CMPSCI 711: "Really Advanced Algorithms"

Lecture 2 – Markov, Chebyshev, & Balls and Bins

Andrew McGregor

Last Compiled: February 4, 2009

Outline

Probability and Random Variables

Markov and Chebyshev

Balls and Bins (and Birthdays and Coupons!)

Puzzle

Probability

- Inclusion-Exclusion: For arbitrary events $A_1, A_2, \ldots, A_n$,

  $$\mathbb{P}\left[\cup_{i=1}^{n} A_i\right] = \sum_{i=1}^{n} \mathbb{P}\left[A_i\right] - \sum_{i<j}^{n} \mathbb{P}\left[A_i \cap A_j\right] + \sum_{i<j<k}^{n} \mathbb{P}\left[A_i \cap A_j \cap A_k\right] - \ldots$$

  Truncating yields upper (or lower) bound if the last term is positive (or negative). Union bound, $\mathbb{P}\left[\cup_{i=1}^{n} A_i\right] \leq \sum_{i=1}^{n} \mathbb{P}\left[A_i\right]$

Probability

- Inclusion-Exclusion: For arbitrary events $A_1, A_2, \ldots, A_n$,

$$\mathbb{P}\left[\cup_{i=1}^{n} A_i\right] = \sum_{i=1}^{n} \mathbb{P}\left[A_i\right] - \sum_{i<j}^{n} \mathbb{P}\left[A_i \cap A_j\right] + \sum_{i<j<k}^{n} \mathbb{P}\left[A_i \cap A_j \cap A_k\right] - \ldots$$

  Truncating yields upper (or lower) bound if the last term is positive (or negative). Union bound, $\mathbb{P}\left[\cup_{i=1}^{n} A_i\right] \leq \sum_{i=1}^{n} \mathbb{P}\left[A_i\right]$
- Conditional Probability: For arbitrary events A and B,

$$\mathbb{P}\left[A|B\right] = \mathbb{P}\left[A \cap B\right] / \mathbb{P}\left[B\right]$$

  and $\Pr(\cap_{i=1}^{n} A_i) = \Pr(A_1)\Pr(A_2|A_1)\ldots\Pr(A_n| \cap_{i=1}^{n-1} A_i)$

Probability

- Inclusion-Exclusion: For arbitrary events $A_1, A_2, \ldots, A_n$,

  $$\mathbb{P}\left[\cup_{i=1}^n A_i\right] = \sum_{i=1}^n \mathbb{P}\left[A_i\right] - \sum_{i<j}^n \mathbb{P}\left[A_i \cap A_j\right] + \sum_{i<j<k}^n \mathbb{P}\left[A_i \cap A_j \cap A_k\right] - \ldots$$

  Truncating yields upper (or lower) bound if the last term is positive (or negative). Union bound, $\mathbb{P}\left[\cup_{i=1}^n A_i\right] \leq \sum_{i=1}^n \mathbb{P}\left[A_i\right]$
- Conditional Probability: For arbitrary events A and B,

  $$\mathbb{P}\left[A|B\right] = \mathbb{P}\left[A \cap B\right] / \mathbb{P}\left[B\right]$$

  and $\Pr(\cap_{i=1}^n A_i) = \Pr(A_1)\Pr(A_2|A_1)\ldots\Pr(A_n| \cap_{i=1}^{n-1} A_i)$
- Independence: A and B are independent is $\mathbb{P}\left[A|B\right] = \mathbb{P}\left[A\right]$ (or equivalently $\mathbb{P}\left[A \cap B\right] = \mathbb{P}\left[A\right]\mathbb{P}\left[B\right]$.)

Random Variables

- Expectation: $\mathbb{E}\left[X\right] = \sum_r r\mathbb{P}\left[X = r\right]$
- Variance: $\mathbb{V}\left[X\right] = \mathbb{E}\left[(X - \mathbb{E}\left[X\right])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2$
- Standard deviation: $\sigma_X = \sqrt{\mathbb{V}\left[X\right]}$

Random Variables

- Expectation: $\mathbb{E}[X] = \sum_r r\mathbb{P}[X = r]$
- Variance: $\mathbb{V}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$
- Standard deviation: $\sigma_X = \sqrt{\mathbb{V}[X]}$

Theorem

- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ *if X and Y independent.*
- $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y]$ *if X and Y independent.*

Moment Generating Functions

Let X be a non-negative integer-valued random variable. The *probability generating function* of X is

$$G_X(z) = \sum_{i=0}^{\infty} z^i \mathbb{P}[X = i]$$

Moment Generating Functions

Let X be a non-negative integer-valued random variable. The *probability generating function* of X is

$$G_X(z) = \sum_{i=0}^{\infty} z^i \mathbb{P}\left[X = i\right]$$

Lemma

- $\mathbb{E}\left[X\right] = G'(1)$.
- $\mathbb{V}\left[X\right] = G'' + G'(1) - G'(1)^2$.

Examples of Random Variables

Example

Let X have the binomial distribution $Bin(n, p)$:

$$\mathbb{P}[X = i] = \binom{n}{i} p^i (1-p)^{n-i}$$

"How many heads do we see when we toss a coin with probability p of heads n times?" This distribution has generating function $G(z) = (1 - p + pz)^n$. $\mathbb{E}[X] = np$ and $\mathbb{V}[X] = np(1-p)$.

Examples of Random Variables

Example

Let X have the binomial distribution $Bin(n, p)$:

$$\mathbb{P}[X = i] = \binom{n}{i} p^i (1-p)^{n-i}$$

"How many heads do we see when we toss a coin with probability p of heads n times?" This distribution has generating function $G(z) = (1 - p + pz)^n$. $\mathbb{E}[X] = np$ and $\mathbb{V}[X] = np(1-p)$.

Example

Let X have the binomial distribution $Geom(p)$:

$$\mathbb{P}[X = i] = (1-p)^{i-1} p$$

"How many times do we toss a coin with probability p of heads until we see a heads." This distribution has generating function $G(z) = pz/(1 - z + pz)$. $\mathbb{E}[X] = 1/p$, $\mathbb{V}[X] = (1-p)/p^2$.

Outline

Probability and Random Variables

Markov and Chebyshev

Balls and Bins (and Birthdays and Coupons!)

Puzzle

Markov Inequality

Theorem (Markov)

Let Y *be a random variable assuming only non-negative values. Then, for all* $t > 0$, $\mathbb{P}\left[Y \geq t\mathbb{E}\left[Y\right]\right] \leq 1/t$.

Markov Inequality

Theorem (Markov)

Let Y be a random variable assuming only non-negative values. Then, for all $t > 0$, $\mathbb{P}\left[Y \geq t\mathbb{E}\left[Y\right]\right] \leq 1/t$.

Proof.

- Define $f(y) = 1$ if $y \geq t\mathbb{E}\left[Y\right]$ and 0 otherwise.

$\square$

Markov Inequality

Theorem (Markov)

Let Y be a random variable assuming only non-negative values. Then, for all $t > 0$, $\mathbb{P}\left[Y \geq t\mathbb{E}\left[Y\right]\right] \leq 1/t$.

Proof.

- Define $f(y) = 1$ if $y \geq t\mathbb{E}\left[Y\right]$ and 0 otherwise.
- Note that $f(y) \leq y/(t\mathbb{E}\left[Y\right])$.

□

Markov Inequality

Theorem (Markov)

Let Y be a random variable assuming only non-negative values. Then, for all $t > 0$, $\mathbb{P}\left[Y \geq t\mathbb{E}\left[Y\right]\right] \leq 1/t$.

Proof.

- Define $f(y) = 1$ if $y \geq t\mathbb{E}\left[Y\right]$ and 0 otherwise.
- Note that $f(y) \leq y/(t\mathbb{E}\left[Y\right])$.
- $\mathbb{P}\left[Y \geq t\mathbb{E}\left[Y\right]\right] = \mathbb{E}\left[f(Y)\right]$

□

Markov Inequality

Theorem (Markov)

Let Y be a random variable assuming only non-negative values. Then, for all $t > 0$, $\mathbb{P}[Y \geq t\mathbb{E}[Y]] \leq 1/t$.

Proof.

- Define $f(y) = 1$ if $y \geq t\mathbb{E}[Y]$ and 0 otherwise.
- Note that $f(y) \leq y/(t\mathbb{E}[Y])$.
- $\mathbb{P}[Y \geq t\mathbb{E}[Y]] = \mathbb{E}[f(Y)]$
- Then, $\mathbb{E}[f(Y)] \leq \mathbb{E}[Y/(t\mathbb{E}[Y])] = 1/t$

☐

Markov Inequality: Questions and Extensions

- Is the inequality ever tight, i.e., $\mathbb{P}[Y \geq t\mathbb{E}[Y]] = 1/t$?

Markov Inequality: Questions and Extensions

- Is the inequality ever tight, i.e., $\mathbb{P}[Y \geq t\mathbb{E}[Y]] = 1/t$?
  - E.g., Y is t with probability $1/t$ and 0 otherwise.

Markov Inequality: Questions and Extensions

- Is the inequality ever tight, i.e., $\mathbb{P}\left[Y \geq t\mathbb{E}\left[Y\right]\right] = 1/t$?
    - E.g., Y is t with probability $1/t$ and 0 otherwise.
- Can we bound $\mathbb{P}\left[Y \geq t\mathbb{E}\left[Y\right]\right]$ below for $t > 1$?

Markov Inequality: Questions and Extensions

- Is the inequality ever tight, i.e., $\mathbb{P}\left[Y \geq t\mathbb{E}\left[Y\right]\right] = 1/t$?
  - E.g., Y is t with probability $1/t$ and 0 otherwise.
- Can we bound $\mathbb{P}\left[Y \geq t\mathbb{E}\left[Y\right]\right]$ below for $t > 1$?
  - E.g., consider a constant random variable $Y = 1$

Markov Inequality: Questions and Extensions

- Is the inequality ever tight, i.e., $\mathbb{P}\left[Y \geq t\mathbb{E}\left[Y\right]\right] = 1/t$?
  - E.g., Y is t with probability $1/t$ and 0 otherwise.
- Can we bound $\mathbb{P}\left[Y \geq t\mathbb{E}\left[Y\right]\right]$ below for $t > 1$?
  - E.g., consider a constant random variable $Y = 1$
- Can we bound $\mathbb{P}\left[Y \leq t\mathbb{E}\left[Y\right]\right]$ for $0 < t < 1$?

Markov Inequality: Questions and Extensions

- Is the inequality ever tight, i.e., $\mathbb{P}\left[Y \geq t\mathbb{E}\left[Y\right]\right] = 1/t$?
  - E.g., Y is t with probability $1/t$ and 0 otherwise.
- Can we bound $\mathbb{P}\left[Y \geq t\mathbb{E}\left[Y\right]\right]$ below for $t > 1$?
  - E.g., consider a constant random variable $Y = 1$
- Can we bound $\mathbb{P}\left[Y \leq t\mathbb{E}\left[Y\right]\right]$ for $0 < t < 1$?
  - If $Y \leq m$, consider the random variable $X = m - Y$.

Chebyshev Inequality

Theorem (Chebyshev)

Let X *be a random variable with expectation* μ_X *and standard deviation* σ_X*. Then for* $t > 0$, $\mathbb{P}\left[|X - \mu_X| \geq t\sigma_X\right] \leq 1/t^2$.

Chebyshev Inequality

Theorem (Chebyshev)

Let X be a random variable with expectation μ_X and standard deviation σ_X. Then for $t > 0$, $\mathbb{P}\left[|X - \mu_X| \geq t\sigma_X\right] \leq 1/t^2$.

Proof.

- Note that $\mathbb{P}\left[|X - \mu_X| \geq t\sigma_X\right] = \mathbb{P}\left[(X - \mu_X)^2 \geq t^2\sigma_X^2\right]$

$\square$

Chebyshev Inequality

Theorem (Chebyshev)

Let X be a random variable with expectation μ_X and standard deviation σ_X. Then for $t > 0$, $\mathbb{P}\left[|X - \mu_X| \geq t\sigma_X\right] \leq 1/t^2$.

Proof.

- Note that $\mathbb{P}\left[|X - \mu_X| \geq t\sigma_X\right] = \mathbb{P}\left[(X - \mu_X)^2 \geq t^2\sigma_X^2\right]$
- Let $Y = (X - \mu_X)^2$ and note $\mathbb{E}\left[Y\right] = \sigma_X^2$

□

Chebyshev Inequality

Theorem (Chebyshev)

Let X be a random variable with expectation μ_X and standard deviation σ_X. Then for $t > 0$, $\mathbb{P}[|X - \mu_X| \geq t\sigma_X] \leq 1/t^2$.

Proof.

- Note that $\mathbb{P}[|X - \mu_X| \geq t\sigma_X] = \mathbb{P}\left[(X - \mu_X)^2 \geq t^2\sigma_X^2\right]$
- Let $Y = (X - \mu_X)^2$ and note $\mathbb{E}[Y] = \sigma_X^2$
- Use Markov's inequality to show $\mathbb{P}\left[Y \geq t^2\mathbb{E}[Y]\right] \leq 1/t^2$

□

Chebyshev Inequality: Questions and Extensions

Theorem

Let $X_1, \ldots, X_n$ *be i.i.d. (independent, identically distributed, random variables) with* $\mathbb{E}[X_i] = \mu$ *and* $\sigma_{X_i} = \sigma$. *Let* $Y = n^{-1} \sum_{1 \le i \le n} X_i$. *Then,*

$$\mathbb{P}[|Y - \mu_Y| \ge t] \le \sigma_Y^2 / t^2 = \sigma^2 / (t^2 n)$$

Chebyshev Inequality: Questions and Extensions

Theorem

Let $X_1, \ldots, X_n$ *be i.i.d. (independent, identically distributed, random variables) with* $\mathbb{E}[X_i] = \mu$ *and* $\sigma_{X_i} = \sigma$*. Let* $Y = n^{-1} \sum_{1 \le i \le n} X_i$*. Then,*

$$\mathbb{P}[|Y - \mu_Y| \ge t] \le \sigma_Y^2 / t^2 = \sigma^2 / (t^2 n)$$

Proof.

- Linearity of expectation implies $\mu_Y = \mu$.
- Linearity of variance implies $\sigma_Y^2 = \sigma^2 / n$.

□

Chebyshev Inequality: Questions and Extensions

Theorem

Let $X_1, \ldots, X_n$ be i.i.d. (independent, identically distributed, random variables) with $\mathbb{E}[X_i] = \mu$ and $\sigma_{X_i} = \sigma$. Let $Y = n^{-1} \sum_{1 \leq i \leq n} X_i$. Then,

$$\mathbb{P}[|Y - \mu_Y| \geq t] \leq \sigma_Y^2 / t^2 = \sigma^2 / (t^2 n)$$

Proof.

- Linearity of expectation implies $\mu_Y = \mu$.
- Linearity of variance implies $\sigma_Y^2 = \sigma^2 / n$.

□

Example

Let $X \sim Bin(n, p)$. Using Chebyshev we deduce,

$$\mathbb{P}[|X - \mu_X| \geq t] \leq (np(1-p))/t^2.$$

Outline

Probability and Random Variables

Markov and Chebyshev

Balls and Bins (and Birthdays and Coupons!)

Puzzle

Balls and Bins

Throw m balls into n bins where each throw is independent. Many questions:

Balls and Bins

Throw m balls into n bins where each throw is independent. Many questions:

- The maximum number of balls that fall into the same bin?

Balls and Bins

Throw m balls into n bins where each throw is independent. Many questions:

- The maximum number of balls that fall into the same bin?
- How large must m be such that there exists a bin with at least two balls? (Birthday Paradox)

Balls and Bins

Throw m balls into n bins where each throw is independent. Many questions:

- The maximum number of balls that fall into the same bin?
- How large must m be such that there exists a bin with at least two balls? (Birthday Paradox)
- How large must m be such that all bins get at least one ball? (Coupon Collecting)

Heaviest Bin (1/2)

Assume $m = n$. Let Y_i be number of balls that fall in i-th bin.

Heaviest Bin (1/2)

Assume $m = n$. Let Y_i be number of balls that fall in i-th bin.

Lemma

Let $k \geq (3 \ln n)/\ln \ln n$. *Then* $\mathbb{P}\left[Y_i \geq k\right] \leq n^{-2}$

Heaviest Bin (1/2)

Assume $m = n$. Let Y_i be number of balls that fall in i-th bin.

Lemma

Let $k \geq (3 \ln n)/\ln \ln n$. *Then* $\mathbb{P}[Y_i \geq k] \leq n^{-2}$

Proof.

- $\mathbb{P}[Y_i = j] = \binom{n}{j}(1/n)^j(1 - 1/n)^{n-j}$

$\square$

Heaviest Bin (1/2)

Assume $m = n$. Let Y_i be number of balls that fall in i-th bin.

Lemma

Let $k \geq (3 \ln n)/\ln\ln n$. *Then* $\mathbb{P}[Y_i \geq k] \leq n^{-2}$

Proof.

- $\mathbb{P}[Y_i = j] = \binom{n}{j}(1/n)^j(1-1/n)^{n-j}$
- Using the bound $\binom{n}{j} \leq (ne/j)^j$:

$$\mathbb{P}[Y_i = j] = \binom{n}{j}(1/n)^j(1-1/n)^{n-j} \leq (e/j)^j$$

□

Heaviest Bin (1/2)

Assume $m = n$. Let Y_i be number of balls that fall in i-th bin.

Lemma

Let $k \geq (3 \ln n)/\ln \ln n$*. Then* $\mathbb{P}\left[Y_i \geq k\right] \leq n^{-2}$

Proof.

- $\mathbb{P}\left[Y_i = j\right] = \binom{n}{j}(1/n)^j(1-1/n)^{n-j}$
- Using the bound $\binom{n}{j} \leq (ne/j)^j$:

$$\mathbb{P}\left[Y_i = j\right] = \binom{n}{j}(1/n)^j(1-1/n)^{n-j} \leq (e/j)^j$$

- By summing up a geometric series:

$$\mathbb{P}\left[Y_i \geq k\right] = \sum_{j \geq k}(e/j)^j \leq (e/k)^k \frac{1}{1-e/k}$$

☐

Heaviest Bin (2/2)

Assume $m = n$. Let Y_i be number of balls that fall in i-th bin.

Lemma

Let $k \geq (3 \ln n) / \ln \ln n$. Then $\mathbb{P}[Y_i \geq k] \leq n^{-2}$

Heaviest Bin (2/2)

Assume $m = n$. Let Y_i be number of balls that fall in i-th bin.

Lemma

Let $k \geq (3 \ln n)/\ln \ln n$. *Then* $\mathbb{P}[Y_i \geq k] \leq n^{-2}$

Theorem

$\mathbb{P}[Y_i < k \text{ for all } i] \geq 1 - 1/n$.

Heaviest Bin (2/2)

Assume $m = n$. Let Y_i be number of balls that fall in i-th bin.

Lemma

Let $k \geq (3 \ln n)/\ln \ln n$. *Then* $\mathbb{P}[Y_i \geq k] \leq n^{-2}$

Theorem

$\mathbb{P}[Y_i < k \text{ for all } i] \geq 1 - 1/n.$

Proof.

Use union bound:

$$\mathbb{P}[Y_i \geq k \text{ for some } i] \leq \sum_i \mathbb{P}[Y_i \geq k] \leq 1/n$$

□

Birthday Paradox

Lemma

$\mathbb{P}[\text{first } m \text{ balls fall in distinct bins}] \leq e^{-m(m-1)/(2n)}$.

Birthday Paradox

Lemma

$\mathbb{P}\left[\text{first } m \text{ balls fall in distinct bins}\right] \leq e^{-m(m-1)/(2n)}$.

Proof.

- Let A_i be event that the i-th ball lands in a bin not containing any of the first $i-1$ balls.

□

Birthday Paradox

Lemma

$\mathbb{P}[\text{first } m \text{ balls fall in distinct bins}] \leq e^{-m(m-1)/(2n)}$.

Proof.

- Let A_i be event that the i-th ball lands in a bin not containing any of the first $i-1$ balls.
- $\mathbb{P}[\cap_{1\leq i\leq m} A_i] = \mathbb{P}[A_1]\,\mathbb{P}[A_2 | A_1] \ldots \mathbb{P}[A_m | \cap_{1\leq i\leq m-1} A_i]$

☐

Birthday Paradox

Lemma

$\mathbb{P}\left[\text{first } m \text{ balls fall in distinct bins}\right] \leq e^{-m(m-1)/(2n)}$.

Proof.

- Let A_i be event that the i-th ball lands in a bin not containing any of the first $i-1$ balls.
- $\mathbb{P}\left[\cap_{1 \leq i \leq m} A_i\right] = \mathbb{P}\left[A_1\right] \mathbb{P}\left[A_2 | A_1\right] \ldots \mathbb{P}\left[A_m | \cap_{1 \leq i \leq m-1} A_i\right]$
- $\mathbb{P}\left[A_i | \cap_{1 \leq i \leq i-1} A_i\right] = 1 - (i-1)/n$

□

Birthday Paradox

Lemma

$\mathbb{P}[\text{first } m \text{ balls fall in distinct bins}] \leq e^{-m(m-1)/(2n)}$.

Proof.

- Let A_i be event that the i-th ball lands in a bin not containing any of the first $i-1$ balls.
- $\mathbb{P}[\cap_{1\leq i\leq m} A_i] = \mathbb{P}[A_1]\,\mathbb{P}[A_2|A_1]\ldots\mathbb{P}[A_m|\cap_{1\leq i\leq m-1} A_i]$
- $\mathbb{P}[A_i|\cap_{1\leq i\leq i-1} A_i] = 1-(i-1)/n$
- Putting it together and using $\sum_{1\leq i\leq a} i = (a+1)a/2$:

$$\mathbb{P}[\cap_{1\leq i\leq m} A_i] = \prod_{1\leq i\leq m}\left(1-\frac{i-1}{n}\right) \leq e^{-m(m-1)/(2n)}$$

□

Birthday Paradox

Lemma

$\mathbb{P}\left[\text{first } m \text{ balls fall in distinct bins}\right] \le e^{-m(m-1)/(2n)}$.

Proof.

- Let A_i be event that the i-th ball lands in a bin not containing any of the first $i-1$ balls.
- $\mathbb{P}\left[\cap_{1 \le i \le m} A_i\right] = \mathbb{P}\left[A_1\right] \mathbb{P}\left[A_2 | A_1\right] \dots \mathbb{P}\left[A_m | \cap_{1 \le i \le m-1} A_i\right]$
- $\mathbb{P}\left[A_i | \cap_{1 \le i \le i-1} A_i\right] = 1 - (i-1)/n$
- Putting it together and using $\sum_{1 \le i \le a} i = (a+1)a/2$:

$$\mathbb{P}\left[\cap_{1 \le i \le m} A_i\right] = \prod_{1 \le i \le m} \left(1 - \frac{i-1}{n}\right) \le e^{-m(m-1)/(2n)}$$

□

With $n = 365$ and $m = 29$, probability $< e^{-1}$. Tighter analysis is possible.

Coupon Collecting (1/2)

Let Z_i be the throw in which exactly i bins become non-empty. Let $X_i = Z_{i+1} - Z_i$. Note that $Z_n = \sum_{0 \le i \le n-1} X_i$

Coupon Collecting (1/2)

Let Z_i be the throw in which exactly i bins become non-empty. Let $X_i = Z_{i+1} - Z_i$. Note that $Z_n = \sum_{0 \le i \le n-1} X_i$

Lemma

$\mathbb{E}[Z_n] = nH_n$ *where* $H_n = 1 + 1/2 + \ldots + 1/n = \ln n + \Theta(n)$.

Coupon Collecting (1/2)

Let Z_i be the throw in which exactly i bins become non-empty. Let $X_i = Z_{i+1} - Z_i$. Note that $Z_n = \sum_{0 \le i \le n-1} X_i$

Lemma

$\mathbb{E}[Z_n] = nH_n$ where $H_n = 1 + 1/2 + \ldots + 1/n = \ln n + \Theta(n)$.

Proof.

- X_i has a geometric distribution:

  $$\mathbb{P}[X_i = j] = p_i (1 - p_i)^{j-1}$$

  where $p_i = 1 - i/n$.

□

Coupon Collecting (1/2)

Let Z_i be the throw in which exactly i bins become non-empty. Let $X_i = Z_{i+1} - Z_i$. Note that $Z_n = \sum_{0 \leq i \leq n-1} X_i$

Lemma

$\mathbb{E}\left[Z_n\right] = nH_n$ *where* $H_n = 1 + 1/2 + \ldots + 1/n = \ln n + \Theta(n)$.

Proof.

- X_i has a geometric distribution:

$$\mathbb{P}\left[X_i = j\right] = p_i \left(1 - p_i\right)^{j-1}$$

  where $p_i = 1 - i/n$.
- $\mathbb{E}\left[X_i\right] = 1/p_i$.

□

Coupon Collecting (1/2)

Let Z_i be the throw in which exactly i bins become non-empty. Let $X_i = Z_{i+1} - Z_i$. Note that $Z_n = \sum_{0 \leq i \leq n-1} X_i$

Lemma

$\mathbb{E}\left[Z_n\right] = nH_n$ where $H_n = 1 + 1/2 + \ldots + 1/n = \ln n + \Theta(n)$.

Proof.

- X_i has a geometric distribution:

$$\mathbb{P}\left[X_i = j\right] = p_i \left(1 - p_i\right)^{j-1}$$

  where $p_i = 1 - i/n$.
- $\mathbb{E}\left[X_i\right] = 1/p_i$.
- $\mathbb{E}\left[Z_n\right] = \sum_{0 \leq i \leq n-1} \mathbb{E}\left[X_i\right] = n/n + n/(n-1) + \ldots + n/1$

□

Coupon Collecting (2/2)

Let Z_i be the throw in which exactly i bins become non-empty. Let $X_i = Z_{i+1} - Z_i$. Note that $Z_n = \sum_{0 \leq i \leq n-1} X_i$

Lemma

$\mathbb{V}\left[Z_n\right] = n^2(\pi^2/6 + o(1)) - nH_n$.

Coupon Collecting (2/2)

Let Z_i be the throw in which exactly i bins become non-empty. Let $X_i = Z_{i+1} - Z_i$. Note that $Z_n = \sum_{0 \leq i \leq n-1} X_i$

Lemma

$\mathbb{V}[Z_n] = n^2(\pi^2/6 + o(1)) - nH_n$.

Proof.

- X_i has a geometric distribution: $\mathbb{V}[X_i] = (1 - p_i)/p_i^2$.

□

Coupon Collecting (2/2)

Let Z_i be the throw in which exactly i bins become non-empty. Let $X_i = Z_{i+1} - Z_i$. Note that $Z_n = \sum_{0 \leq i \leq n-1} X_i$

Lemma

$\mathbb{V}\left[Z_n\right] = n^2(\pi^2/6 + o(1)) - nH_n$.

Proof.

- X_i has a geometric distribution: $\mathbb{V}\left[X_i\right] = (1 - p_i)/p_i^2$.
- The X_i are independent: $\mathbb{V}\left[Z_n\right] = \sum_{0 \leq i \leq n-1} \mathbb{V}\left[X_i\right]$

□

Coupon Collecting (2/2)

Let Z_i be the throw in which exactly i bins become non-empty. Let $X_i = Z_{i+1} - Z_i$. Note that $Z_n = \sum_{0 \le i \le n-1} X_i$

Lemma

$\mathbb{V}[Z_n] = n^2(\pi^2/6 + o(1)) - nH_n$.

Proof.

- X_i has a geometric distribution: $\mathbb{V}[X_i] = (1 - p_i)/p_i^2$.
- The X_i are independent: $\mathbb{V}[Z_n] = \sum_{0 \le i \le n-1} \mathbb{V}[X_i]$
- Therefore, $\mathbb{V}[Z_n]$ equals

$$\sum_{0 \le i \le n-1} \frac{1 - p_i}{p_i^2} = \sum_{0 \le i \le n-1} \frac{ni}{(n-i)^2} = n^2 \sum_{1 \le i \le n} \frac{1}{i^2} - nH_n$$

□

Coupon Collecting (2/2)

Let Z_i be the throw in which exactly i bins become non-empty. Let $X_i = Z_{i+1} - Z_i$. Note that $Z_n = \sum_{0 \le i \le n-1} X_i$

Lemma

$\mathbb{V}[Z_n] = n^2(\pi^2/6 + o(1)) - nH_n$.

Proof.

- X_i has a geometric distribution: $\mathbb{V}[X_i] = (1 - p_i)/p_i^2$.
- The X_i are independent: $\mathbb{V}[Z_n] = \sum_{0 \le i \le n-1} \mathbb{V}[X_i]$
- Therefore, $\mathbb{V}[Z_n]$ equals

$$\sum_{0 \le i \le n-1} \frac{1 - p_i}{p_i^2} = \sum_{0 \le i \le n-1} \frac{ni}{(n-i)^2} = n^2 \sum_{1 \le i \le n} \frac{1}{i^2} - nH_n$$

- Appeal to the fact $\lim_{n \to \infty} \left(\sum_{1 \le i \le n} i^{-2} \right) = \pi^2/6$.

□

Outline

Probability and Random Variables

Markov and Chebyshev

Balls and Bins (and Birthdays and Coupons!)

Puzzle

Clock Solitaire

- Take a standard pack of 52 cards which is randomly shuffled.
- Split into 13 piles of 4 and label piles {A,2,. . . ,10,J,Q,K}.
- Take first card from "K" pile.
- Take next card from "X" pile where X is the face value of the previous card taken.
- Repeat until either all cards are removed (you win) or we get stuck (you lose).

What's the probability you win?