Expectation $E[X] = \sum_{x \in \Omega(X)} x \cdot Pr(X=x) = \sum x \cdot P(x)$

R.V. X requirement values(X) = numbers

intuition

$E[X]$ = weighted average of values (x) weighted

By probabilities of seeing that value

$P(x)$

$P(X=x)$

uniform: $p(x)$ = constant in x $= \frac{1}{|\Omega(X)|} = \frac{1}{n}$

$n \hat{=}$ # of values

$$E[X] = \sum_x x \cdot P(x) = \frac{1}{n} \sum_x x =$$

$$= \frac{x_1 + x_2 + \dots + x_n}{n}$$

arithmetic average

E[ ] rules $\quad c =$ constant

$$E[c \cdot X] = c\,E[X]$$ exercise

$$E[c + X] = c + E[X]$$

X, Y 2 R.V with values = numbers (even DEPENDENT)

Linearity of Expectation

Th $E[X+Y] = E[X] + E[Y]$ $\quad X+Y = $ R.V

proof $E[X+Y] = \sum_{w \in \Omega(X+Y)} w \cdot P(X+Y=w) =$

$$\sum_{\substack{w = x+y \\ \text{values}}} (x+y)\,P(X+Y = x+y) =$$

$$= \sum_{x \in \Omega(x)} \sum_{y \in \Omega(Y)} (x+y)\,P(X=x, Y=y).$$

$$= \sum_x \sum_y x \cdot P(X,Y) + \sum_x \sum_y y \cdot P(X,Y)$$

$$= \sum_x x \sum_y P(X) \cdot P(Y|X) + \sum_y y \sum_x P(Y) \cdot P(X|Y)$$

$$= \sum_x \left[x \, P(x) \cdot \underbrace{\left(\sum_y P(Y|X=x) \right)}_{1} \right] + \sum_y \left[y \cdot P(y) \underbrace{\left(\sum_x P(X|Y=y) \right)}_{1} \right]$$

$$= \sum_x x \cdot P(x) + \sum_y y \cdot P(y)$$

$$= E[X] + E[Y]$$

- $E[X+Y+Z] = E[X+Y] + E[Z] = E[X] + E[Y] + E[Z]$
- $X_1, X_2, \ldots X_n$ indicator R.V $\to$ 1 "present"/T, 0 "not"

$X = X_1 + X_2 + \cdots + X_n$ = count of items present

$E[X] =$ expected # items present

$$= E[X_1 + X_2 + \cdots + X_n] = \sum_i E[X_i]$$

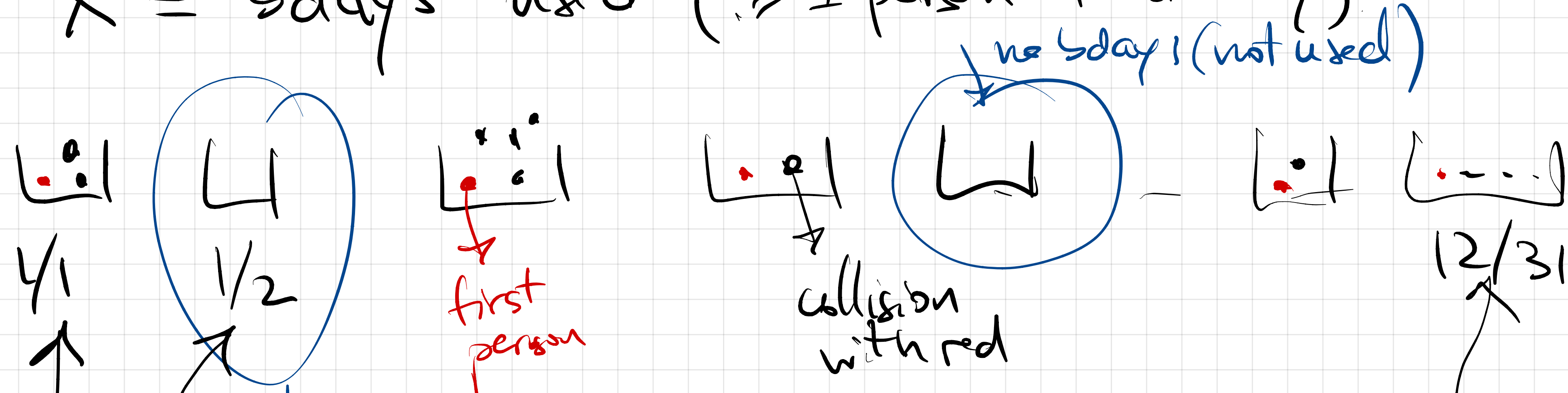
"expected number of ..."

Expected # of collisions? # 2^{nd} bdays in a day with already a first bday.

ex. bday with 3 people → 2 collis

10 people → 9 collis.

X = bdays used (≥ 1 person that day)



RV X_1 1/0 used X_2 1/0 X_3 1/0 — — X_{364} X_{365} 1/0

$X_i = RV$ → 1 if bday i is used (at least 1 person)

→ 0 if nobody has that bday.

$$X = X_1 + X_2 + \dots + X_n$$

$$E[X] = E[X_1 + X_2 \dots + X_n] = E[X_1] + E[X_2] \dots + E[X_n]$$

$E[X_i] = E[X_j]$ (all have the same expectation)

$$E[X_1] = \text{prob}(X_1 = 1) \cdot 1 + \text{prob}(X_1 = 0) \cdot 0 =$$

$$= \text{prob}(X_1 = 1)$$

nobody has X_1 bday

$$P(X_1 = 1) = 1 - \text{prob}(X_1 = 0) =$$

$$1 - \left(1 - \frac{1}{365}\right)^{n}$$

n people

probab of 1 person not have that bday X_1

X, Y R.V

$$E[X \cdot Y] = \sum_{w = x \cdot y} x \cdot y \ \text{Prob}(X \cdot Y = w)$$

$$= E[X] \cdot E[Y] \text{ if } X, Y \text{ independent}$$

$$\neq E[X] \cdot E[Y] \text{ in general } X, Y \text{ might be dependent.}$$

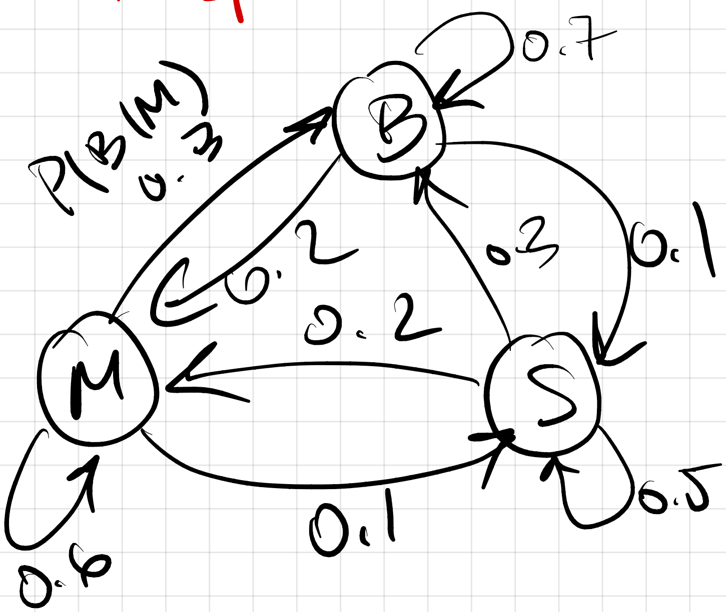
Markov Chains Example: Population in Boston

3 restaurants B, M, S "states"

$$\Pr\left(\underset{\text{restaurant}}{B} \longrightarrow \underset{\text{restaurant}}{M} \right)$$ "transitions"

next day

Markov Chain: transitions are fixed; no memory (history) except last state.



Transition table (matrix)

	B	M	S
B	0.7	0.2	0.1
M	0.3	0.6	0.1
S	0.3	0.2	0.5

→ sum to 1

· Start the first day with a distribution of people

day 0 eating at B, M, S $\pi_0 = (\pi_0^B \ \ \pi_0^M \ \ \pi_0^S)$

usually uniform $\pi_0 = \left(\frac{1}{3} \quad \frac{1}{3} \quad \frac{1}{3}\right)$

- each day we update π according to transition P

	B	M	S
day 1	$\pi_0^B \cdot P(B \to B) + \pi_0^M \cdot P(M \to B) + \pi_0^S \cdot P(S \to B)$	$\pi_0^B \cdot P(B \to M) + \pi_0^M P(M \to M) + \pi_0^S P(S \to M)$	$\pi_0^M \times 0.1 + \pi_0^B \times 0.1$
	$\boxed{\pi_0^B \times 0.7 + \pi_0^M \times 0.3 + \pi_0^S \times 0.3}$	$\pi_0^B \times 0.2 + \pi_0^M \times 0.6 + \pi_0^S \times 0.2$	$+ \pi_0^S \times 0.5$

day$_{k+1}$

$$\underset{1\times 3}{\pi_{k+1}} = \underset{1\times 3}{\pi_k} \cdot \underset{3\times 3}{\underline{P}}$$

$$\begin{pmatrix} \pi_{k+1}^B \\ \pi_{k+1}^M \\ \pi_{k+1}^S \end{pmatrix} = \begin{pmatrix} \pi_k^B \\ \pi_k^M \\ \pi_k^S \end{pmatrix} \cdot \begin{bmatrix} & & \\ & 3\times 3 & \\ & & \end{bmatrix}$$

Markov Th — Most of the time (excluding particular cond)

$\Rightarrow$ STATIONARY DISTRIBUTION

$$\lim_{n \to \infty} \pi_n = \pi^*$$

$$\pi^* = \pi^* \cdot P$$ (convergent by transition P)

$B = \pi_*^B \quad M = \pi_*^M \quad S = \pi_*^S$

$$B = 0.7 \times B + 0.3 \times M + 0.3 \times S$$
$$M = 0.2 \times B + 0.6 \times M + 0.2 \times S$$
$$S = 0.1 \times B + 0.1 \times M + 0.5 \times S$$

Linear 3 eq
3 vars

Solve $\Rightarrow M = \frac{1}{3} \quad B = \frac{1}{2} \quad S = \frac{1}{6}$

$$\pi_* = \left(\tfrac{1}{2}(B) \quad \tfrac{1}{3}(M) \quad \tfrac{1}{6}(S) \right)$$

coin $\Pr(\text{Head}) = p$ $\Pr(\text{Tail}) = 1-p$ Bernoulli $\to$ Binomial distribution

$n = \#$ flips INDEP $X = \#$ of heads in n flips.

k = target

① $P[X=k]$

H/T H/T H/T ... H/T

F_1 F_2 F F_n

exact k heads $\iff$ k flips are "Heads" $n-k$ flips are "tails"

choose k flips $\to$ "heads" $\binom{n}{k}$ sequences H/T with k heads

prob for a fix sequence

example

k heads H T T H . . . H T . . . T

prob $\boxed{p\,(1-p)(1-p)\,p \quad p\,(1-p) \quad (1-p)}$ flips are indep

$p^k (1-p)^{n-k}$

$$P[X=k] = \binom{n}{k} \cdot p^k (1-p)^{n-k}$$

prob of exact k heads and exact $n-k$ tails

Sanity check: $1 = \sum_{k=0}^{n} \underbrace{P[X=k]}_{\text{binomial dist}} = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p+(1-p))^n$

1

$x = p \; ; \; y = 1-p$

② $E[X] = ?$

def: $\sum_{\substack{0 \le k \le n \\ \text{heads}}} k \cdot P[X=k] = \sum_k k \binom{n}{k} p^k (1-p)^{n-k} = \sum_k k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$

$$= np \sum_k \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k}$$

$$= np \sum_{0 \le k-1 \le n-1} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)}$$

$[p+(1-p)]^{n-1}$

exercise

$$= np$$

② $E[X] = E[X_1 + X_2 + \cdots + X_n]$ 1 binary RV per flip

$$X_i = \begin{cases} 1 & \text{if flip } i = H \\ 0 & \text{if flip } i = T \end{cases}$$

$$= n \cdot E[X_1] = n \cdot \text{prob}(X_1 = 1) = n \cdot p$$

$P(H) = p$

Heads = Success

Flip coin every day

Tail = Failure

$P(\text{Tails}) = 1-p$

R.V. X = # flips until first success

distribution(x) ?

$E[x] = ?$

$$P(x=1) = p$$

$$P(x=2) = (1-p)\cdot p$$

$$P(x=3) = (1-p)(1-p)\cdot p = (1-p)^2\cdot p$$

$$P(x=k) = \underbrace{(1-p)\ldots(1-p)}_{k-1 \text{ times}}\cdot \underbrace{p}_{\text{succes}} = (1-p)^{k-1}\cdot p$$

$$P(x = 10^8)$$

Sanity Check: $\sum_{k=1}^{\infty} P[X=k] = \sum_{k=1}^{\infty} p(1-p)^{k-1}$

$\sum$ probab

$$= p \sum_{k=1}^{\infty} (1-p)^{k-1} \qquad x = 1-p$$

$$= p\cdot\left(\underbrace{x^0 + x^1 + x^2 + \dots x^{108} + \dots}_{} \right)$$

$$= p \cdot \frac{1}{1-x}$$

$$= p \cdot \frac{1}{1-(1-p)} = p \cdot \frac{1}{p} = 1$$

$E[X]$ = expected # of trials until first succes
(average)

$V =$

$$= \sum_{k=0}^{\infty} k \cdot P[X=k] = \sum_{k=1}^{\infty} k \cdot p(1-p)^{k-1}$$

Geometric

TELESCOPE V summation

$$V(1-p) = \sum_{k=1}^{\infty} k \cdot p(1-p)^{k}$$

$$V - V(1-p) = \underset{k=1}{p} + \sum_{k=2}^{\infty} kp(1-p)^{k-1} - \sum_{k=2}^{\infty} (k-1)p(1-p)^{k-1}$$

indexed by $k-1$

$$= p + \sum_{k=2}^{\infty} [\underbrace{kp - (k-1)p}_{p}](1-p)^{k-1}$$

$$= p + \sum_{k=2}^{\infty} p(1-p)^{k-1}$$

$$= \sum_{k=1}^{\infty} p(1-p)^{k-1} = 1$$ from before

So $v - v(1-p) = 1$

$$vp = 1$$

$$E[X] = v = \frac{1}{p}$$

Example $p = \text{prob}(\text{succes}) = \frac{1}{5}$

$\Rightarrow E[\#\text{ trials to succes}] = \frac{1}{1/5} = 5$

Balls into Bins — m balls are indep thrown into one of the k bins, uniformly.

Statistics

① What is the expected # balls falling into a bin?

② What is the expected # of empty bins?

① Define R.V. $X_{ij} = \begin{cases} 1 & \text{if ball } i \to \text{bin } j \\ \boxed{0 \text{ if not}} \end{cases}$ Total $m \times k$ R.V.

$$E[\#\text{ of balls in bin } t] = E\Big[\underbrace{\sum_{i=1}^{m} X_{it}}_{\text{"1" for every ball} \to \text{bin } t}\Big] = \sum_{i=1}^{m} E[X_{it}] = m \cdot \frac{1}{k}$$

pick a t $1 \le t \le k$

$$E[X_{it}] = 1 \cdot p(X_{it}=1) + 0 \cdot pr(X_{it}=0) = p(X_{it}=1) = 1/k$$

prob of ball $i \to$ bin t unif over k bins

② Fix a bin t $1 \le t \le k$.

R.V. $Y_t = \begin{cases} 1 & \text{if all balls miss bin } t \Rightarrow \text{empty} \\ 0 & \text{if not} \Rightarrow \text{not empty.} \end{cases}$

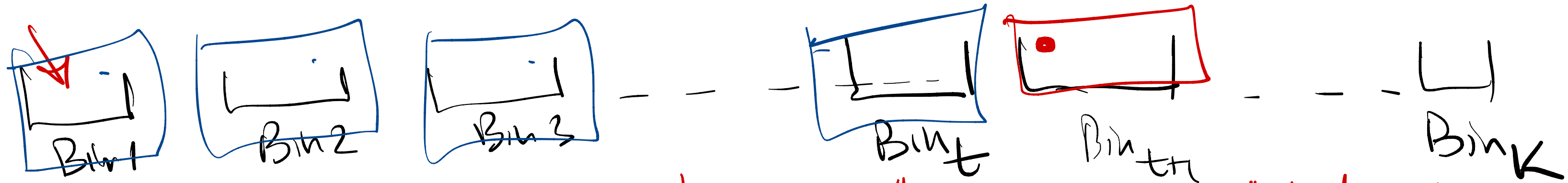
$$P[Y_t = 1] \underset{\text{indep}}{\overset{m \text{ throws}}{=}} \prod_{i=1}^{m} \boxed{\begin{matrix}\text{ball } i \\ \text{miss bin } t\end{matrix}} = \prod_{i=1}^{m} Pr[X_{it}=0] = \prod_{i=1}^{m} \left(1 - \frac{1}{k}\right)$$

$$= \left(1 - \frac{1}{k}\right)^m.$$

$$E[\#\text{empty bins}] = E\Big[\sum_{t=1}^{k} Y_t\Big] = \boxed{\sum_{t=1}^{k} E[Y_t] = k \cdot \left(1 - \frac{1}{k}\right)^m} \sim k \cdot e^{-m/k}$$

optional: "Coupon Collector PB"

(3*) Say k if fixed (#bins), but m is a random variable: we throw m balls until all the bins are non-empty. What is the $E[m]$?



Bins are now in "hit event order" → coupon collected order.

- First ball → Bin 1. — #balls going to bin 1 X_1
- Next bin that gets its first ball Bin 2 — #balls until we hit bin 3 X_2
- Next empty bin gets hit Bin 3 — #balls until we hit bin 4 X_3
- Next Bin 4

\- - - -

Bin t — #balls until we hit (first time) bin t X_t

Bin t+1 — # balls until we hit bin t+1 X_{t+1}

X_t = #balls after hitting bin t (first time) until we hit bin t+1 (first time)

$P(X = \text{ball})$ to hit bin t+1 $= 1 - P(X = \text{ball}$ falls into one of the non-empty bins$)$

$= 1 - \frac{t}{k}$ "chance of success" fixed

$\Longrightarrow$ geometric distribution (exercise)

$E[X_t] = E[X = \text{geometric dist, success} = 1 - \frac{t}{k}]$

$x \in [0,1]$

$1 + x + x^2 + x^3 + x^4 \ldots = \frac{1}{1-x}$

$$= \frac{1}{\text{success rate}} = \frac{1}{1 - t/k} = \frac{k}{k - t + 1}$$

$$\# \text{balls} = 1 + \sum_{t=1}^{k} X_t = \begin{array}{l} \#\text{balls till bin } 1 \rightarrow 1 \\ \#\text{balls till bin } 2 \\ \vdots \\ \#\text{balls till bin } k \end{array}$$

Harmonic H_k

$$= \sum_{t=1}^{k} \frac{k}{k-t+1} = k \sum_{t=1}^{k} \frac{1}{k-t+1} = k \cdot \sum_{t=1}^{k} \frac{1}{t} \approx k \cdot H_k$$

$$H_k = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{k}$$

Variance (R.V X) $= var(X) = avg(\text{distance-to-mean})^2$

$= avg(X - \underbrace{E[X]}_{mean})^2 = E[(X - E[X])^2]$

$E[X] =$ not random value $= \mu$ $\qquad X =$ R.V. random

$var(X) = E[(X - \mu)^2] = E[X^2 - 2X\mu + \mu^2] =$

$= E[X^2] - 2\mu \underbrace{E[X]}_{\mu} + \underbrace{\mu^2}_{E[\mu^2]}$

$= E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - \underbrace{\mu^2}_{E[X]}$

$= E[X^2] - (E[X])^2$

var(die roll)
X

$$\mu = E[X] = 3.5 = \frac{7}{2}$$

$$= \operatorname{avg}\left((X-\mu)^2\right) =$$

$$\frac{1}{6}\left((1-\tfrac{7}{2})^2+(2-\tfrac{7}{2})^2+(3-\tfrac{7}{2})^2+(4-\tfrac{7}{2})^2+(5-\tfrac{7}{2})^2+(6-\tfrac{7}{2})^2\right)$$

$$= \frac{70}{24} \; ?$$

Coin $p(\text{Heads}) = p = 2/3$ $\quad p(\text{tails}) = 1-p = 1/3$

$X = 1$ if heads

$$\mu = E[X] = 1 \cdot \frac{2}{3} + 0 \cdot \frac{1}{3} = \frac{2}{3} =$$

$$\operatorname{var}[X] = \left(1-\frac{2}{3}\right)^2 \cdot p(X=1) + \left(0-\frac{2}{3}\right)^2 \cdot p(X=0) =$$

$$= \left(\frac{1}{3}\right)^2 \cdot \frac{2}{3} + \left(\frac{2}{3}\right)^2 \cdot \frac{1}{3} = \frac{6}{27}$$

$$
\begin{aligned}
\mathsf{E}(XY) &= \sum_{\omega\in\Omega} X(\omega)Y(\omega)\mathsf{Pr}(\omega) \\
&= \sum_x \sum_y xy \cdot \mathsf{Pr}(X = x \text{ and } Y = y) \\
&= \sum_x \sum_y xy \cdot \mathsf{Pr}(X = x)\mathsf{Pr}(Y = y) \\
&= \left(\sum_x x \cdot \mathsf{Pr}(X = x)\right)\left(\sum_y y \cdot \mathsf{Pr}(Y = y)\right) \\
&= \mathsf{E}(X)\mathsf{E}(Y),
\end{aligned}
$$

$$\mathsf{Var}(X) = \mathsf{E}(X^2) - (\mathsf{E}(X))^2.$$

Consequently, $\mathsf{Var}(X) \leq \mathsf{E}(X^2)$.

Proof. Using the linearity of expectation and the fact that the expectation of a constant is itself, we have

$$\begin{aligned}
\mathsf{Var}(X) &= \mathsf{E}(X - \mathsf{E}(X))^2 \\
&= \mathsf{E}(X^2 - 2X\mathsf{E}(X) + (\mathsf{E}(X))^2) \\
&= \mathsf{E}(X^2) - 2\mathsf{E}(X)\mathsf{E}(X) + (\mathsf{E}(X))^2 \\
&= \mathsf{E}(X^2) - (\mathsf{E}(X))^2
\end{aligned}$$

Proposition 6.11. *Given a discrete probability space* (Ω, Pr)*, for any random variable X and Y, if X and Y are independent, then*

$$\mathsf{Var}(X + Y) = \mathsf{Var}(X) + \mathsf{Var}(Y).$$

Proof. Recall from Proposition 6.9 that if X and Y are independent, then $\mathsf{E}(XY) = \mathsf{E}(X)\mathsf{E}(Y)$. Then, we have

$$\begin{aligned}
\mathsf{E}((X+Y)^2) &= \mathsf{E}(X^2 + 2XY + Y^2) \\
&= \mathsf{E}(X^2) + 2\mathsf{E}(XY) + \mathsf{E}(Y^2) \\
&= \mathsf{E}(X^2) + 2\mathsf{E}(X)\mathsf{E}(Y) + \mathsf{E}(Y^2).
\end{aligned}$$

Using this, we get

$$\begin{aligned}
\mathsf{Var}(X+Y) &= \mathsf{E}((X+Y)^2) - (\mathsf{E}(X+Y))^2 \\
&= \mathsf{E}(X^2) + 2\mathsf{E}(X)\mathsf{E}(Y) + \mathsf{E}(Y^2) - ((\mathsf{E}(X))^2 + 2\mathsf{E}(X)\mathsf{E}(Y) + (\mathsf{E}(Y))^2) \\
&= \mathsf{E}(X^2) - (\mathsf{E}(X))^2 + \mathsf{E}(Y^2) - (\mathsf{E}(Y))^2 \\
&= \mathsf{Var}(X) + \mathsf{Var}(Y),
\end{aligned}$$