



AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Large sample properties of Gibbs-type priors**This is a pre print version of the following article:***Original Citation:**Availability:*

This version is available <http://hdl.handle.net/2318/115385> since

Publisher:

CLEUP

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

Large sample properties of Gibbs-type priors

Pierpaolo De Blasi, Antonio Lijoi and Igor Prünster

Abstract In this paper we concisely summarize some recent findings that can be found in [1] and concern large sample properties of Gibbs-type priors. We shall specifically focus on consistency according to the frequentist approach which postulates the existence of a “true” distribution P_0 that generates the data. We show that the asymptotic behaviour of the posterior is completely determined by the probability of obtaining a new distinct observation. Exploiting the predictive structure of Gibbs-type priors, we are able to establish that consistency holds essentially always for discrete P_0 , whereas inconsistency may occur for diffuse P_0 . Such findings are further illustrated by means of three specific priors admitting closed form expressions and exhibiting a wide range of asymptotic behaviours.

Key words: Asymptotics, Bayesian nonparametrics, Gibbs-type priors

1 Gibbs-type priors

In this paper we sketch results that are extensively presented and proved in [1] about the asymptotic posterior behaviour of Gibbs-type priors, a class of discrete nonparametric priors recently introduced in [5]. Gibbs-type priors can be defined through the system of predictive distributions they induce. To this end, let $(X_n)_{n \geq 1}$ be an (ideally) infinite sequence of observations, with each X_i taking values in a complete and separable metric space \mathbb{X} . Moreover, $\mathbf{P}_{\mathbb{X}}$ is the set of all probability measures on \mathbb{X} endowed with the topology of weak convergence. In the most commonly employed Bayesian models $(X_n)_{n \geq 1}$ is assumed to be *exchangeable* which means there exists a

Pierpaolo De Blasi (pierpaolo.deblasi@unito.it), Igor Prünster (igor@econ.unito.it)
University of Torino, Department of Economics and Statistics, corso Unione Sovietica 218/bis,
Torino 10134, Italy

Antonio Lijoi (lijoi@unipv.it)
University of Pavia, Department of Economics and Business, via San Felice 5, Pavia 27100, Italy

probability distribution Q on $\mathbf{P}_{\mathbb{X}}$ such that $X_i|\tilde{p} \stackrel{\text{iid}}{\sim} \tilde{p}$, $\tilde{p} \sim Q$. Hence, \tilde{p} is a random probability measure on \mathbb{X} whose probability distribution Q is also termed *de Finetti measure* and acts as a prior for Bayesian inference. Given a sample (X_1, \dots, X_n) , the predictive distribution coincides with the posterior expected value of \tilde{p} , that is

$$\text{pr}(X_{n+1} \in \cdot | X_1, \dots, X_n) = \int_{\mathbf{P}_{\mathbb{X}}} p(\cdot) Q(dp | X_1, \dots, X_n). \quad (1)$$

A prior Q that selects, almost surely, discrete distributions is said discrete and, in this case, a sample (X_1, \dots, X_n) will feature ties with positive probability: X_1^*, \dots, X_k^* denote the $k \leq n$ distinct observations and n_1, \dots, n_k their frequencies for which $\sum_{i=1}^k n_i = n$. Gibbs-type priors are discrete and characterized by predictive distributions (1) of the form

$$\text{pr}(X_{n+1} \in \cdot | X_1, \dots, X_n) = \frac{V_{n+1,k+1}}{V_{n,k}} P^*(\cdot) + \frac{V_{n+1,k}}{V_{n,k}} \sum_{i=1}^k (n_i - \sigma) \delta_{X_i^*}(\cdot), \quad (2)$$

where $\sigma \in (-\infty, 1)$, $P^*(dx) := E[\tilde{p}(dx)]$ is a diffuse probability measure representing the prior guess at the shape of \tilde{p} and $\{V_{n,k} : k = 1, \dots, n; n \geq 1\}$ is a set of non-negative weights satisfying the recursion

$$V_{n,k} = (n - \sigma k) V_{n+1,k} + V_{n+1,k+1}. \quad (3)$$

Therefore, Gibbs-type priors are characterized by predictive distributions, which are a linear combination of the prior guess and a weighted version of the empirical measure. The most widely known priors within this class are the Dirichlet process [3] and the two-parameter Poisson-Dirichlet process [8].

2 Consistency results

We address posterior consistency according to the “what if” approach of [2], which consists in assuming that the data $(X_n)_{n \geq 1}$ are independent and identically distributed from some “true” $P_0 \in \mathbf{P}_X$ and in verifying whether the posterior distribution $Q(\cdot | X_1, \dots, X_n)$ accumulates in any neighborhood of P_0 , under a suitable topology. Since Gibbs-type priors are defined on $\mathbf{P}_{\mathbb{X}}$ and are discrete, the appropriate notion of convergence is convergence in the weak topology. Therefore, we aim at establishing whether $Q(A_\varepsilon | X_1, \dots, X_n) \rightarrow 1$, a.s.- P_0^∞ , as $n \rightarrow \infty$ and for any $\varepsilon > 0$, where A_ε denotes a weak neighborhood of some $P' \in \mathbf{P}_X$ of radius ε and P_0^∞ is the infinite product measure. Clearly, consistency corresponds to $P' = P_0$.

We prove a general structural result on Gibbs-type priors showing that the posterior distribution converges to a point mass at the weak limit, in an almost sure sense, of the predictive distribution (2). To this aim, we shall assume that the probability of recording a new distinct observation at step $n+1$

$$\frac{V_{n+1, \kappa_n+1}}{V_{n, \kappa_n}} \quad \text{converges} \quad \text{a.s.-}P_0^\infty \quad (\text{H})$$

as $n \rightarrow \infty$, and the limit is identified by some constant $\alpha \in [0, 1]$. We use the notation κ_n in order to make explicit the dependence on n of the number of distinct observations in a sample of size n . Different choices of P_0 yield different limiting behaviours for κ_n : if P_0 is discrete with N point masses, then $P_0^\infty(\lim_n n^{-1} \kappa_n = 0) = 1$ even when $N = \infty$; if P_0 is diffuse, $P_0^\infty(\kappa_n = n) = 1$ for any $n \geq 1$. The following theorem shows that (H) is actually sufficient to establish weak convergence at a certain P' that is explicitly identified. The key ingredient for the proof is represented by an upper bound on the posterior variance $\text{Var}[\tilde{p}(A) | X_1, \dots, X_n]$, which is of independent interest. See [1].

Theorem 1. *Let \tilde{p} be a Gibbs-type prior with prior guess $P^* = E[\tilde{p}]$, whose support coincides with \mathbb{X} , and assume condition (H) holds true. If $(X_i)_{i \geq 1}$ is a sequence of independent and identically distributed random elements from P_0 then the posterior converges weakly, a.s.- P_0^∞ , to a point mass at $\alpha P^*(\cdot) + (1 - \alpha)P_0(\cdot)$.*

According to Theorem 1, weak consistency is achieved in the trivial case of $P^* = P_0$, which will be excluded henceforth, and when $\alpha = 0$: therefore, it is sufficient to check whether the probability of obtaining a new observation, given previously recorded data, converges to 0, a.s.- P_0^∞ . One might also wonder whether there are circumstances leading to $\alpha = 1$, which corresponds to the posterior concentrating around the prior guess P^* , a situation we refer to as “total” inconsistency.

Since a few particular cases of Gibbs-type priors with $\sigma \in (0, 1)$ have already been considered in [7] and [6], attention is focused on the case of $\sigma \in (-\infty, 0)$ for which nothing is known to date. We recall here that, if $\sigma < 0$, Q is a mixture of Poisson-Dirichlet processes with parameters $(\sigma, k|\sigma|)$ and the mixing distribution for k , say π , is supported by the positive integers. Since in the case of negative σ the two-parameter model coincides with a x -variate symmetric Dirichlet distribution, one can describe such Gibbs-type priors as

$$(\tilde{p}_1, \dots, \tilde{p}_k) \sim \text{Dirichlet}(|\sigma|, \dots, |\sigma|) \quad (4)$$

$$k \sim \pi(\cdot)$$

We shall restrict attention to Gibbs-type priors whose realizations are discrete distributions whose support has a cardinality that cannot be bounded by any positive constant, almost surely. This is the same as assuming that the support of π in (4) is \mathbb{N} . Note that for the “parametric” case of $\sigma < 0$ and π supported by a finite subset of \mathbb{N} one immediately has consistency for any P_0 in its support by the results of [4]. Theorem 2 gives neat sufficient conditions for consistency in terms of the tail behaviour of the mixing distribution π on the positive integers \mathbb{N} in (4).

Theorem 2. *Let \tilde{p} be a Gibbs-type prior with parameter $\sigma < 0$, mixing measure π and prior guess P^* whose support coincides with \mathbb{X} . Then the posterior is consistent*

- (i) *at any discrete P_0 if for sufficiently large x*

$$\pi(x+1)/\pi(x) \leq 1; \quad (T1)$$

(ii) at any diffuse P_0 if for sufficiently large x and for some $M < \infty$

$$\pi(x+1)/\pi(x) \leq M/x. \quad (T2)$$

Note that condition (T1) is an extremely mild assumption on the regularity of the tail of the mixing π : it requires $x \mapsto \pi(x)$ to be ultimately decreasing, a condition met by the commonly used probability measures on \mathbb{N} . On the other hand, condition (T2) requires the tail of π to be sufficiently light. This is indeed a binding condition and it is particularly interesting to note that such a condition is also close to being necessary. In [1], three different Gibbs-type priors with $\sigma = -1$ are considered, each prior characterized by a specific choice of the mixing distribution π . These examples show that, according as to heaviness of the tails of π , the value of α in Theorem 1 may actually span the whole interval $[0, 1]$, from situations where consistency holds true ($\alpha = 0$) to cases where “total” inconsistency occurs ($\alpha = 1$). In particular, the heavier the tail of π and the larger α , i.e. the lighter is the weight assigned to the “true” P_0 in the limiting distribution identified in Theorem 1. The first prior is characterized by a heavy-tailed mixing distribution π , which does not admit a finite expected value: condition (T2) is not met and it turns out that $\alpha = 1$ so that the posterior concentrates around the prior guess P^* (“total” inconsistency). The second specific prior, where the mixing π has light tails that satisfy (T2) in Theorem 2, results in a consistent asymptotic behaviour. In the third case α takes values over the whole unit interval $[0, 1]$ according to a parameter that determines the heaviness of the tail of π .

References

1. De Blasi, P., Lijoi, A., Prünster, I.: An asymptotics analysis of a class of discrete nonparametric priors. Technical Report (2012)
2. Diaconis, P., Freedman, D.: On the consistency of Bayes estimates. *Ann. Statist.* **14**, 1-26 (1986)
3. Ferguson, T. S.: A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209-230 (1973)
4. Freedman, D.: On the asymptotic behavior of Bayes’ estimates in the discrete case. *Ann. Math. Statist.* **34**, 1386-1403 (1963)
5. Gnedin, A., Pitman, J.: Exchangeable Gibbs partitions and Stirling triangles. *Journal of Mathematical Sciences* **138**, 5674-5685 (2006)
6. James, L. F.: Large sample asymptotics for the two parameter Poisson Dirichlet process. In: B. Clarke and S. Ghosal (eds.) *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Gosh*, pp. 187-199. IMS, Harward (2008)
7. Jang, G. H., Lee, J., Lee, S.: Posterior consistency of species sampling priors. *Statistica Sinica* **20**, 581-593 (2010)
8. Pitman, J.: Some developments of the Blackwell-MacQueen urn scheme. In: Ferguson, T.S., Shapley, L.S., MacQueen, J.B. (eds.) *Statistics, Probability and Game Theory. Papers in honor of David Blackwell*, pp. 245-267. Lecture Notes, Monograph Series, **30**, IMS, Harward (1996)