

# 醫學統計學

王超辰 Chaochen Wang

最近更新於 2017-11-14

献给……

呃，爱谁谁吧

# 目录



# 表格



## 插图





# 前言

尚未想好寫什麼作前言。我只是默默地想留下一些筆記和思考。

本書用了兩個 R 包編譯，分別是 **knitr** (?) 和 **bookdown** (?)。以下是我的 R 進程信息：

```
sessionInfo()
```

```
## R version 3.4.2 (2017-09-28)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 15063)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=Chinese (Simplified)_China.936
## [2] LC_CTYPE=Chinese (Simplified)_China.936
## [3] LC_MONETARY=Chinese (Simplified)_China.936
## [4] LC_NUMERIC=C
## [5] LC_TIME=Chinese (Simplified)_China.936
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets
## [6] methods    base
##
## other attached packages:
## [1] ggplot2_2.2.1
```

```
##
## loaded via a namespace (and not attached):
## [1] Rcpp_0.12.13      rstudioapi_0.7    knitr_1.17
## [4] magrittr_1.5       munsell_0.4.3     colorspace_1.3-2
## [7] rlang_0.1.4        highr_0.6          stringr_1.2.0
## [10] plyr_1.8.4         tools_3.4.2        grid_3.4.2
## [13] gtable_0.2.0       png_0.1-7          htmltools_0.3.6
## [16] yaml_2.1.14        lazyeval_0.2.1     rprojroot_1.2
## [19] digest_0.6.12      tibble_1.3.4       bookdown_0.5
## [22] evaluate_0.10.1    rmarkdown_1.7      labeling_0.3
## [25] stringi_1.1.5      compiler_3.4.2     scales_0.5.0
## [28] backports_1.1.1
```

王超辰於倫敦

# 我是誰

歡迎參觀我的個人主頁<sup>1</sup>。

---

<sup>1</sup><https://winterwang.github.io>



# 第一部分

## 概率論 Probability



# 第一章 概率論入門：定義與公理

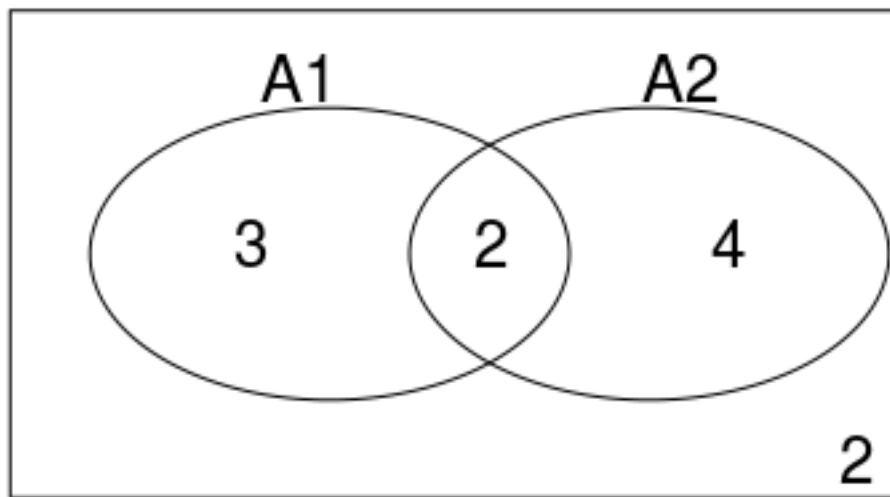
## 1.1 三個概率公理：

1. 對於任意事件  $A$ ，它發生的概率  $P(A)$  滿足這樣的不等式： $0 \leq P(A) \leq 1$
2.  $P(\Omega) = 1$ ， $\Omega$  是全樣本空間 (total sample space)
3. 對於互斥（相互獨立）的事件  $A_1, A_2, \dots, A_n$  有如下的等式關係：
$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

你是不是覺得上面三條公理都是廢話。不用擔心，我也是這麼覺得的。因為所有人都認同的道理，才能成為公理 (axiom)，因為它們是不需要證明的自然而然形成的人人都接受的觀念。(axiom: a saying that is widely accepted on its own merits; its truth is assumed to be self-evident)

然而，正是這樣顯而易見的道理，確是拿來建築理論的基石，千萬不能小看了他們。例如，我們看下面這個看似也應該成為公理的公式，你能證明嗎：

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$



證明：

先考慮  $A_1 \cup A_2$  是什麼（拆分成三個互斥事件）

$$A_1 \cup A_2 = (A_1 \cap \bar{A}_2) \cup (\bar{A}_1 \cap A_2) \cup (A_1 \cap A_2)$$

運用上面的公理 2.3

$$\therefore P(A_1 \cup A_2) = P(A_1 \cap \bar{A}_2) + P(\bar{A}_1 \cap A_2) + P(A_1 \cap A_2) \quad (1)$$

再考慮  $A_1 = (A_1 \cap A_2) \cup (A_1 \cap \bar{A}_2)$  繼續拆分成兩個互斥事件

$$\therefore P(A_1) = P(A_1 \cap A_2) + P(A_1 \cap \bar{A}_2) \text{ 整理一下:}$$

$$P(A_1 \cap \bar{A}_2) = P(A_1) - P(A_1 \cap A_2)$$



同理可得:  $P(\bar{A}_1 \cap A_2) = P(A_2) - P(A_1 \cap A_2)$

代入上面第 (1) 式可得:

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1) - P(A_1 \cap A_2) \\ &+ P(A_2) - P(A_1 \cap A_2) \\ &+ P(A_1 \cap A_2) \\ &= P(A_1) + P(A_2) - P(A_1 \cap A_2) \end{aligned}$$

## 1.2 條件概率 Conditional probability

- $P(A|S) = \frac{P(A \cap S)}{P(S)}$
- $P(A \cap S) = P(A|S)P(S)$

## 1.3 獨立 (independence) 的定義

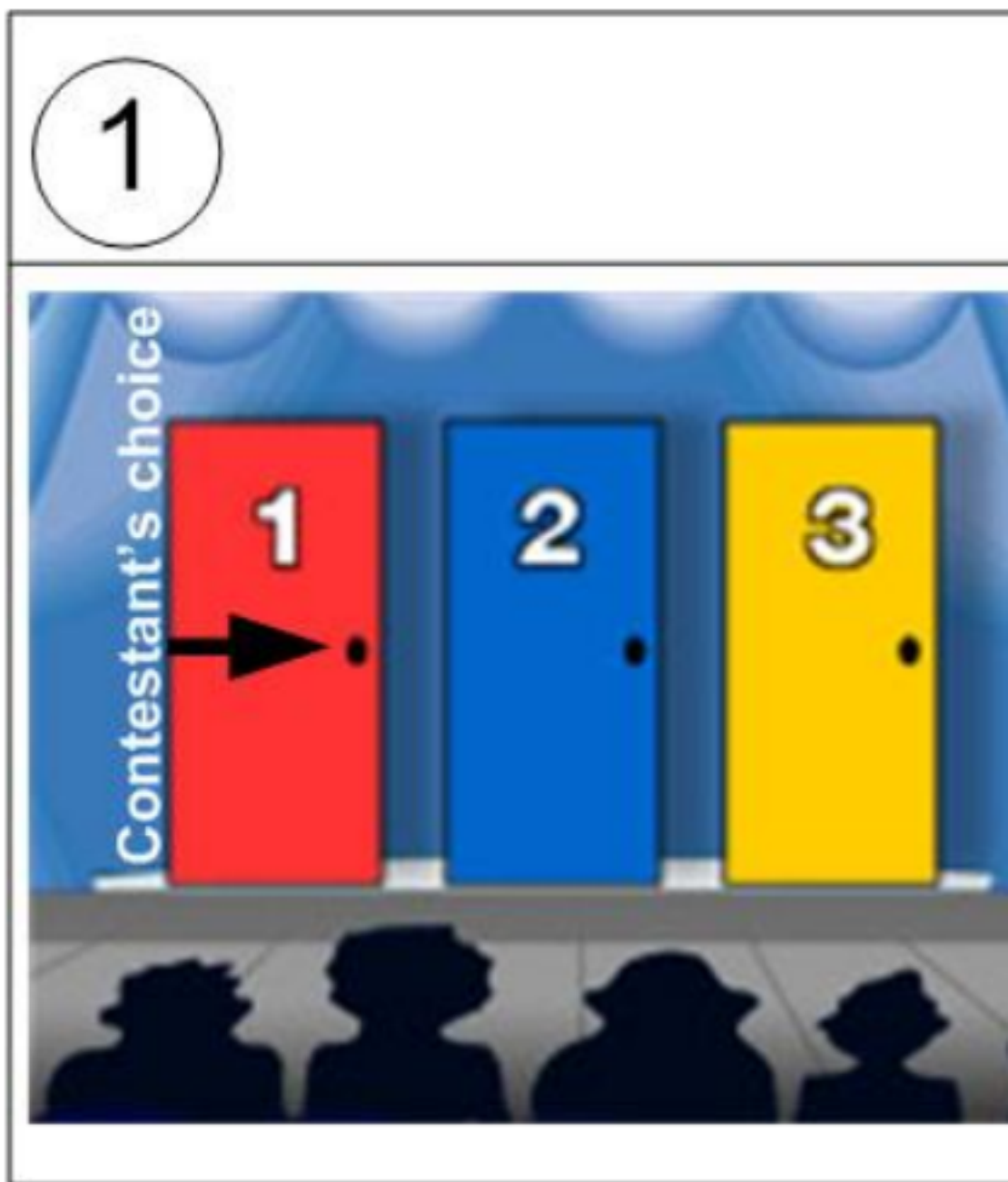
- 兩個事件定義為互為獨立時 ( $A$  and  $B$  are said to be independent **if and only if**)

$$P(A \cap B) = P(A)P(B)$$

- 因為從條件概率的概念我們已知  $P(A \cap B) = P(A|B)P(B)$  所以  $P(A|B) = P(A)$  即: 事件  $B$  無法提供事件  $A$  的任何有效訊息 ( $A, B$  互相獨立)

## 1.4 賭博問題

終於來到本次話題的“重點”了。



假設你在一個電視遊戲節目。有上圖一樣的三扇門。其中一扇門後面有

一輛保時捷，另兩扇門後面則是 (味道奇特的) 山羊<sup>1</sup>。遊戲規則是主持人會讓你先選擇其中一扇門（先不打開你選的這扇門）。主持人隨後打開另外兩扇門中的一扇沒有保時捷的門。主持人問你，你要堅持選擇之前選中的那扇門，還是要改變主意換一扇門去猜是否可以猜中保時捷。請問，堅持選擇之前選中的門猜中保時捷的概率高，還是主持人打開一扇門以後改變主意猜中保時捷的概率更高呢？

## 1.5 賭博問題的答案

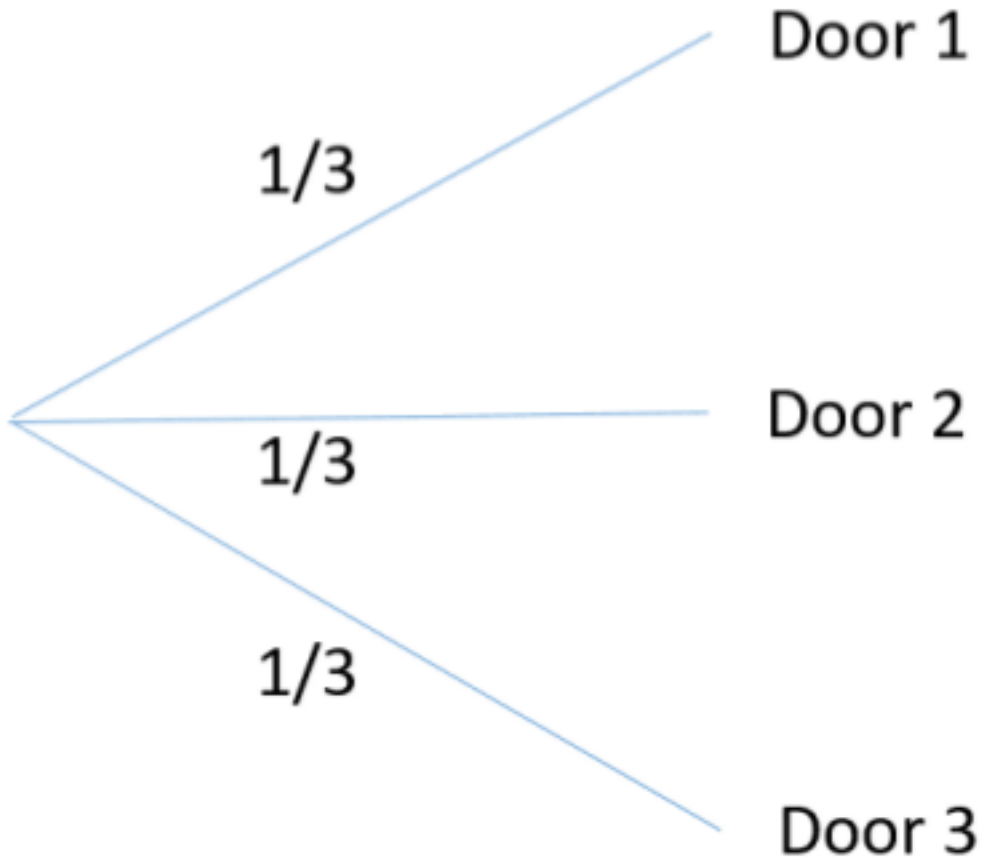
答案是：必須改變主意才能提高中獎概率。

上述情況下，最簡單的是用概率樹 (probability tree) 來做決定：

---

<sup>1</sup><https://winterwang.github.io/post/black-meal/>

Door you initially choose



解說一下：

- 假定保時捷在 1 號門後，你第一次選擇了 1 號門，那麼此時主持人可以任意打開 2 號或者三號門（因為他們後面都沒有保時捷）。
- 假定保時捷在 1 號門後，你第一次選了 2 號門，那麼此時主持人只能打開 3 號門（因為一號門後是保時捷，按照遊戲規則主持人不能打開）。
- 假定保時捷在 1 號門後，你第一次選了 3 號門，那麼此時主持人只能打開 2 號門（因為一號門後是保時捷，按照遊戲規則主持人不能打開）。

所以按照圖中給出的計算概率樹的過程可以得到：

$$P[change] = \frac{1}{3} + \frac{1}{3} = \frac{2}{3} \quad P[not\ change] = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

你是否選擇了改變主意了呢？



## 第二章 Bayes 貝葉斯理論的概念

許多時候，我們需要將概率中的條件相互對調。例如：在已知該人羣中有 20% 的人有吸菸習慣 ( $P(S)$ )，吸菸的人有 9% 的概率有哮喘 ( $P(A|S)$ )，不吸菸的人有 7% 的概率有哮喘 ( $P(A|\bar{S})$ ) 的前提下，有個人前來門診，發現是哮喘患者，那麼這個人有多大的概率是一個菸民？也就是要求  $P(S|A)$

這裏先引入貝葉斯的概念：

我們可以將  $P(A \cap S)$  寫成：

$$P(A \cap S) = P(A|S)P(S) \text{ or } P(A \cap S) = P(S|A)P(A)$$

這兩個等式是完全等價的。我們將他們連起來：

$$P(S|A)P(A) = P(A|S)P(S) \Rightarrow P(S|A) = \frac{P(A|S)P(S)}{P(A)}$$

是不是看起來又像是寫了一堆廢話？沒錯，你看出來是一堆廢話的時候，證明你也同意這背後的簡單邏輯。

再繼續，我們可以利用另外一個廢話： $\because S + \bar{S} = 1$

$$\therefore P(A) = P(A \cap S) + P(A \cap \bar{S})$$

用上面的公式替換掉  $P(A \cap S) + P(A \cap \bar{S})$

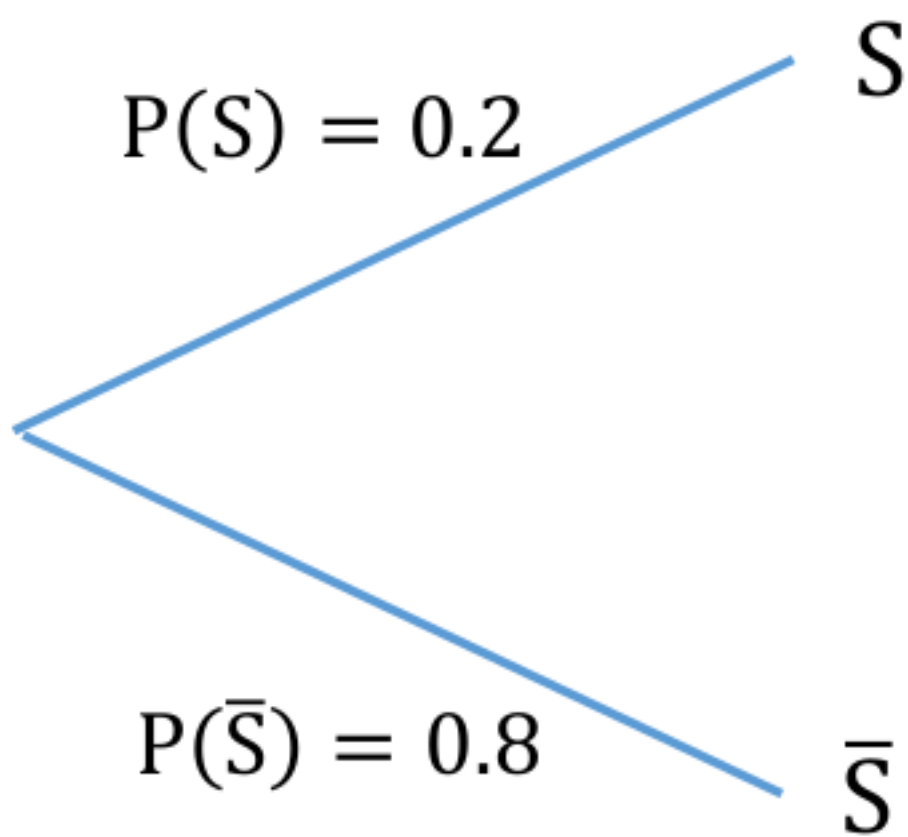
$$\therefore P(A) = P(A|S)P(S) + P(A|\bar{S})P(\bar{S})$$

可以得到貝葉斯理論公式：

$$P(S|A) = \frac{P(A|S)P(S)}{P(A|S)P(S) + P(A|\bar{S})P(\bar{S})}$$

回到上面說到的哮喘人中有多少比例吸菸的問題。可以繼續使用概率樹來方便的計算：





$$P(S|A) = \frac{P(A|S)P(S)}{P(A|S)P(S) + P(A|\bar{S})P(\bar{S})} \quad (2.1)$$

$$= \frac{0.09 \times 0.2}{0.09 \times 0.2 + 0.07 \times 0.8} \quad (2.2)$$

$$= 0.24 \quad (2.3)$$

所以我們的結論就是，在已知該人羣中有 20% 的人有吸菸習慣 ( $P(S)$ )，吸菸的人有 9% 的概率有哮喘 ( $P(A|S)$ )，不吸菸的人有 7% 的概率有哮喘 ( $P(A|\bar{S})$ ) 的前提下，有個人前來門診，發現是哮喘患者，那麼這個人有 24% 的概率是一個菸民 ( $P(S|A)$ )。

## 第三章 期望 Expectation (或均值 or mean) 和方差 Variance

期望（或均值）是用來描述一組數據中心位置的指標（另一個是中位數 Median）。對於離散型隨機變量  $X$  (discrete random variables)，它的期望被定義為：

$$E(X) = \sum_x xP(X = x)$$

所以就是將所有  $X$  可能取到的值乘以相應的概率後求和。這個期望（或均值）常常用希臘字母  $\mu$  來標記。

方差 Variance 是衡量一組數據變化幅度 (dispersion/variability) 的指標之一。方差的定義是：

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

實際上我們更加常用的是它的另外一個公式：

$$Var(X) = E(X^2) - E(X)^2$$

證明上面兩個方差公式相等

$$\text{Var}(x) = E((X - \mu)^2) \quad (3.1)$$

$$= E(X^2 - 2X\mu + \mu^2) \quad (3.2)$$

$$= E(X^2) - 2\mu E(X) + \mu^2 \quad (3.3)$$

$$= E(X^2) - 2\mu^2 + \mu^2 \quad (3.4)$$

$$= E(X^2) - \mu^2 \quad (3.5)$$

$$= E(X^2) - E(X)^2 \quad (3.6)$$

### 3.1 方差的性質：

1.  $\text{Var}(X + b) = \text{Var}(X)$
2.  $\text{Var}(aX) = a^2 \text{Var}(X)$
3.  $\text{Var}(aX + b) = a^2 \text{Var}(X)$

## 第四章 伯努利分佈 Bernoulli distribution

伯努利分佈，說的就是一個簡單的二分變量  $(1, 0)$ ，它取 1 時的概率如果是  $\pi$ 。那麼我們可以計算這個分佈的期望值：

$$E(X) = \sum_x xP(X = x) \quad (4.1)$$

$$= 1 \times \pi + 0 \times (1 - \pi) \quad (4.2)$$

$$= \pi \quad (4.3)$$

由於  $x = x^2$ ，因為  $x = 0, 1$ ，所以  $E[X^2] = E[X]$ ，那麼方差為：

$$Var(X) = E[X^2] - E[X]^2 \quad (4.4)$$

$$= E[X] - E[X]^2 \quad (4.5)$$

$$= \pi - \pi^2 \quad (4.6)$$

$$= \pi(1 - \pi) \quad (4.7)$$

證明， $X, Y$  為互為獨立的隨機離散變量時，**a)**  $E(XY) = E(X)E(Y)$  ；  
**b)**  $Var(X + Y) = Var(X) + Var(Y)$

- a) 證明

$$E(XY) = \sum_x \sum_y xyP(X = x, Y = y) \quad (4.8)$$

$$\because X, Y \text{ are independent to each other} \quad (4.9)$$

$$\therefore = \sum_x \sum_y xyP(X = x)P(Y = y) \quad (4.10)$$

$$= \sum_x xP(X = x) \sum_y yP(Y = y) \quad (4.11)$$

$$= E(X)E(Y) \quad (4.12)$$

- b) 證明根據方差的定義：

$$Var(X + Y) = E((X + Y)^2) - E(X + Y)^2 \quad (4.13)$$

$$\text{Expand} \quad (4.14)$$

$$= E(X^2 + 2XY + Y^2) - (E(X) + E(Y))^2 \quad (4.15)$$

$$= E(X^2) + E(Y^2) + 2E(XY) \quad (4.16)$$

$$- E(X)^2 - E(Y)^2 - 2E(X)E(Y) \quad (4.17)$$

$$\text{We just showed } E(XY) = E(X)E(Y) \quad (4.18)$$

$$= E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 \quad (4.19)$$

$$= Var(X) + Var(Y) \quad (4.20)$$

## 第五章 二項分佈的概念

### Binomial distribution

二項分佈在醫學研究中至關重要，一組二項分佈的數據，指的通常是  $n$  次相互獨立的成功率為  $\pi$  的伯努利實驗<sup>1</sup> ( $n$  independent Bernoulli trials) 中成功的次數。

當  $X$  服從二項分佈，記為  $X \sim \text{binomial}(n, \pi)$  或  $X \sim \text{bin}(n, \pi)$ 。它的 (第  $x$  次實驗的) 概率被定義為：

$$P(X = x) = {}^n C_x \pi^x (1 - \pi)^{n-x} \quad (5.1)$$

$$= \binom{n}{x} \pi^x (1 - \pi)^{n-x} \quad (5.2)$$

$$\text{for } x = 0, 1, 2, \dots, n \quad (5.3)$$

### 5.1 二項分佈的期望和方差

- 期望  $E(X)$ 
  - 若  $X \sim \text{bin}(n, \pi)$ ，那麼  $X$  就是這一系列獨立伯努利實驗中成功的次數。
  - 用  $X_i, i = 1, \dots, n$  標記每個相互獨立的伯努利實驗。
  - 那麼我們可以知道  $X = \sum_{i=1}^n X_i$ 。

---

<sup>1</sup><https://winterwang.github.io/post/probability2-4/>

$$E(X) = E\left(\sum_{i=1}^n X_i\right) \quad (5.4)$$

$$= E(X_1 + X_2 + \cdots + X_n) \quad (5.5)$$

$$= E(X_1) + E(X_2) + \cdots + E(X_n) \quad (5.6)$$

$$= \sum_{i=1}^n E(X_i) \quad (5.7)$$

$$= \sum_{i=1}^n \pi \quad (5.8)$$

$$= n\pi \quad (5.9)$$

- 方差  $Var(X)$

$$Var(X) = Var\left(\sum_{i=1}^n X_i\right) \quad (5.10)$$

$$= Var(X_1 + X_2 + \cdots + X_n) \quad (5.11)$$

$$= Var(X_1) + Var(X_2) + \cdots + Var(X_n) \quad (5.12)$$

$$= \sum_{i=1}^n Var(X_i) \quad (5.13)$$

$$= n\pi(1 - \pi) \quad (5.14)$$

$$(5.15)$$

## 5.2 超幾何分佈 hypergeometric distribution

假設我們從總人數為  $N$  的人羣中，採集一個樣本  $n$ 。假如已知在總體人羣中 ( $N$ ) 有  $M$  人患有某種疾病。請問採集的樣本  $X = n$  中患有這種疾病的人，服從怎樣的分佈？

- 從人羣 ( $N$ ) 中取出樣本 ( $n$ )，有  ${}^N C_n$  種方法。
- 從患病人羣 ( $M$ ) 中取出患有該病的人 ( $x$ ) 有  ${}^M C_x$  種方法。
- 樣本中不患病的人 ( $n - x$ ) 被採樣的方法有  ${}^{N-M} C_{n-x}$  種。
- 採集一次  $n$  人作為樣本的概率都一樣。因此：

$$P(X = x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$



### 5.3 樂透中獎概率問題：

- 從數字 1 ~ 59 中選取 6 個任意號碼
- 開獎時從 59 個號碼球中隨機抽取 6 個
- 如果六個號碼全部猜中 (不分順序)，你可以成為百萬富翁。請問一次猜中全部 6 個號碼的概率是多少？

從 59 個號碼中隨機取出任意 6 個號碼的方法有  ${}^{59}C_6$  種。

$${}^{59}C_6 = \frac{59!}{6!(59-6)!} = 45,057,474$$

每次選取六個號碼做為一組的可能性相同，所以，你買了一組樂透號碼，能中獎的概率就是  $1/45,057,474 = 0.00000002219$ 。你還會再去買彩票麼？

#### 5.3.1 如果我只想中其中的 3 個號碼，概率有多大？

用超幾何分佈的概率公式：

$$P(X=3) = \frac{{}^6C_3 \times {}^{53}C_3}{{}^{59}C_6} \quad (5.16)$$

$$= 0.010 \quad (5.17)$$

你有 1% 的可能中獎。換句話說，如果中三個以上的數字算中獎的話，你買的彩票中獎的概率低於 1%。是不是覺得下次送錢給博彩公司的時候還不如跟我一起喝一杯咖啡划算？



## 第六章 泊松分佈 Poisson Distribution

- 當一個事件，在一段時間 ( $T$ ) 中可能發生的次數是  $\lambda$ 。那麼我們可以認為，經過時間  $T$ ，該時間發生的期望次數是  $E(X) = \lambda T$ 。
- 利用微分思想，將這段時間  $T$  等分成  $n$  個時間段，當  $n \rightarrow \infty$  直到每個微小的時間段內最多發生一次該事件。

那麼

- 每個微小的時間段，可以視為是一個伯努利實驗（有事件發生或者沒有）
- 那麼這整段時間  $T$  內發生的事件可以視為是一個二項分佈實驗。

令  $X$  = 一次事件發生時所經過的所有時間段。

- $X \sim \text{Bin}(n, \pi)$ ，其中  $n \rightarrow \infty$ ， $n$  為時間段。
- 在每個分割好的時間段內，事件發生的概率都是： $\pi = \frac{\lambda T}{n}$
- 期望  $\mu = \lambda T \Rightarrow \pi = \mu/n$
- 所以  $X$  的概率方程就是：

$$P(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x} \quad (6.1)$$

$$= \binom{n}{x} \left(\frac{\mu}{n}\right)^x \left(1 - \frac{\mu}{n}\right)^{n-x} \quad (6.2)$$

$$= \frac{n!}{x!(n-x)!} \left(\frac{\mu}{n}\right)^x \left(1 - \frac{\mu}{n}\right)^{n-x} \quad (6.3)$$

$$= \frac{n!}{n^x(n-x)!} \frac{\mu^x}{x!} \left(1 - \frac{\mu}{n}\right)^{n-x} \quad (6.4)$$

$$\text{when } n \rightarrow \infty \ x \ll n \quad (6.5)$$

$$\frac{n!}{n^x(n-x)!} = \frac{n(n-1)\dots(n-x+1)}{n^x} \rightarrow 1 \quad (6.6)$$

$$\left(1 - \frac{\mu}{n}\right)^{n-x} \approx \left(1 - \frac{\mu}{n}\right)^n \rightarrow e^{-\mu} \quad (6.7)$$

$$\text{the probability function of a Poisson distribution} \quad (6.8)$$

$$P(X = x) \rightarrow \frac{\mu^x}{x!} e^{-\mu} \quad (6.9)$$

當數據服從泊松分佈時，記為  $X \sim \text{Poisson}(\mu = \lambda T)$  or  $X \sim \text{Poi}(\mu)$

證明泊松分佈的參數特徵：

$$1. E(X) = \mu$$

$$E(X) = \sum_{x=0}^{\infty} xP(X=x) \quad (6.10)$$

$$= \sum_{x=0}^{\infty} x \frac{\mu^x}{x!} e^{-\mu} \quad (6.11)$$

$$= 0 + \sum_{x=1}^{\infty} x \frac{\mu^x}{x!} e^{-\mu} \quad (6.12)$$

$$= \sum_{x=1}^{\infty} \frac{\mu^x}{(x-1)!} e^{-\mu} \quad (6.13)$$

$$= \mu \sum_{x=1}^{\infty} \frac{\mu^{x-1}}{(x-1)!} e^{-\mu} \quad (6.14)$$

$$\text{replace } x \text{ with all } i = x - 1 \quad (6.15)$$

$$= \mu \sum_{i=0}^{\infty} \frac{\mu^i}{i!} e^{-\mu} \quad (6.16)$$

$$\text{notice that the right side } \sum_{i=0}^{\infty} \frac{\mu^i}{i!} e^{-\mu} = 1 \text{ is} \quad (6.17)$$

$$\text{the sum of all probability of a Poisson distribution} \quad (6.18)$$

$$= \mu \quad (6.19)$$

2.  $Var(x) = \mu$  爲了找到  $Var(X)$ , 我們用公式  $Var(X) = E(X^2) - E(X)^2$

我們需要找到  $E(X^2)$

$$E(X^2) = \sum_{x=0}^{\infty} x^2 \frac{\mu^x}{x!} e^{-\mu} \quad (6.20)$$

$$= \mu \sum_{x=1}^{\infty} x \frac{\mu^{x-1}}{(x-1)!} e^{-\mu} \quad (6.21)$$

$$\text{replace } x \text{ with all } i = x - 1 \quad (6.22)$$

$$= \mu \sum_{i=0}^{\infty} (i+1) \frac{\mu^i}{i!} e^{-\mu} \quad (6.23)$$

$$= \mu \left( \sum_{i=0}^{\infty} i \frac{\mu^i}{i!} e^{-\mu} + \sum_{i=0}^{\infty} \frac{\mu^i}{i!} e^{-\mu} \right) \quad (6.24)$$

$$= \mu(E(X) + 1) \quad (6.25)$$

$$= \mu^2 + \mu \quad (6.26)$$

$$Var(X) = E(X^2) - E(X)^2 \quad (6.27)$$

$$= \mu^2 + \mu - \mu^2 \quad (6.28)$$

$$= \mu \quad (6.29)$$

## 第七章 正態分佈

### 7.1 概率密度曲線 probability density function, PDF

- 一個隨機連續型變量  $X$  它的性質由一個對應的概率密度方程 (probability density function, PDF) 決定。
- 在給定的範圍區間內，如  $a \sim b, (a < b)$ ，它的概率滿足：

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

- 這個相關的方程，在  $a \sim b$  區間內的積分，就是這個連續變量在這個區間內取值的概率。

```
# R codes for drawing a standard normal distribution by using ggplot2
library(ggplot2)
p <- ggplot(data.frame(x=c(-3,3)), aes(x=x)) +
  stat_function(fun = dnorm)
p + annotate("text", x=2, y=0.3, parse=TRUE, label="frac(1, sqrt(2*pi)) *") +
  theme(plot.subtitle = element_text(vjust = 1),
        plot.caption = element_text(vjust = 1),
        axis.text.x = element_text(size = 12),
        axis.text.y = element_text(size = 12),
        plot.title = element_text(size = 10, face = "bold", hjust = 0.5),
        panel.background = element_rect(fill = "ivory")) +
```

```
labs(title = "Probability density functions \n for standard normal distribut
      x = NULL, y = NULL) +
stat_function(fun = dnorm,
              xlim = c(-1.3,0.4),
              geom = "area",fill="#00688B", alpha= 0.2)
```

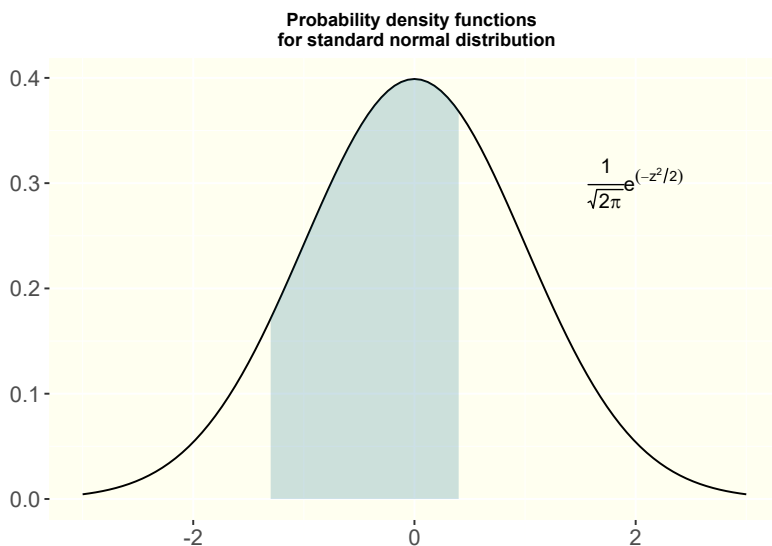


图 7.1: Probability Density Function of a Standard Normal Distribution

注意：整個方程的曲線下面積等於 1：

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

- 期望  $E(X) = \int_{-\infty}^{\infty} x f(x)dx$
- 方差  $Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$

## 7.2 正態分佈

如果一組數據服從正態分佈，我們通常用它的期望（或者叫平均值） $\mu$ ，和它的方差  $\sigma^2$ ，來描述這組數據。記為：

$$X \sim N(\mu, \sigma^2)$$



- 它的概率密度方程可以表述為：

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- $E(x) = \mu$
- $Var(x) = \sigma^2$

## 7.3 標準正態分佈

標準正態分佈的期望（或者均值）為 0，方差為 1

- 記為：  $Z \sim N(0, 1)$
- 它的概率密度方程表述為：

$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$$

- 它的累積分佈方程 (cumulative distribution function, CDF)，是將概率密度方程 (PDF) 積分以後獲得的方程。通常我們記為  $\Phi(z)$

再看一下標準正態分佈的概率密度方程曲線：

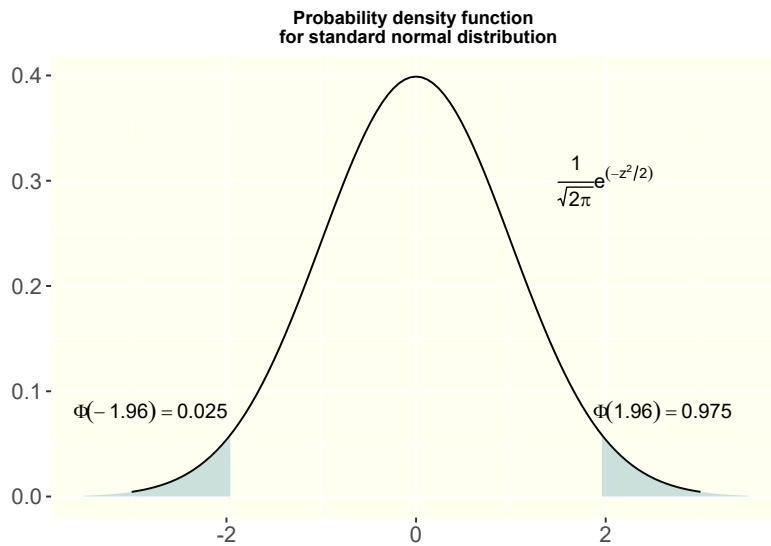


图 7.2: Probability Density function of a Standard Normal Distribution

- 95% 的曲線下面積在標準差 standard deviation  $-1.96 \sim 1.96$  之間的區域。
- 而且,  $\phi(-x) = 1 - \phi(x)$
- 任何一個正態分佈都可以通過下面的公式, 標準化成為標準正態分佈:

$$Z = \frac{X - \mu}{\sigma}$$

## 第八章 中心極限定理 the Central Limit Theorem

最近明顯可以感覺到課程的步驟開始加速。看我的課表：

10月		16		
		月曜日		
10:00	Probability y 5 LG80	Foundations of Medical...	Basic Epidemiology Student Central	
11:00			2B / Jo	
12:00	Foundations of Medical Statistics K-LG80 Bennett Room		Basic John S Theatre	
13:00			Statistical Modelling and Challenges	
14:00				
15:00	Inference 2 LG80 / K-365	Foundations of Me...	Inference 3 MS-42 K-365	