Expectation

Two coins are tossed. How many heads can we *expect* to come up?

Let Y = number of heads. Then $p(y) = \begin{cases} 1/4, & y = 0 \\ 1/2, & y = 1 \\ 1/4, & y = 2 \end{cases}$

The answer seems to be 1 (the middle value).

But what exactly do we mean by "expect"?

A mental experiment: Suppose we toss the two coins 1000 times, and each time record the number of heads, y.

The result would be something like 1,1,2,0,1,2,0,1,1,1,...,1,0.

We'd get about 250 zero's, 500 one's and 250 two's.

So the average of the 1000 values of Y would be approximately

$$\frac{250(0)+500(1)+250(2)}{1000} = 0(1/4) + 1(1/2) + 2(1/4) = 1.$$

This agrees with our intuitive answer.

Observe that 0(1/4) + 1(1/2) + 2(1/4) = $0p(0)+1p(1)+2p(2)=\sum_{y=0}^{2} yp(y)$.

This leads to the following definition.

Suppose Y is a discrete random variable with pdf $p(y)$.
Then the *expected value* (or *mean*) of Y is
$E(Y)=\sum_{y} yp(y)$. (The sum is over all possible values y of the rv Y.)

We may also write Y's mean as EY or μ_Y or μ.

The mean μ is a *measure of central tendency*, in the sense that it represents the average of a hypothetically infinite number of independent realisations of Y.

Example 10 Suppose that *Y* is a random variable which equals 5 with probability 0.2 and 7 with probability 0.8. Find the expected value of *Y*.

$$EY = \sum_y yp(y) = 5(0.2) + 7(0.8) = 6.6\,.$$

This means that if we were to generate many independent realisations of *Y*, so as to get a sequence like 7,7,5,7,5,7,..., the average of these number would be close to 6.6. As the sequence got longer, the average would converge to 6.6. More on this later.

Example 11 Find the mean of the Bernoulli distribution.

Let $Y \sim$ Bern(p). Then $p(y) = \begin{cases} p, & y = 1 \\ 1-p, & y = 0. \end{cases}$

So *Y* has mean $\mu = \sum_{y=0}^{1} yp(y) = 0p(0) + 1p(1) = 0(1-p) + 1p = p$.

Thus for example, if we toss a fair coin thousands of times, and each time write 1 when a head comes up and 0 otherwise, we will get a sequence like 0,0,1,0,1,1,1,0,... The average of these 1's and 0's will be about 1/2, corresponding to the fact that each such number has a Bernoulli distribution with parameter 1/2 and thus a mean of 1/2.

Example 12 Find the mean of the binomial distribution.

Let $Y \sim$ Bin(n,p). Then *Y* has mean

$$\mu = \sum_y yp(y) = \sum_{y=0}^{n} y \binom{n}{y} p^y (1-p)^{n-y}$$

$$= \sum_{y=1}^{n} y \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y} \qquad \text{(the first term is zero)}$$

$$= np \sum_{y=1}^{n} \frac{(n-1)!}{(y-1)!((n-1)-(y-1))!} p^{y-1} (1-p)^{(n-1)-(y-1)}$$

(since $n! = n(n-1)!$ and $p^y = pp^{y-1}$)

$$= np \sum_{x=0}^{m} \frac{m!}{x!(m-x)!} p^x (1-p)^{m-x} \qquad (x = y-1 \text{ and } m = n-1)$$

$= np$ (since the sum equals 1, by the binomial theorem).

This makes sense. Eg, if we roll a die 60 times, we can expect 60(1/6) = 10 sixes.

Expectations of functions of random variables

Suppose that Y is a discrete random variable with pdf $p(y)$, and $g(t)$ is a function.
Then the *expected value* (or *mean*) of $g(Y)$ is defined to be

$$E(g(Y)) = \sum_y g(y)p(y).$$

Note: The text presents this equation as Theorem 3.2 and provides a proof for it. We have instead *defined* the expected value of a function of a rv, with no need for a proof.

Example 13   Suppose that $Y \sim$ Bern(p). Find $E(Y^2)$.

$$E(Y^2) = \sum_y y^2 p(y) = 0^2(1-p) + 1^2 p = p$$

(same as EY; in fact, $EY^k = EY = p$ for all k).

Note: We may also write $E(Y^2)$ as EY^2, which should not be confused with $(EY)^2$.

Laws of expectation   (Theorems 3.3 to 3.5 in text)

1. If c is a constant, then $Ec = c$.

2. $E\{cg(Y)\} = cEg(Y)$.

3. $E\{g_1(Y) + \ldots + g_k(Y)\} = Eg_1(Y) + \ldots + Eg_k(Y)$.

Proof of 1st law: $Ec = \sum_y cp(y) = c\sum_y p(y) = c(1) = c$.

Example 14   Suppose that $Y \sim$ Bern(p). Find $E(3Y^2 + Y - 2)$.

$$E(3Y^2 + Y - 2) = 3EY^2 + EY - 2$$
$$= 3p + p - 2 \quad \text{(recall from Example 13 that } EY^k = p \text{ for all } k\text{)}$$
$$= 4p - 2.$$

Special expectations (definitions)

1.	The kth *raw moment* of Y is $\mu_k' = EY^k$.
2.	The kth *central moment* of Y is $\mu_k = E(Y-\mu)^k$.
3.	The *variance* of Y is $Var(Y) = \sigma^2 = \mu_2 \quad (= E(Y-\mu)^2)$.
4.	The *standard deviation* of Y is $SD(Y) = \sigma \quad (= \sqrt{Var(Y)}\,)$.

We can also write $Var(Y)$ as $V(Y)$ or $VarY$ or VY or σ_Y^2.

Note that $\mu_1' = \mu$. Also, $\mu_1 = E(Y-\mu)^1 = EY - \mu = \mu - \mu = 0$.

Example 15 Suppose that $p(y) = y/3$, $y = 1,2$.
Find: **(a)** μ_3'
(b) σ.

(a) $$p(y) = \begin{cases} 1/3, & y = 1 \\ 2/3, & y = 2 \end{cases}$$

So $$\mu_3' = EY^3 = \sum_y y^3 p(y) = 1^3\left(\frac{1}{3}\right) + 2^3\left(\frac{2}{3}\right) = \frac{17}{3}.$$

(b) First, $$\mu = EY = \sum_y yp(y) = 1(1/3) + 2(2/3) = 5/3.$$

So $$\sigma^2 = \mu_2 = E(Y-\mu)^2 = \sum_y (y-\mu)^2 p(y) = \left(1-\frac{5}{3}\right)^2 \frac{1}{3} + \left(2-\frac{5}{3}\right)^2 \frac{2}{3} = \frac{2}{9}.$$

Hence $\sigma = \sqrt{2}/3 = 0.4714$.

The various moments provide information about the *nature* of a distribution.

We have already seen that the *mean* provides a measure of *central tendency*.

The *variance* and *standard deviation* provide measures of *dispersion*.

Distributions that are highly disperse have a large variance.

Example: Suppose X has pdf $p(x) = 1/2$, $x = 1,3$
and Y has pdf $p(y) = 1/2$, $y = 0,4$.

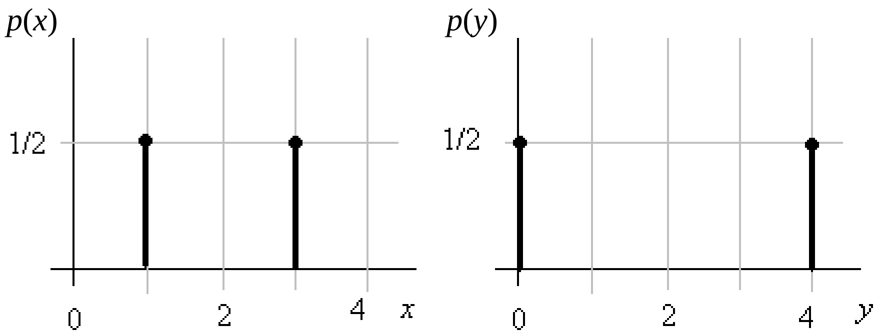
Find *VarX* and *VarY*.

Which distribution is the more disperse?

Both distributions have a mean of 2 (= {average of 1 and 3} = {average of 0 and 4}).

Thus: $VarX = (1-2)^2 0.5 + (3-2)^2 0.5 = 1$
$VarY = (0-2)^2 0.5 + (4-2)^2 0.5 = 4$.

We see that *VarY* > *VarX*.
This corresponds to the fact that *Y*'s distribution is the more disperse of the two.



Two important results (for computing variances)

1.	$VarY = EY^2 - (EY)^2$. (Equivalently, $\sigma^2 = \mu_2' - \mu^2$.)
2.	$Var(a + bY) = b^2 VarY$.

Proof of 1:
LHS = $E(Y-\mu)^2 = E(Y^2 - 2Y\mu + \mu^2) = EY^2 - 2\mu EY + \mu^2$ = RHS.

Proof of 2:
LHS = $E[(a+bY) - E(a+bY)]^2 = E[a + bY - a - b\mu]^2 = b^2 E(Y-\mu)^2$ = RHS.

Example 16 Find the variance of the Bernoulli distribution.

Recall that if $Y \sim$ Bern(p), then $EY = EY^2 = p$ (see Example 13).
Therefore Y has variance $VarY = p - p^2 = p(1-p)$.

What is the variance of $X = 2 - 5Y$? $VarX = (-5)^2 VarY = 25p(1-p)$.

Moment generating functions

The *moment generating function (mgf)* of a random variable Y is defined to be
$m(t) = Ee^{Yt}$.

Mgf's have two important uses:

1. To compute raw moments, according to the formula:

$$\mu_k' = m^{(k)}(0).$$

(See Thm 3.12 in the text for a general proof, and below for the case $k = 1$.)

2. To identify distributions according to the result:

If the mgf of a rv Y is the same as that of another rv X,
we can conclude that Y has the same distribution as X.

(This follows from "the uniqueness theorem", a result in pure mathematics.)

Notes: $m^{(k)}(0)$ denotes the kth derivative of $m(t)$, evaluated at $t = 0$,
and may also be written as $\left[\dfrac{d^k m(t)}{dt^k} \Bigg|_{t=0} \right]$.

We may also write $m^{(1)}(t)$ as $m'(t)$, and $m^{(2)}(t)$ as $m''(t)$, etc.
For all mgf's, $m(t)$, it is true that $m^{(0)}(0) = m(0) = Ee^{Y0} = E1 = 1$.
Proof that $\mu = \mu_1' = m'(0)$:

$$m'(t) = \frac{d}{dt} m(t) = \frac{d}{dt} Ee^{Yt} = \frac{d}{dt} \sum_y e^{yt} p(y) = \sum_y \left(\frac{d}{dt} e^{yt} \right) p(y) = \sum_y ye^{yt} p(y).$$

So $m'(0) = \sum_y ye^{y0} p(y) = \sum_y y \times 1 \times p(y) = EY = \mu$.

Example 17 Use the mgf technique to find the mean and variance of the binomial distribution.

Let Y ~ Bin(n,p). Then Y has mgf

$$m(t) = Ee^{Yt} = \sum_{y=0}^{n} e^{yt} \binom{n}{y} p^y (1-p)^{n-y} = \sum_{y=0}^{n} \binom{n}{y} (pe^t)^y (1-p)^{n-y}$$

$$= \{(pe^t) + (1-p)\}^n \quad \text{by the binomial theorem.}$$

Thus $m(t) = (1-p+pe^t)^n$.

Then, $m'(t) = \dfrac{dm(t)}{dt} = n(1-p+pe^t)^{n-1} pe^t$ by the *chain rule for differentiation*.

(This rule is: $\dfrac{du}{dt} = \dfrac{du}{dv}\dfrac{dv}{dt}$, where here $v = 1-p+pe^t$ and $u = m(t) = v^n$.)

So $\mu = \mu_1' = m'(0) = n(1-p+pe^0)^{n-1} pe^0 = np$ (as before).

Further,

$$m''(t) = \frac{d^2m(t)}{dt^2} = \frac{dm'(t)}{dt} = np\{(1-p+pe^t)^{n-1}e^t + e^t(n-1)(1-p+pe^t)^{n-2}pe^t\}$$

by the *product rule for differentiation*.

(This rule is: $\dfrac{d(uv)}{dt} = u\dfrac{dv}{dt} + v\dfrac{du}{dt}$, where here $u = (1-p+pe^t)^{n-1}$ and $v = e^t$.)

So $\mu_2' = m''(0) = np\{(1-p+pe^0)^{n-1}e^0 + e^0(n-1)(1-p+pe^0)^{n-2}pe^0\}$
$= np\{1+(n-1)p\}$.

Therefore $\sigma^2 = \mu_2' - \mu^2 = np\{1+(n-1)p\} - (np)^2 = np(1-p)$.

Example 18 A random variable Y has the mgf $m(t) = \dfrac{1}{8}(1+e^t)^3$.

Find the probability that Y equals three.

$m(t) = \left(1 - \dfrac{1}{2} + \dfrac{1}{2}e^t\right)^3 = (1-p+pe^t)^n$, where n = 3 and p = 1/2.

Thus $m(t)$ is the mgf of a random variable whose distribution is binomial with parameters 3 and 1/2. Therefore Y ~ Bin(3,1/2), and so $P(Y = 3) = 1/8$.