# An indirect nonparametric regression method for one-dimensional continuous distributions using warping functions

Zhicheng Chen[*a]

[a]School of Civil Engineering, Harbin Institute of Technology, Harbin, 150090, China

**ABSTRACT**

Distributions play a very important role in many applications. Inspired by the newly developed warping transformation of distributions, an indirect nonparametric distribution to distribution regression method is proposed in this article for predicting correlated one-dimensional continuous probability density functions.

**Keywords:** distribution regression, warping transformation, nonparametric regression

## 1. Introduction

In this article, the correlation of two distribution classes means their density functions will vary simultaneously to some extent (see Fig. 1). Note the correlation of two distribution classes is indifferent with the correlation of two random variables. Two correlated distribution classes do not guarantee random variables respectively follow these two distribution classes are also correlated. For instance, suppose probability density functions $g_t$ and $f_t$ are correlated, consider two time-varying random variables $X_t \sim g_t$ and $Y_t \sim f_t$, if the joint distribution of $(X_t, Y_t)$ can be factorized as $p_{X_t Y_t}(x_t, y_t) = g_t(x_t)\, f_t(y_t)$, then $X_t$ and $Y_t$ are independent with each other. Similarly, two correlated random variables also do not guarantee the distributions they follow are also correlated.

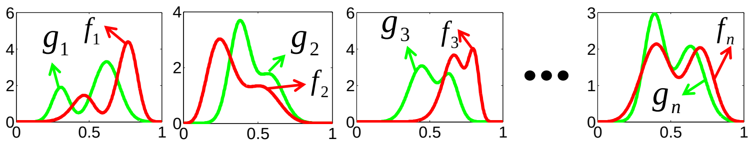


Fig.1 Graphical representation of two correlated distribution families

Inspired by the seminal work done by Dasgupta *et al.* [1], an indirect nonparametric distribution to distribution regression method is proposed in this article for two correlated one-dimensional continuous distribution families. Other related work includes traditional distribution to distribution regression [2] and distribution to real-value or vector-value regression [3-7], etc.

## 2. Introduction to the warping transformation of distributions

The warping transformation of a distribution is a map that used to transform a distribution to another by deforming the original probability density function with a warping function [1, 8]. The newly reported article by Dasgupta *et al.* [1] has given a very detailed discussion of this transformation, where the warping transformation of distributions is used in their proposed two-step distribution estimation framework. Given observed samples of an unknown distribution and an initial estimate of the objective density function, they have developed a maximum-likelihood-based approach to find an optimal warping function to transform the initial estimate to reach the final estimate of the objective density function. Here, we just give a very brief introduction of the warping transformation of distributions, for detailed discussion we refer the reader to the original paper [1].

* 13B933002@hit.edu.cn; phone 86 451 86282068; fax 86 451 86282068.

Here, we restrict the studied distributions have finite support on $[0,1]$. Given a probability density function $g(x)$ with support on $[0,1]$, the warping transformation of $g(x)$ by a warping function $\gamma(x)$ defined on $[0,1]$ is

$$g_{warp}(x) = g(\gamma(x))\dot{\gamma}(x), \quad x \in [0,1] \tag{1}$$

where $\dot{\gamma}(x) = \frac{d}{dx}\gamma(x)$.

## 3. The related work: conditional density estimation using warping functions

The related work is the estimation of the conditional density function given the observation of a correlated random variable. Here, following the description in the original paper [1]. Let $X$ be a d-dimensional random variable, such as $X \sim \mathrm{U}[0,1]^d$, and $Y$ be another random variable that correlated with $X$.

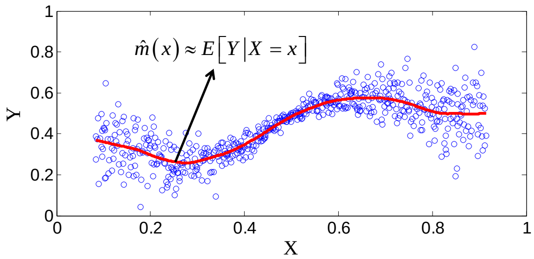


Fig. 2 Observed samples of two correlated random variables $X$ and $Y$ with estimated mean function $\hat{m}(x) \approx E[Y|X=x]$

Suppose we have observed $n$ pairs realizations of $X$ and $Y$, i.e. $\{(X_i, Y_i)\}_{i=1}^n$. Let $f_{y|X=x_0}$ be the conditional density function of $Y$ given $X = x_0$, the task in the original paper is to obtain a warped estimate for $f_{y|X=x_0}$, i.e. $\hat{f}_{y|X=x_0} = f_{w,x_0}$, from the initial density function $f_p(y|\hat{m}(x_0), \hat{\sigma}^2)$, where $\hat{m}(x)$ is the estimated mean function of $Y$ obtained by local linear regression (see Fig. 2), $\hat{\sigma}^2$ is the estimated variance of the residuals $\{Y_i - \hat{m}(X_i)\}_{i=1}^n$. The initial density function $f_p$ can be any parametric family with simple estimation procedure. Suppose $\gamma_0(x)$ is the warping function from $f_p(y|\hat{m}(x_0), \hat{\sigma}^2)$ to $f_{w,x_0}$, i.e.

$$f_{w,x_0}(y|X = x_0) = f_p(\gamma_0(y)|m(x_0), \hat{\sigma}^2)\dot{\gamma}_0(y) \tag{2}$$

The optimal estimate for $\gamma_0$ in the original paper is obtained by the following weighted likelihood method

$$\hat{\gamma}_0 = \underset{\gamma \in \Gamma}{\arg\max}\left(\sum_{i=1}^{n} \log\left(f_p(\gamma(y_i)|\hat{m}(x_i), \hat{\sigma}^2)\dot{\gamma}(y_i)\right) W_{x_0,i}\right) \tag{3}$$

where $f_p(\cdot|\hat{m}(x_i), \hat{\sigma}^2)$ is the initial density function for $Y|X = x_i$, $W_{x_0,i}$ is the local weight at $X_i$, calculated by

$$W_{x_0,i} = \frac{N\left(\|X_i - x_0\|_2 / h | 0,1\right)}{\sum_{j=1}^{n} N\left(\|X_j - x_0\|_2 / h | 0,1\right)} \tag{4}$$

where $N(\cdot \mid 0,1)$ is the standard normal probability density function, $h$ is the bandwidth parameter, $\|\cdot\|_2$ is the 2-norm.

This warping transformation-based approach for conditional distribution estimation in the original paper can be regarded as real value to distribution regression (if $X$ is an univariate random variable) or vector to distribution regression (if $X$ is a multivariate random variable). It can predict the dynamic change of the density function of $Y$ only when $X$ and $Y$ are correlated with each other.

## 4. Extend to distribution to distribution regression

In this section, the warping-transformation-based method is extended to distribution to distribution regression. Here, we also restrict the studied distributions have finite support on $[0,1]$, distributions with general finite supports can be easily tackled by the scale transformation introduced in [1].

For convenience, let $\boldsymbol{\pi}_g$ and $\boldsymbol{\pi}_f$ respectively be two correlated one-dimensional continuous distribution families with finite support on $[0,1]$. Suppose, we have obtained $n$ pairs of probability density functions respectively from $\boldsymbol{\pi}_g$ and $\boldsymbol{\pi}_f$, i.e. $\{g_i, f_i\}_{i=1}^n$, given a new density function $g_0$ from $\boldsymbol{\pi}_g$, the task in this section is to develop a regression model to predict the corresponding density function $f_0$ from $\boldsymbol{\pi}_f$, i.e. use $\{g_i, f_i\}_{i=1}^n \cup g_0$ to predict $f_0$. For this purpose, a nonparametric distribution to warping function regression method is first used to predict the warping function $\gamma_0$, i.e. the map relationship from $g_0$ to $f_0$, then use the predicted warping function to transform $g_0$ to obtain a prediction for $f_0$, i.e.

$$\hat{\gamma}_0 = \sum_{k=1}^{n} \hat{\gamma}_k \frac{K\left(\delta(g_0, g_k)/h\right)}{\sum_{j=1}^{n} K\left(\delta(g_0, g_j)/h\right)} \tag{5a}$$

$$\hat{f}_0(x) = g_0\left(\hat{\gamma}_0(x)\right)\dot{\hat{\gamma}}_0(x) \tag{5b}$$

where, $\hat{\gamma}_k$ is the estimated warping function from $g_k$ to $f_k$, i.e. $f_k(x) \approx g_k\left(\dot{\gamma}_k(x)\right)\dot{\gamma}_k(x)$, $K(\cdot)$ is the kernel function, $h$ is the bandwidth for the kernel regression, $\delta(g_0, g_k)$ is a metric (such as the $L_1$ distance: $\delta(g_0, g_k) = \int \left|g_0(\tau) - g_k(\tau)\right| d\tau$ ) used to measure the similarity between $g_0$ and $g_k$. Note, a convex combination of warping functions is also a warping function [8], thus the regression result being a warping function is guaranteed.

This distribution prediction approach can be regarded as indirect nonparametric distribution to distribution regression. In different with the kernel distribution to distribution regression method in [2], the proposed regression model in Eq. (5) can reflect the true map relationship between two distribution families, thus has more potential for extrapolation forecasting.

## 5. Conclusions

An indirect nonparametric distribution to distribution regression method is proposed in this article, which can reflect the true map relationship between two distribution families. The correlation of two distribution classes is indifferent with the correlation of two random variables. The real value (or vector) to distribution regression and the distribution to distribution regression have different application scopes, the former is for correlated random variables, while the latter is for correlated distributions.

## 6. Acknowledgments

This work is financially supported by the National Basic Research Program of China (Grant No. 2013CB036305), the National Natural Science Foundation of China (Grant No. 51638007), the Ministry of Science and Technology of China (Grant No. 2015DFG82080) and the China Scholarship Council (CSC).

## References

[1] Dasgupta S, Pati D and Srivastava A. A Geometric Framework For Density Modeling. arXiv preprint arXiv:1701.05656; 2017.

[2] Oliva JB, Póczos B and Schneider JG. Distribution to Distribution Regression. In ICML (3) 2013; pp: 1049-1057.

[3] Póczos B, Singh A, Rinaldo A and Wasserman LA. Distribution-Free Distribution Regression. In AISTATS 2013; pp: 507-515.

[4] Oliva JB, Neiswanger W, Póczos B, Schneider JG and Xing EP. Fast Distribution To Real Regression. In AISTATS (Vol. 3, p. 2) 2014.

[5] Szabó Z, Gretton A, Póczos B and Sriperumbudur B. Consistent, two-stage sampled distribution regression via mean embedding. arXiv preprint arXiv:1402.1754; 2014.

[6] Szabó Z, Gretton A, Póczos B and Sriperumbudur BK. Two-stage sampled learning theory on distributions. In AISTATS 2015.

[7] Szabó Z, Sriperumbudur B, Póczos B and Gretton, A. Learning theory for distribution regression. Journal of Machine Learning Research 2016; 17(152), 1-40.

[8] Srivastava A and Klassen EP. Functional and Shape Data Analysis. Springer: New York; 2016.