Some more model descriptions.

Definitions and names.

p(x) = P(X = x)

p(x) is called *Probability Mass Function* (pmf) or the *Probability Density Function* (pdf) of the distribution. The second name is usually used for continuous variables (see later) and is denoted f(x).

F(x) or $F_X(x)$ is called the *Cumulative Distribution Function* (cdf) (or just Distribution Function).

$$F_X(x) = P(X \le x).$$

The cdf is often more useful in calculating probabilities than the pdf.

If X is defined on consecutive integers
(eg 0,1,2,3)

$$F_X(x) = \sum_{k=0}^{x} p(k)$$

also $p(x) = F_X(x) - F_X(x-1)$

For the Binomial and Poisson no analytic expressions exist for the cdf.

For the geometric however we have:

$$F_X(x) = \sum_{k=1}^{x} p(1-p)^{k-1} = p\sum_{n=0}^{x-1}(1-p)^n$$

$$= p\frac{1-(1-p)^x}{1-(1-p)} = p\frac{1-(1-p)^x}{p} = 1-(1-p)^x$$

It just depends whether you can do the summation or not!

Mean, Variance and other moments.

The *mean* of a distribution is the average value of X weighted by the probability.

It is a number denoted by E(X) , it is also called the expectation of X.

$$E(X) = \sum_{all\ x} x\ p(x)$$

Example: X=no. of cars sold by a dealer each day.

X	p(x)
0	0.1
1	0.2
2	0.4
3	0.2
4	0.1

$$E(X) = 0 \times 0.1 + 1 \times 0.2 + 2 \times 0.4 + 3 \times 0.2 + 4 \times 0.1 = 2$$

What does the Mean mean?

The distribution above says that over the long run:

On 10% of days he sells no cars X=0
On 20% of days he sells 1 car X=1
And so on

If we divide the number of cars sold over a long period by the number of days – we get the average number of cars per day; this tends to E(X).

Another dealer has the probability distribution below:

X	p(x)
0	0.1
1	0.1
2	0.4
3	0.4
4	0

Is she a better salesperson?

One criterion we might use is E(X).

Here E(X)=2.1 (verify as exx). So she sells more cars on average.

Linearity of E(X).

Mathematicians call E(.) a linear operator. In a first course there is no need to formally prove this but it is a useful fact to have.

Y is the number of wheels that the (first) dealer sells. Assume each car has 5 wheels.

Y=5X

It is simple to verify that E(Y)=E(5X)=5E(X).

Or more generally that:

$E(aX) = aE(X)$ for a any real number.

Also $E(X+b) = E(X)+b$.

Although a more subtle mathematical argument is required below. But in the example it is obvious that:

Z = the number of cars sold by both dealers.

$$Z = X_1 + X_2$$

$$E(Z) = E(X_1) + E(X_2).$$

This result might be obvious if the numbers of cars sold by each are independent – but that is NOT required it is true anyway irrespective of the relationship between X_1 and X_2.

In summary E() can be taken through any summation (linear combination).

$L = a + \sum_{i=1}^{n} c_i X_i$ then $E(L) = a + c_i E(X_i)$

Means of Standard distributions.

Bernoulli

$$p(x) = p \qquad \text{if } x = 1$$
$$= (1 - p) \qquad \text{if } x = 0$$

$$E(X) = 0 \times (1 - p) + 1 \times p = p$$

We note that for this distribution E(X)=P(X=1).

Binomial

Here we use two facts established earlier:

1. The Binomial is sum of n Bernoullis
2. We can take E(X) through the sum.

Y is B(n,p) then

$Y = \sum_{i=1}^{n} X_i$ where each X_i is Bernoulli(p)

$$E(Y) = E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np$$

Geometric(p)

$$E(Y) = \sum_{x=1}^{n} xp(1 - p)^{x-1} = \frac{1}{p}$$

The last step is not at all trivial – look it up in a series book.

Poisson(λ)

Here again the derivation requires the knowledge of series.

$$E(Y) = \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=1}^{\infty} x \frac{\lambda^x}{x!} = e^{-\lambda} \lambda \sum_{x=1}^{\infty} \frac{\lambda^{(x-1)}}{(x-1)!} = e^{-\lambda} \lambda e^{\lambda} = \lambda$$

So the parameter of the Poisson is its mean.

Variance of a Distribution

A third dealer in this town has the following distribution of sold cars:

x	p(x)
0	0.5
1	0
2	0
3	0
4	0.5

A simple calculation shows that here E(X)=2. So it is the same as dealer 1 , but this dealer has good days (on which he sells 4) and bad days (sales=0).

His sales distribution is more *variable*.

Another useful summary of X is a measure of the variability.

The *variance* written V(X) or Var(X) is such a measure:

V(X) is the average squared deviation from the mean.

Suppose $E(X) = \mu$

$$V(X) = \sum_{allx} (x - \mu)^2 p(x)$$

We look at the squared deviation because the average of $(x - \mu)$ would always be 0. We could use average absolute deviation but that is not nice mathematically.

For dealer 1 we have the distribution of deviations

(x-μ)	p(x)	$(x-\mu)^2 p(x)$
-2	0.1	0.4
-1	0.2	0.2
0	0.4	0
1	0.2	0.2
2	0.1	0.4
	sum	1.2

So V(X)=1.2

For dealer 3 V(X)=4 (verify) so we conclude that 3's sales distribution is more variable.

The computation of variance as written above is tedious (and possibly inaccurate due to rounding).

The formula can be re-written as follows:

$$\begin{aligned} V(X) &= \sum_{allx} (x-\mu)^2 p(x) \\ &= \sum_{allx} (x^2 - 2\mu x + \mu^2) p(x) \\ &= \sum_{allx} x^2 p(x) - 2\mu \sum_{allx} xp(x) + \mu^2 \sum_{allx} p(x) \end{aligned}$$

But $\sum_{allx} xp(x) = \mu$ and $\sum_{allx} p(x) = 1$ so we get:

$$V(X) = \sum_{allx} x^2 p(x) - 2\mu^2 + \mu^2 = \mu_2 - \mu^2$$

Where $\mu_2 = \sum_{allx} x^2 p(x)$, the average X^2, and is conveniently written $E(X^2)$.

For the second dealer the variance is:

$E(X^2) = 0^2 \times 0.1 + 1^2 \times 0.1 + 2^2 \times 0.4 + 3^2 \times 0.4 = 5.3$
$\mu = 2.1$ so $\mu^2 = 4.41$
Thus $V(X) = 5.3 - 4.41 = 0.89$.

Properties of Variance

Although the proofs here are not difficult, this is not a course in algebra so we just state these.

a is a real number.

$V(a+X)=V(X)$ adding a constant to a random quantity does not change its variation.

$V(aX) = a^2V(X)$ variance is scale dependent.

If X_1 and X_2 are independent then:

$$V(X_1 + X_2) = V(X_1) + V(X_2)$$

this extends to arbitrary independent sums.

The only problem with V(X) are the units. The mean in the examples is measured in cars, but the variance is measured in $cars^2$ (whatever they might be).

To get back to the original units we define the *standard deviation.*

$$SD(X) = \sqrt{V(X)}.$$

In fact V(X) is a more natural measure (mathematically) but SD(X) is more interpretable.

The two quantities mean and variance are sufficiently important that traditionally we allocate standard Greek letters to them:

$\mu = E(X)$ and $\sigma^2 = V(X)$

note: $V(X) \geq 0$, in fact if V(X)=0 there is no variation X is just a constant.

Variances of standard distributions

Bernoulli(p).

$$E(X^2) = 0^2 \times (1-p) + 1^2 \times p = p$$

so:

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

Note:

the variance is 0 if p=0 or p=1 because then we just get the same value every time!

The Bernoulli distribution is most variable when p=1/2 .

Binomial

We have shown that the Binomial is a sum of n Bernoulli(p), further these are independent, So:

Y~ B(n,p)

$$V(Y) = \sum_{i=1}^{n} V(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

Geometric

Again a messy series computation gives

$$V(Y) = \frac{1}{p^2}$$

Poisson

Not as nice to compute $E(X^2)$ here. In fact the trick is to compute $E(X(X-1))=E(X^2)-E(X)$ and get $E(X^2)$ from this.

Try it if you like…and if you have done this correctly you will get:

$$V(Y) = \lambda$$

Moments

We came across $E(X^2)$ as a convenient idea in computing variance, we can define general *moments* for any distribution.

$$\mu_p = E(X^p) = \sum_{allx} x^p p(x)$$

p doesn't even have to be an integer, but these are not used in simple statistical ideas so we do not pursue them further.

But it is useful to note that we can take expectations of arbitrary functions of X.

$E(f(X))$, E(ln(X)), $E\left(\frac{1}{X}\right)$ etc. remembering though that we can't take the E() inside a non-linear function.