SECTION 9: SPECIAL DISCRETE DISTRIBUTIONS

9.1. Binomial Distribution

Consider the following random experiments and random variables:

1. Flip a coin 10 times. Let X=number of heads obtained.

2. A worn machine tools produces 1% defective parts. Let X=number of defective parts in the next 25 parts produced.

3. Each sample of air has a 10 % chance of containing a particular rare molecule. Let X=the number of air samples that contain the rare molecule in the next 18 samples analyzed.

4. Of all bits transmitted through a digital transmission channel, 10% are received in error. Let X=the number of bits in error in the next five bits transmitted.

5. A multiple choice test contains 10 questions, each with four choices, and you guess at each question. Let X=the number of questions answered correctly.

6. In the next 20 births at a hospital, let X=the number of female births.

7. Of all patients suffering a particular illness, 35 % experience improvement from a particular medication. In the next 100 patients administered the medication, let X=the number of patients who experience improvement.

These examples illustrate that a general probability model that includes these experiments as particular cases would be very useful.

Each of these random experiments can be thought of as consisting of a series of repeated, random trials: 10 flips of the coin in experiment 1, the production of 25 parts in experiment 2, and so forth. The random variable in each case is a count of the number of trials that meet a specified criterion. The outcome from each trial either meets the criterion that X counts or it does not; consequently, each trial can be summarized as resulting in either a success or a failure. For example, in the multiple choice experiment, for each question, only the choice that is correct is considered a success. Choosing any one of the three incorrect choices results in the trial being summarized as a failure.

The terms *success* and *failure* are just labels. We can just as well use *A* and *B* or 0 or 1. Unfortunately, the usual labels can sometimes be misleading. In experiment 2, because X counts defective parts, the production of a defective part is called a success.

A trial with only two possible outcomes is used so frequently as a building block of a random experiment that it is called a **Bernoulli trial**. It is usually assumed that the trials that constitute the random experiment are **independent.** This implies that the outcome from one trial has no effect on the outcome to be obtained from any other trial. Furthermore, it is often reasonable to assume that the **probability of a success in each trial is constant.** In the multiple choice experiment, if the test taker has no knowledge of the material and just guesses at each question, we might assume that the probability of a correct answer is 1/4 *for each question.*

Definition: A random experiment consists of n Bernoulli trials such that

(1) The trials are independent
(2) Each trial results in only two possible outcomes, labeled as "success" and "failure"
(3) The probability of a success in each trial, denoted as p, remains constant.

The random variable X that equals the number of trials that result in a success has a **binomial random variable** with parameters $0 < p < 1$ and $n = 1, 2, \cdots$. The probability mass function of X is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, \cdots, \text{n} \qquad (1)$$

If X is a binomial random variable with parameters p and n,

$$\mu = E(X) = np \qquad and \qquad \sigma^2 = V(X) = np(1-p) \qquad (2)$$

Example 1: The chance that a bit transmitted through a digital transmission channel is received in error is 0.1. Also, assume that the trials are independent. Let X=the number of bits in error in the next four bits transmitted.

a) Determine $P(X = 2)$.

b) Determine mean and variance of X ($E(X) = ?$ *and* $V(X) = ?$)

Solution:

Let the letter E denote a bit in error, and let the letter O denote the bit is okay, that is, received without error. We can represent the outcomes of this experiment as a list of for letters that indicate the bits that are in error and those that are okay. For example, the outcome OEOE indicates that the second and fourth bits are in error and the other two bits are okay. The corresponding values for x are

Outcome	**x**	**Outcome**	**x**
OOOO	0	EOOO	1
OOOE	1	EOOE	2
OOEO	1	EOEO	2
OOEE	2	EOEE	3
OEOO	1	EEOO	2
OEOE	2	EEOE	3
OEEO	2	EEEO	3
OEEE	3	EEEE	4

a) The event that X=2 consists of the six outcomes:

$\{EEOO, EOEO, EOOE, OEEO, OEOE, OOEE\}$

Using the assumption that the trials are independent, the probability of $\{EEOO\}$ is

$P(EEOO) = P(E)P(E)P(O)P(O) = (0.1)^2(0.9)^2 = 0.0081$

Also, any one of the six mutually exclusive outcomes for which X=2 has the same probability of occurring. Therefore,

$P(X = 2) = 6(0.0081) = 0.0486$

In general, $P(X = x) = (\textit{number of outcomes that result in x errors})\ \textit{times}\ (0.1)^x(0.9)^{4-x}$

To complete a general probability formula, only an expression for the number of outcomes that contain x errors needed. Therefore,

$$P(X = x) = \binom{4}{x}(0.1)^x(0.9)^{4-x}$$

So that $P(X = 2) = \binom{4}{2}(0.1)^2(0.9)^{4-2} = 6(0.0081) = 0.0486$

b) For the number of transmitted bits received in error n=4 and p=0.1, so

$E(X) = 4(0.1) = 0.4 \qquad V(X) = 4(0.1)(0.9) = 0.36$

Example 2: Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant.

- **a)** Find the probability that in the next 18 samples, exactly 2 contain the pollutant.
- **b)** Determine the probability that at least four samples contain the pollutant.
- **c)** Determine the probability that $P(3 \le X < 7)$.

Solution:

Let X=the number of samples that contain the pollutant in the next 18 samples analyzed. Then X is a binomial random variable with p=0.1 and n=18. Therefore,

a) $P(X = 2) = \binom{18}{2}(0.1)^2(0.9)^{16} = 153(0.1)^2(0.9)^{16} = 0.284$

b) The requested probability is $P(X \geq 4)=\sum_{x=4}^{18}\binom{18}{x}(0.1)^x(0.9)^{18-x}$. However, it is easier to use the complementary event,

$$P(X \geq 4)=1-P(X<4)=1-\sum_{x=0}^{3}\binom{18}{x}(0.1)^x(0.9)^{18-x}$$
$$=1-[0.150+0.300+0.284+0.168]=0.098$$

c)
$$P(3 \leq X < 7)=\sum_{x=3}^{6}\binom{18}{x}(0.1)^x(0.9)^{18-x}$$
$$=0.168+0.070+0.022+0.005$$
$$=0.265$$

9.2. Geometric Distribution

Consider a random experiment that is closely related to the one used in the definition of a binomial distribution. Again, assume a series of Bernoulli trials (independent trials with constant probability p of a success on each trial). However, instead of a fixed number of trials, trials are conducted until a success is obtained. Let the random variable X denote the number of trials until the first success. In the transmission of bits, X might be the number of bits transmitted until an error occurs.

Example 3: The probability that a bit transmitted through a digital transmission channel is received in error is 0.1. Assume the transmissions are independent events, and let the random variable X denote the number of bits transmitted until the first error.

Then, P(X=5) is the probability that the first four bits are transmitted correctly and the fifth bit is in error. This event can be denoted as $\{OOOE\}$, where O denotes an okay bit. Because the trials are independent and the probability of a correct transmission is 0.9,

$$P(X=5)=P(OOOOE)=0.9^4 0.1=0.066$$

Note that there is some probability that X will equal any integer value. Also, if the first trial is a success X=1. Therefore, the range of X is $\{1,2,3,\cdots\}$, that is, all positive integers.

Definition: In a series of Bernoulli trials (independent trials with constant probability p of a success), let the random variable X denote the number of trials until the first success. Then X is a geometric random variable with parameter $0<p<1$ and

$$f(x)=(1-p)^{x-1}p \qquad x=1,2,\cdots \tag{3}$$

If X is a **geometric random variable** with parameter p,

$$\mu=E(X)=1/p \qquad and \qquad \sigma^2=V(X)=(1-p)/p^2 \tag{4}$$

Example 4: The probability of a successful optical alignment in the assembly of an optical data storage product is 0.8. Assume the trials are independent.

a) What is the probability that the first successful alignment requires exactly four trials?
b) What is the probability that the first successful alignment requires at most four trials?
c) What is the probability that the first successful alignment requires at least four trials?
d) What is the mean number of trials before the first successful alignment?
e) What is the variance of the number of trials before the first successful alignment?

Solution:

Let the random variable X denote the number of trials until the first successful alignment. Then X is a geometric random variable with $p = 0.8$, $f(x) = (0.2)^{x-1} 0.8 \qquad x = 1, 2, \cdots$

a) $P(X = 4) = (0.2)^3 0.8 = 0.0064$

b) $P(X \le 4) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$

$= 0.8 + (0.2)0.8 + (0.2)^2 0.8 + (0.2)^3 0.8$

$=0.9984$

c) $P(X \ge 4) = 1 - P(X < 4) = 1 - [P(X = 1) + P(X = 2) + P(X = 3)] = 1 - 0.992 = 0.008$

d) $E(X) = 1/p = 1/0.8 = 1.25$

e) $V(X) = (1-p)/p^2 = (1-0.8)/0.8^2 = 0.3125$

9.3. Negative Binomial Distribution

A generalization of a geometric distribution in which the random variable is the number of Bernoulli trials required to obtain r successes results in the **negative binomial distribution**.

Definition: In a series of Bernoulli trials (independent trials with constant probability p of a success), let the random variable X denote the number of trials until the r successes occur. Then X is a **negative binomial random variable** with parameters $0 < p < 1$ and $r = 1, 2, 3, \cdots$, and

$$f(x) = \binom{x-1}{r-1} (1-p)^{x-r} p^r \qquad x = r, r+1, r+2, \cdots \qquad (5)$$

If X is a negative binomial random variable with parameters p and r,

$$\mu = E(X) = r/p \qquad and \qquad \sigma^2 = V(X) = r(1-p)/p^2 \qquad (6)$$

Example 5: A Web site contains three identical computer servers. Only one is used to operate the site, and the other two are spares that can be activated in case the primary system fails. The probability of a failure in the primary computer (or any activated spare system) from a request for service is 0.0005. Assuming that each request represents an independent trial,

a) what is the mean number of requests until failure or all three servers?
b) what is the probability that all three servers fail within five requests?

Solution: Let X denote the number of requests until all three servers fail, and let X_1, X_2 *and* X_3 denote the number of requests before a failure of the first, second and third servers used, respectively. Now, $X = X_1 + X_2 + X_3$. Also, the requests are assumed to comprise independent trials with constant probability of failure $p = 0.0005$. Furthermore, a spare server is not affected by the number of requests before it is activated. Therefore, X has a negative binomial distribution with $p = 0.0005$ and $r = 3$. Consequently,

a) $E(X) = 3/0.0005 = 6000$ *requests*

b) The probability is $P(X \le 5)$ and

$$P(X \le 5) = P(X = 3) + P(X = 4) + P(X = 5) = 0.0005^3 + \binom{3}{2} 0.0005^3 (0.9995) + \binom{4}{2} 0.0005^3 (0.9995)^2$$
$$= 1.25\times10^{-10} + 3.75\times10^{-10} + 7.49\times10^{-10}$$
$$= 1.249\times10^{-9}$$

9.4. Hypergeometric Distribution

Definition:

A set of *N* objects contains
K objects classified as successes
N-K objects classified as failures

A sample of size n objects is selected randomly (without replacement) from the *N* objects, where $K \le N$ and $n \le N$.

Let the random variable X denote the number of successes in the sample. Then X is a **hypergeometric random variable** and

$$f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \qquad x = \max\{0, n+K-N\}\ to \min\{K, n\} \qquad (7)$$

If X is a hypergeometric random variable with parameters *N*, *K*, and *n*, then

$$\mu = E(X) = np \qquad and \qquad \sigma^2 = V(X) = np(1-p)\left(\frac{N-n}{N-1}\right) \qquad (8)$$

where *p=K/N*.

Example 6: A batch of parts contains 100 parts from a local supplier of tubing and 200 parts from a supplier of tubing in the next state. If four parts are selected randomly and without replacement,

a) what is the probability they are all from the local supplier?
b) what is the probability that two or more parts in the sample are from the local supplier?
c) what is the probability that at least one part in the sample is from the local supplier?
d) what is the mean number of parts in the sample from the local supplier?

Solution:

a) Let X equal the number of parts in the sample from the local supplier. Then, X has a hypergeometric distribution and the requested probability is $P(X=4)$. Consequently,

$$P(X=4)=\frac{\binom{100}{4}\binom{200}{0}}{\binom{300}{4}}=0.0119$$

b) $$P(X\geq 2)=\frac{\binom{100}{2}\binom{200}{2}}{\binom{300}{4}}+\frac{\binom{100}{3}\binom{200}{1}}{\binom{300}{4}}+\frac{\binom{100}{4}\binom{200}{0}}{\binom{300}{4}}$$

$$=0.298+0.098+0.0119=0.408$$

c) $$P(X\geq 1)=1-P(X=0)=1-\frac{\binom{100}{0}\binom{200}{4}}{\binom{300}{4}}=0.804$$

d) $E(X)=4(100/300)=1.33$

Example 7: A company employs 800 men under the age of 55. Suppose that 30 % carry a marker on the male chromosome that indicates an increased risk for high blood pressure.

a) If 10 men in the company are tested for the marker in this chromosome, what is the probability that exactly 1 man has the marker?

b) If 10 men in the company are tested for the marker in this chromosome, what is the probability that more than 1 has the marker?

Solution:

Let X denote the number of men have the marker in this chromosome.

a) $P(X=1)=\dfrac{\binom{240}{1}\binom{560}{9}}{\binom{800}{10}}=0.1201$

b) $P(X>1)=1-\left[P(X=0)+P(X=1)\right]=1-\left[\dfrac{\binom{240}{0}\binom{560}{10}}{\binom{800}{10}}+\dfrac{\binom{240}{1}\binom{560}{9}}{\binom{800}{10}}\right]=0.8523$

9.5. Poisson Distribution

Definition: Given an interval of real numbers, assume counts occur at random throughout the interval. If the interval can be partitioned into subintervals of small enough length such that

1) the probability of more than one count in a subinterval is zero,
2) the probability of one count in a subinterval is the same for all subintervals and proportional to the length of the subinterval, and
3) the count in each subinterval is independent of other subintervals, the random experiment is called a **Poisson process**.

The random variable X that equals the number of counts in the interval is a **Poisson random variable** with parameter $\lambda > 0$, and the probability mass function of X is

$$f(x)=\frac{e^{-\lambda}\lambda^{x}}{x!} \qquad x=0,1,2,\cdots \qquad (9)$$

If X is a Poisson random variable with parameter λ, then

$$\mu=E(X)=\lambda \qquad \textit{and} \qquad \sigma^2=V(X)=\lambda \qquad (10)$$

E<u>xperiments yielding numerical values of a random variable X, the number of outcomes occurring during a given time interval or in a specified region, are called **Poisson experiments.**</u> The given time interval may be of any length, such as a minute, a day, a week, a month, or even a year. For example, a Poisson experiment can generate observations for the random variable X representing;

- the number of telephone calls received per hour by an office,
- the number of days school is closed due to snow during the winter,
- the number of games postponed due to rain during a baseball season.

The specified region could be a line segment, an area, a volume, or perhaps a piece of material. In such instances, X might represent;

- the number of field mice per acre,
- the number of bacteria in a given culture,
- the number of typing errors per page.

Example 8: Contamination is a problem in the manufacture of optical storage disks. The number of particles of contamination that occur on an optical disk has a Poisson distribution, and the average number of particles per centimeter squared of media surface is 0.1. The area of a disk under study is 100 squared centimeters.

a) Find the expected number of particles in the area of a disk under study.
b) Find the probability that 12 particles occur in the area of a disk under study.
c) Find the probability that zero particles occur in the area of a disk under study.
d) Determine the probability that 12 or fewer particles occur in the area of the disk under study.

Solution:

a) Let X denote the number of particles in the area of a disk under study. Because the mean number of particles is 0.1 particles per cm^2,

$$E(X) = 100\, cm^2 \times 0.1\, particles / cm^2 = 10\, particles$$

b) $P(X = 12) = \dfrac{e^{-10}10^{12}}{12!} = 0.095$

c) $P(X = 0) = e^{-10} = 4.54 \times 10^{-5}$

d) $P(X \leq 12) = P(X = 0) + P(X = 1) + \cdots + P(X = 12) = \sum_{x=0}^{12} \dfrac{e^{-10}10^{x}}{x!} = 0.791$

Example 9: Suppose that the number of customers that enter a bank in an hour is a Poisson random variable, and suppose that $P(X = 0) = 0.05$. Determine the mean and variance of X.

Solution: $f(x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!} \qquad x = 0, 1, 2, \cdots$ then

$$P(X = 0) = \frac{e^{-\lambda}\lambda^{0}}{0!} = e^{-\lambda} = 0.05 \Rightarrow e^{\lambda} = 20 \Rightarrow \lambda = 2.996$$

$$\mu = E(X) = 2.996 \qquad and \qquad \sigma^2 = V(X) = 2.996$$