Probabilistic Normalizing Flows

Shi Feng, Deqian Kong

December 12, 2021

1 Change of variables

1.1 Single variable

Let X be a continuous random variable with a generic pdf $f_X(x)$ defined over the support $c_1 < x < c_2$, and let $Y = u(x)$ be a continuous monotonically increasing funtion of X with inverse function $X = v(Y)$. We would like to know the pdf $f_Y(y)$ of Y.

We start with the cumulative dsitribution $F_Y(y)$ of Y, written as

$$F_Y(y) = P(Y \leq y) = \int_{d_1}^{y} f_Y(y)dy \tag{1.1}$$

for $d_1 = u(c_1) < y < u(c_2) = d_2$. Since $Y = u(X)$, we have

$$F_Y(y) = P(u(X) \leq y) \tag{1.2}$$

Since the map between X and Y is invertable, the preimage of $u(X) \leq y$ is $X \leq v(y)$, thus

$$F_Y(y) = P(X \leq v(y)) = \int_{c_1}^{v(y)} f_X(x)dx \tag{1.3}$$

Now we can take derivative of $F_Y(y)$ wrt y to get $f_Y(y)$:

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_Y(y)}{dv(y)}\frac{dv(y)}{dy} = f_X[v(y)]v'(y) \tag{1.4}$$

By the same token, one can show that if $v(x)$ is monotonically decreasing, we have

$$f_Y(y) = -f_X[v(y)]v'(y) \tag{1.5}$$

so, inconclusion, for a generic **invertable** funcion $Y = u(X)$, we have

$$f_Y(y) = f_X[v(y)]\big|v'(y)\big| \tag{1.6}$$

1.2 Multi-variable

Now suppose a pair of random variables (X_1, X_2) has joint pdf $f_X(x_1, x_2)$ and support S_X, let (Y_1, Y_2) be some function of (X_1, X_2) defined by $Y_1 = u_1(X_1, X_2)$ and $Y_2 = u_2(X_1, X_2)$ with single-valued inverse given by $X_1 = v_1(Y_1, Y_2)$ and $X_2 = v_2(Y_1, Y_2)$, and let S_Y be the support of Y_1, Y_2.

The cumulative distribution $F(y_1, y_2)$ is

$$F(y_1, y_2) = F(Y_1 < y_2, Y_2 < y_2) = \int_{d_1,d_2}^{y_1,y_2} f_Y(y_1, y_2)dy_1dy_2 \tag{1.7}$$

by the mapping between two sets of variables we have

$$F(Y_1 \le y_1, Y_2 \le y_2) = F[u_1(X_1, X_2) \le y_1, u_2(X_1, X_2) \le y_2] \tag{1.8}$$

Since the map is invertable we have

$$\begin{aligned} F(Y_1 \le y_1, Y_2 \le y_2) &= F[X_1 \le v_1(Y_1, Y_2), X_2 \le v_2(Y_1, Y_2)] \\ &= \int^{v_1(y_1,y_2)} \int^{v_2(y_1,y_2)} f(x_1, x_2) dy_1 dy_2 \end{aligned} \tag{1.9}$$

This gives TODO:

$$g(y_1, y_2) = |J_v| f[v_1(y_1, y_2), v_2(y_1, y_2)] \tag{1.10}$$

where J_v is the Jacobian of the inverse map.

2 Normalizing Flows

Let $\mathbf{x}$ be a D-dimentional continuous random real vector, with a joint distribution $p_x(\mathbf{x})$. Now suppose $\mathbf{x}$ is a variable transformed from another variable $\mathbf{u}$ via transformaion $\mathbf{x} = T(\mathbf{u})$:

$$\mathbf{x} = T(\mathbf{u}), \quad \text{where } \mathbf{u} \sim p_u(\mathbf{u}) \tag{2.1}$$

2.1 Expressive power

Theorem 2.1. *The probabilistic flow can express any distribution $p_x(\mathbf{x})$ regardless of the concrete form of base distribution*

Proof. A joint probability distribution can be expressed as a regression, e.g.

$$p(x_1, x_2, x_3, x_4) = \overbrace{\underbrace{p(x_1)p(x_2|x_1)}_{p(x_1,x_2)} p(x_3|x_1, x_2)}^{p(x_1,x_2,x_3)} p(x_4|x_1, x_2, x_3, x_4)$$

Hence, in a compact form:

$$p_x(\mathbf{x}) = \prod_{i=1}^{D} p_x(x_i|\mathbf{x}_{<i}) \tag{2.2}$$

since $p_x(\mathbf{x})$ is always non-zero, all $p_x(x_i|\mathbf{x}_{<i})$ are also non-zero. Next we define the transformation $F : \mathbf{x} \to \mathbf{z} \in (0, 1)^D$ whose i-th element is given by the cumulative distribution function (CDF) of the i-th conditional pdf:

$$z_i = F_i(x_i, \mathbf{x}_{<i}) = \int_{-\infty}^{x_i} p_x(x_i'|\mathbf{x}_{<i}) dx_i' = Pr(x_i' \le x_i|\mathbf{x}_{<i}) \tag{2.3}$$

note that z_i is a random variable since the upper bound of the integral x_i is a random variable. Since each F_i is differentiable wrt its inputs x_i, F is differentiable wrt $\mathbf{x}$. Moreover, since

$$\frac{\partial F_i}{\partial x_i} = p_x(x_i|\mathbf{x}_{<i}) \ge 0 \tag{2.4}$$

F_i is a continuous monotonic function, thus invertable. Since F_i does not depend on x_j's for $j > i$, we must have

$$\frac{\partial F_i}{\partial x_j} = 0 \text{ for } i < j \tag{2.5}$$

that is, the Jacobian of F is a lower triangular matrix whose determinant is equal to the product of its diagonal elements:

$$\det J_F(\mathbf{x}) = \prod_{i=1}^{D} \frac{\partial F_i}{\partial x_i} = \prod_{i=1}^{D} p_x(x_i|\mathbf{x}_{<i}) = p_x(\mathbf{x}) \tag{2.6}$$

so that Jacobian determinant of F is exactly $p_x(\mathbf{x})$, which is necessarily non-zero. Therefore the inverse of $J_F(\mathbf{x})$ exists, and is equal to the Jacobian of F^{-1} (which is also lower triangular $x_i = F_i^{-1}(\bullet, \mathbf{x}_{<i})(z_i))$. Therefore F is a diffeomorphism. Hence, using $\det J_F = \frac{1}{\det J_{F^{-1}}}$, we get

$$p_z(\mathbf{z}) = p_x[F^{-1}(\mathbf{z})]\big|J_{F^{-1}}[F^{-1}(\mathbf{z})]\big| = p_x(\mathbf{x})|\det J_F(\mathbf{x})|^{-1} = p_x(\mathbf{x})|p_x(\mathbf{x})|^{-1} = 1 \tag{2.7}$$

which implies $\mathbf{z}$ is a uniform distribution in $(0,1)^D$. Thus we have shown that any distribution $p_x(\mathbf{x})$ can be deformed continuously into a uniform distribution. □

References

[1] Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S. & Lakshminarayanan, B. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research* **22**, 1–64 (2021). URL http://jmlr.org/papers/v22/19-1028.html.