常用的離散型機率模式

桑慧敏

清華大學工業工程系

2015.10.07

選擇適當的機率分配與參數

- (1) 1000 晶片不良品個數
- (2) 每一頁書中錯字個數
- (3) 圓週率小數點後的數值
- (4) 每一年新戰爭發生的次數
- (5) 每一年戰爭結束的次數
- (6) 每一天北一高上新車禍的次數
- (7) 電腦資料傳輸系統每 100,000 bits 中錯誤的 bit 次數
- (8) 樂透每期開獎中獎數字的個數
- (9) 在品管圖上, 要做幾次實驗才見到第一個超出上下規格外的點
- (10) 晶片檢驗線上, 檢驗出第二個不良品所需總檢驗數

選擇適當的機率分配與參數

- (1) 1000 晶片不良品個數
- (2) 每一頁書中錯字個數
- (3) 圓週率小數點後的數值
- (4) 每一年新戰爭發生的次數
- (5) 每一年戰爭結束的次數
- (6) 每一天北一高上新車禍的次數
- (7) 電腦資料傳輸系統每 100,000 bits 中錯誤的 bit 次數
- (8) 樂透每期開獎中獎數字的個數
- (9) 在品管圖上, 要做幾次實驗才見到第一個超出上下規格外的點
- (10) 晶片檢驗線上, 檢驗出第二個不良品所需總檢驗數
- 你答對幾題?

選擇適當的機率分配與參數

- (1) 1000 晶片不良品個數
- (2) 每一頁書中錯字個數
- (3) 圓週率小數點後的數值
- (4) 每一年新戰爭發生的次數
- (5) 每一年戰爭結束的次數
- (6) 每一天北一高上新車禍的次數
- (7) 電腦資料傳輸系統每 100,000 bits 中錯誤的 bit 次數
- (8) 樂透每期開獎中獎數字的個數
- (9) 在品管圖上, 要做幾次實驗才見到第一個超出上下規格外的點
- (10) 晶片檢驗線上, 檢驗出第二個不良品所需總檢驗數

- 你答對幾題?

均勻分配 (Uniform Distribution)

- 若一離散型隨機變數 X 的 pmf 為

$$f_X(x) = \begin{cases} 1/n, & x = x_1, \cdots, x_n \\ 0, & \text{其他}, \end{cases}$$

  則稱 X 服從均勻分配, 記做 $X \sim \text{uniform}(x_1, \cdots, x_n)$。

- X 的平均數: $\mathrm{E}(X) = \sum_{i=1}^{n} x_i/n$
- X 的變異數: $\mathrm{V}(X) = \sum_{i=1}^{n} x_i^2/n - [\sum_{i=1}^{n} x_i/n]^2$

- 何謂 “均勻” ?

均匀分配 (Uniform Distribution)

- 若一離散型隨機變數 X 的 pmf 為

$$f_X(x) = \begin{cases} 1/n, & x = x_1, \cdots, x_n \\ 0, & \text{其他,} \end{cases}$$

則稱 X 服從均勻分配, 記做 $X \sim \text{uniform}(x_1, \cdots, x_n)$。

- X 的平均數: $\text{E}(X) = \sum_{i=1}^{n} x_i/n$
- X 的變異數: $\text{V}(X) = \sum_{i=1}^{n} x_i^2/n - [\sum_{i=1}^{n} x_i/n]^2$

- 何謂 "均勻" ?
- 繪圖說明均勻隨機變數的分配 (pmf)

均勻分配 (Uniform Distribution)

- 若一離散型隨機變數 X 的 pmf 為

$$f_X(x) = \begin{cases} 1/n, & x = x_1, \cdots, x_n \\ 0, & \text{其他}, \end{cases}$$

則稱 X 服從均勻分配, 記做 $X \sim \text{uniform}(x_1, \cdots, x_n)$。

- X 的平均數: $\text{E}(X) = \sum_{i=1}^{n} x_i/n$
- X 的變異數: $\text{V}(X) = \sum_{i=1}^{n} x_i^2/n - [\sum_{i=1}^{n} x_i/n]^2$

- 何謂 "均勻" ?
- 繪圖說明均勻隨機變數的分配 (pmf)

圓周率小數點後面的數

- 西元五世紀, 祖沖之 (中國): 6 位數 (i.e., 3.141592)

圓周率小數點後面的數

- 西元五世紀, 祖沖之 (中國): 6 位數 (i.e., 3.141592)
- 1593 年, Roomen(德國): 35 位數

圓周率小數點後面的數

- 西元五世紀, 祖沖之 (中國): 6 位數 (i.e., 3.141592)
- 1593 年, Roomen(德國): 35 位數
- 1873 年, W. Hsanks (英國): 707 位數*

圓周率小數點後面的數

- 西元五世紀, 祖沖之 (中國): 6 位數 (i.e., 3.141592)
- 1593 年, Roomen(德國): 35 位數
- 1873 年, W. Hsanks (英國): 707 位數*
- 1966 年, IBM 7030: 250,000 位數

圓周率小數點後面的數

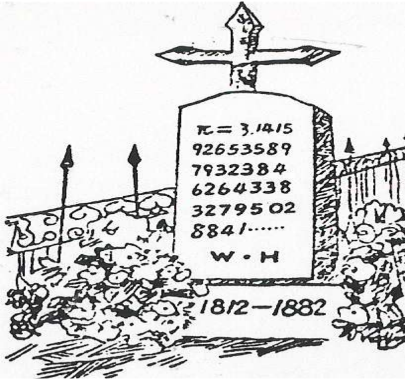
- 西元五世紀, 祖沖之 (中國): 6 位數 (i.e., 3.141592)
- 1593 年, Roomen(德國): 35 位數
- 1873 年, W. Hsanks (英國): 707 位數*
- 1966 年, IBM 7030: 250,000 位數
- 1967 年, CDC6600: 500,000 位數

圓周率小數點後面的數

- 西元五世紀, 祖沖之 (中國): 6 位數 (i.e., 3.141592)
- 1593 年, Roomen(德國): 35 位數
- 1873 年, W. Hsanks (英國): 707 位數*
- 1966 年, IBM 7030: 250,000 位數
- 1967 年, CDC6600: 500,000 位數
- 1999 年, 超過 510 億位數

圓周率小數點後面的數

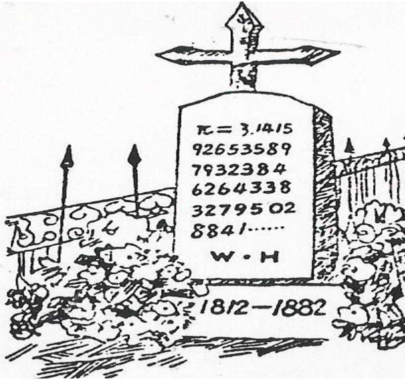
- 西元五世紀, 祖沖之 (中國): 6 位數 (i.e., 3.141592)
- 1593 年, Roomen(德國): 35 位數
- 1873 年, W. Hsanks (英國): 707 位數*
- 1966 年, IBM 7030: 250,000 位數
- 1967 年, CDC6600: 500,000 位數
- 1999 年, 超過 510 億位數



- Q: 算這麼多位數要做什麼?
- A: 數學家說: 人與其他動物最大的不同是: 永遠不止的好奇心
- Q: 如何判斷 W. Hsanks 的 707 位數是否正確?

圓周率小數點後面的數

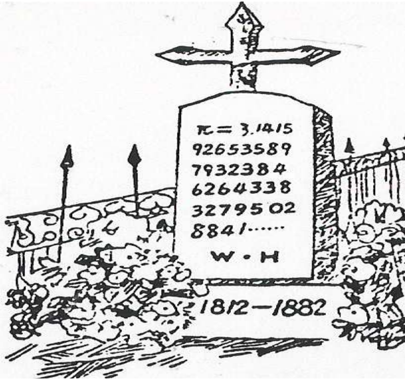
- 西元五世紀, 祖沖之 (中國): 6 位數 (i.e., 3.141592)
- 1593 年, Roomen(德國): 35 位數
- 1873 年, W. Hsanks (英國): 707 位數*
- 1966 年, IBM 7030: 250,000 位數
- 1967 年, CDC6600: 500,000 位數
- 1999 年, 超過 510 億位數



- Q: 算這麼多位數要做什麼?
- A: 數學家說: 人與其他動物最大的不同是: 永遠不止的好奇心
- Q: 如何判斷 W. Hsanks 的 707 位數是否正確?
- A: 均勻分配的妙用

圓周率小數點後面的數

- 西元五世紀, 祖沖之 (中國): 6 位數 (i.e., 3.141592)
- 1593 年, Roomen(德國): 35 位數
- 1873 年, W. Hsanks (英國): 707 位數*
- 1966 年, IBM 7030: 250,000 位數
- 1967 年, CDC6600: 500,000 位數
- 1999 年, 超過 510 億位數

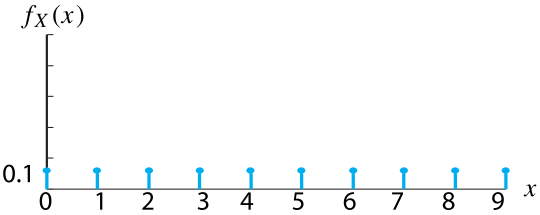


- Q: 算這麼多位數要做什麼?
- A: 數學家說: 人與其他動物最大的不同是: 永遠不止的好奇心
- Q: 如何判斷 W. Hsanks 的 707 位數是否正確?
- A: 均勻分配的妙用

圓周率小數點後面的數 (續)

數學家們以宗教般的狂熱計算圓周率小數點後面的數, 直到1999年, 這些數已超過510億位。你有沒有想過這些數是 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 的機率是多少? 換言之, 如何以機率模式描述圓周率小數點後面的數?

- 令 X 表示圓周率小數點後面的數, 則 X 服從均勻分配 (見下圖)。

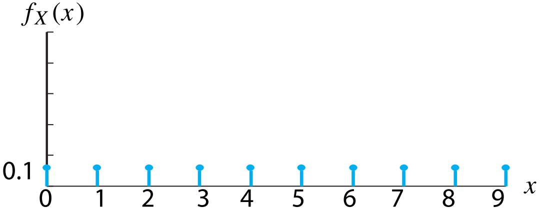


- 到目前為止, 以均勻分配來描述圓周率小數點後面的數是適當的。

圓周率小數點後面的數 (續)

數學家們以宗教般的狂熱計算圓周率小數點後面的數, 直到1999年, 這些數已超過510億位。你有沒有想過這些數是 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 的機率是多少? 換言之, 如何以機率模式描述圓周率小數點後面的數?

- 令 X 表示圓周率小數點後面的數, 則 X 服從均勻分配 (見下圖)。

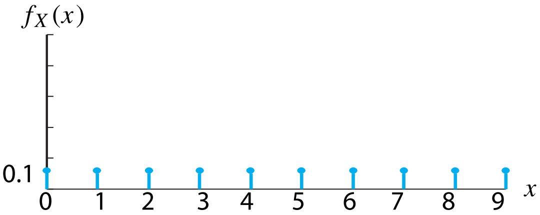


- 到目前為止, 以均勻分配來描述圓周率小數點後面的數是適當的。
- 上帝並不特別偏愛哪一個數字, 讓此世界是「亂中有序」一說再添一例。

圓周率小數點後面的數 (續)

數學家們以宗教般的狂熱計算圓周率小數點後面的數, 直到1999年, 這些數已超過510億位。你有沒有想過這些數是 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 的機率是多少? 換言之, 如何以機率模式描述圓周率小數點後面的數?

- 令 X 表示圓周率小數點後面的數, 則 X 服從均勻分配 (見下圖)。



- 到目前為止, 以均勻分配來描述圓周率小數點後面的數是適當的。
- 上帝並不特別偏愛哪一個數字, 讓此世界是「亂中有序」一說再添一例。

伯努力分配 (Bernoulli Distribution)

- 若一離散型隨機變數 X 的 pmf 為

$$f_X(x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \\ 0, & \text{其他}, \end{cases}$$

 則稱 X 服從伯努力分配, 記做 $X \sim \text{Bernoulli}(p)$

- X 的平均數: $\text{E}(X) = p$
- X 的變異數: $\text{V}(X) = p(1 - p)$

- "Bernoulli" 是...

伯努力分配 (Bernoulli Distribution)

- 若一離散型隨機變數 X 的 pmf 為

$$f_X(x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \\ 0, & \text{其他}, \end{cases}$$

則稱 X 服從伯努力分配, 記做 $X \sim \text{Bernoulli}(p)$

- X 的平均數: $\text{E}(X) = p$
- X 的變異數: $\text{V}(X) = p(1 - p)$

- "Bernoulli" 是...
- 生活中有哪些實例可以用"伯努力分配" 來描述?

伯努力分配 (Bernoulli Distribution)

- 若一離散型隨機變數 X 的 pmf 為

$$f_X(x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \\ 0, & \text{其他}, \end{cases}$$

則稱 X 服從伯努力分配, 記做 $X \sim \text{Bernoulli}(p)$

- X 的平均數: $\text{E}(X) = p$
- X 的變異數: $\text{V}(X) = p(1 - p)$

- "Bernoulli" 是...
- 生活中有哪些實例可以用"伯努力分配"來描述?

新生男女嬰之比

你知道新生男女嬰之比是多少?

新生男女嬰之比

你知道新生男女嬰之比是多少?

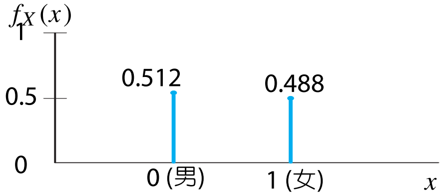
- 1:1
- 25:24
- 22:21

新生男女嬰之比

你知道新生男女嬰之比是多少?

- 1:1
- 25:24
- 22:21

數學家拉普拉斯統計 1745 1784 整 40 年間巴黎新生兒男女嬰的之比為 $25 : 24$。經過修正後, 新生男女嬰之比 $22 : 21 (0.512 : 0.488)$ 這個統計結果第一次出現在法國著名的數學家拉普拉斯之著作「機率的哲學探討」。



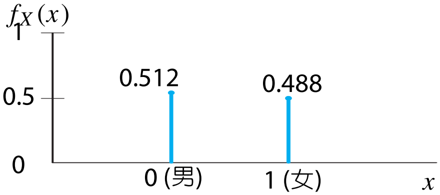
- 以伯努力分配($p = 0.488$) 來描述新生男女嬰是適當的, 讓此世界是「亂中有序」一說再添一例。

新生男女嬰之比

你知道新生男女嬰之比是多少?

- 1:1
- 25:24
- 22:21

數學家拉普拉斯統計 1745 1784 整 40 年間巴黎新生兒男女嬰的之比為 $25:24$。經過修正後, 新生男女嬰之比 $22:21(0.512:0.488)$ 這個統計結果第一次出現在法國著名的數學家拉普拉斯之著作「機率的哲學探討」。



- 以伯努力分配($p = 0.488$) 來描述新生男女嬰是適當的, 讓此世界是「亂中有序」一說再添一例。

伯努力實驗與伯努力程序

- 伯努力實驗 (Bernoulli trial): 若某實驗的結果可分成兩類, 稱其中一類為「成功」, 另一類為「失敗」, 則稱此實驗為伯努力實驗。
- 伯努力程序 (Bernoulli process): 是一連串相同且獨立的伯努力實驗所組成, 具有以下四個特性:
  - 相同的實驗可以重複進行。
  - 每一次實驗的結果可分成兩類, 我們稱其中一類為「成功」, 稱另一類為「失敗」。
  - 每一次實驗成功的機率相同。
  - 每一次實驗都是獨立的, 也就是每一次實驗的結果無法提供另一次實驗結果的任何資訊。

- 舉例說明伯努力程序

伯努力實驗與伯努力程序

- 伯努力實驗 (Bernoulli trial): 若某實驗的結果可分成兩類, 稱其中一類為「成功」, 另一類為「失敗」, 則稱此實驗為伯努力實驗。
- 伯努力程序 (Bernoulli process): 是一連串相同且獨立的伯努力實驗所組成, 具有以下四個特性:
  - 相同的實驗可以重複進行。
  - 每一次實驗的結果可分成兩類, 我們稱其中一類為「成功」, 稱另一類為「失敗」。
  - 每一次實驗成功的機率相同。
  - 每一次實驗都是獨立的, 也就是每一次實驗的結果無法提供另一次實驗結果的任何資訊。

- 舉例說明伯努力程序

兩種情境說明吻合與不吻合伯努力程序

- 投擲一枚相同的錢幣 n 次, 令 X 為人頭面出現的次數, p 為人頭出現機率。

兩種情境說明吻合與不吻合伯努力程序

- 投擲一枚相同的錢幣 n 次, 令 X 為人頭面出現的次數, p 為人頭出現機率。
  - 如果正面是人頭像, 反面是數字, 則 $X \sim \text{binomial}(n, p)$

兩種情境說明吻合與不吻合伯努力程序

- 投擲一枚相同的錢幣 n 次, 令 X 為人頭面出現的次數, p 為人頭出現機率。
  - 如果正面是人頭像, 反面是數字, 則 $X \sim \text{binomial}(n, p)$
  - 假如錢幣正反兩面皆是「人頭」, 則伯努力程序中第 2, 第 3, 與第 4 個假設就不吻合。

兩種情境說明吻合與不吻合伯努力程序

- 投擲一枚相同的錢幣 n 次, 令 X 為人頭面出現的次數, p 為人頭出現機率。
  - 如果正面是人頭像, 反面是數字, 則 $X \sim \text{binomial}(n, p)$
  - 假如錢幣正反兩面皆是「人頭」, 則伯努力程序中第 2, 第 3, 與第 4 個假設就不吻合。
- 檢驗某一生產線上 n 個產品, 令 X 為不良品出現的次數。

兩種情境說明吻合與不吻合伯努力程序

- 投擲一枚相同的錢幣 n 次, 令 X 為人頭面出現的次數, p 為人頭出現機率。
  - 如果正面是人頭像, 反面是數字, 則 $X \sim \text{binomial}(n, p)$
  - 假如錢幣正反兩面皆是「人頭」, 則伯努力程序中第 2, 第 3, 與第 4 個假設就不吻合。
- 檢驗某一生產線上 n 個產品, 令 X 為不良品出現的次數。
  - n 個產品皆在製程正常中產出, 則 $X \sim \text{binomial}(n, p)$

兩種情境說明吻合與不吻合伯努力程序

- 投擲一枚相同的錢幣 n 次, 令 X 為人頭面出現的次數, p 為人頭出現機率。
  - 如果正面是人頭像, 反面是數字, 則 $X \sim \text{binomial}(n, p)$
  - 假如錢幣正反兩面皆是「人頭」, 則伯努力程序中第 2, 第 3, 與第 4 個假設就不吻合。
- 檢驗某一生產線上 n 個產品, 令 X 為不良品出現的次數。
  - n 個產品皆在製程正常中產出, 則 $X \sim \text{binomial}(n, p)$
  - 假如產品不是在製程正常中生產, 則伯努力程序中第 3 與第 3 個假設就不適當, 因為從製程不正常的那時開始起生產的產品皆為不良品。

兩種情境說明吻合與不吻合伯努力程序

- 投擲一枚相同的錢幣 n 次, 令 X 為人頭面出現的次數, p 為人頭出現機率。
  - 如果正面是人頭像, 反面是數字, 則 $X \sim \text{binomial}(n, p)$
  - 假如錢幣正反兩面皆是「人頭」, 則伯努力程序中第 2, 第 3, 與第 4 個假設就不吻合。
- 檢驗某一生產線上 n 個產品, 令 X 為不良品出現的次數。
  - n 個產品皆在製程正常中產出, 則 $X \sim \text{binomial}(n, p)$
  - 假如產品不是在製程正常中生產, 則伯努力程序中第 3 與第 3 個假設就不適當, 因為從製程不正常的那時開始起生產的產品皆為不良品。
- 問卷調查 n 個顧客對某種產品的滿意度, 令 X 為勾選滿意的問卷份數。

兩種情境說明吻合與不吻合伯努力程序

- 投擲一枚相同的錢幣 n 次, 令 X 為人頭面出現的次數, p 為人頭出現機率。
  - 如果正面是人頭像, 反面是數字, 則 $X \sim \text{binomial}(n, p)$
  - 假如錢幣正反兩面皆是「人頭」, 則伯努力程序中第 2, 第 3, 與第 4 個假設就不吻合。
- 檢驗某一生產線上 n 個產品, 令 X 為不良品出現的次數。
  - n 個產品皆在製程正常中產出, 則 $X \sim \text{binomial}(n, p)$
  - 假如產品不是在製程正常中生產, 則伯努力程序中第 3 與第 3 個假設就不適當, 因為從製程不正常的那時開始起生產的產品皆為不良品。
- 問卷調查 n 個顧客對某種產品的滿意度, 令 X 為勾選滿意的問卷份數。
  - 如果 n 個顧客是隨機產生, 則 $X \sim \text{binomial}(n, p)$

兩種情境說明吻合與不吻合伯努力程序

- 投擲一枚相同的錢幣 n 次, 令 X 為人頭面出現的次數, p 為人頭出現機率。
  - 如果正面是人頭像, 反面是數字, 則 $X \sim \text{binomial}(n, p)$
  - 假如錢幣正反兩面皆是「人頭」, 則伯努力程序中第 2, 第 3, 與第 4 個假設就不吻合。
- 檢驗某一生產線上 n 個產品, 令 X 為不良品出現的次數。
  - n 個產品皆在製程正常中產出, 則 $X \sim \text{binomial}(n, p)$
  - 假如產品不是在製程正常中生產, 則伯努力程序中第 3 與第 3 個假設就不適當, 因為從製程不正常的那時開始起生產的產品皆為不良品。
- 問卷調查 n 個顧客對某種產品的滿意度, 令 X 為勾選滿意的問卷份數。
  - 如果 n 個顧客是隨機產生, 則 $X \sim \text{binomial}(n, p)$
  - 假如 n 個顧客都是哈日族, 而問卷調查的產品也是哈日族產品,

兩種情境說明吻合與不吻合伯努力程序

- 投擲一枚相同的錢幣 n 次, 令 X 為人頭面出現的次數, p 為人頭出現機率。
  - 如果正面是人頭像, 反面是數字, 則 $X \sim \text{binomial}(n, p)$
  - 假如錢幣正反兩面皆是「人頭」, 則伯努力程序中第 2, 第 3, 與第 4 個假設就不吻合。
- 檢驗某一生產線上 n 個產品, 令 X 為不良品出現的次數。
  - n 個產品皆在製程正常中產出, 則 $X \sim \text{binomial}(n, p)$
  - 假如產品不是在製程正常中生產, 則伯努力程序中第 3 與第 3 個假設就不適當, 因為從製程不正常的那時開始起生產的產品皆為不良品。
- 問卷調查 n 個顧客對某種產品的滿意度, 令 X 為勾選滿意的問卷份數。
  - 如果 n 個顧客是隨機產生, 則 $X \sim \text{binomial}(n, p)$
  - 假如 n 個顧客都是哈日族, 而問卷調查的產品也是哈日族產品, 則伯努力程序中第 4 個假設可能不適當。

- 吻合與不吻合伯努力程序, 必須視情形而定。

兩種情境說明吻合與不吻合伯努力程序

- 投擲一枚相同的錢幣 n 次, 令 X 為人頭面出現的次數, p 為人頭出現機率。
  - 如果正面是人頭像, 反面是數字, 則 $X \sim \text{binomial}(n, p)$
  - 假如錢幣正反兩面皆是「人頭」, 則伯努力程序中第 2, 第 3, 與第 4 個假設就不吻合。
- 檢驗某一生產線上 n 個產品, 令 X 為不良品出現的次數。
  - n 個產品皆在製程正常中產出, 則 $X \sim \text{binomial}(n, p)$
  - 假如產品不是在製程正常中生產, 則伯努力程序中第 3 與第 3 個假設就不適當, 因為從製程不正常的那時開始起生產的產品皆為不良品。
- 問卷調查 n 個顧客對某種產品的滿意度, 令 X 為勾選滿意的問卷份數。
  - 如果 n 個顧客是隨機產生, 則 $X \sim \text{binomial}(n, p)$
  - 假如 n 個顧客都是哈日族, 而問卷調查的產品也是哈日族產品, 則伯努力程序中第 4 個假設可能不適當。

- 吻合與不吻合伯努力程序, 必須視情形而定。

二項分配 (Binomial Distribution)

- 若一離散型隨機變數 X 的 pmf 為

$$f_X(x) = \begin{cases} C_x^n \, p^x (1-p)^{n-x}, & x = 0, 1, 2, 3, \ldots, n \\ 0, & \text{其他}, \end{cases}$$

則稱 X 服從二項分配, 記做 $X \sim \text{binomial}(n, p)$ 或簡寫成 $X \sim \text{bin}(n, p)$, 其中 $C_x^n = \frac{n!}{x!(n-x)!}$ 為在 n 個不同產品中取 x 個產品的不同取法總數。

- X 的平均數: $\text{E}(X) = np$
- X 的變異數: $\text{V}(X) = np(1-p)$

- 生活中有哪些實例可以用 “二項分配” 來描述?

二項分配 (Binomial Distribution)

- 若一離散型隨機變數 X 的 pmf 為

$$f_X(x) = \begin{cases} C_x^n\, p^x (1-p)^{n-x}, & x = 0, 1, 2, 3, \ldots, n \\ 0, & \text{其他}, \end{cases}$$

則稱 X 服從二項分配, 記做 $X \sim \text{binomial}(n, p)$ 或簡寫成 $X \sim \text{bin}(n, p)$, 其中 $C_x^n = \frac{n!}{x!(n-x)!}$ 為在 n 個不同產品中取 x 個產品的不同取法總數。

- X 的平均數: $\text{E}(X) = np$
- X 的變異數: $\text{V}(X) = np(1-p)$

- 生活中有哪些實例可以用 "二項分配" 來描述?

二項分配應用

- 在通訊學上, 傳輸的資訊是一連串二碼的數字: 0 與 1。因通訊誤差的關係, 每個數字可能傳錯, 假設傳輸錯誤率為 p; 如:「送出為 0, 卻收到 1」的機率為 p。為了降低傳輸錯誤率, "連傳法" 是將每一個數字連續送出 n 次, n 為奇數。以 $n = 3$ 為例, 假設一串 4 個原始數字為 "0, 1, 1, 0 "。連傳法送出的數字為: (0,0,0), (1,1,1), (1,1,1), (0,0,0)。將收到的每 3 個數字的衆數當成收到的數字。如收到 (0,0,1) 或 (0,0,0) 則當成 0。求算連傳法的傳輸錯誤率。
- 令 X 為 n 次傳輸中錯誤的次數, 假設每個數字傳輸結果獨立, 則 $X \sim$ binomial(n, p)。
  連傳法之錯誤率為 P$(X \geq (n+1)/2)$。
  以 $n = 3$, $p = 0.5$ 為例, 連傳法之傳輸錯誤率為
  P$(X \geq 2) = 0.25$, 是原始傳輸錯誤率的一半。

幾何分配 (Geometric Distribution)

- 若一離散型隨機變數 X 的 pmf 為

$$f_X(x) = \begin{cases} p(1-p)^{x-1}, & x = 1, 2, 3, \ldots; 0 < p < 1 \\ 0, & \text{其他}, \end{cases}$$

則稱 X 服從幾何分配, 記做 $X \sim$ geometric(p)。

- X 的平均數: $\mathrm{E}(X) = 1/p$
- X 的變異數: $\mathrm{V}(X) = (1-p)/p^2$

- 生活中有哪些實例可以用 “二項分配” 來描述?

幾何分配 (Geometric Distribution)

- 若一離散型隨機變數 X 的 pmf 為

$$f_X(x) = \begin{cases} p(1-p)^{x-1}, & x = 1, 2, 3, \ldots; 0 < p < 1 \\ 0, & \text{其他}, \end{cases}$$

則稱 X 服從幾何分配, 記做 $X \sim \text{geometric}(p)$。

- X 的平均數: $\mathrm{E}(X) = 1/p$
- X 的變異數: $\mathrm{V}(X) = (1-p)/p^2$

- 生活中有哪些實例可以用 “二項分配” 來描述?

生育政策

假設A市重男輕女風氣很盛,所有的父母都非得生個男孩不可,為了控制人口膨脹,市政府立法:「每對夫婦在生了一個男嬰之後便要節育」。假設每胎生男或生女的機率相等,你認為這個立法會不會導致人口膨脹?會不會導致男多於女?

- (i) 想法一: 第一胎就生男孩的家庭不再繼續生育,所以人口不但不會膨脹,反而會減少。每家一定有男孩,未必有女孩,所以男會多於女。

生育政策

假設A市重男輕女風氣很盛,所有的父母都非得生個男孩不可,為了控制人口膨脹,市政府立法:「每對夫婦在生了一個男嬰之後便要節育」。假設每胎生男或生女的機率相等,你認為這個立法會不會導致人口膨脹? 會不會導致男多於女?

- (i) 想法一: 第一胎就生男孩的家庭不再繼續生育,所以人口不但不會膨脹,反而會減少。每家一定有男孩,未必有女孩,所以男會多於女。
- (ii) 想法二: 此法鼓勵沒生男孩的家庭繼續生育,直到生男孩為止,家庭人口會愈來愈多,且女會多於男。

生育政策

假設A市重男輕女風氣很盛,所有的父母都非得生個男孩不可,為了控制人口膨脹,市政府立法:「每對夫婦在生了一個男嬰之後便要節育」。假設每胎生男或生女的機率相等,你認為這個立法會不會導致人口膨脹? 會不會導致男多於女?

- (i) 想法一: 第一胎就生男孩的家庭不再繼續生育,所以人口不但不會膨脹,反而會減少。每家一定有男孩,未必有女孩,所以男會多於女。
- (ii) 想法二: 此法鼓勵沒生男孩的家庭繼續生育,直到生男孩為止,家庭人口會愈來愈多,且女會多於男。
- (iii) 想法三: 假設每對夫婦皆能生育。令 p 是每胎生男的機率,且每一胎之間彼此獨立。令 X 代表一家庭的子女數目,則 X 服從幾何分配 geometric(p)。若 $p = 1/2$, 則 $\mathrm{E}(X) = 1/p = 2$

生育政策

假設A市重男輕女風氣很盛, 所有的父母都非得生個男孩不可, 為了控制人口膨脹, 市政府立法:「每對夫婦在生了一個男嬰之後便要節育」。假設每胎生男或生女的機率相等, 你認為這個立法會不會導致人口膨脹? 會不會導致男多於女?

- (i) 想法一: 第一胎就生男孩的家庭不再繼續生育, 所以人口不但不會膨脹, 反而會減少。每家一定有男孩, 未必有女孩, 所以男會多於女。
- (ii) 想法二: 此法鼓勵沒生男孩的家庭繼續生育, 直到生男孩為止, 家庭人口會愈來愈多, 且女會多於男。
- (iii) 想法三: 假設每對夫婦皆能生育。令 p 是每胎生男的機率, 且每一胎之間彼此獨立。令 X 代表一家庭的子女數目, 則 X 服從幾何分配 geometric(p)。若 $p = 1/2$, 則 $\mathrm{E}(X) = 1/p = 2$

- 你認同哪個想法?

無記憶性

- 定義: 若隨機變數 X 滿足
  $\mathsf{P}(X \geq t+s|X \geq s) = \mathsf{P}(X \geq t)$, 其中 $t, s \geq 0$, 則稱該隨機變數之分配具有無記憶性。
- 範例: 投擲一個錢幣直到出現正面為止。比較以下兩事件的機率大小:
  事件 A: 假如張生已經投擲了 10 次皆未出現正面, 則再投擲 2 次仍未出現正面的事件。
  事件 B: 李生投擲 2 次未出現正面的事件。

- 無記憶性表示 $\mathsf{P}(X \geq t+s|X \geq s)$ 的機率值與 s 無關, 亦即 s 可為任何數值。
- 但在此需注意的是不要寫成 $\mathsf{P}(X \geq t+s|X \geq s) = \mathsf{P}(X \geq t+s)$
- 哪些分配具有無記憶性?

無記憶性

- 定義: 若隨機變數 X 滿足 $\mathsf{P}(X \geq t+s|X \geq s) = \mathsf{P}(X \geq t)$, 其中 $t, s \geq 0$, 則稱該隨機變數之分配具有無記憶性。
- 範例: 投擲一個錢幣直到出現正面為止。比較以下兩事件的機率大小:

  事件 A: 假如張生已經投擲了 10 次皆未出現正面, 則再投擲 2 次仍未出現正面的事件。

  事件 B: 李生投擲 2 次未出現正面的事件。

- 無記憶性表示 $\mathsf{P}(X \geq t+s|X \geq s)$ 的機率值與 s 無關, 亦即 s 可為任何數值。
- 但在此需注意的是不要寫成 $\mathsf{P}(X \geq t+s|X \geq s) = \mathsf{P}(X \geq t+s)$
- 哪些分配具有無記憶性? 參考書中 25 分配與 Song's 80 分配之關聯圖

負二項分配 (Negative Binomail Dist.)

- 若 X 代表實驗總次數, 且其 pmf 為

$$f_X(x) = \begin{cases} C_{k-1}^{x-1}\, p^k (1-p)^{x-k}, & x = k, k+1, k+2, \ldots \\ 0, & \text{其他}, \end{cases}$$

則稱 X 服從負二項分配, 記做 $X \sim \mathrm{NB}(k, p)$。

- X 的平均數: $\mathrm{E}(X) = k/p$
- X 的變異數: $\mathrm{V}(X) = k(1-p)/p^2$

- 生活中有哪些實例可以用 "負二項分配" 來描述?

負二項分配 (Negative Binomail Dist.)

- 若 X 代表實驗總次數, 且其 pmf 為

$$f_X(x) = \begin{cases} C_{k-1}^{x-1}\, p^k (1-p)^{x-k}, & x = k, k+1, k+2, \ldots \\ 0, & \text{其他}, \end{cases}$$

  則稱 X 服從負二項分配, 記做 $X \sim \mathrm{NB}(k, p)$。

- X 的平均數: $\mathrm{E}(X) = k/p$
- X 的變異數: $\mathrm{V}(X) = k(1-p)/p^2$

- 生活中有哪些實例可以用 "負二項分配" 來描述?

險勝或大勝

假設乒乓球團體錦標決賽之比賽規則為: 比賽9局, 先取得5局勝利的隊伍便是冠軍。假設2000年A隊以5比4的比數取得勝利, 2002年A隊以5比3的比數取得勝利, 我們可不可以說:「2000年A隊險勝B隊, 但2002年A隊大勝B隊」, 或者說以5:3獲勝的A隊比以5:4獲勝更技高一籌呢?

- (i) 想法一: 5:3 代表兩隊實力差距大, 而 5:4 代表兩隊實力比較接近。換言之, 5:3 表示勝利的一方球技更高一籌。
- (ii) 想法二: 令X為A隊第5次獲勝所需之終場局數, X=5, 6, …; p 表示A隊每局獲勝的機率; 則 $X \sim \text{NB}(5, p)$。如果假設兩隊實力相當 $(p = 0.5)$, 則A隊以5:4獲勝機率與以5:3獲勝機率分別為

$$\text{P}(X = 9) = C_4^8\,(0.5)^9 = 70 \times (0.5)^9 \simeq 0.137$$
$$\text{P}(X = 8) = C_4^7\,(0.5)^8 = 35 \times (0.5)^8 = 70 \times (0.5)^9 \simeq 0.137$$

險勝或大勝

假設乒乓球團體錦標決賽之比賽規則為: 比賽9局, 先取得5局勝利的隊伍便是冠軍。假設2000年A隊以5比4的比數取得勝利, 2002年A隊以5比3的比數取得勝利, 我們可不可以說:「2000年A隊險勝B隊, 但2002年A隊大勝B隊」, 或者說以5:3獲勝的A隊比以5:4獲勝更技高一籌呢?

- (i) 想法一: 5:3代表兩隊實力差距大, 而5:4代表兩隊實力比較接近。換言之, 5:3表示勝利的一方球技更高一籌。
- (ii) 想法二: 令X為A隊第5次獲勝所需之終場局數, X=5, 6, …; p表示A隊每局獲勝的機率; 則 $X \sim \text{NB}(5, p)$。如果假設兩隊實力相當 $(p = 0.5)$, 則A隊以5:4獲勝機率與以5:3獲勝機率分別為

$$\text{P}(X = 9) = C_4^8\,(0.5)^9 = 70 \times (0.5)^9 \simeq 0.137$$
$$\text{P}(X = 8) = C_4^7\,(0.5)^8 = 35 \times (0.5)^8 = 70 \times (0.5)^9 \simeq 0.137$$

- 你認同哪個想法?

超幾何分配 (Hyper-Geometric Dist.)

- 若一離散型隨機變數 X 的 pmf 為

$$f_X(x) = \begin{cases} \dfrac{C_x^K C_{n-x}^{N-K}}{C_n^N}, & x = \max\{n-N+K, 0\}, \dots, \min\{K, n\} \\ 0, & \text{其他}, \end{cases} \quad (1)$$

則稱 X 服從超幾何分配, 記做 $X \sim \mathrm{HG}(N, K, n)$。

- X 的平均數: $\mathrm{E}(X) = np$, $p = K/N$
- X 的變異數: $\mathrm{V}(X) = np(1-p)\frac{N-n}{N-1}$

- 生活中有哪些實例可以用 “超幾何分配” 來描述?

超幾何分配 (Hyper-Geometric Dist.)

- 若一離散型隨機變數 X 的 pmf 為

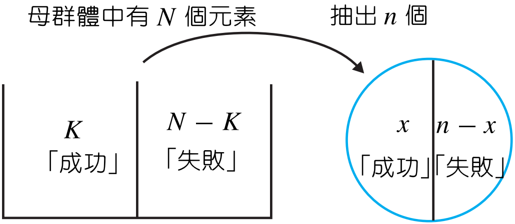
$$f_X(x) = \begin{cases} \dfrac{C_x^K C_{n-x}^{N-K}}{C_n^N}, & x = \max\{n-N+K, 0\}, \ldots, \\ & \quad \min\{K, n\} \\ 0, & \text{其他}, \end{cases} \tag{1}$$

則稱 X 服從超幾何分配, 記做 $X \sim \text{HG}(N, K, n)$。

- X 的平均數: $\text{E}(X) = np$, $p = K/N$
- X 的變異數: $\text{V}(X) = np(1-p)\frac{N-n}{N-1}$

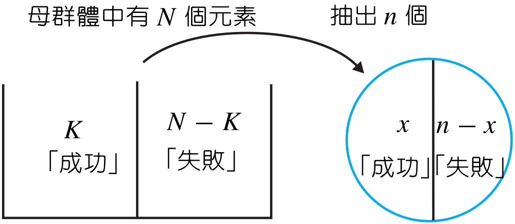
- 生活中有哪些實例可以用 “超幾何分配” 來描述?

超幾何分配示意圖



- 超幾何分配與二項分配關係如何?

超幾何分配示意圖



- 超幾何分配與二項分配關係如何?
- 若 X 服從超幾何分配 (N, K, n), 則當 $\frac{n}{N} \to 0$ 時, 此超幾何分配會趨近二項分配 (n, p), 其中 $p = \frac{K}{N}$
- 超幾何分配與二項分配的平均數皆為 np, 但超幾何分配之變異數比二項分配之變異數多了一項 $\frac{N-n}{N-1}$, 稱為修正項。當 N 很大, 而 n 很小時, 超幾何分配會趨近於二項分配。

台灣樂透彩

張生買了一張台灣樂透彩 (任由42 個數字中抽取 6 個), 中頭獎 (6 個數字完全猜中) 的機率為何?

台灣樂透彩

張生買了一張台灣樂透彩 (任由42 個數字中抽取 6 個), 中頭獎 (6 個數字完全猜中) 的機率為何?

- 令 X 為猜中的個數, 則 $X \sim \mathrm{HG}(N = 42, K = 6, n = 6)$。

台灣樂透彩

張生買了一張台灣樂透彩 (任由42 個數字中抽取 6 個), 中頭獎 (6 個數字完全猜中) 的機率為何?

- 令 X 為猜中的個數, 則 $X \sim \mathrm{HG}(N = 42, K = 6, n = 6)$。
- $\mathrm{P}(X = 6) = \dfrac{C_6^6 C_0^{36}}{C_6^{42}} = 1/5245786$, 大約是五百萬分之一

台灣樂透彩

張生買了一張台灣樂透彩 (任由42 個數字中抽取 6 個), 中頭獎 (6 個數字完全猜中) 的機率為何?

- 令 X 為猜中的個數, 則 $X \sim \text{HG}(N = 42, K = 6, n = 6)$。
- $\text{P}(X = 6) = \dfrac{C_6^6 C_0^{36}}{C_6^{42}} = 1/5245786$, 大約是五百萬分之一
- 猜中5 個數字的機率為何?

台灣樂透彩

張生買了一張台灣樂透彩 (任由42 個數字中抽取 6 個), 中頭獎 (6 個數字完全猜中) 的機率為何?

- 令 X 為猜中的個數, 則 $X \sim \text{HG}(N = 42, K = 6, n = 6)$。
- $\text{P}(X = 6) = \dfrac{C_6^6 C_0^{36}}{C_6^{42}} = 1/5245786$, 大約是五百萬分之一

- 猜中5 個數字的機率為何?
  $\text{P}(X = 5) = \dfrac{C_5^6 C_1^{36}}{C_6^{42}} = 216/5245786 \simeq 0.00004117$

抽籤順序

新竹實驗中學每年新生名額有限, 除了父母親在科學園區上班的學生可直接入學外, 其它的學生必需抽籤決定。張先生認為後抽籤的結果與抽取順序有關, 所以主張先抽取順序籤, 再依順序籤之順序來抽籤。你同意張先生的看法嗎?

- 假設 N 個球中有 K 個白球, 我們從 N 個球中抽出 n 個球 (不放回)。令

$$X_i = \begin{cases} 1, & \text{若第 } i \text{ 次抽中白球} \\ 0, & \text{其他,} \end{cases}$$

- $\mathrm{E}(X_i) = K/N, i = 1, 2, \ldots, N$
- 此例中, 抽中白球就是抽中入學, 此機率與抽取順序無關, 所以抽取順序籤是多此一舉。
- 換一個角度來想, 讓一般大眾瞭解抽取順序無關是不容易的事。若能讓一般大眾求得心安, 抽取順序籤也並非完全是無意義的舉動。

卜瓦松分配 (Poisson Dist.)

- 若一離散型隨機變數 X 的 pmf 為

$$f_X(x) = \begin{cases} \dfrac{\mu^x e^{-\mu}}{x!}, & x = 0, 1, 2, 3, \ldots \\ 0, & \text{其他}, \end{cases}$$

則稱 X 服從卜瓦松分配, 記做 $X \sim \text{Poisson}(\mu)$。

- X 的平均數: $\text{E}(X) = \mu$
- X 的變異數: $\text{V}(X) = \mu$

- 生活中有哪些實例可以用 “卜瓦松分配” 來描述?

卜瓦松分配 (Poisson Dist.)

- 若一離散型隨機變數 X 的 pmf 為

$$f_X(x) = \begin{cases} \dfrac{\mu^x e^{-\mu}}{x!}, & x = 0, 1, 2, 3, \ldots \\ 0, & \text{其他,} \end{cases}$$

則稱 X 服從卜瓦松分配, 記做 $X \sim \text{Poisson}(\mu)$。

- X 的平均數: $\text{E}(X) = \mu$
- X 的變異數: $\text{V}(X) = \mu$

- 生活中有哪些實例可以用 "卜瓦松分配" 來描述?

卜瓦松分配與二項分配

- 若 $X \sim \text{bino}(n, p)$, 且 $\lim\limits_{\substack{n\to\infty \\ p\to 0}} np = \mu, 0 < \mu < \infty$, 則
$$\lim_{n\to\infty} f_X(x) = \frac{\mu^x}{x!}e^{-\mu}$$
- 應用: 假設根據經驗每 1,000 份報稅單中就有一份填寫有錯誤。在國稅局上班的稅務員張先生要審查 10,000 份稅單, 張先生發現填寫錯誤的報稅單數目為 6, 7, 或 8 份的機率是多少?
  - 令 X 為填寫錯誤的報稅單份數, n 為 10,000 份報稅單, p 為每份填錯的機率; 則 $X \sim \text{bin}(n = 10000, p = 0.001)$, 因為 1000 份稅單可看成 1000 個重複伯努力實驗。錯誤的報稅單數目為 6, 7, 或 8 份的機率為

$$\text{P}(X = 6, 7, 8) = \sum_{i=6}^{8} C_i^{10000}\,(0.001)^i (0.999)^{10000-i}$$

要算出最後的答案有實際計算上的困難! 所以...

卜瓦松分配之應用

假如每天在高速公路的車禍發生事件總數服從卜瓦松隨機變數,且每天平均發生車禍3次。求算

- 今天沒有發生車禍事件的機率為何?
- 明後兩天沒有發生車禍事件的機率為何?
- 一星期內發生3次車禍事件的機率為何?

- $X \sim \text{Poisson}(\mu = \lambda t = 3)$,
  $\text{P}(X = 0) = e^{-3} \simeq 0.05$
- $X \sim \text{Poisson}(\mu = \lambda t = 3 \times 2 = 6)$
  $\text{P}(X = 0) = e^{-6} \simeq 0.0025$
- $X \sim \text{Poisson}(\mu = \lambda t = 3 \times 7 = 21)$
  $\text{P}(X = 3) = 21^3 e^{-21}/3! \simeq 1.17 \times 10^{-6}$

- 討論: 高速公路的車禍發生事件總數服從卜瓦松隨機變數

卜瓦松分配之應用

假如每天在高速公路的車禍發生事件總數服從卜瓦松隨機變數,且每天平均發生車禍3次。求算

- 今天沒有發生車禍事件的機率為何?
- 明後兩天沒有發生車禍事件的機率為何?
- 一星期內發生3次車禍事件的機率為何?

- $X \sim \text{Poisson}(\mu = \lambda t = 3)$,
  $\text{P}(X = 0) = e^{-3} \simeq 0.05$
- $X \sim \text{Poisson}(\mu = \lambda t = 3 \times 2 = 6)$
  $\text{P}(X = 0) = e^{-6} \simeq 0.0025$
- $X \sim \text{Poisson}(\mu = \lambda t = 3 \times 7 = 21)$
  $\text{P}(X = 3) = 21^3 e^{-21}/3! \simeq 1.17 \times 10^{-6}$

- 討論: 高速公路的車禍發生事件總數服從卜瓦松隨機變數

卜瓦松分配之應用

假設在統計課本上,一頁中出現排版錯誤的數目服從卜瓦松分配,每一頁平均有1個錯誤。求算

- 至少有一個錯誤發生在第1頁的機率為何?
- 正好有一個錯誤發生在3頁內的機率為何?

- $X \sim \text{Poisson}(\mu = \lambda t = 1)$
- $\text{P}(X \geq 1) = 1 - \text{P}(X = 0) = 1 - e^{-1} \simeq 0.632$
- $X \sim \text{Poisson}(\mu = \lambda t = 3)$
- $\text{P}(X = 1) = 3e^{-3} \simeq 0.149$

卜瓦松分配之應用

假設在統計課本上, 一頁中出現排版錯誤的數目服從卜瓦松分配, 每一頁平均有1個錯誤。求算

- 至少有一個錯誤發生在第1頁的機率為何?
- 正好有一個錯誤發生在3頁內的機率為何?

- $X \sim \text{Poisson}(\mu = \lambda t = 1)$
- $\text{P}(X \geq 1) = 1 - \text{P}(X = 0) = 1 - e^{-1} \simeq 0.632$
- $X \sim \text{Poisson}(\mu = \lambda t = 3)$
- $\text{P}(X = 1) = 3e^{-3} \simeq 0.149$

- 討論: 出現排版錯誤的數目服從卜瓦松分配

卜瓦松分配之應用

假設在統計課本上,一頁中出現排版錯誤的數目服從卜瓦松分配,每一頁平均有1個錯誤。求算

- 至少有一個錯誤發生在第1頁的機率為何?
- 正好有一個錯誤發生在3頁內的機率為何?

- $X \sim \text{Poisson}(\mu = \lambda t = 1)$
- $\text{P}(X \geq 1) = 1 - \text{P}(X = 0) = 1 - e^{-1} \simeq 0.632$
- $X \sim \text{Poisson}(\mu = \lambda t = 3)$
- $\text{P}(X = 1) = 3e^{-3} \simeq 0.149$

- 討論: 出現排版錯誤的數目服從卜瓦松分配