CS5314
Randomized Algorithms

Lecture 4: Discrete Random Variables and Expectation

(Definitions, Basics)

Objectives

- What is a random variable?
- What is expectation?
- Some useful theorems:
  (1) Linearity of Expectations,
  (2) Jensen's inequality
- Binomial random variable

Random Variable

Definition: *A* random variable X on a sample space Ω is a function that maps each outcome of Ω into some real numbers. That is, X: $\Omega \rightarrow \mathcal{R}$.

Definition: *A* discrete random variable is a random variable that takes on only finite or countably infinite number of values.

Example

- Suppose that we throw two dice
- The sample space will be:

$$\Omega = \{ (1,1), (1,2), \ldots, (6,5), (6,6) \}$$

- Define X = sum of outcome of two dice

→ X is a random variable on Ω

(In fact, X is also a discrete random variable)

Discrete Random Variable

- For a discrete random variable X and a value a, the notation

  $$\text{“}X = a\text{”}$$

  denotes the set of outcomes ω in the sample space such that $X(\omega) = a$

  ➔ “$X = a$” is an event

- In previous example,

  “$X = 10$” is the event $\{(4,6), (5,5), (6,4)\}$

Independent

Definition: Two random variables X and Y are independent if and only if

$$\Pr((X=x) \cap (Y=y)) = \Pr(X=x)\Pr(Y=y)$$

for any x and y.

In other words,

the events "$X=x$" and "$Y=y$" are independent, for any x and y.

Independent

Definition: Random variables $X_1, X_2, \ldots, X_k$ are **mutually independent** if and only if for any subset $I \subseteq [1,k]$, and any values x_i

$$\Pr\left(\bigcap_{i\in I} X_i = x_i\right) = \prod_{i\in I} \Pr(X_i = x_i)$$

What does it mean ??

Expectation

Definition: The expectation of a discrete random variable X, denoted by $E[X]$, is

$$E[X] = \sum_i i \Pr(X=i)$$

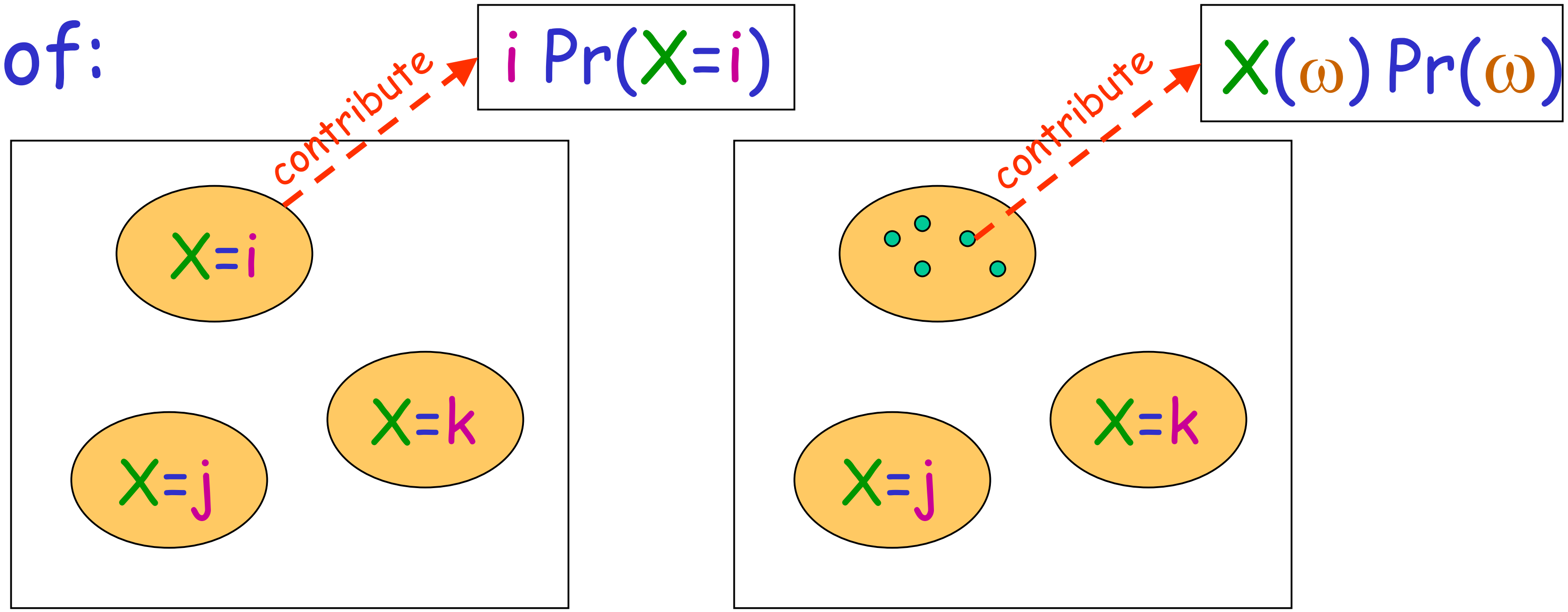
Question:

- X = sum of outcomes of two fair dice
  What is the value of $E[X]$?
- How about the sum of three dice?

Another View of Expectation

Lemma: $E[X] = \sum_{\omega} X(\omega)\Pr(\omega)$

Proof:



$$\sum_i i \Pr(X=i) \qquad\qquad \sum_{\omega} X(\omega)\Pr(\omega)$$

Linearity of Expectation

Theorem: For any finite collection of discrete random variables $X_1, X_2, \ldots, X_k$, each with finite expectation, we have

$$E[\Sigma_i X_i] = \Sigma_i E[X_i]$$

How to prove?

Note: Only need to prove the statement for two random variables

(as general case follows by induction)

Linearity of Expectation

Proof: Let X and Y be two random variables

$$E[X+Y]$$

$$= \sum_{\omega} (X(\omega)+Y(\omega)) \Pr(\omega)$$

$$= \sum_{\omega} X(\omega) \Pr(\omega) + \sum_{\omega} Y(\omega) \Pr(\omega)$$

$$= E[X] + E[Y]$$

Linearity of Expectation (Example)

Let X = sum of outcomes of two fair dice.

Can we compute $E[X]$ with previous theorem?

Let X_i = the outcome of the i^{th} dice

➔ $X = X_1 + X_2$

➔ $E[X] = E[X_1 + X_2] = E[X_1] + E[X_2]$

$= 7/2 + 7/2 = 7$

Can you compute the expectation of the sum of outcomes of three fair dice?

Linearity of Expectation (Remark)

Linearity of expectation does not need to work on independent variables !

E.g., Let X = the outcome of a die throw,

Let $Y = X + X^2$

Clearly, X and X^2 are dependent

However, we can still show that

$$E[Y] = E[X + X^2] = E[X] + E[X^2]$$

Linearity of Expectation [cont.]

Lemma: For any constant c and discrete random variable X,

$$E[cX] = cE[X]$$

The lemma is true when $c = 0$. How about the other cases?

Proof of the Lemma

When $c \neq 0$,

$$E[cX] = \Sigma_{\omega} \left(cX(\omega)\right) \Pr(\omega)$$

$$= c\, \Sigma_{\omega} X(\omega) \Pr(\omega)$$

$$= c\, E[X]$$

Let's Guess

- Which one is larger?

$$(E[X])^2 \text{ or } E[X^2]$$

Let us consider $Y = (X - E[X])^2$

- We notice that $E[Y] \geq 0$. (why??)
- How about $E[Y]$ in terms of $(E[X])^2$ and $E[X^2]$?

Let's Guess

$$E[Y] = E[(X - E[X])^2]$$
$$= E[X^2 - 2X\,E[X] + (E[X])^2]$$
$$= E[X^2] - 2E[X\,E[X]] + (E[X])^2 \quad \text{[why??]}$$
$$= E[X^2] - 2\,E[X]\,E[X] + (E[X])^2 \quad \text{[why??]}$$
$$= E[X^2] - (E[X])^2$$

So, which one is larger? $(E[X])^2$ or $E[X^2]$?

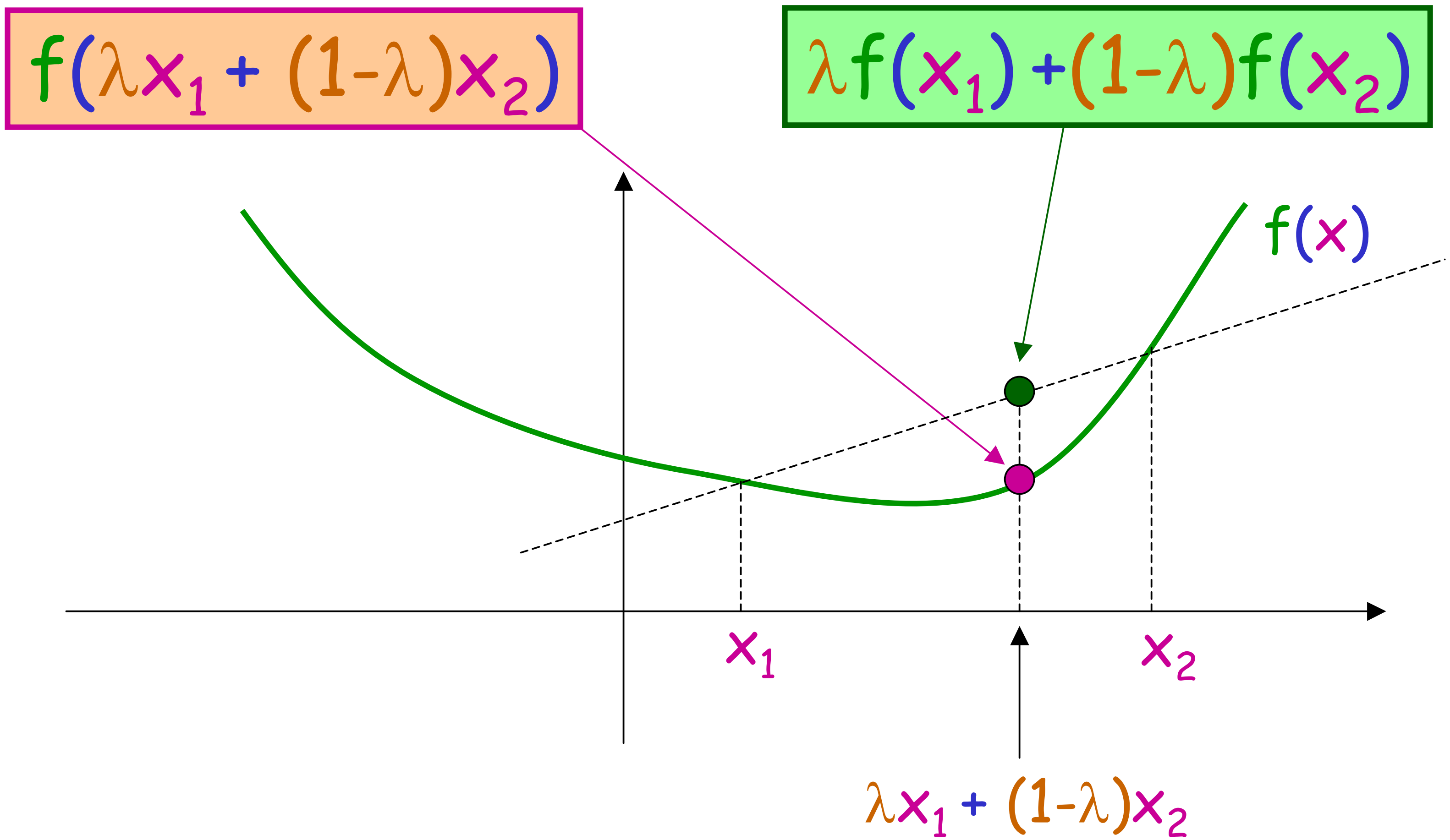
Convex Function

- The previous result is a special case of the Jensen's inequality (which will be described in a moment)

Definition: A function f is convex if for any x_1, x_2 and any $\lambda \in [0,1]$,

$$f(\lambda x_1 + (1-\lambda) x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

Convex Function



Convex Function

Examples of convex function:

$f(x) = x^2$, $f(x) = x^4$, $f(x) = e^x$

Fact: Suppose f is twice differentiable.
Then, f is convex $\Leftrightarrow$ $f''(x) \geq 0$ for all x

Jensen's Inequality

Theorem: If f is convex, then

$$E[f(X)] \geq f(E[X])$$

Proof: Assume f has a Taylor expansion. Then, it is shown that for any μ, there exists c such that $f(x)$ can be written as:

$$f(x) = f(\mu) + f'(\mu)(x-\mu) + f''(c)(x-\mu)^2/2$$

Jensen's Inequality (proof)

Proof [cont.]

Thus, for convex function f, we have

$$f(x) \geq f(\mu) + f'(\mu)(x-\mu) \quad \text{for any } x$$

Now, let $\mu = E[X]$ (so that μ is a constant!)

$$\begin{aligned} E[f(X)] &\geq E[f(\mu) + f'(\mu)(X-\mu)] \\ &= E[f(\mu)] + f'(\mu)(E[X]-\mu) \\ &= f(\mu) + f'(\mu)(0) \\ &= f(\mu) = f(E[X]) \end{aligned}$$

Indicator Random Variable

Suppose that we run an experiment that has only two outcomes: succeeds, or fails

Let Y be a random variable such that

- $Y = 1$ if the experiment succeeds, and
- $Y = 0$ if the experiment fails

→ Y is called an indicator variable (whose values take on either 1 or 0)

If Pr(succeeds) = p, $E[Y] = p = \Pr(Y = 1)$

Binomial Random Variable

Definition: *A* binomial random variable X with parameters n and p, denoted by Bin(n, p), is defined by the following probability distribution on r = 0,1,2,…,n:

$$\Pr(X = r) = C_r^n \, p^r \, (1-p)^{n-r}$$

The event "X = r" represents exactly r successes in n independent experiments, each succeeds with probability p

Expectation of Binomial RV

Question:

What is $E[X]$ of the binomial random variable $X = \text{Bin}(n, p)$?

Ans. np (why??)

Expectation of Binomial RV

First Proof: (using linearity of expectation)

Let $X_1, X_2, \ldots, X_n$ be random variables with

$X_i = 1$ if the i^{th} trial succeeds, and

$X_i = 0$ if the i^{th} trial fails

So, $X = X_1 + X_2 + \ldots + X_n$

Then,

$E[X] = E[X_1 + X_2 + \ldots + X_n]$

$= E[X_1] + E[X_2] + \ldots + E[X_n] = np$

Second Proof: (using definition)

$E[X] = \sum_{r=0 \text{ to } n} r \Pr(X=r)$

$= \sum_{r=1 \text{ to } n} r \Pr(X=r)$

$= \sum_{r=1 \text{ to } n} r\, C(n,r)\, p^r\, (1-p)^{n-r}$

$= np \sum_{r=1 \text{ to } n} (n-1)!/((r-1)!\,(n-r)!)\, p^{r-1}(1-p)^{n-r}$

$= np \sum_{k=0 \text{ to } n-1} (n-1)!/(k!\,(n-1-k)!)\, p^k(1-p)^{n-1-k}$

$= np\, (\, p + (1-p)\,)^{n-1}$

$= np$