1. Random Variables: Brief Review
2. Expectation
3. Important Distributions
Random Variables: Definitions

Definition
A random variable, $X$, for a random experiment with sample space $\Omega$ is a function $X : \Omega \to \mathbb{R}$.

Thus, $X(\cdot)$ assigns a real number $X(\omega)$ to each $\omega \in \Omega$.

Definitions
(a) For $a \in \mathbb{R}$, one defines
$$X^{-1}(a) := \{ \omega \in \Omega \mid X(\omega) = a \}.$$

(b) For $A \subset \mathbb{R}$, one defines
$$X^{-1}(A) := \{ \omega \in \Omega \mid X(\omega) \in A \}.$$

(c) The probability that $X = a$ is defined as
$$Pr[X = a] = Pr[X^{-1}(a)].$$

(d) The probability that $X \in A$ is defined as
$$Pr[X \in A] = Pr[X^{-1}(A)].$$

(e) The distribution of a random variable $X$, is
$$\{(a, Pr[X = a]) : a \in \mathcal{A}\},$$
where $\mathcal{A}$ is the range of $X$. That is, $\mathcal{A} = \{X(\omega), \omega \in \Omega\}$. 
Random Variables: Definitions

**Definition**
Let $X, Y, Z$ be random variables on $\Omega$ and $g : \mathbb{R}^3 \to \mathbb{R}$ a function. Then $g(X, Y, Z)$ is the random variable that assigns the value $g(X(\omega), Y(\omega), Z(\omega))$ to $\omega$.

Thus, if $V = g(X, Y, Z)$, then $V(\omega) := g(X(\omega), Y(\omega), Z(\omega))$.

Examples:

- $X^k$
- $(X - a)^2$
- $a + bX + cX^2 + (Y - Z)^2$
- $(X - Y)^2$
- $X \cos(2\pi Y + Z)$. 
**Definition:** The **expected value** (or mean, or expectation) of a random variable $X$ is

$$E[X] = \sum_a a \times Pr[X = a].$$

**Theorem:**

$$E[X] = \sum_\omega X(\omega) \times Pr[\omega].$$
An Example

Flip a fair coin three times.

Ω = \{HHH, HHT, HTH, THH, HTT, THT, TTH, 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,

\[ \sum_{a} a \times Pr[X = a] = 3 \times \frac{1}{8} + 2 \times \frac{3}{8} + 1 \times \frac{3}{8} + 0 \times \frac{1}{8}. \]
Win or Lose.

Expected winnings for heads/tails games, with 3 flips?
Recall the definition of the random variable $X$:
$\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} \rightarrow \{3, 1, 1, -1, 1, -1, -1, -3\}$.

$$E[X] = 3 \times \frac{1}{8} + 1 \times \frac{3}{8} - 1 \times \frac{3}{8} - 3 \times \frac{1}{8} = 0.$$  

Can you ever win 0?
Apparently: expected value is not a common value, by any means.
The expected value of $X$ is not the value that you expect!
It is the average value per experiment, if you perform the experiment many times:
$$\frac{X_1 + \cdots + X_n}{n}, \text{ when } n \gg 1.$$  

The fact that this average converges to $E[X]$ is a theorem:
the Law of Large Numbers. (See later.)
Law of Large Numbers

An Illustration: Rolling Dice
Indicators

Definition
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,

$$E[X] = 1 \times Pr[X = 1] + 0 \times Pr[X = 0] = Pr[A].$$

This random variable $X(\omega)$ is sometimes written as

$$1\{\omega \in A\} \text{ or } 1_A(\omega).$$

Thus, we will write $X = 1_A$. 
**Linearity of Expectation**

**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:**

\[
E[a_1 X_1 + \cdots + a_n X_n] \\
= \sum_{\omega} (a_1 X_1 + \cdots + a_n X_n)(\omega) Pr[\omega] \\
= \sum_{\omega} (a_1 X_1(\omega) + \cdots + a_n X_n(\omega)) Pr[\omega] \\
= a_1 \sum_{\omega} X_1(\omega) Pr[\omega] + \cdots + a_n \sum_{\omega} X_n(\omega) Pr[\omega] \\
= a_1 E[X_1] + \cdots + a_n E[X_n].
\]

Note: If we had defined \( Y = a_1 X_1 + \cdots + a_n X_n \) has had tried to compute \( E[Y] = \sum_y y Pr[Y = y] \), we would have been in trouble!
Using Linearity - 1: Pips (dots) 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.

\[
E[X] = E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n], \text{ by linearity}
\]

$= nE[X_1]$, because the $X_m$ have the same distribution

Now,

\[
E[X_1] = 1 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = \frac{6 \times 7}{2} \times \frac{1}{6} = \frac{7}{2}.
\]

Hence,

\[
E[X] = \frac{7n}{2}.
\]

Note: Computing $\sum x \, Pr[X = x]$ directly is not easy!
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

\( X_m = 1 \{ \text{student } m \text{ gets his/her own assignment back} \} \).

One has

\[
E[X] = E[X_1 + \cdots + X_n] \\
= E[X_1] + \cdots + E[X_n], \text{ by linearity} \\
= nE[X_1], \text{ because all the } X_m \text{ have the same distribution} \\
= nPr[X_1 = 1], \text{ because } X_1 \text{ is an indicator} \\
= n(1/n), \text{ because student 1 is equally likely} \\
\quad \quad \text{to get any one of the } n \text{ assignments} \\
= 1.
\]

Note that linearity holds even though the \( X_m \) are not independent (whatever that means).

Note: What is \( Pr[X = m] \)? Tricky ....
Using Linearity - 3: Binomial Distribution.

Flip $n$ coins with heads probability $p$. $X$ - number of heads.

Binomial Distribution: $Pr[X = i]$, for each $i$.

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

$$E[X] = \sum_i i \times Pr[X = i] = \sum_i i \times \binom{n}{i} p^i (1 - p)^{n-i}.$$ 

Uh oh. ... Or... a better approach: Let 

$$X_i = \begin{cases} 
1 & \text{if } i\text{th flip is heads} \\
0 & \text{otherwise} 
\end{cases}$$

$$E[X_i] = 1 \times Pr[“heads”] + 0 \times Pr[“tails”] = p.$$ 

Moreover $X = X_1 + \cdots X_n$ and 

$$E[X] = E[X_1] + E[X_2] + \cdots E[X_n] = n \times E[X_i] = np.$$
Assume $A$ and $B$ are disjoint events. Then $1_{A\cup B}(\omega) = 1_A(\omega) + 1_B(\omega)$. Taking expectation, we get


In general, $1_{A\cup B}(\omega) = 1_A(\omega) + 1_B(\omega) - 1_{A\cap B}(\omega)$. Taking expectation, we get $Pr[A \cup B] = Pr[A] + Pr[B] - Pr[A \cap B]$.

Observe that if $Y(\omega) = b$ for all $\omega$, then $E[Y] = b$. Thus, $E[X + b] = E[X] + b$. 
Calculating $E[g(X)]$

Let $Y = g(X)$. Assume that we know the distribution of $X$.

We want to calculate $E[Y]$.

Method 1: We calculate the distribution of $Y$:

$$Pr[Y = y] = Pr[X \in g^{-1}(y)]$$

where $g^{-1}(x) = \{x \in \mathbb{R} : g(x) = y\}$.

This is typically rather tedious!

Method 2: We use the following result.

Theorem:

$$E[g(X)] = \sum_x g(x) Pr[X = x].$$

Proof:

$$E[g(X)] = \sum_{\omega} g(X(\omega)) Pr[\omega] = \sum_x \sum_{\omega \in X^{-1}(x)} g(X(\omega)) Pr[\omega]$$

$$= \sum_x \sum_{\omega \in X^{-1}(x)} g(x) Pr[\omega] = \sum_x g(x) \sum_{\omega \in X^{-1}(x)} Pr[\omega]$$

$$= \sum_x g(x) Pr[X = x].$$
An Example

Let $X$ be uniform in $\{-2, -1, 0, 1, 2, 3\}$.

Let also $g(X) = X^2$. Then (method 2)

$$E[g(X)] = \sum_{x=-2}^{3} x^2 \frac{1}{6}$$

$$= \{4 + 1 + 0 + 1 + 4 + 9\} \frac{1}{6} = \frac{19}{6}.$$

Method 1 - We find the distribution of $Y = X^2$:

$$Y = \begin{cases} 
4, & \text{w.p. } \frac{2}{6} \\
1, & \text{w.p. } \frac{1}{6} \\
0, & \text{w.p. } \frac{1}{6} \\
9, & \text{w.p. } \frac{1}{6}.
\end{cases}$$

Thus,

$$E[Y] = 4 \frac{2}{6} + 1 \frac{2}{6} + 0 \frac{1}{6} + 9 \frac{1}{6} = \frac{19}{6}.$$
Calculating $E[g(X, Y, Z)]$

We have seen that $E[g(X)] = \sum_x g(x) Pr[X = x]$.

Using a similar derivation, one can show that

$$E[g(X, Y, Z)] = \sum_{x,y,z} g(x, y, z) Pr[X = x, Y = y, Z = z].$$

An Example. Let $X, Y$ be as shown below:

$$E[\cos(2\pi X + \pi Y)] = 0.1 \cos(0) + 0.4 \cos(2\pi) + 0.2 \cos(\pi) + 0.3 \cos(3\pi)$$
$$= 0.1 \times 1 + 0.4 \times 1 + 0.2 \times (-1) + 0.3 \times (-1) = 0.$$
Best Guess: Least Squares

If you only know the distribution of $X$, it seems that $E[X]$ is a ‘good guess’ for $X$.

The following result makes that idea precise.

**Theorem**
The value of $a$ that minimizes $E[(X - a)^2]$ is $a = E[X]$.

**Proof 1:**

\[
E[(X - a)^2] = E[(X - E[X] + E[X] - a)^2]
\]

\[
= E[(X - E[X])^2 + 2(X - E[X])(E[X] - a) + (E[X] - a)^2]
\]

\[
= E[(X - E[X])^2] + 2(E[X] - a)E[X - E[X]] + (E[X] - a)^2
\]

\[
= E[(X - E[X])^2] + 0 + (E[X] - a)^2
\]

\[
\geq E[(X - E[X])^2].
\]
If you only know the distribution of $X$, it seems that $E[X]$ is a ‘good guess’ for $X$.
The following result makes that idea precise.

**Theorem**
The value of $a$ that minimizes $E[(X - a)^2]$ is $a = E[X]$.

**Proof 2:**
Let


To find the minimizer of $g(a)$, we set to zero $\frac{d}{da}g(a)$.

We get $0 = \frac{d}{da}g(a) = -2E[X] + 2a$.
Hence, the minimizer is $a = E[X]$. 

\[\square\]
Best Guess: Least Absolute Deviation

Thus $E[X]$ minimizes $E[(X - a)^2]$. It must be noted that the measure of the ‘quality of the approximation’ matters. The following result illustrates that point.

**Theorem**
The value of $a$ that minimizes $E[|X - a|]$ is the *median* of $X$.

The median $\nu$ of $X$ is any real number such that

$$Pr[X \leq \nu] = Pr[X \geq \nu]$$

**Proof:**

$$g(a) := E[|X - a|] = \sum_{x \leq a} (a - x) Pr[X = x] + \sum_{x \geq a} (x - a) Pr[X = x].$$

Thus, if $0 < \varepsilon << 1$,

$$g(a + \varepsilon) = g(a) + \varepsilon Pr[X \leq a] - \varepsilon Pr[X \geq a].$$

Hence, changing $a$ cannot reduce $g(a)$ only if $Pr[X \leq a] = Pr[X \geq a]$. 

\[\square\]
$X \in \{1, 2, 3, 11, 13\}$
equal probabilities

$E[|X - a|]$

$E[(X - a)^2] \times \frac{1}{10}$
Best Guess: Another Illustration

\( X \in \{1, 2, 11, 13\} \)

equal probabilities

\[
E[|X - a|] = \frac{E[(X - a)^2]}{10}
\]

\[
E[X] \quad \text{mean}
\]

\[
\text{median}
\]
The expected value has a *center of mass* interpretation:

\[
\sum_{n} p_n (a_n - \mu) = 0 \\
\iff \mu = \sum_{n} a_n p_n = E[X]
\]
Monotonicity

Definition
Let $X, Y$ be two random variables on $\Omega$. We write $X \leq Y$ if $X(\omega) \leq Y(\omega)$ for all $\omega \in \Omega$, and similarly for $X \geq Y$ and $X \geq a$ for some constant $a$.

Facts
(a) If $X \geq 0$, then $E[X] \geq 0$.
(b) If $X \leq Y$, then $E[X] \leq E[Y]$.

Proof
(a) If $X \geq 0$, every value $a$ of $X$ is nonnegative. Hence,

$$E[X] = \sum_a aPr[X = a] \geq 0.$$ 

(b) $X \leq Y \Rightarrow Y - X \geq 0 \Rightarrow E[Y] - E[X] = E[Y - X] \geq 0$.

Example:

$$B = \bigcup_m A_m \Rightarrow 1_B(\omega) \leq \sum_m 1_{A_m}(\omega) \Rightarrow Pr[\bigcup_m A_m] \leq \sum_m Pr[A_m].$$
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,

$$E[X] = \sum_{m=1}^{n} m Pr[X = m] = \sum_{m=1}^{n} m \times \frac{1}{n} = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}.$$
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
$\omega_n = T \; T \; T \; T \; \cdots \; T \; H$.

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,

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

\[ Pr[X = n] = (1 - p)^{n-1} p, \quad n \geq 1. \]
Geometric Distribution

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

Note that

\[
\sum_{n=1}^{\infty} Pr[X_n] = \sum_{n=1}^{\infty} (1 - p)^{n-1} p = p \sum_{n=1}^{\infty} (1 - p)^{n-1} = p \sum_{n=0}^{\infty} (1 - p)^n.
\]

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

\[
S = 1 + a + a^2 + a^3 + \cdots
\]

\[
aS = a + a^2 + a^3 + a^4 + \cdots
\]

\[
(1-a)S = 1 + a - a + a^2 - a^2 + \cdots = 1.
\]

Hence,

\[
\sum_{n=1}^{\infty} Pr[X_n] = p \frac{1}{1 - (1 - p)} = 1.
\]
Geometric Distribution: Expectation

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

One has

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

Thus,

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

by subtracting the previous two identities

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

Hence,

\[ E[X] = \frac{1}{\rho}. \]
Geometric Distribution: Memoryless

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:**

$$Pr[X > n + m | X > n] = \frac{Pr[X > n + m \text{ and } X > n]}{Pr[X > n]}$$

$$= \frac{Pr[X > n + m]}{Pr[X > n]}$$

$$= \frac{(1 - p)^{n+m}}{(1 - p)^n} = (1 - p)^m$$

$$= Pr[X > m].$$
Geometric Distribution: Memoryless - Interpretation

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

The coin is memoryless, therefore, so is \( X \).
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 $\Pr[X \geq i] = \Pr[X > i - 1] = (1 - p)^{i-1}$.

Hence,

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

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

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

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

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

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

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 $n$.”

We expect $X \ll n$. For $m \ll n$ one has

$$Pr[X = m] = \binom{n}{m} p^m (1 - p)^{n-m}, \text{ with } p = \lambda / n$$

$$= \frac{n(n-1) \cdots (n-m+1)}{m!} \left( \frac{\lambda}{n} \right)^m \left( 1 - \frac{\lambda}{n} \right)^{n-m}$$

$$= \frac{n(n-1) \cdots (n-m+1) \lambda^m}{m!} \frac{1}{n^m} \left( 1 - \frac{\lambda}{n} \right)^{n-m}$$

$$\approx^{(1)} \frac{\lambda^m}{m!} \left( 1 - \frac{\lambda}{n} \right)^{n-m} \approx^{(2)} \frac{\lambda^m}{m!} \left( 1 - \frac{\lambda}{n} \right)^n \approx \frac{\lambda^m}{m!} e^{-\lambda}.$$

For (1) we used $m \ll n; \text{ for (2) we used } (1 - a/n)^n \approx e^{-a}.$
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:**

$$E[X] = \sum_{m=1}^{\infty} m \times \frac{\lambda^m}{m!} e^{-\lambda} = e^{-\lambda} \sum_{m=1}^{\infty} \frac{\lambda^m}{(m-1)!}$$

$$= e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^{m+1}}{m!} = e^{-\lambda} \lambda \sum_{m=0}^{\infty} \frac{\lambda^m}{m!}$$

$$= e^{-\lambda} \lambda e^\lambda = \lambda.$$
Simeon Poisson

The Poisson distribution is named after:
The geometric distribution is named after:

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

- A random variable $X$ is a function $X : \Omega \rightarrow \mathbb{R}$.
- $Pr[X = a] := Pr[X^{-1}(a)] = Pr\{\omega \mid X(\omega) = a\}$.
- $Pr[X \in A] := Pr[X^{-1}(A)]$.
- The distribution of $X$ is the list of possible values and their probability: $\{(a, Pr[X = a]), a \in \mathcal{A}\}$.
- $g(X, Y, Z)$ assigns the value ....
- $E[X] := \sum_a aPr[X = a]$.
- Expectation is Linear.
- $B(n, p), U[1 : n], G(p), P(\lambda)$. 