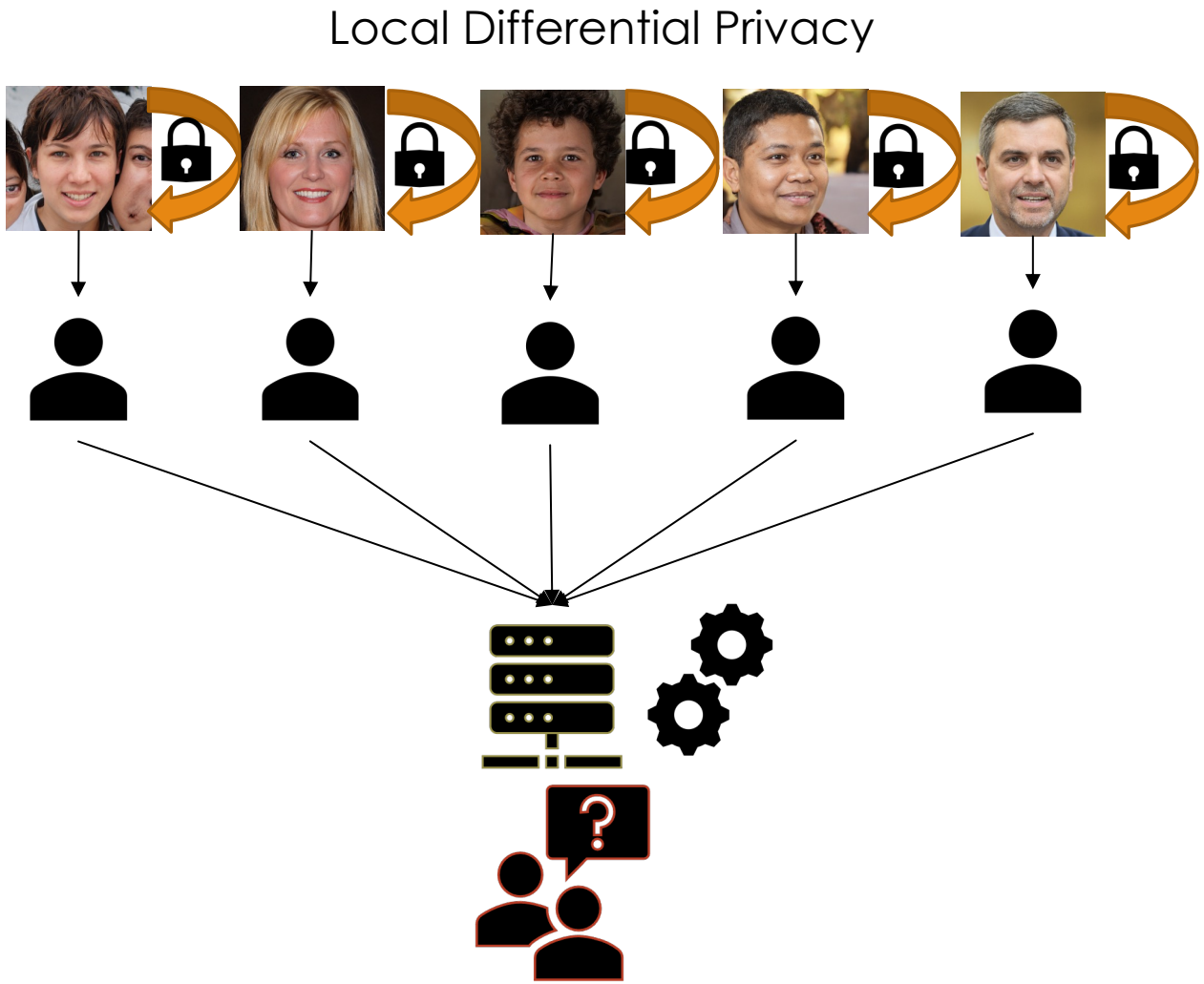
New LDP approach using VAE

Andres Hernandez-Matamoros* and Hiroaki Kikuchi

明治大学
MEIJI UNIVERSITY

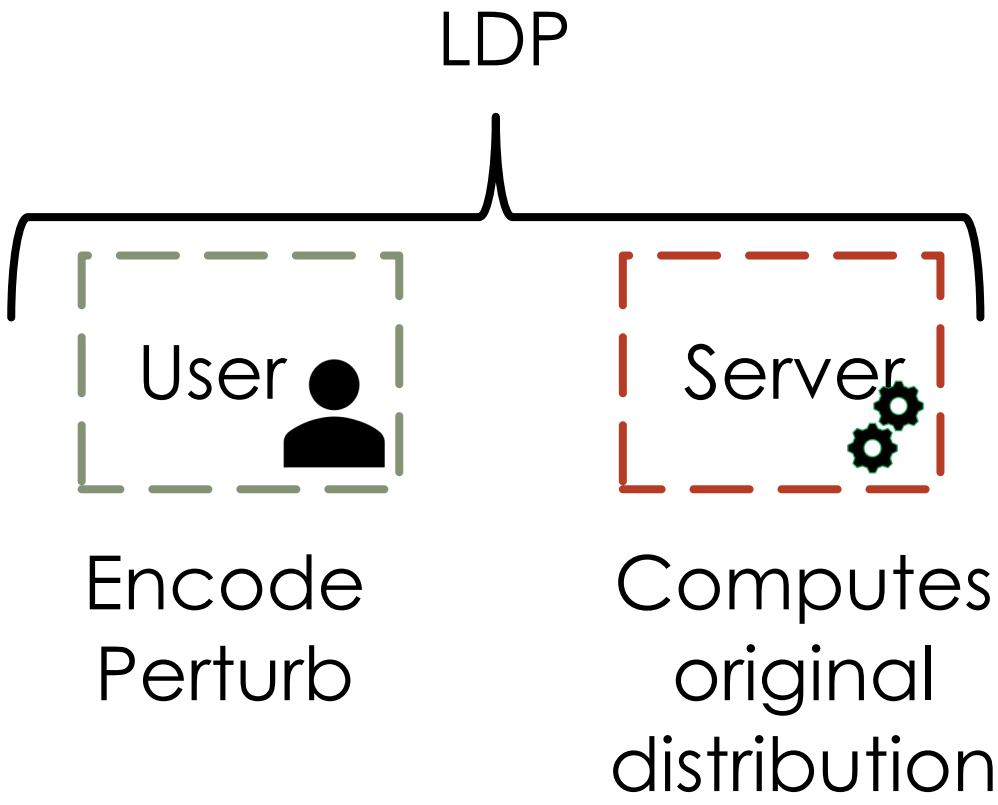
NSS-SocialSec 2023
2023/08/16
*matamoros@meiji.ac.jp

What is LDP?



* Face images were taken from https://thispersondoesnotexist.com/

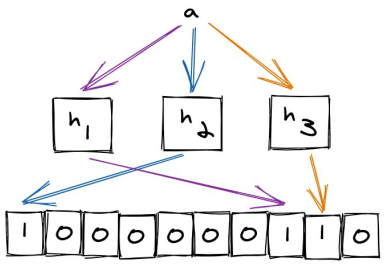
LoPub[1]



[1]Ren, Xuebin and Yu, Chia-Mu and Yu, Weiren and Yang, Shusen and Yang, Xinyu and McCann, Julie A. and Yu, Philip S., IEEE Transactions on Information Forensics and Security, LoPub: High-Dimensional Crowdsourced Data Publication With Local Differential Privacy, 2018, doi=10.1109/TIFS.2018.2812146.

LoPub-Users

4

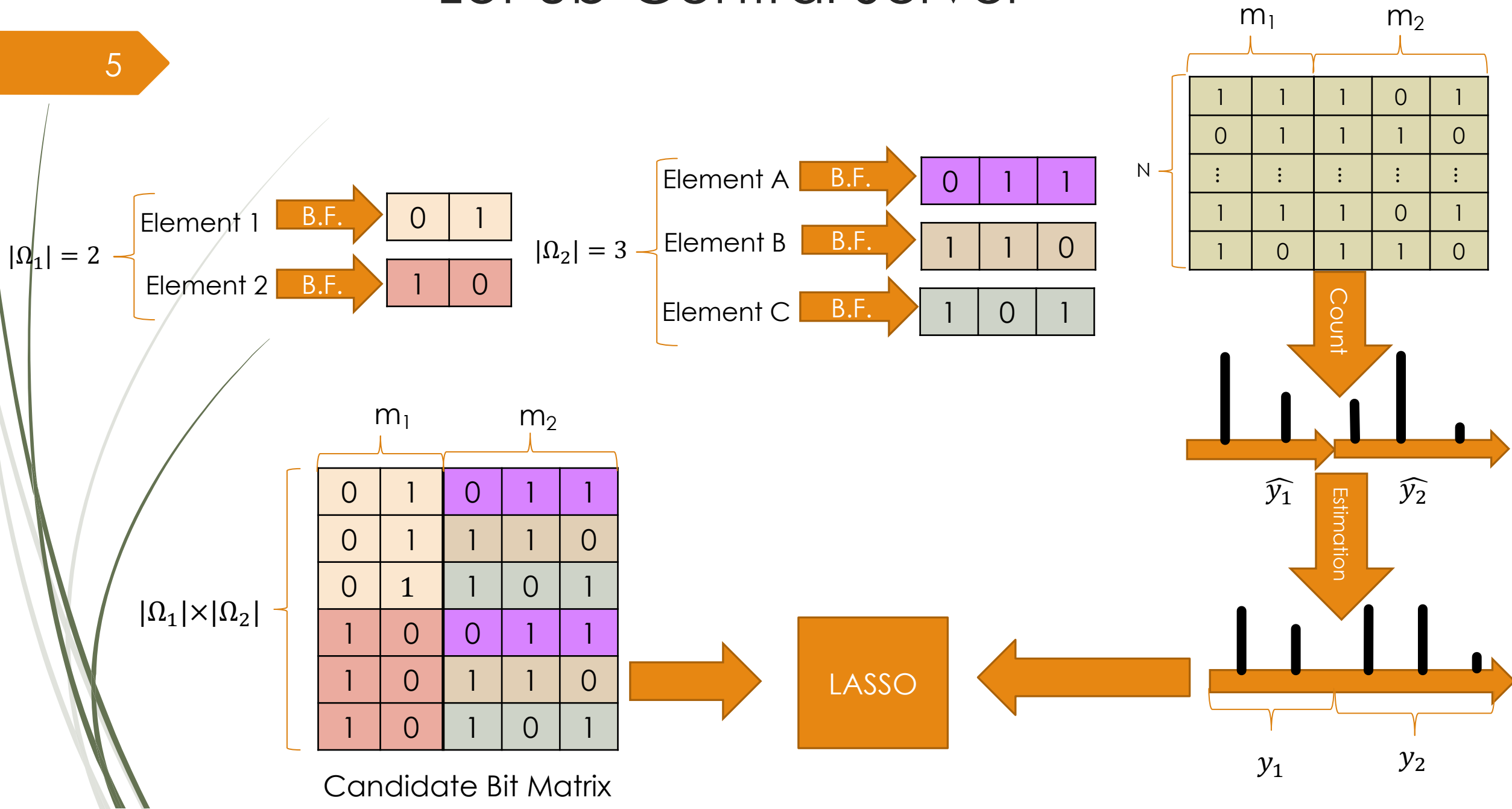


Example of Bloom filters

$$\hat{s}_j^i = \begin{cases} s_j^n \text{ with probability of } 1-f, \\ 1 \text{ with probability of } f/2, \\ 0 \text{ with probability of } f/2 \end{cases}$$

Randomize Response

LoPub-Central Server

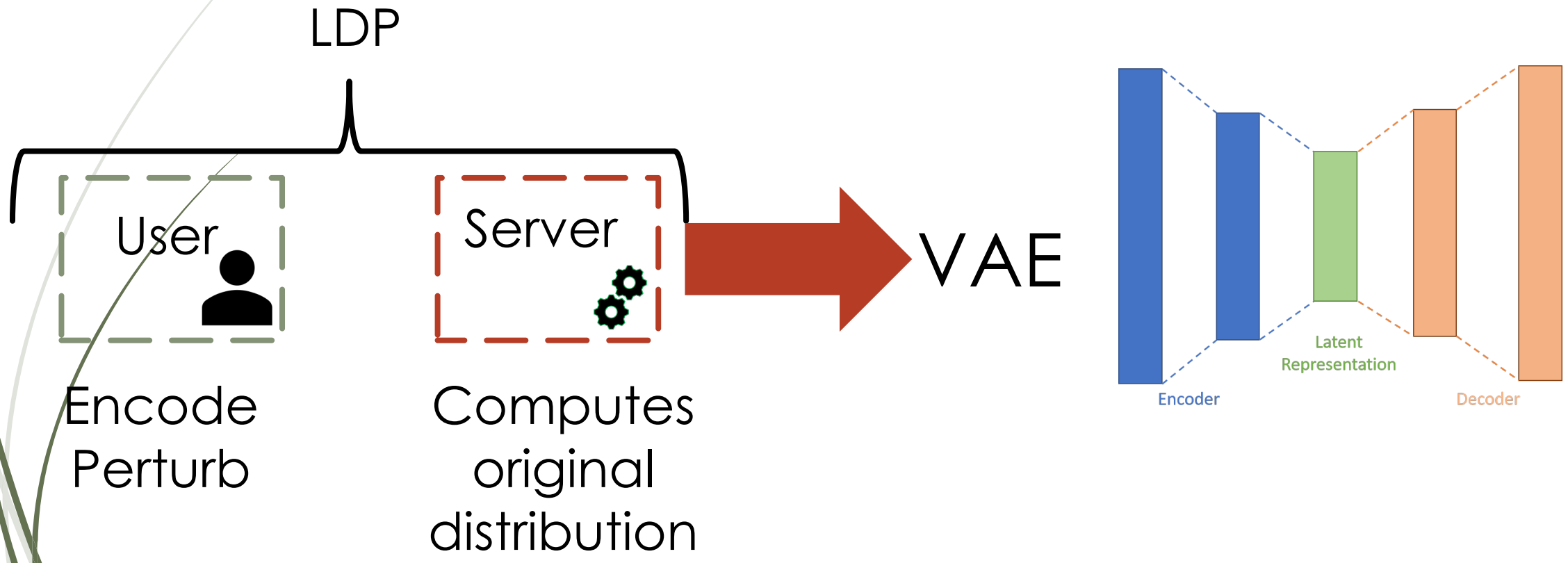


Pro vs Cons

One/two-dimensional probability distributions can be efficiently estimated through the Lasso regression-based algorithm.

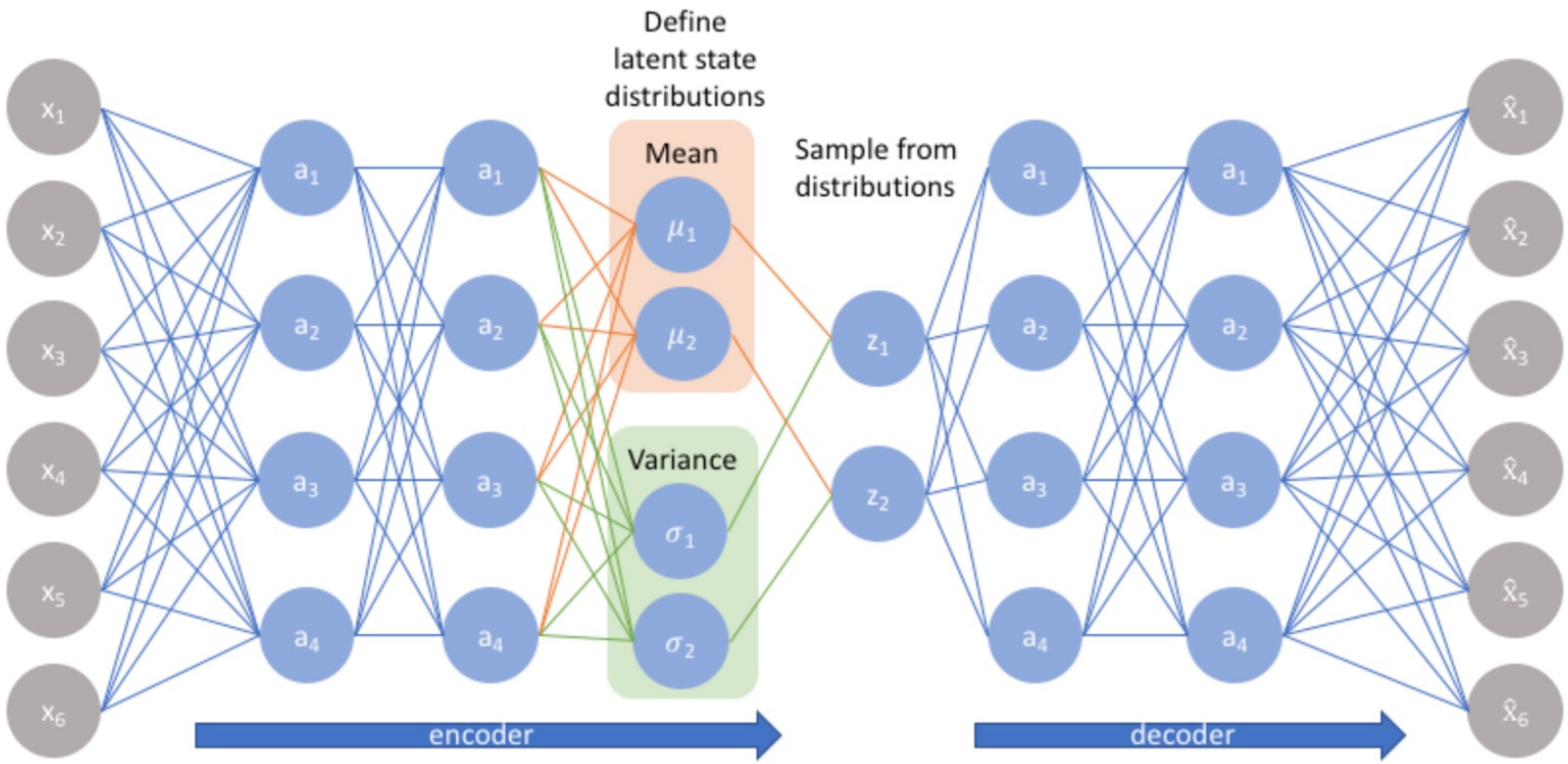
The k-dimensional distribution estimations in LoPub still suffer from the low data utility when k is large.

Proposal

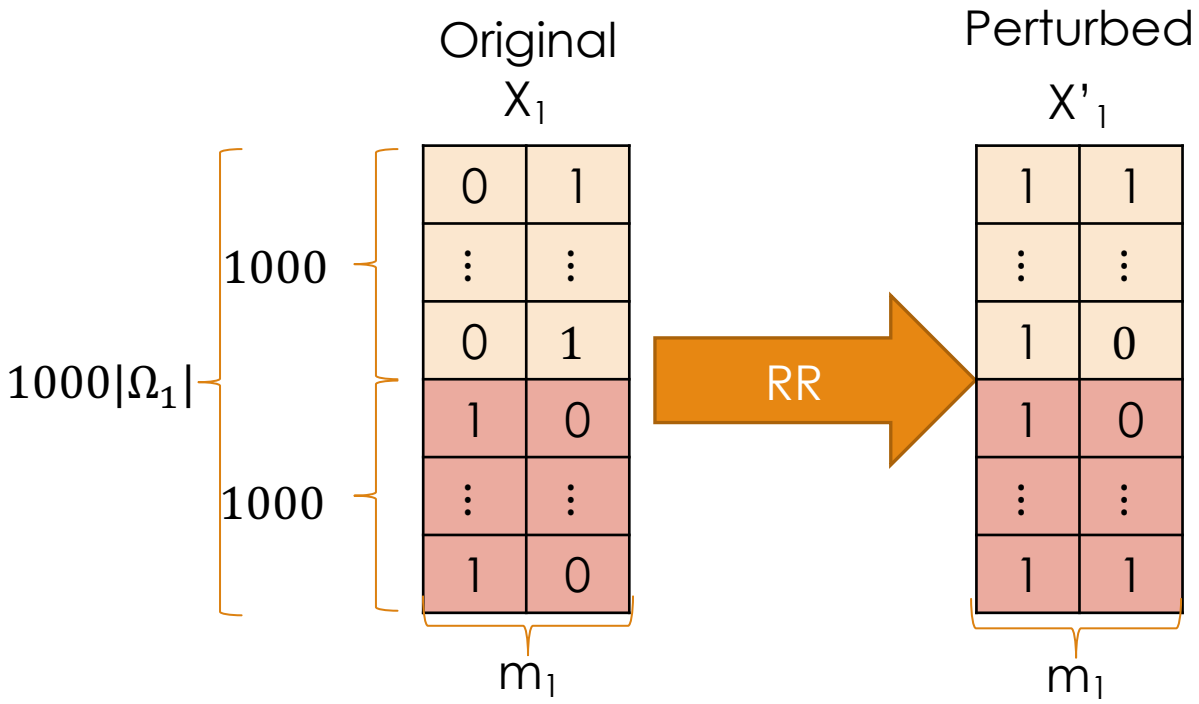

LDP
User
Server
VAE
Encode
Perturb
Computes
original
distribution
Encoder
Latent
Representation
Decoder

What is VAE?

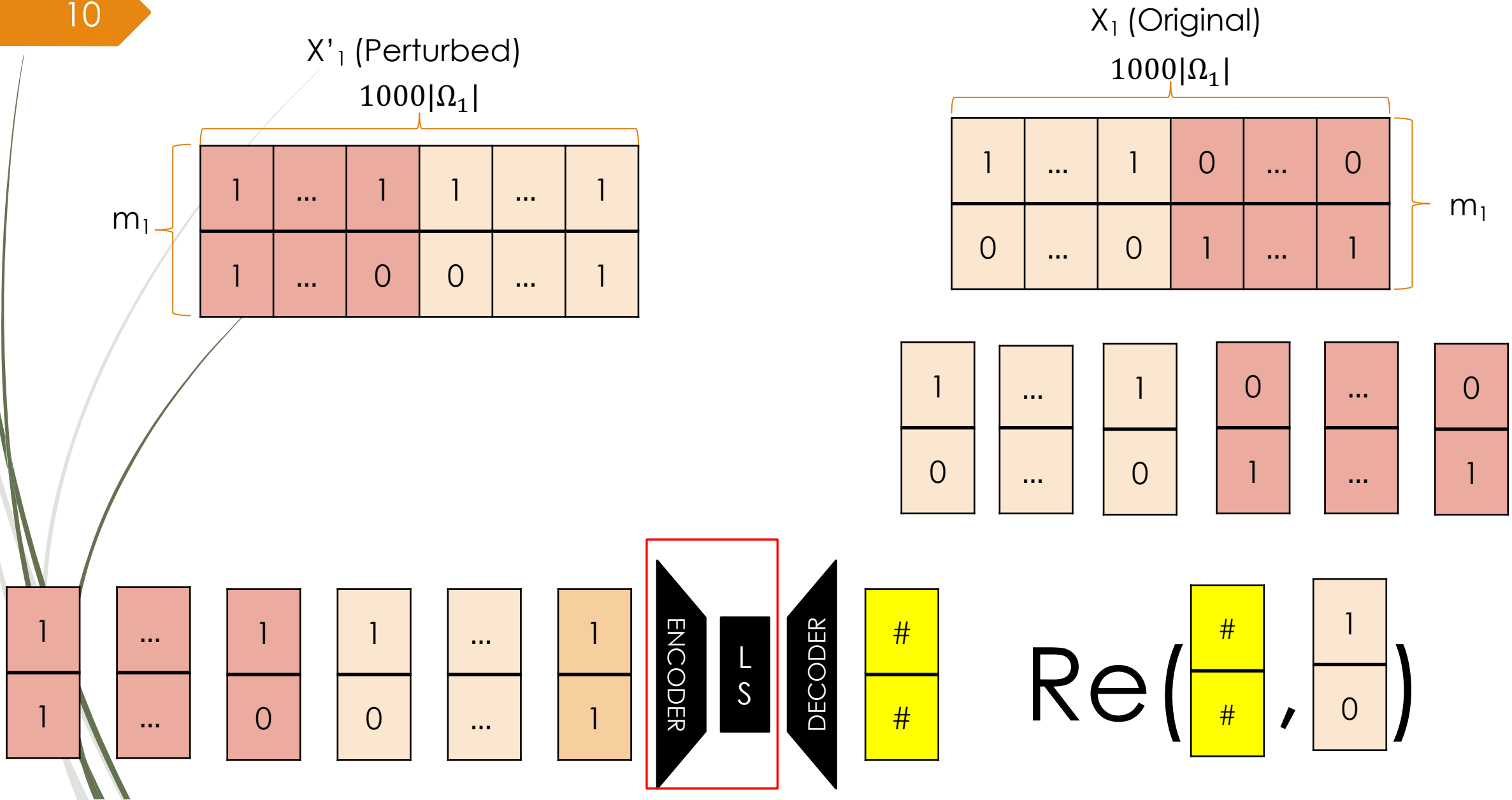
A variational Autoencoder (VAE) provides a probabilistic manner for describing an observation in latent space.



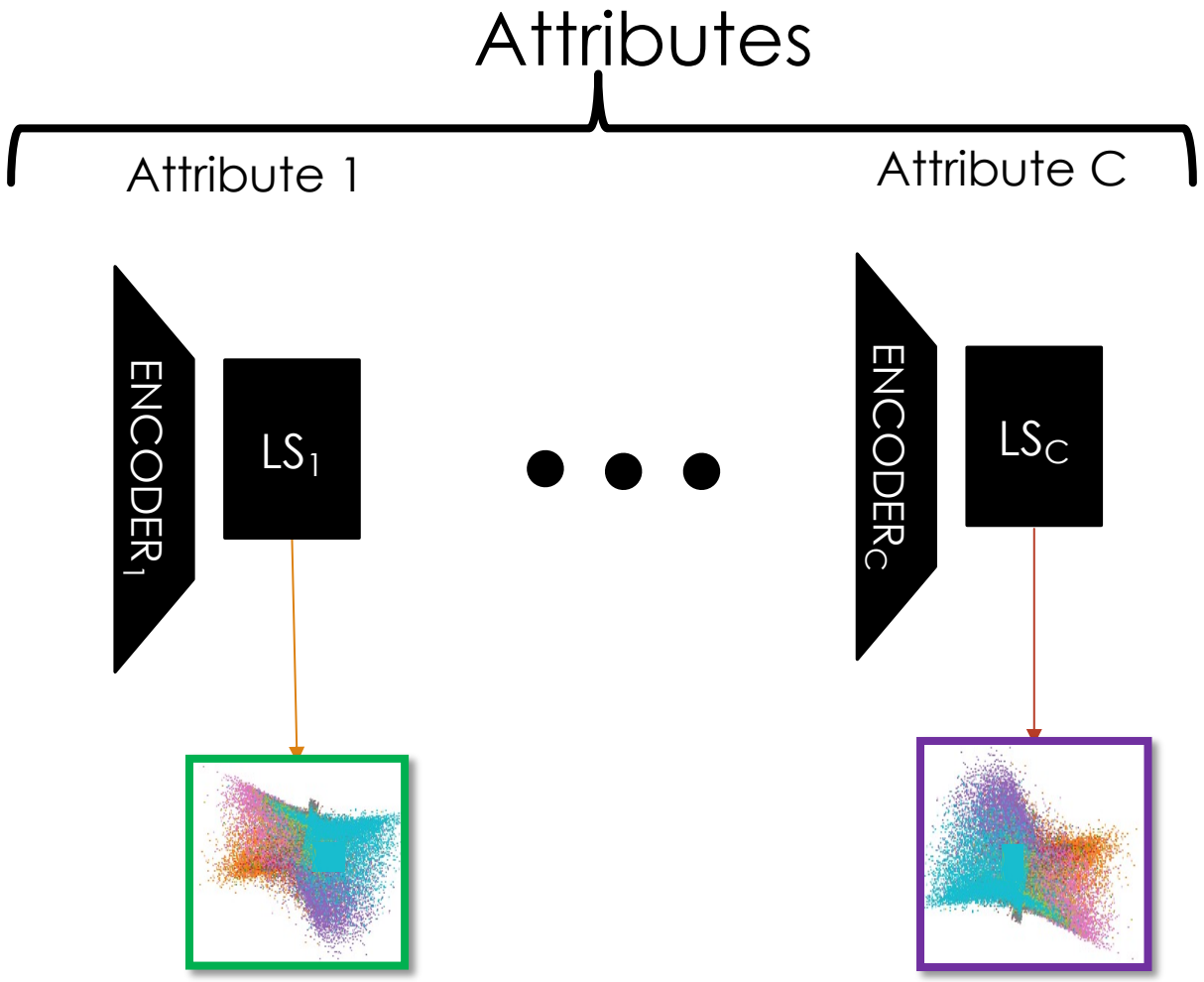
VAE Training-Dataset



VAE Training

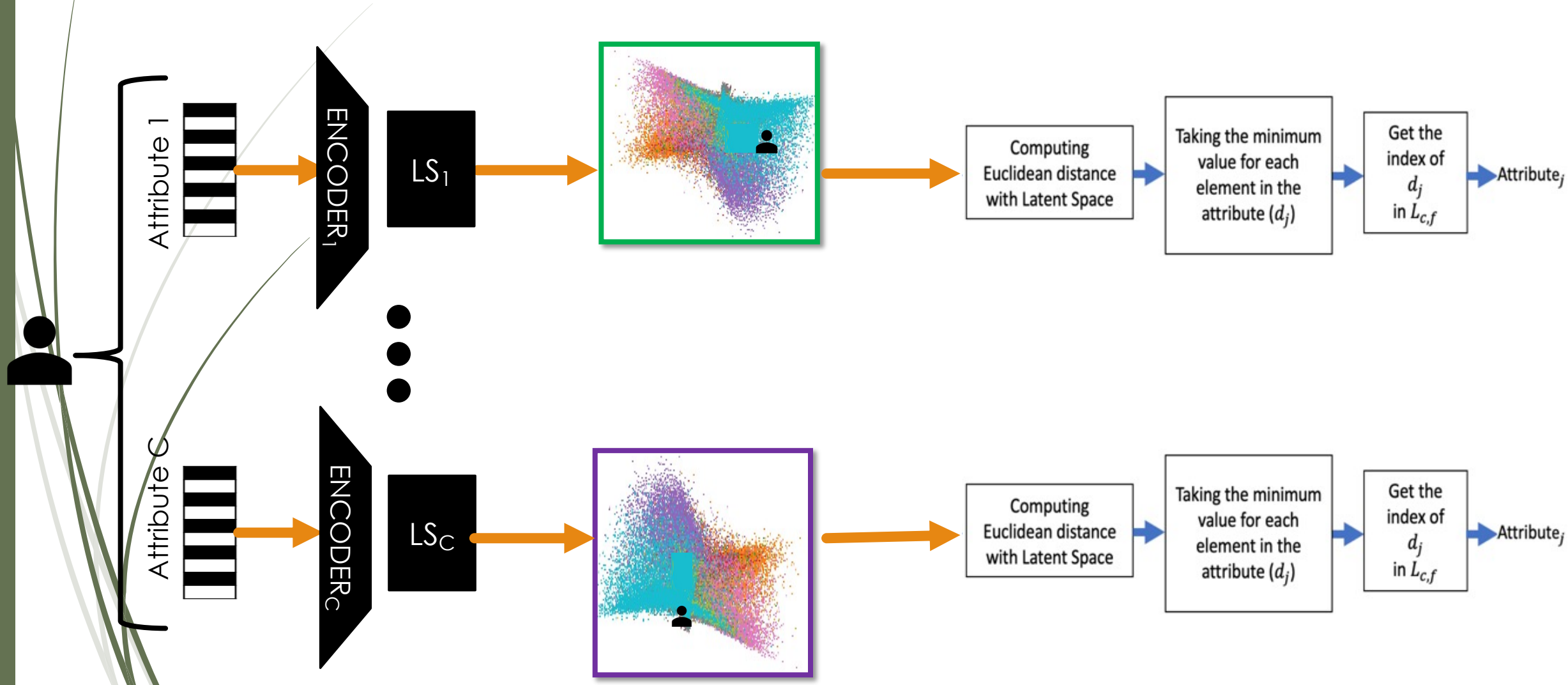


Latent Space's Attributes

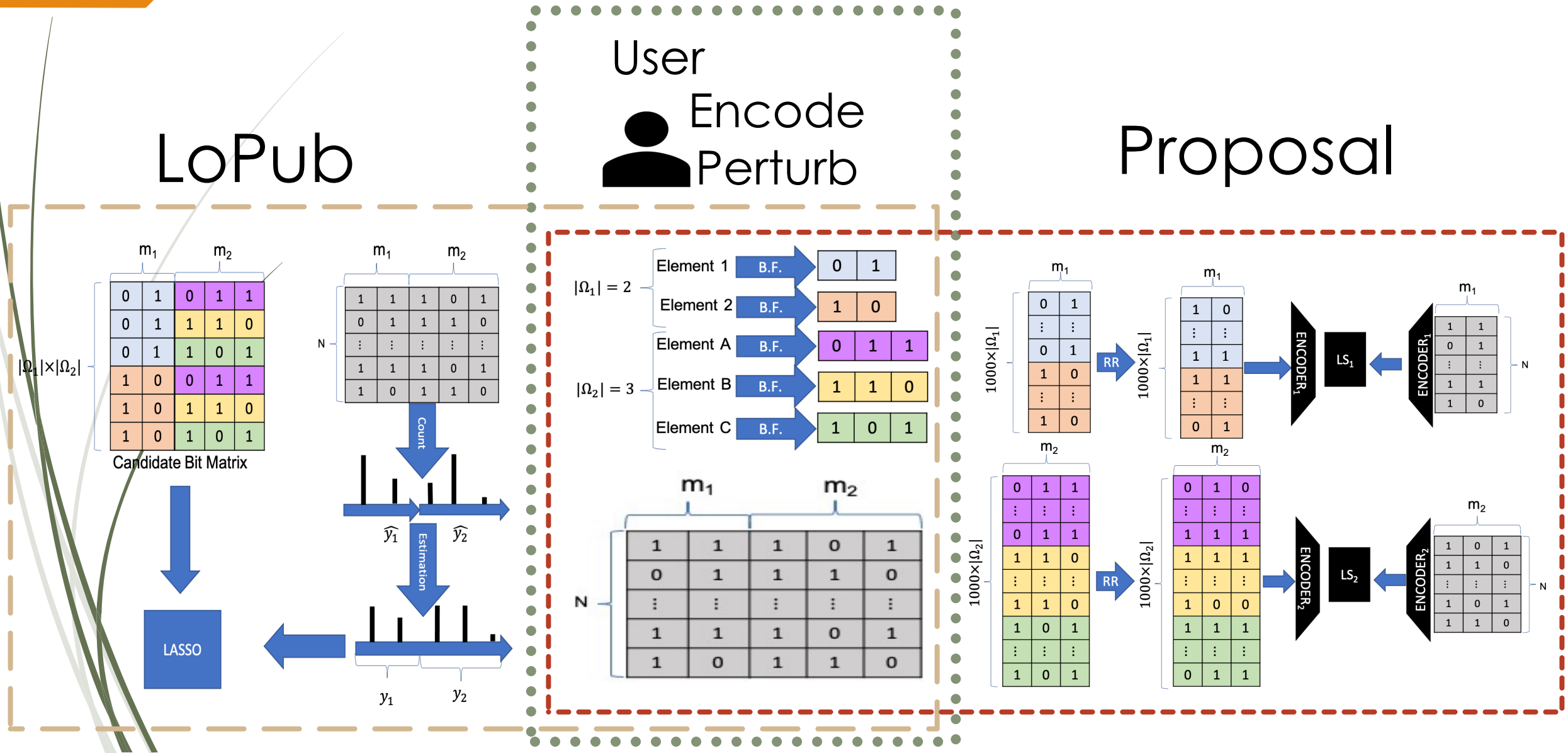


Latent Space Examples

Latent Space's Evaluation



LoPub vs Proposal



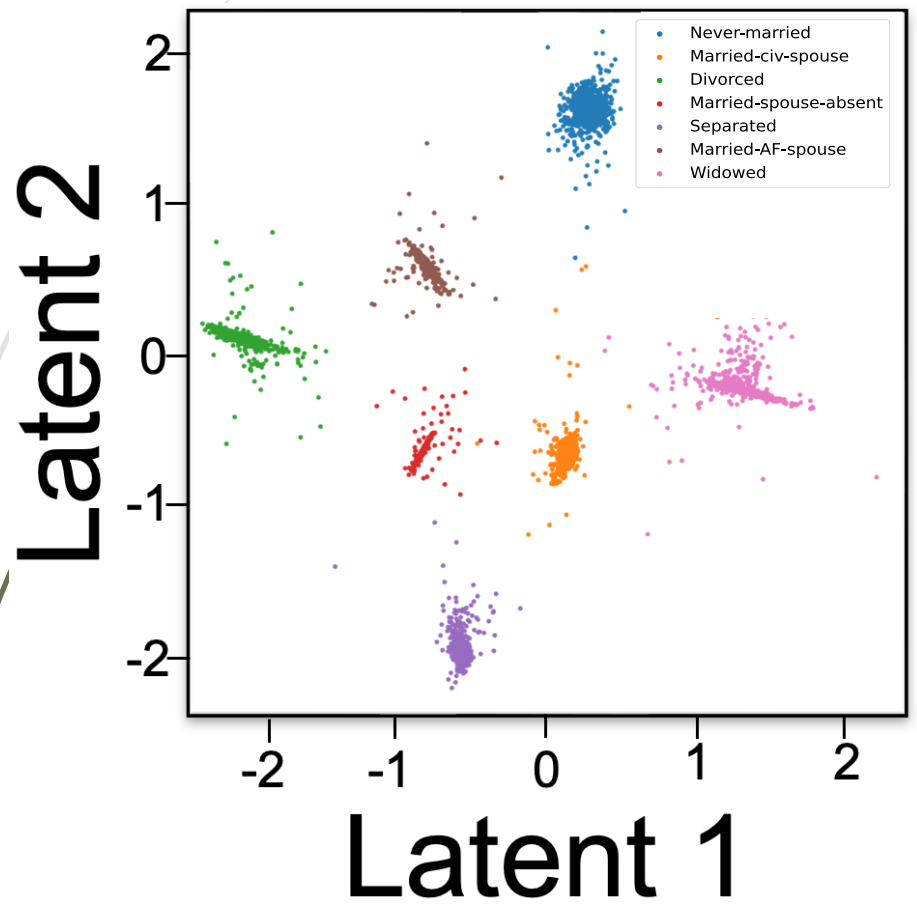
Datasets

Dataset	Users	Attributes	Cardinality		m	
			min	max	min	max
Adult	45223	8	2	16	8	64
Bank	45212	10	2	12	8	47
Nursery	12960	9	2	5	8	20
NHANES	4190	5	2	6	8	23

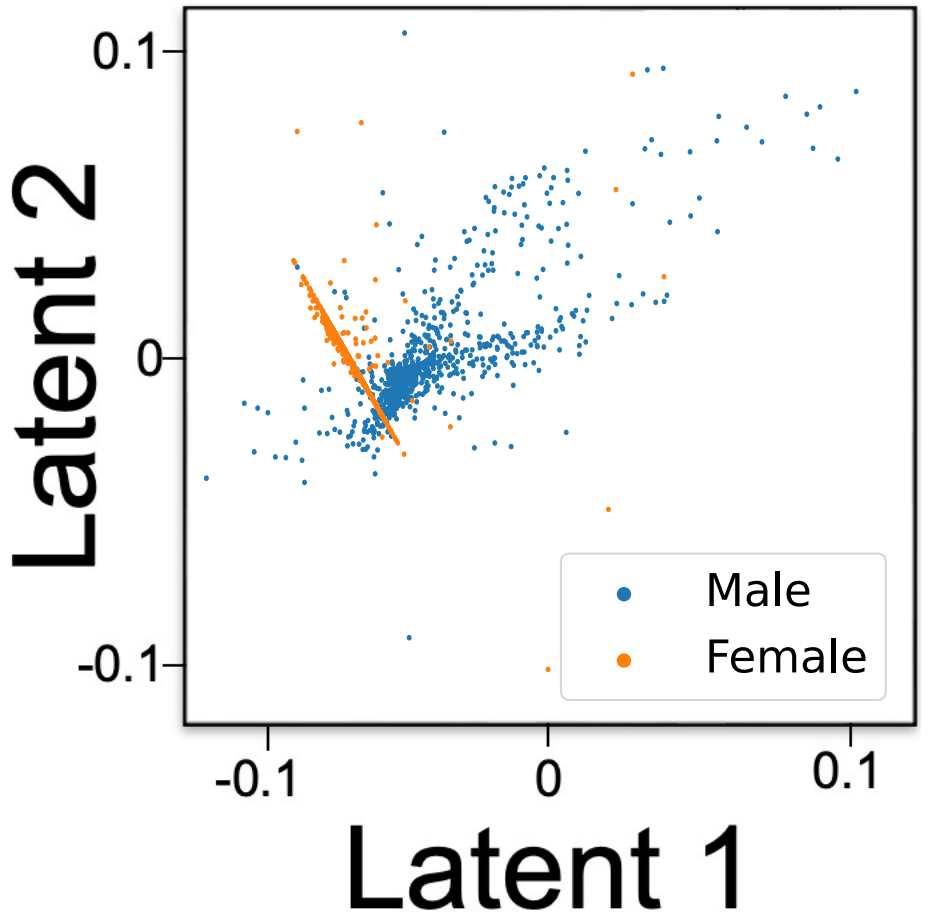
$$h = 5, p = 0.022, m_j = \frac{\ln(\frac{1}{p})}{(\ln 2)^2}|\Omega_j|$$

Latent Space

Examples of Latent Spaces in 2D with f=0.1



Marital Status, Adults Dataset



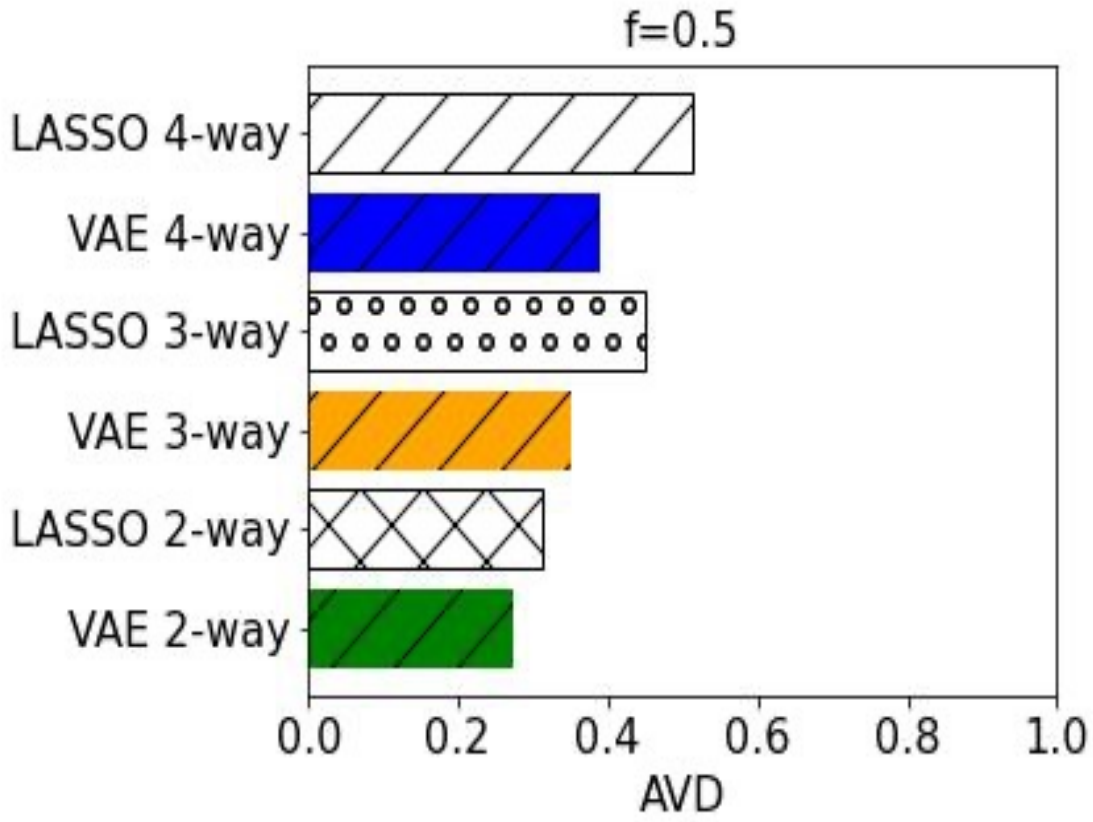
Gender, Adults Dataset

K-way evaluation

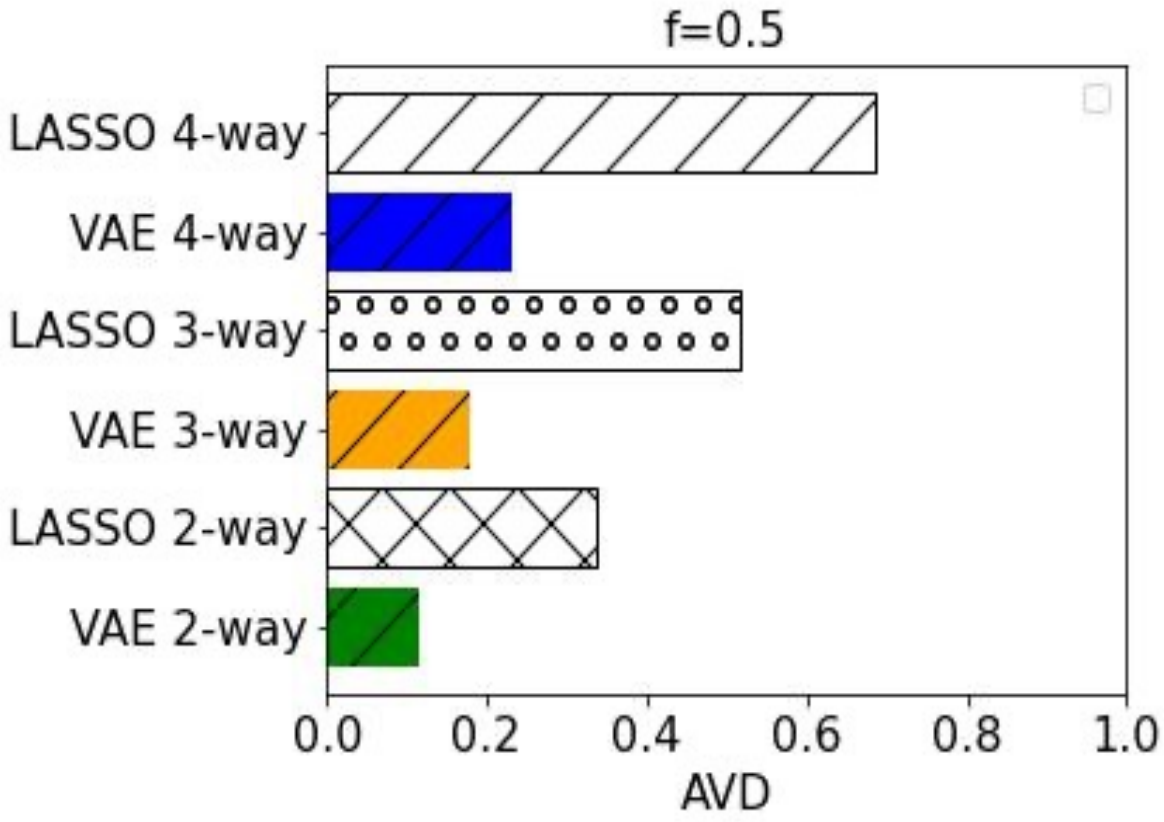
To measure accuracy, we used the distance metric AVD (average variant distance), as suggested in LoPub, to quantify the closeness between the probability distributions *P*(ω) and Q(ω).

$$\mathrm{AVD(P,Q)} = \frac{1}{2}\sum_{\omega\in\Omega} |P(\omega) - Q(\omega)|$$

Accuracy K-way

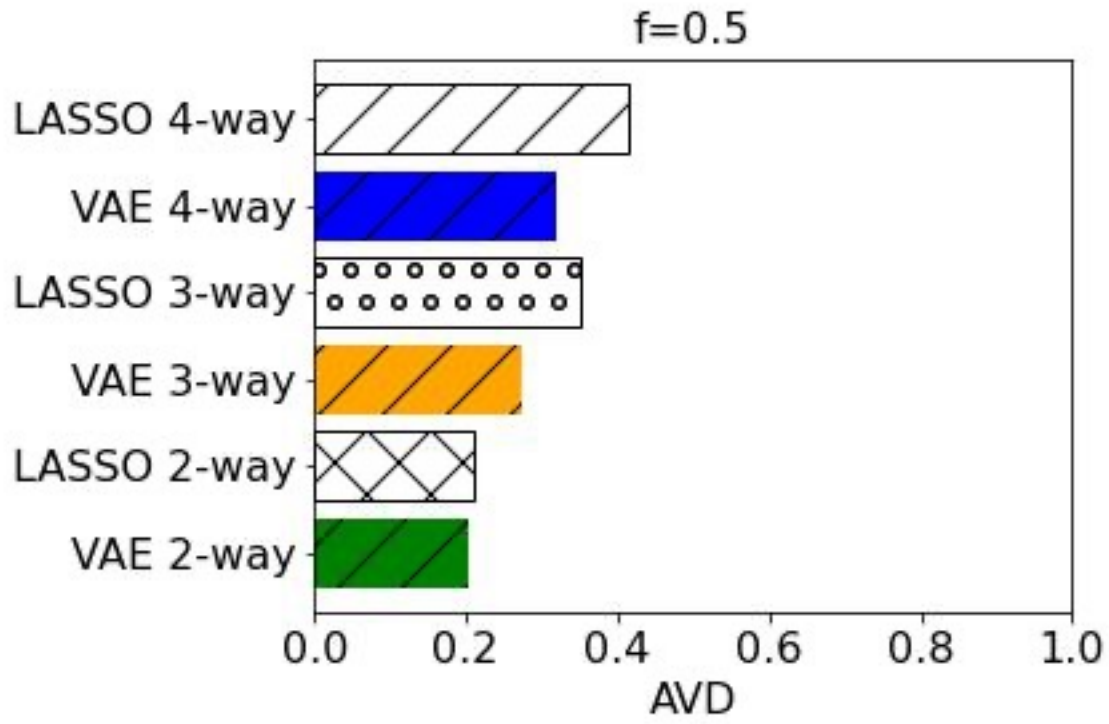


Adult Dataset

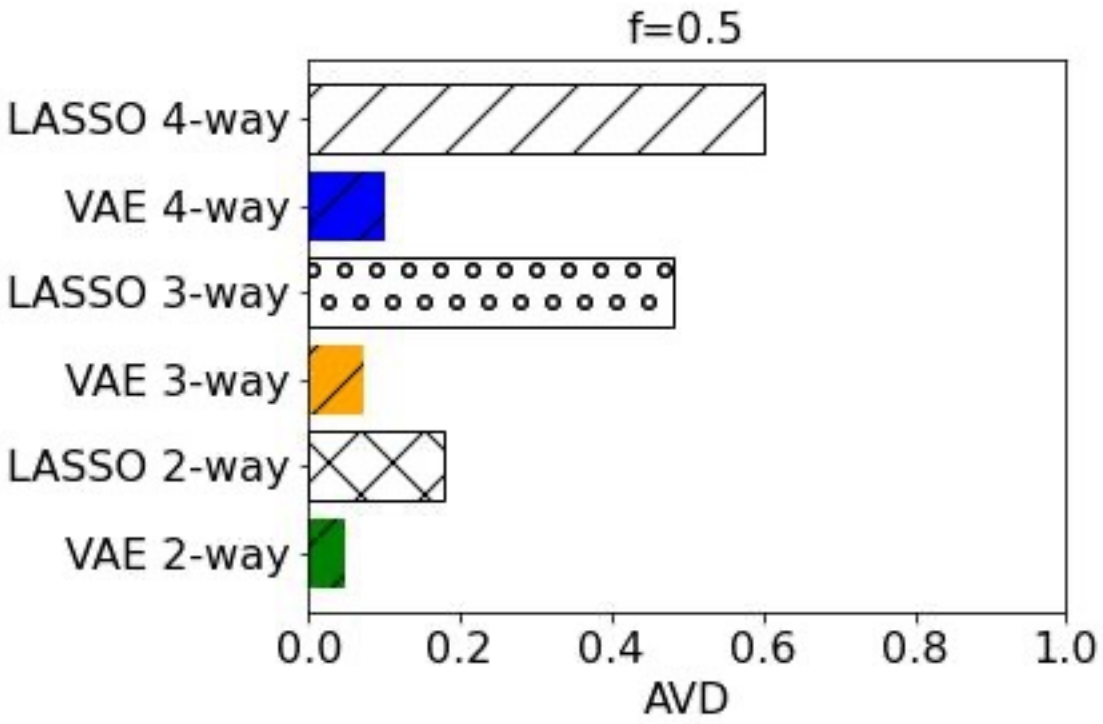


NHANES Dataset

Accuracy K-way

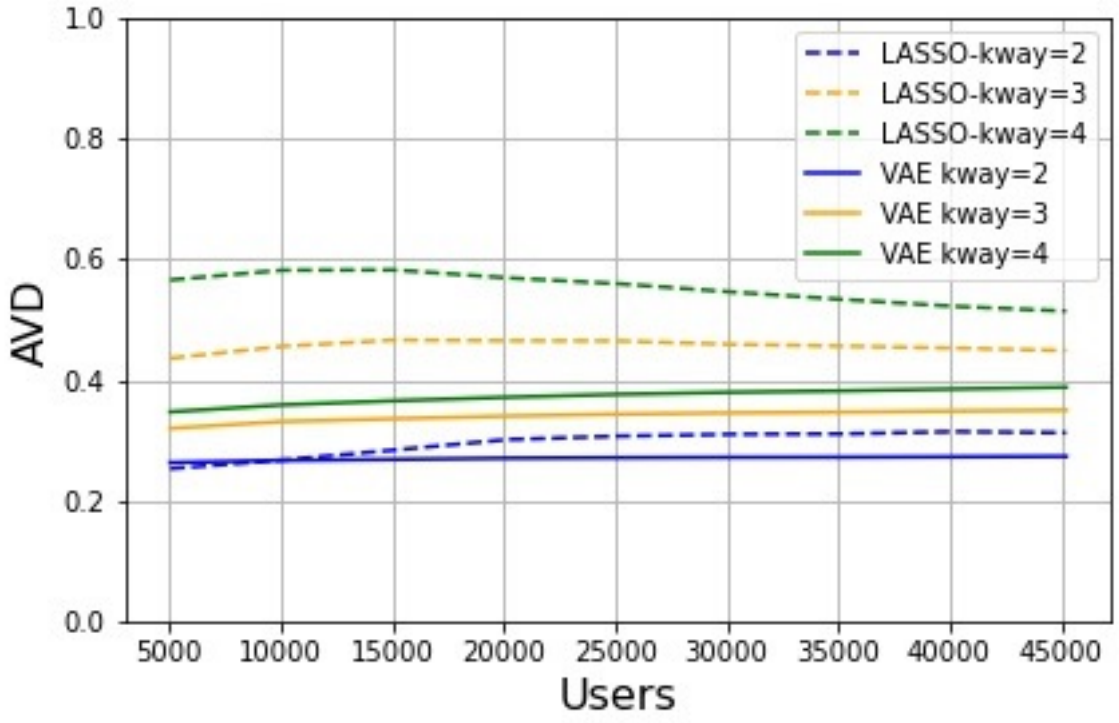


Bank Dataset

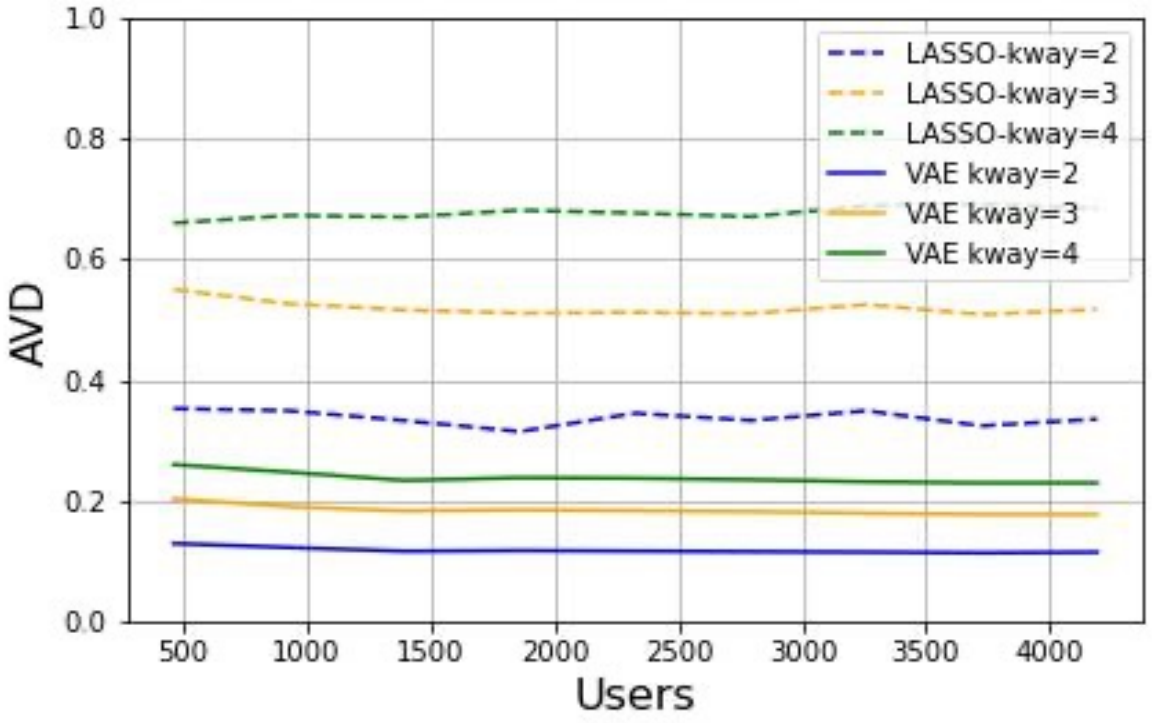


Nursery Dataset

AVD vs N users with f = 0.5

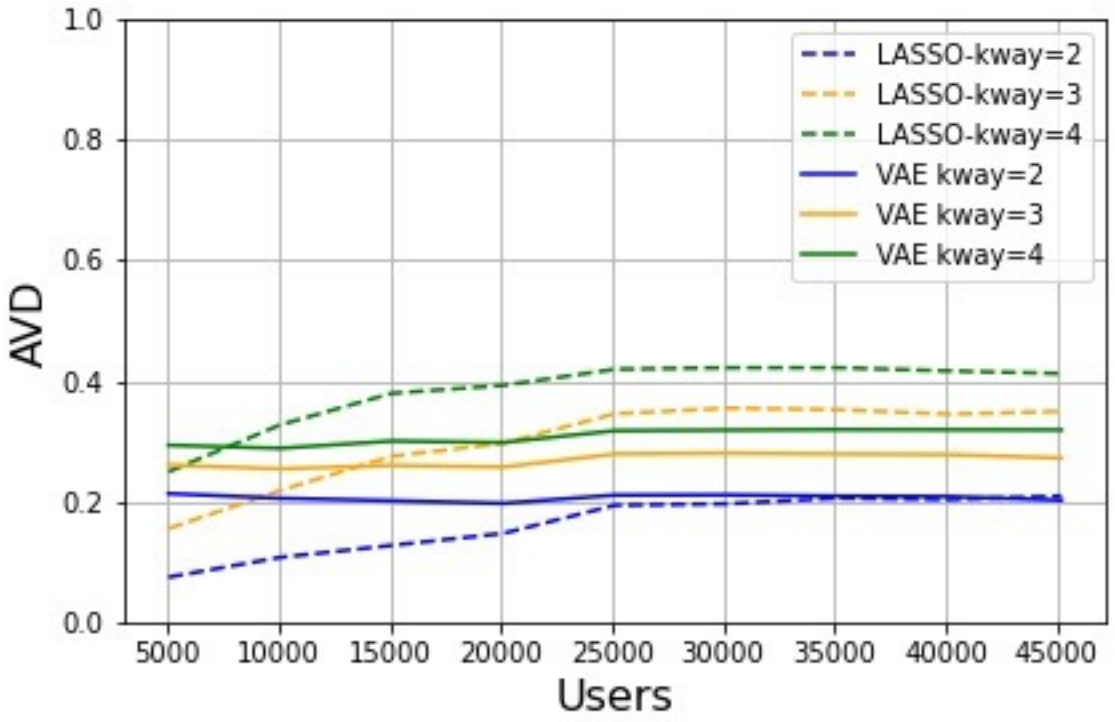


Adult Dataset

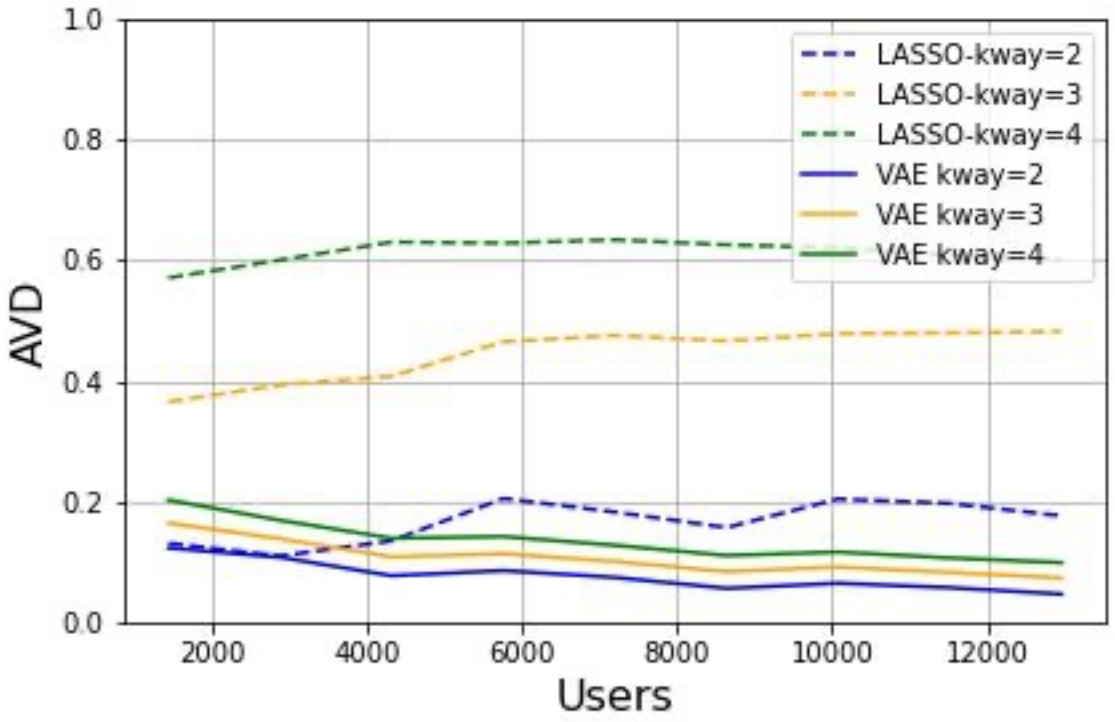


NHANES Dataset

Accuracy vs N users with f = 0.5



Bank Dataset



Nursery Dataset

Conclusions

- This work proposes utilizing the latent space of a VAE in the central server of the LDP scheme.
- The effectiveness is assessed on real datasets encompassing various user counts and attribute cardinalities, all using a single VAE model.
- The findings demonstrate VAE's superiority over LASSO regression by enabling each attribute to possess an independent latent space, mitigating cross-attribute noise interference.
- Future research directions involve exploring the interplay between attribute cardinality and the corresponding latent space.

New LDP approach using VAE

Andres Hernandez-Matamoros* and Hiroaki Kikuchi

Thank You for Your Attention

NSS-SocialSec 2023
2023/08/16
*matamoros@meiji.ac.jp