Random variables

Roll 2 dice, examine probability that sum = k, or = even.

Define random variable X: sum of two dice rolls.

So, \( X[(1,2)] = 3 \)
\( X[(5,5)] = 10 \)

Define random variable Y: parity of two dice rolls.

So, \( Y[(1,2)] = 1 \)
\( Y[(5,5)] = 0 \)
Think of a r.v. as a function, mapping sample space to whatever you like, usually a number.

Then we can express questions neatly:

\[ P(X < 3) = \frac{1}{36} \]
\[ P(Y = 1) = \frac{1}{2} \]

We can also eliminate absurd events, e.g., \[ P(X = 13) = 0 \]
Expected value = weighted average

\[ E(X) = \sum_y y \cdot P(X=y) \]

*over all possible values \( y \), compatible with \( X \).
\[ E(X) = \sum y \cdot P(X = y) \]

Example: roll 2 dice. \[ X = \text{difference between the 2} \]

Possible values of \( X \) → 0  1  2  3  4  5

# outcomes supporting value → 6 5.2 4.2 3.2 2.2 1.2

(for probability, divide by 36)

\[ E(x) = \frac{0 + 10 + 16 + 18 + 16 + 10}{36} \approx 1.944 \]
**EXPECTATION: PROPERTIES**

**LINEARITY OF EXPECTATION** *(important)*

\[ c_1, c_2 \in \mathbb{R} \quad E(c_1X + c_2Y) = c_1 \cdot E(X) + c_2 \cdot E(Y) \]

Generally,

\[ E(c_1X_1 + c_2X_2 + \cdots + c_nX_n) = c_1E(X_1) + c_2E(X_2) + \cdots + c_nE(X_n) \]

\[ E(\sum_c X_i) = \sum_c E(X_i) \rightarrow \text{Does NOT assume independence} \]

Independence:

\[ P(X=a \ & \ Y=b) = P(X=a) \cdot P(Y=b) \]

for all \( a, b \ldots \)

---

2 dice, A, B. \( X = \) result of A, \( Y = \) result of B, \( Z = X+Y \)

\[ E(Z) = E(X+Y) = E(X) + E(Y) = 2 \cdot 3.5 = 7 \]
**EXPECTATION : PROPERTIES**

\[ E(X + Y) = E(X) + E(Y) \]

Linearity of expectation doesn't assume independence

But \[ E(X \cdot Y) \neq E(X) \cdot E(Y) \] in general.

If \( X \) & \( Y \) are independent, then \( E(X \cdot Y) = E(X) \cdot E(Y) \)

However, \( E(X \cdot Y) = E(X) \cdot E(Y) \) does **NOT** imply \( X \) & \( Y \) are independent.
INDICATOR RANDOM VARIABLES

(taking value 0 or 1)

We already saw this: \( Y \): parity of rolling one die.

Another example: flip a coin 10 times.
\[ X = \text{#times we see pattern HT} \]

\[ E(X) = ? \]
INDICATOR RANDOM VARIABLES

Flip a coin 10 times. \( X = \# \) times we see pattern HT

HT could appear at flips 1 & 2, or 2 & 3, \ldots, or 9 & 10

Define r.v. \( X_i = \begin{cases} 1 & \text{if flips } i \text{ & } i+1 \text{ produce HT} \\ 0 & \text{otherwise} \end{cases} \)

\( X = X_1 + X_2 + \cdots + X_9 \)

\[
E(X) = E(X_1 + X_2 + \cdots + X_9) \\
= E(X_1) + E(X_2) + \cdots + E(X_9)
\]

\[
= 9 \cdot \frac{1}{4}
\]

Notice \( X_1 \) & \( X_2 \) are not independent. \( P(X_i=1) = \frac{1}{4} \)

\( P(X_1 \land X_2) = 0 \)

Linearity of expectation

\[
E(X_i) = 0 \cdot P(X_i=0) + 1 \cdot P(X_i=1) = \frac{1}{4}
\]
The hat-check problem (a.k.a. coat-check)

- n people leave their hats with an attendant, & get a ticket = number for retrieval.

- The attendant loses all ticket info & gives hats back randomly.

How many people do we expect to get their own hats back?
INDICATOR RANDOM VARIABLES

\[ X = \# \text{ people who get their own hat back} = \sum_{k=1}^{\hat{n}} X_k \]

\[ X_k = \begin{cases} 1 & \text{if person } k \text{ gets their own hat back} \\ 0 & \text{otherwise} \end{cases} \]

\[ \mathbb{E}(X_k) = \frac{1}{n} \quad \text{(random)} \]

\[ \mathbb{E}(X) = \mathbb{E}\left( \sum_{k=1}^{\hat{n}} X_k \right) = \sum_{k=1}^{\hat{n}} \mathbb{E}(X_k) \quad \text{linearity of expectation} \]

\[ = \sum_{k=1}^{\hat{n}} \frac{1}{n} = \left( \frac{\hat{n}}{n} \right) = 1 \]
The hiring problem: you need one assistant.

- $n$ candidates, interviewed in random order.
- No 2 equally skilled.
- Any time you interview someone better than all previous, you hire the new person & fire the current assistant.

How many people do you expect to hire?
INDICATOR RANDOM VARIABLES

\[ X = \# \text{ people you expect to hire} = \sum_{k=1}^{n} X_k \]

\[ X_k = \begin{cases} 
1 & \text{if you hire candidate } k \\
0 & \text{otherwise} 
\end{cases} \]

\[ E(X_k) = \frac{1}{k} \quad (k \text{ is hired iff better than all } k-1 \text{ previous}) \]

\[ E(X) = E\left( \sum_{k=1}^{n} X_k \right) = \sum_{k=1}^{n} E(X_k) \quad \text{linearity of expectation} \]

\[ = \sum_{k=1}^{n} \frac{1}{k} = \ln n + o(1) < \ln n + 1 \]
The birthday problem

How many people do we need in a room so that we expect to have (at least) one birthday match?
**INDICATOR RANDOM VARIABLES**

\[ X = \# \text{ birthday matches among } n \text{ people} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij} \]

\[ X_{ij} = \begin{cases} 1 & \text{if persons } i \& j \text{ match} \\ 0 & \text{otherwise} \end{cases} \]

\[ E(X_{ij}) = \frac{1}{365} \]

\[ E(X) = E\left( \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij} \right) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E(X_{ij}) \quad \text{(linearity of expectation)} \]

We said we want \( E[X] = 1 \)

\[ = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{365} = \frac{n \cdot (n-1)}{2} \cdot \frac{1}{365} = 1 \quad \Rightarrow n \approx 28 \]