Reference Priors for Discrete Parameter Spaces

José M. Bernardo
Universidad de Valencia, Spain

joint work with
James O. Berger and Dongchu Sun

O-Bayes 09
Philadelphia, USA. June 6th, 2009

Summary

(i) **The Problem**

Conventional priors are often not appropriate.
Embedding into continuous parameter models.

(ii) **Assuming that the parameter is continuous**

Estimating population sizes in biology or reliability

(iii) **Assuming a hierarchical structure**

The hypergeometric model.

(iv) **Reference priors from consistent estimators**

The Binomial-Beta model.

1. The Problem

Conventional priors are often not appropriate

- If the parameter space is finite all proposed solutions (Laplace, Reference, Maximum Entropy...) lead to a uniform prior. This may not be appropriate when there is some structure in the model.

- Example. Hypergeometric distribution for a population of known size N. Unknown parameter is R, the number of items in the population with a property, $R \in \{0, 1, \ldots, N\}$. The data is r, the number of items with the property out of a random sample of size n.

The parameter space finite, but structured: for large N this becomes a binomial with parameter $p = R/N$. The objective prior for R should be compatible with the reference prior for p, the (non-uniform) Jeffreys prior $\pi(p) \propto p^{-1/2}(1-p)^{-1/2}$.

Embedding into continuous parameter models

Propose embedding into continuous parameter models and apply standard reference theory. We consider three possible strategies:

(i) *Treat the parameter as continuous in the original model.*
Fine when it works, but this often introduces a normalizing constant which can cause problems.

(ii) *Introduce a continuous hierarchical hyperparameter describing the problem structure.*
This is the best option when such hierarchical modeling is possible and natural. Unfortunately, this is not a common situation.

(iii) *Apply reference theory with a consistent estimator.*
A very general approach, but the result may weakly depend on the particular estimator used.

2. Assuming the parameter is continuous

- Replace the original model $\mathcal{M} = \{p(\boldsymbol{x} \,|\, \theta), \boldsymbol{x} \in \mathcal{X}, \theta \in \Theta\}$, where Θ is countable, by $\mathcal{M}^* = \{k(\theta)\, p(\boldsymbol{x} \,|\, \theta), \boldsymbol{x} \in \mathcal{X}, \theta \in \Theta^*\}$, where Θ^* is some continuous extension of Θ, and $k(\theta)$ the required normalizing factor to ensure that $\int_{\mathcal{X}} k(\theta)\, p(\boldsymbol{x} \,|\, \theta) d\boldsymbol{x} = 1$, $\forall \theta \in \Theta^*$ (obviously $k(\theta) = 1$ when $\theta \in \Theta$). Use standard reference theory to obtain the continuous reference prior $\pi(\theta)$ for $\mathcal{M}^*$, and use its restriction to the countable Θ as the reference prior for $\mathcal{M}$.

We have found that the procedure only works well when $k(\theta) = 1$ for all $\theta \in \Theta^*$ (for a non-constant normalization factor $k(\theta)$ modifies the structure of the original problem). There are, however important problems in the category, *e.g.*, those of estimating population sizes in biology an number of bugs in reliability.

Estimating a population size

Experiment with Type II censoring; a sample of N with exponential $\mathrm{Ex}(t \,|\, \lambda)$ lifetimes and experiment is stopped when R units have failed; $\{t_1 \leq \ldots \leq t_R\}$ denote the failure times, $N \geq R$ and λ are both unknown, and R is pre-specified. The problem of estimating N has many applications: estimate the number of fish in a lake, where the time to catch a fish is exponential; software reliability, where the number N of bugs is unknown, and the lengths of time to discover the first R bugs are exponential.

From the joint sampling density $p(t_1, \ldots, t_R \,|\, N, \lambda)$, the pair $(v, w) = (t_1 + \ldots + t_R, t_R)$ is found to be sufficient, and the marginal density of v is $p(v \,|\, N) = p(v \,|\, N, \lambda)$ which only depends of N. This also results by integrating out λ in $p(v, w \,|\, N, \lambda)$ with the conditional reference prior $\pi(\lambda \,|\, N) = \lambda^{-1}$. Thus $p(v \,|\, N)$ is all what matters.

Embedding the discrete space for N, which is $\{R, R{+}1, ...\}$, into the continuous space $(R-0.5, \infty)$, $p(v \mid N)$ is still a probability density for the continuous N (with the same normalization). The corresponding reference prior for N is simply Jeffreys prior $\pi(N \mid R) \propto \sqrt{i_R(N)}$, with $N \geq R - 0.5$, where $i_R(N)$ is Fisher information function.

With the convenient reparametrization $\theta = N - R + 1$, $\theta \in \mathcal{N}$ we find the explicit formulae for $\pi(\theta \mid R)$ which are all *proper* priors. For $R = 2$ this is exactly $\pi(\theta \mid R = 2) = \theta^{-1}(1 + \theta)^{-1}$. For other R values they do differ somewhat at $\theta = 1$ but are otherwise remarkably similar. Hence one could just use $\pi(\theta \mid R = 2) \approx \theta^{-1}(1 + \theta)^{-1}$ as the reference prior for any R. although the exact reference prior is not difficult to program and work with.

3. Assuming a hierarchical structure

- In some special models, $\mathcal{M} = \{p(\boldsymbol{x} \,|\, \theta), \boldsymbol{x} \in \mathcal{X}, \theta \in \Theta\}$, where Θ is countable, it is natural to add a hierarchical level of modeling $p(\theta \,|\, \omega)$, $\omega \in \Omega \subset \Re$, to create a continuous hyperparameter ω that can be analyzed with usual reference prior methods.

The problem is then reduced to finding the reference prior $\pi(\omega)$ for the continuous hyperparameter ω in the integrated model $p(\boldsymbol{x} \,|\, \omega) = \sum_{\theta \in \Theta} p(\boldsymbol{x} \,|\, \theta)\, p(\theta \,|\, \omega)$, and this can be used to determine the implied reference prior for θ, $\pi(\theta) = \int_{\Omega} p(\theta \,|\, \omega)\, \pi(\omega)\, d\omega$.

For instance in the hypergeometric model, one can postulate that the unknown number R of successes in the population of size N arises as a Binomial random variable from a binomial $\mathrm{Bi}(R \,|\, N, p)$ distribution, with $p \in (0, 1)$ unknown.

The hypergeometric model

- Consider sampling without replacement from a finite population of know size N, with R positive elements, such that

  $p(r \,|\, n, R, N) = \mathrm{Hy}(r \,|\, n, R, N) = \frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}}, \quad r \in \{0, \ldots, R\},$

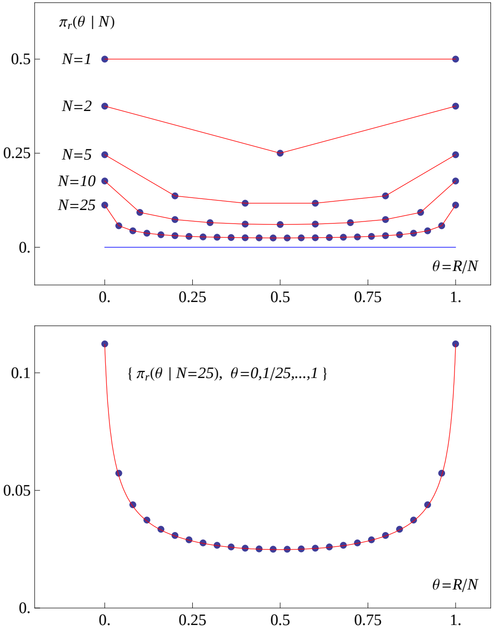
  $p(R \,|\, N, p) = \mathrm{Bi}(r \,|\, n, p) = \binom{n}{r} p^r (1-p)^{n-r}, \quad p \in (0,1).$
- The corresponding integrated model is

  $p(r \,|\, n, N, p) = \sum_{R=0}^{N} \mathrm{Hy}(r \,|\, n, R, N)\, \mathrm{Bi}(r \,|\, n, p) = \mathrm{Bi}(r \,|\, n, p)$

  whose corresponding reference prior is $\pi(p) \propto p^{-1/2}(1-p)^{-1/2}$
- The implied reference prior for R is the (obviously proper) prior

  $\pi(R \,|\, N) = \frac{1}{\pi} \frac{\Gamma(R+\frac{1}{2})\,\Gamma(N-R+\frac{1}{2})}{\Gamma(R+1)\,\Gamma(N-R+1)}\,.$

  For large N this is, in terms of $\theta = R/N$, basically proportional to $\mathrm{Be}(\theta \,|\, \frac{1}{2}, \frac{1}{2})$, the reference prior for the binomial limiting model.



Hypergeometric $\mathrm{Hy}(r \mid n, R, N)$ *model. Reference priors* $\pi(\theta \mid N)$ *for several* N *values, in terms of* $\theta = R/N$.

4. Reference priors from consistent estimators

- Consider model $\mathcal{M} = \{p(\boldsymbol{x} \mid \theta), \boldsymbol{x} \in \mathcal{X}, \theta \in \Theta\}$, where Θ is countable. Let $\boldsymbol{x}^{(k)} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$ denote k conditionally independent observations from $\mathcal{M}$, and consider a *consistent* estimator $\tilde{\theta}_k(\boldsymbol{x}^{(k)})$ of θ (which effectively becomes continuous as $k \to \infty$).
- Let $p(\tilde{\theta}_k \mid \theta)$ be the asymptotic sampling distribution of $\tilde{\theta}_k$ as $k \to \infty$. Treat θ as a continuous parameter in model $p(\tilde{\theta}_k \mid \theta)$, and derive the corresponding reference prior $\pi(\theta)$.
- If, as $k \to \infty$ and for appropriate functions c_k (typically $c_k = \sqrt{k}$), $c_k(\tilde{\theta}_k - \theta)$ has a limiting normal distribution with mean zero and variance $\sigma^2(\theta)$, the reference prior is simply $\pi(\theta) = \sigma(\theta)^{-1}$.
- The resulting reference prior depends of the estimator chosen, but the differences are typically negligible in practice.

The Binomial-Beta model

- Consider data $r \in \{0, \ldots, n\}$ from a binomial-beta distribution $p(r \mid n, a, b) = \int_0^1 \mathrm{Bi}(r \mid n, p)\, \mathrm{Be}(p \mid a, b)\, dp = \binom{n}{r} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(r+a)\Gamma(n-r+b)}{\Gamma(n+a+b)}$, so that $\mathrm{E}[r \mid n, a, b] = n\frac{a}{a+b}$, and $\mathrm{Var}[r \mid n, a, b] = \frac{n(n+a+b)}{(a+b)^2(a+b+1)}$.
- Let $\{r_1, \ldots, r_k\}$ be a sample of size k from $p(r \mid n, a, b)$. and consider the simple (linear) consistent estimator of n is $\tilde{n} = \frac{a+b}{ak} \sum_{j=1}^{k} r_j$
- From the central limit theorem, $p(\tilde{n} \mid n, a, b)$ is approximately normal, with mean n and variance $\mathrm{Var}[\tilde{n} \mid n, a, b] = \frac{n(n+a+b)}{a^2(a+b+1)k}$.
- Hence, the proposed reference prior is $\pi(n) \propto \mathrm{Var}[\tilde{n} \mid n, a, b]^{-1/2} \propto \frac{1}{\sqrt{n(n+a+b)}}$.
- It has been shown by simulation that the posterior induced by this prior has appropriate coverage properties.

Other examples in the supporting paper.

Basic References

Available on line at **www.uv.es/bernardo**

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

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

Berger, J. O., Bernardo, J. M. and Sun, D. (2009c). Reference priors for discrete parameter spaces. Submitted for publication. *Tech. Rep.* available.

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

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

Valencia Mailing List

- The Valencia Mailing List contains about 2,000 entries of people interested in Bayesian Statistics. It sends information about the Valencia Meetings and other material of interest to the Bayesian community.
- If you do not belong to the list and want to be included,
  please send your data to **valenciameeting@uv.es**

9th Valencia International Meeting on Bayesian Statistics

Benidorm (Alicante, Spain), June [3–8], 2010

- Bernardo, J.M., Bayarri, M.J., Berger, J.O. Dawid, A.P. Heckerman, D., Smith, A.F.M. and West, M. (2007). *Bayesian Statistics 8.* Oxford, UK: University Press.

Thanks for your attention!

jose.m.bernardo@uv.es
www.uv.es/bernardo

valenciameeting@uv.es
www.uv.es/valenciameeting