STAT 509 – Section 3.2: Discrete Random Variables

Random Variable: A function that assigns numerical values to all the outcomes in the sample space.

Notation: Capital letters (like *Y*) denote a random variable. Lowercase letters (like *y*) denote possible values of the random variable.

Discrete Random Variable : A numerical r.v. that takes on a countable number of values (there are gaps in the range of possible values).

Examples:

1. **Number of phone calls received in a day by a company**
2. **Number of heads in 5 tosses of a coin**

Continuous Random Variable : A numerical r.v. that takes on an uncountable number of values (possible values lie in an unbroken interval).

Examples:

1. **Length of nails produced at a factory**
2. **Time in 100-meter dash for runners**

Other examples?

The probability distribution of a random variable is a graph, table, or formula which tells what values the r.v. can take and the probability that it takes each of those values.

Example 1: A design firm submits bids for four projects. Let Y = number of successful bids.

y	0	1	2	3	4
P(y)	0.06	0.35	0.43	0.15	0.01

Example 2: Toss 2 coins. The r.v. Y = number of heads showing.

y	0	1	2
P(y)	¼	½	¼

Graph for Example 1:

For any probability distribution:

(1) P(y) is between 0 and 1 for any value of y.
(2) $\sum_y P(y)$ = 1. That is, the sum of the probabilities for all possible y values is 1.

Example 3: P(y) = y / 10 for y = 1, 2, 3, 4.

Valid Probability Distribution?
Property 1?

Property 2?

Cumulative Distribution Function:

If Y is a random variable, then the cumulative distribution function (cdf) is denoted by F(y).
F(y) = P(Y $\leq$ y)

cdf for r.v. in Example 1?

Graph of cdf:

Expected Value of a Discrete Random Variable

The expected value of a r.v. is its mean (i.e., the mean of its probability distribution).

For a discrete r.v. *Y*, the expected value of *Y*, denoted μ or E(*Y*), is:

$$\mu = E(Y) = \sum_{y} yP(y)$$

where $\sum_{y}$ represents a summation over all values of *Y*.

Recall Example 3:

μ =

Here, the expected value of *y* is

Recall Example 1: What is the expected number of successful design bids?

E(*Y*) =

So on average, a firm in this situation would win ______ bids.

The expected value does not have to be a possible value of the r.v. --- it's an average value.

Variance of a Discrete Random Variable

The variance σ^2 is the expected value of the squared deviation from the mean μ; that is, $\sigma^2 = E[(Y - \mu)^2]$.

$$\sigma^2 = \Sigma (y - \mu)^2 P(y)$$

Shortcut formula:

$$\sigma^2 = [\sum_y y^2 P(y)] - \mu^2$$

where $\sum_y$ represents a summation over all values of y.

<u>Example 3</u>: Recall $\mu = 3$ for this r.v.

$\Sigma y^2 P(y) =$

Thus $\sigma^2 =$

Note that the standard deviation σ of the r.v. is the square root of σ^2. So for Example 3, $\sigma =$

<u>Example 1</u>: Recall $\mu = 1.7$ for this r.v.

$\Sigma y^2 P(y) =$

Thus $\sigma^2 =$

and $\sigma =$

Some Rules for Expected Values and Variances

If c is a constant number,

- **$E(c) = c$**
- **$E(cY) = cE(Y)$**

If X and Y are two random variables,

- **$E(X + Y) = E(X) + E(Y)$**

And:

- **$\text{var}(c) = 0$**
- **$\text{var}(cY) = c^2\text{var}(Y)$**

If X and Y are two *independent* random variables,

- **$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$**

The Binomial Random Variable

- **Many experiments have responses with 2 possibilities (Yes/No, Pass/Fail).**
- **Certain experiments called binomial experiments yield a type of r.v. called a binomial random variable.**

Characteristics of a binomial experiment:

(1) The experiment consists of a number (denoted n) of identical trials.

(2) There are only two possible outcomes for each trial – denoted "Success" (S) or "Failure" (F)

(3) The probability of success (denoted p) is the same for each trial.
(Probability of failure = $q = 1 – p$.)

(4) The trials are independent.

Then the binomial r.v. (denoted *Y*) is the number of successes in the *n* trials.

Example 1: A factory makes 25 bricks in an hour. Suppose typically 5% of all bricks produced are nonconforming. Let *Y* = total number of nonconforming bricks. Then *Y* is

Example 2: A student randomly guesses answers on a multiple choice test with 3 questions, each with 4 possible answers. *Y* = number of correct answers. Then *Y* is

What is the probability distribution for *Y* in this case?

Outcome	***y***	**P(outcome)**

Probability Distribution of *Y*

y	$P(y)$

General Formula: (Binomial Probability Distribution)
(n = number of trials, p = probability of success.)
The probability there will be exactly y successes is:

$$p(y) = \binom{n}{y} p^{y} q^{n-y} \qquad (y = 0, 1, 2, \dots, n)$$

where

$$\binom{n}{y} = \text{"}n \text{ choose } y\text{"}$$

$$= \frac{n!}{y!\,(n-y)!}$$

Here, $0! = 1$, $1! = 1$, $2! = 2 \cdot 1 = 2$, $3! = 3 \cdot 2 \cdot 1 = 6$, etc.

Example 1(a): Out of 25 bricks produced, what is the probability of exactly 2 nonconforming bricks?

Example 1(b): Out of 25 bricks produced, what is the probability of at least 1 nonconforming brick?

Example 1(c): Out of 25 bricks produced, what is the probability of at least 2 nonconforming bricks?

- **The mean (expected value) of a binomial r.v. is $\mu = np$.**
- **The variance of a binomial r.v. is $\sigma^2 = npq$.**
- **The standard deviation of a binomial r.v. is $\sigma =$**

Example: What is the mean number of nonconforming bricks that we would expect out of the 25 produced?

$\mu = np =$

What is the standard deviation of this binomial r.v.?

Finding Binomial Probabilities using R

Hand calculations of binomial probabilities can be tedious at times.

Example 1(c): Out of 25 bricks produced, what is the probability of at least 6 nonconforming bricks?

```
> sum(dbinom(6:25, size = 25, prob = 0.05))
[1] 0.001212961
```

Example 4: Historically, 10% of homes in Florida have radon levels higher than that recommended by EPA.

- **In a random sample of 20 homes, find the probability that exactly 3 have radon levels higher than the EPA recommendation.**

- **In a random sample of 20 homes, find the probability that more than 4 have radon levels higher than the EPA recommendation.**

- **In a random sample of 20 homes, find the probability that between 2 and 5 have radon levels higher than the EPA recommendation.**

Sampling without replacement: The Hypergeometric Distribution

• **If we take a sample (without replacement) of size *n* from a finite collection of *N* objects (*r* of which are "successes"), then the number of successes *Y* in our sample is not binomial. Why not?**

• **It follows a hypergeometric distribution. The probability function of *Y* is:**

$$P(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} \quad \text{for } y = 0, 1, 2, \ldots, n$$

• **The mean (expected value) of *Y* is $\mu = E(Y) = n\frac{r}{N}$**

Example: Suppose a factory makes 100 filters per day, 5 of which are defective. If we randomly sample (without replacement) 15 of these filters, what is the expected number of defective filters in the sample?

What is the probability that exactly 1 of the sampled filters will be defective?

Poisson Random Variables

The Poisson distribution can be used to model the number of events occurring in a continuous time or space.

- **Number of telephone calls received per hour**
- **Number of claims received per day by an insurance company**
- **Number of accidents per month at an intersection**
- **Number of breaks per 1000 meters of copper wire.**

The mean number of events (per 1-unit interval) for a Poisson distribution is denoted λ.

Which values can a Poisson r.v. take?

Probability distribution for *Y*
(if *Y* is Poisson with mean λ)

$$P(y) = \frac{\lambda^{y} e^{-\lambda}}{y!} \qquad (\text{for } y = 0, 1, 2, \ldots)$$

Mean of Poisson probability distribution: λ

Variance of Poisson probability distribution: λ

• If we record the number of occurrences *Y* in *t* units of time or space, then the probability distribution for *Y* is:

$$P(y) = \frac{(\lambda t)^y e^{-\lambda t}}{y!} \qquad \text{(for } y = 0, 1, 2, \ldots)$$

Example 1(a): Historically a process has averaged 2.6 breaks in the insulation per 1000 meters of wire. What is the probability that 1000 meters of wire will have 1 or fewer breaks in insulation?

Example 1(b): What is the probability that 3000 meters of wire will have 1 or fewer breaks in insulation?

Now, $E(Y) = \underline{\lambda t}$ =

Finding Poisson Probabilities using R

Hand calculations of Poisson probabilities can be tedious at times.

Example 1(c): What is the probability that 1000 meters of wire will have between 3 and 7 breaks in insulation?

```
> sum(dpois(3:7, lambda = 2.6))
[1] 0.4762367
```

Example 1(d): What is the probability that 2000 meters of wire will have at least 4 breaks in insulation?

```
1 - sum(dpois(0:3, lambda = 2.6 * 2))
```

Conditions for a Poisson Process

1) **Areas of inspection are independent of one another.**
2) **The probability of the event occurring at any particular point in space or time is negligible.**
3) **The mean remains constant over all areas/intervals of inspection.**

Example 2: Suppose we average 5 radioactive particles passing a counter in 1 millisecond. What is the probability that *exactly 10 particles* will pass in the next three milliseconds? *10 or fewer particles*?