Random variable: a variable that assumes its values corresponding to a various outcomes of a random experiment, therefore its value cannot be predicted with certainty.

Discrete r.v.: a r.v. that takes only a finite or countably infinite number of different values.

Continuous r.v.: a. r.v. that can assume any value in some interval.

We denote upper case letters *X, Y, Z* to denote random variables and lower case *x, y, z* to denote particular values assumed by r.v.'s.

$\boldsymbol{P(X = x)}$**:** the probability that a discrete random variable *X* assumes a value *x*, also denoted by $P(x)$

Discrete probability distribution (**Probability mass function**): a table, graph, or a rule that associates a probability *P(x)* with each possible value *x* of a discrete r.v. *X* can assume.

Properties of the discrete probability distribution:

$$0 \le P(x) \le 1$$
$$\sum_x P(x) = 1$$

The **expected value** or **mean** of a random variable *X*, denoted by $E(X)$ or μ_X is

$$E(X) = \mu_X = \sum_{i=1}^{N} x_i P(X = x_i)$$

The expected value of a random variable represents: the center of the probability distribution of the random variable; the "long-run" average of values of the random variable when observed repeatedly; the weighted average of the possible values of *X*; and the value that minimizes the squared loss.

Example 1. Find the expected value of a random variable *X* which is the number of interruptions per day in a large computer network with the following distribution.

x	0	1	2	3	4	5
P(x)	0.35	0.25	0.2	0.1	0.05	0.05

Expected value of $Y = g(X)$, that is, Y is a function of X

$$E(Y) = E\{g(X)\} = \sum_x g(x)P(x)$$

Example 2. Suppose $Y = X^2$, where *X* is as defined in Example 1. Find $E(Y)$.

The **variance** of a r.v. *X*, denoted by $Var(X)$ or σ_X^2, is

$$Var(X) = \sigma_X^2 = E(X - \mu_X)^2 = \sum_x (x - \mu_X)^2 P(x) = \sum_x x^2 P(x) - \mu_X^2 = E(X^2) - \mu_X^2$$

The **standard deviation** of a r.v. X is $\sigma_X = \sqrt{Var(X)}$.

Example 3. For the random variable defined in Example 1, find the standard deviation.

MSL Homework: 5.1, 5.4
HC Homework: 5.5

Example 4. A bag of M&M's contains 6 green and 4 red M&M's. Two M&M's are randomly drawn *without replacement* from the bag. Define X to be the number of green M&M's drawn.

a) Find the discrete probability distribution of X.
b) How many green M&M's do you expect to draw?
c) Find the standard deviation of X.

Suppose $Y = a + bX$, then

$E(Y) = a + bE(X)$, $Var(Y) = b^2 Var(X)$, $\sigma_Y = |b|\sigma_X$

Example 5. Suppose the number of cars sold by a sales person per week, denoted by X, is a random variable with the following distribution.

x	0	1	2	3	4	5
$P(X = x)$	0.08	0.21	0.27	0.24	0.13	0.07

a. Find the mean and the standard deviation of the number of cars cold by the sales person per week.
b. Suppose the sales person gets paid basic salary of \$200 per week plus commission of \$150 per car sold. Find the expected value and the standard deviation of the weekly income of the sales person.

Example 6. Suppose an investor is considering investment of \$10,000 in one of two portfolios, called Portfolio A and Portfolio B whose returns on investment of \$10,000 have the following probability distributions. Discuss which portfolio the investor should choose.

Portfolio A

Return x	Probability $P(X = x)$
\$600	0.4
800	0.3
1000	0.3

Portfolio B

Return y	Probability $P(Y = y)$
\$600	0.1
700	0.4
880	0.5

HC Homework: Do the following written problem

Suppose a portfolio manager charges a fixed fee of \$50 plus 1% of the return for each of the portfolios in Example 6 as the managing fee. Find the expected value and the standard deviation of the managing fee for each of the portfolios.

Bernoulli experiment: a random experiment whose outcome results in one of two possible (mutually exclusive) basic outcomes. For convenience we label these outcomes as success and failure, head and tail, or occurrences and nonoccurrence.

We define π=P(success) and P(failure) $= 1 - \pi$ in a Bernoulli experiment.

A **Bernoulli random variable** Y is such that it takes values either 0 or 1 depending on the outcome of a Bernoulli trial. For example, $Y = 1$ if success and $Y = 0$ if failure.

The (discrete) probability distribution of a Bernoulli random variable with a success probability π is

y	0	1
$P(Y = y)$	$1 - \pi$	π

or

$$P(Y = y) = \pi^y (1 - \pi)^{1-y} \text{ for } y = 0, 1.$$

For the Bernoulli random variable $E(Y) = \pi$ and $Var(Y) = \pi(1 - \pi)$.

When a Bernoulli trial with success probability π is repeated n times independently, we define X to be the number of successes out of these n Bernoulli trials. The X is called a **binomial random variable** with possible values of 0, 1, 2, ..., n.

The (discrete) probability distribution of a binomial r.v. is call the **binomial distribution** and

$$P(X = x) = {}_nC_x \pi^x (1 - \pi)^{n-x} \quad \text{for} \quad x = 0, 1, \dots, n$$

where ${}_nC_x = \frac{n!}{x!(n-x)!}$ and $n! = n \times (n - 1) \times \cdots \times 2 \times 1$.

Note others uses $\binom{n}{x}$ or C_x^n instead of ${}_nC_x$.

For a binomial random variable X with n and π, $E(X) = n\pi$ and $Var(X) = n\pi(1 - \pi)$.

MSL Homework: 5.18, 5.19

We can also obtain cumulative probabilities of binomial random variables using MS Excel For example in order to find the cumulative probabilities of a binomial random variable with n=**20** and π=**0.3** do the following.

1. Enter 0 to **20** in column A,
2. Click b1 assuming 0 is in a1 (or next to 0),
3. **Formulas >f_x (Insert Function) >Statistical>BINOM.DIST**,
4. Enter a1 (assuming 0 is in a1) for *Number_s*, **20** for *Trials*, **0.3** for *Probability_s* and 0 for *ProbabilityMass* 0 and 1 for *Cumulative Probability.*
5. Click the (cumulative) probability just calculated and drag it down.

x	P(X=x)	P(X <=x)
0	0.000798	0.000798
1	0.006839	0.007637
2	0.027846	0.035483
3	0.071604	0.107087
4	0.130421	0.237508
5	0.178863	0.416371
6	0.191639	0.60801
7	0.164262	0.772272
8	0.114397	0.886669
9	0.06537	0.952038
10	0.030817	0.982855
11	0.012007	0.994862
12	0.003859	0.998721
13	0.001018	0.999739
14	0.000218	0.999957
15	3.74E-05	0.999994
16	5.01E-06	0.999999
17	5.05E-07	1
18	3.61E-08	1
19	1.63E-09	1
20	3.49E-11	1

Various forms of probabilities in terms of cumulative probabilities:

$$P(X > a) = 1 - P(X \le a)$$
$$P(X \ge a) = 1 - P(X < a)$$
$$P(a < X \le b) = P(X \le b) - P(X \le a)$$
$$P(a \le X \le b) = P(X \le b) - P(X < a)$$
$$P(a \le X < b) = P(X < b) - P(X < a)$$
$$P(a < X < b) = P(X < b) - P(X \le a)$$

To use PHStat2

1. **Add-Ins > PHStat > Probability & Probability Distributions > Binomial…**
2. Enter **20** for Sample Size (i.e., number of trials) and **0.3** for Prob. of an Event of Interest (i.e., π , probability of occurrence).

Example 9. A bag of M & M's contains 8 green, 6 orange, 6 brown M & M's. Suppose 5 M & M's are randomly taken from the bag with replacement. Let *X* be the number of green M & M's in the sample. Describe the probability distribution of *X*.

Example 10. Only 30 % of the people in a large city feel that its mass transit system is adequate. Twenty people are selected at random for their opinion.

a) Find the probability that exactly six people in the sample feel that the system is adequate.
b) Find the probability that 8 or more people in the sample feel that the system is adequate.
c) Find the expected value and standard deviation of the number of people in the sample who feel than the system is adequate.

Example 11. A test consists of 20 multiple-choice questions. Each question has 4 choices, and only one of them is the correct answer. Suppose a student takes this exam totally unprepared, and thus randomly choosing the answer for each question. In order to pass the test, this student must answer at least 12 questions correctly. Find the probability that this student will pass the test.

MSL Homework: 5.22, 5.23
HC Homework: 5.25, 5.26

The **Poisson distribution** has many application in problems concerning the number of occurrences of some event during some continuous time interval or in some continuous region or space. The variables in the following examples are generally known to be well described by the Poisson distribution:

a) the number of phone calls to a switch board in a given minute,
b) the number of customers entering a checkout line at a supermarket in a given time period,
c) the number of accidents on the Pullman-Moscow highway per week,
d) the number of snowstorms hitting Spokane in a season,
e) the number of chocolate chips in a chocolate chip cookie,
f) the number of scratches on a new car.

The number of occurrence of a phenomenon with a Poisson distribution is called a **Poisson random variable.**

For the number of occurrences of a phenomenon to be a Poisson random variable, it must satisfy the following conditions:

1. The number of occurrences in any interval is independent of the number of occurrences in any other interval.
2. The probability of one occurrence in an interval is the same for all intervals of equal size and is proportional to the size of the interval.
3. The probability that two or more successes will occur in an interval approaches to zero as the interval becomes smaller.

Checking these conditions for a given phenomenon is beyond the scope of this course as it is not an easy task. Therefore, in problems Poisson random variables are explicitly specified.

If we let X denote a Poisson random variable with a mean rate of occurrences λ, the discrete probability distribution of X is

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

where $e \approx 2.718281828\cdots$ is the natural number, and $E(X) = \lambda$ and $Var(X) = \lambda$.

We can also obtain cumulative probabilities of Poisson random variables using MS Excel For example in order to find the cumulative probabilities of a Poisson random variable with λ =**1.5** do the following.

1. Enter 0 to 10 (or more or less) in column A, (Keep in mind that the Poisson random variable assumes values up to infinity.)
2. Click b1 assuming 0 is in a1 (or next to 0),
3. **Formulas >f_x (Insert Function) >Statistical>POISSON.DIST**,
4. Enter a1 (assuming 0 is in a1) for *x*, **1.5** for *Mean*, and 0 or *Probability mass* and 1 for *Cumulative probability.*
5. Click the (cumulative) probability just calculated and drag down.

x	P(X=x)	P(X <=x)
0	0.22313	0.22313
1	0.334695	0.557825
2	0.251021	0.808847
3	0.125511	0.934358
4	0.047067	0.981424
5	0.01412	0.995544
6	0.00353	0.999074
7	0.000756	0.99983
8	0.000142	0.999972
9	2.36E-05	0.999996
10	3.55E-06	0.999999
11	4.84E-07	1
12	6.04E-08	1
13	6.97E-09	1
14	7.47E-10	1
15	7.47E-11	1

MSL Homework: 5.28, 5.29

Example 12. From noon to 4 p.m., the number of airplanes taking off from a certain runway during any one minute interval is a Poisson random variable with a mean of 1.5.

a) Find the probability that during a specific one minute interval no planes take off.
b) Find the probability that during a specific one minute interval two planes take off.
c) Find the probability that during a specific one minute interval at least one plane takes off.
d) Find the probability that exactly two planes take off during a specific three minute interval.

MSL Homework: 5.31, 5.32, 5.33
Hard Copy Homework: 5.37, 5.38

To use PHStat2

1. **Add-Ins > PHStat > Probability & Distributions > Poisson…**
2. Enter **1.5** for Mean/Expected No. of Events of Interest (i.e., λ , mean rate of occurrence).