Normalizing Flows

MUSTAFA HAJIJ

Generative Models

A generative model is a mathematical framework that learns the underlying distribution of a given dataset and then generates new samples that are similar to the original data.

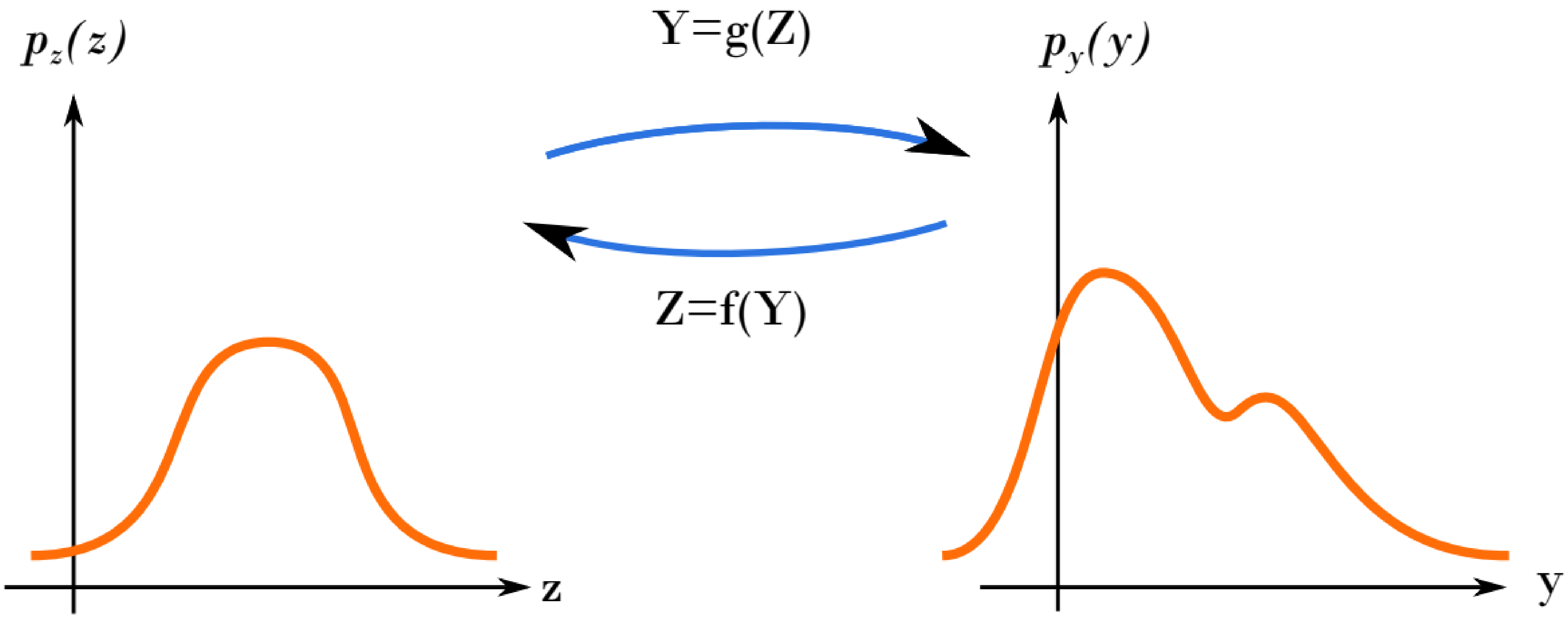
More formally, let's consider a dataset of samples $\mathrm{D} = \{x_1, x_2, \ldots, x_n\}$, where each x_i is a data point. The goal of a generative model is to estimate the underlying probability distribution $p_{data}(x)$ from which the samples in X are drawn.

Generative Models

Explicitly, we are usually given a dataset $D = \{xi \in R^d\}^n{}_{i=1}$, where each x_i is i.i.d sampled from an unknown probability distribution $p_{data} : R^d \rightarrow R$. Within this setting, we are interested in estimating the distribution p_{data} by learning a parameterized density function $p_\theta : R^d \rightarrow R$, where θ is the parameter of p_θ , such that $p_\theta \approx p_{data}$.

The question is how to model the distribution p_θ ?

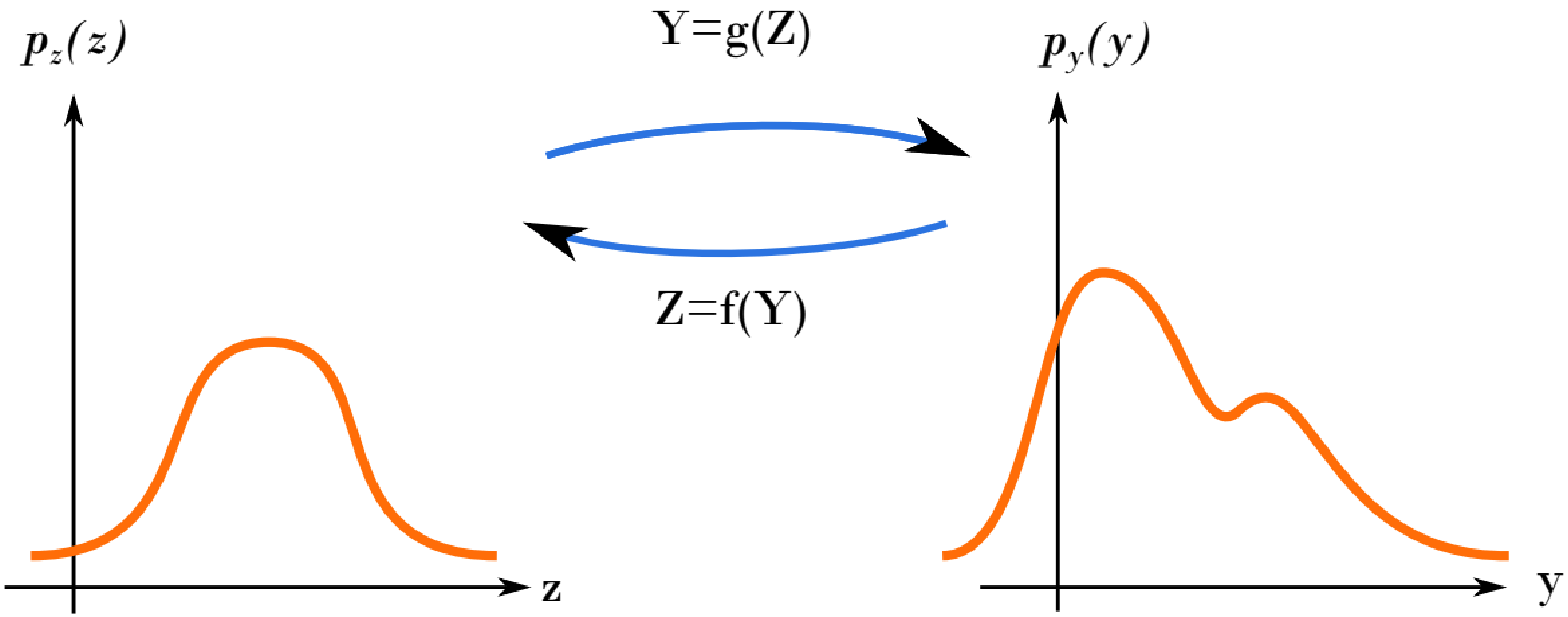
NF main idea



This is the density a known probability distribution, the Gaussian.

This is the density of the probability distribution that we want to estimate $p_y(y) = p_{data}(y)$

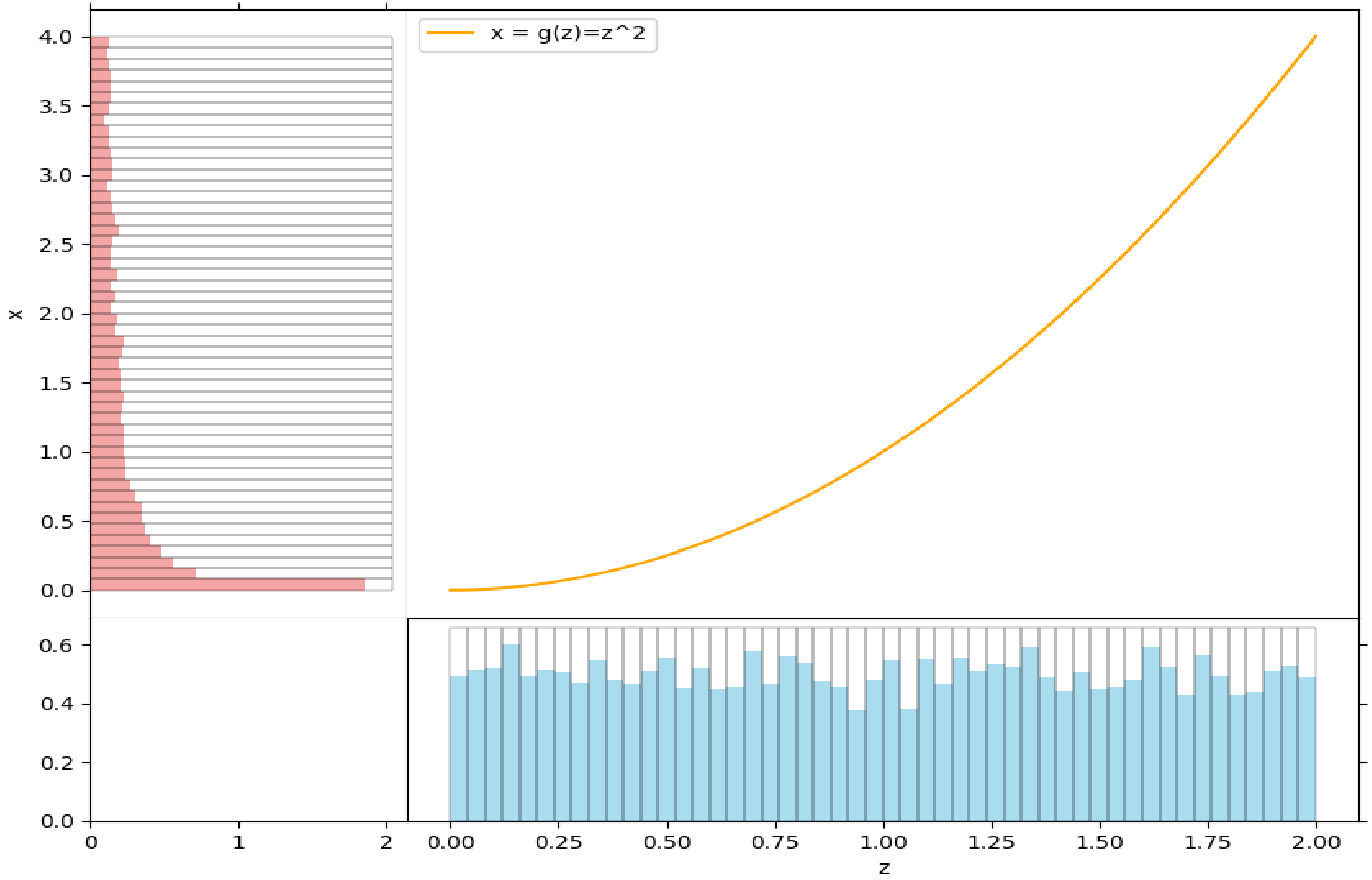
NF main idea



A NF generative model constructs an invertible function $g(z)$ such that g sends the probability distribution p_z to the probability distribution p_y. The direction of g is usually called the **generation direction** whereas the direction of f is usually called the **flow direction.**

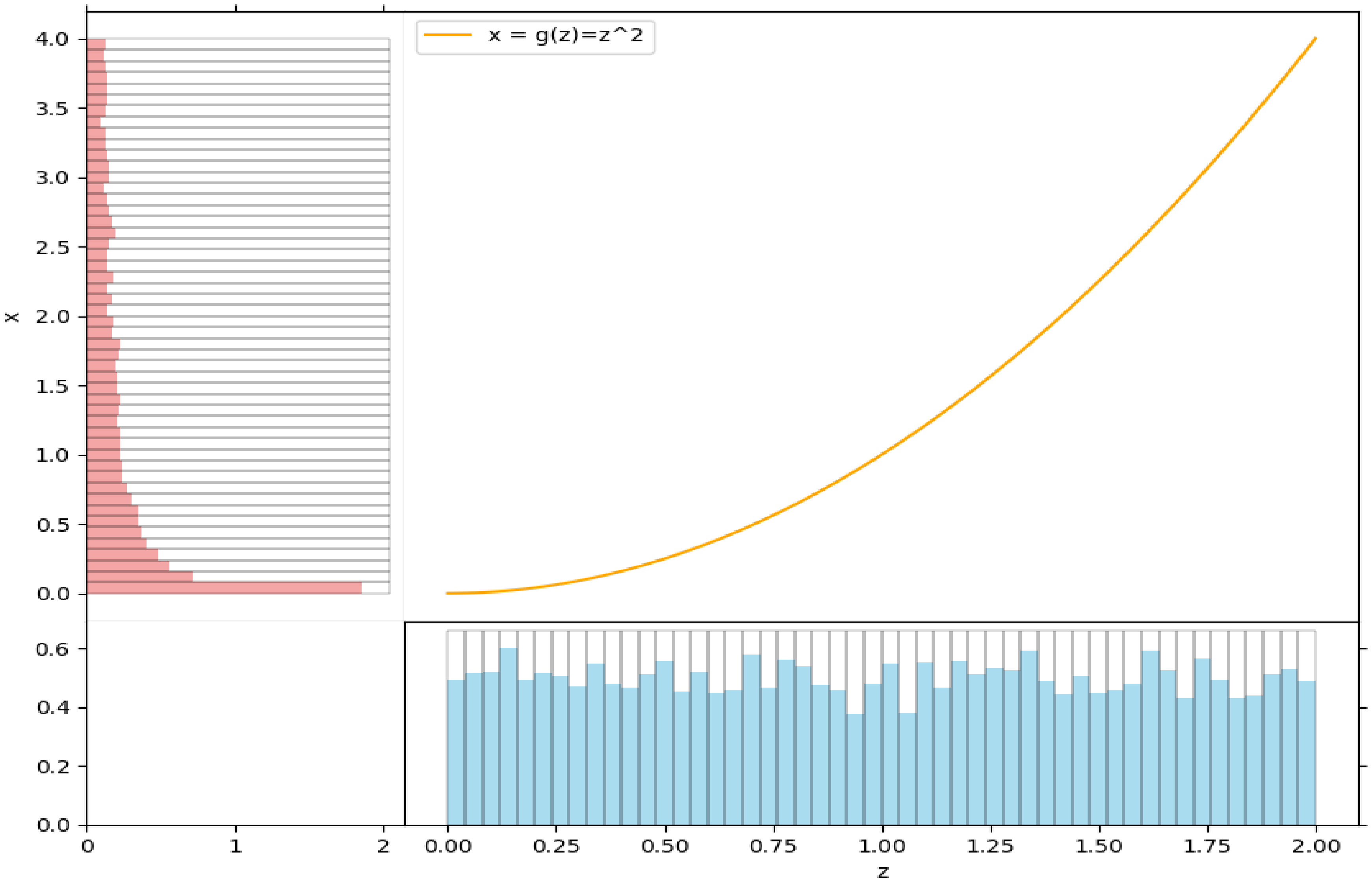
NF main idea

Let z be sampled from uniform (0,2), sample from this distribution and use the function g(z) =z^2 to obtain x :



NF main idea

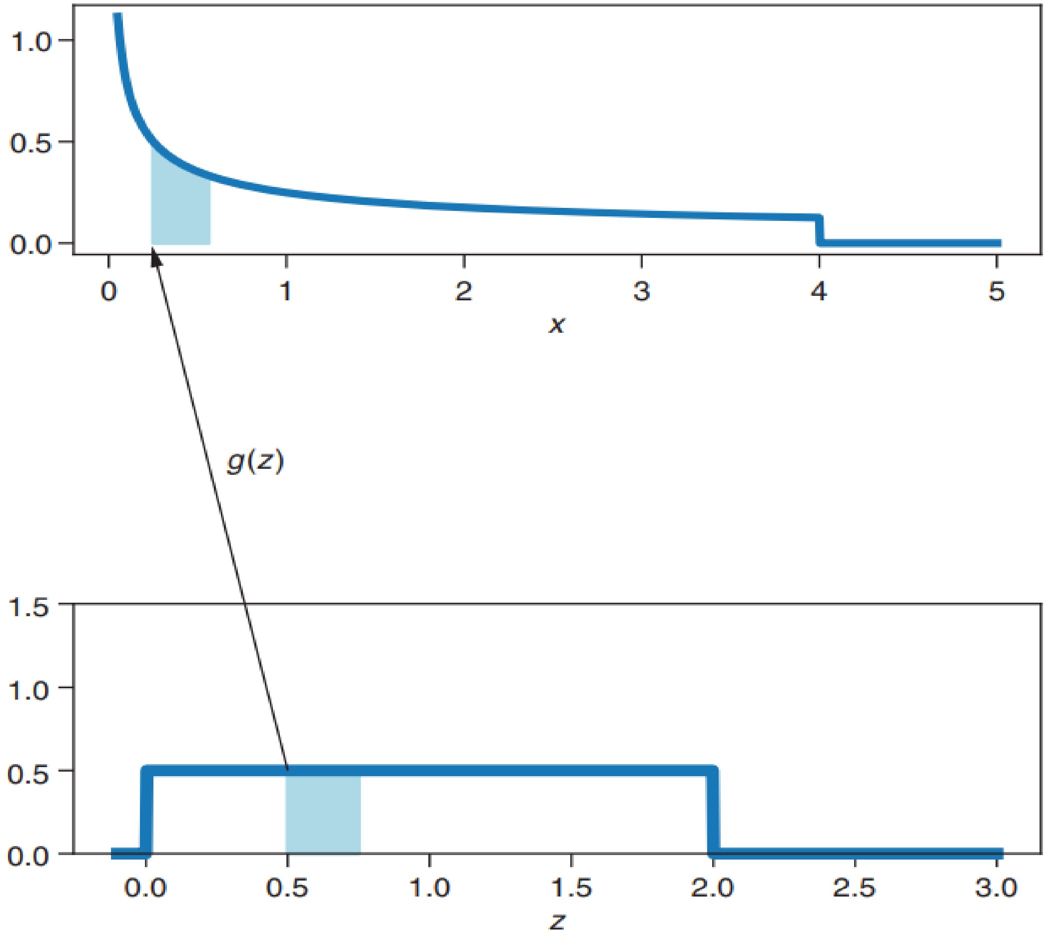
Question : is x is valid distribution ?



NF main idea

Question : is x is valid distribution ?

The area $p_z(z)|dz| = p_x(x)|dx|$ (shaded in the figure) needs to be preserved



Or :

$$p_x(x) = p_z(z) \cdot |\frac{dz}{dx}|$$

$$p_x(x) = p_z(z) \cdot |\frac{dx}{dz}|^{-1}$$

[Volume preservation in normalizing flows - YouTube]

NF main idea

Question : is x is valid distribution ?

The area $p_z(z)|dz| \ = \ p_x(x)|dx|$ (shaded in the figure) needs to be preserved

$$p_x(x) = p_z(z) \cdot |\frac{dz}{dx}|$$

$$p_x(x) = p_z(z) \cdot |\frac{dx}{dz}|^{-1}$$

NF main idea

Question : is x is valid distribution ?

The area $p_z(z)|dz| = p_x(x)|dx|$ (shaded in the figure) needs to be preserved

$$p_x(x) = p_z(z) \cdot |\frac{dz}{dx}|$$

$$p_x(x) = p_z(z) \cdot |\frac{dx}{dz}|^{-1}$$

$$p_x(x) = p_z(z) \cdot |\frac{dg(z)}{dz}|^{-1} \text{ where } x = g(z).$$

$$p_x(x) = p_z(z) \cdot |g'(z)|^{-1}$$

NF main idea

Question : is x is valid distribution ?

The area $p_z(z)|dz| \; = \; p_x(x)|dx|$ (shaded in the figure) needs to be preserved

$$p_x(x) = p_z(z) \cdot |\frac{dz}{dx}|$$

$$p_x(x) = p_z(z) \cdot |\frac{dx}{dz}|^{-1}$$

$$p_x(x) = p_z(z) \cdot |\frac{dg(z)}{dz}|^{-1} \text{ where } x = g(z).$$

$$p_x(x) = p_z(z) \cdot |g'(z)|^{-1}$$

$$p_x(x) = p_z(g^{-1}(x)) \cdot |g'(g^{-1}(x))|^{-1} \text{ where } z = g^{-1}(x).$$

change of variable formula

Fitting an NF to data

$$p_x(x_i) = p_z(g^{-1}(x_i)) \cdot |g'(g^{-1}(x_i))|^{-1}$$

The likelihood of the left handside = The likelihood of the left right-hand side.

Here x_i is a data point.

Assuming we know g, its inverse, how to differentiate it and p_z, we can compute $p_x(x_i)$

Fitting an NF to data

$$p_x(x_i) = p_z(g^{-1}(x_i)) \cdot |g'(g^{-1}(x_i))|^{-1}$$

The likelihood of the left handside = The likelihood of the left right-hand side.

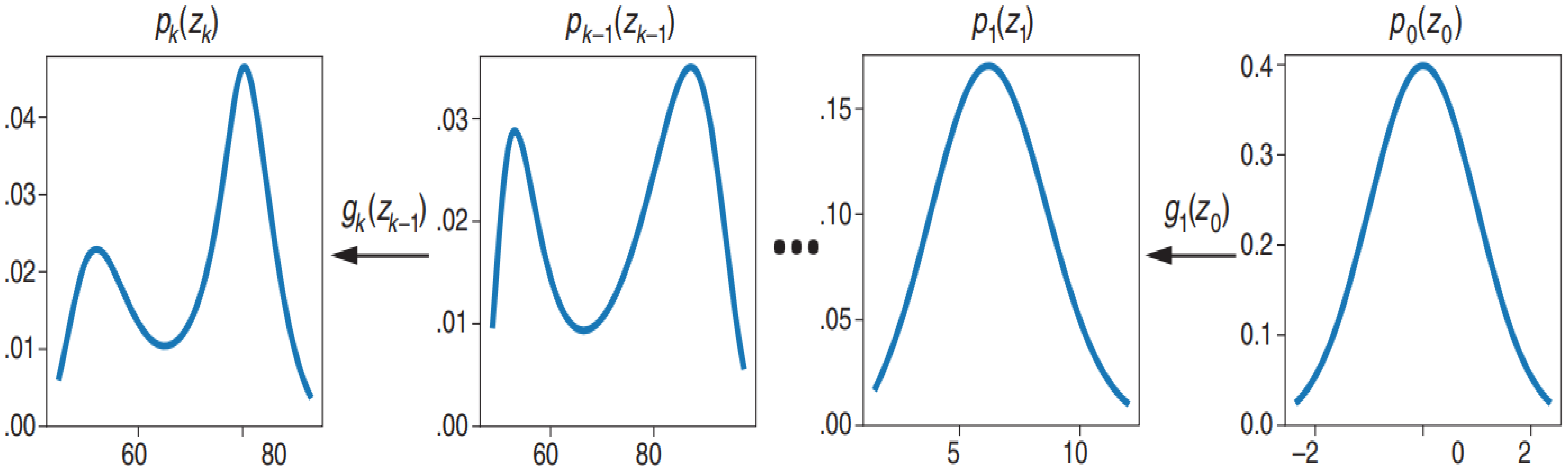
Here xi is a data point.

Assuming we know g, its inverse, how to differentiate it and p_z, we can compute $p_x(xi)$

The joint likelihood of all data: $\prod_{i=1}^{n} p_x(x_i)$

Equivalently we can look for $\sum_{i=1}^{n} \log(p_x(x_i))$

Chain of transformations



A sequence of simple transformations can be combined to create more complex transformations necessary for modeling intricate probability distributions.

By applying a series of these transformations in a cascading manner, starting from a standard Gaussian distribution z0 ~ N(0, 1), we can gradually shape the distribution into a more complex form with distinct modes or peaks. This step-by-step transformation process enables us to model and capture the intricacies of complex distributions, allowing for a more accurate representation of real-world data.

High dimensional data

The same formula works.

$$p_x(x) = p_z(z) \cdot \left| \det \left(\frac{\partial g(z)}{\partial z} \right) \right|^{-1}$$

Jacobi matrix

[Jacobian matrix and determinant - Wikipedia]

High dimensional data

The same formula works.

$$p_x(x) = p_z(z) \cdot \left|\det\left(\frac{\partial g(z)}{\partial z}\right)\right|^{-1}$$

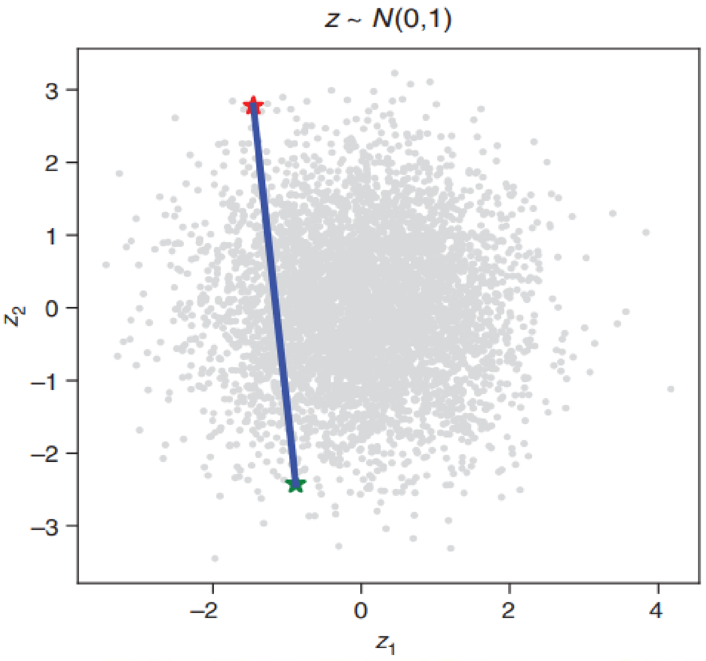
Jacobi matrix

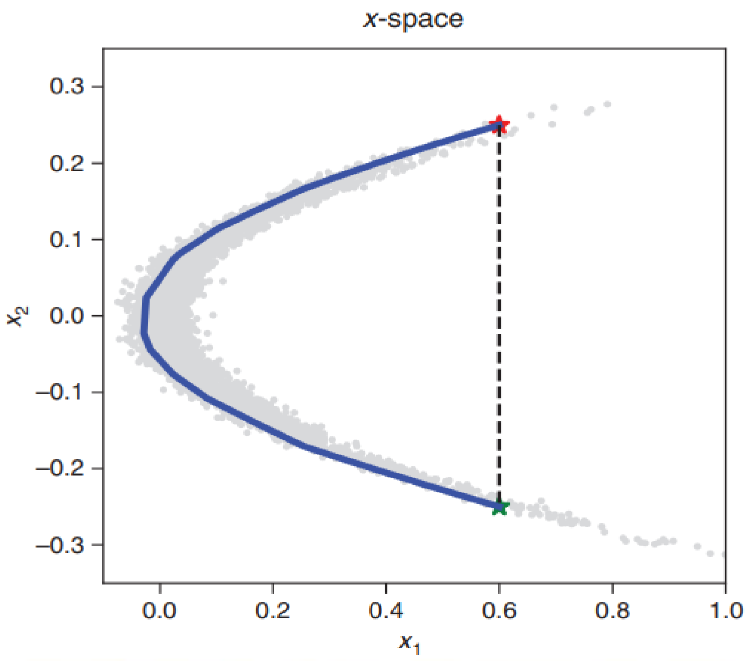
$$\frac{\partial g(z)}{\partial z} = \begin{pmatrix} \frac{\partial g_1(z)}{\partial z_1} & 0 & 0 \\ \frac{\partial g_2(z)}{\partial z_1} & \frac{\partial g_2(z)}{\partial z_2} & 0 \\ \frac{\partial g_3(z)}{\partial z_1} & \frac{\partial g_3(z)}{\partial z_2} & \frac{\partial g_3(z)}{\partial z_3} \end{pmatrix}$$

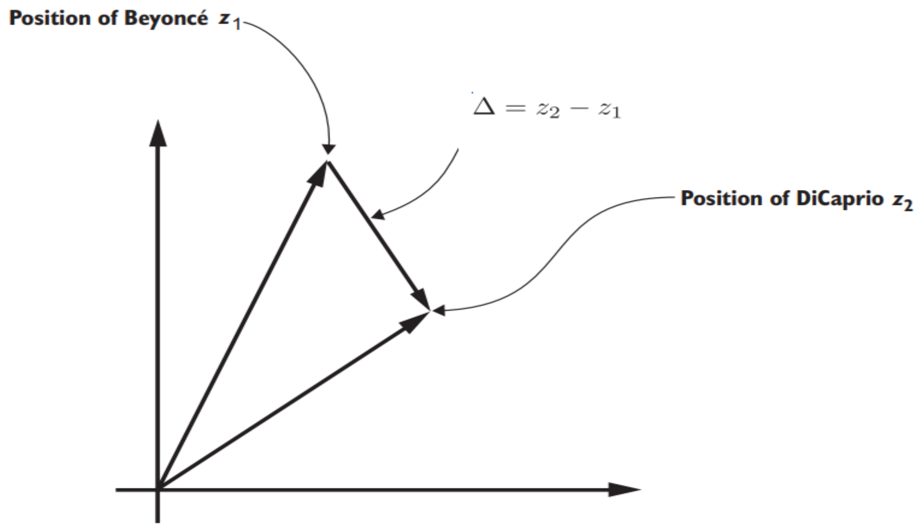
One problem: computing the determinant is expensive! The usual trick is to choose g such that the matrix is upper/lower triangular and hence the determinant is just the multiplication of the diagonal elements

[Jacobian matrix and determinant - Wikipedia]

Applications

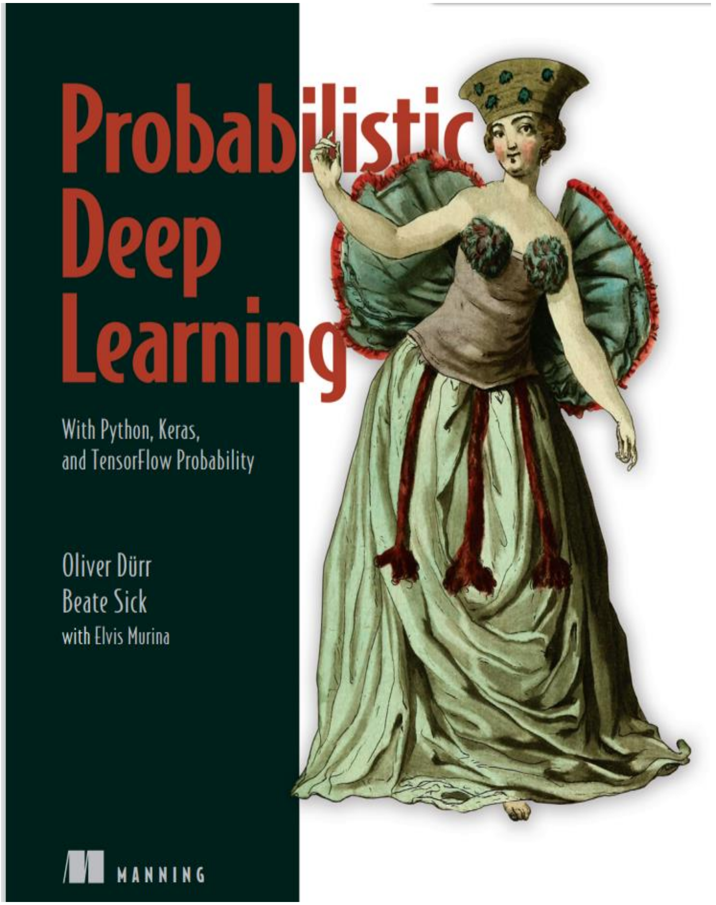






We start from (Beyoncé) and move from there in the direction Δ to DiCaprio. You then use the NF xc = g(zc) to go from the z space to the x space.

Refs


Probabilistic Deep Learning
With Python, Keras, and TensorFlow Probability
Oliver Dürr
Beate Sick
with Elvis Murina
MANNING