Roll 2 dice. Define random variables X, Y

↳ X: sum of two dice. $X[(1,2)] = 3$ $X[(5,5)] = 10$

↳ Y: parity of sum. $Y[(1,2)] = 1$ (odd) $Y[(5,5)] = 0$ (even)

Think of these as functions, mapping sample space to a number.

e.g., $\{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6),$
$(2,1), (2,2), \ldots$
$\vdots$
$(6,1), (6,2), \ldots \quad \ldots (6,6)\}$

$X \to 2 \ldots 12$

$Y \to 0, 1$

Example usage: $P(X<3) = \frac{1}{36}$ $P(Y=1) = \frac{1}{2}$ $P(X=13) = 0$

Expected value = weighted average

$$E[X] = \sum_{*} y \cdot P(X=y)$$

$*$ over all possible values, y, that X could be.

---

example: roll 2 dice. $X = |\text{difference between the two}|$

possible values of X →	0	1	2	3	4	5
# outcomes per value →	6	$5 \cdot 2$	$4 \cdot 2$	$3 \cdot 2$	$2 \cdot 2$	$1 \cdot 2$

(for probability, divide by 36)

$$E[X] = \frac{0 + 10 + 16 + 18 + 16 + 10}{36} \sim 1.944$$

LINEARITY OF EXPECTATION $E[X+Y] = E[X] + E[Y]$

Generally, for $c_i \in \mathbb{R}$

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

$$E\left[\sum X_i\right] = \sum E[X_i]$$

Does NOT assume independence $\longrightarrow$ $P(X=a \;\&\; Y=b) = P(X=a)\cdot P(Y=b)$ for all a, b ...

---

2 dice, A, B. X = result of A. Y = result of B. $Z = X+Y$

$$E[Z] = E[X+Y] = E[X] + E[Y] = 2\cdot 3.5 = 7$$

1000 dice, expected value of sum $= 1000 \cdot 3.5$

EXPECTATION : PROPERTIES

$E[X+Y] = E[X] + E[Y]$ → always true

but $E[X \cdot Y] = E[X] \cdot E[Y]$ → NOT always true

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 (IRV)

(taking value 0 or 1)

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

Another example: flip a coin 10 times.

X = # 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, ..., or 9 & 10

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

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

$P(X_1 \wedge X_2) = 0$

$$X = X_1 + X_2 + \cdots + X_9$$

$$E[X] = E[X_1 + X_2 + \cdots + X_9]$$

linearity of expectation

$$= E[X_1] + E[X_2] + \cdots E[X_9]$$

$$E[X_i] = \cancel{0 \cdot P(X_i = 0) + } 1 \cdot \underbrace{P(X_i = 1)}_{} = \frac{1}{4}$$

$$= 9 \cdot \frac{1}{4}$$

INDICATOR RANDOM VARIABLES

The hat-check problem (a.k.a. coat-check)

- n people at a party leave their hats with an attendant
- The attendant gives hats back randomly.

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

The hat-check problem

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

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

$$E[X] = E\left[\sum_{k=1}^{n} X_k\right]$$

$$= \sum_{k=1}^{n} E[X_k]$$

linearity of expectation

$$= \sum_{k=1}^{n} \frac{1}{n} = 1$$

$$E[X_k] = \frac{1}{n}$$

(random)

$P(X_k = 1)$

The hiring problem: you need one assistant.

- n candidates, interviewed in <u>random</u> 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?

The hiring problem

X = # people you will hire $= \sum_{k=1}^{n} X_k$

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

$$E[X] = E\left[\sum_{k=1}^{n} X_k\right]$$

linearity of expectation

$$= \sum_{k=1}^{n} E[X_k]$$

$E[X_k] = \frac{1}{k}$ $\left\{\begin{array}{l} \text{person } k \text{ is hired} \\ \underline{\text{iff}} \text{ better than} \\ \text{all } k\text{-}1 \text{ previous} \end{array}\right.$

$$= \sum_{k=1}^{n} \frac{1}{k} \leq 1 + \ln n$$

The birthday problem

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

The birthday problem

X = # birthday matches among n people

For what n do we get $E[X] \geqslant 1$?

↳ Set up $E[X]$ as function of n

The birthday problem

X = # birthday matches among n people $= \underbrace{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n}}_{\text{all } \binom{n}{2} \text{ pairs}} X_{ij}$

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

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

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

we said we want $E[X]=1$

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

$\square$

$P(\geq 2$ people in a group of k have same birthday$)$

(new problem)

Not about expected value or IRV

$= 1 - P($nobody has the same birthday$)$

$$= 1 - \overbrace{\frac{365}{365}}^{\text{person \#1}} \cdot \overbrace{\frac{364}{365}}^{\text{person \#2}} \cdot \overbrace{\frac{363}{365}}^{\text{person \#3}} \cdot \;\cdots\; \cdot \overbrace{\frac{365-(k-1)}{365}}^{\text{person \#k}} = 1 - \frac{\frac{365!}{(365-k)!}}{365^k}$$

$P(\geq 2$ people in a group of k have same birthday$)$

$k=2 \rightarrow P \sim 0.27\% \left(\frac{1}{365}\right)$

$k=4 \rightarrow P \sim 1.64\%$

$k=23 \rightarrow P \sim 50.73\%$

$k=30 \rightarrow P \sim 70.6\%$

$k=70 \rightarrow P \sim 99.9\%$

$k \sim 116 \rightarrow P \sim 1 - \frac{1}{10^9}$

$k=300 \rightarrow P \sim 1 - \frac{1}{10^{80}}$

$(10^{80} \sim$ # atoms in universe$)$

$(k > 365 \rightarrow P = 1)$