Today

Balls in Bins.
Random Variables.

Balls in bins
One throws \( m \) balls into \( n > m \) bins.

Theorem:
\[
\Pr[\text{no collision}] \approx \exp\left(-\frac{m^2}{2n}\right),
\]
for large enough \( n \).

In particular,
\[
\Pr[\text{no collision}] \approx \frac{1}{2}
\]
for
\[
m \approx \sqrt{2\ln(2)n} \approx 1.2\sqrt{n}.
\]

E.g., \( 1.2\sqrt{20} \approx 5.4 \).

Roughly,
\[
\Pr[\text{collision}] \approx \frac{1}{2}
\]
for \( m = \sqrt{n} \).

(\*\*) We used \( \ln(1-\epsilon) \approx -\epsilon \) for \( |\epsilon| \ll 1 \).

(\*\*) \( 1 + 2 + \cdots + m - 1 = (m-1)m/2 \).
**Approximation**

\[
\exp(-x) = 1 - x + \frac{1}{2!} x^2 + \cdots \approx 1 - x, \text{ for } |x| \ll 1.
\]

Hence, \(-x \approx \ln(1 - x)\) for \(|x| \ll 1\).

**Coupon Collector Problem.**

There are \(n\) different baseball cards.
(Brian Wilson, Jackie Robinson, Roger Hornsby, ...)

One random baseball card in each cereal box.

**Theorem:** If you buy \(m\) boxes,

(a) \(\Pr[\text{miss one specific item}] = e^{-m/n}\)
(b) \(\Pr[\text{miss any one of the items}] = ne^{-m/n}\).

**Today’s your birthday, it’s my birthday too.**

Probability that \(m\) people all have different birthdays?

With \(n = 365\), one finds

\(\Pr[\text{collision}] \approx 1/2\) if \(m \approx 1.2\sqrt{365} \approx 23\).

If \(m = 60\), we find that

\(\Pr[\text{no collision}] \approx \exp\left(-\frac{m^2}{2n}\right) = \exp\left(-\frac{60^2}{2 \times 365}\right) \approx 0.007\).

If \(m = 366\), then \(\Pr[\text{no collision}] = 0\). (No approximation here!)

**Coupon Collector Problem: Analysis.**

Event \(A_m = \{\text{fail to get Brian Wilson in } m \text{ cereal boxes}\}\)

Fail the first time: \((1 - \frac{1}{n})\)

Fail the second time: \((1 - \frac{1}{n})^2\)

And so on ... for \(m\) times. Hence,

\[
\Pr[A_m] = (1 - \frac{1}{n}) \times \cdots \times (1 - \frac{1}{n}) = \left(1 - \frac{1}{n}\right)^m \approx e^{-m/n}.
\]

\[
\ln(\Pr[A_m]) = \ln(1 - \frac{1}{n}) \approx m \times (-\frac{1}{n}) \approx -\frac{m}{n}.
\]

\[
\Pr[A_m] = \exp\left(-\frac{m}{n}\right).
\]

For \(p_m = \frac{1}{2}\), we need around \(n \ln 2 \approx 0.69n\) boxes.

**Checksums!**

Consider a set of \(m\) files.

Each file has a checksum of \(b\) bits.

How large should \(b\) be for \(\Pr[\text{share a checksum}] \leq 10^{-3}\)?

**Claim:** \(b \geq 2.9 \ln(m) + 9\).

**Proof:**

Let \(n = 2^b\) be the number of checksums.

We know \(\Pr[\text{no collision}] = \exp\left(-\frac{m^2}{2n}\right) \approx 1 - m^2/(2n)\).

Hence,

\[
\frac{m^2}{2n} \approx 10^{-3} \implies \frac{m^2}{2} \approx 10^{-3} \implies 2n = m^2 10^3 \implies 2^{b+1} = m^2 10^3 \implies 2^{b+1} = m^2 10^3 \implies 2^b + 1 \approx 10 + 2 \log_2(m) \approx 10 + 2.9 \ln(m).
\]

Note: \(\log_2(x) = \log_2(e) \ln(x) \approx 1.44 \ln(x)\).

**Collect all cards?**

Experiment: Choose \(m\) cards at random with replacement.

Events: \(E_k = \{\text{fail to get player } k\}, \text{ for } k = 1, \ldots, n\)

Probability of failing to get at least one of these \(n\) players:

\[
p := \Pr[E_1 \cup E_2 \cup \cdots \cup E_n] = \Pr[E_1] + \Pr[E_2] \cdots \Pr[E_n].
\]

How does one estimate \(p\)?

**Union Bound:**

\[
p \leq \Pr[E_1 \cup E_2 \cup \cdots \cup E_n] \leq \Pr[E_1] + \Pr[E_2] \cdots \Pr[E_n].
\]

\[
\Pr[E_k] = e^{-\frac{b}{n}}, \text{ for } k = 1, \ldots, n.
\]

Plug in and get

\[
p \leq ne^{-\frac{b}{n}}.
\]
Random Variables

Collect all cards?

Thus,\[ P_r[\text{missing at least one card}] \leq ne^{-\frac{m}{2n}}.\]
Hence,\[ P_r[\text{missing at least one card}] \leq p \text{ when } m \geq nln\left(\frac{2}{p}\right).\]
To get \( p = 1/2 \), set \( m = nln(2n) \).
(\( n \leq e^{-\frac{m}{2n}} \leq ne^{-nln(\frac{2}{p})} \leq (\frac{p}{2})^n \))
E.g., \( n = 10^2 \Rightarrow m = 530; n = 10^3 \Rightarrow m = 7600.\)

Questions about outcomes ...

Experiment: roll two dice.
Sample Space: \((1,1),(1,2),\ldots,(6,6)\) = \((1,\ldots,6)^2\)
How many pips?
Experiment: flip 100 coins.
Sample Space: \{HHH, H, THH ... H, ... TTT ... T\}
How many heads in 100 coin tosses?
Experiment: choose a random student in cs70.
Sample Space: \(\{\text{Adam, Jin, Bing, ... , Angeline}\}\)
What midterm score?
Experiment: hand back assignments to 3 students at random.
Sample Space: \(\{123, 132, 213, 231, 312, 321\}\)
How many students get back their own assignment?
In each scenario, each outcome gives a number.
The number is a (known) function of the outcome.

Quick Review.

Bayes’ Rule, Mutual Independence, Collisions and Collecting

Main results:
- **Bayes’ Rule:** \( P_r[A_m|B] = \frac{P_r[B|A_m]P_r[A_m]}{P_r[B]} \).
- **Product Rule:** \( P_r[A_1 \cap \cdots \cap A_n] = P_r[A_1]P_r[A_2|A_1] \cdots P_r[A_n|A_1 \cap \cdots \cap A_{n-1}] \).
- **Balls in bins:** \( m \) balls into \( n > m \) bins.
  \( P_r[\text{no collisions}] \approx \exp\left(\frac{-mn}{2n}\right) \).
- **Coupon Collection:** \( n \) items. Buy \( m \) cereal boxes.
  \( P_r[\text{miss one specific item}] \approx e^{-\frac{m}{n}} \); \( P_r[\text{miss any one of the items}] \leq ne^{-\frac{m}{n}} \).

Key Mathematical Fact: \( \ln(1-\varepsilon) \approx -\varepsilon \).

Random Variables

A random variable, \( X \), for an experiment with sample space \( \Omega \) is a function \( X : \Omega \rightarrow \mathbb{R} \).
Thus, \( X(\cdot) \) assigns a real number \( X(\omega) \) to each \( \omega \in \Omega \).

Example 1 of Random Variable

A random variable \( X \) is defined on the outcomes \( \Omega \).
The function \( X(\cdot) \) is not random, not a variable!
What varies at random (from experiment to experiment)? The outcome!
Example 2 of Random Variable

Experiment: flip three coins
Sample Space: \{HHH, THH, HTH, TTH, HHT, THT, HTT, TTT\}
Winnings: if win 1 on heads, lose 1 on tails: \(X\)
\(X(HHH) = 3\) \(X(THH) = 1\) \(X(HTH) = 1\) \(X(TTH) = -1\)
\(X(HHT) = 1\) \(X(THT) = -1\) \(X(HTT) = -1\) \(X(TTT) = -3\)

Handing back assignments

Experiment: hand back assignments to 3 students at random.
Sample Space: \(\Omega = \{123, 132, 213, 231, 312, 321\}\)
How many students get back their own assignment?
Random Variable: values of \(X(\omega)\): \{3, 1, 1, 0, 0, 1\}
Distribution:

<table>
<thead>
<tr>
<th>(X)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>w.p.</td>
<td>1/3</td>
<td>1/2</td>
<td>1/6</td>
<td>0</td>
</tr>
</tbody>
</table>

Flip three coins

Experiment: flip three coins
Sample Space: \{HHH, THH, HTH, TTH, HHT, THT, HTT, TTT\}
Winnings: if win 1 on heads, lose 1 on tails. \(X\)
Random Variable: \{3, 1, 1, -1, -1, -1, -3\}
Distribution:

<table>
<thead>
<tr>
<th>(X)</th>
<th>-3</th>
<th>-1</th>
<th>1</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>w. p.</td>
<td>1/8</td>
<td>3/8</td>
<td>3/8</td>
<td>1/8</td>
</tr>
</tbody>
</table>

Number of pips.

Experiment: roll two dice.

Number of pips in two dice. "What is the likelihood of getting \(n\) pips?"

\(Pr[X = 10] = \frac{3}{36} = Pr[X^{-1}(10)]; Pr[X = 8] = \frac{5}{36} = Pr[X^{-1}(8)].\)

Distribution

The probability of \(X\) taking on a value \(a\).

Definition: The distribution of a random variable \(X\), is
\(\{(a, Pr[X = a]) : a \in \mathcal{A}\}\), where \(\mathcal{A}\) is the range of \(X\).

\(Pr[X = a] := Pr[X^{-1}(a)]\) where \(X^{-1}(a) := \{\omega | X(\omega) = a\}.)
Expectation.  
How did people do on the midterm?  
Distribution.  
Summary of distribution?  
Average!  

An Example  
Flip a fair coin three times.  
Ω = {HHH, HHT, HTH, THH, HTT, TTH, THT, TTT}.  
X = number of H's: {3, 2, 2, 2, 1, 1, 1, 0}.  
Thus,  
\[ \sum_{\omega} X(\omega)Pr[\omega] = (3 + 2 + 2 + 2 + 1 + 1 + 1 + 0) \times \frac{1}{8} \]  
Also,  
\[ \sum_{a} a \times Pr[X = a] = 3 \times \frac{1}{8} + 2 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} + 0 \times \frac{1}{8} \]  
What’s the answer? Uh... 3/2  

Expectation and Average.  
There are n students in the class;  
\[ X(m) = \text{score of student } m, \text{ for } m = 1, 2, \ldots, n. \]  
*Average score* of the n students: add scores and divide by n:  
\[ \text{Average} = \frac{X(1) + X(1) + \ldots + X(n)}{n}. \]  
Experiment: choose a student uniformly at random.  
Uniform sample space: Ω = {1, 2, ..., n}, Pr[ω] = 1/n, for all ω.  
Random Variable: midterm score: X(ω).  
Expectation:  
\[ E(X) = \sum_{\omega} X(\omega)Pr[\omega] = \sum_{\omega} X(\omega)\frac{1}{n}. \]  
Hence,  
\[ \text{Average} = E(X). \]  
This holds for a uniform probability space.  

Expectation - Definition  
Definition: The expected value of a random variable X is  
\[ E[X] = \sum_{a} a \times Pr[X = a]. \]  
The expected value is also called the mean.  
According to our intuition, we expect that if we repeat an experiment a large number N of times and if \( X_1, \ldots, X_N \) are the successive values of the random variable, then  
\[ \frac{X_1 + \ldots + X_N}{N} \approx E[X]. \]  
That is indeed the case, in the same way that the fraction of times that \( X = x \) approaches \( Pr[X = x] \).  
This (nontrivial) result is called the Law of Large Numbers.  
The subjectivist(bayesian) interpretation of \( E[X] \) is less obvious.  

Expectation: A Useful Fact  
Theorem:  
\[ E[X] = \sum_{\omega} X(\omega) \times Pr[\omega]. \]  
Proof:  
\[ E[X] = \sum_{a} a \times Pr[X = a] \]  
\[ = \sum_{a} a \times \sum_{\omega : X(\omega) = a} Pr[\omega] \]  
\[ = \sum_{\omega} \omega \times [X(\omega)Pr[\omega]] \]  
\[ = \sum_{\omega} X(\omega)Pr[\omega] \]  
Distributive property of multiplication over addition.  

Named Distributions.  
Some distributions come up over and over again.  
...like “choose” or “stars and bars”.  
Let’s cover some.
The binomial distribution.

- Flip $n$ coins with heads probability $p$.
- Random variable: number of heads.
- Binomial Distribution: $\Pr[X = i]$, for each $i$.
- How many sample points in event “$X = i$”? $i$ heads out of $n$ coin flips $\implies \binom{n}{i}$
- What is the probability of $\omega$ if $\omega$ has $i$ heads?
  - Probability of heads in any position is $p$.
  - Probability of tails in any position is $(1 - p)$.
- So, we get
  
  \[ \Pr[\omega] = p^i (1 - p)^{n-i}. \]

Probability of “$X = i$” is sum of $\Pr[\omega]$, $\omega \in \{X = i\}$.

\[ \Pr[X = i] = \binom{n}{i} p^i (1 - p)^{n-i}, i = 0, 1, \ldots, n : B(n, p) \text{ distribution} \]

Expectation of Binomial Distribution

Parameter $p$ and $n$. What is expectation?

\[ \Pr[X = i] = \binom{n}{i} p^i (1 - p)^{n-i}, i = 0, 1, \ldots, n : B(n, p) \text{ distribution} \]

\[ E[X] = \sum_i i \times \Pr[X = i]. \]

Uh oh? Well... It is $pn$.

Proof? After linearity of expectation this is easy.

Waiting is good.

The binomial distribution.

| $X$ | $\begin{array}{cccccc} T & T & T & T & T & T \\
1 & 2 & 3 & 4 & 5 & 6 \\
(1-p) & p & p^2 & p^3 & p^4 & p^5 \end{array}$ |
|-----|----------------------------------|
| $X$ | $\Pr[X = i] = \binom{n}{i} p^i (1 - p)^{n-i}, i = 0, 1, \ldots, n : B(n, p) \text{ distribution} \]

Uniform Distribution

Roll a six-sided balanced die. Let $X$ be the number of pips (dots). Then $X$ is equally likely to take any of the values $\{1, 2, \ldots, 6\}$. We say that $X$ is uniformly distributed in $\{1, 2, \ldots, 6\}$.

- More generally, we say that $X$ is uniformly distributed in $\{1, 2, \ldots, n\}$ if $\Pr[X = m] = 1/n$ for $m = 1, 2, \ldots, n$.
- In that case,

\[ E[X] = \sum_{m=1}^{n} m \Pr[X = m] = \sum_{m=1}^{n} m \times \frac{1}{n} = \frac{n(n+1)}{2} = \frac{n+1}{2}. \]

Error channel and...

A packet is corrupted with probability $p$.

Send $n + 2k$ packets.

Probability of at most $k$ corruptions.

\[ \sum_{i=k}^{n} \binom{n + 2k}{i} p^i (1 - p)^{n+2k-i}. \]

Also distribution in polling, experiments, etc.

Geometric Distribution

Let’s flip a coin with $\Pr[H] = p$ until we get $H$.

For instance:

- $\omega_1 = H$, or
- $\omega_2 = T H$, or
- $\omega_3 = T T H$, or
- $\omega_4 = T T T T \ldots T H$.

Note that $\Omega = \{ \omega_n, n = 1, 2, \ldots \}$.

Let $X$ be the number of flips until the first $H$. Then, $X(\omega_n) = n$.

Also,

\[ \Pr[X = n] = (1 - p)^{n-1} p, n \geq 1. \]
Indeed, $\Pr[X = n] = (1 - p)^{n-1}p, n \geq 1$.

Thus, $\Pr[X = n] = (1 - p)^{n-1}p, n \geq 1$.

Note that

$$\sum_{n=1}^{\infty} \Pr[X_n] = \sum_{n=1}^{\infty} (1 - p)^{n-1}p = \frac{p}{1 - (1 - p)} = 1.$$

Now, if $|a| < 1$, then $S := \sum_{n=0}^{\infty} a^n = \frac{1}{1-a}$. Indeed,

$$S = 1 + a + a^2 + a^3 + \cdots,$$

$$aS = a + a^2 + a^3 + \cdots,$$

$$(1-a)S = 1 - a - a^2 - a^3 - \cdots = 1.$$

Hence,

$$\sum_{n=1}^{\infty} \Pr[X_n] = p \frac{1}{1 - (1 - p)} = 1.$$

Poisson

Experiment: flip a coin $n$ times. The coin is such that $\Pr[H] = \lambda/n$.
Random Variable: $X$ - number of heads. Thus, $X = B(n, \lambda/n)$.

**Poisson Distribution** is distribution of $X$ "for large $n$.”

**Poisson Distribution: Expectation**

$$X = G(p), \text{ i.e., } \Pr[X = n] = (1 - p)^{n-1}p, n \geq 1.$$

One has

$$E[X] = \sum_{n=1}^{\infty} n \Pr[X = n] = \sum_{n=1}^{\infty} n(1 - p)^{n-1}p.$$

Thus,

$$E[X] = p + 2(1 - p)p + 3(1 - p)^2p + 4(1 - p)^3p + \cdots$$

$$(1 - p)E[X] = (1 - p)p + 2(1 - p)^2p + 3(1 - p)^3p + \cdots$$

$pE[X] = p + (1 - p)p + (1 - p)^2p + (1 - p)^3p + \cdots$

by subtracting the previous two identities

$$\sum_{n=1}^{\infty} \Pr[X = n] = 1.$$

Hence,

$$E[X] = \frac{1}{p}.$$

Poisson Distribution: Definition and Mean

**Definition Poisson Distribution with parameter $\lambda > 0$**

$$X = P(\lambda) \iff \Pr[X = m] = \frac{\lambda^m}{m!} e^{-\lambda}, m \geq 0.$$

**Fact** $E[X] = \lambda$.

**Proof:**

$$E[X] = \sum_{m=0}^{\infty} m \frac{\lambda^m}{m!} e^{-\lambda} = e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!}$$

$$= e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = e^{-\lambda} \lambda = \lambda e^{\lambda}.$$
The Poisson distribution is named after: Simeon Poisson

The geometric distribution is named after: Equal Time: B.

I could not find a picture of D. Binomial, sorry.

**Summary**

- A random variable $X$ is a function $X : \Omega \to \mathbb{R}$.
- $Pr[X = a] := Pr[X^{-1}(a)] = Pr[\{\omega | X(\omega) = a\}]$.
- $Pr[X \in A] := Pr[X^{-1}(A)]$.
- The distribution of $X$ is the list of possible values and their probability: $\{(a, Pr[X = a]), a \in \mathcal{X}\}$.
- $E[X] := \sum a \cdot Pr[X = a]$.
- Expectation is Linear.
- $\mathcal{B}(n, p), \mathcal{U}[1 : n], \mathcal{G}(p), \mathcal{P}(\lambda)$. 

**Random Variables**

\[
\begin{array}{cccc}
1 & 2 & 3 & 4 & 5 \\
\end{array}
\]  

\[
\begin{array}{cccc}
1 & 2 & 3 & 4 & 5 \\
(1-p)(1-p)(1-p)
\end{array}
\]  

\[
\begin{array}{cccc}
1 & 2 & 3 & 4 \\
(1-p)(1-p)p
\end{array}
\]  

\[
\begin{array}{cccc}
1 & 2 & 3 \\
(1-p)^2p & 2p
\end{array}
\]  

[Image of Simeon Poisson]