Vegas Solutions

written by Alvin Wan . Spr '16 Dis 12a . alvinwan.com/cs70

Thursday, March 3, 2016

On the planet Vegas, everyone carries a coin. Many people are honest and carry a fair coin (heads on one side and tails on the other), but a fraction p of them cheat and carry a trick coin with heads on both sides. You want to estimate p with the following experiment: you pick a random sample of n people and ask each one to flip his or her coin. Assume that each person is independently likely to carry a fair or a trick coin.

1. Given the results of your experiment, how should you estimate p?

   Solution: We are looking for $\tilde{p}$, the fraction of people with trick coins, so let us begin by assuming that the fraction of people with trick coins is $\tilde{p}$.

   Let $\tilde{q}$ be the fraction of people we *observe* with heads, in terms of $\tilde{p}$.

   $$\tilde{q} = (1)\tilde{p} + \frac{1}{2}(1 - \tilde{p})$$

   This implies that

   $$2\tilde{q} = 2\tilde{p} + (1 - \tilde{p})$$
   $$\tilde{p} = 2\tilde{q} - 1$$

   Note that $\tilde{p}$ is the fraction of people we *think* have trick coins. This is different from p, which is the *actual* fraction of people with trick coins. To express the *actual* p in terms of our *actual* q, we rewrite the tilde expressions. Note that this is in theory dependent on the people we sample, so this is only an approximate equality that should be true as we approach an infinite number of samples.

   $$p \approx 2q - 1$$

2. How many people do you need to ask to be 95% sure that your answer is off by at most 0.05?

   Solution:

   We are looking for the difference between $\tilde{p}$ and p to be less than 0.05 with probability 95%. We first note that Chebyshev's inequality naturally follows, as Chebyshev's helps us find distance from the mean with a certain probability. Formally, this is Chebyshev's:

$$Pr[|X - \mu| \geq a] \leq \frac{var(X)}{a^2}$$

However, we are interested in finding an n so that we are off by *at most* 0.05 with probability *95%*. This is equivalent to being off by *at least* 0.05 with probability *5%*. The latter is answerable by Chebyshev's.

Then, we follow three steps.

Step 1 : Fit to $|X - \mu| \geq a$

We first only deal with "your answer is off by at most 0.05". We can re-express this mathematically, with the following:

$$|\tilde{p} - p| < 0.05$$

We don't have $\tilde{p}$, however, so we plug in q for our $\tilde{p}$.

$$|(2\tilde{q} - 1) - (2q - 1)| \leq 0.05$$
$$|2\tilde{q} - 2q| \leq 0.05$$
$$|\tilde{q} - q| \leq 0.025$$

First, note that with infinitely many samples, the fraction $\tilde{q}$ should naturally converge to become the fraction q.

$$\mu_{\tilde{q}} = q$$

We can thus transform this to something closer to our form!

$$|\tilde{q} - \mu_{\tilde{q}}| \leq 0.025$$

.

However, we need to incorporate the number of people we are sampling. So, we multiply all by n.

$$|\tilde{q}n - \mu_{\tilde{q}}n| \leq 0.025n$$

Let us consider again: what is $\tilde{q}$? We know that $\tilde{q}$ was previously defined to be the fraction of people that we *observe* to have heads. We are inherently asking for the number of heads in n trials. In other words, we want k successes among n trials, so this sounds calls for a Bernoulli random variable! We will define X_i to be 1 if the ith person tells us heads. This makes

$$\tilde{q} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

To make our life easier, let us define another random variable $Y = \tilde{q}n$.

$$Y = \tilde{q}n = \sum_{i=1}^{n} X_i$$

.

Seeing that this now matches the format we need, our α is 0.025. Our final form is

$$Pr[|Y - qn| \leq 0.025n] \geq \frac{var(Y)}{(n0.025)^2}$$

Step 2 : Compute $\frac{var(Y)}{a^2}$

We first compute $var(X_i)$.

$$\begin{aligned} var(X_i) &= E[X_i^2] - E[X_i]^2 \\ &= q - q^2 \\ &= q(1-q) \end{aligned}$$

We then compute $var(Y)$.

$$\begin{aligned} var(Y) &= var(\sum_{i=1}^{n} X_i) \\ &= \sum_{i=1}^{n} var(X_i) \\ &= nvar(X_i) \\ &= nq(q-1) \end{aligned}$$

Thus, we have the value of our right-hand-side.

$$\begin{aligned} \frac{var(Y)}{a^2} &= \frac{nq(1-q)}{(n0.025)^2} \\ &= \frac{q(1-q)}{n(0.025)^2} \end{aligned}$$

Step 3 : Compute Bound

We now consider the remainder of our question: "How many people do you need to ask to be 95% sure...". Per the first paragraph right before step 1, we are actually interested in the probability of 5%. Thus, we want the following.

$$\frac{var(Y)}{a^2} = \frac{q(1-q)}{n(0.025)^2} = 0.05$$

.

We have an issue however: there are two variables, and we don't know q. However, we can upper bound the quantity $q(1-q)$. Since Chebyshev's computes an upper bound for the probability, we can substitute $q(1-q)$ for its maximum value.

$$q(1-q) = q^2 - q$$

To find it's maximum, we take the derivative and set equal to 0.

$$q' = 2q - 1 = 0$$
$$q = \frac{1}{2}$$

This means that $q(1-q)$ is maximized at $q = \frac{1}{2}$, making the maximum value for $q(1-q)$, $\frac{1}{2}(1-\frac{1}{2}) = \frac{1}{2}(\frac{1}{2}) = \frac{1}{4}$. We now plug in $\frac{1}{4}$.

$$\frac{q(1-q)}{n(0.025)^2} = \frac{1/4}{n(0.025)^2} = 0.05$$
$$\frac{1}{4n(0.025)^2} = \frac{1}{20}$$
$$\frac{5}{(0.025)^2} = n$$
$$n = 8000$$