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 $X^{-1}(a) := \{ \omega \in \Omega \mid X(\omega) = a \}$.
(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 $\{(a, Pr[X = a]) : a \in \mathscr{A}\}$,
where $\mathscr{A}$ is the range of $X$. That is, $\mathscr{A} = \{X(\omega), \omega \in \Omega\}$.

An Example

Flip a fair coin three times.

$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$.

$X =$ 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,

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

Expected winnings for heads/tails games, with 3 flips?

Recall the definition of the random variable $X$:

$\{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\} \rightarrow \{3, 1, 1, 1, 1, 1, 1, 0\}$.

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

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:

$$X_1 + \cdots + X_n \quad \text{when } n \gg 1.$$ 

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

![Average Dice Value Against Number of Rolls](image-url)
Linearity of Expectation

**Theorem:** Expectation is linear

\[ E[a_1X_1 + \cdots + a_nX_n] = a_1E[X_1] + \cdots + a_nE[X_n]. \]

**Proof:**

\[ E[a_1X_1 + \cdots + a_nX_n] = \sum_i (a_iX_i)Pr[\omega] = \sum_i (a_i)X_1(\omega) + \cdots + a_nX_n(\omega)Pr[\omega]. \]

\[ = a_1E[X_1] + \cdots + a_nE[X_n]. \]

Note: If we had defined \( Y = a_1X_1 + \cdots + a_nX_n \) has had tried to compute \( E[Y] = \sum_x yPr[X = x] \), we would have been in trouble!

Using Linearity - 1: Pips (dots) on dice

Roll a die \( n \) times.

\( X_m = \text{number of pips on roll } m, \)

\[ X = X_1 + \cdots + X_n = \text{total number of pips in } n \text{ rolls}. \]

\[ E[X] = E[X_1 + \cdots + X_n] = nE[X_1], \]

by linearity

\[ X_1 \text{ has the same distribution} \]

\[ = n \times 6 \times \frac{1}{2} \times \frac{1}{6} = 7. \]

Hence,

\[ E[X] = 7n/2. \]

Note: Computing \( \sum_x xPr[X = x] \) directly is not easy!

Using Linearity - 2: Random assignments Example

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

\[ E[X] = E[X_1 + \cdots + X_n] = \sum_i E[X_i] + \cdots + E[X_n], \]

by linearity

\[ = nE[X_1], \]

because all the \( X_m \) have the same distribution

\[ = nPr[X_1 = 1], \]

because \( X_1 \) is an indicator

\[ = n \times 1/n, \]

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.

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 \( i \).

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

**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 \( Y \):

\[ Pr[Y = y] = Pr[X \in g^{-1}(y)] \text{ where } g^{-1}(x) = \{x \in \mathbb{R} : g(x) = y\}. \]

This is typically rather tedious!

**Method 2:** We use the following result.

**Theorem:**

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

**Proof:**

\[ E[g(X)] = \sum_{x \in X} g(x)Pr[X = x] = \sum_{x \in X} \sum_{a \in G} g(x)Pr[\omega]. \]

\[ = \sum_{a \in G} \sum_{x \in X} g(x)Pr[\omega]. \]

\[ = \sum_{a \in G} g(x)Pr[\omega]. \]

\[ = \sum_{a \in G} \sum_{x \in X} Pr[X = x]. \]
An Example
Let $X$ be uniform in $\{-2,-1,0,1,2,3\}$. 
Let also $g(X) = X^2$. Then (method 2)

$$E[g(X)] = \sum_{x=2}^{3} x^2 \frac{1}{6} = \frac{4 + 1 + 0 + 1 + 4 + 9}{6} = \frac{19}{6}.$$ 

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

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

Thus,

$$E[Y] = \frac{2}{6} + \frac{2}{6} + \frac{0}{6} + \frac{1}{6} + \frac{9}{6} = \frac{19}{6}.$$ 

Center of Mass
The expected value has a center of mass interpretation:

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:
$$B = \bigcup_{m} A_m \Rightarrow 1_B(\omega) \leq \sum_m 1_{A_m}(\omega) \Rightarrow Pr[\cup_{m} A_m] \leq \sum_m Pr[A_m].$$

Summary
- A random variable $X$ is a function $X : \Omega \to \Re$.
- $Pr[X = a] = Pr[X^{-1}(a)] = Pr[\{\omega | X(\omega) = a\}]$.
- $Pr[X \in A] = Pr[X^{-1}(A)]$.
- The distribution of $X$ is the list of possible values and their probability: $\{(a,Pr[X = a]), a \in \Re^+\}$.
- $E[X] = \sum_a aPr[X = a]$.
- Expectation is Linear.