COPULAS BY RATIO-ESTIMATION

David Huk

1 Introduction

Copulas are important in multivariate density estimation. [...]

Many methods have been proposed, but many rely on parametric forms or restrictive assumptions. [...]

Here we propose a non-parametric estimation method of copula densities which only relies on samples from the joint density based on density-ratio estimation (DRE). [...]

2 Copulas and Density Ratio Estimation

2.1 Copulas on a data scale

A copula is a multivariate object identified through Sklar's theorem. [...]

2018 Elements of Copula modeling with R - book
Is a book written by the authors of the 'copula' R package. Reading through it.

2.2 DRE by Classification

Ratios of densities are prevalent in statistical models and applications, such as [...]

The estimation of those ratios is thus a crucial inference problem with many proposed approaches, most notably as framed by a classification task. [...]

The usual loss function to train the DRE estimator $\hat{r}$ is the logistic loss of the form:

$$\mathcal{L}(\boldsymbol{\theta}) = -\mathbb{E}_{\mathbf{x}_1 \sim p} \log\left(\frac{r\left(\mathbf{x}_1; \boldsymbol{\theta}\right)}{1 + r\left(\mathbf{x}_1; \boldsymbol{\theta}\right)}\right) - \mathbb{E}_{\mathbf{x}_2 \sim q} \log\left(\frac{1}{1 + r\left(\mathbf{x}_2; \boldsymbol{\theta}\right)}\right) \tag{1}$$

with other alternatives being possible. [...]

2020 Telescoping Density-Ratio Estimation (TRE) - link
Density-ratio estimation focuses on learing p/q by methods such as classification and is widely used. It fails to accurately estimate ratios p/q for which the two densities differ significantly, that is their KL is too large. The classifier then has too simple a task of distinguishing between them and gives no/little information on the ratio. Authors call this phenomenon the *density chasm*.
To remedy this, they construct a telescoping product of bridging density ratios

$$\frac{p_0(\mathbf{x})}{p_m(\mathbf{x})} = \frac{p_0(\mathbf{x})}{p_1(\mathbf{x})} \frac{p_1(\mathbf{x})}{p_2(\mathbf{x})} \cdots \frac{p_{m-2}(\mathbf{x})}{p_{m-1}(\mathbf{x})} \frac{p_{m-1}(\mathbf{x})}{p_m(\mathbf{x})}$$

where each p_k is close to its neighbours so that the classifiers for each ratio are informative. They need (i) to construct waymark samples x_k and (ii) to construct ratio-estimating models for each bridge ratio based on those samples. The waymarks are constructed with simple 'interpolations-like' ideas. The ratio models have a shared NN body with model-specific final layers giving the log-ratios. They are trained by an averaged logistic loss for all models.

2021 Featurized Density Ratio Estimation (F-DRE) - link
Wanting to address the issue of the *density-chasm* in DRE methods, authors propose to first learn a normalising flow embedding f_θ jointly of data from p and q, to then perform the classification-based training of the ratio on the densities in the embedded space (call them p' and q'). By using a normalising flow, its invertibility ensures that the ratio on the data and embedded spaces are equal:

$$\frac{p(\mathbf{x})}{q(\mathbf{x})} = \frac{p'\left(f_\theta(\mathbf{x})\right)}{q'\left(f_\theta(\mathbf{x})\right)}$$

The flow brings the densities closer together, thereby avoiding situations where their dissimilarity causes the ratio to be hard to estimate. This could potentially be useful as a method for sampling from the ratio, by selecting a sample from the correct region in the latent embedding and mapping it to the data-space.

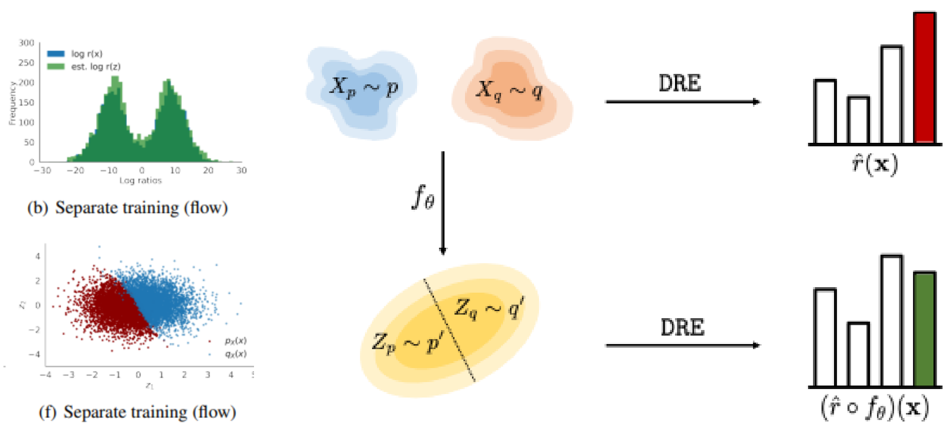

Figure 1: F-DRE figures from their paper.

2022 Density Ratio Estimation via Infinitesimal Classification (DRE -∞) - link
As a follow-up work to TRE, authors propose to interpolate smoothly between the two densities, thereby building a bridge continuum between p and q. As such, DRE-∞ can be seen as the limit case of TRE, when the number of bridges goes to infinity.
The key insight is a well-known identity in path-sampling litterature:

$$\log r(\mathbf{x}) = \log \frac{p_0(\mathbf{x})}{p_1(\mathbf{x})} = \lim_{T \to \infty} \sum_{t=1}^{T} \log \left(\frac{p_{(t-1)/T}(\mathbf{x})}{p_{t/T}(\mathbf{x})} \right) = \int_1^0 \frac{\partial}{\partial \lambda} \log p_\lambda(\mathbf{x}) d\lambda.$$

They begin by constructing way-mark samples between p and q dependent on a time index $t \in [0,1]$, thereby obtaining an infinite number of bridge densities $p_t(\mathbf{x})$ (for example $\mathbf{x}(t) := \sqrt{1-\alpha(t)^2}\mathbf{x}_p + \alpha(t)\mathbf{x}_q$). Then, for each $\mathbf{x}$, they estimate (with a score-matching objective) the rate of change, termed *time-score* $\frac{\partial}{\partial t} \log p_t(\mathbf{x})$ which measures (at $\mathbf{x}$) how each bridging density p_t is changing along a prescribed trajectory in distribution space over time t. With the *time-score*, one can compute the ratio $r(\mathbf{x})$ as a simple $1D$ integral in t, making it very efficient. They show improved performance over TRE and other ratio estimation methods, especially in the case of more complex and multivariate data.

3 DRE for Copula Density Estimation

Consider observations $\mathbf{x}^k = (x_1^k, \ldots, x_n^k) \in \mathbb{R}^n$ for $k \in \{1, \ldots, K\}$ coming from a joint density p with each x_i^k following a given marginal density p_i (with CDFs P_i) with $1 \leq i \leq n$. By Sklar's theorem [ref], it follows that p admits a copula representation of the form:

$$p(\mathbf{x}) = c\left(\mathbf{x}\right) \cdot \prod_{i=1}^{n} \left\{ p_i(x_i) \right\}$$

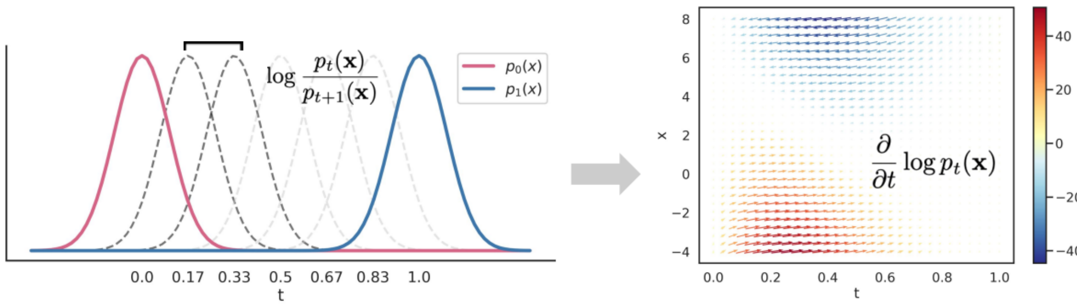

Figure 2: DRE-∞ figure from their paper.

where c is the unique copula corresponding to the joint density, as defined in [ref above]. Notice that the copula density can be rewritten as:

$$c(\mathbf{x}) = \frac{p(\mathbf{x})}{q(\mathbf{x})} \tag{2}$$

with p still being the joint density and $q(\mathbf{x}) := \prod_{i=1}^{n} \{p_i(x_i)\}$ being an independent distribution for $\mathbf{x}$ across dimension.

We take advantage of this identification to rephrase the inference of a copula density into a density-ratio estimation problem. This allows us to make use of the extensive methodology developed for such problems. Furthermore, guarantees on density-ratio estimation ensure convergence to the correct copula expression, as shown in the following result [NCE 2012; LFIRE 2022].

Proposition 1 *Minimising the logistic loss of Equation 1 recovers* $\hat{r}(\mathbf{x}; \boldsymbol{\theta}^*) = \ln c(\mathbf{x})$ *as its unique minimiser.*

4 Bayesian Inference with Copulas

The relationship between copulas and ratios can also be exploited in the other direction, that is treating what was up to now thought to be only ratios as copulas. One such case where this identification is valid is the Bayesian inference framework, where the object of interest is the likelihood-to-evidence ratio $\frac{p(\theta|x)}{p(\theta)} = \frac{p(\theta,x)}{p(\theta)p(x)}$. [...]

Contrastive Neural Ratio Estimation (CNRE) - link

5 Extensions

5.1 Sampling from ratio copulas

While we have convergence guarantees for the ratio copula estimator, we only obtain an approximation of its density. Here are a few ideas to sample from it. [...]

2021 Featurized Density Ratio Estimation (F-DRE) As mentioned above, one could then sample from the latent Gaussian embedding, and conditional on the classifier identifying a sample as coming from p, we map the sample back through the normalising flow to obtain a sample on the data-scale.

2020 Semi-Supervised Learning with Normalizing Flows - link
Authors use a normalising flow to map samples into a latent space where they perform Gaussian Mixture Modelling on the different clusters within the data. For the copula ratio setup, we have 2 clusters, one for the joint and one for independent marginals. then one can sample conditionally from a cluster and map it back to an observation scale.

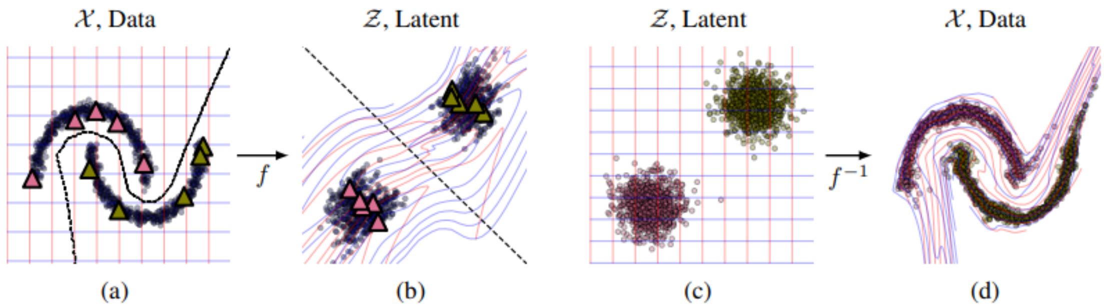

Figure 3: Figure from Semi-Supervised Learning with Normalizing Flows paper.

6 Experiments

6.1 Gaussian copula

6.2 Clayton copula with Gamma and Weibul margins

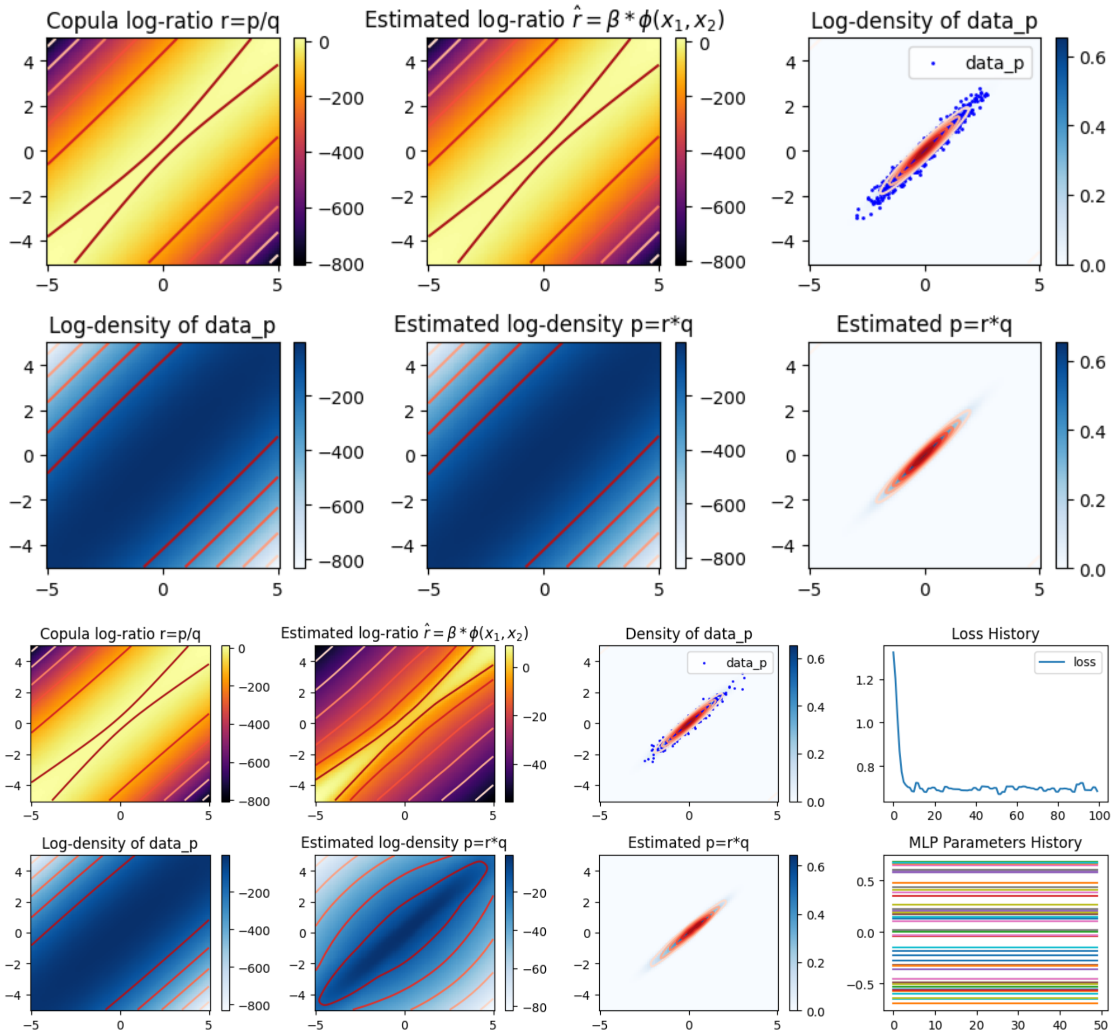

Figure 4

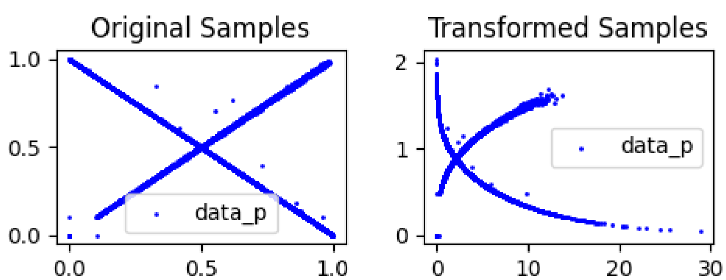

Figure 5

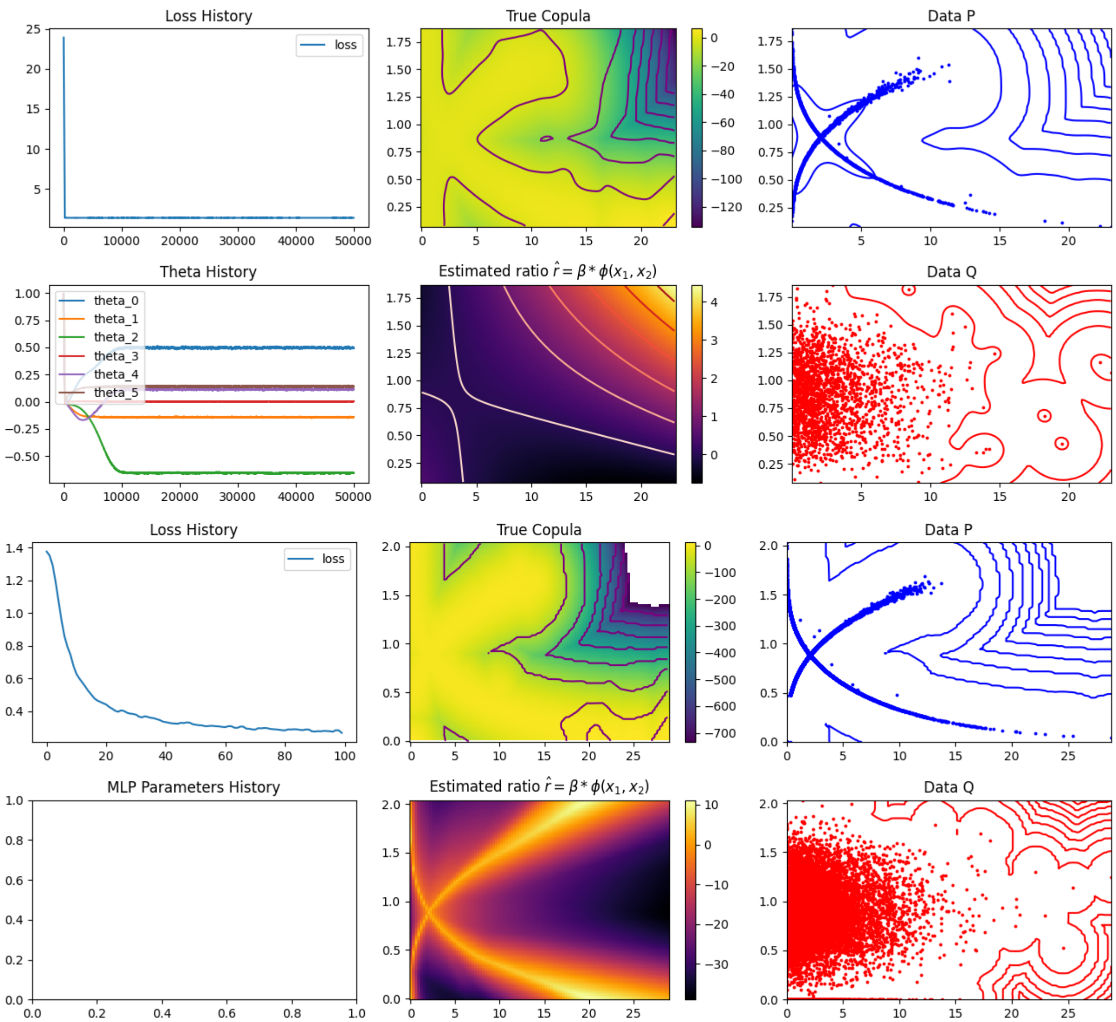
Loss History
loss
True Copula
Data P
Theta History
theta_0
theta_1
theta_2
theta_3
theta_4
theta_5
Estimated ratio $\hat{r} = \beta * \phi(x_1, x_2)$
Data Q
Loss History
loss
True Copula
Data P
MLP Parameters History
Estimated ratio $\hat{r} = \beta * \phi(x_1, x_2)$
Data Q

Figure 6