Outline

Probability and Random Variables

Markov and Chebyshev

Balls and Bins (and Birthdays and Coupons!)

Puzzle
Probability

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

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

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

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

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

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

- **Conditional Probability**: For arbitrary events $A$ and $B$,

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

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

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

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

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

- **Conditional Probability:** For arbitrary events $A$ and $B$,

\[
P\left[ A \mid B \right] = \frac{P\left[ A \cap B \right]}{P\left[ B \right]}
\]

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

- **Independence:** $A$ and $B$ are independent is $P\left[ A \mid B \right] = P\left[ A \right]$ (or equivalently $P\left[ A \cap B \right] = P\left[ A \right] P\left[ B \right]$.)
Random Variables

- **Expectation:** $E[X] = \sum_r rP[X = r]$
- **Variance:** $V[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$
- **Standard deviation:** $\sigma_X = \sqrt{V[X]}$
Random Variables

- **Expectation:** \( \mathbb{E}[X] = \sum_r r \mathbb{P}[X = r] \)
- **Variance:** \( \mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \)
- **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

$$G_X(z) = \sum_{i=0}^{\infty} z^i \mathbb{P}[X = i]$$
Let $X$ be a non-negative integer-valued random variable. The\n*probability generating function* of $X$ is
\[G_X(z) = \sum_{i=0}^{\infty} z^i P[X = i].\]

**Lemma**

- $\mathbb{E}[X] = G'(1)$.
- $\mathbb{V}[X] = G'' + G'(1) - G'(1)^2$. 
Examples of Random Variables

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

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

“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 $t > 0$, $\mathbb{P} [Y \geq t\mathbb{E} [Y]] \leq 1/t$. 
Markov Inequality

Theorem (Markov)

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

Proof.

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

Theorem (Markov)

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

Proof.

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

Can we bound \( \mathbb{P}[Y \geq t\mathbb{E}[Y]] \) below for \( t > 1 \)?
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.
- Can we bound $\mathbb{P}[Y \geq t\mathbb{E}[Y]]$ below for $t > 1$?
  - E.g., consider a constant random variable $Y = 1$. 

If $Y \leq m$, consider the random variable $X = m - Y$. 

Markov Inequality: Questions and Extensions

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

If \( Y \leq m \), consider the random variable \( X = m - Y \).
Markov Inequality: Questions and Extensions

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

- Can we bound $P[Y \geq tE[Y]]$ below for $t > 1$?
  - E.g., consider a constant random variable $Y = 1$

- Can we bound $P[Y \leq tE[Y]]$ for $0 < t < 1$?
  - If $Y \leq m$, consider the random variable $X = m - Y$. 


Theorem (Chebyshev)

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

Theorem (Chebyshev)

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

Proof.

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

Theorem (Chebyshev)

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

Proof.

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

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

Proof.

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

Theorem

Let $X_1, \ldots, X_n$ be i.i.d. \textit{(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,

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

$Y = n^{-1} \sum_{1 \leq i \leq n} X_i$. 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 $\sigma_Y^2 = \sigma^2 / n$. 

\[ \square \]
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,

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

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,

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

▶ 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)
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)
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 $\mathbb{P} [Y_i \geq k] \leq n^{-2}$
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.

$\mathbb{P}[Y_i = j] = \binom{n}{j}(1/n)^j(1 - 1/n)^{n-j}$
Heaviest Bin (1/2)

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

Lemma

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

Proof.

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

Using the bound $\binom{n}{j} \leq (ne/j)^j$:

$\mathbb{P}[Y_i = j] = \binom{n}{j} \left(\frac{1}{n}\right)^j (1 - 1/n)^{n-j} \leq (e/j)^j$
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 $\mathbb{P}[Y_i \geq k] \leq n^{-2}$

**Proof.**

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

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

3. By summing up a geometric series:

   $$\mathbb{P}[Y_i \geq k] = \sum_{j \geq k} (e/j)^j \leq (e/k)^k \frac{1}{1 - e/k}$$
Assume \( m = n \). Let \( Y_i \) be number of balls that fall in \( i \)-th bin.

**Lemma**

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

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

Proof.
\[ P \left[ \bigcap_{1 \leq i \leq m} A_i \right] = P[A_1] \prod_{1 \leq i \leq m} P[A_i | \bigcap_{1 \leq i \leq m-1} A_i] = 1 - \left( \frac{i-1}{n} \right) \]

With \( n = 365 \) and \( m = 29 \), probability < \( e^{-1} \). Tighter analysis is possible.
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
\[ \Pr[ \text{first } m \text{ balls fall in distinct bins} ] \leq e^{-m(m-1)/(2n)}. \]

Proof.

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

- \[ \Pr[ \bigcap_{1 \leq i \leq m} A_i ] = \Pr[A_1] \Pr[A_2|A_1] \ldots \Pr[A_m|\bigcap_{1 \leq i \leq m-1} A_i] \]

With \( n = 365 \) and \( m = 29 \), probability < \( e^{-1} \). Tighter analysis is possible.
Birthday Paradox

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

Proof.

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

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

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

Proof.

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

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

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 $Z_n = \sum_{0 \leq i \leq n-1} X_i$.
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**

$\mathbb{E}[Z_n] = nH_n$ 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 $Z_n = \sum_{0 \leq i \leq n-1} X_i$

**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:

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

  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**

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

**Proof.**

- $X_i$ has a geometric distribution:
  \[ \mathbb{P}[X_i = j] = p_i (1 - p_i)^{j-1} \]
  where $p_i = 1 - i/n$.
- $\mathbb{E}[X_i] = 1/p_i$. 

\[ \square \]
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

$\mathbb{E}[Z_n] = nH_n$ 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 $p_i = 1 - i/n$.

- $\mathbb{E}[X_i] = 1/p_i$.

- $\mathbb{E}[Z_n] = \sum_{0 \leq i \leq n-1} \mathbb{E}[X_i] = n/n + n/(n-1) + \ldots + n/1$
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**

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

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

Proof.

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

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

Lemma

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

Proof.

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

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

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

\[\square\]
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,\ldots,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?