Supplementary Handout for Dis 5: Random Variables; Discrete Probability Distributions

1 Random Variables

- A random variable **assigns a number to each outcome** of an experiment.
  e.g. Let X be a random variable recording the outcome of rolling a fair, six-sided die, then

$$\begin{aligned} \text{Rolling a 1} &\quad\rightarrow\quad X=1 \\ \text{Rolling a 2} &\quad\rightarrow\quad X=2 \\ &\cdots \\ \text{Rolling a 6} &\quad\rightarrow\quad X=6 \end{aligned}$$

- There are two types of random variables:
  1. **Discrete random variable**: numbers assigned to the random variable are countable.
  2. **Continuous random variable**: numbers assigned to the random variable are NOT countable.

> Aside: what does it mean to be countable?
>
> (A pretty loose definition:) as long as you can sequentially count all the numbers assigned – even though it might take forever – then such series of numbers is considered as countable.

> Exercise. Is the following random variable discrete or continuous?
>
> 1. $X =$ whether the result from a fair coin flip is head or not
> 2. $X =$ amount of time it takes for a student to complete a 60-minute exam
> 3. $X =$ the number of rolls it takes to get a 6 from rolling a six-sided die

2 Discrete Probability Distributions

- **Probability distribution**: a table / formula / graph that describes the values of a random variable and the associated probabilities (at these values).

- In the case of **discrete probability distribution** for random variable X, we need some way to describe $P(x)$ at all possible x values.
    * Conditions that a discrete probability distribution needs to satisfy:
        1. $0 \leq P(x) \leq 1$ for all x
        2. $\sum_x P(x) = 1$
- If we already know that some other random variable Y yields outcome y, and want to describe all possible probabilities related to x after y outcome (that is, $P(X = x|Y = y)$ for all x), then **discrete conditional probability distribution** is appropriate.
    * Conditions that a discrete conditional probability distribution needs to satisfy:
        1. $0 \leq P(x|y) \leq 1$ for all x
        2. $\sum_x P(x|y) = 1$

- Since a probability distribution describes the probability at all possible outcomes, this is a representation of the **population**.
    - Thus, one can use a probability distribution to write down the calculation of some **parameters**:

Parameter Name	**Notation**	**Formula**	**Shortcut**
Expected value (mean)	$E(X) = \mu = \mu_X$	$\sum_x xP(x)$	-
Variance	$V(X) = \sigma^2 = \sigma_X^2$	$\sum_x (x - \mu)^2 P(x)$	$E(X^2)$ $-[E(X)]^2$
Standard deviation	$\sigma = \sigma_X$	$\sqrt{V(X)}$	-

    - When there is more than one probability distribution (say, distributions for random variables X and Y), then **parameters** on the relationship between two random variables can be expressed:

Parameter Name	**Notation**	**Formula**	**Shortcut**
Covariance	$Cov(X, Y) = \sigma_{XY}$	$\sum_x \sum_y (x - \mu_X)(y - \mu_Y)P(x, y)$	$E(XY)$ $-E(X)E(Y)$
Correlation (of coefficient)	$Corr(X, Y) = \rho_{XY}$	$\frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$	-

    - Common parameter operations, when the random variable is transformed in some way:
      (Let X and Y be random variables, a, b, c, d be constants.)
        * Expected value (mean):
            · $E(c) =$

            · $E(aX + b) =$

            · $E(X + Y) =$

* Variance:
    * $V(c) =$

    * $V(aX + b) =$

    * $V(X \pm Y) =$

* Covariance:
    * $Cov(a, b) =$

    * $Cov(X, X) =$

    * $Cov(aX + b, cY + d) =$

- Two examples of discrete probability distribution:

  1. **Binomial distribution**: distribution of success among n trials
     - Random variable $X \sim \text{Binomial}(n, p)$ if the following holds:
       * There are fixed number (n) of trials.
       * Each trial has two outcomes: success, and failure.
       * $P(\text{success}) = p$ is constant across all trials.
       * Trials are independent.
     - Once $X \sim \text{Binomial}(n, p)$ is established, then
       * $P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$ for $x = 0, 1, 2, \ldots, n$
       * $E(X) = np$
       * $V(X) = np(1-p)$
  2. **Poisson distribution**: distribution of success within a fixed time period / fixed interval, with success arriving at rate $\mu > 0$
     - Random variable $X \sim \text{Poisson}(\mu)$ if the following holds:
       * Number of success in any interval is independent of the number of success in any other interval.
       * Probability of a success in any equal-size interval is constant.
       * Probability of a success is proportional to the size of the interval.
       * Probability of more than one success in an interval approaches 0 as the interval becomes smaller.
     - Once $X \sim \text{Poisson}(\mu)$ is established, then
       * $P(x) = \frac{e^{-\mu} \mu^x}{x!}$ for $x = 0, 1, 2, \ldots$
       * $E(X) = \mu$
       * $V(X) = \mu$