www.cwthompson.com

Binomial Distribution

Binomial Distribution Properties:

- Fixed number of trials (n)
- Two outcomes for each trial, success or failure
- Trials are independent
- Probability of success, p, at each trial is constant

Always define distribution at start of each question

e.g. Let X = number of successes in n trials

$X \sim B(n, p)$. where n = number of trials

p = probability

Formula:

$$\boxed{p(X=x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}}$$

Mean = $n \cdot p$

Variance = $n \cdot p \cdot (1-p)$

Probability of first success on n^{th} trial

A fair die is rolled 8 times. Find the probability the first 6 occurs on the 4^{th} roll.

Let X = number of 6s

$X \sim B(4, \frac{1}{6})$

$p = \frac{\binom{4}{1}\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^3}{4}$

$= 0.09645...$

$= 0.0965$ (3sf)

Probability of exactly n successes

A cadet fires 10 shots with a 75% chance of hitting the target. Find the probability that exactly 6 shots hit the target.

Let X = number of shots on target

$X \sim B(10, 0.75)$

$p(X=6) = \binom{10}{6}(0.75)^6(0.25)^4$

$= 0.146$ (3sf)

www.cwthompson.com

Cumulative probability tables above 0.5

A cadet fires 10 shots with a 75% chance of hitting the target. Find the probability that at least 8 shots hit the target.

Let X = number of shots on target $\quad X \sim B(10, 0.75)$

Y = number of shots that miss $\quad Y \sim B(10, 0.25)$

$p(X \geq 8) = p(Y \leq 2)$

using tables

$= 0.5256$

Finding independent probability

A cadet fires 20 shots and has an 80% chance of hitting the target at least once. Find the minimum independent probability.

Let X = number of shots on target

$X \sim B(20, p)$

$p(x \geq 1) = 1 - p(x=0)$

$= 0.8$

$p(x=0) = 0.2$

$(1-p)^{20} = 0.2$

$1 - p = 0.2^{1/20}$

$p = 1 - 0.2^{1/20}$

$= 0.0773191\ldots$

$= 0.0773$ (3sf)

Poisson Distribution

Poisson Distribution Properties:

- Events occur at random
- Events occur singly in space or time
- Constant rate/mean
- Independent events

Always define distribution at start of each question

e.g. Let X = number of events in time/space period

$X \sim P_o(\lambda)$ where λ = mean rate

Formula:

$$\boxed{p(X = x) = \frac{e^{-\lambda} \cdot \lambda^{x}}{x!}}$$

mean = λ

variance = λ

Changing the Mean

Patients arrive at a hospital at a random rate of 6 per hour. Define the distribution for over the next 15 minutes.

Let X = number of patients arriving in next hour

$X \sim P_o(6)$

$60/15 = 4$

$6/4 = 1.5$

$\therefore$ Let Y = number of patients arriving in next 15 minutes

$\therefore \; Y \sim P_o(1.5)$

Using Tables

In a village, power cuts occur randomly at a rate of 3 per year. Find the probability that in any given year there will be at least 4 power cuts.

Let X = number of power cuts in a year

$X \sim Po(3)$

$p(X \geq 4) = 1 - p(X \leq 3)$

using tables

$= 1 - 0.6472$

$= 0.3528$

www.cwthompson.com

Continuous Random Variables

- A continuous variable is a variable that can take any value, usually in a given range
- A probability density function (p.d.f), $f(x)$, can be used to calculate the probability in a continuous range

Continuous Random Variable p.d.f. Properties

- $f(x) \geq 0$ for all values of x
- $p(a < x < b) = \int_a^b f(x)\,dx$
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$

e.g. The continuous random variable has the pdf:

$$f(x)\begin{cases} k(4-x) & 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Find the value of k.

$$\int_0^2 k(4-x) = 1$$
$$k\int_0^2 4-x = 1$$
$$k\left[4x - \frac{x^2}{2}\right]_0^2 = 1$$
$$6k = 1$$
$$k = \frac{1}{6}$$

e.g. The continuous random variable has the pdf:

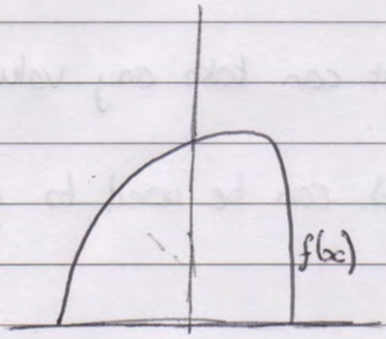
$$f(x)\begin{cases} \frac{1}{6}(4-x) & 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Find $p(0.5 < x < 1)$

$$\int_{0.5}^{1} \frac{1}{6}(4-x) = \frac{1}{6}\left[4x - \frac{x^2}{2}\right]_{0.5}^{1}$$
$$= \frac{1}{6}\left(\frac{7}{2} - \frac{15}{8}\right)$$
$$= \frac{1}{6}\left(\frac{13}{8}\right)$$
$$= \frac{13}{48}$$

www.cwthompson.com

Mean, Variance and Mode



- The mode is where the pdf is at its highest point
- Mean: $\int_a^b x f(x)$
- Variance: $\int_a^b x^2 f(x) - \left[\int_a^b x f(x)\right]^2$

Cumulative Distribution Function

- Gets the cumulative value of a pdf
- $F(x) = \int_a^x f(x)$

e.g. The random variable has pdf

$$f(x) \begin{cases} x/4 & 1 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$

Find the cdf

$$F(x) = \int_1^x \frac{x}{4}$$
$$= \left[\frac{x^2}{8}\right]_1^x$$
$$= \frac{x^2}{8} - \frac{1}{8}$$

$$\therefore F(x) \begin{cases} 0 & x < 1 \\ \frac{x^2}{8} - \frac{1}{8} & 1 \le x \le 3 \\ 1 & x > 3 \end{cases}$$

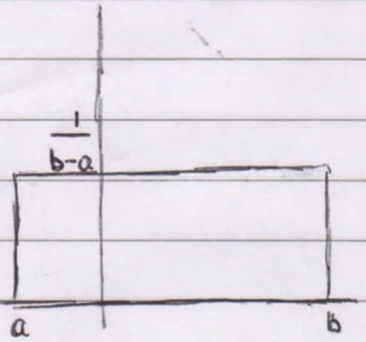
Median and Percentiles

- The c.d.f can be used to find percentiles of functions
- The median is given by $F(m) = 0.5$
- The lower quartile is given by $F(Q_1) = 0.25$
- The upper quartile is given by $F(Q_3) = 0.75$
- This can be done for any percentile

www.cwthompson.com

Continuous Uniform Distribution

- A distribution is uniform if each value in the given range of values has the same probability of occuring

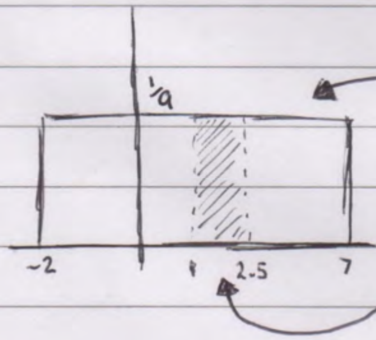


$X \sim U(a, b)$

$$f(x) \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Finding probability in a range

e.g. $X \sim U(-2, 7)$, find $p(1 < x < 2.5)$



- Sketch the distribution
- Find the constant probability

$\frac{1}{b-a} = \frac{1}{7+2} = \frac{1}{9}$

- Find the range of the question's probability

$b_1 - a_1 = 2.5 - 1 = 1.5$

- Multiply the two values together

$\frac{1}{9} \times 1.5 = \frac{1}{6}$

$\therefore p(1 < x < 2.5) = \frac{1}{6}$

The mean of continuous uniform distribution

$$E(x) = \int_{-\infty}^{\infty} x f(x)$$
$$= \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha}$$
$$= \left[\frac{\frac{1}{2}x^2}{\beta - \alpha} \right]_{\alpha}^{\beta}$$
$$= \frac{1}{2} \frac{\beta^2 - \alpha^2}{\beta - \alpha}$$
$$= \frac{1}{2}(\beta + \alpha)$$

The mean is the midpoint of the range, which is also the median.

www.cwthompson.com

The variance of continuous uniform distribution

$$E(x^2) = \int_{\alpha}^{\beta} \frac{x^2}{\beta - \alpha}$$
$$= \left[\frac{\frac{1}{3}x^3}{\beta-\alpha}\right]_{\alpha}^{\beta}$$
$$= \frac{1}{3}\,\frac{\beta^3-\alpha^3}{\beta-\alpha}$$
$$= \frac{1}{3}(\beta^2+\beta\alpha+\alpha^2)$$

$$Var(x) = E(x^2) - [E(x)]^2$$
$$= \frac{1}{3}(\beta^2+\beta\alpha+\alpha^2) - \left[\frac{1}{2}(\beta+\alpha)\right]^2$$
$$= \frac{1}{3}(\beta^2+\beta\alpha+\alpha^2) - \frac{1}{4}(\beta^2+2\beta\alpha+\alpha^2)$$
$$= \frac{1}{12}\left[4(\beta^2+\beta\alpha+\alpha^2) - 3(\beta^2+2\beta\alpha+\alpha^2)\right]$$
$$= \frac{1}{12}\left[4\beta^2+4\beta\alpha+4\alpha^2-3\beta^2-6\beta\alpha-3\alpha^2\right]$$
$$= \frac{1}{12}\left[\beta^2-2\beta\alpha+\alpha^2\right]$$
$$= \frac{1}{12}(\beta-\alpha)^2$$

www.cwthompson.com

Approximations

Continuity Correction

- Used to change values from discrete to continuous
    1) Write probability in terms of $\leq$ or $\geq$
    2a) For $p(X \leq n)$ you approximate to $p(Y \leq n + 0.5)$
    2b) For $p(X \geq n)$ you approximate to $p(Y \geq n - 0.5)$
- Examples:
    - $p(X \geq 11) = p(Y \geq 10.5)$
    - $p(X < 4) = p(X \leq 3) = p(Y \leq 3.5)$
    - $p(X > 7) = p(X \geq 8) = p(Y \geq 7.5)$
    - $p(2 < X < 5) = p(3 \leq X \leq 4) = p(2.5 \leq$

Binomial to Normal

- A binomial to normal approximation can be used if:
    - n is large
    - p is close to 0.5
- $X \sim B(n, p)$ is approximated to $X \sim N(np, np(1-p))$

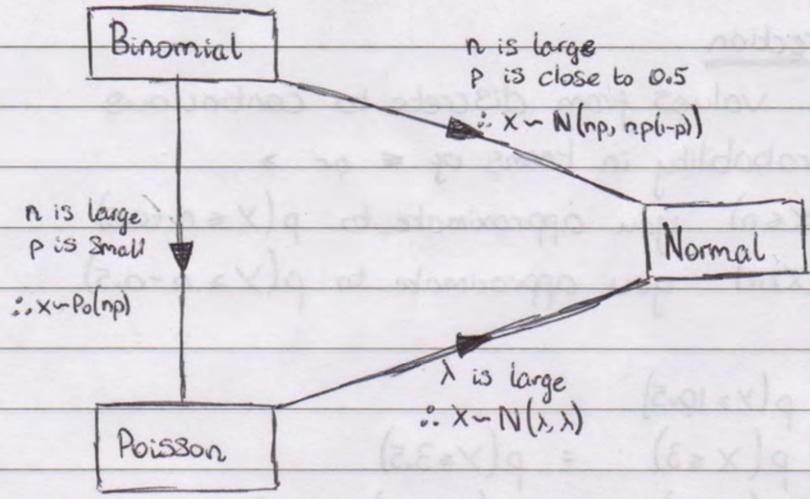
Poisson to Normal

- A poisson to normal approximation can be used if:
    - λ is large
- $X \sim P_o(\lambda)$ is approximated to $X \sim N(\lambda, \lambda)$

Binomial to Poisson

- A binomial to poisson approximation can be used if:
    - n is large
    - p is small
- $X \sim B(n, p)$ is approximated to $X \sim P(np)$

www.cwthompson.com



Sampling

Populations and a Census

A population is a collection of individual people or items, such as members of a club or plants in a store. If a population is limited and its members can be counted, it is known as a finite population. If it is impossible to count the members of a population it is an infinite population. However if in practice it could be counted it is a countably infinite population. An investigation, known as a census, can be used to obtain information from all members of a population.

Advantages of a Census	Disadvantages of a Census
• Every single member of a population is used	• It is very time consuming
• It is unbiased	• It is expensive
• It gives an accurate answer	• It can be difficult to ensure the whole population is surveyed

Samples and Sampling

A sample is a subset of a population, and is easier to survey than a whole population. The individual units of a population are sampling units and a list of all of them is a sampling frame. A sample survey is an investigation using some of the sampling units.

Advantages of Sampling	Disadvantages of Sampling
• If a population is large and well mixed the sample is representative	• Uncertainty due to natural variation in population
• Sampling is cheaper	• There could be bias
• Data is generally more readily available	
• Better than a census if testing of items results in their destruction	

Statistics

A Statistic is a quantity calculated solely from the observations in a sample. It does not involve any unknown parameters.

e.g. A random sample $X_1, X_2, \ldots, X_n$ is taken from a population with unknown mean μ. State whether the following are statistics.

$\frac{X_2 + X_5 + X_8}{3}$ Statistic

$\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$ Statistic

$\frac{\sum X}{n} - \mu^2$ Not Statistic (contains unknown parameter)

Sampling Distribution

The sampling distribution of a statistic gives all the values of a statistic and the probability that each would happen by chance alone.

A manufacturer of light bulbs sells 60W and 100W bulbs is the ratio 3:1. A random sample of 3 light bulbs is taken, find the sampling distribution of the mean.

(60, 60, 60) mean = 60 $p = \left(\frac{3}{4}\right)^3 = \frac{27}{64}$

(100, 60, 60) (60, 100, 60) (60, 60, 100) mean = $\frac{260}{3}$ $p = (3)\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^2 = \frac{27}{64}$

(100, 100, 60) (100, 60, 100) (60, 100, 100) mean = $\frac{280}{3}$ $p = (3)\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right) = \frac{9}{64}$

(100, 100, 100) mean = 100 $p = \left(\frac{1}{4}\right)^3 = \frac{1}{64}$

mean	60	$\frac{260}{3}$	$\frac{280}{3}$	100
p	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

S2 Definitions

Actual Significance Level	The probability of a value falling in the critical region
Census	An investigation of every member of the population
Critical Region	Range of values of a test statistic for which we can reject the null hypothesis
Critical Value	The value on the boundary of the critical region
Hypothesis Test	A statistical test for comparing the value of a population parameter from the null to alternative hypothesis
Null Hypothesis	The hypothesis we assume correct until proven otherwise
Population	All the items from which a sample can be taken
Sample	Chosen members of the population, a subset of the population
Sampling Distribution	The probability distribution of the statistic from all possible samples
Sampling Frame	A list of all the members of a population
Sampling Units	The individual items of a population
Statistic	A random variable calculated solely from the observations in a sample, containing no unknown parameters

www.cwthompson.com

Conditions for Binomial Distribution

- Fixed number of trials
- Trials are independent of each other
- Each trial can only be success or failure
- The probability of success is constant

Conditions for Poisson Distribution

- Events occur randomly
- Events occur singly
- Events occur independently of each other
- Events occur at a constant rate

Approximations