# Homework 4

*By signing my name, I agree to abide by the Stern Code of Conduct* ________________________

**Question 1**

**a. Compute the mean $\bar{x}$ of the list of numbers 4, 3, 4, 4, 5**

$\bar{x} = (4 + 3 + 4 + 4 + 5)/5 = 20/5 = 4$

**b. Consider the random variable $X$ which has probability distribution function $p_X(1) = 1/10, p_X(2) = 1/10, p_X(3) = 2/10, p_X(4) = 4/10, p_X(5) = 2/10$. Draw a graph of the probability distribution function of $X$. Label the axes of your graph.**

The graph should have a horizontal axis with the numbers 1 through 5, and each number should have a bar above it with the height of the bar given by $p_X(x)$.

**c. Is this graph a histogram? Why or why not?**

No. Histograms are for data, not models. (But these graphs are related by sampling!)

**d. Compute $E[X]$.**

$E[X] = 1(1/10) + 2(1/10) + 3(2/10) + 4(4/10) + 5(2/10) = 35/10 = 3.5$

**e. Compute $\mathrm{Var}(X)$. You can use `R`, or any calculator, to simplify the final answer.**

$\mathrm{Var}(X) = E[(X - 3.5)^2] = (1 - 3.5)^2(1/10) + (2 - 3.5)^2(1/10) + (3 - 3.5)^2(2/10) + (4 - 3.5)^2(4/10) + (5 - 3.5)^2(2/10) = 1.45$

**f. Do you think it's plausible that the list from part (a) might have been sampled from the distribution of $X$?**

Yes. The numbers in the list are all possible outcomes for $X$, and their mean is close to the mean of $X$. Later in the course we will have more tools for answering this question rigorously.

**Question 2**

Suppose $Y \sim \text{Bin}(n, p)$, and recall these facts from class:

- $\mu = E[Y] = np$
- $\sigma^2 = \text{Var}(Y) = np(1-p)$
- Chebyshev's inequality: $P(|Y - \mu| \geq a) \leq \sigma^2/a^2$.
- This inequality says the probability that $Y$ falls *outside* the range from $\mu - a$ to $\mu + a$ can't be too high, where the specific meaning of too high depends on $\sigma^2$ and $a^2$.

For this problem, let $p = 1/2$.

**a. Simplify the formulas for $\mu$ and $\sigma^2$ using $p = 1/2$.**

$\mu = n/2$, $\sigma^2 = n/4$

**b. What value of $a$ can you plug in to Chebyshev's inequality to make the right hand side less or equal to 1/2? Write out the inequality for that value of $a$, simplifying as much as you can, and using the formulas above for $\mu$ and $\sigma^2$.**

Set $a = \sqrt{2}\sigma$, so $a^2 = 2\sigma^2$.

Then $P(|Y - n/2| \geq \sqrt{2}n/4) \leq 1/2$

**c. Now let $n = 100$, so $np = 50$. According to Chebyshev's inequality, $P(y_L \leq Y \leq y_U) > 1/2$ for some specific numbers $y_L$ and $y_U$. What are these two numbers? Round up $y_U$ and round down $y_L$ to the nearest integers.**

Now $\mu = 50$, $\sigma^2 = 25$, so $a = \sqrt{2}\sqrt{25} = 5\sqrt{2}$.

The range for Chebyshev's inequality is $50 - 5\sqrt{2}$ to $50 + 5\sqrt{2}$

Rounding down gives $y_L = 42$ and rounding up gives $y_U = 58$.