Monte-Carlo Integration Review

$$X \sim f(x)$$

$$E(h(X)) = \int h(x) f(x)\, dx = I$$

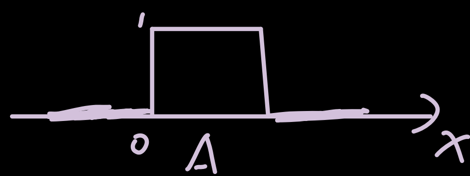
$$X_1, X_2, \dots, X_n \stackrel{iid}{\sim} f(x)$$

iid: independent and identically distributed

$$\hat{I} = \frac{1}{n} \sum_{i=1}^{n} h(X_i)$$

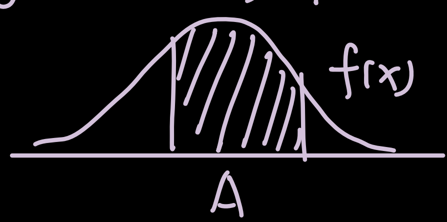
A special Case:

$$h(x) = 1\,(x \in A)$$



$$I = E(h(X)) = \int 1(x \in A)\, f(x)\, dx$$

=



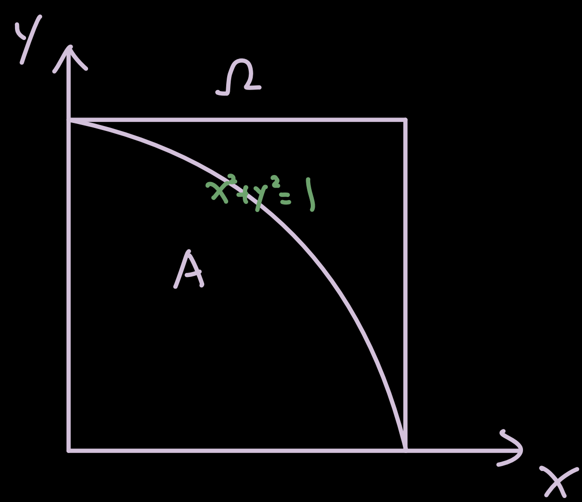
$$= P(A)$$

$\hat{I} = \frac{1}{n} \sum_{i=1}^{n} 1(X_i \in A)$ = frequency

Intuition: population $f(x)$ $\longrightarrow$ Sample $X_1 \cdots X_n$

indicator case: average (proportion) $\xleftarrow{\text{estimate}}$ average (proportion / frequency)

Example: Estimating π



$$E(1(X^2+Y^2 \le 1))$$

$$= \int_0^1 \int_0^1 1(x^2+y^2 \le 1)\, dx dy$$

$$= P(\underbrace{X^2+Y^2 \le 1}_{A})$$

$$= I$$

$$= P(A)$$

$$= \frac{|A|}{|\Omega|}$$

$$= \frac{\pi}{4}$$

$$(X, Y) \sim \underbrace{\mathrm{Unif}[0,1]^2}_{\Omega}$$

So we draw:

$$(X_i, Y_i) \overset{iid}{\sim} \text{Unif}(\Omega)$$

$$i = 1, 2, \ldots, n = 10.000$$

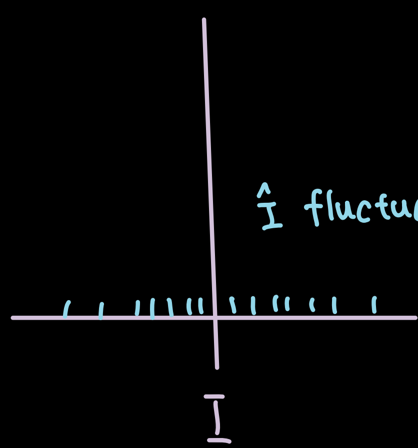
$$\hat{I} = \frac{1}{n} \underbrace{\sum_{i=1}^{n} 1(X_i^2 + Y_i^2 \leq 1)}_{m} = \frac{m}{n}$$

$$\hat{\pi} = \frac{4m}{n}$$

Statistical properties of I

$$E(\hat{I}) = I \quad ?$$

$$\text{Var}(\hat{I}) = ?$$

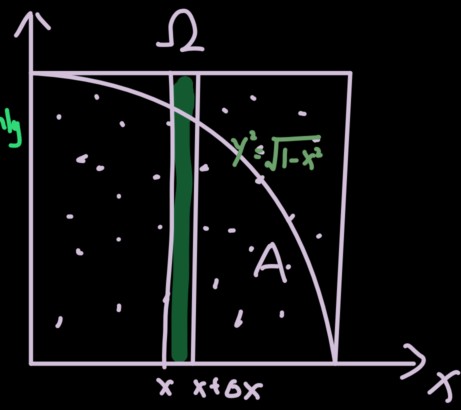


$\hat{I}$ fluctuates around actual I value.
we want $\text{var}(\hat{I})$ to be as small as possible.

Variance Reduction Methods

(1) Conditioning

Points in ▮ are uniformly distributed.



$$E(1(X^2+Y^2 \le 1)) = \frac{\pi}{4}$$

$$E(1(X^2+Y^2 \le 1) \mid X = x)$$

$$= P(X^2+Y^2 \le 1 \mid X = x)$$

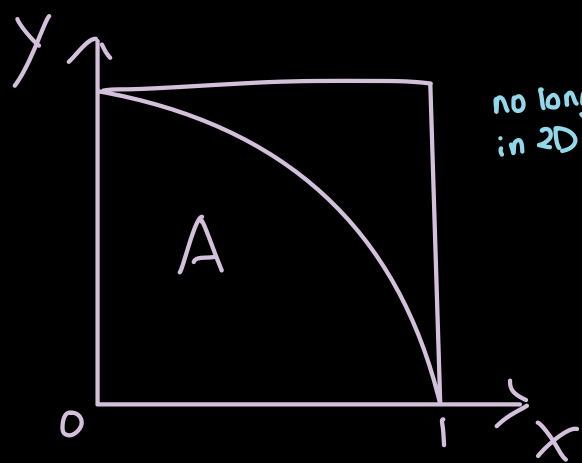
$$= \sqrt{1-x^2}$$

Expectation of conditional probability.

Do not use MC when you can calculate deterministic value.

So we approximate the integral of this expectation.

$$I = \int_0^1 \sqrt{1-x^2}\, dx$$



no longer in 2D

get rid of ~~randomness~~ fluctuation in Y

$$X_1 \dots X_n \overset{iid}{\sim} \text{Unif}[0,1]$$

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} \sqrt{1-X_i^2}$$

$$h(x) = \sqrt{1-x^2}$$

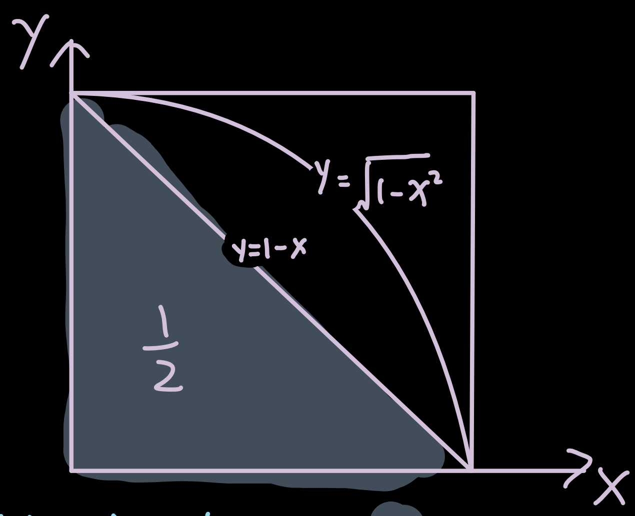
$$f(x) = 1 \quad (x \in [0,1])$$

$$I_1 = \int h(x) f(x)\, dx$$

$$= \underline{E(\sqrt{1-x^2})}$$

Conditioning x such that I does not involve Y

Control Variance (contrast)



We already know so do not need to calculate it → reduce some variability

reduce fluctuations

$$I = \frac{1}{2} + \int_0^1 \left(\sqrt{1-x^2} - (1-x)\right) dx$$

- $h(x) = \sqrt{1-x^2} - (1-x)$
- $f(x) = 1 \ (x \in [0, 1])$

$$I = \frac{1}{2} + E(h(X))$$

$$\hat{I} = \frac{1}{2} + \frac{1}{n}\sum_{i=1}^{n}\left(\sqrt{1-X_i^2} - (1-X_i)\right)$$

If $X_i \sim \text{Unif}[0,1]$ and

$1 - X_i \sim \text{Unif}[0, 1]$

Antithetic variables

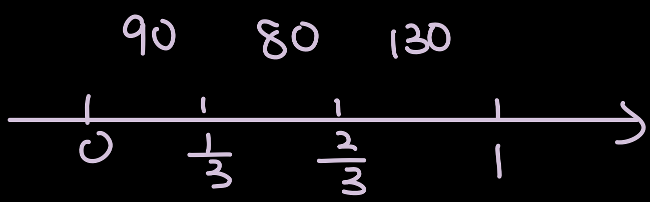
If this is close to 1

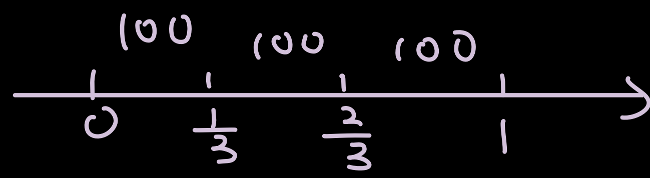
then this is close to 0, canceling out randomness

$$\hat{I} = \frac{1}{2n}\sum_{i=1}^{n}\left(\sqrt{1-X_i^2} + \sqrt{1-(1-X_i)^2}\right)$$

Stratified Sampling

300 points





possible division from fluctuation

So we control # of points in each interval to reduce fluctuation

Stratify interval $[0, 1]$ into

$$X_1, X_2, \ldots X_{1000} \sim \text{Unif}\left[0, \tfrac{1}{10}\right]$$

$$X_{1001}, X_{1002}, \ldots X_{2000} \sim \text{Unif}\left[\tfrac{1}{10}, \tfrac{2}{10}\right]$$

$$\vdots \qquad \vdots$$

$$X_{9001}, X_{9002}, \ldots X_{10000} \sim \text{Unif}\left[\tfrac{9}{10}, 1\right]$$

$$\hat{I} = \frac{1}{10{,}000} \sum_{i=1}^{10000} \sqrt{1 - X_i^2}$$

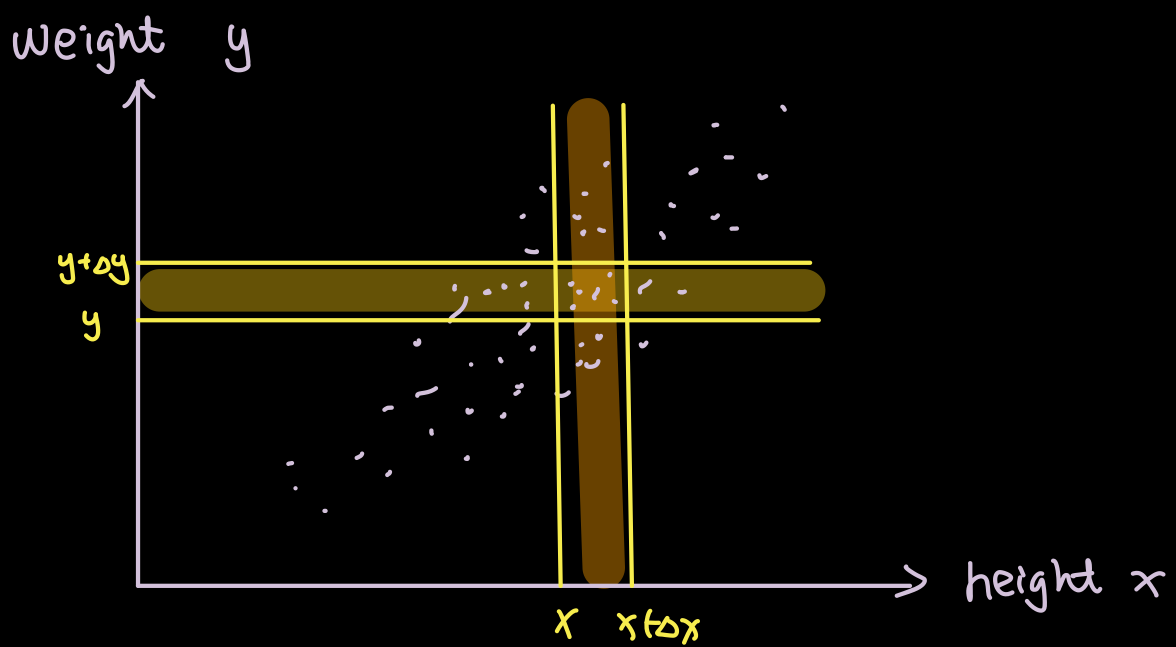
reduce variability by taking average of subintegrals

Background on Probability

$E(\hat{I}) = I$?

$Var(\hat{I}) =$?

Multivariate density



Population of $N (\approx \infty)$ points $(2D, (x, y))$

$N(x,y)$: # of points in $(x, x+\Delta x) * (y, y+\Delta y)$

$$f(x,y) = \frac{N(x,y)/N}{\Delta x \Delta y}$$ ← proportion of ppl in small cell

$$P((X, Y) \in (x, x+\Delta x) * (y, y+\Delta y))$$

$$= \frac{N(x,y)}{N}$$ axiom 0

density function $f(x,y) = \dfrac{P((X,Y) \in (x, x+\Delta x) * (y, y+\Delta y))}{\Delta x \Delta y}$ → area

$\dfrac{\text{probability mass}}{\text{area}}$

Population Average

$$\frac{1}{N} \sum_{i=1}^{n} h((X_i, y_i))$$

$$= \sum_{\substack{\text{cells} \\ (x, x+\Delta x) * (y, y+\Delta y)}} h(x,y) \frac{N(x,y)}{N}$$ # of ppl in cells

$$= \sum h(x,y) f(x,y) \Delta x \Delta y$$

$$\xrightarrow{\Delta x \Delta y \to 0} = \iint h(x,y) f(x,y) \, dx \, dy = E(h(X,Y))$$

$$Var(h(X,Y)) = E\left((h(X,Y) - E(h(X,Y)))^2\right)$$

same

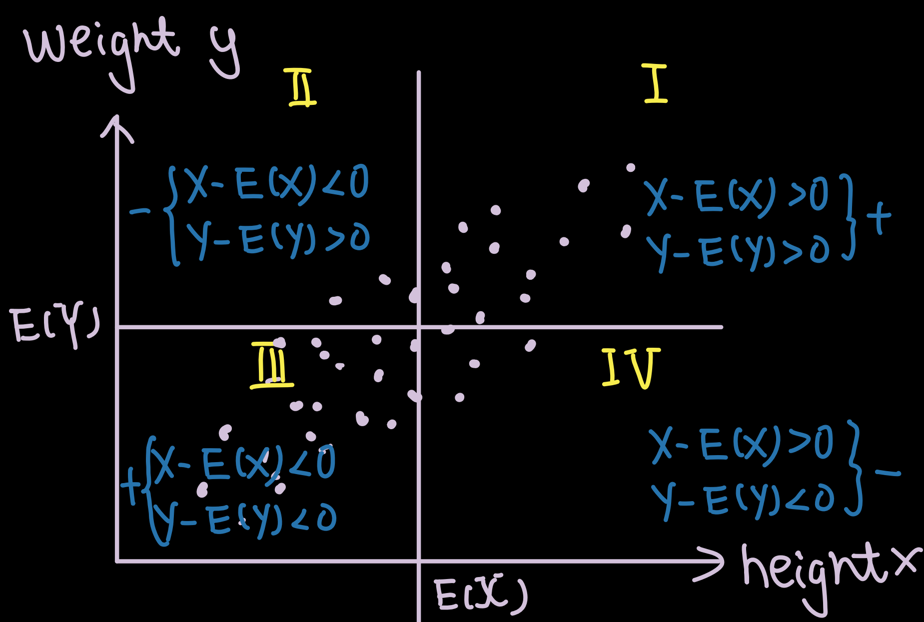
Special Case

$$h(X,Y) = X + Y$$

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

(avg of sums) (sum of averages)

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



+ → positive covariance

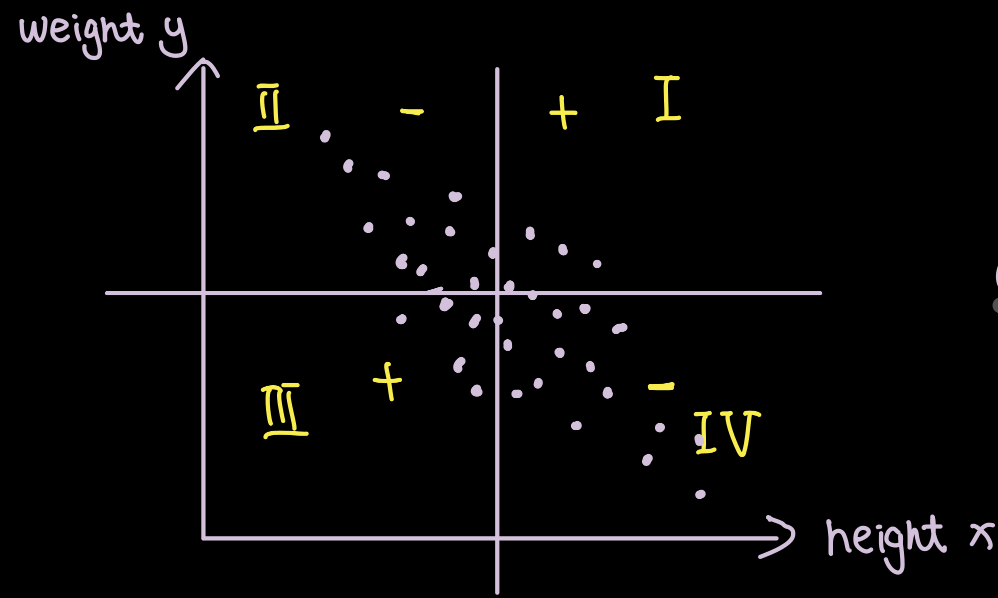
− → negative covariance

more points in "+"

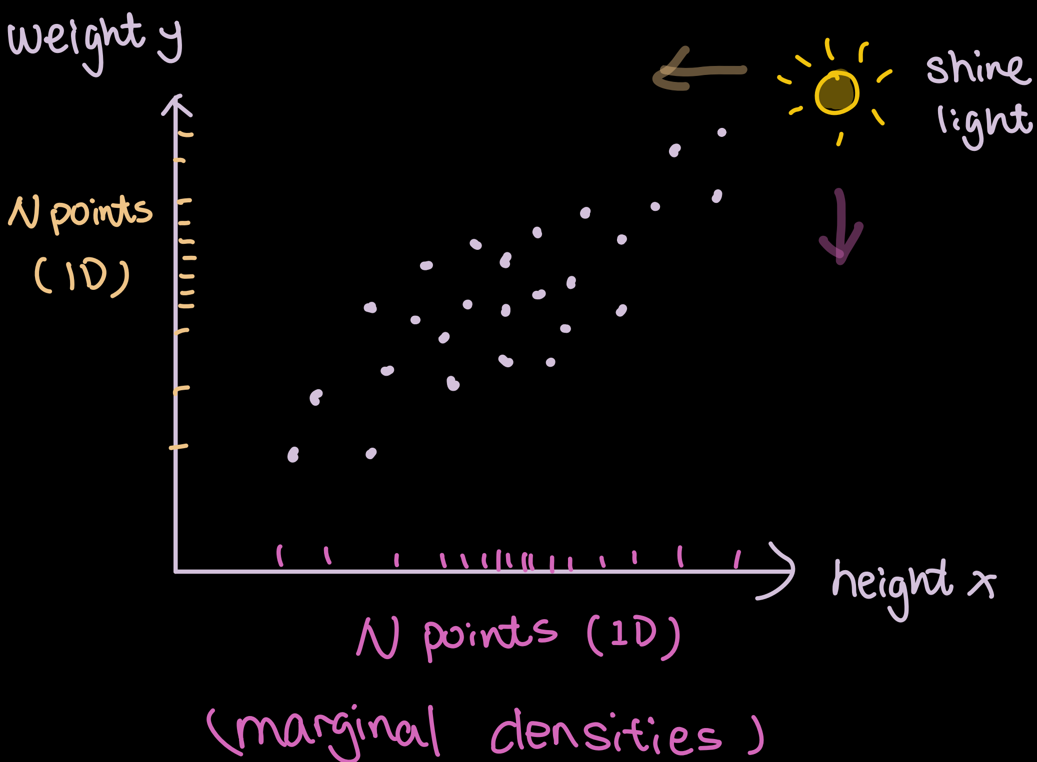
⇒ more positive products than negative

So $cov(X, Y)$

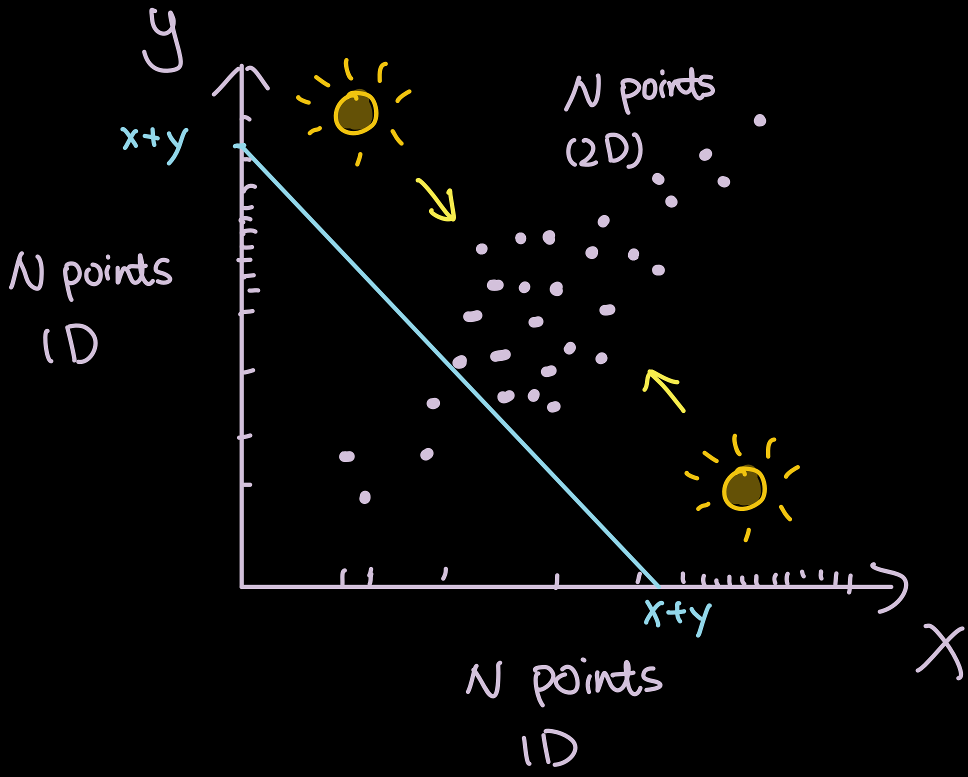
$$= E((X - E(X))(Y - E(Y))) > 0$$



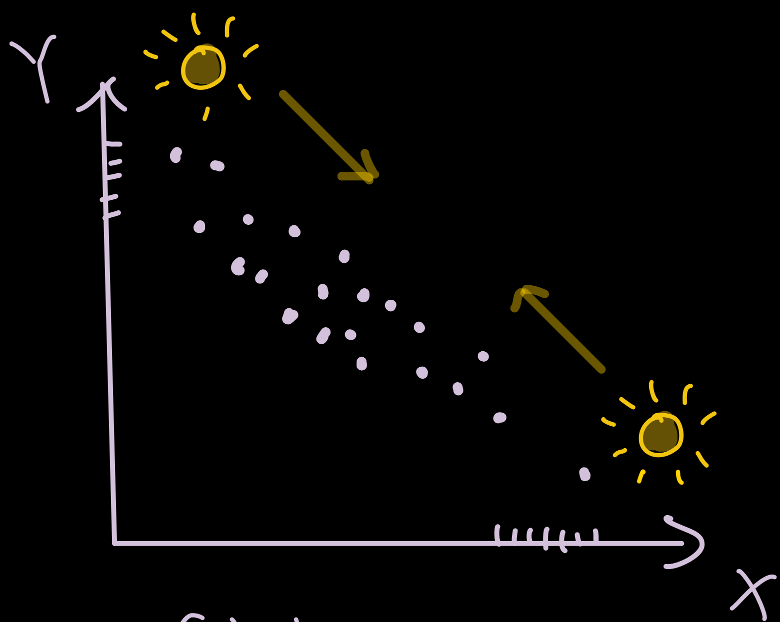
$\text{cov}(X, Y) < 0$



More <u>dispersed</u> when light shines diagonally onto axes since $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X,Y)$. The variances add up.



COV > 0



COV < 0

COV < 0

less dispersed when projected diagonally onto the axes since

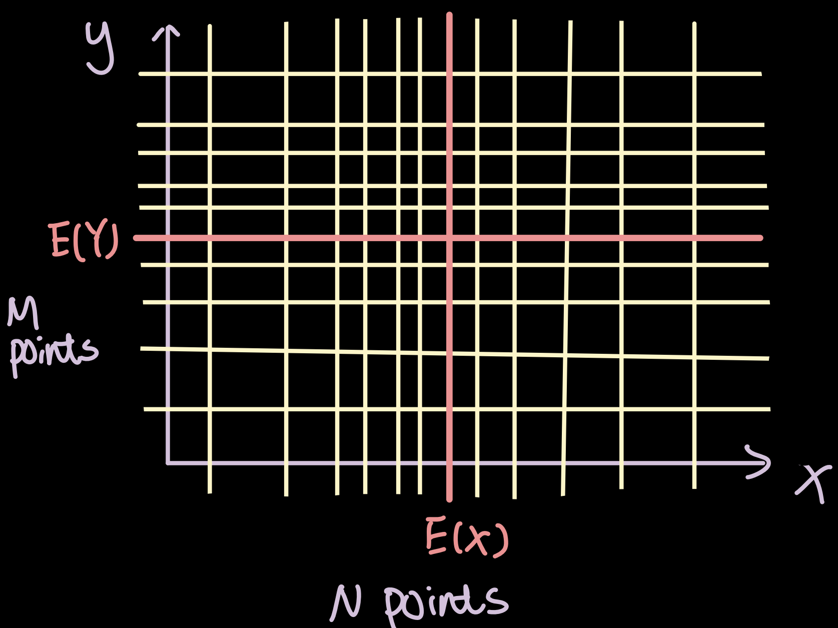
$\text{var}(X-Y) = \underbrace{\text{var}(X) + \text{var}(-Y)}_{\text{cancel out}} + 2\,\text{cov}(X, -Y)$

What if X and Y are independent ?

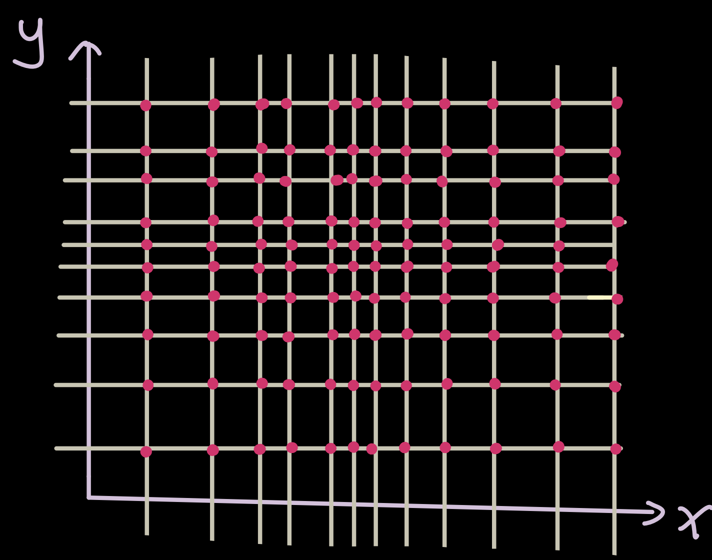
$X \perp Y \Rightarrow$ independent

$X \sim f(x)$ (N points)

$Y \sim g(y)$ (M points)



M x N pairs



$$\mathrm{cov}(X, Y) = 0$$

$$\mathrm{Var}(X, Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

$$\mathrm{cov}(X, Y) = E((\overset{\text{fixed}}{\underline{X - E(X)}})(Y - E(Y)))$$

for each fixed $X = x$,

Distribution of Y is $g(y)$

with fixed X,

$$E(Y - E(Y)) = E(Y) - E(Y) = 0$$

Finally, $\mathrm{cov}(X, Y) = 0$

If $X \perp Y$ independent,

$X_1, X_2, \ldots, X_n \overset{iid}{\sim} f(x)$

$$E(X_i) = \mu$$

$$Var(X_i) = \sigma^2$$

$$S = \sum_{i=1}^{n} X_i$$

$$E(S) = n\mu$$

$$Var(S) = n\sigma^2$$

$$\bar{X} = \frac{S}{n}$$

$$E(\bar{X}) = \frac{n\mu}{n} = \mu$$

$$Var(\bar{X}) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

finally,

$$E(\hat{I}) = I$$

$$var(\hat{I}) = \frac{var(h(x))}{n}$$