The hyper-Dirichlet process and its discrete approximations: The butterfly model

C. Ascia\(^a\), G. Nappo\(^b\), M. Piccioni\(^b,\ast\)

\(^a\)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
\(^b\)Dipartimento di Matematica, Università di Roma “La Sapienza”, piazzale A. Moro 2, I 00185 Roma, Italy

Received 13 February 2004
Available online 25 October 2005

Abstract

The aim of this paper is the study of some random probability distributions, called hyper-Dirichlet processes. In the simplest situation considered in the paper these distributions charge the product of three sample spaces, with the property that the first and the last component are independent conditional to the middle 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, introduced in [A.P. Dawid, S.L. Lauritzen, Hyper-Markov laws in the statistical analysis of decomposable graphical models, Ann. Statist. 21 (3) (1993) 1272–1317], of these two Dirichlet processes. The processes constructed in this way in fact are in fact generalizations of the hyper-Dirichlet laws on contingency tables considered in the above paper. Our main result is the convergence to the 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.

© 2005 Elsevier Inc. All rights reserved.

AMS 1991 subject classification: 60G57; 62G99; 62F15; 60F99

Keywords: Dirichlet laws and processes; Bayesian statistics; Markov distributions; Weak convergence; Coupling

1. Introduction

In a pioneering paper Dawid and Lauritzen [3] defined hyper-Markov combinations of families of random probability measures. Realizations of these processes are random probability measures...
over a sample space $\mathcal{X}_V = \bigotimes_{v \in V} \mathcal{X}_v$ whose realizations are Markov w.r.t. a specified decomposable graph structure imposed on $V$. By this we mean that separation relations on the graph translate into conditional independence relations between the components. The decomposability of the graph implies that, provided some regular conditional probabilities exist, a Markov distribution on the whole sample space can be uniquely derived from the marginal distributions of the components belonging to each of the cliques of the graph. 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 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 has 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, a property which is naturally obtained by “lifting” the Markov property to the higher level of laws for random distributions.

Hyper-Markov laws are natural candidates as conjugate prior laws for the Bayesian analysis of decomposable graphical models. In particular the strong hyper-Markov property allows a separation of inferential problems at the level of each clique, allowing to construct the posterior distribution as the hyper-Markov combination of the posteriors for each clique data.

The hyper-Dirichlet law, which is the hyper-Markov combination of hyperconsistent Dirichlet laws, can be used for graphical models with categorical variables, in which $\mathcal{X}_v$ is finite for every $v \in V$ [3]: it has the strong Markov property referred above.

For graphical models with continuous variables, for instance for $\mathcal{X}_v = \mathbb{R}$ for every $v \in V$, the parametric family of (zero mean) multivariate Gaussian laws has been almost exclusively used [11,6]. These laws are parametrized by the covariance matrix, and the conditional independence relation between two variables given all the others is equivalent to the vanishing of their partial correlation, namely the correlation coefficient between their residuals from the linear regression over the other variables. Since partial correlations are (up to scalings) the extra-diagonal elements of the inverse covariance matrix, prior laws for graphical Gaussian models should produce random matrices with a certain structure of zeros. The hyperinverse Wishart law, which is the hyper-Markov combination of hyperconsistent inverse Wishart laws has been proposed for this purpose in [3]; it is strong hyper-Markov, too.

The validity of inferences for graphical Gaussian models can be extended to more general kind of distributions by replacing conditional independence relations with zero partial correlation relations. As stated before, the two relations are obviously equal in the Gaussian case.

However, these two relations are in general quite different. It is well known that independence implies lack of correlation but not conversely. On the other hand, when dealing with three random variables $X, Y$ and $Z$, we could have not only that

(a) The partial correlation between $X$ and $Z$ given $Y$ is zero without $X$ and $Z$ being independent given $Y$;

but also the converse situation

(b) $X$ and $Z$ are independent given $Y$ without the partial correlation between $X$ and $Z$ given $Y$ being zero (as an example, it suffices to take $(X, Y)$ and $(Z, Y)$ equal in law, without $E(X|Y)$ being linear).

Thus zero partial correlation is a weak surrogate for conditional independence for non-Gaussian models. When lack of Gaussianity is suspected, a non-parametric approach may be considered.
Leaving conditional independence requirements aside for the moment, we recall that the most widely known class of priors for non-parametric problems (i.e. random distributions on general state spaces $X$) is still given by Dirichlet processes. They were introduced by Ferguson [4], and then studied by Blackwell and MacQueen [2], Sethuraman and Tiwari [10] and Sethuraman [9], between others.

In the attempt to incorporate conditional independence requirements (as given by a decomposable graphical model) on the realizations, it is then natural to consider hyper-Dirichlet processes, obtained as hyper-Markov combinations of hyperconsistent Dirichlet processes. The study of these processes is the aim of the present paper. To our knowledge it is the first attempt to introduce graphical concepts in the area of non-parametric Bayesian statistics.

In order to emphasize 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. This means that the set of vertices $V$ is the union of two proper subsets $A$ and $B$, and edges between two vertices $i$ and $j$ are absent from the graph if and only if $i \in A \setminus B$ and $j \in B \setminus A$. By mapping all the vertices in $A \setminus B$, in $A \cap B$ and $B \setminus A$ into single vertices, say 1, 2 and 3, respectively, to which we attach the enlarged sample spaces $X_1 = \otimes_{v \in A \setminus B} X_v$, $X_2 = \otimes_{v \in A \cap B} X_v$ and $X_3 = \otimes_{v \in B \setminus A} X_v$ and joining 1 and 2 and 2 and 3 with an edge, a canonical graphical model is obtained, which is completely equivalent to the previous one. In the following the word Markov or hyper-Markov is always referred to this graph.

A rather famous example of a non-trivial decomposition in two complete subgraphs was considered in the introduction of [11] and called there a butterfly model for its peculiar shape; here we extend this terminology to any non-trivial decomposition in two complete subgraphs. Therefore, when dealing with any butterfly model, we can use the mapping described above to reduce ourselves to the canonical butterfly model. This means that we should produce probability distributions on the product $X \times Y \times Z$ obeying the single conditional independence constraint that the first and the last component are independent conditional to the middle one. Statistical models of this kind can be thought as factorial models in which the middle component is a factor which entirely explains the association between the two measurements represented by the first and the last component.

In order to construct a hyper-Markov law for the canonical butterfly model we need to specify two hyperconsistent laws for the marginal distribution on $X \times Y$, and for the marginal distribution on $Y \times Z$, respectively. When these are the laws of two Dirichlet processes, hyperconsistency is clearly equivalent to consistency of the two parameter measures, i.e. these measures have to induce the same marginal measure on the middle component. Statistical models of this kind can be thought as factorial models in which the middle component is a factor which entirely explains the association between the two measurements represented by the first and the last component.

In order to construct a hyper-Markov law for the canonical butterfly model we need to specify two hyperconsistent laws for the marginal distribution on $X \times Y$, and for the marginal distribution on $Y \times Z$, respectively. When these are the laws of two Dirichlet processes, hyperconsistency is clearly equivalent to consistency of the two parameter measures, i.e. these measures have to induce the same marginal measure on the middle component. In Section 2 the construction of the hyper-Dirichlet process is performed when the middle marginal is diffuse, i.e. it does not have atoms. In this case the 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, firstly in the diffuse case, the relation between a hyper-Dirichlet process and the hyper-Dirichlet laws which are obtained by natural discretizations of their parameter measures. Our main result is that, under suitable regularity assumptions, as the discretizations refine, the hyper-Dirichlet process appears to be the weak limit of the sequence of “discretized” hyper-Dirichlet laws. This is contained in Section 3. 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 middle marginal is diffuse. 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 then prove that hyper-Dirichlet processes are conjugate with respect to random i.i.d. sampling. More precisely, despite hyper-Dirichlet process are not strongly hyper-Markov, the posterior law of a hyper-Dirichlet process is the hyper-Markov combination of the posterior laws for the Dirichlet processes on the first two components and on the last two components, respectively.

As long as Dirac masses coming from distinct observations are added to a diffuse prior parameter measure, the hyper-Dirichlet process remains a Dirichlet process, in agreement with the conjugacy property holding for Dirichlet processes [4,9]. However it is possible to exhibit hyper-Dirichlet processes which are not Dirichlet processes and in fact to characterize them in terms of their parameter measures. The last results presented in Section 4 are the extensions of the limit theorems proved in Section 3 to general hyper-Dirichlet processes for butterfly models.

Section 5 is devoted to a brief discussion concerning the extension of the previous results to general decomposable graphical models.

2. The hyper-Dirichlet law and process for the butterfly model

Dirichlet laws are supported by an unit simplex in some Euclidean space, hence they are used as laws for random probability distributions over a finite set of atoms \( L \). For ease of notation the set \( L \) is always identified with \( \{1, \ldots, |L|\} \), without loss of generality.

First let us suppose that \(|L| > 1\). For a given vector \( \{\phi(l) > 0, l \in L\} \), we say that the random vector \( \{G(l), l \in L\} \) is \( Di \{\phi(l), l \in L\} \) if

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

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

\[
p(g_1, \ldots, g_{|L|-1}) = \frac{\Gamma(\phi(L))}{\prod_{l=1}^{|L|} \Gamma(\phi(l))} \prod_{l=1}^{|L|-1} g_l^{\phi(l)-1} \left( 1 - \sum_{i=1}^{|L|-1} g_i \right)^{\phi(|L)|-1},
\]

where \( \phi(L) = \sum_{l=1}^{|L|} \phi(l) \). In the sequel we will also allow \( \phi(l) = 0 \) for some (but not all) index \( l \), in which case we mean that these \( l \)'s are excluded and \( G(l) \equiv 0 \).

For any non-null contingency table

\[
\phi_{12} = \{\phi_{12}(i, j) > 0, i \in I, j \in J\},
\]
where $I$ and $J$ are finite sets, we use the standard notation for the marginal

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

and for the conditionals

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

By using the representation of the Dirichlet law as a function of independent Gamma variables (see e.g. [5]), the following fundamental result is immediately obtained.

**Proposition 1.** For a random contingency table

$$
G_{12} \sim Di(\varphi_{12}), \quad (1')
$$

the marginal vector $G_2$ and the vectors of conditional distributions $\{G_{1|2}(\cdot|j)\}$ for $j \in I$ such that $\varphi_2(j) > 0$, are mutually independent, with

$$
G_2 \sim Di(\varphi_2), \quad (2)
$$

$$
G_{1|2}(\cdot|j) \sim Di(\varphi_{12}(\cdot, j)), \quad \forall j \in J : \varphi_2(j) > 0. \quad (3)
$$

If $\varphi_2(j) = 0$ for some $j \in J$, then $G_{12}(i, j)$ is identically zero for $i \in I$.

Conversely, if $G_{12}$ is a random table with the properties (2) and (3), such that $G_{12}(i, j) = 0$ whenever $\varphi_2(j) = 0$, then $G_{12} \sim Di(\varphi_{12})$.

Let us consider $\varphi_{12}$ as in $(1')$ and consider another non-null contingency table

$$
\varphi_{23} = \{ \varphi_{23}(j, k), \ j \in J, \ k \in K \}, \quad (4)
$$

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

$$
\sum_{i \in I} \varphi_{12}(i, j) = \sum_{k \in K} \varphi_{23}(j, k) = \varphi_2(j), \quad \text{for any } j \in J. \quad (5)
$$

The Markov combination $\varphi$ of $\varphi_{12}$ and $\varphi_{23}$ 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 $\varphi$ is the only contingency table over $I \times J \times K$ with the properties:

(a) the marginals of $\varphi$ over $I \times J$ and $J \times K$ equal to $\varphi_{12}$ and $\varphi_{23}$, 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 $\varphi$ is normalized it becomes a Markov probability distribution over the product $I \times J \times K$. By this we mean that projections on the first and the last factor are independent conditional to the projection on the middle 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 (1') and consider independent vectors

\[
G_{3|2}(\cdot|j) \sim Di(\phi_{23}(j, \cdot)) , \quad j \in J : \phi_{2}(j) > 0 ,
\]

(6)

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

\[
G(i, j, k) = G_{12}(i, j)G_{3|2}(k|j) = G_{1|2}(i|j)G_{2}(j)G_{3|2}(k|j)
\]

for \( i \in I, j \in J, k \in K \).

(7)

Recalling (2) and (3), we observe that

(i) the laws of the marginals over \( I \times J \) and \( J \times K \) are, respectively,

\[
G_{12} \sim Di(\phi_{12}) , \quad G_{23} = \{ G_{2}(j)G_{3|2}(k|j), j \in J, k \in K \} \sim Di(\phi_{23}) ,
\]

(ii) 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 Markov probability distributions;

(iii) the families of conditional distributions

\[
\{ G_{1|2}(\cdot|j), j \in J \} , \quad \{ G_{3|2}(\cdot|j), j \in J \}
\]

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)–(iii). As such it is called the hyper-Markov combination of the laws \( Di(\phi_{12}) \) and \( Di(\phi_{23}) \). Following [3] we call this law a hyper-Dirichlet law and indicate it by \( HDi(\phi_{12}, \phi_{23}) \).

The following proposition is worth of notice.

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

\[
I_{j} = \{ i : \phi_{12}(i, j) > 0 \} , \quad K_{j} = \{ k : \phi_{23}(j, k) > 0 \} ,
\]

is a singleton. In this case

\[
HDi(\phi_{12}, \phi_{23}) = Di(\phi) ,
\]

where \( \phi \) is the Markov combination of \( \phi_{12} \) and \( \phi_{23} \).

**Proof.** If \( G \) is \( HDi(\phi_{12}, \phi_{23}) \) distributed then \( G_{2} \sim Di(\phi_{2}) \). Hence by Proposition 1 it is clear that \( G \) is \( Di(\bar{\phi}) \) (for some contingency table \( \bar{\phi} \)), if and only if \( \bar{\phi}_{2}(j) = \phi_{2}(j) \) for any \( j \), and each of the conditionals

\[
\{ G_{13|2}(i, k|j) = G_{1|2}(i|j)G_{3|2}(k|j), i \in I, k \in K \} , \quad \forall j : \phi_{2}(j) > 0 ,
\]

(8)

is \( Di((\bar{\phi}(i, j, k), i \in I, k \in K)) \). Then

\[
\bar{\phi}(I \times J \times K) = \bar{\phi}_{2}(J) = \phi_{2}(J) = \phi(I \times J \times K) .
\]

By (1) \( G \sim Di(\bar{\phi}) \) implies

\[
E [G(i, j, k)] = \frac{\bar{\phi}(i, j, k)}{\bar{\phi}(I \times J \times K)} ,
\]
whereas $G \sim HDi(\varphi_{12}, \varphi_{23})$ and $\varphi_2(j) > 0$ imply
\[
E[G(i, j, k)] = E\left[G_{12}(i, j)G_{32}(k|j)\right] = E\left[G_{12}(i, j)\right] E\left[G_{32}(k|j)\right]
\]
\[
= \frac{\varphi_{12}(i, j)}{\varphi_{12}(I \times J)} \frac{\varphi_{23}(j, k)}{\varphi_2(j)} = \frac{\varphi(i, j, k)}{\varphi(I \times J \times K)},
\]

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

Next, whenever $\varphi_2(j) > 0$, $G_{13|2}(\cdot|j)$ is $Di\{\varphi(i, j, k), i \in I, k \in K\}$, from which its law is concentrated on an $I_j \times K_j - 1$ manifold of $R^{I_j \times K_j}$ but not on any submanifold of lower dimension. On the other hand, using (8), $G_{13|2}(\cdot|j)$ is concentrated on an $I_j + K_j - 2$ 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 $\sigma$-algebra which makes the functions $P \in \mathcal{P}(L) \to P(A)$ measurable, for all $A \in \mathcal{L}$. A random distribution $\Gamma$ 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 $\Gamma$ is uniquely determined by the joint law of $(\Gamma(A_1), \ldots, \Gamma(A_m))$ for all finite measurable partitions $\{A_h, h = 1, \ldots, m\}$ of $L$. Given any finite measure $\varphi$ on $(L, \mathcal{L})$, we say that $\Gamma$ is a Dirichlet process with parameter measure $\varphi$ if this joint law is $Di\{\varphi(A_h), h = 1, \ldots, m\}$.

The law of this process will be indicated by capitals $DI(\varphi)$, 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 [10] and further analyzed in [9].

In order to introduce this representation, first define $T : (0, 1)^\infty \to (0, 1)^\infty$ with
\[
T_l(b_l) = b_l \prod_{i=1}^{l-1} (1 - b_l), \quad l = 1, 2, \ldots
\]

with the usual convention that a product over an empty set is equal to 1. If $T$ is applied to a sequence $\beta$ with i.i.d. components $\beta_l \sim Beta(1, \varphi(L))$, $l = 1, 2, \ldots$, then, with probability 1
\[
\gamma = T(\beta) \in \Delta_\infty = \left\{ p \in (R^+)\infty : \sum_l p_l = 1 \right\},
\]

thus it defines a random distribution on $\mathbb{N}^+$. Next let us define the probability distribution
\[
Q = \frac{1}{\varphi(L)} \varphi \text{ on } (L, \mathcal{L}) \text{ and take }
\theta = \{\theta_l, l = 1, 2, \ldots\}
\]

with i.i.d. components $\theta_l \sim Q$. It can be shown that
\[
\kappa(\gamma, \theta) = \sum_l \gamma_l \delta_{\theta_l}
\]

is a measurable mapping of $\Delta_\infty \times L^\infty$ and induces a Dirichlet process with parameter measure $\varphi$. We thus write $\kappa(\gamma, \theta) \sim DI(\varphi)$. Notice that from this representation it is immediately seen that the mean probability measure of $\kappa$ is $Q$, in the sense that, for any $A \in \mathcal{L}$ it holds that $E(\kappa(A)) = Q(A)$.

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 \( \varphi_{12} \) and \( \varphi_{23} \) on \( \mathcal{X} \times \mathcal{Y} \) and \( \mathcal{Y} \times \mathcal{Z} \), respectively, i.e. such that

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

Thus the marginal measure \( \varphi_2 \) 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 \( \varphi \) of \( \varphi_{12} \) and \( \varphi_{23} \) on \( \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \), namely

\[
\varphi(dx, dy, dz) = \varphi_{1|2}(dx | y)\varphi_2(dy)\varphi_{3|2}(dz | y)
\]

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

From (11) we see that

(a) the marginals of \( \varphi \) on \( \mathcal{X} \times \mathcal{Y} \) and \( \mathcal{Y} \times \mathcal{Z} \) are \( \varphi_{12} \) and \( \varphi_{23} \), respectively;

(b) setting \( x = \varphi(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}) \), the normalization \( Q = \frac{\varphi}{x} \) 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 \( \varphi \) uniquely.

Next let us consider a pair of laws \( \Psi_{12} \) and \( \Psi_{23} \) 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 \( \Psi \) on \( \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}) \) such that, if the random distribution \( P \) is drawn from \( \Psi \), 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 Markov w.p. 1,
2. \( P_{12} \sim \Psi_{12} \) and \( P_{23} \sim \Psi_{23} \),
3. \( P_{12} \perp P_{23} | P_2 \).

Properties (i)–(iii) for \( \mathcal{X}, \mathcal{Y}, \mathcal{Z} \) finite obviously imply 1–3, respectively.

Property 3 is called the weak hyper-Markov property and means specifically that the \( \sigma \)-algebra \( \sigma \{ P_{12} \} \) generated by \( P(\mathcal{F} \times \mathcal{Z}) \), for all measurable subsets \( \mathcal{F} \) of \( \mathcal{X} \times \mathcal{Y} \) and the \( \sigma \)-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 \( \sigma \)-algebra \( \sigma \{ P_2 \} \) generated by \( P(\mathcal{X} \times H \times \mathcal{Z}) \), for all measurable subsets \( H \) of \( \mathcal{Y} \). The law \( \Psi \) is then called the hyper-Markov combination of \( \Psi_{12} \) and \( \Psi_{23} \). Notice that property 3 replaces the stronger property (iii), given in the discrete case.

**Proposition 3.** If \( \varphi_2 \) is diffuse, then the law \( DI(\varphi) \) is the hyper-Markov combination of \( DI(\varphi_{12}) \) and \( DI(\varphi_{23}) \).

**Proof.** Write \( \varphi = x \times Q \), and use the Sethuraman representation for \( DI(\varphi) \), i.e.

\[
\Gamma = \kappa(\gamma, (U, V, W)) = \sum_{l} \gamma_l \delta(U_l, V_l, W_l) \sim DI(\varphi)
\]

with \((U_l, V_l, W_l), l = 1, 2, \ldots \) i.i.d. from \( Q \), \( \gamma = T(\beta), \beta \) with i.i.d. components \( \beta_l \sim Beta(1, x), l = 1, 2, \ldots \).
By computing the marginals
\[ \Gamma_{12} = \pi_1(\gamma, (U, V)), \quad \Gamma_{23} = \pi_1(\gamma, (V, W)) \]
we check immediately that \( \Gamma_{12} \sim DI(\varphi_{12}) \) and \( \Gamma_{23} \sim DI(\varphi_{23}) \). Since \( \varphi_2 \) has no atoms, the random sequence \( V = \{V_i, l = 1, 2, \ldots \} \) does not contain repetitions w.p. 1: this immediately implies that \( \Gamma \) is Markovian w.p. 1.

As for the hyper-Markov property we observe that
\[ \sigma \{\Gamma_2\} \subset \sigma \{\gamma, V\}, \quad \sigma \{\Gamma_{12}\} \subset \sigma \{\gamma, U, V\}, \quad \sigma \{\Gamma_{23}\} \subset \sigma \{\gamma, V, W\}, \]
so that by the conditional independence of \( U \) and \( W \) given \( \gamma \) and \( V \),
\[ \Gamma_{12} \perp \Gamma_{23} \mid \gamma, V. \]

Then, for any choice of measurable subsets \( \{F_i, i = 1, \ldots, m\} \) of \( \mathcal{X} \times \mathcal{Y} \) and \( \{G_j, j = 1, \ldots, s\} \) 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,
\[
E [h_{12}(\Gamma_{12}(F_1), \ldots, \Gamma_{12}(F_m))h_{23}(\Gamma_{23}(G_1), \ldots, \Gamma_{23}(G_s))] \mid \sigma \{\gamma, V\} = E [h_{12}(\Gamma_{12}(F_1), \ldots, \Gamma_{12}(F_m))] \mid \sigma \{\gamma, V\}] E [h_{23}(\Gamma_{23}(G_1), \ldots, \Gamma_{23}(G_s))] \mid \sigma \{\gamma, V\}.
\]

Thus it remains to prove that
\[
E [h_{12}(\Gamma_{12}(F_1), \ldots, \Gamma_{12}(F_m))] \mid \sigma \{\gamma, V\} = E [h_{12}(\Gamma_{12}(F_1), \ldots, \Gamma_{12}(F_m))] \mid \sigma \{\Gamma_{23}\}], \quad (13)
\]
\[
E [h_{23}(\Gamma_{23}(G_1), \ldots, \Gamma_{23}(G_s))] \mid \sigma \{\gamma, V\} = E [h_{23}(\Gamma_{23}(G_1), \ldots, \Gamma_{23}(G_s))] \mid \sigma \{\Gamma_{23}\}], \quad (14)
\]

If the function \( h_{12} \) is a polynomial, then \( (13) \) becomes
\[
E \left[ \prod_{i=1}^m \left( \sum_i \gamma_i \delta_{(U_i, V_i)}(F_i) \right)^{d_i} \right] \mid \sigma \{\gamma, V\}
\]
\[
= \sum_{i_1} \cdots \sum_{i_d} \prod_{k=1}^d \gamma_{i_k} \delta_{V_{i_k}}(F_{i_k}) E \left[ \prod_{k=1}^d \delta_{U_{i_k}}(F^{(V_{i_k})}_{i_k}) \right] \mid \sigma \{\gamma, V\}, \quad (15)
\]
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
\[
F_{\mathcal{Y}} = \{y : \exists (x, y) \in F\}, \quad F_{\mathcal{X}}^{(y)} = \{x : (x, y) \in F\},
\]
so that
\[
\delta_{(x, y)}(F) = \delta_y(F_{\mathcal{Y}})\delta_x(F_{\mathcal{X}}^{(y)}).
\]
Then, taking into account that \((U, V)\) are independent of \(\gamma\) and that \((U_i, V_i)\) are i.i.d., we can rewrite the conditional expectation appearing in (15) as
\[
E \left[ \delta_{U_{i_1}} \left( F_{i_1, X}^{(V_{i_1})} \right) \ldots \delta_{U_{i_d}} \left( F_{i_d, X}^{(V_{i_d})} \right) \mid \sigma \{ \gamma, V \} \right]
\]
\[
= E \left[ \delta_{U_{i_1}} \left( F_{i_1, X}^{(V_{i_1})} \right) \ldots \delta_{U_{i_d}} \left( F_{i_d, X}^{(V_{i_d})} \right) \mid \sigma \{ V \} \right]
\]
\[
= \psi \left( V_{i_1}, \ldots, V_{i_d} \right),
\]
where \(\psi\) is the bounded non-negative measurable function
\[
\psi \left( v \right) := E \left[ \delta_{U_{i_1}} \left( \bigcap_{i \in A_1(v)} F_{i_1, X}^{(v_i)} \right) V_1 = v_1' \right] \ldots E \left[ \delta_{U_{i_1}} \left( \bigcap_{i \in A_f(v)} F_{i_1, X}^{(v_f)} \right) V_1 = v_f' \right]
\]
with the definition
\[
A_j(v) = \{ i : v_i = v_j' \}, \quad j = 1, \ldots, f.
\]
where \(\{v_1', \ldots, v_f'\} = \{v_1, \ldots, v_d\}\) is the set whose elements are the distinct components of the vector \(v = (v_1, \ldots, v_d)\).

Summarizing (15) can be written as
\[
\sum_{i_1} \ldots \sum_{i_d} \gamma_{i_1} \ldots \gamma_{i_d} \delta_{V_{i_1}} \left( F_{i_1, \gamma} \right) \ldots \delta_{V_{i_d}} \left( F_{i_d, \gamma} \right) \psi \left( V_{i_1}, \ldots, V_{i_d} \right)
\]
\[
= \int_{F_{i_1, \gamma}} \ldots \int_{F_{i_d, \gamma}} \psi \left( y_1, \ldots, y_d \right) \Gamma_2(dy_1) \ldots \Gamma_2(dy_d).
\]
It is readily proved that the above is a \(\sigma \{ \Gamma_2 \}\)-measurable function, hence this ends the proof of (13). Likewise we prove (14) when \(h_{23}\) is a polynomial. Finally observe that expectations of polynomials determine uniquely the joint conditional law of \(\{\Gamma_{12}(F_i), i = 1, \ldots, m\}\) and \(\{\Gamma_{23}(G_j), j = 1, \ldots, s\}\), hence the whole proof is finished.

Following Dawid and Lauritzen we say that a random distribution is a hyper-Dirichlet process with parameter measures \(\phi_{12}\) and \(\phi_{23}\) if it has the same law as \(\Gamma\) defined in (12). We denote this law by \(HDI(\phi_{12}, \phi_{23})\) in order to distinguish it from the finite case, where we used the symbol \(HDI\) instead. When \(\phi_{2}\) is diffuse, the hyper-Dirichlet process \(HDI(\phi_{12}, \phi_{23})\) is equal to the Dirichlet process \(DI(\phi)\), where \(\phi\) is the Markov combination of \(\phi_{12}\) and \(\phi_{23}\). 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 hyper-Dirichlet process with its discrete approximations

In this section we take \(X = \mathbb{R}^d\), \(Y = \mathbb{R}^d\) and \(Z = \mathbb{R}^d\), and we assume that for each of these spaces we have a sequence of finite measurable partitions
\[
\left\{ B_i^{(n)}, i \in I^{(n)} \right\}, \quad \left\{ C_j^{(n)}, j \in J^{(n)} \right\}, \quad \left\{ D_k^{(n)}, k \in K^{(n)} \right\}, \quad n = 1, 2, \ldots,
\]
respectively. Correspondingly, we discretize the consistent parameter measures $\varphi_{12}$ and $\varphi_{23}$ along the corresponding sequences of product partitions in two-dimensional cells
\[
\left\{ B_i^{(n)} \times C_j^{(n)}, i \in I^{(n)}, j \in J^{(n)} \right\}, \left\{ C_j^{(n)} \times D_k^{(n)}, j \in J^{(n)}, k \in K^{(n)} \right\},
\]
getting the tables
\[
\varphi_{12}^{(n)} = \left\{ \varphi_{12}(B_i^{(n)} \times C_j^{(n)}), i \in I^{(n)}, j \in J^{(n)} \right\}, \quad \varphi_{23}^{(n)} = \left\{ \varphi_{23}(C_j^{(n)} \times D_k^{(n)}), j \in J^{(n)}, k \in K^{(n)} \right\}. \tag{18}
\]
For each integer $n$ consider the three-dimensional random array
\[
G^{(n)} \sim HDi(\varphi_{12}^{(n)}, \varphi_{23}^{(n)}) \tag{20}
\]
representing 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
\[
\Gamma^{(n)}(i, j, k) = \left\{ \Gamma(B_i^{(n)} \times C_j^{(n)} \times D_k^{(n)}), i \in I^{(n)}, j \in J^{(n)}, k \in K^{(n)} \right\}. \tag{21}
\]
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 (12) of the hyper-Dirichlet process $\Gamma$ is defined, coupling them in a suitable way.

First of all notice that whenever $\varphi_{23}^{(n)}(j) = 0$ 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 $\varphi_{23}^{(n)}(j) > 0$ for all $j \in J^{(n)}$. Next observe that
\[
\Gamma^{(n)}(i, j, k) = \sum_l^* \gamma_l, \quad i \in I^{(n)}, j \in J^{(n)}, k \in K^{(n)}, \tag{22}
\]
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
\[
\Gamma_{12}^{(n)} \sim Di(\varphi_{12}^{(n)})
\]
we can define
\[
G^{(n)}(i, j, k) = \Gamma_{12}^{(n)}(i, j)G_{3|2}^{(n)}(k|j), \quad i \in I^{(n)}, j \in J^{(n)}, k \in K^{(n)}, \tag{23}
\]
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 $\Gamma_{12}$, 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 $Z$ with parameter measure $\varphi_{23}(C_j^{(n)} \times \cdot)$ defined through the Sethuraman representation
\[
\sum_r H_{j,r}^{(n)} \delta_{W_{j,r}^{(n)}}, \tag{24}
\]
where \( H_j^{(n)} = \{ H_{j,r}^{(n)}, r = 1, 2, \ldots \} = T(\beta_j^{(n)}), \) with \( \beta_j^{(n)} = \{ \beta_{j,r}^{(n)}, r = 1, 2, \ldots \} \) i.i.d. Beta \((1, \varphi_2^{(n)}(j))\), and \( \{ W_{j,r}^{(n)}, r = 1, 2, \ldots \} \) 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 for \( j \in J^{(n)} \)

\[
G_{3/2}^{(n)}(k | j) = \sum_r H_{j,r}^{(n)} 1_{D_k(W_{j,r}^{(n)})}, \quad k \in K^{(n)}.
\]

All the random variables used to produce the vectors \( \{ G_{3/2}^{(n)}(k | j), k \in K^{(n)} \} \), for \( j \in J^{(n)} \), are drawn independently of \( \varphi_2 \), 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 \( V_l \subset C_j \). We choose the maximal coupling (also known as \( \gamma \)-coupling, see e.g. [7]) conditional to \( V_{l(j)} \), so that

\[
P(W_{j,1}^{(n)} \neq W_{l(j)} | V_{l(j)}) = \frac{1}{2} \| \varphi_{3/2}^{(n)}(\cdot | V_{l(j)}) - \varphi_{3/2}^{(n)}(\cdot | j) \|.
\]

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

**Theorem 4.** Under the following assumptions:

(i) the marginal measure \( \varphi_2 \) has no atoms;

(ii) for any compact set \( A \subset \mathcal{U} \)

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

where \( | \cdot | \) is the diameter;

(iii) the function \( y \in \mathcal{U} \to \varphi_{3/2}^{(n)}(\cdot | y) \) 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

\[
|\Gamma(i, j, k) - G(i, j, k)|
= |\Gamma(i, j, k) - \Gamma_{12}(i, j)G_{3/2}^{(n)}(k | j)|
\leq \gamma_{l(j)} 1_{B_l(U_l(j))} |1_{D_k(W_{l(j)})} - 1_{D_k(W_{j,1})}|
+ \gamma_{l(j)} 1_{B_l(U_l(j))} \sum_{r \geq 2} H_{j,r} \left| 1_{D_k(W_{j,1})} - 1_{D_k(W_{j,r})} \right|
+ \sum_{l > l(j)} \gamma_l 1_{B_l(U_l)} 1_{C_j(V_l)} \left| 1_{D_k(W_l)} - \sum_{r \geq 1} H_{j,r} 1_{D_k(W_{j,r})} \right|,
\]

(27)
from which, by summing over \( i, j, k \), recalling that \( \sum_{r \geq 2} H_{j,r} = 1 - H_{j,1} \) and that for any pair of probability distributions \( \mu \) and \( \nu \) over \( K \)

\[
\sum_k |\mu(k) - \nu(k)| = \|\mu - \nu\| \leq 2,
\]

we get

\[
\sum_{i,j,k} |\Gamma(i, j, k) - G(i, j, k)| \leq \sum_j \gamma_l(j) \sum_k \left| \left(1_D_k(W_l(j)) - 1_D_k(W_{j,1})\right) + 2 \sum_j \gamma_l(j)(1 - H_{j,1}) + 2 \sum_{l > l(j)} \gamma_l C_j(V_l) \right| \leq 2 \sum_j \gamma_l(j)(1 - H_{j,1}) + 2 \sum_{l > l(j)} \gamma_l C_j(V_l). \tag{28}
\]

Let us proceed to bound the first term at the r.h.s. of (28). Since

\[
\sum_k \left| \left(1_D_k(W_l(j)) - 1_D_k(W_{j,1})\right) + 2 \sum_j \gamma_l(j)(1 - H_{j,1}) + 2 \sum_{l > l(j)} \gamma_l C_j(V_l) \right| \leq 2 \cdot 1\{W_l(j) \neq W_{j,1}\},
\]

for any compact set \( A \subseteq Y \) we have

\[
\sum_j \gamma_l(j) \sum_k \left| \left(1_D_k(W_l(j)) - 1_D_k(W_{j,1})\right) + 2 \sum_j \gamma_l(j)(1 - H_{j,1}) + 2 \sum_{l > l(j)} \gamma_l C_j(V_l) \right| \leq 2 \sum_{j: C_j \cap A \neq \emptyset} \gamma_l(j) 1\{W_l(j) \neq W_{j,1}\} + 2 \sum_{j: C_j \cap A \neq \emptyset} \gamma_l 1\{W_l(j) \neq W_{j,1}\} + 2 \Gamma_2(A^c),
\]

and by taking the expected value and using (26), by the independence of \( \gamma, l(j) \) and \( V_l(j) \) we have

\[
E \sum_j \gamma_l(j) \sum_k \left| \left(1_D_k(W_l(j)) - 1_D_k(W_{j,1})\right) + 2 \sum_j \gamma_l(j)(1 - H_{j,1}) + 2 \sum_{l > l(j)} \gamma_l C_j(V_l) \right| \leq \sup_{j: C_j \cap A \neq \emptyset} E \sum_j \gamma_l(j) \left(1\{W_l(j) \neq W_{j,1}\} + 2 \gamma_l 1\{W_l(j) \neq W_{j,1}\} + 2 \Gamma_2(A^c) \right).
\]

Now

\[
\sup_{v \in C_j} \left\| \varphi_{3|2}(\cdot|j) - \varphi_{3|2}(\cdot|v) \right\| = \sup_{v \in C_j} \left\| \frac{\int_{C_j} \varphi_{3|2}(\cdot|y) \varphi_2(dy)}{\varphi_2(C_j)} - \varphi_{3|2}(\cdot|v) \right\| = 2 \sup_{v \in C_j} \left\| \frac{\int_{C_j} (\varphi_{3|2}(D|y) - \varphi_{3|2}(D|v)) \varphi_2(dy)}{\varphi_2(C_j)} \right\| \leq \sup_{v \in C_j} \left\| \varphi_{3|2}(\cdot|y) - \varphi_{3|2}(\cdot|v) \right\| \varphi_2(dy) \leq \sup_{v_1, v_2 \in C_j} \left\| \varphi_{3|2}(\cdot|v_1) - \varphi_{3|2}(\cdot|v_2) \right\|.
\]

Finally select the compact set \( A \) in such a way that \( \varphi_2(A^c) \leq \frac{2 \varepsilon}{\alpha} \) and define the larger compact set

\[
A^1 = \{ t: \text{dist}(t, A) \leq 1 \}.
\]
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

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

From the assumption (ii), for $n$ large enough and any $j$ such that $C^{(n)}_j \cap A \neq \emptyset$, it holds $|C^{(n)}_j| < \delta(\varepsilon)$ so that $C^{(n)}_j \subset A^1$, and consequently the r.h.s. of (29) can be bounded by $\varepsilon$.

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 $\varphi_2$ then we would have a sequence $\{C^{(n)}_{j_n}\}$ 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 (28) can now be bounded by

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

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

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

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

so that, by a trivial computation

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

Finally observe that the expected value of the last sum at the r.h.s. of (28) is

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

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

It is possible to formulate the result of the previous theorem as the convergence of $G^{(n)}$ to $\Gamma$ 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 $\mathbb{R}^d$, but 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 $X \times Y \times Z = \mathbb{R}^d$, with the property that

$$\left\{ G^{(n)}(B^{(n)}_i \times C^{(n)}_j \times D^{(n)}_k), i \in I^{(n)}, j \in J^{(n)}, k \in K^{(n)} \right\}$$

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 $E \subset X$ and $F \subset Z$

$$\lim_{n \to \infty} \max \left\{ \left| B^{(n)}_i \right|, i \in I^{(n)} : B^{(n)}_i \cap E \neq \emptyset \right\} = 0, \quad (30)$$

$$\lim_{n \to \infty} \max \left\{ \left| D^{(n)}_k \right|, k \in K^{(n)} : D^{(n)}_k \cap F \neq \emptyset \right\} = 0. \quad (31)$$

Then $\{ G^{(n)} \}$ converges weakly to $\Gamma$, as random probability distributions on $\mathbb{R}^d$, endowed with the topology of weak convergence.

**Proof.** In order to prove that $\{ G^{(n)} \}$ converges weakly to $\Gamma$, 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 $\Phi$ which are bounded and Lipschitz (BL) w.r.t. a metric $\text{dist}$ that induces the topology of weak convergence on $\mathcal{P}(\mathbb{R}^d)$ (see [1]).

In particular (see [8]) we can choose $\text{dist} = \text{dist}_{BL}$, i.e.

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

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

By assumption

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

where $\hat{\lambda}^{(n)}_{ijk}$ is a probability measure supported by $B^{(n)}_i \times C^{(n)}_j \times D^{(n)}_k$, for $i \in I^{(n)}, j \in J^{(n)}, k \in K^{(n)}$. Likewise we define the intermediate random probability distribution on $\mathbb{R}^d$, namely

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

where we recall that

$$\Gamma^{(n)}(i, j, k) = \Gamma(B^{(n)}_i \times C^{(n)}_j \times D^{(n)}_k). \quad (32)$$

Obviously if $\Phi$ is BL with Lipschitz constant $L_{\Phi} < \infty$

$$|E \left[ \Phi \left( G^{(n)} \right) \right] - E \left[ \Phi \left( \Gamma \right) \right]| \leq L_{\Phi} E \left[ \text{dist}_{BL} \left( G^{(n)}, \Gamma \right) \right].$$
whereas
\[ \text{dist}_{BL} \left( G^{(n)}, \Gamma \right) \leq \frac{1}{2} \left\| G^{(n)} - \Gamma^{(n)} \right\| + \text{dist}_{BL} \left( \Gamma^{(n)}, \Gamma \right). \]

By Theorem 4 the variation distance between \( G^{(n)} \) and \( \Gamma^{(n)} \) tends to zero, hence the first term in the above r.h.s. goes to zero in the mean. The same will be proved for the second, provided we show that for any probability measure \( \Gamma \) 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
\[ \left| \Gamma^{(n)}(f) - \Gamma(f) \right| = \left| \sum_{i,j,k} \Gamma^{(n)}(i,j,k) \int_{B_i^{(n)}} \int_{C_j^{(n)}} \int_{D_k^{(n)}} f(x,y,z) \left( \lambda^{(n)}(dx,dy,dz) - \frac{\Gamma(dx,dy,dz)}{\Gamma^{(n)}(i,j,k)} \right) \right|. \]

Let \( \varepsilon > 0 \) and consider compact sets \( A_h \subset \mathbb{R}^{dh} \) for \( h = 1, 2, 3 \) such that \( \Gamma(A^c) < \varepsilon \), 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
\[ \sum_{i,j,k}^* \Gamma^{(n)}(i,j,k) \left| \int_{B_i^{(n)}} \int_{C_j^{(n)}} \int_{D_k^{(n)}} f(x,y,z) \left( \lambda^{(n)}(dx,dy,dz) - \frac{\Gamma(dx,dy,dz)}{\Gamma^{(n)}(i,j,k)} \right) \right| + 2 \|f\|_\infty \varepsilon \]
\[ \leq \sum_{i,j,k}^* \Gamma^{(n)}(i,j,k) \left( \int_{B_i^{(n)}} \int_{C_j^{(n)}} \int_{D_k^{(n)}} f(x,y,z) - f(\overline{x}_i, \overline{y}_j, \overline{z}_k) \left| \lambda^{(n)}(dx,dy,dz) \right| \right) \]
\[ + \int_{B_i^{(n)}} \int_{C_j^{(n)}} \int_{D_k^{(n)}} f(x,y,z) - f(\overline{x}_i, \overline{y}_j, \overline{z}_k) \left| \frac{\Gamma(dx,dy,dz)}{\Gamma^{(n)}(i,j,k)} \right| + 2\varepsilon, \quad (33) \]
for any choice of \( (\overline{x}_i, \overline{y}_j, \overline{z}_k) \in B_i^{(n)} \times C_j^{(n)} \times D_k^{(n)}. \)

Observe that from the assumption (ii) of the previous Theorem 4, (30) and (31), for \( n \) large enough,
\[ \left| B_i^{(n)} \times C_j^{(n)} \times D_k^{(n)} \right| < \varepsilon, \]
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
\[ \left| f(x,y,z) - f(\overline{x}_i, \overline{y}_j, \overline{z}_k) \right| \leq L(f) \text{dist} \left( (x, y, z), (\overline{x}_i, \overline{y}_j, \overline{z}_k) \right) \leq \varepsilon, \]
which implies that (33) does not exceed \( 4\varepsilon \). Since \( \varepsilon \) is arbitrary, the proof is finished. \( \square \)
Remark 6. Assumption (iii) of Theorem 4 holds true when \( g_{23} \) 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 \( \mu_3 \)-integrable function \( h_y(z) \), such that for any \( y' \) in a neighborhood of \( y \)

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

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 \( \varphi_2 \) w.r.t. \( \mu_2 \) is continuous in \( y \), and by Bayes formula

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

for \( y \) in the open set of full \( \varphi_2 \)-measure where \( g_2(y) > 0 \).

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

4. General hyper-Dirichlet processes for the butterfly model

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 \( \{\rho^{(l)} \}, l = 0, 1, 2, \ldots \) on the measurable space \((E, \mathcal{E})\), and suppose that \( \rho(E) = \sum_{i=0}^{\infty} \rho^{(i)}(E) < +\infty \), in which case \( \rho \) is a finite measure as well. Let \( R = \{R_l, l = 0, 1, 2, \ldots \} \) be a sequence of independent random variables, with

\[
R_l \sim \text{Be} \left( \rho^{(l)}(E), \sum_{i=l+1}^{\infty} \rho^{(i)}(E) \right), \quad l = 0, 1, 2, \ldots
\]  

(34)

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

\[
S_l = R_l \prod_{i=0}^{l-1} (1 - R_i), \quad l = 0, 1, 2, \ldots
\]

Then

(i) \( S \) is a Dirichlet process on \( \mathbb{N} \), with parameter measure \( \sum_{i=0}^{\infty} \rho^{(i)}(E) \delta_i \).

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

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

\[
\sum_{l=0}^{\infty} S_l \Gamma_l
\]

(35)

is a Dirichlet process on \( E \) with parameter measure \( \rho \).
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.

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

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

as \( m \to \infty \). Thus (35) 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 to check that

\[
\left( S_0, \ldots, S_m, 1 - \sum_{l=0}^{m} S_l \right) \sim Di \left( \rho^{(0)}(E), \ldots, \rho^{(m)}(E), \sum_{l=m+1}^{\infty} \rho^{(l)}(E) \right),
\]

for any integer \( m \).

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

\[
\sum_{l=0}^{m} S_l \Gamma_l + \left( 1 - \sum_{l=0}^{m} S_l \right) \delta_{x_0}.
\]

Next consider any finite measurable partition \( \{ A_h, h \in H \} \) of \( E \), and observe that the joint law of

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

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

For this purpose define the following two-dimensional table

\[
D_{12}(h, l) = S_l \Gamma_l(A_h), \quad h \in H, \quad l = 0, \ldots, m,
\]

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

and observe that the marginal \( D_2(\cdot) \) is (36) and the conditionals \( D_{1|2}(\cdot|l) \) are \( \{ \Gamma_l(A_h), h \in H \} \) for \( l = 0, \ldots, m, \) and \( \{ \delta_{x_0}(A_h), h \in H \} \) for \( l = m+1 \), respectively. Therefore by Proposition 1

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

where

\[
d_{12}(h, l) = \rho^{(l)}(A_h), \quad h \in H, \quad l = 0, \ldots, m,
\]

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

Since the marginal \( D_1 \) yields (37) it is enough to apply again Proposition 1 to conclude.
Finally, if we let $m \to \infty$ the random vector (37) converges a.s. to $\{ \sum_{i=0}^{\infty} s_i \Gamma_i(A_h), h \in H \}$, hence it converges in law as well. By the continuity of the Dirichlet laws w.r.t. their parameter vectors, the limit law is $\mathcal{D}l \{ \sum_{i=0}^{\infty} \rho^{(i)}(A_h) \}$. \hfill \Box

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

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

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

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

Then the Dirichlet process with parameter measure $\rho$ can be represented as

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

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

$$
R_0 \sim \text{Be}(\varphi(\mathcal{Y}), \psi(\mathcal{Y})), \quad R_l \sim \text{Be}(\psi_2(\{y_l\}), \sum_{i=l+1}^{\infty} \psi_2(\{y_l\})), \quad l = 1, 2, \ldots,
$$

$\Gamma$ is a Dirichlet process on $\mathcal{X} \times \mathcal{Y}$ with parameter measure $\varphi$, and $\Xi_l$ is a Dirichlet process on $\mathcal{X}$ with parameter measure $\psi^{(l)}(dx) := \psi(dx \times \{y_l\})$, for $l = 1, 2, \ldots$, 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 $\rho = \sum_{i=0}^{\infty} \rho^{(i)}$, where

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

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

Notice that $\rho = \varphi + \psi$ is the unique decomposition where the marginal $\varphi_2$ is diffuse and the marginal $\psi_2$ is purely atomic. Therefore, given two consistent parameter measures $\rho_{12}$ and $\rho_{23}$ on $\mathcal{X} \times \mathcal{Y}$ and $\mathcal{Y} \times \mathcal{Z}$, respectively, they can be uniquely decomposed as

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

where $\varphi_{12}$ and $\varphi_{23}$ are consistent, with common diffuse marginal $\varphi_2$, whereas $\psi_{12}$ and $\psi_{23}$ are consistent, with common purely atomic marginal $\psi_2$.

Assume from now on that $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are Borel spaces. The previous corollary will allow us to construct a general hyper-Dirichlet process, i.e. a process whose law is the hyper-Markov combination of two hyperconsistent laws $\mathcal{D}l(\rho_{12})$ and $\mathcal{D}l(\rho_{23})$ on $\mathcal{X} \times \mathcal{Y}$ and $\mathcal{Y} \times \mathcal{Z}$, respectively.

Let us define

$$
\rho_{2}^{(0)}(dy) = \varphi_2(dy), \quad \rho_{2}^{(l)}(dy) = \psi_2(\{y_l\}) \delta_{y_l}(dy), \quad l = 1, 2, \ldots
$$
so that \( \rho_2 = \varphi_2 + \psi_2 = \sum_{l=0}^{\infty} \rho_2^{(l)} \). Now let \( R \) be a sequence of independent random variables distributed as in (38), and let \( S = T(R) \), so that

\[
S \sim DI(\rho_2^{(l)}(Y), \ l = 0, 1, 2, \ldots).
\]  

Moreover let \( \Gamma \sim HDI(\phi_{12}, \phi_{23}) = DI(\phi) \), and \( \Xi_l^x \) and \( \Xi_l^z \) be Dirichlet processes on \( X \) and \( Z \), 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

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

The measures \( \varphi_2 \) and \( \psi_2 \) being singular, the Markov combination \( \rho \) of \( \rho_{12} \) and \( \rho_{23} \), as defined in (10), is the sum of the Markov combination \( \varphi \) of \( \rho_{12} \) and of \( \rho_{23} \) and the Markov combination \( \psi \) of \( \psi_{12} \) and \( \psi_{23} \), namely

\[
\rho(dx, dy, dz) = \varphi(dx, dy, dz) + \sum_l \frac{1}{\psi_2(\{y_l\})} \psi_{12}^{(l)}(dx) \psi_{23}^{(l)}(dz).
\]

Then define

\[
\Sigma(dx, dy, dz) = \sum_l S_l \Xi_l^x (dx) \Xi_l^y (dz) \delta_{y_l}(dy).
\]  

The following generalization of Theorem 2 holds true.

**Proposition 9.** The law of \( \Sigma \) defined in (40) is the hyper-Markov combination of \( DI(\rho_{12}) \) and \( DI(\rho_{23}) \). The mean probability measure of \( \Sigma \) is the normalized Markov combination of \( \rho_{12} \) and \( \rho_{23} \).

**Proof.** Using the fact that the mean probability measure of a Dirichlet process is its normalized parameter measure, it is immediately computed that

\[
E(\Sigma) = E(S_0) \frac{1}{\varphi(X \times Y \times Z)} \varphi(dx, dy, dz) + \sum_l E(S_l) \delta_{y_l}(dy) \frac{\psi_{12}^{(l)}(dx) \psi_{23}^{(l)}(dz)}{\psi_{12}^{(l)}(X) \psi_{23}^{(l)}(Z)}
\]

\[
= \frac{1}{\rho_2(Y)} \varphi(dx, dy, dz) + \sum_l \delta_{y_l}(dy) \frac{\psi_{12}^{(l)}(dx) \psi_{23}^{(l)}(dz)}{\psi_2^{(l)}(y_l)},
\]

which proves the last statement.

Next, by using the previous Corollary the laws of the marginals \( \Sigma_{12} \) and \( \Sigma_{23} \) are immediately identified to be \( DI(\rho_{12}) \) and \( DI(\rho_{23}) \). Next, independently of \( S \), use independent Sethuraman representations (12)

\[
\Gamma = \kappa(\gamma, (U, V, W)),
\]

\[
\Xi_l^x = \kappa(\gamma_l^x, U_l^x), \quad \Xi_l^z = \kappa(\gamma_l^z, W_l^z), \quad l = 1, 2, \ldots,
\]

where \( \gamma, U, V, W \) are as in Proposition 3, \( \gamma_l^x = (\gamma_l^x, m = 1, 2, \ldots) \) and \( \gamma_l^z = (\gamma_l^z, m = 1, 2, \ldots) \) are independent sequences obtained as the image under \( T \) of i.i.d. sequences

\[
\beta_l^x = (\beta_{l,m}^x \sim Beta(1, \psi_2(\{y_l\})), \ m = 1, 2, \ldots),
\]

\[
\beta_l^z = (\beta_{l,m}^z \sim Beta(1, \psi_2(\{y_l\})), \ m = 1, 2, \ldots),
\]
respectively, and
\[ U_l^x = (U_{l,m}^x \sim \psi_{12} (\cdot \times \{y_l\}) / \psi_2 (\{y_l\}), \ m = 1, 2, \ldots) \]
\[ W_l^x = (U_{l,m}^x \sim \psi_{23} (\{y_l\} \times \cdot) / \psi_2 (\{y_l\}), \ m = 1, 2, \ldots) \]

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

With probability 1 the sequence \( V \) contains neither repetitions nor any of the \( y_l \), for \( l = 1, 2, \ldots \). This implies that the realizations of \( \Sigma \) are Markov with probability 1.

As for the hyper-Markov property observe that
\[ \Sigma_{12} \perp \Sigma_{23} | \Sigma, \gamma, V. \]

The proof is finished as in Proposition 3, the only change being that
\[
E \left[ \prod_{l=1}^{m} (\Sigma_{12}(F_l))^d \bigg| \sigma \{ \Sigma, \gamma, V \} \right] \\
= \int_{F_{1,Y}} \cdots \int_{F_{d,Y}} \tilde{\psi} (y_1, \ldots, y_d) \Sigma_2 (dy_1) \cdots \Sigma_2 (dy_d), \tag{41}
\]

where \( d = d_1 + \cdots + d_m \), and where \( \tilde{\psi} \) is the bounded measurable function
\[ \tilde{\psi} (v) = E \left[ \delta_{U_l} \left( \bigcap_{i \in A_1(v)} F_{l_i, \chi_i}^{(v_i)} \right) | V_1 = v'_1 \right] \cdots E \left[ \delta_{U_f} \left( \bigcap_{i \in A_f(v)} F_{l_i, \chi_i}^{(v'_i)} \right) | V_1 = v'_f \right] \]

\[ \cdots E \left[ \prod_{i \in B_1(v)} \Xi_i^{(y_i)} \left( F_{l_i, \chi_i}^{(y_i)} \right) \right] \cdots E \left[ \prod_{i \in B_e(v)} \Xi_i^{(y_i)} \left( F_{l_i, \chi_i}^{(y_i)} \right) \right] \]

with \( v = (v_1, \ldots, v_d) \), and
\[ \{v_1, \ldots, v_d\} = \{v'_1, \ldots, v'_f, y_j, \ldots, y_k\}. \]

\[ A_1(v), \ldots, A_f(v) \] being defined as in (16), and
\[ B_1(v) = \{i : v_i = y_j\}, \ldots, B_e(v) = \{i : v_i = y_k\}. \]

Extending the definition previously given we call a process with the law of \( \Sigma \) defined in (40) a hyper-Dirichlet process. This law is again denoted by \( \text{HDI}(\rho_{12}, \rho_{23}) \). 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 \( \Sigma \) defined in (40) is a Dirichlet process if and only if for each \( l = 1, 2, \ldots \) either \( \psi_{12}^{(l)} \) or \( \psi_{23}^{(l)} \) is a single Dirac mass. If this happens the parameter measure of \( \Sigma \) is the Markov combination of \( \rho_{12} \) and \( \rho_{23} \).
Proof. If \( \psi^{(l)}_{12} \) or \( \psi^{(l)}_{23} \) is a single Dirac mass it is clear that \( \Xi^f_l(dx)\Xi^f_l(dz)\delta_{y_l}(dy) \) is a Dirichlet process with parameter measure having total mass \( \psi_2(\{y_l\}) \). On the other hand \( \Gamma \) 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, if \( \Gamma \) is a Dirichlet process, by (3) of Proposition 1, for \( l = 1, 2, \ldots \)

\[
\frac{\Sigma(dx \times \{y_l\} \times dz)}{\Sigma(\mathcal{X} \times \{y_l\} \times \mathcal{Z})} = \Xi^f_l(dx)\Xi^f_l(dz)
\]

is a Dirichlet process. 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, necessity is established. \( \square \)

Next our aim is to establish the form of the law of a hyper-Dirichlet process \( \Sigma \) conditional to a random sample \( (X_i, Y_i, Z_i), i = 1, \ldots, n \), drawn from \( \Sigma \) itself. This information is not supplied by the general conjugacy result proved in Dawid and Lauritzen [3, Corollary 5.5], 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 \( \Sigma \) is constructed. This is achieved by using additional label variables.

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

\[
H = \sum_{l=0}^{+\infty} S_l \Gamma_l,
\]

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

Conditionally to \( (S, \Gamma) \), let \( N \) be an integer random variable with law \( (S_l, l \geq 0) \), and let \( \zeta \) have law \( \Gamma_N \), conditionally to \( (S, \Gamma) \) and \( N \). In this way, conditionally to \( \sigma(\mathcal{H}) \), \( \zeta \) has law \( \mathcal{H} \) (see [9, Lemma 4.1]), thus we say that \( \zeta \) 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

\[
\mathcal{L}(S, \Gamma|\zeta) = \mathcal{L}(S|N = n) \otimes \mathcal{L}(\Gamma_l) \otimes \mathcal{L}(\Gamma_n|\zeta), \quad \text{if } \zeta \in A_n.
\]

Proof. The condition on the partition implies that the representation (42) is unique so we may identify \( (S, \Gamma) \) with \( \mathcal{H} \), and moreover

\[
N = \sum_{l=0}^{+\infty} l 1_{A_l}(\zeta).
\]

Hence

\[
\mathcal{L}(S, \Gamma|\zeta) = \mathcal{L}(S, \Gamma|N = n, \zeta), \quad \text{if } \zeta \in A_n.
\]

Since by assumption

\[
P(S \in ds, N = n, \Gamma \in dg, \zeta \in dx) = v(ds)s_n \prod_l \eta_l(dg_l)g_n(dx),
\]

we have

\[
P(S \in ds, N = n, \Gamma \in dg, \zeta \in dx) = v(ds)s_n \prod_l \eta_l(dg_l)g_n(dx),
\]

for \( \zeta \in A_n \).
it is
\[ P(S \in ds, \Gamma \in dg, \xi \in dx | N = n) = \frac{v(ds)s_n}{\int v(dq)q_n} \prod_{l \neq n} \eta_l(dg_l)\eta_n(dg_n)g_n(dx), \]
which ensures that \( S, \{\Gamma_l, l \neq n\} \) and \( \{\Gamma_n, \xi\} \) are conditionally independent given \( N = n \), which implies (43). \( \square \)

This observation is the key to establish the following result.

**Proposition 12.** Consider a hyper-Dirichlet process \( \Sigma \sim HDI(\rho_{12}, \rho_{23}) \). Assume that, conditionally on \( \Sigma \), the vector of observations
\[ ((X_i, Y_i, Z_i), \ i = 1, \ldots, m) \]
has i.i.d. components, each with distribution \( \Sigma \). Then, conditionally on the observations, the law of \( \Sigma \) is \( HDI(\rho_{12} + \sum_{i=1}^{m} \delta_{(X_i, Y_i)}, \rho_{23} + \sum_{i=1}^{m} \delta_{(Y_i, Z_i)}) \). Moreover the mean measure of \( \Sigma \) conditional on the observations (i.e. the predictive measure) is
\[ \frac{1}{\rho_2(\mathcal{Y})} \left( \rho(dx, dy, dz) + \sum_{i=1}^{m} \delta_{(X_i, Y_i, Z_i)} \right). \]

**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. [9, p. 648]). The hyper-Dirichlet process \( \Sigma \) defined in (40) has the form (42) with \( \Gamma_0(dx, dy, dz) = \Gamma(dx, dy, dz) = \Xi(dx)\Xi^T(dy)\delta(dy), \ l = 1, 2, \ldots, \) and in our case \( \xi = (X, Y, Z) \). Moreover \( A_0 = \{(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} : \ y \neq y_i, l = 1, 2, \ldots\} \) and \( A_1 = \mathcal{X} \times \{y_i\} \times \mathcal{Z}, l = 1, 2, \ldots. \)

Since \( S \sim DI(\rho_2^{(1)}(\mathcal{Y})), \ l \in \mathbb{N} \) and \( N \) is drawn from \( S \), the standard result about conjugacy of Dirichlet processes [9, Lemma 4.1] yields that \( S \) is \( DI(\rho_2^{(1)}(\mathcal{Y}) + \delta_n(l), \ l \in \mathbb{N}, \) conditional to \( N = n \). Next, for the same reason, the law of \( \Gamma \) conditional to \( (X, Y, Z) \) (for \( Y \notin \{y_i, l = 1, 2, \ldots\} \)) is \( DI(\rho + \delta_{(X,Y,Z)}) \), where \( \rho \) is the Markov combination of \( \rho_{12} \) and \( \rho_{23} \). 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, using Proposition 10
\[ DI(\rho + \delta_{(X,Y,Z)}) = HDI(\rho_{12} + \delta_{(X,Y)}, \rho_{23} + \delta_{(Y,Z)}). \]

Finally, when \( Y = y_i \), the law of the independent Dirichlet processes \( \Xi^{(l)}_1 \) and \( \Xi^{(l)}_2 \) given samples \( X \) and \( Z \) extracted independently from each of them is the product law of two Dirichlet processes \( \psi^{(l)}_2 + \delta_X \) and \( \psi^{(l)}_3 + \delta_Z \). Putting together all these results in (43) and using Proposition 9 with \( \rho_{12} + \delta_{(X,Y)} \) and \( \rho_{23} + \delta_{(X,Y,Z)} \) replacing \( \rho_{12} \) and \( \rho_{23} \) the statements of the proposition for \( m = 1 \) is immediately obtained. \( \square \)

**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 \( \rho_2 \) diffuse, the probability of observing \( (Y_i, i = 1, \ldots, m) \) with all components distinct is 1. By consequence for any \( l \) the
corresponding components of the posterior parameter measures $\psi_{12}^{(l)}$ and $\psi_{23}^{(l)}$ are Dirac measures, hence, by Proposition 10, the posterior law is not only the a hyper-Dirichlet process but actually a Dirichlet process, as it results directly from the conjugacy property of these latter class of processes.

We now finally pass to a limit result similar to Theorem 4 for general hyper-Dirichlet processes. From now on we take again $X$, $Y$ and $Z$ to be Euclidean spaces with respective dimensions $d_1$, $d_2$ and $d_3$, and assume that for each of these spaces we have a sequence of finite partitions as in (17). Correspondingly, we discretize along the corresponding sequences of partitions in two-dimensional cells (18) 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 (19), and likewise

$$\psi_{12}^{(n)} = \left\{ \psi_{12}(B_i^{(n)} \times C_j^{(n)}), \ i \in I^{(n)}, \ j \in J^{(n)} \right\},$$

$$\psi_{23}^{(n)} = \left\{ \psi_{23}(C_j^{(n)} \times D_k^{(n)}), \ j \in J^{(n)}, \ k \in K^{(n)} \right\}.$$

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

$$\left\{ F^{(n)}(i, j, k), \ i \in I^{(n)}, \ j \in J^{(n)}, \ k \in K^{(n)} \right\}, \quad (46)$$

which will be compared to

$$\left\{ \Sigma^{(n)}(i, j, k) = \Sigma(B_i^{(n)} \times C_j^{(n)} \times D_k^{(n)}), \ i \in I^{(n)}, \ j \in J^{(n)}, \ k \in K^{(n)} \right\}, \quad (47)$$

where $\Sigma \sim \text{HDI}(\varrho_{12}, \varrho_{23})$ is defined in (40). However, for our purposes it is more convenient to use a different representation for the process, which is constructed next.

Let us proceed to construct both (46) and (47) on the same probability space. We start by defining $\Gamma^{(n)}$ and $G_{3/2}^{(n)}$ as in Theorem 4, corresponding to the diffuse component $\varrho_2$ of $\rho_2$. All the other random variables to be defined will be taken independent of these arrays. Let $S_0 \sim \text{Be}(\varrho_2(Y), \psi_2(Y))$ be the weight of the diffuse part in (40). Then we split the atomic part as a mixture of distributions, each concentrated on an element of the partition $\left\{ X \times C_j^{(n)} \times Z, \ j \in J^{(n)} \right\}$. For this we need first to draw, independently of $S_0$, the vector of total masses

$$M^{(n)} = (M_1^{(n)}, \ldots, M_{|J^{(n)}|}^{(n)}) \sim \text{Di}(\varrho_2(C_j^{(n)}), \ j \in J^{(n)}).$$

Next we have to split the mass inside each set of the partition. So for each $j \in J^{(n)}$ such that $\psi_2(C_j^{(n)}) > 0$, define independently the following random variables. First a sequence of independent random variables $L_j^{(n)} = \left\{ L_{j,m}^{(n)}, \ m = 1, 2, \ldots \right\}$ with

$$L_{j,m}^{(n)} \sim \text{Be} \left( \psi_2(y_{j,m}^{(n)}), \sum_{i=m+1}^{\infty} \psi_2(y_{j,i}^{(n)}) \right), \quad m = 1, 2, \ldots,$$
where

\[ \{y_{j,m}^{(n)}, \ m = 1, 2, \ldots\} = \{y_l, \ l = 1, 2, \ldots\} \cap C_j^{(n)} \]

with the points ordered in an arbitrary way. The sequence \( L_j^{(n)} \) is used to obtain the corresponding sequence of weights \( U_j^{(n)} = T(L_j^{(n)}) \), i.e.

\[ U_{j,m} = \prod_{i=1}^{m-1} (1 - L_{j,i}^{(n)}) L_{j,m}^{(n)}, \ m = 1, 2, \ldots, \]

which is a Dirichlet process on \( \mathbb{N}^+ \) with parameter measure \( \sum_m \psi_j(y_{j,m}^{(n)}) \delta_m \), as it results by the application of Proposition 7. Notice that in case the sequence \( \{y_{j,m}^{(n)}, \ m = 1, 2, \ldots\} \) is finite, \( L_j^{(n)} \) and \( U_j^{(n)} \) are both taken to be eventually zero. Observe that, by Proposition 1, the joint law of

\[ \{M_j^{(n)}U_{j,m}^{(n)}, \ m = 1, 2, \ldots, j \in J^{(n)}\} \]

is the same as that of

\[ \left\{ \frac{S_{j,m}^{(n)}}{1 - S_0}, \ m = 1, 2, \ldots, j \in J^{(n)} \right\}, \]

where \( S_{j,m} = S_l \), where \( y_l = y_{j,m}^{(n)} \). Moreover both these sequences are independent of \( S_0 \).

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

\[
\Sigma^{(n)}(i, j, k) = S_0 \Gamma^{(n)}(i, j, k) + (1 - S_0) M_j^{(n)} \sum_m U_{j,m}^{(n)} \Xi_j^{x,(n)}(B_i^{(n)}) \Xi_j^{z,(n)}(D_k^{(n)}). \tag{48}
\]

As far as the approximation is concerned, we also need independent copies of the sequences \( L_j^{(n)} = \{L_{j,m}^{(n)}, m = 1, 2, \ldots\} \), say \( L_j^{(n)} = \{L_{j,m}^{(n)}, m = 1, 2, \ldots\} \), with the corresponding sequence of weights \( U_j^{(n)} = \{U_{j,m}^{(n)}, m = 1, 2, \ldots\} = T(L_j^{(n)}) \), and an independent random vector with independent components \( O_j^{(n)} \sim \text{Be}(\psi_2(C_j^{(n)}), \psi_2(C_j^{(n)})) \), \( j \in J^{(n)} \). Then we define

\[
F^{(n)}(i, j, k) = \left\{ S_0 \Gamma_{12}^{(n)}(i, j) + (1 - S_0) M_j^{(n)} \sum_m U_{j,m}^{(n)} \Xi_j^{x,(n)}(B_i) \right\} \times \left\{ O_j^{(n)} C_{3|2}^{(n)}(k|j) + (1 - O_j^{(n)}) \sum_p U_{j,p}^{(n)} \Xi_j^{z,(n)}(D_k) \right\}. \tag{49}
\]
By using Propositions 1 and 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 (49) from (48), and we bound from above its absolute value by

$$
S_0|\Gamma^{(n)}(i, j, k) - \Gamma^{(n)}(i, j)O_j^{(n)}G_{2|3}^{(n)}(k|j)|
+ (1 - S_0)M^{(n)}_j \sum_m U^{(n)}_{j,m} \Xi^{(n)}_{j,m}(B_i)
\times \left| \Xi^{(n)}_{j,m}(D_k) - (1 - O_j^{(n)}) \sum_p U^{(n)}_{j,p} \Xi^{(n)}_{j,p}(D_k) \right|
+ S_0\Gamma^{(n)}(i, j)(1 - O_j^{(n)}) \sum_p U^{(n)}_{j,p} \Xi^{(n)}_{j,p}(D_k)
+ (1 - S_0)M^{(n)}_j O_j^{(n)}G_{2|3}^{(n)}(k|j) \sum_m U^{(n)}_{j,m} \Xi^{(n)}_{j,m}(B_i).
$$

We then proceed to bound each of the four terms. For the first term (50) notice that

$$
S_0|\Gamma^{(n)}(i, j, k) - \Gamma^{(n)}(i, j)O_j^{(n)}G_{2|3}^{(n)}(k|j)|
\leq S_0|\Gamma^{(n)}(i, j, k) - \Gamma^{(n)}(i, j)G_j^{(n)}(k|j)| + S_0\Gamma^{(n)}(i, j)(1 - O_j^{(n)})G_{2|3}^{(n)}(k|j).
$$

Since $\Gamma^{(n)}(i, j)G_j^{(n)}(k|j) = G_j^{(n)}(i, j, k)$, 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

$$
E \left[ S_0 \sum_{i,j,k} E \left[ \Gamma^{(n)}(i, j, k) \right] E \left[ 1 - O_j^{(n)} \right] E \left[ G_{2|3}^{(n)}(k|j) \right] \right]
= \frac{\varphi_2(\mathcal{Y})}{\varphi_2(\mathcal{Y}) + \psi_2(\mathcal{Y})} \sum_{i,j} \frac{\varphi_{12}(B_i^{(n)} \times C_j^{(n)})}{\varphi_{12}(\mathcal{X} \times \mathcal{Y})} \frac{\psi_2(C_j^{(n)})}{\varphi_2(C_j^{(n)}) + \psi_2(C_j^{(n)})}
= \frac{1}{\varphi_2(\mathcal{Y}) + \psi_2(\mathcal{Y})} \sum_j \frac{\varphi_2(C_j^{(n)})\psi_2(C_j^{(n)})}{\varphi_2(C_j^{(n)}) + \psi_2(C_j^{(n)})} =: Z_n.
$$

In order to deal with the second term (51) observe that

$$
\left| \Xi^{(n)}_{j,m}(D_k) - (1 - O_j^{(n)}) \sum_p U^{(n)}_{j,p} \Xi^{(n)}_{j,p}(D_k) \right|
\leq (1 - O_j^{(n)}) \sum_p U^{(n)}_{j,p} \left| \Xi^{(n)}_{j,m}(D_k) - \Xi^{(n)}_{j,p}(D_k) \right| + O_j^{(n)} \Xi^{(n)}_{j,m}(D_k).
$$
Then, inserting the above estimate in (51), summing both sides over \( i, j, k \), and taking the expected value we get the upper bound

\[
E(1 - S_0) \times \left\{ \sum_j EM_j^{(n)} E(1 - O_j^{(n)}) \sum_m \sum_{p \neq m} EU_{j,m}^{(n)} EU_{j,p}^{(n)} \sum_k E \left| \Xi_{j,m}^{(n)}(D_k) - \Xi_{j,p}^{(n)}(D_k) \right| \right. \\
+ \left. \sum_j EM_j^{(n)} EO_j^{(n)} \right\}.
\]

(54)

The last term in (54) is the easiest to deal with, since it is equal to

\[
E(1 - S_0) \sum_j EM_j^{(n)} EO_j^{(n)} = \chi_n,
\]

whereas the first term in (54) is bounded by

\[
2E(1 - S_0) \sum_j EM_j^{(n)} E(1 - O_j^{(n)}) \sum_m \sum_{p \neq m} EU_{j,m}^{(n)} EU_{j,p}^{(n)} \\
= 2 \frac{\psi_2(\mathcal{Y})}{\phi_2(\mathcal{Y}) + \psi_2(\mathcal{Y})} \sum_j \frac{\psi_2(C_j^{(n)})^2}{\phi_2(C_j^{(n)}) + \psi_2(C_j^{(n)})} \sum_m \sum_{p \neq m} \frac{\psi_2(\{y_{j,m}^{(n)}\})\psi_2(\{y_{j,p}^{(n)}\})}{\phi_2(C_j^{(n)})^2} \\
\leq 4 \frac{\sum_j \sum_{p > 1} \psi_2(\{y_{j,p}^{(n)}\})}{\phi_2(\mathcal{Y}) + \psi_2(\mathcal{Y})} \\
\leq 4 \frac{\sum_{p > 1} \psi_2(\{y_{j,p}^{(n)}\})}{\phi_2(\mathcal{Y}) + \psi_2(\mathcal{Y})} =: \tau_n,
\]

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

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

Likewise, the expected value of the sum of the third terms (52) yields

\[
ES_0 \sum_j E \left( \sum_i \Gamma_i^{(n)}(i, j) \right) E(1 - O_j^{(n)}) \\
= \frac{\phi_2(\mathcal{Y})}{\phi_2(\mathcal{Y}) + \psi_2(\mathcal{Y})} \sum_j \frac{\phi_2(C_j^{(n)})}{\phi_2(C_j^{(n)}) + \psi_2(C_j^{(n)})} = \chi_n
\]

and for the last terms (53) 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 $\psi_2$ is a finite measure, for any $\varepsilon > 0$ there exists an integer $L_\varepsilon$ such that

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

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

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

and analogously, for $n$ large enough,

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

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. $\square$

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

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

has the same law as

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

Furthermore suppose that (30) and (31) hold for all compact sets $E \subset \mathcal{X}$ and $F \subset \mathcal{Z}$. Then $\{F^{(n)}\}$ converges weakly to $\Sigma$, as random probability distributions on $\mathbb{R}^d$, endowed with the topology of weak convergence.

## 5. Extensions to general decomposable graphs

In this section, we discuss briefly the extension of the above results to a general decomposable graph. Both the construction of the Markov combination of a family of consistent clique marginal distributions and the hyper-Markov combination of a family of hyperconsistent laws for random clique marginal distributions can be performed recursively [3]. In order to set this recursion, the existence of a perfect numbering of the cliques $(C_1, \ldots, C_k)$ of the decomposable graph $G$ is used (see e.g. [6]). This means that being

$$H_j = C_1 \cup \cdots \cup C_j, \quad j = 1, \ldots, k,$$

for any $j = 2, \ldots, k$ the separator $S_j = H_{j-1} \cap C_j$ is entirely contained in some $C_i$, for some $i < j$. Then $H_{j-1}$ and $C_j$ form a non-trivial decomposition of the subgraph induced by $H_j$, for
any \( j = 2, \ldots, k \), which means that we are allowed to identify (recursively in \( j = 2, \ldots, k \)) the three disjoint subsets \( H_{j-1}, S_j \) and \( R_j = C_j \setminus H_{j-1} \) with the three vertices 1, 2 and 3 of the canonical butterfly model, in order to construct Markov and hyper-Markov combinations.

The construction of an hyper-Dirichlet process with parameter measures \((\phi_{C_1}, \ldots, \phi_{C_k})\) (i.e. the hyper-Markov combination of Dirichlet processes with parameter measure \(\phi_{C_j}\) for each clique marginal \(C_j, j = 1, \ldots, k\)) is relatively easy provided we assume that each of the marginals \(\phi_{S_j}\) is diffuse. The extension of Proposition 3, namely that the hyper-Dirichlet process is a Dirichlet process having as a parameter measure the hyper-Markov combination of the \(\phi_{C_j}\)’s, holds true. Likewise results analogous to the limit Theorem 4 and Theorem 5 can be proved, by applying recursively the coupling construction needed to prove the convergence of the discretized approximations.

As far as the results of Section 4 are concerned, the construction and the approximation of the hyper-Dirichlet process still hold if for \( j = 1, \ldots, k \) we add to the “diffuse” parameter measure \(\phi_{C_j}\) a sum of Dirac measures

\[
\sum_{i=1}^{m} \delta_{x^{(i)}},
\]

for given vectors \(x^{(i)}, i = 1, \ldots, m\) in the sample space. The representation of their Markov combination as the sum of the Markov combination of the diffuse components and the sum of purely Dirac components still allows to use a simple mixture representation with only two components as before. This is also useful in order to extend Theorems 14 and 15. In this way we cover the a posteriori law of a hyper-Dirichlet process with prior “diffuse” parameter measures (which is still a Dirichlet process with probability 1), which is the most interesting situation for statistical applications.

But for general consistent parameter measures \(\rho_{C_j}\) some complications arise. This is already seen in the construction of their Markov combination. In fact, in order to keep the purely atomic part of each marginal measure \(\rho_{S_j}\) distinct from the diffuse one, for any \( j = 2, \ldots, k \), we may need a decomposition of the Markov combination with a number of components growing along the recursion. Likewise this requires mixture representations for the hyper-Markov combination also of growing size. Thus, in order to construct general hyper-Markov combinations, the Ferguson representation for Dirichlet processes seems to be appropriate [4], since it is directly constructed on the canonical space of the process. We also believe that approximations theorems like Theorems 14 and 15 continue to hold in general.

References


