Randomness and Computation
or, “Randomized Algorithms”

Mary Cryan

School of Informatics
University of Edinburgh
Discrete Random Variables

Most of our reasoning can be cast in terms of *random variables*:

**Definition (2.1)**

A *random variable* over a sample space $\Omega$ is a real-valued function $X : \Omega \rightarrow \mathbb{R}$. It is a *discrete random variable* if it takes on only a finite or countable number of values.

We will write things like $\Pr[X = a]$ to mean (with reference to the event structure of lecture 2)

$$
\sum_{s \in \Omega, X(s) = a} \Pr[s].
$$

**Definition (2.3)**

The *expectation* of a discrete random variable $X : \Omega \rightarrow \mathbb{R}$, denoted $E[X]$, is defined as

$$
E[X] = \sum_a a \Pr[X = a],
$$

where $a$ ranges overall possible values for $X$ (also known as the *support* of $X$).
Discrete Random Variables . . .

Definition (2.2)
A collection $X_1, \ldots, X_k$ of random variables are said to be \textit{mutually independent} if and only if, for every subset $I \subseteq \{1, \ldots, k\}$, and every tuple of values $a_i, i \in I$, we have

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

\textbf{Stronger} definition than “pairwise independent”. Possible for a collection of random variables to be pairwise independent but \textit{not} mutually independent.

\textbf{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 relationships work out (board). $\Pr [(A = 1) \cap (B = 1) \cap (C = 1)] = \, ?.$
Linearity of Expectation

Not all random variables have bounded expectation. Expectation is **finite** if \( \sum_i |i| \Pr[X = i] \) **converges** as a series; otherwise **unbounded**. (note that \( X \) 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,*

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

Note Theorem 2.1 holds regardless of whether the random variables are independent (pairwise, or mutually) or not.

**Lemma (2.2)**

*For any discrete random variable \( X \) and any constant \( c \),*

\[
E[c \cdot X] = c \cdot E[X].
\]
Linearity of Expectation

Not all random variables have bounded expectation. Expectation is *finite* if $\sum_i |i| \Pr[X = i]$ *converges* as a series; otherwise *unbounded*. (note that $X$ 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,

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

Note Theorem 2.1 holds regardless of whether the random variables are independent (pairwise, or mutually) or not.

**Lemma (2.2)**

For any discrete random variable $X$ and any constant $c$,

$$E[c \cdot X] = c \cdot E[X].$$
Linearity of Expectation

Not all random variables have bounded expectation. Expectation is *finite* if \( \sum_i |i| \Pr[X = i] \) *converges* as a series; otherwise *unbounded*. (note that \( X \) 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,

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

Note Theorem 2.1 holds regardless of whether the random variables are independent (pairwise, or mutually) or not.

**Lemma (2.2)**

For any discrete random variable \( X \) and any constant \( c \),

\[
E[c \cdot X] = c \cdot E[X].
\]
Linearity of Expectation

Not all random variables have bounded expectation. Expectation is *finite* if \( \sum_i |i| \Pr[X = i] \) converges as a series; otherwise *unbounded*. (note that \( X \) 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,*

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

Note Theorem 2.1 holds regardless of whether the random variables are independent (pairwise, or mutually) or not.

**Lemma (2.2)**

*For any discrete random variable \( X \) and any constant \( c \),

\[
E[c \cdot X] = c \cdot E[X].
\]
Variance and Second moment

One “partner measure” to expectation (the “first moment”) is variance (or the related measure called the second moment).

Definition
For any discrete random variable $X$, the second moment is defined as $E[X^2]$, ie, $\sum_i i^2 \Pr[X = i]$. The variance is defined as $E[(X - E[X])^2]$, ie, $\sum_i (i - E[X])^2 \Pr[X = i]$.

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

Proof.
Define $Y = (X - E[X])^2$, $Y$ is also a discrete random variable. Also $Y$ only takes non-negative values, hence $E[Y] \geq 0$. $E[Y]$ is $E[X^2 - 2X \cdot E[X] + E[X]^2]$, apply Thm 2.1 to see $E[X^2 - 2X \cdot E[X] + E[X]^2] = E[X^2] - E[2X \cdot E[X]] + E[X]^2$. By Lemma 2.2, $E[X \cdot E[X]] = E[X]^2$. Hence $E[Y] = E[X^2] - E[X]^2$, and by $E[Y] \geq 0$, we have $E[X^2] \geq E[X]^2$. \qed
Variance and Second moment

One “partner measure” to expectation (the “first moment”) is variance (or the related measure called the second moment).

Definition
For any discrete random variable $X$, the second moment is defined as $E[X^2]$, ie, $\sum_i i^2 \Pr[X = i]$. The variance is defined as $E[(X - E[X])^2]$, ie, $\sum_i (i - E[X])^2 \Pr[X = i]$.

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

Proof.
Define $Y = (X - E[X])^2$, $Y$ is also a discrete random variable. Also $Y$ only takes non-negative values, hence $E[Y] \geq 0$.
By Lemma 2.2, $E[X \cdot E[X]] = E[X]^2$.
Hence $E[Y] = E[X^2] - E[X]^2$, and by $E[Y] \geq 0$, we have $E[X^2] \geq E[X]^2$. 

RC (2016/17) – Lecture 5 – slide 5
Variance and Second moment

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

Definition

For any discrete random variable $X$, the second moment is defined as $E[X^2]$, ie, $\sum_i i^2 \Pr[X = i]$. The variance is defined as $E[(X - E[X])^2]$, ie, $\sum_i (i - E[X])^2 \Pr[X = i]$.

Lemma

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

Proof.

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

$E[Y] = E[X^2 - 2X \cdot E[X] + E[X]^2]$, apply Thm 2.1 to see


By Lemma 2.2, $E[X \cdot E[X]] = E[X]^2$.

Hence $E[Y] = E[X^2] - E[X]^2$, and by $E[Y] \geq 0$, we have $E[X^2] \geq E[X]^2$. 

RC (2016/17) – Lecture 5 – slide 5
Variance and Second moment

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

Definition
For any discrete random variable $X$, the second moment is defined as $E[X^2]$, ie, $\sum_i i^2 \Pr[X = i]$. The variance is defined as $E[(X - E[X])^2]$, ie, $\sum_i (i - E[X])^2 \Pr[X = i]$.

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

Proof.
Define $Y = (X - E[X])^2$, $Y$ is also a discrete random variable. Also $Y$ only takes non-negative values, hence $E[Y] \geq 0$. $E[Y]$ is $E[X^2 - 2X \cdot E[X] + E[X]^2]$, apply Thm 2.1 to see $E[X^2] - 2X \cdot E[X] + E[X]^2 = E[X^2] - E[2X \cdot E[X]] + E[X]^2$. By Lemma 2.2, $E[X \cdot E[X]] = E[X]^2$. Hence $E[Y] = E[X^2] - E[X]^2$, and by $E[Y] \geq 0$, we have $E[X^2] \geq E[X]^2$. 

RC (2016/17) – Lecture 5 – slide 5
Variance and Second moment

One “partner measure” to expectation (the “first moment”) is variance (or the related measure called the second moment).

Definition
For any discrete random variable $X$, the second moment is defined as $E[X^2]$, ie, $\sum_i i^2 \Pr[X = i]$. The variance is defined as $E[(X - E[X])^2]$, ie, $\sum_i (i - E[X])^2 \Pr[X = i]$.

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

Proof.
Define $Y = (X - E[X])^2$, $Y$ is also a discrete random variable.
Also $Y$ only takes non-negative values, hence $E[Y] \geq 0$.
$E[Y]$ is $E[X^2 - 2X \cdot E[X] + E[X]^2]$, apply Thm 2.1 to see
By Lemma 2.2, $E[X \cdot E[X]] = E[X]^2$.
Hence $E[Y] = E[X^2] - E[X]^2$, and by $E[Y] \geq 0$, we have $E[X^2] \geq E[X]^2$. 

RC (2016/17) – Lecture 5 – slide 5
Variance and Second moment

One “partner measure” to expectation (the “first moment”) is variance (or the related measure called the second moment).

Definition
For any discrete random variable $X$, the second moment is defined as $E[X^2]$, ie, $\sum_i i^2 \Pr[X = i]$. The variance is defined as $E[(X - E[X])^2]$, ie, $\sum_i (i - E[X])^2 \Pr[X = i]$.

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

Proof.
Define $Y = (X - E[X])^2$, $Y$ is also a discrete random variable. Also $Y$ only takes non-negative values, hence $E[Y] \geq 0$. $E[Y]$ is $E[X^2 - 2X \cdot E[X] + E[X]^2]$, apply Thm 2.1 to see $E[X^2 - 2X \cdot E[X] + E[X]^2] = E[X^2] - E[2X \cdot E[X]] + E[X]^2$. By Lemma 2.2, $E[X \cdot E[X]] = E[X]^2$. Hence $E[Y] = E[X^2] - E[X]^2$, and by $E[Y] \geq 0$, we have $E[X^2] \geq E[X]^2$. □
Variance and Second moment

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

Definition

For any discrete random variable X, the second moment is defined as $E[X^2]$, ie, $\sum_i i^2 \Pr[X = i]$. The variance is defined as $E[(X - E[X])^2]$, ie, $\sum_i (i - E[X])^2 \Pr[X = i]$.

Lemma

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

Proof.

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

$E[Y]$ is $E[X^2 - 2X \cdot E[X] + E[X]^2]$, apply Thm 2.1 to see


By Lemma 2.2, $E[X \cdot E[X]] = E[X]^2$.

Hence $E[Y] = E[X^2] - E[X]^2$, and by $E[Y] \geq 0$, we have $E[X^2] \geq E[X]^2$. 

RC (2016/17) – Lecture 5 – slide 5
Variance and Second moment

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

**Definition**
For any discrete random variable $X$, the *second moment* is defined as $E[X^2]$, ie, $\sum_i i^2 \Pr[X = i]$. The *variance* is defined as $E[(X - E[X])^2]$, ie, $\sum_i (i - E[X])^2 \Pr[X = i]$.

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

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

$E[Y]$ is $E[X^2 - 2X \cdot E[X] + E[X]^2]$, apply Thm 2.1 to see


By Lemma 2.2, $E[X \cdot E[X]] = E[X]^2$.

Hence $E[Y] = E[X^2] - E[X]^2$, and by $E[Y] \geq 0$, we have $E[X^2] \geq E[X]^2$.  \hfill \square
Variance and Second moment

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

Definition
For any discrete random variable $X$, the second moment is defined as $E[X^2]$, ie, $\sum_i i^2 \Pr[X = i]$. The variance is defined as $E[(X - E[X])^2]$, ie, $\sum_i (i - E[X])^2 \Pr[X = i]$.

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

Proof.
Define $Y = (X - E[X])^2$, $Y$ is also a discrete random variable. Also $Y$ only takes non-negative values, hence $E[Y] \geq 0$. $E[Y]$ is $E[X^2] - 2X \cdot E[X] + E[X]^2$, apply Thm 2.1 to see $E[X^2] - 2X \cdot E[X] + E[X]^2 = E[X^2] - E[2X \cdot E[X]] + E[X]^2$. By Lemma 2.2, $E[X \cdot E[X]] = E[X]^2$. Hence $E[Y] = E[X^2] - E[X]^2$, and by $E[Y] \geq 0$, we have $E[X^2] \geq E[X]^2$. 

RC (2016/17) – Lecture 5 – slide 5
Jensen’s Inequality

Definition

A function \( f : \mathbb{R} \rightarrow \mathbb{R} \) is said to be convex if it is the case that for every \( x_1, x_2 \in \mathbb{R} \) and every \( \lambda \in [0, 1] \),

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

Lemma (2.3)

For any \( f \) which is twice differentiable, \( f \) is convex around \( x \) if and only if \( f''(x) \geq 0 \).

Theorem (2.4, Jensen’s Inequality)

If \( f \) is a convex function, then

\[
E[f(X)] \geq f(E[X]).
\]
Jensen’s Inequality

Theorem (2.4, Jensen’s Inequality)

If $f$ is a convex function, then

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

Proof.
Let $\mu = E[X]$.
Assuming that $f$ is twice differentiable on its domain, then Taylor’s theorem implies there is some value $c \in (\mu, x)$ such that

$$f(x) = f(\mu) + f'(\mu)(x - \mu) + f''(c) \frac{(x - \mu)^2}{2}.$$ 

By convexity of $f$, we know $f''(c) \geq 0$ throughout domain, so $f(x) \geq f(\mu) + f'(\mu)(x - \mu)$ for all $x$. Take $E$, apply Thm 2.1, Lem 2.2,

$$E[f(X)] \geq E[f(\mu)] + E[f'(\mu)(x - \mu)] = f(E[X]) + f'(E[X]) \cdot (E[X] - E[X])$$

so the $f'$ term disappears and $E[f(X)] \geq f(E[X])$. 

RC (2016/17) – Lecture 5 – slide 7
Simple distributions

Definition
The *Bernoulli distribution* (biased coin-flip) is the random variable $Y$ such that $Y = 1$ with probability $p$ and $Y = 0$ with probability $1 - p$.

Definition (2.5)
The *binomial distribution* for $n, p$, written $B(n, p)$, is the random variable $X$ which takes values in $\{0, 1, \ldots, n\}$ with the probabilities

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

Notice $E[Y] = p$ when $Y$ is Bernoulli.

We can prove $E[X] = np$ for $X$ being $B(n, p)$ in *two ways*:

- Directly, using the Definition 2.5 and simplifying/summing the series.

- Binomial distribution $B(n, p)$ is the probabilities of getting $j$ flips from $n$ independent trials of a Bernoulli. Then use linearity of expectation.
Conditional Expectation

Definition (2.6)
For two random variables $Y, Z,$

$$E[Y \mid Z = z] = \sum_y y \cdot \Pr[Y = y \mid Z = z],$$

summation being taken over all $y$ in the support of $Y$.

Lemma (2.5)
For any random variables $X$ and $Y$,

$$E[X] = \sum_y \Pr[Y = y] \cdot E[X \mid Y = y],$$

sum taken over the support of $Y$, and we assume every $E[X \mid Y = y]$ is bounded.

Proof.
On board.
Conditional Expectation

Lemma
For any finite collection of discrete random variables $X_1, \ldots, X_n$ with finite expectations, and for any random variable $X$,

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

Definition (2.7)
We will sometimes use the expression $E[Y \mid Z]$, where $Y, Z$ are existing random variables. $E[Y \mid Z]$ itself is a random variable which is a function of $Z$, having the value $E[Y \mid Z = z]$ when applied to $z$. 

RC (2016/17) – Lecture 5 – slide 10
Geometric distributions

Motivated by the scenario where 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 geometric random variable $X$ with parameter $p$ is given by the following probability distribution on $\mathbb{N}$:

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

Should verify that $\sum_{j=1}^{\infty} \Pr[X = j] = 1$ (board).

Geometric random variables are memoryless (like Markov chains ...):

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, we have the following:

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

Proof.

Like this one, will do on board.

Observation

If $X$ is a geometric random variable $X$ with parameter $p$, then for any $i \geq 0$, $\Pr[X \geq i] = (1 - p)^{i-1}$.

Proof.

We have $\Pr[X \geq i] = \sum_{j=i}^{\infty} (1 - p)^{j-1} \cdot p = p \sum_{j=i}^{\infty} (1 - p)^{j-1}$.

Sum $\sum_{j=i}^{\infty} (1 - p)^{j-1}$ as $(1 - p)^{i-1} \frac{1-(1-p)^{\infty}}{1-(1-p)} = (1 - p)^{i-1} p^{-1}$.

Hence $\Pr[X \geq i] = p \cdot (1 - p)^{i-1} p^{-1} = (1 - p)^{i-1}$. 

RC (2016/17) – Lecture 5 – slide 12
Geometric distributions

Lemma (2.9)

For any discrete random variable $X$ that only takes non-negative integer values, we have the following:

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

Proof.
Like this one, will do on board. \qed

Observation

If $X$ is a geometric random variable $X$ with parameter $p$, then for any $i \geq 0$, $\Pr[X \geq i] = (1 - p)^{i-1}$.

Proof.
We have $\Pr[X \geq i] = \sum_{j=i}^{\infty} (1 - p)^{j-1} \cdot p = p \sum_{j=i}^{\infty} (1 - p)^{j-1}$.

Sum $\sum_{j=i}^{\infty} (1 - p)^{j-1}$ as $(1 - p)^{i-1} \frac{1-(1-p)^{\infty}}{1-(1-p)} = (1 - p)^{i-1} p^{-1}$.

Hence $\Pr[X \geq i] = p \cdot (1 - p)^{i-1} p^{-1} = (1 - p)^{i-1}$.
Geometric distributions

Lemma (2.9)

For any discrete random variable $X$ that only takes non-negative integer values, we have the following:

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

Proof.

Like this one, will do on board. \qed

Observation

If $X$ is a geometric random variable $X$ with parameter $p$, then for any $i \geq 0$, $\Pr[X \geq i] = (1 - p)^{i-1}$.

Proof.

We have $\Pr[X \geq i] = \sum_{j=i}^{\infty} (1 - p)^{j-1} \cdot p = p \sum_{j=i}^{\infty} (1 - p)^{j-1}$.

Sum $\sum_{j=i}^{\infty} (1 - p)^{j-1}$ as $(1 - p)^{i-1} \frac{1 - (1 - p)^{\infty}}{1 - (1 - p)} = (1 - p)^{i-1} p^{-1}$.

Hence $\Pr[X \geq i] = p \cdot (1 - p)^{i-1} p^{-1} = (1 - p)^{i-1}$. 

RC (2016/17) – Lecture 5 – slide 12
Geometric distributions

Lemma

If $X$ is a geometric random variable $X$ with parameter $p$, then $E[X] = p^{-1}$.

Proof.

We just apply Lemma 2.9.

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

For a geometric random variable, parameter $p$,

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

which is (using closed form for geometric series again)

$$\frac{1 - (1 - p)^\infty}{1 - (1 - p)} = \frac{1}{p} = p^{-1}.$$
Coupon Collector Problem

“Coupon collecting" is the activity of buying cereal-packets, each of which will have a coupon inside. There are be \( n \) different types of “coupon" (eg photo of different footballers) and the goal is to collect one copy of each ... then stop buying the product.

How many packets do we (expect to) need to buy?

Assumptions:

- Items are randomly and identically distributed in packets (one card per packet). So when buying a box the probability of any particular footballer being inside is \( 1/n \).
Coupon Collector Analysis

- Analyse the “wait time” to get a card not seen before.
- Let $X$ be the number of packets bought to get all cards.
- Let $X_i$ be the number of packets bought (after having $i - 1$ different cards) to get $i$ different cards.
- Clearly $X = \sum_{i=1}^{n} X_i$.
- $X_i$ can be modelled as a geometric random variable where $p_i = \frac{n-(i-1)}{n} = 1 - \frac{i-1}{n}$.

Thm 2.3 (linearity of $E$) says $E[X] = \sum_{i=1}^{n} E[X_i]$.

By Lemma on geometric random variables (2 slides ago) $E[X_i] = \frac{n}{n-(i-1)}$ for every $i$.

Hence $E[X] = \sum_{i=1}^{n} \frac{n}{n-(i-1)} = \sum_{i=1}^{n} \frac{n}{i} = n \sum_{i=1}^{n} \frac{1}{i}$, which is $\sim n \ln(n)$.

We wait $n \ln(n) + \Theta(1)$ to get all cards.
Coupon Collector Analysis

- Analyse the “wait time” to get a card not seen before.
- Let $X$ be the number of packets bought to get all cards.
- Let $X_i$ be the number of packets bought (after having $i - 1$ different cards) to get $i$ different cards.
- Clearly $X = \sum_{i=1}^{n} X_i$.
- $X_i$ can be modelled as a geometric random variable where $p_i = \frac{n-(i-1)}{n} = 1 - \frac{i-1}{n}$.

Thm 2.3 (linearity of $E$) says $E[X] = \sum_{i=1}^{n} E[X_i]$.

By Lemma on geometric random variables (2 slides ago) $E[X_i] = \frac{n}{n-(i-1)}$ for every $i$.

Hence $E[X] = \sum_{i=1}^{n} \frac{n}{n-(i-1)} = \sum_{i=1}^{n} \frac{n}{i} = n \sum_{i=1}^{n} \frac{1}{i}$, which is $\sim n \ln(n)$.

We wait $n \ln(n) + \Theta(1)$ to get all cards.
Coupon Collector Analysis

- Analyse the “wait time” to get a card not seen before.
- Let $X$ be the number of packets bought to get all cards.
- Let $X_i$ be the number of packets bought (after having $i - 1$ different cards) to get $i$ different cards.
- Clearly $X = \sum_{i=1}^{n} X_i$.
- $X_i$ can be modelled as a geometric random variable where $p_i = \frac{n-(i-1)}{n} = 1 - \frac{i-1}{n}$.

Thm 2.3 (linearity of $E$) says $E[X] = \sum_{i=1}^{n} E[X_i]$.

By Lemma on geometric random variables (2 slides ago)
$E[X_i] = \frac{n}{n-(i-1)}$ for every $i$.

Hence $E[X] = \sum_{i=1}^{n} \frac{n}{n-(i-1)} = \sum_{i=1}^{n} \frac{n}{i} = n \sum_{i=1}^{n} \frac{1}{i}$, which is $\sim n \ln(n)$.

We wait $n \ln(n) + \Theta(1)$ to get all cards.
Read parts of Chapter 3 of "Probability and Computing". We will be looking at the Markov, Chebyshev etc inequalities on Friday.