ENM 207

Lecture 9

Some Useful Discrete Distributions

Some Useful Discrete Distributions

- The observations generated by different experiments have
  - the same general type of behavior.
- Consequently, random variables associated with these experiments can be described by
  - the same probability distribution

  and therefore can be represented by
  - a single formula

Discrete Uniform Distribution

- The simplest of all discrete probability distributions is one where the random variable assumes each of its values with an equal probability.
- Such a probability distribution is called as a discrete uniform distribution.

Definition : If the random variable X assumes the values $x_1, x_2, \ldots, x_k$ with equal probabilities, then the discrete uniform distribution is given by

$$p(x) = \begin{cases} \dfrac{1}{k} & x = x_1, x_2, \ldots, x_k \\ 0 & \text{o.w.} \end{cases}$$

Discrete Uniform Distribution

Example

When a die is tossed, each element of the space S = {1, 2, 3, 4, 5, 6} occurs with probability 1/6. Therefore, we have a uniform distribution, with

Discrete Uniform Distribution

Example

When a light bulb is selected at random from a box that contains a 40-watt bulb, a 60-watt bulb, a 75-watt bulb, a 100-watt bulb, each element of the sample space S = {40, 60, 75, 100} occurs with probability ¼. Therefore, we have an uniform discrete distribution, with

Discrete Uniform Distribution

- The mean of the discrete uniform distribution is;

$$\begin{aligned}\mu = E(x) &= \sum_{i=1}^{k} x_i \, p(x_i) \\ &= \sum_{i=1}^{k} \frac{x_i}{k} \\ &= \frac{\sum_{i=1}^{k} x_i}{k}\end{aligned}$$

Discrete Uniform Distribution

- The variance of the discrete uniform distribution is;

$$\sigma^2 = E\left[(X-\mu)^2\right]$$

$$\sigma^2 = \sum_{i=1}^{k}(x_i-\mu)^2 p(x) = \sum_{i=1}^{k}\frac{(x_i-\mu)^2}{k}$$

$$\sigma^2 = \frac{\sum_{i=1}^{k}(x_i-\mu)^2}{k}$$

Discrete Uniform Distribution

Example

For the experiment of tossing a die, find mean and variance of the random variable

Bernoulli Distribution

There are many problems in which the experiment consists of *n* trials or subexperiments.

Here, we are concerned with an individual trial that has two possible outcomes

success {*A*} or

failure {Ā}.

Definition : A bernoulli trial results in one of two outcomes denoted A for success and Ā for failure.

For example , tossing a coin is a Bernoulli trial , since only one of two different outcomes can occur head (H) and tail (T) .

Bernoulli Distribution

Definition : A bernoulli random variable, *X*, is defined as the numerical outcome of a Bernoulli trial where

X = 1 if a success occurs and
X = 0 if a failure occurs

Consequently the probability distribution for *X* is given as follows

$$p(x) = \begin{cases} p^x (1-p)^{1-x} & x = 0, 1 \\ 0 & \text{o.w.} \end{cases}$$

p : probability of success for a Bernoulli trial

Bernoulli Distribution

- The mean of Bernoulli random variable is;

$$\mu = E(X) = \sum_{x=0}^{1} xp(x)$$
$$= 0p(0) + 1p(1)$$
$$= 0p^0(1-p)^{1-0} + 1p^1(1-p)^{1-1}$$
$$= p$$

Bernoulli Distribution

- The variance of Bernoulli random variable is;

$$\sigma^2 = E\left[(X-\mu)^2\right] = E(X^2) - \mu^2$$

$$E(X^2) = \sum_{x=0}^{1} x^2 p(x)$$
$$= (0)^2 p^0(1-p)^{1-0} + (1)^2 p^1(1-p)^{1-1} = p$$

$$\sigma^2 = E(X^2) - \mu^2 = p - p^2 = p(1-p)$$

$$\sigma = \sqrt{p(1-p)}$$

Binomial Distribution

- Consider an experiment consists of *n* independent bernoulli trials.
- There are only two possible outcomes on each trial,
  - A for success and
  - Ā for failure .
- Suppose that the probability of a success is constant from trial to trial and

  p(A) = p,   p(Ā) =1-p .
- Let the random variable X be defined as follows :

  X = number of successes in *n* independent bernoulli trials.
- X is a binomial random variable with parameters *n* and *p*.

Binomial Distribution

Definition : A Bernoulli trial can result in a success with probability p and a failure with probability q=1-p. Then the probability distribution of the binomial random variable, *X*, the number of successes in *n* independent trials, is

$$p(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & x = 0, 1, 2, \ldots, n \\ 0 & \text{o.w.} \end{cases}$$

Binomial Distribution

- The parameters of the binomial distribution are n and p.
- A simple derivation of the binomial distribution is:

Let $p(x) = P\{x$ successes in n trials$\}$

The probability of the particular outcome in S with successes for the first x trials and failures for the last n-x trials is

$$P(\overbrace{AAA....AA}^{x}\overbrace{\bar{A}\bar{A}...\bar{A}\bar{A}\bar{A}}^{n-x}) = p^x q^{n-x}$$

- There are $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ outcomes having exactly x successes and (n-x) failures. Therefore

$$p(x) = \binom{n}{x} p^x q^{n-x}$$

Binomial Distribution

The Binomial distribution is a probability distribution since

$$i)\ 0 \le p(x) \le 1$$

$$ii)\ \sum_{x=0}^{n} p(x) = \sum_{x=0}^{n} P(X = x)$$

$$= \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = [p+q]^n = 1^n = 1$$

Binomial Distribution

- If we want to know the probabilities that
  - 12 flips of a balanced coin will yield 5 heads and 7 tails ,
  - 3 of 10 persons will respond to a given mail questionnarie,
  - 2 of 25 items will be defective ,
  - there are 6 boys and 9 girls among 15 children born on a given day,
- In each case, we are interested in the probability of getting a certain number of successes in a given number of trials .

Binomial Distribution

- The mean of Binomial random variable is;

$$
\begin{aligned}
\mu = E(X) &= \sum_{x=0}^{n} xp(x) \\
&= \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^{x} q^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x} \\
&= np[p+q]^{n-1} \\
&= np
\end{aligned}
$$

Binomial Distribution

- The variance of Binomial random variable is;

$$\sigma^2 = E\left[(X-\mu)^2\right] = E(X^2) - \mu^2$$

$$E(X(X-1)) = \sum_{x=0}^{n} x(x-1)p(x)$$
$$= \sum_{x=0}^{n} x(x-1)\frac{n!}{x!(n-x)!}p^x q^{n-x} = n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!}p^{x-2}q^{n-x}$$
$$= n(n-1)p^2[p+q]^{n-2} = n(n-1)p^2$$

$$E(X(X-1)) = E(X^2 - X) = n(n-1)p \Rightarrow E(X^2) - E(X) = n(n-1)p$$
$$\Rightarrow E(X^2) = n(n-1)p + E(X)$$
$$\Rightarrow E(X^2) = n(n-1)p + np$$

$$\sigma^2 = E(X^2) - \mu^2 = np + n(n-1)p^2 - n^2p^2 = npq$$

Binomial Distribution

- Another approach to find the mean and variance is to consider X as a sum of n independent random variables, each with mean p and variance pq, so that

$$X = X_1 + X_2 + \ldots + X_n$$

$$E(X) = p + p + \ldots + p = np$$
$$\mathrm{Var}(X) = pq + pq + \ldots + pq = npq$$

Binomial Distribution

Example

Suppose that a radio tube inserted into a certain type of set has a probability of 0.2 of functioning more than 500 hours. If we test 20 tubes,

what is the probability that exactly X of them functioning more than 500 hours , X = 0, 1, 2, …, 20

Binomial Distribution

Example

The National Foundation reports that 70% of the U.S. graduate students who earn PhD degrees in engineering are foreign nationals.

Consider the number of foreign students (Y) in a random sample of 25 engineering students who recently earned their PhD

a) Find $P(Y=10)$

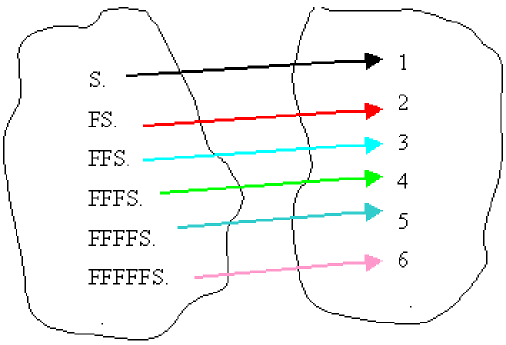
b) Find $P(Y \leq 5)$

c) Find μ and σ.

The Geometric Distribution

- The geometric distribution is also based on a sequence of Bernoulli trials except that the number of trials is not fixed,
- The random variable of interest, denoted by X, is defined to be the number of trials required to reach the first success.
- The sample space and range space for X are illustrated in figure followed.

The Geometric Distribution



The range space for x is Rx={ 1,2,3,....} and the distribution of X given by

$$p(x) = \begin{cases} p\ q^{x-1} & x = 1,2,3,.. \\ 0 & o.w. \end{cases}$$

X is the number of trials until the first success is observed.

The Geometric Distribution

It is easy to verify that this is a probability distribution since ,

$$a)\, p(x) \geq 0 \quad for \;\; all \;\; x$$

$$b) \sum_{x=1}^{\infty} P(x) = \sum_{x=1}^{\infty} pq^{x-1} = p \sum_{x=1}^{\infty} q^{x-1} = p \sum_{x=0}^{\infty} q^{x} = p\left[\frac{1}{1-q}\right] = \frac{p}{p} = 1$$

The mean and variance of geometric distribution are respectively

$$\mu = E(X) = \sum_{x=1}^{\infty} xP(x) = \sum_{x=1}^{\infty} xpq^{x-1}$$

$$= p\frac{d}{dq}\sum_{x=1}^{\infty} q^{x} = p\frac{d}{dq}\left[\frac{q}{1-q}\right] = p\frac{1}{(1-q)^2} = \frac{1}{p}$$

The Geometric Distribution

$$\sigma^2 = E(x^2) - \left[E(x)\right]^2$$

$$E(x^2) = \sum_{x=1}^{\infty} x^2 p(x) = \sum_{x=1}^{\infty} x(x-1)p(x) + \sum_{x=1}^{\infty} xp(x)$$

$$= \sum_{x=1}^{\infty} x(x-1)pq^{x-1} + \sum_{x=1}^{\infty} xpq^{x-1}$$

$$= pq\sum x(x-1)q^{x-2} + p\sum xq^{x-1}$$

$$= pq\sum \frac{d^2}{dq^2}q^x + p\sum \frac{d}{dq}q^x$$

$$= pq\frac{d^2}{dq^2}\sum_{x=1}^{\infty} q^x + p\frac{d}{dq}\sum q^x$$

$$= pq\frac{d^2}{dq^2}\frac{q}{1-q} + p\frac{d}{dq}\frac{q}{1-q}$$

$$= pq\frac{d}{dq}\frac{1}{(1-q)^2} + p\frac{1}{(1-q)^2}$$

$$= \frac{2q+p}{(1-q)^2} = \frac{q+1}{(1-q)^2}$$

$$\sigma^2 = E(x^2) - \left[E(X)\right]^2 = \frac{q+1}{(1-q)^2} - \frac{1}{p^2} = \frac{q+1-1}{p^2} = \frac{q}{p^2}$$

The Geometric Distribution

Example

- A manufacturer uses electrical fuses in an electronic system. The fuses are purchased in large lots and tested sequentially until the first defective fuse observed. Assume that the lot contains 10% defective fuses.
  - What is the probability that the first defective fuse will be one of the first five fuses tested?
  - Find the mean, variance and standard deviation of X, the number of fuses tested until the first defective fuse is observed.

Negative Binomial Distribution

- The negative binomial distribution is also based on the Bernoulli trials.
- It is a logical extension of the geometric distributions.
- Random variable X denotes the number of trials until rth success is observed

$$p(x) = \begin{cases} \binom{x-1}{r-1} p^r q^{x-r} & x = r, r+1, r+2, \ldots \\ 0 & o.w. \end{cases}$$

Negative Binomial Distribution

- The term $p^r q^{x-r}$ arises from the probability associated with exactly one outcome in sample space that has
  - (*x*-*r*) failures
  - *r* success.
- In order for this outcome to occur, there must be *r*-1 successes in the *x*-1 repetitions before the last outcome, which is always success.
- There are $\binom{x-1}{r-1}$ arrangements satisfying this condition.

Negative Binomial Distribution

- The mean of negative binomial distribution;

$$\mu = E(X) = \frac{r}{p}$$

- The variance of negative binomial

$$\sigma^2 = Var(X) = \frac{rq}{p^2}$$

Negative Binomial Distribution

Example

In an NBA championship series, the team which wins four games out of seven will be the winner. Suppose that the team A has probability 0.55 of winning over the team B and both team A and B face each other in the championship games.

a) What is the probability that team A will win the series in six games?

b) What is the probability that team A will win the series?

c) If both teams face each other in a regional playoff series and the winner is decided by winning three out of five games, what is the probability that team A will win a playoff?

Multinomial Distribution

- The binomial experiment becomes multinomial experiment if each trial has more than 2 outcomes.
  - The classification of a manufactured products as being light, heavy, or acceptable
  - The recording of accidents at a certain intersection according to the day of week constitute multinomial experiment.
- If a given trial can result in any one of k possible outcomes

  $E_1, E_2, \ldots, E_k$ with probabilities $p_1, p_2, \ldots, p_k$,

  then the multinomial distribution give the probability that

  E_1 occurs x_1 times, E_2 occurs x_2 times,…, E_k occurs x_k times in n independent trials,

  where

$$x_1 + x_2 + \ldots + x_k = n$$

- $p_1 + p_2 + \ldots + p_k = 1$, since the results of each trial must be one of the k possible outcomes.

Multinomial Distribution

- Since the trails are independent, any specified order yielding
  - x_1 outcomes for E_1, x_2 for E_2,…, x_k for E_k

  will occur with probability p_1^{x1}, p_2^{x2},…, p_k^{xk}.
- The total number of orders yielding similar outcomes for the n trials is equal to the number of partitions of n items into k groups with x_1 in the first group; x_2 in the second group,…, x_k in the kth group. This can be done in

$$\binom{n}{x_1, x_2, ..., x_k} = \frac{n!}{x_1! x_2! ... x_k!}$$

- Since all the partitions are mutually exclusive and occur with equal probability, the multinomial distribution is obtained by multiplying the probability for a specified order by the total number of partitions.

Multinomial Distribution

$$p(x_1, x_2, ... x_k) = \begin{cases} \left[\dfrac{n!}{x_1! x_2! ... x_k!}\right] p_1^{x_1} p_2^{x_2} ... p_k^{x_k} & x_1 = 0,1,2,...;\ x_2 = 0,1,2,...;...;\ x_k = 0,1,2,... \\ 0 & o.w. \end{cases}$$

where

$$\begin{aligned} & x_1 + x_2 + ... + x_k = n \\ & p_1^{x1} + p_2^{x2} + ... + p_k^{xk} = 1 \end{aligned}$$

- The mean and variance of X_i are

$$\begin{aligned} & E(X_i) = np_i \\ & Var(X_i) = np_i(1 - p_i) \end{aligned}$$

Multinomial Distribution

Example

For an airport containing three runways it is known that in the ideal setting the following are the probabilities that the individual runways are accessed by random arriving commercial jet

Runway 1: p_1=2/9
Runway 2: p_2=1/6
Runway 3: p_3=11/18

What is the probability that 6 randomly arriving airplanes are distributed in following fashion?

Runway 1: 2 airplanes
Runway 2: 1 airplanes
Runway 3: 3 airplanes

The Hypergeometric Distribution

- The distinction between the binomial distribution and hypergeometric distribution lies in the way the sampling is done.
- The types of applications of the hypergeometric are very similar to the binomial distribution.
- We are interested in computing probabilities for the number of observations that fall into a particular category.
- But, in the case of the binomial, independence among trials are required.
- If the binomial is applied to sampling from a lot of items (deck of cards, batch of production items), the sampling must be done with replacement of each item after it is observed.
- The hypergeometric distribution does not require independence and is based on the sampling done without replacement.

The Hypergeometric Distribution

- The hypergeometric experiment has two properties:
  - A random sample of size n is selected without replacement from N items
  - k of the N items may be classified as successes and N-k are classified as failures.
- The random variable, X, is the number of successes on a hypergeometric experiment.

$$p(x) = \begin{cases} \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} & x = 0,1,2,\dots, n \\ 0 & o.w. \end{cases}$$

The Hypergeometric Distribution

- The mean and variance of hypergeometric distribution are

$$\mu = E(X) = \frac{nk}{N}$$

$$\sigma^2 = Var(X) = \frac{N-n}{N-1}.n.\frac{k}{n}\left(1-\frac{k}{N}\right)$$

The Hypergeometric Distribution

Example:

Lots of 40 components each are called unacceptable if they contain as many as 3 defectives or more. The procedure for sampling the lot is to select 5 components at random and to reject the lot if a defective is found.

What is the probability that exactly 1 defective is found in the sample if there are 3 defective in the entire lot?

Find mean and variance of defective components.

The Hypergeometric Distribution

Example:

From a lot of 10 missiles, 4 are selected at random and fired. If lot contains 3 defectives missiles that will not fire, what is the probability that

a) All 4 will fire?

b) At most 2 will not fire?

References

- Walpole, Myers, Myers, Ye, (2002),
  - *Probability & Statistics for Engineers & Scientists*
- Dengiz, B., (2004),
  - *Lecture Notes on Probability,* http://w3.gazi.edu.tr/web/bdengiz
- Hines, Montgomery, (1990),
  - *Probability & Statistics in Engineering* & Management *Science*