<u>Defn</u> The <u>covariance</u> of two random variables X & Y is

$$\operatorname{cov}(X,Y) = \mathbb{E}\{[X-\mathbb{E}(X)][Y-\mathbb{E}(Y)]\}$$

<u>Remarks</u>

1. $\operatorname{cov}(X,X) = \operatorname{var}(X)$

2. $\operatorname{cov}(X,Y) = \mathbb{E}\{XY\} - \mathbb{E}(X)\cdot\mathbb{E}(Y)$

<u>Pf</u> For any numbers μ & ν 'lowercase nu'

$$\mathbb{E}\{[X-\mu][Y-\nu]\} = \mathbb{E}\{XY - \mu Y - \nu X + \mu\nu\}$$

linearity of expectation $\quad = \mathbb{E}(XY) - \mu\,\mathbb{E}(Y) - \nu\,\mathbb{E}(X) + \mu\nu$

The result now follows by setting $\mu = \mathbb{E}(X)$ and $\nu = \mathbb{E}(Y)$.

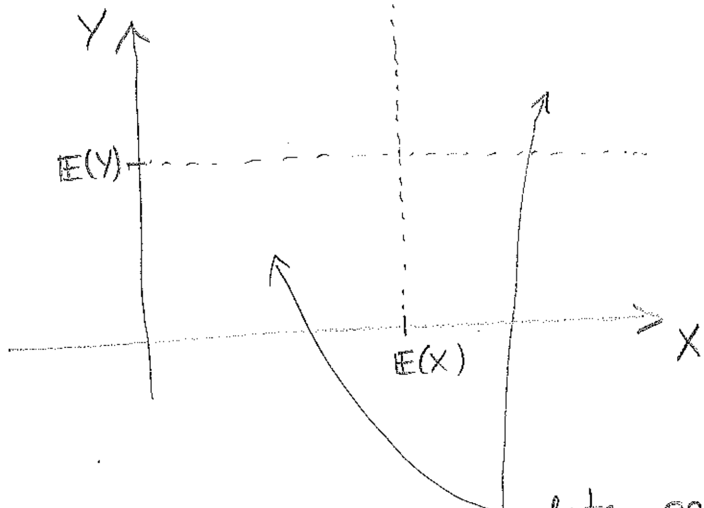
3. If X has dimensions of height, say, and Y of weight, then $\operatorname{cov}(X,Y)$ has dimensions of height×weight. A dimensionless analogue of covariance is the <u>Pearson correlation coefficient</u>:

$$\rho(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)\cdot\operatorname{var}(Y)}}$$

This will be discussed further in 170B.

4. If $cov(X,Y) > 0$ then X & Y are said to be <u>positively correlated</u> If $cov(X,Y) < 0$ then X & Y are <u>negatively correlated</u>.

Positive correlation occurs between variables that tend to be large together, e.g. height, age, math score



data points in these quadrants contribute positively to the covariance.
Points in other two quadrants contribute negatively

Note correlation is very easy to measure — Grab a sample, make some measurements, crunch the numbers. Causation is much more difficult to understand and requires the discovery of an underlying mechanism. Indeed, without an mechanistic understanding, it is very difficult to control for other factors influencing the outcome.

we understand people well enough to understand that an age-related mechanism can explain the positive correlation between height and knowledge of elementary-school math. Conversely, properly evaluating the latest health-food fad is going to require the dissection of a lot of laboratory animals — more than the advocates would probably tolerate....

5. By its definition, covariance is very influenced by outliers. This is also a problem with mean and variance. Imagine what would happen if you forgot the decimal point entering the data ...

180m tall instead of 1.80m ...

It will take a hundred (for mean) or even ten thousand data points (for variance) to drown out the influence of this one error. For comparison, consider how this error would influence the median.

6. As we saw last time for any collection of random variables $X_1, \ldots, X_n$ and numbers $c_1, \ldots, c_n$

$$\mathrm{var}\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i,j=1}^{n} c_i c_j \,\mathrm{cov}(X_i, X_j)$$

— we used this calculation to motivate the defn of covariance.

In this case, it is usual to arrange these numbers into a matrix, the <u>covariance matrix</u>

$$\Sigma_{ij} = \text{cov}(X_i, X_j)$$

It is true that capital sigma could be confused with the summation sign; however linear algebra notation actually gets rid of summation signs. For example

dot product

$$\underbrace{\text{var}(\vec{c}\cdot\vec{X})}_{= \sum_i c_i X_i} = \underbrace{\vec{c}\cdot\Sigma\vec{c}}_{= \sum_{i,j} c_i \,\text{cov}(X_i, X_j) c_j}$$

On to our next topic ...

Defn The conditional PMF of a r.v. X given an event A (with $\mathbb{P}(A) \neq 0$) is defined in the natural way:

$$p_{X|A}(k) = \mathbb{P}(X=k \mid A) = \frac{\mathbb{P}(X=k \ \& \ A)}{\mathbb{P}(A)}$$

Eg Suppose $X \sim \text{Geometric}(p)$. Eg, number of games needed to win. "Given that I lost three games in a row," says the gambler, "I'm going to bet more — I'm overdue for a win."

How much longer must the gambler wait?

We want the law of $Y = X - 3$ given the event $A = \{X \geq 4\}$:

$$p_{Y|A}(k) = \frac{\mathbb{P}(Y=k \ \& \ X \geq 4)}{\mathbb{P}(X \geq 4)} = \frac{\mathbb{P}(X=k+3 \cap X \geq 4)}{\mathbb{P}(X \geq 4)}$$

Recall $\mathbb{P}(X=\ell) = \begin{cases} p(1-p)^{\ell-1} & \ell = 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$

Thus

$$\mathbb{P}(X \geq 4) = 1 - \mathbb{P}(X \leq 3)$$

$$= 1 - \mathbb{P}(X=1) - \mathbb{P}(X=2) - \mathbb{P}(X=3)$$

$$= 1 - p - p(1-p) - p(1-p)^2$$

$$= (1-p)\{1 - p - p(1-p)\} = (1-p)^3$$

Oh Duh = probability of three consecutive failures (among indep. Bernoulli trials).

Looking again to the PMF for X we see that

$$\mathbb{P}(X = k+3 \cap X \geq 4) = \begin{cases} p(1-p)^{k+2} & k = 1,2,3,\ldots \\ 0 & \text{otherwise.} \end{cases}$$

Thus putting the pieces together,

$$P_{Y|A}(k) = \begin{cases} \dfrac{p(1-p)^{k+2}}{(1-p)^3} = p(1-p)^{k-1} & k = 1,2,3,\ldots \\ 0 & \text{otherwise.} \end{cases}$$

Lo and behold, the law of Y conditioned on A is Geometric(p) — the same as X. Losing has in no way affected the waiting time.

On the one hand, this is rather obvious from the fact that the trials are independent. On the other hand the idea expressed by the gambler is sufficiently prevelent to have a name (& a lengthy Wikipedia entry), "The Gambler's Fallacy".

Defn The <u>conditional expectation</u> of a r.v. X given an event A (with $\mathbb{P}(A)>0$) is defined by

$$\mathbb{E}(X \mid A) = \sum_k k \cdot \mathbb{P}(X=k \mid A)$$

$$= \sum_k k \cdot p_{X|A}(k)$$

Remarks As a conditional probability law is still a probability law, conditional expectation obeys the same rules as usual expectation:

$$\mathbb{E}\{g(X) \mid A\} = \sum_k g(k)\, p_{X|A}(k)$$

$$\mathbb{E}\{X+Y \mid A\} = \mathbb{E}\{X \mid A\} + \mathbb{E}\{Y \mid A\}.$$