Inferential statistics

Confidence intervals.

Let us suppose that we have i.i.d. random variables $X_1, \dots, X_n$ each with mean μ and standard deviation $\sigma > 0$. Let $\overline{X}$ be the average, $(X_1 + \cdots + X_n)/n$.

Let Z be the standardization of $\overline{X}$,

$$Z = \frac{\overline{X} - \mu}{(\sigma/\sqrt{n})} \ .$$

Suppose c is a positive number. Then each of the following three conditions is equivalent to the others (hence all four events have the same probability):

$$\begin{aligned}
|Z| &< c \\
\left|\frac{\overline{X} - \mu}{(\sigma/\sqrt{n})}\right| &< c \\
\mu - c(\sigma/\sqrt{n}) &< \overline{X} < \ \mu + c(\sigma/\sqrt{n}) \\
\overline{X} - c(\sigma/\sqrt{n}) &< \mu \ < \ \overline{X} + c(\sigma/\sqrt{n}) \ .
\end{aligned}$$

Let us suppose, for an example, that $c = 2.57$, and n is large, so that by the Central Limit Theorem we approximate the distribution of Z to be $\mathcal{N}(0,1)$. Then $\text{Prob}(|Z| < 2.57) = .99$. Then the interval

$$\left(\overline{X} - 2.57\frac{\sigma}{\sqrt{n}}, \ \overline{X} + 2.57\frac{\sigma}{\sqrt{n}}\right)$$

is called a 99% confidence interval for μ. We interpret this to mean the following: if we repeat the experiment of computing $\overline{X}$ many times, then for about 99 of those times, the computed confidence interval will contain μ.

EXAMPLE. Suppose $X_1, \dots, X_{100}$ are i.i.d random variables which have uniform distribution on $[a - 1/2, a + 1/2]$, where a is unknown. Suppose the random sample produces sample mean equal to 5. Compute a 99% confidence interval for a.

SOLUTION. Each X_i has mean a and variance 1/12. Putting this into the display above with $n = 100$, we get the 99% confidence interval

$$\left(5 - 2.57\frac{\sqrt{1/12}}{\sqrt{100}}, \ 5 + 2.57\frac{\sqrt{1/12}}{\sqrt{100}}\right) \ \approx \ (4.26, 5.74) \ .$$

For a 95% confidence interval, above we'd use 1.96 in place of 2.57.

CLT and standard deviation.

In the example above, using the CLT to get a confidence interval for μ, we had i.i.d. random variables with know standard deviation, and we used that for the computation. Usually, you don't know that standard deviation, and have to estimate it. There are some subtleties to this; we'll just look at two important ways to do it.

Population proportion.

Suppose, as with flipping coins, that the i.i.d. rv's $X_1, \dots, X_n$ equal 1 with probabilty p and equal 0 with probability $(1-p)$.

Suppose we want a confidence interval for p.

Here for each X_k, the mean is p and the standard deviation is $\sqrt{p(1-p)}$. Write $\overline{X}$ as $\widehat{p}$ (the "sample proportion") to emphasize its relation to p. We use $\widehat{p}$ to estimate both the mean and the standard deviation– just substitute the experimental result $\widehat{p}$ for p, giving a 99% confidence interval

$$\left(\widehat{p} - 2.57\frac{\sqrt{\widehat{p}(1-\widehat{p})}}{\sqrt{n}},\ \widehat{p} + 2.57\frac{\sqrt{\widehat{p}(1-\widehat{p})}}{\sqrt{n}}\right)$$

EXAMPLE. 144 voters are randomly surveyed and asked if they will vote for Trump. 48 of those surveyed answer yes. Give a 95% confidence interval for the proportion of voters who would vote for Trump.

ANSWER. We have $\sqrt{n} = 12$, $\widehat{p} = 48/144 = 1/3$, $\sqrt{(\widehat{p})(1-\widehat{p})} = \sqrt{(1/3)(2/3)} \approx .47$. So the confidence interval is

$$\left(.33 - (1.96)\frac{.47}{12},\ .33 + (1.96)\frac{.47}{12}\right) \ \approx\ (.25, .41)$$

Polling is hard!

Sample standard deviation.

Given our i.i.d. rv's $X_1, \dots, X_n$: we assume they have mean μ and positive variance σ^2. For the confidence interval, how should we estimate σ? The natural thing would be to estimate σ^2 and take a square root. If we knew the number μ, then we could estimate σ^2 by $\Big((X_1-\mu)^2+\cdots(X_n-\mu)^2\Big)/n$. Since we don't have μ, we could estimate by replacing μ with $\overline{X}$ in this expression. For reasons we won't go into, a better estimate is

$$s^2 = \frac{\Big((X_1-\overline{X})^2+\cdots(X_n-\overline{X})^2\Big)}{n-1} .$$

(but note, dividing by $n-1$ rather than n has little effect when n is large). The number s^2 (or the random variable producing it) is called the *sample variance*, and its square root s is the *sample standard deviation.*

To get a confidence interval for the mean for a large sample, just use s in place of σ in the formula.

EXAMPLE. Suppose in a random sample of 400 undergraduate men at UMD that the average best time for running a mile is 6 minutes, and the sample standard deviation is 1.2 minute. Compute a 95% confidence interval for the average best time for running a mile for UMD undergraduate men.

ANSWER. The 95% confidence interval (in minutes) is

$$\begin{aligned} &= \left(\overline{X} - 1.96\,\frac{s}{\sqrt{n}},\ \overline{X} + 1.96\,\frac{s}{\sqrt{n}}\right) \\ &= \left(6 - (1.96)\,\frac{1.2}{\sqrt{400}},\ 6 + (1.96)\,\frac{1.2}{\sqrt{400}}\right) \\ &\approx\ (5.88, 6.12)\ . \end{aligned}$$

Hypothesis Testing

To see the basic idea of hypothesis testing, we'll first consider an example.

Example.

You are looking for strong evidence that a certain coin is not fair (probability of heads is not 1/2). So, you set up a probability model ASSUMING THAT THE COIN **IS** FAIR, and run an experiment. If the result is sufficient unlikely (assuming the coin is fair), then you have your evidence.

Your experiment will be to flip the coin 10,000 times, and count the number of heads. Formally, let $X(k) = 1$ if the kth flip is heads; $X(k) = 0$ otherwise. Assume the X_k are i.i.d. with 1 and 0 each having probability $p = 1/2$. Each X_k has standard deviation $\sqrt{p(1-p} = 1/2$. Let $S = X_1 + \cdots + X_{10,000}$. (So, S has distribution Binomial(n, p) with $n = 10,000$ and $p = 1/2$.) Then S has mean $\mu = 5000$ and standard deviation $\sqrt{n}\sigma = 50$.

Let $Z = (S - 5000)/50$; the distribution of Z in the model is very close to $\mathcal{N}(0,1)$ (n is large!).

You decide to reject the "fair coin" hypothesis if you get a sample mean $\overline{X}$ with $|\overline{X}| > c$, where c is the number such that $\big(\text{Prob}|Z| \geq c\big) = .01$. Looking at the tables, determine that c here is approximately 2.6 .
Note, $|Z| > 2.6$ is equivalent to $|S - 5000| > (2.6)(50) = 130$.

Now you do the experiment.

(i) Suppose the outcome is 5100 heads. You may remain suspicious, but you don't abandon the fair coin hypothesis. Put another way, the evidence for unfairness doesn't meet the standard you set.
(ii) Suppose the outcome is 5200 heads. You decide the evidence is good enough to declare the coin unfair. Moreover, $\big(\text{Prob}|S| \geq 5200\big) = \big(\text{Prob}|Z| \geq 4\big) \approx .000\,063$: you would have made the same decision even if you'd de-

manded evidence with .000 063 in place of .01. (Here .000 063 is called the "P-value" of the result.) Very strong evidence!

We considered just the simple example, but this is how hypothesis testing is done in general (with a variety of probability distributions and tests). If the observed result is very unlikely given the assumed model, then that is strong evidence to reject the model. You set up "the model" so that rejecting it is what you are getting evidence for.

Note, when you decide on the cutoff probability for your test (we picked .01), you can make two kinds of mistake: keeping an incorrect assumption of fair coin, or drawing an incorrect conclusion of unfair coin. The cutoff choice represents a tradeoff.

Sounds convoluted, but it makes sense.