Statistics of binomial distribution

If $X \sim Bi(n,p)$ then:

→ Mean $= E(x) = np$

→ Variance $= \sigma^2 = npq$

→ Standard deviation $= \sqrt{npq}$

examples:

① For a random variable $X \sim Bi(20, 0.4)$ calculate $E(x)$, $Var(x)$ and $Sd(x)$

$E(x) = 20 \times 0.4 = 8$

$Var(x) = 8 \times 0.6 = 4.8$

$Sd(x) = \sqrt{4.8} = 2.19$

② If random variable $X \sim (n,p)$ has mean of 12 and a variance of 4, Find n & p

$np = 12$ ①

$np(1-p) = 4$ ②

① ÷ ② → $\frac{1}{1-p} = 3$

$3 - 3p = 1$

$-3p = -2$

$p = \frac{2}{3}$

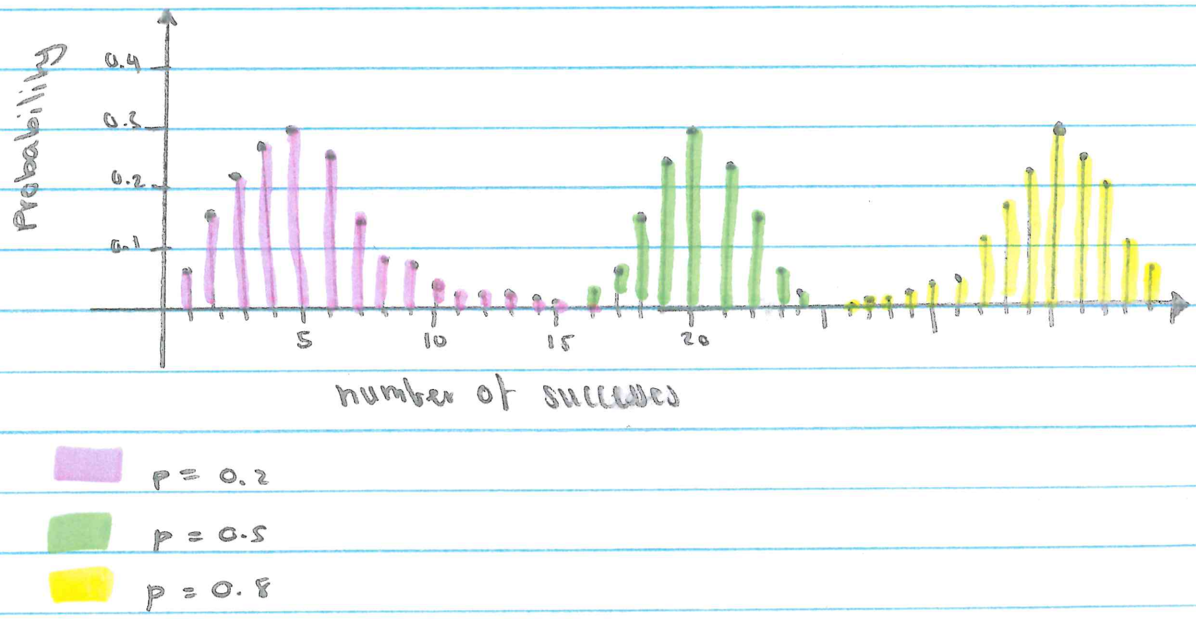
when $p = \frac{2}{3}$, $n = 18$

Effect of varying p:

→ For a small p value ($p < 0.5$) the graph is skewed to the right. This is known as a positive skew. This means there is a greater probability of obtaining lower outcomes and a smaller probability of obtaining higher outcomes

→ When $p = 0.5$ the graph is symmetrical.

→ For a large p value ($p > 0.5$) the graph is skewed to the left. This is known as a negative skew. This means there is a greater probability of obtaining higher outcomes and a smaller probability of obtaining lower outcomes



Finding the sample size

When solving these types of problems with the calculator write the equation, not the inequality.

example:

① For a particular game, the probability of winning is known to be 0.4. What is the least number of times an individual must play to ensure that:

a) The probability of winning at least twice is more than 0.7?

$X \sim Bi(n, 0.4)$

$Pr(x > 1) = Pr(x \geq 2)$

$Pr(x \geq 2) = 1 - Pr(0) - Pr(1)$

$0.7 = 1 - \binom{n}{0}(0.4)^0(0.6)^n - \binom{n}{1}(0.4)^1(0.6)^{n-1}$

$= 1 - 0.6^n - n(0.4)(0.6)^{n-1}$

$n = -1.26$, $n = 5.32$

$\Rightarrow n = 6$ minimum

NOTE:

$\binom{n}{0} = 1$

$\binom{n}{1} = n$

$\binom{n}{n} = 1$

Continuous random variable

A continuous random variable is a variable that can take any value within an interval of the real number line.

For a continuous random variable:

$\Pr(X=x)=0$

Therefore, for $f(x)$ to be a probability density function of a continuous random variable X, we require

→ $f(x) \geq 0$ for all real numbers x

* the curve is positioned above the x-axis and since f does not give probabilities f may take values greater than one.

→ the total area under the graph of $f(x)$ is 1

i.e $\int_{-\infty}^{\infty} (f(x))\,dx = 1$

To find a probability:

$\Pr(a \leq X \leq b) = \int_a^b f(x)\,dx$

example

① A density function has rule: $f(x) = \begin{cases} ke^{-2x} & , x>0 \\ 0 & , x \leq 0 \end{cases}$

a) Find the value of k:

$\int_0^{\infty} ke^{-2x}\,dx = 1$

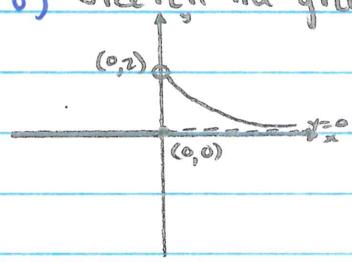
$\left[\frac{-k}{2}e^{-2x}\right]_0^{\infty} = 1$

$\lim_{x\to\infty}\left(\frac{-k}{2}e^{-2x}\right) - \left(\frac{-k}{2}\right) = 1$

$0 + \frac{k}{2} = 1$

$k = 2$

b) Sketch the graph



y int. → $2e^0 = 2$

c) Find $\Pr(X>1)$.

$\Pr(X>1) = 1 - \Pr(X<1)$

$= 1 - \int_0^1 2e^{-2x}\,dx$

$= 1 - \left[-e^{-2x}\right]_0^1$

$= 1 - \left(\left(\frac{-1}{e^2}\right) - (-1)\right)$

$= 1 - \left(\frac{-1}{e^2} + 1\right)$

$= \frac{1}{e^2}$

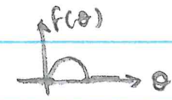
② $f(\theta) = \begin{cases} \sin(2\theta) & 0 \leq \theta \leq \frac{\pi}{2} \\ 0 & , \text{elsewhere} \end{cases}$

Show that $f(\theta)$ is a probability function

$\int_0^{\pi/2} \sin(2\theta)\,d\theta = \left[-\frac{\cos(2\theta)}{2}\right]_0^{\pi/2}$

$= 1$

$f(\theta) \geq 0$ for all θ values



⇒ It is a probability function.

Mean of continuous random variable

The mean or the expected value of a continuous random variable x is given by:

$$\mu = E(x) = \int_{-\infty}^{\infty} x f(x)\, dx$$

Similarly the expected value of $g(x)$ i.e. $E(g(x))$ is given by

$$E(g(x)) = \int_{a}^{\infty} g(x) \times f(x)\, dx$$

i.e. $E(x^2) = \int_a^b x^2 \times f(x)\, dx$

NOTE:

$E(aX + b) = aE(x) + b$ (as for discrete variables)

examples:

① A random variable Y has pdf given by $f(y) = \begin{cases} ky, & 0 \le y \le m \\ 0, & \text{otherwise} \end{cases}$

Find k and m if the mean of Y is 2.

$\int_0^m ky\, dy = 1$

$\left[\frac{k}{2} y^2\right]_0^m = 1$

$\left(\frac{km^2}{2}\right) = 1$

$km^2 = 2$ ①

$\int_0^m ky^2\, dy = 2$

$\left[\frac{k}{3} y^3\right]_0^m = 2$

$\left(\frac{km^3}{3}\right) = 2$

$km^3 = 6$ ②

② ÷ ① $\longrightarrow$ $m = 3$

resub $m = 3$ into ①

$\longrightarrow$ $9k = 2$

$k = 2/9$

② If x is the random variable with probability function f:

$$f(x) = \begin{cases} 0.5x, & 0 \le x \le 2 \\ 0, & x < 0 \text{ or } x > 2 \end{cases}$$

a) Find the expected value of X.

$E(x) = \int_0^2 x \times 0.5x\, dx$

$= \int_0^2 0.5x^2\, dx$

$= \left[\frac{1}{2} \times \frac{1}{3} x^3\right]_0^2$

$= \left[\frac{x^3}{6}\right]_0^2$

$= \frac{8}{6}$

$= 4/3$

b) The expected value of e^x

$E(e^x) = \int_0^2 e^x \times 0.5x\, dx$

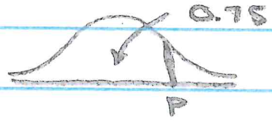
$= 4.1945$

(on calc)

Percentiles and Medians

A percentile is a value of x which bounds a particular area under a pdf. E.g. if X represents student marks then the mark (p) below which lies 75% of all students' marks is called the 75th percentile, and is found by solving:

$$\int_{-\infty}^{p} f(x)\,dx = 0.75$$



whole area = 1

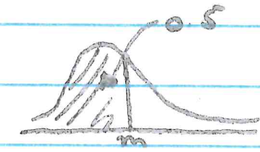
The median which is another measure of centre is also a percentile. Since the median divides the distribution in half (middle), it is referred to as the 50^{th} percentile. That is 50% of all values lie below the median and 50% lie above.

Therefore, the median, m, is the value such that:

$$P_r(X \le m) = \int_{-\infty}^{m} f(x)\,dx = 0.5$$

OR $$P_r(X \ge m) = \int_{m}^{\infty} f(x)\,dx = 0.5$$

OR $$\int_{m}^{\infty} f(x)\,dx = \int_{-\infty}^{m} f(x)\,dx$$



example

① $f(x) = \begin{cases} 2(1-x), & 0 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases}$

②

Find the median value of x

$\int_{0}^{m} 2-2x\,dx = 0.5$

$[2x - x^2]_0^m = 0.5$

$2m - m^2 = 0.5$

$-2m^2 + 4m - 1 = 0$

$2m^2 - 4m + 1 = 0$

$m = 1.702$ or $m = 0.2928$

$m = 1.702$ because $0 \le m \le 1$

Range:

The range is the max x value − min x value it can take

example:

① $f(x) = \begin{cases} 0.5x, & 0 \le x \le 2 \\ 0, & x < 0 \text{ or } x > 2 \end{cases}$

range $= 2 - 0$

$= 2$