DEFAULT BAYESIAN ANALYSIS FOR THE MULTIVARIATE EWENS DISTRIBUTION

ABEL RODRÍGUEZ

Abstract. We derive the Jeffreys prior for the parameter of the Multivariate Ewens Distribution and study some of its properties. In particular, we show that this prior is proper and has no finite moments. We also investigate the impact of this default prior on the a priori distribution of the number of species and the a priori probability of discovery of a new species, which are usually employed in subjective prior elicitation. The effect of the Jeffreys prior for posterior inference is illustrated using examples arising in the context of inference for species sampling models and Dirichlet process mixture models.

1. Introduction

The Multivariate Ewens Distribution (MED), also known as the Ewens Sampling Formula (ESF) (Ewens, 1972; Johnson et al., 1997), is a probability distribution on the partitions of the set $\{1, 2, \dots, n\}$. It appears often in genetics as the distribution of the number of distinct alleles in a sample of size n drawn from an infinite idealized population, or as the limiting distribution of other, more general models (Kingman, 1978; Hoppe, 1987). More generally, the MED belongs to the class of species sampling models, which describe distributions for exchangeable random partitions (Aldous, 1985; Pitman, 1995; Lijoi et al., 2007). The MED also appears in the context of Bayesian nonparametric statistics, where it is related to the number of unique values in a random sample of size n taken from a random distribution that follows a Dirichlet process prior (Antoniak, 1974). Hence, in the context of Dirichlet process mixture models, the MED acts as the prior distribution on the number and size of the clusters imposed by model.

This paper is concerned with Bayesian estimation and prediction under the MED model. Gamma priors, or mixtures thereof, are commonly used as priors in this context because of the existence of simple Gibbs sampling algorithms based on data augmentation (Escobar & West, 1995). Hyperparameters are elicited by either exploiting their link with the expected number of distinct alleles (Escobar & West, 1995) or, in the case of nonparametric mixture models, their link with the mean and variance of the observations (Walker & Mallick, 1997). Alternatively, Carota & Parmigiani (2002) and Griffin & Steel (2004) propose eliciting priors on the probabilities of new alleles, which in turn imply a prior on the parameter of interest. In either case, elicitation can be difficult because of the lack of relevant prior information in specific applications. To deal with the lack of prior information, numerous authors have used very dispersed Gamma priors. In this paper we derive the Jeffreys prior associated with the MED, show that this prior is proper, and investigate some of its properties. Because of its invariance to transformations, the Jeffreys prior provides a natural "default" or "non-informative" alternative to the priors discussed above without a substantial increase in computational complexity.

The remaining of the paper is organized as follows: Section 2 briefly reviews the Multivariate Ewens Distribution and derives the Jeffreys prior associated with its parameter. Section 3 discusses some of the properties of the Jeffreys prior. Section 4 presents some illustrations in the context of species sampling models and Dirichlet process mixture models. We conclude in Section 5 with some brief remarks and future directions.

2. The Jeffreys Prior for the Multivariate Ewens Distribution

Consider a partition of the set $\{1, 2, \dots, n\}$ into $K \leq n$ subsets so that there are r_j subsets of size j, where $\sum_{j=1}^{n} r_j = K$, and $\sum_{j=1}^{n} j r_j = n$. For example, the n elements of the original set might represent individuals being sampled from an infinite population, while the subsets into which they are divided could be interpreted as the species to which these individuals belong. The Multivariate Ewens Distribution (MED) assigns such a partition a probability

$$p(K, r_1, \dots, r_n \mid \beta) = \frac{\Gamma(\beta)\Gamma(n+1)}{\Gamma(\beta+n)} \beta^K \prod_{j=1}^{n} \frac{1}{j^{r_j} \Gamma(r_j+1)}, \tag{1}$$

where $0 < \beta < \infty$ is a parameter controlling the shape of the distribution.

Key words and phrases. Multivariate Ewens Distribution; Jeffreys Prior; Dirichlet Process; Ewens Sampling Formula.

The partitions associated with the MED can alternatively be described in terms of a sequence of exchangeable indicators $\xi_1, \dots, \xi_n$ such that $\xi_i = k$ if the i individual in the population belong to species k. Assuming that the species are labeled consecutively between at 1 and K, and letting $m_k = \sum_{i=1}^{n} I(\xi_i = k)$ be the number of individuals in species k, then

$$p(\xi_1, \dots, \xi_n \mid \beta) = p(K, m_1, \dots, m_K \mid \beta) = \frac{\Gamma(\beta)}{\Gamma(\beta + n)} \beta^K \prod_{k=1}^{K} \Gamma(m_k). \tag{2}$$

From (2) we can compute the probability mass function associated with the species of a new individual

$$p(\xi_{n+1} = k \mid \xi_1, \dots, \xi_n, \beta) = \begin{cases} \dfrac{p(K, m_1, \dots, m_k + 1, \dots, m_K \mid \beta)}{p(K, m_1, \dots, m_k, \dots, m_K \mid \beta)} = \dfrac{m_k}{\beta + n} & k \leq K \\[2ex] \dfrac{p(K + 1, m_1, \dots, m_k, \dots, m_K, 1 \mid \beta)}{p(K, m_1, \dots, m_k, \dots, m_K \mid \beta)} = \dfrac{\beta}{\beta + n} & k = K + 1. \end{cases}$$

This sequence of predictive distributions is sometimes called the Chinese restaurant process.

We are interested in estimating the parameter β based on either an observed sample $\xi_1, \dots, \xi_n$, or the sufficient statistic K. The following lemma provides an expression for a natural default prior for β.

Lemma 1. *For $n \geq 2$, the Jeffreys prior associated with* (1) *and* (2) *is given by*

$$\pi_n^J(\beta) \propto \sqrt{\frac{1}{\beta} \sum_{j=1}^{n-1} \frac{j}{(\beta + j)^2}}. \tag{3}$$

Proof. By definition, $\pi_n^J(\beta) \propto |\mathcal{I}(\beta)|^{1/2}$ where $\mathcal{I}(\beta) = -\mathsf{E}\left[\frac{d^2}{d\beta^2} \log\{p(K, m_1, \dots, m_K \mid \beta)\}\right]$ is the Fisher information associated with β. Now,

$$-\mathsf{E}\left[\frac{d^2}{d\beta^2} \log\{p(K, m_1, \dots, m_K \mid \beta)\}\right] = -\psi'(\beta) + \psi'(\beta + n) + \frac{\mathsf{E}\{K\}}{\beta^2},$$

where ψ' denotes the trigamma function (Abramowitz & Stegun, 1965). Now, using the facts that $\mathsf{E}\{K\} = \sum_{j=0}^{n-1} \frac{\beta}{\beta + j}$ (e.g., see Antoniak, 1974) and $\psi'(\beta + n) = \psi'(\beta) - \sum_{j=0}^{n-1} \frac{1}{(\beta+j)^2}$ (e.g., see Abramowitz & Stegun, 1965) we get

$$\mathcal{I}_E(\beta) = -\sum_{j=0}^{n-1} \frac{1}{(\beta + j)^2} + \frac{1}{\beta^2} \sum_{j=0}^{n-1} \frac{\beta}{\beta + j} = \frac{1}{\beta} \sum_{j=0}^{n-1} \frac{j}{(\beta + j)^2} = \frac{1}{\beta} \sum_{j=1}^{n-1} \frac{j}{(\beta + j)^2},$$

which directly leads to (3). □

In particular, note that, for $n = 2$ (the smallest sample size containing information about β), the Jeffreys prior on β corresponds to a standard Cauchy prior on $\nu = \beta^{1/2}$.

3. Properties of the Jeffreys Prior for the MED

Surprisingly, the Jeffreys prior is proper. Indeed, note that for all $n \geq 2$

$$\int_0^\infty \sqrt{\frac{1}{\beta} \sum_{j=1}^{n-1} \frac{j}{(\beta + j)^2}}\, d\beta \leq \int_0^\infty \sqrt{\frac{1}{\beta} \sum_{j=1}^{n-1} \frac{j}{(\beta + 1)^2}}\, d\beta = \sqrt{\frac{n(n-1)}{2}} \int_0^\infty \frac{1}{\beta^{3/2} + \beta^{1/2}}\, d\beta$$
$$= \sqrt{\frac{n(n-1)}{2}} \int_0^\infty \frac{1}{u^2 + 1}\, du = \pi \sqrt{\frac{n(n-1)}{2}}$$

Although the normalizing constant $C(n) = \int_0^\infty \sqrt{\frac{1}{\beta} \sum_{j=1}^{n-1} \frac{j}{(\beta+j)^2}}\, d\beta$ is not available in closed form, it is easily evaluated numerically using quadrature methods. The prior is decreasing, which can be easily verified,

$$\frac{d}{d\beta} \log\{\pi_n^J(\beta)\} = -\frac{1}{2\beta} - \frac{\sum_{j=1}^{n-1} \frac{j}{(\beta+j)^3}}{\sum_{j=1}^{n-1} \frac{j}{(\beta+j)^2}} < 0.$$

On the other hand, $\pi_n^J(\beta)$ is *log-convex*. Indeed,

$$\frac{\mathrm{d}^2}{\mathrm{d}\beta^2}\log\{\pi_n^J(\beta)\} = \frac{1}{2\beta^2} + \frac{3\left\{\sum_{j=1}^{n-1}\frac{j}{(\beta+j)^4}\right\}\left\{\sum_{j=1}^{n-1}\frac{j}{(\beta+j)^2}\right\} - 2\left\{\sum_{j=1}^{n-1}\frac{j}{(\beta+j)^3}\right\}^2}{\left\{\sum_{j=1}^{n-1}\frac{j}{(\beta+j)^2}\right\}^2} > 0.$$

To show this, note that the first term is clearly positive over the support of the prior. For the second term,

$$\begin{aligned}
3\left\{\sum_{j=1}^{n-1}\frac{j}{(\beta+j)^4}\right\}\left\{\sum_{j=1}^{n-1}\frac{j}{(\beta+j)^2}\right\} - 2\left\{\sum_{j=1}^{n-1}\frac{j}{(\beta+j)^3}\right\}^2 &= \sum_{j=1}^{n-1}\sum_{i=1}^{n-1}\left\{\frac{3ij}{(\beta+j)^4(\beta+i)^2} - \frac{2ij}{(\beta+j)^3(\beta+i)^3}\right\}\\
&= \sum_{j=1}^{n-1}\sum_{i=1}^{n-1}\left\{\frac{ij[3(\beta+i)-2(\beta+j)]}{(\beta+j)^4(\beta+i)^3}\right\}\\
&\geq \frac{1}{(\beta+n)^7}\sum_{j=1}^{n-1}\sum_{i=1}^{n-1}\left\{ij\beta + 3i^2j - 2ij^2\right\}\\
&= \frac{1}{(\beta+n)^7}\left\{\beta\sum_{j=1}^{n-1}\sum_{i=1}^{n-1}ij + \sum_{j=1}^{n-1}\sum_{i=1}^{n-1}i^2j\right\} > 0.
\end{aligned}$$

Figure 1 presents graphs of π_n^J for different values of n.

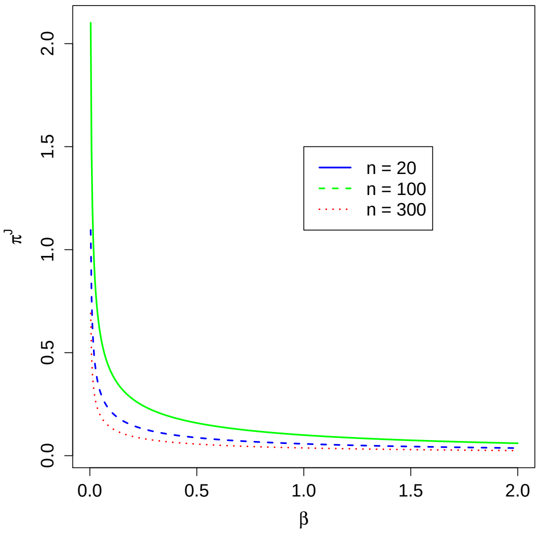

Figure 1. Density of the Jeffreys prior associated with the MED for some selected values of n.

The moments of $\pi_n^J(\beta)$ do not exist. Indeed,

$$\int_0^\infty \beta^s\sqrt{\frac{1}{\beta}\sum_{j=1}^{n-1}\frac{j}{(\beta+j)^2}}\,\mathrm{d}\beta > \int_0^\infty \beta^{s-1/2}\sqrt{\sum_{j=1}^{n-1}\frac{j}{(\beta+n)^2}}\,\mathrm{d}\beta = \sqrt{\frac{n(n-1)}{2}}\int_0^\infty \frac{u^{2s}}{u^2+n}\,\mathrm{d}u = \infty$$

for any $s \geq 1$. Instead, consider the median of $\pi_n^J(\beta)$ as a function of n. Again, although no closed-form expression is available for this median, it can be computed numerically. Figure 2 suggests that the median grows more or less linearly with the sample size n, with $\mathsf{Med}\{\beta\} \approx 0.36n + 1$ for $n \leq 400$.

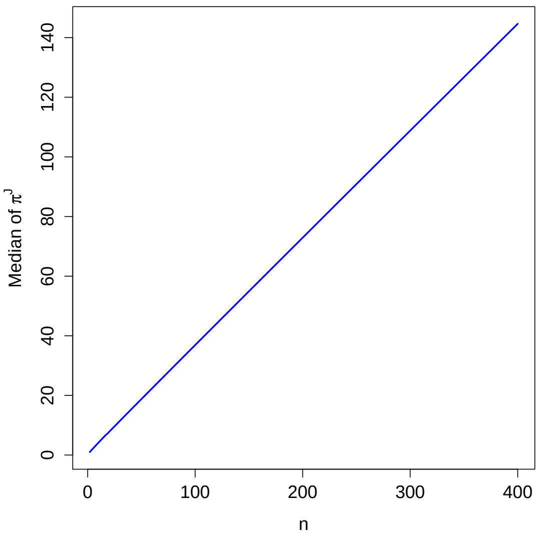

FIGURE 2. Median of $\pi_n^J(\beta)$ as a function of the sample size n.

Alternatively, we can consider the impact of the Jeffreys prior on the prior distribution of the number of species by computing prior moments such as the prior mean,

$$\mathsf{E}\{K \mid n\} = \mathsf{E}\{\mathsf{E}(K \mid \beta, n)\} \int_0^\infty \left(\sum_{j=0}^{n-1} \frac{\beta}{\beta + j} \right) \pi_n^J(\beta) \mathrm{d}\beta,$$

and the prior variance

$$\mathsf{Var}\{K \mid n\} = \int_0^\infty \sum_{j=0}^{n-1} \left(\frac{\beta j}{\{\beta + j\}^2} \right) \pi_n^J(\beta) \mathrm{d}\beta + \int_0^\infty \left(\sum_{j=0}^{n-1} \frac{\beta}{\beta + j} \right)^2 \pi_n^J(\beta) \mathrm{d}\beta - \left\{ \int_0^\infty \left(\sum_{j=0}^{n-1} \frac{\beta}{\beta + j} \right) \pi_n^J(\beta) \mathrm{d}\beta \right\}^2 .$$

The left panel in Figure 3 presents graphs for the prior mean and prior standard deviation as functions of n. Just like the median of $\pi_n^J(\beta)$, both of these quantities grow almost linearly with n. Moreover, note that $\mathsf{E}\{K\} \approx n/2$, and that the prior standard deviation grows somewhat more slowly than $\mathsf{E}\{K\}$. More generally, we can compute the marginal prior probability distribution on the number of species,

$$\mathsf{Pr}(K = k \mid n) = \int \mathsf{Pr}(K = k \mid \beta, n) \pi_n^J(\beta) \mathrm{d}\beta = |S(n, k)| A(n, n, k)$$

where

$$A(n, m, k) = \int_0^\infty \beta^k \frac{\Gamma(\beta)}{\Gamma(\beta + n)} \pi_m^J(\beta) \mathrm{d}\beta. \tag{4}$$

and $|S(n, k)|$ is the absolute value of the Stirling number of the second kind. For illustrative purposes, Figure 3 shows the marginal distribution $\mathsf{Pr}(K = k \mid n = 100)$. The effect of the Jeffrey's prior is striking; the resulting distribution is U-shaped and roughly symmetric around $n/2$ (which is compatible with our observations about $\mathsf{E}(K \mid n)$). Hence, under this prior the model prefers a priori either a very small or a very large number of species, with values around the mean/median actually having very low prior probabilities.

Yet another direction is to consider the effect of $\pi_n^J(\beta)$ on the priori probability of discovery of a new species, $\eta = \beta/(\beta + n + 1)$. In particular, Figures 3 and 3 present graphs of the expected value and the variance of η,

$$\mathsf{E}\{\eta \mid n\} = \int_0^\infty \left(\frac{\beta}{\beta + n + 1} \right) \pi_n^J(\beta) \mathrm{d}\beta, \quad \mathsf{Var}\{\eta \mid n\} = \int_0^\infty \left(\frac{\beta}{\beta + n + 1} \right)^2 \pi_n^J(\beta) \mathrm{d}\beta - \left\{ \int_0^\infty \left(\frac{\beta}{\beta + n + 1} \right) \pi_n^J(\beta) \mathrm{d}\beta \right\}^2 ,$$

as functions of n. Note that the values of both of these summaries are quite stable, with the expected probability of discovery of a new species varying between 0.38 and 0.40 over the range considered here.

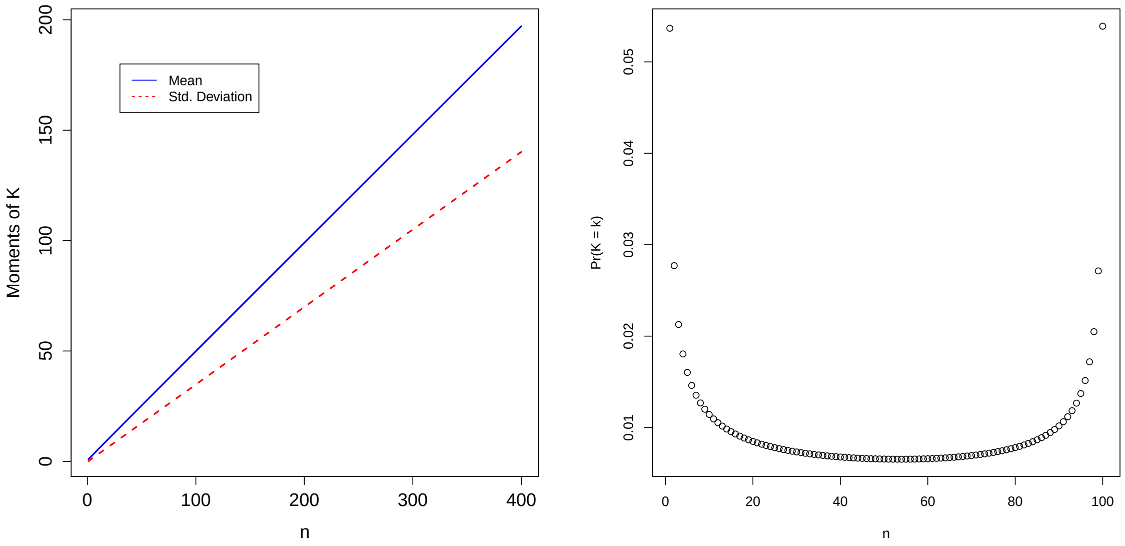

FIGURE 3. Left panel: expectation and standard deviation of the number of unique alleles, $E\{K \mid n\}$ and $\mathsf{Var}\{K \mid n\}$ as a function of the sample size n. Note that they both appear to grow linearly with n. Right panel: full marginal distribution $p(K = k \mid n) = \int_0^\infty p(K = k \mid \beta)\pi_n^J(\beta)\mathrm{d}\beta$ for $n = 100$.

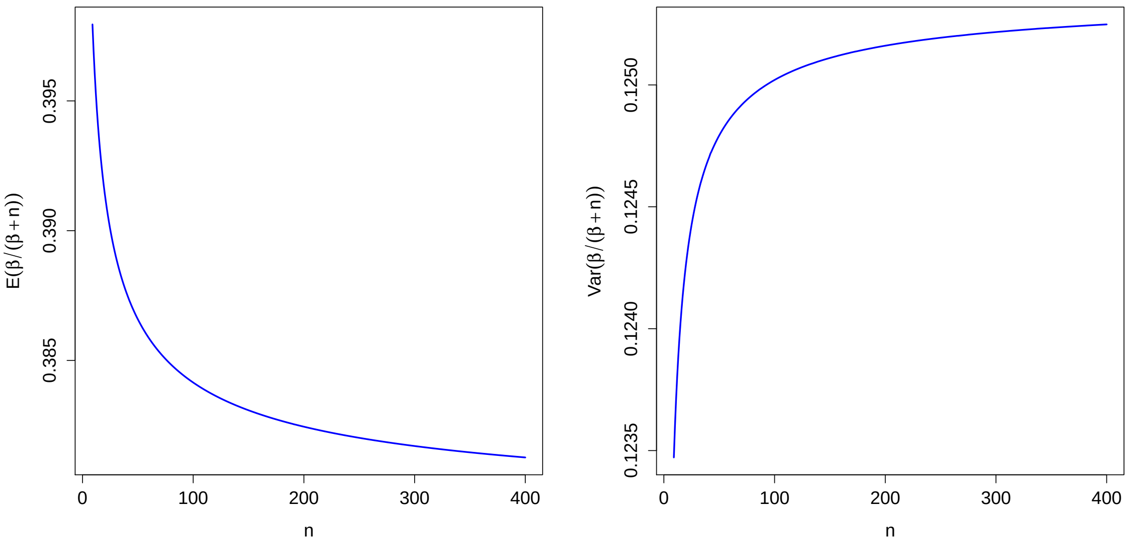

FIGURE 4. Expectation (left panel) and variance (right panel) of the probability of discovery a new species, $\mathsf{E}\{\beta/(\beta + n + 1)\}$, as a function of the sample size n.

4. ILLUSTRATIONS

4.1. **Species sampling models.** In this subsection we consider the use of our Jeffreys prior for inference under the MED model. We consider first a series of simulations where data is generated under a MED with true parameters $\beta_T = 1$ or $\beta_T = 2$. We consider two different scenarios, corresponding to $n = 100$ and $n = 1,000$ individuals, and for each of these scenarios we study the frequentist properties of Bayesian interval estimators generated under two different default priors: the Jeffreys priors derived in this paper, and a diffuse Gamma with shape parameter 0.001 and rate 0.001 (which has mean 1). The choice of a diffuse Gamma prior centered around $\beta = 1$ reflects standard practice in the literature.

β_T	n	α	Summary	Priors Jeffreys	 Gam $(0.001, 0.001)$
1	100	0.90	Coverage Width	0.89 1.85 (0.59)	0.86 1.74 (0.62)
		0.95	Coverage Width	0.95 2.24 (0.71)	0.89 2.11 (0.74)
	1000	0.90	Coverage Width	0.90 1.40 (0.34)	0.86 1.34 (0.35)
		0.95	Coverage Width	0.94 1.67 (0.38)	0.94 1.61 (0.39)
2	100	0.90	Coverage Width	0.91 2.84 (0.75)	0.86 2.74 (0.76)
		0.95	Coverage Width	0.95 3.43 (0.90)	0.94 3.32 (0.91)
	1000	0.90	Coverage Width	0.91 2.07 (0.37)	0.88 2.02 (0.38)
		0.95	Coverage Width	0.94 2.48 (0.44)	0.95 2.42 (0.45)

Table 1. Results of a simulation study to explore the frequentist coverage and expected length of $100\alpha\%$ credible intervals for β generated under three different priors.

	β		$\eta = \beta/(\beta + n + 1)$	
	Post. mean	95% Cred. Interval	Post. mean	95% Cred. Interval
Jeffreys	2763.3	(2526.9, 3015.8)	0.516	(0.494, 0.538)
Gam $(0.001, 0.001)$	2751.3	(2518.1, 3002.4)	0.515	(0.493, 0.537)

Table 2. Posterior credible inferences for β and $\eta = \beta/(\beta + n + 1)$ (the probability of discovery of a new species) under two different "non-informative" prior distributions for the sequencing tag data.

Markov chain Monte Carlo algorithms were used to obtain samples from the posterior distribution under each prior. All inferences are based on 100,000 iterations of the chain obtained after a burn-in period of 5,000 iterations. For the Jeffreys prior, we employed a random walk Metropolis-Hasting algorithm with Gaussian proposals for $\log \beta$; the variance of the proposal was $\tau^2 = 0.05$ for $n = 1,000$ and $\tau^2 = 1$ for $n = 100$, which resulted in average acceptance rates of 70% and 60% respectively. For the Gamma prior, we employed the latent variable approach discussed in Escobar & West (1995), which does not require any tunning parameter.

Table 1 presents empirical coverage probabilities and interval lengths for symmetric credible sets constructed. These summaries were estimated on the basis of $2,000$ randomly generated datasets. As expected, the frequentist coverage probability of 90% and 95% symmetric credible intervals under the Jeffreys prior seems to coincide with their nominal posterior probability. On the other hand, the dispersed Gamma prior produces intervals that are somewhat tighter than the Jeffreys prior, but which they tend to have lower empirical coverage rates (particularly, for 90% coverage).

As a second illustration, we consider a real dataset discussed in Mao & Lindsay (2002), Mao (2004), and Lijoi et al. (2007). The data consists of a $n = 2586$ randomly selected expressed sequence tags taken form a large cDNA library made from the 0 mm to 3 mm buds of tomato flowers. The number of distinct tags observed in this sample is $K = 1825$. Table 2 presents estimates of the MED parameter β under the same two priors discussed in the simulation study. The same MCMC algorithms described in the simulation study were used in this analysis. Note that, although inferences for β differ somewhat among the two priors, inferences for the probability of discovery of a new species, $\eta = \beta/(\beta + n + 1)$, are almost identical in both cases.

4.2. **Dirichlet process mixture models.** The Dirichlet process (DP) (Ferguson, 1973; Antoniak, 1974; Sethuraman, 1994) defines a prior distribution on the space of discrete measures and has been widely used in the context of nonparametric Bayesian inference. However, because of the discrete nature of distributions, the DP is not typically used to model the data directly, but as a prior for the mixing distribution in a kernel convolution. In that case, the data generating process for an independent and identically distributed sample $y_1, \ldots, y_n$ is assumed to be

$$y_i \mid G \sim \int p(y_i \mid \theta) G(\mathrm{d}\theta), \qquad\qquad G \mid \beta \sim \mathsf{DP}(\beta, G_0),$$

where $\mathsf{DP}(\beta, G_0)$ denotes a Dirichlet process prior with centering measure G_0 and precision parameter β, and $p(y_i|\theta)$ is a kernel indexed by the finite-dimensional parameter vector θ. Such a model can alternatively be described in terms of a series of partition indicators $\xi_1, \ldots, \xi_n$ and component-specific parameters $\vartheta_1, \vartheta_2, \ldots$, such that

$$y_i \mid \xi_i, \vartheta_1, \vartheta_2, \ldots \sim p(y_i \mid \vartheta_{\xi_i}), \qquad \xi_1, \ldots, \xi_n \mid \beta \sim \mathsf{MED}(\beta), \qquad \vartheta_k \sim G_0,$$

where $\mathsf{MED}(\beta)$ represents the multivariate Ewens distribution with parameter β.

We ran two simulation studies to investigate the impact of the (marginal) Jeffreys prior on posterior inferences for the DP mixture model. First, a dataset consisting of 50 observations was generated from a negative binomial distribution with mean 20 and variance 220, and a DP mixture of Poisson kernels with an unknown precision parameter β and a Gamma baseline measure was fitted to this data (the baseline measure was selected so that it had mean 20 and variance 200). Note that, because the negative binomial can be represented as a scale mixture of Poissons, the true data generating process corresponds to the limit of the Poisson DP mixture prior when $K = n = 50$ (which can be obtained by letting $\beta \to \infty$).

We considered two different prior distribution for β, namely, the Jeffreys prior introduced in this paper and the "non-informative" $\mathsf{Gam}(0.001, 0.001)$. A variant of the collapsed Gibbs samplers described in Neal (2000) was used to fit the model. In the case of the Jeffreys prior, we used a version of the algorithm that integrates over β, so that the posterior full conditional distribution for ξ_i is given by

$$p(\xi_i = k \mid \xi_1, \ldots, \xi_{i-1}, \xi_{i+1}, \ldots, \xi_n) = \begin{cases} m_k^- \dfrac{A(n, n, K^-)}{A(n-1, n, K^-)} p(y_i | y^-, G_0) & k \leq K^- \\ \\ \dfrac{A(n, n, K^- + 1)}{A(n-1, n, K^-)} p(y_i | G_0) & k = K^- + 1, \end{cases} \tag{5}$$

where $A(n, m, k)$ was defined in (4) and the negative exponent denotes the appropriate quantities computed after excluding observation i.

We focused our analysis on the posterior distribution of the number of occupied mixture components K. The posterior mean for K was very similar in both cases, 9.1309 under the Jeffreys prior and 9.0468 under the diffuse Gamma prior, with configurations including more than 16 mixture components having negligible posterior probability. Having a number of occupied clusters that is smaller than n is not really surprising; because the Dirichlet process prior strongly favors clustering, we expect the model to underestimate the number of mixture components. What is really interesting is that the Jeffreys prior seems to favor a larger number of clusters than the Gamma prior. Indeed, the posterior distribution of K under the Jeffreys prior seems to be stochastically greater than the posterior distribution under the Gamma prior (the same phenomena appeared when we repeated the simulation study with other datasets). This suggest that the Jeffreys prior does a slightly better job at identifying the true number of components in this case.

Finally, in order to evaluate whether the previous behavior is due to a systemic bias in the Jeffreys prior towards larger values of K, we ran a similar experiment where data was generated instead from a Poisson distribution with mean 20. Hence, in this case $K = 1$ corresponds to the truth. In this case, the posterior distribution for K under both models was identical (up to Monte Carlo error), so systematic bias does not seem to be present.

5. Concluding remarks

To the best of our knowledge, this is the first derivation of the Jeffrey's prior associated with the MED. Our numerical evaluations suggest that it might represent a reasonable default prior in situations where little prior information is available, including hierarchical models such as nonparametric mixture models based on the Dirichlet process. However, the Jeffreys prior explicitly depends on the sample size observed. Hence, any statistical procedures derived under this prior will depend on the stopping rule associated with the experiment; for example, the results will vary depending on whether data is analyzed sequentially or in batches.

References

Abramowitz, M. & Stegun, I. A. (Eds.) (1965). *Handbook of Mathematical Functions*. Dover Publication, 1st edition.

Aldous, D. J. (1985). Exchangeability and relatd topics. In *École d'Été de Probabilités de Saint-Flour XIII*, volume 1117 of *Lecture Notes in Mathematics*. Berlin/Heidelberg: Springer.

Antoniak, C. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics* **2**, 1152–1174.

Carota, C. & Parmigiani, G. (2002). Semiparametric regression for count data. *Biometrika* **89**, 265–281.

Escobar, M. D. & West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of American Statistical Association* **90**, 577–588.

Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**, 87–112.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**, 209–230.

Griffin, J. E. & Steel, M. F. J. (2004). Semiparametric Bayesian inference for stochastic frontier models. *Journal of Econometrics* **123**, 121–152.

Hoppe, F. M. (1987). The sampling theory of natural alleles and an urn model in population genetics. *Journal of Mathematical Biology* **23**, 123–159.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. Wiley.

Kingman, J. F. C. (1978). Random partitions in population genetics. *Proceedings of the Royal Society of London, Series B, Mathematical and Physical Sciences* **361**, 1–20.

Lijoi, A., Mena, R. H. & Prünster, I. (2007). Bayesian nonparametric estimation of the probability of discovering new species. *Biometrika* **94**, 769–786.

Mao, C. X. (2004). Prediction of the conditional probability of discovering a new class. *Journal of American Statistical Association* **99**, 1108–1118.

Mao, C. X. & Lindsay, B. G. (2002). A Poisson model for the coverage problem with a genomic application. *Biometrika* **89**, 669–682.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **9**, 249–265.

Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *Probability Theory and Related Fields* **102**, 145–158.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650.

Walker, S. G. & Mallick, B. K. (1997). A note on the scale parameter of the Dirichlet process. *Canadian Journal of Statistics* **25**, 473–479.