Binomial/Poisson Distributions Summary

Binomial	Poisson
Characteristics: (i) n **independent** trials, whereby each trial yields **two possible outcomes**- a fixed constant probability of success p or failure $q = 1 - p$ (ii) Purely **discrete**	Characteristics: (i) Events occur **randomly** and **singly** (ii) A single parameter λ (average number of occurences) defines the distribution and is **proportional** to the frame of measurement (eg $\lambda = 1$ for 1 week and therefore $\lambda = 2$ for 2 weeks) (iii) Purely **discrete**
Definition: $X \sim B(n, p)$	Definition: $X \sim P_0(\lambda)$
Formula: $P(X = r) = {}^nC_r (p)^r (q)^{n-r}$	Formula: $P(X = r) = \dfrac{e^{-\lambda}(\lambda)^r}{r!}$
Mean: $\mu = np$ Variance: $\sigma^2 = npq$	Mean: $\mu = \lambda$ Variance: $\sigma^2 = \lambda$
Separate binomial distributions typically **cannot** be pooled together	For two **independent** poisson distributions defined under the **same** frame of measurement, eg $X \sim P_0(\lambda)$ and $Y \sim P_0(\mu)$, they can be **pooled together** to form a consolidated poisson model whereby $X + Y \sim P_0(\lambda + \mu)$
Example Scenario: number of sixes obtained during ten throws of an unbiased dice.	Example Scenario: number of defects along a 10m long piece of cloth manufactured in a factory.
Graphic calculator commands : $P(X = r) \rightarrow binompdf$ $P(X \leq r) \rightarrow binomcdf$	Graphic calculator commands : $P(X = r) \rightarrow poissonpdf$ $P(X \leq r) \rightarrow poissoncdf$

(Note that *pdf* stands for **probability density function**, while *cdf* stands for **cumulative density function**.)

Binomial to Poisson Approximation:

For a binomial distribution whereby $X \sim B(n, p)$, **IF** $np < 5$ and $p < 0.1$, then

$X \sim P_0(np, npq)$ **approximately**. (Note that there is **NO** poisson to binomial approximation)

NO continuity correction is required since this is a discrete to discrete approximation.

Some important miscellaneous concepts:

(a) It is advisable, amongst all things, to appreciate and convert to memory that $P(x = 0) = q^n$ for a binomial distribution and $P(X = 0) = e^{-\lambda}$ for a poisson distribution.

Example of its relevance:

For $X \sim B(n,\ 0.05)$, find the least value of n such that $P(X \geq 1) > 0.99$ without employing any explicit GC statistical commands.

Workings: $P(X \geq 1) > 0.99 \rightarrow 1 - P(X = 0) > 0.99$

$$1 - (1 - 0.05)^n > 0.99$$

(continue the solving process)

(b) It is very usual for questions to embed a binomial/poisson distribution within another binomial/poisson distribution-students must be sufficiently discerning to relate one part of the context to another. Put it simply, such questions are **multi-layered**.

Example:

Eggs are packed in boxes of 500. On average, **0.8%** of the eggs are found to be broken when the eggs are unpacked.

(i) Find the probability that in a box of **500 eggs**, exactly 3 will be broken.

Let the random variable X denote the number of broken eggs within a box of 500. Then $X \sim B(500,\ 0.008)$

(ii) A supermart unpacks **100 boxes of eggs**. What is the probability that there will be exactly 4 boxes containing no broken eggs?

Now we are **zooming** out and focusing on a more general picture of boxes of eggs. Let the random variable Y denote the number of boxes containing no broken eggs. Then $Y \sim B(100,\ p)$ where $p = P(X = 0)$ based on the random variable X defined in (i).

(c) While less common, finding the **mode** of a distribution via non GC methods can be examined-a detailed worked example will illustrate this better:

The random variable X is the number of successes in 200 independent trials of an experiment in which the probability of success at any one trial is p. Given that $E(X^2) = 10.6008$, find the **exact** value of p and show that

$$\frac{P(X=k+1)}{P(X=k)} = \frac{7(200-k)}{493(k+1)}, \text{ for } k = 0,1,2,\ldots\ldots\ldots,199$$

(i) Hence find the value of $k$such that $P(X = k)$ is the maximum.

(ii) Using a Poisson approximation, find the probability that more than 198 of the 200 trials were not successful.

SOLUTIONS :

$X \sim B(200, p)$

Since $E(X^2) = 10.60008$,

Then $Var(X) = E(X^2) - [E(X)]^2 = 10.6008 - (200\text{p})^2$

$$200(p)(1-p) = 10.6008 - (200\text{p})^2$$

$$200\text{p} - 200p^2 = 10.6008 - 40000\text{p}^2$$

$$39800\text{p}^2 + 200\text{p} - 10.6008 = 0$$

Solving gives $p = 0.014$ (shown)

$$\frac{P(X = k+1)}{P(X = k)} = \frac{\binom{200}{k+1} p^{k+1}(1-p)^{200-k-1}}{\binom{200}{k} p^k (1-p)^{200-k}}$$

$$= \frac{\binom{200}{k+1} 0.014^{k+1}(0.986)^{199-k}}{\binom{200}{k} 0.014^k (0.986)^{200-k}}$$

$$= \frac{\left[\frac{200!}{(k+1)!(199-k)!}\right]0.014^{k+1}(0.986)^{199-k}}{\left[\frac{200!}{(k)!(200-k)!}\right]0.014^{k}(0.986)^{200-k}}$$

$$= \frac{(200-k)}{(k+1)}\frac{0.014}{0.986}$$

$$= \frac{(200\text{-k})}{(\text{k}+1)}\frac{14}{986} = \frac{7(200-k)}{493(k+1)} \text{ (shown)}$$

(i) If $P(X = k + 1) > P(X = k)$,

then $\frac{P(X=k+1)}{P(X=k)} > 1 \Rightarrow \frac{7(200-k)}{493(k+1)} > 1$

where solving the inequality gives $k < 1.814$

Hence, we have $P(X = 2) > P(X = 1) > P(X = 0)$------------------(1)

On the other hand, if $P(X = k + 1) < P(X = k)$,

then $\frac{P(X=k+1)}{P(X=k)} < 1 \Rightarrow \frac{7(200-k)}{493(k+1)} < 1$

where solving the inequality gives $k > 1.814$

Hence, we have

$P(X = 200) < P(X = 199) < P(X = 198)............ < P(X = 4) < P(X = 3) < P(X = 2)$----------(2)

Reconciling (1) and (2) therefore gives $P(X = 2)$as the maximum, ie $k = 2$ (shown)

(ii) Since $np = 200(0.014) = 2.8 < 5$ and $p = 0.014 < 0.1$,

$X \sim P_0(2.8)$ approximately

P(more than 198 of the 200 trials were not successful)

$= P(X \leq 1) = 0.231$(shown)

(Note: Care must be exercised in interpreting the random variable correctly; X was defined from the start as the number of **successful** trials, **NOT unsuccessful** trials .)