Causal inference on networks under continuous treatment interference

Jun Zhao

Nankai University

Aug 20, 2024

Introduction

- Policy interventions may spill over across units and generate indirect effects. These effects, which are pervasive in many economic and social contexts, stem from interference which occurs when an agent's treatment indirectly affects other agents' outcomes (Cox, 1958).
- Understanding the mechanism of interference is therefore crucial for the optimal design of an intervention, because it allows policy-makers to leverage or reduce spillover effects and improve the overall policy effectiveness(Moffitt, 2001).

Notation

- **Players**: Let $\mathcal{N}$ be a sample of N agents or units. We assume that agents are nodes embedded in a network, and a link between two nodes exists if two agents interact in a way that the treatment on an agent has an effect also on the outcome of the other agent.
- **Network**: This interfering network can be represented by the adjacency matrix $\mathbf{A} \in A \subset \mathbb{R}^{\mathbb{N}\times\mathbb{N}}$, with element a_{ij} being a continuous value on the realm of positive real numbers representing the inward relationship intensity from agent i to agent j.
  - $\mathcal{N}_i = \{j \in N : a_{ij} = 1\}$ is the set of neighbours i has in network.
  - denote by $\mathcal{N}_{-i}$ the set containing all nodes other than i that are not in $\mathcal{N}_i$.
  - For each node i, we thus obtain a partition of the set of nodes $\mathcal{N}$ as $(i, \mathcal{N}_i, \mathcal{N}_{-i})$.

Notation

- We now denote by $Y_i \in \mathcal{Y}$ the observed outcome for agent i, and by Y the corresponding vector. We let $Z_i \in \mathcal{Z}$ be the continuous treatment received by agent i. Under the potential outcome framework, $Y_i(Z)$ is the potential outcome of unit i under the treatment vector Z in the whole network.
- Consider $X_i^{ind} \in \mathcal{X}^{ind}$ as the vector of K^{ind} individual-level covariates for agent i.
- Similarly, $X_i^{neigh} \in \mathcal{X}^{neigh}$ denotes the vector of K^{neigh} neighborhood covariates for agent i.
- (i) variables representing the structure of the neighborhood $\mathcal{N}_i$
- (ii) variables representing the composition of the neighborhood $\mathcal{N}_i$
- We further assume that the adjacency matrix A is fixed or does not vary between the time the treatment is measured and the time the outcome is realized.

The stable unit treatment on neighborhood value assumption

Assumption (Consistency)

There are no multiple versions of the treatment. Formally: $Y = Y_i(Z)$.

Assumption (First-Order Interference with Exposure Mapping)

Given a function $g : \mathcal{Z}^{N_i} \times \mathcal{A} \to \mathcal{G}, \forall Z_{\mathcal{N}_{-i}}, Z'_{\mathcal{N}_{-i}}$ *and* $Z_{\mathcal{N}_i}, Z'_{\mathcal{N}_i}$ *such that* $g(Z_{\mathcal{N}_i}; A) = g(Z'_{\mathcal{N}_i}; A)$, *the following equality holds:*
$Y(Z_i, Z_{\mathcal{N}_i}, Z_{\mathcal{N}_{-i}}) = Y(Z_i, Z'_{\mathcal{N}_i}, Z'_{\mathcal{N}_{-i}})$

Neighborhood treatment in a weighted directed network

- The specification of the exposure mapping function $g()$, defining the neighborhood treatment G_i, depends on the mechanism of interference hypothesized for the treatment and outcome of interest.
- Most common definitions of the neighborhood treatment are the number of treated neighbors, i.e., $G_i = \sum_{j \in \mathcal{N}_i} Z_j$, or the proportion, i.e., $G_i = \sum_{j \in \mathcal{N}_i} \frac{Z_j}{\mathcal{N}_i}$.
- Given a weighted directed network, represented by the adjacency matrix A, we can express the neighborhood treatment as the following weighted sum:

$$G_i = \sum_{j \neq i} \frac{\omega_{ij}(A)}{C} Z_j$$

where $\omega_{ij}(A)$ is a weight function depending on the entries of the adjacency matrix A and C is a normalizing constant (e.g., N)

Causal estimands

Definition

Our formalization of the bivariate continuous joint treatment allows to model the potential outcome of unit i $Y_i(z, g)$ as a dose–response function. Therefore, we define the marginal mean of the potential outcome $Y_i(z, g)$, for each value of z and g, as the average dose–response function (aDRF), denoted by $\mu(z, g)$. Formally, let

$$\mu(z, g) = E[Y_i(z, g)]$$

Causal estimands

Definition

$\mu(z, g)$ can be marginalized to get the univariate average dose–response functions

$$\mu^Z(z) = \int_g E[Y_i(z, g)] p^G(g)\, dg$$

$$\mu^G(g) = \int_z E[Y_i(z, g)] p^Z(z)\, dz$$

where $p^G(g)$ and $p^Z(z)$ are the observed marginal densities of the neighborhood and individual treatments.

Causal estimands

Definition

Direct effects of the treatment can be defined as comparisons of the form $\delta(z, z') = \mu^Z(z) - \mu^Z(z')$, or as the first derivative of the average dose–response function $\delta(z, dz) = \frac{d\mu^Z(z)}{dz}$.

Definition

Spillover effects can be defined as the difference between the average potential outcome corresponding to two different levels of the neighborhood treatment g and g: $\delta(g, g') = \mu^G(g) - \mu^G(g')$, or as the first derivative of the average dose–response function $\delta(g, dg) = \frac{d\mu^G(g)}{dg}$.

Unconfoundedness of the joint treatment

Assumption (Unconfoundedness of the Joint Treatment)

Conditional on the vector of covariates X_i, the potential outcome $Y_i(z, g)$ is independent of the level of the treatments Z_i and G_i:

$$Y_i(z, g) \perp Z_i, G_i | X_i, \forall z, g \ \forall i$$

Joint propensity score-based estimator

We now discuss our joint propensity score-based estimator to obtain an unbiased estimate of both the treatment and the spillover effects. This estimator balances individual and neighborhood covariates across agents under different levels of individual and neighborhood treatments by controlling for the joint propensity score.

Given Assumption 3, thanks to the balancing property of the propensity scores, it follows that the assignment to the joint treatment is unconfounded conditional on both the individual and the neighborhood propensity scores (Forastiere et al., 2021). Formally, we can state the following proposition.

Joint propensity score-based estimator

Formally, we define the joint propensity score (JPS) $\psi(z; g; x)$ as the joint density of the individual treatment and network exposure conditional on covariates, that is, the relative likelihood of being subject to direct treatment z and being exposed to a weighted average of the treatments of the agent's connections equal to g, given characteristics $X_i = x$:

$$\psi(z; g; x) = P_{Z,G|X}(z, g|x) = P_{G|Z,X}(g|z, x)P_{Z|X}(z|x)$$

where $\phi(z; x) = P_{Z|X}(z|x)$ is the individual propensity score, i.e., the probability density function (PDF) of the individual treatment conditional on covariates, and $\lambda(g; z; x) = P_{G|Z,X}(g|z, x)$ is the neighborhood propensity score, i.e., the probability density function of the neighborhood treatment conditional on the value z of the individual treatment and on the vector of covariates X_i.

Joint propensity score-based estimator

Proposition

(Unconfoundedness of the Joint Treatment). Under Assumptions 1 and 2, if Assumption 3 holds, then

$$Y_i(z,g) \perp Z_i, G_i | \phi(z; X_i), \lambda(g; z; X_i), \forall z, g \ \forall i$$

Proposition

(Identification of Causal Estimands). Under Assumptions 1, 2 and 3, thanks to Proposition 1, causal quantities are identified from the observed data as follows:

$$\mu(z,g) = E[Y_i | Z_i = z, G_i = g, \phi(z; X_i), \lambda(g; z; X_i)]$$

$$\mu^Z(z) = E[Y_i | Z_i = z, G_i, \phi(z; X_i), \lambda(G_i; z; X_i)]$$

$$\mu^G(g) = E[Y_i | Z_i, G_i = g, \phi(z; X_i), \lambda(g; Z_i; X_i)]$$

Estimation procedure

Consider the following general models for the individual treatment Z, the neighborhood treatment G, and the outcome Y

- (1) $Z_i \sim f^Z(X_i, \theta^Z)$
- (2) $G_i \sim f^G(Z_i, X_i, \theta^G)$
- (3) $Y_i(z, g) \sim f^Y(z, g, \phi(z; X_i), \lambda(g; z; X_i), \theta^Y)$

Estimation procedure

- 1. Estimate the parameters θ^Z and θ^G of the models for the individual treatment in (1) and for the neighborhood treatment in (2);
- 2. Use the estimated parameters in Step 1 to predict for each unit $i \in \mathcal{N}$ the actual individual propensity score $\hat{\Phi} = \phi(z; X_i)$ and the actual neighborhood propensity score $\hat{\Lambda} = \lambda(g; z; X_i)$; that is, the PDFs of the individual treatment and neighborhood treatment, conditional on the covariates X_i, evaluated at the values Z_i and G_i that were actually observed for unit i;
- 3. Estimate the parameters θ^Y of the outcome model in (3) by using the observed data (Y_i, G_i, Z_i, X_i) and the predicted propensity scores $\hat{\Phi}_i$ and $\hat{\Lambda}_i$;

Estimation procedure

- 4. For each level of the joint treatment $(Z_i = z, G_i = g)$, predict for each unit $i \in \mathcal{N}$ the corresponding individual and the neighborhood propensity scores (i.e., $\phi(z; X_i)$ and $\lambda(g; z; X_i)$), and use these predicted values to impute the potential outcome $Y_i(z, g)$:

$$Y_i(z, g) \sim f^Y(z, g, \hat{\phi}(z; X_i), \hat{\lambda}(g; z; X_i), \theta^Y)$$

- 5. To estimate the average dose–response function $\mu(z, g)$, for each level of the joint treatment, take the average of the potential outcomes over all units:

$$\hat{\mu}(z; g) = \frac{1}{N} \sum_{i=1}^{N} \hat{Y}_i(z; g)$$

Estimation procedure

- 6. The univariate average dose–response functions are then obtained by averaging over the marginal densities $\hat{P}^G(g)$ and $\hat{P}^Z(z)$

$$\hat{\mu}^Z(z) = \int_g \hat{\mu}(z; g)\hat{P}^G(g)\, dg$$

$$\hat{\mu}^G(g) = \int_z \hat{\mu}(z; g)\hat{P}^Z(z)\, dz$$

In practice, given the continuous nature of Z and G, we use a grid of values $(\mathcal{Z}^*, \mathcal{G}^*)$, defined by the percentiles of the empirical distributions of Z and G. Therefore, steps 4 and 5 are conducted over the grid $(\mathcal{Z}^*, \mathcal{G}^*)$. The marginalization in step 6 is then performed as follows:

$$\hat{\mu}^Z(z) = \sum_{g \in \mathcal{G}} \hat{\mu}(z; g)\hat{P}(G_i = g)$$

$$\hat{\mu}^G(g) = \sum_{z \in \mathcal{Z}} \hat{\mu}(z; g)\hat{P}(Z_i = z)$$

Estimation procedure

Proposition

(Unbiasedness). If the individual and neighborhood treatment models in (1) and (2) as well as the outcome model in (3) are correctly specified, and an unbiased estimator of the model parameters θ is used in Steps 1 and 3, the estimation procedure, including Steps 1–6,results in an unbiased estimator of the causal quantities $\mu(z;g)$, $\mu^Z(z)$ and $\mu^G(g)$.

Estimation procedure

Propensity scores and outcome models can be estimated in Steps 1 and 3 using maximum likelihood estimation for generalized linear models. Instead, Hirano and Imbens (2004) use a simple linear regression for the generalized propensity score model and a flexible polynomial regression for the outcome model. However, other semi-parametric or non-parametric methods can be used. Zhu et al. (2014) propose the use of a tree-based boosting algorithm to estimate the generalized propensity score of a continuous treatment, while Bia et al. (2011) and Flores et al. (2012) propose penalized splines with tensor products or radial basis functions and a kernel estimator with a polynomial regression.
Standard errors and 95% confidence intervals can be derived using bootstrap methods, taking into account the uncertainty given by both data sampling and estimation of the propensity score models (Efron,1979)

Generalized Propensity Score in Stata

```
clear

use LotteryDataSet.dta, clear
qui gen     cut = 23  if prize<=23
qui replace cut = 80  if prize>23 & prize<=80
qui replace cut = 485 if prize >80

mat def tp = (10\20\30\40\50\60\70\80\90\100)

#delimit ;
doseresponse agew ownhs male tixbot owncoll workthen yearw  yearm1 yearm2 yearm3 yearm4 yearm5 yearm6,
outcome(year6) t(prize) gpscore(pscore) predict(hat_treat) sigma(sd) cutpoints(cut)
index(p50) nq_gps(5)  t_transf(ln)  dose_response(dose_response)
tpoints(tp) delta(1)  reg_type_t(quadratic) reg_type_gps(quadratic)  interaction(1)
bootstrap(yes) boot_reps(100)  filename("output")  analysis(yes) graph("graph_output") det
;
#delimit cr
clear

```

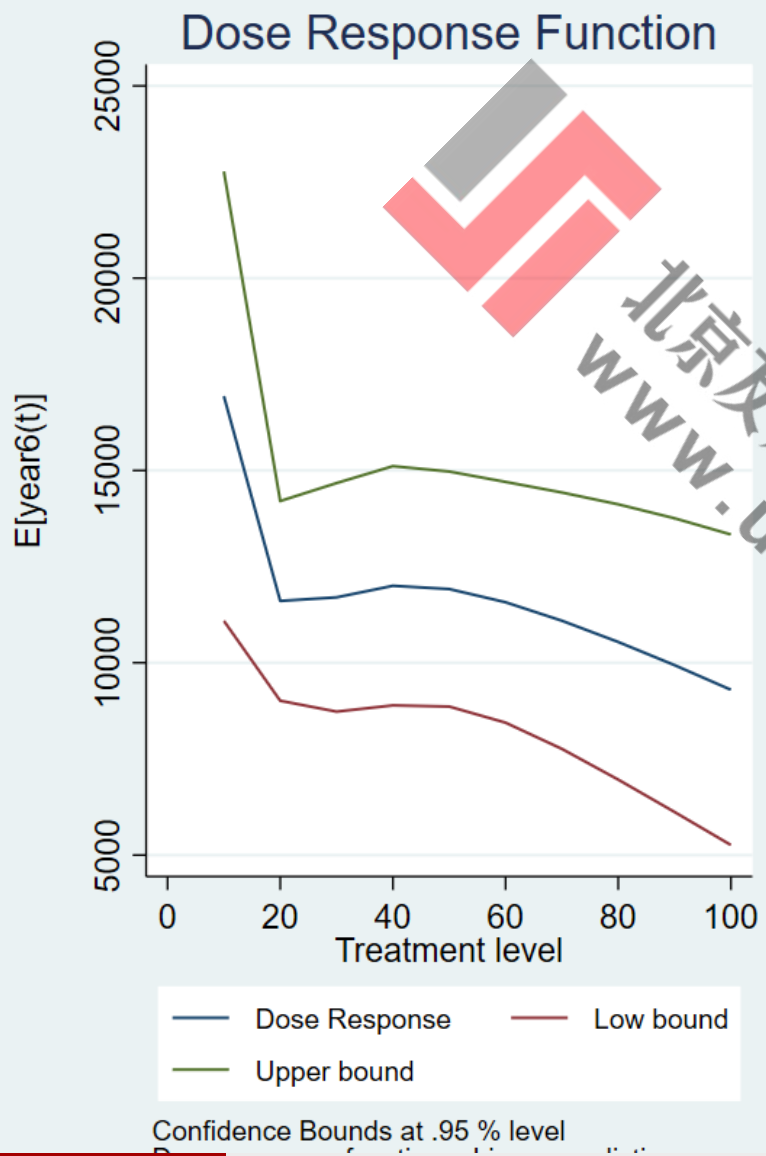
Generalized Propensity Score in Stata

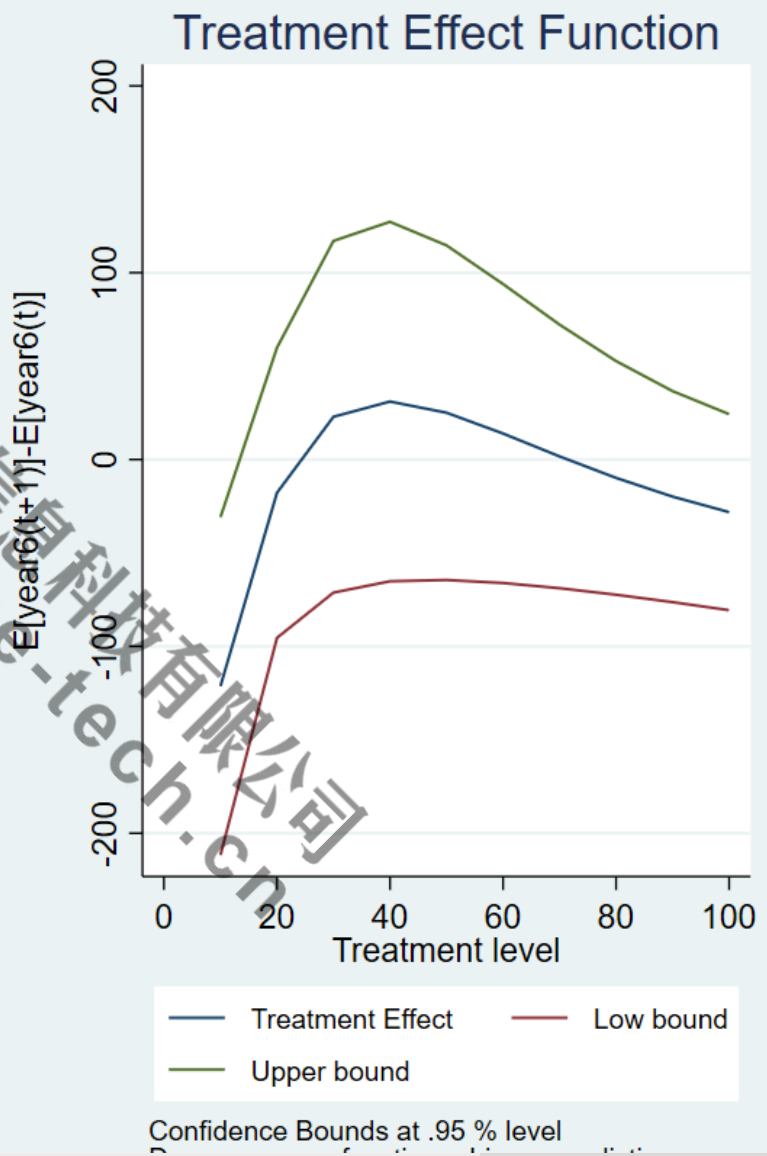
```
The outcome variable ''year6'' is a continuous variable

The regression model is: Y = T + T^2 + GPS +  GPS^2 + T*GPS
```

Source	SS	df	MS		
Model	2.9459e+09	5	589185430	Number of obs	= 202
Residual	3.8379e+10	196	195811039	F(5, 196)	= 3.01
				Prob > F	= 0.0122
				R-squared	= 0.0713
				Adj R-squared	= 0.0476
Total	4.1325e+10	201	205596471	Root MSE	= 13993

year6	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
prize	-225.4371	74.8156	-3.01	0.003	-372.984	-77.89016
prize_sq	.3536809	.1669045	2.12	0.035	.0245218	.6828401
pscore	-103337.3	48370.8	-2.14	0.034	-198731.4	-7943.272
pscore_sq	131949.1	79405.76	1.66	0.098	-24650.25	288548.5
prize_pscore	549.9934	219.7662	2.50	0.013	116.5835	983.4034
_cons	31268.46	6955.423	4.50	0.000	17551.38	44985.54

Generalized Propensity Score in Stata





Generalized Propensity Score in Stata

The nwctcinf command invokes wolframscript.exe. Installation method: wolframengine.exe is part of Mathematica and can be installed directly with Mathematica.
Alternatively, you can install wolframengine.exe separately: https://www.wolfram.com/engine/ (register an account according to the instructions, then download for free).
Specify the paths for wolframscript.exe (not mathematica.exe) in the whereis command. The command is (actual directories may vary depending on your computer and software version):
whereis wolframscript
`"D:\WolframResearch\Mathematica\14.0\wolframscript.exe"`

Empirical application

In this section, we aim to assess the direct effect of national policy interventions for agricultural producers on the country's food security, as well their spillover effects on food security of its commercial partners.

Outcome-Food security. Food security is measured as the level of food availability, that is the supply of food commodities in kilo-calories per person.

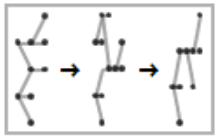
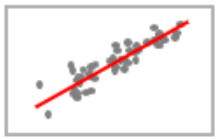
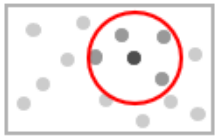
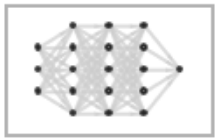
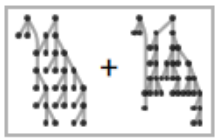
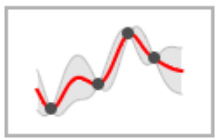
Treatment-Policy intervention. Policy intensity in the agricultural sector is assessed using the Nominal Rate of Assistance (NRA).

Covariates-Country characteristics. These are: real per capita GDP and total population as a proxy of the country demand and size, respectively; per-capita arable land and the agricultural total factor productivity growth index to assess the country's relative agricultural comparative advantage; the ratio of food imports to total exports, net food exports,

Empirical application

```
nwctcinf  foodavalibility nac networknac, ///
ipscovariate(population gdp pcarableland foodimoprttotalexports netexports) ///
npscovariate(population gdp pcarableland foodimoprttotalexports netexports)  ///
ipsmodel(`"Method -> "GradientBoostedTrees""') ///
npsmodel(`"Method -> "RandomForest",PerformanceGoal -> "Quality""') ///
gpsmodel(`"ExponentialFamily -> "Gaussian""') ///
zgrid(10) ggrid(10) ///
regresstypet(`"regressTypet -> "linear""') ///
regresstypegps(`"regressTypegps -> "linear""') ///
interaction(`"interaction -> "1""')  ///
bootstrap(yes) boot_reps(100) ///
zkernel(`"InterpolationPoints -> Automatic"') ///
gkernel(`"MaxMixtureKernels -> Automatic"')
```

Empirical application

	"DecisionTree"	用决策树预测
	"GradientBoostedTrees"	使用梯度提升训练的树群进行预测
	"LinearRegression"	根据特征的线性组合进行预测
	"NearestNeighbors"	根据最近邻接范例预测
	"NeuralNetwork"	利用人工神经网络进行预测
	"RandomForest"	根据 Breiman-Cutler 决策树集合进行预测
	"GaussianProcess"	使用函数的高斯过程先验进行预测

Empirical application

Continous Treatment under Network Inference model:

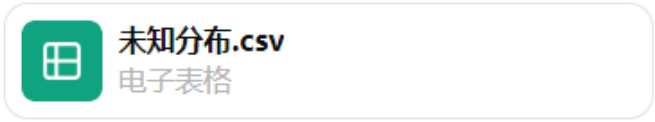
Number of obs.	=	1284	EfronPseudoRSquared	=	0.1878
RSS of IPS Model	=	0.0231	R-squared of IPS Model	=	0.9074
RSS of NPS Model	=	0.0231	R-squared of NPS Model	=	0.4994
AIC	=	-939.4707	BIC	=	-898.2088
Log Likelihood	=	477.7354	LikelihoodRatioStatistic	=	2906103.4635
PearsonChiSquare	=	35.7197			

foodavalib~y	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
intercept	-74.55198	32.33012	-2.31	0.021	-137.9178	-11.18612
z	-1.035487	3.308164	-0.31	0.754	-7.519369	5.448396
g	29.11717	11.4876	2.53	0.011	6.601885	51.63246
phi	7.550476	5.535715	1.36	0.173	-3.299326	18.40028
lambda	23.81052	10.82557	2.20	0.028	2.592794	45.02824
z_phi	.4986167	1.259633	0.40	0.692	-1.970218	2.967451
g_lambda	-11.10295	4.375111	-2.54	0.011	-19.67801	-2.527893
z_g	-.0726029	.0412631	-1.76	0.078	-.1534771	.0082712

Empirical application

z	g	doseresponse	doseresponsell	doseresponsemedian	doseresponseul	zweight	gweight
0.896583	1.020431	7.941159047	7.829405701	7.981085643	8.093293627	1.026364	1.628664
0.896583	1.06635	7.942171939	7.834465149	7.980594039	8.090059617	1.026364	2.651159
0.896583	1.098432	7.941872969	7.837269008	7.98285132	8.088028682	1.026364	3.008529
0.896583	1.131559	7.940767301	7.839576251	7.982239096	8.086167614	1.026364	2.962821
0.896583	1.164145	7.938952653	7.841300433	7.98018143	8.087275838	1.026364	2.80084
0.896583	1.202399	7.935986864	7.842685018	7.98114915	8.088103423	1.026364	2.537385
0.896583	1.243955	7.931854779	7.843473364	7.982252995	8.088479104	1.026364	2.064634
0.896583	1.307556	7.923983847	7.84340913	7.982135731	8.088143215	1.026364	1.345861
0.896583	1.41523	7.907576721	7.840503329	7.979110877	8.085654785	1.026364	0.551751
0.967449	1.020431	7.942864868	7.850237323	7.977411571	8.071623477	1.882149	1.628664
0.967449	1.06635	7.943727694	7.855355228	7.977416522	8.068742735	1.882149	2.651159
0.967449	1.098432	7.943323881	7.858199928	7.978504285	8.066729198	1.882149	3.008529
0.967449	1.131559	7.942109952	7.860549342	7.978108966	8.064999455	1.882149	2.962821
0.967449	1.164145	7.940188814	7.862315007	7.977715632	8.066180777	1.882149	2.80084
0.967449	1.202399	7.937098009	7.863748291	7.977500333	8.067094176	1.882149	2.537385
0.967449	1.243955	7.932830119	7.864589539	7.97808961	8.067563077	1.882149	2.064634
0.967449	1.307556	7.924751336	7.864606271	7.97805693	8.067369862	1.882149	1.345861
0.967449	1.41523	7.907992328	7.860883065	7.975453272	8.065122974	1.882149	0.551751
1.005126	1.020431	7.944921852	7.859294092	7.974611143	8.061476949	2.125079	1.628664
1.005126	1.06635	7.945704892	7.864443077	7.975083576	8.059441189	2.125079	2.651159
1.005126	1.098432	7.945245336	7.867309491	7.975990772	8.057436902	2.125079	3.008529

Empirical application



这是一个1000个样本的数据，请你学习其规律，得到概率密度函数，并告诉我取0的概率，即f[0]的近似值

To find the probability density function (PDF) and calculate the probability density at $x = 0$, I will first examine the data to understand its distribution and then estimate the PDF using kernel density estimation or a similar method.

Let's start by loading and visualizing the data.

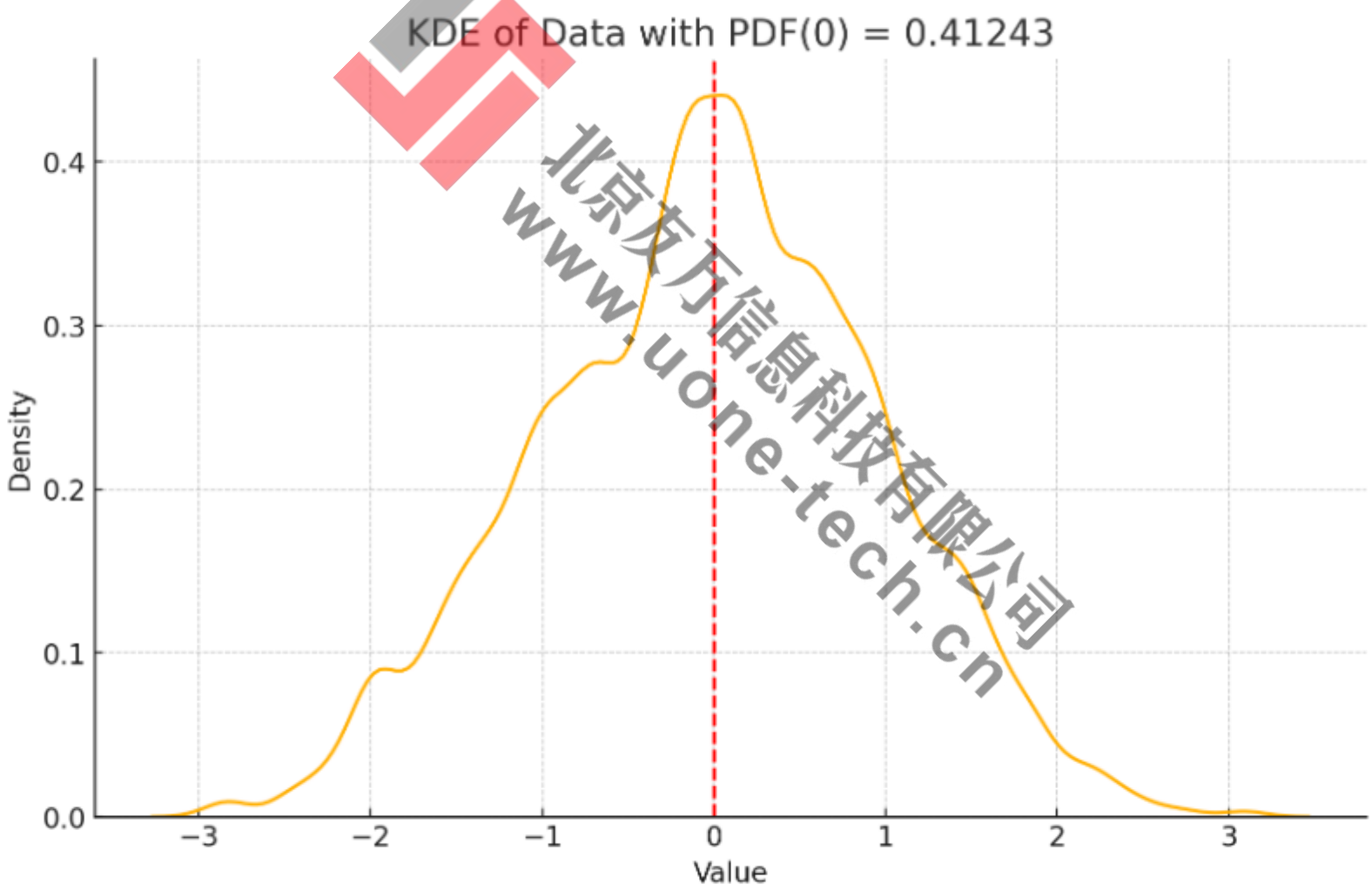
It appears that the first row might contain a header. I'll remove it and then proceed with the analysis.

The data consists of 1,000 samples, and it's numerical with the following summary statistics:

- Mean: approximately -0.024
- Standard Deviation: approximately 0.988
- Minimum value: -2.898
- Maximum value: 3.092

Empirical application

Next, I'll estimate the probability density function (PDF) using kernel density estimation and calculate the value of the PDF at $x = 0$.



Any questions and/or comments welcome.

Thank you!