CS/ECE 473 Fall 2020

Homework 5 Problem 1

Ryan Prendergast (ryanp4)
Noah Watson (nwatson3)
Lawson Probasco (lawsonp2)

Suppose you want to estimate the average of n numbers via sampling, for example the average wealth of people in a town. The average can be very skewed by outliers — perhaps there are a few billionaires that will not make it to the sample but will clearly affect the average. However, we can obtain an accurate estimate if we assume that the numbers are within some limited range. Assume the input numbers $z_1, z_2, ..., z_n$ are from $[a, b]$ where $a, b \in R$ with $a \leq b$. Suppose you sample k input numbers (with replacement) and output their average as the estimate for the true average $\alpha = (\sum_i (z_i)/n$. Let X be the random variable denoting the output value.

Using Chebyshev's inequality, show that for $k \geq \frac{(b-a)^2}{\delta\epsilon^2}$, we have

$$Pr[|X - \alpha| \geq \epsilon] \leq \delta$$

Solution: First we find the variance of one of the samples x_i from the set $z_1, z_2, ...z_n$. Since it is a bounded random variable, the variance cannot be larger than the square of the size of the bounds (the most any samples could deviate by is b-a). We also note that the samples x_i are IID. The random variable $X = \sum_{i=1}^{k}(x_i)/k$.

$$\begin{aligned}
Var[x_i] &\leq (b-a)^2 && \text{upper bound of variance} \\
Var[\sum_{i=1}^{k} x_i] &\leq k(b-a)^2 && \text{linearity of variances of IIDs} \\
Var[X] &\leq \frac{1}{k^2} Var[\sum_{i=1}^{k} x_i] && \text{definition of variance} \\
Var[X] &\leq \frac{(b-a)^2}{k} && \text{simplification}
\end{aligned}$$

By definition, $\alpha = E[X]$. We can use Chebyshev's inequality to bound $Pr[|X - \alpha| \geq \epsilon]$ for $\epsilon > 0$:

$$\begin{aligned}
Pr[|X-\alpha| \geq \epsilon] &\leq \frac{Var[X]}{\epsilon^2} && \text{Definition of Chebyshev} \\
Pr[|X-\alpha| \geq \epsilon] &\leq \frac{(b-a)^2}{k\epsilon^2} && \text{substitution.}
\end{aligned}$$

Given $k \geq \frac{(b-a)^2}{\delta\epsilon^2}$, we substitute for k. This lower bound for k, when substituted for k, serves as an upper bound for the probability, as k is raised to a negative power.

$$\begin{aligned}
Pr[|X-\alpha| \geq \epsilon] &\leq \frac{(b-a)^2\delta\epsilon^2}{(b-a)^2\epsilon^2} && \text{substitution} \\
Pr[|X-\alpha| \geq \epsilon] &\leq \delta
\end{aligned}$$

■

CS/ECE 473 Fall 2020

Homework 5 Problem 1

Ryan Prendergast (ryanp4)
Noah Watson (nwatson3)
Lawson Probasco (lawsonp2)

Using the Chernoff inequality, show that there exists a constant $c > 0$ such that for $k \geq \frac{c(b-a)^2 log(2/\delta)}{\epsilon^2}$, we have

$$Pr[|X - \alpha| \geq \epsilon] \leq \delta$$

Solution: We cite this source http://math.mit.edu/ goemans/18310S15/chernoff-notes.pdf for the Chernoff Bound of a bounded random variable.

We start again by considering the individual random variables X_i. We note that our symmetric analysis $|X - \alpha|$ requires that we add the upper tail and the lower tail Chernoff bounds. This gives us:

$$\begin{aligned}
Pr[|X_i - \alpha| \geq \epsilon] &\leq Pr[X_i - \alpha \geq \epsilon] + Pr[X_i - \alpha \leq -\epsilon] \\
Pr[|\sum_{i=1}^{k}(X_i) - k * \alpha| \geq \epsilon] &\leq 2exp(\frac{-2\epsilon^2}{k(b-a)^2}) && \text{MIT Notes Citation} \\
Pr[|X_i - \alpha| \geq \epsilon] &\leq 2exp(\frac{-2\epsilon^2}{(b-a)^2}) && \text{Single Sample} \\
Pr[|X - \alpha| \geq \epsilon] &\leq 2exp(\frac{-2\epsilon^2 * k}{(b-a)^2}) && \text{Lecture 8}
\end{aligned}$$

We want to find a value c such that $Pr[|X - \alpha| \geq \epsilon] \leq \delta$ given $k \geq \frac{c(b-a)^2 log(2/\delta)}{\epsilon^2}$. So, we solve for k given an upper bound of δ.

$$\begin{aligned}
2exp(\frac{-2\epsilon^2 * k}{(b-a)^2}) &\leq \delta \\
\frac{-2\epsilon^2 * k}{(b-a)^2} &\leq log(\delta/2) \\
\epsilon^2 * k &\geq (b-a)^2(1/2)log(2/\delta) \\
k &\geq \frac{1/2(b-a)^2 log(2/\delta)}{\epsilon^2}
\end{aligned}$$

We can clearly see above that the value $c = 1/2$ gives us a lower bound for k at which $Pr[|X - \alpha| \geq \epsilon] \leq \delta$. Thus such a value for c exists. ■