MTH 224, Spring 2024

Instructor: Bahman Angoshtari

Lecture 8

Section 2.3: variance of discrete r.v.
Section 2.4: Bernoulli and binomial distributions.

8.1. Expectation of discrete random variables (continued)

- **Basic properties of expectation:**
  (1) $\mathbb{E}[a] = a$ for any constant a.
  (2) $\mathbb{E}[aX] = a\mathbb{E}[X]$ for any constant a and any random variable X.
  (3) $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$ for any constants a and b and any random variable X.
  (4) $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ for any constants a and b and any random variables X and Y.
- You can prove (1), (2), and (3) using the definition of expectation. To prove (4), we need the concept of **joint distributions** which we will learn later.

EXAMPLE 8.1. N assignments are returned to N students at random. On average, how many students get their own work back?

SOLUTION. This problem can be solved by a trick using **indicator random variables**. Let X be the number of student who get their own work. Let A_i, $i = 1, ..., N$, be the event that the i-th student gets his/her own work. Define $\mathbb{1}_{A_i}$, the **indicator random variable** for event A_i, as follows

$$\mathbb{1}_{A_i} = \begin{cases} 1 & \text{if } A_i \text{ occurs (the i-th student gets her/his own work)} \\ 0 & \text{otherwise (the i-th student gets someone else's work)} \end{cases}$$

It then follows that

$$X = \mathbb{1}_{A_1} + \mathbb{1}_{A_2} + \cdots + \mathbb{1}_{A_N} = \sum_{i=1}^{N} \mathbb{1}_{A_i}.$$

The question asks for $\mathbb{E}[X]$. By the properties of expected value, we have

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{N} \mathbb{1}_{A_i}\right] = \sum_{i=1}^{N} \mathbb{E}[\mathbb{1}_{A_i}].$$

Note that:

$$\mathbb{E}[\mathbb{1}_{A_i}] = 1 \times \mathbb{P}\Big(\mathbb{1}_{A_i} = 1\Big) + 0 \times \mathbb{P}\Big(\mathbb{1}_{A_i} = 0\Big) = \mathbb{P}(A_i) = \frac{1}{N}.$$

Finally, we obtain that

$$\mathbb{E}[X] = \sum_{i=1}^{N} \mathbb{E}[\mathbb{1}_{A_i}] = \sum_{i=1}^{N} \frac{1}{N} = \frac{N}{N} = 1.$$

On average, only one student gets their own HW back. Interestingly, this average number does not depend on the total number of students N.

8.2. Variance of discrete random variables

- The variance of a random variable provides information of how spread the values taken by the random variable are.

DEFINITION 8.2. Let X be a discrete random variable with $\operatorname{supp}(X) = \{x_1, x_2, \dots\}$ and pmf $p_X(x)$. The **variance** of X, denoted by $\operatorname{Var}(X)$, is

$$\operatorname{Var}(X) = \mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)^2\Big] = \sum_{x\in\mathbb{R}} \big(x - \mathbb{E}[X]\big)^2 p_X(x) = \big(x_1 - \mathbb{E}[X]\big)^2 p_X(x_1) + \big(x_2 - \mathbb{E}[X]\big)^2 p_X(x_2) + \dots$$

The **standard deviation** of X is $\sqrt{\operatorname{Var}(X)}$. It is common to use the notation σ_X^2 for $\operatorname{Var}(X)$, and use σ_X to denote the standard deviation of X.

- Note that $\operatorname{Var}(X)$ is the **weighted average** of the squared distances between values in the support of X and $\mathbb{E}[X]$, with weights equal to the probability of each value.
- If $\mathbb{E}[X]$ does not exist, then $\operatorname{Var}(X)$ does not exist as well. It may be the case that $\mathbb{E}[X]$ exists, but the series in the definition of $\operatorname{Var}(X)$ diverges. In that case, we also say that $\operatorname{Var}(X)$ does not exist.
- **Basic properties of variance:**
  (1) $\operatorname{Var}(a) = 0$ for any constant a.
  (2) $\operatorname{Var}(aX) = a^2\operatorname{Var}(X)$ for any constant a and any random variable X.
  (3) $\operatorname{Var}(aX + b) = a^2\operatorname{Var}(X)$ for any constants a and b and any random variable X.
  (4) $\operatorname{Var}(aX + bY) = a^2\operatorname{Var}(X) + b^2\operatorname{Var}(Y) + 2ab\operatorname{Cov}(X,Y)$ for any constants a and b and any random variable X and Y.
  (5) $\operatorname{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ for any random variable X. In many occasions, it is easier to calculate variance using this formula instead of the definition!
- You can prove (1), (2), and (3) using the definition of variance. To prove (4), we need the concept of **joint distributions** and **covariance** which we will learn later.
- Property (5) is shown as follows. For simplicity, let us define $\mu = \mathbb{E}[X]$. Note that μ is a constant. By the definition of variance,

$$\begin{aligned}\operatorname{Var}(X) &= \mathbb{E}\Big[\big(X-\mu\big)^2\Big] = \mathbb{E}\Big[X^2 + \mu^2 - 2\mu X\Big] \\ &= \mathbb{E}\Big[X^2\Big] + \mathbb{E}\Big[\mu^2\Big] - 2\mu\mathbb{E}\Big[X\Big] \\ &= \mathbb{E}\Big[X^2\Big] + \mu^2 - 2\mu^2 = \mathbb{E}\Big[X^2\Big] - (\mathbb{E}[X])^2 .\end{aligned}$$

EXAMPLE 8.3. N assignments are returned to N students at random. We calculated that, on average, 1 student get their own work back. What is the standard deviation of the number of students who get their work back? Assume that $N \geq 2$.

SOLUTION. **Indicator random variables** are again helpful. Let X be the number of student who get their own work. Let A_i, $i = 1, ..., N$, be the event that the i-th student gets his/her own work. As we saw before,

$$X = \mathbb{1}_{A_1} + \mathbb{1}_{A_2} + \cdots + \mathbb{1}_{A_N} = \sum_{i=1}^{N} \mathbb{1}_{A_i}.$$

We are interested in Var(X). It is easier to use the alternative formula for variance:

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \Big(\mathbb{E}[X]\Big)^2 = \mathbb{E}[X^2] - 1.$$

So, we need to calculate $\mathbb{E}[X^2]$. Note that:

$$\begin{aligned} X^2 &= \Big(\mathbb{1}_{A_1} + \mathbb{1}_{A_2} + \cdots + \mathbb{1}_{A_N}\Big)^2 = \left(\sum_{i=1}^{N} \mathbb{1}_{A_i}\right)^2 \\ &= \sum_{i=1}^{N} (\mathbb{1}_{A_i})^2 + \sum\sum_{i\neq j} \mathbb{1}_{A_i}\mathbb{1}_{A_j} \\ &= \sum_{i=1}^{N} \mathbb{1}_{A_i} + \sum\sum_{i\neq j} \mathbb{1}_{A_i\cap A_j}. \end{aligned}$$

In the last step, we use the facts that $(\mathbb{1}_{A_i})^2 = \mathbb{1}_{A_i}$ and that $\mathbb{1}_{A_i}\mathbb{1}_{A_j} = \mathbb{1}_{A_i\cap A_j}$. By taking the expected value of both sides, we obtain that

$$\mathbb{E}[X^2] = \mathbb{E}\left[\sum_{i=1}^{N} \mathbb{1}_{A_i} + \sum\sum_{i\neq j} \mathbb{1}_{A_i\cap A_j}\right] = \sum_{i=1}^{N} \mathbb{E}[\mathbb{1}_{A_i}] + \sum\sum_{i\neq j} \mathbb{E}[\mathbb{1}_{A_i\cap A_j}].$$

From the calculation for $\mathbb{E}[X]$, we already know that $\mathbb{E}[\mathbb{1}_{A_i}] = \frac{1}{N}$. We also have:

$$\begin{aligned} \mathbb{E}[\mathbb{1}_{A_i\cap A_j}] &= 1 \times \mathbb{P}\Big(\mathbb{1}_{A_i\cap A_j} = 1\Big) + 0 \times \mathbb{P}\Big(\mathbb{1}_{A_i\cap A_j} = 0\Big) \\ &= \mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j|A_i) = \frac{1}{N}\frac{1}{N-1} = \frac{1}{N(N-1)}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{i=1}^{N} \mathbb{E}[\mathbb{1}_{A_i}] + \sum\sum_{i\neq j} \mathbb{E}[\mathbb{1}_{A_i\cap A_j}] \\ &= \sum_{i=1}^{N} \frac{1}{N} + \sum\sum_{i\neq j} \frac{1}{N(N-1)} \\ &= \frac{N}{N} + \frac{N(N-1)}{N(N-1)} = 2. \end{aligned}$$

Finally, we find the variance of X:

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \Big(\mathbb{E}[X]\Big)^2 = 2 - 1^2 = 1.$$

So, the standard deviation is $\sqrt{\mathrm{Var}[X]} = 1$.

8.3. Binomial distribution

- A **Bernoulli trial** is a simple experiment with two possible outcomes, which are usually called success and failure. Example: flip a coin with getting H as success.
- Consider an experiment consisting of n independent Bernoulli trials, each having success with probability p and failure with probability $1 - p$ (here, independence means that the outcomes of each trial are independent of the outcomes of other trials).

- Let X be the number of successes. The pmf of X is then:

$$p_X(k) = \mathbb{P}(X = k) = \mathbb{P}(k \text{ successes in } n \text{ trials}) = \binom{n}{k} p^k (1-p)^{n-k}$$

Explanation: $\binom{n}{k}$ is the number of outcomes in which there are exactly k successes and $n-k$ failures. The probability of one such outcome is $p^k (1-p)^{n-k}$.

Definition 8.4. A random variable X is said to be a **binomial random variable** with parameters n, p if its pmf is given by

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Notation: $X \sim \mathrm{Bin}(n, p)$.

- Note that by the binomial theorem: $\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1^n = 1$.

Example 8.5. Jack hits his target 70% of the time. What is the probability that he hits his target in at least 8 out of 10 shots?

Solution. $X = \#$ of hits. $X \sim \mathrm{Bin}(10, 0.7)$. Then

$$\mathbb{P}(X \geq 8) = p_X(8) + p_X(9) + p_X(10) = \binom{10}{8} 0.7^8 \cdot 0.3^2 + \binom{10}{9} 0.7^9 \cdot 0.3^1 + \binom{10}{10} 0.7^{10} 0.3^0 \approx 0.383.$$