Nonparametric Estimation of Multivariate Distributions with Given Marginals: L_2 Theory

Alessio Sancetta

March 2003

CWPE 0320

Not to be quoted without permission

Nonparametric Estimation of Multivariate Distributions with Given Marginals: L_2 Theory*

Alessio Sancetta

Department of Applied Economics, Cambridge, UK

March 24, 2003

Abstract

Nonparametric estimation of the copula function using Bernstein polynomials is studied. Convergence in the uniform topology is established. From the nonparametric Bernstein copula, the nonparametric Bernstein copula density is derived. It is shown that the nonparametric Bernstein copula density is closely related to the histogram estimator, but has the smoothing properties of kernel estimators. The optimal order of polynomial under the L_2 norm is shown to be closely related to the inverse of the optimal smoothing factor for common nonparametric estimator. In order of magnitude, this estimator has variance equal to the square root of other common nonparametric estimators, e.g. kernel smoothers, but it is biased as a histogram estimator.

Keywords: Bernstein Polynomial, Copula, Curse of Dimensionality, Histogram, Nonparametric Estimator.

JEL Classification: C14, C51.

---

*This is Chapter 2 of my PhD dissertation. The first draft of this paper was written in April 2001. I would like to thank my supervisor Steve Satchell, and my examiners Andrew Harvey and Peter Robinson for helpful comments. Remaining errors are my sole responsibility.

1 Introduction

The use of multivariate distributions with given marginals (i.e. the copula function) have been considerably studied in the last thirty years, but mainly from a mathematical point of view with particular reference to probabilistic metric spaces (e.g. Schweizer and Skalr, 1983). More recently, the study of these distributions has received considerable attention from a statistical point of view, particularly in the case of parametric inference (e.g. Joe, 1997, and references therein). Some exceptions where nonparametric and semiparametric inference is considered are non parametric estimation of some extreme value copulae (Capéraà et al., 1997), the semiparametric estimation of the dependence parameter in parametric families of copulae (Genest et al., 1995), and nonparametric inference for choosing the best Archimedean copula (Genest and Rivest, 1993). Moreover, the need to model complex dependence structures has made the use of the copula an invaluable tool in many applied areas. Recent applications to finance and financial econometrics are given in Li (1999), Patton (2001), Rosenberg (2000), Longin and Solnik (2001), and, with more theoretical emphasis, Sancetta and Satchell (2001).

The use of the copula function allows us to divide the problem of marginal estimation from the one of the copula estimation, e.g. inference functions for marginals; see Joe (1997) and references therein. This provides a considerable advantage in applications as a researcher may fully exploit his limited information in the most efficient way. For example, there could be significant knowledge of the marginal distribution of two variables, while little confidence on their joint distribution. For reasons of this nature, it would be convenient to be able not only to estimate the marginals separately from the whole joint distribution, or copula, but also to find ways that require fewer assumptions on the side of the practitioner. Nonparametric techniques have this purpose in mind.

The aim of this paper is to provide a nonparametric estimation procedure that would allow us to model the copula function nonparametrically in complete independence from the marginal. In particular, we consider estimation via Bernstein polynomials. For this reason, we call our estimator the nonparametric Bernstein copula. This is just a nonparametric estimator for the general family of copulae defined as Bernstein copulae (e.g. Sancetta and Satchell, 2001, and Sancetta 2002). In particular, we will consider the Bernstein operator as a smoother where the coefficients are given by the empirical copula. This has the advantage of replacing a non-smooth distribution estimator with a smooth one. Therefore, simple operations like differentiation and integration can be carried out and even generalized for the case of non differentiable distribution functions.

The main contribution of this paper is to develop the L_2 theory for the nonparametric Bernstein copula density, i.e. convergence of the density of our estimator under the following norm

$$\|f\|_2^2 \equiv \int_{\Omega} |f|^2 \, d\mu,$$

for the following measure space $(\Omega, \mathcal{B}, \mu)$. This estimator does not only add to the large number of nonparametric estimators. It is shown that inside the support of the copula, the variance of this estimator is of lower order than other commonly used nonparametric estimators (e.g. histogram and kernel). The intuition for the result is that Bernstein polynomials do not have a local behaviour as other nonparametric or polynomial estimators; the more global behaviour comes at the expense of the well known slow rate of adjustment. In particular, the Bernstein operator is just the expectation operator with respect to a binomial distribution. In this respect the result represents a partial solution to the so called curse of dimensionality.

The plan for the paper is as follows. Section 2 briefly recalls the ideas behind the copula function and Bernstein polynomials. Section 3 considers the properties of the nonparametric Bernstein copula with given marginals. These marginals can have either a parametric or nonparametric form. Because of the properties of the copula, this is of no concern. Since the copula is defined on the k dimensional unit cube, it is natural to consider convergence under the sup norm. From the nonparametric Bernstein copula we derive and study the nonparametric Bernstein copula density in Section 4. The link between the nonparametric Bernstein copula and other methods of estimation (i.e. histogram and kernel) is established. The main result of convergence under the L_2 norm and weak convergence is stated. Section 5 provides a brief illustrative example of the nonparametric copula and one of its competing alternatives. To make the paper more readable, the proofs of the two theorems in the text are deferred to Section 6. Final remarks can be found in Section 7.

2 The Copula Function and the Bernstein Operator

Recall that if $X_1, ..., X_k$ have joint distribution H and one dimensional margins $F_1, ..., F_k$, their copula function, $C : [0,1]^k \to [0,1]$, is defined as

$$H(X_1, ..., X_k) = C\left(F_1(X_1), ..., F_k(X_k)\right),$$

i.e. it is the joint distribution of uniform $[0,1]$ marginals. An important property used in this paper is that the copula is unique if $F_1, ..., F_k$ are continuous (e.g. Sklar, 1973). Since we will use Bernstein polynomials to estimate the copula, we briefly recall a few facts about them. Let $C_{[0,1]^k}$ be the space of continuous bounded functions in $[0,1]^k$.[1] For any $f \in C_{[0,1]^k}$ its associated Bernstein polynomial is given by

$$(B_m^k f)(X) \equiv \sum_{v_1=0}^{m} ... \sum_{v_k=0}^{m} f\left(\frac{v_1}{m}, ..., \frac{v_k}{m}\right) P_{v_1,m}(x_1) \cdots P_{v_k,m}(x_k), \tag{1}$$

where

$$P_{v_j,m}(x_j) \equiv \binom{m}{v_j} x_j^{v_j}(1-x_j)^{m-v_j},$$

and B_m^k is the k dimensional Bernstein linear operator of order m. The summation in (1) does not need to run over $[0,m]$ for each coordinate, but may vary from one to another. To ease notation, we do not allow for this generality as it will not be

[1] Notice that any compact interval $[a,b]^k$ can be made isomorphic to the unit box.

relevant for our purposes. On the other hand, a property that will be used in the sequel is that Bernstein polynomials are dense in $C_{[0,1]^k}$ (e.g. Devore and Lorentz, 1993).

Before concluding this section we recall the following representation in terms of a Riemann Stieltjes integral of a one dimensional Bernstein polynomial [2],

$$
\begin{aligned}
(B_n f)(x) &\equiv \sum_{v}^{n} f\left(\frac{v}{n}\right) P_{v,n}(x) \\
&= \int_{0}^{1} f(t) d_t Kn(x,t),
\end{aligned} \tag{2}
$$

where

$$
\begin{aligned}
K_n(x,t) &\equiv \sum_{v \le nt} \binom{n}{v} x^v (1-x)^{n-v}, \\
K_n(x,0) &\equiv 0
\end{aligned}
$$

is the kernel function that is constant for $\frac{v}{n} \le t < \frac{v+1}{n}$ and has a jump of $\binom{n}{v} x^v (1-x)^{n-v}$ at $t = \frac{v}{n}$.

3 The Empirical Bernstein Copula

Our purpose is to define a nonparametric estimation procedure for a copula in terms of Bernstein polynomials. To this end, we need to define the empirical copula first.

Definition 1. *For i=1,...,k, let $X_{i,1}, ..., X_{i,n}$ be a sample of n observations with marginal distribution function F_i. The k dimensional empirical copula is given by*

$$
C_n(u_1, ..., u_k) \equiv \frac{1}{n} \sum_{s=1}^{n} I\left\{ \bigcap_{i=1}^{k} [F_i(x_{is}) \le u_i] \right\},
$$

where $I_{\{...\}}$ is the indicator function.

Notice that even when F_i is unknown, by the Glivenko-Cantelli Theorem, (e.g. van der Vaart and Wellner, 2000, ch. 2.4) we can use the empirical distribution as

[2] The k dimensional extension trivially follows.

a uniform strong estimator. Further, by the Glivenko-Cantelli theorem for classes of functions indexed by indicators of half spaces, the empirical copula is a uniform strong estimator of the Copula function (e.g., see van der Vaart and Wellner, 2000, ch. 2.4). In order to make the arguments in the sequel more transparent, we will refrain from pursuing generalization to dependence.

The empirical copula is not a smooth estimator, so it does not posses a density. For this reason, nonparametric kernel estimation and related techniques as spline smoothing are often employed. Nevertheless, the space where a copula is defined and its specific properties make the use of Bernstein polynomials more suitable. In fact, the copula function requires that, integrating out all but one variable, we are left with a one dimensional marginal distribution. Therefore, using the approach in Sancetta and Satchell (2001), we approximate the copula by a Bernstein polynomial. Notice that in the case of the empirical copula, the result of applying this linear operator is to have a smooth function. We have the following definition.

Definition 2. *The k dimensional nonparametric Bernstein copula is given by*

$$\begin{aligned}\left(B_m^k C_n\right)(u_1, ..., u_k) \equiv & \sum_{v_1=0}^{m} \cdots \sum_{v_k=0}^{m} \left[\sum_{s=1}^{n} n^{-1} I \left\{\bigcap_{j=1}^{k} \left[u_{is} \leq t_{v_j}\right]\right\}\right] \\ & \prod_{j=1}^{k} \binom{m}{v_j} u_j^{v_j} (1-u_j)^{m-v_j}, \end{aligned} \tag{3}$$

where $u_{is} \equiv F_i(x_{is})$, $t_{v_j} \equiv \frac{v_j}{m}$, and all other objects are as defined before.

It is simple to see that

$$\left(B_m^k C_n\right)(u_1, ..., u_k) = n^{-1} \sum_{s=1}^{n} \left(\sum_{v_1=0}^{m} \cdots \sum_{v_k=0}^{m} \prod_{j=1}^{k} I_{\left\{u_{js} \leq t_{v_j}\right\}} \binom{m}{v_j} u_j^{v_j} (1-u_j)^{m-v_j}\right).$$

Notice that, so far, we put no restriction on m. For the time being, we can state the following simple consistency result. At first we need to introduce some notation that will be used throughout the paper.

Notation.

$$C_{B_m}(\mathbf{u}) \equiv \left(B_m^k C_n\right)(u_1, ..., u_k),$$

where $\mathbf{u}$ *is a* k *dimensional vector;*

$$\left(B_m^k C_n\right)(\mathbf{u}) \equiv \int C_n(\mathbf{t})\, d_{\mathbf{t}} K_m(\mathbf{u}, \mathbf{t}),$$

where $C_n(...)$ *is the* k*-empirical copula and* $d_{\mathbf{t}} K_m(\mathbf{u}, \mathbf{t})$ *is the* k *dimensional kernel for the Bernstein operator of order* m *with* k *dimensional index* $\mathbf{t}$.

Theorem 1. *Let* $\{\mathbf{u}\}$ $(k \times 1)$ *be a sequence of independent strictly stationary uniform* $[0, 1]$ *random vectors with copula* $C(\mathbf{u})$ *and empirical copula* $C_n(\mathbf{u})$. *Then,*

$$\sup_{\mathbf{u}} \left| \int C_n(\mathbf{t})\, d_{\mathbf{t}} K_m(\mathbf{u}, \mathbf{t}) - C(\mathbf{u}) \right| \to 0, \text{ as } n,\ m \to \infty,$$

and

$$\sup_{\mathbf{u}} \left| \int C(\mathbf{t})\, d_{\mathbf{t}} K_m(\mathbf{u}, \mathbf{t}) - C(\mathbf{u}) \right| = O\left(\frac{k}{m}\right).$$

From the proof of Lemma 1, the reader can see that independence is only used for consistency of the empirical distribution function. Therefore, the result can be easily extended to weak dependence (see Rio, 2000, for the best known results in the case of weakly dependent series). It is clear that if we are interested only in the probability associated to each quantile, there is no advantage in using an empirical Bernstein copula instead of a simple empirical copula as the additional error incurred is $O\left(\frac{k}{m}\right)$. However, by defining the empirical Bernstein copula, we can derive consistent estimators for the density and the conditional distribution. Now we turn to the study of the nonparametric Bernstein copula density.

4 Nonparametric Bernstein Copula Density

By the properties of Bernstein polynomials, the Bernstein copula is absolutely continuous and differentiable in each argument. Differentiating the Bernstein polynomials with respect to each argument and rearranging, we see that they are closed under differentiation (Sancetta and Satchell, 2001, and Sancetta, 2002). Therefore, differentiating, it is easy to see that the coefficients of the polynomial are equivalent to a k dimensional histogram estimator (see Scott, 1992, for details on the histogram estimator),

$$\tilde{c}_B = \sum_{v_1=0}^{m-1} .... \sum_{v_k=0}^{m-1} \triangle_{1,....,k} \left(\frac{m^k}{n} \sum_{s=1}^{n} I \left\{ \bigcap_{j=1}^{k} \left[u_{js} \leq t_{v_j} \right] \right\} \right) \prod_{j=1}^{k} \binom{m-1}{v_j} u_j^{v_j} (1-u_j)^{m-1-v_j}, \tag{4}$$

where we use $\tilde{c}_B$ to stress that it is a particular estimator and $\triangle_{1,....,k}$ is the k dimensional partial forward difference operator, i.e.

$$\triangle_{1,....,k} I \left\{ \bigcap_{j=1}^{k} \left[u_{js} \leq t_{v_j} \right] \right\} = \sum_{l_1=0}^{1} .... \sum_{l_k=0}^{1} (-1)^{l_1+....+l_k} I \left\{ \bigcap_{j=1}^{k} \left[u_{js} \leq t_{v_j} + \frac{l_j}{m} \right] \right\}.$$

4.1 The Bernstein Operator as an Asymptotically Gaussian Convolution Operator

It is evident that the nonparametric Bernstein copula density is a smoothed version of the histogram estimator. It is the expectation of the histogram estimator with respect to the binomial distribution, i.e. the Bernstein operator is a convolution operator. In particular, consider the normal approximation to the binomial distribution, e.g. see Stuart and Ord (1994, p. 138-140). Let

$$P_{v,m}(u) \equiv \binom{m}{v} u^v (1-u)^{m-v},$$

and

$$\mathcal{P}_{v,m}(v) \equiv (2\pi u (1-u) m)^{-\frac{1}{2}} \exp \left\{ -\frac{m}{2u(1-u)} \left(\frac{v}{m} - u \right)^2 \right\}, \tag{5}$$

then

$$\sum_{v=0}^{m} f\left(\frac{v}{m}\right) P_{v,m}(u) \simeq \int_{-\infty}^{\infty} f\left(\frac{v}{m}\right) \mathcal{P}_{v,m}(v)\, dv.$$

Further, the error in this approximation is uniform. A formal proof may be given through the Edgeworth expansion for $z = \left(\frac{v}{m} - u\right)$. Taking squares of the two distributions (i.e. the binomial and the Gaussian),

$$\sum_{v=0}^{m} f\left(\frac{v}{m}\right)\left(P_{v,m}\left(u\right)\right)^2 \simeq \int_{-\infty}^{\infty} f\left(\frac{v}{m}\right)\left(\mathcal{P}_{v,m}\left(v\right)\right)^2 dv, \tag{6}$$

where again the error holds uniformly. This last statement is fundamental in the proof of Theorem 2 below. This result is also valid for higher dimensions. Therefore, using this approximation, we have

$$\tilde{c}_B \simeq \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{m^k}{n} \sum_{s=1}^{n} I\left\{\bigcap_{j=1}^{k}\left[u_{js} \leq \frac{v_j}{m}\right]\right\} \mathcal{P}_{v,m}\left(v\right) dv.$$

This shows that the asymptotic properties of the estimator $\tilde{c}_B$ can also be studied in terms of the properties of the convolution of an histogram with a Gaussian kernel. This shows that is appropriate to consider m^{-1} as smoothing parameter as far as comparisons are concerned. Further, by the properties of convolutions, we should expect $\tilde{c}_B$ to have lower variance than the simple histogram estimator.

4.2 Consistency in MSE of the Nonparametric Bernstein Copula Density

The optimal choice of m depends on the topology we use. We choose m to minimize the mean square error of the density, i.e. $\min \|\tilde{c}_B - c\|_2^2$ where $\|...\|_2$ is the L_2 norm under the true probability measure, and c is the true copula density. Just increasing m will reduce the bias but increase the variance of $\tilde{c}_B$.

We formally state the condition under which the following theorem is derived.

Condition 1. $\mathbf{u}_1, ..., \mathbf{u}_n$ $(k \times 1)$ *is a sequence of independent strictly stationary uniform* $[0,1]^k$ *random vectors with copula* $C(\mathbf{u})$ *and copula density* $c(\mathbf{u})$ *which has a finite first derivative everywhere in the k-cube.*

Remark. *The independence condition is not required, but we use it to make the proof as concise and transparent as possible. From the proof of Theorem 2 it can be seen that, mutatis mutandis, the results are still valid under appropriate mixing conditions by the use of known coupling results. Though we will use differentiability in the proof, this is not required. As long as the copula density is bounded and* *continuous (or has finite discontinuities of the first kind), we could define a generalization of Taylor expansion using Bernstein polynomials (e.g. Lorentz, 1953, p. 12-13, or Feller, 1971, p. 230-232) .*

We state the main result of the paper. Notice that we use $\preceq$ to indicate greater or equal up to a multiplicative constant, e.g. $a \preceq b$ implies $\exists C < \infty$ such that $a \leq Cb$.

Theorem 2. *Let $\tilde{c}_B$ be the k dimensional Bernstein copula density. Under Condition 1*

i. $Bias(\tilde{c}_B) \preceq m^{-1}$;

ii. Let $\lambda_j \equiv [u_j(1-u_j)]^{\frac{1}{2}}$,

(a.) for $u_j \in (0,1), \forall j$,

$$var(\tilde{c}_B) \preceq \left(n \prod_{j=1}^{k} \lambda_j\right)^{-1} m^{\frac{k}{2}} \left(1 + m^{-1}\right),$$

(b.) for $u_j = 0, 1, \forall j$,

$$var(\tilde{c}_B) = \frac{m^k}{n} c(\mathbf{u}) + O\left(\frac{m^{k-1}}{n}\right);$$

iii.

$$\tilde{c}_B(\mathbf{u}) \rightarrow c(\mathbf{u})$$

in mean square error:

(a.) for $u_j \in (0,1), \forall j$, if $\frac{m^{\frac{k}{2}}}{n} \rightarrow 0$ as $m, n \rightarrow \infty$;

(b.) for $u_j = 0, 1, \forall j$, if $\frac{m^k}{n} \rightarrow 0$ as $m, n \rightarrow \infty$;

iv. The optimal order of polynomial in a mean square error sense is:

(a.) $m \preceq n^{\frac{2}{k+4}}$ if $u_j \in (0,1), \forall j$;

(b.) $m \preceq n^{\frac{1}{k+2}}$ if $u_j = 0, 1, \forall j$;

v. If $m - k \geq 2$, then $\tilde{c}_B(\mathbf{u})$ and $\tilde{C}_B(\mathbf{u})$ are Donsker in $(0,1)$, i.e. $z_B(\mathbf{u}) \equiv \sqrt{m^{-\frac{k}{2}} n}\left[\tilde{c}_B(\mathbf{u}) - E\tilde{c}_B(\mathbf{u})\right]$ and $\tilde{Z}_B(\mathbf{u}) \equiv \sqrt{n}\left[\tilde{C}_B(\mathbf{u}) - E\tilde{C}_B(\mathbf{u})\right]$ converge to a zero mean Gaussian process with continuous sample paths and covariance function

$$E\left[z_B(\mathbf{u}_1) z_B(\mathbf{u}_2)\right],$$

and

$$E\left[\tilde{Z}_B\left(\mathbf{u}_1\right)\tilde{Z}_B\left(\mathbf{u}_2\right)\right],$$

respectively.

The weak limit of the infinite dimensional distribution of the nonparametric Bernstein copula density and the empirical Bernstein copula are given because these can be used to devise test statistics for independence based on some norm of the limiting Gaussian process. It is known that if a class is Donsker, then it is Glivenko-Cantelli, i.e. convergence holds uniformly over the class (e.g. van der Vaart and Wellner, 2000, p. 82). Therefore, the results of Theorem 2 hold uniformly.

For comparison purposes, let $h \equiv m^{-1}$ be the smoothing factor in the usual sense. The bias is of the same order as the one for the histogram estimator. In this respect, kernel smoothers would lead to a bias not higher than $O\left(m^{-2}\right)$. The reason for not calculating the constant is that in order to find the term that is $O\left(m^{-1}\right)$, it is necessary to take a Taylor series at least to second order. A Taylor series expansion up to second order requires more stringent conditions than the ones implied by Condition 1. On the other hand, here we do not make explicit use of Taylor expansions.[3] Notice that it is not possible to reduce the bias to $O\left(m^{-2}\right)$ by shifting the histogram, i.e. using frequency polygons (e.g. Scott, 1992, ch. 4, for details on frequency polygons). In this case the first term in the expansion would vanish, but other terms of same order would not. An easier way to see this is to notice that a Bernstein polynomial introduces an error which is $O\left(m^{-1}\right)$; see the proof of Theorem 1 and 2. Therefore, there is no way to reduce the bias unless we improve on the Bernstein approximation first (see Butzer, 1953, for linear combinations of Bernstein polynomials which improve on the rate of approximation).

[3] An earlier version of this paper made use of Taylor expansions in order to derive the exact form of the leading terms. However, the use of Bernstein polynomials led to lengthy calculations that clouded the simplicity of the argument.

While the bias is of the same order as the histogram estimator, the variance is of smaller order (except at the edges of the hypercube): $var\left(\tilde{c}_B\right) = O\left(m^{\frac{k}{2}}\right)$ instead of $O\left(m^k\right)$ as is the case for the histogram and kernel estimators. On the other hand, for $u_j = 0, 1$, for all j's, the variance is of the same order as for these other nonparametric estimators. The case $u_j = 0, 1$ for only some j is not included because the result is just a mixture of the two extreme cases: the variance goes down by a factor that is $O\left(m^{\frac{1}{2}}\right)$ for all the coordinates inside the k-hypercube while for the coordinates on the boundaries the contribution to the variance is $O\left(m\right)$.

As m and n go to infinity, it follows that this estimator has a rate of consistency $\frac{m^{\frac{k}{2}}}{n} \to 0$ inside the hypercube, versus $\frac{m^k}{n} \to 0$ for other common nonparametric estimators, e.g. Gaussian kernel. Inside the hypercube, the optimal order of smoothing is $m = O\left(n^{\frac{2}{k+4}}\right)$ in mean square error sense, versus $m = O\left(n^{\frac{1}{k+2}}\right)$ for the histogram and $m = O\left(n^{\frac{1}{k+4}}\right)$ for a first order kernel.

This implies that the Bernstein polynomials require very little smoothing (i.e. a large order of polynomial). This is due to the fact that Bernstein polynomials are fairly slow to adjust.

5 Illustrative Example

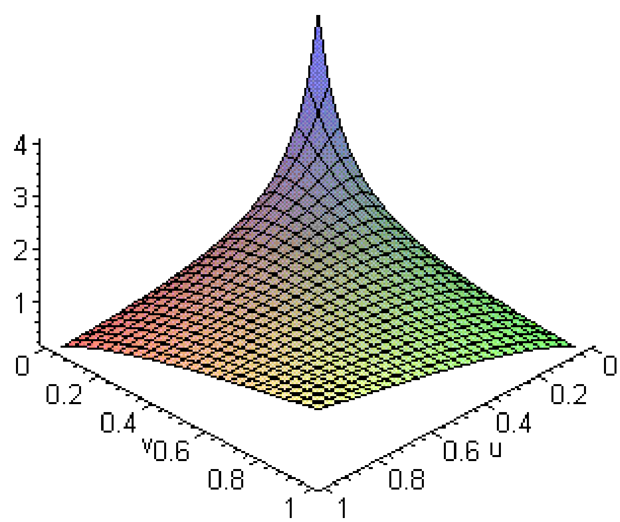
In this section, we use some short simulations to study the finite sample performance of the nonparametric Bernstein copula density. We choose the Kimeldorf and Sampson (KS) copula as the true one, i.e.

$$C\left(u, v\right) = \left(u^{-\gamma} + v^{-\gamma} - 1\right)^{-\frac{1}{\gamma}}.$$

In particular, we let $\gamma = .6$. The dependence parameter $\gamma = .6$ corresponds to a Spearman's rho equal to .34 (e.g., Joe, 1997, p. 32, for a definition of this measure of dependence). Figure I shows the plot for the copula density of the KS copula with dependence parameter $\gamma = .6$. We notice that this copula density exhibits lower tail dependence, e.g., an important property when modelling joint financial returns. However, this copula is singular at the origin. As a condition in Theorem 2 we used the fact that the copula density is nonsingular. Therefore, comparing this copula density with the nonparametric one will be of interest for several reasons.

We estimate the nonparametric Bernstein copula density with $m = 12$ using a simulated sample of 500 observations from a KS copula with $\gamma = .6$ (see Joe, 1997, p. 141, for further details on the KS copula). This boils down to finding the histogram estimator with bin width equal to m^{-1}, and then applying the Bernstein operator to it. We plot the estimated copula in Figure I.

Figure I. Kimeldorf and Sampson Copula Density, $\gamma = .6$.



In order to study more closely the performance of our estimator, we look at the integrated absolute error (IAE) and the integrated square error (ISE). The IAE and ISE are, respectively, the L_1 and the squared L_2 norm with respect to the Lebesgue measure of the difference of the KS copula density and the nonparametric Bernstein copula density, i.e.

$$\|(dC(u,v) - dC_B(u,v))\|_{1,(u,v)} = \int_{[0,1]^2} |(dC(u,v) - dC_B(u,v))|$$

$$\left\|\left(dC\left(u,v\right)-dC_B\left(u,v\right)\right)\right\|_{2,(u,v)}=\left(\int\limits_{[0,1]^2}\left[\left(dC\left(u,v\right)-dC_B\left(u,v\right)\right)\right]^2\right).$$

The KS copula density is singular at the origin, so the ISE would be highly affected by values in the neighborhood of the origin. Since nonparametric estimators cannot detect singularities, it is also informative to produce results for the IAE. The IAE is less sensitive to extreme values over sets of measure close to zero.

The IAE and the ISE are computed for a simulated sample of 500 observations from the KS copula with $\gamma = .6$. Results for different values of m are in Table I.[4] We also report the same results when the copula is estimated using the two dimensional histogram estimator. Further the IAE and the ISE are computed when the copula is assumed to be the independence copula (recall its density is 1). Since the nonparametric Bernstein copula is obtained from the histogram, it is natural to make comparisons with the two dimensional histogram. The values for the deviation of the KS copula density from the independence copula density are given to provide some benchmark values.

[4]The results in Table I should be taken cautiously. Simulated data from a copula are obtained by deterministic transformation of iid uniform $[0, 1]$ random variables. Since pseudo-random numbers are not iid, the implicit dependence (we often found the variables to be correlated as well) in the raw data will produce simulated data which may not be completely consistent with the desired copula.

Figure II. Nonparametric Bernstein Copula Density, $m = 12$.

(Artificially Generated Data, $n = 500$)

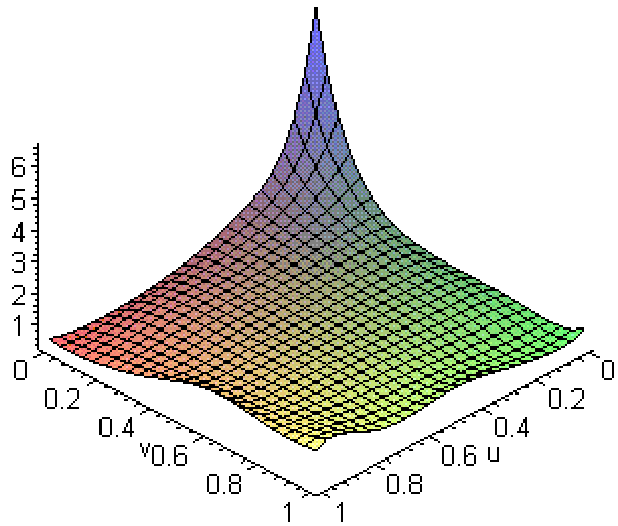


TABLE I. IAE and ISE.

	N.B. Copula		2D Histogram		Indep. Copula	
m	IAE	ISE	IAE	ISE	IAE	ISE
					0.26928	1.67371
4	0.15350	1.5664	0.21968	1.5853		
5	0.14176	1.5568	0.21749	1.5847		
6	0.13721	1.5509	0.24718	1.5969		
7	0.12755	1.5361	0.28075	1.6074		
8	0.11682	1.5251	0.31472	1.6358		
9	0.12404	1.5188	0.34334	1.6597		
10	0.11955	1.5092	0.37963	1.6882		
11	0.11238	1.5058	0.38339	1.7053		
12	0.11507	1.5094	0.39403	1.7189		
13	0.11662	1.5033	0.38975	1.7263		
14	0.11155	1.4929	0.49072	1.8358		
15	0.11437	1.4956	0.53152	1.9026		
16	0.11979	1.4921	0.56853	1.9418		
17	0.12597	1.5076	0.57352	2.03248		
18	0.12677	1.5039	0.65579	2.12442		

The main result of Table I is that applying the Bernstein operator to the 2 dimensional histogram allows us to decrease the size of the mesh, and consequently the bias keeping the uncertainty (i.e. the variance) low. This agrees with the asymptotic results in Theorem 2. Further, the results do not seem to be particularly sensitive to m as opposed to the histogram.

6 Proofs

Proof of Theorem 1. By the triangle inequality,

$$\begin{aligned}
&\sup_{\mathbf{u}}\left|\int C_n(\mathbf{t})\,d_{\mathbf{t}}K_m(\mathbf{u},\mathbf{t})-C(\mathbf{u})\right| \\
\le\ &\sup_{\mathbf{u}}\left|\int C_n(\mathbf{t})\,d_{\mathbf{t}}K_m(\mathbf{u},\mathbf{t})-C(\mathbf{t})\,d_{\mathbf{t}}K_m(\mathbf{u},\mathbf{t})\right| \\
&+\sup_{\mathbf{u}}\left|\int C(\mathbf{t})\,d_{\mathbf{t}}K_m(\mathbf{u},\mathbf{t})-C(\mathbf{u})\right| \\
=\ &\sup_{\mathbf{u}}\left|\int (C_n(\mathbf{t})-C(\mathbf{t}))\,d_{\mathbf{t}}K_m(\mathbf{u},\mathbf{t})\right| \\
&+\sup_{\mathbf{u}}\left|\int C(\mathbf{t})\,d_{\mathbf{t}}K_m(\mathbf{u},\mathbf{t})-C(\mathbf{u})\right|,
\end{aligned}$$

where the equality follows by the properties of linear operators. By the Glivenko-Cantelli Theorem, $(C_n(\mathbf{u})-C(\mathbf{u}))$ converges to zero as $n\to\infty$, it follows that

$$\sup_{\mathbf{u}}\left|\int (C_n(\mathbf{t})-C(\mathbf{t}))\,d_{\mathbf{t}}K_m(\mathbf{u},\mathbf{t})\right|\to 0.$$

By Lemma 1 in Sancetta (2002, ch. 1), for any $\epsilon>0$, $\exists\ m$ such that

$$\sup_{\mathbf{u}}\left|\int C(\mathbf{t})\,d_{\mathbf{t}}K_m(\mathbf{u},\mathbf{t})-C(\mathbf{u})\right|\le\epsilon.$$

Since ϵ can be made arbitrarily small by suitable choice of m, the second term can be made as close as we please to zero. In particular,

$$\sup_{\mathbf{u}}\left|\int C(\mathbf{t})\,d_{\mathbf{t}}K_m(\mathbf{u},\mathbf{t})-C(\mathbf{u})\right|=O\left(\frac{k}{m}\right),$$

from Theorem 2 in Sancetta (2002, ch. 1). ■

Before proving Theorem 2, we need some additional notation.

Notation. *We use $\mathbf{u}_s$ to indicate the vector of r.v.'s. On the other hand $\mathbf{u}$ will denote a fixed, but arbitrary value. This generates no confusion as long as one is willing to look at a function as a point in the space. Moreover,* $\partial_i c(\mathbf{u})\equiv\frac{\partial c(\mathbf{u}_s)}{\partial u_i}|_{\mathbf{u}_s=\mathbf{u}}$.

Proof of Theorem 2. Bias$(\tilde{c}_B) \equiv E(\tilde{c}_B) - c(\mathbf{u})$, which we can rewrite as

$$E\int c_n(\mathbf{t})\, d_{\mathbf{t}}K_m(\mathbf{u},\mathbf{t}) - c(\mathbf{u}).$$

We shall proceed as in the previous proof. Clearly,

$$\begin{aligned} E\int c_n(\mathbf{t})\, d_{\mathbf{t}}K_m(\mathbf{u},\mathbf{t}) - c(\mathbf{u}) &\leq E\int [c_n(\mathbf{t}) - c(\mathbf{t})]\, d_{\mathbf{t}}K_m(\mathbf{u},\mathbf{t}) \\ &\quad + E\left(\int c(\mathbf{t})\, d_{\mathbf{t}}K_m(\mathbf{u},\mathbf{t}) - c(\mathbf{u})\right) \\ &\leq E\int [c_n(\mathbf{t}) - c(\mathbf{t})]\, d_{\mathbf{t}}K_m(\mathbf{u},\mathbf{t}) + O\left(\frac{k}{m}\right), \end{aligned}$$

where the second inequality follows from the same argument as in the previous proof. Since the Bernstein operator is bounded, by Fubini's theorem, it is sufficient to consider

$$E[c_n(\mathbf{t}) - c(\mathbf{t})],$$

i.e. the bias for a histogram estimator. Therefore, (e.g. Scott, 1982, p. 81, or just use (10) below, where no Taylor expansion is required)

$$E[c_n(\mathbf{t}) - c(\mathbf{t})] = O\left(m^{-1}\right).$$

Therefore,

$$Bias(\tilde{c}_B) = O\left(\frac{1}{m}\right).$$

For the variance, notice that the probability of one observation falling inside a subset of the hypercube is equal to the probability of a success in a Bernoulli trial. We know that the probability of n successes, where n is the sample size, is given by a binomial distribution. By the variance of n independent Bernoulli trials

$$var(\tilde{c}_B) = \sum_{v_1=0}^{m-1} \cdots \sum_{v_k=0}^{m-1} \left(\frac{m^{2k}}{n^2}\sum_{s=1}^{n}\left(p_{s,v_1\cdots v_k} - \left(p_{s,v_1\cdots v_k}\right)^2\right)\right)\prod_{j=1}^{k}\left(P_{v_j,m-1}(u)\right)^2, \quad (7)$$

where

$$p_{s,v_1\cdots v_k} \equiv \int_{t_{v_k}}^{t_{v_k}+\frac{1}{m}} \cdots \int_{t_{v_1}}^{t_{v_1}+\frac{1}{m}} c(\mathbf{u}_s)\, du_{1s}\cdots du_{ks}.$$

Consider the following simple identity for any differentiable function f,

$$\int_{t}^{t+\frac{1}{m}} f(u)\,du = \frac{f(t)}{m} - \int_{t}^{t+\frac{1}{m}} \left(u - t - \frac{1}{m}\right) df(u), \tag{8}$$

where the left hand side can be recovered by simple integration of the second term on the right hand side. By direct application of (8) we have

$$\begin{aligned} p_{s,v_1\cdots v_k} &= \int_{t_{v_k}}^{t_{v_k}+\frac{1}{m}} \cdots \int_{t_{v2}}^{t_{v2}+\frac{1}{m}} \Bigg(\frac{c(t_{v_1}, u_{2s}, ..., u_{ks})}{m} \\ &\quad - \int_{t_{v1}}^{t_{v1}+\frac{1}{m}} \left(u_{1s} - t_{v1} - \frac{1}{m}\right) \partial_1 c(u_{1s}, u_{2s}, ..., u_{ks})\,du_{1s} \Bigg) du_{2s} \cdots du_{ks}. \end{aligned} \tag{9}$$

Now, by Condition 1, $\partial_1 c(u_{1s}, u_{2s}, ..., u_{ks}) \leq M$ for some $M < \infty$. Therefore, the last term in (9) is bounded by

$$M \int_{t_{v1}}^{t_{v1}+\frac{1}{m}} \left(t_{v1} + \frac{1}{m} - u_{1s}\right) du_{1s} = O\left(m^{-2}\right).$$

Substituting into (9),

$$p_{s,v_1\cdots v_k} = \int_{t_{v_k}}^{t_{v_k}+\frac{1}{m}} \cdots \int_{t_{v2}}^{t_{v2}+\frac{1}{m}} \left(\frac{c(t_{v_1}, u_{2s}, ..., u_{ks})}{m} + O\left(m^{-2}\right) \right) du_{2s} \cdots du_{ks}.$$

Applying (8) repeatedly, we have

$$p_{s,v_1\cdots v_k} = \frac{c(t_{v_1}, ..., t_{v_k})}{m^k} + O\left(m^{-(k+1)}\right). \tag{10}$$

Since $p_{s,v_1\cdots v_k} \leq 1$, it follows that $\left(p_{s,v_1\cdots v_k}\right)^2 = o\left(p_{s,v_1\cdots v_k}\right)$. Therefore, substituting (10) into (7), we have that

$$var(\tilde{c}_B) = \frac{m^{2k}}{n} \sum_{v_1=0}^{m-1} \cdots \sum_{v_k=0}^{m-1} \left(\frac{c(t_{v_1}, ..., t_{v_k})}{m^k} + O\left(m^{-(k+1)}\right) \right) \prod_{j=1}^{k} \left(P_{v_j,m-1}(u)\right)^2. \tag{11}$$

We use (6) to approximate (11). By Condition 1, $c(t_{v_1}, ..., t_{v_k})$ is bounded, say by $M < \infty$. Recall that $t_{v_j} \equiv \frac{v_j}{m}$, $j = 1, ..., k$. Consequently, solve the following type of integral,

$$\begin{aligned}\Gamma_j &= \int_{\mathbb{R}} c\left(\frac{v_1}{m}, ..., \frac{v_k}{m}\right) \frac{\exp\left\{-\frac{(m-1)}{u_j(1-u_j)}\left(\frac{v_j}{m-1} - u_j\right)^2\right\}}{\left[2\pi(m-1)u_j(1-u_j)\right]} dv_j \\ &\leq M \int_{\mathbb{R}} \frac{\exp\left\{-\frac{(m-1)}{u_j(1-u_j)}\left(\frac{v_j}{m-1} - u_j\right)^2\right\}}{\left[2\pi(m-1)u_j(1-u_j)\right]} dv_j.\end{aligned}$$

Simply make the following change of variable, $x_j = \sqrt{\frac{(m-1)}{u_j(1-u_j)}}\left(\frac{v_j}{m-1} - u_j\right)$, with Jacobian $\sqrt{(m-1)u_j(1-u_j)}$. Then

$$\begin{aligned}\Gamma_j &\leq \int_{\mathbb{R}} \frac{\exp\left\{-x_j^2\right\}}{2\pi\sqrt{(m-1)u_j(1-u_j)}} dx_j \\ &= \frac{1}{\sqrt{2\pi(m-1)u_j(1-u_j)}}.\end{aligned}$$

Therefore, $\Gamma_j = O\left(m^{-\frac{1}{2}}\right)$. This shows that integration leads to a drop in asymptotic magnitude equal to $m^{-\frac{1}{2}}$ for each dimension. Let $\lambda_j \equiv \left[u_j(1-u_j)\right]^{\frac{1}{2}}$, then

$$\begin{aligned}var(\tilde{c}_B) &\preceq \frac{m^{2k}}{n}\left(\left[4\pi(m-1)\right]^{\frac{k}{2}} \prod_{j=1}^{k} \lambda_j\right)^{-1} \left(\frac{M}{m^k} + m^{-(k+1)}\right) \\ &\preceq \left(n \prod_{j=1}^{k} \lambda_j\right)^{-1} m^{\frac{k}{2}}\left(1 + m^{-1}\right).\end{aligned}$$

At the edges of the hypercube, i.e. $u = 0, 1$, $\left(P_{v_j,m-1}(u)\right)^2 = P_{v_j,m-1}(u)$, then

$$\begin{aligned}var(\tilde{c}_B) &= \frac{m^{2k}}{n}\left(\frac{c(\mathbf{u})}{m^k} + O\left(m^{-(1+k)}\right)\right) \\ &= \frac{m^k}{n} c(\mathbf{u}) + O\left(\frac{m^{k-1}}{n}\right).\end{aligned}$$

The mean square error (MSE) convergence simply follows by considering the leading terms for the square bias and the variance for the two distinct cases: $MSE = Bias(\tilde{c}_B)^2 + Var(\tilde{c}_B)$. The optimal order of the polynomial follows by minimization of asymptotic MSE with respect to m. Inside the hypercube we have

$$\begin{aligned}\left(\frac{\partial}{\partial m}\right) MSE &\preceq \left(\frac{\partial}{\partial m}\right)\left(m^{-2}+\frac{m^{\frac{k}{2}}}{n}\right)\\ &\preceq \left(-m^{-3}+\frac{km^{\frac{k}{2}-1}}{2n}\right)\\ &= 0,\end{aligned}$$

which implies $m^{\frac{k+4}{2}} = O(n)$. Similarly, for the density at the boundaries of the k-cube,

$$\begin{aligned}\left(\frac{\partial}{\partial m}\right) MSE &\preceq \left(\frac{\partial}{\partial m}\right)\left(m^{-2}+\frac{m^{k}}{n}\right)\\ &\preceq \left(-m^{-3}+\frac{km^{k-1}}{2n}\right)\\ &= 0,\end{aligned}$$

which implies $m^{k+2} = O(n)$.

The finite dimensional distributions of the nonparametric Bernstein copula density converge to a normal distribution. This follows from the fact that it is the sum of bounded random variables and Condition 1 (weaker conditions than iid are clearly sufficient for the central limit theorem). But the Bernstein copula density has $m-1$ bounded derivatives (recall that Bernstein polynomials are closed under differentiation) and any Bernstein polynomial is Lipschitz. By Theorem 2.7.1 in van der Vaart and Wellner (2000, p. 155) the class of functions that satisfy the properties just mentioned has finite ε bracketing numbers of order $\exp\left\{\varepsilon^{-\frac{k}{m-1}}\right\}$. It follows that their entropy integral with bracketing is finite. This is enough to show (see Ossiander, 1987, for the iid case or Pollard, 2001, for generalizations) that the Bernstein copula density converges to a Gaussian process with continuous sample paths. The $\sqrt{m^{-\frac{k}{2}}n}$ term is required for the leading term in the variance expansion to be independent of n, i.e. $m \to \infty$ as $n \to \infty$ (as usual for other nonparametric estimators). The same condition applies to the copula because it is m times differentiable together with the same properties of the density. Clearly, the simple root-n standardization is employed in this case (integration absorbs the smoothing parameter). ■

From the proof it is clear that what drives the variance down is the fact that approximating the square of $P_{v,m-1}(u)$ leads to a normal approximation times an extra term that is $O\left(m^{-\frac{1}{2}}\right)$. In order to provide more intuition on this result and the difference between the edges of the box and the points inside it, we provide the following heuristic explanation. Bernstein polynomials average the information about the function throughout its support; recall the singular integral representation in (2). On the other hand, the result at the corners of the hypercube is clear: the approximation at these points is exact and it is not influenced by the behaviour of the function in its domain, i.e. it is exactly local so that we just recover the properties of the histogram estimator.

7 Final Remarks

We defined the empirical Bernstein copula, and showed that it converges under the sup norm. We showed that the density of this nonparametric copula shares some of the properties of convolutions with the convolving function being asymptotically Gaussian. In Theorem 2 we collected a complete set of convergence result for this estimator. We provided rates of convergence in L_2, the optimal order of polynomial for the estimator and its weak convergence to a Gaussian process. In particular, while the bias of our estimator is of the same order as the one for the histogram, its variance is of lower order than other common nonparametric estimators. This is a consequence of the convolution properties of the estimator and provides some remedy to the curse of dimensionality. Moreover, the implementation of the estimator is remarkably simple. Further, the simulation showed that the integrated square error is not particularly sensitive to the choice of m as opposed to the histogram. This reduces the risk of choosing the wrong m in empirical work.

References

[1] Butzer, P.L. (1953) Linear Combinations of Bernstein Polynomials. Canadian Journal of Mathematics 5, 559-567.

[2] Capéraà, P., A.-L. Fougères and C. Genest (1997) A Nonparametric Estimation Procedure for Bivariate Extreme Value Copulas. Biometrika 84, 567-577.

[3] Devore, R.A. and G.G. Lorentz (1993) Constructive Approximations. Berlin: Springer.

[4] Feller, W. (1966) An Introduction to Probability Theory and Applications, vol. II. New York: John Wiley & Sons Inc.

[5] Genest, C. and L.P. Rivest (1993) Statistical Inference Procedures for Bivariate Archimedean Copulas. Journal of the American Statistical Association 88, 1034-1043.

[6] Genest, C., K. Ghoudi and L.-P. Rivest (1995) A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions. Biometrika 82, 543-552.

[7] Hall, P. (1990) Using the Bootstrap to Estimate Mean Square Error and Select Smoothing Parameter in Nonparametric Problems. Journal of Multivariate Analysis 32, 177-203.

[8] Joe, H. (1997) Multivariate Models and Dependence Concepts. London: Chapman & Hall Ltd.

[9] Li, D. (1999) On Default Correlation: A Copula Function Approach. The RiskMetrics Group Working Paper 99-07.

[10] Li, X., P. Mikusiński and M.D. Taylor (1998) Strong Approximation of Copulas. Journal of Mathematical Analysis and Applications 225, 608-623.

[11] Li, H., M. Scarsini and M. Shaked (1996) Linkages: a Tool for the Construction of Multivariate Distributions with Given Nonoverlapping Multivariate Marginals. Journal of Multivariate Analysis 56, 20-41.

[12] Longin, F. and B. Solnik (2001) Extreme Correlation of International Equity Markets. Journal of Finance 56, 651-678.

[13] Lorentz, G.G. (1953) Bernstein Polynomials. Toronto: University of Toronto Press.

[14] Ossiander, M. (1987) A Central Limit Theorem under Metric Entropy with L_2 Bracketing. The Annals of Probability 15, 897-919.

[15] Patton, A.J. (2001) Modelling Time-Varying Exchange Rate Dependence Using the Conditional Copula. Department of Economics Discussion Paper 2001-09, UCSD.

[16] Pollard, D. (2001) General Bracketing Inequalities. Statistics Department, Yale University.

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

[18] Rosenberg, J.V. (2000) Nonparametric Pricing of Multivariate Contingent Claims. Working Paper, Stern School of Business, NYU.

[19] Sancetta, A. (2002) Approximation Methods in Financial Econometrics. Unpublished PhD dissertation, Trinity College, University of Cambridge.

[20] Sancetta, A. and S.E. Satchell (2001) Bernstein Approximations to the Copula Function and Portfolio Optimization. DAE Working Paper 0105, University of Cambridge.

[21] Schweizer, B. and A. Sklar (1983) Probabilistic Metric Spaces. New York: Elsevier.

[22] Scott, D.W. (1979) On Optimal and Data-Based Histograms. Biometrika 66, 605-610.

[23] Scott, D.W. (1992) Multivariate Density Estimation: Theory, Practice and Visualization. New York: John Wiley & Sons Inc.

[24] Sklar, A. (1973) Random Variables, Joint Distribution Functions, and Copulas. Kybernetika 9, 449-460.

[25] Stuart A. and J.K. Ord (1994) Kendall's Advanced Theory of Statistics. London: Edward Arnold.

[26] Van der Vaart, A. and J.A. Wellner (2000) Weak Convergence of Empirical Processes. Springer Series in Statistics. New York: Springer.