LECTURE 5: Discrete random variables: probability mass functions and expectations

- Random variables: the idea and the definition
  - **Discrete:** take values in finite or countable set
- Probability mass function (PMF)
- Random variable examples
  - Bernoulli
  - Uniform
  - Binomial
  - Geometric
- Expectation (mean) and its properties
  - The expected value rule
  - Linearity

Random variables: the idea

Random variables: the formalism

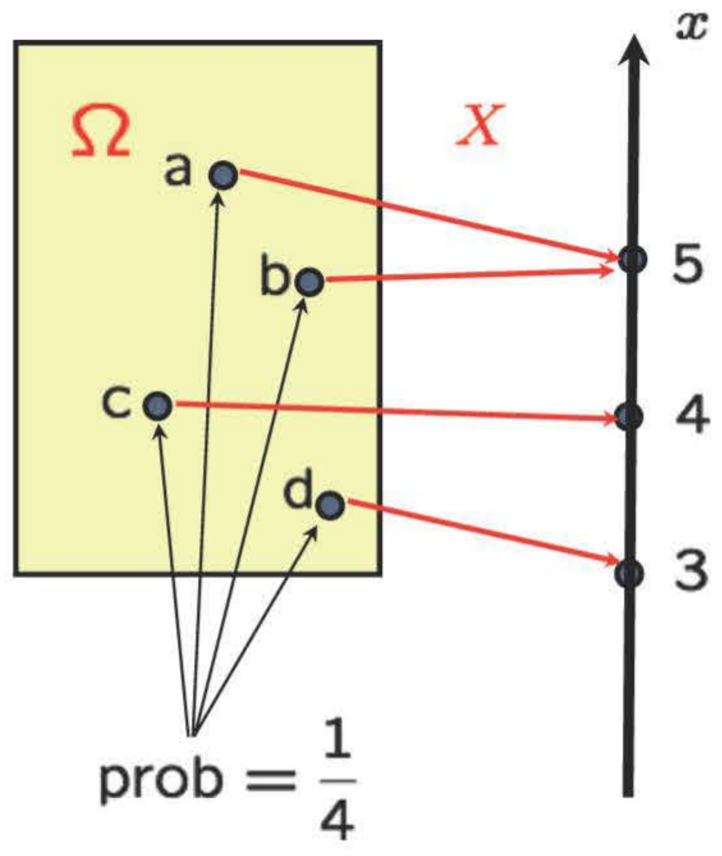
- A random variable ("r.v.") associates a value (a number) to every possible outcome
- Mathematically: A function from the sample space Ω to the real numbers
- It can take discrete or continuous values

Notation: random variable X numerical value x

- We can have several random variables defined on the same sample space
- A function of one or several random variables is also a random variable
  - meaning of $X+Y$:

Probability mass function (PMF) of a discrete r.v. X

- It is the "probability law" or "probability distribution" of X
- If we fix some x, then "$X = x$" is an event

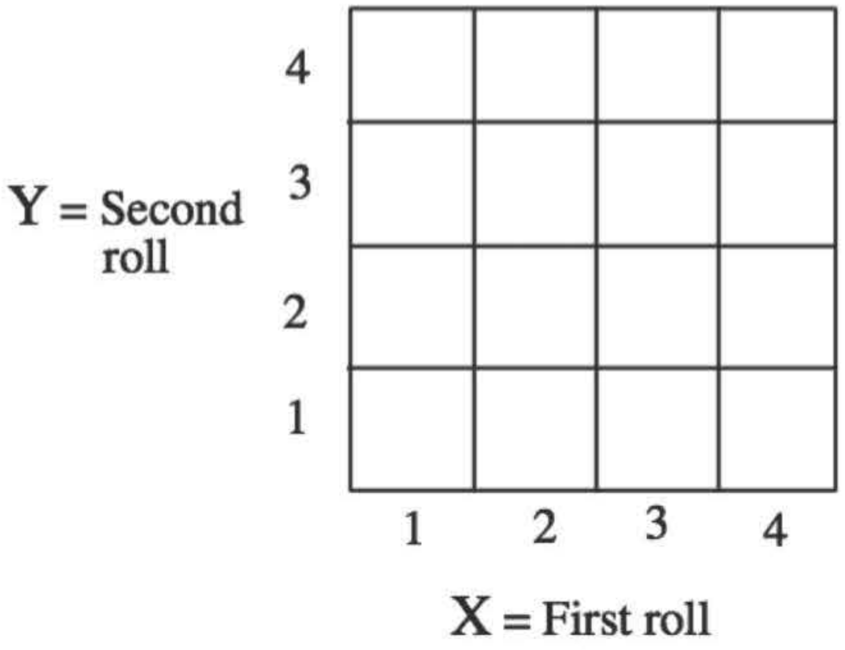


$$p_X(x) = \mathbf{P}(X = x) = \mathbf{P}\Big(\{\omega \in \Omega \text{ s.t. } X(\omega) = x\}\Big)$$

- **Properties:** $p_X(x) \geq 0$ $\qquad \sum_x p_X(x) = 1$

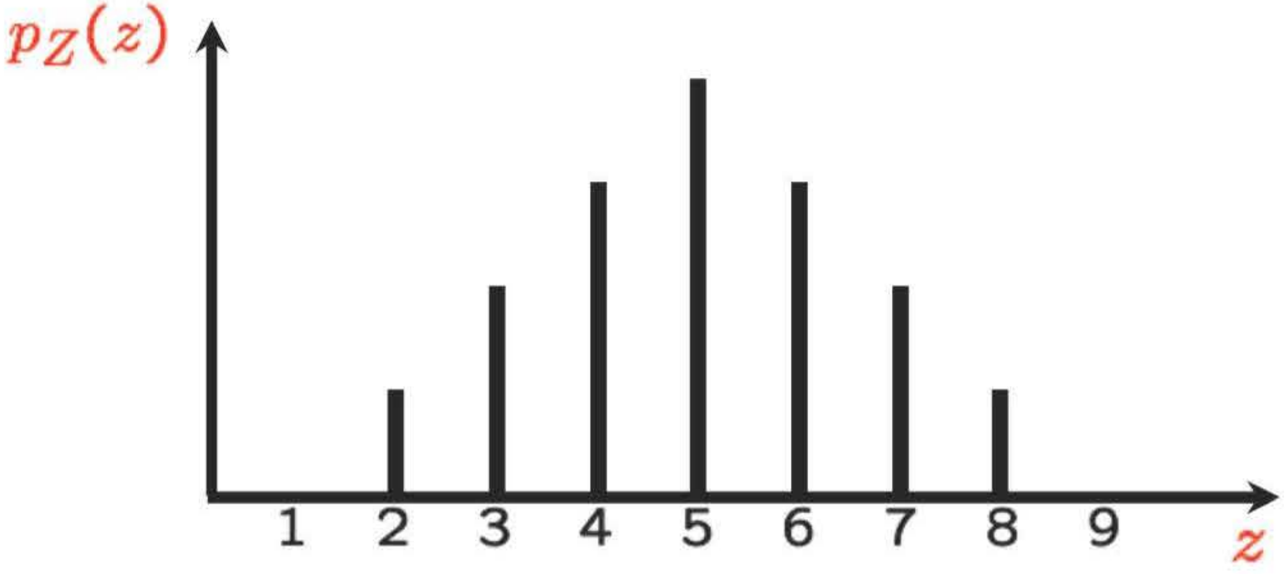
PMF calculation

- Two rolls of a tetrahedral die
- Let every possible outcome have probability 1/16



$Z = X + Y$ Find $p_Z(z)$

- repeat for all z:
  - collect all possible outcomes for which Z is equal to z
  - add their probabilities



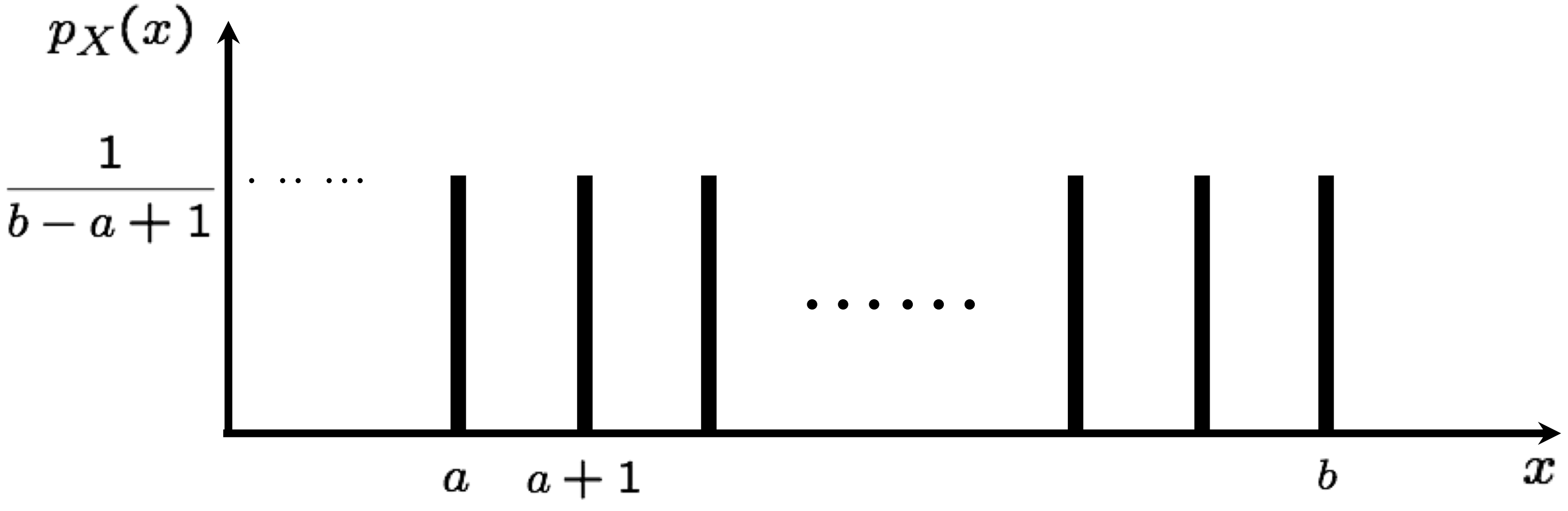
The simplest random variable: Bernoulli with parameter $p \in [0, 1]$

$$X = \begin{cases} 1, & \text{w.p. } p \\ 0, & \text{w.p. } 1-p \end{cases}$$

- Models a trial that results in success/failure, Heads/Tails, etc.
- Indicator r.v. of an event A: $I_A = 1$ iff A occurs

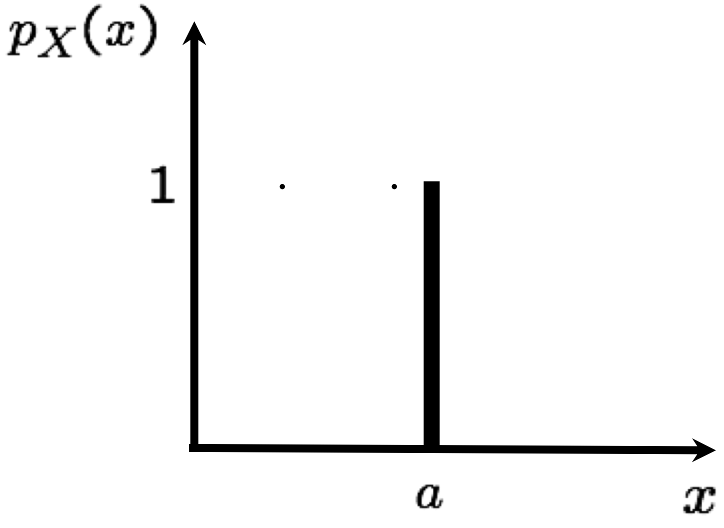
Discrete uniform random variable; parameters a, b

- **Parameters:** integers a, b; $a \leq b$
- **Experiment:** Pick one of $a, a+1, \ldots, b$ at random; all equally likely
- **Sample space:** $\{a, a+1, \ldots, b\}$
- **Random variable X:** $X(\omega) = \omega$
- **Model of:** complete ignorance



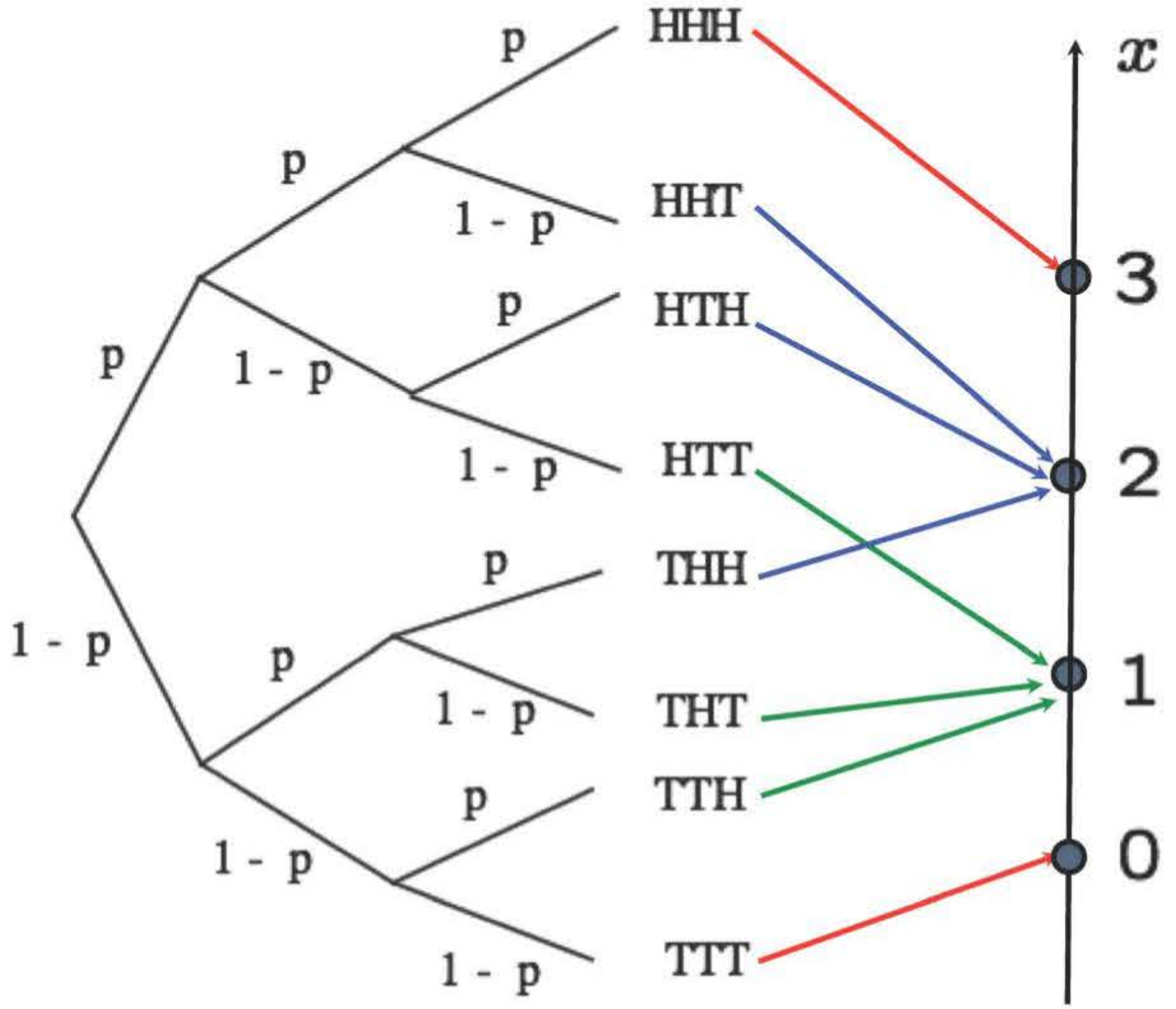
Special case: $a = b$

constant/deterministic r.v.

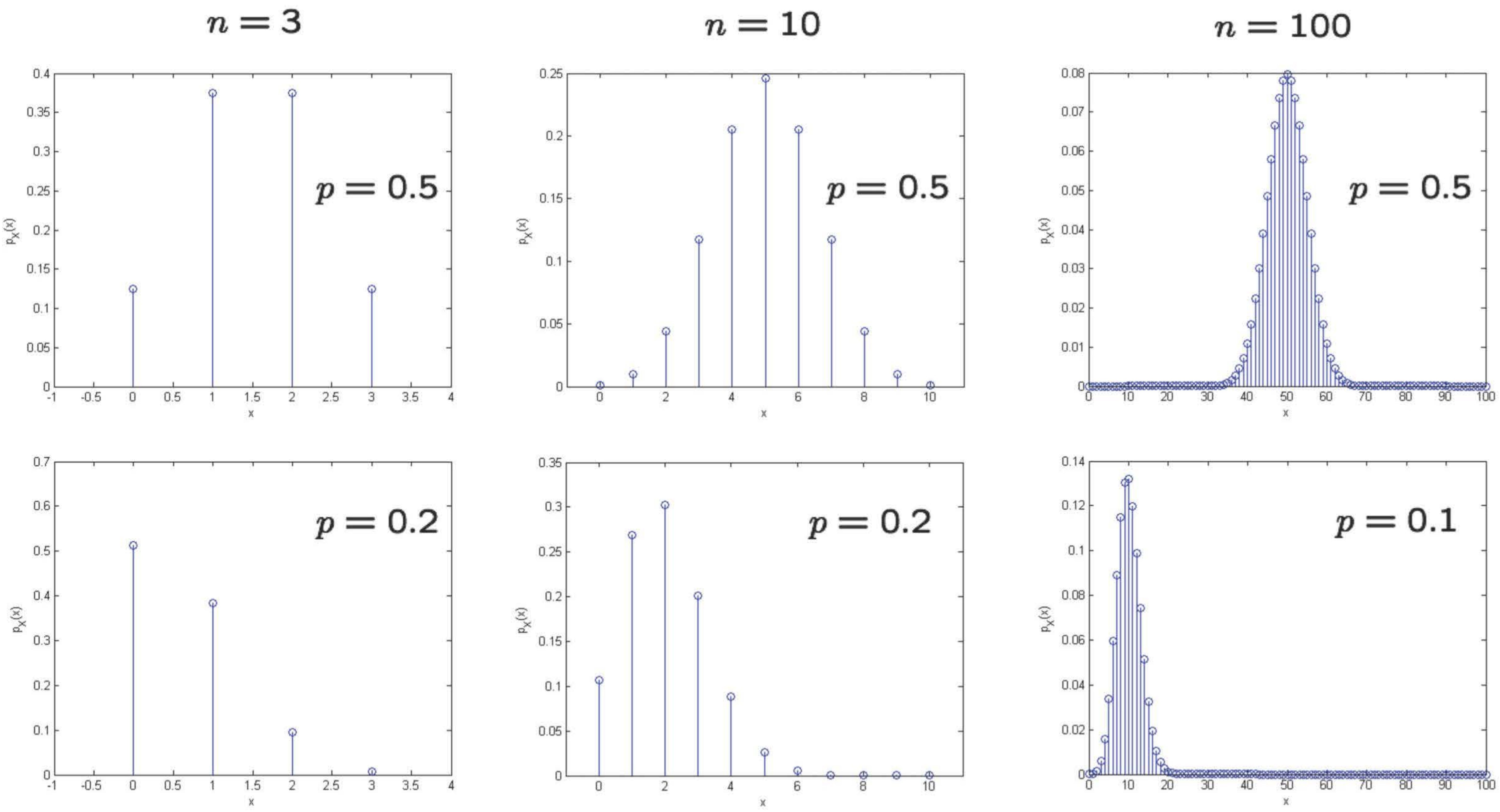


Binomial random variable; **parameters:** positive integer n; $p \in [0,1]$

- **Experiment:** n independent tosses of a coin with $\mathbf{P}(\text{Heads}) = p$
- **Sample space:** Set of sequences of H and T, of length n
- **Random variable X:** number of Heads observed
- **Model of:** number of successes in a given number of independent trials



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



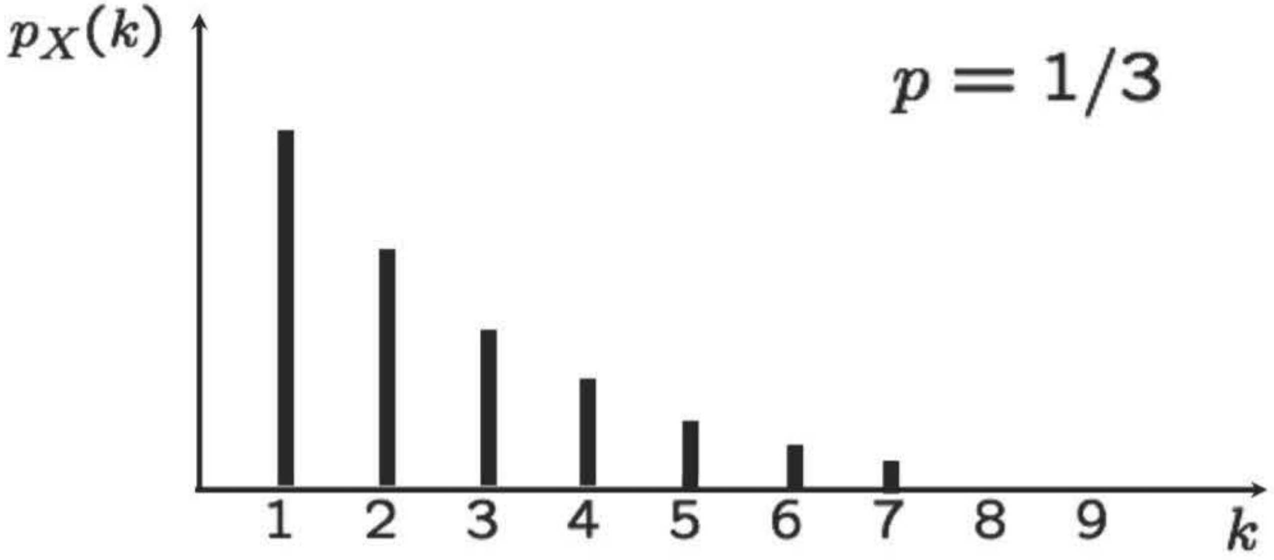
$n = 3$
$n = 10$
$n = 100$
$p = 0.5$
$p = 0.5$
$p = 0.5$
$p = 0.2$
$p = 0.2$
$p = 0.1$
$p_X(x)$
x

Geometric random variable; parameter p: $0 < p \leq 1$

- **Experiment:** infinitely many independent tosses of a coin; $\mathbf{P}(\text{Heads}) = p$
- **Sample space:** Set of infinite sequences of H and T
- **Random variable X:** number of tosses until the first Heads
- **Model of:** waiting times; number of trials until a success

$p_X(k) =$

$\mathbf{P}(\text{no Heads ever})$



Expectation/mean of a random variable

- **Motivation:** Play a game 1000 times. Random gain at each play described by:

$$X = \begin{cases} 1, & \text{w.p. } 2/10 \\ 2, & \text{w.p. } 5/10 \\ 4, & \text{w.p. } 3/10 \end{cases}$$

- "Average" gain:

- **Definition:** $\mathrm{E}[X] = \sum_x x p_X(x)$

- **Interpretation:** Average in large number of independent repetitions of the experiment

- Caution: If we have an infinite sum, it needs to be well-defined. We assume $\sum_x |x|\, p_X(x) < \infty$

Expectation of a Bernoulli r.v.

$$X = \begin{cases} 1, & \text{w.p. } p \\ 0, & \text{w.p. } 1-p \end{cases}$$

If X is the indicator of an event A, $X = I_A$:

Expectation of a uniform r.v.

- Uniform on $0, 1, \ldots, n$

$\mathbf{E}[X] =$

- **Definition:** $\mathbf{E}[X] = \sum_x x p_X(x)$

Expectation as a population average

- n students
- Weight of ith student: x_i
- Experiment: pick a student at random, all equally likely
- Random variable X: weight of selected student
  - assume the x_i are distinct

$p_X(x_i) =$

$\mathbf{E}[X] =$

Elementary properties of expectations

- **Definition:** $\mathbf{E}[X] = \sum_x x p_X(x)$

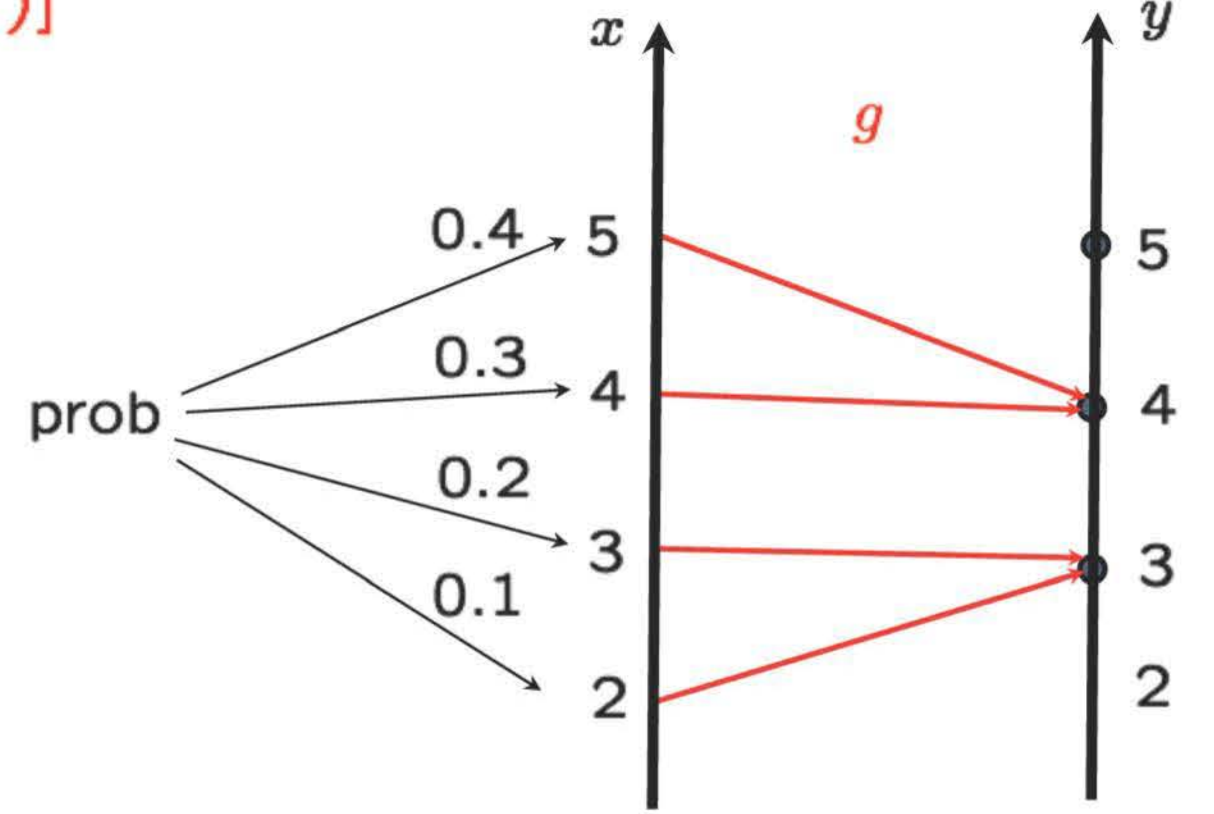
- If $X \geq 0$, then $\mathbf{E}[X] \geq 0$

- If $a \leq X \leq b$, then $a \leq \mathbf{E}[X] \leq b$

- If c is a constant, $\mathbf{E}[c] = c$

The expected value rule, for calculating $\mathbf{E}[g(X)]$

- Let X be a r.v. and let $Y = g(X)$



- Averaging over y: $\mathbf{E}[Y] = \sum_y y p_Y(y)$

- Averaging over x:

$$\mathbf{E}[Y] = \mathbf{E}[g(X)] = \sum_x g(x) p_X(x)$$

Proof:

- $\mathbf{E}[X^2] =$

- Caution: In general, $\mathbf{E}[g(X)] \neq g(\mathbf{E}[X])$

Linearity of expectation: $\mathbf{E}[aX + b] = a\mathbf{E}[X] + b$

- Intuitive

- **Derivation**, based on the expected value rule:

Resource: Introduction to Probability
John Tsitsiklis and Patrick Jaillet