Two Important Distributions

Today we will study various aspects of two important distributions.

- *Binomial Distribution*: This discrete distribution is given by the condition $P(X = r) = \binom{k}{r} p^r (1-p)^{k-r}$, or equivalently:

$$P(X \leq r) = \sum_{s=0}^{r} \binom{k}{s} p^s (1-p)^{k-s}$$

  The random variable $X = B(k; p)$ is the one that counts the number of Heads in k independent coin flips where the coin has probability p of landing Head. There are two parameters k and p in this distribution.

- *Negative Binomial Distribution*: This discrete distribution is given by the condition $P(Y = r) = \binom{k+r-1}{r} p^r (1-p)^k$ or equivalently:

$$P(Y \leq r) = \sum_{s=0}^{r} \binom{k+s-1}{s} p^s (1-p)^k$$

  The random variable $Y = NB(k; p)$ is the one that counts the number Heads that occur *before* k Tails occur in a sequence of independent coin flips where the coin has probability p of landing head. There are two parameters k and p in this distribution.

The second one may seem unfamiliar, but we have seen it before for the case $k = 1$ in a slightly different form. If $Y = B(1; p)$, then $W = Y + 1$ counts the number of flips required in order to get one Tail.

We can see these distributions in other contexts.

Suppose that in a Biology experiment we know that a certain percentage of the population has undergone mutation. We examine the genetic information of 1000 samples and find that 50 carry the mutation marker. Can we justify the guess (estimate) that the mutation percentage is 5 percent? How accurate is this statement? What is the probability that this estimate is off by more than 1 percent?

Suppose that a physics lab has 10 special kinds of diodes. Each time a certain experiment is run, there is a probability of p that the diode burns out and has to be replaced (and the experiment is a failure!). How many times can we expect to run the experiment before we need to go and buy more diodes?

Computational tricks

The names of these distributions come from the generalised Binomial theorem. We know the formula:

$$(1+t)^k = \sum_{r=0}^{k} \binom{k}{r} t^r$$

We can also write this as:

$$(1+t)^k = \sum_{r=0}^{\infty} \binom{k}{r} t^r$$

The explanation is that $\binom{k}{r} = 0$ if $r > k$. In fact, it is convenient to define

$$\binom{x}{r} = \frac{x \cdot (x-1) \cdots (x-r+1)}{r!}$$

for *any* number x and *positive integer* r; we also define $\binom{x}{0} = 1$. Newton recognised (without proof!) that the series also makes sense for x real (or exen complex) provided we restrict $|t| < 1$:

$$(1+t)^x = \sum_{r=0}^{\infty} \binom{x}{r} t^r$$

In particular,

$$\frac{1}{(1+t)^k} = (1+t)^{-k} = \sum_{r=0}^{\infty} \binom{-k}{r} t^r$$

Now, we note that

$$\binom{-k}{r} = \frac{(-k) \cdot (-k-1) \cdots (-k-r+1)}{r!} = (-1)^r \binom{k+r-1}{r}$$

Thus, we get the formula:

$$\frac{1}{(1-t)^k} = \sum_{r=0}^{\infty} \binom{k+r-1}{r} t^r$$

Applying $t(d/dt)$ a number of times to these formulae we get:

$$\left(t\frac{d}{dt}\right)^a (1+t)^k = \sum_{r=0}^{k} \binom{k}{r} t^r r^a$$

and:

$$\left(t\frac{d}{dt}\right)^a (1-t)^{-k} = \sum_{r=0}^{k} \binom{k+r-1}{r} t^r r^a$$

This allows us to compute:

$$E(X^a) = \sum_{r=0}^{k} \binom{k}{r} p^r (1-p)^{k-r} r^a = (1-p)^k \left(t\frac{d}{dt}\right)^a (1+t)^k \bigg|_{t=p/(1-p)}$$

Similarly,

$$E(Y^a) = \sum_{r=0}^{\infty} \binom{k+r-1}{r} p^r (1-p)^k r^a = (1-p)^k \left(t\frac{d}{dt}\right)^a (1-t)^{-k} \bigg|_{t=p}$$

In particular, we get, for $a = 1$, after a little calculation:

$$E(X) = (1-p)^k \frac{p}{1-p} \frac{k}{(1-p)^{k-1}} = kp \text{ and } E(Y) = (1-p)^k p \frac{k}{(1-p)^{k+1}} = \frac{kp}{1-p}$$

For $a = 2$, the calculation is a little more involved:

$$E(X^2) = (1-p)^k \frac{kp}{1-p} \frac{kp + (1-p)}{(1-p)^{k-1}} = kp(kp + (1-p)) = (kp)^2 + kp(1-p)$$

This gives us

$$\sigma^2(X) == E(X^2) - E(X)^2 = (kp)^2 + kp(1-p) - (kp)^2 = kp(1-p)$$

Similarly,

$$E(Y^2) = (1-p)^k kp \frac{kp+1}{(1-p)^{k+2}} = \frac{kp(kp+1)}{(1-p)^2}$$

This gives us

$$\sigma^2(Y) = E(Y^2) - E(Y)^2 = \frac{(kp)^2 + kp}{(1p)^2} - \left(\frac{kp}{1-p}\right)^2 = \frac{kp}{(1-p)^2}$$