Copula Estimation via Maximum Mean Discrepancy

A review of Alquier et al. (2020) and Muandet et al. (2017)

Albert Rapp

WiSe 20/21

Ulm University, Institute of Stochastics

Overview

GOAL

Construct estimator $\hat{\theta}_n = \hat{\theta}(U_1, \ldots, U_n)$ from iid pseudoobservations

$$U_1 = \hat{F}_n(X_1) \sim C_{\theta_0}$$

Parametric family of copulas

We will assume that

$$C_{\theta_0} \in \{C_\theta, \theta \in \Theta\}$$

Possible contamination $\varepsilon \in (0, 1)$

$$C = (1 - \varepsilon) C_{\theta_0} + \varepsilon \tilde{C} \text{ where } \tilde{C} \neq C_{\theta_0}$$

Approach

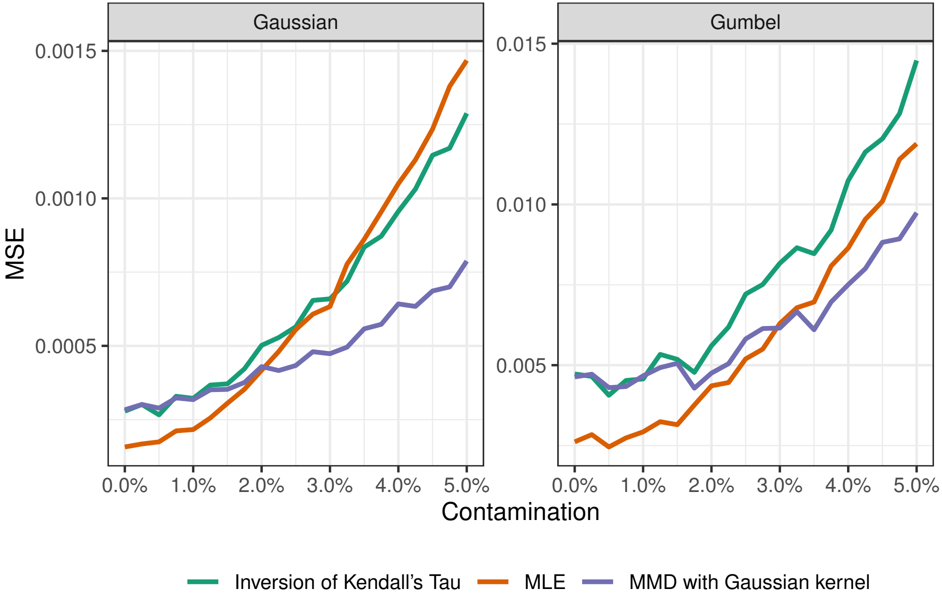
Minimize "distance" of empirical distribution $\hat{\mathbb{P}}_n$ and theoretical distribution $\mathbb{P}_\theta$

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} \mathbb{D}(\hat{\mathbb{P}}_n, \mathbb{P}_\theta)$$

Maximum mean discrepancy (MMD)

$$\mathbb{D}(\mathbb{P},\mathbb{Q}) = \sup_{f\in\mathcal{F}} \left| \int f \, d\mathbb{P} - \int f \, d\mathbb{Q} \right|,$$

where $\mathcal{F}$ is a universal reproducing kernel Hilbert space (RKHS)



Gaussian
Gumbel
MSE
0.0015
0.0010
0.0005
0.015
0.010
0.005
0.0%
1.0%
2.0%
3.0%
4.0%
5.0%
Contamination
Inversion of Kendall's Tau
MLE
MMD with Gaussian kernel

Maximum Mean Discrepancy

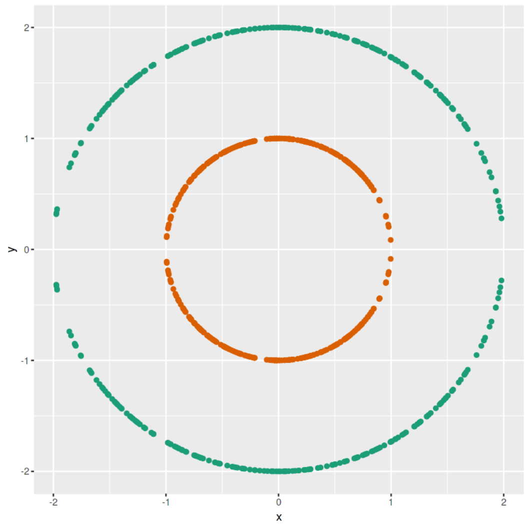
A quick tour through Muandet et al. (2017)

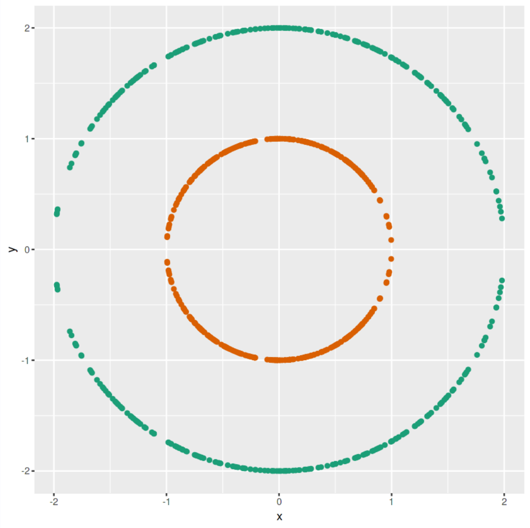
One measure of distance/similarity

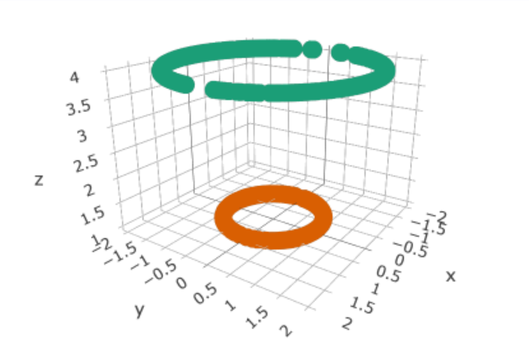
Suppose X and Y are two random variables with $\mathbb{E}X = \mathbb{E}Y = 0$, then

$$\mathrm{Cov}(X, Y) = \mathbb{E}XY =: \langle X, Y \rangle$$

is an inner product.


2
1
0
-1
-2
y
-2
-1
0
1
2
x





$$\phi : \mathbb{R}^2 \to \mathbb{R}^3, \quad (x, y) \mapsto \left(x, y, x^2 + y^2\right)$$

Inner products on suitable spaces

$$\langle x, y \rangle = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}, \text{ for all } x, y \in \mathcal{X}$$

where $\mathcal{H}$ is a Hilbert space.

Inner products on suitable spaces

$$\langle x,\, y \rangle = \langle \phi(x),\, \phi(y) \rangle_{\mathcal{H}}\,, \text{ for all } x, y \in \mathcal{X}$$

where $\mathcal{H}$ is a Hilbert space.

Positive definite kernel

Function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ with

$$k(x, y) = k(y, x) \quad \text{and} \quad \sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) \geq 0.$$

for all $x, y, x_1, \ldots, x_n \in \mathcal{X}$ and $c_1, \ldots, c_n \in \mathbb{R}$.

Reproducing kernel Hilbert spaces (RKHS)

Evaluation functionals $\mathcal{F}_x[f] := f(x)$ fulfill

$$\big|\mathcal{F}_x[f]\big| = |f(x)| \leq C\|f\|_{\mathcal{H}} \quad \text{for all } f \in \mathcal{H}, x \in \mathcal{X}$$

for some constant $C > 0$.

Reproducing kernel Hilbert spaces (RKHS)

Evaluation functionals $\mathcal{F}_x[f] := f(x)$ fulfill

$$\left|\mathcal{F}_x[f]\right| = |f(x)| \leq C\|f\|_{\mathcal{H}} \quad \text{for all } f \in \mathcal{H}, x \in \mathcal{X}$$

for some constant $C > 0$.

Reproducing property

For all $x \in \mathcal{X}$, there is a function $k_x \in \mathcal{H}$ such that

$$f(x) = \langle k_x,\ f\rangle_{\mathcal{H}}\,.$$

Representation of point x by

$$\phi(x) = k_x =: k(x, \cdot).$$

where $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is some two-variable function.

Representation of point x by

$$\phi(x) = k_x =: k(x, \cdot).$$

where $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is some two-variable function.

Reproducing property and feature maps

$$\langle \phi(x), \phi(y) \rangle_{\mathcal{H}} = \langle k_x, k_y \rangle_{\mathcal{H}} = k_y(x) = k(x, y),$$

Theorem

For every positive definite function $k(\cdot, \cdot)$ on $\mathcal{X} \times \mathcal{X}$ there exists a unique RKHS with k as its reproducing kernel.

Conversely, the reproducing kernel of an RKHS is unique and positive definite.

Point representation

$$x \in \mathcal{X} \mapsto k(x, \cdot) \in \mathcal{H}$$

Heuristic extension to Dirac measures

$$\int_{\mathcal{X}} y \; \delta_x(dy) = x \mapsto k(x, \cdot) = \int_{\mathcal{X}} k(y, \cdot) \; \delta_x(dy)$$

Representer of integral evaluation

$$\int_{\mathcal{X}} f(t)\,\delta_x(dt) = \int_{\mathcal{X}} \langle f,\, k(t,\cdot)\rangle_{\mathcal{H}}\,\delta_x(dt) = \left\langle f,\, \int_{\mathcal{X}} k(t,\cdot)\,\delta_x(dt) \right\rangle_{\mathcal{H}}$$

Dirac measure representation

$$\delta_x \in \mathcal{D} \mapsto \int_{\mathcal{X}} k(y,\cdot)\,\delta_x(dy) = k(x,\cdot) \in \mathcal{H}$$

STEP 1: Dirac measures

$$\delta_x \mapsto \int_{\mathcal{X}} k(y, \cdot)\, \delta_x(dy)$$

STEP 1: Dirac measures

$$\delta_x \mapsto \int_{\mathcal{X}} k(y,\cdot)\ \delta_x(dy)$$

STEP 2: Discrete measures

$$\mu = \sum_{i=1}^{n} a_i \delta_{x_i} \mapsto \sum_{i=1}^{n} a_i \int_{\mathcal{X}} k(y,\cdot)\ \delta_{x_i}(dy) = \int_{\mathcal{X}} k(y,\cdot)\ \mu(dy) \qquad (1)$$

STEP 1: Dirac measures

$$\delta_x \mapsto \int_{\mathcal{X}} k(y, \cdot)\ \delta_x(dy)$$

STEP 2: Discrete measures

$$\mu = \sum_{i=1}^{n} a_i \delta_{x_i} \mapsto \sum_{i=1}^{n} a_i \int_{\mathcal{X}} k(y, \cdot)\ \delta_{x_i}(dy) = \int_{\mathcal{X}} k(y, \cdot)\ \mu(dy) \tag{1}$$

STEP 3: In a RKHS $\mathcal{H}$, linear combinations as in (1) form a dense subset of $\mathcal{H}$

STEP 1: Dirac measures

$$\delta_x \mapsto \int_{\mathcal{X}} k(y,\cdot)\ \delta_x(dy)$$

STEP 2: Discrete measures

$$\mu = \sum_{i=1}^{n} a_i \delta_{x_i} \mapsto \sum_{i=1}^{n} a_i \int_{\mathcal{X}} k(y,\cdot)\ \delta_{x_i}(dy) = \int_{\mathcal{X}} k(y,\cdot)\ \mu(dy) \qquad (1)$$

STEP 3: In a RKHS $\mathcal{H}$, linear combinations as in (1) form a dense subset of $\mathcal{H}$

STEP 4: Probability measures

$$\mathbb{P} \mapsto \int_{\mathcal{X}} k(y,\cdot)\ \mathbb{P}(dy)$$

Kernel Mean Embedding

$$\mu : M_+^1(\mathcal{X}) \to \mathcal{H}, \quad \mathbb{P} \mapsto \int_{\mathcal{X}} k(y, \cdot)\ \mathbb{P}(dy)$$

Existence

$$\mathbb{E}_{\mathbb{P}} \sqrt{k(X, X)} < \infty \Longrightarrow \begin{cases} \mu_{\mathbb{P}} \in \mathcal{H} \text{ and} \\ \mathbb{E}_{\mathbb{P}}[f(X)] = \langle f, \mu_{\mathbb{P}} \rangle_{\mathcal{H}} \end{cases}$$

Example

Let $k(x, y) = \exp\{xy\}$, $x, y \in \mathbb{R}$, and $Y \sim \mathbb{P}$

$$\implies \mu_{\mathbb{P}} = \int_{\mathcal{X}} k(y, \cdot)\ \mathbb{P}(dy) = \mathbb{E}k(Y, \cdot) = \mathbb{E}e^{\cdot Y}$$

Thus, $\mu_P(s) = \mathbb{E}e^{sY}$ is moment-generating function.

$$\mathbb{E}_{\mathbb{P}}\sqrt{k(X,X)} < \infty \implies \begin{cases} \mu_{\mathbb{P}} \in \mathcal{H} \text{ and} \\ \mathbb{E}_{\mathbb{P}}[f(X)] = \langle f, \mu_{\mathbb{P}} \rangle_{\mathcal{H}} \end{cases}$$

Expectation is bounded operator

For operator $L_{\mathbb{P}} f := \mathbb{E} f(X)$ where $X \sim \mathbb{P}$ it holds

$$
\begin{aligned}
|L_{\mathbb{P}} f| &= |\mathbb{E} f(X)| \leq \mathbb{E}|f(X)| \\
&= \mathbb{E}|\langle f, k(X, \cdot)\rangle_{\mathcal{H}}| \leq \mathbb{E}\sqrt{k(X, X)}\|f\|_{\mathcal{H}}
\end{aligned}
$$

Expectation is bounded operator

For operator $L_{\mathbb{P}} f := \mathbb{E} f(X)$ where $X \sim \mathbb{P}$ it holds

$$
\begin{aligned}
|L_{\mathbb{P}} f| &= |\mathbb{E} f(X)| \leq \mathbb{E}|f(X)| \\
&= \mathbb{E}|\langle f, k(X, \cdot)\rangle_{\mathcal{H}}| \leq \mathbb{E}\sqrt{k(X, X)}\|f\|_{\mathcal{H}}
\end{aligned}
$$

Riesz representation

For all $f \in \mathcal{H}$, there is a $h \in \mathcal{H}$ such that $L_{\mathbb{P}} f = \langle f, h\rangle_{\mathcal{H}}$

$$
\implies h(x) = \langle k(x, \cdot), h\rangle_{\mathcal{H}} = L_{\mathbb{P}} k(x, \cdot) = \mathbb{E} k(x, X) = \int k(x, y)\, \mathbb{P}(dy)
$$

Expectation is bounded operator

For operator $L_{\mathbb{P}} f := \mathbb{E} f(X)$ where $X \sim \mathbb{P}$ it holds

$$
\begin{aligned}
|L_{\mathbb{P}} f| &= |\mathbb{E} f(X)| \leq \mathbb{E}|f(X)| \\
&= \mathbb{E}|\langle f, k(X,\cdot)\rangle_{\mathcal{H}}| \leq \mathbb{E}\sqrt{k(X,X)}\|f\|_{\mathcal{H}}
\end{aligned}
$$

Riesz representation

For all $f \in \mathcal{H}$, there is a $h \in \mathcal{H}$ such that $L_{\mathbb{P}} f = \langle f, h\rangle_{\mathcal{H}}$

$$
\implies h(x) = \langle k(x,\cdot), h\rangle_{\mathcal{H}} = L_{\mathbb{P}} k(x,\cdot) = \mathbb{E} k(x,X) = \int k(x,y)\,\mathbb{P}(dy)
$$

$$
\implies h = \int k(\cdot,y)\,\mathbb{P}(dy) = \mu_{\mathbb{P}}
$$

Integral Probability Measure (IPM)

$$\gamma[\mathcal{F}, \mathbb{P}, \mathbb{Q}] = \sup_{f \in \mathcal{F}} \left\{ \int_{\mathcal{X}} f(x) \ \mathbb{P}(dx) - \int_{\mathcal{X}} f(x) \ \mathbb{Q}(dx) \right\}$$

where $\mathcal{F}$ is a space of real-valued bounded measurable functions.

Integral Probability Measure (IPM)

$$\gamma[\mathcal{F}, \mathbb{P}, \mathbb{Q}] = \sup_{f \in \mathcal{F}} \left\{ \int_{\mathcal{X}} f(x) \; \mathbb{P}(dx) - \int_{\mathcal{X}} f(x) \; \mathbb{Q}(dx) \right\}$$

where $\mathcal{F}$ is a space of real-valued bounded measurable functions.

Maximum Mean Discrepancy

IPM on $\mathcal{F} = \{f | \; \|f\|_{\mathcal{H}} \leq 1\}$ where $\mathcal{H}$ is a RKHS

$$
\begin{aligned}
\mathrm{MMD}[\mathcal{H}, \mathbb{P}, \mathbb{Q}] &= \sup_{\|f\| \leq 1} \left\{ \int_{\mathcal{X}} f(x) \; \mathbb{P}(dx) - \int_{\mathcal{X}} f(x) \; \mathbb{Q}(dx) \right\} \\
&= \sup_{\|f\| \leq 1} \left\{ \left\langle f, \int_{\mathcal{X}} k(x, \cdot) \; \mathbb{P}(dx) \right\rangle - \left\langle f, \int_{\mathcal{X}} k(x, \cdot) \; \mathbb{Q}(dx) \right\rangle \right\} \\
&= \sup_{\|f\| \leq 1} \left\{ \langle f, \mu_{\mathbb{P}} - \mu_{\mathbb{Q}} \rangle \right\} \\
&= \|\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}\|_{\mathcal{H}}
\end{aligned}
$$

Relating Distance to Expectation

If $X \sim \mathbb{P}$, then

$$\|\mu_{\mathbb{P}}\|_{\mathcal{H}}^2 = \langle \mathbb{E}k(X, \cdot),\, \mu_{\mathbb{P}} \rangle_{\mathcal{H}} = \mathbb{E}k(X, \tilde{X}),$$

where $\tilde{X}$ is an independent copy of X.

Relating Distance to Expectation

If $X \sim \mathbb{P}$, then

$$\|\mu_{\mathbb{P}}\|_{\mathcal{H}}^2 = \langle \mathbb{E}k(X, \cdot), \mu_{\mathbb{P}} \rangle_{\mathcal{H}} = \mathbb{E}k(X, \tilde{X}),$$

where $\tilde{X}$ is an independent copy of X.

$$\begin{aligned} \mathrm{MMD}[\mathcal{H}, \mathbb{P}, \mathbb{Q}]^2 &= \|\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}\|_{\mathcal{H}}^2 \\ &= \langle \mu_{\mathbb{P}} - \mu_{\mathbb{Q}}, \mu_{\mathbb{P}} - \mu_{\mathbb{Q}} \rangle_{\mathcal{H}} \\ &= \|\mu_{\mathbb{P}}\|_{\mathcal{H}}^2 - 2\langle \mu_{\mathbb{P}}, \mu_{\mathbb{Q}} \rangle_{\mathcal{H}} + \|\mu_{\mathbb{Q}}\|_{\mathcal{H}}^2 \\ &= \mathbb{E}k(X, \tilde{X}) - 2\mathbb{E}k(X, Y) + \mathbb{E}k(Y, \tilde{Y}), \end{aligned}$$

where $X, \tilde{X} \sim \mathbb{P}$ and $Y, \tilde{Y} \sim \mathbb{Q}$ are independent copies.

Copula Estimation via MMD

Approach

Minimize

$$\mathrm{MMD}[\mathcal{F}, \hat{\mathbb{P}}_n, \mathbb{P}_\theta]^2 = \int k(u, v)\ \mathbb{P}_\theta(du)\mathbb{P}_\theta(dv)$$

$$- 2\int k(u, v)\ \mathbb{P}_\theta(du)\hat{\mathbb{P}}_n(dv)$$

$$+ \int k(u, v)\ \hat{\mathbb{P}}_n(du)\hat{\mathbb{P}}_n(dv)$$

$$\mathrm{MMD}[\mathcal{H}, \mathbb{P}, \mathbb{Q}]^2 = \mathbb{E}k(X, \tilde{X}) - 2\mathbb{E}k(X, Y) + \mathbb{E}k(Y, \tilde{Y})$$

Empirical Measure for Pseudoobserverations

$$\hat{\mathbb{P}}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\hat{U}_i}$$

Estimator

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} \left\{ \int k(u, v) \; \mathbb{P}_\theta(du)\mathbb{P}_\theta(dv) - \frac{2}{n} \sum_{i=1}^{n} \int k(u, \hat{U}_i)\mathbb{P}(du) \right\}$$

Non-asymptotic guarantees

Let $k \in C^2([0,1]^d)$. Then, with probability $1-\delta-\nu \in (0,1)$ where $\delta, \nu > 0$

$$\text{MMD}[\mathcal{F}, \mathbb{P}_{\hat{\theta}_n}, \mathbb{P}_0] \leq \inf_{\theta \in \Theta} \text{MMD}[\mathcal{F}, \mathbb{P}_\theta, \mathbb{P}_0]$$

$$+ \left(\frac{8}{n} \sup_{u \in [0,1]^d} k(u,v) \right)^{1/2} \left\{ 1 + \sqrt{-\log \delta} \right\}$$

$$+ \left(\frac{4d^2}{n} \| d^{(2)} k \|_\infty \log \frac{2d}{\nu} \right)^{1/2}$$

Under suitable conditions we have

(a) **Strong Consistency**

$$\hat{\theta}_n \xrightarrow{\mathbb{P}_0\text{-a.s.}} \theta_0 \quad \text{as } n \to \infty$$

(b) **Asymptotic Normality**

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\text{d}} \mathcal{N}(0, \Sigma) \quad \text{as } n \to \infty,$$

where Σ is the asymptotic covariance matrix of the limiting distribution.

References

P. Alquier, B.-E. Chérief-Abdellatif, A. Derumigny, and J.-D. Fermanian. Estimation of copulas via maximum mean discrepancy. 2020.

K. Muandet, K. Fukumizu, B. Sriperumbudur, and B. Schölkopf. Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends® in Machine Learning*, 10(1-2):1–141, 2017. ISSN 1935-8245. doi: 10.1561/2200000060. URL http://dx.doi.org/10.1561/2200000060.