RANDOM VARIABLES

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

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

↳ define random variable X : sum of two dice rolls.

So, $X[(1,2)] = 3$

$X[(5,5)] = 10$

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.

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.

$Y[(1,2)] = 1$

$Y[(5,5)] = 0$

Think of a r.v. as a function,

mapping sample space to $\underbrace{\text{whatever you like}}_{\text{usually a number}}$

Think of a r.v. as a function,

mapping sample space to $\underbrace{\text{whatever you like}}_{\text{usually a number}}$

Then we can express questions neatly: $P(X<3) = \frac{1}{36}$ ← sum (2 dice)

$P(Y=1) = \frac{1}{2}$ ← parity

We can also eliminate absurd events, e.g., $P(X=13)=0$

Expected value = weighted average

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

$*$ over all possible values y, compatible with X.

$$E(X) = \sum y \cdot P(X=y)$$

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

$$E(X) = \sum y \cdot P(X=y)$$

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

possible values of $X \rightarrow$ 0 1 2 3 4 5

$$E(X) = \sum y \cdot P(X=y)$$

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

possible values of X →	0	1	2	3	4	5
# outcomes supporting value →	6	?				

$$E(X) = \sum y \cdot P(X=y)$$

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

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

$$E(X) = \sum y \cdot P(X=y)$$

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

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

$$E(X) = \sum y \cdot P(X=y)$$

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

$$E(X) = \sum y \cdot P(X=y)$$

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

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

(for probability, divide by 36)

$$E(X) = \sum y \cdot P(X=y)$$

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

possible values of X →	0	1	2	3	4	5
# outcomes supporting value →	6	$5\cdot2$	$4\cdot2$	$3\cdot2$	$2\cdot2$	$1\cdot2$

(for probability, divide by 36)

$$E(X) = \ ?$$

$$E(X) = \sum y \cdot P(X=y)$$

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)

$$E(x) = \frac{0 + 10 + 16 + 18 + 16 + 10}{36}$$

(~ 1.944)

EXPECTATION : PROPERTIES

$$E(X + Y) = E(X) + E(Y)$$

EXPECTATION : PROPERTIES

LINEARITY OF EXPECTATION (important)

$c_1, c_2 \in \mathbb{R}$

$$E(c_1 X + c_2 Y) = c_1 \cdot E(X) + c_2 \cdot E(Y)$$

EXPECTATION : PROPERTIES

LINEARITY OF EXPECTATION (important)

$\hookrightarrow c_1, c_2 \in \mathbb{R}$

$$E(c_1 X + c_2 Y) = c_1 \cdot E(X) + c_2 \cdot E(Y)$$

Generally, $E(c_1 X_1 + c_2 X_2 + \cdots + c_n X_n) = c_1 E(X_1) + c_2 E(X_2) + \cdots c_n E(X_n)$

EXPECTATION : PROPERTIES

LINEARITY OF EXPECTATION (important)

$\hookrightarrow c_1, c_2 \in \mathbb{R}$ $\quad E(c_1 X + c_2 Y) = c_1 \cdot E(X) + c_2 \cdot E(Y)$

Generally, $E(c_1 X_1 + c_2 X_2 + \dots + c_n X_n) = c_1 E(X_1) + c_2 E(X_2) + \dots c_n E(X_n)$

$E\left(\sum c_i X_i\right) = \sum c_i E(X_i)$ $\longrightarrow$ Does NOT assume independence

EXPECTATION : PROPERTIES

LINEARITY OF EXPECTATION (important)

$c_1, c_2 \in \mathbb{R}$ $\quad E(c_1 X + c_2 Y) = c_1 \cdot E(X) + c_2 \cdot E(Y)$

Generally, $E(c_1 X_1 + c_2 X_2 + \cdots + c_n X_n) = c_1 E(X_1) + c_2 E(X_2) + \cdots c_n E(X_n)$

$E\left(\sum c_i X_i\right) = \sum c_i E(X_i)$ → Does NOT assume independence

Independence: $P(X=a \;\&\; Y=b) = P(X=a) \cdot P(Y=b)$

for all a,b ...

EXPECTATION : PROPERTIES

LINEARITY OF EXPECTATION (important)

$$E(c_1X + c_2Y) = c_1 \cdot E(X) + c_2 \cdot E(Y)$$

$c_1, c_2 \in \mathbb{R}$

Generally, $E(c_1X_1 + c_2X_2 + \cdots + c_nX_n) = c_1E(X_1) + c_2E(X_2) + \cdots c_nE(X_n)$

$$E\left(\sum c_iX_i\right) = \sum c_iE(X_i)$$

→ Does NOT assume independence

Independence: $P(X=a \;\&\; Y=b) = P(X=a) \cdot P(Y=b)$

for all a, b ...

---

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

$$E(Z) = E(X+Y) = E(X) + E(Y) = 2 \cdot 3.5 = 7$$

EXPECTATION : PROPERTIES

$$E(X+Y) = E(X) + E(Y)$$

Linearity of expectation doesn't assume independence

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 $E(X \cdot Y) = E(X) \cdot E(Y)$

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.

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 = # times we see pattern HT

$$E(X) = ?$$

Try solving this "directly"

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

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

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

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

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

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

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

$P(X_1 \wedge X_2) = 0$

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

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

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

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

$P(X_1 \wedge X_2) = 0$

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. $P(X_i=1)=\frac{1}{4}$

$P(X_1 \wedge X_2) = 0$

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

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

$$= E(X_1) + E(X_2) + \cdots E(X_9)$$

linearity of expectation

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. $P(X_i=1)=\frac{1}{4}$

$P(X_1 \wedge X_2) = 0$

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

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

$$= E(X_1) + E(X_2) + \cdots E(X_9)$$

linearity of expectation

$E(X_i) =$?

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. $P(X_i=1)=\frac{1}{4}$

$P(X_1 \wedge X_2) = 0$

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

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

$$= E(X_1) + E(X_2) + \cdots E(X_9)$$

linearity of expectation

$$E(X_i) = 0 \cdot P(X_i=0) + 1 \cdot P(X_i=1)$$

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. $P(X_i=1)=\frac{1}{4}$

$P(X_1 \wedge X_2) = 0$

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

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

$$= E(X_1) + E(X_2) + \cdots E(X_9)$$

linearity of expectation

$$E(X_i) = \cancel{0 \cdot P(X_i=0) + 1} \cdot \underline{P(X_i=1)}$$

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

$$= E(X_1) + E(X_2) + \cdots E(X_9)$$

$$E(X_i) = \cancel{0 \cdot P(X_i = 0)} \cancel{+ 1} \cdot \underline{P(X_i = 1)} = \frac{1}{4}$$

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

INDICATOR RANDOM VARIABLES

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

- n people 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 = # people who get their own hat back

INDICATOR RANDOM VARIABLES

X = # people who get their own hat back

$X_k = ?$

INDICATOR RANDOM VARIABLES

X = # people who get their own hat back

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

INDICATOR RANDOM VARIABLES

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

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

INDICATOR RANDOM VARIABLES

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

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

$E(X_k) = ?$

INDICATOR RANDOM VARIABLES

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

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

$E(X_k) = \frac{1}{n}$ (random)

INDICATOR RANDOM VARIABLES

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

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

$E(X_k) = \frac{1}{n}$ (random)

$E(X) = ?$

INDICATOR RANDOM VARIABLES

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

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

$$E(X_k) = \frac{1}{n}$$ (random)

$$E(X) = E\left(\sum_{k=1}^{n} X_k\right) = ?$$

INDICATOR RANDOM VARIABLES

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

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

$E(X_k) = \frac{1}{n}$ (random)

$E(X) = E\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} E(X_k)$ linearity of expectation

INDICATOR RANDOM VARIABLES

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

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

$E(X_k) = \frac{1}{n}$ (random)

$E(X) = E\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} E(X_k)$ linearity of expectation

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

INDICATOR RANDOM VARIABLES

The hiring problem: you need one assistant.

- n candidates, interviewed in <u>random</u> order.
- No 2 equally skilled.

INDICATOR RANDOM VARIABLES

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?

INDICATOR RANDOM VARIABLES

X = # people you will hire

INDICATOR RANDOM VARIABLES

X = # people you will hire

$X_k = ?$

INDICATOR RANDOM VARIABLES

X = # people you will hire = ?

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

INDICATOR RANDOM VARIABLES

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

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

INDICATOR RANDOM VARIABLES

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

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

$E(X_k) = \ ?$

INDICATOR RANDOM VARIABLES

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

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

$E(X_k) = \frac{1}{k}$ (k is hired <u>iff</u> better than all $k-1$ previous)

INDICATOR RANDOM VARIABLES

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

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

$E(X_k) = \frac{1}{k}$ (k is hired iff better than all k-1 previous)

$E(X) =$?

INDICATOR RANDOM VARIABLES

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

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

$E(X_k) = \frac{1}{k}$ (k is hired <u>iff</u> better than all k-1 previous)

$$E(X) = E\left(\sum_{k=1}^{n} X_k\right) = ?$$

INDICATOR RANDOM VARIABLES

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

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

$E(X_k) = \frac{1}{k}$ (k is hired iff better than all k-1 previous)

$E(X) = E\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} E(X_k)$ linearity of expectation

INDICATOR RANDOM VARIABLES

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

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

$$E(X_k) = \frac{1}{k}$$ (k is hired <u>iff</u> better than all $k-1$ previous)

$$E(X) = E\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} E(X_k)$$ linearity of expectation

$$= \sum_{k=1}^{n} \frac{1}{k} \quad \ldots ?$$

INDICATOR RANDOM VARIABLES

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

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

$$E(X_k) = \frac{1}{k}$$ (k is hired <u>iff</u> better than all $k-1$ previous)

$$E(X) = E\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} E(X_k)$$ linearity of expectation

$$= \sum_{k=1}^{n} \frac{1}{k} = \ln n + O(1) < \ln n + 1$$

INDICATOR RANDOM VARIABLES

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 = # birthday matches among n people

What should our I.R.V. be ? $X_?$

INDICATOR RANDOM VARIABLES

X = # birthday matches among n people

$X_{ij} = \ ?$

INDICATOR RANDOM VARIABLES

X = # birthday matches among n people

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

INDICATOR RANDOM VARIABLES

X = # birthday matches among n people = ?

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

INDICATOR RANDOM VARIABLES

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

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

INDICATOR RANDOM VARIABLES

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

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

$E(X_{ij}) = \ ?$

INDICATOR RANDOM VARIABLES

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

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

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

$$E(X) = \ ?$$

INDICATOR RANDOM VARIABLES

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

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

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

$$E(X) = E\left(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}\right) = \ ?$$

INDICATOR RANDOM VARIABLES

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

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

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

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

INDICATOR RANDOM VARIABLES

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

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

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

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

$$= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{365}$$

INDICATOR RANDOM VARIABLES

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

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

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

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

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

INDICATOR RANDOM VARIABLES

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

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

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

INDICATOR RANDOM VARIABLES

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

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

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

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

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

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

$\square$

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

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

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

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

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

person #1

$= 1 - \frac{365}{365}$

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

$$= 1 - P(\text{nobody has the same birthday})$$

$$= 1 - \overbrace{\frac{365}{365}}^{\text{person \#1}} \cdot \overbrace{\frac{364}{365}}^{\text{person \#2}}$$

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

$= 1 - P(\text{nobody has the same birthday})$

person #1 person #2 person #3

$= 1 - \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365}$

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

$$= 1 - P(\text{nobody has the same birthday})$$

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

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

$= 1 - P(\text{nobody has the same birthday})$

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

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

$k = 2 \rightarrow P \sim 0{,}27\%$ $\left(\frac{1}{365}\right)$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$P(\geq 2$ 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)

P(≥ 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)

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