### Random Variables: Definitions

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

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

**Definitions**
- (a) For $a \in \mathbb{R}$, one defines $X^{-1}(a) := \{\omega \in \Omega | X(\omega) = a\}$.
- (b) For $A \subseteq \mathbb{R}$, one defines $X^{-1}(A) := \{\omega \in \Omega | X(\omega) \in A\}$.
- (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, $\mathcal{A} = \{X(\omega), \omega \in \Omega\}$.

### Win or Lose.

Expected winnings for heads/tails games, with 3 flips?
Recall the definition of the random variable $X$:

$$X = \{(HHH, 3), (HHT, 2), (HTH, 2), (THH, 2), (HTT, 1), (THT, 1), (TTH, 1), (TTT, 0)\}.$$  

$X$ is 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,

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

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 $1\{\omega \in A\}$ or $1_A(\omega)$. Thus, we will write $X = 1_A$.

Linearity of Expectation
Theorem: Expectation is linear
$$E[a_1 X_1 + \cdots + a_n X_n] = a_1 E[X_1] + \cdots + a_n E[X_n].$$

Using Linearity - 1: Pips (dots) on dice
Roll a die $n$ times. $X_\omega = \text{number of pips on roll } \omega$. $X = X_1 + \cdots + X_n = \text{total number of pips in } n \text{ rolls}.$
$$E[X] = E[X_1 + \cdots + X_n]$$
$$= E[X_1] + \cdots + E[X_n], \text{ by linearity}$$
$$= nE[X_1], \text{ because the } X_\omega \text{ have the same distribution}$$
Now,
$$E[X_1] = \sum_{i=1}^6 \frac{i}{6} = 7 \cdot \frac{1}{6} = \frac{7}{2}$$
Hence,
$$E[X] = 7n \cdot \frac{1}{2}$$
Note: Computing $\sum x Pr[X = x]$ directly is not easy!

Using Linearity - 2: Fixed point.
Hand out assignments at random to $n$ students. $X = \text{number of students that get their own assignment back}$. $X = X_1 + \cdots + X_n$ where $X_\omega = \text{1 \ (student } \omega \text{ gets his/her own assignment back)}$.

Using Linearity - 3: Binomial Distribution.
Flip $n$ coins with heads probability $p$. $X = \text{number of heads}$
Binomial Distribution: $Pr[X = i], \text{ for each } i$.
$$Pr[X = i] = \binom{n}{i} p^i (1 - p)^{n-i}.$$ 
Moreover $X = X_1 + \cdots + X_n$ and 
$$E[X] = \sum_i i \times Pr[X = i] = \sum i \times \binom{n}{i} p^i (1 - p)^{n-i}.$$ 

Note: If we had defined $Y = a_1 X_1 + \cdots + a_n X_n$ has had tried to compute $E[Y] = \sum x Pr[Y = x]$, we would have been in trouble!

Using Linearity - 1: Pips (dots) on dice
Roll a die $n$ times. $X_\omega = \text{number of pips on roll } \omega$. $X = X_1 + \cdots + X_n = \text{total number of pips in } n \text{ rolls}.$
$$E[X] = E[X_1 + \cdots + X_n]$$
$$= E[X_1] + \cdots + E[X_n], \text{ by linearity}$$
$$= nE[X_1], \text{ because the } X_\omega \text{ have the same distribution}$$
Now,
$$E[X_1] = \sum_{i=1}^6 \frac{i}{6} = 7 \cdot \frac{1}{6} = \frac{7}{2}$$
Hence,
$$E[X] = 7n \cdot \frac{1}{2}$$
Note: Computing $\sum x Pr[X = x]$ directly is not easy!

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

Note: What is $Pr[X = m]$? Tricky ...

Using Linearity - 3: Binomial Distribution.
Flip $n$ coins with heads probability $p$. $X = \text{number of heads}$
Binomial Distribution: $Pr[X = i], \text{ for each } i$.
$$Pr[X = i] = \binom{n}{i} p^i (1 - p)^{n-i}.$$ 
Moreover $X = X_1 + \cdots + X_n$ and 
$$E[X] = \sum_i i \times Pr[X = i] = \sum i \times \binom{n}{i} p^i (1 - p)^{n-i}.$$ 

Note: If we had defined $Y = a_1 X_1 + \cdots + a_n X_n$ has had tried to compute $E[Y] = \sum x Pr[Y = x]$, we would have been in trouble!
Let $Y = g(X)$. Assume that we know the distribution of $X$.

We want to calculate $E[Y]$. We have that $E[Y] = E[g(X)]$.

Let $g(X) = X^2$. Then $E[Y] = E[g(X)] = \sum_{x} x^2 \Pr[X = x]$.

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:

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

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

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 $a = E[X]$. 

An Example

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

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

$E[g(X)] = \sum_{x=2}^{6} x^2 \frac{1}{6} = \frac{19}{6}$.

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 2:

Let $g(a) := [(X - a)^2] = E[X^2 - 2aX + a^2] = E[X^2] - 2aE[X] + a^2$. 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 $a = E[X]$. 

Calculating $E[g(X)]$

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

We want to calculate $E[Y]$. We have that $E[Y] = E[g(X)]$.

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

$E[g(X)] = \sum_{x=2}^{6} x^2 \frac{1}{6} = \frac{19}{6}$.

An Example

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

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

$E[g(X)] = \sum_{x=2}^{6} x^2 \frac{1}{6} = \frac{19}{6}$.

Calculating $E[g(X)]$

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

We want to calculate $E[Y]$. We have that $E[Y] = E[g(X)]$.

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

$E[g(X)] = \sum_{x=2}^{6} x^2 \frac{1}{6} = \frac{19}{6}$.
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 $E[(X - a)^2]$ is the median of $X$.
The median $v$ of $X$ is any real number such that

\[ Pr[X \leq v] = Pr[X \geq v] \]

**Proof:**
\[ g(a) := E[(X - a)^2] = \sum_{x \in \mathbb{X}} (x-a)^2 \Pr(X=x) = a^2 \Pr(X=a) + \sum_{x \neq a} (x-a)^2 \Pr(X=x). \]

Thus, if $0 < \epsilon < 1$,
\[ g(a + \epsilon) = g(a) + \epsilon \Pr[X \leq a] - \epsilon \Pr[X \geq a]. \]

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

Best Guess: Illustration

Best Guess: Another Illustration

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,
\[ E[X] = \sum_{a \geq 0} a \Pr[X = a] \geq 0. \]

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

Example:
\[ B = \bigcup_{m=1}^{n} A_m \Rightarrow \mathbb{E}1_B(\omega) \leq \sum_{m} \mathbb{E}1_{A_m}(\omega) \Rightarrow Pr[\bigcup_{m} A_m] \leq \sum_{m} Pr[A_m]. \]

Center of Mass

The expected value has a center of mass interpretation:

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 $\{1, 2, ..., 6\}$. We say that $X$ is uniformly distributed in $\{1, 2, ..., 6\}$.

More generally, we say that $X$ is uniformly distributed in $\{1, 2, ..., n\}$ if $Pr[X = m] = 1/n$ for $m = 1, 2, ..., n$. In that case,
\[ E[X] = \sum_{m=1}^{n} m \Pr[X = m] = \sum_{m=1}^{n} m \cdot \frac{1}{n} = \frac{n(n+1)}{2} = \frac{n^2}{2}. \]
Let’s flip a coin with $P[H] = p$ until we get $H$. For instance:

$\omega_1 = H$, or $\omega_2 = TH$, or $\omega_3 = \ldots = 1, 2, \ldots$.

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

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

Geometric Distribution: Expectation

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

One has

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

Thus,

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

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

$pE[X] = p(1 - p)^0 + (1 - p)^1p + (1 - p)^2p + (1 - p)^3p + \cdots$

by subtracting the previous two identities

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

Hence,

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

Geometric Distribution: Memoryless

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

One has

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

Thus,

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

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

$pE[X] = p(1 - p)^0 + (1 - p)^1p + (1 - p)^2p + (1 - p)^3p + \cdots$

by subtracting the previous two identities

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

Hence,

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

Geometric Distribution: Expectation

Let $X \sim G(p)$. Then, for $n \geq 0$,

$P[X > n] = P[\text{first } n \text{ flips are } T] = (1 - p)^n$.

Theorem

$P[X > n + m | X > n] = P[X > m], m, n \geq 0$.

Proof:

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

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

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

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

$= P[X > m]$.

Geometric Distribution: Memoryless - Interpretation

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

For instance:

$\omega_1 = H$, or $\omega_2 = TH$, or $\omega_3 = \ldots = 1, 2, \ldots$.

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

Also, 

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

Note that

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

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

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

$aS = a + a^2 + a^3 + a^4 + \cdots$

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

Hence,

$\sum_{n=1}^{\infty} P[X_n] = P(X > m) = \frac{1}{1-(1-p)} = 1$.
**Geometric Distribution: Yet another look**

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

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

[See later for a proof.]

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

Hence,

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

---

**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 $n$.”

**Expected Value of Integer RV**

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

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

**Proof:** One has

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

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

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

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

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

$$= \Pr[X \geq 1].$$

---

**Poisson Distribution: Definition and Mean**

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

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

**Fact:** $\mathbb{E}[X] = \lambda$.

**Proof:**

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

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

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

---

**Simeon Poisson**

The Poisson distribution is named after:

![Siméon Denis Poisson (1781–1840)](image_url)
Equal Time: B. Geometric

The geometric distribution is named after:

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

Summary

Random Variables

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