Quiz 16 : Inequalities Solutions

written by Alvin Wan . alvinwan.com/cs70 . Wednesday, November 2, 2016

This quiz does not count towards your grade. It exists to simply gauge your understanding. Treat this as though it were a portion of your midterm or final exam.

1 Inequalities

1. (**True** or **False**) For a δ confidence, of an estimate being ϵ distance from the mean, we can take $n \geq \frac{1}{4\epsilon\sigma^2}$ for any set of i.i.d. random variables.

   Solution: False. The i.i.d. random variables must be Bernoulli or, in other words, indicator variables. Recall that to derive $\frac{1}{4\epsilon^2\sigma}$, we take $o^2 = p(1-p)$, which is only true for Bernoulli random variables. The more general form is $\frac{\sigma^2}{\epsilon^2\sigma}$.

2. For any $n > 0$, when is $p^n(1-p)^n$ maximized? (Hint: No calculus is needed.)

   Solution: When $p = \frac{1}{2}$, as then, we have $p^n(1-p)^n = \frac{1}{4^n}$. Any other p will result in a smaller fraction.

2 Confidence Intervals

Consider a game of rock, paper, scissors between Sinho and Forest. With probability p, Sinho can predict Forest's move, and with probability $1-2p$, Forest can predict Sinho's move. If both predict the other, they will tie and neither wins. If only one player predicts successfully, that player wins. As it turns out, without this predictive power, Sinho and Forest are equally likely to win.

1. How can we estimate p?

   Solution: We wish to find the estimate for p. However, we cannot "observe" this probability directly. Instead, we can only observe the number of games that Sinho wins. So, let us consider q to be the fraction of games that we observe Sinho wins.

   Let P be the event that Sinho successfully predicts Forest's move, and let Q be the event that Forest successfully predicts Sinho's move.

$$
\begin{aligned}
q = \Pr(X_i = 1) &= \Pr(X_i = 1 | P, Q^C)\Pr(P, Q^C) + \Pr(X_i = 1 | P^C, Q^C)\Pr(P^C, Q^C) \\
&= (1)p(2p) + \frac{1}{2}(1-p)(2p) \\
&= 2p^2 + p(1-p) \\
&= p(1+p)
\end{aligned}
$$

Our estimate for p is thus the following, in terms of what we observe, q. Plugging into the quadratic formula, we get the following.

$$p = \frac{1}{2}\sqrt{4q+1} - 1)$$

2. How many games (n) will we need to watch, to predict p within 0.05 of the actual value, with 95% confidence?

   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 $|X - \mu| \geq a$

   Since p needs to be within 0.05, q needs to be within $\frac{0.1^2}{4} = 0.0025$. We can alternatively see this as fitting to the form $|X - \mu| \geq a$.

   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 $\tilde{p}$.

$$|(\frac{1}{2}\sqrt{4\tilde{q}+1}-1)-(\frac{1}{2}\sqrt{4q+1}-1))| \leq 0.05$$
$$|\frac{1}{2}\sqrt{4\tilde{q}+1}-\frac{1}{2}\sqrt{4q+1})| \leq 0.05$$
$$|\sqrt{4\tilde{q}+1}-\sqrt{4q+1}| \leq 0.1$$
$$|(4\tilde{q}+1)-(4q+1)| \leq 0.01$$
$$|4\tilde{q}-4q| \leq 0.01$$
$$|\tilde{q}-q| \leq 0.0025$$

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!

$$|\tilde{q}-\mu_{\tilde{q}}| \leq 0.0025$$

.

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

$$|\tilde{q}n-\mu_{\tilde{q}}n| \leq 0.0025n$$

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

$$Pr[|Y - qn| \leq 0.0205n] \geq \frac{var(Y)}{(n0.0025)^2}$$

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.

$$\begin{aligned} \frac{var(Y)}{a^2} &= \frac{nq(1-q)}{(n0.0025)^2} \\ &= \frac{q(1-q)}{n(0.0025)^2} \end{aligned}$$

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.

$$\frac{var(Y)}{a^2} = \frac{q(1-q)}{n(0.0025)^2} = 0.05$$

.

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 $\frac{1}{4}$.

$$\frac{q(1-q)}{n(0.0025)^2} = \frac{1/4}{n(0.0025)^2} = 0.05$$
$$\frac{1}{4n(0.0025)^2} = \frac{1}{20}$$
$$\frac{5}{(0.0025)^2} = n$$
$$n = 800000$$