Introduction to normalizing flows (1/2)

Thorsten Glüsenkamp, 28.3. 2022, Uppsala

Overview

Today:

1) **Introduction to normalizing flows**

2) **Linear/Affine flow (1-d / N-d)**
3) **Hands-on 1**

4) **Non-linear 1-d examples**

5) **Non-analytic inverses: Example: Mixture model**
6) **Hands-on 2**

Wednesday:

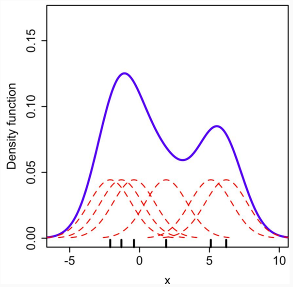
1) **Various N-d non-linear generalizations**

2) **Conditional PDFs: Amortization and the connection to deep learning**
3) **A NF tool: jammy_flows**
4) **Hands-on 1**

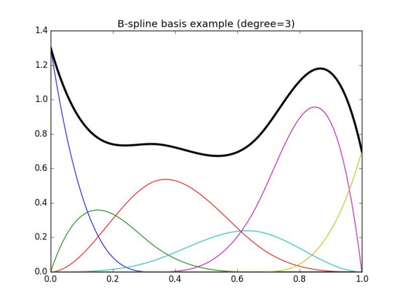
5) **Probabilistic deep learning and variational inference**
6) **Coverage / Systematics / Goodness-of-Fit**
7) **Hands-on 2**

What are normalizing flows? … specific probability density functions

Some well-known techniques to construct PDFs:



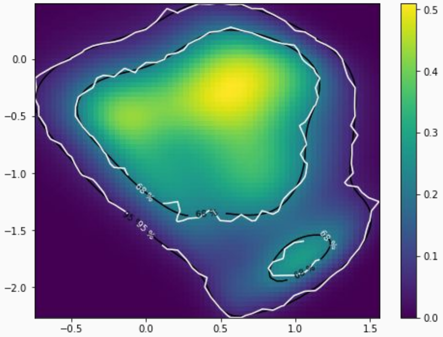
Mixture models /
Kernel Density Functions



B-Spline representation of a PDF

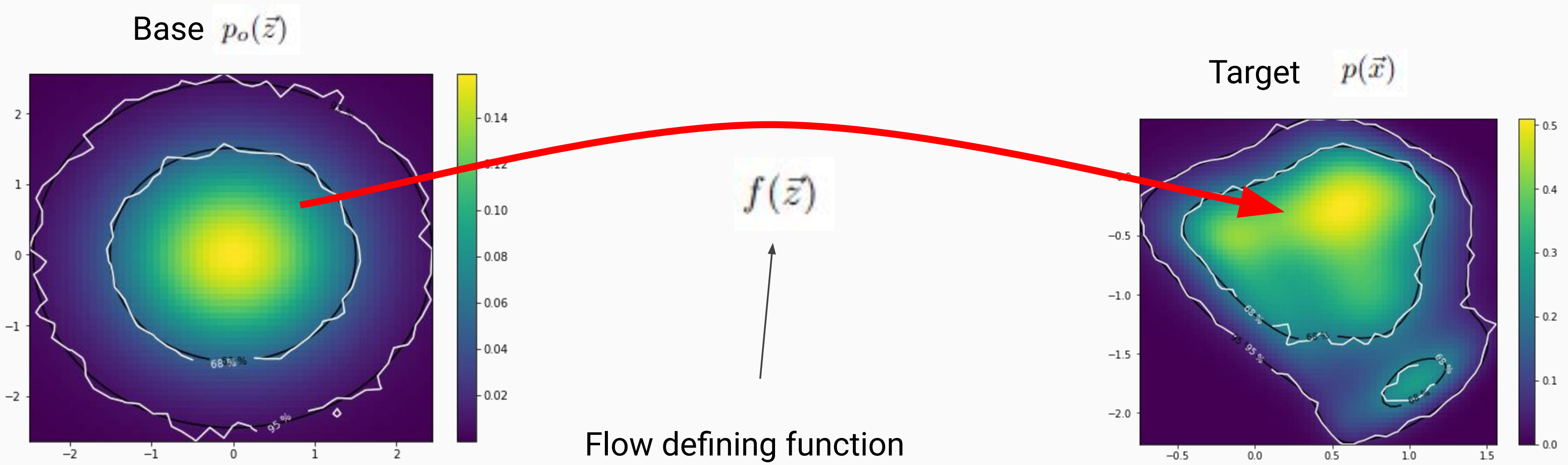
- (arbitrarily) Complex PDF shape
- Evaluate probability analytically
- Works in D > 1

Normalizing flows are also PDFs, but richer functionality:



- (arbitrarily) Complex PDF shape
- Evaluate probability analytically
- Works in D > 1
- Generate differentiable samples (-> differentiable expectation values)
- Coverage of the PDF
- Can be interpreted as generalizations of the Gaussian distribution
- Works on manifolds
- …

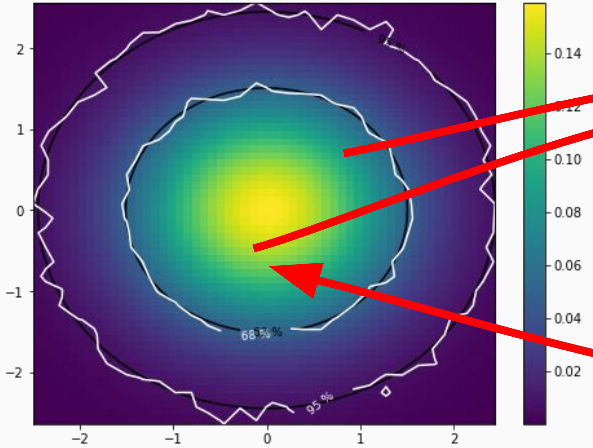
Definition



Definition

$$p(\vec{x}) = p_0(f^{-1}(\vec{x})) \cdot \left|\det \frac{\partial f^{-1}(\vec{x})}{\partial \vec{x}}\right|$$

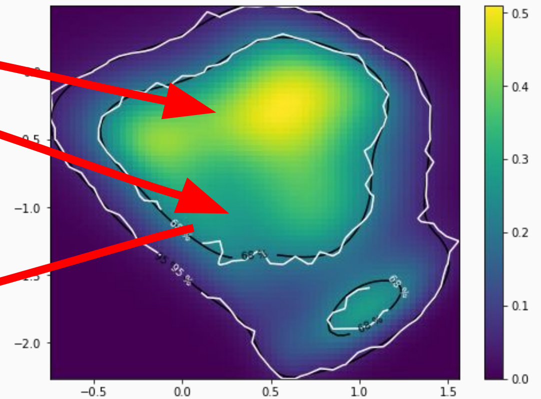
Base $p_o(\vec{z})$



$f(\vec{z})$

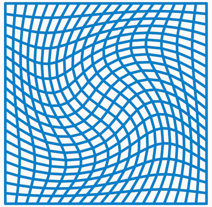
$f^{-1}(\vec{x})$

Target $p(\vec{x})$



1) The function has to be **bijective** for the inverse!
2) The function has to be **differentiable!**

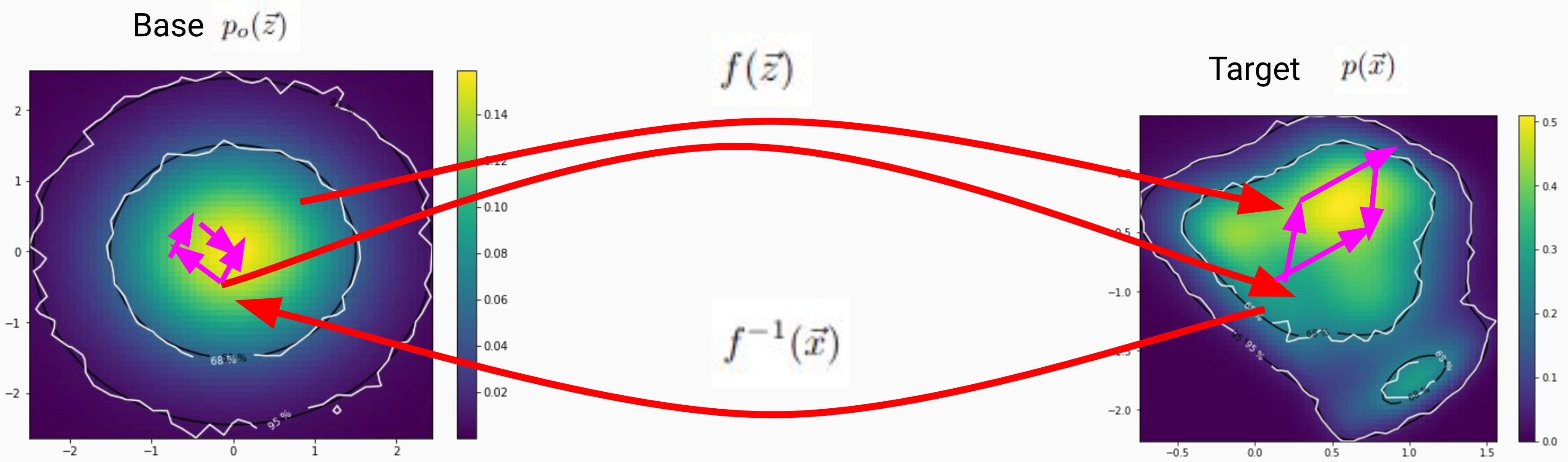
= Diffeomorphism



Definition

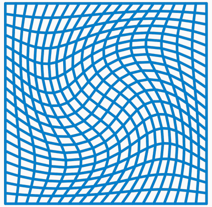
$$p(\vec{x}) = p_0(f^{-1}(\vec{x})) \cdot \left|\det J^{-1}(\vec{x})\right|$$

Log probability includes
Local “stretching”/”squeezing”
Of volume element



1) The function has to be **bijective** for the inverse!
2) The function has to be **differentiable!**

= Diffeomorphism



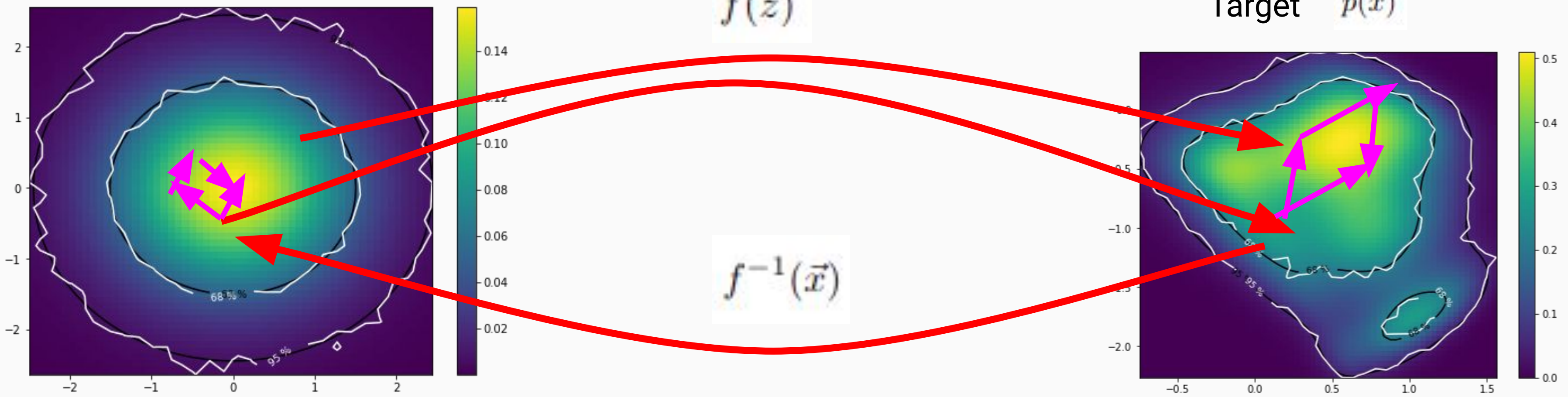
Definition

$$p(\vec{x}) = p_0(f^{-1}(\vec{x}) \cdot \frac{1}{|\det J(f^{-1}(\vec{x}))|}$$

Log probability includes
Local “stretching”/”squeezing”
Of volume element

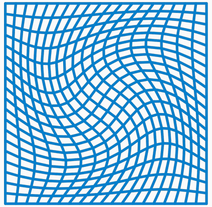
Base $p_o(\vec{z})$

Target $p(\vec{x})$



1) The function has to be **bijective** for the inverse!
2) The function has to be **differentiable!**

= Diffeomorphism



Definition

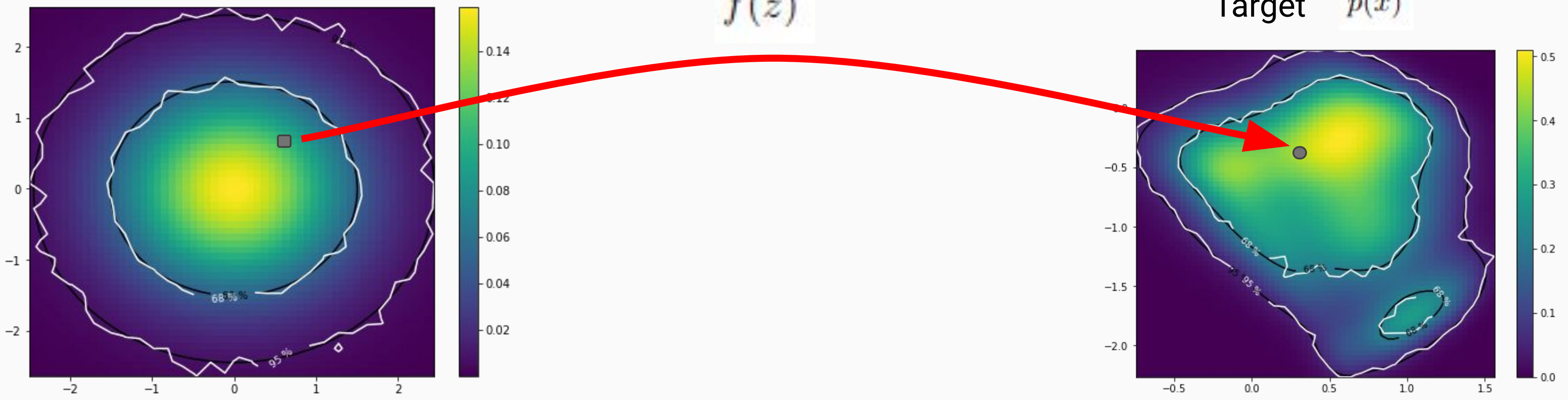
Sampling is also straight forward: $\vec{x} = f(\vec{z})$

Must be able to sample from base distribution!

Base $p_o(\vec{z})$

$f(\vec{z})$

Target $p(\vec{x})$



Definition

Sampling is also straight forward: $\vec{x_\theta} = f_\theta(\vec{z})$

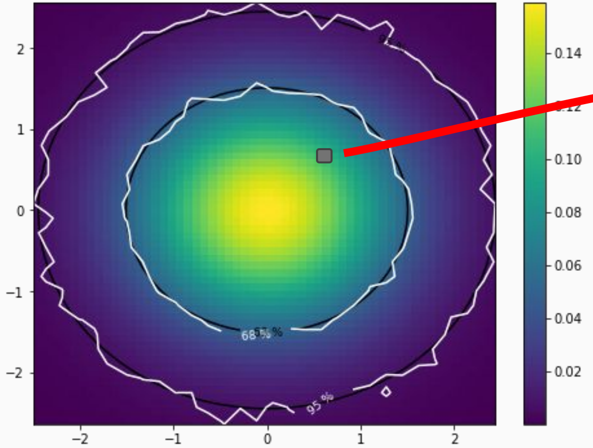
$$p_\theta(\vec{x}) = p_0(f_\theta^{-1}(\vec{x})) \cdot \left|\det \frac{\partial f_\theta^{-1}(\vec{x})}{\partial \vec{x}}\right|$$

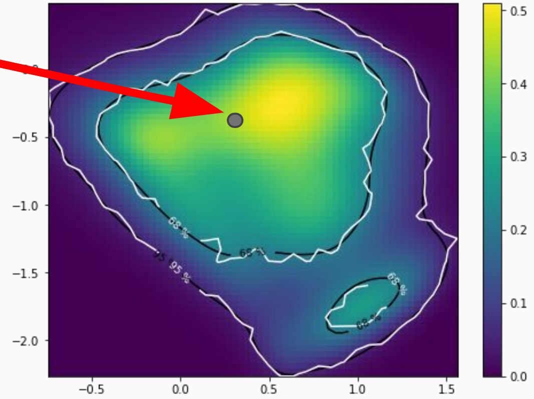
Must be able to sample from base distribution!

Base $p_o(\vec{z})$

$f_\theta(\vec{z})$

Target $p_\theta(\vec{x})$





Definition

Sampling is also straight forward:

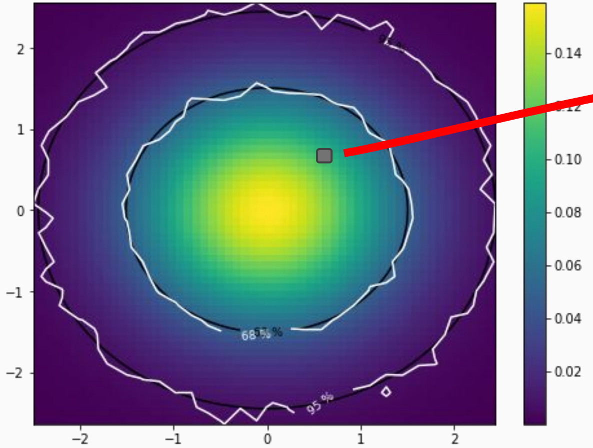
Must be able to sample from base distribution!

$$\vec{x}_\theta = f_\theta(\vec{z})$$

1) Differentiable prob.

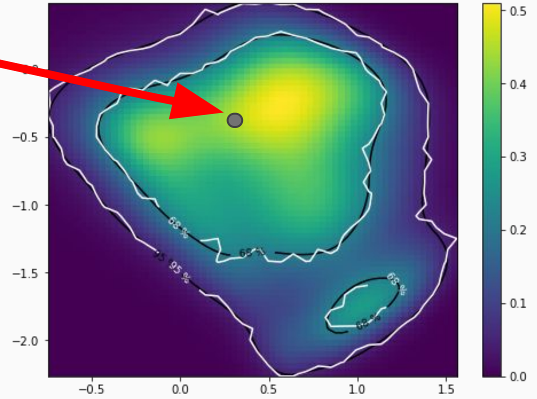
$$p_\theta(\vec{x}) = p_0(f_\theta^{-1}(\vec{x})) \cdot \left|\det \frac{\partial f_\theta^{-1}(\vec{x})}{\partial \vec{x}}\right|$$

Base $p_o(\vec{z})$



$f_\theta(\vec{z})$

Target $p_\theta(\vec{x})$



2) Differentiable samples!

Definition

Sampling is also straight forward:

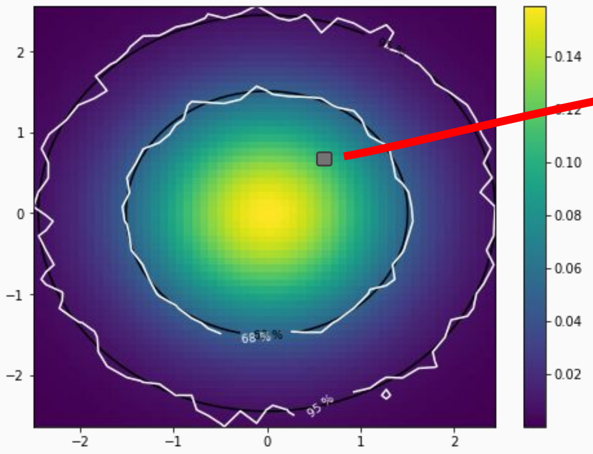
$$\vec{x}_\theta = f_\theta(\vec{z})$$

1) Differentiable prob.

$$p_\theta(\vec{x}) = p_0(f_\theta^{-1}(\vec{x})) \cdot \left|\det \frac{\partial f_\theta^{-1}(\vec{x})}{\partial \vec{x}}\right|$$

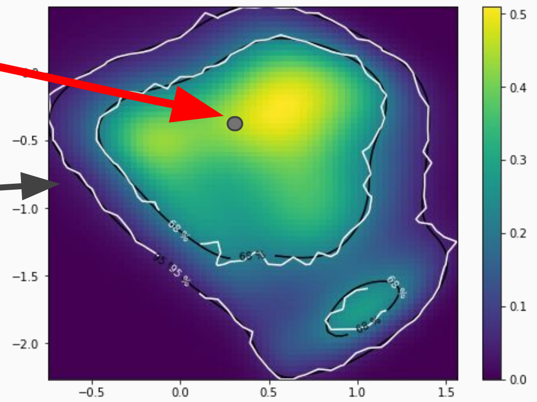
Must be able to sample from base distribution!

Base $p_o(\vec{z})$



$f_\theta(\vec{z})$

Target $p_\theta(\vec{x})$



Crosscheck
with contours

2) Differentiable samples!

Definition

Sampling is also straight forward:

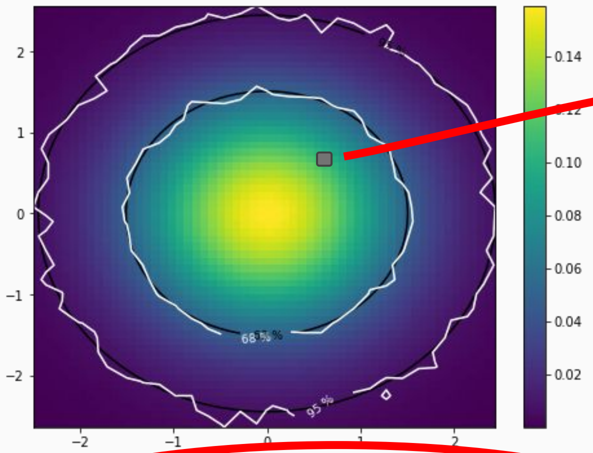
Must be able to sample from base distribution!

$$\vec{x}_\theta = f_\theta(\vec{z})$$

1) Differentiable prob.

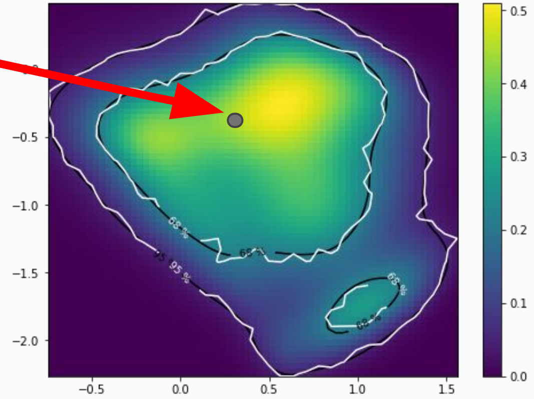
$$p_\theta(\vec{x}) = p_0(f_\theta^{-1}(\vec{x})) \cdot \left|\det \frac{\partial f_\theta^{-1}(\vec{x})}{\partial \vec{x}}\right|$$

Base $p_o(\vec{z})$



$f_\theta(\vec{z})$

Target $p_\theta(\vec{x})$



2) Differentiable samples!

Unique and important!

Definition

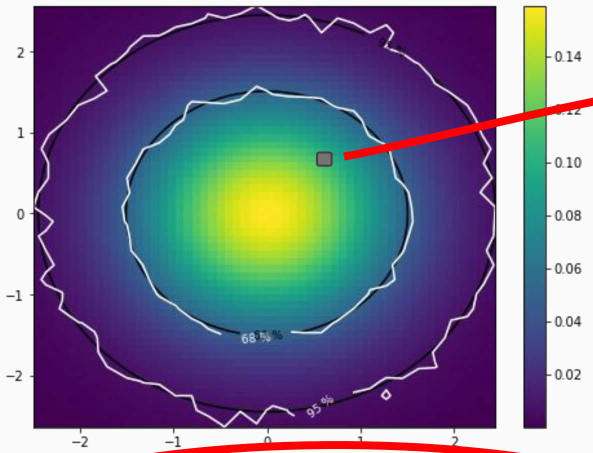
Sampling is also straight forward:

$$\vec{x}_\theta = f_\theta(\vec{z})$$

Must be able to sample from base distribution!

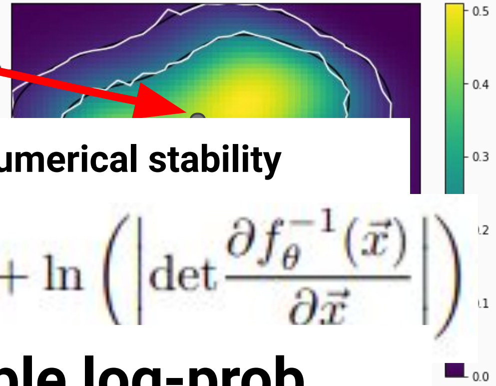
$$p_\theta(\vec{x}) = p_0(f_\theta^{-1}(\vec{x})) \cdot \left|\det \frac{\partial f_\theta^{-1}(\vec{x})}{\partial \vec{x}}\right|$$

Base $p_o(\vec{z})$



$f_\theta(\vec{z})$

Target $p_\theta(\vec{x})$



Typically work with logarithm for numerical stability

$$\ln(p_\theta(\vec{x})) = \ln(p_0(f_\theta^{-1}(\vec{x}))) + \ln\left(\left|\det \frac{\partial f_\theta^{-1}(\vec{x})}{\partial \vec{x}}\right|\right)$$

1) Differentiable log-prob.

2) Differentiable samples!

Unique and important!

Differentiable Samples -> Differentiable expectation values

$$I_\theta = \int p_\theta(x) F_\theta(x) dx \approx \frac{1}{N} \cdot \sum_i^N F_\theta(x_\theta)$$

Examples:

n-th moment of p

$$\int p_\theta(x) x^n dx \approx \frac{1}{N} \cdot \sum_i^N x_\theta^n$$

entropy

$$-\int p_\theta(x) \log\left(p_\theta(x)\right) dx \approx \frac{1}{N} \cdot \sum_i^N -\log p_\theta(x_\theta)$$

…. Many other integrals in information theory

Differentiable Samples -> Differentiable expectation values

$$I_\theta = \int p_\theta(x) F_\theta(x) dx \approx \frac{1}{N} \cdot \sum_i^N F_\theta(x_\theta)$$

Examples:

n-th moment of p

$$\int p_\theta(x) x^n dx \approx \frac{1}{N} \cdot \sum_i^N x_\theta^n$$

entropy

$$-\int p_\theta(x) \log\left(p_\theta(x)\right) dx \approx \frac{1}{N} \cdot \sum_i^N -\log p_\theta(x_\theta)$$

Wrong gradient!

$$\frac{1}{N} \cdot \sum_i^N -\log p_\theta(x)$$

!

…. Many other integrals in information theory

The simplest possible normalizing flow

Let us try to see the distribution of the simplest possible normalizing flow in 1 dimension!

$$x = f(\vec{z}) =$$

The simplest possible normalizing flow

Let us try to see the distribution of the simplest possible normalizing flow in 1 dimension!

$$x = f(\vec{z}) = a \cdot z + b$$

$$\theta = \{a, b\}$$

$$p_\theta(\vec{x}) = p_0(f_\theta^{-1}(\vec{x})) \cdot \left| \det \frac{\partial f_\theta^{-1}(\vec{x})}{\partial \vec{x}} \right|$$

The simplest possible normalizing flow (1-d)

Let us try to see the distribution of the simplest possible normalizing flow in 1 dimension!

$$x = f(\vec{z}) = a \cdot z + b$$

Use Standard normal base distribution (we will always use the standard normal for convenience)

Base distribution

$$p_0(z) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp(-0.5 \cdot z^2)$$

Change of variables formula

$$p_\theta(\vec{x}) = p_0(f_\theta^{-1}(\vec{x})) \cdot \left| \det \frac{\partial f_\theta^{-1}(\vec{x})}{\partial \vec{x}} \right|$$

Inv. Flow function

$$z = f^{-1}(\vec{x}) = \frac{x - b}{a}$$

$$p(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot a^2}} \cdot \exp\left(-0.5 \cdot \left(\frac{x - b}{a}\right)^2\right)$$

The simplest possible normalizing flow (1-d)

Let us try to see the distribution of the simplest possible normalizing flow in 1 dimension!

$$x = f(\vec{z}) = a \cdot z + b$$

Use Standard normal base distribution (we will always use the standard normal for convenience)

Base distribution

$$p_0(z) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp(-0.5 \cdot z^2)$$

Change of variables formula

$$p_\theta(\vec{x}) = p_0(f_\theta^{-1}(\vec{x})) \cdot \left| \det \frac{\partial f_\theta^{-1}(\vec{x})}{\partial \vec{x}} \right|$$

Inv. Flow function

$$z = f^{-1}(\vec{x}) = \frac{x - b}{a}$$

$$p(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot \exp\left(-0.5 \cdot \left(\frac{x - \mu}{\sigma}\right)^2\right)$$

The simplest possible normalizing flow (1-d)

Let us try to see the distribution of the simplest possible normalizing flow in 1 dimension!

$$x = f(\vec{z}) = a \cdot z + b$$

Use Standard normal base distribution (we will always use the standard normal for convenience)

“Linear flow” = “general Gaussian distribution” in 1d

Base distribution

$$p_0(z) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp(-0.5 \cdot z^2)$$

Change of variables formula

$$p_\theta(\vec{x}) = p_0(f_\theta^{-1}(\vec{x})) \cdot \left| \det \frac{\partial f_\theta^{-1}(\vec{x})}{\partial \vec{x}} \right|$$

Inv. Flow function

$$z = f^{-1}(\vec{x}) = \frac{x - b}{a}$$

$$p(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot a^2}} \cdot \exp\left(-0.5 \cdot \left(\frac{x-b}{a}\right)^2\right)$$

The simplest possible normalizing flow (1-d)

Let us try to see the distribution of the simplest possible normalizing flow in 1 dimension!

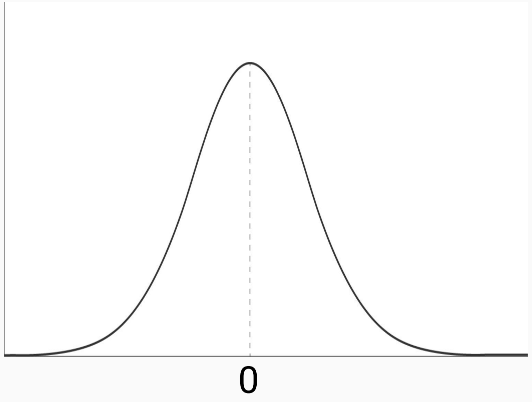
$$x = f(\vec{z}) = a \cdot z + b$$

Use Standard normal base distribution (we will always use the standard normal for convenience)

$\boxed{\text{“Linear flow” = “general Gaussian distribution” in 1d}}$

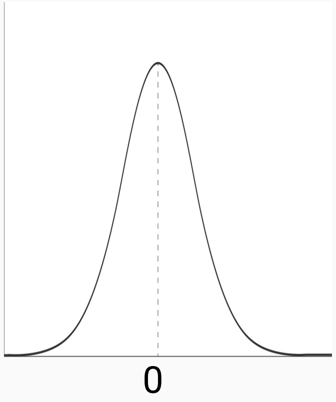
Technically, this is 2-step flow:

$$f(z) = f_2(f_1(z))$$



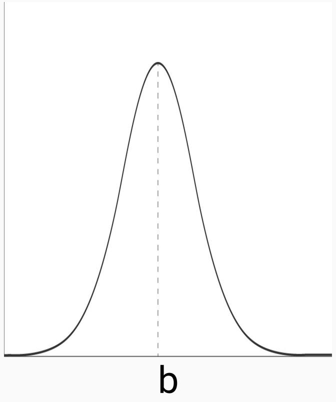
$$f_1(z) = a \cdot z$$

“scale”



$$f_2(z) = z + b$$

“shift”



The simplest possible normalizing flow (n-d)

Let us try to see the distribution of the simplest possible normalizing flow in n dimensions!

$$\vec{x} = f(\vec{z}) = \boldsymbol{L}^{-1} \cdot \vec{z} + \vec{b}$$

Use Standard normal base distribution
(we will always use the standard normal for convenience)

Base distribution

$$p_0(z) = \frac{1}{\sqrt{(2 \cdot \pi)^n}} \cdot \exp(-0.5 \cdot \vec{z}^T \cdot \vec{z})$$

Change of variables formula

$$p_\theta(\vec{x}) = p_0(f_\theta^{-1}(\vec{x})) \cdot \left| \det \frac{\partial f_\theta^{-1}(\vec{x})}{\partial \vec{x}} \right|$$

Inv. Flow function

$$\vec{z} = f^{-1}(\vec{x}) = \boldsymbol{L}(\vec{x} - \vec{b})$$

$$p(x) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot |\det(\boldsymbol{L})| \cdot \exp\left(-0.5 \cdot (\boldsymbol{L}(\vec{x} - \vec{b}))^T \cdot (\boldsymbol{L}(\vec{x} - \vec{b}))\right)$$

The simplest possible normalizing flow (n-d)

Let us try to see the distribution of the simplest possible normalizing flow in n dimensions!

$$\vec{x} = f(\vec{z}) = \boldsymbol{L}^{-1} \cdot \vec{z} + \vec{b}$$

Use Standard normal base distribution (we will always use the standard normal for convenience)

Base distribution

$$p_0(z) = \frac{1}{\sqrt{(2 \cdot \pi)^n}} \cdot \exp(-0.5 \cdot \vec{z}^T \cdot \vec{z})$$

Change of variables formula

$$p_\theta(\vec{x}) = p_0(f_\theta^{-1}(\vec{x})) \cdot \left| \det \frac{\partial f_\theta^{-1}(\vec{x})}{\partial \vec{x}} \right|$$

Inv. Flow function

$$\vec{z} = f^{-1}(\vec{x}) = \boldsymbol{L}(\vec{x} - \vec{b})$$

$$p(x) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot |\det(\boldsymbol{L})| \cdot \exp\left(-0.5 \cdot (\vec{x} - \vec{b})^T \cdot \boldsymbol{L}^T \cdot \boldsymbol{L} \cdot (\vec{x} - \vec{b})\right)$$

The simplest possible normalizing flow (n-d)

Let us try to see the distribution of the simplest possible normalizing flow in n dimensions!

$$\vec{x} = f(\vec{z}) = \boldsymbol{L}^{-1} \cdot \vec{z} + \vec{b}$$

Use Standard normal base distribution (we will always use the standard normal for convenience)

$$\boldsymbol{C}^{-1} = \boldsymbol{L}^T \cdot \boldsymbol{L}$$

$$L = \begin{bmatrix} \ell_{1,1} & & & & 0 \\ \ell_{2,1} & \ell_{2,2} & & & \\ \ell_{3,1} & \ell_{3,2} & \ddots & & \\ \vdots & \vdots & \ddots & \ddots & \\ \ell_{n,1} & \ell_{n,2} & \dots & \ell_{n,n-1} & \ell_{n,n} \end{bmatrix}$$

Base distribution

$$p_0(z) = \frac{1}{\sqrt{(2 \cdot \pi)^n}} \cdot \exp(-0.5 \cdot \vec{z}^T \cdot \vec{z})$$

Change of variables formula

$$p_\theta(\vec{x}) = p_0(f_\theta^{-1}(\vec{x})) \cdot \left|\det \frac{\partial f_\theta^{-1}(\vec{x})}{\partial \vec{x}}\right|$$

Inv. Flow function

$$\vec{z} = f^{-1}(\vec{x}) = \boldsymbol{L}(\vec{x} - \vec{b})$$

$$p(x) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \sqrt{\det(\boldsymbol{C}^{-1})} \cdot \exp\left(-0.5 \cdot (\vec{x} - \vec{b})^T \cdot \boldsymbol{C}^{-1} \cdot (\vec{x} - \vec{b})\right)$$

The simplest possible normalizing flow (n-d)

Let us try to see the distribution of the simplest possible normalizing flow in n dimensions!

$$\vec{x} = f(\vec{z}) = \boldsymbol{L}^{-1} \cdot \vec{z} + \vec{b}$$

Use Standard normal base distribution (we will always use the standard normal for convenience)

$$\boldsymbol{C}^{-1} = \boldsymbol{L}^T \cdot \boldsymbol{L}$$

$\boxed{\text{“Affine flow” = “general Gaussian distribution” in n-d}}$

Base distribution

$$p_0(z) = \frac{1}{\sqrt{(2 \cdot \pi)^n}} \cdot \exp(-0.5 \cdot \vec{z}^T \cdot \vec{z})$$

Change of variables formula

$$p_\theta(\vec{x}) = p_0(f_\theta^{-1}(\vec{x})) \cdot \left| \det \frac{\partial f_\theta^{-1}(\vec{x})}{\partial \vec{x}} \right|$$

Inv. Flow function

$$\vec{z} = f^{-1}(\vec{x}) = \boldsymbol{L}(\vec{x} - \vec{b})$$

$$p(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \det(\boldsymbol{C})}} \exp\left(-0.5 \cdot (\vec{x} - \vec{b})^T \cdot \boldsymbol{C}^{-1} \cdot (\vec{x} - \vec{b})\right)$$

The simplest possible normalizing flow (n-d)

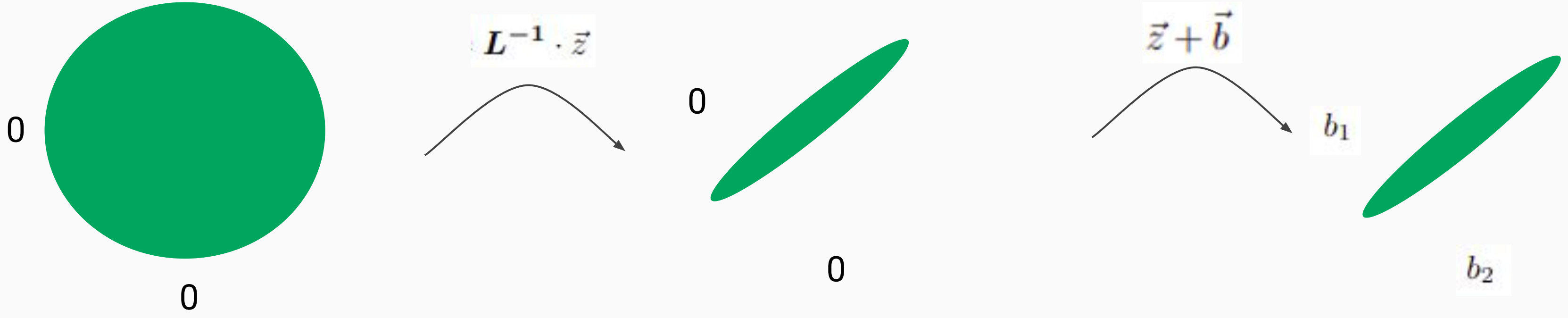
Let us try to see the distribution of the simplest possible normalizing flow in n dimensions!

$$\vec{x} = f(\vec{z}) = \boldsymbol{L}^{-1} \cdot \vec{z} + \vec{b}$$

Use Standard normal base distribution (we will always use the standard normal for convenience)

$$\boldsymbol{C}^{-1} = \boldsymbol{L}^{T} \cdot \boldsymbol{L}$$

“Affine flow” = “general Gaussian distribution” in n-d



Hands on 1

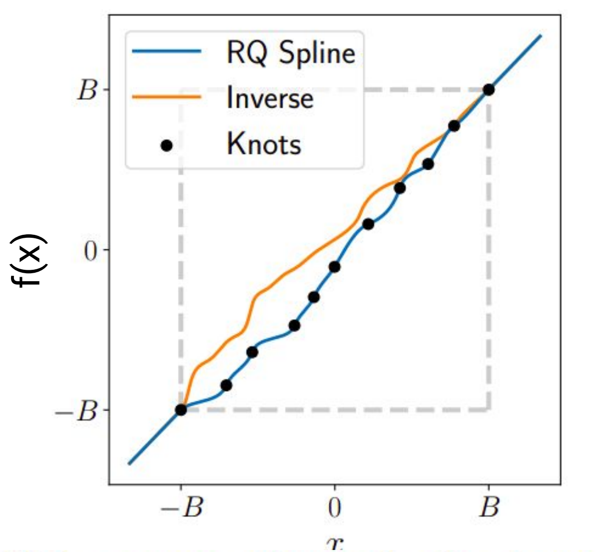
Hands-on 1:

https://colab.research.google.com/drive/1iwCVXF9jeJ8JAqPg_7PV07IPgLk8jEb3?usp=sharing

How can we ensure bijectivity in 1-d for nonlinear functions?

One possibility:
Neural Spline Flows (Durkan et al. 2019)

$$p(\vec{x}) = p_0(f^{-1}(\vec{x})) \cdot \left|\det \frac{\partial f^{-1}(\vec{x})}{\partial \vec{x}}\right|$$



Flow function: $f^{-1}(\vec{x})$

$$\frac{\alpha^{(k)}(\xi)}{\beta^{(k)}(\xi)} = y^{(k)} + \frac{(y^{(k+1)} - y^{(k)})\left[s^{(k)}\xi^2 + \delta^{(k)}\xi(1-\xi)\right]}{s^{(k)} + \left[\delta^{(k+1)} + \delta^{(k)} - 2s^{(k)}\right]\xi(1-\xi)}$$

Derivative: $\frac{\partial f^{-1}(\vec{x})}{\partial \vec{x}}$

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\alpha^{(k)}(\xi)}{\beta^{(k)}(\xi)}\right] = \frac{\left(s^{(k)}\right)^2\left[\delta^{(k+1)}\xi^2 + 2s^{(k)}\xi(1-\xi) + \delta^{(k)}(1-\xi)^2\right]}{\left[s^{(k)} + \left[\delta^{(k+1)} + \delta^{(k)} - 2s^{(k)}\right]\xi(1-\xi)\right]^2}$$

which passes through the knots, with the given derivatives at the knots. Defining $s_k = (y^{k+1} - y^k)/(x^{k+1} - x^k)$ and $\xi(x) = (x - x^k)/(x^{k+1} - x^k)$, the expression for the rational-

How can we ensure bijectivity in 1-d for nonlinear functions?

Another possibility: Use CDFs

$$f(y) = \mathrm{CDF}_2^{-1}(\mathrm{CDF}_1(y))$$

$$f^{-1}(x) = \mathrm{CDF}_1^{-1}(\mathrm{CDF}_2(x))$$

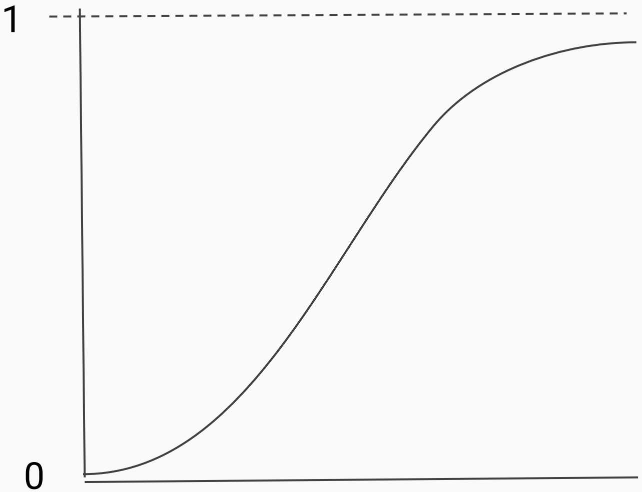
How can we ensure bijectivity in 1-d for nonlinear functions?

Another possibility: Use CDFs

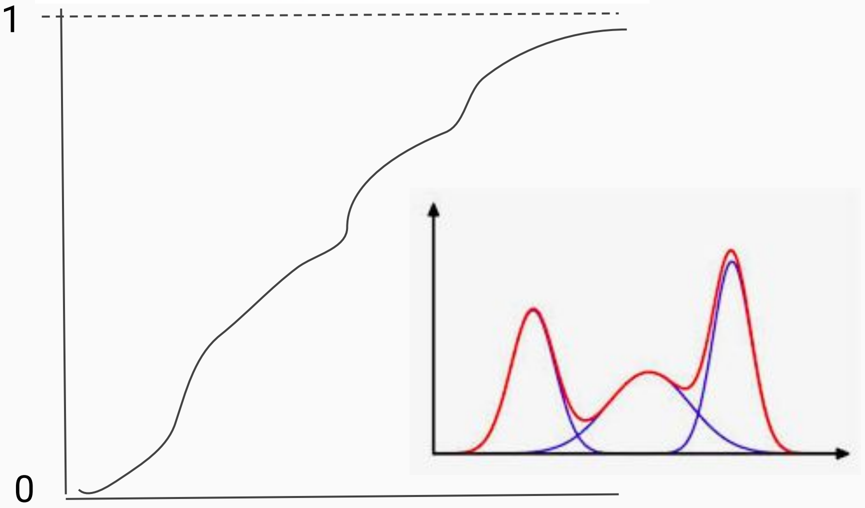
$$f(y) = \mathrm{CDF}_2^{-1}(\mathrm{CDF}_1(y))$$

$$f^{-1}(x) = \mathrm{CDF}_1^{-1}(\mathrm{CDF}_2(x))$$

$$\mathrm{CDF}_1(x) = \mathrm{CDF}_{\mathcal{N}(0;1)}(x)$$



$$\mathrm{CDF}_2(x) = \sum_{i=1}^{N} w_i \cdot \mathrm{CDF}_{\mathcal{N}(\mu_i;\sigma_i)}(x)$$



How can we ensure bijectivity in 1-d for nonlinear functions?

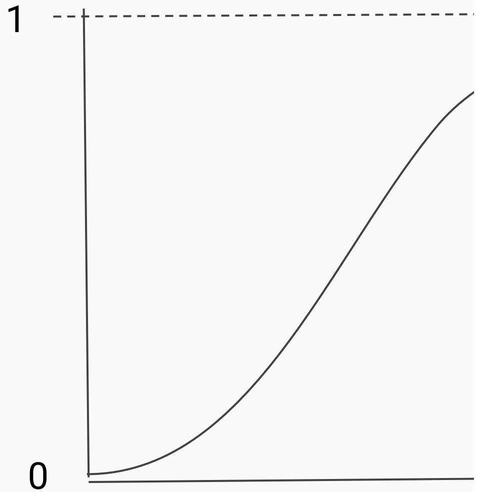
Another possibility: Use CDFs

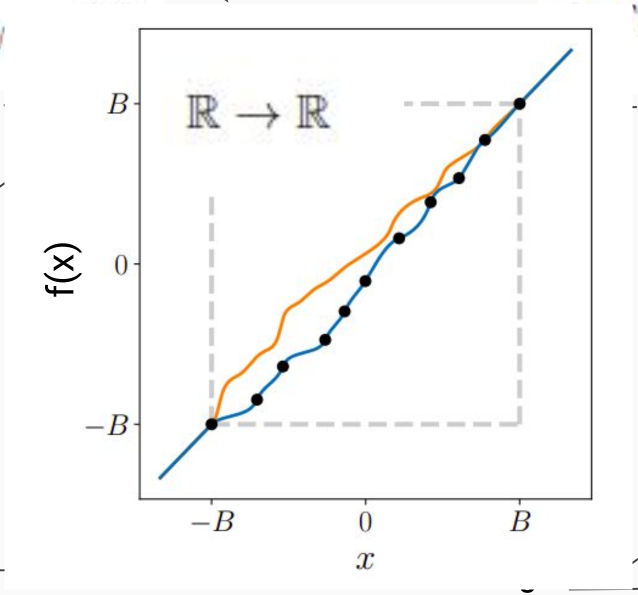
$[0,1] \to \mathbb{R}$ $\quad$ $\mathbb{R} \to [0,1]$

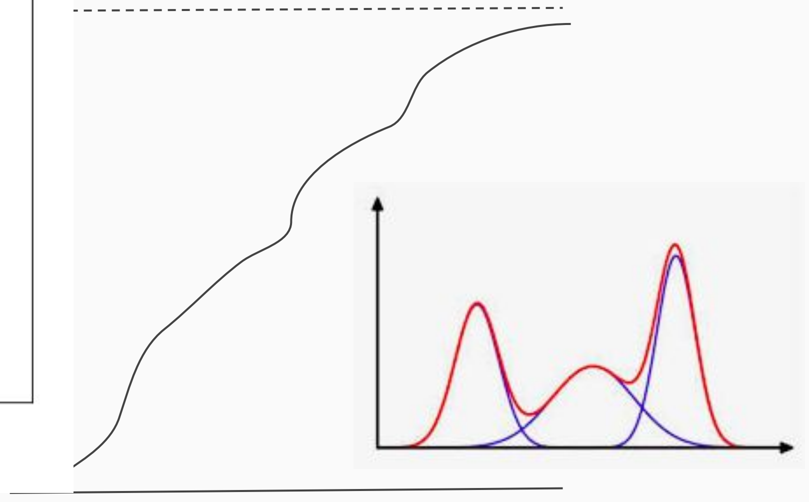
$$f(y) = \mathrm{CDF}_2^{-1}(\mathrm{CDF}_1(y)) \qquad f^{-1}(x) = \mathrm{CDF}_1^{-1}(\mathrm{CDF}_2(x))$$

$$\mathrm{CDF}_1(x) = \mathrm{CDF}_{\mathcal{N}}$$

$$_2(x) = \sum_{i=1}^{N} w_i \cdot \mathrm{CDF}_{\mathcal{N}(\mu_i;\sigma_i)}(x)$$







How can we ensure bijectivity in 1-d for nonlinear functions?

Another possibility: Use CDFs

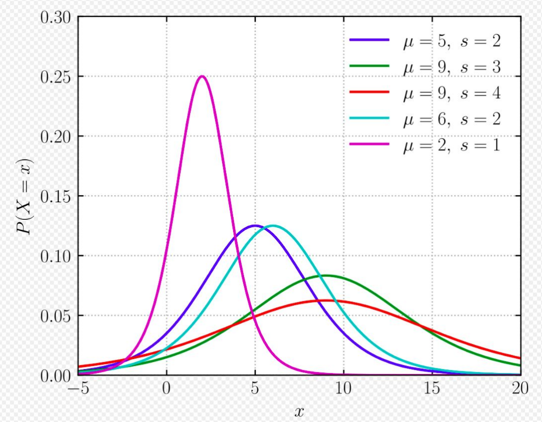
$[0,1] \to \mathbb{R}$ $\quad$ $\mathbb{R} \to [0,1]$

$$f(y) = \mathrm{CDF}_2^{-1}(\mathrm{CDF}_1(y)) \qquad f^{-1}(x) = \mathrm{CDF}_1^{-1}(\mathrm{CDF}_2(x))$$

Example: Use weighted sum of "**logistic distributions**" instead of **Gaussians** for simplicity

$$\mathrm{pdf}_{\mathrm{logistic}} = \frac{1}{s} \cdot \frac{e^{(x-\mu)/s}}{(1+e^{(x-\mu)/s})^2}$$

$$\mathrm{cdf}_{\mathrm{logistic}} = \frac{1}{1+e^{-(x-\mu)/s}}$$



$$\mathrm{cdf}_{\mathrm{logistic}}^{-1}[\mu = 0, s = 1.0](y) = \log(y) - \log(1.0 - y)$$

$$f^{-1}(x) = \mathrm{cdf}_1^{-1}(\mathrm{cdf}_2(x)) = \log(\sum_i w_i \mathrm{cdf}_{\mathrm{logistic},\mu_i,s_i}) - \log(1.0 - \sum_i w_i \mathrm{cdf}_{\mathrm{logistic},\mu_i,s_i})$$

How can we ensure bijectivity in 1-d for nonlinear functions?

Another possibility: Use CDFs

$[0,1] \to \mathbb{R}$ $\mathbb{R} \to [0,1]$

$$f(y) = \mathrm{CDF}_2^{-1}(\mathrm{CDF}_1(y))$$

Not Analytic

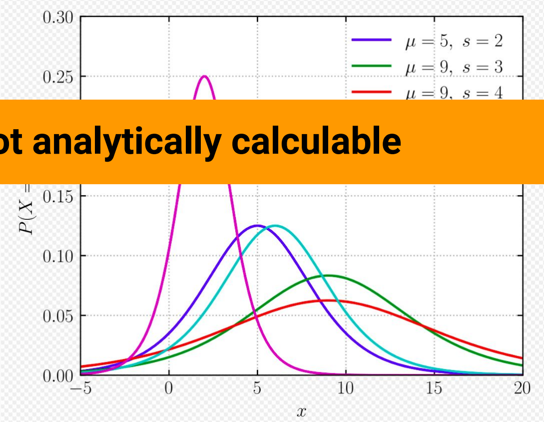
$$f^{-1}(x) = \mathrm{CDF}_1^{-1}(\mathrm{CDF}_2(x))$$

Analytic

Example: Use weighted sum of “**logistic distributions**” instead of **Gaussians** for simplicity

$$\mathrm{pdf}_{\mathrm{logistic}} = \frac{1}{s} \cdot \frac{e^{(x-\mu)/s}}{(1+e^{(x-\mu)/s})^2}$$

$$\mathrm{cdf}_{\mathrm{logistic}} = \frac{1}{1+e^{-(x-\mu)/s}}$$



f ist bijective but f itself is not analytically calculable

$$\mathrm{cdf}^{-1}_{\mathrm{logistic}}[\mu = 0, s = 1.0](y) = \log(y) - \log(1.0 - y)$$

$$f^{-1}(x) = \mathrm{cdf}_1^{-1}(\mathrm{cdf}_2(x)) = \log(\sum_i w_i \mathrm{cdf}_{\mathrm{logistic},\mu_i,s_i}) - \log(1.0 - \sum_i w_i \mathrm{cdf}_{\mathrm{logistic},\mu_i,s_i})$$

How to invert bijective function numerically?

Transform to root finding problem:

Want to find inverse of g(x), but only have access to g(x):

Want: $x = g^{-1}(y)$

solve : $F(x) = g(x) - y = 0$

y is given, vary x until solution found

How to invert bijective function numerically?

Transform to root finding problem:

Want to find inverse of g(x), but only have access to g(x):

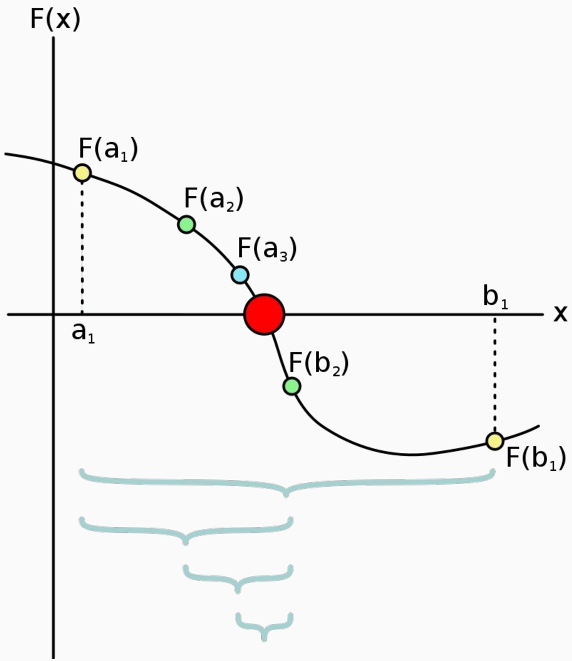
Want: $x = g^{-1}(y)$

solve : $F(x) = g(x) - y = 0$

y is given, vary x until solution found

One option: bisection

Problem: not differentiable



How to invert bijective function numerically?

Transform to root finding problem:

Want to find inverse of g(x), but only have access to g(x):

Want: $x = g^{-1}(y)$

solve : $F(x) = g(x) - y = 0$

y is given, vary x until solution found

Second option: Newton iterations

Actually differentiable

$$x_{n+1} = x_n - \frac{F(x)}{F'(x)} = x_n - \frac{g(x) - y}{g'(x)}$$

How to invert bijective function numerically?

Transform to root finding problem:

Want to find inverse of g(x), but only have access to g(x):

Want: $x = g^{-1}(y)$

solve : $F(x) = g(x) - y = 0$

y is given, vary x until solution found

Second option: Newton iterations

Actually differentiable

$$x_{n+1} = x_n - \frac{F(x)}{F'(x)} = x_n - \frac{g(x) - y}{g'(x)}$$

$$f^{-1}(x) = \mathrm{cdf}_1^{-1}(\mathrm{cdf}_2(x)) = \log(\sum_i w_i \mathrm{cdf}_{\mathrm{logistic},\mu_i,s_i}) - \log(1.0 - \sum_i w_i \mathrm{cdf}_{\mathrm{logistic},\mu_i,s_i})$$

$$\boxed{\text{analytic} \quad f_\theta^{-1}(x), \frac{\partial f_\theta^{-1}(x)}{\partial x}}$$

$$\theta = (s_1, w_1, \mu_1, \ldots, s_N, w_N, \mu_N)$$

How to invert bijective function numerically?

Transform to root finding problem:

Want to find inverse of g(x), but only have access to g(x):

Want: $x = g^{-1}(y)$

solve : $F(x) = g(x) - y = 0$

y is given, vary x until solution found

Second option: Newton iterations

Actually differentiable

$$x_{n+1} = x_n - \frac{F(x)}{F'(x)} = x_n - \frac{g(x) - y}{g'(x)}$$

$$f^{-1}(x) = \mathrm{cdf}_1^{-1}(\mathrm{cdf}_2(x)) = \log(\sum_i w_i \mathrm{cdf}_{\mathrm{logistic},\mu_i,s_i}) - \log(1.0 - \sum_i w_i \mathrm{cdf}_{\mathrm{logistic},\mu_i,s_i})$$

analytic $f_\theta^{-1}(x), \frac{\partial f_\theta^{-1}(x)}{\partial x}$

$$x_{n+1} = x_n - \frac{F(x)}{F'(x)} = x_n - \frac{f^{-1}(x) - y}{f^{-1\prime}(x)}$$

$\theta = (s_1, w_1, \mu_1, \ldots, s_N, w_N, \mu_N)$

numerical $f_\theta(y), \frac{\partial f_\theta(y)}{\partial y}$

$$\rightarrow x_* = f(y) = x_{*,\theta}$$

$$\frac{\partial f(y)}{\partial y}[\theta] = \frac{1.0}{\frac{\partial f_\theta^{-1}(x_{*,\theta})}{\partial x}[\theta]}$$

differentiable

Hands on 2

https://colab.research.google.com/drive/1XPOrfbcbe22octvLLYP_9XQSfEMOFnCg?usp=sharing