The hyper Dirichlet process and its discrete approximations: the single conditional independence model *

C. Asci [†], G. Nappo [‡], M. Piccioni [§]

Abstract

The aim of this paper is the study of some laws of random probability distributions, called hyper Dirichlet processes, charging the product of three sample spaces, with the property that the first and the third components are independent conditional to the second one. The law of the marginals on the first two and on the last two components are specified to be Dirichlet processes with the same marginal parameter measure on the common second component. The joint law is then obtained as the hyper Markov combination of these two laws, as introduced in [3], and in fact these laws are generalizations of the hyper Dirichlet laws on contingency tables considered in the above paper.

Our main result is the convergence to the law of a hyper Dirichlet process of the sequence of hyper Dirichlet laws associated to finer and finer "discretizations" of the two parameter measures, which is proved by means of a suitable coupling construction.

MSC: 60G57; 62G99; 62F15; 60F99.
Keywords: Dirichlet laws and processes; Bayesian statistics; Markov distributions; weak convergence; coupling.

---

*Partially supported by project *"Modelli probabilistici e applicazioni"*

†Dipartimento di Studi geoeconomici, linguistici, statistici, storici per l'analisi regionale - Università di Roma "La Sapienza", via del Castro Laurenziano 9, I 00161 Roma, Italy

‡Dipartimento di Matematica - Università di Roma "La Sapienza", piazzale A. Moro 2, I 00185 Roma, Italy

§Dipartimento di Matematica - Università di Roma "La Sapienza", piazzale A. Moro 2, I 00185 Roma, Italy

1 Introduction

The hyper Markov combination of a family of laws for random probability measures has been defined by Dawid and Lauritzen [3]. Their construction starts from a general decomposable graph, whose vertices indicates the components of the product space supporting these random distributions. This graph represents the Markov structure imposed on the realizations, specifying the conditional independence relations assumed to hold between the components. The fact that the graph is decomposable implies that, under general assumptions guaranteeing the existence of regular conditional probabilities, a Markov distribution on the whole sample space can be uniquely derived from the marginal distributions of the components belonging to each clique: indeed, it is the Markov combination of these distributions, namely the only joint distribution with the specified marginals having the prescribed Markov structure. Clearly the specified marginals have to be consistent, in the sense that when two cliques intersect, the corresponding marginals have to agree on the common margins. It is then natural to build laws of random probability distributions with a prescribed Markov structure from the family of laws of their marginal distributions over each clique. Again, when two cliques intersect, the two corresponding random distributions have to induce the same law on the common margins, i.e. such a family have to be hyperconsistent. Clearly it remains the freedom to specify the joint law of all these marginals, but there is a unique one which has the hyper Markov property. This property is naturally obtained by "lifting" the Markov property from the level of the realizations of random distributions to the level of their laws. In [3] this property is studied from various points of view, including its consequences for the Bayesian analysis of decomposable graphical models.

A particularly relevant example of this construction is the hyper Dirichlet law, which is the hyper Markov combination of hyperconsistent Dirichlet laws. Dirichlet laws are examples of random distributions over discrete sample spaces. For more general sample spaces their generalization is given by Dirichlet processes, introduced by Ferguson [5], and then studied by Blackwell and MacQueen [2] Sethuraman-Tiwari [11] and Sethuraman [10], between others. They still represent the most used class of prior distributions to deal with nonparametric models within the Bayesian framework. To our knowledge, the present paper is a first attempt to introduce genuinely multivariate properties, such as conditional independence relations, in the area of Bayesian nonparametric statistics.

In this paper we study hyper Dirichlet processes, which are hyper Markov combinations of hyperconsistent Dirichlet processes. In order to clarify the main ideas, we limit ourselves to consider a decomposable graph of the simplest kind, which consists in a non-trivial decomposition in two complete subgraphs. By merging vertices and edges in a natural way a graph of this type can be reduced to the graph with three vertices $\{1, 2, 3\}$ with cliques $\{1, 2\}$ and $\{2, 3\}$; hence 2 separates 1 and 3. This means that our aim is the construction of random probability distributions on the product of three sample spaces, with the property that the first and the third components are independent conditional to the second one. In the following the word Markov or hyper Markov is always referred to this graph. In order to construct a hyper Markov combination on this graph, we need to specify two hyperconsistent laws on the first and second components, and on the second and third components, respectively. When these are the laws of two Dirichlet processes, hyperconsistency is clearly equivalent to consistency of the two parameter measures. If the marginal parameter measure on the common second component is diffuse, a hyper Dirichlet process is itself a Dirichlet process (whose parameter measure is the Markov combination of the two given consistent parameter measures). This is surprising since hyper Dirichlet laws are never Dirichlet laws, except in trivial cases. This observation led us to study the relation between a hyper Dirichlet process and the hyper Dirichlet laws which are obtained by a natural discretization of its parameter measures. Our main result is that, under suitable assumptions, as the discretization refines, the hyper Dirichlet process appears to be the weak limit (i.e. in law) of the sequence of "discretized" hyper Dirichlet laws.

Another contribution of the paper is the result that hyper Dirichlet processes are conjugate w.r.t. i.i.d. sampling, despite the fact that they are in general not strongly hyper Markov in the sense of [3]. This property is clearly useful for classes of nonparametric priors in Bayesian inference.

The next step of our analysis would be to consider more general decomposable graphs using the well-known fact that decomposable graphs can be always constructed through a series of decompositions [7]. It is not difficult to generalize the limit result when all the Dirichlet processes to be combined have consistent parameter measures with diffuse common marginals, in which case the hyper Markov combination is again a Dirichlet process, whose the parameter measure is the Markov combination of the given consistent parameter measures. However, in the general case some technical difficulties may arise, because such a Markov combination may have a complicated structure of the marginal atoms on the separators in the perfect decomposition of the graph.

The paper is organized in the following way. In the next section, after having reviewed the main properties of Dirichlet and hyper Dirichlet laws, we introduce the hyper Dirichlet process for pairs of parameter measures which have a common diffuse marginal. In Section 3 the convergence of hyper Dirichlet laws to this process is proved, over finer and finer discretizations. The basic tool is the Sethuraman representation of a Dirichlet process, together with a coupling construction between the limit process and the discretized one. In Section 4 we remove the assumption that the common marginal is diffuse, in which case we deal with hyper Dirichlet processes which are not in general Dirichlet processes. The main tool here is a suitable mixture decomposition of the Markov combination of the two parameter measures, which yields a corresponding mixture structure for the realizations of the process. We finally prove that hyper Dirichlet processes are conjugate with respect to random i.i.d. sampling, and discuss the relation of this conjugacy property with that holding for Dirichlet process ([5], [10]).

2 The hyper Dirichlet law and process for the simplest decomposition

Dirichlet laws are supported by the unit simplices in Euclidean spaces, hence they are used as laws for random variables with values in the set of probability distributions over a finite set of atoms L. For ease of notation the set L is always identified with $\{1, ..., |L|\}$, without loss of generality. Let us suppose first that $|L| > 1$. For a given vector $\{\varphi(l) > 0, l \in L\}$, we say that the random vector $\{G(l), l \in L\}$ is $Di\,\{\varphi(l), l \in L\}$ if

$$G(|L|) = 1 - \sum_{l=1}^{|L|-1} G(l)$$

and the joint density of $(G(1), ..., G(|L|-1))$ w.r.t. the $(|L|-1)$-dimensional Lebesgue measure on the set $\{g_l > 0,\ l = 1, ..., |L|-1$ and $\sum_{i=1}^{|L|-1} g_i < 1\}$ is

$$p(g_1, ..., g_{|L|-1}) = \frac{\Gamma(\varphi(L))}{\prod_{l=1}^{|L|}\Gamma(\varphi(l))} \prod_{l=1}^{|L|-1} g_l^{\varphi(l)-1} \left(1 - \sum_{i=1}^{|L|-1} g_i\right)^{\varphi(|L|)-1},$$

where $\varphi(L) = \sum_{l=1}^{|L|}\varphi(l)$. In the sequel we will also allow $\varphi(l) = 0$ for some (but not all) index l, in which case we mean that $G(l) \equiv 0$. If there is only one index $m \in L$ such that $\varphi(m) > 0$, then we define $G(l) = \delta_{l,m}$, whatever the value $\varphi(m)$ is. It is well known that if $\{G(l), l \in L\}$ is $Di\,\{\varphi(l), l \in L\}$ then, for any $l \in L$, $G(l)$ is $Beta(\varphi(l), \varphi(L) - \varphi(l))$ (with the convention that a $Beta(c, 0)$ is identically 1 for any $c > 0$ and a $Beta(0, c)$ random variable is identically 0 for any $c > 0$), hence

$$E(G(l)) = \frac{\varphi(l)}{\varphi(L)} =: q(l), \quad l \in L. \tag{1}$$

Next consider a non null contingency table

$$\varphi_{12} = \{\varphi_{12}(i,j) \geq 0, i \in I, j \in J\}$$

and a random table

$$G_{12} \sim Di\left(\varphi_{12}\right) \tag{2}$$

over a finite product set $I \times J$. We use the standard notation for the marginal

$$\varphi_2\left(j\right) = \sum_{i \in I} \varphi_{12}(i,j)$$

and for the conditionals

$$\varphi_{1|2}(i|j) = \begin{cases} \frac{\varphi_{12}(i,j)}{\varphi_2(j)} & \text{if } \varphi_2(j) > 0, \\ 0 & \text{if } \varphi_2(j) = 0. \end{cases}$$

These notations are used also for random probability distributions over $I \times J$.

By using the representation of the Dirichlet law as a function of independent Gamma variables (see e.g. Kotz, Balakrishnan, and Johnson [6]), the following fundamental result is immediately obtained.

Proposition 1 *Let the table $G_{12} \sim Di(\varphi_{12})$. Then the marginal vector G_2 and the vectors of conditional distributions $\left\{G_{1|2}(\cdot|j),\ j \in I : \varphi_2(j) > 0\right\}$ are mutually independent, with*

$$G_2 \sim Di\left(\varphi_2\right), \tag{3}$$

$$G_{1|2}(\cdot|j) \sim Di\left(\varphi_{12}(\cdot,j)\right),\ \forall j \in J : \varphi_2(j) > 0. \tag{4}$$

If $\varphi_2(j) = 0$ for some $j \in J$, then $G_{12}(i,j)$ is identically zero for $i \in I$, in which case $G_{1|2}(i|j)$ can be defined to be an arbitrary probability on I.

Conversely, if G_{12} is a random table with the properties (3) and (4), such that $G_{12}(i,j) = 0$ whenever $\varphi_2(j) = 0$, then $G_{12} \sim Di(\varphi_{12})$.

Next consider another non null contingency table

$$\varphi_{23} = \{\varphi_{23}(j,k),\ j \in J,\ k \in K\} \tag{5}$$

where K is another finite set. We assume that φ_{12} and φ_{23} are ***consistent***, i.e.

$$\varphi_2(j) = \sum_{i \in I} \varphi_{12}(i,j) = \sum_{k \in K} \varphi_{23}(j,k), \qquad \text{for any } j \in J. \tag{6}$$

Then the Markov combination φ of φ_{12} and φ_{23} is defined as

$$\varphi(i,j,k) = \begin{cases} \frac{\varphi_{12}(i,j)\varphi_{23}(j,k)}{\varphi_2(j)} & \text{if } \varphi_2(j) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Notice φ is the only contingency table over $I \times J \times K$ with the properties:

a) the marginals of φ over $I \times J$ and $J \times K$ equal to φ_{12} and φ_{23}, respectively;

b) $\varphi(i,j,k)$ factorizes in a product of a function of (i,j) and one of (j,k).

Property **b)** implies that if φ is normalized it becomes a probability distribution over the product $I \times J \times K$ under which the projections on the first and the third factor are independent conditional to the projection on the second one.

Now we construct a random probability distribution on the set $I \times J \times K$ in the following way. Let G_{12} be distributed as in (2) and consider independent vectors

$$G_{3|2}(\cdot|j) \sim Di\left(\varphi_{23}(j,\cdot)\right), \quad j \in J : \varphi_2(j) > 0, \tag{7}$$

independent of G_{12}. Then define the three-dimensional random array

$$G = \left\{ G_{12}(i,j) G_{3|2}(k|j) \quad \text{for } i \in I, j \in J, k \in K \right\}. \tag{8}$$

Since $G_{12}(i,j) = G_2(j) G_{1|2}(i|j)$, recalling (3) and (4), we observe that
i) the laws of the marginals over $I \times J$ and $J \times K$ are respectively

$$G_{12} \sim Di\left(\varphi_{12}\right), \quad G_{23} = \left\{ G_2(j) G_{3|2}(k|j), j \in J, k \in K \right\} \sim Di\left(\varphi_{23}\right),$$

ii) the realizations $\{G(i,j,k), i \in I, j \in J, k \in K\}$ factorize in a product of a function of (i,j) and one of (j,k), hence they are Markovian probability distributions;
iii) the families of conditional distributions

$$\left\{ G_{1|2}(\cdot|j), j \in J \right\}, \qquad \left\{ G_{3|2}(\cdot|j), j \in J \right\}$$

and the marginal distribution G_2 are mutually independent.

Property **iii)** is called the strong hyper Markov property [3].

The law of G is clearly uniquely determined by the properties **i)**, **ii)** and **iii)**. As such it is called the hyper Markov combination of the laws $Di\,(\varphi_{12})$ and $Di\,(\varphi_{23})$. Following [3] we call this law a hyper-Dirichlet law and indicate it by $HDi(\varphi_{12}, \varphi_{23})$.

The following proposition is worth of notice.

Proposition 2 *The hyper Dirichlet law $HDi(\varphi_{12}, \varphi_{23})$ is a Dirichlet law if and only if, for any j such that $\varphi_2(j) > 0$, at least one of the following sets*

$$I_j = \{i : \varphi_{12}(i,j) > 0\}\,, \qquad K_j = \{k : \varphi_{23}(j,k) > 0\}\,,$$

is a singleton. In this case

$$HDi(\varphi_{12}, \varphi_{23}) = Di(\varphi).$$

Proof. If G is defined by (2), (7) and (8) then $G_2 \sim Di\,(\varphi_2)$. Hence by Proposition 1 it is clear that G is $Di(\widetilde{\varphi})$, for some contingency table $\widetilde{\varphi}$, if and only if $\widetilde{\varphi}_2(j) = \varphi_2(j)$ for any j, and each of the conditionals

$$\left\{G_{13|2}(i,k|j) = G_{1|2}(i|j) G_{3|2}(k|j),\, i \in I, k \in K\right\}, \quad j\,:\,\varphi_2(j) > 0, \qquad (9)$$

is $Di(\{\widetilde{\varphi}(i,j,k), i \in I, k \in K\})$. Then

$$\widetilde{\varphi}(I \times J \times K) = \widetilde{\varphi}_2(J) = \varphi_2(J) = \varphi(I \times J \times K).$$

By (1), for $G \sim Di(\widetilde{\varphi})$

$$E\,[G(i,j,k)] = \frac{\widetilde{\varphi}(i,j,k)}{\widetilde{\varphi}(I \times J \times K)},$$

and, for $G \sim HDi(\varphi_{12}, \varphi_{23})$ and $\varphi_2(j) > 0$

$$\begin{aligned} E\,[G(i,j,k)] &= E\left[G_{12}(i,j) G_{3|2}(k|j)\right] = E\,[G_{12}(i,j)]\, E\left[G_{3|2}(k|j)\right] \\ &= \frac{\varphi_{12}(i,j)}{\varphi_{12}(I \times J)} \frac{\varphi_{23}(j,k)}{\varphi_2(j)} = \frac{\varphi(i,j,k)}{\varphi(I \times J \times K)}, \end{aligned}$$

hence it is immediately seen that necessarily $\widetilde{\varphi} = \varphi$.

Therefore, whenever $\varphi_2(j) > 0$, (9) is $Di\,\{\varphi(i,j,k), i \in I, k \in K\}$, from which its law is concentrated on a $I_j \times K_j - 1$ manifold of $R^{I_j \times K_j}$ but not on any submanifold of lower dimension. On the other hand, by (4) and (7), (9) is concentrated on a $I_j + K_j - 2$ dimensional manifold. Thus it has to be either $I_j = 1$ or $K_j = 1$, or both. The sufficiency of this condition is obvious. ■

It is clear that Dirichlet laws can be used only as laws for random distributions on a finite sample space L. For a general measurable sample space $(L, \mathcal{L})$ we consider the space $\mathcal{P}(L)$ of all probability measures on $(L, \mathcal{L})$, endowed with the smallest σ-algebra which makes the functions $P \in \mathcal{P}(L) \to P(A)$ measurable, for all $A \in \mathcal{L}$. A random distribution Γ on $(L, \mathcal{L})$ is then a measurable mapping defined on some probability space with values in $\mathcal{P}(L)$. By Dynkin's lemma the law of Γ is uniquely determined by the joint law of $(\Gamma(A_1), ..., \Gamma(A_m))$ for all finite measurable partitions $\{A_h, h = 1, ..., m\}$ of L. Given any finite measure φ on $(L, \mathcal{L})$, we say that Γ is a Dirichlet process with parameter measure φ if this joint law is $Di\,\{\varphi(A_h), h = 1, ..., m\}$. The law of this process will be indicated by $DI(\varphi)$, in order to distinguish it from the case of a finite space. The existence of a random distribution with this property is ensured by the following representation, introduced in [11] and further analyzed in [10].

In order to introduce this representation, first define $\mathbf{T} : (0,1)^\infty \to (0,1)^\infty$ with

$$T_l(\boldsymbol{b}) = b_l \prod_{i=1}^{l-1}(1 - b_i), \qquad l = 1, 2, ...$$

with the usual convention that a product over an empty set is equal to 1. If $\mathbf{T}$ is applied to a sequence $\boldsymbol{\beta}$ with i.i.d. components $\beta_l \sim Beta(1, \varphi(L))$, $l = 1, 2, ...$, then with probability 1

$$\boldsymbol{\gamma} = \mathbf{T}(\boldsymbol{\beta}) \ \in \Delta_\infty = \left\{ \mathbf{p} \in \left(\mathbf{R}^+\right)^\infty : \sum_l p_l = 1 \right\},$$

thus it defines a random distribution on $\mathbb{N}^+$. Next let us define the probability distribution $Q = \frac{1}{\varphi(L)}\varphi$ on $(L, \mathcal{L})$ and take

$$\boldsymbol{\theta} = \{\theta_l, l = 1, 2, ...\},$$

with i.i.d. components $\theta_l \sim Q$. It can be shown that

$$\varkappa(\boldsymbol{\gamma}, \boldsymbol{\theta}) = \sum_l \gamma_l \delta_{\theta_l} \tag{10}$$

is a Dirichlet process with parameter measure φ, where $\varkappa$ is a measurable mapping of $\Delta_\infty \times L^\infty$ into $\mathcal{P}(L)$. We thus write $\varkappa(\boldsymbol{\gamma}, \boldsymbol{\theta}) \sim DI(\varphi)$.

Next we extend the notion of Markov combination to distributions over more general spaces. In the sequel we assume that $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are Borel spaces, i.e. they are measurably isomorphic to a Borel subset of a Polish space. This assumption ensures the existence of regular versions of conditional probabilities.

Next consider two consistent finite measures φ_{12} and φ_{23} on $\mathcal{X} \times \mathcal{Y}$ and $\mathcal{Y} \times \mathcal{Z}$, respectively, i.e. such that

$$\varphi_2(dy) = \varphi_{12}(\mathcal{X} \times dy) = \varphi_{23}(dy \times \mathcal{Z}).$$

Thus the marginal measure φ_2 is unambiguously defined. Our assumption implies that there exists two families of conditional measures $\{\varphi_{1|2}(\cdot|y), y \in \mathcal{Y}\}$ and $\{\varphi_{3|2}(\cdot|y), y \in \mathcal{Y}\}$ on $\mathcal{X}$ and $\mathcal{Z}$ respectively, such that

$$\varphi_{12}(dx, dy) = \varphi_2(dy)\varphi_{1|2}(dx|y), \quad \varphi_{23}(dx, dy) = \varphi_2(dy)\varphi_{3|2}(dz|y).$$

Now we can construct the Markov combination φ of φ_{12} and φ_{23} on $\mathcal{X} \times \mathcal{Y}$ $\times \mathcal{Z}$, namely

$$\varphi(dx, dy, dz) = \varphi_{1|2}(dx|y)\varphi_2(dy)\varphi_{3|2}(dz|y) \tag{11}$$

$$= \varphi_{12}(dx, dy)\varphi_{3|2}(dz|y) = \varphi_{1|2}(dx|y)\varphi_{23}(dy, dz). \tag{12}$$

From (12) we see that the marginals of φ on $\mathcal{X} \times \mathcal{Y}$ and $\mathcal{Y} \times \mathcal{Z}$ are φ_{12} and φ_{23}, respectively. Furthermore, for $\alpha = \varphi(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$, the normalization $Q = \frac{\varphi}{\alpha}$ is Markovian, that is a draw $\theta = (U, V, W)$ from Q has the property that U and W are conditionally independent given V. These properties characterize φ uniquely.

Next let us consider a pair of laws $\mathfrak{P}_{12}$ and $\mathfrak{P}_{23}$ on $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ and $\mathcal{P}(\mathcal{Y} \times \mathcal{Z})$, respectively, which are hyperconsistent, in the sense they induce the same law of the marginal on $\mathcal{Y}$. Then [3] there exists a unique law $\mathfrak{P}$ on $\mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$ such that, if the random distrubution P is drawn from $\mathfrak{P}$, and P_{12}, P_{23} and P_2 are the corresponding marginals on $\mathcal{X} \times \mathcal{Y}$, $\mathcal{Y} \times \mathcal{Z}$ and $\mathcal{Y}$

1. P is Markovian w.p. 1;
2. $P_{12} \sim \mathfrak{P}_{12}$ and $P_{23} \sim \mathfrak{P}_{23}$;
3. $P_{12} \perp P_{23}|P_2$.

Properties **i)**, **ii)** and **iii)** for $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ finite obviously imply 1, 2 and 3, respectively.

Property 3 is called the *weak hyper Markov* property and means specifically that the σ-algebra $\sigma\{P_{12}\}$ generated by $P(F \times \mathcal{Z})$, for all measurable subsets F of $\mathcal{X} \times \mathcal{Y}$ and the σ-algebra $\sigma\{P_{23}\}$ generated by $P(\mathcal{X} \times G)$, for all measurable subsets G of $\mathcal{Y} \times \mathcal{Z}$ are independent conditionally to the σ-algebra $\sigma\{P_2\}$ generated by $P(\mathcal{X} \times H \times \mathcal{Z})$, for all measurable subsets H

of $\mathcal{Y}$. The law $\mathfrak{P}$ is then called the *hyper Markov combination* of $\mathfrak{P}_{12}$ and $\mathfrak{P}_{23}$. Notice that property 3 replaces the stronger property **iii)**, given in the discrete case.

Proposition 3 *If φ_2 has no atoms, then the law $DI(\varphi)$ is the hyper Markov combination of $DI(\varphi_{12})$ and $DI(\varphi_{23})$.*

Proof. We use the Sethuraman representation for $DI(\varphi)$, i.e.

$$\Gamma = \varkappa(\boldsymbol{\gamma}, (\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W})) = \sum_l \gamma_l \delta_{(U_l, V_l, W_l)} \sim DI(\varphi), \tag{13}$$

with (U_l, V_l, W_l), $l = 1, 2, ...$ i.i.d. from Q, $\boldsymbol{\gamma} = \mathbf{T}(\boldsymbol{\beta})$, $\boldsymbol{\beta}$ with i.i.d. components $\beta_l \sim Beta(1, \alpha)$, $l = 1, 2, ...$

By computing the marginals

$$\Gamma_{12} = \varkappa(\boldsymbol{\gamma}, (\boldsymbol{U}, \boldsymbol{V})), \qquad \Gamma_{23} = \varkappa(\boldsymbol{\gamma}, (\boldsymbol{V}, \boldsymbol{W}))$$

we check immediately that $\Gamma_{12} \sim DI(\varphi_{12})$ and $\Gamma_{23} \sim DI(\varphi_{23})$. Since φ_2 has no atoms, the random sequence $\boldsymbol{V} = \{V_l,\ l = 1, 2, ...\}$ does not contain repetitions w.p. 1: this immediately implies that Γ is Markovian w.p. 1.

As for the hyper Markov property we observe that

$$\sigma\{\Gamma_2\} \subset \sigma\{\boldsymbol{\gamma}, \mathbf{V}\}, \ \sigma\{\Gamma_{12}\} \subset \sigma\{\boldsymbol{\gamma}, \mathbf{U}, \mathbf{V}\}, \ \sigma\{\Gamma_{23}\} \subset \sigma\{\boldsymbol{\gamma}, \mathbf{V}, \mathbf{W}\},$$

so that by the conditional independence of $\boldsymbol{U}$ and $\boldsymbol{W}$ given $\boldsymbol{\gamma}$ and $\boldsymbol{V}$,

$$\Gamma_{12} \perp \Gamma_{23} \mid \boldsymbol{\gamma}, \mathbf{V}.$$

Then, for any choice of measurable subsets $\{F_i, i = 1, ..., m\}$ of $\mathcal{X} \times \mathcal{Y}$ and $\{G_j, j = 1, ..., s\}$ of $\mathcal{Y} \times \mathcal{Z}$, for any pair h_{12} and h_{23} of bounded measurable functions defined on $[0, 1]^m$ and $[0, 1]^s$ respectively

$$E\left[h_{12}(\Gamma_{12}(F_1), ..., \Gamma_{12}(F_m))h_{23}(\Gamma_{23}(G_1), ..., \Gamma_{23}(G_s)) \mid \sigma\{\boldsymbol{\gamma}, \mathbf{V}\}\right]$$
$$= E\left[h_{12}(\Gamma_{12}(F_1), ..., \Gamma_{12}(F_m)) \mid \sigma\{\boldsymbol{\gamma}, \mathbf{V}\}\right] E\left[h_{23}(\Gamma_{23}(G_1), ..., \Gamma_{23}(G_s)) \mid \sigma\{\boldsymbol{\gamma}, \mathbf{V}\}\right].$$

Thus it remains to prove that

$$E\left[h_{12}(\Gamma_{12}(F_1), ..., \Gamma_{12}(F_m)) \mid \sigma\{\boldsymbol{\gamma}, \mathbf{V}\}\right] = E\left[h_{12}(\Gamma_{12}(F_1), ..., \Gamma_{12}(F_m)) \mid \sigma\{\Gamma_2\}\right], \tag{15}$$

$$E\left[h_{23}(\Gamma_{23}(G_1), ..., \Gamma_{23}(G_s)) \mid \sigma\{\boldsymbol{\gamma}, \mathbf{V}\}\right] = E\left[h_{23}(\Gamma_{23}(G_1), ..., \Gamma_{23}(G_s)) \mid \sigma\{\Gamma_2\}\right]. \tag{16}$$

If the function h_{12} is a polynomial, then (15) becomes

$$E\left[\prod_{l=1}^{m}\left(\sum_{i}\gamma_i\delta_{(U_i,V_i)}(F_l)\right)^{d_l}\middle|\sigma\left\{\boldsymbol{\gamma},\boldsymbol{V}\right\}\right]$$
$$=\sum_{i_1}\cdots\sum_{i_d}\prod_{k=1}^{d}\gamma_{i_k}\delta_{V_{i_k}}\left(F_{l_k,\mathcal{Y}}\right)E\left[\prod_{k=1}^{d}\delta_{U_{i_k}}\left(F_{l_k,\mathcal{X}}^{\left(V_{i_k}\right)}\right)\middle|\sigma\left\{\boldsymbol{\gamma},\boldsymbol{V}\right\}\right] \qquad (17)$$

where $d = d_1 + \cdots + d_m$,

$$\begin{cases} l_1,\ldots,l_{d_1} = 1 \\ l_{d_1+1},\ldots,l_{d_1+d_2} = 2 \\ \ldots \\ l_{d_1+\ldots+d_{m-1}+1},\ldots,l_d = m, \end{cases}$$

and

$$F_{\mathcal{Y}} = \left\{y : \exists\,(x,y)\in F\right\}, \quad F_{\mathcal{X}}^{(y)} = \left\{x : (x,y)\in F\right\},$$

so that

$$\delta_{x,y}(F) = \delta_y(F_{\mathcal{Y}})\delta_x(F_{\mathcal{X}}^{(y)}).$$

Then, taking into account that $(\mathbf{U},\mathbf{V})$ are independent of $\boldsymbol{\gamma}$ and that (U_i,V_i) are i.i.d., we can rewrite the conditional expectation appearing in (17) as

$$E\left[\delta_{U_{i_1}}\left(F_{l_1,\mathcal{X}}^{\left(V_{i_1}\right)}\right)\ldots\delta_{U_{i_d}}\left(F_{l_d,\mathcal{X}}^{\left(V_{i_d}\right)}\right)\middle|\sigma\left\{\boldsymbol{\gamma},\boldsymbol{V}\right\}\right]$$
$$=E\left[\delta_{U_{i_1}}\left(F_{l_1,\mathcal{X}}^{\left(V_{i_1}\right)}\right)\ldots\delta_{U_{i_d}}\left(F_{l_d,\mathcal{X}}^{\left(V_{i_d}\right)}\right)\middle|\sigma\left\{\boldsymbol{V}\right\}\right]$$
$$=\psi\left(V_{i_1},\cdots,V_{i_d}\right),$$

where ψ is the bounded measurable function

$$\psi\left(\boldsymbol{v}\right) := E\left[\delta_{U_1}\left(\bigcap_{i\in A_1(\boldsymbol{v})}F_{l_i,\mathcal{X}}^{\left(v_1'\right)}\right)\middle|V_1=v_1'\right]\ldots E\left[\delta_{U_1}\left(\bigcap_{i\in A_f(\boldsymbol{v})}F_{l_i,\mathcal{X}}^{\left(v_f'\right)}\right)\middle|V_1=v_f'\right],$$

with the definition

$$A_j(\boldsymbol{v}) = \{i : v_i = v_j'\}, \qquad j = 1,\cdots,f \qquad (18)$$

where $\{v_1',\cdots,v_f'\} = \{v_1,\cdots,v_d\}$ is the set whose elements are the distinct components of the vector $\boldsymbol{v} = (v_1,\cdots,v_d)$.

Summarizing (17) can be written as

$$\sum_{i_1}...\sum_{i_d}\gamma_{i_1}...\gamma_{i_d}\delta_{V_{i_1}}\left(F_{l_1,\mathcal{Y}}\right)...\delta_{V_{i_d}}\left(F_{l_d,\mathcal{Y}}\right)\psi\left(V_{i_1},\cdots,V_{i_d}\right)$$
$$=\int_{F_{l_1,\mathcal{Y}}}\cdots\int_{F_{l_d,\mathcal{Y}}}\psi\left(y_1,\cdots,y_d\right)\Gamma_2(dy_1)\cdots\Gamma_2(dy_d).$$

Arguing as in [4], it is readily proved that the above is a $\sigma\left\{\Gamma_2\right\}-$measurable function because the Borel σ-algebra of a Borel space is countably generated. This ends the proof of (15). Likewise we prove (16) when h_{23} is a polynomial. Finally observe that expectations of polynomials determine uniquely the joint conditional law of $\{\Gamma_{12}(F_i),\, i=1,...,m\}$ and $\{\Gamma_{23}(G_j),\, j=1,...,s\}$, hence the whole proof is finished. ■

Following Dawid and Lauritzen we say that a random distribution is a hyper Dirichlet process with parameter measures φ_{12} and φ_{23} if it has the same law as Γ defined in (13). We denote this law by $HDI(\varphi_{12},\varphi_{23})$ in order to distinguish it from the finite case, where we used the symbol HDi instead. The striking feature of the hyper Dirichlet process, when φ_2 is diffuse, is that it is a Dirichlet process, namely $HDI(\varphi_{12},\varphi_{23})=DI(\varphi)$, where φ is the Markov combination of φ_{12} and φ_{23}. On the contrary we saw that the hyper Dirichlet law is practically never a Dirichlet law. The aim of the next section is to prove that nonetheless, under suitable regularity conditions, we can exhibit a sequence of hyper Dirichlet laws converging in a natural sense to the law of a hyper Dirichlet process. In Section 4 we will turn our attention to more general hyper Dirichlet processes which are not Dirichlet process anymore.

3 Coupling the limiting process with the discrete approximations

In this section we take $\mathcal{X}=\mathbb{R}^{d_1}$, $\mathcal{Y}=\mathbb{R}^{d_2}$ and $\mathcal{Z}=\mathbb{R}^{d_3}$, and we assume that for each of these spaces we have a sequence of finite measurable partitions

$$\left\{B_i^{(n)},i\in I^{(n)}\right\},\ \left\{C_j^{(n)},j\in J^{(n)}\right\},\ \left\{D_k^{(n)},k\in K^{(n)}\right\},\quad n=1,2,...,\quad (19)$$

respectively. Correspondingly, we discretize the consistent parameter measures φ_{12} and φ_{23} along the corresponding sequences of product partitions in two-dimensional cells

$$\left\{B_i^{(n)}\times C_j^{(n)},i\in I^{(n)},j\in J^{(n)}\right\},\ \left\{C_j^{(n)}\times D_k^{(n)},j\in J^{(n)},k\in K^{(n)}\right\},\quad (20)$$

getting the tables

$$\varphi_{12}^{(n)} = \left\{\varphi_{12}(B_i^{(n)} \times C_j^{(n)}), i \in I^{(n)}, j \in J^{(n)}\right\}$$
$$\varphi_{23}^{(n)} = \left\{\varphi_{23}(C_j^{(n)} \times D_k^{(n)}), j \in J^{(n)}, k \in K^{(n)}\right\}. \tag{21}$$

For each integer n consider a random array

$$G^{(n)} \sim HDi(\varphi_{12}^{(n)}, \varphi_{23}^{(n)}) \tag{22}$$

of probabilities $G^{(n)}(i,j,k)$ randomly assigned to the three-dimensional cells $B_i^{(n)} \times C_j^{(n)} \times D_k^{(n)}$ for $i \in I^{(n)}, j \in J^{(n)}, k \in K^{(n)}$. It is then natural to compare its law with that of the corresponding array of probabilities

$$\Gamma^{(n)}(i,j,k) = \left\{\Gamma(B_i^{(n)} \times C_j^{(n)} \times D_k^{(n)}), i \in I^{(n)}, j \in J^{(n)}, k \in K^{(n)}\right\}. \tag{23}$$

assigned by the process $\Gamma \sim HDI(\varphi_{12}, \varphi_{23})$, which is a random distribution on $\mathbb{R}^d$, with $d = d_1 + d_2 + d_3$. This will be done by constructing both three-dimensional random arrays on the same probability space where the representation (13) of the hyper Dirichlet process Γ is defined, coupling them in a suitable way. First of all notice that whenever $\varphi_2^{(n)}(j) = 0$ it has to be $\Gamma^{(n)}(i,j,k) = G^{(n)}(i,j,k) = 0$ for any $i \in I^{(n)}$ and $k \in K^{(n)}$, so we can assume w.l.o.g. that $\varphi_2^{(n)}(j) > 0$ for all $j \in J^{(n)}$. Next observe that

$$\Gamma^{(n)}(i,j,k) = {\sum_l}^{*} \gamma_l, \quad i \in I^{(n)},\ j \in J^{(n)},\ k \in K^{(n)}, \tag{24}$$

where in ${\sum_l}^{*}$ the index l ranges in $\{l : (U_l, V_l, W_l) \in B_i^{(n)} \times C_j^{(n)} \times D_k^{(n)}\}$. Since

$$\Gamma_{12}^{(n)} \sim Di\left(\varphi_{12}^{(n)}\right)$$

we can define

$$G^{(n)}(i,j,k) = \Gamma_{12}^{(n)}(i,j) G_{3|2}^{(n)}(k|j), \quad i \in I^{(n)},\ j \in J^{(n)},\ k \in K^{(n)}, \tag{25}$$

provided the vector $\left\{G_{3|2}^{(n)}(k|j), k \in K^{(n)}\right\}$ is defined, independently for all $j \in J^{(n)}$ and independently of Γ_{12}, with the law $Di\left\{\varphi_{23}^{(n)}(j,k), k \in K^{(n)}\right\}$. This can be achieved by using mutually independent (w.r.t. $j \in J^{(n)}$) Dirichlet processes on $\mathcal{Z}$ with parameter measure $\varphi_{23}(C_j^{(n)} \times \cdot)$ defined through the Sethuraman representation

$$\sum_r H_{j,r}^{(n)} \delta_{W_{j,r}^{(n)}}, \tag{26}$$

where $\boldsymbol{H}_j^{(n)} = \left\{H_{j,r}^{(n)},\ r = 1, 2, ...\right\} = \boldsymbol{T}(\boldsymbol{\beta}_j^{(n)})$, with $\boldsymbol{\beta}_j^{(n)} = \left\{\beta_{j,r}^{(n)},\ r = 1, 2, ...\right\}$ i.i.d. $Beta(1, \varphi_2^{(n)}(j))$, and $\left\{W_{j,r}^{(n)},\ r = 1, 2, ...\right\}$ is a sequence of i.i.d. random variables, with common distribution $\varphi_{3|2}^{(n)}(\cdot|j)$ given by

$$\varphi_{3|2}^{(n)}(dz|j) = \frac{\varphi_{23}(C_j^{(n)} \times dz)}{\varphi_2(C_j^{(n)})}.$$

With these positions we define

$$G_{3|2}^{(n)}(k|j) = \sum_r H_{j,r}^{(n)} 1_{D_k}(W_{j,r}^{(n)}),\ k \in K^{(n)}, \tag{27}$$

All the random variables used to produce the vectors $\left\{G_{3|2}^{(n)}(k|j), k \in K^{(n)}\right\}$, for $j \in J^{(n)}$, are drawn independently of Γ, except for the mutually independent random variables $W_{j,1}^{(n)}$, $j \in J^{(n)}$. Each of these random variables is coupled with $W_{l(j)}$, where $l(j)$ is the first index l (existing w.p. 1) such that $V_l \in C_j$. We choose the maximal coupling (also known as γ-coupling, see e.g. Lindvall [8]), conditional to $V_{l(j)}$, so that

$$P(W_{j,1}^{(n)} \neq W_{l(j)}|V_{l(j)}) = \frac{1}{2}\left\|\varphi_{3|2}(\cdot|V_{l(j)}) - \varphi_{3|2}^{(n)}(\cdot|j)\right\|. \tag{28}$$

Having completed the construction, we are now ready to prove the following result.

Theorem 4 *Under the following assumptions:*

i) *the marginal measure φ_2 has no atoms;*

ii) *for any compact set $A \subset \mathcal{Y}$*

$$\lim_{n\to\infty} \max\left\{\left|C_j^{(n)}\right|, j \in J^{(n)} : C_j^{(n)} \cap A \neq \emptyset\right\} = 0,$$

where $|\cdot|$ is the diameter;

iii) *the function $y \in \mathcal{Y} \to \varphi_{3|2}(\cdot|y)$ has a version which is continuous in total variation;*

then

$$\lim_{n\to\infty} \sum_{i\in I^{(n)}} \sum_{j\in J^{(n)}} \sum_{k\in K^{(n)}} E|\Gamma^{(n)}(i,j,k) - G^{(n)}(i,j,k)| = 0.$$

Proof. Along the whole proof we will drop the superscript n, except when we exhibit the final estimates. By standard manipulations we get

$$
\begin{aligned}
|\Gamma(i,j,k) - G(i,j,k)| &= \left|\Gamma(i,j,k) - \Gamma_{12}(i,j) G_{3|2}(k|j)\right| \qquad (29)\\
&\le \gamma_{l(j)} 1_{B_i}(U_{l(j)}) \left|1_{D_k}(W_{l(j)}) - 1_{D_k}(W_{j,1})\right| \\
&+ \gamma_{l(j)} 1_{B_i}(U_{l(j)}) \sum_{r\ge 2} H_{j,r} \left|1_{D_k}(W_{j,1}) - 1_{D_k}(W_{j,r})\right| \\
&+ \sum_{l>l(j)} \gamma_l 1_{B_i}(U_l) 1_{C_j}(V_l) \left|1_{D_k}(W_l) - \sum_{r\ge 1} H_{j,r} 1_{D_k}(W_{j,r})\right|,
\end{aligned}
$$

from which, by summing over i, j, k, recalling that $\sum_{r\ge 2} H_{j,r} = 1 - H_{j,1}$ and that for any pair of probability distributions μ and ν over K

$$
\sum_k |\mu(k) - \nu(k)| = \|\mu - \nu\| \le 2,
$$

we get

$$
\begin{aligned}
\sum_{i,j,k} |\Gamma(i,j,k) - G(i,j,k)| &\le \sum_j \gamma_{l(j)} \sum_k \left|1_{D_k}(W_{l(j)}) - 1_{D_k}(W_{j,1})\right| \\
&+ 2\sum_j \gamma_{l(j)}(1 - H_{j,1}) + 2\sum_j \sum_{l>l(j)} \gamma_l 1_{C_j}(V_l). \qquad (30)
\end{aligned}
$$

Let us proceed to bound the first term at the r.h.s. of (30). Since

$$
\sum_k \left|1_{D_k}(W_{l(j)}) - 1_{D_k}(W_{j,1})\right| \le 2 \cdot 1_{\{W_{l(j)} \ne W_{j,1}\}},
$$

for any compact set $A \subset \mathcal{Y}$ we have

$$
\sum_j \gamma_{l(j)} \sum_k \left|1_{D_k}(W_{l(j)}) - 1_{D_k}(W_{j,1})\right| \le 2 \sum_{j: C_j \cap A \ne \emptyset} \gamma_{l(j)} 1_{\{W_{l(j)} \ne W_{j,1}\}} + 2\Gamma_2(A^c),
$$

and by taking the expected value and using (28), by the independence of $\boldsymbol{\gamma}$, $l(j)$ and $V_{l(j)}$ we have

$$
\begin{aligned}
&E \sum_j \gamma_{l(j)} \sum_k \left|1_{D_k}(W_{l(j)}) - 1_{D_k}(W_{j,1})\right| \\
&\le \sum_{j: C_j \cap A \ne \emptyset} E\left[\gamma_{l(j)}\right] E\left[\|\varphi^{(n)}_{3|2}(\cdot|j) - \varphi_{3|2}(\cdot|V_{l(j)})\|\right] + 2E(\Gamma_2(A^c)) \\
&\le \sup_{j: C_j \cap A \ne \emptyset} \sup_{v \in C_j} \left\|\varphi^{(n)}_{3|2}(\cdot|j) - \varphi_{3|2}(\cdot|v)\right\| + \frac{2}{\alpha}\varphi_2(A^c). \qquad (31)
\end{aligned}
$$

Now

$$\sup_{v\in C_j}\left\|\varphi^{(n)}_{3|2}(\cdot|j)-\varphi_{3|2}(\cdot|v)\right\| = \sup_{v\in C_j}\left\|\frac{\int\limits_{C_j}\varphi_{3|2}(\cdot|y)\varphi_2(dy)}{\varphi_2\left(C_j\right)}-\varphi_{3|2}(\cdot|v)\right\|$$

$$= 2\sup_{v\in C_j}\sup_{D}\frac{\left|\int\limits_{C_j}\left(\varphi_{3|2}(D|y)-\varphi_{3|2}(D|v)\right)\varphi_2(dy)\right|}{\varphi_2\left(C_j\right)}$$

$$\leq \sup_{v\in C_j}\frac{\int\limits_{C_j}\left\|\varphi_{3|2}(\cdot|y)-\varphi_{3|2}(\cdot|v)\right\|\varphi_2(dy)}{\varphi_2\left(C_j\right)}$$

$$\leq \sup_{v_1,v_2\in C_j}\left\|\varphi_{3|2}(\cdot|v_1)-\varphi_{3|2}(\cdot|v_2)\right\|.$$

Finally select the compact set A in such a way that $\varphi_2(A^c)\leq\frac{\alpha\varepsilon}{4}$ and define the larger compact set

$$A^1\equiv\{t: dist(t,A)\leq 1\}.$$

From the assumption iii), on the compact set A^1 the function $y\to\varphi_{3|2}(\cdot|y)$ is uniformly continuous in total variation, so let $\delta(\varepsilon)<1$ such that

$$\sup_{\substack{v_1,v_2\in A^1\\|v_1-v_2|<\delta(\varepsilon)}}\left\|\varphi_{3|2}(\cdot|v_1)-\varphi_{3|2}(\cdot|v_2)\right\|<\frac{\varepsilon}{2}.$$

From the assumption ii), for n large enough and any j such that $C^{(n)}_j\cap A\neq\emptyset$, it holds $\left|C^{(n)}_j\right|<\delta(\varepsilon)$ so that $C^{(n)}_j\subset A^1$, and consequently the r.h.s. of (31) can be bounded by ε.

For the remaining terms we need first to observe that

$$\lambda_n:=\max_{j\in J^{(n)}}\varphi_2(C^{(n)}_j)$$

tends to zero as $n\to\infty$. If this is not true, by tightness of φ_2 then we would have a sequence $\left\{C^{(n)}_{j_n}\right\}$ of Borel sets of diameter shrinking to zero such that $\varphi_2(C^{(n)}_{j_n})>\varepsilon>0$ all contained in a given compact set. But then this sequence would accumulate at some point, thus denying the absence of atoms.

The expected value of the second term at the r.h.s. of (30) can now be bounded by

$$2\sum_j E\left[\gamma_{l(j)}\right](1-E\left[H_{j,1}\right]) = 2\sum_j E\left[\gamma_{l(j)}\right]\frac{\varphi_2(C_j)}{1+\varphi_2(C_j)}$$
$$\leq 2\max_j \frac{\varphi_2(C_j^{(n)})}{1+\varphi_2(C_j^{(n)})} \leq 2\lambda_n \ ,$$

which tends to zero as $n\to\infty$.

Observe that $l(j)$ has a geometric distribution with probability of success $p_j = \frac{\varphi_2(C_j)}{\alpha}$, whereas

$$E(\gamma_{l(j)}|l(j)=h) = E(\gamma_h) = \frac{1}{1+\varphi_2(C_j)}\left(\frac{\varphi_2(C_j)}{1+\varphi_2(C_j)}\right)^{h-1}, h=1,2,...$$

so that, by a trivial computation

$$E(\gamma_{l(j)}) = \frac{\varphi_2(C_j)}{\alpha+\varphi_2(C_j)^2} \geq \frac{\varphi_2(C_j)}{\alpha\left(1+\varphi_2(C_j)\right)}.$$

Finally observe that the expected value of the last sum at the r.h.s. of (30) is

$$2E\left[\sum_j\sum_{l>l(j)}\gamma_l 1_{C_j}(V_l)\right] = 2E(1-\sum_j\gamma_{l(j)})$$
$$\leq \frac{2}{\alpha}\sum_j\left(\varphi_2(C_j)-\frac{\varphi_2(C_j)}{1+\varphi_2(C_j)}\right) = \frac{2}{\alpha}\sum_j\frac{\varphi_2(C_j)^2}{1+\varphi_2(C_j)}$$
$$\leq \frac{2}{\alpha}\max_j\varphi_2(C_j^{(n)})\sum_j\frac{\varphi_2(C_j^{(n)})}{1+\varphi_2(C_j^{(n)})} \leq 2\lambda_n,$$

which again tends to zero as $n\to\infty$. ■

The reader will expect that it is possible to formulate the result of the previous theorem as the convergence of $G^{(n)}$ to Γ in a suitable sense. In order to use weak convergence we have to extend $G^{(n)}$ to be a random distribution on the whole space $\mathbb{R}^d$. However in order to obtain the next result such an extension is essentially arbitrary.

Theorem 5 *Under the assumption of the previous theorem, let $G^{(n)}$ be any random probability distribution on $\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}=\mathbb{R}^d$, with the property that*

$$\left\{G^{(n)}(B_i^{(n)}\times C_j^{(n)}\times D_k^{(n)}), i\in I^{(n)}, j\in J^{(n)}, k\in K^{(n)}\right\}$$

has the same law as

$$\left\{G^{(n)}(i,j,k), i\in I^{(n)}, j\in J^{(n)}, k\in K^{(n)}\right\}.$$

Furthermore suppose that for all compact sets $E\subset\mathcal{X}$ and $F\subset\mathcal{Z}$

$$\lim_{n\to\infty}\ \max\left\{\left|B_i^{(n)}\right|, i\in I^{(n)}: B_i^{(n)}\cap E\neq\emptyset\right\}=0 \tag{32}$$

$$\lim_{n\to\infty}\ \max\left\{\left|D_k^{(n)}\right|, k\in K^{(n)}: D_k^{(n)}\cap F\neq\emptyset\right\}=0. \tag{33}$$

Then $\left\{G^{(n)}\right\}$ converges weakly to Γ, as random variables with values in the space of probability distributions on $\mathbb{R}^d$, endowed with the topology of weak convergence.

Proof. In order to prove that $\left\{G^{(n)}\right\}$ converges weakly to Γ, it suffices to show that

$$\lim_{n\to\infty}E\left[\Phi\left(G^{(n)}\right)\right]=E\left[\Phi\left(\Gamma\right)\right],$$

for all functions Φ which are bounded and Lipschitz (BL) w.r.t. a metric *dist* that induces the topology of weak convergence on $\mathcal{P}(\mathbf{R}^d)$, i.e. with the property that for all probability distributions μ and ν

$$\left|\Phi\left(\mu\right)-\Phi\left(\nu\right)\right|\leq L_\Phi dist\left(\mu,\nu\right),$$

with $L_\Phi<\infty$ (see [1]).

In particular (see [9]) we can choose $dist=dist_{BL}$, i.e.

$$dist_{BL}\left(\mu,\nu\right)=\sup_{\|f\|_\infty\leq 1, L(f)\leq 1}\left|\mu(f)-\nu(f)\right|,$$

where the supremum is taken over all bounded Lipschitz functions on $\mathbf{R}^d$ such that both the L^∞ norm $\|f\|_\infty$ and the Lipschitz constant $L(f)$ are bounded by 1.

By assumption

$$G^{(n)}(dx,dy,dz)=\sum_{i,j,k}G^{(n)}(i,j,k)\lambda_{ijk}^{(n)}(dx,dy,dz),$$

where $\lambda_{ijk}^{(n)}$ is a probability measure supported by $B_i^{(n)} \times C_j^{(n)} \times D_k^{(n)}$, for $i \in I^{(n)}, j \in J^{(n)}, k \in K^{(n)}$. Likewise we define the intermediate random probability distribution on $\mathbf{R}^d$, namely

$$\Gamma^{(n)}(dx, dy, dz) = \sum_{i,j,k} \Gamma^{(n)}(i, j, k) \lambda_{ijk}^{(n)}(dx, dy, dz),$$

where we recall that

$$\Gamma^{(n)}(i, j, k) = \Gamma(B_i^{(n)} \times C_j^{(n)} \times D_k^{(n)}). \tag{34}$$

Obviously

$$\begin{aligned} &\left| E\left[\Phi\left(G^{(n)}\right)\right] - E\left[\Phi\left(\Gamma\right)\right] \right| \\ &\leq \left| E\left[\Phi\left(G^{(n)}\right) - \Phi\left(\Gamma^{(n)}\right)\right] \right| + \left| E\left[\Phi\left(\Gamma^{(n)}\right) - \Phi\left(\Gamma\right)\right] \right|. \end{aligned}$$

If Φ is BL, then

$$\begin{aligned} \left| E\left[\Phi\left(G^{(n)}\right)\right] - E\left[\Phi\left(\Gamma^{(n)}\right)\right] \right| &\leq E\left[L_\Phi dist_{BL}\left(G^{(n)}, \Gamma^{(n)}\right)\right] \\ &\leq \frac{1}{2} L_\Phi E\left[\left\| G^{(n)} - \Gamma^{(n)} \right\|\right]. \end{aligned}$$

By Theorem 4 the variation distance between $G^{(n)}$ and $\Gamma^{(n)}$ tends to zero, hence it follows

$$\lim_{n\to\infty} \left| E\left[\Phi\left(G^{(n)}\right)\right] - E\left[\Phi\left(\Gamma^{(n)}\right)\right] \right| = 0.$$

Furthermore

$$\begin{aligned} \left| E\left[\Phi\left(\Gamma^{(n)}\right)\right] - E\left[\Phi\left(\Gamma\right)\right] \right| &\leq L_\Phi E\left[dist_{BL}\left(\Gamma^{(n)}, \Gamma\right)\right] \\ &= L_\Phi E\left[\sup_{\|f\|_\infty \leq 1, L(f) \leq 1} \left| \Gamma^{(n)}(f) - \Gamma(f) \right| \right]. \end{aligned}$$

Since the quantity inside the expectation is bounded by 2, by the bounded convergence theorem, the statement will be proved once we show that for any probability measure Γ on $\mathbb{R}^d$

$$\lim_{n\to+\infty} \sup_{\|f\|_\infty \leq 1, L(f) \leq 1} \left| \Gamma^{(n)}(f) - \Gamma(f) \right| = 0.$$

For this observe that

$$\begin{aligned} &\left| \Gamma^{(n)}(f) - \Gamma(f) \right| \\ &= \left| \sum_{i,j,k} \Gamma^{(n)}(i, j, k) \int_{B_i^{(n)}} \int_{C_j^{(n)}} \int_{D_k^{(n)}} f(x, y, z) \left(\lambda^{(n)}(dx, dy, dz) - \frac{\Gamma(dx, dy, dz)}{\Gamma^{(n)}(i, j, k)} \right) \right|. \end{aligned}$$

Let $\epsilon > 0$ and consider compact sets $A_h \subset \mathbf{R}^{d_h}$ for $h = 1, 2, 3$ such that $\Gamma(A^c) < \epsilon$, where $A = A_1 \times A_2 \times A_3$. Then, by splitting the sum as the sum $\sum_{i,j,k}^{*}$ over those indices (i, j, k) such that $B_i^{(n)} \cap A_1 \neq \emptyset$, $C_j^{(n)} \cap A_2 \neq \emptyset$ and $D_k^{(n)} \cap A_3 \neq \emptyset$, and the sum over the remaining indices, the above display is bounded from above by

$$\sum_{i,j,k}^{*} \Gamma^{(n)}(i,j,k) \left| \int\limits_{B_i^{(n)}} \int\limits_{C_j^{(n)}} \int\limits_{D_k^{(n)}} f(x,y,z) \left(\lambda^{(n)}(dx,dy,dz) - \frac{\Gamma(dx,dy,dz)}{\Gamma^{(n)}(i,j,k)} \right) \right| + 2 \|f\|_\infty \epsilon$$

$$\leq \sum_{i,j,k}^{*} \Gamma^{(n)}(i,j,k) \left(\int\limits_{B_i^{(n)}} \int\limits_{C_j^{(n)}} \int\limits_{D_k^{(n)}} \left| f(x,y,z) - f\left(\overline{x}_i, \overline{y}_j, \overline{z}_k\right) \right| \lambda^{(n)}(dx,dy,dz) \right.$$
$$\left. + \int\limits_{B_i^{(n)}} \int\limits_{C_j^{(n)}} \int\limits_{D_k^{(n)}} \left| f(x,y,z) - f\left(\overline{x}_i, \overline{y}_j, \overline{z}_k\right) \right| \frac{\Gamma(dx,dy,dz)}{\Gamma^{(n)}(i,j,k)} \right) + 2\epsilon, \tag{35}$$

for any choice of $\left(\overline{x}_i, \overline{y}_j, \overline{z}_k\right) \in B_i^{(n)} \times C_j^{(n)} \times D_k^{(n)}$.

Observe that from the assumption ii) of the previous Theorem 4, (32) and (33), for n large enough,

$$\left| B_i^{(n)} \times C_j^{(n)} \times D_k^{(n)} \right| < \epsilon,$$

for all the indices (i, j, k) considered in the above sum. Hence for any $(x, y, z) \in B_i^{(n)} \times C_j^{(n)} \times D_k^{(n)}$ as above

$$\left| f(x,y,z) - f\left(\overline{x}_i, \overline{y}_j, \overline{z}_k\right) \right| \leq L(f) dist\left((x,y,z), (\overline{x}_i, \overline{y}_j, \overline{z}_k)\right) \leq \epsilon,$$

which implies that (35) does not exceed 4ϵ. Since ϵ is arbitrary, the proof is finished. ■

Remark 6 *We remark that assumption iii) of Theorem 4 holds true when φ_{23} has a jointly continuous density $g_{23}(y,z)$ w.r.t. to a product measure $\mu_2 \times \mu_3$, with the property that for any $y \in \mathbb{R}^{d_2}$ there exists a μ_3-integrable function $h_y(z)$, such that for any y' in a neighborhood of y*

$$g_{23}(y', z) \leq h_y(z) \quad \textit{for all } z \in \mathbb{R}^{d_3}.$$

In fact this implies that as y_n tends to y,

$$\lim_{n\to\infty}\int |g_{23}(y_n,z)-g_{23}(y,z)|\mu_3(dz)=0,$$

the density $g_2(y)=\int g_{23}(y,z)\mu_3(dz)$ of φ_2 w.r.t. μ_2 is continuous in y, and by Bayes formula

$$\|\varphi_{3|2}(\cdot|y_n)-\varphi_{3|2}(\cdot|y)\|\le\frac{2}{g_2(y)}\int |g_{23}(y_n,z)-g_{23}(y,z)|\mu_3(dz),$$

for y in the open set of full φ_2-measure where $g_2(y)>0$.

In the following sections we will prove results analogous to Theorem 4 and Theorem 5 for more general hyper Dirichlet processes.

4 More general hyper Dirichlet processes

In this section we construct a general hyper Dirichlet process by removing the assumption that the common marginal of its two parameter measures is diffuse. We start with the following general result about Dirichlet processes.

Proposition 7 *Let us consider a sequence of finite measures $\{\rho^{(l)},\ l=0,1,2,..\}$ on the measurable space $(E,\mathcal{E})$, and suppose that $\rho(E)=\sum_l\rho^{(l)}(E)<+\infty$, in which case ρ is a finite measure as well. Let $\mathbf{R}=\{R_l,\ l=0,1,2,...\}$ be a sequence of independent random variables, with*

$$R_l\sim Be\big(\rho^{(l)}(E),\sum_{i=l+1}^{\infty}\rho^{(i)}(E)\big),\qquad l=0,1,2,... \tag{36}$$

and define the random sequence $\mathbf{S}=\mathbf{T}(\mathbf{R})$, i.e.

$$S_l=R_l\prod_{i=0}^{l-1}(1-R_i),\quad l=0,1,2,...$$

Then

i) $\mathbf{S}$ is a Dirichlet process on $\mathbb{N}$, with parameter measure $\sum_l\rho^{(l)}(E)\delta_l$.

Moreover let $\{\Gamma_l\sim DI\left(\rho^{(l)}\right),\ l=0,1,2,...\}$ be a sequence of independent Dirichlet processes on E, all independent of $\mathbf{S}$. Then

ii) the random probability distribution on $(E,\mathcal{E})$

$$\sum_{l=0}^{\infty}S_l\Gamma_l \tag{37}$$

is a Dirichlet process on E with parameter measure ρ.

Proof. A Dirichlet process whose parameter measure is a single mass on some (fixed) point x is with probability 1 a Dirac probability distribution on x. As a consequence statement *i)* is included in statement *ii)*, when $\rho^{(l)} = \rho^{(l)}(E)\delta_l$, and therefore we prove only the latter.

Next observe that the residual masses $1-\sum_{l=0}^{m} S_l$ of the sequence $\mathbf{S}$ decrease to zero a.s. as $m \to \infty$, since

$$E\left[1-\sum_{l=0}^{m} S_l\right] = \prod_{l=0}^{m} E(1-R_l) = \prod_{l=0}^{m} \frac{\sum_{j=l+1}^{\infty} \rho^{(j)}(E)}{\sum_{j=l}^{\infty} \rho^{(j)}(E)} = \frac{\sum_{j=m+1}^{\infty} \rho^{(j)}(E)}{\rho(E)} \to 0$$

as $m \to \infty$. Thus (37) is well defined as a random variable with values in the set of probability measures over $(E, \mathcal{E})$.

Using the converse part of Proposition 1, and induction on m, it is not difficult check that

$$(S_0, ..., S_m, 1-\sum_{l=0}^{m} S_l) \sim Di\left(\rho^{(0)}(E), ..., \rho^{(m)}(E), \sum_{l=m+1}^{\infty} \rho^{(l)}(E)\right), \qquad (38)$$

for any integer m.

Let x_0 be any fixed element of E and define the random probability distribution

$$\left\{\sum_{l=0}^{m} S_l\Gamma_l + \left(1-\sum_{l=0}^{m} S_l\right)\delta_{x_0},\ h \in H\right\} \qquad (39)$$

Next consider any finite measurable partition $\{A_h, h \in H\}$ of E, and suppose for the moment that the joint law of

$$\left\{\sum_{l=0}^{m} S_l\Gamma_l(A_h) + \left(1-\sum_{l=0}^{m} S_l\right)\delta_{x_0}(A_h), h \in H\right\} \qquad (40)$$

is $Di\left(\sum_{l=0}^{m} \rho^{(l)}(A_h) + \left(\sum_{l=m+1}^{\infty} \rho^{(l)}(E)\right)\delta_{x_0}(A_h),\ h \in H\right)$.

If we let $m \to \infty$ the random vector (40) converges a.s. to $\left\{\sum_{l=0}^{\infty} S_l\Gamma_l(A_h),\right.$ $\left. h \in H\right\}$, hence it converges in law as well. By the continuity of the Dirichlet laws w.r.t. their parameter vectors, the limit law is $Di\left\{\sum_{l=0}^{\infty} \rho^{(l)}(A_h)\right\}$. This ends the proof of statement *ii)*.

In order to see that (40) has the correct law, define the following two-dimensional table

$$D_{12}(h,l) = S_l \Gamma_l(A_h), \qquad\qquad h \in H, \quad l = 0, ..., m,$$
$$D_{12}(h, m+1) = \left(1 - \sum_{l=0}^{m} S_l\right) \delta_{x_0}(A_h), \qquad h \in H,$$

and observe that the marginal $D_2(\cdot)$ is (38) and the conditionals $D_{1|2}(\cdot|l)$ are $\{\Gamma_l(A_h), h \in H\}$ for $l = 0, ..., m$, and $\{\delta_{x_0}(A_h), h \in H\}$ for $l = m+1$. Therefore by Proposition 1

$$D_{12} \sim Di(d_{12}(h,l),\ h \in H,\ l = 0, 1, ..., m+1),$$

where

$$d_{12}(h,l) = \rho^{(l)}(A_h), \qquad\qquad h \in H, \quad l = 0, ..., m,$$
$$d_{12}(h, m+1) = \left(\sum_{l=m+1}^{\infty} \rho^{(l)}(E)\right) \delta_{x_0}(A_h), \qquad h \in H.$$

Finally observe that the marginal D_1 coincides with (40) and apply again Proposition 1 to conclude.

■

The consequence of the above proposition is the following convenient representation of a general two-dimensional Dirichlet process.

Corollary 8 *Let ρ be a finite measure on the product space $\mathcal{X} \times \mathcal{Y}$, and let $\{y_l, l = 1, 2, ...\}$ be the set of the atoms of $\rho_2(dy)$. Define the measure ψ as*

$$\psi(dx, dy) := \sum_l \delta_{\{y_l\}}(dy) \rho(dx \times \{y_l\})$$

and the measure φ in such a way that $\rho = \varphi + \psi$.

Then the Dirichlet process with parameter measure ρ can be represented as

$$S_0 \Gamma(dx, dy) + \sum_l S_l \Xi_l(dx) \delta_{y_l}(dy),$$

where the sequence $\mathbf{S} = \mathbf{T}(\mathbf{R})$, $\mathbf{R}$ being a sequence of independent random variables with

$$R_0 \sim Be(\varphi_2(\mathcal{Y}), \psi_2(\mathcal{Y})), R_l \sim Be(\psi_2(\{y_l\}), \sum_{i=l+1}^{\infty} \psi_2(\{y_i\})), l = 1, 2, ..., \quad (41)$$

Γ is a Dirichlet process on $\mathcal{X} \times \mathcal{Y}$ with parameter measure φ, and Ξ_l is a Dirichlet process on $\mathcal{X}$ with parameter measure $\psi^{(l)}(dx) := \psi(dx \times \{y_l\})$, for $l = 1, 2....$, all mutually independent.

Proof. It suffices to apply the representation of Proposition 7 to the Dirichlet process on $E = \mathcal{X} \times \mathcal{Y}$ with parameter measure $\rho = \sum_{l=0}^{\infty} \rho^{(l)}$, where

$$\rho^{(0)}(dx, dy) = \varphi(dx, dy), \quad \rho^{(l)}(dx, dy) = \psi^{(l)}(dx)\delta_{y_l}(dy), \qquad l = 1, 2, ...$$

with $\Gamma_0 = \Gamma$ and $\Gamma_l \sim DI\left(\psi^{(l)}(dx)\delta_{y_l}(dy)\right)$, which can be represented as $\Xi_l(dx)\delta_{y_l}(dy)$, for $l = 1, 2, ...$ ■

Notice that $\rho = \varphi + \psi$ is the unique decomposition where the marginal φ_2 is diffuse and the marginal ψ_2 is purely atomic. Therefore, given two consistent parameter measures ρ_{12} and ρ_{23} on $\mathcal{X} \times \mathcal{Y}$ and $\mathcal{Y} \times \mathcal{Z}$ respectively, they can be uniquely decomposed as

$$\rho_{12} = \varphi_{12} + \psi_{12}, \qquad \rho_{23} = \varphi_{23} + \psi_{23},$$

where φ_{12} and φ_{23} are consistent, with diffuse marginal φ_2, whereas ψ_{12} and ψ_{23} are consistent, with purely atomic marginal ψ_2.

Assume from now on that $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are Borel spaces. Then the previous corollary allows to construct a general hyper Dirichlet process, i.e. a process whose law is the hyper Markov combination of two hyperconsistent laws $DI(\rho_{12})$ and $DI(\rho_{23})$ on $\mathcal{X} \times \mathcal{Y}$ and $\mathcal{Y} \times \mathcal{Z}$, respectively. The measures φ_2 and ψ_2 being singular, the Markov combination ρ of ρ_{12} and ρ_{23}, as defined in (11), is the sum of the Markov combination φ of φ_{12} and of φ_{23} and the Markov combination ψ of ψ_{12} and ψ_{23}. Let us define

$$\rho_2^{(0)}(dy) = \varphi_2(dy), \ \rho_2^{(l)}(dy) = \psi_2(\{y_l\})\delta_{y_l}(dy), \quad l = 1, 2, ...$$

so that $\rho_2 = \varphi_2 + \psi_2 = \sum_{l=0}^{\infty} \rho_2^{(l)}$. Now let $\mathbf{R}$ be a sequence of independent random variables distributed as in (41), and let $\mathbf{S} = \mathbf{T}(\mathbf{R})$, so that

$$\mathbf{S} \sim DI(\rho_2^{(l)}(\mathcal{Y}), \ l = 0, 1, 2, ...) \tag{42}$$

Moreover let $\Gamma \sim HDI(\varphi_{12}, \varphi_{23})$, and Ξ_l^x and Ξ_l^z be Dirichlet processes on $\mathcal{X}$ and $\mathcal{Z}$ respectively, with parameter measures $\psi_{12}^{(l)}(dx) = \psi_{12}(dx \times \{y_l\})$ and $\psi_{23}^{(l)}(dz) = \psi_{23}(\{y_l\} \times dz)$, all mutually independent. Notice that hyperconsistency implies that

$$\psi_{12}^{(l)}(\mathcal{X}) = \psi_{23}^{(l)}(\mathcal{Z}) = \psi_2(\{y_l\}).$$

Then define

$$\Sigma(dx,dy,dz) = S_0\Gamma(dx,dy,dz) + \sum_l S_l \Xi_l^x(dx)\Xi_l^z(dz)\delta_{y_l}(dy). \tag{43}$$

The following generalization of Theorem 2 holds true.

Proposition 9 *The law of Σ defined in (43) is the hyper Markov combination of $DI(\rho_{12})$ and $DI(\rho_{23})$.*

Proof. By using the previous Corollary the laws of the marginals Σ_{12} and Σ_{23} are immediately identified to be $DI(\rho_{12})$ and $DI(\rho_{23})$. Next, independently of $\boldsymbol{S}$, use independent Sethuraman representations (13)

$$\begin{aligned} \Gamma &= \varkappa\big(\boldsymbol{\gamma}, (\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W})\big) \\ \Xi_l^x &= \varkappa(\boldsymbol{\gamma}_l^x, \boldsymbol{U}_l^x), \qquad \Xi_l^z = \varkappa(\boldsymbol{\gamma}_l^z, \boldsymbol{W}_l^z), \quad l = 1,2,\dots, \end{aligned}$$

where $\boldsymbol{\gamma}$, $\boldsymbol{U}$, $\boldsymbol{V}$, $\boldsymbol{W}$ are as in Proposition 3, $\boldsymbol{\gamma}_l^x = (\gamma_{l,m}^x,\ m = 1,2,...)$ and $\boldsymbol{\gamma}_l^z = (\gamma_{l,m}^z,\ m = 1,2,...)$ are independent sequences obtained as the image under $\mathbf{T}$ of i.i.d. sequences

$$\begin{aligned} \boldsymbol{\beta}_l^x &= (\beta_{l,m}^x \sim Beta(1, \psi_2(\{y_l\})),\ m = 1,2,...), \\ \boldsymbol{\beta}_l^z &= (\beta_{l,m}^z \sim Beta(1, \psi_2(\{y_l\})),\ m = 1,2,...), \end{aligned}$$

respectively, and

$$\begin{aligned} \boldsymbol{U}_l^x &= (U_{l,m}^x \sim \psi_{12}(\cdot \times \{y_l\})/\psi_2(\{y_l\}),\ m = 1,2,...) \\ \boldsymbol{W}_l^z &= (U_{l,m}^z \sim \psi_{23}(\{y_l\} \times \cdot)/\psi_2(\{y_l\}),\ m = 1,2,...) \end{aligned}$$

are sequences of i.i.d. random variables.

With probability 1 the sequence $\boldsymbol{V}$ contains neither repetitions nor any of the y_l, for $l = 1,2,\dots$. This implies that the realizations of Σ are Markovian with probability 1.

As for the hyper Markov property observe that

$$\begin{aligned} &\sigma\{\Sigma_2\} \subset \sigma\{\boldsymbol{S}, \boldsymbol{\gamma}, \mathbf{V}\}, \\ &\sigma\{\Sigma_{12}\} \subset \sigma\{\boldsymbol{S}, \boldsymbol{\gamma}, \mathrm{U}, \mathbf{V}, \boldsymbol{\gamma}_l^x, \boldsymbol{U}_l^x, l = 1,2,\dots\}, \\ &\sigma\{\Sigma_{23}\} \subset \sigma\{\boldsymbol{S}, \boldsymbol{\gamma}, \mathbf{V}, \mathbf{W}, \boldsymbol{\gamma}_l^z, \boldsymbol{W}_l^z, l = 1,2,\dots\}, \end{aligned}$$

so that, by the conditional independence of $\boldsymbol{U}$ and $\boldsymbol{W}$ given $\boldsymbol{S}$, $\boldsymbol{\gamma}$ and $\boldsymbol{V}$, and the independence of $\boldsymbol{\gamma}_l^x, \mathbf{U}_l^x, \boldsymbol{\gamma}_l^z, \mathbf{W}_l^z, l = 1,2,\dots$ from all the remaining sequences, it is immediately obtained that

$$\Sigma_{12} \perp \Sigma_{23} \mid \boldsymbol{S}, \boldsymbol{\gamma}, \mathbf{V}.$$

The proof is finished as in Proposition 3, the only change being that

$$E\left[\prod_{l=1}^{m}\left(\Sigma_{12}(F_l)\right)^{d_l}\middle|\,\sigma\left\{\boldsymbol{S},\boldsymbol{\gamma},\boldsymbol{V}\right\}\right]$$
$$=\int_{F_{l_1,\mathcal{Y}}}\cdots\int_{F_{l_d,\mathcal{Y}}}\widetilde{\psi}\left(y_1,\cdots,y_d\right)\Sigma_2(dy_1)\cdots\Sigma_2(dy_d), \qquad (44)$$

where $d=d_1+\cdots+d_m$, and where $\widetilde{\psi}$ is the bounded measurable function

$$\widetilde{\psi}\left(\boldsymbol{v}\right)=E\left[\delta_{U_1}\left(\bigcap_{i\in A_1(\boldsymbol{v})}F_{l_i,\mathcal{X}}^{\left(v_1'\right)}\right)\middle|V_1=v_1'\right]\ldots E\left[\delta_{U_f}\left(\bigcap_{i\in A_f(\boldsymbol{v})}F_{l_i,\mathcal{X}}^{\left(v_f'\right)}\right)\middle|V_1=v_f'\right]$$
$$\ldots E\left[\prod_{i\in B_1(\boldsymbol{v})}\Xi_{j_1}^{x}\left(F_{l_i,\mathcal{X}}^{\left(y_{j_1}\right)}\right)\right]\ldots E\left[\prod_{i\in B_e(\boldsymbol{v})}\Xi_{j_e}^{x}\left(F_{l_i,\mathcal{X}}^{\left(y_{j_e}\right)}\right)\right],$$

with $\boldsymbol{v}=(v_1,\ldots v_d)$, and

$$\left\{v_1,\ldots,v_d\right\}=\left\{v_1',\ldots,v_f',y_{j_1},\ldots,y_{j_e}\right\},$$

$A_1(\boldsymbol{v}),\ldots,A_f(\boldsymbol{v})$ being defined as in (18), and

$$B_1(\boldsymbol{v})=\{i:v_i=y_{j_1}\},\ldots,B_e(\boldsymbol{v})=\{i:v_i=y_{j_e}\}.$$

■

We call a process having the same law as Σ defined in (43) a hyper Dirichlet process, extending the previously given definition. This law is again denoted by $HDI(\rho_{12},\rho_{23})$. The reader will then expect that in this more general case the hyper Dirichlet process is not necessarily a Dirichlet process. In fact the following result is easily established.

Proposition 10 *The process Σ defined in (43) is a Dirichlet process if and only if for each $l=1,2,...$ either $\psi_{12}^{(l)}$ or $\psi_{23}^{(l)}$ is a single mass. If this happens the parameter measure of Σ is the Markov combination of ρ_{12} and ρ_{23}.*

Proof. If $\psi_{12}^{(l)}$ or $\psi_{23}^{(l)}$ is a single mass it is clear that $\Xi_l^x(dx)\Xi_l^z(dz)\delta_{y_l}(dy)$ is a Dirichlet process with parameter measure having total mass $\psi_2(\{y_l\})$. On the other hand Γ is always a Dirichlet process with parameter measure having total mass $\varphi_2(\mathcal{Y})$. Hence, by Proposition 7, the sufficiency of the condition is established.

The necessity of the condition is obtained from the observation that, by (4) of Proposition 1, for $l = 1, 2, \cdots$

$$\frac{\Sigma(dx \times \{y_l\} \times dz)}{\Sigma(\mathcal{X} \times \{y_l\} \times \mathcal{Z})} = \Xi_l^x(dx)\Xi_l^z(dz)$$

is a Dirichlet process. Conversely, by taking any finite partition $\{B_i\}$ of $\mathcal{X}$ and $\{D_k\}$ of $\mathcal{Z}$, and arguing as in the proof of Proposition 2 for the finite case, the necessity follows.

■

Next our aim is to establish the form of the law of a hyper Dirichlet process Σ conditional to a random sample (X_i, Y_i, Z_i), $i = 1, .., n$, drawn from Σ itself. This information is not supplied by the general conjugacy result proved in Dawid and Lauritzen (Corollary 5.5 of [3]), which holds only for strong hyper Markov processes. We begin by constructing a single observation $(X, Y, Z) = (X_1, Y_1, Z_1)$ in the same space where Σ is constructed. This is achieved by using additional label variables.

Consider first a random probability distribution on some measurable space $(E, \mathcal{E})$ of the form

$$\mathcal{H} = \sum_{l=0}^{+\infty} S_l \Gamma_l, \tag{45}$$

where $\mathbf{\Gamma} = (\Gamma_l, l \geq 0)$ is a sequence of independent random probability distributions on $(E, \mathcal{E})$ with $\Gamma_l \sim \eta_l$, and $\mathbf{S} = (S_l, l \geq 0)$ is a random probability distribution on $\mathbb{N}$ independent of $\mathbf{\Gamma}$, with law ν.

Conditionally to $(\mathbf{S}, \mathbf{\Gamma})$, let N be an integer random variable with law $(S_l, l \geq 0)$, and let ξ have law Γ_N, conditionally to $(\mathbf{S}, \mathbf{\Gamma})$ and N. In this way, conditionally to $\sigma(\mathcal{H})$, ξ has law $\mathcal{H}$ (see Lemma 4.1 in[10]), thus we say that ξ is a sample from $\mathcal{H}$.

Lemma 11 *Suppose there exists a family $\{A_l, l \geq 0\}$ of disjoint subsets of $\mathcal{E}$ such that $\Gamma_l(A_l) = 1$ with probability 1. Then*

$$\mathcal{L}(\mathbf{S}, \mathbf{\Gamma}|\xi) = \mathcal{L}(\mathbf{S}|N = n) \underset{l \neq n}{\otimes} \mathcal{L}(\Gamma_l) \otimes \mathcal{L}(\Gamma_n|\xi), \textit{ if } \xi \in A_n. \tag{46}$$

Proof. The condition on the partition implies that the representation (45) is unique so we may identify $(\mathbf{S}, \mathbf{\Gamma})$ with $\mathcal{H}$, and moreover

$$N = \sum_{l=0}^{\infty} l\, 1_{A_l}(\xi), \tag{47}$$

Hence

$$\mathcal{L}\left(\mathbf{S},\mathbf{\Gamma}|\xi\right)=\mathcal{L}\left(\mathbf{S},\mathbf{\Gamma}|N=n,\xi\right),\quad \text{if } \xi\in A_n, \tag{48}$$

Since by assumption

$$P\left(\mathbf{S}\in d\mathbf{s},N=n,\mathbf{\Gamma}\in d\mathbf{g},\xi\in dx\right)=\nu(d\mathbf{s})s_n\prod_l \eta_l(dg_l)g_n(dx),$$

it is

$$P\left(\mathbf{S}\in d\mathbf{s},\mathbf{\Gamma}\in d\mathbf{g},\xi\in dx|N=n\right)=\frac{\nu(d\mathbf{s})s_n}{\int\nu(d\mathbf{q})q_n}\cdot\prod_{l\neq n}\eta_l(dg_l)\cdot\eta_n(dg_n)g_n(dx),$$

which ensures that $\mathbf{S}$, $\{\Gamma_l,\, l\neq n\}$ and (Γ_n,ξ) are conditionally independent given $N=n$, which implies (46). ■

This observation is the key to establish the following result.

Proposition 12 *Consider a hyper Dirichlet process* $\Sigma\sim HDI\left(\rho_{12},\rho_{23}\right)$*. Assume that, conditionally on* Σ*, the vector of observations*

$$\left((X_i,Y_i,Z_i),\ i=1,...,m\right)$$

has i.i.d. components, each with distribution Σ*. Then, conditionally on the observations, the law of* Σ *is* $HDI\left(\rho_{12}+\sum_{i=1}^m\delta_{(X_i,Y_i)},\rho_{23}+\sum_{i=1}^m\delta_{(Y_i,Z_i)}\right).$

Proof. Let us consider the case of a sample of size $m=1$. For larger values of m the result is obtained through the standard recursion argument for constructing posterior distributions (see e.g. Sethuraman [10], page 648). The hyper Dirichlet process Σ defined in (43) has the form (45) with $\Gamma_0(dx,dy,dz)=\Gamma(dx,dy,dz)$ and $\Gamma_l(dx,dy,dz)=\Xi_l^x(dx)\Xi_l^z(dz)\delta_{y_l}(dy)$, $l=1,2,...$, and in our case $\xi=(X,Y,Z)$. Moreover $A_0=\big\{(x,y,z)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}:\ y\neq y_l, l=1,2,...\big\}$ and $A_l=\mathcal{X}\times\{y_l\}\times\mathcal{Z},\ l=1,2,...$

Since $\mathbf{S}\sim DI(\rho_2^{(l)}(\mathcal{Y}),\ l\in\mathbb{N})$ and N is drawn from $\mathbf{S}$, the standard result about conjugacy of Dirichlet processes ([10], Lemma 4.1) yields that $\mathbf{S}$ is $DI(\rho_2^{(l)}(\mathcal{Y})+\delta_n(l),\ l\in\mathbb{N})$, conditional to $N=n$. Next, for the same reason, the law of Γ conditional to (X,Y,Z) (for $Y\notin\{y_l,l=1,2,...\}$) is $DI(\rho+\delta_{(X,Y,Z)})$, where ρ is the Markov combination of ρ_{12} and ρ_{23} . It is easily seen that $\rho+\delta_{(X,Y,Z)}$ is the Markov combination of $\rho_{12}+\delta_{(X,Y)}$ and $\rho_{23}+\delta_{(Y,Z)}$, thus

$$DI(\rho+\delta_{(X,Y,Z)})=HDI(\rho_{12}+\delta_{(X,Y)},\rho_{23}+\delta_{(Y,Z)}).$$

Finally, when $Y = y_l$, the law of the independent Dirichlet processes Ξ_l^x and Ξ_l^z given samples X and Z extracted independently from each of them is the product law of two Dirichlet processes $\psi_{12}^{(l)} + \delta_X$ and $\psi_{23}^{(l)} + \delta_Z$. Putting together all these results in (46) and using Proposition 9 with $\rho_{12} + \delta_{(X,Y)}$ and $\rho_{23} + \delta_{(X,Y,Z)}$ replacing ρ_{12} and ρ_{23} the statement of the proposition for $m = 1$ is immediately obtained. ■

Remark 13 *The previous theorem states that the class of hyper Dirichlet processes is conjugate w.r.t. random sampling, as it happens for the class of Dirichlet processes. Now observe that the law of a hyper Dirichlet process conditional to the observations has always parameter measures with atomic components. But if the prior law is $HDI(\rho_{12}, \rho_{23})$ with $\rho_2 = \varphi_2$ diffuse, the probability of observing $(Y_i,\ i = 1, ..., m)$ with all distinct components is 1. By consequence the atomic component of the posterior parameter measure is a sum of Dirac measures, hence, by Proposition 10, the posterior law is not only the law of a hyper Dirichlet process but actually the law of a Dirichlet process, as it results directly from the conjugacy property of these latter processes.*

We now finally pass to a limit result similar to Theorem 4 for general Hyper Dirichlet processes. From now on we take again $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ to be Euclidean spaces with dimensions d_1, d_2 and d_3, respectively, and assume that for each of these spaces we have a sequence of finite partitions as in (19). Correspondingly, we discretize along the corresponding product sequences of partitions in two-dimensional cells (20) the parameter measures $\rho_{12} = \varphi_{12} + \psi_{12}$ and $\rho_{23} = \varphi_{23} + \psi_{23}$, thereby getting the tables

$$\rho_{12}^{(n)} = \varphi_{12}^{(n)} + \psi_{12}^{(n)}, \quad \rho_{23}^{(n)} = \varphi_{23}^{(n)} + \psi_{23}^{(n)},$$

where $\varphi_{12}^{(n)}$ and $\varphi_{23}^{(n)}$ are defined in (21), and

$$\begin{aligned} \psi_{12}^{(n)} &= \left\{ \psi_{12}(B_i^{(n)} \times C_j^{(n)}),\ i \in I^{(n)}, j \in J^{(n)} \right\}, \\ \psi_{23}^{(n)} &= \left\{ \psi_{23}(C_j^{(n)} \times D_k^{(n)}),\ j \in J^{(n)}, k \in K^{(n)} \right\}. \end{aligned}$$

For each integer n consider a $HDi(\rho_{12}^{(n)}, \rho_{23}^{(n)})$ distributed random array

$$\left\{ F^{(n)}(i,j,k),\ i \in I^{(n)}, j \in J^{(n)}, k \in K^{(n)} \right\} \tag{49}$$

which has to be compared with

$$\left\{ \Sigma^{(n)}(i,j,k) = \Sigma(B_i^{(n)} \times C_j^{(n)} \times D_k^{(n)}),\ i \in I^{(n)}, j \in J^{(n)}, k \in K^{(n)} \right\} \tag{50}$$

where $\Sigma \sim HDI\left(\rho_{12}, \rho_{23}\right)$ is defined in (43). However, for our next result it is more convenient to use a different representation for this process which is constructed next.

Let us proceed to construct both (49) and (50) on the same probability space. We start by defining $\Gamma^{(n)}$ and $G^{(n)} = G^{(n)}_{3|2}$ as in Theorem 4. All the other random variables to be defined will be taken independent of these two arrays. Let $S_0 \sim Be(\varphi_2(\mathcal{Y}), \psi_2(\mathcal{Y}))$ be the weight of the diffuse part in (43). Then we split the atomic part as a mixture of distributions, each concentrated on a set of the partition $\left\{\mathcal{X} \times C^{(n)}_j \times \mathcal{Z}\ , j \in J^{(n)}\right\}$. For this we need first an independent vector of masses

$$(M^{(n)}_1, ..., M^{(n)}_{\left|J^{(n)}\right|}) \sim Di(\psi_2(C^{(n)}_j), j \in J^{(n)})$$

Next we have to split the mass inside each set of the partition. So for each $j \in J^{(n)}$ such that $\psi_2(C^{(n)}_j) > 0$, define the following random variables. First an independent sequence of independent random variables $\mathbf{L}^{(n)}_j = \left\{L^{(n)}_{j,m}, m = 1, 2, ...\right\}$ with

$$L^{(n)}_{j,m} \sim Be\left(\psi_2(y^{(n)}_{j,m}), \sum_{i=m+1}^{\infty} \psi_2(y^{(n)}_{j,i})\right), \quad m = 1, 2, ...,$$

where $\left\{y^{(n)}_{j,m},\ m = 1, 2, ...\right\}$ is the subset $\left\{y_l,\ l = 1, 2, ...\right\} \cap C^{(n)}_j$, with the points ordered in an arbitrary way. The sequence $\mathbf{L}^{(n)}_j$ is used to obtain the corresponding sequence of weights $\mathbf{U}^{(n)}_j = \mathbf{T}(\mathbf{L}^{(n)}_j)$, i.e.

$$U^{(n)}_{j,m} = \prod_{i=1}^{m-1}(1 - L^{(n)}_{j,i})L^{(n)}_{j,m}, \quad m = 1, 2, ...$$

which is a Dirichlet process on $\mathbb{N}$ with parameter measure $\sum_m \psi_2(y^{(n)}_{j,m})\delta_m$, as it results by the application of Proposition 7. Notice that in case the sequence $\left\{y^{(n)}_{j,m},\ m = 1, 2, ...\right\}$ is finite, $\mathbf{L}^{(n)}_j$ and $\mathbf{U}^{(n)}_j$ are both taken to be eventually zero. Observe that, by Proposition 1, the joint law of

$$\left\{M^{(n)}_j U^{(n)}_{j,m},\ m = 1, 2, ..., j \in J^{(n)}\right\}$$

is the same as that of

$$\left\{\frac{S^{(n)}_{j,m}}{1 - S_0},\ m = 1, 2, ..., j \in J^{(n)}\right\},$$

where $S_{j,m}^{(n)}$ is defined to be S_l as given by (42) for the index $l \geq 1$, such that $y_l = y_{j,m}^{(n)}$. Moreover both these sequences are independent of S_0.

Finally, for each j such that $\psi_2(C_j^{(n)}) > 0$, mutually independent Dirichlet processes $\Xi_{j,m}^{x,(n)}$ and $\Xi_{j,m}^{z,(n)}$ are defined on $\mathcal{X}$ and $\mathcal{Z}$, with parameter measures $\psi_{12}(dx \times \{y_{j,m}^{(n)}\})$ and $\psi_{23}(\{y_{j,m}^{(n)}\} \times dz),\ m = 1, 2, ...$ We have thus completed the construction of the probability space (depending on n) over which we define

$$\Sigma^{(n)}(i,j,k) = S_0\Gamma^{(n)}(i,j,k) + (1-S_0)M_j^{(n)} \sum_m U_{j,m}^{(n)}\Xi_{j,m}^{x,(n)}(B_i^{(n)})\Xi_{j,m}^{z,(n)}(D_k^{(n)}) \quad (51)$$

As far as the approximation is concerned, we also need independent copies of the sequences $\mathbf{L}_j^{(n)} = \left\{L_{j,m}^{(n)}, m = 1, 2, ...\right\}$, say $\mathbf{L}_j'^{(n)} = \left\{L_{j,m}'^{(n)}, m = 1, 2, ...\right\}$, with the corresponding sequence of weights $\mathbf{U}_j'^{(n)} = \left\{U_{j,m}'^{(n)}, m = 1, 2, ...\right\} = \mathbf{T}(\mathbf{L}_j'^{(n)})$, and an independent random vector with independent components $O_j^{(n)} \sim Be(\varphi_2(C_j^{(n)}), \psi_2(C_j^{(n)}))$, $j \in J^{(n)}$. Then we define

$$F^{(n)}(i,j,k) = \left\{ S_0\Gamma_{12}^{(n)}(i,j) + (1-S_0)M_j^{(n)} \sum_m U_{j,m}^{(n)}\Xi_{j,m}^{x,(n)}(B_i) \right\} \cdot \left\{ O_j^{(n)}G^{(n)}(k|j) + (1-O_j^{(n)}) \sum_p U_{j,p}'^{(n)}\Xi_{j,p}^{z,(n)}(D_k) \right\} \quad (52)$$

By using Proposition 1 and Proposition 7, it is easy to check that the marginal laws of the arrays $\Sigma^{(n)}$ and $F^{(n)}$ are the required ones. Moreover the following result holds.

Theorem 14 *Under the assumptions* **ii)** *and* **iii)** *of Theorem 4*

$$\lim_{n\to\infty} \sum_{i\in I^{(n)}} \sum_{j\in J^{(n)}} \sum_{k\in K^{(n)}} E|\Sigma^{(n)}(i,j,k) - F^{(n)}(i,j,k)| = 0.$$

Proof. We subtract (52) from (51), and we bound from above its absolute value by

$$S_0|\Gamma^{(n)}(i,j,k) - \Gamma^{(n)}(i,j)O_j^{(n)}G^{(n)}(k|j)| \tag{53}$$

$$+ (1-S_0)M_j^{(n)}\sum_m U_{j,m}^{(n)}\Xi_{j,m}^x(B_i)\left|\Xi_{j,m}^{z,(n)}(D_k) - (1-O_j^{(n)})\sum_p U_{j,p}'^{(n)}\Xi_{j,p}^{z,(n)}(D_k)\right| \tag{54}$$

$$+ S_0\Gamma^{(n)}(i,j)(1-O_j^{(n)})\sum_p U_{j,p}'^{(n)}\Xi_{j,p}^{z,(n)}(D_k) \tag{55}$$

$$+ (1-S_0)M_j^{(n)}O_j^{(n)}G^{(n)}(k|j)\sum_m U_{j,m}^{(n)}\Xi_{j,m}^{x,(n)}(B_i). \tag{56}$$

We then proceed to bound each of the four terms. For the first term (53) notice that

$$\begin{aligned}
&S_0|\Gamma^{(n)}(i,j,k) - \Gamma^{(n)}(i,j)O_j^{(n)}G^{(n)}(k|j)| \\
&\le S_0|\Gamma^{(n)}(i,j,k) - \Gamma^{(n)}(i,j)G^{(n)}(k|j)| + S_0\Gamma^{(n)}(i,j)(1-O_j^{(n)})G^{(n)}(k|j)
\end{aligned}$$

Since $\Gamma^{(n)}(i,j)G^{(n)}(k|j) = G^{(n)}(i,j,k)$, by summing over i,j,k and taking the expected value, the first term goes to zero in the mean by Theorem 4. For the second term, by summing and taking the mean we get

$$\begin{aligned}
&E\left[S_0\right]\sum_{i,j,k} E\left[\Gamma^{(n)}(i,j)\right] E\left[1-O_j^{(n)}\right] E\left[G^{(n)}(k|j)\right] \\
&\le \frac{\varphi_2(\mathcal{Y})}{\varphi_2(\mathcal{Y})+\psi_2(\mathcal{Y})}\sum_{i,j}\frac{\varphi_{12}(B_i^{(n)}\times C_j^{(n)})}{\varphi_{12}(\mathcal{X}\times\mathcal{Y})}\frac{\psi_2(C_j^{(n)})}{\varphi_2(C_j^{(n)})+\psi_2(C_j^{(n)})} \\
&= \frac{1}{\varphi_2(\mathcal{Y})+\psi_2(\mathcal{Y})}\sum_j \frac{\varphi_2(C_j^{(n)})\psi_2(C_j^{(n)})}{\varphi_2(C_j^{(n)})+\psi_2(C_j^{(n)})} =: \chi_n.
\end{aligned}$$

In order to deal with the second term (54) observe that

$$\begin{aligned}
&\left|\Xi_{j,m}^{z,(n)}(D_k) - (1-O_j^{(n)})\sum_p U_{j,p}'^{(n)}\Xi_{j,p}^{z,(n)}(D_k)\right| \\
&\le (1-O_j^{(n)})\sum_p U_{j,p}'^{(n)}\left|\Xi_{j,m}^{z,(n)}(D_k) - \Xi_{j,p}^{z,(n)}(D_k)\right| + O_j^{(n)}\Xi_{j,m}^{z,(n)}(D_k).
\end{aligned}$$

Then, inserting the above estimate in (54), summing both sides over i, j, k, and taking the expected value we get the upper bound

$$E(1-S_0)\cdot$$
$$\cdot\Big\{\sum_j EM_j^{(n)}E(1-O_j^{(n)})\sum_m\sum_{p\neq m}EU_{j,m}^{(n)}EU_{j,p}^{\prime(n)}\sum_k E\left|\Xi_{j,m}^{z,(n)}(D_k)-\Xi_{j,p}^{z,(n)}(D_k)\right|$$
$$+\sum_j EM_j^{(n)}EO_j^{(n)}\Big\} \tag{57}$$

The last term in (57) is the easiest to deal with, since it is equal to

$$E(1-S_0)\sum_j EM_j^{(n)}EO_j^{(n)}$$
$$=\frac{\psi_2(\mathcal{Y})}{\varphi_2(\mathcal{Y})+\psi_2(\mathcal{Y})}\sum_j\frac{\psi_2(C_j^{(n)})}{\psi_2(\mathcal{Y})}\frac{\varphi_2(C_j^{(n)})}{\varphi_2(C_j^{(n)})+\psi_2(C_j^{(n)})}=\chi_n,$$

whereas the first term in (57) is bounded by

$$2E(1-S_0)\sum_j EM_j^{(n)}E(1-O_j^{(n)})\sum_{m\neq p}\sum EU_{j,m}^{(n)}EU_{j,p}^{\prime(n)}$$
$$=2\frac{\psi_2(\mathcal{Y})}{\varphi_2(\mathcal{Y})+\psi_2(\mathcal{Y})}\sum_j\frac{\psi_2(C_j^{(n)})^2}{\psi_2(\mathcal{Y})\left(\varphi_2(C_j^{(n)})+\psi_2(C_j^{(n)})\right)}\sum_m\sum_{p\neq m}\frac{\psi_2(\{y_{j,m}^{(n)}\})\psi_2(\{y_{j,p}^{(n)}\})}{\psi_2(C_j^{(n)})^2}$$
$$\leq\frac{4}{\varphi_2(\mathcal{Y})+\psi_2(\mathcal{Y})}\sum_j\frac{\psi_2(C_j^{(n)})}{\varphi_2(C_j^{(n)})+\psi_2(C_j^{(n)})}\sum_{p>1}\psi_2(\{y_{j,p}^{(n)}\})$$
$$\leq\frac{4}{\varphi_2(\mathcal{Y})+\psi_2(\mathcal{Y})}\sum_j\sum_{p>1}\psi_2(\{y_{j,p}^{(n)}\})=:\tau_n,$$

where we have used the fact that, for any sequence $a_n\geq 0$

$$\sum_n\sum_{m\neq n}a_na_m=2\sum_n\sum_{m>n}a_na_m\leq\sum_n a_n\sum_{m>1}a_m.$$

Likewise, the expected value of the sum of the third terms (55) yields

$$ES_0\sum_j E(\sum_i\Gamma^{(n)}(i,j))E(1-O_j^{(n)})$$
$$=\frac{\varphi_2(\mathcal{Y})}{\varphi_2(\mathcal{Y})+\psi_2(\mathcal{Y})}\sum_j\frac{\varphi_2(C_j^{(n)})}{\varphi_2(\mathcal{Y})}\frac{\psi_2(C_j^{(n)})}{\varphi_2(C_j^{(n)})+\psi_2(C_j^{(n)})}=\chi_n,$$

and for the last terms (56) we bound the sum of the expected values by

$$E(1-S_0)\sum_j EM_j^{(n)}EO_j^{(n)} = \chi_n.$$

We finally end the proof by noticing that, since ψ_2 is a finite measure, for any $\epsilon > 0$ there exists an integer L_ϵ such that

$$\sum_{l=1}^{L_\epsilon} \psi_2(\{y_l\}) \geq \psi_2(\mathcal{Y}) - \epsilon;$$

moreover, for n large enough each of the $C_j^{(n)}$ will contain either one or none of the $y_l,\ l = 1, 2, ..., L_\epsilon$. Therefore, for n large enough, $\sum_j \sum_{p>1} \psi_2(\{y_{j,p}^{(n)}\}) \leq \epsilon$, so that

$$\tau_n \leq \frac{4\epsilon}{\varphi_2(\mathcal{Y}) + \psi_2(\mathcal{Y})}$$

and analogously, for n large enough,

$$\chi_n \leq \frac{1}{\varphi_2(\mathcal{Y}) + \psi_2(\mathcal{Y})}\left(\epsilon + L_\epsilon \max_{j\in J^{(n)}} \varphi_2(C_j^{(n)})\right) \leq \frac{2\epsilon}{\varphi_2(\mathcal{Y}) + \psi_2(\mathcal{Y})}$$

since $\max_{j\in J^{(n)}} \varphi_2(C_j^{(n)}) = \lambda_n$ goes to zero as $n \to \infty$, as it has been proved in Theorem 4. Thus the proof is finished. ∎

As a corollary of the above result we get immediately the announced generalization of Theorem 5.

Theorem 15 *Under the assumption of the previous theorem, let $F^{(n)}$ be any random probability distribution on $\mathcal{X}\times\mathcal{Y}\times\mathcal{Z} = \mathbb{R}^d$, with the property that*

$$\left\{F^{(n)}(B_i^{(n)} \times C_j^{(n)} \times D_k^{(n)}), i \in I^{(n)}, j \in J^{(n)}, k \in K^{(n)}\right\}$$

has the same law as

$$\left\{F^{(n)}(i,j,k), i \in I^{(n)}, j \in J^{(n)}, k \in K^{(n)}\right\}.$$

Furthermore suppose that (32) and (33) hold for all compact sets $E \subset \mathcal{X}$ and $F \subset \mathcal{Z}$. Then $\left\{F^{(n)}\right\}$ converges weakly to Σ, as probability measures on $\mathbb{R}^d$, endowed with the topology of weak convergence.

References

[1] Billingsley, P. *Convergence of probability measures*, second ed. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, 1999. A Wiley-Interscience Publication.

[2] Blackwell, D., and MacQueen, J. B. Ferguson distributions via Pólya urn schemes. *Ann. Statist. 1* (1973), 353–355.

[3] Dawid, A. P., and Lauritzen, S. L. Hyper-Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist. 21*, 3 (1993), 1272–1317.

[4] Dubins, L., and Freedman, D. Measurable sets of measures. *Pacific J. Math. 14* (1964), 1211–1222.

[5] Ferguson, T. S. A Bayesian analysis of some nonparametric problems. *Ann. Statist. 1* (1973), 209–230.

[6] Kotz, S., Balakrishnan, N., and Johnson, N. L. *Continuous multivariate distributions. Vol. 1*, second ed. Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience, New York, 2000. Models and applications.

[7] Lauritzen, S. L. *Graphical models*, vol. 17 of *Oxford Statistical Science Series.* The Clarendon Press Oxford University Press, New York, 1996. Oxford Science Publications.

[8] Lindvall, T. *Lectures on the coupling method.* Dover Publications Inc., Mineola, NY, 2002. Corrected reprint of the 1992 original.

[9] Rachev, S. T. *Probability metrics and the stability of stochastic models.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1991.

[10] Sethuraman, J. A constructive definition of Dirichlet priors. *Statist. Sinica 4*, 2 (1994), 639–650.

[11] Sethuraman, J., and Tiwari, R. C. Convergence of Dirichlet measures and the interpretation of their parameter. In *Statistical decision theory and related topics, III, Vol. 2 (West Lafayette, Ind., 1981).* Academic Press, New York, 1982, pp. 305–315.