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, $X(\omega)$ 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

$\{(a, Pr[X = a]) : a \in \sigma'\}$,

where $\sigma'$ is the range of $X$. That is, $\sigma' = \{X(\omega), \omega \in \Omega\}$.

Some Distributions.

Binomial Distribution: $B(n, p)$, For $0 \leq i \leq n$, $Pr[X = i] = \binom{n}{i} p^i (1-p)^{n-i}$.

Geometric Distribution: $G(p)$, For $i \geq 1$, $Pr[X = i] = (1-p)^{i-1} p$.

Poisson: Motivation and derivation.

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 $n$”

McDonalds: How many arrive at McDonalds in an hour?

Know: average is $\lambda$.

What is distribution?

Example: $Pr[2 \lambda, \text{arrivals}]$?

Assumption: “arrivals are independent.”

Derivation: cut hour into $n$ intervals of length $1/n$.

$Pr[\text{two arrivals}]$ is $(\lambda/n)^2$ or small if $n$ is large.

Model with binomial.

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 $X < n$. For $m \ll n$ one has

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

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

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

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

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

Definition: The expected value (or mean, or expectation) of a random variable $X$ is

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

**Theorem:**

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

**Proof:**

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

$$ = \sum_a a \times \sum_{\omega : X(\omega) = a} Pr[\omega] $$

$$ = \sum_\omega X(\omega) Pr[\omega] $$

**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:**

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

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

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

**Recall: An Example**

Flip a fair coin three times.

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

$$ X = \text{number of } Hs: \{3, 2, 2, 1, 1, 0\}. $$

Thus,

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

Also,

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

**Simeon Poisson**

The Poisson distribution is named after:

Siméon Denis Poisson (1781–1840)

**Win or Lose.**

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

Recall the definition of the random variable $X:

$$ \{HHH, HHT, HTT, THH, THH, HTT, TTT\} \rightarrow \{3, 1, 1, -1, -1, -1, -3\}. $$

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

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.)

**Equal Time: B. Geometric**

The geometric distribution is named after:

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

Win or Lose.

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

Recall the definition of the random variable $X:

$$ \{HHH, HHT, HTT, THH, THH, HTT, TTT\} \rightarrow \{3, 1, 1, -1, -1, -1, -3\}. $$

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

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.)
Multiple Random Variables.

Experiment: toss two coins. \( \Omega = \{ HH, TH, HT, TT \} \).

\[
X_1(\omega) = \begin{cases} 
1 & \text{if coin 1 is heads} \\
0 & \text{otherwise} 
\end{cases} \quad X_2(\omega) = \begin{cases} 
1 & \text{if coin 2 is heads} \\
0 & \text{otherwise} 
\end{cases}
\]

Independent Random Variables.

Definition: Independence
The random variables \( X \) and \( Y \) are independent if and only if
\[
Pr(Y = b | X = a) = Pr(Y = b), \text{ for all } a \text{ and } b.
\]

Fact:
\( X, Y \) are independent if and only if
\[
Pr(X = a, Y = b) = Pr(X = a)Pr(Y = b), \text{ for all } a \text{ and } b.
\]

Follows from \( Pr[A \cap B] = Pr[A]Pr[B] \) (Product rule.)

Multiple Random Variables: setup.

Joint Distribution: \( \{(a,b) Pr(X = a, Y = b) : a \in \Omega, b \in \Omega\} \), where \( \Omega \) is possible values of \( Y \).

\[
\sum_{a \in \Omega, b \in \Omega} Pr(X = a, Y = b) = 1
\]

Marginal for \( X \): \( Pr(X = a) = \sum_{b \in \Omega} Pr(X = a, Y = b) \).
Marginal for \( Y \): \( Pr(Y = b) = \sum_{a \in \Omega} Pr(X = a, Y = b) \).

<table>
<thead>
<tr>
<th>X/Y</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>0</td>
<td>-2</td>
<td>3</td>
</tr>
<tr>
<td>Y</td>
<td>-3</td>
<td>2</td>
<td>-5</td>
<td>3</td>
</tr>
</tbody>
</table>

Conditional Probability: \( Pr(X = a | Y = b) = \frac{Pr(X = a, Y = b)}{Pr(Y = b)} \).

Review: Independence of Events

- Events \( A, B \) are independent if \( Pr[A \cap B] = Pr[A]Pr[B] \).
- Events \( A, B, C \) are mutually independent if \( A, B \) are independent, \( A, C \) are independent, \( B, C \) are independent and \( Pr[A \cap B \cap C] = Pr[A]Pr[B]Pr[C] \).
- Events \( \{A_n, n \geq 0\} \) are mutually independent if …
- Example: \( X, Y \in \{0,1\} \) two fair coin flips \( \rightarrow X, Y \) are pairwise independent but not mutually independent.
- Example: \( X, Y, Z \in \{0,1\} \) three fair coin flips are mutually independent.

Independence: Examples

Example 1
Roll two die. \( X, Y \) = number of pips on the two dice. \( X, Y \) are independent.
Indeed: \( Pr[X = a, Y = b] = \frac{3}{a} \times \frac{2}{b} = \frac{3}{a} \times \frac{2}{b} \).

Example 2
Roll two die. \( X \) = total number of pips, \( Y \) = number of pips on die 1 minus number on die 2. \( X \) and \( Y \) are not independent.
Indeed: \( Pr[X = 12, Y = 1] = 0 \neq Pr[X = 12]Pr[Y = 1] > 0 \).

Example 3
Flip a fair coin five times, \( X \) = number of HS in first three flips, \( Y \) = number of HS in last two flips. \( X \) and \( Y \) are independent.
Indeed:
\[
Pr[X = a, Y = b] = \left( \frac{3}{a} \right) \left( \frac{2}{3} \right)^{a-2} \times \left( \frac{2}{3} \right)^{2-b} = Pr[X = a]Pr[Y = b].
\]

Linearity of Expectation

Theorem:
\[
E[X + Y] = E[X] + E[Y]
\]
\[
E[cX] = cE[X]
\]
Proof: \( E[X] = \sum_{\omega \in \Omega} X(\omega) \times Pr[\omega] \).

\[
E[X + Y] = \sum_{\omega \in \Omega} (X(\omega) + Y(\omega))Pr[\omega]
\]
\[
= \sum_{\omega \in \Omega} X(\omega)Pr[\omega] + Y(\omega)Pr[\omega]
\]
\[
= \sum_{\omega \in \Omega} X(\omega)Pr[\omega] + \sum_{\omega \in \Omega} Y(\omega)Pr[\omega]
\]
\[
= E[X] + E[Y]
\]
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 \( X = 1_A \).

Using Linearity - 2: Fixed point.

Hand out assignments at random to n students.

\( X = \text{number of students that get their own assignment back}. \)

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

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

\[ E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \left( \frac{1}{6} \right) p(1-p)^{6-i}. \]

\[ E[X] = \sum_{i=1}^{7} E[X_i]. \]

Note: Computing \( \sum_{i=1}^{7} x \cdot Pr[X = x] \) directly is not easy!

Using Linearity - 3: Binomial Distribution.

Flip n coins with heads probability \( p \). \( X = \text{number of heads}. \)

Binomial Distribution: \( Pr[X = i] \), for each i.

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

\[ E[X] = \sum_{i=1}^{n} i \cdot Pr[X = i] = \sum_{i=1}^{n} \left( \frac{n!}{i!(n-i)!} \right) p^i (1-p)^{n-i}. \]

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

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

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

Note: What is \( Pr[X = m]? \) Tricky....

Using Linearity - 4

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, \( 1_{A \cap B}(\omega) = 1_A(\omega) + 1_B(\omega) - 1_{A \cup B}(\omega). \)

Taking expectation, we get

\[ Pr[A \cap B] = Pr[A] + Pr[B] - Pr[A \cup B]. \]

Observe that if \( Y(\omega) = b \) for all \( \omega \), then \( E[Y] = b \).

Thus, \( E[X + b] = E[X] + b \).
Empty Bins

Experiment: Throw \( m \) balls into \( n \) bins.

- \( Y \) - number of empty bins.
- Distribution is horrible.
- Expectation? \( X_i \) - indicator for bin \( i \) being empty.

\[ E(Y) = X_1 + \ldots + X_n. \]

For \( n = m \) and large \( n, (1 - 1/n)^m \approx 1/2. \)

\[ \frac{m}{n} \approx 0.368n \text{ empty bins on average}. \]

Coupon Collectors Problem.

Experiment: Get random coupon from \( n \) until get all \( n \) coupons.

Outcomes: \{123145...,56765... \}

Random Variable: \( X \) - length of outcome.

Today: \( E[X] \)?

Time to collect coupons.

- \( X \) - time to get \( n \) coupons.
- \( X_i \) - time to get first coupon. Note: \( X_1 = 1. \) \( E(X_1) = 1. \)
- \( X_2 \) - time to get second coupon after getting first.

\[ Pr[\text{"second coupon"} \cap \text{"first coupon"}] = \frac{1}{n}. \]

\[ E[X_2] = \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n^2}. \]

\[ E[X] = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \frac{1}{i}. \]

Geometric Distribution: Expectation

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

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

Hence, \( E[X] = \frac{1}{p}. \)

Review: Harmonic sum

\[ H(n) = 1 + \frac{1}{2} + \ldots + \frac{1}{n} = \int_1^n \frac{1}{x} \, dx = \ln(n). \]

A good approximation is

\[ H(n) \approx \ln(n) + \gamma \text{ where } \gamma \approx 0.58 \text{ (Euler-Mascheroni constant)}. \]

Harmonic sum: Paradox

Consider this stack of cards (no glue!):

If each card has length 2, the stack can extend \( H(n) \) to the right of the table. As \( n \) increases, you can go as far as you want!
Paradox

par·a·dox
/ pə-ˈrä-dəks/
noun
-a statement or proposition that, despite sound (or apparently sound) reasoning from acceptable premises, leads to a conclusion that seems senseless, logically unacceptable, or self-contradictory.

*a potentially serious conflict between quantum mechanics and the general theory of relativity known as the information paradox*

-a seemingly absurd or self-contradictory statement or proposition that when investigated or explained may prove to be well founded or true.

*a conclusion that is ultimately refuted by evidence*

-a situation, person, or thing that combines contradictory features or qualities.

The cards have width 2. Induction shows that the center of gravity after \( n \) cards is \( H(n) \) away from the right-most edge.

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.

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

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

Thus,

\[
E[Y] = \frac{4}{6}, \text{ w.p. } \frac{2}{6}, \text{ w.p. } \frac{1}{6}, \text{ w.p. } \frac{9}{6} = 19 \frac{1}{6}.
\]

Stacking

The cards have width 2. Induction shows that the center of gravity after \( n \) cards is \( H(n) \) away from the right-most edge.

Summary

Probability Space: \( \Omega, \Pr[\omega] \geq 0, \sum \Pr[\omega] = 1 \).

Random Variable: Function on Sample Space.

Distribution: Function \( \Pr[X = a] \geq 0, \sum \Pr[X = a] = 1 \).

Expectation: \( E[X] = \sum a \Pr[a] = \sum \Pr[X = a] \).

Many Random Variables: each one function on a sample space.

Joint Distributions: Function \( \Pr[X = a, Y = b] \geq 0, \sum \Pr[X = a, Y = b] = 1 \).


Applications: compute expectations by decomposing.

Indicators: Empty bins. Fixed points.

Time to Coupon: Sum times to “next” coupon.

\( Y = f(X) \) is Random Variable.

Distribution of \( Y \) from distribution of \( X \).