Expectation
Recall: \( X : \Omega \to \Re; \Pr[X = a] = \Pr[X^{-1}(a)] \);
Definition: The expectation \( E[X] \) of a random variable \( X \) is
\[
E[X] = \sum_{\omega} a \times \Pr[X = a].
\]
Indicator:
Let \( A \) be an event. The random variable \( X \) defined by
\[
X(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{if } \omega \notin A \end{cases}
\]
is called the indicator of the event \( A \).
Note that \( \Pr[X = 1] = \Pr[A] \) and \( \Pr[X = 0] = 1 - \Pr[A] \).
Hence,
\[
E[X] = 1 \times \Pr[X = 1] + 0 \times \Pr[X = 0] = \Pr[A].
\]
The random variable \( X \) is sometimes written as
\( 1_{\{\omega \in A\}} \) or \( 1_A(\omega) \).

Using Linearity - 1: Pips on dice
Roll a die \( n \) times.
\( X_m = \) number of pips on roll \( m \).
\( X = X_1 + \cdots + X_n = \) total number of pips in \( n \) rolls.
\[
E[X] = E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n], \quad \text{by linearity}
\]
\( = nE[X_1], \) because the \( X_m \) have the same distribution
Now,
\[
E[X_1] = 1 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = \frac{7}{2}.
\]
Hence,
\[
E[X] = \frac{7n}{2}.
\]

Using Linearity - 2: Fixed point.
Hand out assignments at random to \( n \) students.
\( X = \) number of students that get their own assignment back.
\( X = X_1 + \cdots + X_n, \) where
\( X_m = 1 \) (student \( m \) gets his/her own assignment back).
One has
\[
E[X] = E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n], \quad \text{by linearity}
\]
\( = nE[X_1], \) because all the \( X_m \) have the same distribution
\( = n \times 1, \) because \( X_1 \) is an indicator
\( = n(1/n), \) because student 1 is equally likely
to get any one of the \( n \) assignments
\( = 1. \)
Note that linearity holds even though the \( X_m \) are not independent (whatever that means).

Using Linearity - 3: Binomial Distribution.
Flip \( n \) coins with heads probability \( p \). \( X = \) number of heads
Binomial Distribution: \( \Pr[X = i], \) for each \( i \).
\[
\Pr[X = i] = \binom{n}{i} p^i (1-p)^{n-i}.
\]
Furthermore,
\[
E[X] = \sum_i i \times \Pr[X = i] = \sum_i \binom{n}{i} i p^i (1-p)^{n-i}.
\]
Uh oh. ... Or... a better approach: Let
\[
X_i = \begin{cases} 1, & \text{if } \text{ith flip is heads} \\ 0, & \text{otherwise} \end{cases}
\]
\( E[X_i] = \frac{1}{2}. \)
Moreover \( X = X_1 + \cdots + X_n \) and
\[
E[X] = E[X_1] + E[X_2] + \cdots + E[X_n] = n \times E[X_1] = np.
\]

Linearity of Expectation
Theorem:
\[
E[X] = \sum_{\omega} X(\omega) \times \Pr[\omega].
\]
Theorem: Expectation is linear
\[
E[a_1 X_1 + \cdots + a_n X_n] = a_1 E[X_1] + \cdots + a_n E[X_n].
\]
Proof:
\[
E[a_1 X_1 + \cdots + a_n X_n] = \sum_{\omega} (a_1 X_1(\omega) + \cdots + a_n X_n(\omega)) \Pr[\omega]
\]
\( = a_1 \sum_{\omega} X_1(\omega) \Pr[\omega] + \cdots + a_n \sum_{\omega} X_n(\omega) \Pr[\omega].
\]
\( = a_1 E[X_1] + \cdots + a_n E[X_n]. \quad \Box \)
**Conditional Expectation**

How do observations affect expectation?

**Example 1:**
Roll one die. You are told that the outcome $X$ is at least 3. What is the expected value of $X$ given that information?

Given that $X \geq 3$, we know that $X$ is uniform in $\{3, 4, 5, 6\}$. Hence, the mean value is 4.5. We write

$$E[X|X \geq 3] = 4.5.$$  

Similarly, we have

$$E[X|X < 3] = 1.5.$$  

because, given that $X < 3$, $X$ is uniform in $\{1, 2\}$. Note that

$$E[X|X \geq 3] \times Pr[X \geq 3] + E[X|X < 2] \times Pr[X < 2] = 4.5 \times 4/6 + 1.5 \times 2/6 \approx 3.5 = E[X].$$

Is this a coincidence?

---

**Conditional Expectation**

How do observations affect expectation?

**Example 2:**
Roll two dice. You are told that the total number $X$ of pips is at least 8. What is the expected value of $X$ given that information?


Given that $X \geq 8$, the distribution of $X$ becomes

$$\{(8,5/15), (9,4/15), (10,3/15), (11,2/15), (12,1/15)\}.$$  

For instance,

$$Pr[X = 8|X \geq 8] = Pr[X = 8] \times Pr[X \geq 8] = 5/36$$

Hence,

$$E[X|X \geq 8] = 5 \times 15/36 + 4 \times 10/36 + 3 \times 11/36 + 2 \times 12/36 = 140/15 \approx 9.33.$$  

Coincidence? Probably not.

---

**Conditional Probability**

**Definition**

Let $X$ be a RV and $A$ an event. Then

$$E[X|A] := \sum_a \frac{a \times Pr[X = a|A]}{Pr[A]}.$$  

It is easy (really) to see that

$$E[X] = \sum_a X(a) \times Pr[X = a] = \sum_a \frac{X(a) \times Pr[X = a]}{Pr[A]} \times Pr[A].$$

**Theorem** Conditional Expectation is linear

$$E[a_1 X_1 + \cdots + a_n X_n|A] = a_1 E[X_1|A] + \cdots + a_n E[X_n|A].$$

**Proof**

The law of total probability says that

$$Pr[a] = Pr[a|A]Pr[A] + Pr[a|A^c]Pr[A^c].$$

Hence,

$$E[X] = \sum_a X(a) \times Pr[a] = \sum_a [X(a) \times Pr[a|A] \times Pr[A] + \sum_a X(a) \times Pr[a|A^c] \times Pr[A^c]]$$

$$= E[X|A]Pr[A] + E[X|A^c]Pr[A^c].$$

---

**Conditional Probability**

**Geometric Distribution**

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

For instance:

$$\omega_1 = H,$$  

$$\omega_2 = T H,$$  

$$\omega_3 = \cdots = 1, 2, \ldots \}.$$  

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

$$Pr[X = n] = (1 − p)^{n−1} p, \ n \geq 1.$$
Let \( X \) be \( G(p) \). Then, for \( n \geq 0 \),
\[
\Pr[X > n] = \Pr[ \text{first } n \text{ flips are T} ] = (1 - p)^n.
\]

**Theorem**
\[
\Pr[X > n + m | X > n] = \Pr[X > m], m,n \geq 0.
\]

**Proof:**
\[
\Pr[X > n + m | X > n] = \frac{\Pr[X > n + m \text{ and } X > n]}{\Pr[X > n]}
= \frac{\Pr[X > n + m]}{\Pr[X > n]}
= \frac{p + (1 - p)(1 + E[Y])}{1 - p}
= \frac{p + (1 - p)(1 - p)^m}{1 - p}
= \frac{(1 - p)^{m+1}}{1 - p}
= \Pr[X > m].
\]

Hence, \( E[X] = 1/p \).

**Geometric Distribution: Memoryless - Interpretation**
\[
\Pr[X > n + m | X > n] = \Pr[X > m], m,n \geq 0.
\]

The coin is memoryless, therefore, so is \( X \).
Expected Value of Integer RV

**Theorem:** For a r.v. $X$ that takes values in $\{0,1,2,\ldots\}$, one has

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

**Proof:** One has

$$E[X] = \sum_{i=1}^{\infty} i \times \Pr[X = i] = \sum_{i=1}^{\infty} \left( \frac{i}{i-1} \times \Pr[X \geq i] - (i-1) \times \Pr[X \geq i+1] \right)$$

$$= \sum_{i=1}^{\infty} \left( i \times \Pr[X \geq i] - i \times \Pr[X \geq i+1] \right)$$

$$= \sum_{i=1}^{\infty} \left( i \times \Pr[X \geq i] - (i-1) \times \Pr[X \geq i+1] \right)$$

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

Hence,

$$E[X] = \sum_{i=1}^{\infty} \frac{(1-\bar{p})^{i-1}}{1-\bar{p}} = \frac{1}{\bar{p}}$$

Riding the bus.

$n$ buses arrive uniformly at random throughout a 24 hour day. What is the time between buses? What is the time to wait for a bus? Here are typical arrival times, independent and uniform in $[0,24]$.

Here is an alternative picture (left)

Summary.

**Expectation; Conditional Expectation:** $B(n,p); G(p)$

**Expectation:** $E[X] = \sum_{x \in \mathbb{X}} a \times \Pr[X = a] = \sum_{x \in \mathbb{X}} X(a) \Pr[a]$.  

**Linearity:** $E[a_1X_1 + \cdots + a_nX_n] = a_1E[X_1] + \cdots + a_nE[X_n]$.  

**Binomial:** $\Pr[X = \ell] = \binom{n}{\ell} \bar{p}^{\ell} (1-\bar{p})^{n-\ell}; E(X) = np$.  

**Geometric:** $\Pr[X = \ell] = (1-\bar{p})^{\ell-1} \bar{p}; E(X) = \frac{1}{\bar{p}}$; memoryless.  

**Condition Expectation:** $E[X|A]$. Linear and  

$$E[X] = E[X|A] \Pr[A] + E[X|\overline{A}] \Pr[\overline{A}]$$

Geometric Distribution: Yet another look

**Theorem:** For a r.v. $X$ that takes the values $\{0,1,2,\ldots\}$, one has

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

[See later for a proof.]

If $X = G(p)$, then $\Pr[X \geq i] = \Pr[X > i - 1] = (1-\bar{p})^{i-1}$.

Hence,

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

Riding the bus.

Add the black dot uniformly at random and pretend that it represents $0/24$.

This is legitimate, because given the black dot, the other dots are uniform at random.

Then,

$$24 = E[X_1 + \cdots + X_n] = 5E[X_1]$$

Hence,

$$E[X_1] = E[X_n] = \frac{24}{5} = \frac{24}{n+1} \text{ for } n \text{ busses.}$$