Random Points on a Donut

If you want random points on a line segment, you just open python and type

```
import random
random.random()
```

but what if you want random points from a more exotic object, like a donut, or an ellipse? In this article I will develop a method for picking random points from these objects in the general setting of k-dimensional submanifolds of $\mathbb{R}^n$.

A naive method for curves

If you draw a curve on a piece of paper you have to move a pencil across it. This movement can be described by specifying where the tip of your pencil has to be at any given time. We can describe this mathematically via a function $\alpha : (0, T) \to \mathbb{R}^2$ where T is the time it takes you to draw the curve and $\mathbb{R}^2$ represents our flat paper. This function will be called a *parametrization* for the curve.

An example of a curve is the ellipse, it can be drawn according to the following parametrization $\alpha(t) = (a\cos t, b\sin t)$ for $t \in (0, 2\pi)$. Obviously, there are multiple ways of drawing the same curve. For example, you can move your pencil twice as fast and draw the exact same ellipse, but your parametrization will be $\alpha(t) = (a\cos(2t), b\cos(2t))$ for $t \in (0, \pi)$. If we have any two parameterizations α and β, the function $\alpha^{-1} \circ \beta$ will be called a *reparametrization* because it specifies how to translate between α and β. This raises an important point: *geometric properties of curves should be the same regardless of what parametrization is used.*

How do we sample random points from an ellipse? A promising strategy is to put random dots on the curve as you're drawing it. Mathematically this is equivalent to taking a random sample $x_1, \ldots, x_n$ from $(0, 2\pi)$ and returning $\alpha(x_1), \ldots, \alpha(x_n)$.

The problem with this approach is that we move our pencil with different speeds as we draw the curve. We jot down more points in areas were the pencil moves slowly. We need to find a way to find a parametrization β that has the same speed, lets say 1, everywhere: $|\beta'(t)| = 1$ for all t. Notice that if we move with unit speed, the length of the curve we draw in t units of time will always be equal to t, and so it takes L units of time to draw the whole curve, where L is the length of the curve. For this reason any parametrization satisfying $|\beta'(t)| = 1$ is called an *arclength parametrization.*

To find the arclength parametrization we need to find a suitable *reparametrization* $t : (0, L) \to (0, T)$ such that $\beta = \alpha \circ t$ satisfies the ordinary differential equation

$$|\beta'(s)| = |\alpha'(t(s))|\, t'(s) = 1.$$

The solution to this equation is given by the inverse of the function

$$s(t) = \int_0^t |\alpha'(t)|\, dt$$

which is called the arclength function.

If we use β in place of α in our sampling scheme, we will get an adequate sampling algorithm for curves. Another advatange of this scheme is that it doesn't depend on the parametrization, since the arclength parametrization is unique.

Our sampling algorithm for a curve parametrized by α is

1. Find its arclength parametrization β and the length of the curve.
2. Sample $x_1, \ldots, x_n$ uniformly from $[0, L)$.
3. Return $\alpha(x_1), \ldots, \alpha(x_n)$.

For reasons that will become clear later I will refer to this as the *naive* method.

Manifolds and integration on manifolds

I have included a section with a brief outline of elementary differential geometry. Those who are familiar with the concepts of manifolds and volume form can safely skip.

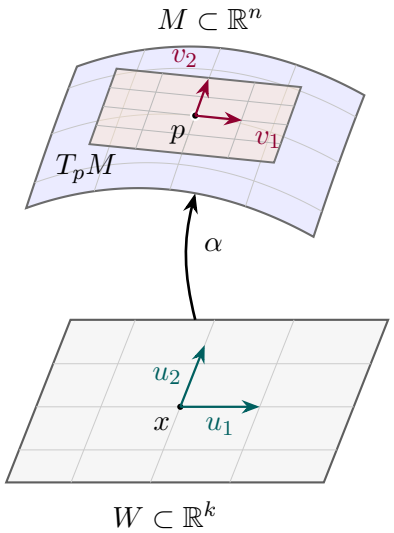

Figure 1: A parametrization α maps a point $x \in W \subset \mathbb{R}^k$ and tangent vectors $u_1, u_2 \in T_x\mathbb{R}^k$ to a point $p \in M$ and tangent vectors $v_1, v_2 \in T_pM$.

A similar description can be given for compact orientable k-dimensional submanifolds $M \subset \mathbb{R}^n$.

Definition. For each $x \in M$ there is a neighborhood U of x and an open set $W \subset \mathbb{R}^k$ with a function $\alpha : W \to U$ such that

1. $\alpha(W) = M \cap U$
2. $\alpha'(y)$ has rank k for each $y \in W$.
3. $\alpha^{-1} : (M \cap U) \to W$ is continuous.

Such a function is called a *coordinate system* or *parametrization* of M around x. It can be thought of as specifying a way of deforming a sheet of paper into a particular shape.

In particular any open subset of $\mathbb{R}^n$ will be an n-dimensional manifold with a coordinate system given by the identity $\alpha(x) = x$. Intuitively this means that, to wrap a sheet of paper around this manifold, we don't have to deform it at all! However, It is not the only coordinate system, for example we can parametrize $(0, \infty) \subset \mathbb{R}$ by $\beta(x) = \sqrt{x}$. This time we have to imagine that we are stretching the sheet of paper outward. As

with curves, the map $\alpha^{-1} \circ \beta$—which we will call a *change of coordinate* map—tells us how to translate between the parameter spaces of α and β.

Another example would be the sphere $S^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}$. In the case of $n = 2$ we get the circle, which is typically parametrized by $\alpha(\theta) = (\cos\theta, \sin\theta)$ with $\theta \in (0, 2\pi)$. Notice that this doesn't include the point $(0, -1)$, therefore to completely specify the circle we need at least two coordinate systems: one which covers the bottom half and one which covers the top half.

A result in differential geometry states that a compact manifold M will admit a coordinate system $\alpha : W \to U$ such that $M - (M \cap U)$ will have "measure zero" in M. For this reason, we will only deal with one coordinate system that describes M "almost everywhere."

I will denote by v_x—called the vector v tangent at x—a vector $v \in \mathbb{R}^k$ emanating from $x \in \mathbb{R}^k$. The set of all such vectors $T_x\mathbb{R}^k = \{v_x : v \in \mathbb{R}^k\}$ is called the tangent space at x.

The plane that around a point $p \in M$ best approximates the manifold M will be called the *tangent space* and denoted T_pM. If M is parametrized by $\alpha : U \to M$ and $p = \alpha(x)$ then

$$T_pM = \{\alpha'(x)u : u \in T_x\mathbb{R}^k\}$$

The tangent space $T_x\mathbb{R}^k$ can be understood as a special case of this, where the best linearly approximating subspace is just $\mathbb{R}^k$.

We wish to find the volume/surface area of a manifold M. We will do this by first introducing a way of measuring volumes of parallelepipeds on the tangent spaces T_pM. Then we will show how we can express this information in terms of $x = \alpha^{-1}(p)$ and the tangent vectors at x. Finally we will integrate over x, adding up the volumes of infinitesimally small parallelepipeds on the tangent spaces to obtain the volume of the entire manifold.

Volume form. The object which measures the volumes of parallelepipeds on the tangent spaces is called the *volume form* and is denoted by dV. A standard way to obtain such an object is by introducting another object g which assigns to each tangent space T_pM an inner product $g_p : T_pM \times T_pM \to \mathbb{R}$ in a smooth way. A standard result says that the parallelepiped with sides $v_1, \ldots, v_k$ on T_pM has volume

$$dV_p(v_1, \ldots, v_k) = \det((g_p(v_i, v_j))_{ij})^{1/2}.$$

In particular, since our manifolds are subsets of $\mathbb{R}^n$, we can take g to be the usual inner product on $\mathbb{R}^n$, i.e., $g_p(u, v) = u^T v$.

Example. As an example, lets consider the volume form of $\mathbb{R}^n$ is going to be the volume of the n-dimensional parallelepiped defined by the n-dimensional vectors $v_1, \ldots, v_n \in T_x\mathbb{R}^n$ is

$$dV_x(v_1, \ldots, v_n) = (dx_1 \ldots dx_n)_x(v_1, \ldots, v_n) = \det V$$

where $V = (v_1, \ldots, v_n)$. We will refer to this object as $dx_1 \ldots dx_n$ from here on out so that it matches with the usual notation for multiple integrals.

Pullback. If x is a point in the parameter space, then we can express this object in terms of x via the *the pullback of the volume form*

$$(\alpha^* dV)_x(u_1, \ldots, u_k) = dV_{\alpha(x)}(\alpha'(x)u_1, \ldots, \alpha'(x)u_k)$$

In the case where g is the usual inner product we will have that

$$\begin{aligned} dV_{\alpha(x)}(\alpha'(x)u_1, \ldots, \alpha'(x)u_k) &= \det\Big((\alpha'(x)U)^T(\alpha'(x)U)\Big)^{1/2} \\ &= \det\Big(\alpha'(x)^T\alpha'(x)\Big)^{1/2} \det U \end{aligned}$$

where U is the $k \times k$ matrix whose columns are $u_1, \ldots, u_k$. Since $\det U$ is the volume form of $\mathbb{R}^k$ applied to $u_1, \ldots, u_k$ we can say that

$$(\alpha^* dV)_x = \det\Big(\alpha'(x)^T\alpha'(x)\Big)^{1/2} dx_1 \ldots dx_k$$

and, as such, we define the volume of M to be

$$\int_M dV = \int_U (\alpha^* dV) = \int_U \det\Big(\alpha'(x)^T\alpha'(x)\Big)^{1/2} dx_1 \ldots dx_k.$$

Example. In particular, that if $M = \mathbb{R}^n$ and $dV = dx_1 \ldots dx_n$ is the volume form, then

$$\theta^*(dx_1 \ldots dx_n) = |\det \theta'| dx_1 \ldots dx_n$$

which corresponds to the usual change of variables formula in multivariable calculus.

We define the integration of functions $f : M \to \mathbb{R}$ over M by claryfing what $f\,dV$ and $\alpha^*(f\,dV)$ mean. Firstly, we define

$$(f\,dV)_p(v_1, \ldots, v_k) = f(p)\,dV_p(v_1, \ldots, v_k)$$

so, in a sense, integrating $f\,dV$ means summing up the areas of infinitesimally small parallelepipeds multiplied by values of f. The pullback $\alpha^*(f\,dV)$ is therefore $(f \circ \alpha)\,\alpha^* dV$. The integral of f over M is thus defined to be

$$\int_M f\,dV = \int_U (f \circ \alpha)\,\alpha^* dV$$

Its worth noting again that this result is analagous to the change of variables formula when $M = \mathbb{R}^n$ since

$$\int_{\alpha(U)} f(x)\,dx_1 \ldots dx_n = \int_U f(\alpha(x))|\det \alpha'(x)|\,dx_1 \ldots dx_n$$

As an exercise, you should try to show that the usual formulas for line and surface integrals in vector calculus are special cases of what we've discussed here.

We will need the following fact in the sequel

Proposition. *If $\alpha : W \to M$ is a coordinate system on M and $\theta : U \to W$ is a change of coordinates map then $(\alpha \circ \theta)^* dV = \theta^*(\alpha^* dV)$.*

Proof.

$$\begin{aligned} &(\theta^*(\alpha^* dV))_x(u_1, \ldots, u_k) \\ &\quad = (\alpha^* dV)_{\theta(x)}(\theta'(x)u_1, \ldots, \theta'(x)u_1) \\ &\quad = dV_{\alpha(\theta(x))}(\alpha'(\theta(x))\theta'(x)u_1, \ldots, \alpha'(\theta(x))\theta'(x)u_1) \\ &\quad = dV_{\alpha(\theta(x))}((\alpha \circ \theta)'(x)u_1, \ldots, (\alpha \circ \theta)'(x)u_k) \\ &\quad = ((\alpha \circ \theta)^* dV)_x(u_1, \ldots, u_k) \end{aligned}$$

□

Probability on manifolds

To avoid unnecessary mathematical machinery (geometric measure theory) I will be informal in this section and only focus on densities.

A function $f : M \to \mathbb{R}$ is said to be a *density* on M if $f \geq 1$ and $\int_M f\, dV = 1$. The probability of a subset A of M is determined by

$$P(A) = \int_A f\, dV.$$

In particular, for compact manifolds, we can define the uniform density $f(p) = 1/\operatorname{vol}(M)$ where $\operatorname{vol} M = \int_M dV$ is the total volume of the manifold.

As usual, given another probability space (Ω, P), a random variable $X : \Omega \to M$ is said to follow a density $X \sim f$ if $P \circ X^{-1}$ has that density.

Generalizing the arclength parametrization

Let $\alpha : W \to M$ be an almost-everywhere coordinate system of the compact manifold M. If X is a uniform random variable on W, will $\alpha(X)$ be uniform on M? This turns out not to be the case, as α might shrink or expand some neighborhoods of W and hence change their volume. Just like in the one-dimensional case, we need coordinate systems that don't change volume. A coordinate system $\beta : W \to M$ is said to be *volume-preserving* if

$$\int_{\beta(U)} dV = \int_U dx_1 \dots dx_k$$

for any measurable $U \subset W$. That is, U has the same volume in $\mathbb{R}^k$ that its image $\beta(U)$ has in M.

Proposition. *If X is a uniform random variable on W, then $\beta(X)$ is uniform on M if and only if β is a volume-preserving parametrization.*

Proof.

$$\int_{\beta(U)} dV = P(\beta(X) \in \beta(U)) = P(X \in U) = \int_U dx_1 \dots dx_k$$

□

Since volume-preservation is something that holds for all U, we can restate it in terms of the differentials alone: β is volume-preserving if $\beta^* dV = dx_1 \dots dx_k$. In the one-dimensional case this is equivalent to $|\beta'(t)|dt = dt$ which immediately implies that $|\beta'(t)| = 1$.

We'd like to find $\theta : V \to W$ such that $\beta = \alpha \circ \theta$ is volume-preserving. This means that

$$\begin{aligned}\beta^* dV &= \theta^*(\alpha^* dV) = \theta^*(f\, dx_1 \dots dx_k) \\ &= (f \circ \theta) \det \theta'\, dx_1 \dots dx_k = dx_1 \dots dx_k\end{aligned}$$

where $f(x) = \det\Big(\alpha'(x)^T \alpha'(x)\Big)^{1/2}$. Thus, to find β we have to solve the partial differential equation $(f \circ \theta) \det \theta' = 1$ for θ, called the *Monge-Ampere equation*, under some suitable boundary conditions. One can easily show that in one-dimension this equation is equivalent to the ODE for arclength reparametrization.

Our problem is thus reduced to solving the Monge-Ampere equation. There are results in the mathematical literature that prove the existence

of solutions, and there are numerical procedures for solving it in special cases, but implementing any of them was such a daunting task to me that I decided to look for other approaches to sampling from manifolds. Luckily I found something much simpler.

A change of perspective

So far, we've been focusing on taking uniform samples from some subset of $\mathbb{R}^k$ and passing them through a volume-preserving coordinate system. Another approach is to find a density f on W that is pushed forward to a uniform distribution on M by α. We can obtain such an f by pulling back dV along α

$$\alpha^* dV = f\, dx_1 \dots dx_k.$$

More precisely, if $X \sim f$ and B is a measurable subset of M then

$$P(\alpha(X) \in B) = P(X \in \alpha^{-1}(B)) = \int_{\alpha^{-1}(B)} f\, dx_1 \dots dx_k = \int_B dV$$

I almost can't believe I thought of this *after* the complicated approach of the previous section. This motivates a new approach:

1. Take random samples $\{x_1, \dots, x_n\}$ from W using a distribution proportional to $f(x) = \det\Big(\alpha'(x)^T \alpha'(x)\Big)^{1/2}$.
2. Pass those samples through α and return $\{\alpha(x_1), \dots, \alpha(x_n)\}$.

whereas before the main difficulty lay with finding a good parametrization, now the main difficulty is sampling from f. Fortunately this problem has well known and simple solutions. In particular I will use the method of *rejection sampling*.

Proposition. *Suppose that f is (a possibly unnormalized) probability density on $\mathbb{R}^k$ that we can't easily sample from. If g is a density on $\mathbb{R}^k$ that we can sample from and there exists a constant c such that $f \le cg$ then the following generates $X \sim f$:*

1. *Generate $U \sim \mathrm{U}(0,1)$ and $X \sim g$ independently of each other.*
2. *If $Ucg(X) > f(X)$ then go back to one, otherwise return X.*

Proof. Exericse. □

Under certain reasonable assumptions we can show that this method is suitable for what we want to do. In particular, we assume that

1. The parameter space W is a bounded open rectangle. This is done to simplify the representation on the computer.
2. The coordinate system $\alpha : W \to M$ is an almost-everywhere coordinate system.
3. The function f is bounded on W. This is not the case in general (e.g., parametrizing $(0,1)$ using $\sqrt{x}$, in which case $f(x) = x^{-1/2}/2$ is not bounded), but will hold if α can be extended continuously to some compact set containing W. I am not sure if this assumption is always compatible with (1), though I cannot produce a counterexample. I believe that it is reasonable.

Since f is bounded on the rectangle W, then there is a constant c such that $f \le c \cdot \mathrm{vol}(W) \cdot g$ where g is the uniform density on W. This justifies the use of rejection sampling for this purpose, though the performance might not be so great for very high-dimensional manifolds. A possible next step is applying Markov Chain Monte Carlo methods to this problem.

Application: Estimating mean Hausdorff distance

A common approach for measuring the geometric similarity between two objects is the *mean Hausdorff distance*. It is used in medicine for image segmentation, dentistry, etc.

The mean Hausdorff distance between manifolds M and N is defined to be

$$d_H(M,N) = \frac{1}{\operatorname{vol} M}\int_M d(x,N)\,dV_M + \frac{1}{\operatorname{vol} N}\int_N d(x,M)\,dV_N$$

where $d(x,N) = \min_{y\in N}|x-y|$ is the shortest distance from x to the manifold M. Computing this distance involves (a) finding $d(x,N) = \min_{y\in M}|x-y|$, i.e., the shortest distance from a fixed $x \in M$ to N, and (b) evaluating the integrals above.

The first part can be estimated by

$$d(x,N) = \sqrt{\min_{w\in W}|x-\alpha(w)|^2}$$

where $\alpha : W \to N$ is an almost-everywhere parameterization for N. The second part, i.e., the integral, can be estimated using Monte Carlo methods and manifold sampling. Let $X_1,\dots,X_n$ be a uniform sample from the manifold M. This means that $d(X_1,N),\dots,d(X_n,N)$ is a sequence of iid real-valued random variables. The law of large numbers states that

$$\frac{1}{n}\sum_{i=1}^{n} d(X_i,N) \xrightarrow{\text{a.s.}} \frac{1}{\operatorname{vol} M}\int_M d(x,N)\,dV_M$$

as $n\to\infty$. We will use the partial averages as estimates of this integral.

Application: Non-informative priors in Bayesian statistics

A typical task in statistics is to use i.i.d. observations $X = (X_1,\dots,X_n)$ to figure out what distribution they were generated from. To this end we specify a set $\mathcal{P} = \left\{p_\theta : \theta\in\Theta\subset\mathbb{R}^k\right\}$, called the *model*, of candidate distributions with the idea that the true distribution of the data is in $\mathcal{P}$. Under certain regularity conditions, we can view $\mathcal{P}$ as a k-dimensional manifold parametrized by the map

$$\Theta \ni \theta \mapsto \alpha(\theta) = p_\theta \in \mathcal{P}.$$

Moreover, we can specify a metric on this manifold that compares how distinguishable nearby points—i.e., probability distributions—are from each other based on the data. This is called the Fisher-Rao metric:

$$g_\theta = \left(E_\theta\left[\frac{\partial}{\partial\theta_i}\log p_\theta(X)\,\frac{\partial}{\partial\theta_j}\log p_\theta(X)\right]\right)_{ij} = I(\theta)$$

Notice that this is the Fisher information matrix. Whenever it is s.p.d., it will define an inner product on the tangent spaces $T_p\mathcal{P}, p\in\mathcal{P}$.

The Bayesian approach to statistical inference requires us so specify a probability distribution π on the model $\mathcal{P}$, called *the prior*, which describes our prior beliefs about the true distribution of the data. In the cases were we don't possess any accurate knowledge about the data distribution, or simply don't want to incorporate too many prior beliefs, we can choose a

prior that assigns the same density to every candidate distribution, i.e., the uniform distribution on the manifold $\mathcal{P}$.

$$\Pi(U) = \int_U dV, \quad U \subset \mathcal{P}$$

This choice of prior is called *Jeffreys' prior*. In coordinates, this prior can be expressed as

$$\alpha^*(dV) = \sqrt{\det g_\theta}\, d\theta_1 \dots d\theta_k$$

by a previous result.

We can apply the sampling algorithm to compute the marginal distribution

$$m(x) = \int_{\mathcal{P}} p(x)\, dV_p$$

under the Jeffreys' prior. In particular, if $p_1, \dots, p_n$ is a uniform sample from the model $\mathcal{P}$ then

$$m(x) \approx \frac{1}{n} \sum_{i=1}^{n} p_i(x)$$