LECTURE 13

Agenda:

1. Geometric distribution continued
2. Negative Binomial distribution

GEOMETRIC RANDOM VARIABLE

Let us recollect that a geometric random variable arises out of an experiment which consists of repeating a Bernoulli trial until the first success. Let p denote the probability of success of the Bernoulli trial.
Let X = # of failures before the first success.

We saw that,

$$\mathcal{X} = \text{Range}(X) = \{0, 1, 2, \ldots\}$$

$$P(X = x) = p(1-p)^x \quad \text{for } x = 0, 1, 2, \ldots$$

$$P(X \le x) = \left(1 - (1-p)^{x+1}\right) \quad \text{for } x = 0, 1, 2, \ldots$$

$\boxed{\text{RESULT:}}$ $\quad E(X) = \dfrac{1-p}{p}$

Proof: $E(X) = \displaystyle\sum_{x=0}^{\infty} x p (1-p)^x$

$$= p(1-p)\left(1 + 2(1-p) + 3(1-p)^2 + \ldots\right)$$

$$= p(1-p)\left(\begin{array}{l} 1 + (1-p) + (1-p)^2 + \cdots\cdots \\ \quad + (1-p) + (1-p)^2 + \cdots\cdots \\ \quad\quad + (1-p)^2 + \cdots\cdots \\ \quad\quad\quad + \cdots\cdots \end{array}\right)$$

Recall that if $|x| < 1$, then

$$a + ax + ax^2 + \cdots\cdots = \frac{a}{1-x}.$$

Hence,

$$E(X) = p(1-p)\left(\frac{1}{1-(1-p)} + \frac{q}{1-(1-p)} + \frac{q^2}{1-(1-p)} + \cdots\cdots\right)$$

(q denotes $1-p$)

$$= (1-p)\left(1 + q + q^2 + \cdots\cdots\right)$$

$$= \frac{(1-p)}{1-(1-p)}$$

$$= \frac{1-p}{p}.$$

Along identical lines, it can be proved that

$\boxed{\text{RESULT:}}$ $$V(X) = \frac{1-p}{p^2}$$

<u>Example:</u> A firm has a new position which needs fluency in both English and Spanish. Applicants are selected randomly from (a very large) pool, and interviewed until the first applicant who is fluent in both English and Spanish is found. If there are roughly 20% applicants in the pool who are fluent in both English and Spanish, what is the expected number of unqualified applicants who will be interviewed before a qualified applicant is found?

Since each applicant is either qualified or unqualified, each interview is a Bernoulli trial.

HOWEVER, EACH BERNOULLI TRIAL IS NOT STRICTLY INDEPENDENT AND IDENTICAL. BUT IF THE POPULATION IS VERY LARGE, WE CAN APPROXIMATELY ASSUME THAT.

Let X = # of unqualified applicants interviewed before the first qualified applicant

Then X is a geometric random variable, where the success probability of the Bernoulli experiment is 0.2. Hence, $E(X) = \frac{1-0.2}{0.2} = 4$.

MEMORYLESS PROPERTY OF THE GEOMETRIC DISTRIBUTION

Let j, k be positive integers.

$\boxed{\text{RESULT:}}$ $P(X \geq j+k \mid X \geq j) = P(X \geq k)$.

Proof: $P(X \geq j+k \mid X \geq j)$

$$= \frac{P\left(\{X \geq j+k\} \cap \{ X \geq j\}\right)}{P(\{X \geq j\})}$$

$$= \frac{P\left(\{X \geq j+k\}\right)}{P(\{ X \geq j\})}$$

$$= \frac{\sum_{x=j+k}^{\infty} p(1-p)^x}{\sum_{x=j}^{\infty} p(1-p)^x}$$

$$= \frac{p(1-p)^{j+k} \sum_{x=0}^{\infty} (1-p)^x}{p(1-p)^{j} \sum_{x=0}^{\infty} (1-p)^x}$$

$$= (1-p)^k$$

$$P(X \geq k) = \sum_{x=k}^{\infty} p(1-p)^x$$

$$= p(1-p)^k \sum_{x=0}^{\infty} (1-p)^x$$

$$= \frac{p(1-p)^k}{1-(1-p)}$$

$$= (1-p)^k .$$

In words, this property means that given that there have been j failures the chance of atleast k more failures before the first success is exactly the same as if we are just beginning the experiment and want to know the probability of having atleast k failures before the first success.

THE NEGATIVE BINOMIAL RANDOM VARIABLE

The geometric random variable corresponds to the number of failures before the first success in a sequence of independent Bernoulli trials. But what if we are interested in the number of failures before the r^{th} success for some positive integer r?

Let X = # of failures observed before the r^{th} success in the Bernoulli trials

X is called as the NEGATIVE BINOMIAL RANDOM VARIABLE.

Clearly, the random variable X can take any non-negative integer as a value, i.e,

$$\mathcal{X} = \text{Range}(X) = \{0, 1, 2, \text{-----------}\}.$$

$$P(X = x) = P\left(\begin{array}{l}\text{The first } x+r-1 \text{ trials contain} \\ x \text{ failures and } r-1 \text{ successes,} \\ (x+r)^{th} \text{ trial is a success}\end{array}\right)$$

$$= P\left(\begin{array}{l}\text{The first } x+r-1 \text{ trials contain } x \\ \text{failures and } r-1 \text{ successes}\end{array}\right) P\left((x+r)^{th} \text{ trial is a success}\right)$$

$$= \Big\{ \underbrace{\binom{x+r-1}{x}}_{\substack{\text{\# of ways of choosing the} \\ x \text{ trials with success}}} p^{r-1} (1-p)^x \Big\} p$$

$$= \binom{x+r-1}{x} p^r (1-p)^x .$$

Hence, the probability mass function of a negative binomial random variable is given by

$$p_X(x) = \binom{x+r-1}{x} p^r (1-p)^x , \quad x = 0, 1, 2, \text{---------}$$