Probability and Random Variable (2)

- **Random Variables**
  - Consider a function that assigns real numbers to events (outcomes) in Ω. Such real-valued function is a *random variable*.
  - E.g., when rolling two fair dice, define X as the sum of the two dice. Then, X is a random variable with $P\{X = 2\} = P\{(1,1)\}=1/36$, $P\{X = 3\} = P\{(1, 2), (2, 1)\}=2/36=1/18$, etc.
  - E.g., the salvage value of a machine, S, is \$1,500 if the market goes up (with probability 0.4) and \$1,000 if the market goes down (with probability 0.6). Then, S is a random variable with $P\{S = 1500\} = 0.4$ and $P\{S = 1000\} = 0.6$.
  - If the random variable can take on a limited number of values. Then, this is a *discrete random variable*. E.g., the random variable X representing the sum of two dice.
  - If the random variable can take on an uncountable number of values. Then, this is a continuous random variable. E.g., the random variable H representing height of an AUB student.
  - If X is a discrete random variable, the function $f_X(x) = P\{X = x\}$ is the *probability mass function* (pmf) of X .

- The function $F_X(x) = \mathrm{P}\{X \le x\} = \sum_{x_i \le x} f_X(x_i)$ is the *cumulative distribution function* (cdf) of X.
- E.g., for the random variable S representing salvage value of a machine above,

$$f_S(s) = \begin{cases} 0.6 & \text{if } s = 1000 \\ 0.4 & \text{if } s = 1500 \\ 0 & \text{othewise} \end{cases} \quad , \quad F_S(s) = \begin{cases} 0 & \text{if } \; s < 1000 \\ 0.6 & \text{if } \; 1000 \le s < 1500 \\ 1 & \text{if } \; s \ge 1500 \end{cases}.$$

- For a continuous random variable, X, the cdf is defined based on a function $f_X(x)$ called the *density function*, where

$$P\{X \le x\} = F_X(x) = \int_{-\infty}^{x} f_X(t)dt.$$

Fact. *For a discrete random variable* $\sum_{x_i} f_X(x_i) = 1$. *For a continuous random variable*, $\int_{-\infty}^{\infty} f_X(x) = 1$.

- **Independent Random variables**
  - Two random variables X and Y are said to be independent if

$$P\{X \le x, Y \le y\} = P\{X \le x\}P\{Y \le y\} = F_X(x)F_Y(y) \; .$$

- **Expectation of a random variable**
  - The *expectation* of a discrete random variable X is

$$E[X] = \sum_{x_i} x_i P\{X = x_i\} = \sum_{x_i} x_i f_X(x_i).$$

- The expectation of a continuous random variable X is

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx \, .$$

- The expectation of a random variable is the value obtained if the underlying experience is repeated for a number of times which is large enough and the resulting values are averaged.
- The expectation is "linear." That is, for two random variables X and Y, $E[aX + bY] = aE[X] + bE[Y]$.
- The expectation of a function of random variable X, $g(X)$, is

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx \, .$$

- An important measure is the n^{th} moment of X, $n = 1, 2, \ldots$

$$E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x) dx \, .$$

- **Measures of variability**
  - The *variance* of a random variable X is

$$\mathrm{Var}[X] = E[(X - E[X])^2] = E[X^2] - \left(E[X]\right)^2 .$$

  - The *standard deviation* of a random variable X is $\sigma_X = \sqrt{\mathrm{Var}[X]}$.
  - The *coefficient of variation* of a random variable X is $\mathrm{CV}[X] = \sigma_X / E[X]$.

- The variance (standard deviation) measures the spread of the random variable around the expectation.
- The coefficient of variation is useful when comparing variability of different alternatives.
- Note that Var[$aX+b$] $=a^2$ Var[X], for any real numbers a and b and random variable X .
- In addition, if X and Y are two *independent* random variables, then Var[$X + Y$] = Var[X] + Var[Y] .

- **The Bernoulli Random Variable**
  - Suppose an experiment can result in success with probability p and failure with probability (w.p.) $1-p$.
  - We define a *Bernoulli* random variable X as X =1 if the experiment outcome is a success and X = 0, otherwise.
  - The pmf of X is

$$f_X(x) = P\{X = x\} = \begin{cases} 1-p & \text{if } x = 0 \\ p & \text{if } x = 1 \end{cases}.$$

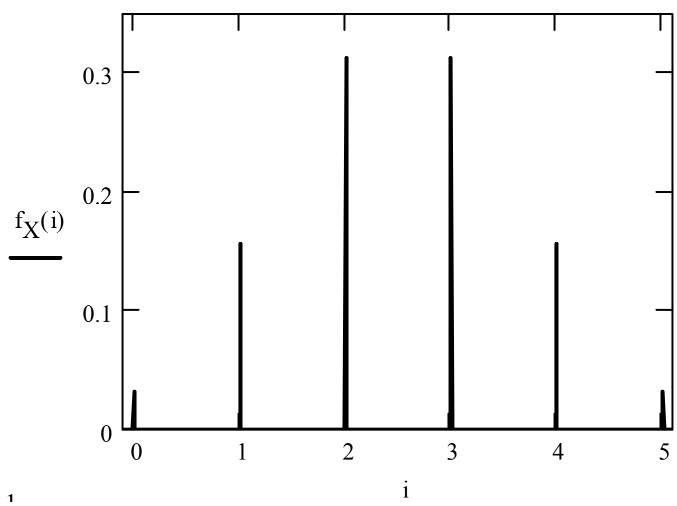
  - The expected value of X is E[X] = $0(1-p) + 1(p) = p$.
  - The second moment of X is E[X^2] = $0^2(1-p) + 1^2(p) = p$.
  - The variance of X is Var[X] = E[X^2] −(E[X])^2

$$= p - p^2 = p(1-p).$$

- **The Binomial Random Variable**
  - Consider n independent trials, each of which can results in a success w.p. p and failure w.p. $1-p$.
  - We define a Binomial random variable, X, as the number of successes in the n trials.
  - The pmf of X is defined as

$$f_X(i) = P\{X = i\} = \binom{n}{i} p^i (1-p)^{n-i} \text{ , } i = 0,1,\ldots n$$

where $\binom{n}{i} = \frac{n!}{(n-i)!i!}$.



  - Note that $\sum_{i=1}^{n} f_X(i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = 1, i = 0,1,\ldots n$.

Fact. *Let* X_i = 1, *i* =1, …, *n, if the* i^{th} *trial results in success and* X_i = 0, *otherwise. Then* $X = \sum_{i=1}^{n} X_i$.

- Note that X_i are independent and identically distributed (iid) Bernoulli random variable with parameter p.
- Therefore,

$$E[X] = \sum_{i=1}^{n} E[X_i] = np, \quad Var[X] = \sum_{i=1}^{n} Var[X_i] = np(1-p).$$

- **Example 9**
  - A fair coin is flipped 5 times.
  - What is the probability that two heads are obtained?
  - The number of heads, X, is binomial random variable with parameters $n = 5$ and $p = 0.5$. Then, the desired probability is $P\{X = 2\} = [5!/(2!\times3!)]\times(0.5)^2(0.5)^3 = 0.313$.

- **Example 10**
  - It is known that any item produced by a machine will be defective with probability 0.1 independently of any other item.
  - What is the probability that in a sample of three items at most one is defective?
  - If X is the number of defective items than X is binomial with parameters $n = 3$ and $p = 0.1$. Then, the desired probability is

$$P\{X \le 1\} = P\{X = 0\} + P\{X = 1\} = 0.9^3 + 3\times0.1\times0.9 = 0.972.$$

- **Example 11**
  - Suppose that an airplane engine will fail, when in flight, with probability $1 - p$ independently from engine to engine. Suppose that the airplane will make a successful flight if at least 50% of its engines remain operative.

- For what value of p is a two-engine plane preferable to a four-engine plane?
- The number of engines which do not fail, X, is a binomial random variable with $n = 4$ (2) for a four (two) engine plane and p.
- P{4-engine plane flies successfully}

  $= \text{P}\{X = 2\} + \text{P}\{X = 3\} + \text{P}\{X = 3\}$

  $$= \binom{4}{2} p^2(1-p)^2 + \binom{4}{3} p^3(1-p) + \binom{4}{4} p^4 = 6p^2(1-p)^2 + 4p^3(1-p) + p^4 \ .$$
- P{2-engine plane flies successfully} $= \text{P}\{X = 1\} + \text{P}\{X = 2\}$

  $= 2p(1-p) + p^2$.
- Then, the two-engine plane is preferred if

  $$2p(1-p) + p^2 \geq 6p^2(1-p)^2 + 4p^3(1-p) + p^4 \Rightarrow -3p^3 + 8p^2 - 7p + 2 \geq 0$$
  $$\Rightarrow (p-1)^2(2-3p) \geq 0 \Rightarrow p \leq 2/3 .$$