Ch 7 實習

Agenda

- 隨機變數與機率分配
  - Random variable and Probability Distribution
- 學習目標
  - The Mean and the Variance
  - 特定值的機率
- 一個隨機變數時
- 兩個隨機變數（以上）時
  - Bivariate Distributions (Joint distributions)
  - Portfolio concept
- 特殊分配
  - Overall review of Chapter 7 and 8
  - Binomial distribution
  - Poisson Distribution

一、 隨機變數與機率分配
Random Variables and Probability Distributions

- 骰子出現: 1,3,4,5,2,4,5,6,3,2, 1,1,2,3
- 若有100次結果，你要如何告訴別人?
- 三種機率呈現方式：圖、表格、pf or pdf

呈現方式：表格

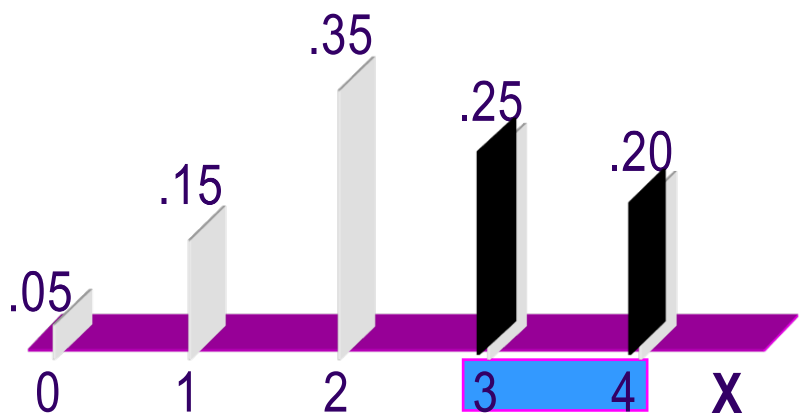
- NTU Company tracks the number of desktop computer systems it sells over a year (360 days), assume they take the two-day order policy, please use the following information to decide the best order number so that they can satisfy 70% customers' need. Note: For example, 26 days out of 360, 2 desktops were sold

兩天 demand	兩天# of days	相對機率	累加機率
0	25	0.069444	0.069444
1	35	0.097222	0.166667
2	26	0.072222	0.238889
3	50	0.138889	0.377778
4	63	0.175	0.552778
5	58	0.161111	0.713889
6	54	0.15	0.863889
7	39	0.108333	0.972222
8	10	0.027778	1
	360	1	

呈現方式：圖

- From the table of frequencies we can calculate the relative frequencies, which becomes our estimated probability distribution

Daily sales	Relative Frequency
0	5/100=.05
1	15/100=.15
2	35/100=.35
3	25/100=.25
4	20/100=.20
	1.00



P(X>2) = P(X=3) + P(X=4) =
.25 + .20 = .45

呈現方式: pf, pdf

$$P(X = x) = p(x) = \frac{e^{-\mu}\mu^{x}}{x!} \qquad x = 0,1,2...$$

$$E(X) = V(X) = \mu$$

pf :不連續分配 (C7)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty \le x \le \infty$$

$$where\ \pi = 3.14159... \ \ and \ \ e = 2.71828...$$

pdf :連續分配 (C8)

一、 隨機變數與機率分配
Random Variables and Probability Distributions

- A random variable is a function or rule that assigns a numerical value to each simple event in a sample space.
  - 骰子出現: 1,3,4,5,2,4,5,6,3,2, 1,1,2,3……
  - 為了降低分析的複雜性，將所有可能結果加以數值化
    - 例如投骰子十次，出現六點的次數，就是random variable
  - 短期不知道是什麼，長期下來會呈現某種分配

一、隨機變數與機率分配
Random Variables and Probability Distributions

- There are two types of random variables:
  - Discrete random variable (C7)
  - Continuous random variable (C 8)

1.1 Discrete Probability Distribution

- A table, formula, or graph that lists all possible values a discrete random variable can assume, together with associated probabilities, is called ***a discrete probability distribution***

- To calculate the probability that the random variable **X** assumes the value x, P(**X** = x),
  - add the probabilities of all the simple events for which **X** is equal to x, or
  - Use probability calculation tools (tree diagram),
  - Apply probability definitions

- EX:骰子出現: 1,3,4,5,2,4,5,6,3,2, 1,1,2,3

二、學習目標

- 不管是一個隨機變數、兩個隨機變數、或是特定分配
  - The population mean
  - The population variance or standard deviation
  - 特定值的機率 ex: P(X=3)

- 將題目公式化
- The number of students who seek assistance with their statistics assignments is Poisson distributed with a mean of three per day.
  - What is the probability that no student seek assistance tomorrow?
  - P(X=0)
  - What is the probability that two student seek assistance tomorrow?
  - P(X=2)
  - What is the probability two or more than two student seek assistance tomorrow?
  - P(X>=2)
  - What is the probability at least two student seek assistance tomorrow?
  - P(X>=2)
  - What is the probability less than two student seek assistance tomorrow?
  - P(X<2)

1.1 Population Mean (Expected Value) and Population Variance

- Given a discrete random variable **X** with values x_i, that occur with probabilities $p(x_i)$, the population mean of **X** is.
  - 加權平均概念(權數是機率)

$$E(X) = \mu = \sum_{\text{all } x_i} x_i \cdot p(x_i)$$

$$\mu = \frac{\sum X}{n}$$

$$\sigma = \sqrt{\frac{\Sigma(X-u)}{n}}$$

- Let **X** be a discrete random variable with possible values x_i that occur with probabilities $p(x_i)$, and let $E(x_i) = \mu$. The variance of **X** is defined by

$$V(X) = \sigma^2 = E\left[(X-\mu)^2\right] = \sum_{\text{all } x_i} (x_i - \mu)^2 p(x_i)$$

The standard deviation is

$$\sigma = \sqrt{\sigma^2}$$

1.1 The Mean and the Variance

- The variance can also be calculated as follows:

$$V(X) = \sigma^2 = E(X^2) - \mu^2 = \sum_{all\ x_i} x_i^2 p(x_i) - \mu^2$$

- Proof

$$E[(X-\mu)^2]$$

$$= E[X^2 - 2X\mu + \mu^2]$$

$$= E(X^2) - 2\mu E(X) + E(\mu^2)$$

$$= E(X^2) - \mu^2$$

1.1 Laws of Expected Value and Variance

- **Laws of Expected Value**
  - $E(c) = c$
  - $E(X + c) = E(X) + c$
  - $E(c\mathbf{X}) = cE(\mathbf{X})$

全班+5分
~ 平均 +5

- **Laws of Variance**
  - $V(c) = 0$
  - $V(X + c) = V(X)$
  - $V(cX) = c^2V(X)$

為什麼加一個常數，常數的變異數為0?

全班+5分 ，彼此差距有變小（大）嘛？

Example 2

- We are given the following probability distribution:

x	0	1	2	3
P(x)	0.4	0.3	0.2	0.1

- a. Calculate the mean, variance, and standard deviation
- b. Suppose that Y=3X+2. For each value of X, determine the value of Y. What is the probability distribution of Y?
- c. Calculate the mean, variance, and standard deviation from the probability distribution of Y.
- d. Use the laws of expected value and variance to calculate the mean, variance, and standard deviation of Y from the mean, variance, and standard deviation of X. Compare your answers in Parts c and d. Are they the same (except for rounding)

Example 2

- We are given the following probability distribution:

x	0	1	2	3
P(x)	0.4	0.3	0.2	0.1

- a. Calculate the mean, variance, and standard deviation
- a.
  - $E(X) = 0(0.4) + 1(0.3) + 2(0.2) + 3(0.1) = 1$
  - $V(X)=E(X^2)-u^2$

    $=(0)^2(0.4)+(1)^2(0.3)+(2)^2(0.2)+(3)^2(0.1)-(1)^2 =1$
  -

Example 2

- We are given the following probability distribution:

x	0	1	2	3
P(x)	0.4	0.3	0.2	0.1

- b. Suppose that Y=3X+2. For each value of X, determine the value of Y. What is the probability distribution of Y?
- c. Calculate the mean, variance, and standard deviation from the probability distribution of Y.
-

x	0	1	2	3
y	2	5	8	11
P(x)	0.4	0.3	0.2	0.1

2=3*0+2

E(Y) = 2(0.4) + 5(0.3) + 8(0.2) + 11(0.1) = 5

$V(Y)=E(Y^2)-\{E(Y)\}^2$

$=(2)^2(0.4)+(5)^2(0.3)+(8)^2(0.2)+(11)^2(0.1)-(5)^2 =9$

Example 2

- We are given the following probability distribution:

x	0	1	2	3
P(x)	0.4	0.3	0.2	0.1

- d. Use the laws of expected value and variance to calculate the mean, variance, and standard deviation of Y from the mean, variance, and standard deviation of X. Compare your answers in Parts c and d. Are they the same (except for rounding)

$Y=3X+2$, $E(X)=1$, $V(X)=1$

d. $E(Y) = E(3X+2)$

$= 3E(X)+2 = 5$

$V(Y) = V(3X+2)$

$= 9V(X) = 9$

三、兩個變數（以上）時：Bivariate Distributions

- The **bivariate (or joint) distribution** is used when the relationship between two random variables is studied.
  - 也就是第六章所看到的聯合機率分配
- The probability that X assumes the value x, and Y assumes the value y is denoted

$$p(x,y) = P(X=x \text{ and } Y = y)$$

The joint probability function satisfies the following conditions:

1. $0 \le p(x,y) \le 1$
2. $\sum_{\text{all } x}\sum_{\text{all } y} p(x,y) = 1$

三、兩個變數（以上）時：Bivariate Distributions

- 有一個分配如下方，請求他的平均數與標準差

Y	X		
	0	1	2
0	.12	.42	.06
1	.21	.06	.03
2	.07	.02	.01

Step1: 先求marginal dis.
Step2: 畫成機率的表格
Step3:代入公式

三、兩個變數（以上）時：Bivariate Distributions

- The joint distribution can be described by the mean, variance, and standard deviation of each variable.
- This is done using the marginal distributions.

Joint pro.

p(0,0)
p(0,1)
p(0,2)

Y	X			p(y)
	0	1	2	
0	.12	.42	.06	.60
1	.21	.06	.03	.30
2	.07	.02	.01	.10
p(x)	.40	.50	.10	1.00

Step1: 先求marginal dis.

P(Y=1), the marginal probability.

The marginal probability P(X=0)

1.3 Describing the Bivariate Distribution

- The joint distribution can be described by the mean, variance, and standard deviation of each variable.
- This is done using the marginal distributions.

Y	X 0	X 1	X 2	p(y)
0	.12	.42	.06	.60
1	.21	.06	.03	.30
2	.07	.02	.01	.10
p(x)	.40	.50	.10	1.00

Step2:畫成機率的表格

x	p(x)	y	p(y)
0	.4	0	.6
1	.5	1	.3
2	.1	2	.1

E(X) = E(Y) =

V(X) = V(Y) =

1.3 Describing the Bivariate Distribution

- The joint distribution can be described by the mean, variance, and standard deviation of each variable.
- This is done using the marginal distributions.

Y	X 0	X 1	X 2	p(y)
0	.12	.42	.06	.60
1	.21	.06	.03	.30
2	.07	.02	.01	.10
p(x)	.40	.50	.10	1.00

Step3:代入公式

x	p(x)	y	p(y)
0	.4	0	.6
1	.5	1	.3
2	.1	2	.1

$E(X) = .7$ $E(Y) = .5$

$V(X) = .41$ $V(Y) = .45$

$E(X)=0*0.4+1*0.5+2*0.1=0.7$

$V(X)= E(X^2)-u^2$

$=(0)^2(0.4)+(1)^2(0.5)+(2)^2(0.1)-(0.7)^2=0.41$

三.特殊分配

Overall review of Chapter 7 and 8

Discrete Probability Distributions	Continuous Probability Distributions
• Binomial distribution • Poisson Distribution	• Uniform Distribution • Normal Distribution • Standard Normal Distribution • Exponential Distribution • t Distribution • F Distribution

三、特殊分配: Binomial distribution

- The binomial experiment can result in only one of two possible outcomes. (成功 or失敗）
- Binomial Experiment
  - There are n trials (n is finite and fixed).
  - Each trial can result in a success or a failure.
  - The probability p of success is the same for all the trials.
  - All the trials of the experiment are independent
- Binomial Random Variable
  - The binomial random variable ***counts*** the number of successes in n trials of the binomial experiment.
  - By definition, this is a discrete random variable (因為數的完).

三、特殊分配: Binomial distribution

In general, The binomial probability is calculated by:

$$P(X = x) = p(x) = C_x^n p^x (1 - p)^{n-x}$$

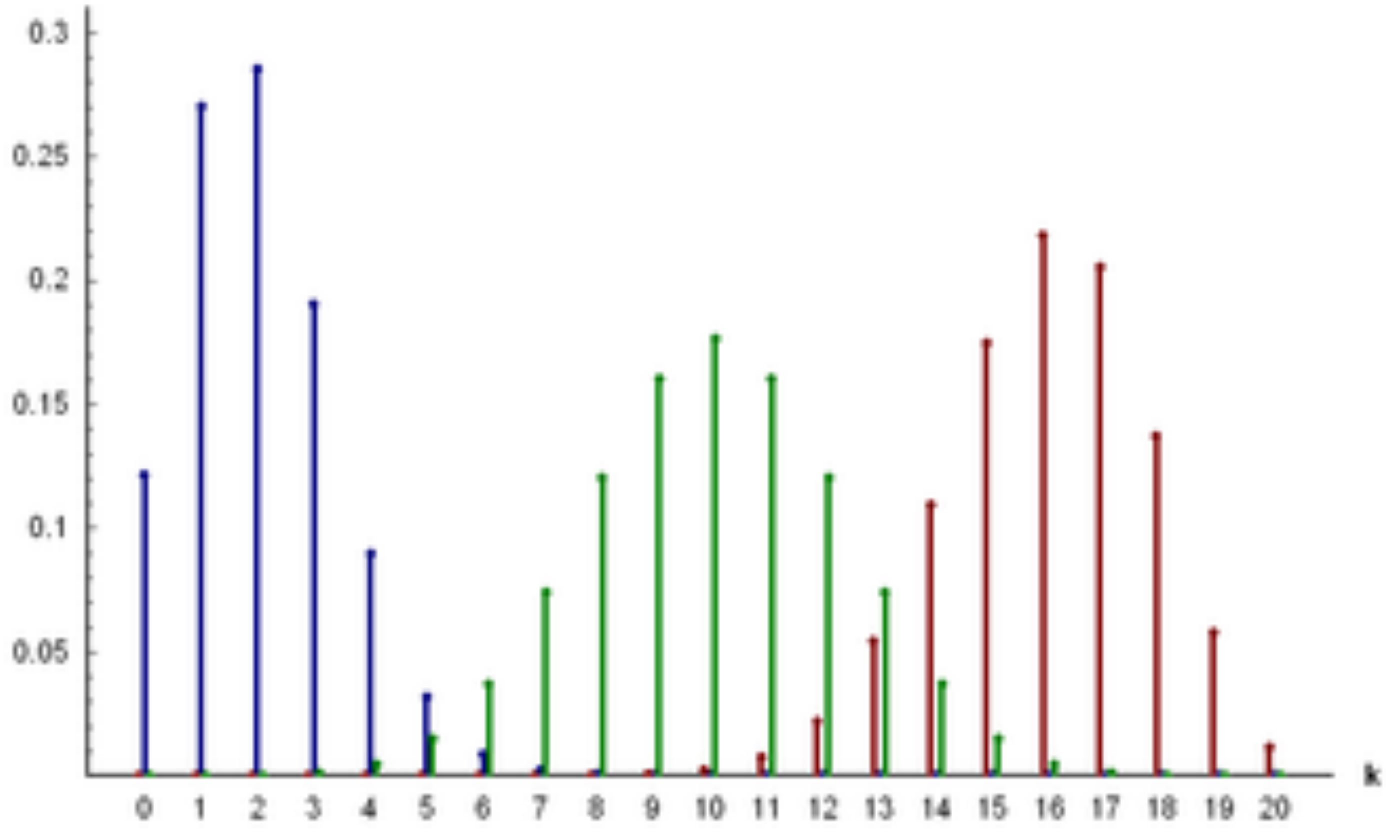
$$\text{where } C_x^n = \frac{n!}{x!(n - x)!}$$

Mean and Variance of Binomial Variable

$$E(X) = \mu = np$$

$$V(X) = s^2 = np(1 - p)$$

3. Binomial distribution



- Binomial distribution for $n = 20$
  $p = 0.1$ (blue), $p = 0.5$ (green) and $p = 0.8$ (red)

Example 4 （類似4,5,6)

- In the game of blackjack as played in casinos in Las Vegas, Atlantic City, Niagara Falls, as well as many other cities, the dealer has the advantages. Most players do not play very well. As a result, the probability that the average player wins a about 45%.Find the probability that an average player wins

a. twice in 5 hands

b. ten or more times in 25 hands

Step 1:找到 p (成功的機率）, n (樣本數）

Step 2:依照題目找到x？P(X=2)=?

Step3:代入公式（or要學會用excel求解）

Example 4 （類似4,5,6)

- In the game of blackjack as played in casinos in Las Vegas, Atlantic City, Niagara Falls, as well as many other cities, the dealer has the advantages. Most players do not play very well. As a result, the probability that the average player wins a about 45%.Find the probability that an average player wins

a. twice in 5 hands

b. ten or more times in 25 hands

Step 1:找到 p (成功的機率）,n (樣本數）

→ p=0.45, q=1-0.45=0.55

$C_X^n * P^X * (1-P)^{n-x}$

Step 2:依照題目找到x？

(a) P(X=2)

*這裡n是5

$(a)\ P(X=2) = C_2^5 * 0.45^2 * (1-0.45)^3$ =0.3369

(b) P(X>=10)

*這裡n是25

$$(b)\ P(X \geq 10) = 1 - P(X \leq 9)$$

$$= 1 - (\sum_{x=i}^{9} C_i^{25} * 0.45^i * (1-0.45)^{n-i})$$

$$= 0.7576$$

Step3:代入公式（or要學會用excel求解）

Example 4 （類似4,5,6)

- In the game of blackjack as played in casinos in Las Vegas, Atlantic City, Niagara Falls, as well as many other cities, the dealer has the advantages. Most players do not play very well. As a result, the probability that the average player wins a about 45%.Find the probability that an average player wins

a. twice in 5 hands

b. ten or more times in 25 hands

Step 1:找到 p (成功的機率）,n (樣本數）
→ p=0.45, q=1-0.45=0.55

Step 2:依照題目找到x？
(a) P(X=2)
*這裡n是5

(b) P(X>=10)
*這裡n是25

BINOM.DIST(X, N, p, 1)→累積機率
- P(X<=6)

BINOM.DIST(X, N, p, 0)→單點值機率
- P(X=6)

(a)=BINOM.DIST(2,5,0.45,0)
=0.3369

(b)=1-(BINOM.DIST(9,25,0.45,1))
=0.7576

Step3:代入公式（or要學會用excel求解）

HW#6. When a customer places an order with Rudy's On-Line Office Supplies, a computerized accounting information system (AIS) automatically checks to see if the customer has exceeded his or her credit limit. Past records indicate that the probability of customers exceeding their credit limit is 0.05. Suppose that, on a given day, 20 customers place orders. Assume that the number of customers that the AIS detects as having exceeded their credit limit is distributed as a binomial random variable. (會不會有人信用破產）～Binomial dist.

1. What are the mean and standard deviation of the number of customers exceeding their credit limits?
2. What is the probability that zero customers will exceed their limits?
3. What is the probability that one customer will exceed his or her limit?
4. What is the probability that two or more customers will exceed their limits?

Step 1:找到 p (成功的機率）, n (樣本數）
--> p=0.05, n=20

Step 2:依照題目找到x？
(a) u=np, var=npq,
(b)P(X=0)
(c) P(X=1)
(d) P(X>=2)

6. When a customer places an order with Rudy's On-Line Office Supplies, a computerized accounting information system (AIS) automatically checks to see if the customer has exceeded his or her credit limit. Past records indicate that the probability of customers exceeding their credit limit is 0.05. Suppose that, on a given day, 20 customers place orders. Assume that the number of customers that the AIS detects as having exceeded their credit limit is distributed as a binomial random variable.

1. What are the mean and standard deviation of the number of customers exceeding their credit limits?
2. What is the probability that zero customers will exceed their limits?
3. What is the probability that one customer will exceed his or her limit?
4. What is the probability that two or more customers will exceed their limits?

Step 1:找到 p (成功的機率）, n (樣本數）
--> p=0.05, n=20

BINOM.DIST(X, N, p, 1)→累積機率
BINOM.DIST(X, N, p, 0)→單點值機率

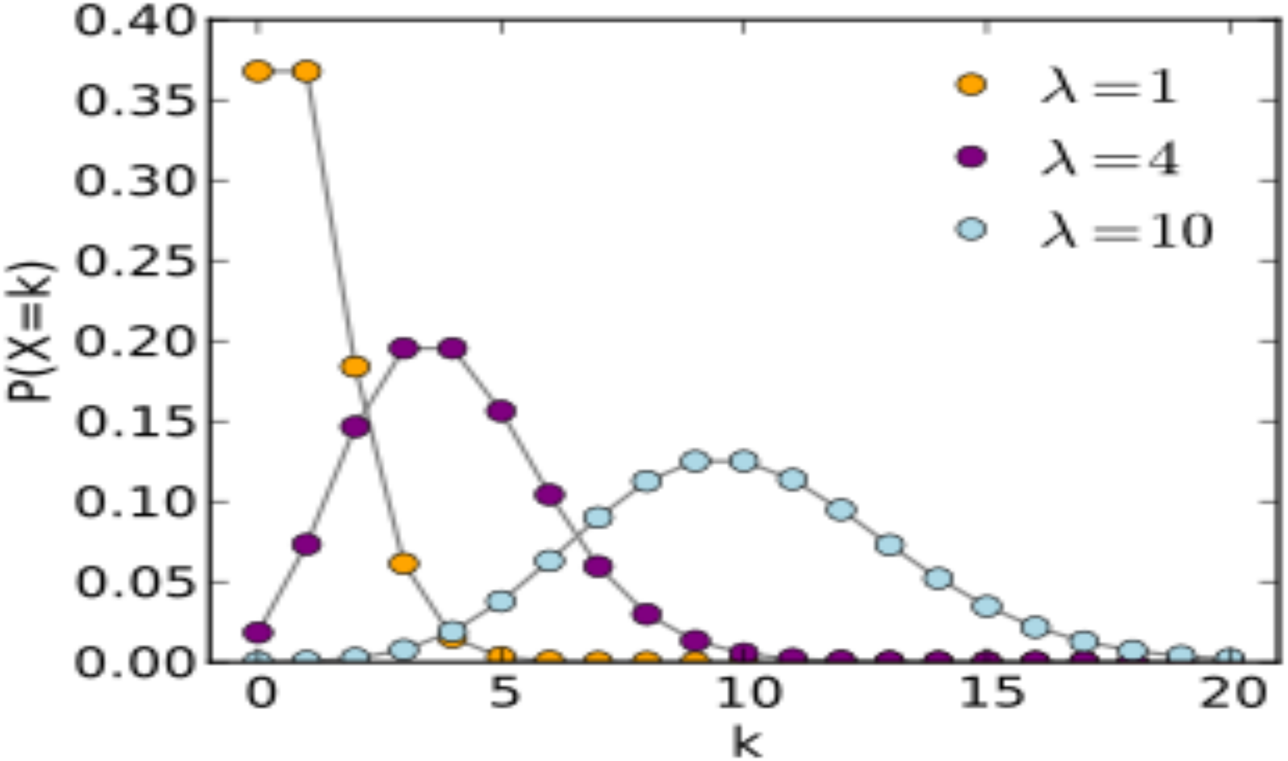
Step 2:依照題目找到x？P(X=6)
(a) u=np,
(b)P(X=0)
(c) P(X=1)
(d) P(X>=2)

(a) u=np=0.05*20=1
sd=sqrt(npq)= sqrt (20*0.05*0.95)=0.9747
(b)=BINOM.DIST(0,20,0.05,0)=0.3858

(c)=BINOM.DIST(1,20,0.05,0)=0.3774

(d)=1-(BINOM.DIST(1,20,0.05,1))=0.2642

三、特殊分配: Poisson Distribution



- A famous quote of Poisson is: "Life is good for only two things: Discovering mathematics and teaching mathematics."

三、特殊分配: Poisson Distribution

- **The Poisson Random Variable (單位時間的來客數)**
  - The Poisson variable indicates the number of successes that occur during a given time interval or in a specific region in a Poisson experiment

- **Probability Distribution of the Poisson Random Variable.**

$$P(X = x) = p(x) = \frac{e^{-\mu}\mu^{x}}{x!} \qquad x = 0,1,2\ldots$$

$$E(X) = V(X) = \mu$$

Step 1:找到 u (單位時間的來客數）
Step 2:依照題目找到x？（P(X=6)=?
Step3:代入公式（or要學會用excel求解）

Example 5 (類似7.8)

- The number of students who seek assistance with their statistics assignments is Poisson distributed with a mean of three per day.
  a. What is the probability that no student seek assistance tomorrow?
  b. Find the probability that 10 students seek assistance in a week.

Example 5 (類似7.8)

- The number of students who seek assistance with their statistics assignments is Poisson distributed with a mean of three per day.
  a. What is the probability that no student seek assistance tomorrow?
  b. Find the probability that 10 students seek assistance <u>in a week</u>.

Step 1:找到 u (單位時間的來客數）
→一天有三個學生（u=3)
Step 2:依照題目找到x？
(a) P(X=0)=
(b) P(X=10)=
Step3:代入公式

- a. P(X = 0 with μ = 3)= $\dfrac{e^{-\mu}\mu^{x}}{x!} = \dfrac{e^{-3}(3)^{0}}{0!}$ = 0.0498
- b. P(X = 10 with μ = 21)= $\dfrac{e^{-\mu}\mu^{x}}{x!} = \dfrac{e^{-21}(21)^{10}}{10!}$ = 0.0035

HW# 8. Snowfalls occur randomly and independently over the course of winter in a Minnesota city. The average is one snowfall every 3 days.

1. What is the probability of five snowfalls in 2 weeks?
2. Find the probability of a snowfall today.

Step 1:找到 u (單位時間的來客數）

→一天有1/3下雪（u=1/3)

Step 2:依照題目找到x？

(a) P(X=5)=

*此時的u=14/3

$$\text{a } P(X = 5 \text{ with } \mu = 14/3) = \frac{e^{-\mu}\mu^{x}}{x!} = \frac{e^{-14/3}(14/3)^{5}}{5!}$$

（b) P(X=1)=

*此時的u=1/3

$$\text{b. } P(X = 1 \text{ with } \mu = 1/3) = \frac{e^{-\mu}\mu^{x}}{x!} = \frac{e^{-1/3}(1/3)^{1}}{1!}$$

Step3:代入公式

HW# 8. Snowfalls occur randomly and independently over the course of winter in a Minnesota city. The average is one snowfall every 3 days.

1.What is the probability of five snowfalls in 2 weeks?

2.Find the probability of a snowfall today.

Step 1:找到 u (單位時間的來客數）

→一天有1/3下雪（u=1/3)

Step 2:依照題目找到x？

(a) P(X=5)=

*此時的u=14/3

（b) P(X=10)=

*此時的u=1/3

Step3:代入公式

POISSON.DIST(X, u,1)→累積機率

- P(X<=6)

POISSON.DIST(X, u,0) →單點值機率

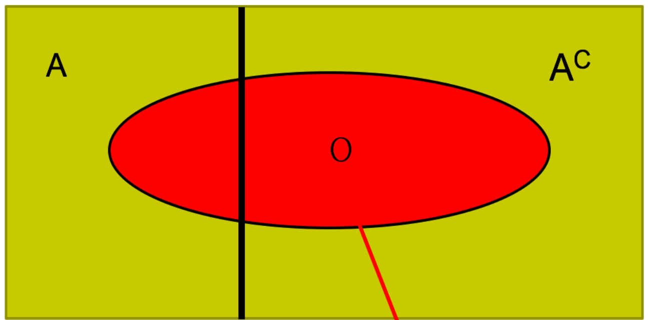
- P(X=6)

(a) POISSON.DIST(5, 14/3,0)=0.1734

(b) POISSON.DIST(1, 1/3,0)=0.2388

4. 研究者若善用統計技巧，將可降低受訪者心防，使調查結果更加可靠。例如某位老師想了解班上 60 名同學中，約有多少比率在國中前談過戀愛，於是他設計了 A、B 兩份問卷(其中 A 卷 40 張，B 卷 20 張)及不記名答案卷 60 張，隨機發給班上同學(同學不是拿到 A 卷就是 B 卷，至於 A、B 卷接不回收，答案卷則要回收)。其中 A 卷的題目是我曾經在國中以前談過戀愛，因此拿到 A 卷的同學若國中前曾談戀愛，就要在答案卷上劃 O，若不曾談過戀愛，則劃 X；反之，B 卷的題目是我不曾在國中以前談過戀愛，因此拿到 B 卷的同學若國中前曾談戀愛，就要在答案卷上劃 X，若不曾談過戀愛，則劃 O。如此設計，不論某位同學回答的結果是 O 或 X，老師都無法得知該生在國中前是否談過戀愛，但卻可大致了解全班的狀況。答案卷經回收後，老師發現 60 張答案卷中有 25 張劃 O，35 張劃 X。請你求出在本班同學中，約有多少比例的同學曾在國中前談過戀愛?

A：拿A卷　　　A^c：拿B卷



B: 畫圈圈的人 (after)

a. P(B)=

P(B)= $P(A) * P(B|A) + P(A^c) * P(B|A^c)$

$$\frac{25}{60} = \frac{40}{60} * P(B|A) + \frac{20}{60} * P(B|A^c)$$

$$\frac{25}{60} = \frac{40}{60} * P(X) + \frac{20}{60} **(1-P(X))$$

＊假設 拿A卷與拿 B卷，要獨立事件（只能拿一種考卷）