Suppose X and Y are independent random variables on (S, P), and $a, b \in \mathbb{R}$ constants, then for $i, j \in \mathbb{R}$

$$P(X-a=i,\ Y-b=j) = P(X=a+i,\ Y=b+j)$$
$$= P(X=a+i)P(Y=b+j) \quad \left(\text{since } X, Y \text{ are independent}\right)$$
$$= P(X-a=i)\ P(Y-b=j).$$

We have therefore proved the following:

Lemma: If X and Y are independent random variables, then $X-a$ and $Y-b$ are also independent for any real constants a and b.

Example: Suppose a total of $n+1$ lottery tickets are sold in a town and the tickets are labelled $\{0, 1, 2, \dots, n\}$. Assume that every number has an equal chance of being drawn.

The expected winning ticket number $= \sum_{i=0}^{n} i \cdot \frac{1}{n+1}$

$$= \frac{1}{n+1} \sum_{i=0}^{n} i$$

$$= \frac{1}{n+1} \cdot \frac{n(n+1)}{2}$$

$$= \frac{n}{2}.$$

We will use this result to illustrate the notion of variance.

The Variance of a random variable

Consider the following two situations:

(a) A small town conducts a lottery and sells 10,001 tickets. The Tickets are labelled $\{0, 1, \dots, 10{,}000\}$. A box contains duplicates of every ticket sold and only these.

The lottery is conducted by drawing a ticket from the box with every ticket having the same chance of being drawn. For a ticket t in the box, let $X(t)$ be its number. From the example we just did

$$EX = \frac{10{,}000}{2} = 5000.$$

(b) Suppose a fair coin is tossed 10,000 (fair means $p = \frac{1}{2}$) and Y is the random variable which counts the number of heads. We saw in the last lecture (with $n = 10{,}000$ and $p = \frac{1}{2}$) that

$$EY = 10{,}000 \left(\tfrac{1}{2}\right) = 5000.$$

In each case the image of the random variable is $\{0, 1, \ldots, 10000\}$ and so in each case these random variables make $\{0, 1, \ldots, 10000\}$ into the sample space of a probability distribution. The expected value of the random variables is the same, namely 5000.

Now suppose after 10,000 coin tosses, you are told that on at least 7,500 occasions the coin turned up heads. What would your reaction be?

On the other hand, after the drawing of the winning ticket in the lottery, it was announced (to prolong the suspence) that the winning ticket number was at least 7,500. What would your reaction be?

Chances are you were shocked by the first statement but took the second one in your stride. Nevertheless 7500 is equally from the expected value of X as it is from the expected value of Y — after all the expected value of both is 5000 and the set of outcomes of X as well as that for Y is $\{0, 1, \ldots, n\}$.

The idea of variance and standard deviation will help us understand that indeed the possibility of Y being over 7500 is shocking while it is not too shocking that X is larger than 7500.

Definition: Let X be a random variable with expectation $E(X)=\mu$. The variance of X, denoted $\mathrm{var}(X)$, is defined as

$$\mathrm{var}(X) = E(X-\mu)^2.$$

Note that

$$\mathrm{var}(X) = \sum_{i \in X(S)} (i-\mu)^2 P(X=i).$$

Now

$$\left.\begin{aligned} E(X-\mu)^2 &= E(X^2 - 2X\mu + \mu^2) \\ &= E(X^2) - 2\mu E(X) + E(\mu^2) \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2. \end{aligned}\right\} \quad (*)$$

We are using the fact that if c is a constant, then $E(c)=c$. This is because $E(c) = \sum_{i \in X(S)} c\, P(X=i) = c$.

Another way of expressing $(*)$ is

$$\boxed{\mathrm{var}(X) = E(X^2) - E(X)^2}$$

The standard deviation of X, denoted σ_X, is the square root of var(X), i.e.

$$\boxed{\sigma_X = \sqrt{\operatorname{var}(X)}}$$

Here are some basic properties of variance and standard deviations.

1. Let X be a random variable and $c \in \mathbb{R}$ a constant. Then
   (a) $\operatorname{var}(cX) = c^2 \operatorname{var}(X)$
   (b) $\sigma_{cX} = |c|\, \sigma_X$

2. If X, Y are random variables then $\operatorname{var}(X+Y)$ need not equal $\operatorname{var}(X) + \operatorname{var}(Y)$. However if X and Y are independent random variables with $E(X) = \mu$ and $E(Y) = \nu$ (say), then

$$\begin{aligned}
\operatorname{var}(X+Y) &= E(X+Y)^2 - (\mu+\nu)^2 \\
&= E(X^2) + E(Y^2) + 2E(XY) - \mu^2 - \nu^2 - 2\mu\nu \\
&= E(X^2) + E(Y^2) + 2E(X)E(Y) - \mu^2 - \nu^2 - 2\mu\nu \\
&\qquad \text{(since X and Y are indep.)} \\
&= E(X^2) + E(Y^2) - \mu^2 - \nu^2 \\
&= \operatorname{var}(X) + \operatorname{var}(Y).
\end{aligned}$$

The same argument gives us:

Theorem: If $X_1, \ldots, X_n$ are independent random variables, then

$$\operatorname{var}(X_1 + \ldots + X_n) = \operatorname{var}(X_1) + \ldots + \operatorname{var}(X_n).$$

This gives us the following result about Binomial random variables:

Theorem: Let X be the Binomial random variable with parameters (n,p). Then

$$\mathrm{var}(X) = np(1-p).$$

Proof:

We know that

$$X = X_1 + \ldots + X_n$$

where each X_i is Bernoulli with parameter p, and $X_1, X_2, \ldots, X_n$ are independent random variables. It follows that

$$\mathrm{var}(X) = \mathrm{var}(X_1) + \ldots + \mathrm{var}(X_n).$$

Thus it is enough for us to prove that if Y is a Bernoulli random variable, then $\mathrm{var}(Y) = p(1-p)$. Now

$$\begin{aligned} \mathrm{var}(Y) &= 0 \cdot P(Y=0) + 1 \cdot P(Y=1) \\ &= 0 + 1 \cdot p \\ &= p. \end{aligned}$$

This proves the theorem. q.e.d.

Remark: An immediate consequence is that the standard deviation of a Bernoulli random variable X, with parameters (n,p), is $\sqrt{np(1-p)}$.

Variance and standard deviation as measures of dispersion

Intuitively, both $\mathrm{var}(X)$ and σ_X measure the possible dispersion of the outcome from the expected value of the outcome and this will help us analyse the difference between the lottery example and the repeated coin toss example above.

Here is the first step towards this analysis of dispersion. Suppose $\sigma_X = 0$. From what we said, we expect $X(s) = \mu$ for every outcome s which has a positive probability of occurrence. Here are the precise calculations.

Let $\mu = E(X)$. Then

$$\sum_{s \in S} (X(s)-\mu)^2 P(s) = \operatorname{var}(X) = \sigma_X^2 = 0.$$

Since $(X(s)-\mu)^2 P(s) \geq 0 \ \forall s \in S$, the above means that

$$(X(s)-\mu)^2 \cdot P(s) = 0 \qquad \forall s \in S.$$

This means that $X(s) = \mu \ \forall s$ such that $P(s) > 0$. If $P(s) = 0$, the outcome s is not going to occur in any experiment (since S is a finite set). Thus $X(s) = \mu$ for every outcome s which could occur. In other words we have:

Lemma: Let X be a random variable with mean μ such that $\sigma_X = 0$. Then

$$P(X=\mu) = 1.$$

The following famous inequality will help us analyse why it is very surprising if we have 7,500 or more heads in 10,000 tosses of a fair coin but that is not too surprising if a lottery number drawn at random from $\{0, 1, \ldots, 10,000\}$ is larger than 7,500.

Theorem (Chebyshev's inequality): Let X be a random variable on a probability space (S, P), and let $\mu = E(X)$. Then, for every $k > 0$, we have

$$P\left(|X-\mu| \leq k\,\sigma_X\right) \geq 1 - \frac{1}{k^2}.$$

Proof:

If $\sigma_X = 0$ then we have seen that $P(X=\mu) = 1$. Clearly $(X=\mu) = \left(|X-\mu| \leq k \cdot \sigma_X\right)$ if $\sigma_X = 0$, and therefore the theorem is trivially true.

We will now consider the case where the standard deviation does not vanish.

Suppose $\sigma_X \neq 0$. Then $\sigma_X > 0$.

Let
$$A = \{s \in S \mid |X-\mu| > k\sigma_X\}$$
and
$$B = \{s \in S \mid |X-\mu| \leq k\sigma_X\}.$$

Then A and B are disjoint and
$$S = A \cup B.$$
Thus
$$P(A) = 1 - P(B).$$
We have to find a lower bound for $P(B)$. This amounts to finding an upper bound for $P(A)$ and that is what we now proceed to do.

We have:
$$\begin{aligned}
\operatorname{var}(X) &= E(X-\mu)^2 \\
&= \sum_{s \in S} (X(s)-\mu)^2 P(s) \\
&\geq \sum_{s \in A} (X(s)-\mu)^2 \cdot P(s) \\
&\geq \sum_{s \in A} k^2 \sigma_X^2 \cdot P(s) \qquad \left(\text{since } |X(s)-\mu| > k\sigma_X \text{ for } s \in A\right) \\
&= k^2 \sigma_X^2 \sum_{s \in A} P(s) \\
&= k^2 \sigma_X^2 P(A).
\end{aligned}$$

This would be strict if $P(s) > 0 \ \forall s \in A$.

In other words,
$$\sigma_X^2 \geq k^2 \sigma_X^2 P(A) \qquad (\text{for } \operatorname{var}(X) = \sigma_X^2)$$
Since we assumed $\sigma_X > 0$, so the above gives $1 \geq k^2 P(A)$, i.e.
$$P(A) \leq \frac{1}{k^2}.$$

This means
$$1 - P(B) \leq \frac{1}{k^2}$$
which in turn gives
$$P(B) \geq 1 - \frac{1}{k^2}.$$
This is what we had to prove. q.e.d.

A return to the discussion on lottery tickets vs. fair coin tosses

The probability that a lottery ticket has value greater than or equal to 7,500 is
$$\frac{2,500}{10,001} \approx \frac{1}{4}$$

and so the fact that the winning number is greater than or equal to 7,500 is not too surprising (certainly nothing to be shocked about).

As for tossing a fair coin 10,000 is concerned, we know that the number of successes follows a Binomial distribution with parameters (n,p) where
$$n = 10,000 \quad \text{and} \quad p = \frac{1}{2}.$$
It follows that if Y is a Binomial random variable with parameters (n,p) as above, then
$$E(Y) = 5000 \quad \text{and} \quad \sigma_Y = 50$$
By Chebyshev's inequality we therefore have

$$P\left(|Y - 5000| \leq 50k\right) \leq 1 - \frac{1}{k^2}$$

for every $k > 0$. For $k = 50$, the above translates to
$$P\left(|Y - 5000| \leq 2500\right) \leq 1 - \frac{1}{50^2} = \frac{2499}{2500} \approx 0.9996.$$

This means the chances of there being 7,500 or more heads in 10,000 tosses is very very slim. A good reason to be shocked if that indeed does happen.