LECTURE 5

- Readings: Sections 2.1-2.3, start 2.4

Lecture outline

- Random variables
- Probability mass function (pmf)
  - Binomial Random Variable
- Expectation
  - Example

Random Variables - 1

- An assignment of a value (number) to every possible outcome.
- Mathematically: A function from the sample space to the real numbers:
  - Discrete or Continuous
- Can have several random variables defined on the same sample space

Random Variables - 2

- Notation:
  - Random Variable X
  - Experimental Value x
- Example: 1 coin toss. Define X:

$$X(H) = 1,\ X(T) = 0$$

- Example: $Y = g(X)$

Random Variables - 3

- Temperature in Boston on Feb 22.
- Length of queue at Laverde
- Amount of water in a “tall Americano”
- The number of points the Celtics score in a game they win
- The number of words in your emails

Probability mass function (pmf)

- ("probability law",
  "probability distribution")
- Notation: $p_X(x) = \mathbf{P}(X = x)$
- **Example**: X = number of coin tosses until first head
  - Assume independent tosses, $\mathbf{P}(H) = p > 0$

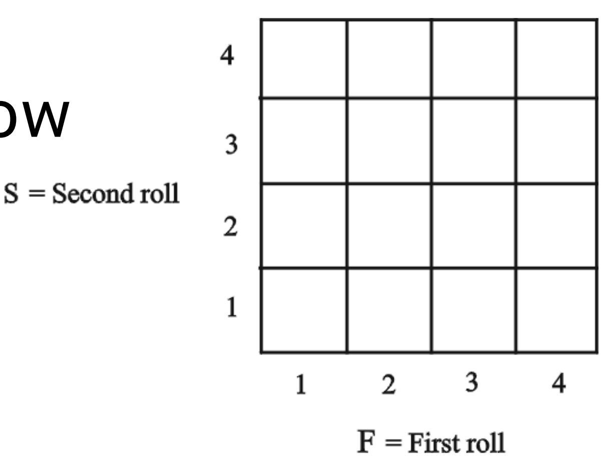
$$
\begin{aligned}
p_X(k) &= \mathbf{P}(X = k) \\
&= \mathbf{P}(TT \cdots TH) \\
&= (1-p)^{k-1} p, \qquad k = 1, 2, \cdots
\end{aligned}
$$

How to compute a pmf $p_X(x)$

- Collect all possible outcomes fro which X is equal to x: $\{w \in \text{Sample Space} | X(w) = x\}$
- Add their probabilities.
- Repeat for all x.

- **Example**: Two independent throws of a fair tetrahedral die:
  - F: outcome of first throw
  - S: outcome of second throw

  $L = \min(F, S)$



$$p_L(2) = \frac{5}{16}$$

Binomial pmf

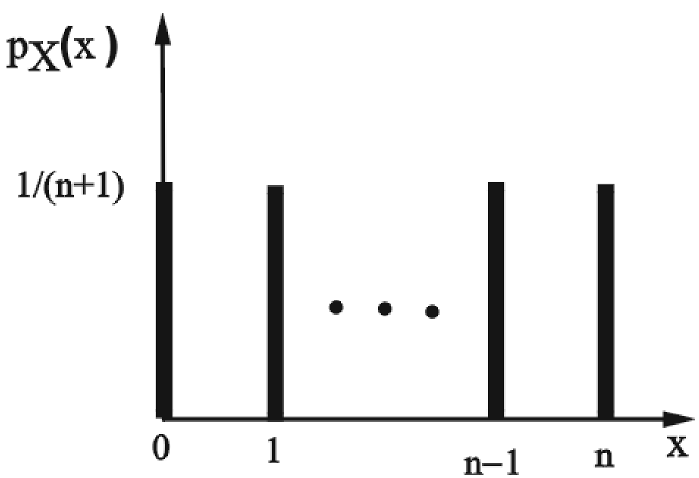
- X: number of heads in n independent coin tosses
- $\mathbf{P}(H) = p$
- Let $n = 4$
- $$\begin{aligned} p_X(2) &= \mathbf{P}(HHTT) + \mathbf{P}(HTHT) + \mathbf{P}(HTTH) \\ &\quad + \mathbf{P}(THHT) + \mathbf{P}(THTH) + \mathbf{P}(TTHH) \\ &= 6p^2(1-p)^2 \\ &= \binom{4}{2} p^2 (1-p)^2 \end{aligned}$$

In general:

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \cdots, n$$

Expectation

- Definition: $$\mathbf{E}[X] = \sum_x x \cdot p_X(x)$$
- Interpretations:
  - Center of gravity of pmf.
  - Average in large number of repetitions of the experiment. (to be substantiated later in this course)
- Example: Uniform on 0, 1, …, n



$$\mathbf{E}[X] = 0 \times \frac{1}{n+1} + 1 \times \frac{1}{n+1} + \cdots + n \times \frac{1}{n+1}$$

$$=$$

Properties of Expectations - 1

- Let X be a r.v. and let $Y = g(X)$
  - Hard: $\mathbf{E}[Y] = \sum_y y \cdot p_Y(y)$
  - Easy: $\mathbf{E}[Y] = \sum_x g(x) \cdot p_X(x)$

- "Second Moment": $\mathbf{E}[X^2]$
- Caution: In general, $\mathbf{E}[g(X)] \neq g(E[X])$
- Variance: $\mathrm{var}(X) = \mathbf{E}[(X - \mathbf{E}[X])^2]$

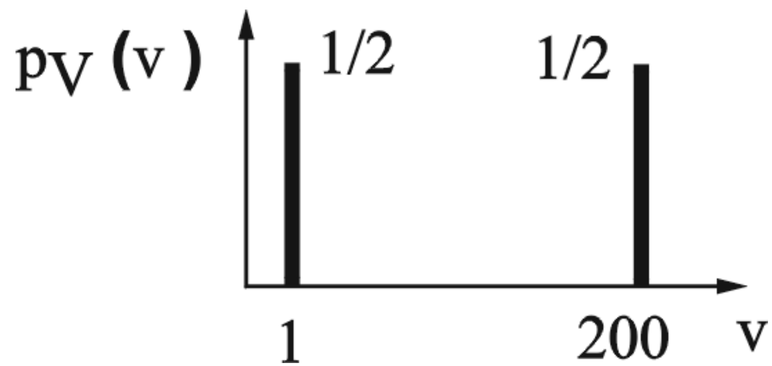
$$= \sum_x (x - \mathbf{E}[X])^2 \cdot p_X(x)$$

Properties of Expectations - 2

- If α, β are constants, then:
  - $\mathbf{E}[\alpha] =$
  - $\mathbf{E}[\alpha X] =$
  - $\mathbf{E}[\alpha X + \beta] =$

Average Speed vs. Average Time - 1

- Traverse a 200 mile distance at constant but random speed V:



- $d = 200, T = t(V) = 200/V$
  - $\mathbf{E}[V] = 1 \cdot (1/2) + 200 \cdot (1/2) = 100.5$
  - $$\begin{aligned} \mathbf{E}[T] &= \mathbf{E}[t(V)] = \sum_v t(v) \cdot p_V(v) \\ &= \frac{200}{1} \cdot \frac{1}{2} + \frac{200}{200} \cdot \frac{1}{2} = 100.5 \end{aligned}$$

Average Speed vs. Average Time - 2

- $\mathbf{E}[T] \cdot \mathbf{E}[V] \neq 200 = d = \mathbf{E}[TV]$
- $\mathbf{E}[T] \neq 200/\mathbf{E}[V]$.

$$\begin{aligned} \text{var}(V) = \textstyle\sum_v (v - \mathbf{E}[V])^2 \cdot p_V(v) \\ &= (1 - 100.5)^2 \frac{1}{2} + (200 - 100.5)^2 \frac{1}{2} \\ &\approx 10,000 \end{aligned}$$

- Standard Deviation $\sigma_V = \sqrt{\text{var}(V)} \approx 100$.