Objective Bayesian Hypothesis Testing

José M. Bernardo
Universitat de València, Spain
jose.m.bernardo@uv.es

Statistical Science and Philosophy of Science
London School of Economics (UK), June 21st, 2010

Summary

(i) **Hypothesis testing: Foundations**

(ii) **Bayesian Inference Summaries**

(iii) **Loss Functions**

(iv) **Objective Bayesian Methods**

(v) **Integrated Reference Analysis**

(vi) **Basic References**

Hypothesis Testing: Foundations

Time to revise foundations?

- No obvious agreement on the appropriate solution to even simple (textbook) stylized problems:

  Testing compatibility of the normal mean with a precise value

  Comparing two normal means or two binomial proportions

- Let alone in more complex problems:

  Testing a population data for Hardy-Weingerg equilibrium

  Testing for independence in contingency tables

  Our proposal: Use Bayesian decision-theoretic machinery with reference (objective) priors.

Bayesian Inference Summaries

- Assume data $\boldsymbol{z}$ have been generated as one random observation form $\mathcal{M}_{\boldsymbol{z}} = \{p(\boldsymbol{z} \mid \boldsymbol{\theta}, \boldsymbol{\lambda}), \boldsymbol{z} \in \boldsymbol{\mathcal{Z}}, \boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{\lambda} \in \boldsymbol{\Lambda}\}$, where $\boldsymbol{\theta}$ is the vector of interest and $\boldsymbol{\lambda}$ a nuisance parameter vector.
- Assume a joint prior $p(\boldsymbol{\theta}, \boldsymbol{\lambda}) = p(\boldsymbol{\lambda} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})$ (more later).
- Given data $\boldsymbol{z}$, model $\mathcal{M}_{\boldsymbol{z}}$ and prior $p(\boldsymbol{\theta}, \boldsymbol{\lambda})$, the complete solution to all inference questions about $\boldsymbol{\theta}$ is contained in the marginal posterior $p(\boldsymbol{\theta} \mid \boldsymbol{z})$, derived by standard use of probability theory.
- Appreciation of $p(\boldsymbol{\theta} \mid \boldsymbol{z})$ may be enhanced by providing both point and region estimates of the vector of interest $\boldsymbol{\theta}$, and by declaring whether or not some context-suggested specific value $\boldsymbol{\theta}_0$ (or maybe a set of values $\boldsymbol{\Theta}_0$), is (are) compatible with the observed data $\boldsymbol{z}$. All of these provide useful (and often required) summaries of $p(\boldsymbol{\theta} \mid \boldsymbol{z})$.

Decision-theoretic structure

- All these summaries may be framed as different decision problems which use precisely the same loss function $\ell\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\}$ describing, as a function of the (unknown) $(\boldsymbol{\theta}, \boldsymbol{\lambda})$ values which have generated the data, the loss to be suffered if, working with model $\mathcal{M}_{\boldsymbol{z}}$, the value $\boldsymbol{\theta}_0$ were used as a proxy for the unknown value of $\boldsymbol{\theta}$.

- The results dramatically depend on the choices made for both the prior and the loss functions but, given $\boldsymbol{z}$, only depend on those through the expected loss, $\overline{\ell}(\boldsymbol{\theta}_0 \,|\, \boldsymbol{z}) = \int_{\boldsymbol{\Theta}} \int_{\boldsymbol{\Lambda}} \ell\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\}\, p(\boldsymbol{\theta}, \boldsymbol{\lambda} \,|\, \boldsymbol{z})\, d\boldsymbol{\theta} d\boldsymbol{\lambda}$.

- As a function of $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, $\overline{\ell}(\boldsymbol{\theta}_0 \,|\, \boldsymbol{z})$ is a measure of the unacceptability of all possible values of the vector of interest. This provides a dual, complementary information on all $\boldsymbol{\theta}$ values (on a loss scale) to that provided by the posterior $p(\boldsymbol{\theta} \,|\, \boldsymbol{z})$ (on a probability scale).

Point estimation

To choose a point estimate for $\boldsymbol{\theta}$ is a decision problem where the action space is the class $\boldsymbol{\Theta}$ of all possible $\boldsymbol{\theta}$ values.

Definition 1 *The* Bayes estimator $\boldsymbol{\theta}^*(\boldsymbol{z}) = \arg\inf_{\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}} \overline{\ell}(\boldsymbol{\theta}_0 \,|\, \boldsymbol{z})$ *is that which minimizes the posterior expected loss.*

- Conventional examples include the ubiquitous quadratic loss $\ell\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\} = (\boldsymbol{\theta}_0 - \boldsymbol{\theta})^t(\boldsymbol{\theta}_0 - \boldsymbol{\theta})$, which yields the posterior mean as the Bayes estimator, and the zero-one loss on a neighborhood of the true value, which yields the posterior mode a a limiting result.
- Bayes estimators with conventional loss functions are typically not invariant under one to one transformations. Thus, the Bayes estimator under quadratic loss of a variance s not the square of the Bayes estimator of the standard deviation. This is rather difficult to explain when one merely wishes to report an estimate of some quantity of interest.

Region estimation

Bayesian region estimation is achieved by quoting posterior credible regions. To choose a q-credible region is a decision problem where the action space is the class of subsets of $\boldsymbol{\Theta}$ with posterior probability q.

Definition 2 *(Bernardo, 2005). A* Bayes q-credible region $\boldsymbol{\Theta}^*_q(\boldsymbol{z})$ *is a q-credible region where any value within the region has a smaller posterior expected loss than any value outside the region, so that* $\forall \boldsymbol{\theta}_i \in \boldsymbol{\Theta}^*_q(\boldsymbol{z}),\ \forall \boldsymbol{\theta}_j \notin \boldsymbol{\Theta}^*_q(\boldsymbol{z}),\quad \bar{\ell}(\boldsymbol{\theta}_i \,|\, \boldsymbol{z}) \leq \bar{\ell}(\boldsymbol{\theta}_j \,|\, \boldsymbol{z})$.

- The quadratic loss yields credible regions with those $\boldsymbol{\theta}$ values closest, in the Euclidean sense, to the posterior mean. A zero-one loss function leads to highest posterior density (HPD) credible regions.
- Conventional Bayes regions are often not invariant: HPD regions in one parameterization will not transform to HPD regions in another.

□ Precise hypothesis testing

• Consider a value $\boldsymbol{\theta}_0$ which deserves special consideration. Testing the hypothesis $H_0 \equiv \{\boldsymbol{\theta} = \boldsymbol{\theta}_0\}$ is as a decision problem where the action space $\mathcal{A} = \{a_0, a_1\}$ contains only two elements: to accept (a_0) or to reject (a_1) the hypothesis H_0.

• Foundations require to specify the loss functions $\ell_h\{a_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\}$ and $\ell_h\{a_1, (\boldsymbol{\theta}, \boldsymbol{\lambda})\}$ measuring the consequences of accepting or rejecting H_0 as a function of $(\boldsymbol{\theta}, \boldsymbol{\lambda})$. The optimal action is to reject H_0 iif

$$\int_{\boldsymbol{\Theta}} \int_{\boldsymbol{\Lambda}} [\ell_h\{a_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\} - \ell_h\{a_1, (\boldsymbol{\theta}, \boldsymbol{\lambda})\}]\, p(\boldsymbol{\theta}, \boldsymbol{\lambda} \,|\, \boldsymbol{z})\, d\boldsymbol{\theta} d\boldsymbol{\lambda} > 0.$$

• Hence, only $\Delta\ell_h\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\} = \ell_h\{a_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\} - \ell_h\{a_1, (\boldsymbol{\theta}, \boldsymbol{\lambda})\}$, which measures the conditional advantage of rejecting, must be specified.

- Without loss of generality, the function $\Delta\ell_h$ may be written as

$$\Delta\ell_h\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\} = \ell\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\} - \ell_0$$

where (precisely as in estimation), $\ell\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda})\}$ describes, as a function of $(\boldsymbol{\theta}, \boldsymbol{\lambda})$, the non-negative loss to be suffered if $\boldsymbol{\theta}_0$ were used as a proxy for $\boldsymbol{\theta}$, and the constant $\ell_0 > 0$ describes (in the same loss units) the context-dependent non-negative advantage of accepting $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ when it is true.

Definition 3 *(Bernardo, 1999; Bernardo and Rueda, 2002). The* Bayes test criterion *to decide on the compatibility of* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ *with available data* $\boldsymbol{z}$ *is to reject* $H_0 \equiv \{\boldsymbol{\theta} = \boldsymbol{\theta}_0\}$ *if (and only if),* $\bar{\ell}(\boldsymbol{\theta}_0 \,|\, \boldsymbol{z}) > \ell_0$*, where* ℓ_0 *is a context dependent positive constant.*

- The compound case may be analyzed by separately considering each of the values which make part of the compound hypothesis to test.

• Using a zero-one loss function, so that the loss advantage of rejecting $\boldsymbol{\theta}_0$ is equal to one whenever $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ and zero otherwise, leads to rejecting H_0 if (and only if) $\Pr(\boldsymbol{\theta} = \boldsymbol{\theta}_0 \,|\, \boldsymbol{z}) < p_0$ for some context-dependent p_0. Use of this loss requires the prior probability $\Pr(\boldsymbol{\theta} = \boldsymbol{\theta}_0)$ to be *strictly positive*. If $\boldsymbol{\theta}$ is a continuous parameter this forces the use of a non-regular "sharp" prior, concentrating a positive probability mass at $\boldsymbol{\theta}_0$, the solution early advocated by Jeffreys.

This formulation (i) implies the use of radically different priors for hypothesis testing than those used for estimation, (ii) precludes the use of conventional, often improper, 'noninformative" priors, and (iii) may lead to the difficulties associated to Jeffreys-Lindley paradox.

• The quadratic loss function leads to rejecting a $\boldsymbol{\theta}_0$ value whenever its Euclidean distance to $\mathrm{E}[\boldsymbol{\theta} \,|\, \boldsymbol{z}]$, the posterior expectation of $\boldsymbol{\theta}$, is sufficiently large.

- The use of continuous loss functions (such as the quadratic loss) permits the use in hypothesis testing of precisely the same priors that are used in estimation.
- With conventional losses the Bayes test criterion is not invariant under one-to-one transformations. Thus, if $\boldsymbol{\phi}(\boldsymbol{\theta})$ is a one-to-one transformation of $\boldsymbol{\theta}$, rejecting $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ does not generally imply rejecting $\boldsymbol{\phi}(\boldsymbol{\theta}) = \boldsymbol{\phi}(\boldsymbol{\theta}_0)$.
- The threshold constant ℓ_0, which controls whether or not an expected loss is too large, is part of the specification of the decision problem, and should be context-dependent. However a judicious choice of the loss function leads to calibrated expected losses, where the relevant threshold constant has an immediate, operational interpretation.

Loss Functions

- A dissimilarity measure $\delta\{p_{\boldsymbol{z}}, q_{\boldsymbol{z}}\}$ between two probability densities $p_{\boldsymbol{z}}$ and $q_{\boldsymbol{z}}$ for a random vector $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$ should be
  (i) non-negative, and zero if (and only if) $p_{\boldsymbol{z}} = q_{\boldsymbol{z}}$ a.e.,
  (ii) invariant under one-to-one transformations of $\boldsymbol{z}$,
  (iii) symmetric, so that $\delta\{p_{\boldsymbol{z}}, q_{\boldsymbol{z}}\} = \delta\{q_{\boldsymbol{z}}, p_{\boldsymbol{z}}\}$,
  (iv) defined for densities with strictly nested supports.

Definition 4 *The intrinsic discrepancy* $\delta\{p_1, p_2\}$ *is*

$$\delta\{p_1, p_2\} = \min\left[\,\kappa\{p_1 \,|\, p_2\},\ \kappa\{p_2 \,|\, p_1\}\,\right]$$

where $\kappa\{p_j \,|\, p_i\} = \int_{\boldsymbol{\mathcal{Z}}_i} p_i(\boldsymbol{z}) \log[p_i(\boldsymbol{z})/p_j(\boldsymbol{z})]\, d\boldsymbol{z}$ *is the* (KL) *divergence of* p_j *from* p_i. *The intrinsic discrepancy between* p *and a family* $\mathcal{F} = \{q_i, i \in I\}$ *is the intrinsic discrepancy between* p *and the closest of them,* $\delta\{p, \mathcal{F}\} = \inf_{q,\in\mathcal{F}} \delta\{p, q\}$.

The intrinsic loss function

Definition 5 *Consider* $\mathcal{M}_{\boldsymbol{z}} = \{p(\boldsymbol{z} \,|\, \boldsymbol{\theta}, \boldsymbol{\lambda}), \boldsymbol{z} \in \boldsymbol{\mathcal{Z}}, \boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{\lambda} \in \boldsymbol{\Lambda}\}$. *The intrinsic loss of using* $\boldsymbol{\theta}_0$ *as a proxy for* $\boldsymbol{\theta}$ *is the intrinsic discrepancy between the true model and the class of models with* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, $\mathcal{M}_0 = \{p(\boldsymbol{z} \,|\, \boldsymbol{\theta}_0, \boldsymbol{\lambda}_0), \boldsymbol{z} \in \boldsymbol{\mathcal{Z}}, \boldsymbol{\lambda}_0 \in \boldsymbol{\Lambda}\}$,

$$\ell_\delta\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda}) \,|\, \mathcal{M}_{\boldsymbol{z}}\} = \inf_{\boldsymbol{\lambda}_0 \in \boldsymbol{\Lambda}} \delta\{p_{\boldsymbol{z}}(\cdot \,|\, \boldsymbol{\theta}, \boldsymbol{\lambda}), p_{\boldsymbol{z}}(\cdot \,|\, \boldsymbol{\theta}_0, \boldsymbol{\lambda}_0)\}.$$

□ Invariance

- For any one-to-one reparameterization $\boldsymbol{\phi} = \boldsymbol{\phi}(\boldsymbol{\theta})$ and $\boldsymbol{\psi} = \boldsymbol{\psi}(\boldsymbol{\theta}, \boldsymbol{\lambda})$,

$$\ell_\delta\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda}) \,|\, \mathcal{M}_{\boldsymbol{z}}\} = \ell_\delta\{\boldsymbol{\phi}_0, (\boldsymbol{\phi}, \boldsymbol{\psi}) \,|\, \mathcal{M}_{\boldsymbol{z}}\}.$$

This yields invariant Bayes point and region estimators, and invariant Bayes hypothesis testing procedures.

□ Reduction to sufficient statistics

• If $\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{z})$ is a sufficient statistic for model $\mathcal{M}_{\boldsymbol{z}}$, one may also work with marginal model $\mathcal{M}_{\boldsymbol{t}} = \{p(\boldsymbol{t} \,|\, \boldsymbol{\theta}, \boldsymbol{\lambda}), \boldsymbol{t} \in \boldsymbol{\mathcal{T}}, \boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{\lambda} \in \boldsymbol{\Lambda}\}$ since

$$\ell_\delta\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda}) \,|\, \mathcal{M}_{\boldsymbol{z}}\} = \ell_\delta\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda}) \,|\, \mathcal{M}_{\boldsymbol{t}}\}.$$

□ Additivity

• If data consist of a random sample $\boldsymbol{z} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ from some model $\mathcal{M}_{\boldsymbol{x}}$, so that $\boldsymbol{\mathcal{Z}} = \boldsymbol{\mathcal{X}}^n$, and $p(\boldsymbol{z} \,|\, \boldsymbol{\theta}, \boldsymbol{\lambda}) = \prod_{i=1}^n p(\boldsymbol{x}_i \,|\, \boldsymbol{\theta}, \boldsymbol{\lambda})$,

$$\ell_\delta\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda}) \,|\, \mathcal{M}_{\boldsymbol{z}}\} = n\, \ell_\delta\{\boldsymbol{\theta}_0, (\boldsymbol{\theta}, \boldsymbol{\lambda}) \,|\, \mathcal{M}_{\boldsymbol{x}}\}.$$

This considerably simplifies frequent computations.

Objective Bayesian Methods

- The methods described above may be used with any prior. However, an "objective" procedure, where the prior function is intended to describe a situation where there is no relevant information about the quantity of interest, is often required.
- Objectivity is an emotionally charged word, and it should be explicitly qualified. No statistical analysis is really objective (both the experimental design and the model have strong subjective inputs). However, frequentist procedures are branded as "objective" just because their conclusions are only conditional on the model assumed and the data obtained. Bayesian methods where the prior function is derived from the assumed model are objective is this limited, but precise sense.

□ Development of objective priors

• Vast literature devoted to the formulation of objective priors.

• Reference analysis, (Bernardo, 1979; Berger and Bernardo, 1992; Berger, Bernardo and Sun, 2009), has been a popular approach.

Theorem 1 *Let* $\boldsymbol{z}^{(k)} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k\}$ *denote* k *conditionally independent observations from* $\mathcal{M}_{\boldsymbol{z}}$. *For sufficiently large* k

$$\pi(\theta) \propto \exp\left\{\mathrm{E}_{\boldsymbol{z}^{(k)}\,|\,\theta}[\log p_h(\theta \mid \boldsymbol{z}^{(k)})]\right\}$$

where $p_h(\theta \mid \boldsymbol{z}^{(k)}) \propto \prod_{i=1}^{k} p(\boldsymbol{z}_i \mid \theta)\, h(\theta)$ *is the posterior which corresponds to some arbitrarily chosen prior function* $h(\theta)$ *which makes the posterior proper for any* $\boldsymbol{z}^{(k)}$.

• The reference prior at θ is proportional to the logarithmic sampling average of the posterior densities of θ that would be obtained if this where the true parameter value.

Approximate reference priors

• Reference priors are derived for an ordered parameterization. Given $\mathcal{M}_{\boldsymbol{z}} = \{p(\boldsymbol{z} \,|\, \boldsymbol{\omega}), \boldsymbol{z} \in \boldsymbol{\mathcal{Z}}, \boldsymbol{\omega} \in \boldsymbol{\Omega}\}$ with m parameters, the reference prior with respect to $\boldsymbol{\phi}(\boldsymbol{\omega}) = \{\phi_1, \ldots, \phi_m\}$ is sequentially obtained as $\pi(\boldsymbol{\phi}) = \pi(\phi_m \,|\, \phi_{m-1}, \ldots, \phi_1) \times \cdots \times \pi(\phi_2 \,|\, \phi_1)\, \pi(\phi_1)$.

• One is often simultaneously interested in several functions of the parameters. Given $\mathcal{M}_{\boldsymbol{z}} = \{p(\boldsymbol{z} \,|\, \boldsymbol{\omega}), \boldsymbol{z} \in \boldsymbol{\mathcal{Z}}, \boldsymbol{\omega} \in \boldsymbol{\Omega} \subset \Re^m\}$ with m parameters, consider a set $\boldsymbol{\theta}(\boldsymbol{\omega}) = \{\theta_1(\boldsymbol{\omega}), \ldots, \theta_r(\boldsymbol{\omega})\}$ of $r > 1$ functions of interest; Berger, Bernardo and Sun (work in progress) suggest a procedure to select a joint prior $\pi_{\boldsymbol{\theta}}(\boldsymbol{\omega})$ whose corresponding marginal posteriors $\{\pi_{\boldsymbol{\theta}}(\theta_i \,|\, \boldsymbol{z})\}_{i=1}^r$ will be close, for all possible data sets $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$, to the set of reference posteriors $\{\pi(\theta_i \,|\, \boldsymbol{z})\}_{i=1}^r$ yielded by the set of reference priors $\{\pi_{\theta_i}(\boldsymbol{\omega})\}_{i=1}^r$ derived under the assumption that each of the θ_i's is of interest.

Definition 6 *Consider model* $\mathcal{M}_{\boldsymbol{z}} = \{p(\boldsymbol{z} \,|\, \boldsymbol{\omega}), \boldsymbol{z} \in \mathcal{Z}, \boldsymbol{\omega} \in \boldsymbol{\Omega}\}$ *and* $r > 1$ *functions of interest,* $\{\theta_1(\boldsymbol{\omega}), \ldots, \theta_r(\boldsymbol{\omega})\}$*. Let* $\{\pi_{\theta_i}(\boldsymbol{\omega})\}_{i=1}^{r}$ *be the relevant reference priors, and* $\{\pi_{\theta_i}(\boldsymbol{z})\}_{i=1}^{r}$ *and* $\{\pi(\theta_i \,|\, \boldsymbol{z})\}_{i=1}^{r}$ *the corresponding prior predictives and marginal posteriors. Let* $\mathcal{F} = \{\pi(\boldsymbol{\omega} \,|\, \boldsymbol{a}), \boldsymbol{a} \in \mathcal{A}\}$ *be a family of prior functions. For each* $\boldsymbol{\omega} \in \boldsymbol{\Omega}$*, the best approximate joint reference prior within* $\mathcal{F}$ *is that which minimizes the average expected intrinsic loss*

$$d(\boldsymbol{a}) = \frac{1}{r} \sum_{i=1}^{r} \int_{\mathcal{Z}} \delta\{\pi_{\theta_i}(\cdot \,|\, \boldsymbol{z}),\, p_{\theta_i}(\cdot \,|\, \boldsymbol{z}, \boldsymbol{a})\}\, \pi_{\theta_i}(\boldsymbol{z})\, d\boldsymbol{z}, \quad \boldsymbol{a} \in \mathcal{A}.$$

• Example. Use of the Dirichlet family in the m-multinomial model (with $r = m + 1$ cells) yields $\mathrm{Di}(\boldsymbol{\theta} \,|\, 1/r, \ldots, 1/r)$, with important applications to sparse multinomial data and contingency tables.

Integrated Reference Analysis

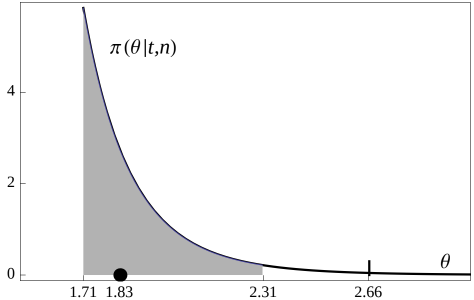
• We suggest a systematic use of the intrinsic loss function and an appropriate joint reference prior for an integrated objective Bayesian solution to both estimation and hypothesis testing in pure inference problems.

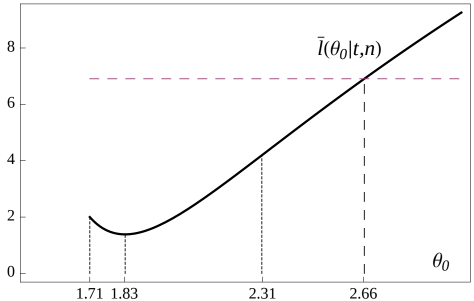
• We have stressed foundations-like decision theoretic arguments, but a large collection of detailed, non-trivial examples prove that the procedures advocated lead to attractive, often novel solutions. Details in Bernardo (2010) and references therein.

□ Estimation of the normal variance

• The intrinsic (invariant) point estimator of the normal standard deviation is is $\sigma^* \approx \frac{n}{n-1}\, s$. Hence, $\sigma^2{*} \approx \frac{n}{n-1}\,\frac{ns^2}{n-1}$, larger than both the mle s^2 and the unbiased estimator $ns^2/(n-1)$.

□ Uniform model $\mathrm{Un}(x \,|\, 0, \theta)$





$$\ell_\delta\{\theta_0, \theta \,|\, \mathcal{M}_{\boldsymbol{z}}) = n \begin{cases} \log(\theta_0/\theta), & \text{if } \theta_0 \geq \theta, \\ \log(\theta/\theta_0, & \text{if } \theta_0 \leq \theta. \end{cases}$$

$\pi(\theta) = \theta^{-1}, \quad \boldsymbol{z} = \{x_1, \ldots, x_n\},$

$t = \max\{x_1, \ldots, x_n\}, \quad \pi(\theta \,|\, \boldsymbol{z}) = n\, t^n \theta^{-(n+1)}$

The q-quantile is $\theta_q = t\,(1-q)^{-1/n}$;

Exact probability matching.

$\theta^* = t\, 2^{1/n}$ (posterior median)

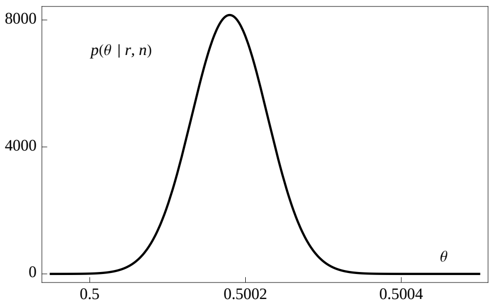
$\mathrm{E}[\overline{\ell}_\delta(\theta_0 \,|\, t, n) \,|\, \theta] = (\theta/\theta_0)^n - n \log(\theta/\theta_0)$;

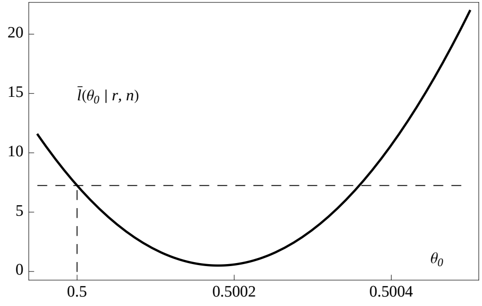
this is equal to 1 if $\theta = \theta_0$,

and increases with n otherwise.

- Simulation: $n = 10$ with $\theta = 2$ which yielded $t = 1.71$;
  $\theta^* = 1.83$, $\Pr[t < \theta < 2.31 \,|\, \boldsymbol{z}] = 0.95$, $\overline{\ell}_\delta(2.66 \,|\, \boldsymbol{z}) = \log 1000$.

□ Extra Sensory Power (ESP) testing





Jahn, Dunne and Nelson (1987)

Binomial data. Test $H_0 \equiv \{\theta = 1/2\}$ with $n = 104,490,000$ and $r = 52,263,471$.

For any sensible continuous prior $p(\theta)$,

$p(\theta \mid \boldsymbol{z}) \approx \mathrm{N}(\theta \mid m_{\boldsymbol{z}}, s_{\boldsymbol{z}})$,

with $m_{\boldsymbol{z}} = (r + 1/2)/(n + 1) = 0.50018$,

$s_{\boldsymbol{z}} = [m_{\boldsymbol{z}}(1 - m_{\boldsymbol{z}})/(n+2)]^{1/2} = 0.000049$.

$\bar{\ell}(\theta_0 \mid \boldsymbol{z}) \approx \frac{n}{2} \log[1 + \frac{1}{n}(1 + t_{\boldsymbol{z}}(\theta_0)^2)]$,

$t_{\boldsymbol{z}}(\theta_0) = (\theta_0 - m_{\boldsymbol{z}})/s_{\boldsymbol{z}}$, $t_{\boldsymbol{z}}(1/2) = 3.672$.

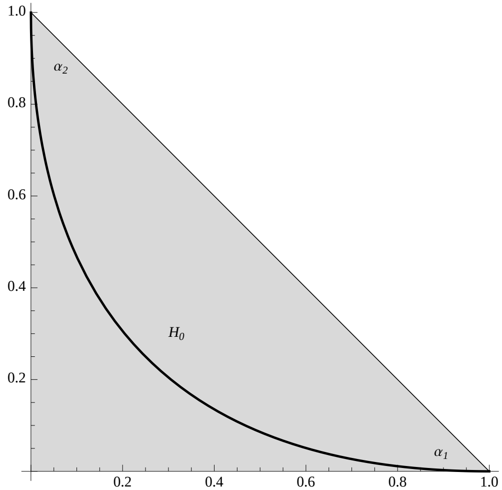
$\bar{\ell}(\theta_0 \mid \boldsymbol{z}) = 7.24 = \log 1400$: Reject H_0

- Jeffreys-Lindley paradox: With any "sharp" prior, $\Pr[\theta = 1/2] = p_0$, $\Pr[\theta = 1/2 \mid \boldsymbol{z}] > p_0$ (Jefferys, 1990) suggesting data support H_0 !!!

□ Trinomial data: Testing for Hardy-Weinberg equilibrium

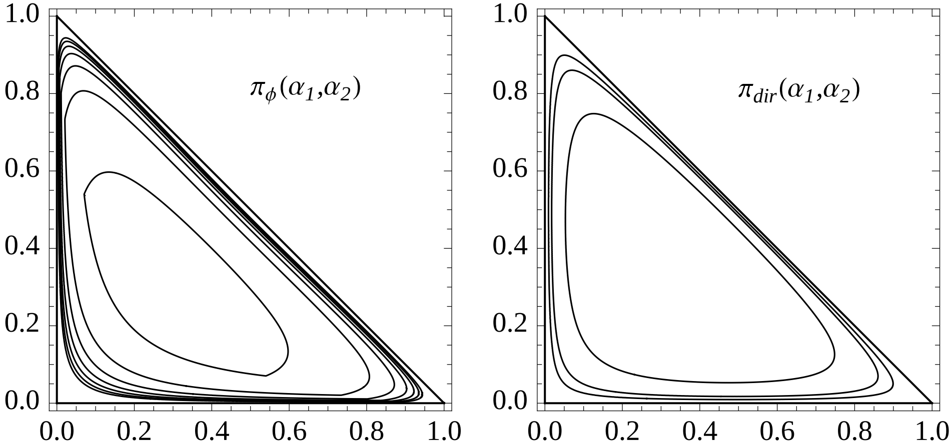
- To determine whether or not a population mates randomly.
- At a single autosomal locus with two alleles, a diploid individual has three possible genotypes, $\{AA, aa, Aa\}$, with (unknown) population frequencies $\{\alpha_1, \alpha_2, \alpha_3\}$, where $0 < \alpha_i < 1$ and $\sum_{i=1}^{3} \alpha_i = 1$.
- Hardy-Weinberg (HW) equilibrium iff $\exists\, p = \Pr(A)$, such that $\{\alpha_1, \alpha_2, \alpha_3\} = \{p^2, (1-p)^2, 2p(1-p)\}$.
- Given a random sample of size n from the population, and observed $\boldsymbol{z} = \{n_1, n_2, n_3\}$ individuals (with $n = n_1 + n_2 + n_3$) from each of the three possible genotypes $\{AA, aa, Aa\}$, the question is whether or not these data support the hypothesis of HW equilibrium.
- This is a good example of *precise* hypothesis in the sciences, since HW equilibrium corresponds to a zero measure set within the original simplex parameter space:

• The null is $H_0 = \{(\alpha_1, \alpha 2);\ \sqrt{\alpha_1} + \sqrt{\alpha_2} = 1\}$, a zero measure set within the (simplex) parameter spate of a trinomial distribution.



• The parameter of interest is is the intrinsic divergence of H_0 from the model, $\phi(\alpha_1, \alpha_2) = \delta\{H_0, \mathrm{Tri}(r_1, r_2, r_3 \,|\, \alpha_1, \alpha_2)\}$

- The reference prior when $\theta(\alpha_1, \alpha_2)$ is the quantity of interest is $\pi_\phi(\alpha_1, \alpha_2) \approx \mathrm{Di}[\alpha_1, \alpha_2 \,|\, 1/3, 1/3, 1/3]$.



- $\overline{\ell}(H_0 \,|\, \boldsymbol{z}) = \int_{\mathcal{A}} \delta\{H_0, \mathrm{Tri}(r_1, r_2, r_3 \,|\, \alpha_1, \alpha_2)\}\, \pi_\phi(\alpha_1, \alpha_2 \,|\, \boldsymbol{z}) d\alpha_1 d\alpha_2,$
  $\approx \int_{\mathcal{A}} \pi_\phi(\alpha_1, \alpha_2)\, \mathrm{Di}[\alpha_1, \alpha_2 \,|\, r_1 + 1/3, r_2 + 1/3, r_3 + 1/3]\, d\alpha_1 d\alpha_2.$

- Sample of size $n = 30$ simulated from a population in HW equilibrium with $p = 0.3$, so that $\{\alpha_1, \alpha_2\} = \{p^2, (1-p)^2\} = \{0.09, 0.49\}$, yielded $\{n_1, n_2, n_3\} = \{2, 15, 13\}$.

  This gives $\bar{\ell}(H_0 \mid \boldsymbol{z}) = 0.321 = \log[1.38]$, so that the likelihood ratio against the null is expected to be only about 1.38, and the null is accepted. One may proceed under the assumption that the population is in HW equilibrium, suggesting random mating.

- Sample of size $n = 30$ simulated from a trinomial with $\{\alpha_1, \alpha_2\} = \{0.45, 0.40\}$, so that $\sqrt{\alpha_1} + \sqrt{\alpha_2} = 1.303 \neq 1$, and population is *not* in HW equilibrium, yielded $\{n_1, n_2, n_3\} = \{12, 12, 6\}$.

  This gives $\bar{\ell}(H_0 \mid \boldsymbol{z}) = 5.84 \approx \log[344]$, so that the likelihood ratio against the null is expected to be about 344. Thus, the null should be *rejected*, and one should proceed under the assumption that the population is *not* in HW equilibrium, suggesting non random mating.

□ Contingency tables: Testing for independence

Data $\boldsymbol{z} = \{\{n_{11},\ldots,n_{1b}\},\ldots,\{n_{a1},\ldots,n_{ab}\}\}$, $k = a\times b$,
$\bar{\ell}(H_0\,|\,\boldsymbol{z}) \approx \int_{\boldsymbol{\Theta}} n\,\phi(\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}\,|\,\boldsymbol{z})\,d\boldsymbol{\theta}$, $\quad \phi(\boldsymbol{\theta}) = \sum_{i=1}^{a}\sum_{j=1}^{b}\theta_{ij}\log\big[\frac{\theta_{ij}}{\alpha_i\,\beta_j}\big]$,
where $\alpha_i = \sum_{j=1}^{b}\theta_{ij}$ and $\beta_j = \sum_{i=1}^{a}\theta_{ij}$ are the marginals, and
$\pi(\boldsymbol{\theta}\,|\,\boldsymbol{z}) = \mathrm{Di}_{k-1}(\boldsymbol{\theta}\,|\,n_{11}+1/k,\ldots,n_{ab}+1/k)$.

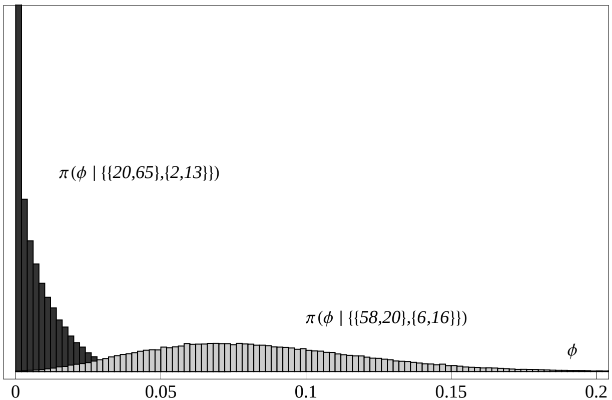
• *Simulation under independence.* Observations ($n = 100$) simulated from a contingency table with cell probabilities
$\boldsymbol{\theta} = \{\{0.24, 0.56\},\{0.06, 0.14\}\}$,
an independent contingency table with marginals $\boldsymbol{\alpha} = \{0.8, 0.2\}$ and $\boldsymbol{\beta} = \{0.3, 0.7\}$. This yielded data $\boldsymbol{z} = \{\{20, 65\},\{2, 13\}\}$.

This produces $\bar{\ell}(H_0\,|\,\boldsymbol{z}) = 0.80 = \log[2.23]$, suggesting that the observed data are indeed compatible with the independence hypothesis.

- *Simulation under non independence.* Observations ($n = 100$) simulated from a non independent contingency table with cell probabilities $\boldsymbol{\theta} = \{\{0.60, 0.20\}, \{0.05, 0.15\}\}$, yielding data $\boldsymbol{z} = \{\{58, 20\}, \{6, 16\}\}$.

  This produces $\overline{\ell}(H_0 \,|\, \boldsymbol{z}) = 8.35 = \log[4266]$, implying that the observed data are *not* compatible with the independence assumption.

- Posterior distributions of $\phi(\boldsymbol{\theta})$ for the two simulations:



Basic References

(In chronological order)

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. B* **41**, 113–147 (with discussion).

Berger, J. O. and Bernardo, J. M. (1992). On the development of reference priors. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) Oxford: University Press, 35–60 (with discussion).

Bernardo, J. M. (1997). Noninformative priors do not exist *J. Statist. Planning and Inference* **65**, 159–189 (with discussion).

Bernardo, J. M. (1999). Nested hypothesis testing: The Bayesian reference criterion. *Bayesian Statistics 6* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) Oxford: University Press, 101–130 (with discussion).

Bernardo, J. M. and Rueda, R. (2002). Bayesian hypothesis testing: A reference approach. *Internat. Statist. Rev.* **70**, 351–372.

Bernardo, J. M. (2005a). Reference analysis. *Bayesian Thinking: Modeling and Computation, Handbook of Statistics* **25** (Dey, D. K. and Rao, C. R., eds). Amsterdam: Elsevier, 17–90.

Bernardo, J. M. (2005b). Intrinsic credible regions: An objective Bayesian approach to interval estimation. *Test* **14**, 317–384 (with discussion).

Berger, J. O. (2006). The case for objective Bayesian analysis. *Bayesian Analysis* **1**, 385–402 and 457–464, (with discussion).

Bernardo, J. M. (2007). Objective Bayesian point and region estimation in location-scale models. *Sort* **31**, 3–44, (with discussion).

Berger, J. O., Bernardo, J. M. and Sun, D. (2009). Natural induction: An objective Bayesian approach. *Rev. Acad. Sci. MadridA* **103**, 125–159, (with discussion).

Berger, J. O., Bernardo, J. M. and Sun, D. (2009). The formal definition of reference priors. *Ann. Statist.* **37**, 905–938.

Bernardo, J. M. and Tomazella, V. (2010). Bayesian reference analysis of the Hardy-Weinberg equilibrium. *Frontiers of Statistical Decision Making and Data Analysis. In Honor of James O. Berger* (M.-H. Chen, D. K. Dey, P. Müller, D. Sun and K. Ye, eds.) New York: Springer, (to appear).

Berger, J. O., Bernardo, J. M. and Sun, D. (2010). Reference priors for discrete parameters. *J. Amer. Statist. Assoc.* (under revision).

Bernardo, J. M. (2010). Objective Bayesian estimation and hypotheis testing. *Bayesian Statistics 9* (to appear).