5 Frequentist Inference

One of the main reasons we do statistics is to make inferences about a population given data from a subset of that population. For example, suppose there are two candidates running for office. We could be interested in finding out the true proportion of the population that supports a particular political candidate. Instead of asking every single person in the country their preferred candidate, we could randomly select a couple thousand people from across the country and record their preference. We could then estimate the true proportion of the population that supports the candidate using this sample proportion. Since each person can only prefer one of two candidates, we can model this person's preference as a coin flip with bias $p =$ the true proportion that favors candidate 1.

In the following two sections, we discuss two main types of estimation: *point* and *interval* estimation. In the above paragraph, we discussed estimating a number p, the true proportion of voters that favor candidate 1. This is called a **point estimate**, since we are using data to estimate a point on the real line, i.e. the true proportion of people who supported a particular candidate. An **interval estimate** uses the data to come up with an interval, within which we are fairly confident the true parameter p lies. For example, we could come up with a 95% confidence interval for p using the sample data. We would construct it in such a way that p lies within this subinterval of $[0, 1]$ with probability 0.95.

5.1 Point Estimation

5.1.1 Estimating the Bias of a Coin

Suppose now that we are again flipping a coin, this time with bias p. In other words, our coin can be thought of as a random quantity X defined

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

where 1 represents `H` and 0 represents `T`. If we were just handed this coin, and told that it has some bias $0 \leq p \leq 1$, how would we estimate p? One way would be to flip the coin n times, count the number of heads we flipped, and divide that number by n. Letting X_i be the outcome of the i^{th} flip, our estimate, denoted $\hat{p}$, would be

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

As the number of samples n gets bigger, we would expect $\hat{p}$ to get closer and closer to the true value of p.

5.1.2 Estimating π

In the website's visualization, darts are thrown uniformly at a square, and inside that square is an inscribed circle. If the side length of the square that inscribes the circle is L, then the radius of the circle is $R = \frac{L}{2}$, and its area is $A = \pi(\frac{L}{2})^2$. At the i^{th} dart throw, we can define

$$X_i = \begin{cases} 1 & \text{if the dart lands in the circle} \\ 0 & \text{otherwise} \end{cases}$$

The event "dart lands in the circle" has probability

$$p = \frac{\text{Area of Circle}}{\text{Area of Square}} = \frac{\pi\left(\frac{L}{2}\right)^2}{L^2} = \frac{\pi}{4}$$

So with probability $p = \frac{\pi}{4}$, a dart lands in the circle, and with probability $1 - \frac{\pi}{4}$, it doesn't.

By the previous section, we can estimate p using $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$ so that for large enough n, we have

$$\hat{p} \approx p = \frac{\pi}{4}$$

so that rearranging for π yields

$$\pi \approx 4\hat{p}$$

Hence as the number of samples n goes to ∞, the estimator $\hat{p}$ approaches p, and thus $4\hat{p} \to 4p = \pi$. This means that if we continue throwing darts at the board, the value $4\hat{p}$ will get closer and closer to π.

5.1.3 Consistency

What exactly do we mean by "closer and closer"? In this section, we describe the concept of **consistency** in order to make precise this notion of convergence. Our estimator in the last section, $4\hat{p}$ is itself random, since it depends on the n sample points we used to compute it. If we were to take a different set of n sample points, we would likely get a different estimate. Despite this randomness, intuitively we believe that as the number of samples n tends to infinity, the estimator $4\hat{p}$ will converge, at least in some probabilistic sense, to π.

Another way to formulate this is to say, no matter how small a number we pick, say 0.001, we should always be able to conclude that the probability that our estimate differs from π by more than 0.001, goes to 0 as the number of samples goes to infinity. We chose 0.001 in this example, but this notion of probabilistic convergence should hold for any positive number, no matter how small. This leads us to the following definition.

Definition 5.1. *We say an estimator $\hat{p}$ is a* ***consistent*** *estimator of p if for any $\varepsilon > 0$,*

$$\lim_{n\to\infty} P(|\hat{p} - p| > \varepsilon) = 0.$$

Let's show that $4\hat{p}$ is a *consistent* estimator of π.

Proof. Choose any $\varepsilon > 0$. By Chebyshev's inequality (Corollary 2.13),

$$\begin{aligned}
P(|4\hat{p} - \pi| > \varepsilon) &\le \frac{\mathrm{Var}(4\hat{p})}{\varepsilon^2} \\
&= \frac{\mathrm{Var}\Big(4 \cdot \frac{1}{n}\sum_{i=1}^{n} X_i\Big)}{\varepsilon^2} && \text{(Definition of } \hat{p}\text{)} \\
&= \frac{\frac{16}{n^2}\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big)}{\varepsilon^2} && (\mathrm{Var}(cY) = c^2\mathrm{Var}(Y)) \\
&= \frac{\frac{16}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i)}{\varepsilon^2} && (X_i\text{'s are independent}) \\
&= \frac{\frac{16}{n^2}\cdot n \cdot \mathrm{Var}(X_1)}{\varepsilon^2} && (X_i\text{'s are identically distributed}) \\
&= \frac{\frac{16}{n}\cdot p(1-p)}{\varepsilon^2} && (\mathrm{Var}(X_i) = p(1-p)) \\
&= \frac{16 \cdot \frac{\pi}{4}\Big(1 - \frac{\pi}{4}\Big)}{n\varepsilon^2} && (p = \frac{\pi}{4}) \\
&\to 0
\end{aligned}$$

as $n \to \infty$. Hence we have shown that $4\hat{p}$ is a consistent estimator of π. $\square$