PNAS
www.pnas.org

Supplementary Information for
Density estimation using deep generative neural networks.

Qiao Liu[a,b], Jiaze Xu[c,b], Rui Jiang[a,*], and Wing Hung Wong[b,*]
[a]Ministry of Education Key Laboratory of Bioinformatics, Research Department of Bioinformatics at the Beijing National Research Center for Information Science and Technology, Center for Synthetic and Systems Biology, Department of Automation, Tsinghua University, Beijing 100084, China; [b]Department of Statistics, Department of Biomedical Data Science, Bio-X Program, Stanford University, Stanford, CA 94305; and [c]Center for Statistical Science and Department of Industrial Engineering, Tsinghua University, Beijing 100084, China

[*]To whom correspondence may be addressed. Email: ruijiang@tsinghua.edu.cn or whwong@stanford.edu.

This PDF file includes:

Supplementary texts S1-S4
Figures S1 to S4
Tables S1 to S3
SI References

Supplementary Information Texts

Text S1. The derivation of density estimation based on Laplace approximation

In the Methods section, density is modeled as

$$p_{\mathrm{x}}(\mathbf{x}) = (\tfrac{1}{\sqrt{2\pi}})^{m+n}\sigma^{-n}\int e^{-\frac{v(\mathbf{x},\mathbf{z})}{2}}d\mathbf{z} \qquad \textbf{[1]}$$

where $v(\mathbf{x},\mathbf{z}) = ||\mathbf{z}||_2^2 + \sigma^{-2}||\mathbf{x} - G(\mathbf{z})||_2^2$. Note that $\mathbf{z}$ is a latent variable which follows a standard Gaussian distribution. First, we expand $G(\mathbf{z})$ around $\tilde{\mathbf{z}} = H(\mathbf{x})$ to obtain a quadratic approximation, which can be represented as

$$\mathbf{x} - G(\mathbf{z}) \approx \mathbf{x} - G(\tilde{\mathbf{z}}) - \nabla G(\tilde{\mathbf{z}})(\mathbf{z} - \tilde{\mathbf{z}}) \qquad \textbf{[2]}$$

where $\nabla G(\tilde{\mathbf{z}}) \in \mathbb{R}^{n\times m}$ is the Jacobian matrix of $G(\cdot)$ at $\tilde{\mathbf{z}}$. Substitute **[2]** into $||\mathbf{x} - G(\mathbf{z})||_2^2$, we have

$$\left\|\mathbf{x} - G(\mathbf{z})\right\|_2^2 = \left(\mathbf{x} - G(\mathbf{z})\right)^T\left(\mathbf{x} - G(\mathbf{z})\right) = ||\mathbf{x} - G(\tilde{\mathbf{z}})||_2^2 - 2\left(\mathbf{x} - G(\tilde{\mathbf{z}})\right)^T\nabla G(\tilde{\mathbf{z}})(\mathbf{z} - \tilde{\mathbf{z}}) + (\mathbf{z} - \tilde{\mathbf{z}})^T\nabla G^T(\tilde{\mathbf{z}})\nabla G(\tilde{\mathbf{z}})(\mathbf{z} - \tilde{\mathbf{z}}) \qquad \textbf{[3]}$$

Next, we made variable substitutions as

$$\begin{cases} \mathbf{A} = \nabla G^T(\tilde{\mathbf{z}})\nabla G(\tilde{\mathbf{z}}) \ \in \mathbb{R}^{m\times m} \\ \mathbf{b} = \nabla G^T(\tilde{\mathbf{z}})(\mathbf{x} - G(\tilde{\mathbf{z}})) \ \in \mathbb{R}^{m} \\ \mathbf{w} = \mathbf{z} - \tilde{\mathbf{z}} \ \in \mathbb{R}^{m} \\ \lambda = \sigma^{-2} \end{cases} \qquad \textbf{[4]}$$

Taking equations **[3]** and **[4]** into $v(\mathbf{x},\mathbf{z})$, we can get

$$v(\mathbf{x},\mathbf{z}) = \tilde{v}(\mathbf{x},\mathbf{w}) = ||\mathbf{w}||_2^2 + 2\mathbf{w}^T\tilde{\mathbf{z}} + ||\tilde{\mathbf{z}}||_2^2 + \lambda(||\mathbf{x} - G(\tilde{\mathbf{z}})||_2^2 - 2\mathbf{b}^T\mathbf{w} + \mathbf{w}^T\mathbf{A}\mathbf{w}) = \mathbf{w}^T(\mathbf{I} + \lambda\mathbf{A})\mathbf{w} - 2(\lambda\mathbf{b} - \tilde{\mathbf{z}})^T\mathbf{w} + c_1(\mathbf{x}) \qquad \textbf{[5]}$$

where $\mathbf{I} \in \mathbb{R}^{m\times m}$ is the identity matrix and $c_1(\mathbf{x}) = ||\tilde{\mathbf{z}}||_2^2 + \lambda||\mathbf{x} - G(\tilde{\mathbf{z}})||_2^2$. The integral in **[1]** *w.r.t* $\mathbf{z}$ can now be solved by constructing a multivariate Gaussian distribution *w.r.t* $\mathbf{w}$ in **[5]** as the following

$$\int e^{-\frac{v(\mathbf{x},\mathbf{z})}{2}}d\mathbf{z} = \int e^{-\frac{\tilde{v}(\mathbf{x},\mathbf{w})}{2}}d\mathbf{w} = \int e^{-\frac{(\mathbf{w}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\boldsymbol{\mu})+c(\mathbf{x})}{2}}d\mathbf{w} = e^{-\frac{c(\mathbf{x})}{2}}\int e^{-\frac{(\mathbf{w}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\boldsymbol{\mu})}{2}}d\mathbf{w} = e^{-\frac{c(\mathbf{x})}{2}}\sqrt{(2\pi)^m\det(\boldsymbol{\Sigma})} \qquad \textbf{[6]}$$

where $c(\mathbf{x}) = c_1(\mathbf{x}) - \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, $\det(\boldsymbol{\Sigma})$ denotes the determinant of the covariance matrix $\boldsymbol{\Sigma}$. The constructed mean and covariant matrix of the multivariate Gaussian are formulated as

$$\begin{cases} \boldsymbol{\Sigma} = (\mathbf{I} + \lambda\mathbf{A})^{-1} \\ \boldsymbol{\mu} = \boldsymbol{\Sigma}(\lambda\mathbf{b} - \tilde{\mathbf{z}}) \end{cases} \qquad \textbf{[7]}$$

Substitute **[6]** into **[1]**, then we can get the final closed-form solution for density of $\mathbf{x}$ as

$$p^{LP}(\mathbf{x}) = (\tfrac{1}{\sqrt{2\pi}})^n\sigma^{-n}\sqrt{\det(\boldsymbol{\Sigma})}e^{-\frac{c(\mathbf{x})}{2}} \qquad \textbf{[8]}$$

Note that the equation **[8]** is not an accurate probability distribution as the approximation error comes from the quadratic approximation in **[2]**. Besides, compared to previous methods that require the calculation of the determinant and the inverse of a $n \times n$ Jacobian matrix. Laplace approximation is flexible in setting the dimension of the base density m. The computation in Laplace approximation only involves the determinant and inverse of a $m \times m$ matrix $\mathbf{A}$, which can achieve a faster computation when $m < \mathrm{n}$.

Text S2. The *change of variable rule* as a special case in proposed framework.
Assume that we have the following three conditions: 1) $m = n$ 2) $H(\cdot) = G^{-1}(\cdot)$ 3) $\sigma \to 0$. We first rephrase the density of Laplace approximation in **[8]** as the following

$$\begin{aligned} p^{LP}(\mathbf{x}) &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \sigma^{-n} \sqrt{\det\big(\mathrm{inv}(\mathbf{I} + \sigma^{-2}\mathbf{A})\big)}\, e^{-\frac{c(\mathbf{x})}{2}} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \sigma^{-n} \sqrt{\det\big(\sigma^2 \mathrm{inv}(\mathbf{A} + \sigma^2\mathbf{I})\big)}\, e^{-\frac{c(\mathbf{x})}{2}} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \sigma^{-n} \sqrt{\sigma^{2m} \det\big(\mathrm{inv}(\mathbf{A} + \sigma^2\mathbf{I})\big)}\, e^{-\frac{c(\mathbf{x})}{2}} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \sigma^{m-n} \sqrt{\det\big(\mathrm{inv}(\mathbf{A} + \sigma^2\mathbf{I})\big)}\, e^{-\frac{c(\mathbf{x})}{2}} \\ &= (\frac{1}{\sqrt{2\pi}})^n \sqrt{\det\,(\mathrm{inv}(\mathbf{A} + \sigma^2\mathbf{I}))} e^{-\frac{c(\mathbf{x})}{2}} \end{aligned}$$

[9]

where $\det\,(\cdot)$ and $\mathrm{inv}(\cdot)$ denotes the determinant and inverse of a matrix. Using condition 2), we have $\mathbf{x} - G(\tilde{\mathbf{z}}) = \mathbf{x} - G\big(H(\mathbf{x})\big) = \mathbf{x} - G\big(G^{-1}(\mathbf{x})\big) = \mathbf{0}$, $\mathbf{b} = \nabla G^T(\tilde{\mathbf{z}})\big(\mathbf{x} - G(\tilde{\mathbf{z}})\big) = \mathbf{0}$, $\boldsymbol{\mu} = \boldsymbol{\Sigma}(\lambda\mathbf{b} - \tilde{\mathbf{z}}) = -\boldsymbol{\Sigma}\tilde{\mathbf{z}}$. So, we have

$$\begin{aligned} c(\mathbf{x}) &= \big|\big|\tilde{\mathbf{z}}\big|\big|_2^2 + \lambda\big|\big|\mathbf{x} - G(\tilde{\mathbf{z}})\big|\big|_2^2 - \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\ &= \big|\big|\tilde{\mathbf{z}}\big|\big|_2^2 + \mathbf{0} - (-\boldsymbol{\Sigma}\tilde{\mathbf{z}})^T\boldsymbol{\Sigma}^{-1}(-\boldsymbol{\Sigma}\tilde{\mathbf{z}}) \\ &= \big|\big|\tilde{\mathbf{z}}\big|\big|_2^2 - \tilde{\mathbf{z}}^T\boldsymbol{\Sigma}\tilde{\mathbf{z}} = \big|\big|\tilde{\mathbf{z}}\big|\big|_2^2 - \tilde{\mathbf{z}}^T(\mathbf{I} + \lambda\mathbf{A})^{-1}\tilde{\mathbf{z}} \\ &= \big|\big|\tilde{\mathbf{z}}\big|\big|_2^2 - \sigma^2\tilde{\mathbf{z}}^T(\mathbf{A} + \sigma^2\mathbf{I})^{-1}\tilde{\mathbf{z}} \end{aligned}$$

[10]

Finally, we take the limit of σ by condition 3), we have $\lim\limits_{\sigma\to 0} c(\mathbf{x}) = ||\tilde{\mathbf{z}}||_2^2$, and

$$\begin{aligned} &\operatorname*{Lim}_{\sigma\to 0} p^{LP}(\mathbf{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \sqrt{\det\big(\mathrm{inv}(\mathbf{A})\big)}\, e^{-\frac{||\tilde{\mathbf{z}}||_2^2}{2}} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \sqrt{\det\big(\mathrm{inv}(\mathbf{J}_{\tilde{\mathbf{z}}}^T\mathbf{J}_{\tilde{\mathbf{z}}})\big)}\, e^{-\frac{||\tilde{\mathbf{z}}||_2^2}{2}} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \sqrt{\det(\mathbf{J}_{\tilde{\mathbf{z}}}^{-1}\mathbf{J}_{\tilde{\mathbf{z}}}^{-T})}\, e^{-\frac{||\tilde{\mathbf{z}}||_2^2}{2}} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{||\tilde{\mathbf{z}}||_2^2}{2}} |\det(\mathbf{J}_{\tilde{\mathbf{z}}})|^{-1} \\ &= p_{\tilde{\mathbf{z}}}(\tilde{\mathbf{z}})|\det\,(\mathbf{J}_{\tilde{\mathbf{z}}})|^{-1} \end{aligned}$$

[11]

where $p(\tilde{\mathbf{z}})$ denotes a base density (standard Gaussian) for $\tilde{\mathbf{z}}$, and $\mathbf{J}_{\tilde{\mathbf{z}}}$ denotes the Jacobian matrix $\nabla G(\tilde{\mathbf{z}})$ or $\frac{\partial G(\tilde{\mathbf{z}})}{\partial \tilde{\mathbf{z}}^T}$.

To sum up, we proved that under the three conditions 1) $m = n$ 2) $H(\cdot) = G^{-1}(\cdot)$ 3) $\sigma \to 0$, the proposed Laplace approximation is degraded into the *change of variable rule,* which is the principle of previous neural density estimators. Our Laplace approximation approach can be considered as an extension of the *change of variable rule* which requires equal dimension in base density and target density.

Text S3. The details of data preprocessing of datasets for density estimation used in our study

AreM. The Activity Recognition system based on Multisensor data fusion (AReM) (1) dataset contains temporal data from a Wireless Sensor Network worn by an actor performing the activities: bending, cycling, lying down, sitting, standing, walking. The time-domain features including 3 mean values and 3 standard deviations were collected from the multisensor system during a period of time. Although it is time-series data but we treat it as if each example was drawn from an *i.i.d.* distribution from the target distribution. Then raw data was first applied a feature scaling through a min-max normalization and then randomly split into 90% training set and 10% test. Note that for neural density estimators, 10% of the training set will be kept for validation.

CASP. The CASP dataset contains the physicochemical properties of the protein tertiary structure. Each example denotes an individual residue which has 9 features, including total surface area, non-polar exposed area, fractional area of exposed non-polar residue, fractional area of exposed non-polar part of the residue, molecular mass weighted exposed area, Euclidian distance, secondary structure penalty and spacial distribution constraints (N.K Value). The same data normalization and split were used as AreM dataset.

HEPMASS. HEPMASS (2) dataset describes the particle collisions signatures of exotic particles in high energy physics. We preprocessed this dataset following the same strategy suggested by (3). Examples from the "1000" dataset were collected where the particle mass is 1000 and five features were removed due to too many reoccurring values.

BANK. BANK dataset (4) is related to a marketing campaign of a Portuguese banking institution where the goal is to predict whether the client will subscribe a deposit. The label encoding was used for discrete features in the raw data with values between 0 and `n_classes`. Then a uniform noise of (-0.2,0.2) was added to each feature. At last, the same data normalization and split were used as AreM dataset.

YPMSD. YPMSD (http://millionsongdataset.com/) is a dataset that contains the audio features of songs from different years ranging from 1922 to 2011. Each song has 90 features which relate to 12 timbre average and 78 timbre covariance. The same data normalization and split were used as AreM dataset.

The descriptions of the five UCI datasets and the two image datasets (MNIST and CIFAR-10), including feature dimension and sample size, were summarized in Table S1.

Text S4. The details of data preprocessing of datasets for outlier detection used in our study

Shuttle. Shuttle (http://odds.cs.stonybrook.edu/shuttle-dataset/) dataset contains 9 numerical features. The smallest five classes, i.e. 2, 3, 5, 6, 7 are combined to form the class of outliers, while class 1 forms the inlier class. Data for class 4 is discarded. All inlier and outlier data were first mixed together and then randomly split into 90% training set and 10% test set. For neural density estimators, 10% of the training set were kept for validation.

Mammography Mammography (http://odds.cs.stonybrook.edu/mammography-dataset/) dataset describes the characteristics of 260 calcifications. The minority class of calcification is considered as an outlier class and the non-calcification class as inliers. The same data split strategy was used for Shuttle dataset.

ForestCover ForestCover (http://odds.cs.stonybrook.edu/forestcovercovertype-dataset/) dataset is used in predicting forest cover type from cartographic variables. Outlier detection dataset is created using only 10 quantitative attributes. Instances from class 2 are considered as normal points and instances from class 4 are anomalies. The same data split strategy was used for Shuttle dataset.

The descriptions of the three ODDS datasets are summarized in Table S2.

Supplementary Figures

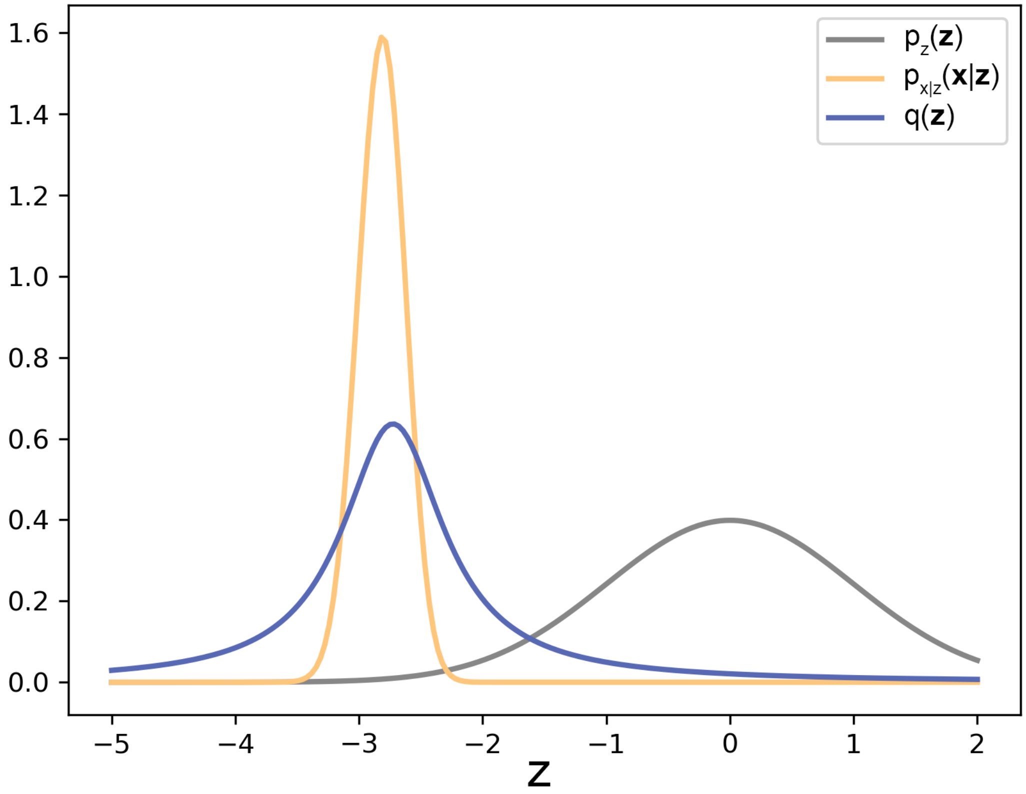


Fig. S1. To make the importance sampling strategy more understandable, we illustrated an example based on the simulation study here. We take the *Involute* simulation dataset for an example. After model training, we visualize $\boldsymbol{p}_{\mathbf{z}}(\mathbf{z})$, $\boldsymbol{p}_{\mathbf{x}|\mathbf{z}}(\mathbf{x} = (\mathbf{3}, \mathbf{3})|\mathbf{z})$ and $\boldsymbol{q}(\mathbf{z})$ for the first dimension. As $\boldsymbol{p}_{\mathbf{x}|\mathbf{z}}(\mathbf{x}|\mathbf{z})$ typically decays much faster than $\boldsymbol{p}_{\mathbf{z}}(\mathbf{z})$, we chose $\boldsymbol{q}(\mathbf{z})$ in which the center is close to the center of $\boldsymbol{p}_{\mathbf{x}|\mathbf{z}}(\mathbf{x}|\mathbf{z})$ as much as possible. To sum up, in the importance sampling strategy, $\boldsymbol{G}(\mathbf{z})$ network was used for generating samples while $\boldsymbol{H}(\mathbf{x})$ network was used for determining the center of importance distribution $\boldsymbol{q}(\mathbf{z})$.

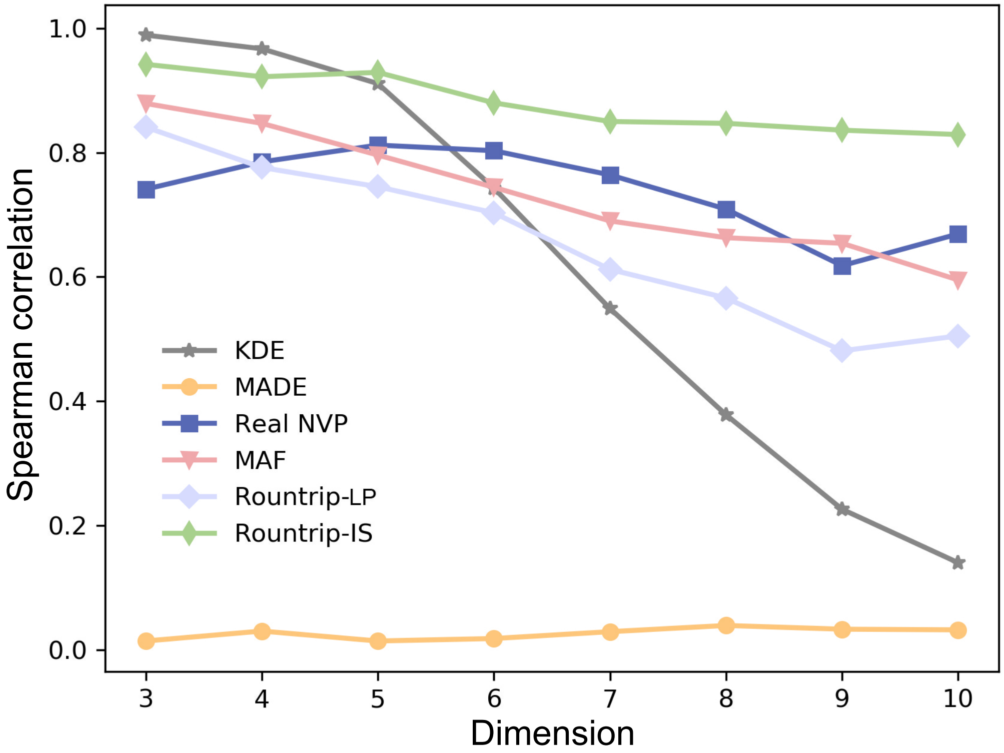
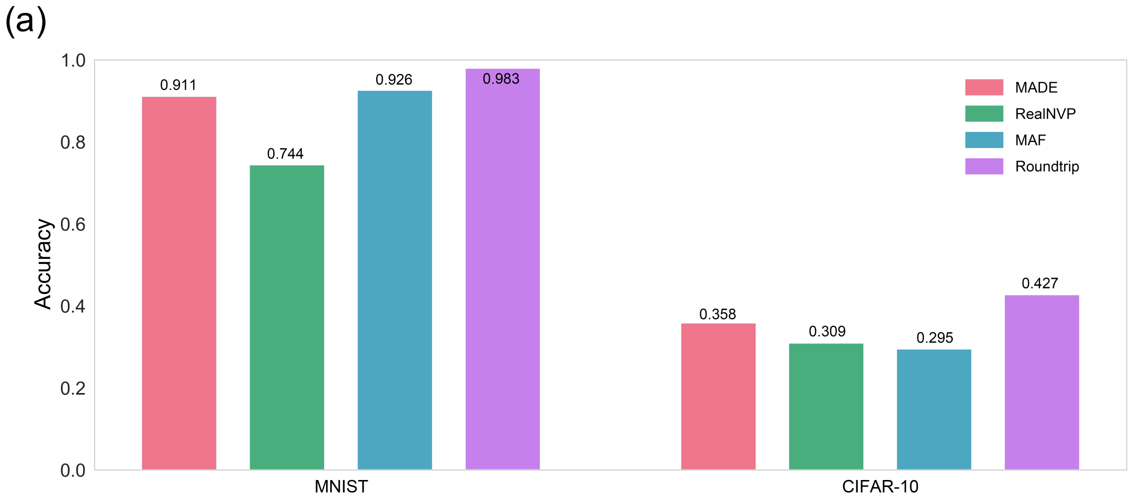


Fig. S2. we took the case (a) independent Gaussian mixture for a further study by increasing the dimension up to 10 (containing $\mathbf{3^{10}}$ modes). The Spearman correlation between estimated density and true density of the test set is calculated. The kernel density estimator (KDE) performs comparable or even better when the dimension is less than 5. But the performance of KDE decreases sharply when the dimension is larger than 5. Our Roundtrip model with the importance sampling (Roundtrip-IS) strategy can achieve a consistently better performance than other neural density estimators at different dimensions. We also note that the performance of Roundtrip model with Laplace approximation (Roundtrip-LP) outperforms MADE but not as good as MAF and RealNVP in most cases. The theoretical guarantees (Text S1) on the Laplacian solution suggested that the success of Roundtrip-LP requires that the high order terms in equation [2] is negligible, which may introduce bias related to the data distribution. Therefore, we reported all results of density estimation using the more robust Roundtrip-IS model (default setting).



(b)

	MADE	RealNVP	MAF	Roundtrip
MNIST	-2.66 ± 0.133	-4.82 ± 2.21	-1.47 ± 0.06	-0.93 ± 0.08
CIFAR-10	-33.27 ± 0.49	-23.69 ± 1.41	-62.30 ± 0.87	-19.53 ± 0.74

Fig. S3. Conditional density estimation with different neural density models. (a) After training the models with labelled image data, a Bayesian posterior probability was calculated for image classification on test images. The test accuracy of different methods on both MNIST and CIFAR-10 image databases were shown. (b) The average log Bayesian posterior probability (Bayesian score) for test data and the corresponding standard deviation were illustrated in the table.

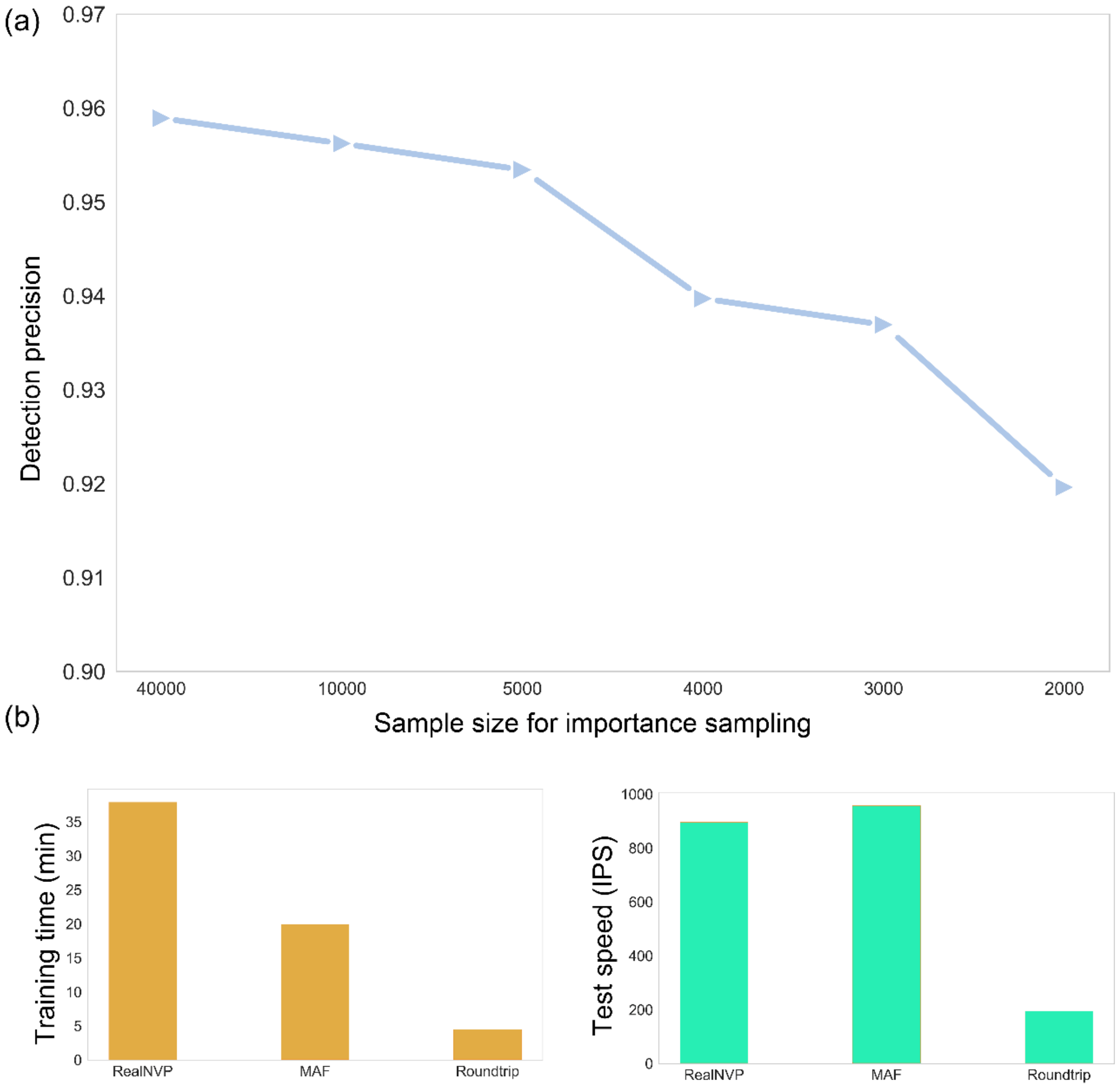


Fig. S4. The impact of sample size for importance sampling (IS) and the running time comparison. The results were based on the outlier detection "Shuttle" dataset. (a) It is noted that the performance (precision at *k*) will slowly decrease if the IS sampling size becomes smaller. The precision at *k* only decreases about 0.27% when sample size changes from 40000 to 10000. (b) The IS sample size has a significant impact on the speed of Roundtrip. There is a trade-off between fast density evaluation and accuracy of estimated densities. Roundtrip has an advantage in model training (around 4.6 mins, 10 epochs for Roundtrip), which is about 4x faster than MAF and 8x faster than RealNVP. In the density evaluation stage, the IS sample size was set to 2500 for Roundtrip, which is enough for achieving the best performance among comparison models. Roundtrip achieves a test speed of 197 instances per second (IPS), which is around 4.5 and 4.8 times slower than RealNVP and MAF, respectively. All the experiments were carried on a Linux platform with an AMD EPYC 7502P CPU (32 cores), 250 Gb memory and a single GeForce RTX 2080 Ti GPU was used.

Supplementary Tables

Table S1. Basic descriptions of 7 datasets for density estimation used in our study.

Dataset	Domain	Dim($\mathbf{z}$)	Dim($\mathbf{x}$)	# of examples Train	 Validation	 Test
AReM	Social science	3	6	34215	3801	4223
CASP	Chemistry	5	9	37042	4115	4573
HEPMASS	Physics	8	21	315123	35013	174987
BANK	Finance	8	17	36621	4069	4521
YPMSD	Audio	20	90	417430	46381	51534
MNIST	Image	100	784	50000	10000	10000
CIFAR-10	Image	100	3072	45000	5000	10000

Table S2. Basic descriptions of 3 datasets for outlier detection used in our study

Dataset	Dim($\mathbf{z}$)	Dim($\mathbf{x}$)	Outliers(%)	# of examples Train	 Validation	 Test
Shuttle	3	9	7	39770	4418	4909
Mammograph	3	6	2.32	9059	1006	1118
ForestCover	4	10	0.9	231700	25744	28604

Table S3. The network architecture for conditional image generation and density estimation. Take the MNIST database for an example. The one-hot encoded label $\mathbf{y}$ will be fed to both G network and D_x network. Note that $\mathbf{yb}$ is a reshape of $\mathbf{y}$, which is convenient for channel-wise concatenation. In the generator G, the transposed convolutions (Upconv) were used for up-sampling. For CIFAR-10 database, the network architecture used in Roundtrip is exactly the same except that the image size will be $32 \times 32 \times 3$ and the hidden unit number in the second fully-connected layer of G network will be $8\times 8 \times 128$.

Generator G	**Discriminator** D_x
Inputs: $\mathbf{z} \in \mathbb{R}^{100}$ and $\mathbf{y} \in \mathbb{R}^{10}$	Inputs: flattened image $\in \mathbb{R}^{784}$ and $\mathbf{y} \in \mathbb{R}^{10}$
Concat($\mathbf{z}, \mathbf{y}$)	Reshape, $28 \times 28 \times 1$, $1\times 1 \times 10$
FC, 1024 batchnorm, LRelu	4×4 conv, kernels 32, stride 2, batchnorm, LRelu
Concat(FC, $\mathbf{y}$)	Concat(Conv, $\mathbf{yb}$)
FC, $7\times 7 \times 28$, batchnorm, LRelu	4×4 conv, kernels 64, stride 2, batchnorm, LRelu
Reshape, $7\times 7 \times 28$	Flatten, 1568
Concat(Reshape, $\mathbf{yb}$)	Concat(Flat, $\mathbf{yb}$)
4×4 Upconv, kernels 64, stride 2, LRelu	FC, 1024 batchnorm, LRelu
4×4 Upconv, kernels 64, stride 2, Sigmoid	Concat(FC, $\mathbf{y}$)
Flatten 784	FC, 1
Generator H	**Discriminator** D_z
Inputs: flattened image $\in \mathbb{R}^{784}$	Inputs: $\mathbf{z} \in \mathbb{R}^{100}$
Reshape, $7\times 7 \times 28$	FC, 128, LRelu
4×4 Conv, kernels 64, stride 2, LRelu	FC, 128 batchnorm, Tanh
4×4 Conv, kernels 64, stride 2, LRelu	FC, 1
FC, 1024	
FC, 100	

References

1. F. Palumbo, C. Gallicchio, R. Pucci, A. Micheli, Human activity recognition using multisensor data fusion based on reservoir computing. *Journal of Ambient Intelligence and Smart Environments* **8**, 87-107 (2016).
2. P. Baldi, K. Cranmer, T. Faucett, P. Sadowski, D. Whiteson, Parameterized machine learning for high-energy physics. *arXiv preprint arXiv:1601.07913* (2016).
3. G. Papamakarios, T. Pavlakou, I. Murray (2017) Masked autoregressive flow for density estimation. in *Advances in Neural Information Processing Systems*, pp 2338-2347.
4. S. Moro, P. Cortez, P. Rita, A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems* **62**, 22-31 (2014).