Bayesian Semiparametric Estimation of Elliptic Copulae

Alessio Sancetta*
University of Cambridge

January 18, 2008

Abstract

The copula function is the joint distribution function of a random vector with uniform marginals. It is used to define multivariate distributions with given marginals. The elliptic class of copulae is a subset of the class of copula functions that is fully determined by a scaling matrix and a function with domain in the positive reals. It contains the normal copula and the t-copula often used in applications as well as many other ones. This class is somehow "parsimonious" relative to the whole class of copulae, yet attention is usually confined to the Gaussian and Student case only. Here we establish weak conditions for posterior consistency (in a frequentist sense) of the Bayesian semiparametric estimator of copulae in the elliptic class when the scaling matrix is full rank. Conditions for consistency when the marginals are given but subject to estimation error are provided.

Key Words: Copula, copula modelling, Dirichlet prior process, Kullback-Leibler support, Mixture of Gaussian Densities.

1 Introduction

Multivariate distributions are complex objects and it is often conceptually difficult to model directly the whole distribution F of a multivariate random variable $X = (X_1, ..., X_K)$, while it can be conceptually easier to model the univariate marginals $F_1, ..., F_K$ first and then try to model their dependence structure. The formal validity of this approach relies on Sklar's theorem (e.g. Schweizer and Sklar, 1983, Scarsini, 1989, for extensions) that provides a representation of any joint distribution function in terms of a function of the marginals

$$F(x) = C(F_1(x_1), ..., F_K(x_K)),$$

where C is called copula function ($x := (x_1, ..., x_K)$). When the marginals are continuous, the copula is unique and $F_k(X_k)$ is $[0, 1]$ uniformly distributed. If this is not the case, there is still a transformation that maps the original variables into uniform $[0, 1]$. Hence, we can assume that the marginals are continuous, so that the copula is simply the joint distribution of a $[0, 1]^K$random vector with uniform marginals.

*Address for Correspondence: Faculty of Economics, Sidgwick Avenue, University of Cambridge, CambridgeCB3 9DE, UK. E-mail: alessio.sancetta@econ.cam.ac.uk. Acknowledgment: I thank Stephen G. Walker for explaining me the extension of one of his results.

Lately, there has been considerable interest in the copula as a tool for modelling in financial applications (e.g. Embrechts et al., 2002, Patton, 2006, see also Mikosch, 2006, for a critique) as well as in other areas (e.g. Hoff, 2007, and references therein) as opposed to its first use as a tool to describe probabilistic metric spaces (e.g. Schweizer and Sklar, 1983) and dependence conditions (e.g. Joe, 1997).

Modelling and inference for joint distributions via the copula is often parametric, where the copula function is assumed to be known up to a finite dimensional parameter. Nonparametric estimation and inference has also been considered by several authors (e.g. Fermanian et al., 2004, Fermanian and Scaillet, 2003, Sancetta and Satchell, 2004, Sancetta, 2007, and references therein). Parametric methods are often subject to bias, while nonparametric methods are subject to the well known "curse of dimensionality" but can consistently estimate almost any copula density function (Sancetta, 2007, provides almost universal consistency using Bernstein/Kantarovich polynomials). A middle ground alternative is semiparametric estimation by restriction of the class of copulae (e.g. Gagliardini and Gouriéroux, 2007, for the class of Archimedean copulae).

Here, the class of possible copulae will be restricted to the class of elliptic copulae. Elliptic copulae are related to elliptic distributions, routinely used in multivariate analysis (e.g. Fang et al., 2002, Hult and Lindskog, 2002). We shall restrict attention to elliptic copulae with full rank scaling matrix Σ, which is also an unknown parameter. The goal of the paper is to define a Bayesian semiparametric estimator for this class of copulae and to provide conditions under which the posterior is strongly consistent. There is a rich literature on Bayesian estimation for infinite dimensional parameters (e.g. Barron et al, 1999, Ghosal et al., 1999, 2000, Ghosal and van der Vaart, 2007a, 2007b, Kleijn and van der Vaart, 2006, Lijoi et al., 2005, Walker, 2004, Walker et al., 2007); see Walker 2004 for a review of results up to 2004. These results are directly relevant to this study. However, considerable effort is required to establish primitive weak and easy to verify conditions that lead to consistency for the Bayesian semiparametric estimator of an elliptic copula. Mutatis mutandis, it is possible to draw some parallels with the problem of consistency of the Bayesian nonparametric density estimator via Gaussian mixtures. However, the details differ and previous results cannot be directly used (e.g. Ghosal et al., 1999, Lijoi et al., 2005). Moreover, we shall allow the marginals to be subject to estimation error and show that consistency is still possible.

Below, we shall review some basic definitions concerning elliptic copulae. Section 2 formally introduces the estimation problem and studies the frequentist properties of the Bayesian estimator of the copula when the true marginals are known. Conditions for posterior strong consistency when the marginals are subject to estimation error are also established. Virtually all the proof are deferred to Section 3.

1.1 Background on Elliptic Copulae

Let X be a K dimensional elliptically distributed random vector with mean zero, full rank scaling matrix Σ and generator g. Then, X has density $|\Sigma|^{-1/2} g\left(X'\Sigma^{-1}X\right)$, where the prime stands for the transpose. If Σ is not full rank, this representation does not hold (see Hult and Lindskog, 2002, for the general case). In this paper, attention is restricted to the subclass of elliptic copulae with full rank scaling matrix Σ. The class of elliptic copula densities is defined as a transformation of the class of elliptic densities where Σ is a positive definite matrix with diagonal entries equal to one. Then, the univariate marginal distributions F_g are all the same and only depend on g (Fang et al., 2002, eq. 1.5, for the exact form). Define the quantile function $Q_g(u) := \inf_{x\in\mathbb{R}} \{F_g(x) > u\}$. Let $U = (U_1, ..., U_K)$ be a random variable with values in the unit K dimensional cube and with uniform $[0,1]$ marginals. Let $\mathcal{Q}_g : [0,1]^K \to \mathbb{R}^K$ be an operator such that

$$\mathcal{Q}_g(U) = (Q_g(U_1), ..., Q_g(U_K)). \tag{1}$$

Then, U has elliptic copula with full rank scaling matrix Σ if and only if its density at u is $|\Sigma|^{-1/2} g\left(\mathcal{Q}_g(u)' \Sigma^{-1} \mathcal{Q}_g(u)\right) J_g(u)$ where $J_g(u)$ is the Jacobian of the transformation $x \mapsto \mathcal{Q}_g(u)$. In compact notation, we just write

$$c_g(u|\Sigma) := |\Sigma|^{-1/2} g\left(\mathcal{Q}_g(u)' \Sigma^{-1} \mathcal{Q}_g(u)\right) J_g(u). \tag{2}$$

By the properties of elliptic random variables (Hult and Lindskog, 2002), deduce that U has elliptic copula (2) if and only if the following stochastic representation holds

$$\mathcal{Q}_g(U) \stackrel{d}{=} RAS \tag{3}$$

where $\stackrel{d}{=}$ is equality in distribution, S is uniformly distributed in the unit hypersphere $\left\{s \in \mathbb{R}^K : s's \leq 1\right\}$, A is full rank such that $AA' = \Sigma$ and R is a positive real random variable. The generator g is uniquely determined by R up to a scaling factor, i.e. X_1 with scaling matrix Σ and generator $g(x)$ has same distribution as X_2 with scaling matrix $v\Sigma$ and generator $g(xv)$. For this reason, there is no loss of generality in restricting the diagonal entries of Σ to be equal to one.

1.2 Example of Copula Modelling

An elliptic copula can be used to construct parsimonious multivariate models for high dimensional data even in a time series context. To provide some motivation for the estimation of a copula and for the sake of definiteness we remark on a natural extensions of the Constant Conditional Correlation GARCH of Bollerslev (1990). Clearly, this is one of many important problems (e.g. Hoff, 2007, for survey data modelling), but it seems instructive to dwell on this specific time series example. To this end, suppose $X_t = (X_{t1}, ..., X_{Kt})$ has components each following a univariate t-GARCH model with possibly different degrees of freedom. Denote the resulting univariate conditional distributions by $F_k(x|\mathcal{F}_{t-1})$ where $\mathcal{F}_{t-1}$ is the sigma algebra generated by past observations of X_t. For simplicity, just assume that $F_k(x|\mathcal{F}_{t-1})$ is known. The practical case when the marginals need to be estimated will also be considered in this paper. By the univariate GARCH assumption, each X_{tk} depends on $\mathcal{F}_{t-1}$ only through its past values. Define $U_{tk} = F_k(X_{tk}|\mathcal{F}_{t-1})$ and $U_t = (U_{t1}, ..., U_{tK})$. Then, U_{tk} is a uniform $[0,1]$ random variable independent of $\mathcal{F}_{t-1}$ (e.g. Rio, 2000, Lemma F.1). Assuming that the "true" copula is in the elliptic class, this paper shows that we can consistently estimate the posterior distribution of the joint density of U_t by Bayesian semiparametric methods. Note that independence between U_t and $\mathcal{F}_{t-1}$ implies that both the scaling matrix Σ and the generator g are independent of the past, i.e. independent of $\mathcal{F}_{t-1}$. Hence, (2) can be directly used in a time series context as well (under time homogeneity of Σ and g).

By the mentioned independence of U_t with $\mathcal{F}_{t-1}$, the two stage modelling of marginals and copula requires extra care in order to avoid common mistakes found in applied work.

For example, Engle (2002) assumes that $F_k(x|\mathcal{F}_{t-1})$ is the distribution of a GARCH process and uses a Gaussian copula to join these marginals (he does not uses the terminology used here, but the result is the same; see his eq. 26). The novelty of that paper is to make the scaling matrix dependent on $\mathcal{F}_{t-1}$ (he uses the notation R_t for the scaling matrix), but by the previous remarks, the scaling matrix cannot depend on $\mathcal{F}_{t-1}$. Basically, Engle's Dynamic Conditional Correlation model violates Kolmogorov consistency of probabilities (i.e. such process does not exist). This mistake is common in related work that uses the same terminology used here (e.g. Patton, 2006, eq.'s 6-11,16,17, among others). Clearly, the scaling matrix (or copula parameters in general) can be time inhomogeneous. Frequentist methods that assume local stationarity often allow to deal with time inhomogeneous parameters under suitable conditions (e.g. Dahlhaus, 1997). Time inhomogeneous parameters will not be considered here.

2 Estimation Problem

The goal of this paper is give weak conditions for posterior consistency of the elliptic copula densities $c_g(u|\Sigma)$ without making any assumption on the generator g, though some regularity condition will be used. The estimated copula needs to satisfy the usual properties of copulae (e.g. Schweizer and Sklar, 1983) and one notable property is that the marginals need to be uniform. This requires some structure on the estimation procedure. We address this issue via the next result that provides a representation of elliptic copulae via mixtures of Gaussian copulae.

Theorem 1 *Any elliptic copula with full rank scaling matrix Σ can be written as*

$$C_g(u|\Sigma) = \mathbb{E}C_\phi(u|\Sigma/V) \tag{4}$$

where C_ϕ is the Gaussian copula and V is some positive random variable (expectation is w.r.t. V).

Proof. From (3), $\mathcal{Q}_g(U) \stackrel{d}{=} \sqrt{\chi_K^2/V}AS$, where $V \stackrel{d}{=} \chi_K^2/R^2$ and χ_K^2 is a Chi square random variable with K degrees of freedom. Conditioning on $V = v$, $\mathcal{Q}_g(U) \stackrel{d}{=} N(0_K, \Sigma/v)$ where $N(0_K, \Sigma/v)$ is a Gaussian random variable with mean vector zero and covariance matrix Σ/v. Then, take expectation. ■

Hence, the problem of consistently estimating the density of $C_g(u|\Sigma)$ is equivalent to the problem of estimating the finite dimensional parameter Σ and the law of V, which we denote by P, an infinite dimensional parameter. Next we define the Bayesian semiparametric estimator and show that its posterior is consistent.

2.1 Nonparametric Bayesian Estimation

By Theorem 1, the problem reduces to the estimation of a mixture of Gaussian copulae. Bayesian nonparametric estimation of mixtures models requires to select a prior measure Π on the set $\mathcal{P}$ of distributions with support, in this case, in the positive real line. If also Σ is unknown, the prior Π is assumed to have support in $\Theta = \mathcal{C} \times \mathcal{P}$ where $\mathcal{C}$ is a suitable subset of the set of correlation matrices (any positive definite matrix Σ with diagonal entries equal to one is a correlation matrix). The prior Π through the map $(\Sigma, P) =: \theta \mapsto c_\theta$ induces a prior on the set of elliptic copula densities

$$\left\{ c_\theta (u) = \int_0^\infty c_\phi (u|\Sigma/v) P(dv) : (\Sigma, P) \in \mathcal{C} \times \mathcal{P} = \Theta \right\}. \tag{5}$$

Then, the posterior Π_n induces a random copula as follows

$$c_n (u) = \int_\Theta c_\theta (u) \Pi_n (d\theta) \tag{6}$$

where, for $A \subseteq \Theta$,

$$\Pi_n (A) := \frac{\int_A \prod_{i=1}^n c_\theta (U_i) \Pi (d\theta)}{\int_\Theta \prod_{i=1}^n c_\theta (U_i) \Pi (d\theta)}$$

and $(U_i)_{i>0}$ are iid random variables with values in $[0,1]^K$ and joint density in (5). By the remarks in Section 1.2, if the marginals are known, there is no loss to assume that the data are iid with uniform marginals (e.g. Rio, 2000, Lemma 1F for the general construction). The case when the marginals are subject to estimation error will be considered later.

2.2 Consistency of Posterior

We need to recall some definitions. For two arbitrary measures P and Q, on some set $\mathcal{X}$ with densities p and q w.r.t. some dominating measure μ, their Kullback-Leibler distance is defined as $D(P,Q) = \int_\mathcal{X} \ln (p/q) dP$ while their Hellinger and total variation distance are defined as $d_H (P,Q) = \left| \int_\mathcal{X} \left| \sqrt{p} - \sqrt{q} \right|^2 d\mu \right|^{1/2}$ and $d_{TV} (P,Q) = \int_\mathcal{X} |p-q| \, d\mu$ and they are just the L_2 distance of the square root of two densities and the L_1distance of densities. The following relations are known (e.g. Pollard, 2002, p. 61-62):

$$d_H (P,Q)^2 \leq d_{TV} (P,Q) \leq d_H (P,Q) \leq D(P,Q)^{1/2} \tag{7}$$

showing that d_{TV} and d_H are topologically equivalent. The relation between total variation and Kullback-Leibler distance is often called Pinsker inequality. We define the following sets

$$K_\epsilon := \{\theta \in \Theta : D(C_{\theta_0}, C_\theta) \leq \epsilon\} \tag{8}$$

$$A_\epsilon := \{\theta \in \Theta : d_{TV} (C_{\theta_0}, C_\theta) > \epsilon\}. \tag{9}$$

The support of Π is in $\Theta = (\mathcal{C}, \mathcal{P})$ and we write $\Pi_\Sigma (\bullet) = \Pi (\bullet, \mathcal{P})$ and $\Pi_P (\bullet) = \Pi (\mathcal{C}, \bullet)$ for the marginals. An element $P_0 \in \mathcal{P}$ is said to be in the support of Π_P if every open neighborhood of P_0 in the weak topology is given positive Π_P measure. We shall establish a.s. convergence to zero of the posterior over A_ϵ with respect to the "true" ∞-fold product measure $C_{\theta_0}^\infty$, the infinite product of the copula measure induced by the copula density c_{θ_0}, assumed to be the true density of the data. We shall use linear functional notation for the expectation w.r.t. Π. Hence, $\Pi(A)$ means expectation of A under the prior Π, where A can be a set or some other object for which the expectation makes sense (i.e. measurable). Introduce the following conditions that will be discussed at length after the statement of the main results. Recall that K is the dimension of U_i.

Condition 2 $\int_0^\infty v^{K/2+\alpha} P_0 (dv) < \infty$ *for some* $\alpha > 0$.

Condition 3 *Σ_0 is full rank.*

Condition 4 *(i.) For $a \to \infty$, $a^{(K/2+\alpha)}\Pi_P\left(P\left([a,\infty)\right)\right) \to 0$ for some $\alpha > 0$; (ii.) for $M \to \infty$, $\exp\left\{\alpha_1 M^{\alpha_2}\right\}\Pi_\Sigma\left(\left\{\Sigma \in \mathcal{C} : |\Sigma|^{-1} > M\right\}\right) \to 0$ for some $\alpha_1, \alpha_2 > 0$ ($|\Sigma|$ is the determinant of Σ).*

These conditions are sufficient for strong consistency of the posterior.

Theorem 5 *Suppose that θ_0 is in the support of Π. Under Conditions 2, 3, and 4, $\Pi_n\left(A_\epsilon\right) \to 0$, $C^\infty_{\theta_0}$-a.s.*

Remark 6 *The above conditions imply that if θ_0 is in the support of Π then $\Pi\left(K_\epsilon\right) > 0$, i.e. θ_0 is also in the Kullback-Leibler support of Π.*

A common example of nonparametric prior is the Dirichlet process $\mathcal{D}\left(\nu, G\right)$ where $\nu \in (0, \infty)$ is a scaling parameter controlling the confidence on the mean prior probability measure G with support $[0, \infty)$. In particular, we derive consistency using $\mathcal{D}\left(\nu, G\right)$ as prior. It is convenient to use following hierarchical representation

$$\begin{array}{rcl} U_i|V_i, \Sigma & \sim & C_\phi\left(u|\Sigma/V_i\right) \\ V_i|P & \sim & P \\ P & \sim & \mathcal{D}\left(\nu, G\right) \\ \Sigma & \sim & P_\Sigma. \end{array}$$

Corollary 7 *For the above hierarchical representation, suppose that $v^{(K/2+\alpha)}G\left([v,\infty)\right) \to 0$, as $v \to \infty$, for some $\alpha > 0$, and P_Σ is a distribution having support $\mathcal{C}_s := \{\Sigma \in \mathcal{C} : |\Sigma| \geq s\}$ for some $s > 0$. Then, $\Pi_n\left(A_\epsilon\right) \to 0$, $C^\infty_{\theta_0}$-a.s. for any P_0 satisfying Condition 2 and any $\Sigma_0 \in \mathcal{C}_\epsilon$.*

2.2.1 Remarks on Condition 2 and 3

Conditions 2 and 3 restrict the class of elliptic copulae. In particular, Condition 2 restricts the mixing measure but does not rule out strong tail dependence (e.g. Joe, 1997 for definitions). To appreciate this point, note that if X is a K dimensional student t random variable with r degrees of freedom and scaling matrix Σ, then $X \stackrel{d}{=} N\left(0, \Sigma\right)/\sqrt{V}$ where $V \stackrel{d}{=} \chi^2_{(r)}/r$. Since P_0 is the law of V, then $\int_0^\infty v^{K/2+\alpha} P_0\left(dv\right) < \infty$ for any $r > 0$ and K (finite). Finally, the full rank condition of Σ is required for the representation in (4) to be true and cannot be weakened.

2.2.2 Remarks on Condition 4($i.$)

It seems reasonable that if the true mixing measure P_0 needs to satisfies Condition 2, then the mean of P under the prior Π_P also satisfies Condition 2, as implied by Condition 4($i.$). Under the Dirichlet process prior $\mathcal{D}\left(\nu, G\right)$, this implies that G satisfies Condition 2.

2.2.3 Remarks on Condition 4($ii.$)

Checking Condition 4($ii.$) can be challenging for two reasons. First, it can be difficult to find a prior on $\mathcal{C}$ or some suitable subset of it. It can be possible to simulate correlation matrices via different approaches. For example simulate a covariance matrix from a Wishart distribution and then transform to a correlation matrix. To the author's knowledge, these approaches do not lead to closed form expressions for the distribution of the correlation matrix. The second related problem is that it can be difficult to verify the exponential tail condition on $|\Sigma|^{-1}$. Only ad hoc solutions seem to exist to these problems. One of these ad hoc possibilities is to use a factor representation of the correlation matrix. We give some details next.

2.2.4 Calculations for a One Factor Representation of the Scaling Matrix

Suppose $Y = \gamma F + \sigma Z$ where Y is a $K \times 1$ vector, F is a mean zero variance one random variable uncorrelated with Z, a K dimensional vector of iid mean zero variance one random variables. The parameters γ and σ are $K \times 1$ and one dimensional, respectively. Defining

$$\Sigma(\sigma,\gamma) := diag\left(\mathbb{E}YY'\right)^{-1/2}\mathbb{E}YY' diag\left(\mathbb{E}YY'\right)^{-1/2}$$

it is easy to see that

$$\Sigma_{ij}(\sigma,\gamma) = \frac{\sigma^2\delta_{ij} + \gamma_i\gamma_j}{\sqrt{\sigma^2+\gamma_i^2}\sqrt{\sigma^2+\gamma_j^2}}$$

where $\delta_{ij} = 1$ if $i = j$, zero otherwise (Kronecker delta), and γ_i is the i^{th} entry of γ. Clearly $\mathcal{C}(\sigma,\gamma) := \left\{\Sigma(\sigma,\gamma) \in \mathcal{C} : \sigma > 0, \gamma \in \mathbb{R}^K\right\} \subset \mathcal{C}$. The stringent restrictions imposed on $\mathcal{C}$ are compensated by the great tractability for the Bayesian problem considered here. In particular, we have the following.

Lemma 8 *Suppose $\Sigma \in \mathcal{C}(\sigma,\gamma)$. Let $e_1, ..., e_{K-1}$ be arbitrary but fixed orthogonal K dimensional vectors. Then,*

$$\Sigma^{-1} = D\Gamma\Lambda^{-1}\Gamma' D$$

where:
Λ is the diagonal matrix of eigenvalues with $\Lambda_{11} = \sigma^2 + \gamma'\gamma$ and $\Lambda_{kk} = \sigma^2$, for $k > 1$;
$\Gamma := (\Gamma_1, ..., \Gamma_K)$ is a $K \times K$ matrix of orthonormal vectors, $\Gamma_1 = \gamma/|\gamma|$ ($|\gamma|$ is the Euclidean norm of γ) and for $k > 1$ $\Gamma_k = \Delta_k e_k/|\Delta_k e_k|$, $\Delta_1 = I_K$, $\Delta_k = \left(I_K - \Gamma_{k-1}\Gamma'_{k-1}\right)\Delta_{k-1}$, I_K is the K-dimensional identity matrix;
D is a diagonal matrix with $D_{kk} = \sqrt{\sigma^2+\gamma_k^2}$.
It also follows that $|\Sigma|^{-1} = \left(\sigma^{2K} + \gamma'\gamma\sigma^{2(K-1)}\right)^{-1}\prod_{k=1}^{K}\left(\sigma^2+\gamma_k^2\right)$.

From Lemma 8, we can find a prior that satisfies Condition 4($ii.$). However, we will need to further restrict the class of scaling matrices to the following $\mathcal{C}_a(\sigma,\gamma) := \left\{\Sigma(\sigma,\gamma) \in \mathcal{C} : \sigma \geq a > 0, \gamma \in \mathbb{R}^K\right\}$. Hence, a prior with support in $\mathcal{C}_a(\sigma,\gamma)$ only allows for $\sigma \in [a,\infty)$. This is strictly stronger than the assumption that the scaling matrices are full rank.

Lemma 9 *Let $\Sigma \in \mathcal{C}_a(\sigma,\gamma)$ and $\Pi_\Sigma = \Pi_\sigma \times \Pi_\gamma$ where $\Pi_\sigma([a,\infty)) = 1$ and Π_γ is a K dimensional Gaussian distribution with finite mean and finite diagonal covariance matrix. Then, Condition 4($ii.$) is satisfied.*

2.3 Consistency when the Marginals are Estimated

Statistical interest in the copula is when the marginals are given. However, in applied problems, it is often the case that marginals have to be estimated. Then, it is natural to ask if the results of this paper continue to hold. The answer is yes, clearly under suitable conditions. To set the scene for the affirmative solution of the problem, let $(X_i)_{i>0}$ be a sequence of random variables with values in $\mathbb{R}^K$ with univariate marginals $F_1(x_1), ..., F_K(x_K)$ and elliptic copula. Suppose that the marginals are unknown, and we use instead $F_{1,i}(x_1), ..., F_{K,i}(x_K)$ for the marginals of X_i. We call these the surrogate marginals. These may depend on the sample size n, though they not need to, as it is the case in a prequential framework (e.g. Dawid, 1997). From the true and surrogate marginals we derive $(U_i)_{i>0}$ and $\left(\hat{U}_i\right)_{i>0}$, respectively as random variables with values in $[0,1]^K$ where $(U_i)_{i>0}$ are assured to have uniform marginals (e.g. Rio, 2000, Lemma F.1). While continuity of the marginals is not required, in practice problems may arise depending on the estimator to be used (see Hoff, 2007, for problems that arise in some circumstances when the marginals are discontinuous and estimated by the empirical distribution). Define the posterior based on $\left(\hat{U}_i\right)_{i>0}$ as

$$\hat{\Pi}_n(A) := \frac{\int_A \prod_{i=1}^n c_\theta\left(\hat{U}_i\right)\Pi(d\theta)}{\int_\Theta \prod_{i=1}^n c_\theta\left(\hat{U}_i\right)\Pi(d\theta)}.$$

Given that both $(U_i)_{i>0}$ and $\left(\hat{U}_i\right)_{i>0}$ are defined on the same probability space, denote this probability space by $(\Omega, \mathbb{P})$. The following conditions together with the ones used previously are sufficient for strong consistency of $\hat{\Pi}_n(A)$.

Condition 10 *For any $\epsilon > 0$,*

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n\to\infty}\sup_{i\geq n}\left|U_i(\omega) - \hat{U}_i(\omega)\right| > \epsilon\right\}\right) = 0.$$

Condition 11 *There is an $\alpha > 0$, such that for Π- almost all θ ,*

$$\sup_{i>0}\mathbb{E}\left|\ln\frac{c_\theta\left(\hat{U}_i(\omega)\right)}{c_\theta\left(U_i(\omega)\right)}\right|^{1+\alpha} < \infty.$$

Theorem 12 *Suppose that θ_0 is in the support of Π. Under Conditions 2, 3, 4, 10 and 11, $\hat{\Pi}_n(A_\epsilon) \to 0$, $\mathbb{P}$-a.s. .*

2.3.1 Checking Conditions 10 and 11

To avoid peripheral arguments, suppose that the marginals are continuous. Suppose also that these marginals are indexed by some finite dimensional parameter, continuous with respect to it and that this parameter is unknown and needs to be either estimated or calibrated. Then, by the continuous mapping theorem, Condition 10 is satisfied if the unknown parameter converges a.s. to the true one. Conditions for this to hold under maximum likelihood estimation are well known (e.g. van der Vaart and Wellner, 2000). Alternatives based on sequential estimation like Bayesian or prequential estimation usually require weaker conditions; indeed for Condition 10 we only need prediction consistency which is weaker than consistency of an estimator (see Dawid, 1997 and references therein).

Finally, Condition 11 can be checked using the following.

Lemma 13 *Suppose that Π_P is a Dirichlet prior process with base probability measure G such that for some set $[a_1, a_2] \subset (0, 1)$, $G([a_1, a_2]) > 0$. Suppose that for any $k = 1, ..., K$, $\sup_{i>0} \mathbb{E} \left| \ln \left(1 - \hat{U}_{ik}(\omega)\right) \right|^{(1+\alpha)} < \infty$ for some $\alpha > 0$. Then, under Conditions 3 and 4, Condition 11 is satisfied.*

2.4 Further Remarks

One of the fastest areas of research of Bayesian nonparametrics is concerned with computational issues. In particular the method of Escobar (1994) opened the way for Markov Chain Monte Carlo (MCMC) estimation using Dirichlet processes (see also Walker, 2005, and references therein for a review). A recent appealing approach is to use Sethuraman's constructive definition of a Dirichlet process:

$$P(dv) = \sum_{s>0} W_s \delta_{V_s}(dv)$$

where $(W_s)_{s>0}$ are random variables in the infinite unit simplex derived by stick breaking construction (e.g. Sethuraman, 1994), $(V_s)_{s>0}$ is a sequence of iid random variables from the measure G of the Dirichlet process $\mathcal{D}(\nu, G)$, while $\delta_V(\bullet)$ is the point measure at V. Given that it is simple to simulate from $P_N(dv) = \sum_{s=1}^{N} W_s \delta_{V_s}(dv)$, when N is finite, a natural approach is to approximate the posterior using P_N as prior (Ishwaran and James, 2002, for details). This approach can be used here as well.

In some applications it can be interesting to allow for unknown time inhomogeneous parameters. For the simpler problem of predictive density estimation, it is possible to modify the posterior to account for time inhomogeneous finite dimensional parameters (e.g. Sancetta, 2007, using the results of Bousquet and Warmuth, 2002). It would be interesting to investigate notions of posterior consistency in this more general framework.

3 Proofs

We find convenient to collect here the notation used in the lemmata and the proofs and refer to it when required. The reader can just browse to it whenever unfamiliar/undefined notation is found in the proofs.

Notation 14 *$\mathcal{P}$ is the set of measures on the positive reals;*
$\mathcal{C}$ is the set of full rank K dimensional correlation matrices;
$\lesssim$ and $\gtrsim$ stand for inequality up to a finite absolute constant, while $\asymp$ implies that the right hand side is proportional to the left hand side;
Φ is the standard Gaussian distribution and ϕ its density; if the argument of ϕ is a vector, then, ϕ will denote the standard multivariate Gaussian density;

For a positive definite matrix Σ, write $\Sigma^{1/2}$ for the matrix such that $\left(\Sigma^{1/2}\right)\left(\Sigma^{1/2}\right)' = \Sigma$, $|\Sigma|$ stands for its determinant, $|\Sigma|_\infty := \max_{1\leq i,j\leq K}|\Sigma_{ij}|$, $\Lambda(\Sigma)$ is the matrix of eigenvalues of Σ and $\lambda_k(\Sigma)$ is its k^{th} eigenvalue;
For $a>0$, $\delta\in(0,1)$and constants $m<M$, define the the following classes of mixtures of normals

$$\mathfrak{F} := \left\{\int_0^\infty v^{K/2}\phi\left(v^{1/2}\Sigma^{-1/2}x\right)P(dv) : P\in\mathcal{P}, \Sigma\in\mathcal{C}\right\}$$

$$\mathfrak{F}_P := \left\{\int_0^\infty v^{K/2}\phi\left(v^{1/2}\Sigma^{-1/2}x\right)P(dv) : \Sigma\in\mathcal{C}\right\}$$

$$\mathfrak{F}_a^{\Sigma,} := \left\{\int_0^\infty v^{K/2}\phi\left(v^{1/2}\Sigma^{-1/2}x\right)P(dv) : P((0,a]) = 1\right\}$$

$$\mathfrak{F}_{a,\delta}^{\Sigma} := \left\{\int_0^\infty v^{K/2}\phi\left(v^{1/2}\Sigma^{-1/2}x\right)P(dv) : P((0,a]) > 1-\delta\right\}$$

$$\mathfrak{F}_P^{M} := \left\{\int_0^\infty v^{K/2}\phi\left(v^{1/2}\Sigma^{-1/2}x\right)P(dv) : |\Sigma|^{-1}\leq M\right\}$$

$$\mathfrak{F}_P^{m,M} := \left\{\int_0^\infty v^{K/2}\phi\left(v^{1/2}\Sigma^{-1/2}x\right)P(dv) : m<|\Sigma|^{-1}\leq M\right\}$$

$$\mathfrak{F}_{a,\delta}^{m,M} := \left\{\int_0^\infty v^{K/2}\phi\left(v^{1/2}\Sigma^{-1/2}x\right)P(dv) : P((0,a]) > 1-\delta, m<|\Sigma|^{-1}\leq M\right\};$$

For any set of functions $\mathfrak{G}$, and $\delta>0$, $N(\delta,\mathfrak{G})$ is the δ-covering number of $\mathfrak{G}$ under the L_1 norm (van der Vaart and Wellner, 2000, for details).

The lemmata are numbered sequentially and proofs may refer to technical results that are only stated subsequently in order not to interrupt the the flow of the main steps of each proof.

3.1 Proof of Theorem 5

Proof of Theorem 5. According to a slight extension of Theorem 4 in Walker (2004) (see the proof of Theorem 1 in Lijoi et al., 2005), it is enough to show that: (1.) $\Pi(K_\epsilon)>0$ and (2.) for any $\delta>0$ there is a countable partition $(A_j)_{j>0}$ of A_ϵ such that $A_j := \left\{\theta\in\Theta : d_{TV}\left(C_{\theta_j},C_\theta\right)<\delta\right\}$ where $d_{TV}\left(C_{\theta_0},C_{\theta_j}\right)>\epsilon$ for any j and $\sum_j \Pi^\beta(A_j)<\infty$ for some $\beta\in(0,1)$. This is accomplished using Lemmata 15 and 18, respectively. ■

3.1.1 Statement and Proof of Lemma 15

At some stage in the proof we recall the following fact: an open set centered at P_0 in the weak topology of $\mathcal{P}$ can be metricized by the Bounded Lipschitz Metric (Dudley metric)

$$\left\{P\in\mathcal{P} : \sup_{f\in BL_1}\left|\int f d(P-P_0)\right|\leq\epsilon\right\}, \tag{10}$$

where BL_1 is the class of functions where each element f satisfies $\|f\|_{BL} = \|f\|_L + \|f\|_\infty \leq 1$, $\|f\|_L$ being the Lipschitz constant of f and $\|f\|_\infty$ its L_∞ norm (Dudley, 2002, ch. 11.2 for further details).

Lemma 15 *Under Conditions 2, 3 and 4(i.), if* $\theta_0 := (P_0, \Sigma_0)$ *is in the support of* Π, *then* $\Pi(K_\epsilon) > 0$, K_ϵ *as in (8).*

Proof. Define

$$f_0(x) := \int_0^\infty v^{K/2}\phi\left(v^{1/2}\Sigma_0^{-1/2}x\right)P_0(dv),$$

$$f_P(x) := \int_0^\infty v^{K/2}\phi\left(v^{1/2}\Sigma_0^{-1/2}x\right)P(dv),$$

$$f_{P\Sigma}(x) := \int_0^\infty v^{K/2}\phi\left(v^{1/2}\Sigma^{-1/2}x\right)P(dv).$$

By the change of variables $u_k \mapsto \Phi(x_k)$ for $k = 1, ..., K$,

$$\begin{aligned}
D(C_{\theta_0}, C_\theta) &= \int_{\mathbb{R}^K} \frac{f_0(x)}{|\Sigma_0|^{1/2}} \ln \frac{f_0(x)|\Sigma|^{1/2}}{f_{P\Sigma}(x)|\Sigma_0|^{1/2}} dx \\
&= \int_{\mathbb{R}^K} \frac{f_0(x)}{|\Sigma_0|^{1/2}} \ln \frac{f_0(x)}{f_{P\Sigma}(x)} dx + \ln \frac{|\Sigma|^{1/2}}{|\Sigma_0|^{1/2}} \\
&= \int_{\mathbb{R}^K} \frac{f_0(x)}{|\Sigma_0|^{1/2}} \ln \frac{f_0(x)}{f_P(x)} dx + \int_{\mathbb{R}^K} \frac{f_0(x)}{|\Sigma_0|^{1/2}} \ln \frac{f_P(x)}{f_{P\Sigma}(x)} dx + \ln \frac{|\Sigma|^{1/2}}{|\Sigma_0|^{1/2}} \\
&\leq \sup_{x\in\mathbb{R}^K} |f_0(x) - f_P(x)| + \int_{\mathbb{R}^K} \frac{f_0(x)}{|\Sigma_0|^{1/2}} \ln \frac{f_P(x)}{f_{P\Sigma}(x)} dx + \ln \frac{|\Sigma|^{1/2}}{|\Sigma_0|^{1/2}} \\
&\quad \text{[by Lemma 16]} \\
&= \mathrm{I} + \mathrm{II} + \mathrm{III}
\end{aligned}$$

and we bound each term separately.

Control over I.

Define

$$\Gamma_0 := \left\{\gamma > 0 : \int_\gamma^\infty v^{K/2} P_0(dv) < \epsilon\right\}, \Gamma_P := \left\{\gamma > 0 : \int_\gamma^\infty v^{K/2} P(dv) < \epsilon\right\}.$$

Condition 2 assures that Γ_0 is not empty, and, by Lemma 17, Γ_P is also not empty Π-a.s., but for convenience we will suppress the a.s. qualifier throughout as, by the statement of the lemma, we are only interested in Π-non-null sets. Hence, define

$$a = \inf\{\gamma > 0 : \gamma \in \Gamma_0 \cap \Gamma_P\}. \tag{11}$$

Then,

$$\begin{aligned}
\mathrm{I} &= \sup_{x\in\mathbb{R}^K} \left| \left(\int_0^a + \int_a^\infty\right) v^{K/2}\phi\left(v^{1/2}\Sigma_0^{-1/2}x\right)(P - P_0)(dv) \right| \\
&\leq \sup_{x\in\mathbb{R}^K} \left| \int_0^a v^{K/2}\phi\left(v^{1/2}\Sigma_0^{-1/2}x\right)(P - P_0)(dv) + \int_0^a v^{K/2}(P + P_0)(dv) \right| \\
&\quad \text{[because } \phi < 1\text{]} \\
&\leq \sup_{x\in\mathbb{R}^K} \left| \int_0^a v^{K/2}\phi\left(v^{1/2}\Sigma_0^{-1/2}x\right)(P - P_0)(dv) \right| + 2\epsilon
\end{aligned}$$

by definition of a. To bound the remaining term, note that the family of functions

$$v \mapsto \left\{ v^{K/2} \phi \left(v^{1/2} \Sigma_0^{-1/2} x \right) \{ v \in [0, a] \} : x \in \mathbb{R}^K \right\} \tag{12}$$

is bounded by some constant proportional to $a^{K/2}$ and Lipschitz with Lipschitz constant less or equal to

$$\begin{aligned} & \sup_{v \in [0,a], x \in \mathbb{R}^K} \left| \left(\frac{d}{dv} \right) v^{K/2} \phi \left(v^{1/2} \Sigma_0^{-1/2} x \right) \right| \\ \leq & \sup_{v \in [0,a], x \in \mathbb{R}^K} \frac{v^{(K-2)/2}}{2} \phi \left(v^{1/2} \Sigma_0^{-1/2} x \right) \left(v^2 x' \Sigma_0^{-1} x + K \right) \\ & \text{[differentiating and rearranging]} \\ \lesssim & \; a^{(K-2)/2} \end{aligned}$$

because $\phi \left(v^{1/2} \Sigma_0^{-1/2} x \right) \left(v^2 x' \Sigma_0^{-1} x + K \right)$ is bounded for any $v \in [0, a]$ and $x \in \mathbb{R}^K$ as long as $K \geq 2$. Then, each element, say f, in (12) satisfies $\|f\|_{BL_1} \lesssim a^{(K-2)/2} + a^{K/2} \lesssim a^{K/2}$. This implies that the family of functions in (12) is equicontinuous. Define weak neighbors of P_0 of diameter δ in terms of the Dudley metric in (10). Choosing $\delta \lesssim \epsilon a^{-K/2}$ assures that we can find a P in the support of Π_P such that

$$\sup_{x \in \mathbb{R}^K} \left| \int_0^\infty v^K \phi \left(v^{1/2} \Sigma_0^{-1/2} x \right) \{ v \in (0, a] \} \left(P - P_0 \right) (dv) \right| \leq \epsilon,$$

implying I $\lesssim \epsilon$.

Control over II

Let Σ_{ij} be the i, j entry of Σ and similarly for Σ_{0ij}. Since Σ_0 is in the support of Π_Σ, we can choose $\Sigma_{ij} = (1 + \epsilon) \Sigma_{0ij}$, $\epsilon > 0$. Then, $\Sigma^{-1} = (1 + \epsilon)^{-1} \Sigma_0^{-1}$, implying

$$\Sigma_0^{-1} - \Sigma^{-1} = \frac{\epsilon}{(1 + \epsilon)} \Sigma_0^{-1}. \tag{13}$$

Hence,

$$\begin{aligned} \ln \frac{f_P(x)}{f_{P\Sigma}(x)} & \leq \sup_{v \geq 0} \ln \left(\frac{\phi \left(v^{1/2} \Sigma_0^{-1/2} x \right)}{\phi \left(v^{1/2} \Sigma^{-1/2} x \right)} \right) \\ & = \sup_{v \geq 0} - \frac{v}{2} x' \left(\Sigma_0^{-1} - \Sigma^{-1} \right) x \\ & = \sup_{v \geq 0} - \frac{v \epsilon}{2 (1 + \epsilon)} x' \Sigma_0^{-1} x \\ & \quad \text{[by (13)]} \\ & = 0 \end{aligned}$$

because $\epsilon > 0$, and $x' \Sigma_0^{-1} x \geq 0$, Σ_0^{-1} being positive definite. This implies that II ≤ 0.

Control over III.

By choice of Σ as in Control over II, since each term of Σ is a $(1+\epsilon)$ multiple of Σ_0, we have $|\Sigma| = (1+\epsilon)^K |\Sigma_0|$ implying III $\lesssim \epsilon$. Putting together I, II and III, we deduce that the set K_ϵ is not empty under Π. ■

Lemma 16 *Suppose F_0 and F are two distributions with densities f_0/c_0 and f/c w.r.t. some dominating measure μ with support $\mathcal{X}$, and c_0, c are constants of integration. Then,*

$$D(F_0, F) \le \frac{c_0 + c}{2c_0} \sup_{x\in\mathcal{X}} |f_0(x) - f(x)| + \ln\frac{c}{c_0}.$$

Proof. By Taylor expansion with integral reminder

$$\ln\left(\frac{x}{y}\right) = \int_0^1 \frac{x+\tau(y-x)}{x} d\tau (y-x). \tag{14}$$

Then, by (14), and the definition of the K-L distance,

$$\begin{aligned}
D(F_0, F) &= \int_{\mathbb{R}^K} \frac{f_0(x)}{c_0} \ln\frac{f_0(x)}{f(x)} + \ln\frac{c}{c_0} \\
&\le \frac{1}{c_0}\int_{\mathbb{R}^K}\int_0^1 f_0(x)\frac{(1-\tau) f_0(x) + \tau f(x)}{f_0(x)} d\tau dx \sup_{x\in\mathcal{X}} |f_0(x) - f(x)| + \ln\frac{c}{c_0} \\
&\quad \text{[by (14) and a simple upperbound]} \\
&= \int_0^1 \frac{(1-\tau)c_0 + \tau c}{c_0} d\tau \sup_{x\in\mathcal{X}} |f_0(x) - f(x)| + \ln\frac{c}{c_0},
\end{aligned}$$

using Fubini's Theorem because the integrals are finite. Computing the integral w.r.t. τ gives the result.

■

Lemma 17 *Under Condition 4, for any $\epsilon > 0$ there is a $\gamma < \infty$ such that*

$$\int_\gamma^\infty v^{K/2} P(dv) < \epsilon, \Pi_P - a.s.$$

Proof. It is sufficient to show that $\int_0^\infty v^{K/2} P(dv) < \infty$, Π_P-a.s. because integrability would imply that eventually for γ large enough $\int_\gamma^\infty v^{K/2} P(dv) < \epsilon$, Π_P-a.s.. By Condition 4 $\Pi_P(P([v,\infty))) \lesssim v^{-K/2-\alpha}$ for some $\alpha > 0$. Then,

$$\begin{aligned}
\Pi_P\left(\int_0^\infty v^{K/2} P(dv)\right) &= \Pi_P\left(\frac{K}{2}\int_0^\infty v^{K/2-1} P([v,\infty)) dv\right) \\
&\quad \text{[integrating by parts, e.g. Petrov (1995, Lemma 2.4)]} \\
&\le \frac{K}{2}\int_0^\infty v^{K/2-1} v^{-K/2-\alpha} dv \\
&\quad \text{[by Condition 4 } (i.)\text{]} \\
&< \infty.
\end{aligned}$$

Since for any random variable Y, $\mathbb{E}Y < \infty$ implies that $Y < \infty$ a.s., then the result follows. ■

3.1.2 Statement and Proof of Lemma 18

Lemma 18 *Under Conditions 3 and 4, for any $\delta > 0$ there exists a δ-cover $(A_j)_{j>0}$ of A_ϵ, as in (9), such that, for some $\beta \in (0,1)$,*

$$\sum_{j>0} \Pi^\beta (A_j) < \infty.$$

The poof of Lemma 18 is long and requires several intermediate results, which are sequentially derived next. We shall make heavy use of Notation 14 particularly to define classes of functions that will be related to the mixture of Gaussian copulae. The reader is recommended to consult Notation 14 while reading the statement of each lemma.

Lemma 19 *For any $\delta \in (0,1)$, there exists a finite $p > 0$, depending on δ and K only, such that $N\left(\delta, \mathfrak{F}^{\Sigma}_{a,\delta}\right) \leq a^p |\Sigma|^{-p/K}$.*

Proof. For $\delta \in (0,1)$ define a sequence $(r_i)_{i\geq 0}$ such that $r_0 = \left(\delta |\Sigma|^{1/2}/2\right)^{2/K}$ and $r_i = r_0 \exp\{i\delta/K\}$. From this sequence define a countable partition $(A_i)_{i\geq 0}$ of $(0,\infty)$ as follows: $A_0 = \{v : 0 < v \leq r_0\}$, and $A_i = \{v : r_{i-1} < v \leq r_i\}$ for $i > 0$. Then, we will need the following estimate for $v \in A_0$:

$$\begin{aligned}
\mathrm{I} &:= |\Sigma|^{-1/2} \int_{\mathbb{R}^K} \left| v^{K/2}\phi\left(v^{1/2}\Sigma^{-1/2}x\right) - r_0^{K/2}\phi\left(r_0^{1/2}\Sigma^{-1/2}x\right)\right| dx \\
&\leq |\Sigma|^{-1/2} \sup_{x\in\mathbb{R}^K} \left| v^{K/2}\phi\left(v^{1/2}\Sigma^{-1/2}x\right) + r_0^{K/2}\phi\left(r_0^{1/2}\Sigma^{-1/2}x\right)\right| \\
&\leq \frac{2r_0^{K/2}}{|\Sigma|^{1/2}}
\end{aligned}$$

by definition of A_0, and for r_0 as defined above, $\mathrm{I} \leq \delta$. We also need an estimate when $v \in A_{i+1}$ for $i \geq 0$. In this case, by Taylor expansion with integral reminder, setting $r(\tau) = r_i + \tau(v - r_i)$ with $\tau \in [0,1]$, we have the first equality in the next display

$$\begin{aligned}
\mathrm{II} &:= |\Sigma|^{-1/2} \int_{\mathbb{R}^K} \left| v^{K/2}\phi\left(v^{1/2}\Sigma^{-1/2}x\right) - r_i^{K/2}\phi\left(r_i^{1/2}\Sigma^{-1/2}x\right)\right| dx \\
&= |\Sigma|^{-1/2} \int_{\mathbb{R}^K} \left| \int_0^1 r(\tau)^{K/2}\,\phi\left(r(\tau)^{1/2}\Sigma^{-1/2}x\right)\left(\frac{K}{2r(\tau)} - \frac{x'\Sigma^{-1}x}{2}\right)(v - r_i)\,d\tau \right| dx \\
&\leq |\Sigma|^{-1/2}(v - r_i) \int_{\mathbb{R}^K} \int_0^1 r(\tau)^{K/2}\,\phi\left(r(\tau)^{1/2}\Sigma^{-1/2}x\right)\left(\frac{K}{2r(\tau)} + \frac{x'\Sigma^{-1}x}{2}\right) d\tau dx \\
&\quad \text{[because } \Sigma \text{ is positive definite and } r_i \leq v \in A_{i+1}] \\
&\leq |\Sigma|^{-1/2}(v - r_i) \int_0^1 \int_{\mathbb{R}^K} r(\tau)^{K/2}\,\phi\left(r(\tau)^{1/2}\Sigma^{-1/2}x\right)\left(\frac{K}{2r(\tau)} + \frac{x'\Sigma^{-1}x}{2}\right) dx d\tau \\
&\quad \text{[by Fubini's Theorem because the double integral is finite]} \\
&= (v - r_i)\int_0^1 \frac{K}{r(\tau)} d\tau \\
&\quad \text{[performing integration with the change of variable } r(\tau)^{1/2}\Sigma^{-1/2}x \mapsto y] \\
&= K \ln\left(\frac{v}{r_i}\right),
\end{aligned}$$

by direct integration and algebraic simplification. Hence, when $v \in A_{i+1}$ with $i \geq 0$, we have II $\leq \delta$. Using the same arguments as in Ghosal et al. (1999, proof of Lemma 1, p. 157) together with the estimates I and II, we deduce that we can find a 2δ-cover of $\mathfrak{F}_a^{\Sigma}$ consisting of discrete probabilities with atoms at $(r_i)_{i\in\{0,\dots,I\}}$ where I is the smallest integer greater or equal than $\min\{i>0 : r_i > a\}$ and we can choose

$$I = 1 + \left\lfloor \frac{K}{\delta} \ln\left(\frac{a\left|\Sigma\right|^{-1/K}}{(\delta/2)^{2K}}\right)\right\rfloor,$$

$\lfloor x \rfloor$ is the integer part of x. By Ghosal et al. (1999, p.157, see also Barron et al., 1999) for the set $\mathcal{P}_I$ of discrete probabilities with I atoms $N(\delta, \mathcal{P}_I) \leq \exp\left\{\left(1+\ln\frac{1+\delta}{\delta}\right)I\right\}$. By these remarks, $N\left(2\delta, \mathfrak{F}_a^{\Sigma}\right) \leq N(\delta, \mathcal{P}_I) \leq a^p \left|\Sigma\right|^{-p/K}$ for some some $p<\infty$ depending only on K and $\delta > 0$. Moreover, following the proof of Lemma 2 in Ghosal et al. (1999), we also deduce that $N\left(2\delta, \mathfrak{F}_{a,\delta}^{\Sigma}\right) \leq N\left(\delta, \mathfrak{F}_a^{\Sigma}\right)$ and the lemma is proved because $\delta \in (0,1)$ is arbitrary. ■

The next two lemmata will be used in the proof of Lemma 22 and the reader may wish to skip them and look at them while reading the proof of Lemma 22.

Lemma 20 *Fix an arbitrary $\delta \in (0,1)$. Let $A \subset \mathcal{C}$ be a set such that for any $\Sigma \in A$, there are constants $s_k^{(1)}, s_k^{(2)}$ satisfying*

$$[1-(\delta/4)]^{2/K} \leq s_k^{(1)}/s_k^{(2)} \tag{15}$$

such that $s_k^{(1)} \leq \lambda_k\left(\Sigma^{-1}\right) \leq s_k^{(2)}$, $k=1,\dots,K$. Moreover, if $\Sigma_1, \Sigma_2 \in A$, their matrices of orthonormal eigenvectors D_1, D_2 are assumed to satisfy $\left|D_1 - D_2\right|_\infty \leq \delta^2/\left(4K^3 \max_k s_k^{(2)}\right)$. Then, for any $f_1, f_2 \in \{\mathfrak{F}_P \cap A\} = \{\mathfrak{F}_P : \Sigma \in A\}$,

$$\int_{\mathbb{R}^K} \left|f_1(x) - f_2(x)\right| dx \leq \delta.$$

Proof. For any two matrices $\Sigma_1, \Sigma_3 \in A$, let $\Sigma_2 \in A$ be a matrix similar to Σ_1 and with same eigenvectors as Σ_3, i.e.

$$\Sigma_1^{-1} = D_1\Lambda_1 D_1',\ \Sigma_2^{-1} = D_2\Lambda_1 D_2',\ \Sigma_3^{-1} = D_2\Lambda_2 D_2' \tag{16}$$

where D_1 and D_2 are matrices of orthonormal eigenvectors. For $r = 1, 2, 3$, define

$$f_r(x|v) = \frac{v^{K/2}}{\left|\Sigma_r\right|^{1/2}}\phi\left(v^{1/2}\Sigma_r^{-1/2}x\right)$$

where $\Sigma_1, \Sigma_2, \Sigma_3$ are as in (16). We claim that for any $v \geq 0$,

$$\int_{\mathbb{R}^K} \left|f_1(x|v) - f_3(x|v)\right| dx \leq \delta. \tag{17}$$

Assuming for the moment that (17) holds, then for any $\Sigma_1, \Sigma_3 \in A$, we deduce the following bound using the triangle inequality and Fubini's Theorem, which can be applied because the double integral below is finite,

$$
\begin{aligned}
\int_{\mathbb{R}^K}\left|\int_0^{\infty} f_1(x|v)\,P(dv)-\int_0^{\infty} f_3(x|v)\,P(dv)\right|dx &\leq \int_0^{\infty}\int_{\mathbb{R}^K}\left|f_1(x|v)-f_3(x|v)\right|dx P(dv)\\
&\quad \text{[by Jensen Inequality and Fubini's Theorem]}\\
&\leq \delta\int_0^{\infty} P(dv)\\
&\quad \text{[by (17)]}\\
&= \delta
\end{aligned}
$$

and the lemma is proved. Hence, it remains to show that (17) holds. If $v=0$, (17) is obvious, as the densities are both zero. Hence, we assume $v>0$. Then, by the triangle inequality,

$$
\begin{aligned}
\int_{\mathbb{R}^K}\left|f_1(x|v)-f_3(x|v)\right|dx &\leq \int_{\mathbb{R}^K}\left|f_1(x|v)-f_2(x|v)\right|dx+\int_{\mathbb{R}^K}\left|f_2(x|v)-f_3(x|v)\right|dx\\
&= \mathrm{I}+\mathrm{II}.
\end{aligned}
$$

We shall control each term separately.

Control over I.

The eigenvectors are orthonormal, hence their entries are bounded by one in absolute value. By this remark derive the following inequality,

$$
\begin{aligned}
\left|\Sigma_1^{-1}-\Sigma_2^{-1}\right|_{\infty} &= \left|D_1\Lambda_1(D_1-D_2)+(D_1-D_2)\Lambda_1 D_2\right|_{\infty}\\
&\leq 2K\max_{1\leq k\leq K}\lambda_k(\Lambda_1)\left|D_1-D_2\right|_{\infty}\\
&\quad \text{[by the previous remark about the eigenvectors]}\\
&\leq 2K\max_{1\leq k\leq K}s_k^{(2)}\left|D_1-D_2\right|_{\infty}\\
&\quad \text{[by the eigenvalues for matrices in } A\text{]}\\
&\leq \frac{1}{2K^2}\delta^2 \qquad (18)
\end{aligned}
$$

by the eigenvectors for matrices in A. Then,

$$\begin{aligned}
\mathrm{I}^2 &\leq \int_{\mathbb{R}^K} |\Sigma_1|^{-1/2} v^{K/2} \phi\left(v^{1/2}\Sigma_1^{-1/2}x\right) \ln \frac{|\Sigma_1|^{-1/2} v^{K/2}\phi\left(v^{1/2}\Sigma_1^{-1/2}x\right)}{|\Sigma_2|^{-1/2} v^{K/2}\phi\left(v^{1/2}\Sigma_2^{1/2}x\right)} dx \\
&\quad \text{[by (7)]} \\
&= \int_{\mathbb{R}^K} |\Sigma_1|^{-1/2} v^{K/2}\phi\left(v^{1/2}\Sigma_1^{-1/2}x\right)\left[-\frac{1}{2}vx'\left(\Sigma_1^{-1}-\Sigma_2^{-1}\right)x\right]dx + \ln\frac{|\Sigma_1|^{-1/2}}{|\Sigma_2|^{-1/2}} \\
&= \int_{\mathbb{R}^K} \phi(y)\left[-\frac{1}{2}\left(\Sigma_1^{1/2}y\right)'\left(\Sigma_1^{-1}-\Sigma_2^{-1}\right)\Sigma_1^{1/2}y\right]dy \\
&\quad \text{[by the change of variables } v^{1/2}\Sigma_1^{-1/2}x \mapsto y \text{ and noting that } \Sigma_1, \Sigma_2 \text{ have same eigenvalues]} \\
&\leq \frac{K^2}{2}\left|\Sigma_1^{-1}-\Sigma_2^{-1}\right|_\infty \\
&\quad \text{[computing the integral noting that } \left|\Sigma_1^{1/2}\right|_\infty \leq 1\text{: } \Sigma_1 \text{ is a correlation matrix]} \\
&\leq \delta^2/4
\end{aligned}$$

by (18) so that $\mathrm{I} \leq \delta/2$.

Control over II.

Define S_1 as the matrix with eigenvalues equal to $\left\{s_k^{(1)}, k=1,...,K\right\}$ and matrix of eigenvectors equal to D_2: $S_1 = D_2\Lambda(S_1)D_2'$. Define S_2 similarly, but with eigenvalues $\left\{s_k^{(2)}, k=1,...,K\right\}$. For $\Sigma_2, \Sigma_3 \in A$ as in (16), $\left(S_2 - \Sigma_2^{-1}\right)$, $\left(S_2 - \Sigma_3^{-1}\right)$ are positive definite. Moreover, $|S_1| \leq |\Sigma_2|^{-1}$ and $|S_1| \leq |\Sigma_3|^{-1}$. These two remarks imply

$$|\Sigma_3|^{-1/2} v^{K/2}\phi\left(v^{1/2}\Sigma_3^{-1/2}x\right) \geq |S_1|^{1/2} v^{K/2}\phi\left(v^{1/2}S_2^{1/2}x\right)$$

and

$$|\Sigma_2|^{-1/2} v^{K/2}\phi\left(v^{1/2}\Sigma_2^{-1/2}x\right) \geq |S_1|^{1/2} v^{K/2}\phi\left(v^{1/2}S_2^{1/2}x\right).$$

By the last displays, and the triangle inequality,

$$\begin{aligned}
\mathrm{II} &\leq \int_{\mathbb{R}^K}\left|f_2(x|v) - |S_1|^{1/2} v^{K/2}\phi\left(v^{1/2}S_2^{1/2}x\right)\right|dx \\
&\quad + \int_{\mathbb{R}^K}\left|f_3(x|v) - |S_1|^{1/2} v^{K/2}\phi\left(v^{1/2}S_2^{1/2}x\right)\right|dx \\
&= \int_{\mathbb{R}^K}\left[f_3(x|v) - |S_1|^{1/2} v^{K/2}\phi\left(v^{1/2}S_2^{1/2}x\right)\right]dx \\
&\quad + \int_{\mathbb{R}^K}\left[f_3(x|v) - |S_1|^{1/2} v^{K/2}\phi\left(v^{1/2}S_2^{1/2}x\right)\right]dx \\
&= 2\left(1 - \frac{|S_1|^{1/2}}{|S_2|^{1/2}}\right) \\
&= \delta/2 \qquad (19)
\end{aligned}$$

because the determinant is the product of the eigenvalues, which are chosen to satisfy (15) in the statement of the lemma. Putting together I and II proves (17) as required.

■

Lemma 21 *Under Condition 3, if* $\Sigma \in \mathcal{C}$ *then* $\min_k \lambda_k \left(\Sigma^{-1}\right) \geq 1/K$, $|\Sigma|^{-1} > K^{-K}$, *and if* $|\Sigma|^{-1} \leq M$, *then* $\max_k \lambda_k \left(\Sigma^{-1}\right) \leq MK^{K-1}$.

Proof. The sum of the eigenvalues of a K dimensional full rank correlation matrix is K and these eigenvalues must be all positive. Hence, the largest eigenvalue of Σ is bounded above by K implying that the smallest eigenvalue of Σ^{-1} is bounded below by K^{-1}. Since the determinant of a matrix is the product of its eigenvalues, we deduce that $|\Sigma|^{-1} > \prod_{k=1}^{K} K^{-1}$. Finally, if the smallest eigenvalue of Σ^{-1} is $1/K$, the largest eigenvalue of Σ^{-1} must be smaller or equal than MK^{K-1} if we want $|\Sigma|^{-1} \leq M$. ■

Lemma 22 *Let* $(M_i)_{i\geq 0}$ *be any strictly increasing sequence such that* $M_0 = K^{-K}$. *Then,* $\mathfrak{F}_P \subseteq \bigcup_{i>0} \mathfrak{F}_P^{M_{i-1},M_i}$ *and*

$$N\left(\delta, \mathfrak{F}_P^M\right) \lesssim \ln^K (M) M^{K-1}$$

for any $P \in \mathcal{P}$ *and* $M > 0$.

Proof. By definition of $(M_i)_{i\geq 0}$, Lemma 21 implies that $\mathfrak{F}_P \subseteq \bigcup_{i>0} \mathfrak{F}_P^{M_{i-1},M_i}$ and the first part is proved. To show the cardinality of a δ cover of $\mathfrak{F}_P^M$ we shall eventually use Lemma 20. To this end, for arbitrary, but fixed $\delta \in (0,1)$ define the sequences $\left(s_{j(k)}\right)_{j(k)\geq 0}$ where $s_0 = 1/K$, $s_{j(k)} := s_0 \left[1 - (\delta/4)\right]^{-2j(k)/K}$, $k = 1, ..., K$. Let

$$B_j := \left\{\Sigma \in \mathcal{C} : s_{j(k)} \leq \lambda_k \left(\Sigma^{-1}\right) < s_{j(k)+1}, k = 1, ..., K\right\}. \tag{20}$$

Note that $(B_j)_{j\in\mathbb{N}^K}$ is a countable cover of $\mathcal{C}$ because, as mentioned above, the smallest eigenvalue of Σ^{-1} is bounded below by $s_0 = K^{-1}$. Let $\left(E_r^{B_j}\right)_{r>0}$ be a countable cover for B_j ($j \in \mathbb{N}^K$) such that if $\Sigma_1, \Sigma_2 \in E_r^{B_j}$, their orthonormal matrices of eigenvectors D_1, D_2 satisfy $|D_1 - D_2|_\infty \leq \delta^2 / \left(4K^3 \max_k s_{j(k)+1}\right) =: \delta'$. We claim that there is a finite $N_j = N(B_j)$ such that $B_j = \bigcup_{r=1}^{N_j} E_r^{B_j}$, implying that $\left(E_r^{B_j}\right)_{r\in\{1,...,N_j\}}$ is a finite cover of B_j. To find this N_j, consider the set of orthonormal matrices

$$\mathcal{E} := \left\{D \in \mathbb{R}^{K\times K} : DD' = D'D = I_K\right\}$$

where I_K is the K dimensional identity matrix. For each given matrix of eigenvalues, the set of possible eigenvectors of a correlation matrix is a strict subset of $\mathcal{E}$. To see this, recall that the eigenvectors need also to satisfies the constraint that the diagonal elements of the correlation matrix are all one. Moreover, given any K dimensional vector e, there exist only 2^{K-1} possible sets of orthonormal vectors orthogonal to e. This is clear for $K = 2$ and for $K > 2$ it follows by induction. Hence, the cardinality of a δ' cover for the set of unit vectors in $\mathbb{R}^K$ is proportional to the cardinality of a δ' cover for $\mathcal{E}$. Given that a unit vector in $\mathbb{R}^K$ is a point on the unit dimensional sphere in $\mathbb{R}^K$, it is sufficient to find the cardinality of a δ' cover for the the surface of the unit sphere in $\mathbb{R}^K$, proportionally equivalent to the cardinality of a δ' cover for $[-1,1]^{K-1}$. By these remarks it follows that $N_j \asymp (1/\delta')^{K-1} \asymp \left(\max_k s_{j(k)+1}/\delta^2\right)^{(K-1)}$ when $\left(E_r^{B_j}\right)_{r>0}$ is as defined above; for ease of notation we have partially suppressed dependence of N_j on K hence on s_0 because finite and not required for the final result.

Proven that $N_j < \infty$, we then claim that there is a finite integer J depending on M only, such that $\left\{\Sigma \in \mathcal{C} : |\Sigma|^{-1} \leq M\right\}$ can be covered by $\left\{\left(E_r^{B_j}\right)_{r\leq N_j} : j \leq (J, ..., J)\right\}$. Choose J as the smallest integer such that $J \geq \min\left\{j(k) > 0 : s_{j(k)+1} \geq M/s_0^{K-1}\right\}$ for any $k = 1, ..., K$. In fact, by Lemma 21 if $|\Sigma|^{-1} \leq M$ then $\max_k s_{j(k)+1} \leq M/s_0^{K-1} = MK^{K-1}$. Hence,

$$J = 1 + \left\lfloor \frac{K(\ln M + K \ln K)}{2[\ln 4 - \ln(4-\delta)]} \right\rfloor.$$

Therefore, our cover has cardinality

$$\#\left\{\left(E_r^{B_j}\right)_{r\leq N_j} : j \leq (J, ..., J)\right\} < J^K N_J \asymp \ln^K(M) M^{K-1}$$

using the fact that $\max_{j\leq(J,...,J)} \max_k s_{j(k)+1} \asymp M$ suppressing dependence on δ and K. To finish the proof, note that $\mathfrak{F}_P^M \subseteq \bigcup_{j\leq(J,...,J)} \bigcup_{r\leq N_j} \left(\mathfrak{F}_P \cap E_r^{B_j}\right)$ and that $E_r^{B_j}$ satisfies the conditions of Lemma 20 (i.e. set $E_r^{B_j} = A$ where A is as in Lemma 20). Hence, $\left\{\left(\mathfrak{F}_P \cap E_r^{B_j}\right)_{r\leq N_j} : j \leq (J, ..., J)\right\}$ is a δ-cover in total variation for $\mathfrak{F}_P^M$. ■

We can now prove Lemma 18.

Proof Lemma 18. The distances in (7) are all invariant w.r.t. the dominating measure. Hence, map a Gaussian copula density to a Gaussian density (see proof of Lemma 15). We show that we can find a δ-cover of $\mathfrak{F}$ satisfying the statement of the lemma. Let $(a_i)_{i>0}$ and $(M_j)_{j>0}$ be sequences of strictly increasing positive numbers. Note that

$$\mathfrak{F} \subseteq \bigcup_{i,j>0} \mathfrak{F}_{a_i,\delta}^{M_{j-1},M_j}$$

and

$$\mathfrak{F}_{a,\delta}^{m,M} = \left(\bigcup_{P:P((0,a])>1-\delta}\right)\left(\bigcup_{\Sigma:m<|\Sigma|^{-1}\leq M}\right)\left\{\int_0^\infty v^{K/2}\phi\left(v^{1/2}\Sigma^{-1/2}x\right)P(dv)\right\},$$

where, by Lemmata 19 and 22, and the triangle inequality,

$$N\left(\delta, \mathfrak{F}_{a,\delta}^{m,M}\right) \lesssim a^p M^{(p/K)+K-1} \ln^K(M) \tag{21}$$

for some positive finite p. Define the following sequence of sets $(\mathfrak{G}_{i,j,\delta})_{i,j>0}$

$$\mathfrak{G}_{i,j,\delta} := \left\{\int_0^\infty v^{K/2}\phi\left(v^{1/2}\Sigma^{-1/2}x\right)P(dv) : P((0,a_{i-1}]) \leq 1-\delta, P((0,a_i]) > 1-\delta, M_{j-1} < |\Sigma|^{-1} \leq M_j\right\}.$$

Mutatis mutandis, by the arguments in Lijoi et al. (2005, p.1295), for any $\delta \in (0,1)$, there is an integer N large enough such that for $i > 0$, $N(\delta, \mathfrak{G}_{i,j,\delta}) \leq N\left(\delta, \mathfrak{F}_{a_N,\delta}^{M_{j-1},M_j}\right) \lesssim a_N^p M_j^{(p/K)+K-1} \ln^K(M_j)$ using (21) for the right hand term. Then, note that $\mathfrak{F} \subseteq$

$\bigcup_{i,j>0}\mathfrak{G}_{i,j,\delta}$. Hence, there exists a countable δ-cover $(A_s)_{s>0}$ of A_ϵ such that, for some $\beta \in (0,1)$,

$$
\begin{aligned}
& \sum_{s>0}\Pi^{\beta}(A_s) \\
\leq\; & \sum_{i,j>0} a_N^p M_j^{(p/K)+K-1}\ln^K(M_j)\,\Pi^{\beta}(\mathfrak{G}_{i,j,\delta}) \\
\leq\; & \sum_{i,j>0} a_N^p M_j^{(p/K)+K-1}\ln^K(M_j)\,\Pi_P^{\beta}\left(P([a_{i-i},\infty))>\delta\right)\Pi_\Sigma^{\beta}\left(|\Sigma|^{-1}>M_{j-1}\right) \\
& \text{[by independence of } P \text{ and } \Sigma \text{ and obvious set inequalities]} \\
& \sum_{i>0} a_N^p \frac{\Pi_P^{\beta}\left(P([a_{i-i},\infty))\right)}{\delta^{\beta}}\sum_{j>0} M_j^{(p/K)+K-1}\ln^K(M_j)\,\Pi_\Sigma^{\beta}\left(|\Sigma|^{-1}>M_{j-1}\right) \\
& \text{[by Markov inequality for } P\text{]} \\
<\; & \infty
\end{aligned}
$$

by Condition 4. ■

3.2 Proof of Theorem 12

Proof of Theorem 12. By Lemma 23 (below), the set $\{\omega\in\Omega : \Pi(B_\delta(\omega))=1\}$, where $B_\delta(\omega)$ is as in (22), has $\mathbb{P}$-probability one for any $\delta>0$. Hence, multiplying and dividing by $c_\theta(U_i)$,

$$
\begin{aligned}
\hat{\Pi}_n(A_\epsilon) &= \frac{\int_{A_\epsilon}\left[\prod_{i=1}^n c_\theta\left(\hat{U}_i\right)/c_\theta(U_i)\right]\left[\prod_{i=1}^n c_\theta(U_i)\right]\Pi(d\theta)}{\int_{\Theta}\left[\prod_{i=1}^n c_\theta\left(\hat{U}_i\right)/c_\theta(U_i)\right]\left[\prod_{i=1}^n c_\theta(U_i)\right]\Pi(d\theta)} \\
&= \frac{\int_{A_\epsilon\cap B_\delta(\omega)}\left[\prod_{i=1}^n c_\theta\left(\hat{U}_i\right)/c_\theta(U_i)\right]\left[\prod_{i=1}^n c_\theta(U_i)\right]\Pi(d\theta)}{\int_{\Theta\cap B_\delta(\omega)}\left[\prod_{i=1}^n c_\theta\left(\hat{U}_i\right)/c_\theta(U_i)\right]\left[\prod_{i=1}^n c_\theta(U_i)\right]\Pi(d\theta)} \\
&\quad \text{[by the previous remarks about } B_\delta(\omega)\text{]} \\
&\leq \frac{\sup_{\theta\in B_\delta(\omega)}\left[\prod_{i=1}^n c_\theta\left(\hat{U}_i\right)/c_\theta(U_i)\right]}{\inf_{\theta\in B_\delta(\omega)}\left[\prod_{i=1}^n c_\theta\left(\hat{U}_i\right)/c_\theta(U_i)\right]} \\
&\quad \times\frac{\int_{A_\epsilon}\prod_{i=1}^n c_\theta(U_i)\,\Pi(d\theta)}{\int_{\Theta}\prod_{i=1}^n c_\theta(U_i)\,\Pi(d\theta)} \\
&\leq \exp\{2n\delta\}\,\Pi_n(A_\epsilon)
\end{aligned}
$$

$\mathbb{P}$-a.s. by the remarks about $B_\delta(\omega)$ for all but finitely many n. By Theorem 5 and Walker (2004, proof of Theorem 4, p. 2036), deduce that $\Pi_n(A_\epsilon)\lesssim\exp\{-n\gamma\}$ a.s. for some $\gamma>0$. Then, choose $2\delta<\gamma$ in the definition of $B_\delta(\omega)$. ■

Lemma 23 *Under Conditions 10 and 11, for any $\delta>0$,*

$$\mathbb{P}\left(\{\omega\in\Omega : \Pi(B_\delta(\omega))=1\}\right)=1,$$

where

$$B_{\delta}(\omega) := \left\{ \theta \in \Theta : \lim_{n\to\infty} \left[\prod_{i=1}^{n} c_{\theta}\left(\hat{U}_i(\omega)\right) / c_{\theta}(U_i(\omega)) \right]^{1/n} \in [\exp\{-\delta\}, \exp\{\delta\}] \right\}. \quad (22)$$

Proof. We argue along the lines of Lemma 3 in Barron et al. (1999). Let

$$B_{\delta} := \left\{ \omega \in \Omega, \theta \in \Theta : \lim_{n\to\infty} \left[\prod_{i=1}^{n} c_{\theta}\left(\hat{U}_i(\omega)\right) / c_{\theta}(U_i(\omega)) \right]^{1/n} \in [\exp\{-\delta\}, \exp\{\delta\}] \right\}.$$

Obvious readaptations of Lemma 11 and Lemma 10 in Barron et al. (1999), using Condition 11 imply that $\frac{1}{n}\sum_{i=1}^{n} \ln\left(c_{\theta}\left(\hat{U}_i(\omega)\right)/c_{\theta}(U_i(\omega))\right)$ is a measurable function from $\Omega \times \Theta$ to $\mathbb{R}$. A continuous function of a measurable function is measurable, then $\left[\prod_{i=1}^{n} c_{\theta}\left(\hat{U}_i(\omega)\right)/c_{\theta}(U_i(\omega))\right]^{1/n}$ is also measurable. Then, the set B_{δ} is measurable and we can apply Fubini's Theorem. Let $B_{\delta}(\theta)$ be as in (23) below. By Lemma 24, $\mathbb{P}(B_{\delta}(\theta)) = 1$ for every $\theta \in \Theta$ and integrating both sides of this equality w.r.t. Π,

$$\begin{aligned} 1 &= \int_{\Theta} \mathbb{P}(B_{\delta}(\theta))\,\Pi(d\theta) \\ &= \int_{\Omega} \Pi(B_{\delta}(\omega))\,\mathbb{P}(d\omega) \end{aligned}$$

by Fubini's Theorem. Since Π and $\mathbb{P}$ have range $[0,1]$, the last display implies the statement of the lemma. ■

Lemma 24 *For any* $\delta > 0$,

$$\mathbb{P}(B_{\delta}(\theta)) = 1,$$

where

$$B_{\delta}(\theta) := \left\{ \omega \in \Omega : \lim_{n\to\infty} \left[\prod_{i=1}^{n} \frac{c_{\theta}\left(\hat{U}_i(\omega)\right)}{c_{\theta}(U_i(\omega))} \right]^{1/n} \in [\exp\{-\delta\}, \exp\{\delta\}] \right\}. \quad (23)$$

Proof. Define $N := N_n \asymp \ln n$. Then,

$$\begin{aligned} \left[\prod_{i=1}^{n} \frac{c_{\theta}\left(\hat{U}_i\right)}{c_{\theta}(U_i)} \right]^{1/n} &= \exp\left\{ \frac{1}{n} \left(\sum_{i\le N} + \sum_{i>N} \right) \ln \frac{c_{\theta}\left(\hat{U}_i\right)}{c_{\theta}(U_i)} \right\} \\ &= \exp\{\mathrm{I} + \mathrm{II}\} \end{aligned}$$

and we shall upper and lowerbound each term separately. To this end, there is an $\alpha > 0$ such that

$$\begin{aligned}
\mathbb{P}(|\mathrm{I}| > \epsilon) &\leq \frac{N^{1+\alpha}}{(\epsilon n)^{1+\alpha}} \mathbb{E}\left|\frac{1}{N}\sum_{i\leq N} \ln \frac{c_\theta\left(\hat{U}_i\right)}{c_\theta(U_i)}\right|^{1+\alpha} \\
&\leq \frac{N^{1+\alpha}}{(\epsilon n)^{1+\alpha}} \sup_{i>0} \mathbb{E}\left|\ln \frac{c_\theta\left(\hat{U}_i\right)}{c_\theta(U_i)}\right|^{1+\alpha} \\
&\to 0.
\end{aligned}$$

using Condition 11. Since, by definition of N, $(N/n)^{1+\alpha}$ is summable in n, the Borel-Cantelli lemma gives $|\mathrm{I}| \leq \epsilon$, $\mathbb{P}$-a.s.. Finally,

$$\begin{aligned}
|\mathrm{II}| &\leq \lim_{N\to\infty} \sup_{i>N} \left|\ln c_\theta\left(\hat{U}_i(\omega)\right) - \ln c_\theta(U_i(\omega))\right| \\
&\to 0
\end{aligned}$$

$\mathbb{P}$-a.s. by Condition 10, using the Continuous Mapping Theorem as the copula density is continuous in its argument u for any θ. Since $\epsilon > 0$ is arbitrary, the lemma follows. ■

3.3 Proof of Remaining Results

Proof of Lemma 8. Note that

$$D\Sigma D = \sigma^2 I_K + \gamma\gamma'$$

and we shall find the eigenvalues of $D\Sigma D$. Note that D is just the diagonal matrix of standard deviations, so that decomposition of $D\Sigma D$ allows us to recover all the required information for Σ. Hence, by direct calculation $\left(\sigma^2 I_K + \gamma\gamma'\right)\Gamma_1 = \Lambda_{11}\Gamma_1$ confirming that Λ_{11} is an eigenvalue with eigenvector Γ_1 as stated in the lemma. Since $\gamma\gamma'$ has rank one, all the other eigenvalues are σ^2 with eigenvectors that only need to be orthogonal to γ and to each other. Hence, by Gram-Schmidt orthogonalization, the eigenvectors are found as given in the lemma for any initial basis $e_1, ..., e_{K-1}$. The determinant is directly found from the eigenvalues of $D\Sigma D$ using the fact that D is diagonal. ■

Proof of Lemma 9. From Lemma 8

$$\begin{aligned}
|\Sigma|^{-1} &\leq \max_{k\in\{1,...,K\}} \frac{\left(\sigma^2 + \gamma_k^2\right)^K}{\sigma^{2K} + \sigma^{2(K-1)}\gamma'\gamma} \\
&= \max_{k\in\{1,...,K\}} \frac{\sum_{i=0}^K \binom{K}{i}\sigma^{2(K-i)}\gamma_k^{2i}}{\sigma^{2K} + \sigma^{2(K-1)}\sum_{k=1}^K \gamma_k^2}.
\end{aligned}$$

Since the prior has support in $\mathcal{C}_a(\sigma,\gamma)$, we have $\sigma \in [a,\infty]$. If $\sigma \to \infty$, $|\Sigma|^{-1} \to 1$, and if $\sigma = a$ the magnitude of $|\Sigma|^{-1}$ is determined by $|\gamma| \to \infty$, because for small γ, $|\Sigma|^{-1}$ is small. Hence,

$$|\Sigma|^{-1} \leq \max_{k\in\{1,...,K\}} \frac{1 + 2^K\gamma_k^{2K}}{\min\left(a^{2K}, 1\right)},$$

by some bounding arguments. Let σ_γ^2 be the largest variance of the entries in γ. Then, γ^2/σ_γ^2 is a non central Chi-square. By these remarks, as $M \to \infty$,

$$\begin{aligned}
\Pr\left(|\Sigma|^{-1} > M\right) &\leq K \max_{k\in\{1,\ldots,K\}} \Pr\left(\frac{1+2^K\gamma_k^{2K}}{\min\left(a^{2K},1\right)} > M\right) \\
&\leq K \max_{k\in\{1,\ldots,K\}} \Pr\left(\left(\frac{\gamma_k}{\sigma_\gamma}\right)^2 > \frac{\left(2^{-K}\min\left(a^{2K},1\right)M-1\right)^{1/K}}{\sigma_\gamma^2}\right) \\
&\lesssim \exp\left\{-\alpha\frac{\left(2^{-K}\min\left(a^{2K},1\right)M\right)^{1/K}}{\sigma_\gamma^2}\right\}
\end{aligned}$$

for any $\alpha \in (0,1/2)$ using finiteness of the moment generating function of γ^2/σ_γ^2 and the exponential Markov inequality. Condition $4(ii.)$ is then satisfied. ■

Proof of Lemma 13. Define

$$f(x|v) = v^{K/2}\phi\left(v^{1/2}\Sigma^{-1/2}x\right).$$

Let $\mathcal{Q}_\phi$ be as in (1) using the standard normal quantile function. By change of variables and Minkowski inequality,

$$\begin{aligned}
\mathbb{E}\left|\ln\frac{c_\theta\left(\hat{U}_i(\omega)\right)}{c_\theta\left(U_i(\omega)\right)}\right|^{1+\alpha} &= \mathbb{E}\left|\ln\frac{\int_0^\infty f\left(\mathcal{Q}_\phi\left(\hat{U}_i(\omega)\right)|v\right)P(dv)}{\int_0^\infty f\left(\mathcal{Q}_\phi\left(U_i(\omega)\right)|v\right)P(dv)}\right|^{1+\alpha} \\
&\quad +\mathbb{E}\left|\sum_{k=1}^K \ln\frac{\phi\left(\Phi^{-1}\left(U_{ik}(\omega)\right)\right)}{\phi\left(\Phi^{-1}\left(\hat{U}_{ik}(\omega)\right)\right)}\right|^{1+\alpha} \\
&= \mathrm{I}+\mathrm{II}.
\end{aligned}$$

For any $x, y > 0$, $|\ln(x+y)| \leq |\ln(x)| + y$. By this remark and by similar arguments as in Control over I in the proof of Lemma 15, for some finite a and $a_1 < a_2$ as in the statement of the lemma,

$$\begin{aligned}
\mathrm{I} &\leq \mathbb{E}\left|\ln\frac{\int_0^a f\left(\mathcal{Q}_\phi\left(\hat{U}_i(\omega)\right)|v\right)P(dv)}{\int_0^\infty f\left(\mathcal{Q}_\phi\left(U_i(\omega)\right)|v\right)P(dv)}\right|^{1+\alpha}+\mathbb{E}\left|\frac{\epsilon}{\int_0^\infty f\left(\mathcal{Q}_\phi\left(U_i(\omega)\right)|v\right)P(dv)}\right|^{1+\alpha}\\
&\leq \mathbb{E}\left|\ln\frac{\int_0^a f\left(\mathcal{Q}_\phi\left(\hat{U}_i(\omega)\right)|v\right)P(dv)}{\int_0^a f\left(\mathcal{Q}_\phi\left(U_i(\omega)\right)|v\right)P(dv)}\right|^{1+\alpha}+\mathbb{E}\sup_{v\in[a_1,a_2]}\left|\frac{\epsilon}{f\left(\mathcal{Q}_\phi\left(U_i(\omega)\right)|v\right)P([a_1,a_2])}\right|^{1+\alpha}\\
&\leq \mathbb{E}\sup_{v\in[0,a]}\left|\ln\frac{f\left(\mathcal{Q}_\phi\left(\hat{U}_i(\omega)\right)|v\right)}{f\left(\mathcal{Q}_\phi\left(U_i(\omega)\right)|v\right)}\right|^{1+\alpha}+\mathbb{E}\sup_{v\in[a_1,a_2]}\left|\frac{\epsilon}{f\left(\mathcal{Q}_\phi\left(U_i(\omega)\right)|v\right)P([a_1,a_2])}\right|^{1+\alpha}\\
&= \mathbb{E}\sup_{v\in[0,a]}\left(\frac{v}{2}\right)^{1+\alpha}\left|\mathcal{Q}_\phi(U_i)\Sigma^{-1}\mathcal{Q}_\phi(U_i)-\mathcal{Q}_\phi\left(\hat{U}_i\right)\Sigma^{-1}\mathcal{Q}_\phi\left(\hat{U}_i\right)\right|^{1+\alpha}\\
&\quad+\mathbb{E}\sup_{v\in[a_1,a_2]}\left|\epsilon\frac{\exp\left\{\frac{(1+\alpha)v}{2}\mathcal{Q}_\phi(U_i)\Sigma^{-1}\mathcal{Q}_\phi(U_i)\right\}}{v^{K/2}P([a_1,a_2])}\right|^{1+\alpha}\\
&\leq \left(\frac{a}{2}\right)^{1+\alpha}\left(\mathbb{E}\left|\mathcal{Q}_\phi(U_i)\Sigma^{-1}\mathcal{Q}_\phi(U_i)\right|^{1+\alpha}+\mathbb{E}\left|\mathcal{Q}_\phi\left(\hat{U}_i\right)\Sigma^{-1}\mathcal{Q}_\phi\left(\hat{U}_i\right)\right|^{1+\alpha}\right)\\
&\quad+\left|\frac{\epsilon}{a_1^{K/2}P([a_1,a_2])}\right|^{1+\alpha}\mathbb{E}\exp\left\{\frac{(1+\alpha)a_2}{2}\mathcal{Q}_\phi(U_i)\Sigma^{-1}\mathcal{Q}_\phi(U_i)\right\}.
\end{aligned}$$

Noting that $\mathcal{Q}_\phi(U_i)$ is a Gaussian random vector with covariance matrix Σ, it follows that the first expectation is finite. If Π_P is a Dirichlet process prior with base probability measure G such that $G([a_1,a_2])>0$, then, $\Pi_P(P[a_1,a_2]>0)=1$ (Ferguson, 1973, Proposition 1). Hence choose a_2 such that $(1+\alpha)a_2<1$. Then, also the third expectation is finite Π_P-a.s., as $(1+\alpha)a_2<1$, Π_P-a.s.. To bound the second expectation note that

$$\begin{aligned}
\mathbb{E}\left|\mathcal{Q}_\phi\left(\hat{U}_i\right)\Sigma^{-1}\mathcal{Q}_\phi\left(\hat{U}_i\right)\right|^{1+\alpha} &\lesssim \sum_{k=1}^K\mathbb{E}\left|\Phi^{-1}\left(\hat{U}_{ik}\right)\right|^{2(1+\alpha)}\\
&\lesssim \sum_{k=1}^K\mathbb{E}\left|\ln\left(1-\hat{U}_{ik}\right)\right|^{(1+\alpha)}
\end{aligned}$$

because for any $u\in[0,1]$, $\left|\Phi^{-1}(u)\right|\lesssim|\ln(1-u)|^{(1/2)}$. This term is finite by the conditions of the lemma. The bound for II follows by similar but simpler arguments. ■

References

[1] Barron A.R. (1998) Information-Theoretic Characterization of Bayes Performance and the Choice of Priors in Parametric and Nonparametric Problems. In J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds), Bayesian Statistics 6, 27-52. Oxford University Press.

[2] Barron, A., M.J. Schervish and L. Wasserman (1999) The Consistency of Posterior Distributions in Nonparametric Problems. Annals of Statistics 27, 536-561.

[3] Bollerslev, T. (1990) Modeling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH Model. Review of Economics and Statistics 72, 498-505.

[4] Bousquet, O. and M.K. Warmuth (2002) Tracking a Small Set of Experts by Mixing Past Posteriors. Journal of Machine Learning Research 3, 363-396.

[5] Dahlhaus, R. (1997) Fitting Time Series Models to Nonstationary Processes. Annals of Statistics 25, 1-37.

[6] Dawid, A.P. (1997) Prequential analysis. In S. Kotz, C.B. Read and D.L. Banks (eds.), Encyclopedia of Statistical Sciences Volume 1, 464-470. Wiley.

[7] Engle, R. (2002) Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. Journal of Business and Economic Statistics 20, 339-350.

[8] Embrechts, P., A. McNeil and D. Straumann (2002) Correlation and Dependence in Risk Management: Properties and Pitfalls In M.A.H. Dempster (ed.), Risk Management: Value at Risk and Beyond, 176-223. Cambridge: Cambridge University Press.

[9] Escobar, M.D. (1994) Estimating Normal Means with a Dirichlet Process Prior. Journal of the American Statistical Society 89, 268-277.

[10] Fang, H.-B., K.-T. Fang and S. Kotz (2002) The Meta-Elliptical Distributions with Given Marginals. Journal of Multivariate Analisis 82, 1-16.

[11] Ferguson, T.S. (1973) A Bayesian Analysis of Some Nonparametric Problems. Annals of Statistics 1, 209-230.

[12] Fermanian, J.D. and O. Scaillet (2003) Nonparametric Estimation of Copulas for Time Series. Journal of Risk 5, 25-54.

[13] Fermanian, J.-D., D. Radulovic and M. Wegkamp (2004) Weak Convergence of Empirical Copula Processes. Bernoulli 10, 847-860.

[14] Gagliardini, P. and C. Gouriéroux (2007) An Efficient Nonparametric Estimator for Models with Nonlinear Dependence. Journal of Econometrics 137, 189-229.

[15] Ghosal, S., J.K. Ghosh and R.V. Ramamoorthi (1999) Posterior Consistency of Dirichlet Mixtures in Density Estimation. Annals of Statistics 27, 143-158.

[16] Ghosal, S., J.K. Ghosh and A. van der Vaart (2000) Convergence Rates of Posterior Distributions. Annals of Statistics 28, 500-531.

[17] Ghosal, S. and A. van der Vaart (2007a) Convergence Rates of Posterior Distributions for Noniid Observations. Annals of Statistics 35, 192-223.

[18] Ghosal, S. and A. van der Vaart (2007b) Posterior Convergence Rates of Dirichlet Mixtures at Smooth Densities. Annals of Statistics 35, 697-723.

[19] Hoff, P.H. (2007) Extending the Rank Likelihood for Semiparametric Copula Estimation. Annals of Applied Statistics 1, 265-283.

[20] Hult, H. and F. Lindskog (2002) Multivariate Extremes, Aggregation and Dependence in Elliptical Distributions. Advances in Applied Probability 34, 587-608.

[21] Ishwaran, H. and L.F. James (2002) Approximate Dirichlet Process Computing in Finite Normal Mixtures: Smoothing and Prior Information. Journal of Computational and Graphical Statistics 11, 508-532.

[22] Lijoi, A., I. Prünster and S.G. Walker (2005) On Consistency of Nonparametric Normal Mixtures for Bayesian Density Estimation. Journal of the American Statistical Association 100, 1292-1296.

[23] Kleijn, B.J.K. and A. van der Vaart (2006) Misspecification in Infinite-Dimensional Bayesian Statistics. Annals of Statistics 34, 837-877.

[24] Mikosch, T. (2006) Copulas: Tales and Facts. Extremes 9, 3-20, and rejoinder, 55-62.

[25] Patton, A. (2006) Modelling Asymmetric Exchange Rate Dependence. International Economic Review 47, 527-556.

[26] Rio, E. (2000) Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants. Paris: Springer.

[27] Sancetta, A. and S.E. Satchell (2004) The Bernstein Copula and its Applications to Modeling and Approximations of Multivariate Distributions. Econometric Theory 20, 535-562.

[28] Sancetta, A. (2007) Nonparametric Estimation of Distributions with Given Marginals via Bernstein-Kantorovich Polynomials: L_1 and Pointwise Convergence Theory. Journal of Multivariate Analysis 98, 1376-1390.

[29] Sancetta (2007) Universality of Bayesian Predictions. Preprint: http://www.econ.cam.ac.uk/dae/repec/cam/pdf/cwpe0755.pdf.

[30] Scarsini, M. (1989) Copulae of Probability Measures on Product Spaces. Journal of Multivariate Analysis 31, 201-219.

[31] Schweizer, B. and A. Sklar (1983) Probabilistic Metric Spaces. New York: North-Holland Publishing Co.

[32] Sethuraman, J. (1994) A Constructive Definition of Dirichlet Priors. Statistica Sinica 4, 639-650.

[33] Walker, S. (2004a) New Approaches to Bayesian Consistency. Annals of Statistics 32, 2028-2043.

[34] Walker, S.G. (2004b) Modern Bayesian Asymptotics. Statistical Science 19, 111-117.

[35] Walker, S.G. (2005) Bayesian Nonparametric Inference. In D.K. Dey and C.R. Rao (eds.) Handbook of Statistics 25, 339-371. Amsterdam: Elsevier.

[36] Walker, S.G., A. Lijoi and I. Prünster (2007) On Rates of Convergence for Posterior Distributions in Infinite-Dimensional Models. Annals of Statistics 35, 738-746.