Journal of Official Statistics, Vol. 25, No. 3, 2009, pp. 405–413

Beyond Objective Priors for the Bayesian Bootstrap Analysis of Survey Data

Cinzia Carota[1]

This article provides reasonable answers to the problems left unsolved in Aitkin (2008), a recent paper on the Bayesian bootstrap in finite population inference. These problems are essentially two: the choice of the population parameter cannot be discussed from within the Aitkin's Bayesian bootstrap approach, which is based on a multinomial likelihood with unconstrained parameters; assumptions such as model constraints on the multinomial probabilities are difficult to implement in such a Bayesian framework. The answers are obtained by assigning suitable informative priors to the population proportions involved in the analysis.

Key words: Survey sampling; Bayes factors; population parameter selection; post-data Dirichlet priors; constraints on the multinomial probabilities.

1. Introduction

New possibilities for applications of the Bayesian method to real problems with a high degree of complexity are delineated by Aitkin (2008) in a recent article on the Bayesian bootstrap in finite population inference. Starting from the consideration that in a population of known dimension N the values of a numerical response Y are always measured with finite precision, δ say, and can be tabulated by the distinct values $Y_1 < \cdots < Y_D$ which Y can take, Aitkin denotes the corresponding counts by $N_1, \ldots, N_D$ $\left(Y_{j+1} - Y_j = \delta, \sum_1^D N_j = N\right)$ and assumes as natural model for Y the multinomial distribution with parameters given by the population proportions $p_j = N_j/N$, $j = 1, \ldots, D$. Any population parameter (as, for instance, the mean $\mu = \sum_1^D Y_j p_j$ or the variance $Var(Y) = \sum_1^D (Y_j - \mu)^2 p_j$) is a particular function of the population proportions and, given a prior for $p_1, \ldots, p_D$, its posterior can be derived from the posterior distribution of $p_1, \ldots, p_D$. *A* practical and flexible prior for the multinomial probabilities is the Dirichlet distribution $D(a_1, \ldots, a_D)$, whose density is

$$\pi(p_1, \ldots, p_D) = \frac{a!}{\prod_{j=1}^D a_j!} \prod_{j=1}^D p_j^{a_j}$$

[1] Dipartimento di Statistica e Matematica Applicata "Diego De Castro", Via Maria Vittoria 38, 10123, Torino, Italy. Email: carota@econ.unito.it
Acknowledgments: I would like to thank the organizers of the 6th International Workshop on Objective Bayesian Methodology (Rome, 8–12 June 2007) who invited me to discuss the paper by Professor Aitkin, and Piercesare Secchi for the useful references he kindly sent me.

where $a = \sum_{j=1}^{D} a_j$ is called the total prior weight. Given an n-dimensional simple random sample $\mathbf{y}$ (n small compared to N), the multinomial likelihood turns out to be

$$L(p_1, \ldots, p_D|\mathbf{y}) \propto \prod_{j=1}^{D} p_j^{n_j} = \prod_{j=1}^{d} p_j^{n_j}$$

where the n_j are the sample counts and d denotes the number of distinct values observed in the sample; consequently the posterior of $p_1, \ldots, p_D$ is the updated Dirichlet distribution $\mathcal{D}(a_1 + n_1, \ldots, a_D + n_D)$, with $n_j > 0$ only if Y_j has been observed in the sample. The Bayesian bootstrap approach consists in simulating from this posterior to obtain the posterior distribution of any population parameter.

Special cases of the Dirichlet prior are the Ericson prior $\mathcal{D}(\epsilon_1, \ldots, \epsilon_D)$, with parameters all near to zero and such that $\sum_1^D \epsilon_j$ is small, and the Haldane prior $\mathcal{D}_H = \mathcal{D}(0, \ldots, 0)$, which produces a posterior Dirichlet distribution with nonintegrable spikes at the zero values of the p_j corresponding to any unobserved value Y_j, unless all the possible distinct values of Y are assumed to be observed ($d = D$). Under this last assumption one can simulate, according to the Aitkin prescriptions, from a Dirichlet posterior with d parameters given by the sample counts, from now on denoted by $\mathcal{D}_{H|\mathbf{y}} = \mathcal{D}(n_1, \ldots, n_d)$. "Both the Bayesian bootstrap and the bootstrap operate under this assumption" (Rubin 1981, p. 133).

Given this premiss, Aitkin first of all compares the results obtained by the Bayesian bootstrap approach with an Ericson-type Dirichlet prior and with a Haldane prior and the results obtained from different methods, in particular by the classical bootstrap approach or derived under the assumption of suitable parametric distributions for Y. The comparison follows the precept (Rubin 1987, p. 62) that, to be useful, Bayes procedures need to be well calibrated in the frequentist sense. Surprisingly, the improper Haldane prior turns out to produce better results than the Ericson-type Dirichlet prior which is a proper distribution. Then, throughout the paper the Haldane prior is adopted and similar comparisons are made in a number of increasingly complex cases. The computational effort required by the Bayesian bootstrap approach appears to be moderate and interestingly insertable into standard statistical packages. However, there are two open problems outlined by the author himself (Section 8):

- the choice of the population parameter cannot be discussed from within this Bayesian bootstrap approach, since the comparison of different choices would require different likelihoods for different models whereas there is only one multinomial likelihood with unconstrained parameters;
- assumptions such as model constraints on the multinomial probabilities are difficult to implement in such a Bayesian framework.

In this article we provide reasonable solutions to these open problems by assigning a suitable informative prior to the population proportions. Moreover we clarify the reason for the better performance of the Haldane prior with respect to the Ericson-type Dirichlet prior and build a bridge between the choice of the multinomial model for the response Y and the choice of the population parameter, i.e., of the feature of Y taken to be the object of interest. Through examples drawn from the Aitkin's paper itself this last point is illustrated in the rest of this section and the remaining two points in the next one; section 3 contains some concluding remarks.

When it comes to the choice of the population parameter Aitkin (2008) in complex problems adopts the so-called "working model" strategy (Valliant, Dorfman, and Royall 2000) according to which a "working" probability model leads to an optimal estimator under the model, which is then used to define the population parameter without the working model being assumed to hold. For instance, in the last example in section 7 it is declared: "We adopt the maximum likelihood estimators of fixed effects and variance components (from the usual two level normal model) as defining the population parameters . . ., but without the assumption of normality." Moreover, in all cases discussed by Aitkin (2008), specific parametric models play a further role: not only they can be seen as "inspiring" the choice of the population parameter, they are also systematically assumed as relevant reference models in the comparison between Bayesian bootstrap analysis results and results obtained by different methods (see, e.g., Table 1, Table 3 and comments in Sections 6 and 7). These facts allow us to realize that under all choices of the population parameters there are indeed precise parametric models which are totally ignored in the specification of the statistical model for the response Y. The proposal of a suitable informative prior for the multinomial probabilities originates from this consideration. It will be illustrated in the next section.

2. A Solution to the Open Problems

2.1. The Comparison of Different Choices of the Population Parameter

We reconsider the well-known short-stay hospital example (Herson 1976, discussed by Aitkin in Section 3) where the response Y denotes the number of patients discharged by hospitals with fewer than 1,000 beds in one year and the explanatory variable X is the known number of hospital beds in that year. Let us suppose we have doubts about two possible choices of the population parameter: the ratio regression coefficient,

$$B_1 = \frac{\sum_j^D Y_j p_j}{\sum_j^D X_j p_j}$$

and the different regression coefficient

$$B_2 = \sum_j^D \frac{Y_j}{X_j} p_j$$

From a model-based viewpoint B_1, or B_2, is the optimal choice (in the least squared sense) when the set of assumptions (1), or (2), holds

$$E(Y) = BX \text{ and } Var(Y) = \sigma^2 X \tag{1}$$

$$E(Y) = BX \text{ and } Var(Y) = \sigma^2 X^2 \tag{2}$$

Furthermore, under the assumption of normality (made by Aitkin (2008) in order to make comparisons at the end of the section) B_1 and B_2 appear as suggested by their maximum likelihood estimators respectively under the normal model $M_1 = N(B_1X, \sigma^2X)$ and the normal model $M_2 = N(B_2X, \sigma^2X^2)$, whose difference lies in the structure of the response variance.

A natural way to declare this uncertainty about $Var(Y)$, without assuming the full parametric model M_1 or M_2, is by embedding them in two different Dirichlet priors for the population proportions $p_1, \ldots, p_D$,

$$\mathcal{D}_1 = \mathcal{D}(a_{11}, \ldots a_{1D}) \text{ and } \mathcal{D}_2 = \mathcal{D}(a_{21}, \ldots a_{2D})$$

through a suitable specification of the D-dimensional parameters $(a_{11}, \ldots, a_{1D})$ and $(a_{21}, \ldots, a_{2D})$, respectively. This, of course, implies a contextual "discretization" of the models M_1 and M_2.

Some interesting specifications of the parameters $a_{i,j}$, $j = 1, \ldots, D$, of the prior $\mathcal{D}_i$, $i = 1, 2$, are given below:

$$\text{U)} \quad a_{i,j}(B_i, \sigma^2) = a_{M_i}\left[\Phi\left(Y_j + \frac{\delta}{2}|B_i, \sigma^2\right) - \Phi\left(Y_j - \frac{\delta}{2}|B_i, \sigma^2\right)\right] = a_{M_i}\pi_j(B_i, \sigma^2)$$

where (B_i, σ^2) is assigned the prior $p(B_i, \sigma^2)$, $\Phi()$ is the distribution function corresponding to the normal model M_i, δ is the distance between two consecutive possible values of Y and $\sum_{j=1}^{D}\pi_j \simeq 1$. So doing, $\pi_j()$, derived from the normal model M_i (or more generally from any underlying parametric model), turns out to be the prior mean of the population proportion p_j, $j = 1, \ldots, D$, and the parameters $a_{i,j}()$ can be interpreted in terms of counts of "conceptual observations" from the discretized version $\pi_j()$, $j = 1, \ldots, D$, of the model M_i, while the total prior weight $a_{M_i} = \sum_1^D a_{i,j}$, which controls the prior variance of p_j, turns out to express our prior degree of belief in this discretized version. Note also that a_{M_i} is consequently interpretable as the "size" of the above-mentioned conceptual sample (this interpretation will appear particularly appropriate later in the discussion of the posterior corresponding to this prior specification) and that it is independent of (B_i, σ^2).

$$\text{V)} \quad a_{i,j}(\hat{B}_i, \hat{\sigma}^2) = a_{M_i}\left[\Phi\left(Y_j + \frac{\delta}{2}\middle|\hat{B}_i, \hat{\sigma}^2\right) - \Phi\left(Y_j - \frac{\delta}{2}\middle|\hat{B}_i, \hat{\sigma}^2\right)\right] = a_{M_i}\pi_j(\hat{B}_i, \hat{\sigma}^2)$$

where, all other things being equal, $(\hat{B}_i, \hat{\sigma}^2)$ is the maximum likelihood estimate of (B_i, σ^2).

$$U^*) \; a_{i,j}(B_i, \sigma^2) = a_{M_i}[\delta\phi(Y_j|B_i, \sigma^2)], \quad \text{or} \quad \text{V}^*) \; a_{i,j}(\hat{B}_i, \hat{\sigma}^2) = a_{M_i}[\delta\phi(Y_j|\hat{B}_i, \hat{\sigma}^2)]$$

where $\phi()$ is the probability density function corresponding to the normal model M_i, so that, all other things being equal, these are histogram-based approximations of the specifications U and V, respectively.

In all these cases, from the informative Dirichlet priors, $\mathcal{D}_1$ and $\mathcal{D}_2$, and the multinomial model for Y we can obtain two different likelihoods for the different choices of the population parameter B_1 or B_2 (inspired by the different underlying models M_1 and M_2),

$$L_1 = \frac{\Gamma(a_{M_1})}{\Gamma(a_{M_1}+n)} \iint \prod_{1}^{d} \frac{\Gamma(a_{1,j}+n_j)}{\Gamma(a_{1,j})} p(B_1,\sigma^2)dB_1 d\sigma^2$$

$$L_2 = \frac{\Gamma(a_{M_2})}{\Gamma(a_{M_2}+n)} \iint \prod_{1}^{d} \frac{\Gamma(a_{2,j}+n_j)}{\Gamma(a_{2,j})} p(B_2,\sigma^2)dB_2 d\sigma^2$$

For instance, with the short-stay hospital data, if we assume that $a_{M_1} = a_{M_2} = a$ and adopt the specification V* (according to which $p(\hat{B}_i, \hat{\sigma}^2) = 1$), the Bayes factor in favour of B_1 against B_2 is given by

$$\begin{aligned} BF_{V^*,12} &= \frac{L_1}{L_2} = \frac{\prod_{j=1}^{n=32} a\delta\phi(y_j|\hat{B}_1,\hat{\sigma}^2) + \prod_{j=1}^{d=31} a\delta\phi(y_j|\hat{B}_1,\hat{\sigma}^2)}{\prod_{j=1}^{n=32} a\delta\phi(y_j|\hat{B}_2,\hat{\sigma}^2) + \prod_{j=1}^{d=31} a\delta\phi(y_j|\hat{B}_2,\hat{\sigma}^2)} \\ &= \frac{(a\delta)^{32} 5.837\,10^{-71}/\pi^{16} + (a\delta)^{31} 1.740\,10^{-68}/\pi^{15.5}}{(a\delta)^{32} 1.198\,10^{-72}/\pi^{16} + (a\delta)^{31} 4.640\,10^{-70}/\pi^{15.5}} \\ &= \frac{48.70 + (a\delta)^{-1} 14516.89\sqrt{\pi}}{1 + (a\delta)^{-1} 387.18\sqrt{\pi}} \end{aligned}$$

since $(\hat{B}_1, \hat{\sigma}^2) = (3.200, 176.316)$ and $(\hat{B}_2, \hat{\sigma}^2) = (3.349, 1.189)$. Analogously, if we specify the parameters $a_{i,j}$ according to U* and assume that $(B_i,\ \sigma^2)$ is distributed according to the noninformative prior $p(B_i, \sigma^2) \propto 1/\sigma^2$, all other things being equal, we obtain the following exact expression for the Bayes factor

$$BF_{U^*,12} = \frac{L_1}{L_2} = \frac{\iint \prod_{j=1}^{n=32} a\delta\phi(y_j|B_1,\sigma^2)\frac{1}{\sigma^2}dB_1 d\sigma^2 + \iint \prod_{j=1}^{d=31} a\delta\phi(y_j|B_1,\sigma^2)\frac{1}{\sigma^2}dB_1 d\sigma^2}{\iint \prod_{j=1}^{n=32} a\delta\phi(y_j|B_2,\sigma^2)\frac{1}{\sigma^2}dB_2 d\sigma^2 + \iint \prod_{j=1}^{d=31} a\delta\phi(y_j|B_2,\sigma^2)\frac{1}{\sigma^2}dB_2 d\sigma^2}$$

$$= \frac{\dfrac{(a\delta)^{32}\left(\sum_{j=1}^{n=32} x_j \prod_{j=1}^{n=32} x_j\right)^{-\frac{1}{2}} \Gamma\left(\frac{n-1}{2}\right)}{(2\pi)^{\frac{n-1}{2}} \left\{\frac{1}{2}\left[\sum_{j=1}^{n=32} \frac{y_j^2}{x_j} - \frac{\left(\sum_{j=1}^{n=32} y_j\right)^2}{\sum_{j=1}^{n=32} x_j}\right]\right\}^{\frac{n-1}{2}}} + \dfrac{(a\delta)^{31}\left(\sum_{j=1}^{d=31} x_j \prod_{j=1}^{d=31} x_j\right)^{-\frac{1}{2}} \Gamma\left(\frac{d-1}{2}\right)}{(2\pi)^{\frac{d-1}{2}} \left\{\frac{1}{2}\left[\sum_{j=1}^{d=31} \frac{y_j^2}{x_j} - \frac{\left(\sum_{j=1}^{d=31} y_j\right)^2}{\sum_{j=1}^{d=31} x_j}\right]\right\}^{\frac{d-1}{2}}}}{\dfrac{(a\delta)^{32} n^{-\frac{1}{2}}\left(\prod_{j=1}^{n=32} x_j\right)^{-1} \Gamma\left(\frac{n-1}{2}\right)}{(2\pi)^{\frac{n-1}{2}} \left\{\frac{1}{2}\left[\sum_{j=1}^{n=32} \frac{y_j^2}{x_j^2} - \frac{1}{n}\left(\sum_{j=1}^{n=32} \frac{y_j}{x_j}\right)^2\right]\right\}^{\frac{n-1}{2}}} + \dfrac{(a\delta)^{31} d^{-\frac{1}{2}}\left(\prod_{j=1}^{d=31} x_j\right)^{-1} \Gamma\left(\frac{d-1}{2}\right)}{(2\pi)^{\frac{d-1}{2}} \left\{\frac{1}{2}\left[\sum_{j=1}^{d=31} \frac{y_j^2}{x_j^2} - \frac{1}{d}\left(\sum_{j=1}^{d=31} \frac{y_j}{x_j}\right)^2\right]\right\}^{\frac{d-1}{2}}}}$$

$$= \frac{(a\delta)^{32} 9.965\,10^{-86}/\pi^{15.5} + (a\delta)^{31} 3.256\,10^{-83}/\pi^{15}}{(a\delta)^{32} 2.670\,10^{-87}/\pi^{15.5} + (a\delta)^{31} 1.180\,10^{-84}/\pi^{15}} = \frac{37.33 + (a\delta)^{-1} 12196.64\sqrt{\pi}}{1 + (a\delta)^{-1} 442.16\sqrt{\pi}}$$

In both cases the huge values of these Bayes factors for all choices of δ and a strongly induce us to select the ratio regression coefficient as the parameter of interest for inference. Their expressions are noticeably simple since the counts in the considered data set are “small”, that is to say since each ratio of Gamma functions in the two likelihoods L_1 and L_2 reduces to a "small" number of terms $(a_{i,j}+n_j-1)\cdots(a_{i,j}+1)a_{i,j}$, $j=1,\ldots,d$, $i=1,2$. These results are special cases of the Bayes factor for nonparametric model selection (Carota 1999) where the unknown distribution function of the response variable under both the null and the alternative hypothesis is assumed to be a mixture of Dirichlet processes (see also Carota and Parmigiani 1996). In a similar way we can obtain the Bayes factor corresponding to the specifications V) and U) of the parameters $a_{i,j}$ or corresponding to different underlying models as it is illustrated below.

A visual inspection of the data suggests that probably a valid alternative structure for the response variance is $Var(Y)=\sigma^2X^r$ with $r<1$, instead of $r=2$ as mentioned by Aitkin. If we denote by $M_3=N(B_3X,\sigma^2X^r)$ the normal model "working" in this case and suggesting the choice of the population parameter

$$\mathrm{B}_3=\frac{\sum_j^D Y_jX_j^{1-r}p_j}{\sum_j^D X_j^{2-r}p_j}$$

by assuming the specification V* and $a_{M_1}=a_{M_3}=a$, we can obtain the Bayes factor in favour of B_1 against B_3 for different values of r. For instance, for $r=0.9$ and $r=0.8$ we have

$$BF_{V^*,13}(r=0.9)=\frac{L_1}{L_2}=\frac{1.31+(\delta a)^{-1}389.77\sqrt{\pi}}{1+(\delta a)^{-1}296.41\sqrt{\pi}}$$

$$BF_{V^*,13}(r=0.8)=\frac{L_1}{L_2}=\frac{1.85+(\delta a)^{-1}551.47\sqrt{\pi}}{1+(\delta a)^{-1}295.46\sqrt{\pi}}$$

whose values, though still in favour of the ratio regression coefficient B_1, are far from being strongly conclusive for all values of δ and a as the values of $BF_{V^*,12}$.

It is also worth noting that the parameters of the posterior corresponding to the informative prior $\mathcal{D}_i$, denoted by $\mathcal{D}_{i|\mathbf{y}}=\mathcal{D}(a_{i1}+n_1,\ldots,a_{i\mathcal{D}}+n_{\mathcal{D}})$, $i=1,2$, are a compromise between the "conceptual" counts $a_{i,j}$, based on the normal model M_i, and the sample counts n_j, and that the total posterior weight $a_{Mi}+n$, in turn, is a compromise between sizes of the "conceptual" sample and the real sample.

Let us now consider the impact of the informative prior $\mathcal{D}_i$ on the Bayesian bootstrap. If one simulates from the posterior $\mathcal{D}_{i|\mathbf{y}}$ according to the Aitkin prescriptions the additional computational effort required with respect to simulating from $\mathcal{D}_{H|\mathbf{y}}$ ranges from none (if the $a_{i,j}$ are specified according to V^* and the assumption $\mathcal{D}=d$ were true) to the inclusion of one more step in the simulation (if the $a_{i,j}$ are specified according to U or U*). The right balance between additional computational effort and degree of approximation in the estimation of the discretized normal model $\pi_j(B_i,\sigma^2)$, $j=1,\ldots,D$, $(B_i,\sigma^2)\sim p(B_i,\sigma^2)$ nested in the Dirichlet prior distribution should be found through a sensible and practical choice of the $a_{i,j}$; often, however, the very simple specification V* turns out to be appropriate. As regards the advantages of using the informative prior $\mathcal{D}_i$ instead of the Haldane prior, for all specifications of the $a_{i,j}$, $\mathcal{D}_i$ allows us not only to compare different choices of the population parameter, but also to handle predictive problems without the well-known difficulties outlined by Rubin (1981) and to avoid the dichotomy between working model strategy (or survey sampling approach in simple cases) for the choice of the population parameter and Bayesian bootstrap approach for the analysis of its posterior distribution.

2.2. *Informative or Noninformative Prior?*

Let us now turn to explain the increasingly poor performance of the informative Ericson-type Dirichlet prior, from now on denoted by $\mathcal{D}_E$, as the prior weight on the unobserved values increases with respect to the noninformative prior $\mathcal{D}_H$. This aspect is discussed by Aitkin (2008) in the income population example and judged to be "surprising".

We recall that the structure of $\mathcal{D}_E$ is assumed to be

$$\mathcal{D}_E = \mathcal{D}(\epsilon, \ldots, \epsilon)$$

where all parameters are given by the same "small" constant ϵ and their number, d^* say, is the number of distinct values which the response Y can take in the observed range $[y_{(1)},y_{(n)}]$, where $y_{(1)}$ and $y_{(n)}$ respectively denote the smallest and the largest observation in the sample. Recalling also that in the income population example the underlying ignored parametric model is a slightly perturbed normal or a gamma distribution, the reason for the increasingly poor performance of $\mathcal{D}_E$, as $\sum_1^{d^*} \epsilon = \epsilon \times d^* = e$ increases, appears to be that $\mathcal{D}_E$ becomes a more and more "misinformative" prior. Actually, its parameters (constrained all to take the same value and corresponding to a uniform distribution on $[y_{(1)},y_{(n)}]$) more and more conflict with the genuine prior information suggesting to assume them as roughly proportional to the probability assigned to the d^* intervals $(Y_j - \delta/2, Y_j + \delta/2]$ by a suitable normal or gamma distribution.

By way of contrast, let us suppose we adopt an informative prior $\mathcal{D}_i$ with parameters $a_{i,j}$ proportional to the probability assigned to the D intervals $(Y_j - \delta/2, Y_j + \delta/2]$ by the same normal model, denoted by M_1, and, successively, by the same gamma model, denoted by M_2, employed by Aitkin (2008) to obtain the confidence intervals which are compared in Table 1 to the intervals derived from the Bayesian bootstap approach with an Ericson-type Dirichlet prior and with a Haldane prior. If we denote the two proportionality constants, or prior degree of belief in the discretized version of the normal model and of the gamma model, respectively by a_{M_1} and a_{M_2}, then as a_{M_i}, $i = 1, 2$, decreases to zero, the Bayesian bootstrap intervals corresponding to $\mathcal{D}_i$ approximate the intervals obtained by the Bayesian bootstrap with the Haldane prior, while as a_{M_i}, $i = 1, 2$, increases, they respectively approximate the intervals obtained under the assumption of the normal model and under the assumption of the gamma model. This conclusion can be easily drawn by interpreting the Dirichlet distribution $\mathcal{D}_i$ as a special case of the more general Dirichlet process prior for a response variable which is discrete and finitely supported. Under more general assumptions an interesting Bayesian resampling plan (proper Bayesian bootstrap) is described in Muliere and Secchi (1995, 1996).

In other words, the true misleading feature of the informative prior $\mathcal{D}_E$ is the structure of its parameters, constrained all to take the same value ϵ, before the amount of the total prior weight e or the amount of weight assigned to the $(d^* - d)$ unobserved values of Y. When a suitable informative Dirichlet prior, like $\mathcal{D}_i$, is adopted the total prior weight simply states how large is the dispersion around the discretized version of the parametric model employed to specify the parameters $a_{i,j}$, and the range of the corresponding Bayesian bootstrap results is roughly bounded by the results obtained by assuming the underlying parametric model, on the one hand, and the results obtained from the Bayesian bootstrap approach with the Haldane prior, on the other.

One more point to be clarified concerns the appellative of post-data prior reserved to $\mathcal{D}_E$ when, as a matter of fact, the Haldane prior used by Aitkin (2008) to implement the Bayesian bootstrap is itself a post-data prior. The specification of $\mathcal{D}_H$ requires the number of distinct values d in the sample to be known in order to realize the great saving of not considering all the D possible values of Y and, simultaneously, avoid the nonintegrable spikes in the corresponding posterior, just as the specification of $\mathcal{D}_E$ requires the observed range to be known in order to derive the value of d^*. Analogously, the informative prior $\mathcal{D}_i$ turns out to be a post-data prior whenever the $a_{i,j}$ are specified by using the maximum likelihood estimate of the underlying parametric model, while $\mathcal{D}_i$ is entirely assigned before looking at the data when we adopt, for instance, the specifications U or U^*.

2.3. Constraints on the Multinomial Probabilities

The difficulty of considering constraints on the multinomial probabilities is illustrated by Aitkin (2008) by reconsidering the normal variance component model assumed by Box and Tiao (1973, p. 246) for the dyestuff data. This is a model with only two components of variance, the "within-batch" variances and the "among-batches" variance where, moreover, the within-batch variances are taken to be the same across batches. For the same data Aitkin adopts the working model strategy to define the parameters of interest, derives their estimates from the Bayesian bootstrap approach (with the Haldane prior) and observes that there is no way in this framework to take the variance homogeneity into account. Here the arguments in Subsection 2.2 suggest that a way to introduce the variance homogeneity assumption (and more generally constraints on the multinomial probabilities) in our Bayesian bootstrap approach (we assume the informative prior $\mathcal{D}_i$) is by graduating the total prior weight $a_{M_i} = \sum_1^D a_{i,j}$, provided that the parameters $a_{i,j}$ of $\mathcal{D}_i$ are specified according to the probabilities assigned to $(Y_j - \delta/2, Y_j + \delta/2]$, $j = 1, \ldots, D$, by the Box and Tiao normal variance component model. This implies that the more the total prior weight a_{M_i} increases, the more the homogeneity variance assumption is close to be true since it is inherent in the Box and Tiao model and, consequently, in its discretized version. Of course this is an indirect way of introducing constraints on the multinomial probabilities, but it represents one step ahead with respect to the situation described by Aitkin (2008).

3. Concluding Remarks

This article shows that by connecting the parametric model underlying the choice of the population parameter and the multinomial model for the response variable Y (that is to say by nesting such a parametric model in the Dirichlet prior for the multinomial probabilities) we can give a completely satisfying answer to the first open problem outlined by Aitkin (2008) and a partial indirect answer to the second one. We have in fact provided a flexible method to obtain:

- different likelihoods for different choices of the population parameter which, thus, can be compared from within our modified Bayesian bootstrap approach, and
- a way to introduce model constraints on the multinomial probabilities.

Moreover the informative prior obtained through the nesting described above, $\mathcal{D}_i$, can be contrasted with the Haldane prior $\mathcal{D}_H$ more interestingly than the Ericson-type Dirichlet prior $\mathcal{D}_E$. It allows us to explain the poor performance of $\mathcal{D}_E$ with respect to $\mathcal{D}_H$, to overcome the severe limitations of the Bayesian bootstrap due to the unrealistic assumption that all the possible distinct values of the response are observed in the sample ($d = D$) and to bridge the gap between the Bayesian bootstrap analysis and the fully parametric model analysis. From a logical point of view the informative prior $\mathcal{D}_i$ also bridges the gap between the choice of multinomial model and choice of the population parameter which represent two completely separate starting points in Aitkin (2008) where the Bayesian bootstrap analysis is founded on this dichotomy.

Finally, as regards the saving of computational effort, probably a useful compromise between $\mathcal{D}_i$ and $\mathcal{D}_E$ can be realized by suitably reducing from D to d^* the number of parameters in the informative Dirichlet prior $\mathcal{D}_i$.

4. References

Aitkin, M. (2008). Applications of the Bayesian Bootstrap in Finite Population Inference. Journal of the Official Statistics, 24, 21–51.

Box, G.E.P. and Tiao, G.C. (1973). Bayesian Inference in Statistical Analysis. New York: Wiley.

Carota, C. and Parmigiani, G. (1996). On Bayes Factors for Nonparametric Alternatives. In Bayesian Statistics V: Fifth Valentia International Meeting on Bayesian Statistics, J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith (eds). London: Oxford University Press.

Carota C. (1999). Some Results on Bayes Factors in a Nonparametric Context. In Proceedings of the Section on Bayesian Statistical Science at the Annual Meeting of the American Statistical Association, Baltimore.

Herson, J. (1976). An Investigation of Relative Efficiency of Least Squares Prediction to Conventional Probability Sampling Plans. Journal of the American Statistical Association, 71, 700–703.

Muliere, P. and Secchi, P. (1995). Bayesian Nonparametric Predictive Inference and Bayesian Techniques. Annals of the Institute of Statistical Mathematics, 48(4), 663–673.

Muliere, P. and Secchi, P. (1996). Alcune osservazioni sul metodo bootstrap in ambito bayesiano. Studi in onore di Giampiero Landenna. Bologna: Giuffré [In Italian].

Rubin, D.B. (1981). The Bayesian Bootstrap. Annals of Statistics, 9, 130–134.

Rubin, D.B. (1987). Multiple Imputation for Nonresponses in Surveys. New York: Wiley.

Valliant, R., Dorfman, A.H., and Royall, R.M. (2000). Finite Population Sampling and Inference: A Prediction Approach. New York: Wiley.

Received November 2007
Revised December 2008