Lecture 8 – Agenda & Examples

Agenda

1. Review Questions
2. Geometric Distribution – Context & Proof of Valid PDF ($X \sim Geo(p) \rightarrow p(x) = (1-p)^{x-1}p$)
3. Negative Binomial Distribution – Context & Formulaically $\left(X \sim Nbin(r,p) \rightarrow p(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}\right)$
4. Expectation & Variance – Proof of Geometric Mean $\left(X \sim Nbin(r,p) \rightarrow E[X] = \frac{r}{p} \text{ and } V[X] = \frac{r(1-p)}{p^2}\right) \left(X \sim Geo(p) \rightarrow E[X] = \frac{1}{p} \text{ and } V[X] = \frac{1-p}{p^2}\right)$
5. Examples

Review

1. Show that if $Y \sim Bin(n,p)$ and $Z \sim Bin(n, 1-p)$, then $p(y = n - y) = P(z = y)$.
2. Consider the RV $Y \sim Bin(n,p)$. Recursively define the $p(y)$.
3. Consider the RV $W \sim Bin(n,p)$. Find the $E[W(W-1)(W-2)(W-3)]$.
4. Consider the fact that of the 30,000 students at UConn, 3000 have blood type O. Now, let us randomly select 200 students from UConn, and let the RV X be the number of students with blood type O.
    a. Properly label the distribution of X.
    b. We know (or will know from the homework) that the Binomial distribution is approximately Normal when the sample size reaches a certain number. The general rule of thumb is that if $np \geq 5$ and $n(1-p) \geq 5$, we can use the Normal approximation where our random variable X can be said to be $Normal\big(np, np(1-p)\big)$. Using this information, what is the median of the distribution of X?

Lecture

1. The memoryless property for the Geometric distribution states that if $X \sim Geo(p)$, with i, j positive numbers, then $P(X \geq i + j | X \geq i) = P(X \geq j)$.
    a. First, determine the cumulative distribution function for a geometric distribution (hint: the cumulative distribution function $F(x) = P(X \leq x) = P(you\ get\ a\ success\ at\ or\ before\ the\ x^{th} trial) = 1 - P(first\ x\ trials\ are\ all\ failures)$)
    b. Now, prove that the Geometric distribution indeed exhibits the memoryless property.
2. Properly label the distribution for each scenario below:
    a. We are looking at baseball hits for Aaron Judge. We know since his career started in 2016, he has a batting average of .274 (meaning that he gets a hit

27.4% of the time he bats). Let the RV X be the number of at bats until he gets his first hit.

b. Suppose that we are now rolling a die and are interested in studying the number of “3”s that come up. In this case, however, the die is weighted in that 40% of the time we will roll a “2”, but the rest of the outcomes are equally likely. Let the RV V be the number of rolls until we get 4 “3”s.
c. Consider the case where we are flipping a coin once. Let the RV X be that we flip a head.
d. Consider the situation now where we are flipping a coin. Let the RV X be the number of flips until we get 2 heads.
e. Consider the situation where we are flipping a coin 10 times. Let the RV Z be the number of heads flipped.

3. Recursively define the $p(x)$ if $X \sim Nbin(r, p)$.