Stats100A **Summer 2024**

Week 3: Extra Problems - Solutions

Author: Andrew Lizarraga

Solutions

3.1 Basic Examples

Problem 1 (Determining Coin Bias): I give you a coin and I want you to determine if it is biased or not. What can you do to try and determine the coins bias?

Solution: The simplest thing we can do is flip the coin many times. Ideally if the coin is heavily biased (let's say to heads), then we should see the majority of the coin flips coming up heads.

Remark: If the coin is only slightly biased in favor of heads, let's say a 51% chance of heads, we would need to run many simulations of coin flips in order to determine the bias. What makes this trickier is that if the coin is fair, there is always going to be variation in the coin flips in say 1000 coin flips. **Can you think of how we can resolve this?**

Problem 2 (Darts In One-Half): I have a unit square $[0, 1]^2$ and I toss 10 darts in the unit square. How many darts would you expect to land in the left-half of the square?

Solution: Firstly, we express this problem as sampling a point $(X, Y) \sim U[0, 1]^2$, or equivalently $X, Y \sim^{ind} U[0, 1]$. Solving the problem corresponds to the probability: $P(X \leq \frac{1}{2})$. Since the area of the entire square is 1, and $X \leq \frac{1}{2}$ corresponds to half of the square (which has area $\frac{1}{2}$), then we have a $\frac{1}{2}$ chance a dart lands in the left-half square, hence we expect *about* 5 of the darts to land in the left-half.

Problem 3 (Measuring a Cancer Cell): I have a $2D$ slice of a cell on a plane and I want to measure a cross-sectional area of interest, let's call it $|A|$. The issue is that $|A|$ is highly irregular and my microscopes are too weak to view it. However, I can shoot X-rays at the cell and the area of interest give a signal response that I can measure. What can I do with this setup to measure $|A|$, given I know that the cells entire cross-sectional area is $|\Omega|$?

Solution: I can emit many X-rays uniformly at random on the entire cell which has area $|\Omega|$. Let's say I do this n times and I get m signal responses back indicating an X-ray hit area $|A|$. Then we'd expect for large n that $\frac{m}{n} \to \frac{|A|}{|\Omega|}$.

3.2 Tricky To Deduce Probabilistic Phenomena

In the following problems provide and experimental setup that will allow you to deduce an approximate answer for otherwise very tricky and counterintuitive probabilistic problems.

Problem 4 (HH v.s. HT): I have a fair coin and I keep flipping until I see HH or HT. Should it take more flips on average until I see HH, or more flips for HT, or should it be the same?

Solution: There are clever and clean ways to deduce that the average number of flips until we see HH is 6 and the average until we see HT is 4. People are often surprised to learn, that they are unequal, you would think that HH should have equivalent probability to HT, so why would one take longer? **I leave this for the reader to figure out**.

However in the meantime, let's think instead of how we can use a Monte Carlo approach to deduce the solution. One way to do this is to run an experiment. Let's say we run an experiment N times: In each trial we keep flipping the coin until we see HH and keep record in each trial how many flips it took. So for trial 1, we record n_1 flips, for trial 2, we record n_2 flips and we continue in this fashion until we reach trial N and record n_N flips. Then we suspect that $\frac{n_1+n_2+\cdots+n_N}{N}$ to be close to $\mathbb{E}(HH)$ for large N. We can run a similar experiment for $\mathbb{E}[HT]$.

Remark: One can confirm (and I highly suggest you do), that running this simulation gives us $\mathbb{E}[HH] \approx 6$ and $\mathbb{E}[HT] \approx 4$.

Problem 5 (Jane Street's Some Off Square): A circle is randomly generated by sampling two points uniformly and independently from the interior of a square and using these points to determine its diameter. Approximate the probability that the circle has a part of it that is off the square?

Solution: Again, there is a clever way to setup and integral and get the exact answer. However, let's think about an experimental setup: Let's say the the square is $[0,1]^2$, then we sample two points $(X_1, Y_1), (X_2, Y_2) \sim^{ind} U[0,1]^2$. These two points then form the diameter of the circle. If the circle intersects the square at any point, then we count this. Now suppose that we run this simulation n times and count the circles intersecting the square m times then we expect as n gets large that the fraction $\frac{m}{n}$ approaches the true probability.

3.3 Estimators

In the following problems, we will deduce some famous estimates of π.

Problem 6 (Buffon's Needle): We have parallel lines that are separated 1-unit away from each other and they span the entire plane. We toss a needle of length 1 onto the plane, what is the probability that the needle crosses a line?

Solution: The key observation is determining a good representation of the needle. Before we do that let's make a couple of observations:

1. With probability 1 the center of the needle must land in some parallel strip. So it's enough to focus on a single strip **(Why?)**.
2. More so, within the strip that contains the center of the needle, we also know with probability 1 that the needle must be on one-half of the strip **(Why?)**.

So it's enough to solve the problem in this case. We denote the distance from the center of the needle to the nearest parallel line by x, and denote the acute angle formed by the needle and the orthogonal projection of the center of the needle and the parallel line, by θ.

So the position of the needle is completely captured by (x, θ) where $x \sim U[0, \frac{1}{2}]$ and $\theta \sim U[0, \frac{\pi}{2}]$. So we may think of the sample space as $\Omega \equiv [0, \frac{1}{2}] \times [0, \frac{\pi}{2}]$. Now we must determine what region corresponds to the needle crossing the parallel line. Let's call this region A. Once we find this region, then we can determine the probability of a needle crossing a parallel line by computing $\frac{|A|}{|\Omega|}$.

Notice that we can relate x, θ, and hypotenuse (half of the needle) via $\cos\theta = \frac{x}{\frac{1}{2}} = 2x$. Thus $x = \frac{\cos\theta}{2}$. Notice that if $\theta = 0$, then the needle is orthogonal to the parallel line on the plane, so any x in the range from 0 up to $\frac{1}{2}$ indicate the needle is intersecting the parallel line. Also since $\frac{\cos\theta}{2}$ is a concave function, thus we can conclude that $x \leq \frac{\cos\theta}{2}$ is the region corresponding to the needle crossing.

As mentioned early, the probability corresponds to computing $\frac{|A|}{|\Omega|}$. Since $\Omega \equiv [0, \frac{1}{2}] \times [0, \frac{\pi}{2}]$, thus $|\Omega| = \frac{1}{2} \times \frac{\pi}{2} = \frac{\pi}{4}$. To compute $|A|$ corresponds computing the area under the curve $x = \frac{\cos\theta}{2}$ which is given by

$$|A| = \int_0^{\frac{\pi}{2}} \frac{\cos\theta}{2} d\theta = \frac{1}{2}$$

Therefore the probability of a needle crossing is $\frac{|A|}{|\Omega|} = \frac{2}{\pi}$.

Now to run a Monte Carlo Simulation, we can note the following. Toss n needles, and suppose that m of them cross a parallel line. Then by what we've shown, we expect that $\frac{m}{n} \to \frac{2}{\pi}$. This can be reexpressed as: $\frac{2n}{m} \to \pi$. So by computing the fraction $\frac{2n}{m}$, we are computing an approximation of π.

Problem 7 (Laplace's Needle): We have a set of parallel lines that are separated 1-unit away from each other. Additionally, we have another set of parallel lines, orthogonal to the first set of parallel lines, that are also spaced 1-unit apart. So the plane is effectively covered by a grid of unit squares. What is the probability that when tossing a needle that it crosses a line?

Solution: Consider the center of the needle. We know with probability 1 that it lands in some unit square. Now, we consider how to parameterize the needle. One intuitive parameterization is letting (x, y) denote the center of the needle. By a similar argument as the Buffon Needle problem, we need only focus on the sample space given by with $x \sim U[0, \frac{1}{2}]$ and $y \sim U[0, \frac{1}{2}]$. We can let θ be the acute angle formed by the needle and the x-axis, so $\theta \sim U[0, \frac{\pi}{2}]$.

So our sample space consists of points, (x, y, θ), belonging to the cube $\Omega \equiv [0, \frac{1}{2}] \times [0, \frac{1}{2}] \times [0, \frac{\pi}{2}]$. Notice that the volume of the cube is given by $|\Omega| = \frac{\pi}{8}$. So to determine the probability of a needle crossing a grid-line in Laplace's experiment, it suffices to compute the volume of the region $A \subset \Omega$ that corresponds to the needle crossing and then dividing this volume by the total volume, $|\Omega|$.

We let $I_A(x, y, \theta)$ denote the indicator function, which evaluates to 1 when the tuple (x, y, θ) corresponds to values indicating the needle has crossed a grid-line in the plane (otherwise it is 0). So the volume of A is given by $|A| = \iiint_\Omega I_A(x, y, \theta) dxdyd\theta$. Recall from Buffon's experiment that a condition for the needle to cross a parallel line with respect to x is given by $x \leq \frac{\cos\theta}{2}$. Similarly, by use of the complementary angle, we have that the needle crossing condition with respect to y is given by $y \leq \frac{\cos(\frac{\pi}{2} - \theta)}{2} = \frac{\sin\theta}{2}$. Keeping in mind that it's possible for the needle to cross orthogonal lines. Thus we may express the volume $|A|$ as follows:

$$\int_0^{\frac{\pi}{2}} \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} I\left(x \leq \frac{\cos\theta}{2}\right) + I\left(y \leq \frac{\sin\theta}{2}\right) - I\left(x \leq \frac{\cos\theta}{2}\right) I\left(y \leq \frac{\sin\theta}{2}\right) dxdyd\theta$$

where $I(\cdot)$ indicates on the region where the respective condition is true. We may solve the integral in it's three terms:

$$\begin{aligned}
&\int_0^{\frac{\pi}{2}} \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} I\left(x \leq \frac{\cos\theta}{2}\right) dxdyd\theta \\
&= \int_0^{\frac{\pi}{2}} \int_0^{\frac{1}{2}} \int_0^{\frac{\cos\theta}{2}} dxdyd\theta \\
&= \int_0^{\frac{\pi}{2}} \int_0^{\frac{1}{2}} \frac{\cos\theta}{2} dyd\theta \\
&= \int_0^{\frac{\pi}{2}} \frac{\cos\theta}{4} d\theta \\
&= \frac{1}{4}
\end{aligned}$$

Similarly,

$$\int_0^{\frac{\pi}{2}} \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} I\left(y \leq \frac{\sin\theta}{2}\right) dxdyd\theta$$

$$= \int_0^{\frac{\pi}{2}} \int_0^{\frac{\sin\theta}{2}} \int_0^{\frac{1}{2}} dxdyd\theta$$

$$= \int_0^{\frac{\pi}{2}} \int_0^{\frac{\sin\theta}{2}} \frac{1}{2} dyd\theta$$

$$= \int_0^{\frac{\pi}{2}} \frac{\sin\theta}{4} d\theta$$

$$= \frac{1}{4}$$

Lastly,

$$\int_0^{\frac{\pi}{2}} \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} -I\left(x \leq \frac{\cos\theta}{2}\right) I\left(y \leq \frac{\sin\theta}{2}\right) dxdyd\theta$$

$$= \int_0^{\frac{\pi}{2}} \int_0^{\frac{\sin\theta}{2}} \int_0^{\frac{\cos\theta}{2}} -1 \; dxdyd\theta$$

$$= \int_0^{\frac{\pi}{2}} \int_0^{\frac{\sin\theta}{2}} -\frac{\cos\theta}{2} dyd\theta$$

$$= \int_0^{\frac{\pi}{2}} -\frac{\cos\theta\sin\theta}{4} d\theta$$

$$= \int_0^{\frac{\pi}{2}} -\frac{\sin 2\theta}{8} d\theta$$

$$= -\frac{1}{8}$$

Therefore $|A| = \frac{1}{4} + \frac{1}{4} - \frac{1}{8} = \frac{3}{8}$. Now recall that $|\Omega| = \frac{\pi}{8}$, thus the probability of a needle intersecting is given by $\frac{|A|}{|\Omega|} = \frac{3}{\pi}$. Again, among n tosses if we count m the number of needle crosses, then we expect $\frac{m}{n} \to \frac{3}{\pi}$ as n gets large. If we ran Laplace's experiment, then we expect $\frac{3n}{m} \to \pi$.