CENG 222
Statistical Methods for Computer Engineering

Week 2

Chapter 3

Discrete Random Variables and Their Distributions

Random Variables

- A random variable (r.v.) associates a unique numerical value with each outcome in the sample space. It is a real (or integer)-valued function from a sample space Ω into real (or integer) numbers.

- Similar to events it is denoted by an uppercase letter (e.g., X or Y) and a particular value taken by a r.v. is denoted by the corresponding lowercase letter (e.g., x or y).

Examples

- Toss three coins. X = number of heads
- Pick a student from the Computer Engineering Department.

  X = age of the student

- Observe lifetime of a light bulb

  X = lifetime in minutes

- X may be discrete or continuous

Random variables and probabilities

- Consider the three coin tosses example where X = number of heads

- Each value of a random variable is an event
  - E.g. $P(X = 2)$

- We may compute probabilities of these events:
  - $P(X = 2) = P(\text{HHT}) + P(\text{HTH}) + P(\text{THH}) = 3/8$

Discrete random variables

- If the values a r.v. can take is finite or countably infinite then the r.v. is discrete

- Suppose that we can calculate $P(X=x)$ for every value of x. The collection of these probabilities can be viewed as a function of X. The probability mass function (p.m.f) of a discrete r.v. is given by

$$f_X(x) = P(X = x)$$

Distribution

- Collection of all probabilities related to X is the distribution of X.

- $f_X(x) = P(X = x)$ is the probability mass function (pmf)

- Cumulative distribution function (cdf)
  - $F(x) = P(X \leq x) = \sum_{y \leq x} f_X(y)$

- The set of all possible values of X is called the **support** of the distribution F.

Continuous random variables

- A r.v. *X* is continuous if it can take any value from one or more intervals of real numbers.

- We cannot use p.m.f because $P(X=x) = 0$ since there are infinitely many possible outcomes. Instead we use a *probability density function* (p.d.f), $f_X(x)$, such that the areas under the curve represent probabilities

Continuous random variable

- The p.d.f ($f_X(x)$) of a continuous r.v. X is the function that satisfies

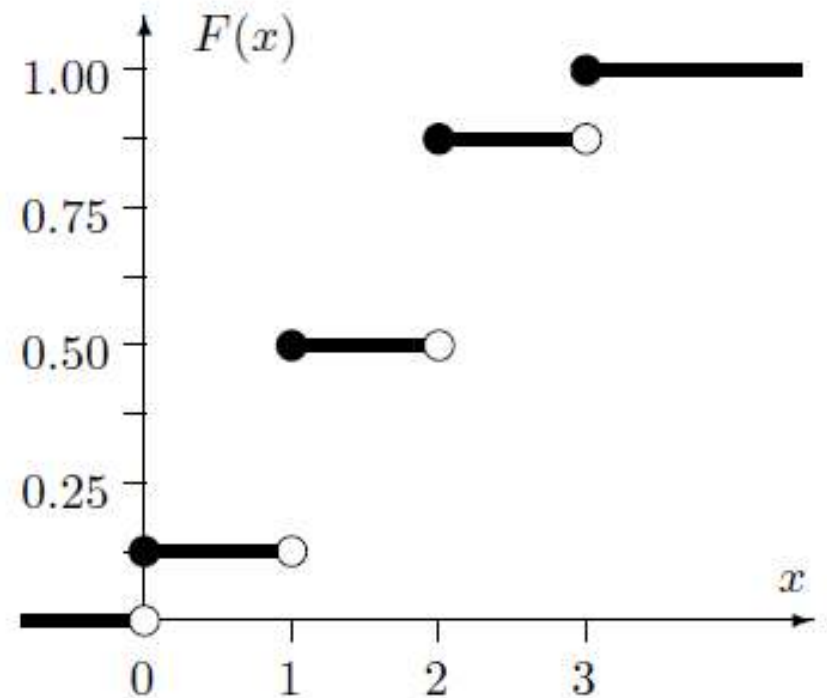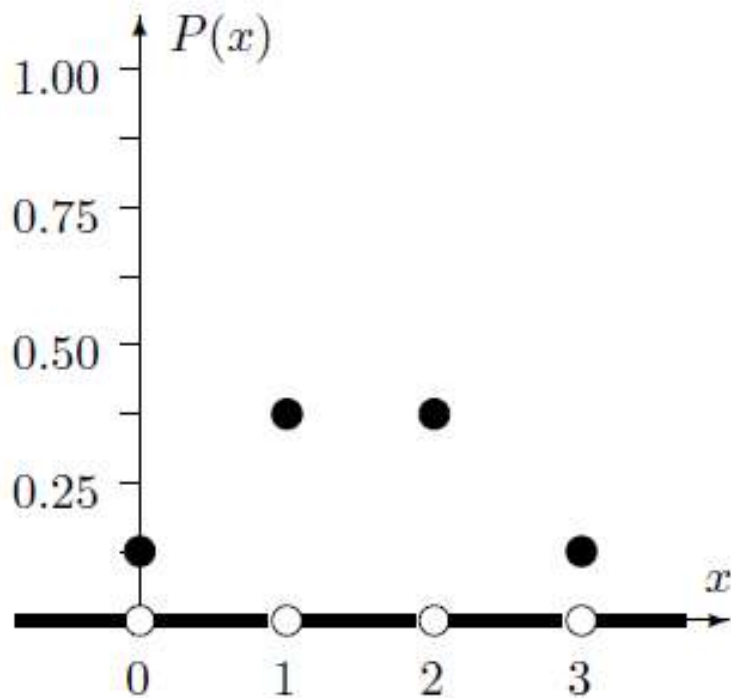$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \quad \text{for all } x$$

where $F_X(x)$ is the cumulative distribution function (c.d.f) of a r.v. X defined by

$$F_X(x) = P_X(X \leq x) \quad \text{for all } x$$

We will discuss continuous random variables in detail in Chapter 4
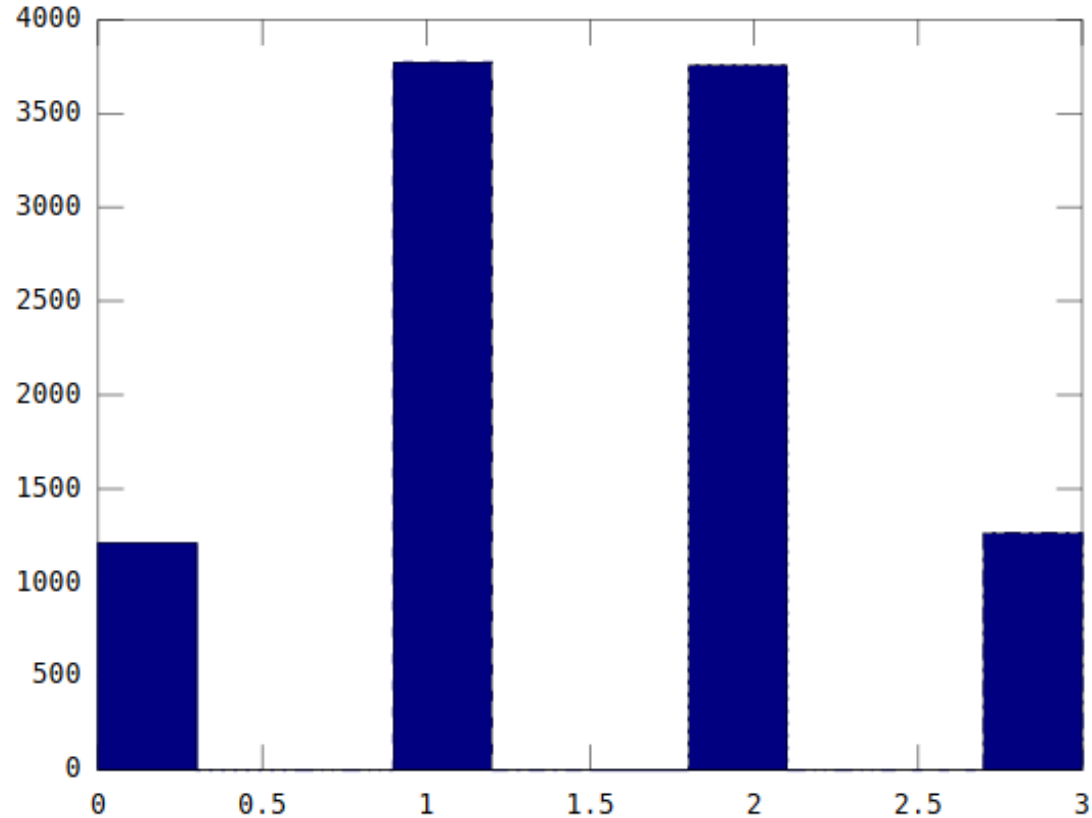
pmf and cdf

Back to toss of
3 coins

Octave Online

```
octave:1> N = 10000;
octave:2> U = rand(3,N);
octave:3> Y = (U<0.5);
octave:4> X = sum(Y);
octave:5> hist(X);
```

Example 3.3: Errors in two modules

x	$P_1(x)$	$P_2(x)$
0	0.5	0.7
1	0.3	0.2
2	0.1	0.1
3	0.1	0

- Y = total number of errors
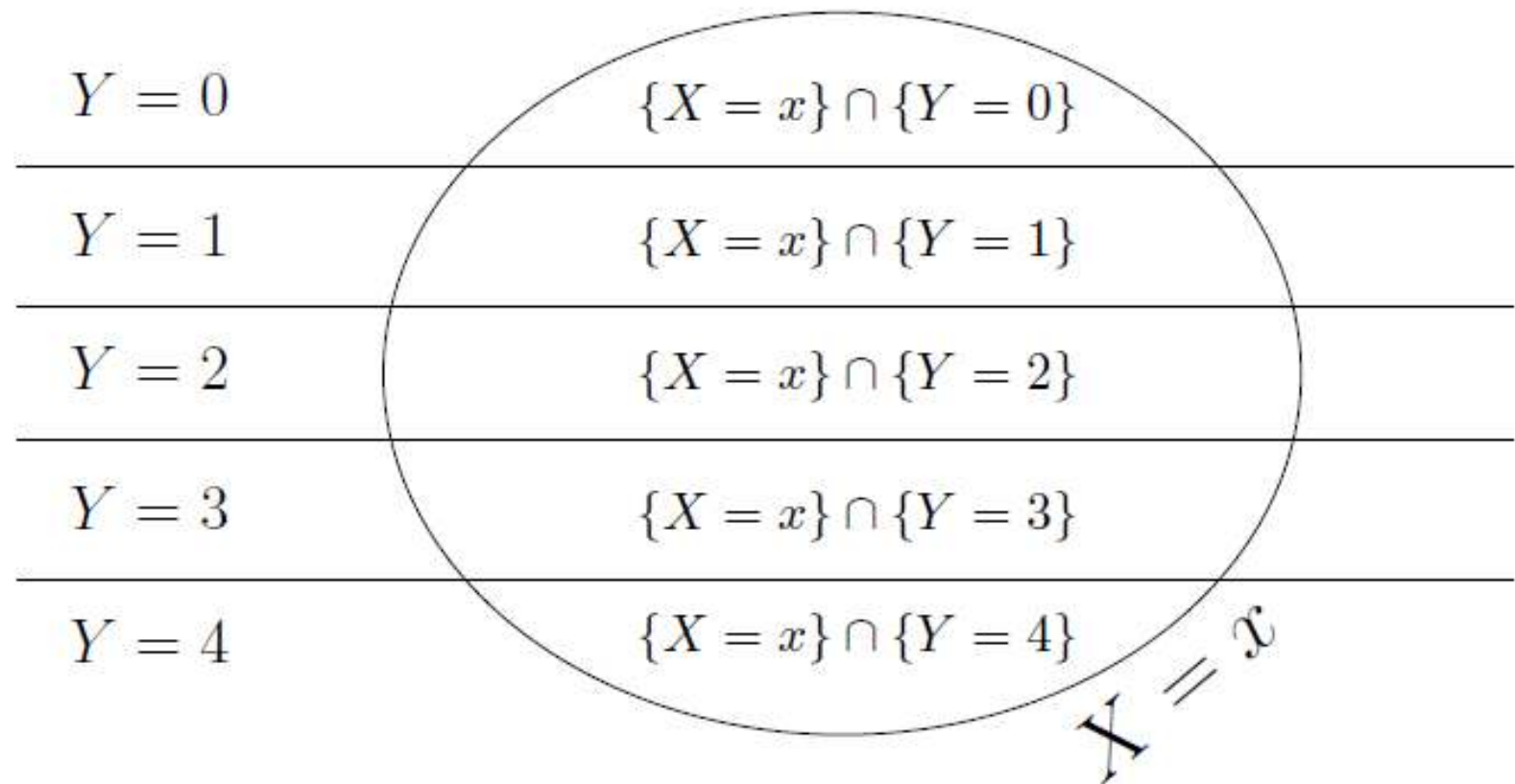- What is the pmf and cdf of Y?

Random Vectors

- Study of several random variables simultaneously

- If *X* and *Y* are two random variables, (*X*,*Y*) is a random vector

- The distribution of (*X*,*Y*) is called the *joint distribution* of *X* and *Y*.

- Given the joint distribution, the individual distribution of *X* and *Y* are called *marginal distributions*.

Joint p.m.f

- $f_{X,Y}(x, y) = P\big((X, Y) = (x, y)\big) = P(X = x \cap Y = y)$
- For all the possible and different pairs of (x,y) {(X,Y)=(x,y)} are mutually exclusive and exhaustive events.
  - In other words:
    - $\sum_x \sum_y f_{X,Y}(x, y) = 1$
- The joint distribution has all the information about the random variables
  - Marginal (i.e., individual) probabilities can be computed

Addition Rule

- $f_X(x) = P(X = x) = \sum_y f_{X,Y}(x, y)$

$Y = 0$	$\{X = x\} \cap \{Y = 0\}$
$Y = 1$	$\{X = x\} \cap \{Y = 1\}$
$Y = 2$	$\{X = x\} \cap \{Y = 2\}$
$Y = 3$	$\{X = x\} \cap \{Y = 3\}$
$Y = 4$	$\{X = x\} \cap \{Y = 4\}$

$X = x$

Independent Random Variables

- In general, the joint pmf cannot be computed from marginal pmfs

- This is only possible if the random variables are independent

    - $f_{X,Y}(x,y) = f_X(x)f_Y(y)$

- Given the joint pmf, we can check independence of two random variables by checking whether the product of marginal pmfs is equal to the joint pmf for every pair of (*x*,*y*)

Example 3.6

- Are *X* and *Y* independent?

$P_{(X,Y)}(x,y)$		y				$P_X(x)$
		0	1	2	3	
x	0	0.20	0.20	0.05	0.05	0.50
	1	0.20	0.10	0.10	0.10	0.50
$P_Y(y)$		0.40	0.30	0.15	0.15	1.00

Expected Value

- Expected value or the mean of a discrete r.v.

$$E(X) = \mu = \sum_x xf(x) = x_1 f(x_1) + x_2 f(x_2).......$$

- Expected value of a continuous r.v.

$$E(X) = \mu = \int_x xf(x)dx$$

- *E(X)* can be thought of as the center of gravity of the distribution of *X*

17

Expected Value

- Expected value of a function of X

$$E[g(x)] = \begin{cases} \sum_x g(x)f(x) & \text{if } X \text{ is discrete} \\ \int_x g(x)f(x)dx & \text{if } X \text{ is continuous} \end{cases}$$

- For example if $g(X) = X^2$

$$E[X^2] = \begin{cases} \sum_x X^2 f(x) & \text{if } X \text{ is discrete} \\ \int_x X^2 f(x)dx & \text{if } X \text{ is continuous} \end{cases}$$

Linearity of Expected Values

- For any two random variables X_1 and X_2 and any constants $c_1, c_2 \in \Re$

$$E(c_1 X_1 + c_2 X_2) = c_1 E(X_1) + c_2 E(X_2)$$

Expected value of a product

- In general $E(XY)$ is not equal to $E(X)E(Y)$

$$E(XY) = \int_y \int_x xy \, j(x, y) \, dx \, dy$$

- where $j(x,y)$ is the joint distribution which is NOT equal to $f(x)g(y)$ if X and Y are not independent.

Variance and Standard Deviation

- Variance of a r.v. is denoted by $Var(X)$ or σ^2

$$Var(X) = E[(X - \mu)^2] = E(X^2) - \mu^2$$

$$\sigma = \sqrt{Var(X)}$$

Example

- Experiment: two fair dice are tossed

- What is the expected value of the r.v. X?

  X = sum of two dice

- What is the variance?

Example

- Let X be a continuous r.v. with a p.d.f $f(X) = \frac{1}{2}$ for $0<x<2$.

- What is the expected value of X?

- What is the variance of X?

Covariance

- A measure of strength of a relationship between two random variables

- E.g., X = height of a person p, Y = weight of the same person

- X and Y are paired random variables

- Is there a relationship between them?

$$Cov(X,Y) = E[(X - \mu_x)(Y - \mu_y)] = E(XY) - E(X)E(Y)$$

Correlation

- The correlation (or correlation coefficient) is simply the covariance standardized to the range of [-1,1]

$$Corr(X,Y) = \rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$