The Petrov-Galerkin projection for copula density estimation isn't counting

D. Uhlig, R. Unger

Preprint 2014-03

Preprintreihe der Fakultät für Mathematik
ISSN 1614-8835

Impressum:
Herausgeber:
Der Dekan der
Fakultät für Mathematik
an der Technischen Universität Chemnitz
Sitz:
Reichenhainer Strae 39
09126 Chemnitz
Postanschrift:
09107 Chemnitz
Telefon: (0371) 531-22000
Telefax: (0371) 531-22009
E-Mail: dekanat@mathematik.tu-chemnitz.de
Internet:
http://www.tu-chemnitz.de/mathematik/
ISSN 1614-8835 (Print)

The Petrov-Galerkin projection for copula density estimation isn't counting

February 12, 2014

Dana Uhlig
TU Chemnitz
Department of Mathematics
09107 Chemnitz, Germany
dana.uhlig@mathematik.tu-chemnitz.de

Roman Unger
TU Chemnitz
Department of Mathematics
09107 Chemnitz, Germany
roman.unger@mathematik.tu-chemnitz.de

Abstract

Non-parametric copula density estimation in the $d-$dimensional case is a big challenge in particular if the dimension d of the problem increases. In [UU13] we proposed to solve the d-dimensional Volterra integral equation $\int\limits_0^{\mathbf{u}} \mathrm{c}(\mathbf{s}) \mathrm{d}\mathbf{s} = \mathrm{C}(\mathbf{u})$ for a given copula C. In the statistical framework the copula C is unobservable and hence we solved the linear integral equation for the empirical copula. For the numerical computation we used a Petrov-Galerkin projection for the approximated piecewise constant function $\mathrm{c}_h = \sum\limits_{j=1}^{N} c_j \phi_j$. Other than might be expected, the vector $\mathbf{c} = (c_1, \ldots, c_N)^T$ doesn't count the number of samples in the elements of the discretized grid, even the approximated solution c_h is a piecewise constant function on the elements. We will establish that solving the Volterra integral equation by a Petrov-Galerkin projection is not simple counting.

1 Introduction

In the non-parametric copula density estimation from T given $d-$dimensional pseudo samples $\hat{\mathbf{U}}_1, \hat{\mathbf{U}}_2, \dots, \hat{\mathbf{U}}_T$ there is no particular information about the structure of the copula density. The copula density of an absolutely continuous copula C is the derivative

$$\mathrm{c}(u_1, \dots, u_d) = \frac{\partial^d \mathrm{C}}{\partial u_1 \dots \partial u_d} \tag{1}$$

of the given copula C. Unfortunately the copula C is not observable, but we can treat the empirical copula

$$\hat{\mathrm{C}}(\mathbf{u}) = \frac{1}{T} \sum_{j=1}^{T} \mathbb{1}_{\hat{\mathbf{U}}_j \leq \mathbf{u}} = \frac{1}{T} \sum_{j=1}^{T} \prod_{k=1}^{d} \mathbb{1}_{\hat{U}_{kj} \leq u_k} \tag{2}$$

as a noisy version of C and solve the linear Volterra integral equation

$$\int_0^{u_1} \cdots \int_0^{u_d} \mathrm{c}(s_1, \dots, s_d) \mathrm{d}s_1 \cdots \mathrm{d}s_d = \mathrm{C}(u_1, \dots, u_d) \quad \forall \mathbf{u} = (u_1, \dots, u_d)^T \in [0,1]^d \tag{3}$$

as an inverse problem. For the sake of convenience we write

$$\int_{\mathbf{0}}^{\mathbf{u}} \mathrm{c}(\mathbf{s}) d\mathbf{s} = \mathrm{C}(\mathbf{u}) \quad \forall \mathbf{u} = (u_1, \dots, u_d)^T \in \Omega = [0,1]^d$$

for equation (3) as a short form. In the working paper [UU13] we decomposed the $d-$dimensional hypercube $\Omega = [0,1]^d$ into $N = n^d$ elements $e_1, \dots, e_N$ (see figure 1) and proposed a Petrov-Galerkin projection for the ansatz

$$\mathrm{c}_h(\mathbf{s}) = \sum_{j=1}^{N} c_j \phi_j(\mathbf{s}) \tag{4}$$

with $N = n^d$ ansatz functions

$$\phi_j(\mathbf{u}) = \begin{cases} 1 & \mathbf{u} \in e_i \\ 0 & \mathbf{u} \notin e_i \end{cases} \tag{5}$$

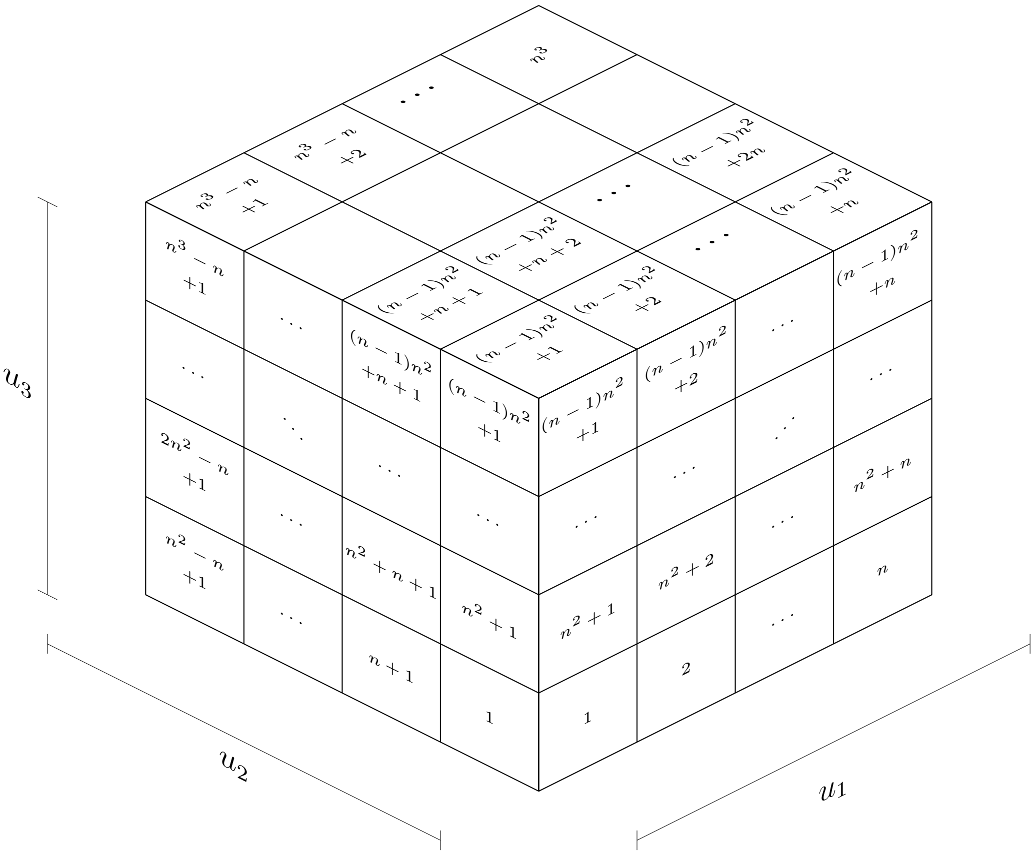


Figure 1: Discretization of the unit hypercube $\Omega = [0,1]^d$ for $d = 3$

and N test functions $\psi_1, \ldots, \psi_N$. The Petrov-Galerkin projection

$$\int_\Omega \int_{\mathbf{0}}^{\mathbf{u}} \mathrm{c}_h(\mathbf{s}) \mathrm{d}\mathbf{s} \psi_j(\mathbf{u}) \mathrm{d}\mathbf{u} = \int_\Omega \mathrm{C}(\mathbf{u}) \psi_j(\mathbf{u}) \mathrm{d}\mathbf{u} \quad j = 1, \ldots, N\,.$$

leads to a linear system $Kc = C$ with right hand side

$$C_i = \int_\Omega \mathrm{C}(\mathbf{u}) \psi_i(\mathbf{u}) \mathrm{d}\mathbf{u}, \quad i = 1, \ldots, N \tag{6}$$

and the $N \times N$ matrix K with

$$K_{ij} = \int_{\Omega} \int_{\mathbf{0}}^{\mathbf{u}} \phi_j(\mathbf{s}) \mathrm{d}\mathbf{s} \psi_i(\mathbf{u}) \mathrm{d}\mathbf{u} \,. \tag{7}$$

In [UU13] it is shown, that if the test functions are chosen as the integrated ansatz functions, the system matrix has a special structure and can be written as a $d-$times Kronecker product ${}^{(d)}K = {}^{(1)}K \otimes {}^{(1)}K \otimes \ldots \otimes {}^{(1)}K$ of the one-dimensional problem, such that the solution of the linear system $Kc = C$ is

$$c = {}^{(d)}K^{-1}C = \left({}^{(1)}K^{-1} \otimes {}^{(1)}K^{-1} \otimes \ldots \otimes {}^{(1)}K^{-1}\right) C \,. \tag{8}$$

It becomes apparent that the product structure of the ansatz functions

$$\phi_i(\mathbf{u}) = \prod_{k=1}^{d} \phi_i^k(u_k) \tag{9}$$

with the one-dimensional ansatz functions[1]

$$\phi_i^k = \mathbb{1}_{[b_k^i, b_k^i + h]}$$

as well as the product structure of the test functions

$$\psi_i(\mathbf{u}) = \prod_{k=1}^{d} \psi_i^k(u_k) \tag{10}$$

with the one-dimensional test functions

$$\psi_i^k(u) = \int_0^u \phi_i^k(s) \mathrm{d}s$$

is responsible for the special structure of the system matrix K and proposition 2.5 can be generalized.

Theorem 1.1. *If the test functions and ansatz functions have the product structure (9) and (10) then the system matrix for the $(d+1)-$dimensional case can be extracted from the one and $d-$dimensional system matrices.*

$${}^{(d+1)}K = {}^{(1)}K \otimes {}^{(d)}K \tag{11}$$

[1] The vector $\mathbf{b}^i$ is the lowest corner of the $i-$th element e_i and $h = 1/n$.

Hence (8) is also the solution of the linear system $Kc = C$ if arbitrary test and ansatz functions decompose into a product of one-dimensional test and ansatz functions.

It seems reasonable to suppose that the estimated density (4) equates the empirical density of the samples $\hat{\mathbf{U}}_1, \hat{\mathbf{U}}_2, \ldots, \hat{\mathbf{U}}_T$ if the right hand side (6) of the linear system $Kc = C$ is based on the empirical copula. However, even if we use piecewise constant test functions, that is $\psi_i = \phi_i$ we do not count the number of samples in the elements $e_1, \ldots, e_N$.

2 The empirical copula

In the following we assume for simplicity that the pseudo samples lie at the grid (see figure 2) and define the counting vector A with

$$A_i = \# \left\{ \hat{\mathbf{U}}_1, \ldots, \hat{\mathbf{U}}_T \in e_i \right\} .$$

In this case the empirical copula $\hat{\mathrm{C}}$ is constant on the elements e_i and

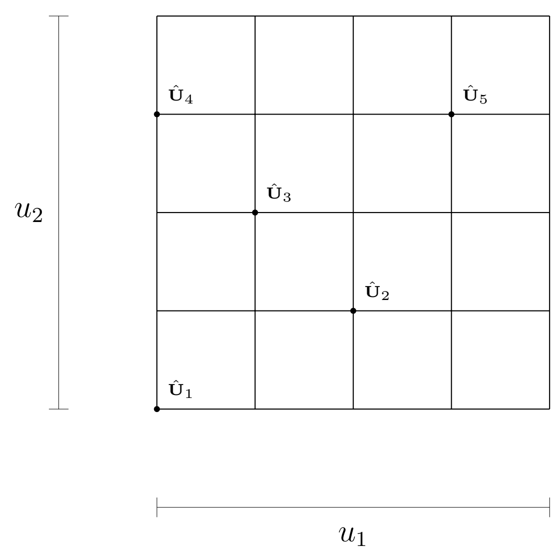


Figure 2: Pseudo samples on the grid

therefore we can write it as the vector

$$\tilde{C} = \frac{1}{T} \left(D \otimes \cdots \otimes D \right) A \qquad (12)$$

that is the $i-$th component $\tilde{C}_i$ describes the value of the empirical copula over the $i-$th element e_i using the $d-$times Kronecker product of the $n \times n$ matrix

$$D = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} .$$

We will analyze the structure of the right hand side C^δ using formula (6) for the empirical copula, that is

$$C_i^\delta = \int_\Omega \hat{\mathrm{C}}(\mathbf{u})\psi_i(\mathbf{u})\mathrm{d}\mathbf{u}\,, \tag{13}$$

in order to show, that the solution vector $c = K^{-1}C^\delta$ does not count the number of samples lying on the elements $e_1, \dots, e_N$.

2.1 Solution of the Petrov-Galerkin projection for piecewise constant test functions

In the simplest case the test functions ψ_i are chosen such that they are equal to the ansatz functions ϕ_i, that is

$$\psi_i(\mathbf{u}) = \begin{cases} 1 & \mathbf{u} \in e_i \\ 0 & \mathbf{u} \notin e_i \end{cases} . \tag{14}$$

Then it yields for our assumption

$$C_i^\delta = \int_\Omega \hat{\mathrm{C}}(\mathbf{u})\psi_i(\mathbf{u})\mathrm{d}\mathbf{u} = \int_{e_i} \hat{\mathrm{C}}(\mathbf{u})\mathrm{d}\mathbf{u} = \tilde{C}_i h^d$$

and hence

$$C^\delta = \frac{h^d}{T}\left(D \otimes \dots \otimes D\right) A\,. \tag{15}$$

Moreover, we get

$$\begin{aligned} c = K^{-1}C^\delta &= \frac{h^d}{T}\left({}^{(1)}K^{-1} \otimes {}^{(1)}K^{-1} \otimes \ldots \otimes {}^{(1)}K^{-1}\right)\left(D \otimes \dots \otimes D\right) A \\ &= \frac{h^d}{T}\left(\left({}^{(1)}K^{-1}D\right) \otimes \dots \otimes \left({}^{(1)}K^{-1}D\right)\right) A \end{aligned} \tag{16}$$

for the solution vector. It is an easy computation to see that the $n \times n$ matrix

$$^{(1)}K^{-1}D = \frac{2}{h^2}\begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ -2 & 1 & 0 & \dots & 0 & 0 \\ 2 & -2 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ (-1)^{n+1}2 & (-1)^{n+2}2 & (-1)^{n+3}2 & \dots & -2 & 1 \end{pmatrix}$$

has not a diagonal structure. However, only if the $n \times n$ matrix $^{(1)}K^{-1}D$ would have a diagonal structure, the $N \times N$ matrix $K^{-1}(D \otimes \cdots \otimes D)$ also would have a diagonal structure and hence the solution vector c would be a multiple of the counting vector A, such that we would count the number of samples in each element.

2.2 Solution of the Petrov-Galerkin projection for integrated ansatz functions as test functions

In [UU13] we proposed to choose the test functions ψ_i as integrated ansatz functions, that is

$$\psi_i(\mathbf{u}) = \int_{\mathbf{0}}^{\mathbf{u}} \phi_i(\mathbf{s})\mathrm{d}\mathbf{s}\,.$$

In this case we obtain for our assumption

$$C_i^\delta = \int_\Omega \hat{\mathrm{C}}(\mathbf{u})\psi_i(\mathbf{u})\mathrm{d}\mathbf{u} = \sum_{l=1}^{N}\int_{e_l} \hat{\mathrm{C}}(\mathbf{u})\psi_i(\mathbf{u})\mathrm{d}\mathbf{u} = \sum_{l=1}^{N}\tilde{C}_l\int_{e_l} \psi_i(\mathbf{u})\mathrm{d}\mathbf{u}$$

and hence

$$C^\delta = \frac{h^{2d}}{T}\left(FD \otimes \cdots \otimes FD\right)A \tag{17}$$

with the $n \times n$ matrix

$$F = \begin{pmatrix} \frac{1}{2} & 1 & \dots & 1 \\ 0 & \frac{1}{2} & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{2} \end{pmatrix}.$$

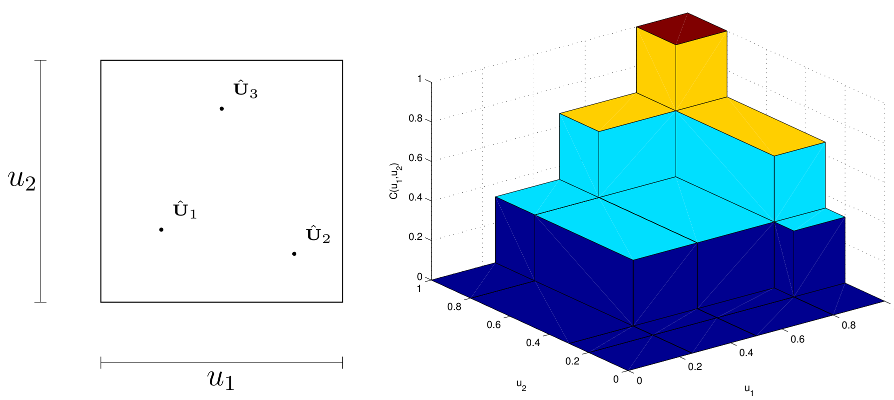
Using equation (17) we obtain

$$c = K^{-1}C^{\delta} = \frac{h^{2d}}{T}\left({}^{(1)}K^{-1} \otimes {}^{(1)}K^{-1} \otimes \cdots \otimes {}^{(1)}K^{-1}\right)(FD \otimes \cdots \otimes FD)\,A$$
$$= \frac{h^{2d}}{T}\left(\left({}^{(1)}K^{-1}FD\right) \otimes \cdots \otimes \left({}^{(1)}K^{-1}FD\right)\right)A \tag{18}$$

for the solution vector. Again it is an easy computation to see that the $n \times n$ matrix ${}^{(1)}K^{-1}FD$ is not diagonal such that also in this second case we do not count the number of samples in each element.

2.2.1 Illustration

In order to illustrate the statement we reconstruct the density for $T = 3$ samples for different discretizations. Figure 3 shows the samples and the corresponding empirical copula $\hat{\mathrm{C}}$. Note that the assumption, the samples lie at the grid, is not fulfilled but this is not essential, because this was only for convenience. Figures 4, 5 and 6 show the reconstructed piecewise



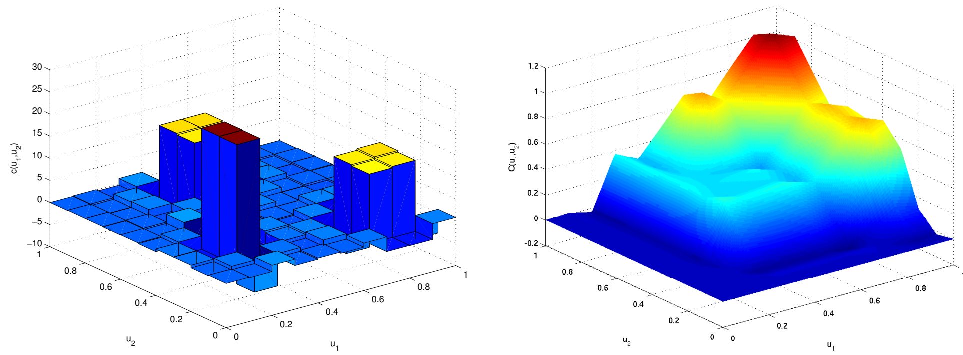
(a) Pseudo samples (b) empirical copula $\hat{\mathrm{C}}$

Figure 3: Input data

constant densities (4) for $n = 10$, $n = 30$ and $n = 100$ and the corresponding reconstructed copulas

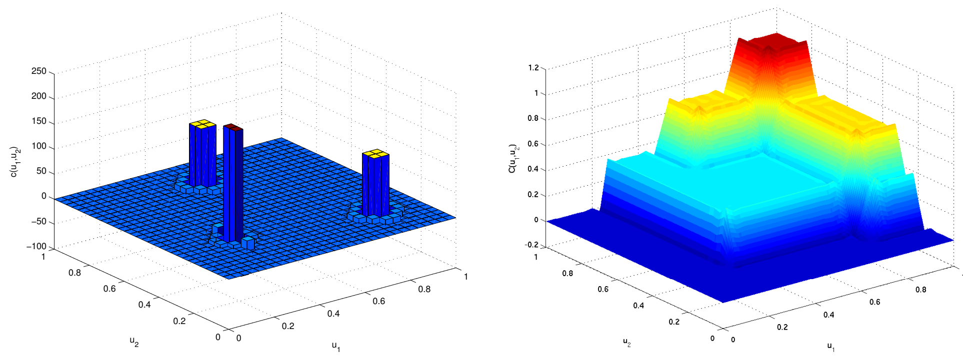
$$\mathrm{C}_h(\mathbf{u}) = \int_{\mathbf{0}}^{\mathbf{u}} \mathrm{c}_h(\mathbf{s})\mathrm{d}\mathbf{s} = \sum_{j=1}^{N} c_j \psi_j(\mathbf{u})\,, \tag{19}$$

if we use the integrated ansatz functions as test functions.



(a) c_h (b) C_h

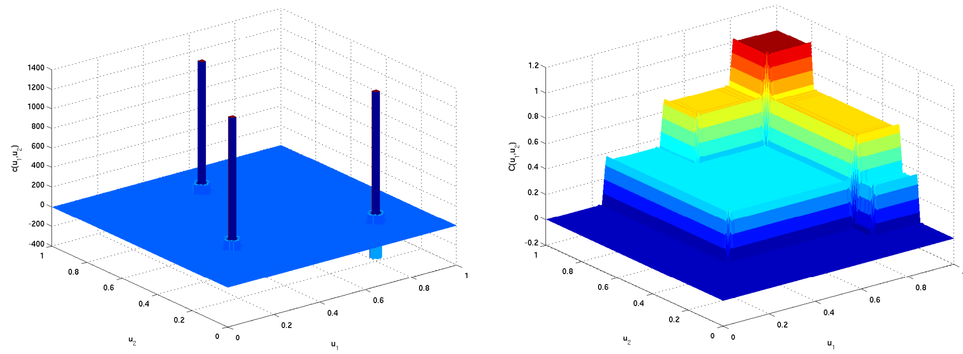
Figure 4: Reconstructed density and copula for $n = 10$



(a) c_h (b) C_h

Figure 5: Reconstructed density and copula for $n = 30$

The reconstructed copula (19) is a linear combination of the test functions, if they are chosen as the integrated ansatz functions. As a direct consequence the approximated copula is smoother than the approximated density, which is a natural property. Therefore, the choice (14) is not appropriate to the interrelation of a copula and their density. Actually, the numerical results are unstable if the computation is based on test functions (14). However, the



(a) c_h (b) C_h

Figure 6: Reconstructed density and copula for $n = 100$

purpose of this paper was to find arguments that the Galerkin projection is no simple counting algorithm, even if we use very simple test and ansatz functions.

References

[UU13] D. Uhlig and R. Unger. A Petrov-Galerkin projection for copula density estimation. Technical report, TU Chemnitz, 2013.