Correlations

For the Binomial distribution

$X = \sum_{i=1}^{n} B_i$ B_i : Random variable following the Bernoulli dist.

Toss a coin N times

B_i → follows the same distribution
↳ the same object

$P(B_i = 0) = P(B_i = 1) = 1/2$ true for every $i = 1, \dots, N$

Why don't we use the same variable?
↳ we're repeating the experiment.

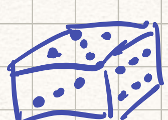
$X = \sum_{i=1}^{N} B = NB \begin{cases} 0 \\ N \end{cases}$ we're missing all the values in between

If each B_i is independent → we can have $0 \leq X \leq N$

the outcome of a given experiment is completely independent from the outcome of the previous one

Random variables do NOT need to be independent.

Let's consider a dice



X : One side of the dice

Y : opposite side of X

$X + Y = 7$ ⇒ we know by construction

X:	2	5	6	1	1	3	4
Y:	5	2	1	6	6	4	3
$X+Y$:	7	7	7	7	7	7	7

Not random anymore

Perfect correlation: we can infer the value of Y by only measuring X.

X:	2	5	6	1	1	3	4
$X+Y$:	7	6	7	7	8	9	7

There's a degree a correlation, but it is not perfect

Correlations are not restricted to the sum of two (or more) random variables

↳ In general, we see correlations in the output of an operation

Joint probability distribution

X: it follows a distribution $P(k)$: $\sum_k P(k) = 1$

Y: it follows a distribution $P'(l)$: $\sum_l P'(l) = 1$

k are the possible values that X can take

l are the possible values that Y can take

We need a joint prob. dist.

$P(X=k, Y=l)$ Probability that $X=k$ while $Y=l$.

The normalization becomes

$$\boxed{\begin{array}{c} \sum_k \sum_l P(X=k; Y=l) = 1 \\ \sum_{k,l} P(k,l) = 1 \end{array}}$$

If the two random variables are independent

$$P(X=k, Y=l) = P(X=k)\, P'(Y=l)$$

Coin 1		Coin 2	
H	T	H	T
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

H, H	$\frac{1}{2} \cdot \frac{1}{2}$	$= 1/4$
H, T	$1/2 \cdot 1/2$	$= 1/4$
T, H	$1/2 \cdot 1/2$	$= 1/4$
T, T	$1/2 \cdot 1/2$	$= 1/4$

In general, we have that

$$P(X=k, Y=l) \neq P(X=k)\, P'(Y=l)$$

we have correlated random variables

There are correlations between the variables

Reduced probability distribution

$$\sum_{l} P(k,l) = P_X(k)$$: Probability of $X=k$ independently of the value of Y

$$\sum_{k} P(k,l) = P_Y(l)$$ Probability of $Y=l$ independently of the value of X.

Note that in general $P(k,l) \neq P_X(k)\, P_Y(l)$

This is a good approximation

"we're neglecting the correlations between X and Y"

Average is linear: Proof

$$\langle X + Y \rangle = \sum_{k,l} (k+l)\, P(k,l)$$

$$= \sum_{k,l} k\, P(k,l) + \sum_{k,l} l\, P(k,l)$$

$$= \sum_k k \sum_l P(k,l) + \sum_l l \sum_k' P(k,l)$$

$$= \sum_k' k\, P_X(k) + \sum_l' l\, P_Y(l)$$

$$= \langle X \rangle + \langle Y \rangle$$

This also applies to the scaling of the random variable $\langle \alpha X + \beta Y \rangle$ $\alpha, \beta \in \mathbb{R}$.

$$\langle \alpha X + \beta Y \rangle = \alpha \langle X \rangle + \beta \langle Y \rangle$$

In contrast, we don't have $\langle XY \rangle = \langle X \rangle \langle Y \rangle$

This is not true in general.

$\langle XY \rangle = \langle X \rangle \langle Y \rangle$ only if X and Y are uncorrelated

$\boxed{\text{Proof:}}$

$$\langle XY \rangle = \sum_{k,l}' k l\, P(k,l)$$

$$= \sum_{k,l}' k l\, P(k) P'(l)$$

$$= \sum_k' \sum_l' k l\, P(k) P'(l)$$

$$= \sum_k' k\, P(k) \sum_l' l\, P'(l) = \langle X \rangle \langle Y \rangle$$

If $P(k,l) \neq P(k) P'(l)$

$$\langle XY \rangle = \sum_{k,l}' k l\, P(k,l) = \sum_k' k \sum_l' l\, P(k,l)$$

$$\langle Y \rangle_k = \sum_l' l\, P(k,l) \qquad = \sum_k' k \langle Y \rangle_k = \langle XY \rangle$$