Randomness and Computation

or, “Randomized Algorithms”

Mary Cryan

School of Informatics
University of Edinburgh

Discrete Random Variables

Much of our reasoning in RC is in terms of *random variables*, especially *discrete random variables* (when X can take on a finite or countable number of values). I assume standard definitions (and known facts!).

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,

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

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

Discrete Random Variables

Much of our reasoning in RC is in terms of *random variables*, especially *discrete random variables* (when X can take on a finite or countable number of values). I assume standard definitions (and known facts!).

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,

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

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

Discrete Random Variables

Much of our reasoning in RC is in terms of *random variables*, especially *discrete random variables* (when X can take on a finite or countable number of values). I assume standard definitions (and known facts!).

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,

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

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

Discrete Random Variables

Much of our reasoning in RC is in terms of *random variables*, especially *discrete random variables* (when X can take on a finite or countable number of values). I assume standard definitions (and known facts!).

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,

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

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, $\mathrm{E}[c \cdot X] = c \cdot \mathrm{E}[X]$.

Definition (2.2)

A collection $X_1, \ldots, X_k$ of random variables are said to be *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

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

Stronger than “pairwise independent" - 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 relationships work out but $\Pr[(A = 1) \cap (B = 1) \cap (C = 1)] = ?$.

Variance and Second moment

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, the *second moment* is defined as $\mathrm{E}[X^2]$, ie, $\sum_i i^2 \Pr[X = i]$ (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, $\mathrm{E}[X^2] \geq \mathrm{E}[X]^2$.

Proof.

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

Variance and Second moment

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, the *second moment* is defined as $\mathrm{E}[X^2]$, ie, $\sum_i i^2 \Pr[X = i]$ (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, $\mathrm{E}[X^2] \geq \mathrm{E}[X]^2$.

Proof.

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

Variance and Second moment

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, the *second moment* is defined as $\mathrm{E}[X^2]$, ie, $\sum_i i^2 \Pr[X = i]$ (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, $\mathrm{E}[X^2] \geq \mathrm{E}[X]^2$.

Proof.

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

Variance and Second moment

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, the *second moment* is defined as $\mathrm{E}[X^2]$, ie, $\sum_i i^2 \Pr[X = i]$ (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, $\mathrm{E}[X^2] \geq \mathrm{E}[X]^2$.

Proof.

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

Variance and Second moment

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, the *second moment* is defined as $\mathrm{E}[X^2]$, ie, $\sum_i i^2 \Pr[X = i]$ (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, $\mathrm{E}[X^2] \geq \mathrm{E}[X]^2$.

Proof.

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

Variance and Second moment

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, the *second moment* is defined as $\mathrm{E}[X^2]$, ie, $\sum_i i^2 \Pr[X = i]$ (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, $\mathrm{E}[X^2] \geq \mathrm{E}[X]^2$.

Proof.

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

□

Variance and Second moment

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, the *second moment* is defined as $\mathrm{E}[X^2]$, ie, $\sum_i i^2 \Pr[X = i]$ (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, $\mathrm{E}[X^2] \geq \mathrm{E}[X]^2$.

Proof.

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

Variance and Second moment

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, the *second moment* is defined as $\mathrm{E}[X^2]$, ie, $\sum_i i^2 \Pr[X = i]$ (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, $\mathrm{E}[X^2] \geq \mathrm{E}[X]^2$.

Proof.

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

Variance and Second moment

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, the *second moment* is defined as $\mathrm{E}[X^2]$, ie, $\sum_i i^2 \Pr[X = i]$ (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, $\mathrm{E}[X^2] \geq \mathrm{E}[X]^2$.

Proof.

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

Jensen's Inequality

Definition

A function $f : \mathbb{R} \to \mathbb{R}$ 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

$$\mathrm{E}[f(X)] \; \geq \; f(\mathrm{E}[X]).$$

Jensen's Inequality

Theorem (2.4, Jensen's Inequality)

If f is a convex function, then

$$\mathrm{E}[f(X)] \;\geq\; f(\mathrm{E}[X]).$$

Proof.

Let $\mu = \mathrm{E}[X]$.

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 $f''(\cdot) \geq 0$ throughout domain, so $f(x) \geq f(\mu) + f'(\mu)(x - \mu)$ for all x. Take E, apply Thm 2.1, Lem 2.2,

$$\mathrm{E}[f(X)] \;\geq\; \mathrm{E}[f(\mu)] + \mathrm{E}[f'(\mu)(X - \mu)] \;=\; f(\mathrm{E}[X]) + f'(\mathrm{E}[X]) \cdot (\mathrm{E}[X] - \mathrm{E}[X])$$

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

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 $1 - p$.

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

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

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

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

- Directly, using 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,

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

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

Lemma (2.5)

For any random variables X and Y,

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

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

Proof.

On visualiser. □

Conditional Expectation

Observation

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

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

Definition (2.7)

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

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 *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 $\sum_{j=1}^{\infty} \Pr[X = j] = 1$ (on visualiser).

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:*

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

Proof.

Like this one, will do on visualiser. □

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:*

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

Proof.

Like this one, will do on visualiser. □

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:*

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

Proof.

Like this one, will do on visualiser. □

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

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

Proof.

We just apply Lemma 2.9.
We have $\mathrm{E}[X] = \sum_{i=1}^{\infty} \Pr[X \geq i]$.
For a geometric random variable, parameter p,

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

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 cards with a photo of a foot-baller) and the goal is to collect one copy of each . . . then stop buying.*
>
> *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 card being inside is $1/n$.

Coupon Collector Analysis

How to analyse the process?

Could evaluate *expected number of purchases* to get card j (that particular footballer), for any $1 \le i \le n$. The "number of steps" Y_j is a *geometric random variable* with parameter $1/n$. By our Lemma from Tuesday, $\mathrm{E}[Y_j] = \frac{1}{(1/n)} = n$ for any j.

But this is n Y_j variables in total, which are **not** independent (WHY?). Hard to combine them for a tight estimate. Better to find another angle ... not focused on any particular card. So ...

- At any stage of the process (having found some cards already), analyse the *"further purchases"* to get a *card not seen before*.
- Let X_i be the number of packets bought (after having $i-1$ different cards) to get the ith new card.
- Let X be the number of packets bought to get all cards.
- Clearly $X = \sum_{i=1}^{n} X_i$.

Coupon Collector Analysis

X_i can also be modelled as a *geometric random variable*; if we own $i-1$ different cards, and buy one more packet, the (conditional) probability p_i that we get a *new* card is $p_i = \frac{n-(i-1)}{n} = 1 - \frac{i-1}{n}$.

Linearity of $\mathrm{E}[\cdot]$ says $\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{E}[X_i]$.

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

Hence $\mathrm{E}[X] = \sum_{i=1}^{n} \frac{n}{n-(i-1)} = \sum_{i=n}^{1} \frac{n}{i} = n(\sum_{i=1}^{n} \frac{1}{i})$.

$H(n) = \sum_{i=1}^{n} \frac{1}{i}$ is a crude "Riemann sum" to approximate $\int_{x=1}^{n} \frac{1}{x}$.

Can show $\int_{x=1}^{n} \frac{1}{x} < \sum_{i=1}^{n} \frac{1}{i}$ and $\sum_{i=2}^{n} \frac{1}{i} < \int_{x=1}^{n} \frac{1}{x}$ *(Fig 2.1 in book)*.

Hence $\ln(n) < \sum_{i=1}^{n} \frac{1}{i} \leq \ln(n) + 1$.

So the expected time $\mathrm{E}[X]$ to collect all cards is $\sim n \ln(n)$.

Coupon Collector Analysis

X_i can also be modelled as a *geometric random variable*; if we own $i-1$ different cards, and buy one more packet, the (conditional) probability p_i that we get a *new* card is $p_i = \frac{n-(i-1)}{n} = 1 - \frac{i-1}{n}$.

Linearity of $\mathrm{E}[\cdot]$ says $\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{E}[X_i]$.

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

Hence $\mathrm{E}[X] = \sum_{i=1}^{n} \frac{n}{n-(i-1)} = \sum_{i=n}^{1} \frac{n}{i} = n(\sum_{i=1}^{n} \frac{1}{i})$.

$H(n) = \sum_{i=1}^{n} \frac{1}{i}$ is a crude "Riemann sum" to approximate $\int_{x=1}^{n} \frac{1}{x}$.

Can show $\int_{x=1}^{n} \frac{1}{x} < \sum_{i=1}^{n} \frac{1}{i}$ and $\sum_{i=2}^{n} \frac{1}{i} < \int_{x=1}^{n} \frac{1}{x}$ *(Fig 2.1 in book)*.

Hence $\ln(n) < \sum_{i=1}^{n} \frac{1}{i} \leq \ln(n) + 1$.

So the expected time $\mathrm{E}[X]$ to collect all cards is $\sim n \ln(n)$.

Coupon Collector Analysis

X_i can also be modelled as a *geometric random variable*; if we own $i-1$ different cards, and buy one more packet, the (conditional) probability p_i that we get a *new* card is $p_i = \frac{n-(i-1)}{n} = 1 - \frac{i-1}{n}$.

Linearity of $\mathrm{E}[\cdot]$ says $\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{E}[X_i]$.

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

Hence $\mathrm{E}[X] = \sum_{i=1}^{n} \frac{n}{n-(i-1)} = \sum_{i=n}^{1} \frac{n}{i} = n(\sum_{i=1}^{n} \frac{1}{i})$.

$H(n) = \sum_{i=1}^{n} \frac{1}{i}$ is a crude "Riemann sum" to approximate $\int_{x=1}^{n} \frac{1}{x}$.

Can show $\int_{x=1}^{n} \frac{1}{x} < \sum_{i=1}^{n} \frac{1}{i}$ and $\sum_{i=2}^{n} \frac{1}{i} < \int_{x=1}^{n} \frac{1}{x}$ *(Fig 2.1 in book)*.

Hence $\ln(n) < \sum_{i=1}^{n} \frac{1}{i} \leq \ln(n) + 1$.

So the expected time $\mathrm{E}[X]$ to collect all cards is $\sim n\ln(n)$.

Coupon Collector Analysis

X_i can also be modelled as a *geometric random variable*; if we own $i-1$ different cards, and buy one more packet, the (conditional) probability p_i that we get a *new* card is $p_i = \frac{n-(i-1)}{n} = 1 - \frac{i-1}{n}$.

Linearity of $\mathrm{E}[\cdot]$ says $\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{E}[X_i]$.

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

Hence $\mathrm{E}[X] = \sum_{i=1}^{n} \frac{n}{n-(i-1)} = \sum_{i=n}^{1} \frac{n}{i} = n(\sum_{i=1}^{n} \frac{1}{i})$.

$H(n) = \sum_{i=1}^{n} \frac{1}{i}$ is a crude “Riemann sum" to approximate $\int_{x=1}^{n} \frac{1}{x}$.

Can show $\int_{x=1}^{n} \frac{1}{x} < \sum_{i=1}^{n} \frac{1}{i}$ and $\sum_{i=2}^{n} \frac{1}{i} < \int_{x=1}^{n} \frac{1}{x}$ *(Fig 2.1 in book)*.

Hence $\ln(n) < \sum_{i=1}^{n} \frac{1}{i} \leq \ln(n) + 1$.

So the expected time $\mathrm{E}[X]$ to collect all cards is $\sim n \ln(n)$.

Coupon Collector Analysis

X_i can also be modelled as a *geometric random variable*; if we own $i-1$ different cards, and buy one more packet, the (conditional) probability p_i that we get a *new* card is $p_i = \frac{n-(i-1)}{n} = 1 - \frac{i-1}{n}$.

Linearity of $\mathrm{E}[\cdot]$ says $\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{E}[X_i]$.

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

Hence $\mathrm{E}[X] = \sum_{i=1}^{n} \frac{n}{n-(i-1)} = \sum_{i=n}^{1} \frac{n}{i} = n(\sum_{i=1}^{n} \frac{1}{i})$.

$H(n) = \sum_{i=1}^{n} \frac{1}{i}$ is a crude "Riemann sum" to approximate $\int_{x=1}^{n} \frac{1}{x}$.

Can show $\int_{x=1}^{n} \frac{1}{x} < \sum_{i=1}^{n} \frac{1}{i}$ and $\sum_{i=2}^{n} \frac{1}{i} < \int_{x=1}^{n} \frac{1}{x}$ *(Fig 2.1 in book)*.

Hence $\ln(n) < \sum_{i=1}^{n} \frac{1}{i} \leq \ln(n) + 1$.

So the expected time $\mathrm{E}[X]$ to collect all cards is $\sim n \ln(n)$.

Coupon Collector Analysis

X_i can also be modelled as a *geometric random variable*; if we own $i-1$ different cards, and buy one more packet, the (conditional) probability p_i that we get a *new* card is $p_i = \frac{n-(i-1)}{n} = 1 - \frac{i-1}{n}$.

Linearity of $\mathrm{E}[\cdot]$ says $\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{E}[X_i]$.

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

Hence $\mathrm{E}[X] = \sum_{i=1}^{n} \frac{n}{n-(i-1)} = \sum_{i=n}^{1} \frac{n}{i} = n(\sum_{i=1}^{n} \frac{1}{i})$.

$H(n) = \sum_{i=1}^{n} \frac{1}{i}$ is a crude "Riemann sum" to approximate $\int_{x=1}^{n} \frac{1}{x}$.

Can show $\int_{x=1}^{n} \frac{1}{x} < \sum_{i=1}^{n} \frac{1}{i}$ and $\sum_{i=2}^{n} \frac{1}{i} < \int_{x=1}^{n} \frac{1}{x}$ *(Fig 2.1 in book)*.

Hence $\ln(n) < \sum_{i=1}^{n} \frac{1}{i} \leq \ln(n) + 1$.

So the expected time $\mathrm{E}[X]$ to collect all cards is $\sim n\ln(n)$.

Coupon Collector Analysis

X_i can also be modelled as a *geometric random variable*; if we own $i-1$ different cards, and buy one more packet, the (conditional) probability p_i that we get a *new* card is $p_i = \frac{n-(i-1)}{n} = 1 - \frac{i-1}{n}$.

Linearity of $\mathrm{E}[\cdot]$ says $\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{E}[X_i]$.

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

Hence $\mathrm{E}[X] = \sum_{i=1}^{n} \frac{n}{n-(i-1)} = \sum_{i=n}^{1} \frac{n}{i} = n(\sum_{i=1}^{n} \frac{1}{i})$.

$H(n) = \sum_{i=1}^{n} \frac{1}{i}$ is a crude "Riemann sum" to approximate $\int_{x=1}^{n} \frac{1}{x}$.

Can show $\int_{x=1}^{n} \frac{1}{x} < \sum_{i=1}^{n} \frac{1}{i}$ and $\sum_{i=2}^{n} \frac{1}{i} < \int_{x=1}^{n} \frac{1}{x}$ *(Fig 2.1 in book)*.

Hence $\ln(n) < \sum_{i=1}^{n} \frac{1}{i} \leq \ln(n) + 1$.

So the expected time $\mathrm{E}[X]$ to collect all cards is $\sim n\ln(n)$.

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 know how likely a example “run" of the process is to come close to that value.

Inequalities like *Markov’s Inequality*, *Chebyshev’s Inequality*, *Chernoff/Hoeffding Bounds* help us bound *deviation from the mean*.

Have a look in Chapter 3 of “Probability and Computing" to remind yourself of the Markov, Chebyshev etc inequalities before Friday.