BINOMIAL DISTRIBUTION

Consider an experiment consisting of n independent "trials", each of which has 2 possible outcomes:

E.g. a series of n coin flips : H or T

... dice rolls : 6 or not-6

Students taking a test : Pass/Fail

Molecules of a gas inside/outside some volume v within a vessel.

The binomial distribution gives the probability of k "successes" in n trials.

EXAMPLE 1 : # of heads in 3 coin tosses

outcomes	# heads, k	P_k
TTT	0	$(\frac{1}{2})^3$
TTH, THT, HTT	1	$3\times(\frac{1}{2})^2\times(\frac{1}{2})$
THH, HTH, HHT	2	$3\times(\frac{1}{2})\times(\frac{1}{2})^2$
HHH	3	$(\frac{1}{2})^3$

[BINOMIAL DIST, CONTINUED]

EXAMPLE 2 : # of sixes in 2 dice rolls

outcomes	# of sixes, k	P_k
XX	0	$(\frac{5}{6})^2$
6X, X6	1	$2\times\frac{1}{6}\times\frac{5}{6}$
66	2	$(\frac{1}{6})^2$.

GENERAL CASE : Each trial has probability p of "success" and q [$\equiv 1-p$] of "failure". One possibility is that we have k successes followed by $(n-k)$ failures, with prob. $\underbrace{p\times p\times\cdots p}_{k\text{ times}}\times\underbrace{q\times\cdots q}_{n-k\text{ times}}$

But the successes and failures could occur in any one of $\frac{n!}{k!(n-k)!}=\binom{n}{k}$ ways,

so the total probability of k successes will be

$$\boxed{P_k=\binom{n}{k}p^k q^{n-k}}$$ BINOMIAL DISTRIBUTION

[BINOMIAL DISTRIBUTION, CONT.]

In EXAMPLE 1, $n=3$ and $p=q=\frac{1}{2}$.

Then, for example, $P_2 = \binom{3}{2} \times \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{2}\right) = \frac{3}{8}$.

In EXAMPLE 2, $n=2$ and $p=\frac{1}{6}$, $q=\frac{5}{6}$

$$\Rightarrow P_1 = \binom{2}{1} \times \frac{1}{6} \times \frac{5}{6} = \frac{5}{18}.$$

CHECK: As always for a probability distribution, the probabilities must add up to 1.

$$\sum_{k=0}^{n} P_k = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$$

$$= (p+q)^n \quad \text{[BINOMIAL THEOREM]}$$

$$= 1 \quad [\text{since } p+q=1]$$

We often need to know the expected [mean] number of successes $\langle k \rangle$ and the variance $\sigma_k^2 = \mathrm{var}(k) = \langle k^2 \rangle - \langle k \rangle^2$.

MEAN OF BINOMIAL DIST.:

Expect this to be np, as the relative frequency of success in repeated single trials is p.

CHECK:
$$\langle k \rangle = \sum_{k=0}^{n} k P_k$$

$$= \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k}$$

$$= q^n \sum_{k=0}^{n} \binom{n}{k} k x^k \quad \text{where } x = \frac{p}{q}$$

But $k x^k = x \frac{d}{dx}(x^k)$,

$$\Rightarrow \langle k \rangle = q^n \times \frac{d}{dx} \sum_{k=0}^{n} \binom{n}{k} x^k 1^{n-k}$$

$$= q^n \times \frac{d}{dx} (x+1)^n \quad \text{BINOMIAL THEOREM}$$

$$= q^n x \cdot n (x+1)^{n-1}$$

$$= q^n \cdot \frac{p}{q} \cdot n \left(\frac{p}{q} + 1\right)^{n-1}$$

$$= p \cdot n (p+q)^{n-1}$$

$$= np \quad (\text{since } p+q=1)$$

VARIANCE OF BINOMIAL DIST.:

We follow the same approach that we used with the Poisson distribution, i.e., look at

$$\langle k(k-1)\rangle = \langle k^2\rangle - \langle k\rangle$$
$$= [\sigma_k^2 + \langle k\rangle^2] - \langle k\rangle.$$

<u>Details</u>:

$$\langle k(k-1)\rangle = \sum_{k=0}^{n} k(k-1)\binom{n}{k} p^k q^{n-k}$$
$$= q^n \sum_{k=0}^{n} \binom{n}{k} \underbrace{k(k-1)\,x^k}_{x^2 \frac{d^2}{dx^2} x^k}$$
$$= q^n x^2 \frac{d^2}{dx^2} \sum_{k=0}^{n} \binom{n}{k} x^k$$
$$= q^n x^2 \cdot n(n-1)(x+1)^{n-2}$$
$$= p^2 \cdot n(n-1)(p+q)^{n-2}$$
$$= n(n-1)p^2$$

Hence

$$\sigma_k^2 = n(n-1)p^2 - \langle k\rangle^2 + \langle k\rangle$$
$$= (n^2p^2 - np^2) - n^2p^2 + np$$
$$= n(p-p^2) = npq.$$

$$\boxed{\text{SUMMARY:}\quad \langle k\rangle = np \qquad \sigma_k^2 = npq}$$

Note that mean is proportional to n and the "width" of the distribution, σ_k, varies as $\sqrt{n}$ so that "fractional width" $\sigma_k/\langle k\rangle \propto \frac{1}{\sqrt{n}}$ decreases with increasing n: the binomial distribution becomes [relatively] sharply peaked about its mean.

For large n, P_k can be approximated by a GAUSSIAN DISTRIBUTION of mean np and s.d. $(npq)^{1/2}$.

[Proof is similar to large λ limit of Poisson, with Stirling's approx. used for each of the factorials in the binomial coefficients $\frac{n!}{k!\,(n-k)!}$.]

EXAMPLE 3: Each of the 3 lifts in the Schuster lab may fail with probability 0·1 on a given day. Calculate:

(a) the probability that only one lift is working all day;

(b) the probability that at least one lift fails;

(c) the mean number of failed lifts.

SOLUTION: $n = 3$, $p = 0{\cdot}1$, $q = 0{\cdot}9$

(a) $P_2 = \binom{3}{2} p^2 q$

$= 3 \times 0{\cdot}1^2 \times 0{\cdot}9 = 0{\cdot}027$

(b) $P_1 + P_2 + P_3 = 1 - P_0$

$= 1 - (0{\cdot}9)^3 = 0{\cdot}271$

(c) $np = 3 \times 0{\cdot}1 = 0{\cdot}3$