Sampling and transport maps

The first two problems in this tutorial will revolve around the concept of *transport maps*, which is a very important idea in modern statistics and machine learning. In particular, most generative modelling techniques such as diffusion models, GANs, flow matching, etc... can be seen as particular instances of this idea.

A map $T : \mathcal{X} \to \mathcal{Y}$ is said to be a *transport map* between two densities $f_X : \mathcal{X} \to \mathbb{R}$ and $f_Y : \mathcal{Y} \to \mathbb{R}$ if for any $X \sim f_X$, the random variable $Y = T(X)$ has density f_Y.

The core application of this idea is in generating samples from "weird" densities. Suppose that f_X is some "nice" density that we can sample from, such as Uniform$(0, 1)$, whereas f_Y is a more complicated density which we don't yet know how to sample from. If we have a transport map T between these two, then we can convert a sample $X \sim f_X$ into a sample from f_Y by simply letting $Y = T(X)$.

1 (Box-Muller method)**.** Suppose that the bivariate random variable $X = (X_1, X_2)$ is uniform over the unit disk $D = \{x_1^2 + x_2^2 \leq 1\}$, i.e.,

$$f_X(x_1, x_2) = \frac{1}{\pi} 1_D(x)$$

and that $f_Y : \mathbb{R}^2 \to \mathbb{R}$ is the standard bivariate normal density

$$f_Y(y_1, y_2) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(y_1^2 + y_2^2)\right\}$$

Show that the map $T : D \to \mathbb{R}^2$ given by

$$T(x_1, x_2) = \sqrt{\frac{-2 \log r^2}{r^2}} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad r^2 = x_1^2 + x_2^2$$

is a transport map between f_X and f_Y. That is to say, that if $X \sim f_X$ then $T(X) \sim f_Y$. This gives us a way to sample from the standard normal distribution.

Solution. The change of variables theorem states that

$$f_Y(y_1, y_2) = \frac{f_X(T^{-1}(y_1, y_2))}{\det T'(T^{-1}(y_1, y_2))}.$$

Therefore we aim to find T^{-1} and the Jacobian matrix J.

To find the inverse map, note that

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \sqrt{\frac{-2 \log r^2}{r^2}} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

which means that

$$y_1^2 + y_2^2 = \frac{-2 \log r^2}{r^2}(x_1^2 + x_2^2) = -2 \log r^2$$

resulting in

$$\frac{-2 \log r^2}{r^2} = (y_1^2 + y_2^2) \exp\left\{(y_1^2 + y_2^2)/2\right\}$$

and so

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = T^{-1}(y_1, y_2) = \frac{\exp\{-(y_1^2 + y_2^2)/4\}}{\sqrt{y_1^2 + y_2^2}} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

which is the first piece that we wanted.

For the second piece, we can take a short-cut by noting that

$$T'(T^{-1}(y_1, y_2)) = (T^{-1})'(y_1, y_2)^{-1}$$

which results in

$$f_Y(y_1, y_2) = f_X(T^{-1}(y_1, y_2)) \det\Big[(T^{-1})'(y_1, y_2)\Big]$$

so actually all we need to do is find the Jacobian of T^{-1}. Many sources use the notation

$$f_Y(y_1, y_2) = f_X(x_1, x_2) \left|\det \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)}\right|$$

with $x = T^{-1}(y)$. This computation is horrendously ugly so I will save all of us some hassle and just put the answer down

$$\det \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = -\frac{1}{2} \exp\Big\{-(y_1^2 + y_2^2)/2\Big\}$$

I suggest you guys to go over this differentiation in your own time if you'd like some practice with finding Jacobian determinants.

Putting the pieces together results in

$$f_Y(y_1, y_2) = \frac{1}{2\pi} \exp\Big\{-(y_1^2 + y_2^2)/2\Big\}$$

which is what we wanted to show. □

For example, if f_Y is the density over all images of cats, then we can generate a new cat picture by generating $X \sim f_X$ and letting $Y = T(X)$. In fact, diffusion models and GANs do exactly this by learning a transport map T from data. In what follows, we'll find transport maps T when f_X is uniform over the unit square and f_Y is "nice enough".

2 (Inverse transform method)**.** In this problem, we show a general method for finding transport maps between Uniform$(0, 1)$ and arbitrary univariate distributions.

(a) Suppose that $X \sim \text{Uniform}(0, 1)$ and F is a c.d.f. on $\mathbb{R}$ that is invertible. Show that $T = F^{-1}$ is a transport map between Uniform$(0, 1)$ and F, i.e., that the c.d.f. of $Y = T(X)$ is F.

(b) Find a transport map between Uniform$(0, 1)$ and the Cauchy distribution with c.d.f. $F(y) = \frac{1}{2} + \frac{1}{\pi}\arctan(y)$.

Solution. For part (a), Let's find the c.d.f. of Y

$$\mathrm{P}(Y \le y) = \mathrm{P}(F^{-1}(X) \le y)$$

and since F^{-1} is increasing this is equal to

$$\mathrm{P}(X \le F(y)) = F(y)$$

since the c.d.f. of $X \sim \text{Uniform}(0,1)$ is the identity function. That is to say, $\mathrm{P}(Y \le y) = F(y)$ so the c.d.f. of Y is F as we wanted.

Part (b) is a simple application of this idea. From part (a) we know that F^{-1} is the desired transport map, so all we need to do is invert

$$x = F(y) = \frac{1}{2} + \frac{1}{\pi}\arctan(y)$$

which is $y = \tan[\pi(x - 1/2)]$. Thus $T(x) = \tan[\pi(x - 1/2)]$ is the transport map. □

Order statistics

3. Let X_1 and X_2 be independent $\text{Uniform}(0,1)$ random variables and let $M = \min\{X_1, X_2\}$. Find the c.d.f. and p.d.f. of M.

Solution. For the c.d.f., note that

$$\mathrm{P}(M \le m) = \mathrm{P}(X_1 \le m, X_2 \le m) = \mathrm{P}(X_1 \le m)\,\mathrm{P}(X_2 \le m) = m^2$$

whenever $m \in [0,1]$ and 0 otherwise. Thus the p.d.f. is

$$f_M(m) = \frac{d}{dm}\mathrm{P}(M \le m) = 2m$$

on $[0,1]$ and 0 otherwise. □

4. Let X and Y be independent $\text{Exp}(\lambda)$ random variables. Find the density of $Z = X - Y$.

Solution. Notice that the set

$$\{Z = z\} = \{X - Y = z\} = \{X = Y + z\} = \{Y = X + z\}$$

□

Bonus: the Knothe map

The following problem, adapted from Figalli and Glaudo 2023, is a generalization of the inverse transform method to bivariate distributions.

5 (Knothe map)**.** Suppose f_X is the uniform density over the unit cube $U = [0,1]^2$, i.e.,

$$f_X(x_1, x_2) = 1_U(x_1, x_2)$$

and $f_Y(y_1, y_2)$ be any bivariate density over $\mathbb{R}^2$. Let

$$F(y_1) = \int_{-\infty}^{y_1} f_{Y_1}(y_1')\,dy_1', \quad f_{Y_1}(y_1) = \int_{\mathbb{R}} f_Y(y_1, y_2)\,dy_2$$

be the c.d.f. of Y_1 and

$$F_{y_1}(y_2) = \int_{-\infty}^{y_2} f_{Y_2|Y_1}(y_2'|y_1)\,dy_2', \quad f_{Y_2|Y_1}(y_2|y_1) = f_Y(y_1, y_2)/f_{Y_1}(y_1)$$

be the c.d.f. of $Y_2 \mid Y_1 = y_1$.

(a) Find a transport map T_1 between f_{X_1} and f_{Y_1}.

(b) For (almost) every x_1 find a transport map $T_2(x_1, \cdot)$ between f_{X_2} and $f_{Y_2|Y_1}(\cdot \mid T_1(x_1))$.

(c) Find a transport map T between f_X and f_Y by combining the previous two results.

Solution. Part (a) is an easy consequence of the inverse transform method, $T_1 = F^{-1}$. Part (b) can be done similarly, we just invert the conditional c.d.f. $T_2(x_1, x_2) = (F_{T_1(x_1)})^{-1}(x_2)$. For part (c), take

$$T(x_1, x_2) = \begin{pmatrix} T_1(x_1) \\ T_2(x_1, x_2) \end{pmatrix}$$

Clearly, if $X_1, X_2 \sim \text{Uniform}(0, 1)$, then we need to verify that

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} T_1(X_1) \\ T_2(X_1, X_2) \end{pmatrix} \sim f_Y.$$

By part (a), $Y_1 = T_1(X_1) \sim f_{Y_1}$. Part (b) also tells us that $Y_2 \mid Y_1 = y_1 \sim f_{Y_2|Y_1}(\cdot \mid y_1)$. Putting it all together means that the joint distribution is

$$f_{Y_1}(y_1) f_{Y_2|Y_1}(y_2 \mid y_1) = f_Y(y_1, y_2)$$

which is what we wanted to show. □

6. Find the Knothe map between f_X and f_Y where

$$F_Y(y_1, y_2) = \frac{y_1 y_2}{y_1 + y_2 + y_1 y_2} 1_{\{y_1, y_2 \geq 0\}}$$

is the c.d.f. corresponding to f_Y.

Solution. To find the Knothe map we first need to find the marginal c.d.f. of Y_1 which is

$$F(y_1) = \lim_{y_2 \to \infty} F_Y(y_1, y_2) = \frac{y_1}{1 + y_1}$$

whose inverse is $T_1(x_1) = x_1/(1 - x_1)$. The c.d.f. of the conditional distribution $Y_2 \mid Y_1 = y_1$ is

$$F_{y_1}(y_2) = \frac{\partial}{\partial y_1} F_Y(y_1, y_2) / f_{Y_1}(y_1) = \left(\frac{(1 + y_1) y_2}{y_1 + (1 + y_1) y_2} \right)^2$$

with inverse

$$(F_{y_1})^{-1}(x_2) = \frac{y_1 \sqrt{x_2}}{(1 + y_1)(1 - \sqrt{x_2})}$$

which means that

$$T_2(x_1, x_2) = \frac{x_1 \sqrt{x_2}}{1 - \sqrt{x_2}}$$

and thus

$$T(x_1, x_2) = \begin{pmatrix} \frac{x_1}{1 - x_1} & \frac{x_1 \sqrt{x_2}}{1 - \sqrt{x_2}} \end{pmatrix}^T$$

is what we wanted to find. □