1 Unbiased Sampling of Latitudes

1.0.1 Motivation

Nearly everyone has engaged in the fun pasttime of spinning a globe with their eyes closed and striking their fingertip down on a "random" point. Of course, there is little reason to believe that such a practice results in an unbiased sampling method, even if describing the exact nature of the bias isn't all that important. An technologically-updated version of "spinning the globe" might make use of random digits generated from some random source or oracle to represent latitude and longitude. For instance, one could take a random number from -180 to 179 to represent a randomly chosen meridian (line of longitude). Then, a second number from -90 to 90 could represent a randomly chosen latitude. But, if you think about it, you may realize that this approach has its flaws. In particular, while any given latitude is equally likely under this scheme, that is actually not a desirable property: it is biased towards the poles. That is to say, realizing that it is just as likely to choose a locale on the Arctic Circle as it is one on the Equator under this method, this is not something you would prefer to have happen, as the Equator is much longer then the Arctic Circle. The result, again, is a bias towards the poles—a bias entirely parallel to that seen in maps which utilize a projection that exaggerates the areas of regions closer to the poles, leading many to believe, for instance, that Greenland is approaching the size of Australia, when, in fact, the latter is over 3.5 times as big. But choosing a random locale is fun! So, let's fix our method.

1.0.2 Heuristic

First of all, note that picking a random longitude is not inherently biased: after all, each meridian, or line of longitude, can be regarded as equal in length to all others. (One might certainly complain about questions of topography and also the fact that the Earth is not perfectly spherical, but we will regard it as a sphere for the purpose of choosing latitude and longitude.) So, the question is really how to choose lines of latitude in such a way that eliminates the bias described above.

As an initial heuristic, let's imagine that we are going to use some random source of digits. A nifty strategy is actually just to use an oracle that will give you randomly chosen values between 0 and 1 (out to some specified precision, which we can imagine to be very fine). A value, say, $p \in [0, 1]$ can be interpreted as a probability. Taking the set of all possible lines of latitude (down to some certain precision, which we can worry about later), there is (by tautology) a one-to-one correspondence between these lines of latitude taken in order from the south pole to the north pole on the one hand and the probability of choosing it (using an arbitrary method) or one lower in latitide. (We will use negatives for degrees South of the Equator.) Thus, what we need is some function P which will give us $P(\Phi \leq \phi)$ for any given position of latitude *phi*. Note that Φ symbolizes a formal random variable (I am assuming a familiarity with this concept and this sort of notation), namely position of latitide *as a variable* (rather than referring to a specific value of latitude, for which I use the lower case, ϕ.) If we imagine that we will use a large number of possible lines of latitude for our sample space, then, continuing the heuristic thinking, we can try a continuous model for this, P, our cumulative probability function. Thus, in all, we consider the notation, the probability of choosing ϕ or a lower line of latitude might be expressed as follows:

$$p = P(\Phi \leq \phi) = \int_{-\frac{\pi}{2}}^{\phi} f\left(\phi'\right)\, d\phi'$$

As you can see, I have chosen to measure latitude in radians rather than degrees in anticipation of doing some calculus with it. (The derivative rules for trig functions only work as advertised when the independent variable is measured in radians.) Note, too, that ϕ' is just the variable of integration; I couldn't use ϕ since it was already taken.

So, what I have done with this notation is to narrow the focus of our need on a new item. We can instead focus on the *probability densities* of possible latitudes. The function f appearing in the equation above would indeed be termed a probability density function, or pdf, for short. That is, what we need is a probability density function that fixes the problem of overrepresentation of higher latitudes—that is, those near the poles—(or underrepresentation of middle latitudes). But our description of the problem contains within it already the germ of the solution: Our concern is that higher latitudes represent smaller distances—let's call them circumferences to be more suggestive—than the middle latitudes. What we want is for the probability of choosing any given latitiude, then, is just for it to be proprtional to the circumferntial distance occupied by that line. That's no problem at all! Doing just a little trignonometry (staring at a picture that helps one to understand the definition of latitude) you will hopefully agree that the circumferential distance d occupied by line of latitude at position ϕ is

$$d = 2\pi R \cos(\phi)$$

where R is the radius of the Earth in whatever units you want to choose for that. We want the probability of choosing any given latitude—ideally, choosing one among very many, that is, at high resolution—to be at least approximately proportional to the circumference a that latitude. Thus, we can disregard all the constants there for the moment and write

$$f(\phi) = k\cos(\phi)$$

Further confirmation of this notion can be obtained by observing that, for any given *phi*,

$$\begin{aligned} P(\Phi = \phi) &\approx \int_{\phi}^{\phi+\Delta\phi} f\left(\phi'\right)\, d\phi' \\ &\approx f(\phi)\Delta\phi \end{aligned}$$

assuming, say, that f is a continuous function. (Basically, just regard f as approximately constant over that very small interval of length $\Delta\phi$. So, in particular, note that we are regarding $\Delta\phi$ as some very small increment of latitude—say, one small enough to only encompass one of our very many possible candidates for random choice. The important takeaway here is that, since,

$$P(\Phi = \phi) \approx f(\phi)\Delta\phi,$$

we see that the probability of choosing ϕ is proportional to $f(\phi)$; thus, if we make sure that $f(\phi) = k\cos(\phi)$ for some ϕ, for some k, then we should be in good shape. (A concern may occur to you as to whether we intend to exert control over $\Delta\phi$—that is, to use the same spacing between possible latitudes. In fact, it would be more convenient to use uniform increments between our possible values of p, which would leave the uniformity of $\Delta\phi$ in doubt. But, this only a heuristic. We shall verify that the method that this thinking suggests is indeed sound—in the sense that we will check that different regions of same *area* chosen using this method will be chosen with equal likelihood.)

Now, let's just find what k ought to be. We compute:

$$\begin{aligned} 1 &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} k\cos\left(\phi'\right)\, d\phi' \\ &= k\sin\left(\phi'\right)\Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \\ &= k\left(\sin\left(\frac{\pi}{2}\right) - \sin\left(-\frac{\pi}{2}\right)\right) \\ &= 2k \end{aligned}$$

Thus, we get that $k = \frac{1}{2}$.

Thus, in all, we suggest that

$$\begin{aligned} p = P(\Phi = \phi) &\approx \int_{-\frac{\phi}{2}}^{\phi} \frac{1}{2}\cos\left(\phi'\right)\, d\phi' \\ &= \frac{1}{2}\sin\left(\phi'\right)\Big|_{-\frac{\pi}{2}}^{\phi} \\ &= \frac{1}{2}\big(\sin(\phi) + 1\big) \end{aligned}$$

Solving, we obtain,

$$\phi(p) = \sin^{-1}(2p - 1)\,, p \in [0, 1]$$

Thus, for instance, if our random number generator gives us $p = 0$, then we choose

$$\phi = \sin^{-1}(-1) = -\frac{\pi}{2}$$

which we have interpreted as the South pole (a conversion from radians to degrees, etc., should now take place.) More plausibly (in a rough sense), if our random number generator gives us $p = 0.123456789$, then

$$\phi = \sin^{-1}(2 \cdot 0.123456789 - 1) \approx -0.852740737864701$$

Evidently, this is in the Southern Hemisphere, and, should we convert to degrees, we find that this represents 48.85844529851904 degrees South. Now, we might as well choose a random longitude. Asking `random.org`, I obtained 0.4296521693. Converted to degrees East (noting that is less than 0.5), I obtained 25.325219051999994 degrees East. In all, here is the location that was chosen.

1.0.3 Verification

Next, let's verify that this sampling scheme has the desired property that motivated it. Namely, let's check to see if "circumferential" regions of the same area have the same probability of being chosen. That is, we can regard any latitude randomly chosen via this method as the representative of a thin band of the Earth consisting of that latitude and a little more due North of it, encircling the globe at that small interval of latitudes.

Suppose that random decimal r is chosen from $[0,1]$ with some high precision. Specifically, if we allow for n many possible values of r to be chosen, then let's write $\Delta r = \frac{1}{n}$. Of course, this Δr thus becomes the probability p of choosing any given latitude, $\phi = \phi(r) = \sin^{-1}(2r-1)$. In this section, we will recap an approximation of this probability (albeit from a reverse perspective), as well as derive an approximation for the area represented between latitudes ϕ and $\phi + \Delta\phi$, where $\Delta\phi = \phi(r+\Delta r) - \phi(r)$. our goal is to show that this approximate probability is proportional to the approximate area.

Writing $\phi = \phi(r)$, we get

$$\begin{aligned} \phi + \Delta\phi &= \phi(r+\Delta r) \\ &= \sin^{-1}(2(r+\Delta r)-1) \end{aligned}$$

From this, we obtain

$$\sin(\phi+\Delta\phi) = 2r - 1 + 2\Delta r$$

Writing $\sin(\phi) = 2r-1$, we thus get

$$\begin{aligned} & \sin(\phi+\Delta\phi) = \sin(\phi) + 2\Delta r \\ \Rightarrow\quad & \sin(\phi+\Delta\phi) - \sin(\phi) = 2\Delta r \\ \Rightarrow\quad & \frac{\sin(\phi+\Delta\phi)-\sin(\phi)}{\Delta\phi} = \frac{2\Delta r}{\Delta\phi} \end{aligned}$$

Now, it is intuitively clear (but also not hard to show) that $\Delta\phi$ gets smaller as Δr gets smaller. (This might be a nice exercise for you to demonstrate.) Thus, assuming we are choosing r from very many (uniformly spaced) choices, we are justified in adopting the approximation

$$\frac{\sin(\phi+\Delta\phi)-\sin(\phi)}{\Delta\phi} \approx \phi$$

Thus,

$$\cos(\phi) \approx \frac{2\Delta r}{\Delta\phi}$$

Solving, we obtain an approximation for the probability p of choosing ϕ:

$$p = \Delta r \approx \frac{1}{2}\cos(x)\Delta\phi$$

We will next show that the approximate area encircling the globe between latitude ϕ and $\phi + \Delta\phi$ is proprtional to this. We will use "cone frustums" to approximate these areas. (We expect it is intuitively clear that these will do a good job at approximating the true areas, but, as an exercise, a verification of this fact will be offered.) To that end, let $A(\phi)$ represent the area of the cone with base, the cross-sectional area of the Earth running through the Equator and tangent to the Earth at latitude ϕ, with its vertex on the Earth's rotational axis. [Pic, please?] The area of the frustum approximating the true area between latitudes ϕ and $\phi + \Delta\phi$ will this be $\Delta A = A(\phi + \Delta\phi) - A(\phi)$ in the Southern Hemisphere and $-\Delta A = A(\phi + \Delta\phi) - A(\phi)$ in the Northern. To that end, it is helpful to know that the (outer) surface area of a cone with base radius r and slant height s is

$$A = \pi r s$$

Noting that phi determines each such cone uniquely, we write s for the slant height and r (not to be confused with our use of r above, which does not come into the present analysis) for the base radius associated with phi. Likewise, $s + \Delta s$ and $r + \Delta r$ are those associated with $\phi + \Delta\phi$. Thus, we have that

$$\begin{aligned}\Delta A &= \pi\left(r + \Delta r\right)\left(r + \Delta r\right) - \pi r s \\ &= \pi\left(r\Delta s + s\Delta r\right)\end{aligned}$$

Now, while the values r, s, Δr, and Δs are all determined by ϕ and $Delta\phi$, it remains to see exactly how. A diagram would be pretty helpful here! Using R for the radius of the Earth,

$$r = R\cos(\phi)$$

while

$$s = R\cot(\phi)$$

in the Northern Hemisphere. From here on, we will only consider the case of the Northern Hemisphere, accepting that the approximations made in the Souther Hemisphere will be congruent for corresponding latitudes. Also, observe (using a picure!) that

$$\Delta s = \csc(\phi)\Delta r$$

Finally, we use an approximation (based again on the assumed smallness of $\Delta\phi$:

$$\Delta r = R\left(cos\left(\phi + \Delta\phi\right) - \cos(\phi)\right) \approx -R\sin(\phi)\Delta\phi$$

Putting these to immediate work, we obtain,

$$\begin{aligned}\Delta A &= \pi\left(r\Delta s + s\Delta r\right)\\ &= \pi\left(r\csc(\phi)\Delta r + s\Delta r\right)\\ &= -\pi R\sin(\phi)\Delta\phi\left(r\csc(\phi) + s\right)\\ &= -\pi R\sin(\phi)\Delta\phi\left(R\cos(\phi)\csc(\phi) + R\cot(\phi)\right)\\ &= -2\pi R^2\cos(\phi)\Delta\phi\end{aligned}$$

Thus, we see that the area of this frustum located above ϕ is indeed proportional to the probability of choosing ϕ. Since this frustum approximates the true area enircling the globe at latitude ϕ, we are satisfied that our sampling procedure has ameliorated the problem of the bias of sampling latitudes at uniform widths. Indeed!—note that the formula

$$\Delta\phi = \sin^{-1}\left(2(r+\Delta r) - 1\right) - \sin^{-1}\left(2r - 1\right)$$

reveals that, when the values $r \in [0,1]$ are chosen uniformly, the values ϕ will not be.

1.0.4 Exercises.

1. Let $dA = 2\pi R^2\cos(\phi)d\phi$ and integrate $\int_{-\pi/2}^{\pi/2} dA$. Interpret your results.