COMP 480 HW 1 Q 2

Gyuheon Oh

September 10, 2023

1. Can you find an interesting a (use Chebyshev). It could be of the form contant $\times$ std (where std is the standard deviation or square root of some variance)

The goal is to find a high probability confidence interval for the estimator $\hat{n}$, which is defined as:

$$\hat{n} = \frac{k_1 k_2}{\frac{1}{R}\sum_{i=1}^{R} M_i}$$

A good idea is to find confidence interval for:

$$\hat{n} = \frac{1}{R}\sum_{i=1}^{R} M_i$$

We want to find a suitable value 'a' such that:

$$P(E(M) - a \leq \hat{n} \leq E(M) + a) \geq 1 - \delta$$

Let's use Chebyshev's inequality, which states that for any random variable with finite mean (μ) and variance (σ^2):

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

In our case, X represents M_i (each measurement), and we want to find 'a' such that:

$$P(|\hat{n} - E(M)| \geq a) \leq \delta$$

Since we want to bound the absolute difference between $\hat{n}$ and $E(M)$, we can write this as:

$$P\left(\left|\frac{1}{R}\sum_{i=1}^{R} M_i - E(M)\right| \geq a\right) \leq \delta$$

Now, let's apply Chebyshev's inequality:

$$\frac{1}{k^2} = \delta$$

Solving for k:

$$k = \frac{1}{\sqrt{\delta}}$$

Now, let's replace k with $\frac{1}{\sqrt{\delta}}$ in the inequality:

$$P\left(\left|\frac{1}{R}\sum_{i=1}^{R} M_i - E(M)\right| \geq \frac{1}{\sqrt{\delta}}\sigma\right) \leq \delta$$

Here, σ represents the standard deviation of the measurements M_i.
Now, we can define 'a' in terms of δ and σ:

$$a = \frac{1}{\sqrt{\delta}}\sigma$$

So, $a =$ constant $\times$ std, where the constant is $\frac{1}{\sqrt{\delta}}$.

2. Comment on how you will use the formula to find the number of repetitions to get to a particular fraction f of error, compared to E(M), with a probability of failure less than 0.05

To find the number of repetitions required to achieve a particular fraction f of error compared to the expected value $E(M)$ with a probability of failure less than 0.05, we can use the confidence interval formula we've derived.

The formula you've derived for 'a' is:

$$a = \frac{1}{\sqrt{\delta}}\sigma$$

Here, 'a' represents the margin of error, 'δ' represents the probability of failure (in this case, $\delta = 0.05$ for a 95% confidence interval), and 'σ' is the standard deviation of the measurements 'M_i'. The margin of error 'a' represents the maximum acceptable difference between your estimator $\hat{n}$ and the true value $E(M)$.

Now, to find the number of repetitions 'R' required to achieve a fraction 'f' of error, we'll set up an inequality based on your confidence interval:

$$\left|\frac{1}{R}\sum_{i=1}^{R} M_i - E(M)\right| \leq f \cdot E(M)$$

we want the probability of failure (the probability that the absolute difference between our estimator and the true value exceeds 'f' times $E(M)$) to be less than 0.05:

$$P\left(\left|\frac{1}{R}\sum_{i=1}^{R} M_i - E(M)\right| \leq f \cdot E(M)\right) \geq 1 - 0.05$$

Now, we can use the Chebyshev's inequality we mentioned earlier to set up the inequality:

$$\frac{1}{f^2} = 0.05$$

Solving for 'f':

$$f = \frac{1}{\sqrt{0.05}} = \frac{1}{0.2236} \approx 4.47$$

So, we want the fraction 'f' of error to be approximately 4.47 times the standard deviation 'σ'.

Now, we can use this 'f' value to calculate the number of repetitions 'R' needed to achieve this level of accuracy:

$$\left| \frac{1}{R} \sum_{i=1}^{R} M_i - E(M) \right| \leq 4.47 \cdot \sigma$$

This gives us a specific criterion for the margin of error. To find 'R', we would need to consider the known or estimated value of 'σ' (the standard deviation of the measurements) and solve for 'R' based on our desired level of accuracy 'f'.

Keep in mind that the actual calculation of 'R' would depend on the specific values of 'σ' and 'f' for your problem, but this approach provides us with the framework to set up the inequality and calculate the number of repetitions required.

3. How does the above translate into an interval for ˆn. Can we argue some cases where estimation is really hard.

To translate the analysis into a confidence interval for $\hat{n}$, you can follow these steps:

1. Using the estimated value of 'R' calculated based on the desired accuracy 'f' and standard deviation 'σ', we can find the margin of error 'a' as discussed earlier:

$$a = 4.47 \cdot \sigma$$

2. Now, we can construct a confidence interval for $\hat{n}$ as follows:

$$\hat{n} \pm a$$

This confidence interval represents a range of values within which we can be reasonably confident (with a probability of 95% or greater, based on the choice of δ) that the true value of $\hat{n}$ lies.

Now, let's discuss some scenarios where estimation can be particularly challenging:

1. High Variability: When the measurements (values of M_i) have high variability or a large standard deviation 'σ', it becomes challenging to estimate $\hat{n}$ accurately with a small margin of error. In such cases, you would need a larger sample size 'R' to achieve the desired level of accuracy.

2. Small Effect Size: If the true value of $\hat{n}$ is very close to the expected value $E(M)$, it can be challenging to estimate it accurately, especially with a limited number of repetitions. Small changes in the measurements can lead to large relative errors.

3. Complex Dependencies: When the measurements have complex dependencies or are influenced by many factors that are difficult to control or measure, it can be challenging to model the system accurately and estimate $\hat{n}$.

4. Non-Normal Distributions: If the distribution of measurements is not close to normal, Chebyshev's inequality may not be as effective, and other statistical methods may be required for accurate estimation.

In such challenging cases, you may need to use more advanced statistical techniques, consider increasing the sample size 'R,' or explore alternative methods for estimation to achieve the desired level of accuracy in estimating $\hat{n}$.