Lecture No. 5
Introduction to Random Variables

Azeem Iqbal

University of Engineering and Technology, Lahore
(Faisalabad Campus)

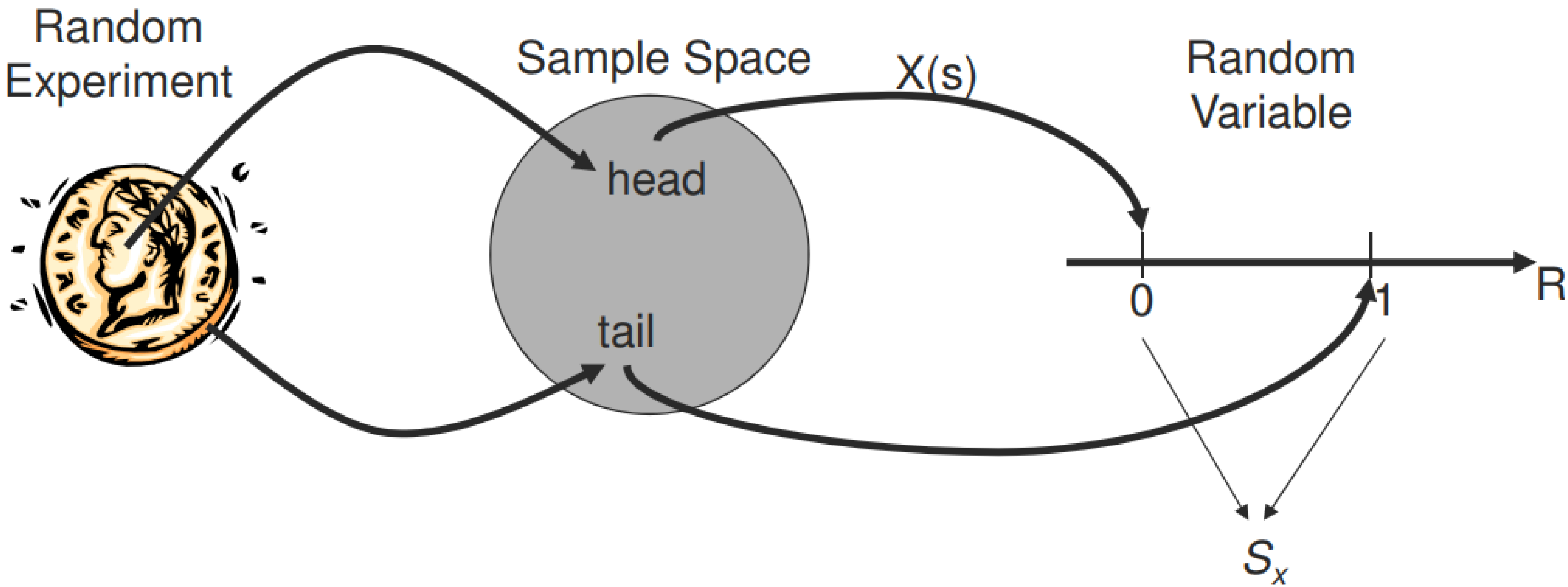
Definition of a Random Variable

Random variable is a mapping from an outcome s of a random experiment to a real number.

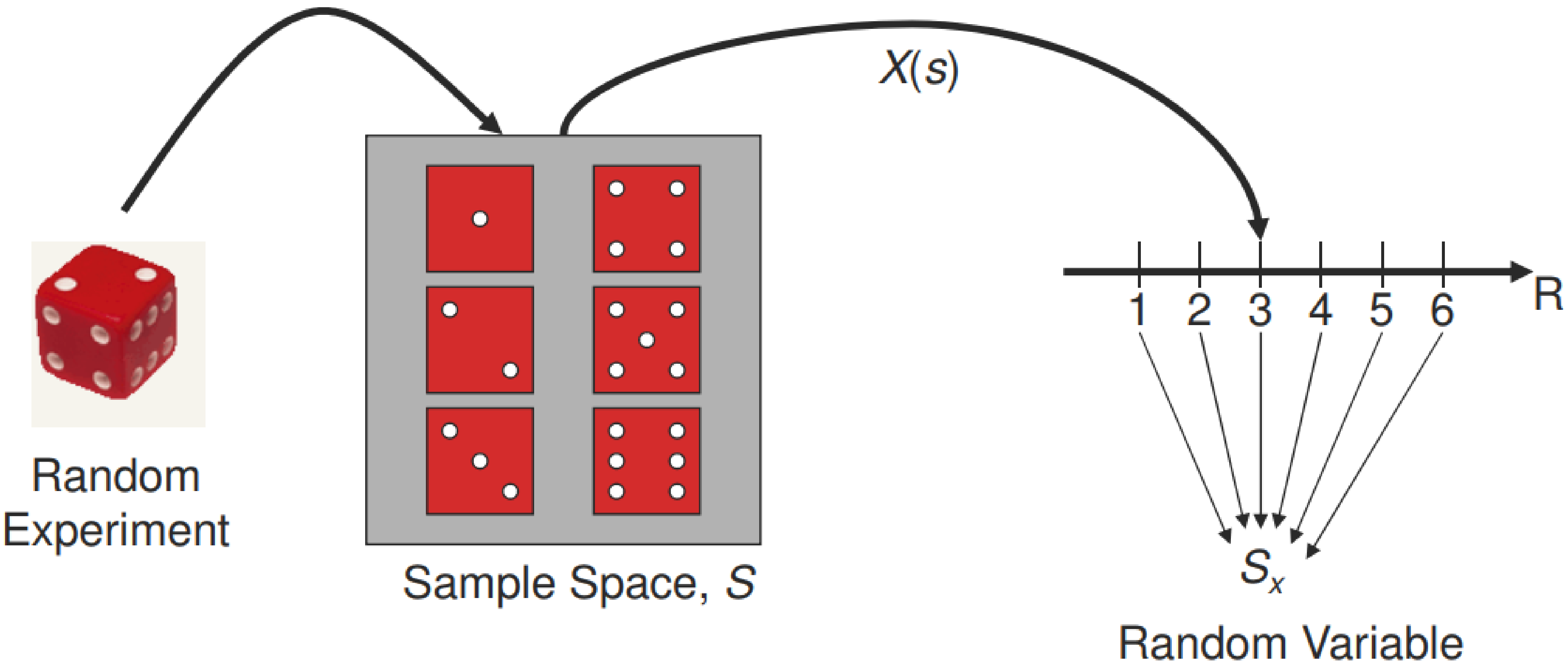
$$X : \underbrace{S}_{\text{domain}} \to \underbrace{S_X}_{\text{range}} \subset \mathbb{R}$$

S_X is called the image of X

Definition of a Random Variable

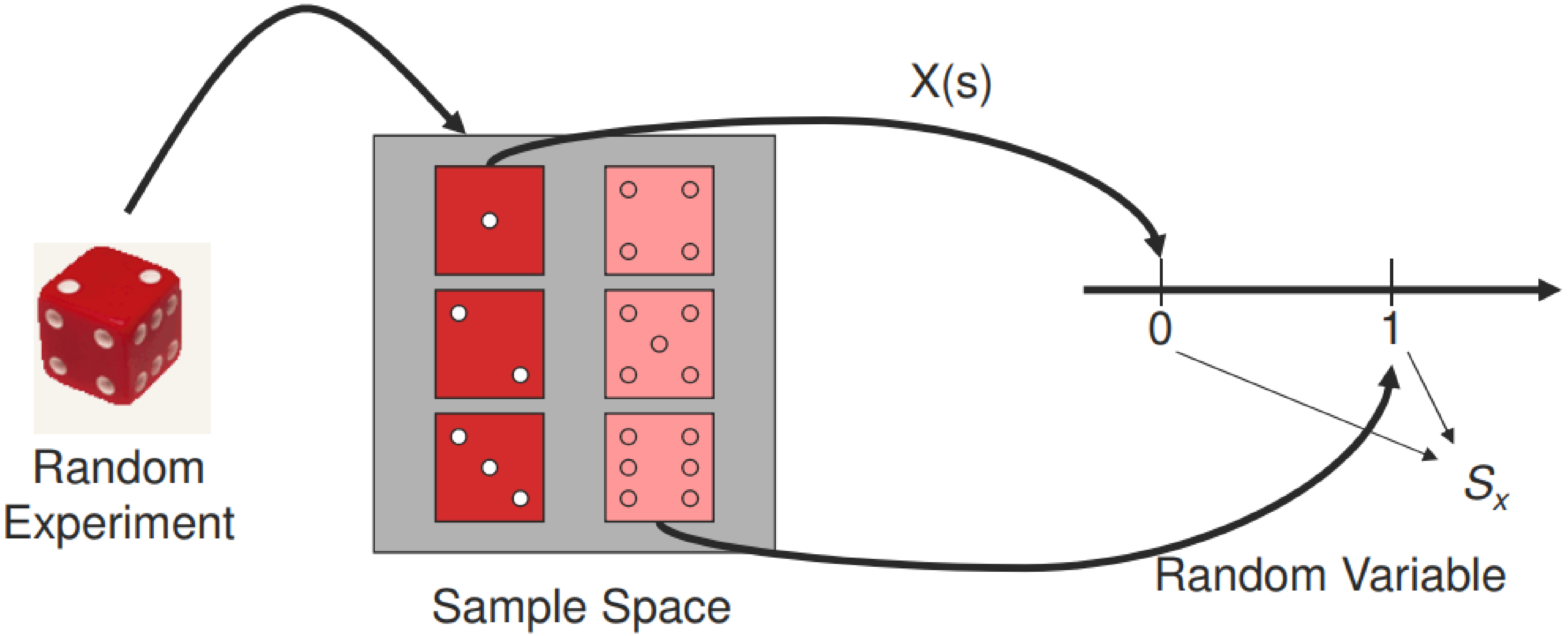


Definition of a Random Variable



Definition of a Random Variable

More than one outcomes can be mapped to the same real number to a Random Experiment.



Types of Random Variables

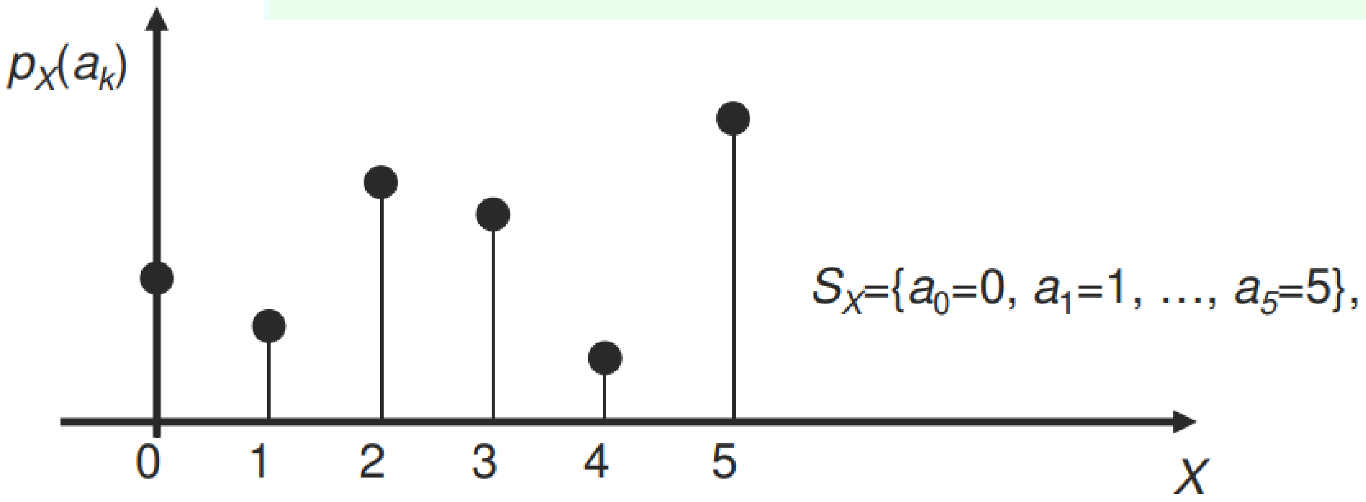
- **Discrete Random Variables:** have a countable (finite or infinite) image
  - $S_X = \{0,1\}$
  - $S_X = \{1,2,3,4 \;\dots\dots\dots\}$
- **Continuous Random Variables:** have an uncountable image
  - $S_X = (0,1]$
  - $S_X = \mathbb{R}$
- **Mixed Random Variables:** have an image which contains continuous and discrete parts
  - $S_X = \{0\} \cup (0,2]$

Discrete Random Variables

Probability Mass Function (PMF)

- The **Probability Mass Function** (PMF) provides the probability of a particular point in the sample space of a discrete random variable (R.V).
- For a countable $S_X = \{a_0, a_2, \dots, a_n\}$, the PMF is the set of probabilities.

$$p_X(a_k) = \Pr\{X = a_k\}, \;\; k = 1, 2, \dots, n$$



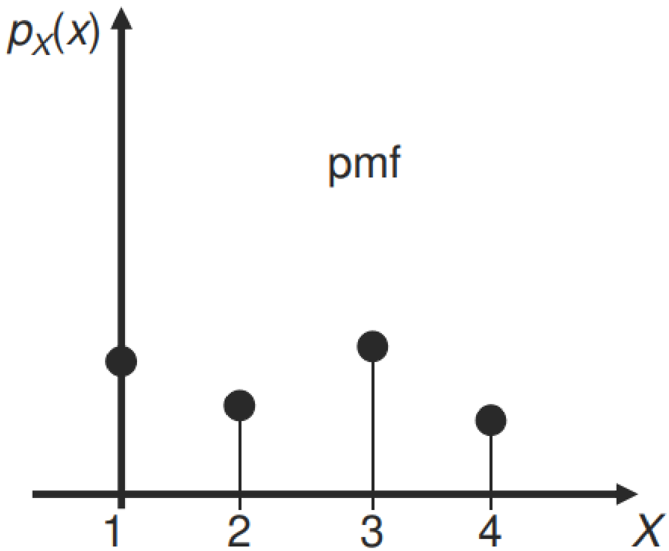
Properties of a PMF

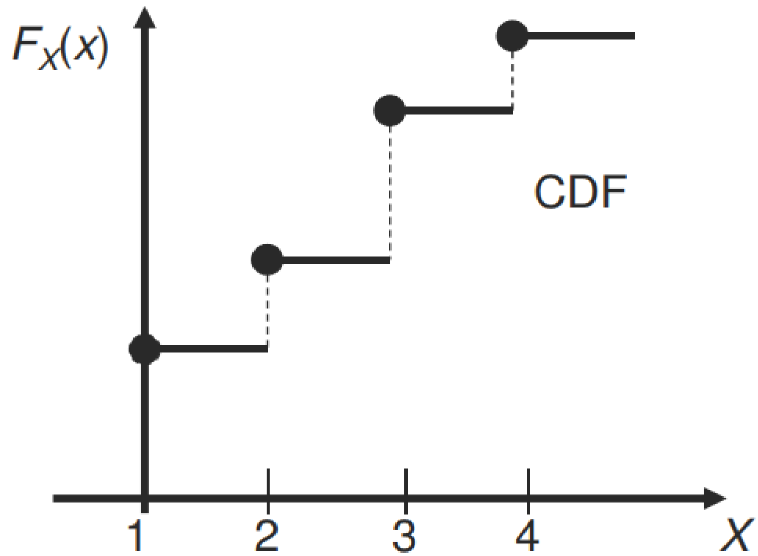
- **P1:** $0 \leq P_X(x) \leq 1$
- **P2:** $\sum_{x \in S_X} P_X(x) = 1$

Cumulative Distribution Function (CDF) of Discrete Random Variable

- The Cumulative Distribution Function (CDF) for a discrete R.V is defined as:

$$F_X(t) = \Pr\{X \le t\} = \sum_{x \le t} p_X(x)$$





Cumulative Distribution Function (CDF) of Discrete Random Variable

- CDF can be used to find the probability of a range of values in a R.Vs image:

$$\Pr\{a < X \le b\} = \Pr\{X \le b\} - \Pr\{X \le a\}$$
$$= F_X(b) - F_X(a)$$

Properties of CDF

- **P1:** $0 \leq F_X(x) \leq 1$
- **P2:** $\mathrm{a} \leq b \ \rightarrow F_X(a) \leq F_X(b)$
- **P3:** $\lim_{x \to -\infty} F_X(x) = 0$
- **P4:** $\lim_{x \to \infty} F_X(x) = 1$
- **P5:** $F_X(x_{i+1}) = \mathrm{F_X(x_i)} + \mathrm{p_X(x_{i+1})}$

Expected Value of a Discrete Random Variable

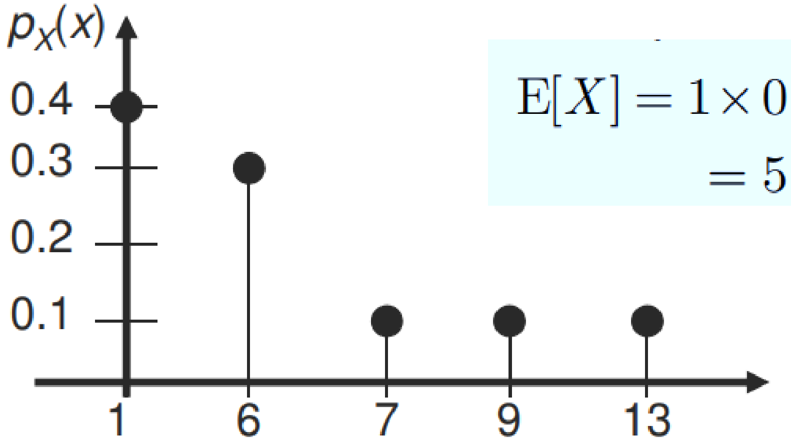
- The expected value, expectation or mean of a discrete R.V is the "average" value of the random variable
- What is the average value of the following random variable whose image is $S_X = \{1, 6, 7, 9, 13\}$

If your answer is **7.2** then you assumed that all of the values in the R.V image have equal probabilities

$$1 \times \frac{1}{5} + 6 \times \frac{1}{5} + 7 \times \frac{1}{5} + 9 \times \frac{1}{5} + 13 \times \frac{1}{5} = (1 + 6 + 7 + 9 + 13)\frac{1}{5} = 7.2$$

Expected Value of a Discrete Random Variable

- Expectation assigns appropriate weights to each value in the R.V's image
- Expectation is computed by using the probability of each point in the R.V's image as the weight of that point
- Hence, the expectation of the random variable given below is:



$$
\begin{aligned}
\mathrm{E}[X] &= 1\times 0.4 + 6\times 0.3 + 7\times 0.1 + 9\times 0.1 + 13\times 0.1 \\
&= 5.1
\end{aligned}
$$

Expected Value of a Discrete Random Variable

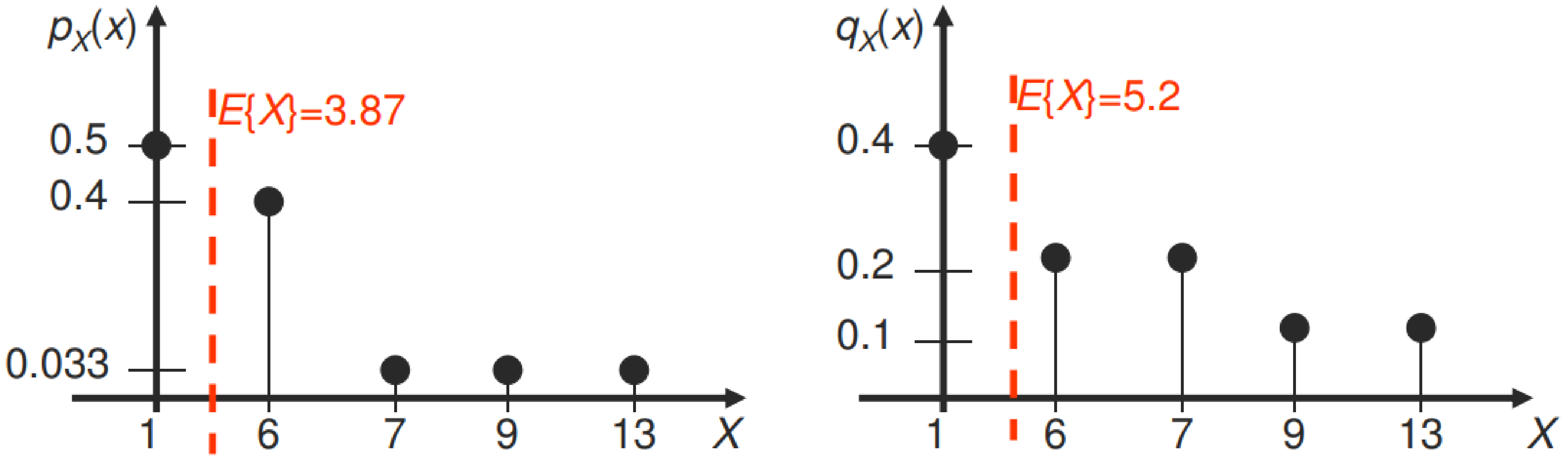
- Mathematically, the expected value of a discrete random variable is:

$$\mathrm{E}[X] = \mu_X = \sum_{x \in S_X} x \times p_X(X = x)$$

- In some cases, the expected value does not converge
- In such cases, we say that the expected value does not exist

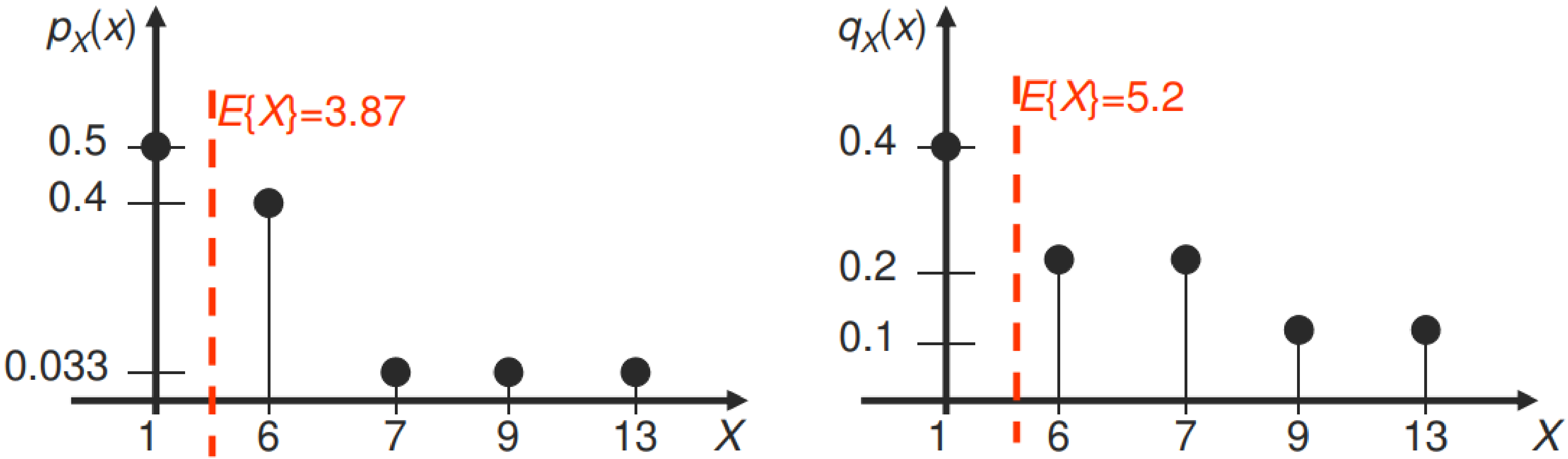
Variance of a Discrete Random Variable

- Variance of a R.V is a measure of “the amount of variation of a R.V around its mean”
- Intuitively, which of the following discrete R.Vs has a higher variance?



Variance of a Discrete Random Variable

- Variance of a R.V is a measure of "the amount of variation of a R.V around its mean"
- Intuitively, which of the following discrete R.Vs has a higher variance?



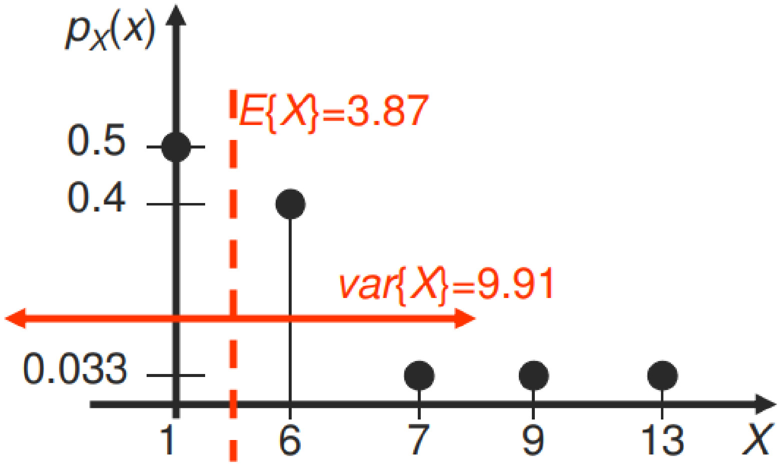
$q_X(x)$ has a higher variance because it varies more around its mean than $p_X(x)$

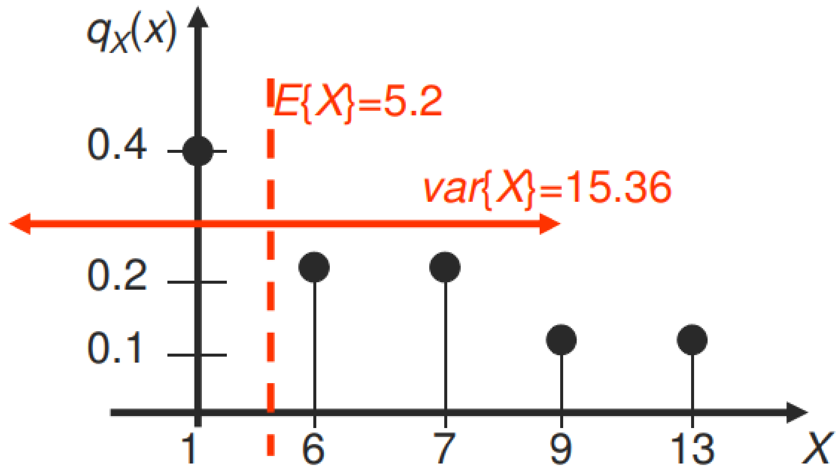
Variance of a Discrete Random Variable

- Can you come up with an expression / function that would be a good measure of variance of a random variable.
  - Should measure the spread around the 'center' of the PMF
  - Measure should be shift invariant
  - Greater where outcomes are further spaced apart
  - Should take into account different frequencies of occurrence of different outcomes
- Mathematically, the variance of a discrete rv is defined as:

$$\mathrm{var}[X] = \sigma_X^2 = \sum_{x \in S_X} \left(x - \mathrm{E}[X]\right)^2 p_X(x)$$

Variance of a Discrete Random Variable





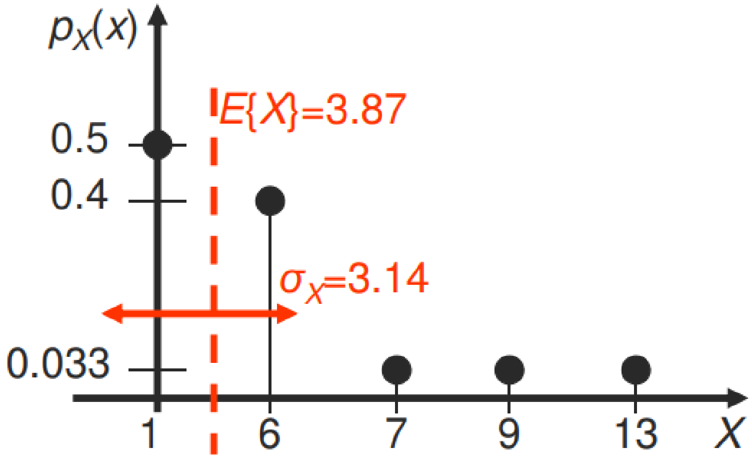
$$\begin{aligned}\operatorname{var}[X] &= (1-3.87)^2 \times 0.5 + \\ &(6-3.87)^2 \times 0.4 + (7-3.87)^2 \times 0.033 + \\ &(9-3.87)^2 \times 0.033 + (13-3.87)^2 \times 0.033 \\ &= 9.91\end{aligned}$$

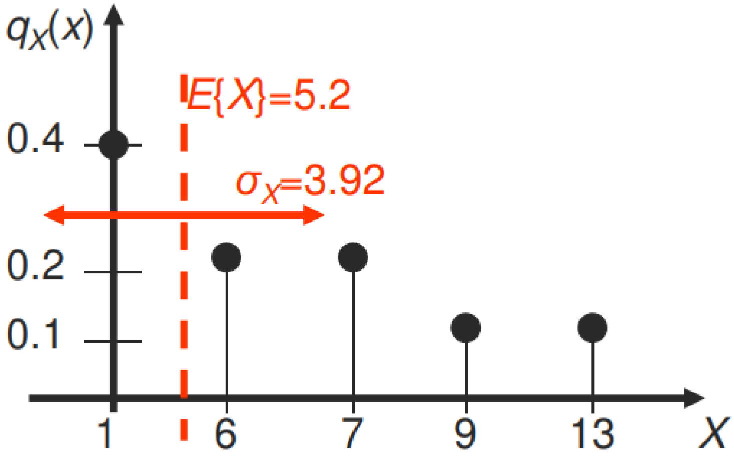
$$\begin{aligned}\operatorname{var}[X] &= (1-5.2)^2 \times 0.4 + \\ &(6-5.2)^2 \times 0.2 + (7-5.2)^2 \times 0.2 + \\ &(9-5.2)^2 \times 0.1 + (13-5.2)^2 \times 0.1 \\ &= 15.36\end{aligned}$$

Standard Deviation of a Random Variable

- In many scenarios, we use the square root of the variance called its standard deviation

$$\sigma_X = \sqrt{\mathrm{var}[X]}$$





$$\sigma_X = \sqrt{\mathrm{var}[X]} = \sqrt{9.91} = 3.14$$

$$\sigma_X = \sqrt{\mathrm{var}[X]} = \sqrt{15.36} = 3.92$$