Deep Generative Quantile Bayes

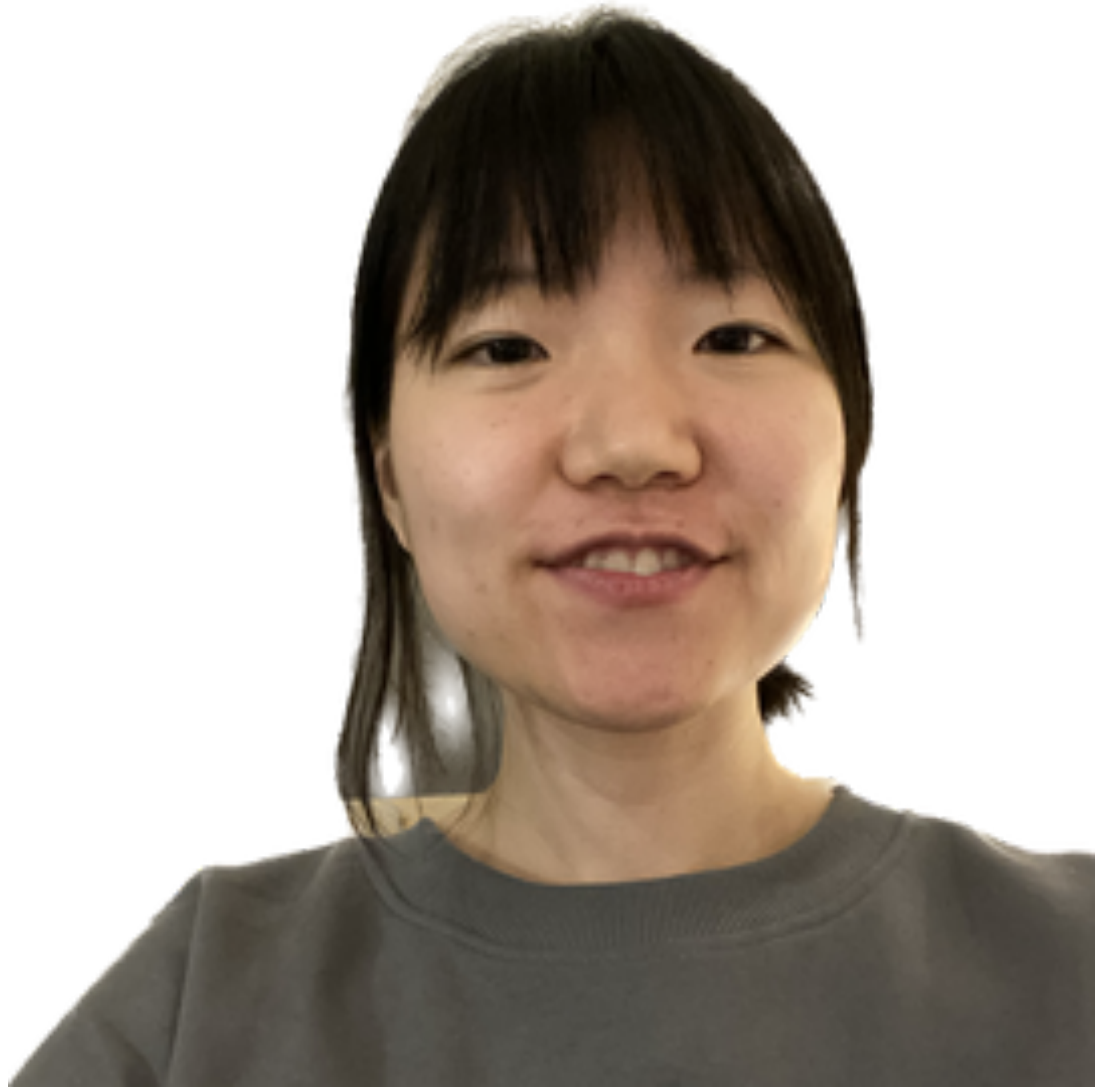
Jungeum Kim, Percy Zhai, and Veronika Rockova

https://jungeumkim.com/

https://arxiv.org/abs/2410.08378

Authors

Jungeum Kim (Me)



https://jungeumkim.com/

Percy Zhai

Veronika Rockova

Short Introduction

Multi-dimensional Posterior Sampling with Likelihood-Free (simulation based) Models

Computational Physics, Mechanical Engineering, Bio Engineering,…

Challenge in Posterior Sampling

Through MCMC: Need likelihood

Through ABC (Approximate Bayesian Computation): Sampling Efficiency (low acceptance rate)

Through Deep Generative Models

Variational Approaches:

Maceda el al. (2024), ...

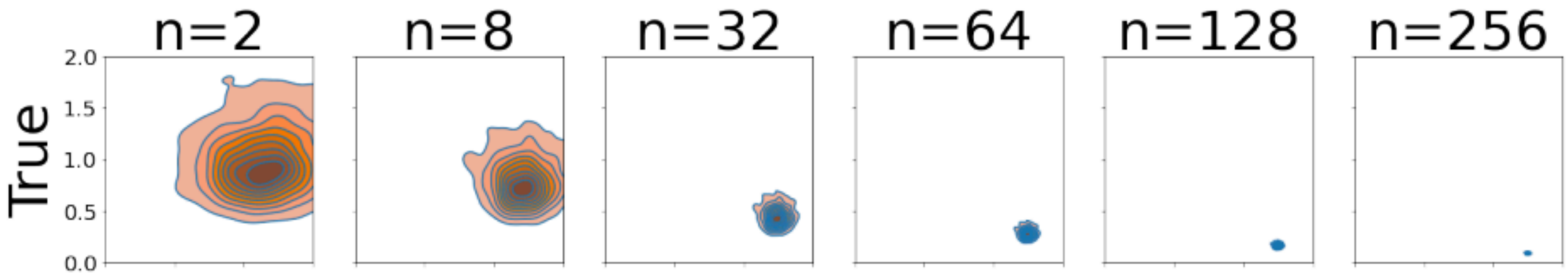
Direct Approaches:

Wang and Rockova (2022): W-1 GAN framework with adversarial training

Polson and Sokolov (2023): Quantile modeling for one dimensional setting

Short Introduction

Support Shrinkage (Posterior Contraction)



iid Model: $X \mid \mu, \sigma \sim N(\mu, \sigma^2)$

$$\mu \mid \sigma^2 \sim N(\mu_0, \sigma^2/\kappa)$$

$$\nu_0\sigma_0^2/\sigma^2 \sim \chi^2(\nu_0)$$

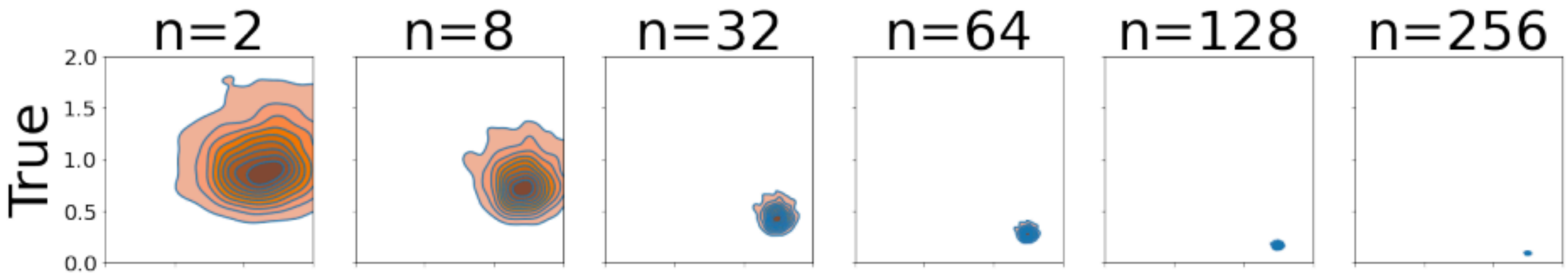
$$\mu_0 = 0,\ \sigma_0 = 1,\ \kappa = 2,\ \nu_0 = 25$$

n: sample size

$$data = (X_1, \ldots, X_n)$$

Short Introduction

Support Shrinkage (Posterior Contraction)

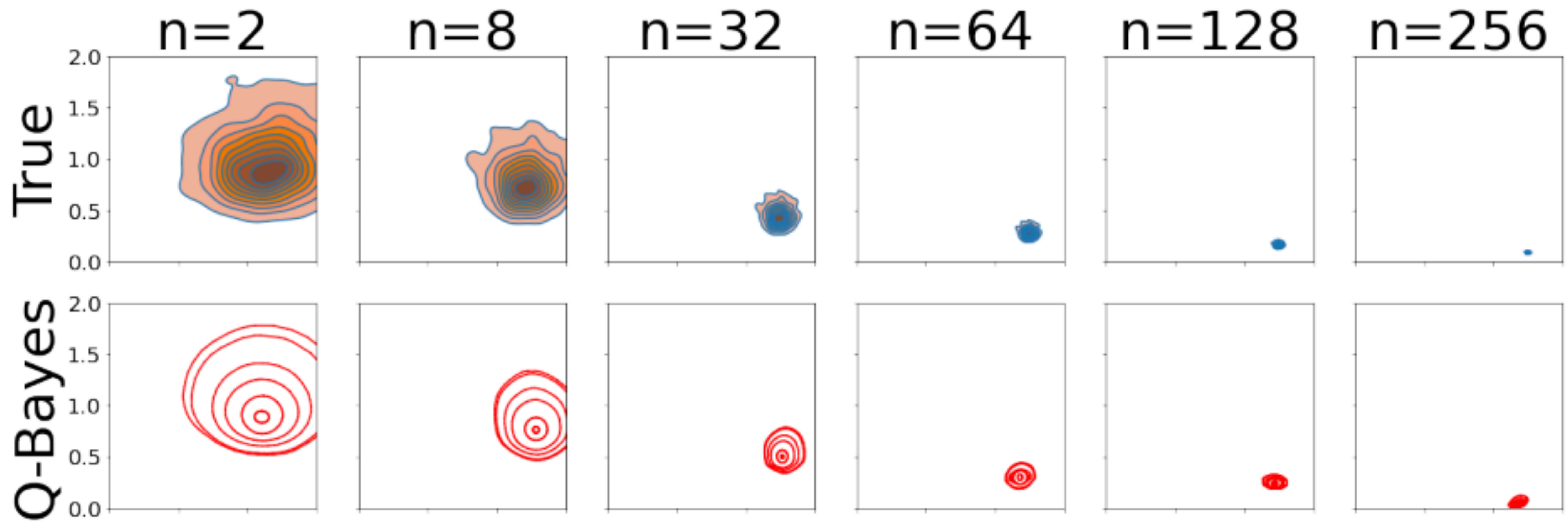


Challenge:

- 🤔 Stable Learning
- 🤔 Automatic Summary Statistics

Short Introduction

Support Shrinkage (Posterior Contraction)



Our method: **Posterior Quantile Learning**

- ✅ Stable Learning (No adversarial training)
- ✅ Automatic Summary Statistics

Byproduct:

- 🤓 Credible Set Sampling (red lines: convex hull)

Deep Generative Quantile Bayes

Deep Generative Modeling

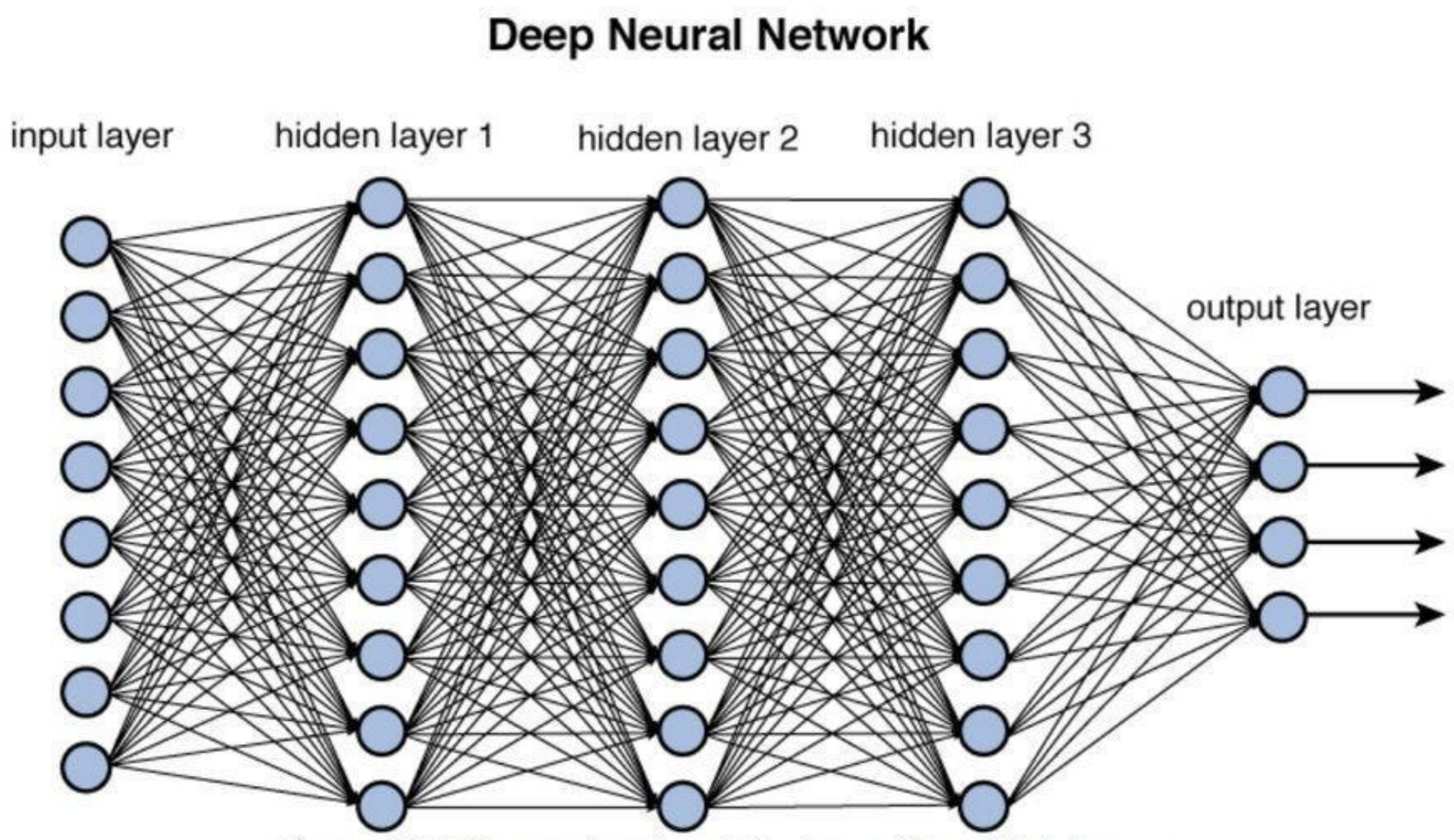


Figure 12.2 Deep network architecture with multiple layers.

Image: https://botpenguin.com/glossary/deep-neural-network

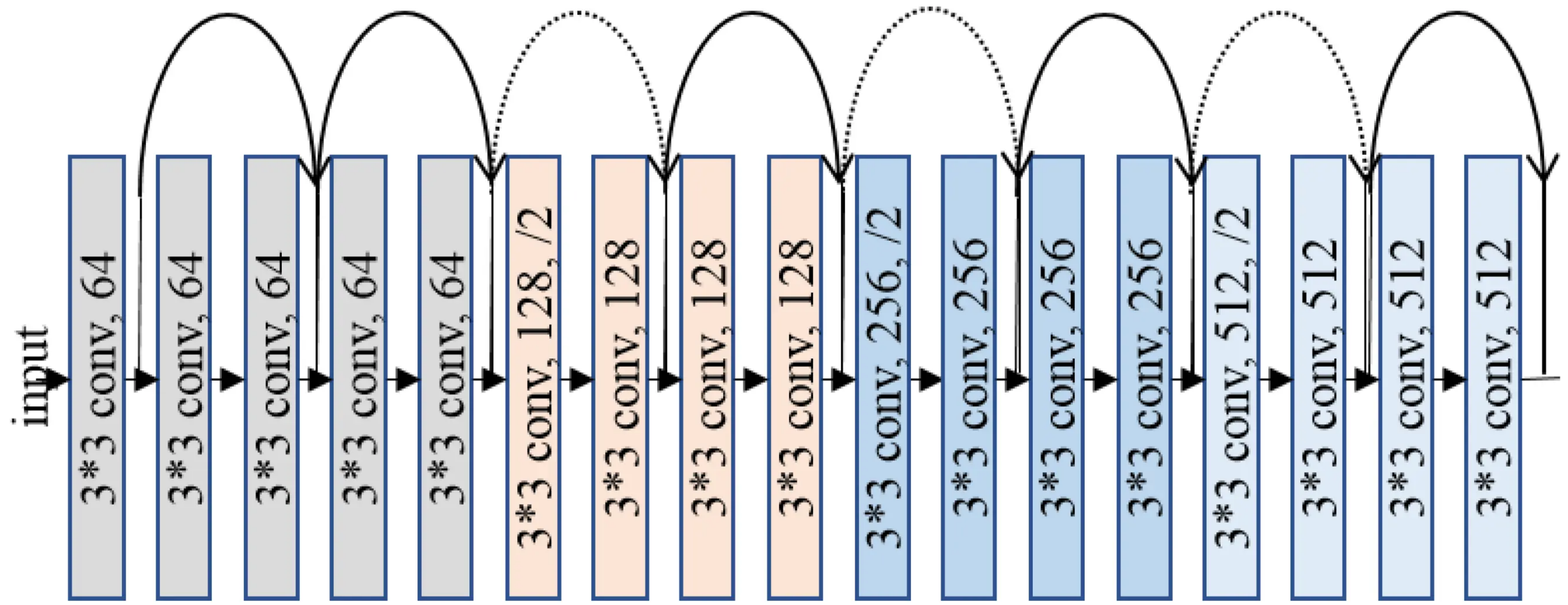
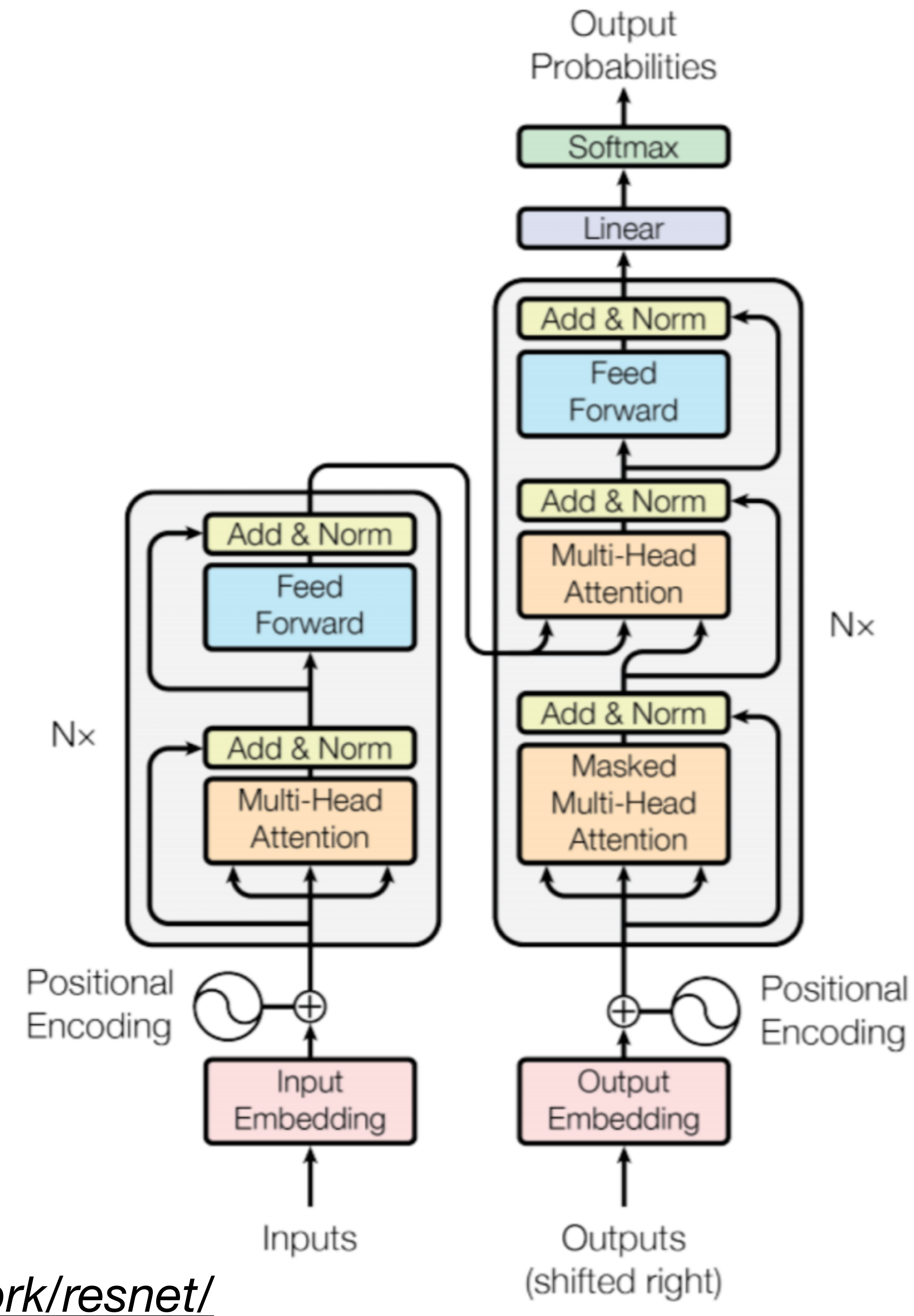
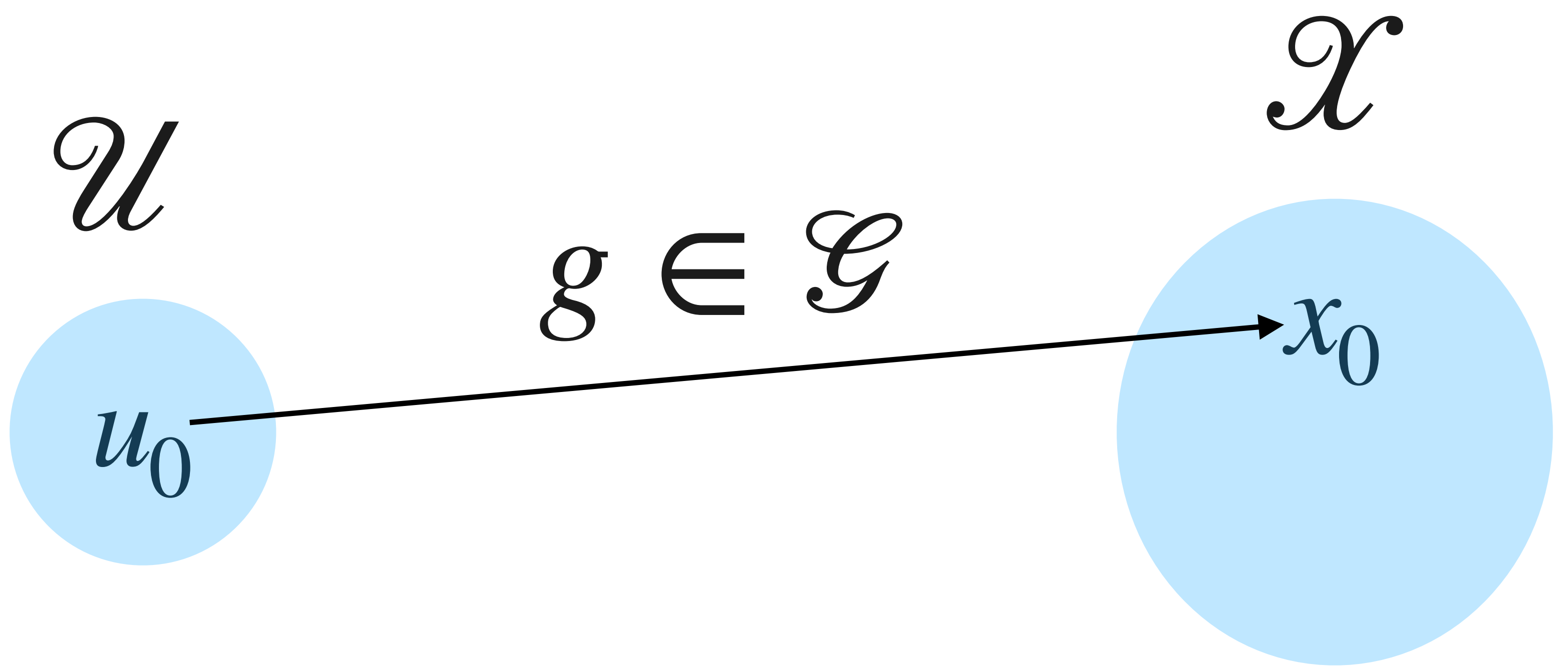


Image: https://insightfultscript.com/collections/programming/neural-network/resnet/

Image:https://arxiv.org/abs/1706.03762



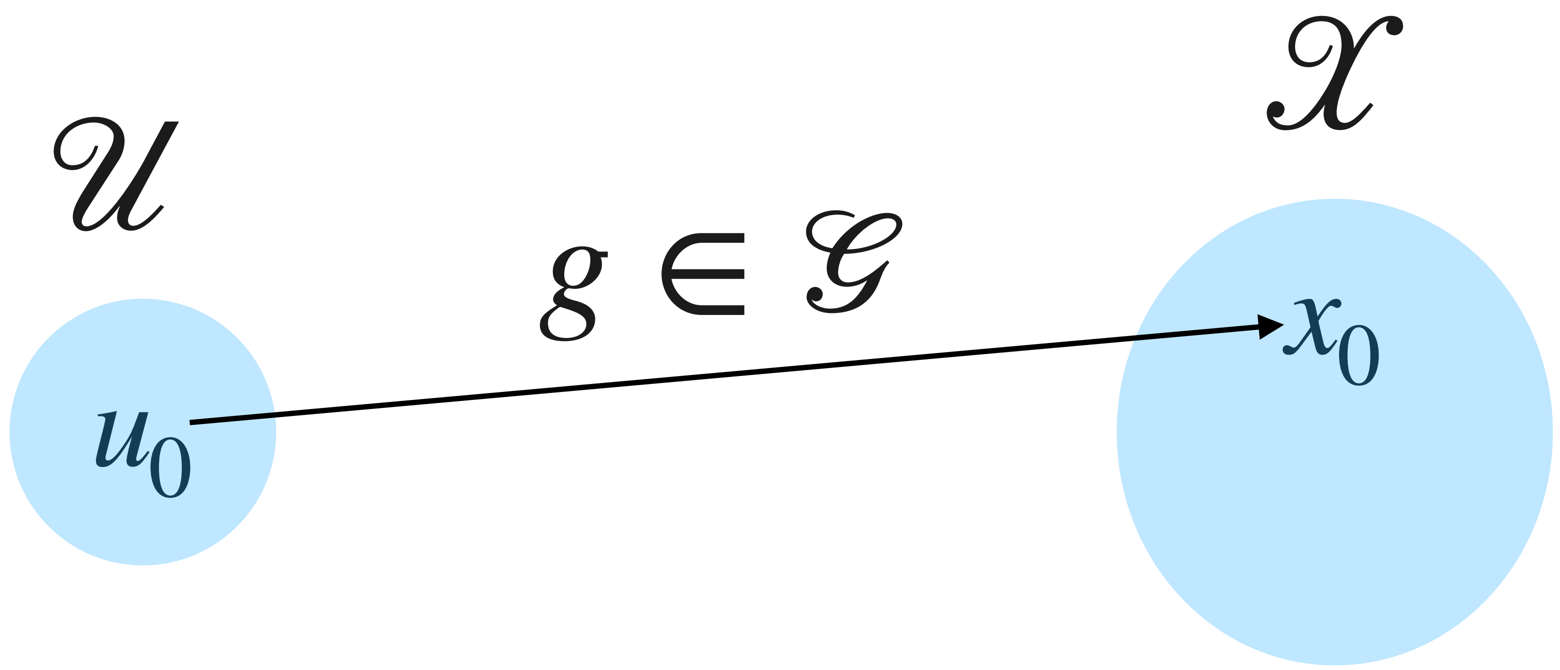
Deep Generative Modeling



Generators $\mathscr{G} = \{g : \mathscr{U} \mapsto \mathscr{X} \mid \text{defined by DNN architectures}\}$

Generation $\tilde{X} = g^*(u),\ u \sim \text{random}(\mathscr{U})$

Deep Generative Modeling



Generators $\mathscr{G} = \{g : \mathscr{U} \mapsto \mathscr{X} \,|\, \text{defined by DNN architectures}\}$

Generation $\tilde{X} = g^*(u),\ u \sim \ \text{random}(\mathscr{U})$

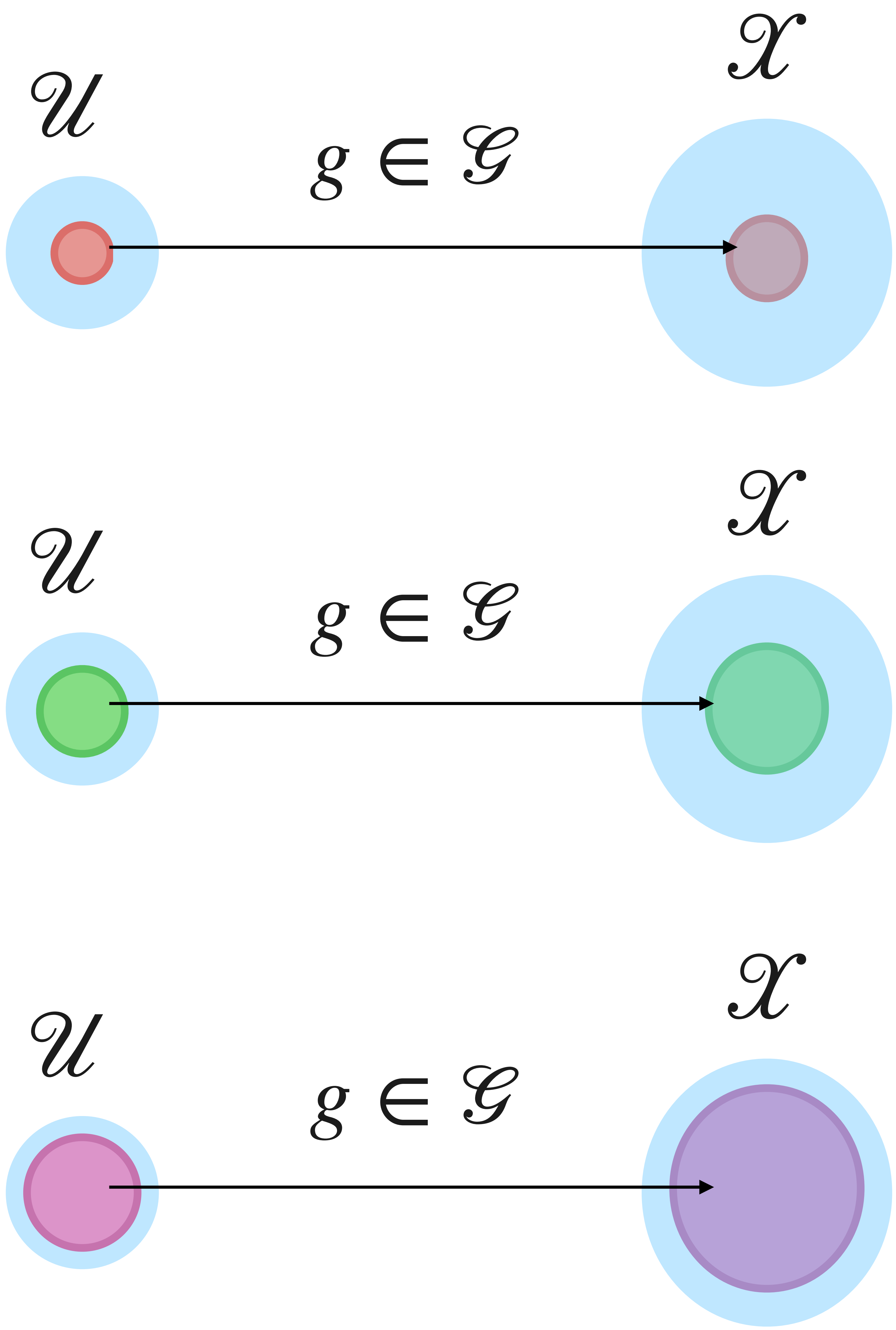
Goal $P_{\tilde{X}} \overset{d}{=} P_{\text{data}}$

Training $\mathscr{D}_{\text{train}} = \{x_i\}_{i=1}^{N} \subset \mathscr{X},\ x_i \sim P_{\text{data}}$

$$g^* = \min_{g \in \mathscr{G}} \mathscr{L}(g, \mathscr{D}_{\text{train}})$$

Deep Generative Quantile Bayes

Generative Quantile Modeling



Generative Quantile Modeling

Quantile for univariate continuous X:

F_X : the cdf of $P_X \Rightarrow F_X(X) \sim U(0,1),\ for\ X \sim P_X$

F_X^{-1} : quantile function (left continuous inverse)

$F_X^{-1}(U) \sim P_X,\ for\ U \sim Unif(0,1)$

Generative Quantile Modeling

Quantile for univariate continuous X:

F_X : the cdf of $P_X \Rightarrow F_X(X) \sim U(0,1),\ for\ X \sim P_X$

F_X^{-1} : quantile function (left continuous inverse)

$$F_X^{-1}(U) \sim P_X,\ for\ U \sim Unif(0,1)$$

Quantile Map:

$Q : \mathscr{U} \to \mathscr{X}$ **s.t.** 1) Monotonic non-decreasing

2) $Q(U) \stackrel{d}{=} X,\ \ for\ U \sim Unif(0,1)$

is $F_X^{-1}\ a.e.$

Generation:

$$F_X^{-1} \# P_U = P_X$$

$$X^{(j)} = F_x^{-1}(U^{(j)}),\ for\ U^{(j)} \sim Unif(0,1)$$

Generative Quantile Modeling

Quantile for multivariate continuous $X \in \mathbb{R}^d \sim P$:

Vector quantile notion (Carlier et al., (2016) and Chernozhukov et al., (2017))

Definition:

$Q : \mathcal{U} \to \mathcal{X}$ **s.t.** 1) A gradient of a convex function, say, $Q_p(u) = \nabla \psi(u)$

2) $Q_P(U) \stackrel{d}{=} X, \;\; for \;\; U \sim S^d(1)$

$U \sim S^d(1): \;\; U = r\phi, \; r \sim Unif[0,1], \phi \sim Unif(S^{d-1}(1)),$

where $r \perp \phi$ and $S^{d-1}(1)$ is the (d-1) dimensional unit sphere

$\rightarrow$ Such a map uniquely exists almost everywhere w.r.t $S^d(1)$

Generative Quantile Modeling

Quantile for multivariate continuous $X \in \mathbb{R}^d \sim P$:

Vector quantile notion (Carlier et al., (2016) and Chernozhukov et al., (2017))

Definition:

$Q : \mathcal{U} \to \mathcal{X}$ **s.t.** 1) A gradient of a convex function, say, $Q_p(u) = \nabla \psi(u)$

2) $Q_P(U) \stackrel{d}{=} X, \ \ for\ U \sim S^d(1)$

$U \sim S^d(1): \ \ U = r\phi,\ r \sim Unif[0,1], \phi \sim Unif(S^{d-1}(1)),$
where $r \perp \phi$ and $S^{d-1}(1)$ is the (d-1) dimensional unit sphere

Q_P as a gradient of a convex function implies:

$[Q_P(u) - Q_P(u')]^T(u - u') \geq 0$ for all $u, u' \in \mathcal{U}$

It can be seen as a generalization of the monotonicity requirement

Generative Quantile Modeling

When P has finite second moments, Q_P is the **Monge-Kantorovich** optimal transport map,

$$Q_P = \arg \min_{Q:Q\#F_U=P} \mathbb{E}_{U\sim F_U}\|Q(U) - U\|^2$$

Dual problem:

$$\psi = \arg \inf_{\varphi:\mathcal{U}\to\mathbb{R}\cup\{+\infty\}} \left(\int \varphi dF_U + \int \varphi^* dP \right)$$

where $\varphi^*(x) := \sup_{u\in\mathcal{U}} [x^T u - \psi(u)]$ is the convex conjugate

By Brenier Theorem (1991), we know $\nabla\psi(u) = Q_p(u)$

It is known that this optimal $\psi(u)$ is convex

Generative Quantile Modeling

Extension to the conditional case

For a fixed $X = x$, the conditional quantile map is a gradient of a convex function with respect to u,

$$Q_{\theta|X=x}(u) = \nabla_u \psi(u, x)$$

Primal problem:

$$\min_{U} \mathbb{E}||\theta - U||^2 : U \sim F_U, U \perp X$$

Dual problem:

$$\psi = \arg\inf_{\varphi} \left(\int \varphi(x, u) F_X(dx) F_U(du) + \int \varphi^*(x, \theta) F_{X,\theta}(dx, d\theta) \right)$$

where $\varphi^*(x, \theta)$ is the convex conjugate

Deep Generative Quantile Bayes

Vector Quantile for Generative Bayes

Assumption 1 : We assume that the quantile map Q_p is affine $f(X)$

There exists some summary statistics $f(X)$ that satisfies $X \perp \theta \mid f(X)$ and

$$\psi(u, x) = \varphi(u) + b(u)^\top f(x)$$

Relaxation of the condition: from $f(X) \perp U$ to $\mathbb{E}[f(X) \mid U] = \mathbb{E}[f(X)]$

-> Dual:
$$\inf_{b,\varphi} \left(\int \varphi(u) F_U(du) + \int \tilde{\varphi}(f(x), \theta) F_{X,\theta}(dx, d\theta) \right)$$

$$\tilde{\varphi}(f(x), \theta) := \max_{u \in \mathcal{U}} [u^\top \theta - \varphi(u) - b(u)^\top f(x)]$$

$$\varphi : \mathcal{U} \to \mathbb{R}$$

$$b : \mathcal{U} \to \mathbb{R}^q$$

Deep Generative Quantile Bayes

The Objective:

$$\mathscr{L}_1(\varphi, b, f \mid X, \theta, U) = \sum_{i=1}^{N} \left(\varphi(U_i) + \max_{j \in \{1,\ldots,N\}} \left\{ U_j^\top \theta_i - \varphi(U_j) - b(U_j)^\top f(X_i) \right\} \right),$$

where the data is generated by $\theta_i \sim \pi(\theta),\ X_i \sim P(x \mid \theta_i)$

and independently, $U_j \sim S^d(1)$

We optimize the summary statistics f and the transport φ, b together

Automatic Learning for Summary Statistics

Summary Statistics design:

Deep Set :

A neural network design for permutation invariance among the observations

LSTM:

A neural network design to reflect dependence among the observations

Credible Set Computation

Credible Set Computation:

Monge-Kantrovich Data Depth: the Tukey depth on the reversed distribution $Q_p^{-1}(X)$

the MK-depth region with probability content $\tau \in (0,1)$ is identical to $Q_p(S^d(\tau))$.

Nested structure: $\tau \leq \tau'$, $Q_p(S^d(\tau)) \subset Q_p(S^d(\tau'))$.

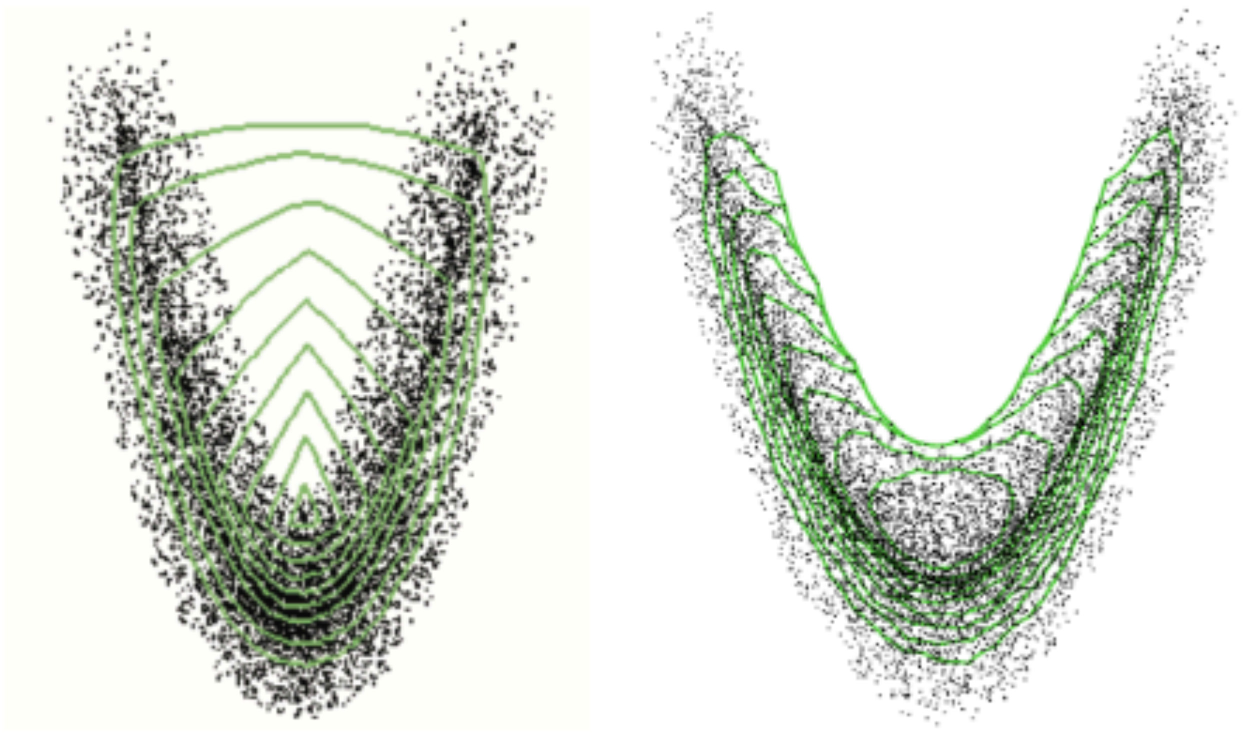
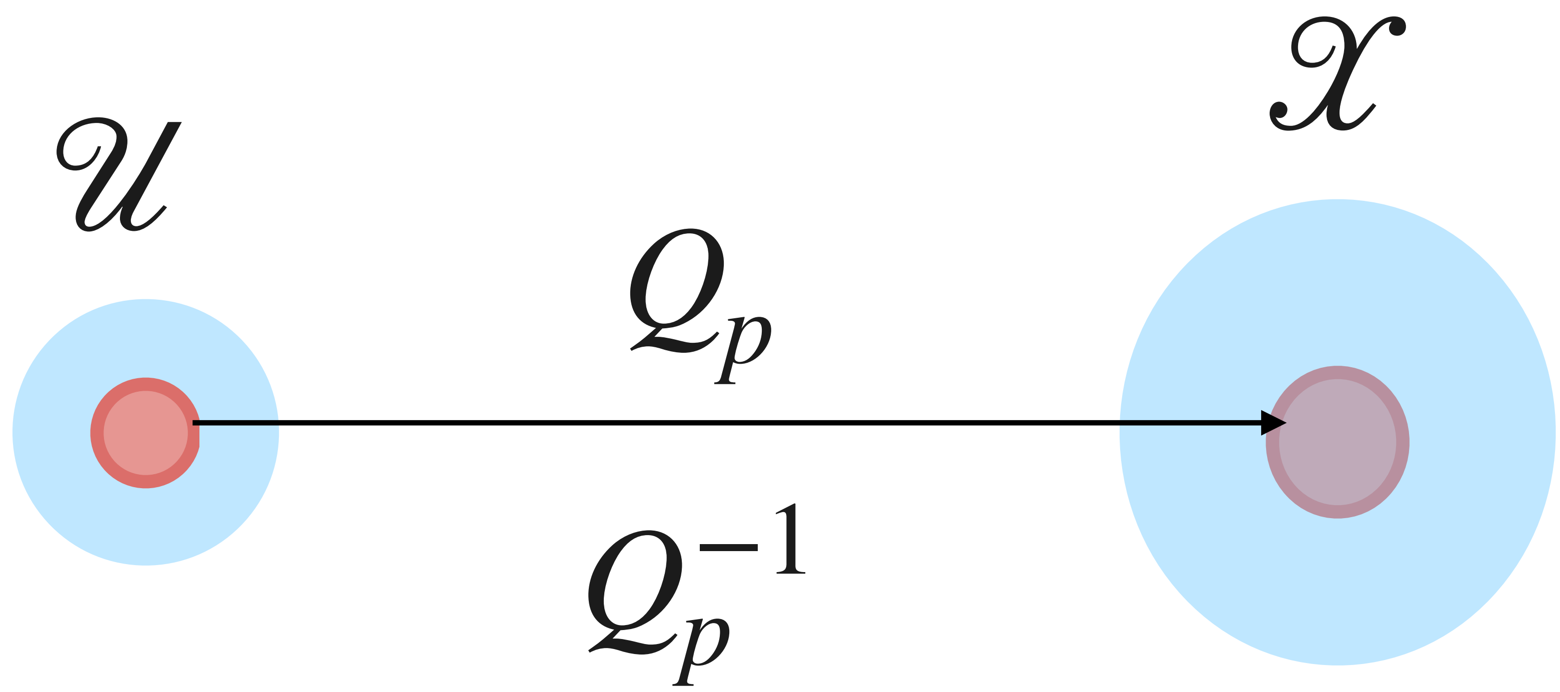


Figure 1: The data depth on a banana shaped distribution. Left: Halfspace depth (Tukey, 1975). Right: Monge-Kantorovich depth (Chernozhukov et al., 2017). The right one aligns more with our natural idea of credible set. Copied from Chernozhukov et al. (2017).



Convergence Theorem

Under mild regularity conditions and Assumption 1,

For any closed subset $K_U \subset \mathscr{U}$ *and* $K'_X \subset \mathscr{X}$ *as* $N \to \infty$,

$$\sup_{u \in K_U, x \in K_X} \|\hat{Q}^N_{\theta|X=x}(u) - Q_{\theta|X=x}(u)\| \to 0$$

Furthermore, for any closed subset $K_X \subset \mathscr{X}$,

$$\sup_{x \in K_X} W_2(\hat{\pi}^N(\theta \mid X = x), \pi(\theta \mid X = x)) \to 0.$$

Similar convergence results for the credible set.

Experiment

Brock and Hommes (1998):

$$x_{t+1} = \frac{1}{R}\sum_{h=1}^{H} n_{h,t+1}(g_h x_t + b_h) + \epsilon_{t+1},$$

$$n_{h,t+1} = \frac{\exp(\beta U_{h,t})}{\sum_{h=1}^{H}\exp(\beta U_{h,t})}$$

$$U_{h,t} = (x_t - Rx_{t-1})(g_h x_{t-2} + b_h - Rx_{t-1})$$

$$\epsilon_t \sim \mathcal{N}(0,\sigma^2)$$

$$\beta = 120,\ H = 4,\ R = 1.01,\ \sigma = 0.04,\ g_1 = b_1 = b_4 = 0,\ and\ g_4 = 1.01$$

$\theta = (g_2, b_2, g_3, b_3)$, $\mathbf{y} := (y_1, y_2, \ldots, y_T) \sim p(\mathbf{x} \mid \theta^*)$ represents the pseudo-observation, with T = 100, and $\theta^* = (g_2^*, b_2^*, g_3^*, b_3^*) = (0.9,0.2,0.9,-0.2)$ as the parameters used to generate $\mathbf{y}$. The priors are specified as $g_2, b_2, g_3 \sim \mathcal{U}(0,1)$, while $b_3 \sim \mathcal{U}(-1,0)$.

Experiment

Brock and Hommes (1998):

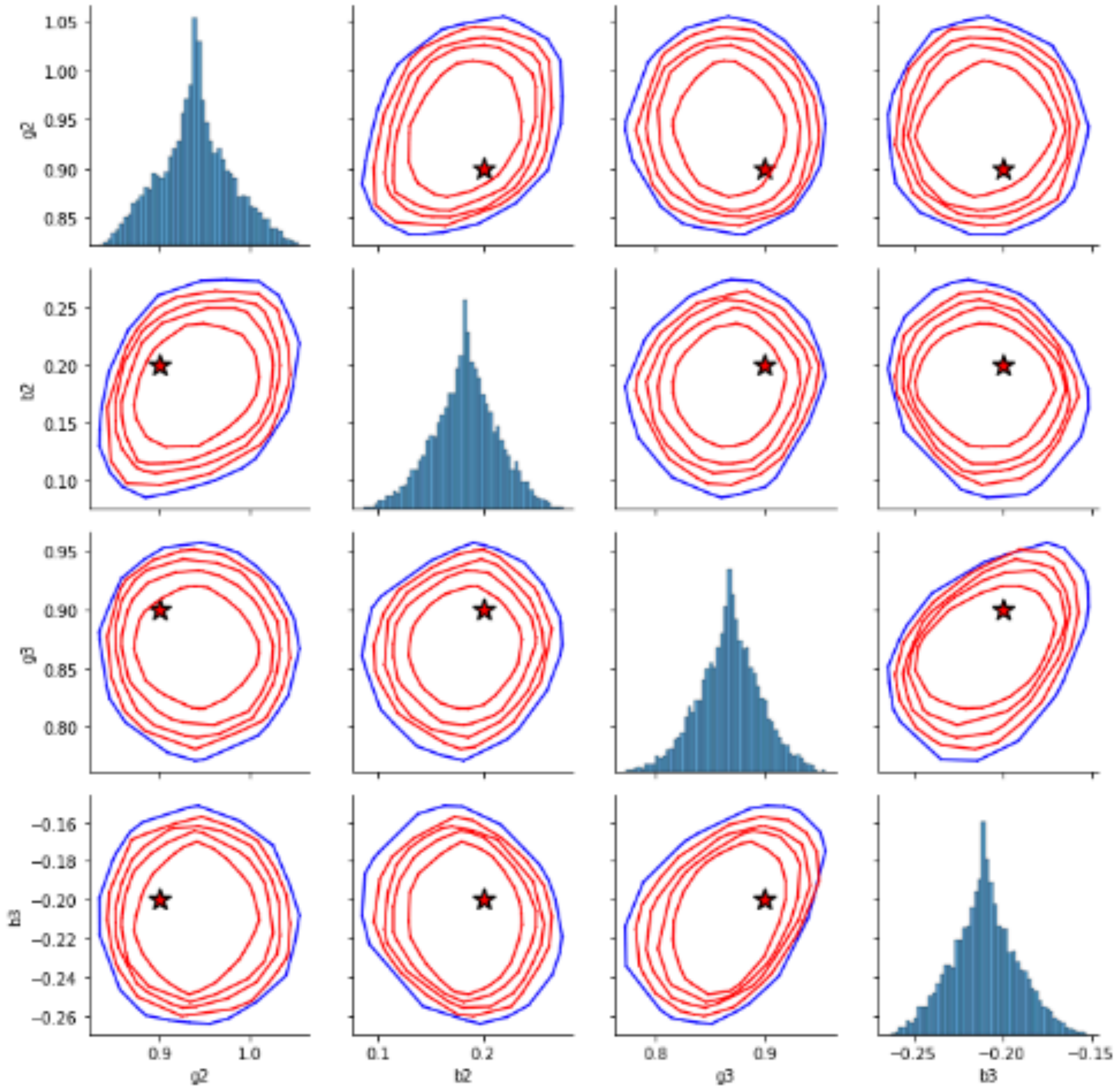


Figure 3: The Brienier map results on Brock Hommes Model, with convex hull corresponding to $\tau \in [0.5, 0.6, 0.7, 0.8, 0.9]$ (red) and $\tau = 1$ (blue). Red stars: the true parameters.

Comparison with Existing Works

Wang and Rockova (2022) :

The first work that applies deep generative models to likelihood-free posterior sampling

It learns the joint distribution of (θ, X)

It optimizes W-1 distance through adversarial training

Our work optimizes W-2 distance: no adversarial training

Once the distance is computed we instantly optain the optimal generator

Polson and Sokolov (2023): Training of the quantile map —> One dimensional case

Limitations

Efficiency: issues common in deep learning based posterior sampling

Not learning $P(\theta \mid X_0)$ but $P(\theta \mid X)$ for any generic $X \sim P_X$

When $P(\theta \mid X)$ is very complex in X, we invest too much computational power and memory to learn $P(\theta \mid \tilde{X})$, $\tilde{X} \neq X_0$ that will not be used

Need for a large number of training data

When simulation takes long time (one point generation takes five days), generating sufficient amount of data for training is not plausible

Credible Set:

Not the standard credible set if the posterior is not unimodal and symmetric

Summary

Generative Bayes Through Quantiles

Deep Generative Modeling

Deep Generative Quantile Bayes

Posterior Sampling Without Likelihood

Summary

Our Deep Generative Quantile Bayes

Vector Quantile Formulation (Monge-Kantrovich Data Depth): push-forwawrd from spherical distribution

Optimizing 2-Wasserstein distance

-> No adversarial training: Stable Optimization

Sampling from the Credible Set (estimation: convex hull)

Automatic Summary Statistics Learning

-> No need to manually choose summary statistics

-> Acheives Support Shrinkage

-> Scalable Learning

Thank you!

For more theoretical/numerical results:
Please check out https://arxiv.org/abs/2312.05411

About me: https://jungeumkim.com/