Randomized Algorithms
Lecture 4

Kousha Etessami

Discrete Random Variables

A main focus for us in RA is on *random variables*, X, especially *discrete random variables*, where X can take on a finite or countable number of values.

The *expectation*, $\mathrm{E}[X]$, of a discrete random variable X can be defined as $\mathrm{E}[X] \doteq \sum_i i \Pr[X = i]$ where the summation is over all values in the range of X.

Not all random variables have a well-defined finite expectation. Expectation is ***defined*** and ***finite*** if $\sum_i |i| \Pr[X = i]$ *converges* as a series; otherwise it is called ***unbounded*** (or simply *undefined*).

(note that $\mathrm{E}[X]$ cannot be unbounded unless it has infinite support).

Discrete Random Variables

A main focus for us in RA is on *random variables*, X, especially *discrete random variables*, where X can take on a finite or countable number of values.

The *expectation*, $\mathrm{E}[X]$, of a discrete random variable X can be defined as $\mathrm{E}[X] \doteq \sum_i i \Pr[X = i]$ where the summation is over all values in the range of X.
Not all random variables have a well-defined finite expectation. Expectation is ***defined*** and ***finite*** if $\sum_i |i| \Pr[X = i]$ *converges* as a series; otherwise it is called ***unbounded*** (or simply *undefined*).
(note that $\mathrm{E}[X]$ cannot be unbounded unless it has infinite support).

Theorem (2.1, Linearity of Expectation)

For any finite collection of discrete random variables $X_1, \ldots, X_k$ with finite expectations,

$$E\left[\sum_{j=1}^{k} X_j\right] = \sum_{j=1}^{k} E[X_j].$$

Theorem 2.1 holds regardless of whether the random variables are independent or not.

Discrete Random Variables . . .

Lemma (2.2)

For any discrete random variable X, any constant c, $E[c \cdot X] = c \cdot E[X]$.

Definition (2.2)

A collection $X_1, \ldots, X_k$ of random variables are said to be ***mutually independent*** if for every subset $I \subseteq \{1, \ldots, k\}$, and every tuple of values a_i, $i \in I$, we have

$$\Pr\left[\cap_{i \in I}(X_i = a_i)\right] = \prod_{i \in I} \Pr[X_i = a_i].$$

They are called ***pairwise independent*** if
$\Pr[X_i = a_i \cap X_j = a_j] = \Pr[X_i = a_i] \cdot \Pr[X_j = a_j]$, $\forall$ values a_i, a_j and $i, j \in I$.

NOTE: *mutual independence* is **stronger**: a collection of random variables can be pairwise independent but *not* mutually independent:

Example

Two fair coins, values 1 and 0. A "value of first flip", B "value of second flip", C "absolute difference of two values". Pairwise independence works out but $\Pr[(A = 1) \cap (B = 1) \cap (C = 1)] = ?$

Variance and k-th moments

A "partner measure" to *expectation* (the "first moment") is *variance* (or the related measure called the *second moment*).

Variance and k-th moments

A "partner measure" to *expectation* (the "first moment") is *variance* (or the related measure called the *second moment*).

Definition

For any discrete random variable X, and integer $k \geq 1$, the k-th moment is defined as $\mathrm{E}[X^k]$, ie, $\sum_i i^k \Pr[X = i]$ (i ranging over the support of X).
The variance is defined as $\mathrm{E}[(X - \mathrm{E}[X])^2]$, ie, $\sum_i (i - \mathrm{E}[X])^2 \Pr[X = i]$.

Variance and k-th moments

A "partner measure" to *expectation* (the "first moment") is *variance* (or the related measure called the *second moment*).

Definition

For any discrete random variable X, and integer $k \geq 1$, the k-th moment is defined as $\mathrm{E}[X^k]$, ie, $\sum_i i^k \Pr[X = i]$ (i ranging over the support of X).
The variance is defined as $\mathrm{E}[(X - \mathrm{E}[X])^2]$, ie, $\sum_i (i - \mathrm{E}[X])^2 \Pr[X = i]$.

Lemma

For any discrete random variable X, $E[X^2] \geq E[X]^2$.

Variance and k-th moments

A "partner measure" to *expectation* (the "first moment") is *variance* (or the related measure called the *second moment*).

Definition

For any discrete random variable X, and integer $k \geq 1$, the k-th moment is defined as $\mathrm{E}[X^k]$, ie, $\sum_i i^k \Pr[X = i]$ (i ranging over the support of X).
The variance is defined as $\mathrm{E}[(X - \mathrm{E}[X])^2]$, ie, $\sum_i (i - \mathrm{E}[X])^2 \Pr[X = i]$.

Lemma

For any discrete random variable X, $E[X^2] \geq E[X]^2$.

Proof.

Define $Y = (X - \mathrm{E}[X])^2$, Y is also a discrete random variable. Also Y only takes non-negative values, hence $\mathrm{E}[Y] \geq 0$. Moreover,

Variance and k-th moments

A "partner measure" to *expectation* (the "first moment") is *variance* (or the related measure called the *second moment*).

Definition

For any discrete random variable X, and integer $k \geq 1$, the k-th moment is defined as $\mathrm{E}[X^k]$, ie, $\sum_i i^k \Pr[X = i]$ (i ranging over the support of X). The variance is defined as $\mathrm{E}[(X - \mathrm{E}[X])^2]$, ie, $\sum_i (i - \mathrm{E}[X])^2 \Pr[X = i]$.

Lemma

For any discrete random variable X, $E[X^2] \geq E[X]^2$.

Proof.

Define $Y = (X - \mathrm{E}[X])^2$, Y is also a discrete random variable. Also Y only takes non-negative values, hence $\mathrm{E}[Y] \geq 0$. Moreover,

$$\begin{aligned}
\mathrm{E}[Y] &= \mathrm{E}[X^2 - 2\mathrm{E}[X] \cdot X + \mathrm{E}[X]^2] \\
&= \mathrm{E}[X^2] - \mathrm{E}[2\mathrm{E}[X] \cdot X] + \mathrm{E}[X]^2 && \text{(Thm 2.1)} \\
&= \mathrm{E}[X^2] - 2\mathrm{E}[X] \cdot \mathrm{E}[X] + \mathrm{E}[X]^2 && \text{(Lemma 2.2)} \\
&= \mathrm{E}[X^2] - \mathrm{E}[X]^2.
\end{aligned}$$

By $\mathrm{E}[Y] \geq 0$, we have $\mathrm{E}[X^2] \geq \mathrm{E}[X]^2$. □

An application of Linearity of Expectation

Consider a uniformly at random permutation, σ, of n elements. What is the expected number of fixed points of σ (i.e., expected number of $i \in \{1, \ldots, n\}$ such that $\sigma(i) = i$)?

An application of Linearity of Expectation

Consider a uniformly at random permutation, σ, of n elements. What is the expected number of fixed points of σ (i.e., expected number of $i \in \{1, \ldots, n\}$ such that $\sigma(i) = i$)?

Let X_i be the indicator variable of the event $\sigma(i) = i$. In other words:

$$X_i = \begin{cases} 1 & \text{if } \sigma(i) = i \\ 0 & \text{otherwise} \end{cases}$$

Let $X = \sum_{i=1}^{n} X_i$. Then

$$\mathrm{E}X_i = \Pr[X_i = 1] = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

Thus, by linearity of expectation,

$$\mathrm{E}X = \sum_{i=1}^{n} \mathrm{E}X_i = \sum_{i=1}^{n} \frac{1}{n} = 1.$$

Convex functions, and Jensen's Inequality

Definition

A function $f : \mathbb{R} \rightarrow \mathbb{R}$ is called *convex* if for all $x_1, x_2 \in \mathbb{R}$ and all $\lambda \in [0, 1]$,

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

Equivalently, f is convex if for any $k \geq 2$, for all $x_1, \ldots, x_k \in \mathbb{R}$, and all $\lambda_1, \ldots, \lambda_k \in [0, 1]$ such that $\sum_{i=1}^{k} \lambda_i = 1$

$$f(\sum_{i=1}^{k} \lambda_i x_i) \leq \sum_{i=1}^{k} \lambda_i f(x_i).$$

In other words, f applied to any **convex combination** of values is $\leq$ that same convex combination applied to f applied to those values.

Jensen's Inequality

Theorem (2.4, Jensen's Inequality)

If $f : \mathbb{R} \to \mathbb{R}$ is a convex function, and X is a random variable with finite expectation, then

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

We will not provide a proof.

It is not difficult to prove for discrete random variables.
<u>Inuition</u>: an "expectation" is nothing other than a "weighted average", i.e., it is a "convex combination".
(The book provides a proof, but only for differentiable f.)

Simple distributions

Definition

The *Bernoulli distribution* ("biased coin-flip") is given by a random variable Y such that $Y = 1$ ("success") with probability p and $Y = 0$ ("failure") with probability $q = 1 - p$.

Notice $\mathrm{E}[Y] = p$ when Y is Bernoulli.

Definition (2.5)

The *binomial distribution*, written $B(n, p)$, for integer $n \geq 1$ and probability $p \in [0, 1]$, is given by a random variable X which takes values in $\{0, 1, \ldots, n\}$ with the probabilities $\mathrm{Pr}[X = j] = \binom{n}{j} p^j (1-p)^{n-j}$.
$\mathrm{Pr}[X = j]$ is precisely the probability of j successes in n mutually independent Bernoulli trials, each with probability of success p.

Claim: $\mathrm{E}[X] = np$ for a binomially distributed $B(n, p)$ random variable X.

Proof.

Note that we can write $X = \sum_{i=1}^{n} X_i$, where X_i is the Bernoulli random variable representing the i'th Bernoulli trial. Then by linearity of expectation $\mathrm{E}[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np$. □
(Note that we didn't even use independence of the different Bernoulli trials for this!!)

Conditional Expectation

Definition (2.6)

For two random variables X, Y,

$$\mathrm{E}[X \mid Y = y] \;=\; \sum_{x} x \cdot \Pr[X = x \mid Y = y],$$

the summation being taken over all x in the support of X, and we asssume y is in the support of Y for this to be well-defined.

Definition (2.7)

We use $\mathrm{E}[X \mid Y]$, where X, Y are random variables. to denote a new random variable, a function of Y, which takes on the value $\mathrm{E}[X \mid Y = y]$ whenever event $[Y = y]$ holds.

Conditional Expectation

Observation

For any finite collection of discrete random variables $X_1, \ldots, X_n$ with finite expectations, and for any random variable Y, and value y in the support of Y,

$$E\left[\left(\sum_{i=1}^{n} X_i\right) \mid Y = y\right] = \sum_{i=1}^{n} E[X_i \mid Y = y].$$

Lemma (2.5)

For any random variables X and Y, such that $E[X \mid Y = y]$ is always bounded

$$E[E[X \mid Y]] = E[X].$$

Conditional Expectation

Proof.

$$
\begin{aligned}
\mathrm{E}[\mathrm{E}[X \mid Y]] &= \sum_y \Pr[Y = y]\mathrm{E}[X \mid Y = y] \\
&= \sum_y \Pr[Y = y] \sum_x x \Pr[X = x \mid Y = y] \\
&= \sum_y \sum_x x \Pr[Y = y] \Pr[X = x \mid Y = y] \\
&= \sum_y \sum_x x \Pr[X = x \cap Y = y] \\
&= \sum_x \sum_y x \Pr[X = x \cap Y = y] \\
&= \sum_x x \Pr[X = x] = \mathrm{E}[X]. \qquad \square
\end{aligned}
$$

Geometric distributions

Imagine we flip a biased coin many times (success with prob. p), and stop when we see the first success (heads, or alternatively 1). What is the distribution of the number of flips?

Definition (2.8)

A *geometrically distributed* random variable X with parameter $p \in (0, 1)$ has the following probability distribution on $\mathbb{Z}_+ = \{1, 2, 3, \ldots\}$:

$$\Pr[X = j] \;=\; (1-p)^{j-1}p.$$

Let's confirm that this does indeed define a probability distribution on $\mathbb{Z}_+$: $\sum_{j=1}^{\infty} \Pr[X = j] = \sum_{j=1}^{\infty}(1-p)^{j-1}p = p\sum_{j=0}^{\infty}(1-p)^{j} = p(\frac{1}{p}) = 1.$

Lemma (2.8)

For a geometric random variable X with parameter p, and for any $j > 0$, $k \geq 0$,

$$\Pr[X = j + k \mid X > k] \;= \Pr[X = j].$$

Geometric distributions

Lemma (2.9)

For any discrete random variable X that only takes non-negative integer values:

$$E[X] = \sum_{i=1}^{\infty} \Pr[X \geq i].$$

Proof.

$\sum_{i=1}^{\infty} \Pr[X \geq i] = \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \Pr[X = j] = \sum_{j=1}^{\infty} \sum_{i=1}^{j} \Pr[X = j] = \sum_{j=1}^{\infty} j \Pr[X = j] = \mathrm{E}[X]$. □

Observation

If X is a geometric random variable X with parameter p, then for any $i \geq 1$, $\Pr[X \geq i] = (1-p)^{i-1}$.

Proof.

The event that $X \geq i$ is exactly the event that the first $(i-1)$ trials all fail. (You can also directly calculate it.) □

Geometric distributions

Lemma
For a geometric random variable X with parameter p, $E[X] = \frac{1}{p}$.

Proof.
By Lemma 2.9, we have $\mathrm{E}[X] = \sum_{i=1}^{\infty} \Pr[X \geq i]$. Thus

$$\mathrm{E}[X] = \sum_{i=1}^{\infty} (1-p)^{i-1} = \sum_{i=0}^{\infty} (1-p)^{i}$$
$$= \frac{1}{1-(1-p)} = \frac{1}{p}.$$

□

Coupon Collector Problem

"Coupon collecting": *suppose that when you buy a Kellogg's cornflakes cereal box, each box contains a "coupon" inside, e.g., an action hero card. Suppose there are n different types of "coupons", i.e., n different action hero cards, and each box contains a card chosen independently and uniformly at random from the n possibilities.*

Suppose your goal is to collect at least one of each action hero card, after which you will stop buying.

What is the (expected) number of packets you need to buy to complete your collection of all n cards?

- Note that when buying a box the probability it contains any particular card inside is $1/n$.

Coupon Collector Analysis

- Let X be a random variable denoting the total number of boxes bought to get all n cards.
- For each $i = 1, \ldots, n$, let X_i be a random variable denoting the number of boxes bought while already having exactly $i - 1$ different cards, and just until you get the ith different card.
- Clearly $X = \sum_{i=1}^{n} X_i$.

Coupon Collector Analysis

Key observation: X_i is also a *geometrically distributed* random variable: if we already own $i-1$ different cards, when we buy a box the probability that we get a *new* card is $p_i = 1 - \frac{i-1}{n} = \frac{n-i+1}{n}$.
Hence X_i is a geometric r.v. with parameter p_i.
Thus, $\mathrm{E}[X_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$.

Coupon Collector Analysis

Key observation: X_i is also a *geometrically distributed* random variable: if we already own $i-1$ different cards, when we buy a box the probability that we get a *new* card is $p_i = 1 - \frac{i-1}{n} = \frac{n-i+1}{n}$.
Hence X_i is a geometric r.v. with parameter p_i.
Thus, $\mathrm{E}[X_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$.

By linearity of expectation, $\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{E}[X_i] = \sum_{i=1}^{n} \frac{n}{n-(i-1)} = n \sum_{i=1}^{n} \frac{1}{i}$.

Coupon Collector Analysis

Key observation: X_i is also a *geometrically distributed* random variable: if we already own $i-1$ different cards, when we buy a box the probability that we get a *new* card is $p_i = 1 - \frac{i-1}{n} = \frac{n-i+1}{n}$.
Hence X_i is a geometric r.v. with parameter p_i.
Thus, $\mathrm{E}[X_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$.

By linearity of expectation, $\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{E}[X_i] = \sum_{i=1}^{n} \frac{n}{n-(i-1)} = n \sum_{i=1}^{n} \frac{1}{i}$.

Fact: For all $n \geq 1$, the *"Harmonic number"*, $H(n) = \sum_{i=1}^{n} \frac{1}{i}$, satisfies
$$\ln(n) + \frac{1}{2} \leq H(n) \leq \ln(n) + 1.$$

Coupon Collector Analysis

Key observation: X_i is also a *geometrically distributed* random variable: if we already own $i-1$ different cards, when we buy a box the probability that we get a *new* card is $p_i = 1 - \frac{i-1}{n} = \frac{n-i+1}{n}$.
Hence X_i is a geometric r.v. with parameter p_i.
Thus, $\mathrm{E}[X_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$.

By linearity of expectation, $\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{E}[X_i] = \sum_{i=1}^{n} \frac{n}{n-(i-1)} = n \sum_{i=1}^{n} \frac{1}{i}$.

Fact: For all $n \geq 1$, the *"Harmonic number"*, $H(n) = \sum_{i=1}^{n} \frac{1}{i}$, satisfies
$$\ln(n) + \frac{1}{2} \leq H(n) \leq \ln(n) + 1.$$

So, the expected number $\mathrm{E}[X]$ of boxes needed to collect all n cards, is $n \ln(n) + \Theta(n)$, or more specifically $n \ln(n) + \frac{n}{2} \leq \mathrm{E}[X] \leq n \ln(n) + n$.

Is “expected” the same as “typical”?

All we know (for Coupon collecting) is the “average” (weighted over random choices) number of cards.

We don’t (yet) know how likely one “run” of the process is to come close to that expected value.

“Concentration inequalities” help us bound the *deviation from the mean.* Next time, we start talking about such inequalities:

- Markov’s Inequality;
- Chebyshev’s Inequality;
- Chernoff Bound / Hoeffding inequality.