Neural Inverse Transform Sampler

Henry Li, Yuval Kluger
Density Estimation
Density Estimation
Density Estimation
Ideal Properties
Ideal Properties

Flexible Density Approximation
Ideal Properties

Flexible Density Approximation

Tractable Evaluation

Input → Output $p(\cdot)$
Ideal Properties

Flexible Density Approximation

Tractable Evaluation

Fast Sampling

Input

Output \( p(\cdot) \)

Generate, , or
Flexible Density Approximation

\[ p(x = \theta) \propto f(x, \theta) \]
Flexible Density Approximation

\[ p(x = \text{cat}) \propto f(x, \theta) \]

Positivity

\[ p(x) \geq 0 \quad x \in A \]

Easy
Flexible Density Approximation

\[ p(x = \quad ) \propto f(x, \theta) \]

\[ p(x) \geq 0 \quad x \in A \]

\[ \int_A p(x) dx = 1 \]

Positivity

Integration to Unity

General case: NP Hard!
Integration Trick (in 1D)

Fundamental Theorem of Calculus. Let $F$ be such that

$$F(x) = \frac{d}{dx} f(x) \quad \text{for all } x \in A.$$
Integration Trick (in 1D)

Fundamental Theorem of Calculus. Let $F$ be such that

$$F(x) = \frac{d}{dx} f(x) \quad \text{for all } x \in A.$$ 

Then, if $A \in [a, b]$,

$$\int_{a}^{b} f(x) \, dx = F(b) - F(a).$$
Integration Trick (in 1D)

Fundamental Theorem of Calculus. Let $F$ be such that

$$F(x) = \frac{d}{dx} f(x) \quad \text{for all } x \in A.$$ 

Then, if $A \in [a, b],$

$$\int_a^b f(x) dx = F(b) - F(a).$$
Integration Trick (in 1D)

Fundamental Theorem of Calculus. Let $F$ be such that

$$F(x) = \frac{d}{dx} f(x) \quad \text{for all } x \in A.$$  

Then, if $A \in [a, b],$

$$\int_a^b f(x) \, dx = F(b) - F(a).$$
Integration Trick

$$CDF(x = \cdots) \propto \mathbf{F}(x, \theta)$$

$$\int_a^b p(x) \, dx = \int_a^b \frac{F'(x)}{F(b) - f(a)}$$
Tractable Evaluation

\[ \text{cdf}(x = \_\_\_\_) \propto F(x, \theta) \]

Input \rightarrow Output \ p(\_\_\_\_\_\_) \ 

\[
\int_a^b p(x) dx = \int_a^b \frac{F'(x)}{F(b) - f(a)}
\]
Fast Sampling

(Inverse Transform Method.) Two step process:

1. draw $z \sim \text{Unif}[0, 1]$

2. compute $x = \text{cdf}^{-1}(z)$
Lemma 3. (Gradient Theorem) Let $F : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuously differentiable function and $\varphi : [a, b] \rightarrow \mathbb{R}^n$ be a curve in $\mathbb{R}^n$, where $a, b \in \mathbb{R}$ and $\varphi(a), \varphi(b)$ are the endpoints of the curve. Then

$$\int_{\varphi[a,b]} \nabla F \cdot dr = F(\varphi(b)) - F(\varphi(a)). \quad (12)$$
Neural Inverse Transform Sampler

\[ \text{cdf}(x = \text{cat}) \propto F(x, \theta) \]

\[ F_\theta(x, y) \]

\[ \nabla_x | y = c \]

\[ y = c \]

\[ x = A \]

\[ x = B \]

\[ f_\theta(x, y) \]

\[ Z_{f_\theta | y = c}^{-1} \]

\[ \nu_\theta(x | y) \]
### Results

*Table 2.* Test log likelihood for UCI datasets and BSDS300, with error bars corresponding to two standard deviations. The table is split into two halves: the upper half denotes flow-based models, and the lower half denotes autoregressive continuous density models. NITS-CONV is only applied to BSDS300, as the convolutional architecture is only readily applicable to images.

| Model        | POWER    | GAS      | HEPMASS  | MINIBOONE | BSDS300  |
|--------------|----------|----------|----------|-----------|----------|
| MAF          | 0.30 ± 0.01 | 9.59 ± 0.02 | -17.39 ± 0.02 | -11.68 ± 0.44 | 156.36 ± 0.28 |
| TAN          | 0.48 ± 0.01 | 11.19 ± 0.02 | -15.12 ± 0.02 | -11.01 ± 0.48 | 157.03 ± 0.07 |
| NAF          | 0.62 ± 0.02 | 11.91 ± 0.13 | -15.09 ± 0.40 | **-8.86 ± 0.15** | 157.73 ± 0.04 |
| B-NAF        | 0.61 ± 0.01 | 12.06 ± 0.02 | -14.71 ± 0.02 | -8.95 ± 0.07 | 157.36 ± 0.03 |
| FFJORD       | 0.46 ± 0.01 | 8.59 ± 0.12 | -14.92 ± 0.08 | -10.43 ± 0.04 | 157.40 ± 0.19 |
| SOS          | 0.60 ± 0.01 | 11.99 ± 0.41 | -15.15 ± 0.10 | -8.90 ± 0.11 | 157.48 ± 0.41 |
| NSF          | **0.66 ± 0.01** | 13.09 ± 0.02 | -14.01 ± 0.03 | -9.22 ± 0.48 | 157.31 ± 0.28 |
| RealNVP      | 0.17 ± 0.01 | 8.33 ± 0.14 | -18.71 ± 0.02 | -13.84 ± 0.52 | 153.28 ± 1.78 |
| MADE MoG     | 0.40 ± 0.01 | 8.47 ± 0.02 | -15.15 ± 0.02 | -12.27 ± 0.47 | 153.71 ± 0.28 |
| NITS-MLP (Ours) | **0.66 ± 0.01** | **13.20 ± 0.01** | **-12.93 ± 0.02** | -10.85 ± 0.02 | 155.91 ± 0.21 |
| NITS-CONV (Ours) | - | - | - | - | **163.35 ± 0.22** |
Results

Table 2. Test log likelihood for UCI datasets and BSDS300, with error bars corresponding to two standard deviations. The table is split into two halves: the upper half denotes flow-based models, and the lower half denotes autoregressive continuous density models. NITS-CONV is only applied to BSDS300, as the convolutional architecture is only readily applicable to images.

| Model               | POWER  | GAS     | HEPMASS | MINIBOONE | BSDS300  |
|---------------------|--------|---------|---------|-----------|----------|
| MAF                 | 0.30 ± 0.01 | 9.59 ± 0.02 | -17.39 ± 0.02 | -11.68 ± 0.44 | 156.36 ± 0.28 |
| TAN                 | 0.48 ± 0.01 | 11.19 ± 0.02 | -15.12 ± 0.02 | -11.01 ± 0.48 | 157.03 ± 0.07 |
| NAF                 | 0.62 ± 0.02 | 11.91 ± 0.13 | -15.09 ± 0.40 | -8.86 ± 0.15 | 157.73 ± 0.04 |
| B-NAF               | 0.61 ± 0.01 | 12.06 ± 0.02 | -14.71 ± 0.02 | -8.95 ± 0.07 | 157.36 ± 0.03 |
| FFJORD              | 0.46 ± 0.01 | 8.59 ± 0.12 | -14.92 ± 0.08 | -10.43 ± 0.04 | 157.40 ± 0.19 |
| SOS                 | 0.60 ± 0.01 | 11.99 ± 0.41 | -15.15 ± 0.10 | -8.90 ± 0.11 | 157.48 ± 0.41 |
| NSF                 | 0.66 ± 0.01 | 13.09 ± 0.02 | -14.01 ± 0.03 | -9.22 ± 0.48 | 157.31 ± 0.28 |
| REALNVP             | 0.17 ± 0.01 | 8.33 ± 0.14 | -18.71 ± 0.02 | -13.84 ± 0.52 | 153.28 ± 1.78 |
| MADE MOG            | 0.40 ± 0.01 | 8.47 ± 0.02 | -15.15 ± 0.02 | -12.27 ± 0.47 | 153.71 ± 0.28 |
| NITS-MLP (OURS)     | 0.66 ± 0.01 | 13.20 ± 0.01 | -12.93 ± 0.02 | -10.85 ± 0.02 | 155.91 ± 0.21 |
| NITS-CONV (OURS)    | -       | -        | -       | -         | 163.35 ± 0.22 |
Results

Table 1. Negative log likelihood (in bits/dim) for CIFAR-10. The table is split into halves, with discretized density models above and continuous density models below. We obtain competitive results among both types of models.

| Model                          | CIFAR-10 |
|-------------------------------|----------|
| PIXEL CNN                     | 3.14     |
| GATED PIXEL CNN               | 3.03     |
| ROW PIXEL RNN                 | 3.00     |
| PIXEL CNN++                   | 2.92     |
| IMAGE TRANSFORMER             | 2.90     |
| PIXELSNAIL                    | 2.85     |
| DISCRETE NITS-CONV (Ours)     | 2.94     |
| REALNVP                       | 3.49     |
| GLOW                          | 3.35     |
| FLOW++                        | 3.08     |
| NITS-CONV (Ours)              | 2.97     |

Figure 2. Randomly generated images from DISCRETE NITS-CONV (top left) and NITS-CONV (top right). Compare with competing discretized and continuous density models, Pixel CNN (bottom left) and Flow++ (bottom right), respectively.
Results

Table 1. Negative log likelihood (in bits/dim) for CIFAR-10. The table is split into halves, with discretized density models above and continuous density models below. We obtain competitive results among both types of models.

| Model                        | CIFAR-10 |
|-------------------------------|----------|
| Pixel CNN                     | 3.14     |
| Gated Pixel CNN              | 3.03     |
| Row Pixel RNN                | 3.00     |
| Pixel CNN++                  | 2.92     |
| Image Transformer            | 2.90     |
| PixelSNAIL                   | 2.85     |
| Discrete NITS-CONV (Ours)    | 2.94     |
| REALNVP                      | 3.49     |
| Glow                         | 3.35     |
| Flow++                       | 3.08     |
| **NITS-CONV (Ours)**         | **2.97** |

Figure 2. Randomly generated images from DISCRETE NITS-CONV (top left) and NITS-CONV (top right). Compare with competing discretized and continuous density models, Pixel CNN (bottom left) and Flow++ (bottom right), respectively.
In Summary: Integration Trick

\[ \int_a^b p(x) \, dx = \int_a^b \frac{F'(x)}{F(b) - f(a)} \]
In Summary: Inverse Transform Method

1. draw $z \sim \text{Unif}[0, 1]$

2. compute $x = \text{cdf}^{-1}(z)$