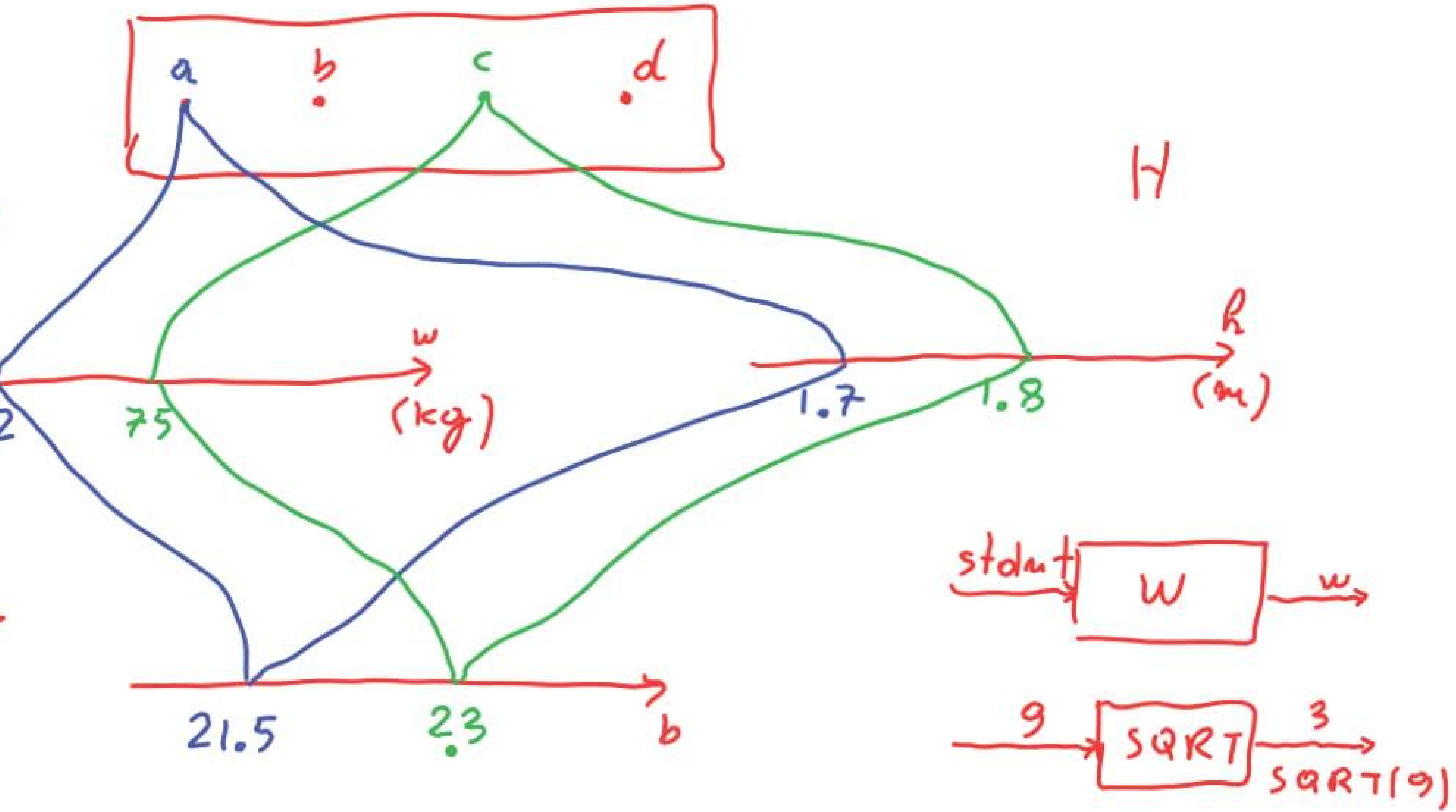
LECTURE 5: Discrete random variables: probability mass functions and expectations

- 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



$$B = \frac{W}{H^2}$$

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 $X+Y$: r.v takes value $x+y$, when X takes value x, Y takes value y

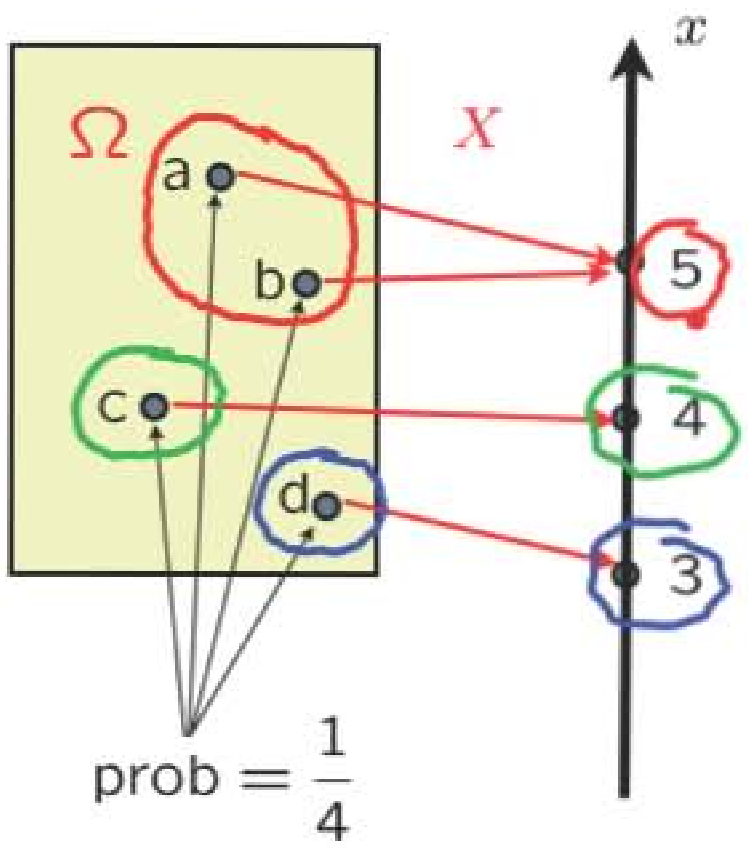
Probability mass function (PMF) of a discrete r.v. X

- It is the "probability law" or "probability distribution" of X

- If we fix some x, then "$X = x$" is an event

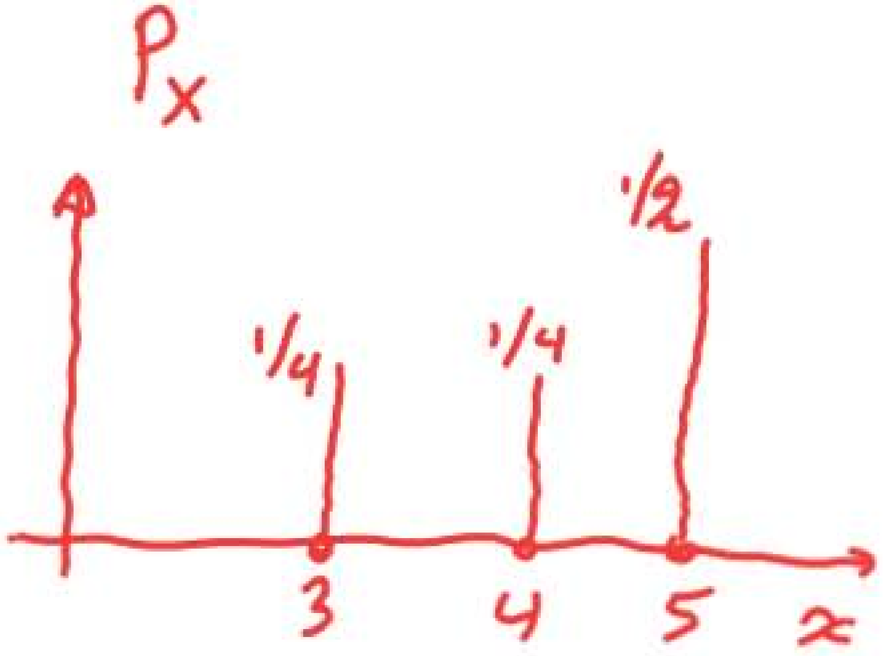
$x=5 \quad X=5 \quad \{\omega : X(\omega)=5\} = \{a, b\}$

$p_X(5) = 1/2$



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

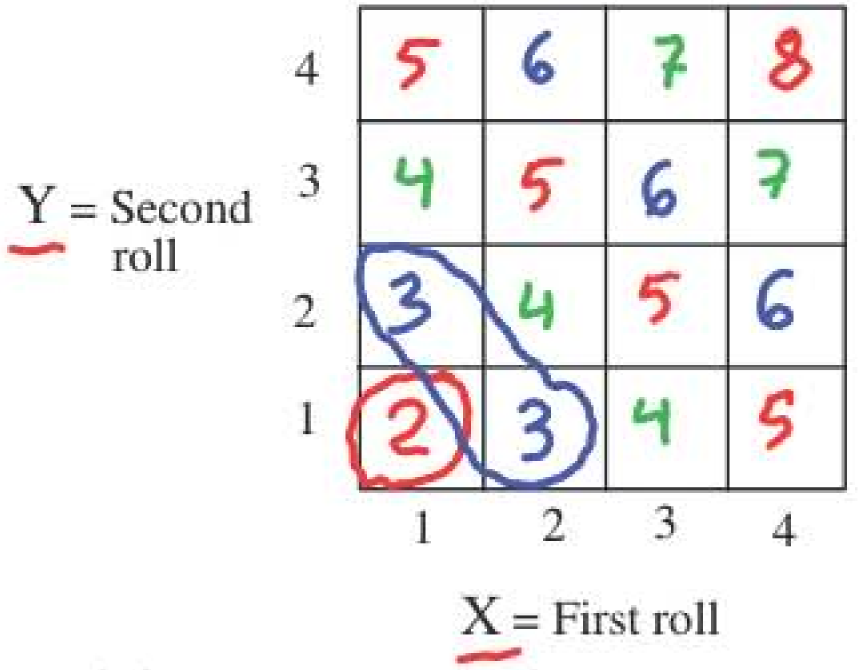
$p_Y(y)$



- **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 $p_Z(z)$ for all z

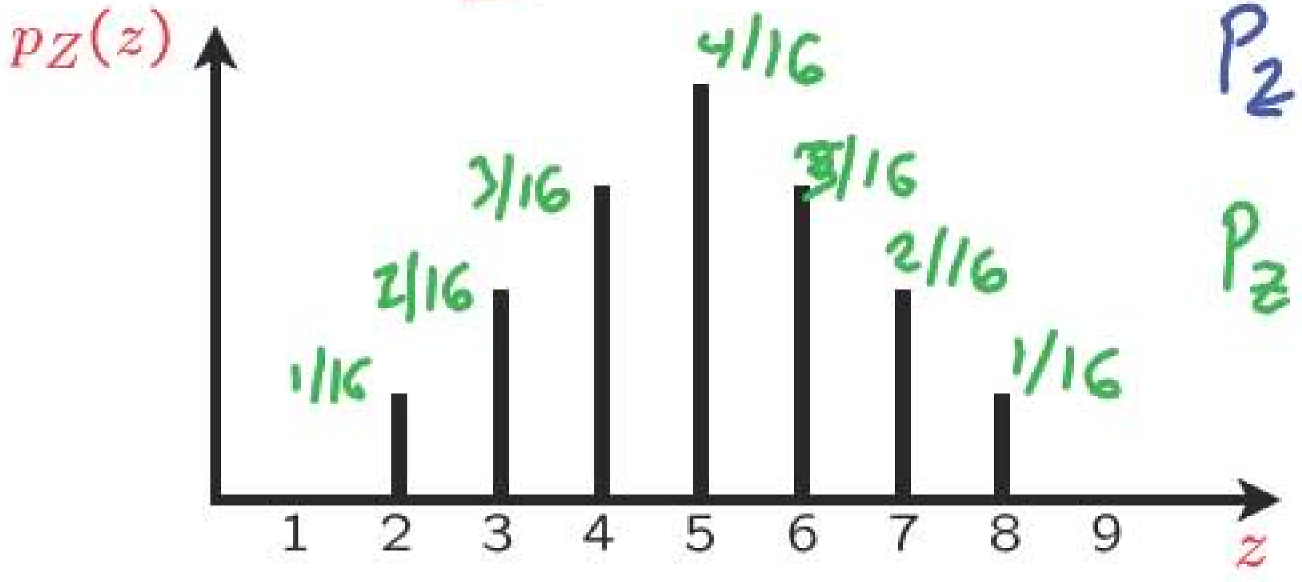
- repeat for all z:
  - collect all possible outcomes for which Z is equal to z
  - add their probabilities

$p_Z(2) = \mathbf{P}(Z = 2) = 1/16$

$p_Z(3) = \mathbf{P}(Z = 3) = 2/16$

$p_Z(4) = \mathbf{P}(Z = 4) = 3/16$

⋮

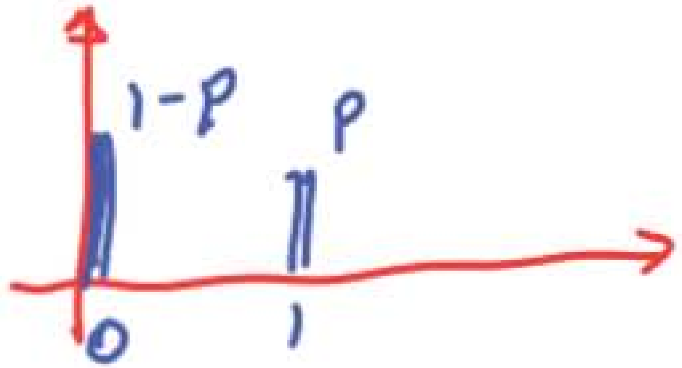


The simplest random variable: Bernoulli with parameter $p \in [0,1]$

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

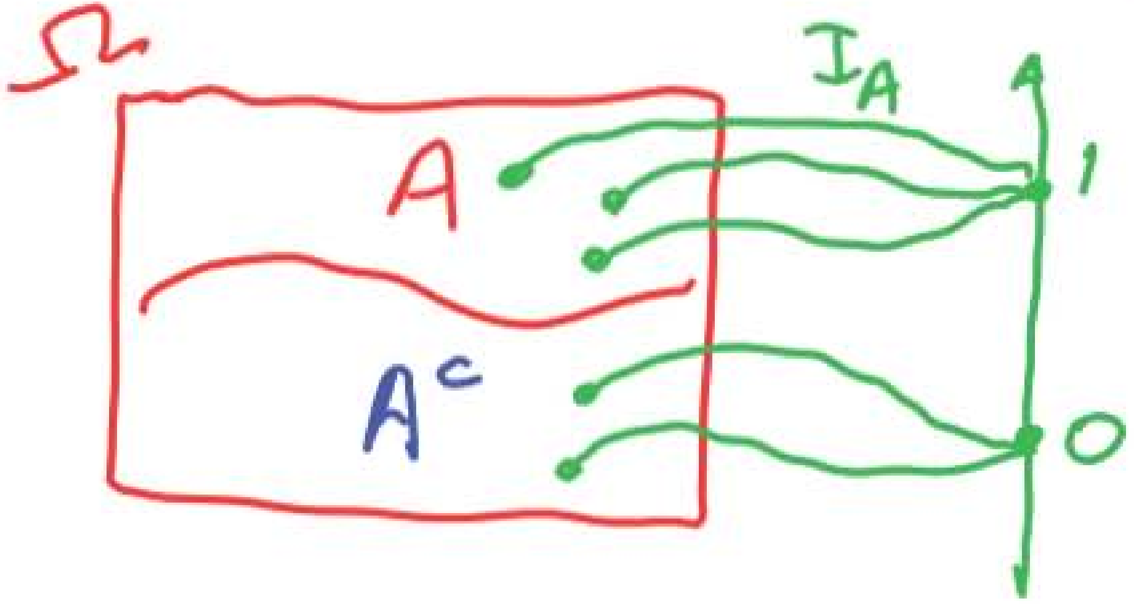
$p_X(0) = 1 - p$

$p_X(1) = p$



- Models a trial that results in success/failure, Heads/Tails, etc.

- Indicator r.v. of an event A: $I_A = 1$ iff A occurs



$p_{I_A}(1) = \mathbf{P}(I_A = 1) = \underline{\underline{\mathbf{P}(A)}}$

p

Discrete uniform random variable; parameters a, b

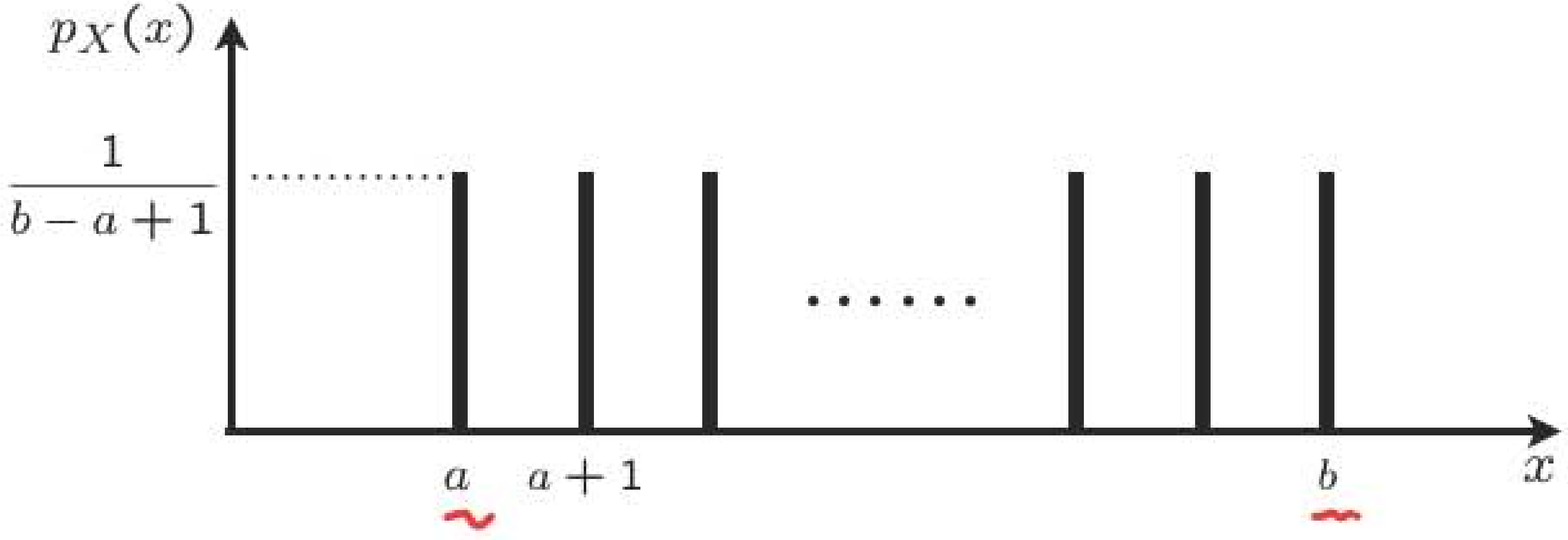
- **Parameters:** integers a, b; $a \leq b$
- **Experiment:** Pick one of $a, a+1, \ldots, b$ at random; all equally likely
- **Sample space:** $\{a, a+1, \ldots, b\}$
- **Random variable X:** $X(\omega) = \omega$
- **Model of:** complete ignorance

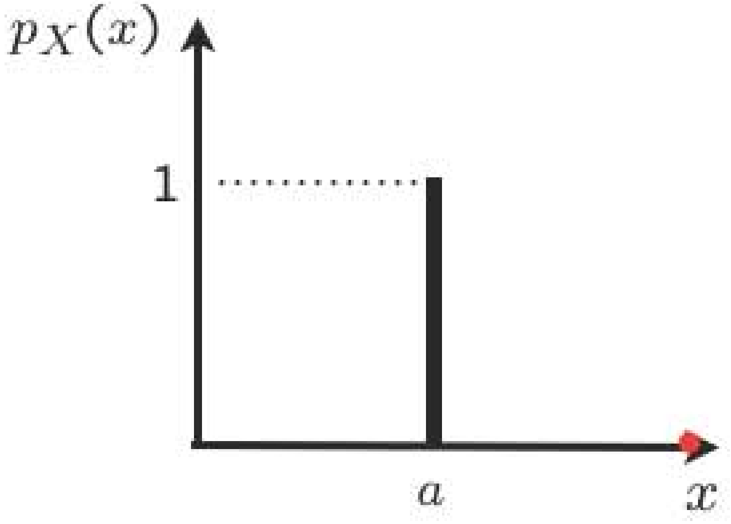
$b - a + 1$ possible values

$11:52:\underline{26}$ $\{0, 1, \ldots, 59\}$

Special case: $a = b$

constant/deterministic r.v.

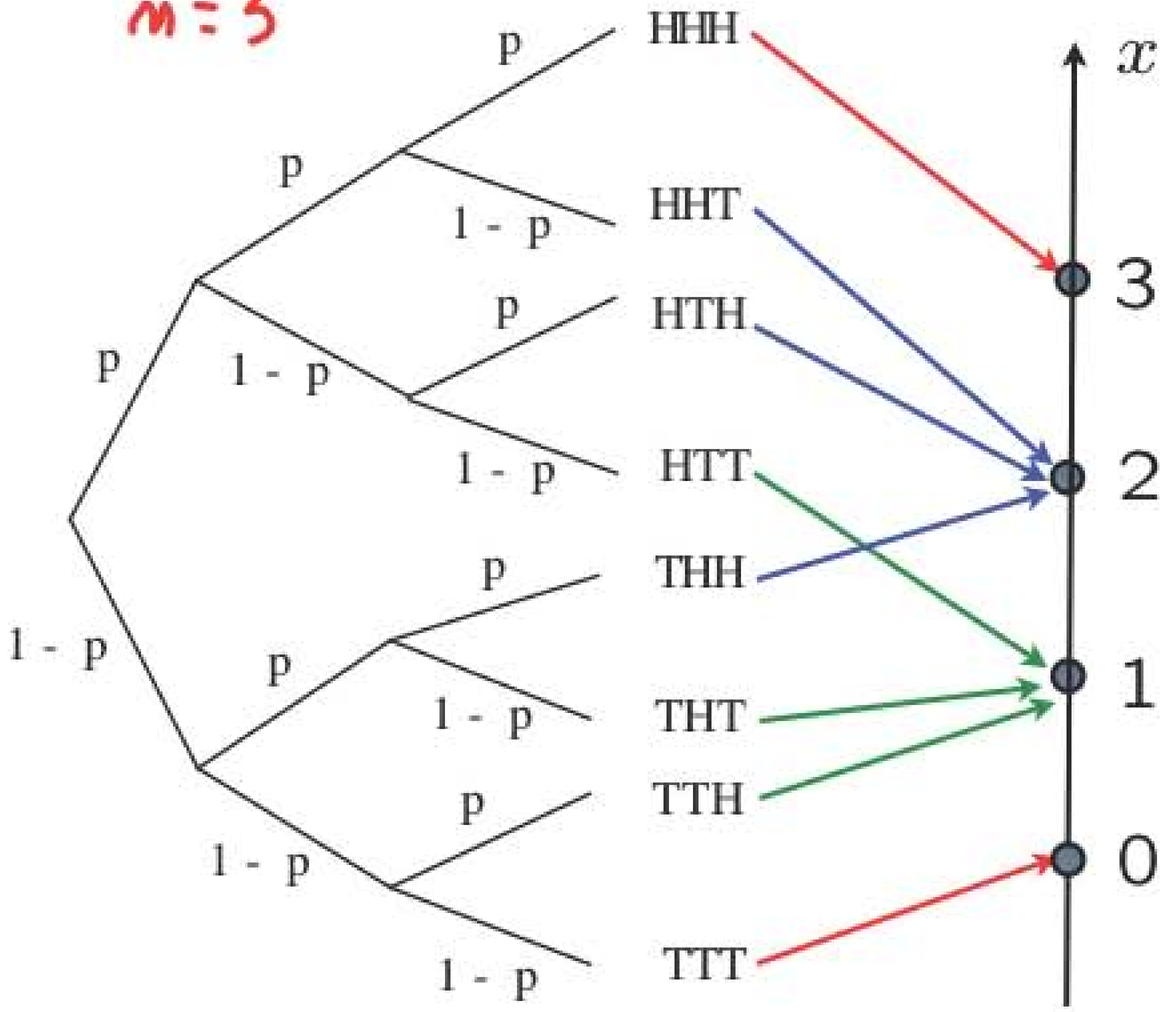




Binomial random variable; **parameters:** positive integer n; $p \in [0,1]$

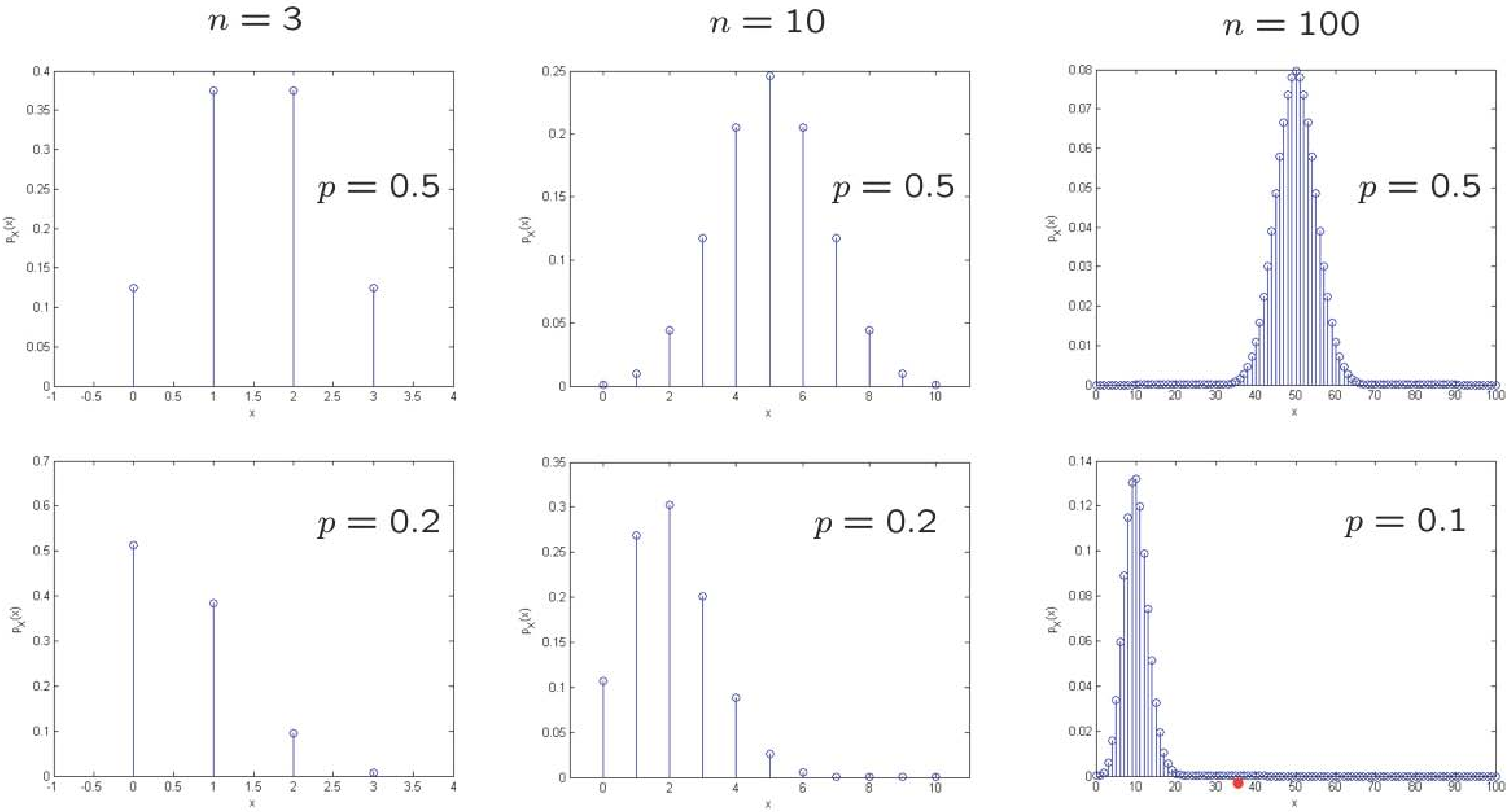
- **Experiment:** n independent tosses of a coin with $\mathbf{P}(\text{Heads}) = p$
- **Sample space:** Set of sequences of H and T, of length n
- **Random variable X:** number of Heads observed
- **Model of:** number of successes in a given number of independent trials

$n=3$



$$p_X(2) = \mathbf{P}(X=2)$$
$$= \mathbf{P}(HHT) + \mathbf{P}(HTH) + \mathbf{P}(THH)$$
$$= 3p^2(1-p) = \binom{3}{2} p^2(1-p)$$

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


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

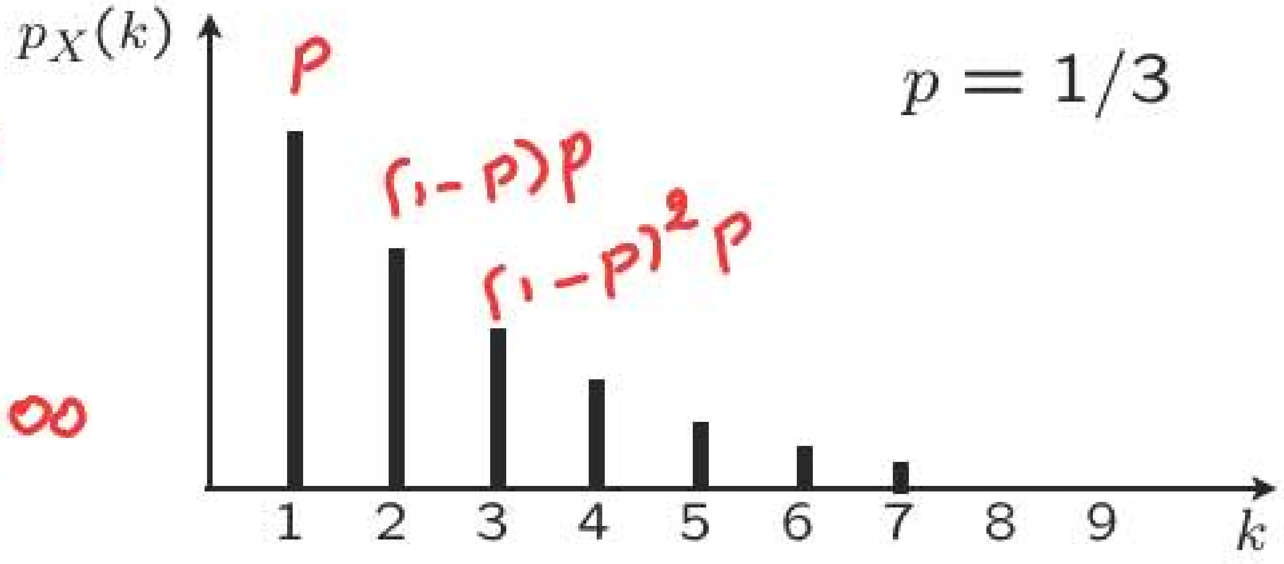
Geometric random variable; parameter p: $0 < p \leq 1$

- **Experiment:** infinitely many independent tosses of a coin; $\mathbf{P}(\text{Heads}) = p$
- **Sample space:** Set of infinite sequences of H and T $\quad \underbrace{TTTTH}HT\ldots$
- **Random variable X:** number of tosses until the first Heads $\quad X = 5$
- **Model of:** waiting times; number of trials until a success

$$p_X(k) = \mathbf{P}(X=k) = \mathbf{P}(\underbrace{T \ldots T}_{k-1} H) = (1-p)^{k-1}p \qquad k = 1, 2, 3, \ldots$$

$\mathbf{P}(\text{no Heads ever})$ $\leq \mathbf{P}(\underbrace{T \ldots T}_{k}) = \underbrace{(1-p)^k}$ $\downarrow k \to \infty$ 0

$(TTT\ldots)$ "$X = \infty$" $= 0$



Expectation/mean of a random variable

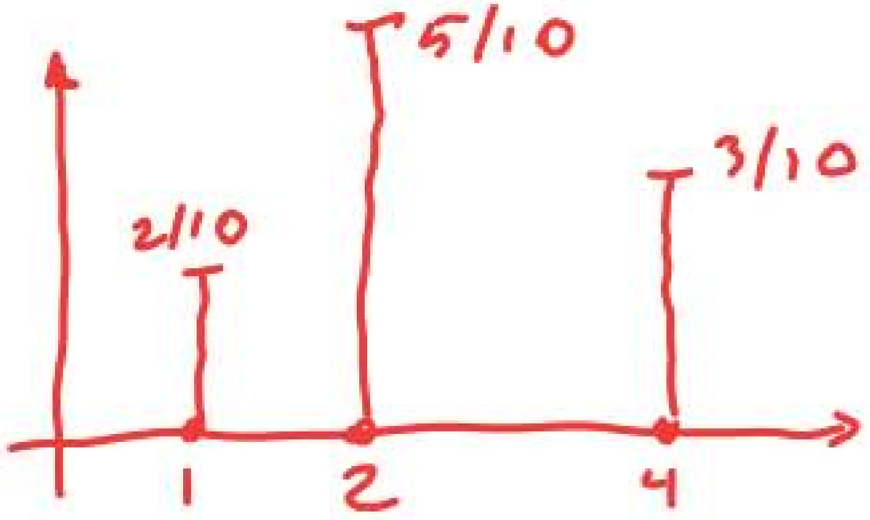
- **Motivation:** Play a game 1000 times.
  Random gain at each play described by:

$$X = \begin{cases} 1, & \text{w.p. } 2/10 \quad \sim 200 \\ 2, & \text{w.p. } 5/10 \quad \sim 500 \\ 4, & \text{w.p. } 3/10 \quad \sim 300 \end{cases}$$

- "Average" gain:

$$\frac{1 \cdot 200 + 2 \cdot 500 + 4 \cdot 300}{1000}$$

$$= 1 \cdot \frac{2}{10} + 2 \cdot \frac{5}{10} + 4 \cdot \frac{3}{10}$$



- **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.

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

$E[X] = 1 \cdot p + 0 \cdot (1-p) = p$

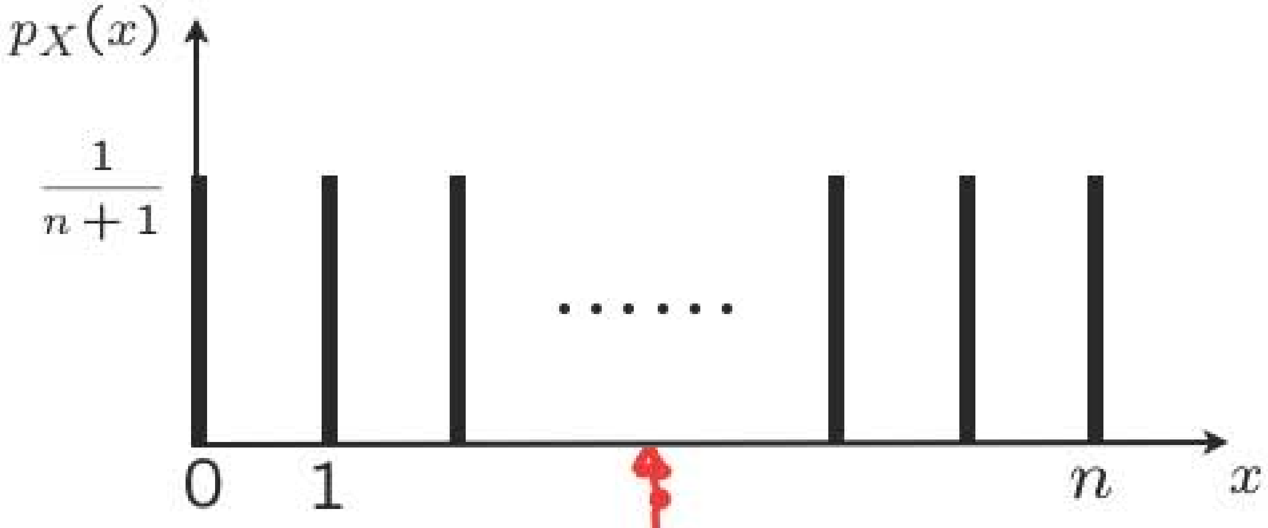
If X is the indicator of an event A, $X = I_A$:

$X = 1$ iff A occurs $\qquad p = P(A)$

$E[I_A] = P(A)$

Expectation of a uniform r.v.

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



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

$$\mathrm{E}[X] = 0 \cdot \frac{1}{n+1} + 1 \cdot \frac{1}{n+1} + \cdots + n \frac{1}{n+1}$$

$$= \frac{1}{n+1}(0 + 1 + \cdots + n) = \frac{1}{n+1} \cdot \frac{n(n+1)}{2} = \frac{n}{2}$$

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

$p_X(x_i) = \frac{1}{n}$

$\mathbf{E}[X] = \sum_i x_i \frac{1}{n} = \frac{1}{n} \sum x_i$

Elementary properties of expectations

- **Definition:** $\mathbf{E}[X] = \sum_x \underbrace{x}_{\geq 0}\, \underbrace{p_X(x)}_{\geq 0}$

≥ 0

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

for all ω: $X(\omega) \geq 0$

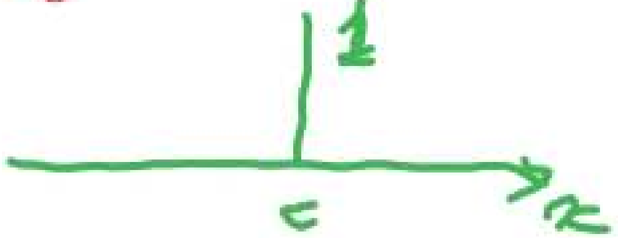
- If $a \leq X \leq b$, then $a \leq \mathbf{E}[X] \leq b$

for all ω: $a \leq X(\omega) \leq b$

$$E[X] = \sum_x x\, p_X(x) \geq \sum_x a\, p_X(x)$$

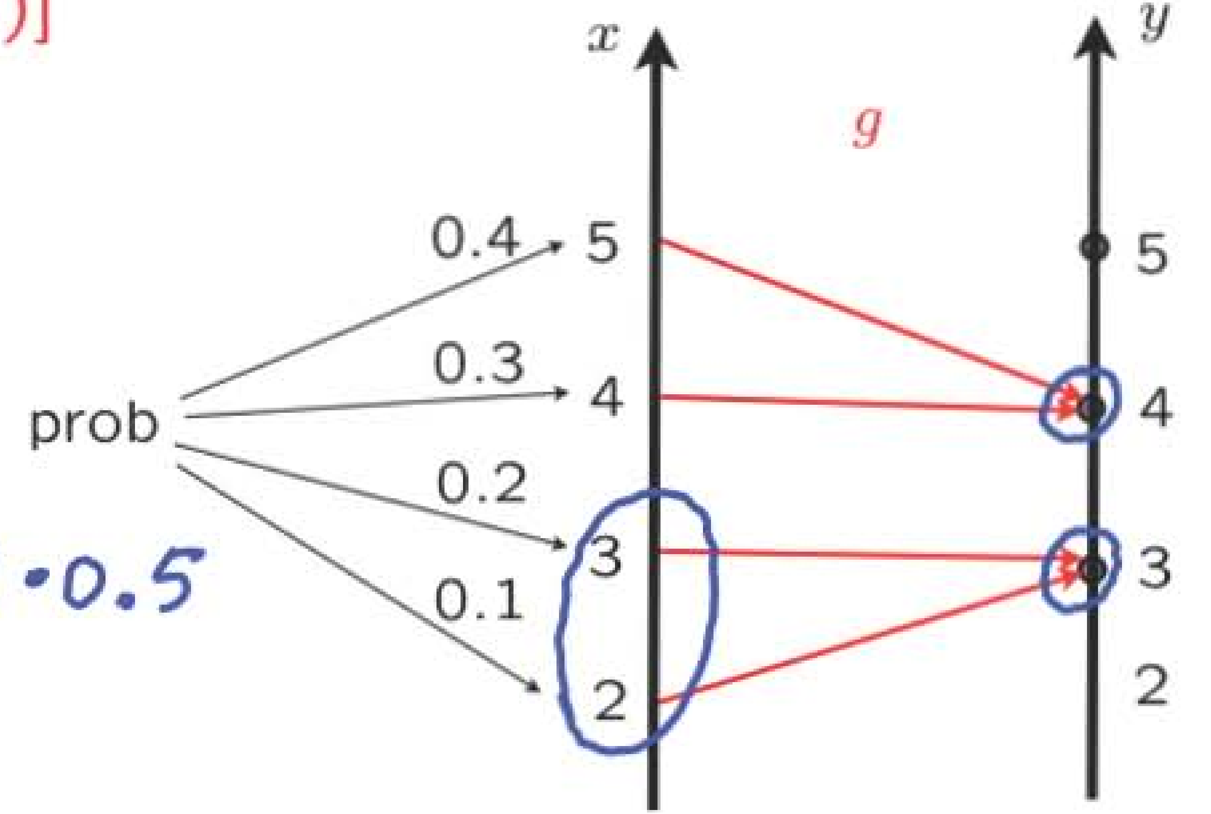
$$= a \sum_x p_X(x) = a \cdot 1 = a$$

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



$$E[c] = c \cdot p(c) = c$$

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



- Let X be a r.v. and let $Y = g(X)$
- Averaging over y: $\mathbf{E}[Y] = \sum_y y p_Y(y)$

$3 \cdot (0.1 + 0.2) + 4 \cdot (0.3 + 0.4)$

- Averaging over x: $3 \cdot 0.1 + 3 \cdot 0.2 + 4 \cdot 0.3 + 4 \cdot 0.5$

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

Proof: $\sum_y \sum_{x: g(x)=y} g(x) p_x(x)$

$= \sum_y \sum_{x: g(x)=y} y \, p_x(x) = \sum_y y \underbrace{\sum_{x: g(x)=y} p_x(x)}$

$= \sum_y y \, p_Y(y) = E[Y]$

- $\mathbf{E}[X^2] = \sum_x x^2 p_x(x)$

$g(x) = x^2$

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

$E[X^2] \neq (E[X])^2$

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

X = salary $\quad E[X]$ = average salary

Y = new salary $= 2X+100 \quad E[Y] = E[2X+100] = 2E[X]+100$

- Intuitive

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

$g(x) = ax+b$

$Y = g(X)$

$$E[Y] = \sum_x g(x)\,p_X(x)$$

$$= \sum_x (ax+b)\,p_X(x) = a\sum_x x\,p_X(x) + b\underbrace{\sum_x p_X(x)}_{1}$$

$$= aE[X]+b$$

$$E[g(X)] = g(E[X])?$$

exceptional g