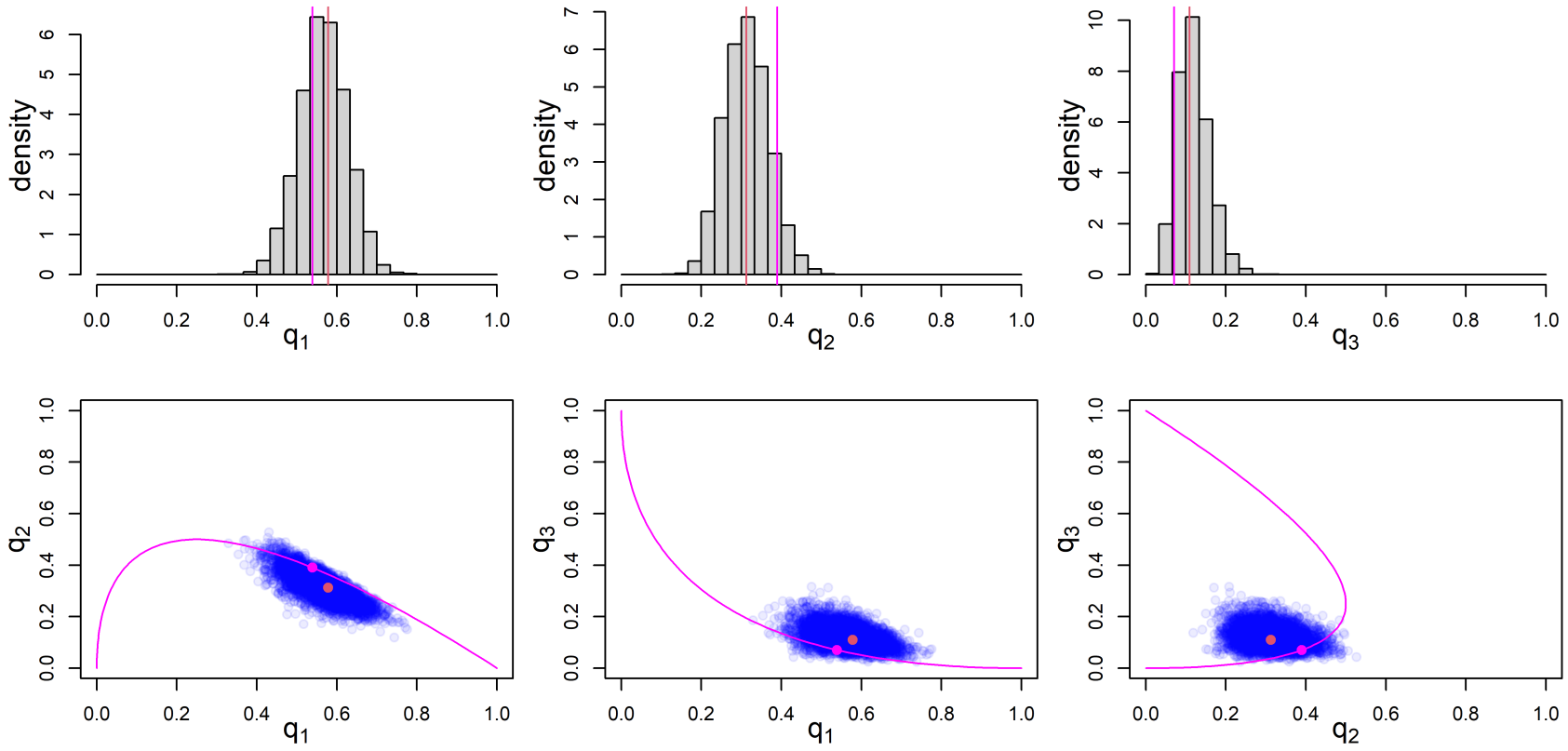
The magenta line shows all parameter values where exact HWE holds – so strong support for being at least *close* to HWE.

Note: an approximate frequentist 95% confidence region is bounded by the contour $\times 20$ **lower** than the likelihood's peak.



Otters again

Summaries of 1 and 2 parameters, with **MLE**, and **MLE under HWE**.



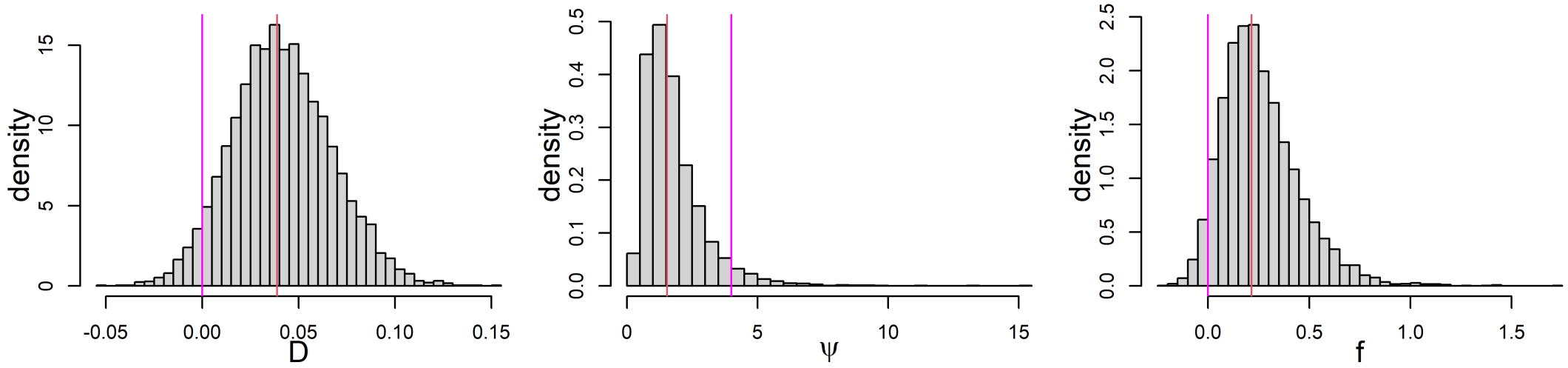
Otters again

Notes:

- As expected with a sample size of $n = 64$ and a flat prior (broadly equivalent to $n = 3$), the MLE is in the posterior's high support region
- The posterior is a little asymmetric, and its contours are not (quite) ellipses
- Asymptotic confidence intervals/regions (e.g. $\widehat{q}_i \pm 1.96 \times \text{se}(\widehat{q}_i)$, or dropping down $\times 20$ from the likelihood peak) would rely on ≈symmetry & ≈elliptical contours in the likelihood
- Credible intervals/regions are 'exact', in the sense of exactly summarizing the posterior. This differs from 'exact' frequentist coverage, or control of T1ER under the null

Otters again

For the more complex 1D measures of HWE violation:



- Again, there are no closed forms for these densities
- In all cases, $\times 2$ the minimum tail area (a Bayesian analog of the two-sided p-value) is 0.096, from 5000 posterior samples. (Similar to exact test's p=0.11)
- Bayes Factors are available in packaged code — note results are sensitive to the priors on 'nuiscance' parameters, a known issue with Bayes Factors

INLA

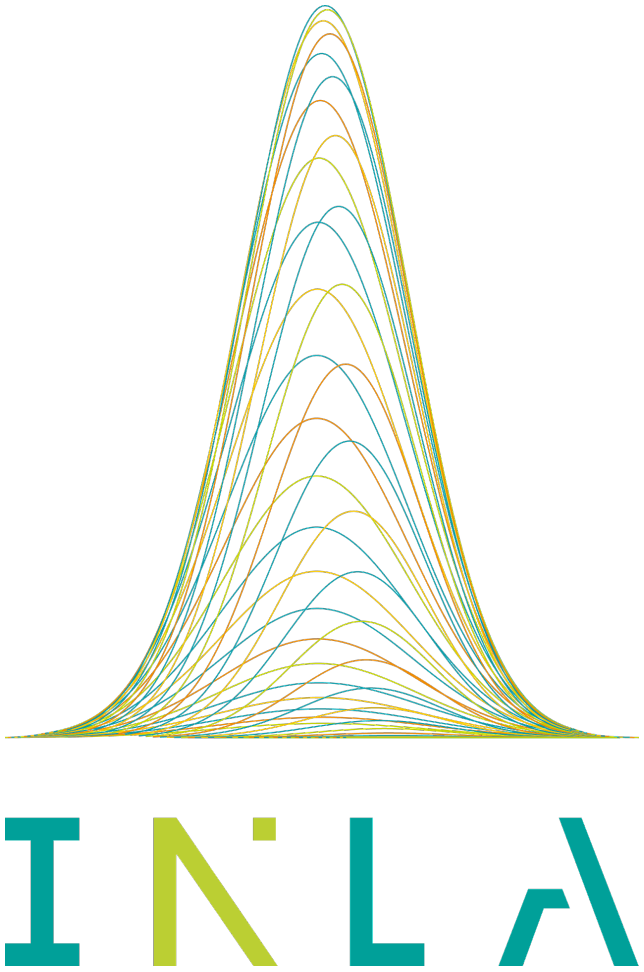
To calculate posteriors, so far we've seen

- Conjugate analyses: particularly priors for particular models give posteriors without any special calculation
- Direct sampling: knowing the posterior and sampling from it (particularly convenient for functions of parameters)
- Rejection sampling: Sampling points from the prior and (perhaps!) accepting them, based on likelihood **known [only] up to a constant**
- Approximations: e.g. assuming point estimate and standard error give a $\approx$Normal likelihood – formally known as *Laplace approximation*

For many widely-used models (i.e. those used in linear regression, logistic regression, Poisson regression, linear mixed models etc) *Integrated Nested Laplace Approximation* provides a a sophisticated **and fast** way to get the posterior.

INLA method

- INLA is provided via a non-CRAN R package, at http://www.r-inla.org/home
- No version for R 4.3 yet, current version is for R 4.2
- It's a large download! With many dependencies – updating can be an issue
- The INLA site has many examples, FAQs, other useful material
- The method is becoming increasingly popular as a Bayesian computational tool, in large part because of the package



INLA method

Warning: math ahead, but it's optional

- INLA combines Laplace approximations and numerical integration in a very efficient manner – first introduced by Rue *et al* (2009)
- The method is designed for *latent Gaussian models* (LGMs), i.e. models with Normal likelihoods and/or priors, but this is a huge set
- Suppose the model has the form

$$\begin{aligned} y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_1 &\sim p(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_1) \qquad \text{(Likelihood Function)} \\ \boldsymbol{x}|\boldsymbol{\theta}_2 &\sim N(\mathbf{0}, \boldsymbol{Q}(\boldsymbol{\theta}_2)^{-1}) \end{aligned}$$

  where $\boldsymbol{x}$ denotes a vector of variables with normal priors, for example, regression coefficients and random effects and $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are variance components.

INLA method

- We also have a prior, $\pi(\boldsymbol{\theta})$, for $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2]$ — non-normal, because the variance component parameters have to be positive (among other constraints)
- The posterior has the form:

$$\begin{aligned}
\pi(\boldsymbol{x}, \boldsymbol{\theta} \mid \boldsymbol{y}) &\propto \pi(\boldsymbol{\theta}) \pi(\boldsymbol{x} \mid \boldsymbol{\theta}_2) \prod_i p(y_i \mid \boldsymbol{x}_i, \boldsymbol{\theta}_1) \\
&\propto \pi(\boldsymbol{\theta}) \mid \boldsymbol{Q}(\boldsymbol{\theta}_2) \mid^{p/2} \exp\left\{ -\frac{1}{2} \boldsymbol{x}^T \boldsymbol{Q}(\boldsymbol{\theta}_2) \boldsymbol{x} + \sum_i \log p(\boldsymbol{y}_i \mid \boldsymbol{x}_i, \boldsymbol{\theta}_1) \right\}
\end{aligned}$$

INLA method

INLA calculates the *univariate posterior's marginals*:

$$\begin{aligned}
\pi(\theta_j|\mathbf{y}) &= \int\int \pi(\mathbf{x},\boldsymbol{\theta}|\mathbf{y})\ d\mathbf{x}d\boldsymbol{\theta}_{-j} \\
&= \int \pi(\boldsymbol{\theta}|\mathbf{y})\ d\boldsymbol{\theta}_{-j} \\
\pi(x_i|\mathbf{y}) &= \int\int \pi(\mathbf{x},\boldsymbol{\theta}|\mathbf{y})\ d\mathbf{x}_{-i}d\boldsymbol{\theta} \\
&= \int \left[\int \pi(x_i,\mathbf{x}_{-i}|\boldsymbol{\theta},\mathbf{y})d\mathbf{x}_{-i}\right] \pi(\boldsymbol{\theta}|\mathbf{y})\ d\boldsymbol{\theta} = \int \pi(x_i|\boldsymbol{\theta},\mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y})\ d\boldsymbol{\theta}
\end{aligned}$$

The *latent field* $\mathbf{x}$ and the variance components $\boldsymbol{\theta}$ are treated differently by INLA, because the latter are less normal-like in general, even after reparameterization.

The *nested* part of INLA reflects that given values of $\boldsymbol{\theta}$ Laplace approximations are carried out for $\mathbf{x}$, and these are averaged over using numerical integration techniques.

INLA method: calculating posteriors

We now describe the various approximations used in INLA.

The marginal posterior for $\boldsymbol{\theta}$ is, for any value of $\mathbf{x}$,

$$\begin{aligned}\pi(\boldsymbol{\theta}|\mathbf{y}) &= \frac{\pi(\mathbf{x},\boldsymbol{\theta}|\mathbf{y})}{\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})} \\ &\propto \frac{p(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})}\end{aligned}$$

The numerator is available, while the denominator is in general not. The Laplace approximation instead uses

$$\widehat{\pi}(\boldsymbol{\theta}^k|\mathbf{y}) \propto \frac{p(\mathbf{y}|\mathbf{x},\boldsymbol{\theta}^k)p(\mathbf{x}|\boldsymbol{\theta}^k)\pi(\boldsymbol{\theta}^k)}{\widehat{\pi}_G(\mathbf{x}|\boldsymbol{\theta}^k,\mathbf{y})}$$

where $\widehat{\pi}_G(\mathbf{x}|\boldsymbol{\theta}^k,\mathbf{y})$ is the Gaussian approximation to the conditional which is obtained by matching the mode and the curvature at the mode.

INLA method: calculating posteriors

The marginal $\pi(x_i|\mathbf{y})$ needs to be calculated for a potentially very long vector $\mathbf{x}$. We *could* take the marginal from $\widehat{\pi}_G(\mathbf{x}|\boldsymbol{\theta}^k, \mathbf{y})$ but this is generally not very accurate.

As an alternative, rewrite as

$$\begin{aligned} \pi(x_i|\mathbf{y}) &= \frac{\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})}{\pi(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} \\ &\propto \frac{p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})p(\mathbf{x}|\boldsymbol{\theta})\pi(\mathbf{x}, \boldsymbol{\theta})}{\pi(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} \end{aligned}$$

and the denominator can again be estimated estimated using a density approximation due to Tierney & Kadane (1986).

INLA method: calculating posteriors

Rue *et al* (2009) describe a third approximation, the *simplified Laplace* which corrects the Gaussian approximation for location and skewness using a Taylor series about the mode. INLA's algorithm (Martino & Riebler 2019) consists of

1. **Explore** the $\boldsymbol{\theta}$ space via the approximation $\widehat{\pi}(\boldsymbol{\theta}^k|\mathbf{y})$. Specifically, find the mode of $\widehat{\pi}(\boldsymbol{\theta}^k|\mathbf{y})$ and identify a set of points $\{\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^K\}$ in the areas of high density.
2. **Compute** $\widehat{\pi}(\boldsymbol{\theta}^k|\mathbf{y})$ for k=1... K, using the denominator approximation above
3. **Calculate** $\widehat{\pi}(x_i|\boldsymbol{\theta}^k, \mathbf{y})$ for $k = 1, \ldots, K$ using one of Gaussian, Laplace, simplified Laplace.
4. Use **numerical integration** to approximate the marginal,

$$\widehat{\pi}(x_i|\mathbf{y}) = \sum_{k=1}^{K} \widehat{\pi}(x_i|\boldsymbol{\theta}^k, \mathbf{y}) \times \widehat{\pi}(\boldsymbol{\theta}^k|\mathbf{y})\Delta_k,$$

using points and weights $\{\boldsymbol{\theta}^k, \Delta_k, k = 1, \ldots, K\}$.

Exploring the θ space

First, a "good" parameterization is found (often this is achieved by simply transforming to the real line), we assume that $\boldsymbol{\theta}$ satisfies this; also let $\dim(\boldsymbol{\theta}) = m$.

Second, find the mode, $\boldsymbol{\theta}^\star$, and the Hessian matrix $\mathbf{H}$; let $\mathbf{H}^{-1} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}$ be the eigen decomposition, then form the new standardized variable:

$$\mathbf{z} = (\mathbf{V}\boldsymbol{\Lambda}^{1/2})^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}^\star),$$

which adjusts for location, scale, and rotation.

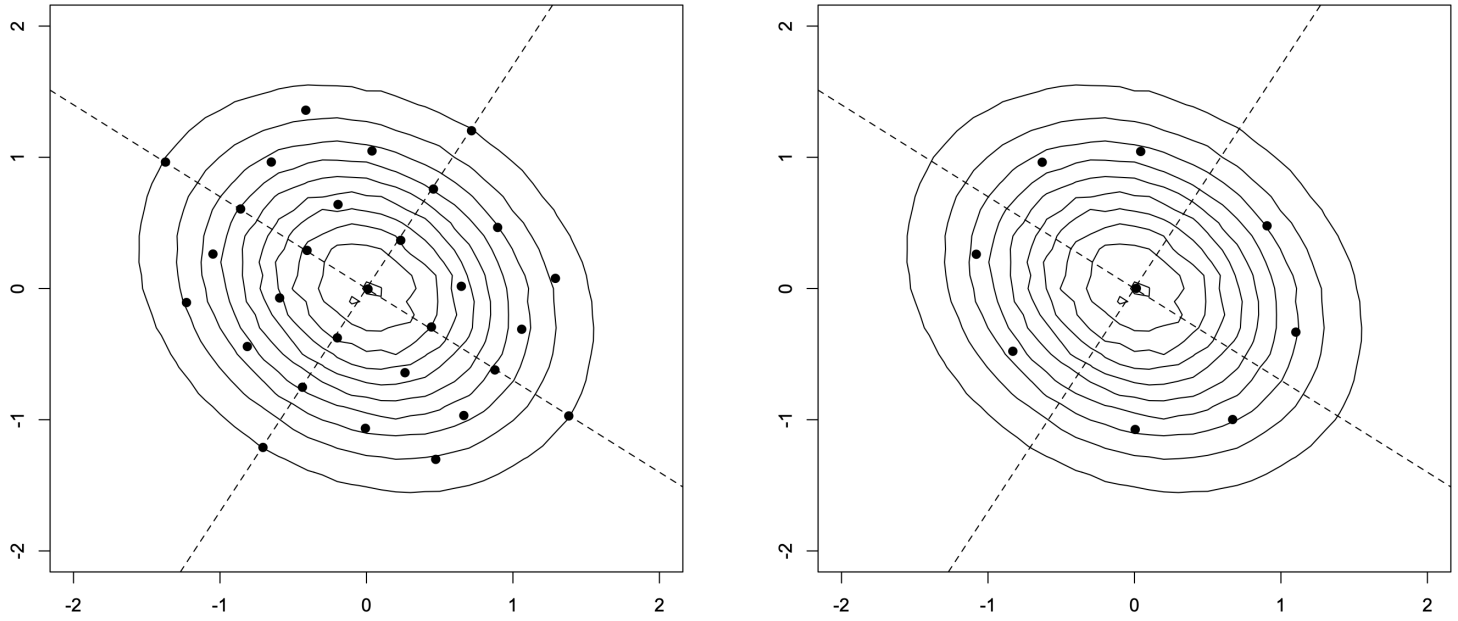
Exploring the θ space

Rue *et al* (2009) describe three methods for exploration:

- grid: This approach builds a grid for the standardized variable $\mathbf{z}$. Unfortunately the number of points grows exponentially with m; if we use p points in each dimension, p^m are required in total
- empirical Bayes: just take the posterior mode only, i.e., a single point
- CCD: use a classical design, specifically the central composite design (CCD) – integration points are placed on spheres

Grid versus CCD

Grid (left) and CCD (right) points for numerical integration, from Wang *et al*'s free book.



INLA: Posterior sampling

Marginals are the standard output of INLA, but various operations may be carried out using the functions:

- `inla.dmarginal` for density values
- `inla.pmarginal` for the CDF
- `inla.qmarginal` for quantiles
- `inla.rmarginal` for random samples
- `inla.hpdmarginal` for highest posterior density (HPD) credible regions
- `inla.emarginal` computes the expected values of a function of a parameter
- `inla.tmarginal` calculates the marginal distribution of a transformation of a latent variable or hyperparameter.

INLA: Practical Advice

Some functionals cannot be obtained using these functions, so samples may be drawn from an approximation to the posterior*, and manipulated:

- `inla.posterior.sample()` draws samples from the approximate posterior distribution of β and θ.
- To make use of this function, use `control.compute = list(config = TRUE)` in the INLA model fit.
- Included in the arguments is `selected` which allows only specific components to be sampled.
- In general, the returned sample contains

```
"hyperpar" "latent" "logdens"
```

*for the latent field $\mathbf{x}$ we sample from a mixture of multivariate Gaussians, where the weights correspond to the integration weights (for the grid and CCD options).

INLA: LHON again

We return to the LHON example, analyzed in Session 3 with rejection sampling.

```
# setup data
cc.dat <- data.frame(x=c(0,1,2), success=c(6,8,75), fail=c(10,66,163))

# non-Bayes analysis
logitmod <- glm(cbind(success,fail)~x,family="binomial", data=cc.dat)
coef(summary(logitmod))
             Estimate Std. Error   z value     Pr(>|z|)
(Intercept) -1.8076928  0.4553938 -3.969515 7.201898e-05
x            0.4787428  0.2504594  1.911459 5.594568e-02
confint.default(logitmod)
                  2.5 %     97.5 %
(Intercept) -2.70024830 -0.9151373
x           -0.01214865  0.9696342
```

INLA: LHON again

Recall we set up the diffuse prior to have 95% point log(5): telling INLA about this and getting the posterior:

```
Upper95 <- log(5)
sigma <- Upper95/qnorm(0.95)
cc.inla <- inla(success~x,family="binomial",data=cc.dat,Ntrials=success+fail,
                control.fixed=list(mean.intercept=c(0),prec.intercept=c(1/10),
                mean=c(0),prec=c(1/sigma^2)))
summary(cc.inla)
        Time used:    Pre = 0.236, Running = 0.107, Post = 0.0166, Total = 0.359
Fixed effects:
              mean    sd 0.025quant 0.5quant 0.975quant   mode kld
(Intercept) -1.760 0.431     -2.605   -1.760     -0.916 -1.760   0
x            0.449 0.237     -0.016    0.449      0.914  0.449   0
```

- Non-Bayes gave 0.48 (-0.01, 0.97), here we get 0.45 (-0.02, 0.91), rejection sampling with $B = 50,000$ samples gives 0.45 (-0.01, 0.93).
- The `kld` column indicates a distance between the posterior approximated in two ways, i.e. they agree.

INLA: LHON again

And for the more informative prior:

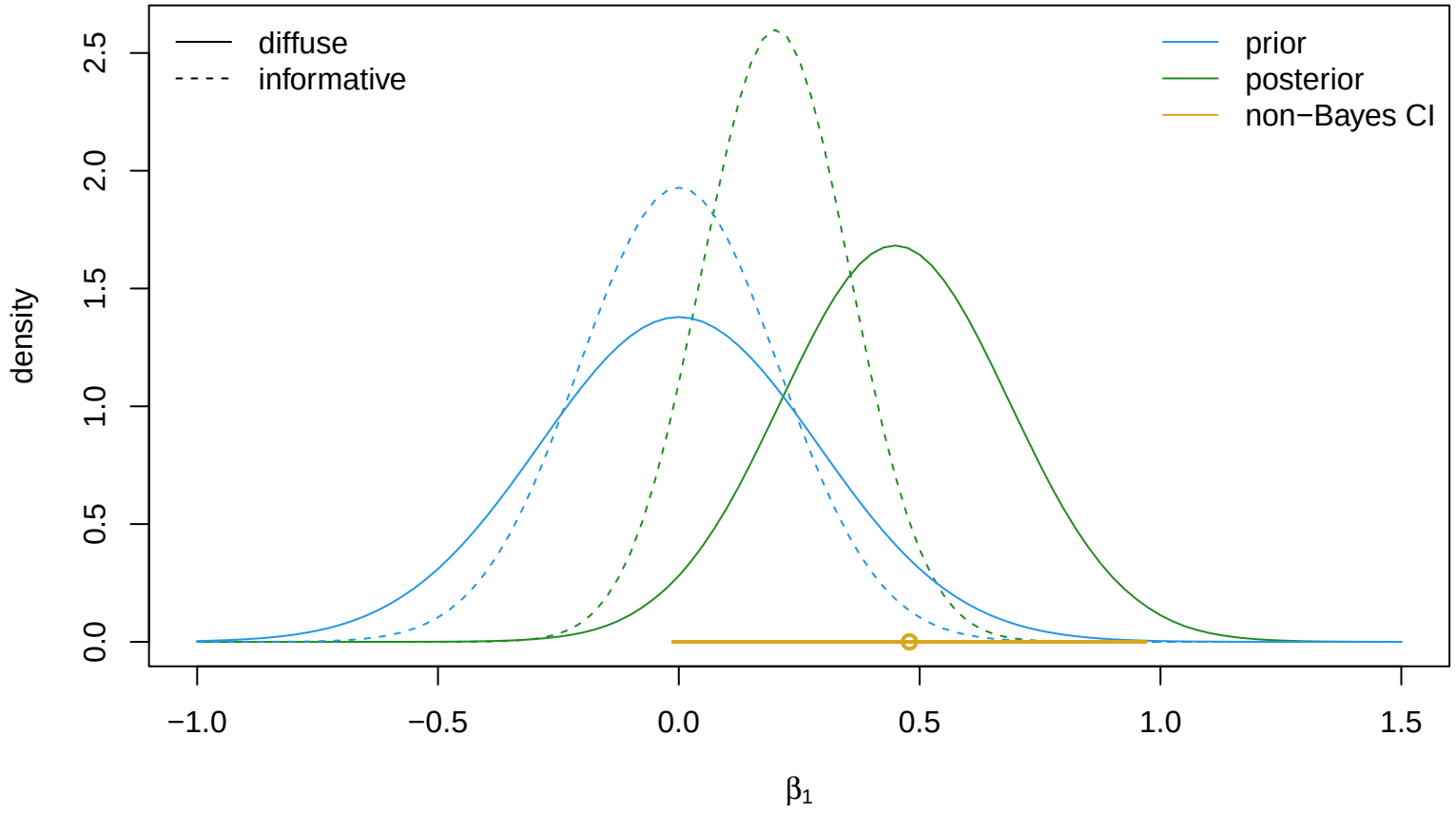
```
Upper975 <- 1.5
sigma <- log(Upper975)/qnorm(0.975)
cc.inf.inla < inla(success~x,family="binomial",data=cc.dat,Ntrials=success+fail,
            control.fixed=list(mean.intercept=c(0),prec.intercept=c(1/10),
            mean=c(0),prec=c(1/sigma^2)))

summary(cc.inf.inla)
Time used:
    Pre = 0.269, Running = 0.19, Post = 0.0167, Total = 0.476
Fixed effects:
              mean    sd 0.025quant 0.5quant 0.975quant   mode kld
(Intercept) -1.332 0.290     -1.899   -1.332     -0.764 -1.332   0
x            0.201 0.154     -0.100    0.201      0.502  0.201   0
```

- Here we get 0.20 (-0.10, 0.50) for the log odds ratio, versus rejection sampling's 0.20 (-0.09, 0.51)
- The Monte Carlo error/approximation error are **massively** smaller than uncertainty due to the limited sample size. Does 0.50 vs 0.51 matter?

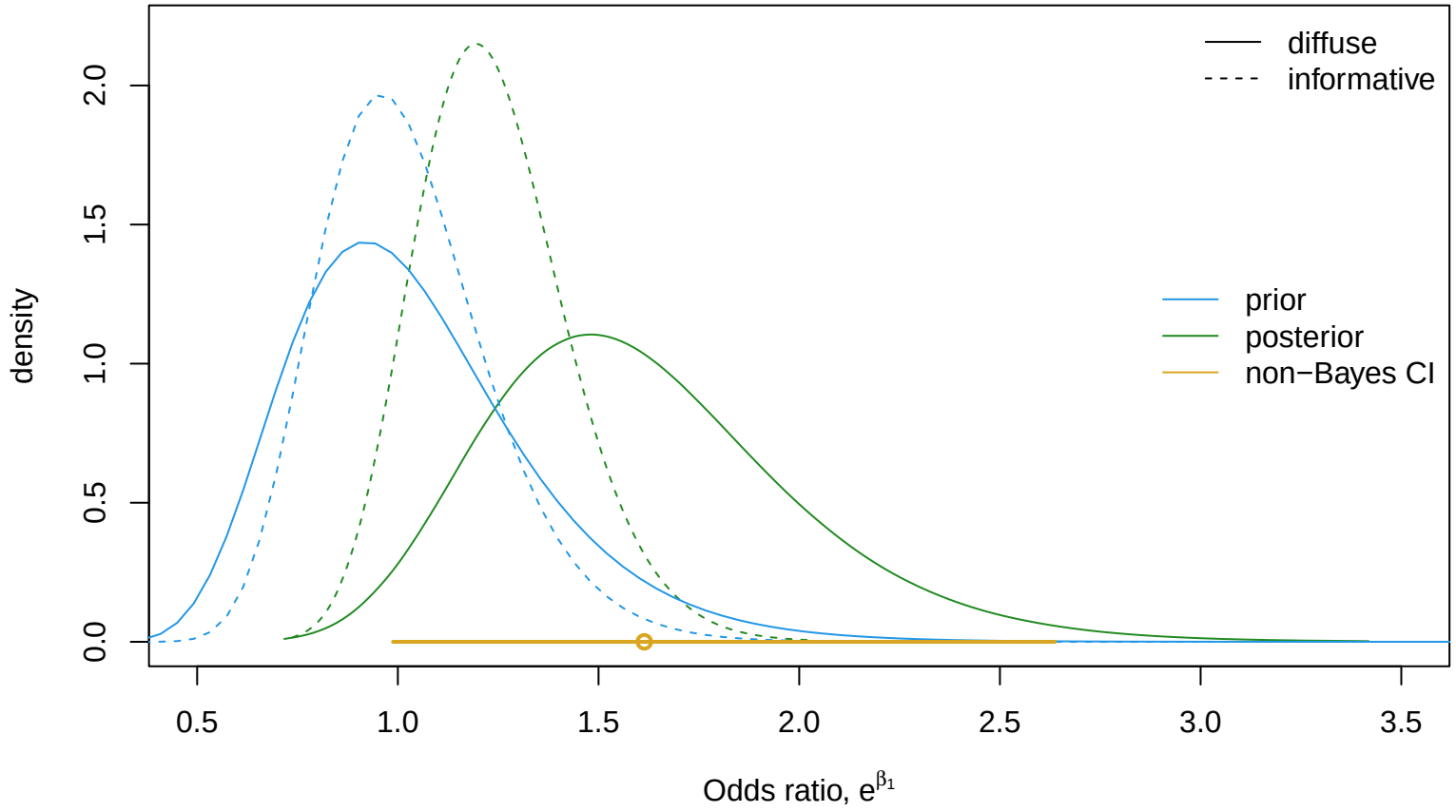
INLA: LHON again

The `inla.dmarginal()` function can be used with e.g. `curve()` to plot posteriors:



INLA: LHON again

And using `inla.tmarginal(fun=exp)` to get odds ratio's posterior, not the log OR;



Pros and Cons of INLA

Advantages:

- Quite widely applicable: Generalized Linear Mixed Models (GLMMs) including temporal and spatial error terms – many book-length treatments now available
- Very fast — enabling bootstrapping, leave-one-out, etc
- Works from within R

Disadvantages:

- Restricted to models with Gaussian random effects – Template Model Builder is more flexible, but the `TMB` package needs you to write your own C++
- Spotting with INLA's approximation fails takes experience, but though lots of empirical evidence is being gathered

Summary

- With higher-dimensional parameters – such as compositional vectors modeling options become more limited, and priors harder to think about
- INLA provides a hugely flexible system for evaluating posteriors. Understanding exactly what was done is more work than e.g. rejection sampling
- INLA's speed makes sensitivity analysis (to the prior, or individual data points) much more plausible