

## Randomness and Computation

or, “Randomized Algorithms”

Heng Guo

(Based on slides by M. Cryan)

RC (2019/20) – Lecture 5 – slide 1

## Coupon Collector Problem

“Coupon collecting” is the activity of buying packs, each of which will have a uniform at random coupon inside. There are  $n$  different types of “coupon” and the goal is to collect one copy of each ... then stop buying.

Last time we have showed that the expected number  $E[X]$  of purchases to collect all cards is  $nH(n) \sim n \ln(n)$ .

Today we examine how likely a example “run” of the purchasing process is to come close to that expectation.

Concentration inequalities will be vital:

- ▶ Markov Inequality;
- ▶ Chebyshev Inequality;
- ▶ Chernoff Bound / Hoeffding inequality.

RC (2019/20) – Lecture 5 – slide 2

## Markov Inequality

The simplest one.

### Theorem (3.1, Markov Inequality)

Let  $X$  be any random variable that takes only non-negative values. Then for any  $a > 0$ ,

$$\Pr[X \geq a] \leq \frac{E[X]}{a}.$$

RC (2019/20) – Lecture 5 – slide 3

## Markov Inequality

The simplest one.

### Theorem (3.1, Markov Inequality)

Let  $X$  be any random variable that takes only non-negative values. Then for any  $a > 0$ ,

$$\Pr[X \geq a] \leq \frac{E[X]}{a}.$$

**Proof.**

Define the indicator function  $I = I(X)$  by

$$I(x) = \begin{cases} 0 & x < a; \\ 1 & x \geq a. \end{cases}$$

RC (2019/20) – Lecture 5 – slide 3

## Markov Inequality

The simplest one.

### Theorem (3.1, Markov Inequality)

Let  $X$  be any random variable that takes only non-negative values. Then for any  $a > 0$ ,

$$\Pr[X \geq a] \leq \frac{E[X]}{a}.$$

### Proof.

Define the indicator function  $I = I(X)$  by

$$I(x) = \begin{cases} 0 & x < a; \\ 1 & x \geq a. \end{cases}$$

Then  $X \geq a \cdot I(X)$ , and hence  $I(X) \leq \frac{X}{a}$ .

RC (2019/20) – Lecture 5 – slide 3

## Markov Inequality

The simplest one.

### Theorem (3.1, Markov Inequality)

Let  $X$  be any random variable that takes only non-negative values. Then for any  $a > 0$ ,

$$\Pr[X \geq a] \leq \frac{E[X]}{a}.$$

### Proof.

Define the indicator function  $I = I(X)$  by

$$I(x) = \begin{cases} 0 & x < a; \\ 1 & x \geq a. \end{cases}$$

Then  $X \geq a \cdot I(X)$ , and hence  $I(X) \leq \frac{X}{a}$ .

Taking expectation of both sides, and using  $E[I] = \Pr[X \geq a]$ , we have

$$\Pr[X \geq a] = E[I] \leq \frac{1}{a}E[X]. \quad \square$$

RC (2019/20) – Lecture 5 – slide 3

## Bounding Coupon Collector purchases - Markov

Recall that  $X$  is the number of purchases of the coupon collector problem and  $E[X] = n \ln n + \Theta(n)$ .

Say we want a bound  $T$  so that the probability of  $X \geq T$  is at most  $\frac{1}{n}$ .

By Markov ineq.,  $\Pr[X \geq T] \leq \frac{E[X]}{T}$ . Thus, we need  $T$  to be at least  $n^2 \ln n$ .

RC (2019/20) – Lecture 5 – slide 4

## Bounding Coupon Collector purchases - Markov

Recall that  $X$  is the number of purchases of the coupon collector problem and  $E[X] = n \ln n + \Theta(n)$ .

Say we want a bound  $T$  so that the probability of  $X \geq T$  is at most  $\frac{1}{n}$ .

By Markov ineq.,  $\Pr[X \geq T] \leq \frac{E[X]}{T}$ . Thus, we need  $T$  to be at least  $n^2 \ln n$ .

This is far from tight!

The power of Markov ineq. is that it does not require any other knowledge of the random variable. However for specific problems, we can often do better.

For example, we can bound the variance.

RC (2019/20) – Lecture 5 – slide 4

## Variance, Moments of a Random Variable

### Definition (3.1)

The *kth moment* of a random variable  $X$  is defined to be  $E[X^k]$ .

### Definition (3.2)

The *variance* of a random variable is defined to be

$$\text{Var}[X] := E[(X - E[X])^2] = E[X^2] - E[X]^2.$$

RC (2019/20) – Lecture 5 – slide 5

## Covariance of two Random Variables

### Definition (3.3)

The *covariance* of two random variables  $X$  and  $Y$  is defined as

$$\text{Cov}[X, Y] = E[(X - E[X])(Y - E[Y])].$$

### Theorem (3.2)

For any two random variables  $X, Y$ , we have

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y].$$

RC (2019/20) – Lecture 5 – slide 6

## Covariance of two Random Variables

### Definition (3.3)

The *covariance* of two random variables  $X$  and  $Y$  is defined as

$$\text{Cov}[X, Y] = E[(X - E[X])(Y - E[Y])].$$

### Theorem (3.2)

For any two random variables  $X, Y$ , we have

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y].$$

**Proof.**

$$\begin{aligned}\text{Var}[X + Y] &= E[(X + Y)^2] - E[X + Y]^2 \\ &= E[X^2] + E[Y^2] + 2E[XY] - E[X]^2 - E[Y]^2 - 2E[X]E[Y] \\ &= \text{Var}[X] + \text{Var}[Y] + 2(E[XY] - E[X]E[Y]).\end{aligned}$$

□

RC (2019/20) – Lecture 5 – slide 6

## Covariance of two Random Variables

### Definition (3.3)

The *covariance* of two random variables  $X$  and  $Y$  is defined as

$$\text{Cov}[X, Y] = E[(X - E[X])(Y - E[Y])].$$

### Theorem (3.2)

For any two random variables  $X, Y$ , we have

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y].$$

**Proof.**

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2(E[XY] - E[X]E[Y]).$$

□

RC (2019/20) – Lecture 5 – slide 6

## Covariance of two Random Variables

### Definition (3.3)

The *covariance* of two random variables  $X$  and  $Y$  is defined as

$$\text{Cov}[X, Y] = E[(X - E[X])(Y - E[Y])].$$

### Theorem (3.2)

For any two random variables  $X, Y$ , we have

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y].$$

**Proof.**

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2(E[XY] - E[X]E[Y]).$$

$$\begin{aligned}\text{Cov}[X, Y] &= E[XY - E[X]Y - E[Y]X + E[X]E[Y]] \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y]\end{aligned}$$

□

RC (2019/20) – Lecture 5 – slide 6

## (pairwise) Independent Random Variables

### Theorem (3.3)

If  $X, Y$  are a pair of independent random variables, then

$$E[XY] = E[X] \cdot E[Y].$$

RC (2019/20) – Lecture 5 – slide 7

## (pairwise) Independent Random Variables

### Theorem (3.3)

If  $X, Y$  are a pair of independent random variables, then

$$E[XY] = E[X] \cdot E[Y].$$

### Corollary (3.4)

If  $X, Y$  are a pair of independent random variables, then

$$\text{Cov}[X, Y] = 0$$

and

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y].$$

**Proof** is straightforward application of Thm 3.3.

RC (2019/20) – Lecture 5 – slide 7

## Chebyshev Inequality

### Theorem (3.2, Chebyshev Inequality)

For every  $a > 0$ ,

$$\Pr[|X - E[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2}.$$

RC (2019/20) – Lecture 5 – slide 8

## Chebyshev Inequality

### Theorem (3.2, Chebyshev Inequality)

For every  $a > 0$ ,

$$\Pr[|X - E[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2}.$$

#### Proof.

First we claim for any  $a > 0$ ,

$$|X - E[X]| \geq a \Leftrightarrow (X - E[X])^2 \geq a^2$$

RC (2019/20) – Lecture 5 – slide 8

## Chebyshev Inequality

### Theorem (3.2, Chebyshev Inequality)

For every  $a > 0$ ,

$$\Pr[|X - E[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2}.$$

#### Proof.

First we claim for any  $a > 0$ ,

$$|X - E[X]| \geq a \Leftrightarrow (X - E[X])^2 \geq a^2$$

Applying **Markov Ineq.** to the random variable  $(X - E[X])^2$ , we know

$$\Pr[(X - E[X])^2 \geq a^2] \leq \frac{E[(X - E[X])^2]}{a^2},$$

RC (2019/20) – Lecture 5 – slide 8

## Chebyshev Inequality

### Theorem (3.2, Chebyshev Inequality)

For every  $a > 0$ ,

$$\Pr[|X - E[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2}.$$

#### Proof.

First we claim for any  $a > 0$ ,

$$|X - E[X]| \geq a \Leftrightarrow (X - E[X])^2 \geq a^2$$

Applying **Markov Ineq.** to the random variable  $(X - E[X])^2$ , we know

$$\Pr[(X - E[X])^2 \geq a^2] \leq \frac{E[(X - E[X])^2]}{a^2},$$

and by definition of  $\text{Var}(\cdot)$ , this gives

$$\Pr[|X - E[X]| \geq a] = \Pr[(X - E[X])^2 \geq a^2] \leq \frac{\text{Var}[X]}{a^2}. \quad \square$$

RC (2019/20) – Lecture 5 – slide 8

## Bounding Coupon Collector purchases - Markov

Recall that  $X$  is the number of purchases of the coupon collector problem and  $E[X] = n \ln n + \Theta(n)$ .

Using Markov ineq., we can get an upper bound of a “typical” number of the order  $n^2 \ln n$ , which is not particularly interesting.

We can do better with Chebyshev ineq. ...

RC (2019/20) – Lecture 5 – slide 9

## Bounding Coupon Collector purchases - Chebyshev

$$\Pr[|X - E[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2}.$$

RC (2019/20) – Lecture 5 – slide 10

## Bounding Coupon Collector purchases - Chebyshev

$$\Pr[|X - E[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2}.$$

- ▶ Need to evaluate  $\text{Var}[X]$ , which is  $\text{Var}[X_1 + \dots + X_n]$ .  
Recall that  $X_i$  is the *number of packets* bought to get the  $i$ -th new card.

RC (2019/20) – Lecture 5 – slide 10

## Bounding Coupon Collector purchases - Chebyshev

$$\Pr[|X - E[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2}.$$

- ▶ Need to evaluate  $\text{Var}[X]$ , which is  $\text{Var}[X_1 + \dots + X_n]$ .  
Recall that  $X_i$  is the *number of packets* bought to get the  $i$ -th new card.
- ▶ Corollary 3.4: for independent  $Y, Z$ ,  $\text{Var}[Y + Z] = \text{Var}[Y] + \text{Var}[Z]$ .  
Are these  $X_i$ 's independent?

RC (2019/20) – Lecture 5 – slide 10

## Bounding Coupon Collector purchases - Chebyshev

$$\Pr[|X - E[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2}.$$

- ▶ Need to evaluate  $\text{Var}[X]$ , which is  $\text{Var}[X_1 + \dots + X_n]$ .  
Recall that  $X_i$  is the *number of packets* bought to get the  $i$ -th new card.
- ▶ Corollary 3.4: for independent  $Y, Z$ ,  $\text{Var}[Y + Z] = \text{Var}[Y] + \text{Var}[Z]$ .  
Are these  $X_i$ 's independent?
- ▶  $X_i$  is independent of the value of  $X_{i-1}$  or any of the earlier values.  $X_i$  only depends on the values  $n$  and  $i$  (and not on what cards we have collected or how long it takes to collect them).

RC (2019/20) – Lecture 5 – slide 10

## Bounding Coupon Collector purchases - Chebyshev

$$\Pr[|X - E[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2}.$$

- ▶ Need to evaluate  $\text{Var}[X]$ , which is  $\text{Var}[X_1 + \dots + X_n]$ .  
Recall that  $X_i$  is the *number of packets* bought to get the  $i$ -th new card.
- ▶ Corollary 3.4: for independent  $Y, Z$ ,  $\text{Var}[Y + Z] = \text{Var}[Y] + \text{Var}[Z]$ .  
Are these  $X_i$ 's independent?
- ▶  $X_i$  is independent of the value of  $X_{i-1}$  or any of the earlier values.  $X_i$  only depends on the values  $n$  and  $i$  (and not on what cards we have collected or how long it takes to collect them).
- ▶ Hence the random variables  $X_1, \dots, X_n$  are all mutually independent, and

$$\text{Var}[X] = \text{Var}[X_1] + \text{Var}[X_2] + \dots + \text{Var}[X_n].$$

RC (2019/20) – Lecture 5 – slide 10

## Bounding Coupon Collector purchases - Chebyshev

Each  $X_i$  is a geometric random variable with parameter  $\frac{n-(i-1)}{n}$ .

RC (2019/20) – Lecture 5 – slide 11

## Bounding Coupon Collector purchases - Chebyshev

Each  $X_i$  is a geometric random variable with parameter  $\frac{n-(i-1)}{n}$ .

### Lemma (3.8)

For any geometric random variable  $X$  with parameter  $p$ ,  $E[X] = p^{-1}$  and  $\text{Var}[X] = \frac{1-p}{p^2}$ .

RC (2019/20) – Lecture 5 – slide 11

## Bounding Coupon Collector purchases - Chebyshev

Each  $X_i$  is a geometric random variable with parameter  $\frac{n-(i-1)}{n}$ .

### Lemma (3.8)

For any geometric random variable  $X$  with parameter  $p$ ,  $E[X] = p^{-1}$  and  $\text{Var}[X] = \frac{1-p}{p^2}$ .

### Proof.

We have  $\text{Var}[X] = E[X^2] - E[X]^2$ . For geometric variable,  $E[X]^2 = p^{-2}$ .

RC (2019/20) – Lecture 5 – slide 11

## Bounding Coupon Collector purchases - Chebyshev

Each  $X_i$  is a geometric random variable with parameter  $\frac{n-(i-1)}{n}$ .

### Lemma (3.8)

For any geometric random variable  $X$  with parameter  $p$ ,  $E[X] = p^{-1}$  and  $\text{Var}[X] = \frac{1-p}{p^2}$ .

### Proof.

We have  $\text{Var}[X] = E[X^2] - E[X]^2$ . For geometric variable,  $E[X]^2 = p^{-2}$ . For  $E[X^2]$ , we could do direct calculation. We can also consider conditional expectation. Once again, let  $Y$  indicate the outcome of the first trial.

RC (2019/20) – Lecture 5 – slide 11

## Bounding Coupon Collector purchases - Chebyshev

Each  $X_i$  is a geometric random variable with parameter  $\frac{n-(i-1)}{n}$ .

### Lemma (3.8)

For any geometric random variable  $X$  with parameter  $p$ ,  $E[X] = p^{-1}$  and  $\text{Var}[X] = \frac{1-p}{p^2}$ .

### Proof.

We have  $\text{Var}[X] = E[X^2] - E[X]^2$ . For geometric variable,  $E[X]^2 = p^{-2}$ . For  $E[X^2]$ , we could do direct calculation. We can also consider conditional expectation. Once again, let  $Y$  indicate the outcome of the first trial.

$$\begin{aligned} E[X^2] &= \Pr[Y = 1]E[X^2 | Y = 1] + \Pr[Y = 0]E[X^2 | Y = 0] \\ &= p + (1-p) \sum_{x \geq 1} x^2 \Pr[X = x | Y = 0] \\ &= p + (1-p) \sum_{x \geq 2} x^2 \Pr[X = x - 1] \\ &= p + (1-p) \sum_{x \geq 1} (x+1)^2 \Pr[X = x] = p + (1-p)E[(X+1)^2] \end{aligned}$$

RC (2019/20) – Lecture 5 – slide 11

## Bounding Coupon Collector purchases - Chebyshev

### Lemma (3.8)

For any geometric random variable  $X$  with parameter  $p$ ,  $E[X] = p^{-1}$  and  $\text{Var}[X] = \frac{1-p}{p^2}$ .

### Proof (cont'd).

RC (2019/20) – Lecture 5 – slide 12

## Bounding Coupon Collector purchases - Chebyshev

### Lemma (3.8)

For any geometric random variable  $X$  with parameter  $p$ ,  $E[X] = p^{-1}$  and  $\text{Var}[X] = \frac{1-p}{p^2}$ .

### Proof (cont'd).

$$\begin{aligned} E[X^2] &= p + (1-p)E[(X+1)^2] \\ &= p + (1-p)E[X^2] + 2(1-p)E[X] + 1 - p \\ &= 1 + \frac{2(1-p)}{p} + (1-p)E[X^2]. \end{aligned}$$

RC (2019/20) – Lecture 5 – slide 12



## Bounding Coupon Collector purchases - Chebyshev

### Lemma (3.8)

For any geometric random variable  $X$  with parameter  $p$ ,  $E[X] = p^{-1}$  and  $\text{Var}[X] = \frac{1-p}{p^2}$ .

Proof (cont'd).

$$\begin{aligned} E[X^2] &= p + (1-p)E[(X+1)^2] \\ &= p + (1-p)E[X^2] + 2(1-p)E[X] + 1-p \\ &= 1 + \frac{2(1-p)}{p} + (1-p)E[X^2]. \end{aligned}$$

We can solve that  $E[X^2] = \frac{2-p}{p^2}$ . Thus,

$$\text{Var}[X] = E[X^2] - E[X]^2 = \frac{1-p}{p^2}. \quad \square$$

RC (2019/20) – Lecture 5 – slide 12

## Bounding Coupon Collector purchases - Chebyshev

$$\Pr[|X - E[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2} = \frac{\sum_{j=1}^n \text{Var}[X_j]}{a^2}.$$

RC (2019/20) – Lecture 5 – slide 13

## Bounding Coupon Collector purchases - Chebyshev

$$\Pr[|X - E[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2} = \frac{\sum_{j=1}^n \text{Var}[X_j]}{a^2}.$$

Each individual  $X_j$  is geometric with parameter  $\frac{n-(j-1)}{n}$ , so each  $X_j$  has

$$\text{Var}[X_j] = \frac{j-1}{n} \left( \frac{n}{n+1-j} \right)^2 \leq \left( \frac{n}{n+1-j} \right)^2.$$

RC (2019/20) – Lecture 5 – slide 13

## Bounding Coupon Collector purchases - Chebyshev

$$\Pr[|X - E[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2} = \frac{\sum_{j=1}^n \text{Var}[X_j]}{a^2}.$$

Each individual  $X_j$  is geometric with parameter  $\frac{n-(j-1)}{n}$ , so each  $X_j$  has

$$\text{Var}[X_j] = \frac{j-1}{n} \left( \frac{n}{n+1-j} \right)^2 \leq \left( \frac{n}{n+1-j} \right)^2.$$

Hence, using the Euler's series for  $\frac{\pi^2}{6}$ ,

$$\text{Var}[X] \leq n^2 \sum_{j=1}^n \left( \frac{1}{n+1-j} \right)^2 = n^2 \sum_{j=n}^1 \left( \frac{1}{j} \right)^2 \leq \frac{\pi^2 n^2}{6}.$$

RC (2019/20) – Lecture 5 – slide 13

## Bounding Coupon Collector purchases - Chebyshev

We know  $\text{Var}[X] \leq \frac{\pi^2 n^2}{6}$  for our coupon collector process.

## Bounding Coupon Collector purchases - Chebyshev

We know  $\text{Var}[X] \leq \frac{\pi^2 n^2}{6}$  for our coupon collector process.

Suppose we are willing to make  $2E[X]$  (about  $2n \ln(n)$ ) purchases.

The probability we fail to get all cards is

$$\begin{aligned}\Pr[X > 2E[X]] &= \Pr[X - E[X] > E[X]] \\ &= \Pr[|X - E[X]| > E[X]].\end{aligned}\quad (\text{as } X \geq 0)$$

RC (2019/20) – Lecture 5 – slide 14

RC (2019/20) – Lecture 5 – slide 14

## Bounding Coupon Collector purchases - Chebyshev

We know  $\text{Var}[X] \leq \frac{\pi^2 n^2}{6}$  for our coupon collector process.

Suppose we are willing to make  $2E[X]$  (about  $2n \ln(n)$ ) purchases.

The probability we fail to get all cards is

$$\begin{aligned}\Pr[X > 2E[X]] &= \Pr[X - E[X] > E[X]] \\ &= \Pr[|X - E[X]| > E[X]].\end{aligned}\quad (\text{as } X \geq 0)$$

We can upper bound the probability of the bad event  $|X - E[X]| > E[X]$  using Chebyshev Inequality with  $a = E[X]$ :

$$\begin{aligned}\Pr[|X - E[X]| \geq E[X]] &\leq \frac{\text{Var}[X]}{E[X]^2} \leq \frac{\pi^2 n^2}{6n^2 H(n)^2} \\ &= \frac{\pi^2}{6H(n)^2} \leq \frac{2}{\ln(n)^2}.\end{aligned}$$

This improves over  $\frac{1}{2}$ , which is what Markov gives us.

RC (2019/20) – Lecture 5 – slide 14

## Bounding Coupon Collector purchases - Union bound

### Theorem (1.2, Union bound)

Let  $E_1, E_2, \dots$  be a finite or countably infinite sequence of events,

$$\Pr\left[\bigcup_{i \geq 1} E_i\right] \leq \sum_{i \geq 1} \Pr[E_i].$$

Similar to Markov ineq., there is almost no requirement to the union bound!

RC (2019/20) – Lecture 5 – slide 15

## Bounding Coupon Collector purchases - Union bound

Let  $E_i$  be the “bad” event where card  $i$  is still missing at time  $T$ .

$$\Pr[E_i] \leq \left(1 - \frac{1}{n}\right)^T.$$

Thus, by a union bound,

$$\Pr[X \geq T] = \Pr\left[\bigcup_{i \geq 1} E_i\right] \leq n \left(1 - \frac{1}{n}\right)^T.$$

RC (2019/20) – Lecture 5 – slide 16

## Bounding Coupon Collector purchases - Union bound

Once again we use  $(1 - 1/n)^n \leq 1/e$ . If  $T = (1 + \varepsilon)n \ln n$ ,

$$\begin{aligned} n \left(1 - \frac{1}{n}\right)^T &\leq n \left(\left(1 - \frac{1}{n}\right)^n\right)^{(1+\varepsilon) \ln n} \\ &\leq n(e^{-1})^{(1+\varepsilon) \ln n} = n^{-\varepsilon}. \end{aligned}$$

Thus, for example if  $\varepsilon = 1$ ,

$$\Pr[X \geq 2n \ln n] \leq n^{-1}.$$

As  $E[X] \geq n \ln n$ ,

$$\Pr[X \geq 2E[X]] \leq \Pr[X \geq 2n \ln n] \leq n^{-1}.$$

RC (2019/20) – Lecture 5 – slide 17

## Coupon collector bounds

$$\Pr[X \geq 2E[X]] \leq \frac{1}{2} \quad (\text{Markov})$$

$$\Pr[X \geq 2E[X]] \leq \frac{2}{\ln(n)^2} \quad (\text{Chebyshev})$$

$$\Pr[X \geq 2E[X]] \leq \frac{1}{n} \quad (\text{Union bound})$$

The stronger the bounds are, the more information we use.

RC (2019/20) – Lecture 5 – slide 18

## Coupon collector bounds

$$\Pr[X \geq 2E[X]] \leq \frac{1}{2} \quad (\text{Markov})$$

$$\Pr[X \geq 2E[X]] \leq \frac{2}{\ln(n)^2} \quad (\text{Chebyshev})$$

$$\Pr[X \geq 2E[X]] \leq \frac{1}{n} \quad (\text{Union bound})$$

The stronger the bounds are, the more information we use.

Chebyshev also gives (weak) lower bound. Using Chernoff bound for negatively correlated rv, one can show

$$\Pr[X \leq (1 - \varepsilon)(n - 1) \ln n] \leq e^{-n^\varepsilon}.$$

However this is beyond the scope of our course. Check out Chapter 9 and 10 of <https://arxiv.org/abs/1801.06733> if you are interested.

RC (2019/20) – Lecture 5 – slide 18

## Wrapping up today

Next week we will continue the theme of “bounding deviation from the mean” by introducing some stronger concentration inequalities called Chernoff bounds/Hoeffding ineq.

First, on Friday (to give a break) we will look at a simple random algorithm to approximately calculate **Max** Cut, and show how to *derandomize* it.

- ▶ Coursework 1 will be available on Thursday.
- ▶ Tutorials are starting next week. The first tutorial sheet will also be available soon.