ryNoSmBk

ryCh 1 【第 1 章】

1.0.0.1 概率基礎

本章讓我們了解概率的基礎知識。我們將涵蓋以下主題，目的是介紹對課程剩餘部分有用的基本工具。

1.1 簡介並誘發動機示例

1.2 數學訓練營

1.3 實驗和概率空間

1.4 有限樣本空間

1.1 簡介並誘發動機示例

由於這是一門關於隨機性的課程，讓我們從一些充滿隨機性的日常現象的例子開始。

- 下次我拋硬幣會是正面還是反面？
- 明天會降多少雪？
- IBM 今年會盈利嗎？
- 我應該購買看漲期權還是看跌期權？
- 我可以贏二十一點嗎？
- 一種新藥的成本效益如何？
- 哪匹馬會贏得 肯塔基德比 (Kentucky Derby)？
- 我的飛機會準時到達目的地嗎？
- 政治候選人在演講中會撒謊多少次？
- 某項槍支管制法會挽救生命嗎？
- 考慮到 綠野仙踪 (The Land of Oz)（那是我工作的地方）目前的交通狀況，今天返回 堪薩斯 (Kansas) 的最佳路線是什麼？
- 如果電路正在經歷隨機的電氣波動，它的平均電阻是多少？[1]

另一方面，還有其他一些事情似乎根本沒有表現出任何隨機性，例如，

- 一條裝配線每小時生產 20 件產品。在 t 小時後，我們有 $20t$ 的物品。
- 伽利略從 h_0 高度從比薩斜塔投下一個球。t 秒後，它的高度為 $h(t) = h_0 - 16t^2$。
- 將 \$1000 存入一個連續複利的 2% 支票賬戶。在時間 t，它的價值正好是 $\$1000e^{0.02t}$。

或者他們可能會這樣做 - 大多數現實世界的過程至少涉及某種程度的隨機性，並且可能必須使用概率模型進行分析。讓我們仔細看看前面的例子。

- 如果裝配線出現隨機故障怎麼辦？這將如何影響生產能力？
- 如果伽利略在丟球時錯誤地推了一下球怎麼辦？
- 如果銀行違約怎麼辦？

世界上真的存在隨機性和不確定性嗎？實際上，有許多可觀察到的系統似乎是隨機的——也許我們對管理這些系統的基本規則的了解是不完整的；也許真正的參數是未知的、不可知的或無法測量的；也許有混沌效應或量子效應。在許多此類情況下，我們無法創建可接受的確定性模型，但我們仍然可以創建有用的概率模型。這裡有一些激勵我們前進的例子。

示例：生日問題。

有時，概率論的結果可能出乎意料。

一個班級（或任何小組）中的兩個人生日相同的概率是多少？

假設生日在一年中平均分佈 [2] 大多數人會驚訝地發現：

[1] 沒有像歐姆這樣的地方。

[2] 假設沒有怪人在 2 月 29 日出生

- 如果房間裡只有 23 人，那麼會有一次匹配的勝率高於 $50-50$。
- 如果有 50 人，概率大約是 97% ！

示例：大富翁遊戲。從長遠來看，您登陸 伊利諾伊大道 (Illinois Ave) 的概率最高。

示例：撲克。 從一副標準紙牌中挑選五張牌。然後，

- 你正好有 兩對 (two pairs) 的概率大約是 0.0475
- 滿屋 (full house) 的概率約為 0.00144 $。
- 同花 (flush) 的概率約為 0.00198.

示例：股票市場。 猴子隨機選擇股票應該能夠跑贏大多數市場分析師 [3]

示例： 一對夫婦有兩個孩子，其中至少一個是男孩。兩個都是男孩的概率是多少？

顯然，（等可能的）可能性是 GG、GB、BG 和 BB。消除 GG，因為我們知道至少有一個男孩。那麼兩個都是男孩的概率是 1/3。

示例：來自 Ask Marilyn [4] 假設您在舊電視遊戲節目 讓我們來場交易！ (Let’s Make a Deal)中，主持人 Monty Hall 向您展示了三個門：A、B 和 C。一扇門後面是一輛車；另外兩個門後面卻是羊。你選擇 A 門。蒙蒂打開 B 門，露出一隻羊。然後他給你一個機會決定是否「切換」到 C 門。你應該怎麼做？ 答：你「要切換」！

請繼續關注以了解原因！

示例：越南戰爭選秀彩票 - 它並不像您想像的那麼 “公平”！

示例：哪種軟性飲料最受歡迎？好吧，我們這些住在亞特蘭大的人都知道這個問題的答案！

示例：為什麼有些選舉民意調查結果如此錯誤？

示例：隨著籃球比賽的進行，他們如何實時更新獲勝概率？

示例：您如何在計算機上模擬隨機性，您可以使用它來做什麼？

示例：您如何判斷您的製造工廠生產的產品的質量是否開始下降？

[3] www.telegraph.co.uk/finance/personalfinance/investing/9971683/Monkeys-beat-the-stockmarket.html

[4] 這就是所謂的“蒙蒂霍爾”問題。讓我們從簡單的、一句話的概率和統計定義開始。

概率——一種描述系統隨機變化的方法。（我們將花費大約 50 $ %$ 的時間在這上面。）

統計學 - 應用數學的一個分支，它使用樣本數據得出關於樣本總體的一般結論。(50% 我們的時間。)

在開始討論概率之前，我們將從一個小訓練營開始回顧一些背景材料。

1.2 數學訓練營

我們將在 §1.2.2 中回顧 §1.2.1 中的一些基本集合論，然後在 $1.2.3 中回顧證明事物的藝術。包含此材料是為了使本書或多或少獨立。如果記憶猶新，讀者可以跳過本節。

1.2.1 集合的喜悅

定義：集合是可能以某種方式相關的不同對象的集合。集合的成員稱為元素。

這是您在小學時可能見過的一些標準符號。

- 集合的大寫字母（例如，A、X、Ω 等）。
- 集合元素的小寫字母 (a, x, ω, etc.)。
- $\in$ 表示集合成員關係，例如，$x \in X$ 表示x 是集合X 的一個元素。
- $\notin$ 用於非會員，例如，$x \notin X$。
- Ω 通常表示“通用”集合（即所有感興趣的集合）。
- $\emptyset$ 是空集（又名空集），即不包含任何元素的集合 - 根本沒有元素。

1.2.1.1 示例/符號：

- $A=\{1,2,3,4,5\}$，即從 1 到 5 的整數集合。集合中元素的順序無關緊要，所以你也可以寫成 $A=\{5,4,3,2,1\}$，這樣就可以了。
- $B=\{x \mid -1 \leq x \leq 1\}$，即所有$x$的集合，使得$x$在$-1$和$1$之間，包括。- $C=\left\{x \mid x^2-3x+2=0\right\}=\{1,2\}$.（兩者都可以。）
- $D=\{$ Justin Bieber 的好歌 $\}=\emptyset$.（他真的沒有！）
- \$E=2^{C} C的 $\equiv$ 冪集 $C=\{\emptyset,\{1\},\{2\}, C\}$ 的所有子集的 \$ 集。
- $\mathbb{R}$ 是所有實數的集合。注意 $\sqrt{-1} \notin \mathbb{R}$.
- $\mathbb{R}^2$ 是所有有序實數對的集合。
- $\mathbb{Q}$ 是所有有理數的集合。注意 $3.14159 \in \mathbb{Q}$，但是 $\pi \notin \mathbb{Q}$。

定義：如果集合S 的每個元素都是集合T 的一個元素，那麼S 稱為T 的子集，我們用$S \subseteq T$ 表示。當然，如果$T \subseteq S$也是這樣，那麼$S=T$

1.2.1.2 示例：

- $\{a, e, i, o, u\}$ 是字母表的一個子集。
- $\mathbb{Q} \subseteq \mathbb{R}$
- 對於任何集合 S 和全集 Ω，我們有 $\emptyset \subseteq S \subseteq S \subseteq \Omega$。因此，空集是任何其他集的子集；任何集合都是它自己的一個子集，也是Ω 的一個子集。
- 子集的傳遞性：如果 $X \subseteq Y$ 和 $Y \subseteq Z$，那麼 $X \subseteq Z$。

定義： 集合S 的基數是S 中元素的個數，用$|S|$ 表示。如果 $|S|<\infty$，則 S 是有限的。例如，$S=\{3,4\}$ 是有限的，因為 $|S|=2$

自然數集合 $\mathbb{N} \equiv\{0,1,2, \ldots\}$ 是可數無限的；無限基數 $|\mathbb{N}|$ 有時用 $\aleph_0$（“aleph-naught”）表示。與 $\mathbb{N}$ 的子集具有相同基數的集合稱為可數集合。可數集可以是有限的也可以是無限的。

任何定義區間的實數集合，例如 $\{x \mid 0 \leq x \leq 1\}$，都是不可數無限的；這個基數用 $\aleph_1$ 表示。

備註：關於無窮大的一些事實令人驚訝，例如： (i) 有理數集 $\mathbb{Q}$ 具有基數 $\aleph_0$ - 與整數的基數相同！(ii) $\aleph_1$ 是比 $\aleph_0$“更大的無窮大”；事實上，我們有時會寫成 $\aleph_1=2^{\aleph_0}$。(iii) 集合論之父是 Georg Cantor；他的工作在當時引起了極大的爭議，以至於他多年來一直患有嚴重的健康問題。

1.2.1.3 更多定義：

- 集合X 的補集[5] 是所有不在X 中的元素的集合，即$\bar{X} \equiv \{x \mid x \notin X\}$。

5“你是一個很好的組合，X！” - 集合 X 和 Y 的交集是兩個集合中所有元素的集合，即 $X \cap Y \equiv \{z \mid z \in X$ 和 $z \in Y\}$。如果$X \cap Y = \emptyset$（即沒有公共元素），那麼X 和Y 是不相交或互斥的集合。

- sets X 和 Y 的並集是包含在一個或兩個集合中的所有元素的集合，即 $X \cup Y \equiv \{z \mid z \in X$ 和 $ /$ 或 $z \in Y\}$。

- 集合 X 和 Y 的減法運算是 $XY \equiv X \cap \bar{Y}$，即當一個人從 X 中刪除任何也恰好在 Y中的東西時得到的集合。

- 兩組之間的對稱差異或 XOR（“異或”）是 X 或 Y 中的所有內容，但不是兩者都有，

$$
\begin{aligned}
X \Delta Y &\equiv (X - Y) \cup (Y - X) \\
&= (X \cap \bar{Y}) \cup (\bar{X} \cap Y) \\
&= (X \cup Y) - (X \cap Y)
\end{aligned}
$$

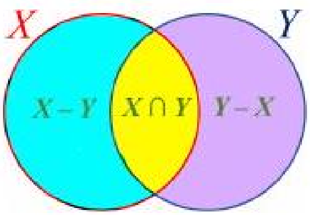
示例：考慮以下通用的有趣符號集，

$$\Omega = \{\not\chi, \infty, \backsim, \nabla, \Delta, \$, £, ¥\}$$

以及子集 $X =$ 類型的貨幣 $= \{\$$、£、¥$\}$ 和 $Y = \{\varnothing$、∇、Δ、$\$\}$。

- $\bar{X} = \{\not\chi, \infty, \backsim, \infty, \nabla, \Delta\}.$
- $X \cap Y = \{\$\}$
- $X \cup Y = \{\bigcirc, \nabla, \Delta, \$, £, ¥\}$。
- $XY = \{£, ¥\}$。
- $X\Delta Y = \{\bigcirc, \nabla, \Delta, £, ¥\}$。

備註：可以通過維恩圖來說明各種集合操作。下面是減號和相交操作的樣子。



一些涉及集合操作的定律：

- 補碼：$X \cup \bar{X} = \Omega, \quad X \cap \bar{X} = \emptyset, \quad$ 和 $\quad \bar{\bar{X}} = X$。
- 交換：$X \cap Y = Y \cap X, \quad$ 和 $X \cup Y = Y \cup X$。
- 聯想：$X \cap (Y \cap Z) = (X \cap Y) \cap Z = X \cap Y \cap Z$，$X \cup (Y \cup Z) = (X \cup Y) \cup Z = X \cup Y \cup Z$。- 分配式：$X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$，$X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$。
- 德摩根的：$\overline{X \cap Y} = \bar{X} \cup \bar{Y}$，和 $\overline{X \cup Y} = \bar{X} \cap \bar{Y}$。

證明：這些結果的證明可以使用維恩圖或其他基本方法進行。為了說明，我們將做 DeMorgan 的後半部分。

$$
\begin{aligned}
z \in \overline{X \cup Y} &\Leftrightarrow z \notin X \cup Y \\
&\Leftrightarrow z \notin \bar{X} \notin \bar{X} \text{ and } z \notin Y \\
&\Leftrightarrow z \in \bar{X} \text{ and } z \in \bar{Y} \\
&\Leftrightarrow z \in \bar{X} \cap \bar{Y}. \square
\end{aligned}
$$

示例：假設$\Omega = \{1, 2, \ldots, 10\}$，$X$是素數集合$(\leq 10)$，$Y = \{1, 2, 3, 4, 6\}$。讓我們驗證 DeMorgan 是否有效（我們將從末端開始並在中間相遇）：

1.2.2 微積分入門

在這個訓練營中，我們將從一些基礎知識開始，然後繼續討論衍生品、集成和其他感興趣的項目。

1.3 2,1 基礎知識

首先，假設 $f(x)$ 是一個函數，它將 x 的值從某個域 X 映射到某個範圍 Y。我們用簡寫 $f : X \to Y$ 來表示它。

示例：如果 $f(x) = x^2$，則函數將 x-值從實線 $\mathbb{R}$（域 X）到實線的非負部分$\mathbb{R}^+$(範圍 Y)。

定義：我們說 $f(x)$ 是一個連續函數，如果對於任何 x_0 和 $x \in X$，我們有 $\lim_{x \to x_0} f(x) = f(x_0)$，其中“lim”表示一個極限，假設所有 $x \in X$ 都存在 $f(x)$。

示例：函數 $f(x) = 3x^2$ 對於所有 x 都是連續的。函數 $f(x) = \lfloor x \rfloor$（向下舍入到最接近的整數，例如 $\lfloor 3.4 \rfloor = 3$) 在任何整數 x 處都有“跳躍”不連續性。

定義：函數 $f : X \to Y$ 的逆函數是（非正式地）“反向”映射 $g : Y \to X$，使得 $f(x) = y$ 當且僅當 $g(y) = x$，對於所有適當的

$$\begin{aligned}
\overline{X \cup Y} &= \overline{\{2,3,5,7\} \cup \{1,2,3,4,6\}} \\
&= \overline{\{1,2,3,4,5,6,7\}} \\
&= \{8,9,10\} \\
&= \{1,4,6,8,9,10\} \cap \{5,7,8,9,10\} \\
&= \overline{\{2,3,5,7\}} \cap \overline{\{1,2,3,4,6\}} \\
&= \bar{X} \cap \bar{Y}.
\end{aligned}$$

x 和 y。f 的逆函數通常寫為 f^{-1}，如果 $f(x)$ 是嚴格遞增或嚴格遞減函數，則特別有用。注意$f^{-1}(f(x)) = x$。

示例：如果 $f(x) = x^3$，那麼我們有 $f^{-1}(y) = y^{1/3}$。如果$f(x) = e^x$，則$f^{-1}(y) = \ell n(y)$。但遺憾的是，$f(x) = x^2$ 沒有唯一的逆。

1.3.1 區分

定義：如果 $f(x)$ 是連續的，那麼它是可微的（有導數）如果

$$\frac{d}{dx} f(x) \equiv f'(x) \equiv \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

存在並且對於任何給定的 x 都是明確定義的。將導數視為函數的斜率。

示例：一些著名的衍生物是：

$$\begin{aligned}
\left[x^k\right]' &= kx^{k-1} \\
[e^x]' &= e^x \\
[\sin(x)]' &= \cos(x) \\
[\cos(x)]' &= -\sin(x) \\
[\ln(x)]' &= \frac{1}{x} \\
[\arctan(x)]' &= \frac{1}{1+x^2}
\end{aligned}$$

定理：導數的一些眾所周知的性質是：

$$\begin{aligned}
[af(x) + b]' &= af'(x) \\
[f(x) + g(x)]' &= f'(x) + g'(x) \\
[f(x)g(x)]' &= f'(x)g(x) + f(x)g'(x) \quad \text{(product rule)} \\
\left[\frac{f(x)}{g(x)}\right]' &= \frac{g(x)f'(x) - f(x)g'(x)}{g^2(x)} \quad \text{(quotient rule)} \\
[f(g(x))]' &= f'(g(x))g'(x) \quad \text{(chain rule for compositions).}
\end{aligned}$$

示例：假設 $f(x) = x^2$，並且 $g(x) = \ln(x)$。然後

$$
\begin{aligned}
[f(x)g(x)]' &= f'(x)g(x) + f(x)g'(x) \\
&= 2x\ell\,\mathrm{n}(x) + x^2(1/x) \\
&= 2x\ell\,\mathrm{nn}(x) + x \\
\left[\frac{f(x)}{g(x)}\right]' &= \frac{g(x)f'(x) - f(x)g'(x)}{g^2(x)} \\
&= \frac{[\ell\,\mathrm{n}(x)]2x - x^2(1/x)}{\ln^2(x)} \\
&= \frac{2x\ell n(x) - x}{\ln^2(x)}, \text{ and}
\end{aligned}
$$

$$
[f(g(x))]' = \left[(g(x))^2\right]' = 2g(x)g'(x) = \frac{2\ell\ln(x)}{x}
$$

備註：二階導數 $f''(x) \equiv \frac{d}{dx}f'(x)$ 是"斜率的斜率"。如果 $f(x)$ 是"位置"，那麼 $f'(x)$ 可以看作是"速度"，而 $f''(x)$ 可以看作是"加速度"。 " $f(x)$ 的最小值或最大值只有在 $f(x)$ 的斜率為 0 時才會出現，即只有當 $f'(x) = 0$ 時，比如說在臨界點$x = x_0$。例外：還要檢查您感興趣的區間的端點。那麼如果 $f''(x_0) < 0$，你得到一個最大值；如果 $f''(x_0) > 0$，你得到一個最小值；如果 $f''(x_0) = 0$，你得到一個拐點。

示例：找到使 $f(x) = e^{2x} + e^{-x}$ 最小化的 x 的值。只有當 $f'(x) = 2e^{2x} - e^{-x} = 0$ 時才會出現最小值。經過一點代數，我們發現這發生在 $x_0 = -(1/3)\ln(2) \doteq -0.231$。也很容易證明對於所有 x，$f''(x) > 0$，所以 x_0 產生最小值。

1.3.2 整合

定義：具有導數 $f(x)$ 的函數 $F(x)$ 稱為反導數（或不定積分）。它由 $F(x) = \int f(x)dx$ 表示。

微積分基本定理：如果 $f(x)$ 是連續的，則 $x \in [a, b]$ 的曲線下面積由定積分]^6 表示和給出

$$
\int_a^b f(x)dx \equiv F(x)\Big|_a^b \equiv F(b) - F(a)
$$

示例：一些著名的不定積分是：

$$
\begin{gathered}
\int x^k dx = \frac{x^{k+1}}{k+1} + C, \text{ for } k \neq -1 \\
\int \frac{dx}{x} = \ln|x| + C \\
\int e^x dx = e^x + C \\
\int \cos(x)dx = \sin(x) + C \\
\int \frac{dx}{1+x^2} = \arctan(x) + C
\end{gathered}
$$

其中 C 是一個任意常數。

示例：很容易看出

$$
\int \frac{d\,\text{cabin}}{\text{cabin}} = \ell n \mid \text{ cabin } \mid + C = \text{ houseboat.}
$$

6"我真的是不可或缺的！" 定理：定積分的一些眾所周知的性質是：

$$
\begin{gathered}
\int_a^a f(x)dx = 0 \\
\int_a^b f(x)dx = -\int_b^a f(x)dx \\
\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx
\end{gathered}
$$

定理：一般積分的其他一些性質是：

$$\int [f(x)+g(x)]dx = \int f(x)dx + \int g(x)dx$$
$$\int f(x)g'(x)dx = f(x)g(x) - \int g(x)f'(x)dx \quad \text{(integration by parts)}$$
$$\int f(g(x))g'(x)dx = \int f(u)du \quad \text{(substitution rule).}$$

示例：為了演示定積分上的分部積分，令 $f(x)=x$ 和 $g'(x)=e^{2x}$，使得 $g(x)=e^{2x}/2$。然後

1.3.3 系列

定義：任意階 k 的導數可以寫成 $f^{(k)}(x)$ 或 $\frac{d^k}{dx^k}f(x)$。按照慣例，$f^{(0)}(x)=f(x)$。

$f(x)$ 關於點 a 的泰勒級數展開式由下式給出

$$f(x)=\sum_{k=0}^{\infty}\frac{f^{(k)}(a)(x-a)^k}{k!}.$$

麥克勞林級數只是泰勒在 $a=0$ 周圍展開的。

$$\begin{aligned}\int_0^1 xe^{2x}dx &= \int_0^1 f(x)g'(x)dx \\ &= f(x)g(x)|_0^1 - \int_0^1 g(x)f'(x)dx \quad (\text{ parts }) \\ &= \left.\frac{xe^{2x}}{2}\right|_0^1 - \int_0^1 \frac{e^{2x}}{2}dx \\ &= \frac{e^2}{2} - \left.\frac{e^{2x}}{4}\right|_0^1 \\ &= \frac{e^2+1}{4}.\end{aligned}$$

示例：這裡有一些著名的麥克勞林系列：

$$\sin(x)=\sum_{k=0}^{\infty}\frac{(-1)^k x^{2k+1}}{(2k+1)!}$$
$$\cos(x)=\sum_{k=0}^{\infty}\frac{(-1)^k x^{2k}}{(2k)!}$$
$$e^x=\sum_{k=0}^{\infty}\frac{x^k}{k!}$$

示例：在我們討論的時候，這裡有一些你應該知道的雜項：

$$\sum_{k=1}^{n}k=\frac{n(n+1)}{2}$$
$$\sum_{k=1}^{n}k^2=\frac{n(n+1)(2n+1)}{6}$$
$$\sum_{k=0}^{\infty}p^k=\frac{1}{1-p}(\text{ for } -1<p<1).$$

1.3.4 去醫院

定理：有時，我們在取 $0/0$ 或 ∞/∞ 形式的不確定比率時會遇到麻煩。在這種情況下，L'Hôpital 規則很有用：

如果 $\lim_{x\to a} f(x)$ 和 $\lim_{x\to a} g(x)$ 都趨於 0 或都趨於 ∞，那麼

$$\lim_{x\to a}\frac{f(x)}{g(x)}=\lim_{x\to a}\frac{f'(x)}{g'(x)}.$$

示例：L’Hôpital 表明

$$\lim_{x\to 0}\frac{\sin(x)}{x}=\lim_{x\to 0}\frac{\cos(x)}{1}=1$$

1.3.5 雙重積分

我們將有機會計算幾個雙積分。雖然我們通常的（單）積分得到曲線下的面積，但雙積分錶示三維函數下的體積。

示例：$f(x,y)=8xy$ 下的交易量，區域 $0<x<y<1$，由下式給出

$$\int_0^1\int_0^y f(x,y)dxdy=\int_0^1\int_0^y 8xydxdy=\int_0^1 4y^3dy=1$$

我們通常可以交換積分順序以獲得相同的答案（稱為“Fubini 魔術”的操作）：

$$\int_0^1\int_x^1 8xydydx=\int_0^1 4x\left(1-x^2\right)dx=1$$

通過從 (x,y) 平面到極坐標 (r,θ) 的轉換可以更容易地解決一些雙重積分問題，其中我們設置 $x=r\cos(\theta)$和 $y=r\sin(\theta)$。那麼可以證明

$$\iint_A f(x,y)dxdy=\iint_B f(r\cos(\theta),r\sin(\theta))rdrd\theta$$

對於適當的集成區域 A 和 B.

示例：我們可以使用極坐標和 Fubini 來計算具有挑戰性的量，

$$\begin{aligned}\left(\int_{\mathbb{R}}e^{-x^2/2}dx\right)^2&=\int_{\mathbb{R}}\int_{\mathbb{R}}e^{-\left(x^2+y^2\right)/2}dxdy\\&=\int_0^{2\pi}\int_0^{\infty}e^{-r^2\left(\cos^2(\theta)+\sin^2(\theta)\right)/2}rdrd\theta\\&=\int_0^{\infty}re^{-r^2/2}\int_0^{2\pi}d\theta dr=2\pi\end{aligned}$$

1.3.6 由零保存！（如何求根）

假設我們想為根 $r^\star$ 求解某個方程 $g(r)=0$，其中 $g(r)$ 是一個表現良好的連續函數。我們可以用：

- 試錯法或某種線性搜索 - 這是為失敗者準備的！
- 平分法 - 相當快！=
- 牛頓法 - 真快！（一世）

首先，一個有用的定理...

中值定理 (IVT)：如果 $g(\ell)g(u)<0$，則存在零（根）$r^\star\in[\ell,u]$。換句話說，如果（i）$g(\ell)<0$ 並且 $g(u)>0$，或者 (ii) $g(\ell)>0$ 並且 $g(u)<0 $，然後 $g(r)$ 在 ℓ 和 u 之間的某處穿過 0。

Bisection 使用 IVT 通過順序二等分來磨練零：

- 初始化：求上下界$\ell_0<u_0$，使得$g\left(\ell_0\right)g\left(u_0\right)<0$. 那麼 IVT 意味著 $r^\star\in\left[\ell_0,u_0\right]$。
- 對於 $j=0,1,2,\ldots,$
- 令當前區間的中點為 $r_j\leftarrow\left(\ell_j+u_j\right)/2$。- 如果 $g\left(r_j\right)$ 足夠接近 0 ，或者區間寬度 $u_j-\ell_j$ 足夠小，或者超出了您的迭代預算，則設置$r^\star\leftarrow r_j$ 和停止。

- 如果 $g\left(r_j\right)$ 的符號與 $g\left(\ell_j\right)$ 的符號匹配，這意味著 $r^\star \in \left[r_j, u_j\right]$; 所以設置 $\ell_{j+1} \leftarrow r_j$ 和 $u_{j+1} \leftarrow u_j$。否則，$r^\star \in \left[\ell_j, r_j\right]$; 所以設置 $\ell_{j+1} \leftarrow \ell_j$ 和 $u_{j+1} \leftarrow r_j$

該算法的每次迭代都將搜索區域切成兩半，因此很快收斂到 $r^\star$。

示例：通過求解 $g(x)=x^2-2=0$，使用二分法求 $\sqrt{2}$。為了初始化二分算法，我們注意到 $g(1)=-1$ 和 $g(2)=2$。所以在 $[1,2]$ 中有一個根源，只是很想找到！

您可以看到 r_j 似乎正在收斂到 $r^\star=\sqrt{2} \doteq 1.4142$。

牛頓法。該技術使用有關函數 g 及其導數 g' 的信息，以便更有效地將搜索指向正確的方向。因此，它通常比二分法快得多。這是一個合理的實現。

1. 初始化 r_0 作為對根的初步猜測。設置 $j \leftarrow 0$。
2. 更新$r_{j+1} \leftarrow r_j-g\left(r_j\right) / g'\left(r_j\right)$。
3. 如果 $\left|g\left(r_{j+1}\right)\right|$ 或 $\left|r_{j+1}-r_j\right|$ 或者你的預算適當小，然後停止並設置 $r^\star \leftarrow r_{j+1}$。否則，設 $j \leftarrow j+1$ 並返回步驟 2

示例：使用牛頓通過求解 $g(x)=x^2-2=0$ 來找到 $\sqrt{2}$。為此，請注意

$$r_{j+1} \leftarrow r_j-\frac{g\left(r_j\right)}{g'\left(r_j\right)}=r_{j+1} \leftarrow r_j-\frac{r_j^2-2}{2 r_j}=\frac{r_j^2+2}{2 r_j}$$

如果$r_0=1$，那麼我們發現$r_1=3/2, r_2=17/12 \doteq 1.4167, r_3=577/408 \doteq 1.4142, \ldots$.那是快速收斂！

1.3.7 證明事物

我們在本書中做了一些證明（其中一些你可以隨意跳過）。當前小節列出了您在旅行中可能遇到的有用證明技術的一小部分。如果你喜歡這個話題，一個很棒的參考是 Polya 的經典文本 [5]

1.3.7.1 §1.2.31 很明顯

有時事情是如此明顯，以至於假設它們是真實的可能是可以的——所以你根本不需要“證明”它們！比如$1+1=2$，對於直角三角形$a^2+b^2=c^2$，$x^2 \geq 0$，賈斯汀比伯很煩，等等。大多數時候，假設一切都很好，這樣的結果很好，我們的生活不受干擾，這是完全可以的。

然而，你應該時刻注意尋找麻煩。畢竟，以 2 為底的 $1+1=10$（另請參閱 Enderton [2] 的第 5 章，了解如何嚴格證明 $1+1=2$）；$a^2+b^2 \neq c^2$ 在球體上；$i^2=-1<0,$ 其中 $i \equiv \sqrt{-1};$ 還有賈斯汀比伯——嗯，實際上，他很煩人！更重要的是，我們將在本章後面遇到一些非常不直觀的示例，因此您可能需要注意自己的步驟。

最重要的是，你真的不應該擔心大多數顯而易見的事情——儘管有點偏執也沒有什麼壞處，因為也許他們是來抓你的！

1.3.7.2 $1.2.3，恐嚇（又名，“你最好相信我！”）

我們都聽過人們告訴我們“這是對的，因為我說這是對的！” 這也發生在數學中。比如，你應該從小學就知道以下這些，你不敢讓我證明它！

1.3.7.3 二項式定理：

$$(x+y)^n=\sum_{k=0}^n\binom{n}{k} x^k y^{n-k},$$

where the binomial coefficient is $\binom{n}{k} \equiv \frac{n!}{k!(n-k)!}$, for $k=1,2, \ldots, n$.

我們可能每隔一段時間就會潛入這種可疑的證明技術，但我們會始終警告您，以便您可以以正確的心態繼續進行。僅僅因為我們很棒並不意味著您必須始終相信我們 - 但我們會盡量不讓您誤入歧途！

1.3.7.4 $1.2.3 打滑證明

一些有問題的證據不會上升到恐嚇的程度，但它們會讓你走上一個滑坡。例如，您可能會看到諸如“無需詳述”或“在某些條件下”或“相信我……”之類的短語，我們有時會因為這種證明性而感到內疚，通常是為了說明清楚。例如，為了避免堵塞原本乾淨的證明，眾所周知，我們會進行非法交換，例如 $\frac{d}{dt}\sum_{k=1}^{\infty} f_n(t) = \sum_{k=1}^{\infty} \frac{d}{dt} f_n(t)$，由於無限和，這不一定是真的。好消息是，我們會在採取此類行動時通知您。

1.3.7.5 $1.2.3.4 完整枚舉

到目前為止，我們已經討論了一些影子證明技術。現在我們終於開始評論更嚴格的方法論了。

您有時可以簡單地通過列舉每一種可能性並證明結果對每一種可能性都是正確的來證明一個結果。這種方法可能適用於較小的問題，但隨著問題規模的擴大，很快就會變得不可能。

示例：正好有八個小於 20 的素數。為了證明這一點，我們可以繁瑣地寫出相關的素數分解。

示例：旅行推銷員問題挑戰您找到通過集合 $\{1, 2, \ldots, n\}$ 中每個城市的最短距離路線，只要您在同一個城市，並且不要多次訪問任何城市（除了您的起點）。假設 i 到 j 城市的距離與 j 到 i 的距離相同，那麼很容易顯示（尤其是在你完成 $1.5.2 之後，你必須檢查$(n-1)!/2$ 行程的距離。人們可能會想列舉所有的可能性，但這變得非常乏味，即使對於適度的 n 也是如此（參見 Cook [1].

1.3.7.6 $1.2.3.5 磨出來

完全枚舉的家庭親屬通常是在涉及微積分和代數等載體的大量血液、汗水和眼淚之後才得出答案。這種方法的例子在正文中到處都是榮譽結果，例如，我們在 §1.3.4 中對一般包容-排除原則的證明

在這種熟練的方法中確實沒有什麼可羞恥的，儘管它可能會找到節省努力的優雅捷徑。

1.3.7.7 $1.2.3.6 更大的特殊情況

經常發生的情況是，您已經擁有所需的工具，就像綠野仙踪中的多蘿西一樣！事實上，通過發現已知結果的正確特殊情況，您可以節省大量工作。

示例：令人驚訝的事實：

$$\sum_{k=0}^{n}\binom{n}{k} = 2^n$$

證明：由二項式定理 1.1，

$$2^n = (1+1)^n = \sum_{k=0}^{n}\binom{n}{k} 1^k 1^{n-k}.$$

1.3.7.8 § 1.2.3.7 反例

為了使陳述為真，它必須始終為真。如果您只能找到一種情況，那麼當該陳述不成立時，它就是錯誤的。反例是解決真/假參數的好方法。一旦你找到一個，你可以宣布這個陳述是假的，然後走開。當然，反例不是證明（恰恰相反），但如果您試圖找到錯誤結果的證明，它會為您節省很多精力。

示例：我們可以反駁賈斯汀比伯的所有歌曲都很棒的瘋狂說法，只需找到一首糟糕的歌曲……這很容易做到。 $\not{x}\times$

示例：如果 p 是質數，那麼 $2^p - 1$ 也是質數的說法怎麼樣？從表面上看，這看起來很合理。畢竟，$2^2 - 1 = 3$ 是素數，$2^3 - 1 = 7, 2^5 - 1 = 31$ 和 $2^7 - 1 = 127$也是素數。但是，遺憾的是，$2^{11} - 1 = 2047 = 23 \cdot 89$ 反駁了我的說法。岡×

另一方面，事實證明，已知最大的素數（截至 2018 年底）確實具有 $2^p - 1$ 的形式，其中 $p = 282,589,933$。所以現在你可以打動你的朋友了！

1.3.7.9 $1.2.3, 8$ 矛盾

矛盾證明是您駕駛室中的絕佳工具。假設您想證明索賠 A。矛盾要求您在 A 是錯誤的（不正確的）假設下繼續進行。如果該假設導致根本矛盾，那麼 A 為假的假設一定是不正確的；因此，唯一剩下的選擇是 A 必須為真！讓我們用幾個

著名的例子來說明這個巧妙的技巧。

示例：有無數個素數。為了證明這一點，假設只有有限個質數 n，例如 $p_1, p_2, \ldots, p_n$。現在讓我們考慮數字 $P = 1 + \prod_{k=1}^{n} p_k$，即我們所有 n 素數的乘積加一。請注意，$p_1, p_2, \ldots, p_n$ 都不是 P 的因數（因為我們已將 $+1$ 添加到產品中）。因此，P 必須是素數——但這將是我們的 $(n+1)^{\text{st}}$ 素數，這與只有 n 素數的假設相矛盾。所以有無窮多個素數。$\not\times$

示例：$\sqrt{2}$ 是非理性的。為了證明這一點，假設它是合理的。這意味著（根據有理數的定義）存在沒有公因數的整數 p 和 q，例如 $\sqrt{2} = p/q$，或者等價地 $p^2 = 2q^2$。反過來，這意味著 p^2 是偶數；這意味著 p 必須是偶數。因此，p^2 實際上可以被 4 整除，因此我們可以為某個整數 k 寫成 $p^2 = 4k$。因此，$4k = 2q^2$，因此 $q^2 = 2k$。現在這意味著 q^2 然後 q 是偶數。所以我們現在有 p 和 q 偶數。但這與 p 和 q 沒有公因數的要求相矛盾！所以我們假設 $\sqrt{2}$ 是理性的是錯誤的。所以 $\sqrt{2}$ 是不合理的。$\times\times$

1.3.7.10 &1.2.3.9 歸納

當我們想要證明所有整數 $n \geq 1$ 的形式為 $\mathcal{S}(n)$ 的陳述時，歸納法是一種強大的技術。這個想法是

1. 建立一個“基本情況”的結果，比如 $\mathcal{S}(1)$。
2. 假設 $\mathcal{S}(n)$ 為真。
3. 證明$\mathcal{S}(n)$ 的真值蘊含$\mathcal{S}(n+1)$ 的真值。

把這些步驟放在一起，我們將有

$$\mathcal{S}(0) \Rightarrow \mathcal{S}(1) \Rightarrow \mathcal{S}(2) \Rightarrow \cdots \Rightarrow \mathcal{S}(n) \Rightarrow \mathcal{S}(n+1) \Rightarrow \mathcal{S}(n+2) \Rightarrow \cdots,$$

即，$\mathcal{S}(n)$ 對所有 $n \geq 1$ 都是正確的。幾個例子將說明這是如何工作的。

示例：我們將使用歸納來證明 $1.2.24 的著名代數結果 $_{k=1}^{n} k^{2}=n(n+1)(2 n+1) / 6 $。

基本情況 $(n=1)$：$\sum_{k=1}^{1} k^2 = 1 = 1(1+1)(2(1)+1)/6$。

歸納步驟：假設 $\sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6$ 為真。

現在是“ $n+1$ ”步驟：

$$\begin{aligned}\sum_{k=1}^{n+1} k^2 &= \sum_{k=1}^{n} k^2 + (n+1)^2 \\ &= \frac{n(n+1)(2n+1)}{6} + (n+1)^2 \quad \text{(inductive step)} \\ &= \frac{(n+1)[n(2n+1)+6(n+1)]}{6} \\ &= \frac{(n+1)((n+1)+1)(2(n+1)+1)}{6}.\end{aligned}$$

示例：這是通過歸納二項式定理 1.1) 的證明。基本情況 $(n=0)$ ：注意到 $0! = 1$，我們有

$$\sum_{k=0}^{0} \binom{0}{k} x^k y^{0-k} = \binom{0}{0} x^0 y^{0-0} = 1 = (x+y)^0.$$

歸納步驟：假設等式 1.1 為真。

當我們證明“$n+1$”的情況時，我們就完成了，即，

$$(x+y)^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} x^k y^{n+1-k}.$$

在我們跳入火中之前，請注意

$$\begin{aligned}\binom{n+1}{k} &= \frac{(n+1)!}{k!(n+1-k)!} \\ &= \frac{n!}{(k-1)!(n-k)!}\left[\frac{n!}{k(n+1-k)}\right] \\ &= \frac{n+1}{(k-1)!(n-k)!}\left[\frac{n!}{n+1-k}+\frac{1}{k}\right] \\ &= \frac{n!}{(k-1)!(n+1-k)!}+\frac{n!(n-k)!}{k}\Bigg] \\ &= \binom{n}{k-1}+\binom{n}{k}\end{aligned}$$

Since we need $k \geq 1$ for the $\binom{n}{k-1}$ term and $k \leq n$ for the $\binom{n}{k}$ term, we have

$$\begin{aligned}\sum_{k=0}^{n+1}\binom{n+1}{k} x^k y^{n+1-k} &= \sum_{k=1}^{n+1}\binom{n}{k-1} x^k y^{n+1-k}+\sum_{k=0}^{n}\binom{n}{k} x^k y^{n+1-k} \\ &= x\sum_{k=0}^{n}\binom{n}{k} x^k y^{n-k}+y\sum_{k=0}^{n+1}\binom{n}{k} x^k y^{n-k} \\ &= x(x+y)^n+y(x+y)^n=(x+y)^{n+1}\end{aligned}$$

1.4 實驗和概率空間

在本節中，我們最終將開始討論隨機實驗，我們如何表示這些實驗的可能結果，然後是一些涉及與結果相關的概率的問題。

首先，這裡有一些可以進行的隨機實驗的例子。

- 擲硬幣並觀察結果：正面 (H) 或反面 (T)。
- 擲硬幣兩次，觀察 H 和 T 的順序。
- 擲一個六面骰子並觀察結果。
- 隨機大學生的 GPA 是多少？
- 燈泡能用多久？
- 詢問 10 個人更喜歡可樂還是百事可樂。

為了分析此類實驗，我們需要一種通用語言來描述"玩家"。實驗的概率空間的概念符合要求，我們在這裡談到組成樣本空間、事件和概率函數。

定義：與實驗相關的概率空間是由以下組件組成的三元組。1. 一個樣本空間，它是實驗所有可能結果的集合。它通常用 S 或 Ω 表示。

2. 一組 $\mathcal{F}$ 事件，其中每個事件本身都是 Ω 的子集。我們可以非正式地將 $\mathcal{F}$ 視為 Ω 的所有子集的集合（即 Ω 的冪集）；見問題 2]

3. 一個概率函數 $\mathrm{P}: \mathcal{F} \rightarrow [0,1]$ 將概率分配給事件。

讓我們更詳細地看一下這些組件中的每一個。

1.4.1 樣本空間

示例：以下是上述每個實驗的樣本空間。

- 擲硬幣：$\Omega=\{H, T\}$。
- 擲硬幣兩次：$\Omega=\{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$。

- 擲骰子：$\Omega = \{1, 2, 3, 4, 5, 6\}$。
- 大學生 GPA：$\Omega = \{0, 0.1, 0.2, \ldots, 3.9, 4.0\}$。
- 燈泡壽命：$\Omega = \{x \mid x \geq 0\}$。
- 喜歡可樂勝過百事可樂：$\Omega = \{0, 1, 2, \ldots, 10\}$。

備註：樣本空間不必是唯一定義的——它取決於我們感興趣的內容。例如，我們擲硬幣兩次的實驗的替代樣本空間可以是 $\Omega' = \{0, 1, 2\}$，可以解釋為我們觀察到的 H 的數量。

1.4.2 事件

事件只是一組可能的結果。因此，

- 樣本空間Ω 的任何子集都是一個事件。
- 空集 $\emptyset$ 是 Ω 的一個事件（“沒有觀察到任何可能的實驗結果”）。
- 樣本空間 Ω 本身就是一個事件（“來自樣本空間的事情發生了”）。
- 如果$A \subseteq \Omega$ 是一個事件，那麼$\bar{A}$ 是互補（相反）事件（“A 沒有發生”）。
- 如果A 和B 是事件，那麼$A \cup B$ 是一個事件（“A 或B 或兩者都發生”）。
- 如果A 和B 是事件，那麼$A \cap B$ 是一個事件（“A 和B 都發生”）。示例：投擲一個八面的龍與地下城死亡。$\Omega = \{1, 2, \ldots, 8\}$。如果 A 是“出現奇數”事件，那麼 $A = \{1, 3, 5, 7\}$，即擲骰子時，我們得到 1 或 3 或 5 或 7 。此外，$\bar{A} = \{2, 4, 6, 8\}$，即偶數發生的互補事件。

例：擲三枚硬幣。

$$\begin{aligned} A &= \text{"exactly one T was observed"} = \{\mathrm{HHT}, \mathrm{HTH}, \mathrm{THH}\} \\ B &= \text{"no T's observed"} = \{\mathrm{HHH}\} \\ C &= \text{"first coin is H}' = \{\mathrm{HHH}, \mathrm{HHT}, \mathrm{HTH}, \mathrm{HTT}\} \end{aligned}$$

然後，

$$\begin{aligned} A \cup B &= \text{"at most one T observed"} = \{\mathrm{HHT}, \mathrm{HTH}, \mathrm{THH}, \mathrm{HHH}\} \\ A \cap C &= \{\mathrm{HHT}, \mathrm{HTH}\}. \quad \square \end{aligned}$$

1.4.3 什麼是概率？

我們都對概率是什麼有一個直觀的認識。在本小節中，我們將嘗試在一定程度上形式化這個概念。

我們可以將概率視為將事件 A 從樣本空間 Ω 映射到區間 $[0, 1]$ 的函數。（顯然，一個事件的概率不能小於 0 或大於 1。）A 發生的概率用 $\mathrm{P}(A)$ 表示。

示例：如果我們拋一枚公平的硬幣，那麼當然是 $\Omega = \{H, T\}$。H 出現的概率是多少（即事件 $A = \{\mathrm{H}\}$ 發生）？我們都知道答案是 $\mathrm{P}(\{\mathrm{H}\}) = \mathrm{P}(\mathrm{H}) = 1/2$（當意思很明顯）。

這是什麼意思？

概率論的觀點認為，如果實驗要重複 n 次，其中 n 非常大，那麼我們預計大約 $1/2$ 的投擲是 H：

$$\frac{\text{Total number of H's out of } n \text{ tosses}}{n} \doteq 1/2.$$

示例：擲一個公平的骰子。$\Omega = \{1, 2, 3, 4, 5, 6\}$，其中每個單獨結果的概率為 $1/6$。那麼 $\mathrm{P}(1$ 或 $2) = 1/3$。

更正式的定義（例如，參見 Meyer 4]）：一般事件 $A \subseteq \Omega$ 的概率是一個遵循以下公理的函數：

1. $0 \leq \mathrm{P}(A) \leq 1$（概率總是在 0 和 1 之間）。

2. $\mathrm{P}(\Omega)=1$（來自 Ω 的某些結果的概率為 1 ）。

示例：擲骰子。$\mathrm{P}(\Omega)=\mathrm{P}(1,2,3,4,5,6)=1$。(3) 如果$A$和$B$是不相交事件，即$A\cap B=\emptyset$，則$\mathrm{P}(A\cup B)=\mathrm{P}(A)+\mathrm{P}(B)$。

示例：P(1 或 2)$=\mathrm{P}(1)+\mathrm{P}(2)=1/6+1/6=1/3$

備註：在註意到 $\Omega=A\cup\bar{A}$ 之後，這個公理立即暗示互補概率 $\mathrm{P}(\bar{A})=1-\mathrm{P}(A)$。

示例：明天會下雨的概率是 1 減去不會下雨的概率。

備註：特別是，Ω 沒有結果的概率是（直觀地）$\mathrm{P}(\emptyset)=1-\mathrm{P}(\Omega)=0$。

備註/示例：但反過來是錯誤的：$\mathrm{P}(A)=0$ 並不意味著 $A=\emptyset$。作為反例，選擇一個介於 0 和 1 之間的隨機數。稍後，我們將展示為什麼任何特定結果的概率實際上都是 0 ！××

4. 假設$A_1,A_2,\ldots$是一系列不相交的事件（即$A_i\cap A_j=\emptyset$，對於$i\neq j$）。然後

$$\mathrm{P}\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}\mathrm{P}\left(A_i\right)$$

示例：擲一個公平的骰子，直到第一次出現 3。我們定義不相交事件$A_i=$“3第一次出現在拋擲i”上，對於$i=1,2,\ldots$，即，

$$A_1=\{3\},A_2=\{\bar{3}3\},A_3=\{\overline{3}\overline{3}3\},\ldots$$

那麼公理 (2) 和 (4) 意味著

$$1=\mathrm{P}(\Omega)=\mathrm{P}\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}\mathrm{P}\left(A_i\right)$$

稍後，我們將了解到 $\mathrm{P}\left(A_i\right)=5^{i-1}/6^i, i=1,2,\ldots$

備註：A_1、A_2、$\ldots$、A_n 的有限版本作為特例如下。

這些公理將使我們能夠構建討論概率論所需的所有技術。

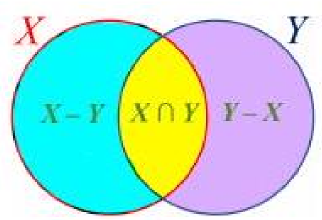
1.4.4 一些涉及聯合的例子

下一個結果涉及幾個事件聯合的概率。

定理：對於任意兩個事件 A 和 B（不一定不相交），

$$\mathrm{P}(A\cup B)=\mathrm{P}(A)+\mathrm{P}(B)-\mathrm{P}(A\cap B)$$

證明：接下來是一個簡單的維恩圖論證。（減去 $\mathrm{P}(A\cap B)$ 以避免重複計算。）



備註：公理 (3) 是這個定理的一個“特例”，$A\cap B=\emptyset$。

示例：假設有一個...

40% 天氣變冷的機率

10% 下雨和寒冷天氣的機率

80% 下雨或更冷天氣的機會。

那麼下雨的機會是

$$\mathrm{P}(R)=\mathrm{P}(R\cup C)-\mathrm{P}(C)+\mathrm{P}(R\cap C)=0.8-0.4+0.1=0.5.$$

定理：對於任意三個事件 A、B 和 C，

$$\begin{aligned}\mathrm{P}(A\cup B\cup C)=\mathrm{P}(A)+\mathrm{P}(B)+\mathrm{P}(C)\\ -\mathrm{P}(A\cap B)-\mathrm{P}(A\cap C)-\mathrm{P}(B\cap C)\\ +\mathrm{P}(A\cap B\cap C).\end{aligned}$$

證明：您可以嘗試通過維恩圖進行非正式證明，但要小心雙重和三重計數事件。這是建立在前一個定理之上的證明。

$$\begin{aligned}\mathrm{P}((A\cup B)\cup C)=&\mathrm{P}(A\cup B)+\mathrm{P}(C)-\mathrm{P}((A\cup B)\cap C)\\ =&\mathrm{P}(A\cup B)+\mathrm{P}(C)-\mathrm{P}((A\cap C)\cup(B\cap C))\\ =&\mathrm{P}(A\cup B)+\mathrm{P}(C)\\ &-[\mathrm{P}(A\cap C)+\mathrm{P}(B\cap C)-\mathrm{P}((A\cap C)\cap(B\cap C))]\\ =&\mathrm{P}(A)+\mathrm{P}(B)-\mathrm{P}(A\cap B)+\mathrm{P}(C)\\ &-\mathrm{P}(A\cap C)-\mathrm{P}(B\cap C)+\mathrm{P}(A\cap B\cap C).\quad\square\end{aligned}$$

示例：在某個音樂愛好者群體中，60 % 喜歡披頭士樂隊，50 % 喜歡滾石樂隊，40 % 喜歡殭屍樂隊，40 % 喜歡披頭士樂隊和滾石樂隊，30 % The Beatles and The Zoms，35% The Stones and The Zoms，以及 30% 三者。一個隨機的人不是這三個群體中任何一個的粉絲的概率是多少？

解決方案：首先，我們將計算隨機人至少喜歡其中一個組的概率，

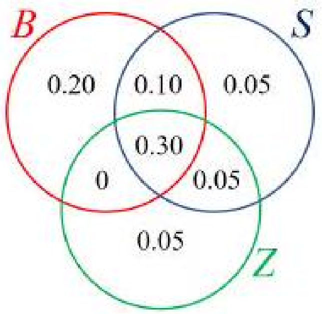
$$\begin{aligned}\mathrm{P}(B\cup S\cup Z)=&\mathrm{P}(B)+\mathrm{P}(S)+\mathrm{P}(Z)-\mathrm{P}(B\cap S)-\mathrm{P}(B\cap Z)-\mathrm{P}(S\cap Z)\\ &+\mathrm{P}(B\cap S\cap Z)\\ =&0.60+0.50+0.40-0.40-0.30-0.35+0.30=0.75.\end{aligned}$$

那麼這個人不喜歡任何群體的期望概率是

$$\mathrm{P}(\overline{B\cup S\cup Z})=1-\mathrm{P}(B\cup S\cup Z)=0.25.$$

現在找出一個隨機的人恰好喜歡三個組中的兩個的概率。

為了解決這個問題，我們可以使用維恩圖參數，從中心開始（因為 $\mathrm{P}(B\cap S\cap Z)=0.30$ ）並向外擴展。



然後我們有

$$\begin{aligned}&\mathrm{P}(\text{ only } B \text{ and } S)+\mathrm{P}(\text{ only } B \text{ and } Z)+\mathrm{P}(\text{ only } S \text{ and } Z)\\ &=\mathrm{P}(B\cap S\cap\bar{Z})+\mathrm{P}(B\cap\bar{S}\cap Z)+\mathrm{P}(\bar{B}\cap S\cap Z)\\ &=0.10+0+0.05=0.15.\end{aligned}$$

這是一般結果。

定理：n 事件並集的包含-排除原則：

$$
\begin{aligned}
&\mathrm{P}\left(A_1 \cup A_2 \cup \cdots \cup A_n\right) \\
&\qquad =\sum_{i=1}^n \mathrm{P}\left(A_i\right)-\sum_{i<j} \mathrm{P}\left(A_i \cap A_j\right)+\sum_{i<k} \sum_{i<j<k} \mathrm{P}\left(A_i \cap A_j \cap A_k\right) \\
&\qquad \cdots+(-1)^{n-1} \mathrm{P}\left(A_1 \cap A_2 \cap \cdots \cap A_n\right) .
\end{aligned}
$$

備註：您“包括”所有“單”事件，“排除”雙事件，包括三事件等。顯然，前兩個定理是特例。

1.4.4.1 榮譽證明

我們將通過歸納進行。我們定義 $P_i \equiv \mathrm{P}\left(A_i\right), P_{ij} \equiv \mathrm{P}\left(A_i \cap A_j\right), P_{ijk} \equiv \mathrm{P}\left(A_i \cap A_j \cap A_k\right)$ 等，對於所有 $i, j, k, \ldots$

另外，對於所有的 i，令 $B_i \equiv A_i \cap A_{n+1}$，使得 $\mathrm{P}\left(B_i\right)=P_{i, n+1}, \mathrm{P}\left(B_i \cap B_j\right)=P_{ij, n+1}, \mathrm{P}\left(B_i \cap B_j \cap B_k\right)=P_{ijk, n+1}$ 等。

讓 $\mathcal{A}_n \equiv \bigcup_{i=1}^n A_i$ 和 $\mathcal{B}_n \equiv \bigcup_{i=1}^n B_i$，對於 $n \geq 1$。

最後，我們可以開始了！歸納的基本情況很簡單（只需 $\mathrm{P}\left(A_1\right)=\mathrm{P}\left(A_1\right)$ ），然後是歸納步驟假設

$\mathrm{P}\left(\mathcal{A}_n\right)=\sum_{i=1}^n P_i-\sum_{i=1}^{n-1} \sum_{j=i+1}^n P_{ij}+\sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^n P_{ijk}-\cdots+(-1)^{n-1} P_{12 \cdots n}.$

當我們為 ” $n+1$ ” 案例建立結果時，歸納將完成。為此，

$$
\begin{aligned}
\mathrm{P}\left(\mathcal{A}_{n+1}\right) & =\mathrm{P}\left(\mathcal{A}_n \cup A_{n+1}\right) \\
& =\mathrm{P}\left(A_{n+1}\right)+\mathrm{P}\left(\mathcal{A}_n\right)-\mathrm{P}\left(\mathcal{A}_n \cap A_{n+1}\right) \\
& =P_{n+1}+\mathrm{P}\left(\mathcal{A}_n\right)-\mathrm{P}\left(\mathcal{B}_n\right)
\end{aligned}
$$

通過 $\mathrm{P}\left(\mathcal{A}_n\right)$ 的歸納步驟，然後再為 $\mathrm{P}\left(\mathcal{B}_n\right)$，我們有

$$
\begin{aligned}
&\mathrm{P}\left(\mathcal{A}_{n+1}\right)= \\
&\qquad \left[\sum_{i=1}^{n+1} P_i-\sum_{i=1}^{n-1} \sum_{j=i+1}^n P_{ij}+\sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^n P_{ijk}-\cdots+(-1)^{n-1} P_{12} \cdots n\right. \\
&\qquad \left.-\left[\sum_{i=1}^{n-1} P_{i, n+1}-\sum_{i=1}^n \sum_{j=i+1} P_{ij, n+1}-\cdots+(-1)^{n-1} P_{12 \cdots n+1}\right]\right. .
\end{aligned}
$$

煙霧從代數中清除後，我們得到

$$
\mathrm{P}\left(\mathcal{A}_{n+1}\right)=\sum_{i=1}^{n+1} P_i-\sum_{i=1}^n \sum_{j=i+1}^{n+1} P_{ij}+\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{k=j+1}^{n+1} P_{ijk}-\cdots+(-1)^n P_{12} \cdots n+1 .
$$

哇！

1.5 有限樣本空間

假設樣本空間 Ω 是有限的，比如 $\Omega=\left\{\omega_1, \omega_2, \ldots, \omega_n\right\}$。有限樣本空間通常允許我們更有效地計算某些事件的概率。

為了說明，讓$A \subseteq \Omega$ 是任何事件。那麼$\mathrm{P}(A)=\sum_{\omega_i \in A} \mathrm{P}\left(\omega_i\right)$。

示例：您有兩張紅牌、一張藍牌和一張黃牌。隨機選擇一張牌，其中“隨機”表示四張牌中的每一張都有相同的概率（1/4）被選中。

樣本空間$\Omega=\{$ red, blue, yellow $\}=\left\{\omega_1, \omega_2, \omega_3\right\}$。

$\mathrm{P}\left(\omega_1\right)=1/2, \mathrm{P}\left(\omega_2\right)=1/4, \mathrm{P}\left(\omega_3\right)=1/4$。

那麼，P(red or yellow\$)=(*{1})+(*{3})= 3 / 4 \$。

定義：簡單樣本空間 (SSS) 是一個有限樣本空間，其中所有結果的可能性均等。事實上，對於 SSS，所有結果的概率為 $1/|\Omega|$。

注意：在上面的例子中，Ω 不是 simple，因為 $\mathrm{P}(\omega_1) \neq \mathrm{P}(\,omega_2)$。

示例：投擲兩個公平的硬幣。

$\Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$ 是一個 SSS（所有概率都是 1/4）。

$\Omega' = \{0, 1, 2\}$（H' 的數量）不是 SSS。為什麼不？

下一個定理確立了當我們有 SSS 時計算某些概率的容易性。定理：對於 SSS Ω 中的任何事件 A，

$$\mathrm{P}(A) = \frac{|A|}{|\Omega|} = \frac{\#\text{ of elements in } A}{\#\text{ of elements in } \Omega}.$$

示例：擲骰子。令$A = \{1, 2, 4, 6\}$。每個結果的概率為 $1/6$，因此 $\mathrm{P}(A) = 4/6$

示例：現在讓我們擲一對骰子。以下是可能的結果（每個有序對的概率為 $1/36$）：

從這個 SSS 中，我們很容易得到關於兩次投擲之和的以下結果。

例如，$\mathrm{P}(\mathrm{Sum} = 4) = \mathrm{P}((1, 3)) + \mathrm{P}((2, 2)) + \mathrm{P}((3, 1)) = 3/36$

考慮到這些材料，我們現在可以繼續解決更複雜的計數問題。

1.6 計數技巧

想法：計算來自 SSS 的事件中的元素，以便有效地計算某些概率。我們將研究各種有用的規則/技術，包括 (i) 一些直觀的嬰兒示例，(ii) 排列和 (iii) 組合。

1.6.1 嬰兒示例

我們提供了一些直觀的例子——只是為了說明一些簡單的計數規則和由此產生的概率。

示例：假設您可以以 n_{A} 種方式選擇 A，並以 n_{B} 種方式選擇 B。如果只能做出一種選擇，那麼您有 $n_{\mathrm{A}} + n_{\mathrm{B}}$ 種可能的方式來選擇。例如，您去星巴克並決定吃鬆餅（藍莓或燕麥片）或百吉餅（芝麻、原味、鹽、大蒜），但不能兩者兼而有之。您總共有 $2 + 4 = 6$ 的選擇。

示例：假設有 n_{AB} 種方式從 City A 到 City B，n_{BC} 種方式從城市 B 到城市 C。然後你可以在 $n_{} 中從 A 到 C(via B) n_{}$ 種方式。例如，如果您可以從 A 步行、騎自行車或開車到 B，並且如果您可以從 開車、飛行、乘船或乘火車B 到 C，那麼你有 12 條可能的路線，從 A 到 B 到 C。

示例：擲兩個骰子。有多少結果？（假設 $(3, 2) \neq (2, 3)$。）答案是 $6 \times 6 = 36$

示例：擲 n 骰子。有 6^n 個可能的結果。

示例：翻轉三個硬幣。$2 \times 2 \times 2 = 8$ 可能的結果，

$$\{\mathrm{HHH}, \mathrm{HHT}, \mathrm{HTH}, \mathrm{HTT}, \mathrm{THH}, \mathrm{THT}, \mathrm{TTH}, \mathrm{TTT}\}.$$

第三枚硬幣是正面的概率是多少？答案（通過查看列表 ）$= 4/8$

備註：在前面的示例中，我們隱含地指定選擇順序很重要（例如，$(3, 2) \neq (2, 3)$）。此外，當我們從有限集合中隨機選擇項目時，重要的是要指定我們是選擇有放回還是無放回。替換選擇意味著被選擇的元素在下一次選擇之前返回到集合中，因此可以多次選擇相同的元素。不替換選擇意味著每次選擇都會從集合中刪除一個元素，因此每個項目最多可以選擇一次。

示例：從一副牌中選擇兩張牌，不考慮順序（即 $(Q\phi, 7\kappa\beta) \neq (7\boldsymbol{\kappa}, Q(\phi))$. 我們有多少種方法可以做到這一點？不難看出答案是

$(\#$ 的選擇方式 1^{st} card $)\cdot(\#$ 的選擇方式 2^{nd}（後 1^{st}）$)$,

這只是 $52 \cdot 51 = 2652$。

示例：給定一個包含 10 只襪子的盒子 - 兩隻紅色和八隻黑色 - 選擇兩隻，無需更換。

- 選擇的兩隻襪子都是紅色的概率是多少？讓A 成為兩者都是紅色的事件。然後

$$\mathrm{P}(A)=\frac{\#\text{ of ways to pick two reds}}{\#\text{ of ways to pick two socks}}=\frac{2\cdot 1}{10\cdot 9}=\frac{1}{45}.$$

- 兩隻襪子都是黑色的概率是多少？讓 B 成為兩者都是黑色的事件。那麼$\mathrm{P}(B)=\frac{8\cdot 7}{10\cdot 9}=\frac{28}{45}$。

- 一隻襪子是紅色的，另一隻是黑色的概率是多少？讓 C 是每種顏色都有一個的事件。由於 A 和 B 是不相交的，我們有

$$\mathrm{P}(C)=1-\mathrm{P}(\bar{C})=1-\mathrm{P}(A\cup B)=1-\mathrm{P}(A)-\mathrm{P}(B)=\frac{16}{45}$$

1.6.2 排列

現在樂趣開始了。假設我們想計算從 n 個對像中選擇 r 的方式的數量。讓我們從一個特殊情況開始。

定義：按一定順序排列的n個不同符號是n個符號的排列。

示例：數字$\{1,2,3\}$有多少種排列方式？

答：有 $3!=6$ 的方式。這很容易通過列出所有可能性來驗證：$123,132,213,231,312,321$。

在大多數實際示例中，所涉及的元素數量過多，無法使完全枚舉成為可行的策略，因此我們需要一種更通用的方法。

示例：數字 $\{1,2,\ldots,n\}$ 有多少種不同的排列方式？等效地，您可以從 n 個不同的字母中拼出多少個單詞，每個字母只使用一次？或者您如何安排一系列 n 工作？或者參觀 n 個城市的行程？答：稍微反思一下，這只是一個沒有替換的順序採樣的例子。

$$\begin{aligned}&(\text{visit 1st city})\ (\text{then visit 2nd city})\cdots(\text{then visit last city})\\&=n(n-1)(n-2)\cdots 2\cdot 1=n!\end{aligned}$$

因此，如果您想訪問 9 個城市，您可以在 $9!=362880$ 的訂單中完成。

定義：一個 r 元組是一個基數 r 的有序列表（即它有 r 個元素）。例如，$(5,6,1,3)$ 是一個 4 元組。

定義/定理：我們可以從 n 個不同的符號（每個最多使用一次）組成的 r-tuples 的數量稱為一次 r 次採取的 n 個事物的排列數，並且

$$P_{n,r}\equiv\frac{n!}{(n-r)!}.$$

（我們處理了上面 $r=n$ 的特殊情況，導致 $P_{n,n}=n!/0!=n!$ ）

證明：你可以從 n 個不同的字母組成的大小為 r 的單詞的數量（每個最多使用一次）是

$$\begin{aligned}P_{n,r}&=\left(\text{ choose } 1^{\text{st}}\text{ letter }\right)\left(\text{ choose } 2^{\text{nd}}\right)\cdots\left(\text{ choose } r^{\text{th}}\right)\\&=n(n-1)(n-2)\cdots(n-r+1)\\&=\frac{n(n-1)\cdots(n-r+1)(n-r)\cdots 2\cdot 1}{(n-r)\cdots 2\cdot 1}\\&=\frac{n!}{(n-r)!}.\end{aligned}$$

示例：您可以從 $\{1,2,3,4,5\}$ 組成多少個兩位數？

答案：$P_{5,2}=5 ！/(5-2) !=20 $。讓我們列出它們：

12	13	14	15	21	23	24	25	31	32
34	35	41	42	43	45	51	52	53	54

有多少個三位數？答案：$P_{5,3}=5 ！/ 2 !=60 $。（如果您真的無事可做，請隨意列出它們。）

示例：有八名候選人競選吉利根島鑑賞協會 (GIAS) 的三個職位（總裁、副總裁和財務主管）。這些高管職位有多少種選擇方式？答案：$P_{8,3}=8！/(8-3)!=336$ 種方式。

示例：在這 336 種方式中，有多少種方式讓瓊斯擔任總統？

方法1：前三個位置看起來像：（瓊斯，？，？）。這相當於從剩下的七名候選人中選出兩名。所以 $P_{7,2}=7！/(7-2)!=42$ 種方式。

方法2：解決問題的另一種方法是簡單地認識到八位候選人中的每一位都“同等可能”成為總統。因此，Pres 的方式數量。瓊斯等於 $336/8=42$。

示例：由數字 $\{1,2,\ldots,9\}\ldots$ 可以組成多少個六位數的車牌

- 沒有重複？（例如，123465 可以，但 133354 不行。）$P_{9,6}=9 ！/ 3 !=60480 $。
- 允許重複？（什麼都好。）$9\times 9\times\cdots\times 9=9^6=531441$。
- 包含重複？$531441-60480=470961$。

1.6.3 組合

我們可能還希望計算從 n 個對像中選擇 r 的方法的數量，而不考慮順序。在這種情況下，我們問：可以從一組 n 個對像中選擇多少個大小為 r 的不同子集？（回想一下，集合成員的順序沒有意義。）

示例：$\{1,2,3\}$ 有多少個子集正好包含兩個元素？（順序並不重要。）

答案：三個子集$-\{1,2\},\{1,3\},\{2,3\}$。

定義：具有n 個元素的集合中具有r 個元素的子集的數量稱為n 個事物的組合數r-at-a-time。

Notation: $\binom{n}{r}$ or $C_{n,r}$ (read as " n choose r "). These are also called binomial coefficients, and have already made an appearance in the Binomial Theorem, Equation 1.1. We will see below that $C_{n,r}=\frac{n!}{r!(n-r)!}$.

排列組合的區別： - 組合不關心元素的順序，即$\{a,b,c\}=\{b,a,c\}$

- 排列與元素的順序有關，即 $(a,b,c)\neq(b,a,c)$
- 每次r 所取的n 事物的排列數總是至少與組合數一樣大。實際上， …

選擇排列與首先從 n 中選擇 r 個對象的組合，然後將 r 元素按順序排列是一樣的，即

$$\frac{n!}{(n-r)!}=\binom{n}{r}r!$$

所以，

$$C_{n,r}\equiv\binom{n}{r}\equiv\frac{n!}{(n-r)!r!}$$

以下結果都應該是直觀的：

$$\binom{n}{r}=\binom{n}{n-r},\quad\binom{n}{0}=\binom{n}{n}=1,\quad\binom{n}{1}=\binom{n}{n-1}=n$$

事實上，您可能會在帕斯卡三角的背景下認出這些二項式係數，您可能早在小學就已經看到了。

n	$\binom{n}{r},\ r=0,1,\ldots,n$
0	1
1	1 1
2	1 2 1
3	1 3 3 1
4	1 4 6 4 1
5	1 5 10 10 5 1
6	1 6 15 20 15 6 1

示例：一個辦公室排球隊有 14 名成員。教練有多少種選擇首發六人的方法？（順序無關緊要。）

$$\binom{14}{6}=\frac{14!}{6!8!}=3003$$

示例：史密斯是球隊中的一名球員。3003首發陣容中有多少人包括她？

$$\binom{13}{5}=\frac{13!}{5!8!}=1287$$

（史密斯免費獲得了六個職位之一。現在剩下的 13 名球員可以填補五個職位。）

示例：假設您自豪地擁有七隻相同的紅色鞋子和五隻相同的藍色鞋子。您收藏的可能安排的總數是多少？這是一種這樣的安排：

$$\mathrm{R}\quad \mathrm{B}\quad \mathrm{R}\quad \mathrm{R}\quad \mathrm{B}\quad \mathrm{B}\quad \mathrm{R}\quad \mathrm{R}\quad \mathrm{R}\quad \mathrm{B}\quad \mathrm{R}\quad \mathrm{B}.$$

Answer: You can think of this as the number of ways to put seven red shoes into 12 slots (or, equivalently, the number of ways to put five blue shoes into 12 slots), which is simply equal to $\binom{12}{7}$.

1.7 計數應用程序

在本節中，我們將討論計數技術的各種應用。這是即將發生的事情。

1.6.1 超幾何問題

1.6.2 二項式問題

$1.6.3 多項式係數

1.6.4 排列與組合

1.6.5 生日問題

$1.6.6 信封問題

1.6.7 撲克概率

1.7.1 超幾何分佈

假設您有一個由類型 1 的 a 對象和類型 2 的 b 對象組成的集合。從這個 $a+b$ 對象的集合中，您將選擇總共 n 個對象而不進行替換。被選中的對像中有 k 屬於類型 1 的概率是多少，其中 $k=\max\{0,nb\},\ldots,\min\{a,n\}$。

我們有

$$
\begin{aligned}
&\mathrm{P}(k \text{ type 1's are picked })\\
&=\frac{(\# \text{ ways to choose } k \text{ type 1 's from } a)(\text{ choose } n-k \text{ type 2 's from } b)}{\# \text{ ways to choose } n \text{ out of } a+b}\\
&=\frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}} \cdot \square
\end{aligned}
$$

選擇的類型 1 的數量被稱為具有超幾何分佈。稍後我們將對“分佈”進行非常徹底的討論。示例：假設您的襪子抽屜包含 15 隻紅色襪子和 10 只藍色襪子，總共 25 只襪子。隨機抽取七個，不更換。你選擇三個紅色和四個藍色的概率是多少？回答：

$$
\mathrm{P}(\text{ exactly three reds are picked })=\frac{\binom{15}{3}\binom{10}{4}}{\binom{25}{7}}=0.1988.
$$

示例：從同一個襪子抽屜（包含全部 25 只襪子）開始，假設一覺醒來（在你喝完早晨的咖啡之前），你盲目地挑選了兩隻襪子而沒有更換。你選擇配對的概率是多少？回答：

$$
\begin{aligned}
&\mathrm{P}(\text{ matching pair })\\
&\quad=\mathrm{P}(\text{ exactly two reds are picked })+\mathrm{P}(\text{ exactly two blues are picked})\\
&\quad=\frac{\binom{15}{2}\binom{10}{0}}{\binom{25}{2}}+\frac{\binom{15}{0}\binom{10}{2}}{\binom{25}{2}}=\frac{\binom{15}{2}+\binom{10}{2}}{\binom{25}{2}}=0.5 . \square
\end{aligned}
$$

1.7.2 二項分佈

讓我們從 \$1.6.1 中相同的對象集合開始。由類型 1 的 a 對象和類型 2 的 b 對象組成。但是現在從 $a+b$ 中選擇 n 個帶有替換的對象。然後，

$$
\begin{aligned}
&\mathrm{P}(k \text{ type 1 's are picked })\\
&\quad=(\# \text{ of ways to choose } k \text{ type 1 's and } n-k \text{ type 2 's })\\
&\quad\times \mathrm{P}(\text{ a particular selection of } k \text{ type 1 's and } n-k \text{ type 2 's })\\
&=\binom{n}{k} \mathrm{P}(\text{ choose } k \text{ type 1 's in a row, then } n-k \text{ type 2 's in a row })\\
&\quad=\binom{n}{k}\left(\frac{a}{a+b}\right)^k\left(\frac{b}{a+b}\right)^{n-k}, \quad k=0,1, \ldots, n . \quad \square
\end{aligned}
$$

選擇的類型 1 的數量被稱為具有二項分佈。

示例：一個盒子裡有 25 只襪子，其中 15 隻紅色和 10 只藍色。選擇七個替換。

$$
\mathrm{P}(\text{ exactly three reds are picked })=\binom{7}{3}\left(\frac{15}{25}\right)^3\left(\frac{10}{25}\right)^{7-3}=0.1936
$$

請務必將此答案與類似的超幾何示例的答案進行比較。我們將在接下來的章節中詳細討論二項分佈。

1.7.3 多項係數

在本小節中，我們將通過添加更多類別來擴展超幾何和二項式分佈的概念。以前，我們手頭有類型 1 的 a 對象和類型 2 的 b 對象，我們從中採樣了 n 個對象。這裡我們考慮更一般的情況，其中我們有 i 類型的 a_i 對象，其中 $i=1,2,\ldots,c$，因此我們現在有 c 類別而不是只有兩個。令 $A=\sum_{i=1}^c a_i$ 表示所有類型對象的總數，假設我們抽取其中的 n 個樣本。我們感興趣的是獲得選擇類型為 $i=1,2,\ldots,c$ 的 k_i 個對象的概率，其中 \$n=*{i=1}^{c}* k{i} \$，並且 k_i 都是“合法的”，例如 $k_i \leq a_i$

1.7.3.1 \$ 1.6.3.1 無替換採樣

在無放回抽樣的情況下，超幾何分佈的直接概括表明

$$\mathrm{P}\left(\text{ Select } k_i \text{ type } i\text{ 's }, i=1,2,\ldots,c\right)=\frac{\binom{a_1}{k_1}\binom{a_2}{k_2}\cdots\binom{a_c}{k_c}}{\binom{A}{n}}=\frac{\prod_{i=1}^{c}\binom{a_i}{k_i}}{\binom{A}{n}}.$$

示例：假設您的襪子抽屜包含 15 隻紅色襪子、10 只藍色襪子和 12 只綠色襪子，總共 37 只襪子。隨機抽取九個，不更換。您選擇三個紅色、四個藍色和兩個綠色的概率是多少？回答：

$$\mathrm{P}\left(3\mathrm{R}'\mathrm{s}, 4\ \mathrm{B}'\mathrm{s}, 2\mathrm{G}'\mathrm{s}\right)=\frac{\binom{15}{3}\binom{10}{4}\binom{12}{2}}{\binom{37}{9}}=0.0507$$

1.7.3.2 &1.6.3.2 替換採樣

在帶放回抽樣的情況下，二項分佈的類似直接推廣產生多項分佈，

$$\mathrm{P}\left(\text{ Select } k_i \text{ type } i\text{ 's }, i=1,2,\ldots,c\right)=\binom{n}{k_1,k_2,\ldots,k_c}\prod_{i=1}^{c}\left(a_i/A\right)^{k_i}$$

在哪裡

$$\binom{n}{k_1,k_2,\ldots,k_c}\equiv\frac{n!}{k_1!k_2!\cdots k_c!}$$

是所謂的多項式係數，它表示您可以從所有 n 個對像中選擇 k_i 個類型為 $i=1,2,\ldots,c$ 的對象的方式的數量。當然，二項式係數對應於$c=2$的情況。

示例：從數字 $1,2,2,3,3,3,3,4,4,4,4$ 可以形成多少個不同的 11 位數字？這只是一個多項式係數，

$$\frac{\#\text{ of permutations of 11 digits}}{(\#1's)!\,(\#2's)!(\#3's)!(\#3's)!}=\frac{11!}{1!2!4!4!}=34,650.$$

示例：同樣，考慮有 15 隻紅色襪子、10 只藍色襪子和 12 只綠色襪子的襪子抽屜，總共 37 只襪子。隨機挑選九個，有替換。現在你選擇三個紅色、四個藍色和兩個綠色的概率是多少？回答：

$$\mathrm{P}\left(3\mathrm{R}'\mathrm{s}, 4\ \mathrm{B}'\mathrm{s}, 2\mathrm{G}'\ \mathrm{s}\right)=\binom{9}{3,4,2}(15/37)^3(10/37)^4(12/37)^2=0.0471$$

1.7.4 排列與組合

到現在為止，您可能對何時考慮排列以及何時考慮組合有點困惑。實際上，您通常可以使用多種方法來解決特定問題。下面是一個可以通過排列或組合來解決的問題的示例。這兩種方法都會給你正確的答案，但是（在這種情況下，無論如何）一種方法最終會簡單得多。

示例：假設您有四顆紅色彈珠和兩顆白色彈珠。出於某種原因（只有您自己知道），您希望將它們隨機排列成一排。尋找：

a. P（兩端的彈珠都是W）。

b. P（兩端的彈珠不都是W）。

c. P（兩個 W 並排）。

方法一（使用排列）：讓樣本空間

$$\Omega=\{\text{ every random ordering of the six marbles }\}.$$

a. 定義事件A：兩端彈珠是W，即WRRRRW。注意

$$\begin{aligned}|A| =& \left(\# \text{ of ways to permute the } 2W' \text{ s in the end slots }\right)\\ &\times \left(\# \text{ of ways to permute the } 4R' \text{ 's in the middle slots }\right)\\ =& 2! \times 4! = 48.\end{aligned}$$

這意味著

$$\mathrm{P}(A) = \frac{|A|}{|\Omega|} = \frac{48}{720} = \frac{1}{15}$$

b. 事件“結束的彈珠不是都是白色的”只是A 的補集。因此，$\mathrm{P}(\bar{A}) = 1 - \mathrm{P}(A) = 14/15$。

c. 定義事件B：兩個W 並排，即WWRRRR 或RWWRRR 或$\cdots$ 或RRRRWW。然後

$$\begin{aligned}|B| =& \left(\# \text{ of ways to select a pair of side-by-side slots for 2 W's}\right)\\ &\times (\# \text{ of ways to insert W's into a pair of slots })\\ &\times (\# \text{ of ways to insert } R \text{ 's into the remaining slots })\\ =& 5 \times 2! \times 4! = 240.\end{aligned}$$

因此，

$$\mathrm{P}(B) = \frac{|B|}{|\Omega|} = \frac{240}{720} = \frac{1}{3}$$

但是 - 上述方法花費了太多時間！這裡有一個更簡單的方法…

方法 2（使用組合）：在這種方法中，我們首先重新定義樣本空間，如下所示：

$$\Omega = \{ \text{ possible pairs of slots that the white marbles can occupy } \}.$$

Clearly, $|\Omega| = \binom{6}{2} = 15$. (a) Since the W's must occupy the end slots in order for A to occur, then

$$|A| = 1 \Rightarrow \mathrm{P}(A) = |A|/|\Omega| = 1/15.$$

b. $\mathrm{P}(\bar{A}) = 14/15$

c. $|B| = 5 \Rightarrow \mathrm{P}(B) = 5/15 = 1/3$。（那好多了！）

1.7.5 生日問題

“生日問題”是傳統上介紹概率類的經典示例。結果可能會讓你吃驚！

假設一個房間裡有 n 人。求他們中至少有兩個生日相同的概率。為了簡化分析，我們忽略了 2 月 29 日出生的任何人，並假設所有 365 天的概率相等。

（簡單）樣本空間是

$$\Omega = \{(x_1, \ldots, x_n) : x_i = 1, 2, \ldots, 365, \forall i\}$$

其中 x_i 是人 i 的生日，注意 $|\Omega| = (365)^n$。

讓 A 成為“所有生日都不同”的事件。那麼A的大小就是n不同天的排列數，簡單來說就是

$$|A| = P_{365,n} = (365)(364)\cdots(365 - n + 1).$$

因此，我們有

$$\begin{aligned}P(A) &= \frac{(365)(364)\cdots(365 - n + 1)}{(365)^n}\\ &= 1 \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{365 - n + 1}{365}.\end{aligned}$$

但我們想要

$$\mathrm{P}(\bar{A})=1-\left(1\cdot\frac{364}{365}\cdot\frac{363}{365}\cdots\frac{365-n+1}{365}\right)$$

1.7.5.1 備註：

- 當 $n=366$時，$\mathrm{P}(\bar{A})=1$。
- 對於 $\mathrm{P}(\bar{A})$ 為 $>1/2$，n 必須為 ≥ 23。換句話說，一個只有 23 名學生的班級有兩個生日相同的學生比平均概率要高。（奇怪！）
- 當 $n=50$時，$\mathrm{P}(\bar{A})=0.97$。

1.7.6 信封問題

一群 n 人收到了 n 個寫有他們名字的信封——但有人完全把信封弄混了！找出至少一個人會收到正確信封的概率。（這個故事有很多變化，在數學上是等價的。）

讓A_i 表示i 人收到了正確的信封。那麼我們顯然想要 $\mathrm{P}\left(A_1\cup A_2\cup\cdots\cup A_n\right)$

回想一下包含-排除的一般原理，等式（1.2，它指出

$$\begin{aligned}\mathrm{P}\left(A_1\cup A_2\cup\cdots\cup A_n\right)&\\&=\sum_{i=1}^n\mathrm{P}\left(A_i\right)-\sum_{i<j}\mathrm{P}\left(A_i\cap A_j\right)+\sum_{i<j}\sum_{i<j<k}\mathrm{P}\left(A_i\cap A_j\cap A_k\right)\\-\cdots+(-1)^{n-1}\mathrm{P}\left(A_1\cap A_2\cap\cdots\cap A_n\right).\end{aligned}$$

由於所有的 $\mathrm{P}(A_i)'$ 都是相同的，所以所有的 $\mathrm{P}\left(A_i\cap A_j\right)$ 是相同的，等等，我們有

$$\begin{aligned}\mathrm{P}\left(A_1\cup A_2\cup\cdots\cup A_n\right)&\\&=n\mathrm{P}\left(A_1\right)-\binom{n}{2}\mathrm{P}\left(A_1\cap A_2\right)+\binom{n}{3}\mathrm{P}\left(A_1\cap A_2\cap A_3\right)\\&\quad-\cdots+(-1)^{n-1}\mathrm{P}\left(A_1\cap A_2\cap\cdots\cap A_n\right)\end{aligned}$$

最後，$\mathrm{P}\left(A_1\right)=1/n,\mathrm{P}\left(A_1\cap A_2\right)=1/(n(n-1))$ 等（為什麼？），暗示

$$\begin{aligned}&\mathrm{P}(A_1\cup A_2\cup\cdots\cup A_n)\\&\quad=\frac{n}{n}-\binom{n}{2}\frac{1}{n}\cdot\frac{1}{n-1}+\binom{n}{3}\frac{1}{n}\cdot\frac{1}{n-1}\cdot\frac{1}{n-2}-\cdots+(-1)^{n-1}\frac{1}{n!}\\&\quad=1-\frac{1}{2!}+\frac{1}{3!}-\cdots+(-1)^{n-1}\frac{1}{n!}\doteq 1-\frac{1}{e}\doteq 0.6321.\square\end{aligned}$$

示例：如果只有 $n=4$ 個信封，那麼

$$\mathrm{P}\left(A_1\cup A_2\cup A_3\cup A_4\right)=1-\frac{1}{2!}+\frac{1}{3!}-\frac{1}{4!}=0.625,$$

這在漸近貨幣上是正確的！

1.7.7 撲克問題

賭博問題是概率論早期發展的重要動力。我們將看幾個與撲克遊戲相關的例子。

First of all, we draw a “hand” of five cards at random from a standard deck of 52 cards. The sample space Ω is the set of all possible hands, so that the number of possible hands is $|\Omega|=\binom{52}{5}=2,598,960$. Here is some standard terminology (not that we’re advocating gambling, but if you aren’t familiar with poker, take a few minutes to learn the basics):

排名 $=2,3,\ldots,10$, Jack (11), Queen (12), King (13), Ace (14, 有時 1) 花色 $=\boldsymbol{\Omega}($ clubs),\$ (\$ 鑽石), $\vee$ (紅心), (黑桃)。

我們現在將計算獲得各種特殊牌的概率。

a. 一對，例如，$K\diamond, A\kappa, K\boldsymbol{A}, 7\nabla, 6\diamond$ 恰好是一對 K，但沒有比這更好的了（例如兩對或三對，如下所述）。

We can pick the rank of the pair (a K here) in 13 ways, and then the suits of the two kings (here $\diamond$ and $\mathbf{A}$) in $\binom{4}{2} = 6$ ways.

We now have to deal with the remaining three cards (here $A\mathbf{s}, 7\oslash, 6\diamond$). The restrictions are that no more K 's are allowed, and that no other pairs (or a three-of-a-kind) are allowed. Thus, we must choose those cards from three different ranks (from the remaining 12 non- K 's). Those three ranks can be chosen in $\binom{12}{3}$ ways, and then the suits of the corresponding three cards can be chosen in $4 \cdot 4 \cdot 4 = 4^3$ ways.

把所有這些放在一起，我們得到

$$\mathrm{P}(\text{ one pair }) = \frac{13\binom{4}{2}\binom{12}{3}4^3}{\binom{52}{5}} = \frac{1,098,240}{2,598,960} \doteq 0.4226.$$

b. 兩對，例如，$K\diamond, 6\mathbf{c}, K\mathbf{A}, 7\otimes, 6\diamond$。這手牌正好有兩對（$K$ 和 6's）和一張垃圾牌（但不包括接下來會出現的葫蘆）。

Since one pair is as good as the other, the order in which we pick the ranks of the pairs doesn't matter. Thus, we can pick the ranks of the two pairs (K and 6 here) in $\binom{13}{2}$ ways.

Then we select the suits of the two kings (here $\diamond$ and $\mathbf{0}$) in $\binom{4}{2} = 6$ ways, and the suits of the two 6 's (here $\boldsymbol{s}$ and $\diamond$) in $\binom{4}{2} = 6$ ways.

剩下的牌（這裡是 7ʊ ）可以從除了 K 和 6 之外的任何東西中選擇，這可以以 44 種方式發生。

把所有這些放在一起，我們得到

$$\mathrm{P}(\text{ two pairs }) = \frac{\binom{13}{2}\binom{4}{2}\binom{4}{2}44}{\binom{52}{5}} = \frac{123,552}{2,598,960} \doteq 0.0475.$$

c. 滿堂彩（一對 + 一對三），例如 $K\diamond, 6\mathbf{c}, K, \mathbf{0}, 6\mathrm{D}, 6\diamond$。與前面涉及兩個對的示例不同，我們注意到在選擇對和三元組的等級時順序很重要（因為對和三元組是不同的）。事實上，這個選擇可以通過 $P_{13,2} = 13 \cdot 12$ 的方式來完成。

Once we have the rank of the three-of-a-kind, we can select the corresponding suits in $\binom{4}{3}$ ways; and, similarly, $\binom{4}{2}$ ways for the pair. Then

$$\mathrm{P}(\text{ full house }) = \frac{P_{13,2}\binom{4}{3}\binom{4}{2}}{\binom{52}{5}} = \frac{3744}{2,598,960} \doteq 0.00144.$$

d. 直線（連續五個等級），例如 $3\diamond, 4\kappa, 5\boldsymbol{A}, 6 \curvearrowright, 7\diamond$。我們首先選擇直線 ($A, 2, 3, \ldots, 10$，其中約定允許 Ace 將雙重任務作為"1") 的起點；這可以通過 10 種方式完成。然後我們只需為順子中的每張牌選擇一套花色，這可以以 4^5 的方式完成。因此，

$$\mathrm{P}(\text{ straight }) = \frac{10 \cdot 4^5}{2,598,960} \doteq 0.00394$$

e. Flush (all five cards from same suit), e.g., $3\diamond, 7\diamond, 9\diamond, J\diamond, A\diamond$. We start by selecting a suit for the flush (4 ways), and then five cards from that suit, which can be achieved in $\binom{13}{5}$ ways. Then

$$\mathrm{P(\ flush\ )} = \frac{5148}{2,598,960} \doteq 0.00198$$

f. 同花順，例如 3 ◇ 、4 ◇ 、5 ◇ 、6 ◇ 、7◇。選擇順子的起點（10 種方式），然後選擇花色（4 種方式）。這產生

$$\mathrm{P(\ straight\ flush\ )} = \frac{40}{2,598,960} \doteq 0.0000154$$

這個概率真的很小！什麼同花順？！

備註：涉及撲克等遊戲的組合問題是無限的。你能做橋樑問題嗎？快艇？

Example: Yahtzee! Toss five six-sided dice. What's the probability that you'll see a full house, e.g., 2,5,5,2,5? This is actually easier to analyze than the corresponding poker problem. Select the numbers that will correspond to the pair and triple (order matters), which can be done in $P_{6,2} = 6 \cdot 5$ ways. Then choose the positions in the group of five tosses that the pair occupies, which can be done in $\binom{5}{2}$ ways; and finish by choosing the remaining slots for the triple, which can be done in $\binom{3}{3}$ ways. Thus,

$$\mathrm{P(\ Yahtzee\ full\ house\ )} = \frac{P_{6,2}\binom{5}{2}\binom{3}{3}}{6^5} = \frac{300}{7776} \doteq 0.03858.$$

如需更多 Yahtzee 樂趣，請參閱

$$\mathrm{https://datagenetics.com/blog/january42012/index.\ html}\ .$$

1.8 條件概率和獨立性

想法：當我們獲得更多信息時，我們經常可以更新事件的概率。有時概率會改變，有時不會。在本節中，我們將探討這個問題並提出許多有趣的發現。

1.8.1 條件概率

例子：如果 A 是一個人體重至少 150 磅的事件，那麼 $\mathrm{P}(A)$ 肯定與這個人的身高有關，例如，如果 B 是這個人是至少 6 英尺高，而 B 是指該人的高度為 <5 英尺。

示例：對於標準擲骰子，定義事件 $A=\{2,4,6\}$ 和 $B=\{1,2,3,4,5\}$。那麼 $\mathrm{P}(A)=1/2$ 和 $\mathrm{P}(B)=5/6$。

假設我們知道出現了 B（這樣就不可能出現“6”）。那麼在 B 出現的情況下 A 出現的概率是

$$\mathrm{P}(A \mid B) = \frac{2}{5} = \frac{|A \cap B|}{|B|}$$

其中符號“$A \mid B$” 可以被解讀為事件“A given B”。

因此，A 的概率取決於您擁有的信息！B 出現的信息使我們能夠將 B 視為一個新的、受限的樣本空間。所以，假設我們有一個簡單的樣本空間Ω，那麼

$$\mathrm{P}(A \mid B) = \frac{|A \cap B|}{|B|} = \frac{|A \cap B|/|\Omega|}{|B|/|\Omega|} = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}.$$

最後，這裡是條件概率的（更一般的）定義，它不依賴於對 SSS 的需求。

定義：如果$\mathrm{P}(B)>0$，給定B，A的條件概率為

$$\mathrm{P}(A \mid B) \equiv \mathrm{P}(A \cap B)/\mathrm{P}(B).$$

備註：如果A和B不相交，則$\mathrm{P}(A \mid B)=0$。（如果出現了 B，那麼 A 也不會出現。）

如果 $\mathrm{P}(B)=0$ 會發生什麼？不用擔心！在這種情況下，考慮 $\mathrm{P}(A \mid B)$ 是沒有意義的

示例：擲兩個骰子並觀察總和。定義事件A：奇數，即$\{3,5,7,9,11\}$和$B:\{2,3\}$。然後

$$
\begin{aligned}
\mathrm{P}(A)&=\mathrm{P}(3)+\mathrm{P}(5)+\cdots+\mathrm{P}(11)=\frac{2}{36}+\frac{4}{36}+\cdots+\frac{2}{36}=\frac{1}{2},\\
\mathrm{P}(B)&=\mathrm{P}(2)+\mathrm{P}(3)=\frac{1}{36}+\frac{2}{36}=\frac{1}{12},\quad \text{and}\\
\mathrm{P}(A\mid B)&=\frac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)}=\frac{\mathrm{P}(3)}{\mathrm{P}(B)}=\frac{2/36}{1/12}=\frac{2}{3}.
\end{aligned}
$$

因此，根據 B 提供的信息，我們看到 $\mathrm{P}(A)=1/2$ 增加到 $\mathrm{P}(A\mid B)=2/3$

例子：假設你的襪子抽屜裡只有四隻白襪子和八隻紅襪子。（毫無疑問，您經常會發現自己處於一種相似的情況。）選擇兩隻襪子，無需更換。定義以下事件： $A: 1^{\text{st}}$ 襪子是白色的。

$B: 2^{\text{nd}}$ 襪子是白色的。

C : 兩隻襪子都是白色的 $(=A\cap B)$。

讓我們找到 $\mathrm{P}(C)$ 和 $\mathrm{P}(B)$，後者在直覺上會變得微不足道（但無論如何我們都會小心處理）。首先，根據條件的定義，

$$
\mathrm{P}(C)=\mathrm{P}(A\cap B)=\mathrm{P}(A)\mathrm{P}(B\mid A)=\frac{4}{12}\cdot\frac{3}{11}=\frac{1}{11}
$$

此外，很容易看出 $B=(A\cap B)\cup(\bar{A}\cap B)$，其中並集的兩個分量是不相交的。所以，

$$
\begin{aligned}
\mathrm{P}(B)&=\mathrm{P}(A\cap B)+\mathrm{P}(\bar{A}\cap B)\\
&=\mathrm{P}(A)\mathrm{P}(B\mid A)+\mathrm{P}(\bar{A})\mathrm{P}(B\mid \bar{A})\\
&=\frac{4}{12}\cdot\frac{3}{11}+\frac{8}{12}\cdot\frac{4}{11}=\frac{1}{3}.\square
\end{aligned}
$$

你能不假思索地得到這個結果嗎？（答案：是的，只要稍加練習，你就會非常擅長！）

示例：一對夫婦有兩個孩子，其中至少一個是男孩。兩個都是男孩的概率是多少？7^7

樣本空間和相關事件是：

- $\Omega=\{GG,GB,BG,BB\}$（例如，“BG”的意思是“男孩然後女孩”）。
- C : 都是男孩 $=\{BB\}$。
- D : 至少有一個是男孩 $=\{\mathrm{GB},\mathrm{BG},\mathrm{BB}\}$。

我們需要計算

$$
\mathrm{P}(C\mid D)=\frac{\mathrm{P}(C\cap D)}{\mathrm{P}(D)}=\frac{\mathrm{P}(\mathrm{BB})}{\mathrm{P}(\mathrm{GB},\mathrm{BG},\mathrm{BB})}=\frac{1/4}{3/4}=1/3.
$$

（幾年前，當我還是一名高中數學家時，我的直覺是 1 $ / 2 $——錯誤的答案！問題是我們不知道 D 是指第一個孩子還是第二個孩子。）

隨著您獲得更多信息，您可以做出一些更令人驚訝的發現。.

1.8.1.1 榮譽示例

一對夫婦有兩個孩子，其中至少一個是周二出生的男孩。假設任何孩子有 1/14 的機會出現任何性別/星期幾組合，那麼這對夫婦的兩個孩子都是男孩的概率是多少？

[7] 這是一個有趣而棘手的問題，引發了許多熱烈的討論；直覺有時會愚弄你！讓事件 $\mathrm{B}_x\,[\mathrm{G}_x]=\mathrm{Boy}$ [Girl] 出生於 $x, x=1,2, , 7 （例如，x=3$ 是星期二）。包含有序孩子對的可行樣本空間是：

$$
\Omega=\left\{\left(\mathrm{G}_x,\mathrm{G}_y\right),\left(\mathrm{G}_x,\ \mathrm{B}_y\right),\left(\mathrm{B}_x,\mathrm{G}_y\right),\left(\mathrm{B}_x,\ \mathrm{B}_y\right),x,y=1,2,\ldots,7\right\}
$$

所以$|\Omega|=4\times 49=196$。同時，定義以下事件：

- C : 都是男孩（至少有一個在星期二出生）

$$=\{(\mathrm{B}_x,\ \mathrm{B}_3), x=1,2,\ldots,7\}\cup\left\{\left(\mathrm{B}_3,\ \mathrm{B}_y\right), y=1,2,\ldots,7\right\}$$

注意 $|C|=13$ (以避免重複計算 $(\mathrm{B}_3,\ \mathrm{B}_3)$)。

- D : 至少有一個男孩在星期二出生

$$=C\cup\left\{(\mathrm{G}_x,\ \mathrm{B}_3), \left(\mathrm{B}_3, \mathrm{G}_y\right), x,y=1,2,\ldots,7\right\}$$

所以$|D|=27$（如果你不相信我，就把它們列出來）。

然後

$$\mathrm{P}(C\mid D)=\frac{\mathrm{P}(C\cap D)}{\mathrm{P}(D)}=\frac{\mathrm{P}(C)}{\mathrm{P}(D)}=\frac{13/196}{27/196}=13/27$$

與上一個示例中的 $1/3$ 的答案相比，為什麼這個答案更接近於 $1/2$，您有什麼直覺嗎？

條件概率的屬性 - 條件概率只是有限樣本空間上事件的概率。因此，條件概率具有類似於 1.3.3 中的概率公理的性質。為了完整起見，我們在此處列出。

$\left(1'\right)$ $0\leq\mathrm{P}(A\mid C)\leq 1$

$\left(2'\right)$ $\mathrm{P}(\Omega\mid C)=1$

$\left(3'\right)$ 如果 A 和 B 不相交，則 $\mathrm{P}(A\cup B\mid C)=\mathrm{P}(A\mid C)+\mathrm{P}(B\mid C)$。

$\left(4'\right)$ 如果 $A_1, A_2, \ldots$ 都是不相交的，那麼 $\mathrm{P}\left(\bigcup_{i=1}^{\infty} A_i\mid C\right)=\sum_{i=1}^{\infty}\mathrm{P}\left(A_i\mid C\right)$

subsubsection {獨立}

我們現在討論獨立性 .[8] 的概念以及它與條件概率的關係。

簡而言之，任何不相關的事件都是獨立的。例如，

A：明天火星會下雨，B : 一枚硬幣落在人頭上。

定義：當 $\mathrm{P}(A\cap B)=\mathrm{P}(A)\mathrm{P}(B)$ 時，A 和 B 是獨立的。

[8] 7 月記住這個部分 4^{th} ！示例：如果 $\mathrm{P(A)}=0.2$，$\mathrm{P}(B)=0.5$，並且 $\mathrm{P}(A$ 和 $B)=\mathrm{P}(A)\mathrm{P}(B)=0.1$，則 A 和 B 是獨立的。

備註：如果 $\mathrm{P}(A)=0$，則 A 獨立於任何其他事件。

備註：事件不必在物理上無關才能獨立。

示例：擲骰子，並定義 $A=\{2,4,6\}$ 和 $B=\{1,2,3,4\}$。那麼$A\cap B=$$\{2,4\}$，使得$\mathrm{P}(A)=1/2$，$\mathrm{P}(B)=2/3$，並且$ (A B)=1 / 3$。

那麼 $\mathrm{P}(A)\mathrm{P}(B)=1/3=\mathrm{P}(A\cap B)\Rightarrow A,B$ 是獨立的。

下一個定理對獨立性提供了更自然的解釋。

定理：假設 $\mathrm{P}(B)>0$。則 A 和 B 是獨立的 $\Leftrightarrow\mathrm{P}(A\mid B)=\mathrm{P}(A)$。

證明：根據獨立性和條件概率的定義，

$$\begin{aligned}A,B\text{ independent } &\Leftrightarrow\mathrm{P}(A\cap B)=\mathrm{P}(A)\mathrm{P}(B)\\&\Leftrightarrow\mathrm{P}(A\cap B)/\mathrm{P}(B)=\mathrm{P}(A)\\&\Leftrightarrow\mathrm{P}(A\mid B)=\mathrm{P}(A).\end{aligned}$$

備註：定理說如果A和B是獨立的，那麼A的概率不取決於B是否出現。

重要提示：不要將獨立性與不相交混為一談！事實上，以下定理表明獨立性和不相交性幾乎是對立的。直觀地說，如果 A 和 B 不相交並且出現了 A，那麼你就有了 B 不可能出現的信息——所以 A 和 B 不可能是獨立的！

定理：如果$\mathrm{P}(A)$和$\mathrm{P}(B)>0$，則A和B不能同時獨立和不相交。

證明：假設A、B不相交（即$A \cap B=\emptyset$）。那麼$\mathrm{P}(A \cap B)=0<\mathrm{P}(A) \mathrm{P}(B)$。因此，$A$、$B$ 不是獨立的。同樣，獨立性意味著 A 和 B 不是不相交的。

以下定義將獨立性的概念擴展到兩個以上的事件。

定義：A, B, C是獨立的$\Leftrightarrow$

a. $\mathrm{P}(A \cap B \cap C)=\mathrm{P}(A) \mathrm{P}(B) \mathrm{P}(C)$ 和

b. 所有對都是獨立的：

$$\mathrm{P}(A \cap B)=\mathrm{P}(A) \mathrm{P}(B), \mathrm{P}(A \cap C)=\mathrm{P}(A) \mathrm{P}(C), \text { and } \mathrm{P}(B \cap C)=\mathrm{P}(B) \mathrm{P}(C)$$

備註：小心！請注意，條件 (a) 本身是不夠的。示例：設 $\Omega=\{1,2, \ldots, 8\}$，其中每個元素的概率為 $1/8$。定義事件 $A=\{1,2,3,4\}$、$B=\{1,5,6,7\}$ 和 $C=\{1,2,3,8\}$。

a. $A \cap B \cap C=\{1\}$。因此，$\mathrm{P}(A \cap B \cap C)=\mathrm{P}(A) \mathrm{P}(B) \mathrm{P}(C)=1/8$，所以 (a) 滿意。但是，(b) 不是...

b. $A \cap B=\{1\}$。$\mathrm{P}(A \cap B)=1/8 \neq 1/4=\mathrm{P}(A) \mathrm{P}(B)$。所以，$A$ 和 B 不是獨立的！

保持警惕，因為，如下例所示，(b) 本身也不夠！

示例：設 $\Omega=\{1,2,3,4\}$（每個元素 wp 1/4）。定義事件 $A=\{1,2\}$、$B=\{1,3\}$ 和 $C=\{1,4\}$。

b. $\mathrm{P}(A \cap B)=1/4=\mathrm{P}(A) \mathrm{P}(B)$。與 A、C 和 B、C 的處理相同。所以（ b ）沒問題。但是（ a ）不是...

c. $\mathrm{P}(A \cap B \cap C)=1/4 \neq 1/8=\mathrm{P}(A) \mathrm{P}(B) \mathrm{P}(C)$.

一般定義：我們可以將獨立性的概念擴展到兩個以上的事件。實際上，$A_1, A_2 \ldots, A_n$ 是獨立的 $\Leftrightarrow$

$$\mathrm{P}\left(A_1 \cap A_2 \cap \cdots \cap A_n\right)=\mathrm{P}\left(A_1\right) \mathrm{P}\left(A_2\right) \cdots \mathrm{P}\left(A_n\right)$$

並且由 $\left\{A_1 、 A_2 \ldots 、 A_n\right\}$ 的任何子集組成的 A_i 是獨立的。

定義：如果對一個實驗進行了 n 次試驗，使得一個試驗的結果獨立於其他試驗的結果，則稱它們是獨立試驗。

示例：獨立翻轉三個硬幣。

a. $\mathrm{P}\left(1^{\text {st }} \text { 硬幣是 } \mathrm{H}\right)=1/2$。不要擔心另外兩個硬幣，因為它們獨立於 $1^{\text {st }}$。

b. $\mathrm{P}\left(1^{\text {st }} \text { coin } \mathrm{H}, 3^{\text {rd }} \mathrm{T}\right)=\mathrm{P}\left(1^{\text {st }} \text { 硬幣 } \mathrm{H}\right) \mathrm{P}\left(3^{\text {rd }} \mathrm{T}\right)=1/4$。

備註：對於獨立試驗，只需將個體概率相乘即可。

示例：無限次翻轉硬幣（每次翻轉都獨立於其他翻轉）。

$$\begin{aligned}
p_n & \equiv \mathrm{P}\left(1^{\text {st }} \mathrm{H} \text { on } n^{\text {th }} \text { trial }\right) \\
& =\mathrm{P}(\underbrace{\mathrm{T} T \cdots \mathrm{T}}_{n-1} \mathrm{H}) \\
& =\underbrace{\mathrm{P}(\mathrm{T}) \mathrm{P}(\mathrm{T}) \cdots \mathrm{P}(\mathrm{T})}_{n-1} \mathrm{P}(\mathrm{H})=1/2^n,
\end{aligned}$$

以便

$$\mathrm{P}(\mathrm{H} \text { eventually })=\sum_{n=1}^{\infty} p_n=\sum_{n=1}^{\infty} 2^{-n}=1$$

1.9 貝葉斯定理

貝葉斯定理允許我們根據可用的某些信息來調整事件的概率。當我們能夠將樣本空間劃分為不同的部分（即分區）時，它特別有用。隨後的討論部分遵循 Meyer 4] 中的介紹。

定義：樣本空間的分區Ω 將樣本空間分成不相交的、包羅萬象的子集。具體來說，事件 $A_1, A_2, \ldots, A_n$ 形成 Ω 的一個分區，如果：

a. $A_1, A_2, \ldots, A_n$ 是不相交的。

b. $\bigcup_{i=1}^{n} A_i = \Omega$。

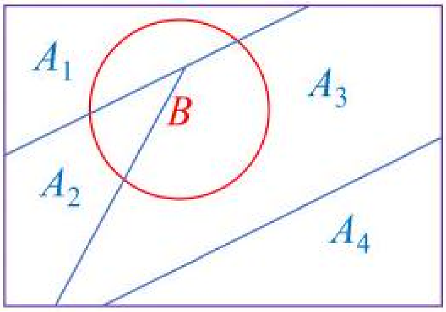
備註：執行實驗時，恰好出現 A_i 之一。

示例：A 和 $\bar{A}$ 形成一個分區，對於任何事件 A

示例："元音"和"輔音"形成字母的分區（如果您假設只有 a,e,i,o,u 是元音）。

備註：選擇所有的 A_i 使得 $\mathrm{P}\left(A_i\right) > 0$ 通常很方便，但這實際上不是必需的。

假設 A_1、A_2、 ... 、A_n 形成樣本空間 Ω 的分區，並且 B 是某個任意事件。然後（見圖）$B = \bigcup_{i=1}^{n}\left(A_i \cap B\right)$。



因此，如果 $A_1, A_2, \ldots, A_n$ 是一個分區，那麼我們可以將 B 分解為 A_i 的片段，

$$
\begin{aligned}
\mathrm{P}(B) &= \mathrm{P}\left(\bigcup_{i=1}^{n}\left(A_i \cap B\right)\right) \\
&= \sum_{i=1}^{n} \mathrm{P}\left(A_i \cap B\right) \quad \left(\text{ since } A_1, A_2, \ldots, A_n \text{ are disjoint}\right) \\
&= \sum_{i=1}^{n} \mathrm{P}\left(A_i\right) \mathrm{P}\left(B \mid A_i\right) \quad (\text{definition of conditional probability}).
\end{aligned}
$$

這就是全概率法則——這不僅是個好主意，還是法則！例子：$\mathrm{P}(B) = \mathrm{P}(A)\mathrm{P}(B \mid A) + \mathrm{P}(\bar{A})\mathrm{P}(B \mid \bar{A})$，我們在 §1.7 的示例之一中看到

示例：假設我們有 10 名來自聰明大學的學生和 20 名來自天才大學的學生參加考試。聰明的學生有 95 $的機會通過考試，但有天賦的學生不知何故只有 50 $的機會通過考試。隨機選擇一個學生，並確定他/她通過的概率。

根據全概率定律，

$$
\begin{aligned}
\mathrm{P}(\text{ passes }) &= \mathrm{P}(\mathrm{C})\mathrm{P}(\text{ passes } \mid \mathrm{C}) + \mathrm{P}(\mathrm{G})\mathrm{P}(\text{ passes } \mid \mathrm{G}) \\
&= (1/3)(0.95) + (2/3)(0.5) = 0.65. \quad \square
\end{aligned}
$$

這是全概率定律的一個重要直接結果，我們將同時陳述和證明（因為我們可以）。

貝葉斯定理：如果A_1、A_2、 ... 、A_n構成樣本空間Ω的一個分區，並且B是任意事件，那麼

$$
\mathrm{P}\left(A_j \mid B\right) = \frac{\mathrm{P}\left(A_j \cap B\right)}{\mathrm{P}(B)} = \frac{\mathrm{P}\left(A_j\right)\mathrm{P}\left(B \mid A_j\right)}{\sum_{i=1}^{n} \mathrm{P}\left(A_i\right)\mathrm{P}\left(B \mid A_i\right)}
$$

- $\mathrm{P}\left(A_j\right)'$ 稱為先驗概率（“在 B 之前”）。
- $\mathrm{P}\left(A_j \mid B\right)$ 被稱為後驗概率（“after B ”)。
- $\mathrm{P}\left(A_j \mid B\right)$ 的總和為 1，從等式的右側可以看出。這就是為什麼我們有看起來很有趣的分母。
- 結果 - 可追溯到 1763 年 - 以牧師 Thomas Bayes 的名字命名。

示例：有兩盒襪子。盒子 A 包含一隻紅色襪子和一隻藍色襪子。框 B 包含兩個紅色。你的朋友隨機挑選了一個盒子，然後給你看一個隨機的襪子，結果是紅色的。它來自方框 A 的概率是多少？

讓R 表示抽出一隻紅色襪子的事件。相關分區為$\{A, B\}$。

$$\begin{aligned}
\mathrm{P}(A \mid R) & =\frac{\mathrm{P}(A) \mathrm{P}(R \mid A)}{\mathrm{P}(A) \mathrm{P}(R \mid A)+\mathrm{P}(B) \mathrm{P}(R \mid B)} \\
& =\frac{(0.5)(0.5)}{(0.5)(0.5)+(0.5)(1.0)}=1 / 3
\end{aligned}$$

請注意方框 A 的概率如何從 0.5（之前）變為 $1/3$（後驗）——它肯定受到我們收到的信息的影響。

示例：兩位政治候選人正在辯論。候選人 A 被問到 60% 的問題，而候選人 B 只被問到 40%（出於某種原因）。A 很可能在 20 \$的時間內做出愚蠢的回答，而 B 在 50 \$的時間裡做出愚蠢的回答。其中一位候選人被問到一個問題並做出愚蠢的回答。是A的概率是多少？

（我們預計概率為 $<60\%$，因為愚蠢的答案有利於 B。）

讓 D 表示該人做出愚蠢回答的事件。相關分區為$\{A, B\}$。然後，

$$\begin{aligned}
\mathrm{P}(A \mid D) & =\frac{\mathrm{P}(A) \mathrm{P}(D \mid A)}{\mathrm{P}(A) \mathrm{P}(D \mid A)+\mathrm{P}(B) \mathrm{P}(D \mid B)} \\
& =\frac{(0.6)(0.2)}{(0.6)(0.2)+(0.4)(0.5)}=0.375
\end{aligned}$$

示例：您是舊電視節目讓我們做個交易的參賽者！三扇門中的一扇後面是一輛車；另外兩個後面是山羊。你選擇 1 號門。Monty Hall 打開 2 號門，露出一隻山羊。Monty 為您提供了切換到 3 號門的機會。您應該怎麼做？（注意：如果汽車確實在您的 1 號門後面，蒙蒂會隨機選擇向您顯示 2 號門或 3 號門。）

根據貝葉斯，我們有

$$\begin{aligned}
& \mathrm{P} \text{ Car behind Door } 1 \mid \text{ Monty shows Door } 2) \\
& \quad=\frac{\mathrm{P}(\text{ Monty shows Door } 2 \mid \text{ Car behind } 1) \mathrm{P}(\text{ Car behind } 1)}{\sum_{i=1}^3 \mathrm{P}(\text{ Monty shows Door } 2 \mid \text{ Car behind } i) \mathrm{P}(\text{ Car behind } i)} \\
& \quad=\frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3}+0 \cdot \frac{1}{3}+1 \cdot \frac{1}{3}}=1 / 3 .
\end{aligned}$$

另一方面，

$$\begin{aligned}
& \mathrm{P}(\text{ Car behind Door } 3 \mid \text{ Monty shows Door } 2) \\
& \quad=\frac{\mathrm{P}(\text{ Monty shows Door } 2 \mid \text{ Car behind } 3) \mathrm{P}(\text{ Car behind } 3)}{\sum_{i=1}^3 \mathrm{P}(\text{ Monty shows Door } 2 \mid \text{ Car behind } i) \mathrm{P}(\text{ Car behind } i)} \\
& \quad=\frac{1 \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3}+0 \cdot \frac{1}{3}+1 \cdot \frac{1}{3}}=2 / 3 .
\end{aligned}$$

因此，謹慎的行動是切換到 3 號門！

如果您不完全相信這一點，那麼您並不孤單。但是考慮一下如果有 1000 扇門，而蒙蒂透露了其中的 998 扇，你會怎麼做——當然，你會從你的門切換到剩下的那扇門！

1.10 練習

1. ($1.2) 假設我們定義如下線段：$U=[0,2], A=[0.5,1]$, and $B=[0.5,1.5]$。$\bar{A}$、$\overline{A\cup B}$ 和 $A\cup\bar{B}$ 是什麼？

2. ($1.2) 集合$S$ 的所有子集的集合稱為S 的冪集，記為2^S。例如，如果 $S=\{a,b\}$，則 $2^S=\{\emptyset,\{a\},\{b\},S\}$。如果 $|S|=n$，求 $\left|2^S\right|$ 3。 (1.2) 證明德摩根定律之一：$\overline{A\cap B}=\bar{A}\cup\bar{B}$。您可以使用維恩圖或進行數學論證。

3. ($8) 如果$f(x)=\ln(2x-3)$，求導數$f'(x)$。

4. ($1.2) 如果$f(x)=\cos(1/x)$，求導數$f'(x)$。

5. ($1.2) 如果$f(x)=\sin(\ln(x)\ )$，求導數$f'(x)$。

6. ($1.2) 求$\int_0^1(2x+1)^2dx$。

7. ($1.2) 求 $\int_1^2 e^{2x}dx$。

8. ($) 1.2) 求 $\int_1^2\ln(x)dx$。

9. ($1.2) 計算限額

$$\lim_{x\to 0}\frac{e^x-1}{\sin(x)}.$$

（提示：如果這個問題讓你生病了，你必須去...？）

11. ($1.2) 計算限額

$$\lim_{x\to 0}\frac{\sin(x)-x}{x}.$$

12. ($1.2) 用歸納法證明眾所周知的結果$\sum_{k=1}^n k=n(n+1)/2$。

13. ($1.3) 一個盒子包含四顆彈珠（一顆紅色、兩顆綠色和一顆紫色）。

a. 考慮一個實驗，從盒子裡取出一顆彈子，然後把它放回盒子裡，然後從盒子裡取出第二顆彈子。什麼是樣本空間？

b. 當第二個彈子被拉出而不更換第一個彈子時，重複上述操作。

14. ($1.3) 設$A$、$B$和$C$為三個事件，假設樣本空間$\Omega=A\cup B\cup C$。

a. 找出B 和C 發生但A 不發生的事件的表達式。

b. 為僅發生 B 的事件找到一個表達式。

15. ($1.3) 好像無事可做，擲骰子 6,000 次。（我想你可以在 Excel 或其他軟件中做到這一點。）每個數字出現多少次？你預計大約有多少？

16. ($1.3) 有趣的運動！擲硬幣直到看到序列 HT。嘗試了多少次？（例如，序列“TTHHHT”在你停止之前需要投擲六次。）現在投擲一枚硬幣，直到你看到連續出現兩個 H。那花了多長時間？（例如，序列 TTHTTTHH” 將需要八次拋擲。）想法？

17. ($1.3) 讓我們看看情侶旅遊俱樂部會員的偏好。事實證明，60 $的人喜歡安道爾，60 $的人喜歡拉斯維加斯，70 $的人偏愛巴黎。此外，30 $ %$ 像安道爾和拉斯維加斯，40 $ %$ 安道爾和巴黎，50 $ %$ 拉斯維加斯和巴黎，以及 20 $ %$ 三者。(a) 詹姆斯和喬伊斯是俱樂部裡的一對隨機情侶。他們喜歡去三個目的地中的至少一個的概率是多少？

b. Ricky 和 Nelson 是俱樂部中的另一對隨機情侶。他們是旅行的人，他們在世界各地做了很多停留。找出他們恰好享受三個目的地中的兩個的概率。

18. (1.3 證明 Bonferroni 不等式：$\mathrm{P}(A\cap B)\geq\mathrm{P}(A)+\mathrm{P}(B)-1$。

19. ($1.4) 假設我擲硬幣 3 次。我得到兩個頭和一個尾的概率是多少？

20. (1.5) 如果我們需要，可以從字母表中組成多少個五個字母的單詞

a. 第二個和第四個字母是元音（a, e, i, o, u）？

b. 至少一個元音？

21. ($1.5 TRUE or FALSE? $P_{n,k} = \binom{n}{k} k\,!$ for all appropriate (n, k).

22. ($1.5 考慮一支正好有九名球員的棒球隊。

a. 你有多少種方法可以填補擊球順序的前四個位置？

b. 有多少種方式是史密斯先擊球的？

23. (1.5) 假設 12 個小丑正試圖進入一輛最多只能容納 7 個小丑的小汽車。七個小丑有多少種可能的選擇？

24. ($1.5) 用你喜歡的語言編寫一個計算機程序來計算組合。在 $C_{100,50}$ 上演示您的程序。

25. (1.6) 擲三個骰子。三個骰子中的兩個恰好出現相同數字的概率是多少？

26. ($1.6) “SYZYGY”中的字母有幾種排列方式？（你知道這個詞是什麼意思嗎？）

27. ($1.6) 橋牌包含來自標準牌組的 13 張牌。找出橋牌包含的概率…

a. 正好是三個國王。

b. 相同花色的所有 13 張牌。

28. ($1.6) 一個六面骰子被擲了七次。尋找

a. $P\left(3'\right.$ 至少出現一次)。

b. P（每張臉至少出現一次）。

29. (1.6) 格魯布諾星球有 50 天年。

a. 假設房間裡有兩個格魯布諾人。他們生日相同的概率是多少？(b) 假設房間裡有 3 個格魯布諾人。（他們很大，所以房間越來越擁擠。）他們中至少有兩個生日相同的概率是多少？

30. ($1.6) 婚禮請柬！我們有四張邀請卡和隨附的信封。但是哎呀——我們隨機混合了卡片和信封！我們得到至少一個正確匹配的概率是多少？

31. (1.6) 從標準套牌中選擇六張牌。你得到三對的概率是多少？

32. ($1.6) 快艇！擲五個骰子。

a. 你觀察到所謂的長度為 5 的“大直線”的概率是多少？（例如，這樣的順子來自拋擲 $3, 2, 6, 4, 5$。請注意，您只能從 1 或 2 開始獲得順子。）

b. 你正好看到兩對和一擲骰子的概率是多少？（請注意，這排除了滿堂彩的可能性。此外，也許令人驚訝的是，這不會讓您在 Yahtzee 中獲得任何分數！）

33. ($1.7) 你有 20 顆彈珠 - 12 顆紅色和 8 顆藍色。假設您隨機選擇其中兩個彈珠（假設您沒有丟失彈珠）。概率是多少…

a. 兩者都是紅色的？

b. 一個是紅色的，一個是藍色的？

34. ($1.7) 證明：如果 $\mathrm{P}(A \mid B) > \mathrm{P}(A)$，則 $\mathrm{P}(B \mid A) > \mathrm{P}(B)$。

35. ($1.7) 假設 A 和 B 是獨立的，$\mathrm{P}(A) = 0.4$，並且 $\mathrm{P}(A \cup B) = 0.6$。求 $\mathrm{P}(B)$

36. (1.7) 假設 $\mathrm{P}(A) = 0.4$，$\mathrm{P}(A \cup B) = 0.7$，並且 $\mathrm{P}(B) = x$。

a. 對於 x 的哪一個選擇是 A 和 B 不相交的？

b. 獨立？

37. ($1.7) 如果$\mathrm{P}(A)=\mathrm{P}(B)=\mathrm{P}(C)=0.6$，並且$A$、$B$和$C$是獨立，找出恰好出現 A、B 和 C 之一的概率。

38. ($1.7) 證明：$C$ 和 D 是獨立的 $\Leftrightarrow$ C 和 $\bar{D}$ 是獨立的。

39. (1.8 考慮兩個盒子。盒子 I 包含一個藍色大理石和一個白色大理石。盒子 II 包含兩個藍色和一個白色。隨機選擇一個盒子，並從選定的盒子中隨機抽取一個大理石。

a. 找到 P（大理石是藍色的）。

b. 假設大理石是白色的，從方框 I 中選擇大理石的概率是多少？

40. ($1.8) 我有一枚公平的硬幣和一枚雙頭硬幣。(a) 我隨機選擇一個，當我翻轉它時，它會出現正面。硬幣公平的概率是多少？

b. 我擲同一枚硬幣，再次出現正面。同樣的問題。

c. 我擲硬幣八次，得到 HHHHHHHT。同樣的問題。

41. (1.8) 斯普林菲爾德市的警察具有極好的判斷力 - 以至於

P（任何受審的被告實際上都是有罪的）$=0.99$。

斯普林菲爾德的法院系統也相當不錯。事實上，在任何審判中，

P（如果被告實際上是無辜的，則陪審團宣告被告無罪）$=0.95$

和

P(如果被告確實有罪，則陪審團定罪)$=0.95$。

找到 P（被告是無辜的|陪審團釋放了他）。

42. ($1.8) 一個棘手的問題！三名囚犯 A、B 和 C 被獄卒告知，其中一名被隨機選擇處決，另外兩名將被釋放. 犯人 A 要求獄卒私下告訴他哪些獄友將被釋放（假設獄卒將在 B 和 C 之間隨機化，如果 $ $ 是不幸的靈魂）。A 聲稱洩露這些信息不會有任何害處，因為他已經知道這兩者中至少有一個會被釋放。獄卒拒絕辯稱，如果 A 知道，那麼 A 被處決的概率將從 $1/3$ 上升到 $1/2$（即只剩下兩個囚犯）。誰是正確的？

ryNoSmBk

ryCh 2 【第 2 章】

2.0.0.1 隨機變量

本章介紹了隨機變量的重要概念，它可用於計算概率，以及中心性和變異性的度量。

2.1 - 介紹和定義

$2.2 - 離散隨機變量

2.3 - 連續隨機變量

2.4 - 累積分佈函數

2.5 - 遠大的期望

2.6 - 矩生成函數

$2.7 - 一些概率不等式

$2.8 - 隨機變量的函數

$2.9 - 練習

2.1 簡介和定義

定義：隨機變量（RV）X 是一個從樣本空間到實線的函數，$X : \Omega \to \mathbb{R}$。

示例：翻轉兩個硬幣。樣本空間為 $\Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$。假設$X$ 是對應於H 數量的RV。然後（抑制無關的“$\{.\}$”符號），

$$X(\mathrm{TT}) = 0, \quad X(\mathrm{HT}) = X(\mathrm{TH}) = 1, \quad X(\mathrm{HH}) = 2.$$

這導致

$$\mathrm{P}(X = 0) = \frac{1}{4}, \quad \mathrm{P}(X = 1) = \frac{1}{2}, \quad \mathrm{P}(X = 2) = \frac{1}{4}.$$

符號：X、Y、Z、U、V、W 等大寫字母通常代表 RV。x、y、z、u、v、w 等小寫字母通常代表 RV 的特定值。因此，您可以說諸如 $\mathrm{P}(X=x)$ 這樣的量。

示例：假設 X 是兩次擲骰子的總和。樣本空間是擲兩個骰子的所有方法的集合。然後，例如，$(4,6)$ 是來自樣本空間的結果，當然 $X((4,6))=10$。此外，我們可以計算每個可能結果的概率：

$$\mathrm{P}(X=x)=\begin{cases}1/36 & \text{if } x=2\\ 2/36 & \text{if } x=3\\ \vdots & \\ 6/36 & \text{if } x=7\\ \vdots & \\ 1/36 & \text{if } x=12\\ 0 & \text{otherwise.}\end{cases}$$

例：擲硬幣。

$$X\equiv\begin{cases}0 & \text{if T}\\ 1 & \text{if H}\end{cases}$$

示例：擲骰子。

$$Y\equiv\begin{cases}0 & \text{if }\{1,2,3\}\\ 1 & \text{if }\{4,5,6\}\end{cases}$$

出於所有意圖和目的，RV 的 X 和 Y 是相同的，因為 $\mathrm{P}(X=0)=\mathrm{P}(Y=0)=\frac{1}{2}$ 和 $\mathrm{P}(X=1)=\mathrm{P}(Y=1)=\frac{1}{2}$

示例：在 0 和 1 之間隨機選擇一個實數。這個實驗有無數個“同樣可能”的結果。

結論：P（我們選擇單個點x）$=\mathrm{P}(X=x)=0$，信不信由你！但請注意 $\mathrm{P}(X\leq 0.65)=0.65$，並且 $\mathrm{P}(X\in[0.3,0.7])=0.4$。事實上，如果 A 是 $[0,1]$ 中的任意區間，那麼 $\mathrm{P}(X\in A)$ 就是 A 的長度。

定義：如果隨機變量 X 的可能值的數量是有限的或可數無限的，則 X 是離散隨機變量。否則， $\ldots$

連續隨機變量是在每一點的概率為 0 的隨機變量。

示例：擲硬幣 - 結果是 H 或 T。離散的。

示例：在 $[0,1]$ 中隨機選取一個點。連續的。

示例：您在一行中等待的時間為 0（概率為正）或某個正實數 - 一個組合的離散連續隨機變量！

2.2 離散隨機變量

定義：如果 X 是一個離散隨機變量，它的概率質量函數 (pmf) 是 $f(x)\equiv\mathrm{P}(X=x)$，對於所有 x。

請注意，如果 X 是離散的，則對於所有 x，$0\leq f(x)\leq 1$，並且 $\sum_x f(x)=1$。

示例：翻轉兩個硬幣。設 X 為正面數量。

$$f(x)=\begin{cases}1/4 & \text{if } x=0 \text{ or } 2\\ 1/2 & \text{if } x=1\\ 0 & \text{otherwise.}\end{cases}$$

示例/定義：假設 X 可以等於 $1,2,\ldots,n$，每個概率為 $1/n$，即 $f(i)=1/n, i=1,2,\ldots,n$。那麼我們說X 具有離散均勻$\{1,2,\ldots,n\}$ 分佈。

示例：離散 RV 可以具有任何值。例如，讓 X 表示庫存策略的可能利潤，其中 $f(-5.1)=0.2($ 虧損 $)$，$f(1.3)=0.5$（盈虧平衡）和 \$f(11) = 0.3 \$（大筆錢）。

示例/定義：令 X 表示來自 n 個獨立試驗的"成功"數，使得每次試驗的成功概率為 $p(0\leq p\leq 1)$。然後 X 具有二項分佈，參數為 n 和 p。這些試驗被稱為伯努利試驗，以 Jakob Bernoulli $(1655-1705)$ 命名。

符號：$X\sim \operatorname{Bin}(n,p)$。

示例：擲骰子三個獨立的時間。找到 P（正好是兩個 6）。

我們將每次拋擲解釋為伯努利試驗，其中 6 表示成功，其他任何 $(1,2,3,4,5)$ 都是失敗。所有三個試驗都是獨立的，並且 P(success)$=1/6$ 不會因試驗而異。

令 $X=$ 為 6 的個數。然後 $X\sim \operatorname{Bin}\left(3,\frac{1}{6}\right)$。

定理：如果 $X\sim \operatorname{Bin}(n,p)$，則在 n 次試驗中 k 次成功的概率為

$$\mathrm{P}(X=k)=\binom{n}{k}p^kq^{n-k},\quad k=0,1,\ldots,n,\quad \text{with } q=1-p$$

證明：考慮成功和失敗的特定順序：

$$\underbrace{\mathrm{SS}\cdots\mathrm{S}}_{k \text{ successes}}\ \underbrace{\mathrm{FF}\cdots\mathrm{F}}_{n-k \text{ fails}}\left(\text{probability } =p^kq^{n-k}\right).$$

The number of ways to arrange the sequence is $\binom{n}{k}$. Done. Example (cont'd): Back to the dice example, where $X\sim \operatorname{Bin}\left(3,\frac{1}{6}\right)$, and we want P(exactly two 6 's). We have $n=3, k=2, p=1/6$, and $q=5/6$. Then

$$\mathrm{P}(X=2)=\binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1=\frac{15}{216}$$

事實上，我們可以得到整個 pmf，

示例：擲兩個骰子並得到總和。重複這個實驗 12 次。求 P（總和將是 7 或 11 正好 3 次）。

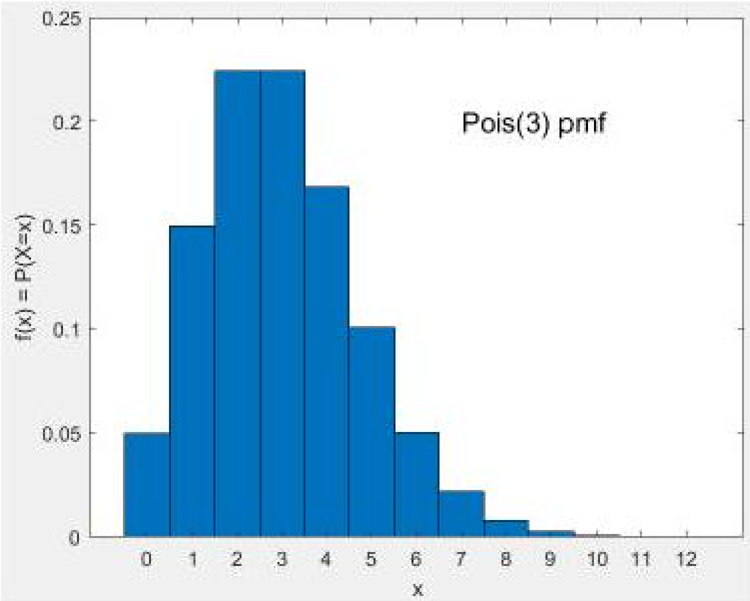
讓 $X=$ 我們得到 7 或 11 的次數。然後

$$\mathrm{P}(7 \text{ or } 11)=\mathrm{P}(7)+\mathrm{P}(11)=\frac{6}{36}+\frac{2}{36}=\frac{2}{9}.$$

所以 $X\sim \operatorname{Bin}(12,2/9)$，然後

$$\mathrm{P}(X=3)=\binom{12}{3}\left(\frac{2}{9}\right)^3\left(\frac{7}{9}\right)^9$$

定義：如果$\mathrm{P}(X=k)=e^{-\lambda}\lambda^k/k!, k=0,1,2,\ldots,$ and \$>0 \$，我們說 X 具有參數 λ 的泊松分佈，因此以 Siméon Denis Poisson $(1781-1840)$ 命名



符號：$X\sim \mathrm{Pois}(\lambda)$

示例：假設一杯餅乾麵團中的葡萄乾數量為 Pois (10)。找出一杯麵團至少有四個葡萄乾的概率。

$$\begin{aligned}\mathrm{P}(X\geq 4)&=1-\mathrm{P}(X=0,1,2,3)\\&=1-e^{-10}\left(\frac{10^0}{0!}+\frac{10^1}{1!}+\frac{10^2}{2!}+\frac{10^3}{3!}\right)\\&=0.9897.\quad\square\end{aligned}$$

2.3 連續隨機變量

示例：在 0 和 1 之間隨機選取一個點 X，並定義連續函數

$$f(x)=\begin{cases}1 & \text{if } 0\leq x\leq 1\\ 0 & \text{otherwise}\end{cases}$$

直觀地說，如果 $0\leq a\leq b\leq 1$，那麼

$$\mathrm{P}(a<X<b)=\text{ the "area" under } f(x), \text{ from } a \text{ to } b=b-a.$$

定義：假設 X 是一個連續的 RV。魔術函數 $f(x)$ 是概率密度函數 (pdf) 如果

- $\int_{\mathbb{R}} f(x)dx=1$ （$f(x)$ 下的面積為 1 ），
- $f(x)\geq 0,\forall x$ （函數總是非負的），和
- 如果 $A\subseteq\mathbb{R}$，則 $\mathrm{P}(X\in A)=\int_A f(x)dx$（$X$ 在 A 中的概率）。

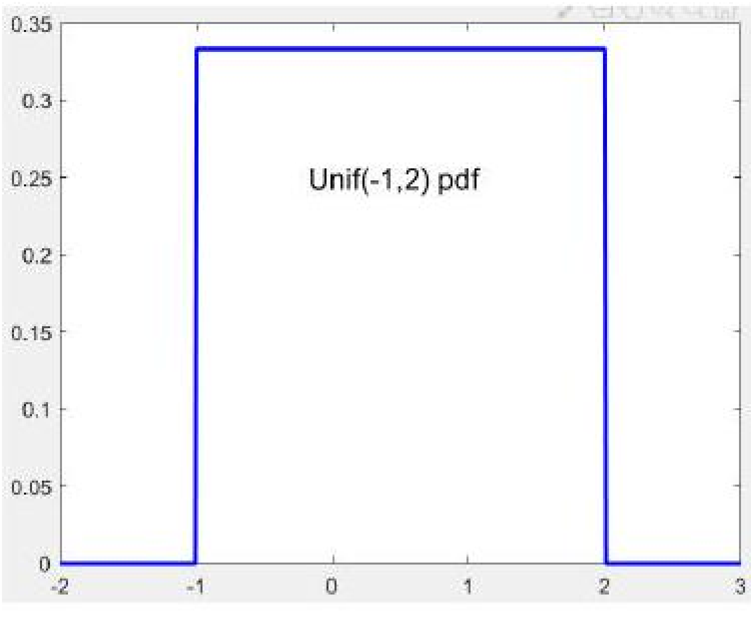
備註：如果X是一個連續的RV，則$\mathrm{P}(a < X < b) = \int_a^b f(x)dx$。作為一種特殊情況，任何單個點的概率為零，即$\mathrm{P}(X = a) = \int_a^a f(x)dx = 0$。這意味著 $\mathrm{P}(a < X < b) = \mathrm{P}(a \le X \le b)$，所以我們可以對連續情況的不等式有點鬆懈。

備註：請注意 $f(x)$ 表示 pmf（離散情況）和 pdf（連續情況）- 但它們是不同的：

- 如果 X 是離散的，則 $f(x) = \mathrm{P}(X = x)$ 並且我們必須有 $0 \le f(x) \le 1$。
- 如果X 是連續的，那麼
- $f(x)$ 不是概率，但它用於計算概率。
- 相反，將 $f(x)dx$ 視為 $\doteq \mathrm{P}(x < X < x + dx)$。
- 必須有 $f(x) \ge 0$（可能還有 > 1）。
- 通過積分 $\int_A f(x)dx$ 計算事件 A 的概率。

示例：如果 X “等可能”在 a 和 b 之間，則 X 在 (a, b) 上具有均勻分佈。

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



符號：$X \sim \mathrm{Unif}(a, b)$。

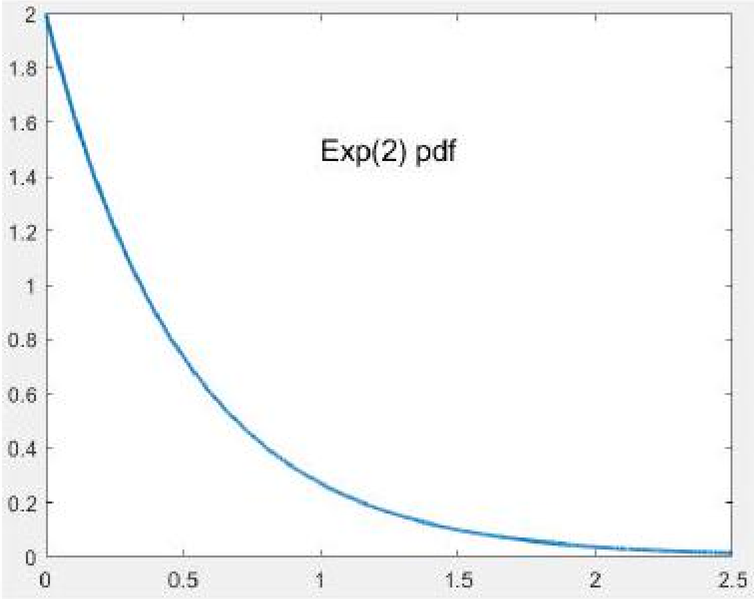
備註：注意 $\int_{\mathbb{R}} f(x)dx = \int_a^b \frac{1}{b-a}dx = 1$（根據需要）。

示例：如果 $X \sim \mathrm{Unif}(-2, 8)$，那麼

$$\mathrm{P}(-1 < X < 6) = \int_{-1}^{6} \frac{1}{8-(-2)}dx = 0.7$$

示例：X 具有參數 $\lambda > 0$ 的指數分佈，如果它有 pdf

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & \text{if } x\geq 0\\ 0 & \text{otherwise}\end{cases}$$



符號：$X\sim \text{Exp}(\lambda)$

備註：$\int_{\mathbb{R}} f(x)dx=\int_0^\infty \lambda e^{-\lambda x}dx=1$（根據需要）。

示例：假設 $X\sim \text{Exp}(1)$。然後

$\text{P}(X\leq 3)=\int_0^3 e^{-x}dx=1-e^{-3}.$

$\text{P}(X\geq 5)=\int_5^\infty e^{-x}dx=e^{-5}$ $\text{P}(2\leq X<4)=\text{P}(2\leq X\leq 4)=\int_2^4 e^{-x}dx=e^{-2}-e^{-4}$

$\text{P}(X=3)=\int_3^3 e^{-x}dx=0$

示例：假設 X 是帶有 pdf 的連續 RV

$$f(x)=\begin{cases}cx^2 & \text{if } 0<x<2\\ 0 & \text{otherwise}\end{cases}$$

首先，讓我們找到c。注意到 pdf 必須集成到 1 ，我們有

$$1=\int_{\mathbb{R}} f(x)dx=\int_0^2 cx^2dx=8c/3$$

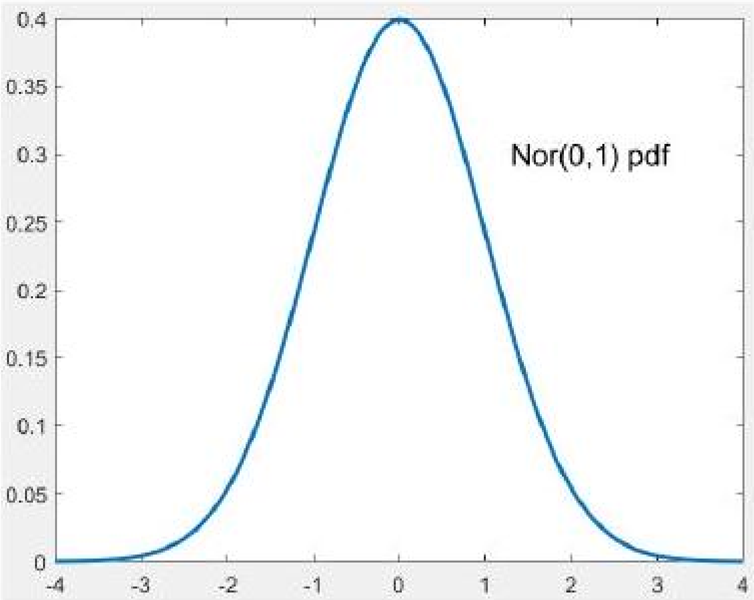
所以 $c=3/8$。現在我們可以計算任何合理的概率，例如，

$$\text{P}(0<X<1)=\int_0^1 \frac{3}{8}x^2dx=1/8$$

還有更複雜的，例如，

$$
\begin{aligned}
&\mathrm{P}(0<X<1 \mid 1/2<X<3/2) \\
=&\frac{\mathrm{P}(0<X<1 \text{ and } 1/2<X<3/2)}{\mathrm{P}(1/2<X<3/2)} \\
=&\frac{\mathrm{P}(1/2<X<1)}{\mathrm{P}(1/2<X<3/2)} \\
=&\frac{\int_{1/2}^{1} \frac{3}{8}x^2dx}{\int_{1/2}^{3/2} \frac{3}{8}x^2dx} = 7/26.
\end{aligned}
$$

示例：如果 X 的 pdf 為 $\phi(x) \equiv$ \$ e$^{\{-x}$$\{2\}$ / 2$\}$，則 X 具有標準正態分佈\$, 對於所有 $x \in \mathbb{R}$。這就是著名的“鐘形曲線”分佈。



符號：$X \sim \mathrm{Nor}(0,1)$。

2.4 累積分佈函數

定義：對於任意隨機變量X（離散或連續），累積分佈函數（cdf）為$F(x) \equiv P(X \leq x), \forall x \in \mathbb{R}$。

對於 X 離散的，

$$F(x) = \sum_{y|y\leq x} f(y) = \sum_{y|y\leq x} \mathrm{P}(X=y).$$

對於 X 連續，

$$F(x) = \int_{-\infty}^{x} f(y)dy.$$

2.4.0.1 離散 cdf 的

示例：擲硬幣兩次。讓X =個H。然後，

$$X = \begin{cases} 0 \text{ or } 2 & \text{w.p. } 1/4 \\ 1 & \text{w.p. } 1/2 \end{cases}$$

並且 cdf 是以下階躍函數：

$$F(x)=\mathrm{P}(X\le x)=\begin{cases}0 & \text{if } x<0\\ 1/4 & \text{if } 0\le x<1\\ 3/4 & \text{if } 1\le x<2\\ 1 & \text{if } x\ge 2\end{cases}$$

備註：警告！對於離散 RV，您必須注意間隔端點處的“$\le$” 與“$<$”（階躍函數跳躍的位置）。

2.4.0.2 連續 cdf’s

定理：如果X是一個連續隨機變量，那麼$f(x)=F'(x)$（假設導數存在）。

證明：$F'(x)=\frac{d}{dx}\int_{-\infty}^{x}f(t)dt=f(x)$，由基本定理微積分 $(\{1.2.2,3)$。

示例：$X\sim\mathrm{Unif}(0,1)$。pdf 和 cdf 分別是

$$f(x)=\begin{cases}1 & \text{if } 0<x<1\\ 0 & \text{otherwise}\end{cases}\quad\text{and}\quad F(x)=\begin{cases}0 & \text{if } x\le 0\\ x & \text{if } 0<x<1\\ 1 & \text{if } x\ge 1\end{cases}$$

示例：$X\sim\mathrm{Exp}(\lambda)$。回想一下，pdf 是 $f(x)=\lambda e^{-\lambda x},x>0$。那麼cdf是

$$F(x)=\int_{-\infty}^{x}f(t)dt=\begin{cases}0 & \text{if } x\le 0\\ 1-e^{-\lambda x} & \text{if } x>0\end{cases}$$

我們可以使用 cdf 找到 X 的中位數，即點 m 使得

$$0.5=\mathrm{P}(X\le m)=F(m)=1-e^{-\lambda m}$$

求解，我們得到 $m=(1/\lambda)\ln(2)$

示例：如果 $X\sim\mathrm{Nor}(0,1)$ 與 pdf $\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$，對於所有 $x\in\mathbb{R}$，則 cdf 為 $\Phi(x)\equiv\int_{-\infty}^{x}\phi(t)dt$，其中沒有封閉的形式，但可以在表格中廣泛使用（例如，我們在附錄中的表 B.1）或通過無處不在的軟件。我們將在續集中一遍又一遍地訪問這個分佈。

所有 cdf 的屬性：任何 cdf 都具有以下直觀的屬性，我們將在本書中使用這些屬性。

- $F(x)=\mathrm{P}(X\le x)$ 在 x 中不減，即 $a<b$ 意味著 $F(a)\le F(b)$。鑑於$F(x)$ 是直到x 點的概率的總和，這當然是有道理的。
- $\lim_{x\to-\infty}F(x)=0$ 和 $\lim_{x\to\infty}F(x)=1.$

定理：$\mathrm{P}(X>x)=1-F(x)$。

證明：通過補碼，$\mathrm{P}(X>x)=1-\mathrm{P}(X\le x)=1-F(x)$。

定理：$a<b\Rightarrow\mathrm{P}(a<X\le b)=F(b)-F(a)$。

證明：由於$a<b$，我們有

$$\begin{aligned}\mathrm{P}(a < X \le b) &= \mathrm{P}(X > a \cap X \le b)\\ &= \mathrm{P}(X > a) + \mathrm{P}(X \le b) - \mathrm{P}(X > a \cup X \le b)\\ &= 1 - F(a) + F(b) - 1.\end{aligned}$$

2.5 遠大前程

我們從 2.5.1 $開始，定義一個隨機變量的期望值（也就是平均值），這是 RV 的“集中趨勢”的度量。然後，我們在 $2.5.2 中引入了非常普遍的無意識統計學家定律，並使用它來定義其他 RV 度量，例如方差和矩。我們將用一些有用的近似值來結束討論，該近似值是 X 在 $2.5.3 中的複雜函數的期望值和方差。

2.5.1 期望值

定義：RV X 的平均值或期望值或平均值是

$$\mu \equiv \mathrm{E}[X] \equiv \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} x f(x) dx & \text{if } X \text{ is continuous.} \end{cases}$$

平均值表示 RV 的集中趨勢。它可以被認為是可能的 x 的加權平均值，其中權重由 $f(x)$ 給出。

示例：離散 RV X 具有伯努利分佈，參數 p，如果它具有非常簡單的 pmf $\mathrm{P}(X=0)=1-p$ 和 $\mathrm{P}(X=1)=p$。我們將 $X=1$ 視為“成功”，將 $X=0$ 視為“失敗”。$\mathrm{Bern}(p)$RV 與 $\mathrm{Bin}(1,p)$ 相同。然後

$$\mathrm{E}[X] = \sum_x x f(x) = (0 \cdot (1-p)) + (1 \cdot p) = p$$

示例：擲骰子。那麼 $X = 1, 2, \ldots, 6$，每個概率為 $1/6$，並且

$$\mathrm{E}[X] = \sum_x x f(x) = 1 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = 3.5$$

備註：上面的幾個例子表明 $\mathrm{E}[X]$ 不必等於 RV 的潛在值之一 - 不用擔心！

定義：我們說X 具有幾何分佈，參數p，如果X 定義為$\mathrm{Bern}(p)$ 試驗的次數，直到您獲得第一次成功。（例如，FFFFS 將給出 $X=5$。）然後 X 有 pmf

$$f(x) = (1-p)^{x-1} p, x = 1, 2, \ldots$$

符號：$X \sim \mathrm{Geom}(p)$

擴展示例：這是一個要求我承認我不是一個非常好的籃球運動員的應用程序。假設我獨立投籃，但投中任何特定投籃的概率只有 0.4。我至少要嘗試三次才能成功投籃的概率是多少？

答案：在我第一次成功之前的嘗試次數是 $X \sim \mathrm{Geom}(0.4)$。因此，

$$\begin{aligned}\mathrm{P}(X \ge 3) &= 1 - \mathrm{P}(X \le 2)\\ &= 1 - \mathrm{P}(X=1) - \mathrm{P}(X=2)\\ &= 1 - 0.4 - (0.6)(0.4) = 0.36\end{aligned}$$

現在，讓我們找到 $X \sim \mathrm{Geom}(p)$ 的期望值。讓$q = 1-p$，我們有

$$
\begin{aligned}
\mathrm{E}[X] &= \sum_x x f(x) = \sum_{x=1}^{\infty} x q^{x-1} p \\
&= p \sum_{x=1}^{\infty} \frac{d}{dq} q^x = p \frac{d}{dq} \sum_{x=1}^{\infty} q^x \quad \text{(carefully swap derivative and sum)} \\
&= p \frac{d}{dq} \frac{q}{1-q} (\text{ geometric sum}) \\
&= p \left[\frac{(1-q) - q(-1)}{(1-q)^2} \right] = 1/p
\end{aligned}
$$

因此，由於我的籃球技術有問題，平均而言，在我第一次投籃之前，我需要 $\mathrm{E}[X] = 1/p = 1/0.4 = 2.5$ 次投籃。

示例： $X \sim \operatorname{Exp}(\lambda)$。$f(x) = \lambda e^{-\lambda x}, x \geq 0$。然後

$$
\begin{aligned}
\mathrm{E}[X] &= \int_{\mathbb{R}} x f(x) dx \\
&= \int_0^{\infty} x \lambda e^{-\lambda x} dx \\
&= -x e^{-\lambda x} \Big|_0^{\infty} - \int_0^{\infty} \left(-e^{-\lambda x}\right) dx \quad \text{(by parts)} \\
&= \int_0^{\infty} e^{-\lambda x} dx \quad (\text{ 'Hôpital rules!}) \\
&= 1/\lambda.
\end{aligned}
$$

2.5.2 LOTUS、矩和方差

下一個定理通常以神秘的綽號“無意識統計學家定律”（LOTUS）命名，並概括了我們的期望值概念。

定理 (LOTUS)：X 的函數的期望值，比如 $h(X)$，是

$$
\mathrm{E}[h(X)] = \begin{cases} \sum_x h(x) f(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} h(x) f(x) dx & \text{if } X \text{ is continuous.} \end{cases}
$$

證明：見 82.8.3

$\mathrm{E}[h(X)]$ 是 $h(x)$ 的加權函數，其中權重是 $f(x)$ 值。

備註：看起來像一個定義，但實際上是一個定理——他們稱它為 LOTUS，因為一些統計老師沒有註意並將其作為定義呈現！

示例：LOTUS 適用於任何合理的函數（假設生成的預期值是明確定義的且有限的）。以下是具有 pdf $f(x) = 3x^2, 0 \leq x \leq 1$ 的連續隨機變量的幾個示例

- $\mathrm{E}\left[X^2\right] = \int_{\mathbb{R}} x^2 f(x) dx = \int_0^1 3x^4 dx = 3/5.$
- $\mathrm{E}[1/X] = \int_{\mathbb{R}} (1/x) f(x) dx = \int_0^1 3x dx = 3/2$。
- 如果我們定義符號 $y^+ \equiv \max\{y, 0\}$，那麼我們有

$$\mathrm{E}\left[(X-0.5)^{+}\right]=\int_{\mathbb{R}}(x-0.5)^{+} f(x) d x=\int_{0.5}^{1}(x-0.5) 3 x^{2} d x=\frac{17}{64}$$

- 零件產量集成的應用

$$\begin{aligned}\mathrm{E}\left[\frac{\sin (X)}{X}\right] & =\int_{\mathbb{R}} \frac{\sin (x)}{x} f(x) d x=\int_{0}^{1} 3 x \sin (x) d x \\ & =\left.(-x \cos (x)+\sin (x))\right|_{0}^{1}=0.3012 .\end{aligned}$$

- 最後，這是一個可愛且有時有用的結果，適用於任何連續的 RV X。假設 A 是一個表現良好的集合，定義指標函數，

$$h(X)=1_{A}(X) \equiv \begin{cases}1 & \text { if } X \in A \text { occurs } \\ 0 & \text { otherwise. }\end{cases}$$

然後

$$\mathrm{E}\left[1_{A}(X)\right]=\int_{\mathbb{R}} 1_{A}(x) f(x) d x=\int_{A} f(x) d x=\mathrm{P}(A) .$$

請稍等一下。. 現在我們將討論幾個特別重要的 LOTUS 特殊情況。

定義：X 的 k^{th} 矩是

$$\mathrm{E}\left[X^{k}\right]=\begin{cases}\sum_{x} x^{k} f(x) & \text { if } X \text { is discrete } \\ \int_{\mathbb{R}} x^{k} f(x) d x & \text { if } X \text { is continuous. }\end{cases}$$

示例：假設 $X \sim \operatorname{Bern}(p)$，因此 $f(1)=p$ 和 $f(0)=q=1-p$。

$$\mathrm{E}\left[X^{k}\right]=\sum_{x} x^{k} f(x)=\left(0^{k} \cdot q\right)+\left(1^{k} \cdot p\right)=p, \text { for all } k! \text { Cool! }$$

示例： $X \sim \operatorname{Exp}(\lambda)$。$f(x)=\lambda e^{-\lambda x}, x \geq 0$。然後

$$\begin{aligned}\mathrm{E}\left[X^{k}\right] & =\int_{\mathbb{R}} x^{k} f(x) d x \\ & =\int_{0}^{\infty} x^{k} \lambda e^{-\lambda x} d x \\ & =\int_{0}^{\infty}(y / \lambda)^{k} \lambda e^{-y}(1 / \lambda) d y \quad(\text {substitute } y=\lambda x\Big) \\ & =\frac{1}{\lambda^{k}} \int_{0}^{\infty} y^{(k+1)-1} e^{-y} d y \\ & =\frac{\Gamma(k+1)}{\lambda^{k}} \quad \text { (by definition of the gamma function) } \\ & =\frac{k!}{\lambda^{k}} \square\end{aligned}$$

定義：X 的 k^{th} 中心矩為

$$\mathrm{E}\left[(X-\mu)^{k}\right]=\begin{cases}\sum_{x}(x-\mu)^{k} f(x) & X \text { is discrete } \\ \int_{\mathbb{R}}(x-\mu)^{k} f(x) d x & X \text { is continuous. }\end{cases}$$

定義：X的方差是第二個中心矩，即RVX與其均值μ的平方差的期望值。換句話說，

$$\mathrm{Var}(X) \equiv \mathrm{E}\left[(X-\mu)^2\right]$$

方差是散佈或分散的量度 - 緊密分佈的方差低，而散佈分佈的方差高。

符號：$\sigma^2 \equiv \mathrm{Var}(X)$。

定義：X 的標準差為 $\sigma \equiv +\sqrt{\mathrm{Var}(X)}$。

示例：假設 $X \sim \mathrm{Bern}(p)$，因此 $f(1)=p$ 和 $f(0)=q=1-p$。

回想一下 $\mu=\mathrm{E}[X]=p$。然後

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}\left[(X-\mu)^2\right] \\
&= \sum_x (x-p)^2 f(x) \\
&= (0-p)^2 q + (1-p)^2 p \\
&= p^2 q + q^2 p = pq(p+q) \\
&= pq. \square
\end{aligned}$$

接下來的結果證實了期望值算子可以通過X的某些線性函數，然後可以用於獲得其他期望值和方差的令人愉快的表達式。

定理：對於 X 的任何函數，例如 $h(X)$，以及常數 a 和 b，我們有 [1]

$$\mathrm{E}[ah(X)+b] = a\mathrm{E}[h(X)]+b$$

證明（只做連續案例）：通過 LOTUS，

$$\begin{aligned}
\mathrm{E}[ah(X)+b] &= \int_{\mathbb{R}} (ah(x)+b) f(x) dx \\
&= a\int_{\mathbb{R}} h(x) f(x) dx + b\int_{\mathbb{R}} f(x) dx \\
&= a\mathrm{E}[h(X)]+b.
\end{aligned}$$

推論：特別是，

$$\mathrm{E}[aX+b] = a\mathrm{E}[X]+b.$$

相似地，

$$\mathrm{E}[g(X)+h(X)] = \mathrm{E}[g(X)]+\mathrm{E}[h(X)]$$

1 “移位”（即 b ）發生 定理：這是有時更簡單的計算方差的方法：

$$\mathrm{Var}(X) = \mathrm{E}\left[X^2\right]-(\mathrm{E}[X])^2$$

證明：由公式 2.1

$$\begin{aligned} \operatorname{Var}(X) &= \mathrm{E}\left[(X-\mu)^2\right] \\ &= \mathrm{E}\left[X^2-2\mu X+\mu^2\right] \\ &= \mathrm{E}\left[X^2\right]-2\mu \mathrm{E}[X]+\mu^2 \\ &= \mathrm{E}\left[X^2\right]-\mu^2. \end{aligned}$$

示例：假設 $X \sim \operatorname{Bern}(p)$。回想一個很酷的事實，$\mathrm{E}\left[X^k\right]=p$，對於所有 $k=1,2$， $\ldots$ So

$$\operatorname{Var}(X)=\mathrm{E}\left[X^2\right]-(\mathrm{E}[X])^2=p-p^2=pq$$

示例：$X \sim \operatorname{Unif}(a, b)$，因此 $f(x)=1/(ba), a<x<b$。然後

$$\begin{aligned} \mathrm{E}[X] &= \int_{\mathbb{R}} x f(x) dx=\int_a^b \frac{x}{b-a} dx=\frac{a+b}{2}, \\ \mathrm{E}\left[X^2\right] &= \int_{\mathbb{R}} x^2 f(x) dx=\int_a^b \frac{x^2}{b-a} dx=\frac{a^2+ab+b^2}{3}, \end{aligned}$$

和

$$\operatorname{Var}(X)=\mathrm{E}\left[X^2\right]-(\mathrm{E}[X])^2=\frac{(a-b)^2}{12} \quad \text{(after algebra)}$$

通常情況下，我們需要 X 的線性函數的方差。

定理：對於常數 a 和 b，

$$\operatorname{Var}(aX+b)=a^2 \operatorname{Var}(X)$$

證明：根據方差的定義，並利用前定理，我們有

$$\begin{aligned} \operatorname{Var}(aX+b) &= \mathrm{E}\left[(aX+b-\mathrm{E}[aX+b])^2\right] \\ &= \mathrm{E}\left[(aX+b-a\mathrm{E}[X]-b)^2\right]\right] \\ &= a^2 \mathrm{E}\left[(X-\mathrm{E}[X])^2\right]. \end{aligned}$$

因此，將 RV 乘以常數 a 會導致方差的乘積增加 a^2；但是恆定的加性偏移 b 根本不會影響 RV 的可變性。

示例：$X \sim \operatorname{Bern}(0.3)$。回顧

$$\mathrm{E}[X]=p=0.3 \quad \text{and} \quad \operatorname{Var}(X)=pq=(0.3)(0.7)=0.21.$$

設 $Y=4X+5$。然後

$$\mathrm{E}[Y]=\mathrm{E}[4X+5]=4\mathrm{E}[X]+5=6.2$$

和

$$\operatorname{Var}(Y)=\operatorname{Var}(4X+5)=16 \operatorname{Var}(X)=3.36$$

2.5.3 LOTUS 通過泰勒級數

有時函數 $h(X)$ 很混亂，我們可以通過泰勒級數方法來評估 $\mathrm{E}[h(X)]$，而不是直接使用 LOTUS，如 Meyer 4 中所述。

2.5.3.1 §2.5.3.1 精確泰勒級數

回想一下，實函數 $h(x)$ 的泰勒級數由下式給出

$$h(x)=\sum_{k=0}^{\infty}\frac{h^{(k)}(a)}{k!}(x-a)^k \quad (\text{ for any constant } a)$$

其中$h^{(k)}(a)=\left.\frac{d^k}{dx^k}h(x)\right|_{x=a}$。假設一切都是明確定義的，並且我們可以使用等式（2.1）將期望值移動到（無限）和內，我們有

$$\mathrm{E}[h(X)]=\sum_{k=0}^{\infty}\frac{h^{(k)}(a)}{k!}\mathrm{E}\left[(X-a)^k\right].$$

示例：假設 $h(X)=e^{tX}$。然後通過方程 2.2，其中 $a=0$ 並且 $h^{(k)}(x)=t^k e^x$，我們有

$$\mathrm{E}\left[e^{tX}\right]=\sum_{k=0}^{\infty}\frac{h^{(k)}(0)}{k!}\mathrm{E}\left[X^k\right]=\sum_{k=0}^{\infty}\frac{t^k}{k!}\mathrm{E}\left[X^k\right]$$

這是一個非常普遍的結果，但現在讓我們具體假設 $X\sim\operatorname{Exp}(\lambda=1)$。然後通過 §2.5.2 中的示例，我們知道 $\mathrm{E}\left[X^k\right]=k!$，所以

$$\mathrm{E}\left[e^{tX}\right]=\sum_{k=0}^{\infty}\frac{t^k}{k!}k!=\sum_{k=0}^{\infty}t^k=\frac{1}{1-t}, \quad \text{for } t<1$$

這與我們在 \$2.6. 中通過稍微不同的方法獲得的答案相同

當我們在做的時候，請注意

$$\begin{aligned}\operatorname{Var}\left(e^{tX}\right)&=\mathrm{E}\left[e^{2tX}\right]-\left(\mathrm{E}\left[e^{tX}\right]\right)\\&=\frac{1}{1-2t}-\frac{1}{(1-t)^2}\\&=\frac{t^2}{(1-2t)(1-t)^2}\end{aligned}$$

2.5.3.2 &2.5.3,2 截斷泰勒級數逼近

在 \$2.5.3.1 中計算 $\mathrm{E}[h(X)]$ 的成功取決於我們是否願意計算所有矩 $\mathrm{E}\left[(Xa)^k\right], k=1,2,\ldots$ 但是如果我們想稍微懶一點，那麼可以說只有泰勒級數的前三個項的平均值為 $\mu=\ mathrmE[X]$ 代替 a 可以作為粗略的近似值，即

$$h(X)\approx h(\mu)+h'(\mu)(X-\mu)+\frac{h''(\mu)}{2}(X-\mu)^2$$

然後

$$\begin{aligned}\mathrm{E}[h(X)]&\doteq h(\mu)+h'(\mu)\mathrm{E}[X-\mu]+\frac{h''(\mu)}{2}\mathrm{E}\left[(X-\mu)^2\right]\\&=h(\mu)+\frac{h''(\mu)\sigma^2}{2}\end{aligned}$$

其中 $\sigma^2 = \operatorname{Var}(X) = \mathrm{E}\left[(X-\mu)^2\right]$，我們使用 "$\approx$" 表示具有大致相同分佈的隨機變量，而 "$\doteq$" 表示實數的近似相等。

此外，如果我們碰巧想要更粗略的近似，那麼我們可以只使用泰勒級數的前兩項來近似 $h(X)$ 的方差

$$\begin{aligned}\operatorname{Var}(h(X)) &\doteq \operatorname{Var}\big(h(\mu)+h'(\mu)(X-\mu)\big)\\ &= \big(h'(\mu)\big)^2 \operatorname{Var}(X) = \big(h'(\mu)\big)^2\sigma^2\end{aligned}$$

示例：再次假設 $h(X)=e^{tX}$，$X \sim \operatorname{Exp}(\lambda=1)$。我們現在將擴展 $a=\mu=1/\lambda=1$（而不是前面例子中的 $a=0$）並註意 $\sigma^2=1/\lambda^2=1$，獲得近似值

$$\mathrm{E}\left[e^{tX}\right] \doteq h(\mu)+\frac{h''(\mu)\sigma^2}{2} = h(1)+\frac{h''(1)}{2} = e^t+\frac{t^2e^t}{2}$$

和

$$\operatorname{Var}(h(X)) \doteq \big(h'(\mu)\big)^2\sigma^2 = \big(h'(1)\big)^2 = t^2e^t$$

如果我們回憶一下 §2.5.3 的確切答案，即 $\mathrm{E}\left[e^{tX}\right]=\frac{1}{1-t}$ 和 $operatornameVar\left(e^{tX}\right)=\frac{t^2}{(1-2t)(1-t)^2}$，那麼我們看到我們對 $\mathrm{E}\left[e^{tX}\right]$ 的近似值在 $t<0.4$ 和 $\operatorname{Var}\big(e^{tX}\big)$ 在 $t<0.2$ 的情況下確實很好，然後事情開始變得混亂。

示例：假設 X 有 pdf $f(x)=3x^2, 0\le x\le 1$，我們想測試我們對"複雜"隨機變量 $Y=h(X)=X^{3/4}$。好吧，它並沒有那麼複雜，因為我們可以計算出確切的時刻，我們現在將這樣做以便以後進行比較：

$$\begin{aligned}\mathrm{E}[Y] &= \int_{\mathbb{R}} x^{3/4} f(x)dx = \int_0^1 3x^{11/4}dx = 4/5\\ \mathrm{E}\left[Y^2\right] &= \int_{\mathbb{R}} x^{6/4} f(x)dx = \int_0^1 3x^{7/2}dx = 2/3, \quad \text{and}\\ \operatorname{Var}(Y) &= \mathrm{E}\left[Y^2\right]-(\mathrm{E}[Y])^2 = 2/75 = 0.0267\end{aligned}$$

在我們進行近似之前，請注意

$$\begin{aligned}\mu = \mathrm{E}[X] &= \int_{\mathbb{R}} x f(x)dx = \int_0^1 3x^3dx = 3/4,\\ \mathrm{E}\left[X^2\right] &= \int_{\mathbb{R}} x^2 f(x)dx = \int_0^1 3x^4dx = 3/5, \quad \text{and}\\ \sigma^2 = \operatorname{Var}(X) &= \mathrm{E}\left[X^2\right]-(\mathrm{E}[X])^2 = 3/80 = 0.0375.\end{aligned}$$

現在我們可以開始練習的近似部分了。我們有

$$\begin{aligned}h(\mu) &= \mu^{3/4} = (3/4)^{3/4} = 0.8059\\ h'(\mu) &= (3/4)\mu^{-1/4} = (3/4)(3/4)^{-1/4} = 0.8059\\ h''(\mu) &= -(3/16)\mu^{-5/4} = -0.2686.\end{aligned}$$

因此，

$$\mathrm{E}[Y] \doteq h(\mu)+\frac{h''(\mu)\sigma^2}{2} = 0.8059-\frac{(0.2686)(0.0375)}{2} = 0.8009$$

和

$$\mathrm{Var}(Y) \doteq \left[h'(\mu)\right]^2\sigma^2 = (0.8059)^2(0.0375) = 0.0243$$

它們相當接近它們的真實值，分別為 $\mathrm{E}[Y] = 0.8$ 和 $\mathrm{Var}(Y) = 0.0267$。

2.6 矩生成函數

我們現在將討論一個有很多非常好的用途的工具。回想一下 $\mathrm{E}\left[X^k\right]$ 是 X 的 k^{th} 矩。

定義：隨機變量X的矩生成函數（mgf）為

$$M_X(t) \equiv \mathrm{E}\left[e^{tX}\right]$$

注：$M_X(t)$ 是 t的函數， not of X ！

示例：$X \sim \mathrm{Bern}(p)$，因此 $X = 1$ 概率為 p，0 概率為 q。然後

$$M_X(t) = \mathrm{E}\left[e^{tX}\right] = \sum_x e^{tx} f(x) = e^{t\cdot 1}p + e^{t\cdot 0}q = pe^t + q$$

示例：$X \sim \mathrm{Exp}(\lambda)$。對於 $x \geq 0$，pdf 是 $f(x) = \lambda e^{-\lambda x}$。然後

$$\begin{aligned} M_X(t) &= \mathrm{E}\left[e^{tX}\right] \\ &= \int_{\mathbb{R}} e^{tx} f(x) dx \quad \text{(LOTUS)} \\ &= \lambda \int_0^\infty e^{(t-\lambda)x} dx \\ &= \frac{\lambda}{\lambda - t} \quad \text{if } \lambda > t \end{aligned}$$

定理（為什麼稱為矩生成函數）：在某些技術條件下（例如，對於所有 $t \in (-\epsilon, \epsilon)$ 必須存在 $M_X(t)$，對於某些 $\epsilon > 0$），我們有

$$\mathrm{E}\left[X^k\right] = \left.\frac{d^k}{dt^k} M_X(t)\right|_{t=0}, \quad k = 1, 2, \ldots$$

因此，您可以從 mgf 生成 X 的矩。（有時，以這種方式獲得時刻比直接獲得更容易。）

“證明”（有點不嚴謹）：通過指數函數的泰勒級數表達式以及期望值和總和的一些巧妙交換，我們有

$$M_X(t) = \mathrm{E}\left[e^{tX}\right] \text{"="} \mathrm{E}\left[\sum_{k=0}^\infty \frac{(tX)^k}{k!}\right] \text{"="} \sum_{k=0}^\infty \mathrm{E}\left[\frac{(tX)^k}{k!}\right] = \sum_{k=0}^\infty \frac{t^k}{k!} \mathrm{E}\left[X^k\right]$$

這（有點）暗示

$$\frac{d}{dt} M_X(t) \text{"="} \sum_{k=0}^\infty \frac{d}{dt} \frac{t^k}{k!} \mathrm{E}\left[X^k\right] = \sum_{k=1}^\infty \frac{t^{k-1}}{(k-1)!} \mathrm{E}\left[X^k\right] = \mathrm{E}[X] + t\mathrm{E}\left[X^2\right] + \cdots$$

所以

$$\left.\frac{d}{dt} M_X(t)\right|_{t=0} = \mathrm{E}[X]$$

高階時刻的相同交易。

示例：$X \sim \operatorname{Bern}(p)$。那麼$M_X(t) = pe^t + q$，並且

$$\mathrm{E}[X] = \left.\frac{d}{dt} M_X(t)\right|_{t=0} = \left.\frac{d}{dt}\left(pe^t + q\right)\right|_{t=0} = \left.pe^t\right|_{t=0} = p$$

其實很容易看出 $\mathrm{E}\left[X^k\right] = \left.\frac{d^k}{dt^k} M_X(t)\right|_{t=0} = p$，對於所有 k。

示例：$X \sim \operatorname{Exp}(\lambda)$。那麼$M_X(t) = \lambda/(\lambda - t)$，對於$\lambda > t$。所以

$$\mathrm{E}[X] = \left.\frac{d}{dt} M_X(t)\right|_{t=0} = \left.\frac{\lambda}{(\lambda - t)^2}\right|_{t=0} = 1/\lambda.$$

更遠，

$$\mathrm{E}\left[X^2\right] = \left.\frac{d^2}{dt^2} M_X(t)\right|_{t=0} = \left.\frac{2\lambda}{(\lambda - t)^3}\right|_{t=0} = 2/\lambda^2$$

因此，

$$\operatorname{Var}(X) = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = 1/\lambda^2$$

你可以用 mgf 做很多其他的好事：

- 求 X 的線性函數的 mgf（下一步）。
- 確定分佈（下）。
- 概率不等式應用($2.7)。
- 求獨立隨機變量之和的 mgf ($3.6)。
- 隨機變量的收斂證明（另一門課程）。

Theorem (mgf of a linear function of X) : 假設 X 有 mgf $M_X(t)$，令 $Y = aX + b$。那麼 $M_Y(t) = e^{tb} M_X(at)$。

證明：我們有

$$M_Y(t) = \mathrm{E}\left[e^{tY}\right] = \mathrm{E}\left[e^{t(aX+b)}\right] = e^{tb}\mathrm{E}\left[e^{(at)X}\right] = e^{tb} M_X(at)$$

示例：設 $X \sim \operatorname{Exp}(\lambda)$，且 $Y = 3X + 2$。然後

$$M_Y(t) = e^{2t} M_X(3t) = e^{2t}\frac{\lambda}{\lambda - 3t}, \quad \text{if } \lambda > 3t$$

定理（識別分佈）：在本文中，每個分佈都有一個唯一的 mgf。

證明：不在這裡！

示例：假設 Y 有 mgf

$$M_Y(t)=e^{2t}\frac{\lambda}{\lambda-3t},\quad \text{for } \lambda>3t$$

那麼通過前面的例子和mgf的唯一性，肯定是$Y\sim 3X+2$，其中$X\sim \operatorname{Exp}(\lambda)$

2.7 一些概率不等式

目標：給出提供一般概率界限的結果。如果我們想要粗略估計概率，這些很有用，但也可以應用於許多類型的證明。

定理（馬爾科夫不等式）：如果X是一個非負隨機變量並且$c>0$，那麼$\mathrm{P}(X\geq c)\leq \mathrm{E}[X]/c$。（這是一個非常粗略的上限。）證明：因為 X 是非負的，我們有

$$\begin{aligned}
\mathrm{E}[X] &= \int_{\mathbb{R}}^{\infty} xf(x)dx \\
&= \int_{0}^{\infty} xf(x)dx \\
&\geq \int_{c}^{\infty} xf(x)dx \\
&\geq c\int_{c}^{\infty} f(x)dx \\
&= c\mathrm{P}(X\geq c)
\end{aligned}$$

Theorem (Chebychev's Inequality) 2 ，假設$\mathrm{E}[X]=\mu$，並且$\operatorname{Var}(X)=\sigma^2$。然後對於任何 $c>0$

$$\mathrm{P}(|X-\mu|\geq c)\leq \sigma^2/c^2$$

證明：由馬爾可夫，用$|X-\mu|^2$代替X，用c^2代替c，我們有

$$\mathrm{P}(|X-\mu|\geq c)=\mathrm{P}\left((X-\mu)^2\geq c^2\right)\leq \frac{\mathrm{E}\left[(X-\mu)^2\right]}{c^2}=\sigma^2/c^2$$

2.7.0.1 備註：

- 我們也可以寫成 $\mathrm{P}(|X-\mu|<c)\geq 1-\sigma^2/c^2$
- 或者，如果 $c=k\sigma$，則 $\mathrm{P}(|X-\mu|\geq k\sigma)\leq 1/k^2$
- 就常數和方差而言，切比雪夫給出了 X 偏離均值超過一個常數的概率的界限。您總是可以使用 Chebychev，但它很粗糙。

示例：假設 $X\sim \operatorname{Unif}(0,1)$，因此 $f(x)=1$，對於 $0<x<1$。回想一下，對於 $\operatorname{Unif}(a,b)$ 分佈，我們有 $\mathrm{E}[X]=(a+b)/2=1/2$ 和 $\operatorname{Var}(X)=(ba)^2/12=1/12$。那麼切比雪夫暗示

$$\mathrm{P}\left(\left|X-\frac{1}{2}\right| \geq c\right) \leq \frac{1}{12c^2}$$

特別是對於 $c=1/3$，

$$\mathrm{P}\left(\left|X-\frac{1}{2}\right| \geq \frac{1}{3}\right) \leq \frac{3}{4} \quad \text{(Chebychev upper bound)}$$

2 拼寫“Chebychev”的方式有很多種。讓我們將上述綁定與確切答案進行比較：

$$\begin{aligned}
&\mathrm{P}\left(\left|X-\frac{1}{2}\right| \geq \frac{1}{3}\right) \\
&= 1-\mathrm{P}\left(\left|X-\frac{1}{2}\right| < \frac{1}{3}\right) \\
&= 1-\mathrm{P}\left(-\frac{1}{3} < X-\frac{1}{2} < \frac{1}{3}\right) \\
&= 1-\mathrm{P}\left(\frac{1}{6} < X < \frac{5}{6}\right) \\
&= 1-\int_{1/6}^{5/6} f(x)dx \\
&= 1-\frac{2}{3} = 1/3
\end{aligned}$$

因此，相比之下，切比雪夫界限相當高。

獎勵定理（切爾諾夫不等式）{3 對於任何 c，

$$\mathrm{P}(X \geq c) \leq e^{-ct} M_X(t)$$

證明：由 Markov 用 e^{tX} 代替 X 和 e^{tc} 代替 c，我們有

$$\mathrm{P}(X \geq c) = \mathrm{P}\left(e^{tX} \geq e^{tc}\right) \leq e^{-ct}\mathrm{E}\left[e^{tX}\right] = e^{-ct}M_X(t)$$

示例：假設 X 具有標準正態分佈，pdf $\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$，對於所有 $x \in \mathbb{R}$。很容易證明（通過一些涉及完成平方的微積分肘部潤滑脂）標準法線的 mgf 是

$$M_X(t) = \mathrm{E}\left[e^{tX}\right] = \int_{\mathbb{R}} e^{tx}\phi(x)dx = e^{t^2/2}$$

然後使用帶有 $t=c$ 的 Chernoff 立即產生“尾”概率，

$$\mathrm{P}(X \geq c) \leq e^{-c^2} M_X(c) = e^{-c^2/2}.$$

2.8 隨機變量的函數

在 \$2.5.2 中，我們了解了 LOTUS，它為我們提供了隨機變量函數的期望值。現在我們將討論該函數的整個分佈。例如，如果 X 是指數的，那麼 X^2 的分佈是什麼？這種類型的問題在各地都有巨大的應用，而且它經常會隨著我們的進行而彈出。

我們將從 &2.8.1 中的介紹性材料和基本示例開始討論，然後 §2.8.2 涉及所謂的逆變換定理，該定理廣泛用於計算機模擬問題（以及其他應用程序）這需要我們生成各種隨機變量。\$2.8.3 以涉及

LOTUS 的性質的幾個“榮譽”結果結束了討論。

3^3 這些天似乎每個人都有自己的不平等（只是說……）

2.8.1 簡介和嬰兒示例

2.8.1.1 問題陳述：

- 你有一個隨機變量 X，並且你知道它的 pmf/pdf $f(x)$。
- 定義 $Y \equiv h(X)$ (X 的一些函數)。
- 找到 $g(y)$，Y 的 pmf/pdf。

我們將從 X 是離散 RV 的情況開始，然後我們將轉到連續 X 的情況。

離散情況：X 離散隱含 Y 離散隱含

$$g(y) = \mathrm{P}(Y = y) = \mathrm{P}(h(X) = y) = \mathrm{P}(\{x \mid h(x) = y\}) = \sum_{x \mid h(x)=y} f(x).$$

示例：假設 X 是兩次拋硬幣中 H 的數量。假設我們想要 $Y = h(X) = X^3 - X$ 的 pmf。

這立即給了我們

$$g(0) = \mathrm{P}(Y = 0) = \mathrm{P}(X = 0 \text{ or } 1) = 3/4, \text{ and}$$
$$g(6) = \mathrm{P}(Y = 6) = \mathrm{P}(X = 2) = 1/4$$

換句話說，

$$g(y) = \begin{cases} 3/4 & \text{if } y = 0 \\ 1/4 & \text{if } y = 6 \end{cases}$$

示例：假設 X 是離散的

$$f(x) = \begin{cases} 1/8 & \text{if } x = -1 \\ 3/8 & \text{if } x = 0 \\ 1/3 & \text{if } x = 1 \\ 1/6 & \text{if } x = 2 \\ 0 & \text{otherwise.} \end{cases}$$

令 $Y = X^2$（所以 Y 只能等於 0,1 或 4）。然後

$$g(y) = \begin{cases} \mathrm{P}(Y = 0) = f(0) = 3/8 \\ \mathrm{P}(Y = 1) = f(-1) + f(1) = 11/24 \\ \mathrm{P}(Y = 4) = f(2) = 1/6 \end{cases}$$

連續情況：X 連續意味著 Y 可以是連續的或離散的。

示例：$Y = X^2$（顯然是連續的）。

Example: $Y=\begin{cases}0 & \text{if } X<0\\ 1 & \text{if } X\geq 0\end{cases}$ is not continuous.

方法：如果Y是連續的，計算它的cdf然後微分得到它的pdf。

- $G(y)\equiv \mathrm{P}(Y\leq y)=\mathrm{P}(h(X)\leq y)=\int_{x\mid h(x)\leq y} f(x)dx$

$g(y)=\frac{d}{dy}G(y)$

示例：假設 X 有 pdf $f(x)=|x|,-1\leq x\leq 1$。求隨機變量 $Y=h(X)=X^2$ 的 pdf。我們首先獲取Y的cdf。

$$G(y)=\mathrm{P}(Y\leq y)=\mathrm{P}\left(X^2\leq y\right)=\begin{cases}0 & \text{if } y\leq 0\\ 1 & \text{if } y\geq 1\\ (\star) & \text{if } 0<y<1\end{cases}$$

在哪裡

$$(\star)=\mathrm{P}(-\sqrt{y}\leq X\leq\sqrt{y})=\int_{-\sqrt{y}}^{\sqrt{y}} f(x)dx=\int_{-\sqrt{y}}^{\sqrt{y}}|x|dx=y$$

因此，

$$G(y)=\begin{cases}0 & \text{if } y\leq 0\\ 1 & \text{if } y\geq 1\\ y & \text{if } 0<y<1\end{cases}$$

這意味著

$$g(y)=G'(y)=\begin{cases}0 & \text{if } y\leq 0 \text{ or } y\geq 1\\ 1 & \text{if } 0<y<1\end{cases}$$

這是 $\mathrm{Unif}(0,1)$ 分佈！

2.8.2 青少年逆變換定理示例

在本小節中，我們將討論一個有很多應用的了不起的結果。首先，一個激勵的例子。示例：假設 $U\sim\mathrm{Unif}(0,1)$。求 $Y=-\ln(1-U)$ 的 pdf。使用通常的配方，我們有

$$\begin{aligned}G(y)&=\mathrm{P}(Y\leq y)\\&=\mathrm{P}(-\ell\mathrm{n}(1-U)\leq y)\\&=\mathrm{P}\left(U\leq 1-e^{-y}\right)\\&=\int_0^{1-e^{-y}} f(u)du\\&=1-e^{-y}(\text{ since } f(u)=1)\end{aligned}$$

因此，$g(y)=G'(y)=e^{-y},y>0$，所以 $Y\sim\mathrm{Exp}(\lambda=1)$。

哇！我們將製服代入一個簡單的方程並獲得指數！在下面的討論中，我們將推廣這個結果，這樣我們就可以通過將統一代入適當的方程來按需獲得幾乎任何合理的連續隨機變量。

逆變換定理： 4 假設 X 是具有 cdf $F(x)$ 的連續隨機變量。然後隨機變量$F(X) \sim \text{Unif}(0,1)$。（注意 $F(X)$ 是 RV，而 $F(x)$ 是 x的實函數。）

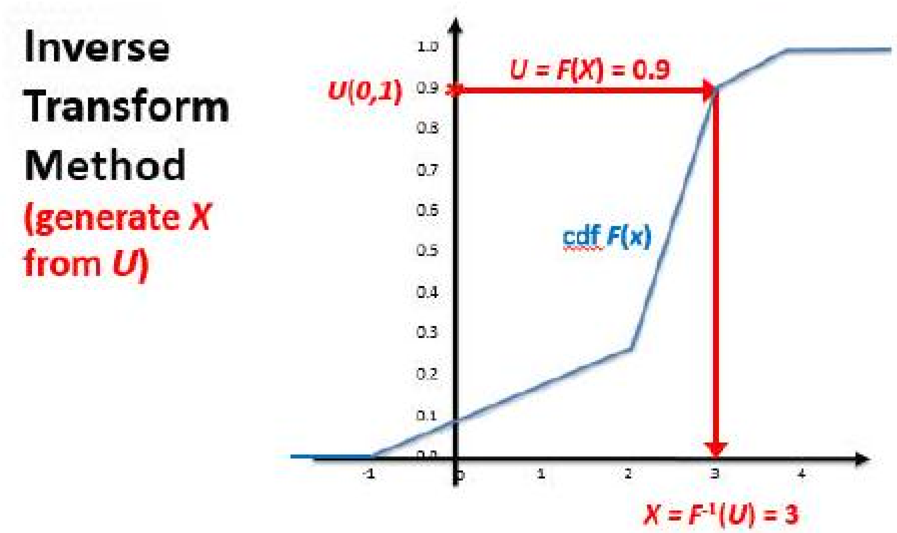
證明：令$Y = F(X)$。那麼 Y 的 cdf 是

$$\begin{aligned} G(y) &= \mathrm{P}(Y \leq y) = \mathrm{P}(F(X) \leq y) \\ &= \mathrm{P}\left(F^{-1}(F(X)) \leq F^{-1}(y)\right) \text{ (the cdf is monotone increasing)} \\ &= \mathrm{P}\left(X \leq F^{-1}(y)\right) \left(F^{-1} \text{ and } F \text{ go poof!}\right) \\ &= F\left(F^{-1}(y)\right) (F(x) \text{ is the cdf of } X) \\ &= y. \text{ Uniform!} \end{aligned}$$

備註：多麼好的定理！它適用於任何連續隨機變量 X ！

推論：$X = F^{-1}(U)$，因此您可以將 Unif $(0,1)$RV 插入逆 cdf 以生成 RV的實現 具有 X 的分佈。

方法：設$F(X) = U$，求解$X = F^{-1}(U)$，生成X。

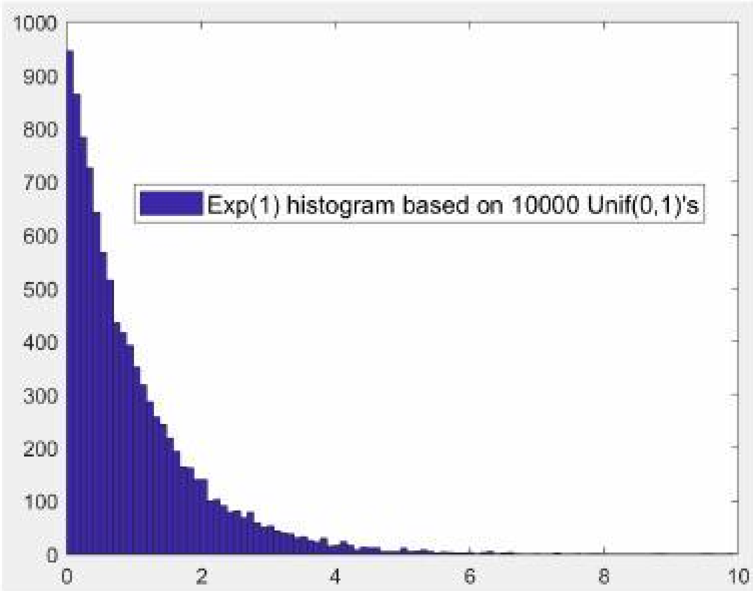


[4] 也稱為概率積分變換。示例：假設 X 是 $\text{Exp}(\lambda)$，因此它有 cdf $F(x) = 1 - e^{-\lambda x}$。與前面的示例類似，設置 $F(X) = 1 - e^{-\lambda X} = U$，並通過求解生成 Exp (λ) RV

$$X = F^{-1}(U) = -\frac{1}{\lambda}\ln(1-U) \sim \text{Exp}(\lambda)$$

備註：那麼這個結果是什麼意思？如果你想獲得一個漂亮的 Exp (λ) pdf，你所要做的就是。.

- 在計算機上生成 10000 個 Unif $(0,1)'$（例如，使用 Excel 中的 rand 函數或 Matlab 中的 unifrnd），
- 將 10000 個值代入上述 X 的方程，
- 繪製 X' 的直方圖，並且
- 欣賞你的出色工作。



備註：這個技巧在計算機模擬領域有著巨大的應用，我們經常需要生成隨機變量，例如客戶到達時間、資源服務時間、機器故障時間等。

2.8.3 成人榮譽示例

這是另一種更直接的方法來查找連續隨機變量的函數的 pdf。榮譽定理：假設 $Y = h(X)$ 是具有 pdf $f(x)$ 和 cdf $F(x)$ 的連續 RV X 的單調函數。那麼 Y 的 pdf $g(y)$ 由下式給出

$$g(y) = f\left(h^{-1}(y)\right)\left|\frac{dh^{-1}(y)}{dy}\right|.$$

證明：根據定義，我們有

$$\begin{aligned} g(y) &= \frac{d}{dy}G(y) = \frac{d}{dy}\mathrm{P}(Y \leq y) \\ &= \frac{d}{dy}\mathrm{P}(h(X) \leq y) \\ &= \frac{d}{dy}\mathrm{P}\left(X \leq h^{-1}(y)\right) \quad (h(x) \text{ is monotone}) \\ &= \frac{d}{dy}F\left(h^{-1}(y)\right) \\ &= f\left(h^{-1}(y)\right)\left|\frac{dh^{-1}(y)}{dy}\right| \quad \text{(chain rule).} \end{aligned}$$

示例：假設 $f(x) = 3x^2, 0 < x < 1$。設$Y = h(X) = X^{1/2}$，單調遞增。那麼 Y 的 pdf 是

$$\begin{aligned} g(y) &= f\left(h^{-1}(y)\right)\left|\frac{dh^{-1}(y)}{dy}\right| = f\left(y^2\right)\left|\frac{d\left(y^2\right)}{dy}\right| \\ &= 3y^4(2y) = 6y^5, \quad 0 < y < 1. \quad \square \end{aligned}$$

備註：這個結果的推廣可用於隨機變量的非單調函數，但我們將把這個討論留到另一天。

最後，這就是 LOTUS 工作的原因。

榮譽定理：如果 $h(\cdot)$ 是單調遞增的，那麼

$$\begin{aligned}\mathrm{E}[h(X)]=\mathrm{E}[Y]&=\int_{\mathbb{R}} y g(y) d y \\ &=\int_{\mathbb{R}} y f\left(h^{-1}(y)\right)\left|\frac{d h^{-1}(y)}{d y}\right| d y \\ &=\int_{\mathbb{R}} h(x) f(x)\left|\frac{d x}{d y}\right| d y \\ &=\int_{\mathbb{R}} h(x) f(x) d x .\end{aligned}$$

備註：同樣，單調性假設不是必需的，但它使證明和方法更容易。

2.9 2.9 練習

1. (2.2) 假設我們將一個六面骰子擲了五次。讓 X 表示您看到 1,2 或 3 的次數。求 $\mathrm{P}(X=4)$。

2. ($2.2) 假設 $X \sim \operatorname{Pois}(2)$。求 $\mathrm{P}(X>3)$。

3. (2.2) 我媽媽從 A 或 B 兩個盒子中的一個給我糖果。盒子 A 中的糖果一半好吃，一半好吃。但盒子 B 中的情況更糟，因為只有 1 $ / 4 $是美味的，而 3 $ / 4 $是噁心的。唯一的好消息是，當媽媽抓起一把糖果時，她有 90 % 可能會從盒子 A 中取出。

考慮到這一點，假設媽媽從其中一個盒子裡給我帶來了五顆糖果。如果其中三個糖果很好吃（兩個很噁心），那麼它們來自盒子 A 的可能性有多大？

提示：使用二項式條件概率嘗試貝葉斯定理。

4. ($2.3) 假設 X 是燈泡的壽命，並且 $X \sim \operatorname{Exp}(2 / \text { year })$

a. 求燈泡至少持續一年的概率，$\mathrm{P}(X>1)$。

b. 假設燈泡已經存活了兩年。它還能再存活一年的概率是多少，即 $\mathrm{P}(X>3 \mid X>2)$ ？

5. ($2.4) 假設連續隨機變量 X 對於 $0 \leq x \leq 3$ 有 pdf $f(x)=x^2 / 9$。它的 cdf 是什麼？

6. ($2.5) 我經營一家時髦的汽車經銷店。今天我打算以 0.5 $的概率不賣汽車，以 0.4 $的概率賣一輛汽車，以 0.1 $的概率賣兩輛汽車。我預計會賣出多少輛汽車？

7. (2.5) 假設連續隨機變量 X 對於 $0 \leq x \leq 3$ 有 pdf $f(x)=x^2 / 9$。它的期望值是多少？

8. (2.5) 證明：如果X是一個非負連續隨機變量（總是≥ 0），那麼$\mathrm{E}[X]=\int_0^{\infty} \mathrm{P}(X>x) d x$。

9. ($2.5) 假設 X 具有以下離散分佈。

a. 找到使 pmf 總和為 1 的 c 的值。

b. 計算 $\mathrm{P}(1 \leq X \leq 2)$。

c. 找出所有 x 的 cdf $F(x)$。

d. 計算 E[X]。

e. 計算 $\mathrm{Var}(X)$

10. (2.5 假設 X 是連續的，與 pdf $f(x)=cx^2, 0 \leq x \leq 1$ (a) 找到 c 的值，這將使 pdf積分為 1 。

b. 計算 $\mathrm{P}(0 \leq X \leq 1/2)$。

c. 找出所有 x 的 cdf $F(x)$。

d. 計算 $\mathrm{E}[X]$。

e. 計算 $\mathrm{Var}(X)$。

11. ($2.5 令 $\mathrm{E}[X]=4, \mathrm{Var}(X)=3$, and $Z=-4X+7$。求 $\mathrm{E}[-3Z]$ 和 $\mathrm{Var}(-3Z)$

12. ($2.5) 假設 X 是一個離散隨機變量，$X=-1$ 的概率為 0.2，$X=3$ 的概率為 0.8。

a. 求 E[X]。

b. 查找 $\mathrm{Var}(X)$

c. 求 $\mathrm{E}\left[3-\frac{1}{X}\right]$。

13. (2.5) 假設 X 是一個連續隨機變量，pdf $f(x)=4x^3$，對於 $0 \leq x \leq 1$。求 $\mathrm{E}\left[1/X^2\right]$

14. ($2.5) 今天外面很熱，我的家人想喝檸檬水。已知檸檬水的需求 X 有 pdf $f(x)=2x, 0 \leq x \leq 1$，其中 x 以加侖為單位。

a. 對於某個集合 $A \subseteq [0,1]$，定義指標函數，

$$1_{X \in A} \equiv \begin{cases} 1 & \text{if } X \in A \\ 0 & \text{otherwise.} \end{cases}$$

求 $\mathrm{E}\left[1_{X \in [0.5,1]}\right]$。

b. 不幸的是，我今天家裡只有 1/2 加侖，所以可能無法滿足需求。數量 $Y=(0.5-X)^{+}=\max\{0, 0.5-X\}$ 清楚地表示在漫長而炎熱的一天結束時將剩下多少檸檬水。找到 E[Y]。

15. (2.5) 假設X是一個連續隨機變量，均值$\mu=2$，方差$\sigma^2=10$。考慮隨機函數 $Y=h(X) \equiv e^X$。使用本課中的泰勒級數方法找到 $\mathrm{E}[Y]$ 的近似值，即 $\mathrm{E}[Y] \doteq h(\mu)+h''(\mu)\sigma^2/2$

16. ($2.6) 假設 $X=-2$ 的概率為 0.6，$X=4$ 的概率為 0.4。求X的矩生成函數，用它得到$\mathrm{E}[X]$和 $\mathrm{Var}(X)$

17. ($2.6) 假設 $X \sim \mathrm{Pois}(\lambda)$。$X$ 的矩生成函數是什麼？用它來查找 $\mathrm{E}[X]$

18. ($2.6) 假設 $X \sim \operatorname{Exp}(\lambda)$。使用 X 的 mgf 求 $\mathrm{E}\left[X^k\right]$。19. ($2.6) 假設 X 是一個離散隨機變量，其唯一可能的值是非負整數。我們將 X 的概率生成函數 (pgf) 定義為

$$g_X(s) \equiv \mathrm{E}\left[s^X\right]=\sum_{k=0}^{\infty} s^k \mathrm{P}(X=k)$$

a. 假設 pgf 存在，證明 $\mathrm{E}[X] = \frac{d}{ds} g_X(s)\Big|_{s=1}$。

b. 如果$X \sim \mathrm{Pois}(\lambda)$，求$g_X(s)$。（可能有用的事實：$\sum_{k=0}^{\infty} y^k/k! = e^y$。）

c. 假設 $X \sim \mathrm{Pois}(\lambda)$。使用 (a) 和 (b) 求 $\mathrm{E}[X]$。

20. (2.6) 如果 Y 有 mgf $M_Y(t) = 3e^{5t}/(3-4t)$，對於 $t < 3/4$，什麼是 Y 的分佈嗎？

21. ($2.7) 假設 $X \sim \mathrm{Unif}(-1, 6)$。比較從切比雪夫不等式獲得的概率 $\mathrm{P}(|X-\mu| \geq 1.5\sigma)$ 的上界與精確概率。

22. (2.8) 假設 X 是擲五面骰子的結果，邊數為 $-2, -1, 0, 1, 2$。求 $Y = X^2$ 的概率質量函數。

23. (2.8) 假設 $X \sim \mathrm{Unif}(1, 3)$。找到 $Z = e^{3X}$ 的 pdf。

24. ($2.8) 假設 X 有 pdf $f(x) = 3x^2 e^{-x^3}, x \geq 0$。找到 $Z = X^2$ 的 pdf。

25. (2.8) 假設 X 是一個連續隨機變量，pdf $f(x) = 2x$，對於 $0 < x < 1$。求 $Y = X^2$ 的 pdf $g(y)$。（這可能比您想像的要容易。）

26. (2.8) 假設 X 是一個連續隨機變量，pdf $f(x) = 2x$，對於 $0 < x < 1$。查找 $Y = \sqrt{X}$ 的 pdf

27. ($2.8) 計算機練習 - 隨機變量生成

a. 讓我們從簡單的東西開始 - Unif $(0, 1)$ 分佈。要在 Excel 中生成 Unif $(0, 1)$ 隨機變量，您只需使用 rand () 。（我們在本練習中使用 Excel 作為示例，但您可以隨意使用您選擇的軟件。）複製一整列包含 10,000 人的這些人並製作直方圖。

b. 在 Excel 中通過逆變換技術很容易生成 Exp(1) 隨機變量。只需使用

$$-\ln(\mathrm{rand}()) \quad \text{or} \quad -\ln(1-\mathrm{rand}()).$$

無論如何，生成 10,000 個左右的這些傢伙並製作一個漂亮的直方圖。

c. 在 Excel 中，您可以使用生成 Normal$(0, 1)$ 隨機變量

$$\text{norminv } (\text{rand}(), 0, 1) \text{ (inverse transform method).}$$

使用這個方程生成一堆法線並製作直方圖。(d) 好像無事可做，擲骰子 10,000 次。每個數字出現多少次？你預計大約有多少？

如果您在 Excel 中執行此操作，請嘗試使用

$$\mathrm{int}(6 * \mathrm{rand}()) + 1$$

為什麼這行得通？

e. 好像你還無事可做，擲兩個骰子 10,000 次。每個可能的總和出現多少次？你預計大約有多少？

f. 三角分佈。生成兩列 Unif $(0, 1)'$ s。在第三列中，將前兩列中的相應條目相加，例如 $\mathrm{C1} = \mathrm{A1} + \mathrm{B1}$ 等。製作第三列的直方圖. 猜猜你得到了什麼？

g. 來自中心極限定理的正態分佈（我們稍後會了解）。生成 12 列 Unif $(0,1)'$。在 13^{th} 列中，將前 12 列中的相應條目相加。製作 13^{th} 列的直方圖。猜猜這次你得到了什麼？

28. (2.8) 假設 $f(x) = 4x^3$，對於 $0 < x < 1$。令$Y = h(X) = X^2$，$X > 0$ 時單調遞增。使用“直接”方程找到 Y 的 pdf，

$$g(y) = f\left(h^{-1}(y)\right)\left|\frac{dh^{-1}(y)}{dy}\right|$$

ryNoSmBk

ryCh 3 【第 3 章】

3.0.0.1 二元隨機變量

在本章中，我們將探討同時考慮兩個隨機變量時會發生什麼。例如，假設我們隨機選擇一個人並查看她的身高和體重 (X, Y)。顯然，X 和 Y 會以某種方式相關，並且在任何後續分析中都必須考慮到這種關係。我們將研究前一章工作的各種概括，以及獨立性和相關性等新概念。

$3.1 - 介紹和定義

$3.2 - 條件分佈

$3.3 - 獨立隨機變量

$3.4 - 條件分佈的擴展

$3.5 - 協方差和相關性

$3.6 - 矩生成函數，再訪

3.7 - 隨機變量的二元函數

3.1 簡介和定義

我們考慮離散和連續的二元隨機變量。我們將在第 2 章中對單變量隨機變量的討論中自然地概括事物

3.1.1 離散案例

定義：如果X和Y是離散隨機變量，則(X, Y)稱為聯合離散二元隨機變量。聯合（或二元）pmf 是

$$f(x, y) = \mathrm{P}(X = x, Y = y), \quad \forall x, y.$$

$f(x, y)$ 的屬性：

- $0 \leq f(x, y) \leq 1, \forall x, y$。

- $\sum_x \sum_y f(x,y) = 1$。
- $A \subseteq \mathbb{R}^2 \Rightarrow \mathrm{P}((X,Y) \in A) = \sum \sum_{(x,y)\in A} f(x,y)$。

示例：考慮一個盒子裡的三隻襪子（編號為 1、2、3）。假設我們在沒有放回的情況下隨機抽取兩隻襪子。

令X 表示第一隻襪子的號碼，Y 表示第二隻襪子的號碼。聯合 pmf $f(x,y)$ 如下表所示。

我們說（稍後將正式定義）底行 $f_X(x) \equiv \mathrm{P}(X=x)$ 定義了 X 的邊際 pmf；最右邊的列 $f_Y(y) \equiv \mathrm{P}(Y=y)$ 定義了 Y 的邊際 pmf。當然，所有內容加起來都是 1（右下條目），我們覺得這很令人滿意，同時也讓人放心。

繼續這個例子，注意根據全概率定律，

$$\mathrm{P}(X=1) = \sum_{y=1}^{3} \mathrm{P}(X=1, Y=y) = 1/3.$$

此外，

$$\begin{aligned}
&\mathrm{P}(X \geq 2, Y \geq 2) \\
&\quad = \sum_{x\geq 2}\sum_{y \geq 2} f(x,y) \\
&\quad = f(2,2) + f(2,3) + f(3,2) + f(3,3) \\
&\quad = 0 + 1/6 + 1/6 + 0 = 1/3.
\end{aligned}$$

3.1.2 連續案例

定義：如果X和Y是連續隨機變量，那麼如果存在一個魔函數$f(x,y)$，則(X,Y)是一個聯合連續二元隨機變量，使得

- $f(x,y) \geq 0, \forall x, y$
- $\iint_{\mathbb{R}^2} f(x,y) dx dy = 1$。- 如果 $A \subset \mathbb{R}^2$，則 $\mathrm{P}(A) = \mathrm{P}((X,Y) \in A) = \iint_A f(x,y) dx dy$。$\mathrm{P}(A)$ 是 $f(x,y)$ 和 A 之間的體積。

在這種情況下，$f(x,y)$ 稱為 (X,Y) 的聯合 pdf。考慮到

$$f(x,y) dx dy \doteq \mathrm{P}(x < X < x + dx, y < Y < y + dy)$$

很容易看出所有這些如何概括一維 pdf，$f(x)$。

示例：在單位正方形內接圓的內部隨機選擇一個點(X,Y)，即$C \equiv \left(x-\frac{1}{2}\right)^2 + \left(y - \frac{1}{2}\right)^2 \leq \frac{1}{4}$. 由於圓的面積是 $\pi/4$，並且我們在該區域內均勻地選擇一個點，(X,Y) 的聯合 pdf 是

$$f(x,y) = \begin{cases} 4/\pi & \text{if } (x,y) \in C \\ 0 & \text{otherwise.} \end{cases}$$

我們可以使用這個公式來計算概率。例如，

$$\begin{aligned}\mathrm{P}(X \geq 1/2, Y \geq 1/2) &= \int_{1/2}^{1}\int_{1/2}^{1} f(x,y)dxdy \\ &= \int_{x\geq 1/2, y\geq 1/2, (x,y)\in C}\int_{x} (4/\pi)dxdy \\ &= 1/4\end{aligned}$$

我們只是使用了對稱性參數而不是嘗試正式整合（看起來比實際更混亂）。

應用：將 n 飛鏢隨機投擲到單位方格中。任何單個飛鏢落入圓圈的概率是 $\pi/4$。按理說，飛鏢落在圓圈中的比例 $\hat{p}_n$ 大約為 $\pi/4$。所以你可以用 $4\hat{p}_n$ 來估計 π ！例如，如果我們投擲 1000 個飛鏢和 752 個土地，那麼我們對 π 的估計是 $4(752/1000) = 3.08$。

示例：假設

$$f(x,y) = \begin{cases} 4xy & \text{if } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise.}\end{cases}$$

讓我們求區域 $0 \leq y \leq 1 - x^2$ 的概率（體積）。

$$V = \int_0^1 \int_0^{1-x^2} 4xydydx = \int_0^1 \int_0^{\sqrt{1-y}} 4xydxdy = 1/3$$

道德：如果你小心限制，你可以檢查你的答案！

3.1.3 雙變量 cdf

定義：X 和 Y 的聯合（二元）cdf 是 $F(x,y) \equiv P(X \leq x, Y \leq y)$，對於所有 x, y。換句話說，

$$F(x,y) = \begin{cases} \sum_{s\leq x, t\leq y} f(s,t) & \text{discrete} \\ \int_{-\infty}^{y}\int_{-\infty}^{x} f(s,t)dsdt & \text{continuous.}\end{cases}$$

3.1.3.1 cdf 的屬性：

- $F(x,y)$ 在 x 和 y 中均不減少。
- $\lim_{x\to-\infty} F(x,y) = \lim_{y\to-\infty} F(x,y) = 0$。
- $\lim_{x\to\infty} F(x,y) \equiv F_Y(y) = \mathrm{P}(Y \leq y) \quad$ (Y 的邊際 cdf)。
- $\lim_{y\to\infty} F(x,y) \equiv F_X(x) = \mathrm{P}(X \leq x) \quad$ (X 的邊際 cdf)。
- $\lim_{x\to\infty}\lim_{y\to\infty} F(x,y) = 1$。

在連續情況下，從 cdf 到 pdf 很容易 - 只需取導數：

- 一維：$f(x) = F'(x) = \frac{d}{dx}\int_{-\infty}^{x} f(t)dt$。
- 二維：$f(x,y) = \frac{\partial^2}{\partial x\partial y}F(x,y) = \frac{\partial^2}{\partial x\partial y}\int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)dtds$。

示例：假設

$$F(x, y)= \begin{cases}1-e^{-x}-e^{-y}+e^{-(x+y)} & \text { if } x \geq 0, y \geq 0 \\ 0 & \text { if } x<0 \text { or } y<0 .\end{cases}$$

X 的邊際 cdf 是

$$F_X(x)=\lim _{y \rightarrow \infty} F(x, y)= \begin{cases}1-e^{-x} & \text { if } x \geq 0 \\ 0 & \text { if } x<0 .\end{cases}$$

聯合pdf是

$$\begin{aligned} f(x, y) & =\frac{\partial^2}{\partial x \partial y} F(x, y) \\ & =\frac{\partial}{\partial y}\left(e^{-x}-e^{-y} e^{-x}\right) \\ & = \begin{cases}e^{-(x+y)} & \text { if } x \geq 0, y \geq 0 \\ 0 & \text { if } x<0 \text { or } y<0 .\end{cases} \end{aligned}$$

3.1.4 邊際分佈

我們可以使用來自 $f(x, y)$ 和/或 $F(x, y)$ 的信息來獲得 X 和 Y 的單獨分佈。定義：如果 X 和 Y 是聯合離散的，那麼 X 和 Y 的邊際 pmf 分別是，

$$f_X(x)=\mathrm{P}(X=x)=\sum_y \mathrm{P}(X=x, Y=y)=\sum_y f(x, y)$$

和

$$f_Y(y)=\mathrm{P}(Y=y)=\sum_x \mathrm{P}(X=x, Y=y)=\sum_x f(x, y)$$

備註：這些定義只是全概率定律的應用。

示例：考慮以下聯合 pmf $f(x, y)=\mathrm{P}(X=x, Y=y)$。

例如，通過總概率，

$$\mathrm{P}(X=1)=\mathrm{P}(X=1, Y=\text { any number })=0.3$$

示例：考慮另一個離散 pmf。

備註：嗯…與上一個示例相比，這具有相同的邊緣但不同的聯合分佈！這是因為聯合分佈包含的信息遠不止邊際，因此，許多聯合分佈可以產生相同的邊際。

定義：如果 X 和 Y 是聯合連續的，那麼 X 和 Y 的邊際 pdf 分別是，

$$f_X(x)=\int_{\mathbb{R}} f(x, y) d y \quad \text { and } \quad f_Y(y)=\int_{\mathbb{R}} f(x, y) d x .$$

備註：這些定義顯然是離散情況的連續類比（我們現在有積分而不是和）。請注意，例如，$f_X(x)$ 是 x 單獨的函數，因為 y 已全部集成。示例：假設 (X, Y) 具有雙變量 pdf

$$f(x, y)= \begin{cases}e^{-(x+y)} & \text { if } x \geq 0, y \geq 0 \\ 0 & \text { otherwise }\end{cases}$$

那麼 X 的邊際 pdf 是

$$f_X(x) = \int_{\mathbb{R}} f(x,y)dy = \int_0^\infty e^{-(x+y)}dy = \begin{cases} e^{-x} & \text{if } x \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

所以 $X \sim \text{Exp}(1)$。也很容易證明 $Y \sim \text{Exp}(1)$。

這是一個更棘手的例子，我們偶爾會重溫一下。

示例：考慮聯合 pdf

$$f(x,y) = \frac{21}{4}x^2y, \quad \text{if } x^2 \leq y \leq 1,$$

並註意 pdf 為正的非矩形有趣限制，即 $x^2 \leq y \leq 1$。經過一些仔細的整合，我們有

$$f_X(x) = \int_{\mathbb{R}} f(x,y)dy = \int_{x^2}^1 \frac{21}{4}x^2ydy = \frac{21}{8}x^2\left(1-x^4\right), \quad \text{if } -1 \leq x \leq 1; \quad \text{and}$$
$$f_Y(y) = \int_{\mathbb{R}} f(x,y)dx = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4}x^2ydx = \frac{7}{2}y^{5/2}, \quad \text{if } 0 \leq y \leq 1. \quad \square$$

3.2 條件分佈

在我們開始任何新工作之前，回顧一下基於事件 A 和
B的條件概率的舊定義：$\text{P}(A \mid B) = \text{P}(A \cap B)/\text{P}(B)$，如果 $\text{P}(B) > 0$。在本節中，我們將把這個概念應用於概率分佈。為此，讓我們使用涉及隨機變量的事件建立適當的類比。

假設 X 和 Y 是聯合離散的 RV 。那麼如果$\text{P}(X=x) > 0$，

$$\text{P}(Y=y \mid X=x) = \frac{\text{P}(X=x \cap Y=y)}{\text{P}(X=x)} = \frac{f(x,y)}{f_X(x)}.$$

因此，例如，$\text{P}(Y=y \mid X=2)$ 定義了 Y 的條件概率分佈，給定 $X=2$ 的信息。

現在我們準備繼續進行一般定義。

定義：如果$f_X(x) > 0$，那麼給定$X=x$的Y的條件pmf/pdf是

$$f_{Y|X}(y \mid x) \equiv \frac{f(x,y)}{f_X(x)}, \text{ for all } y$$

備註：- 為了節省寶貴的墨水，我們通常只寫 $f(y \mid x)$ 而不是 $f_{Y|X}(y \mid x)$。

- 有條件的 pmf/pdf 是合法的 pmf/pdf。因此，如果我們認為 x 是固定的，對於離散情況，我們有 $\sum_y f(y \mid x) = 1$，並且 $\int_{\mathbb{R}} f(y \mid x)$對於連續情況，$dy = 1$。
- 當然，通過符號對稱，我們有 $f_{X|Y}(x \mid y) = f(x \mid y) = f(x,y)/f_Y(y)$

示例：考慮下面給出的（離散）聯合 pmf $f(x,y) = \text{P}(X=x, Y=y)$。

然後，例如，

$$f(x \mid y=60)=\frac{f(x,60)}{f_Y(60)}=\frac{f(x,60)}{0.8}=\begin{cases}\frac{29}{80} & \text{if } x=1\\ \frac{3}{80} & \text{if } x=2\\ \frac{48}{80} & \text{if } x=3\end{cases}$$

它代表 X 的更新 pmf，假設 $Y=60$。

舊的連續示例：假設

$$f(x,y)=\frac{21}{4}x^2y, \quad \text{if } x^2 \leq y \leq 1,$$

我們再次注意到有趣的限制並回憶起

$$f_X(x)=\tfrac{21}{8}x^2\left(1-x^4\right), \quad \text{if } -1 \leq x \leq 1; \quad \text{and}$$
$$f_Y(y)=\tfrac{7}{2}y^{5/2}, \quad \text{if } 0 \leq y \leq 1$$

那麼 Y 的條件 pdf 給定 $X=x$ 是

$$f(y \mid x)=\frac{f(x,y)}{f_X(x)}=\frac{\frac{21}{4}x^2y}{\frac{21}{8}x^2\left(1-x^4\right)}=\frac{2y}{1-x^4}, \quad \text{if } x^2 \leq y \leq 1,$$

我們指出有趣的限制已被保留（因為 x 仍在發揮作用）。請注意，在 $f(y \mid x)$ 的一般表達式中，數量 $2/\left(1-x^4\right)$ 是關於 y 的常數，我們可以檢查看到 $f(y \mid x)$ 是合法的條件pdf：

$$\int_{\mathbb{R}} f(y \mid x)dy=\int_{x^2}^{1}\frac{2y}{1-x^4}dy=1$$

繼續這個例子，讓我們看看當我們有$X=1/2$的特定條件信息時會發生什麼，即

$$f(y \mid 1/2)=\frac{2y}{1-\frac{1}{16}}=\frac{32y}{15}, \quad \text{if } \frac{1}{4} \leq y \leq 1$$

這個條件pdf允許我們計算任何相關的概率，例如，

$$\mathrm{P}\left(\frac{3}{4} \leq Y \leq 1 \mid X=\frac{1}{2}\right)=\int_{3/4}^{1} f(y \mid 1/2)dy=\int_{3/4}^{1}\frac{32y}{15}dy=\frac{7}{15}$$

通用示例：給定 $f_X(x)$ 和 $f(y \mid x)$，找到 $\mathrm{E}[Y]$。

遊戲計劃是找到 $f(x,y)=f_X(x)f(y \mid x)$，然後 $f_Y(y)=\int_{\mathbb{R}} f(x,y)dx$，最後是 $\mathrm{E}[Y]=\int_{\mathbb{R}} y f_Y(y)dy$。例如，..

示例：假設 $f_X(x)=2x$，對於 $0<x<1$。鑑於 $X=x$，假設 $Y \mid x \sim \mathrm{Unif}(0,x)$。現在找到 $\mathrm{E}[Y]$

解決方案：$Y \mid x \sim \mathrm{Unif}(0,x)$ 意味著 $f(y \mid x)=1/x$，對於 $0<y<x$（注意有趣的限制）。所以

$$\begin{aligned} f(x,y) &= f_X(x)f(y \mid x)\\ &= 2x \cdot \frac{1}{x}, \quad \text{for } 0<x<1 \text{ and } 0<y<x\\ &= 2, \quad \text{for } 0<y<x<1 \quad \text{(still have the funny limits).} \end{aligned}$$

因此，

$$f_Y(y)=\int_{\mathbb{R}} f(x,y)dx=\int_y^1 2dx=2(1-y), 0<y<1$$

接著

$$\mathrm{E}[Y]=\int_{\mathbb{R}} y f_Y(y)dy=2\int_0^1\left(y-y^2\right)dy=1/3$$

3.3 獨立隨機變量

正如兩個事件可以相互依賴一樣，我們可以定義關於隨機變量的獨立性。

3.3.1 定義和基本結果

回想一下，如果 $\mathrm{P}(A\cap B)=\mathrm{P}(A)\mathrm{P}(B)$，則兩個事件 A 和 B 是獨立的。然後

$$\mathrm{P}(A\mid B)=\frac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)}=\frac{\mathrm{P}(A)\mathrm{P}(B)}{\mathrm{P}(B)}=\mathrm{P}(A)$$

同樣，$\mathrm{P}(B\mid A)=\mathrm{P}(B)$。

現在我們要定義 RV 的獨立性，在這種情況下 X 的結果不會影響 Y 的結果，反之亦然。

定義：X 和 Y 是獨立隨機變量，如果

$$f(x,y)=f_X(x)f_Y(y) \text{ for all } x,y$$

事實證明，以下是等價的定義：

$$\begin{gathered}F(x,y)=F_X(x)F_Y(y),\quad \forall x,y;\quad \text{or}\\ \mathrm{P}(X\le x, Y\le y)=\mathrm{P}(X\le x)\mathrm{P}(Y\le y),\quad \forall x,y\end{gathered}$$

如果 X 和 Y 不是獨立的，那麼它們是依賴的。

直觀定理：當 $f(y\mid x)=f_Y(y), \forall x,y$ 時，X 和 Y 是獨立的。

證明：通過假設和條件的定義，我們有

$$f_Y(y)=f(y\mid x)=\frac{f(x,y)}{f_X(x)}\Leftrightarrow f(x,y)=f_X(x)f_Y(y)\Leftrightarrow \text{ indep}$$

類似地，X 和 Y 獨立意味著 $f(x\mid y)=f_X(x)$

換句話說，X 和Y 的結果不會相互影響，例如火星上的溫度和IBM 當前的股價是不相關的。

示例：考慮下表給出的（離散）二元 pmf $f(x,y)=\mathrm{P}(X=x,Y=y)$。

那麼 X 和 Y 是獨立的，因為 $f(x,y)=f_X(x)f_Y(y), \forall x,y$

示例：有時，您必須小心確保每個 (x,y) 對的行為。

即使 $f(1,2)$ 服從因素，X 和 Y 也不是獨立的，因為例如 $f(5,2) \neq f_X(5) f_Y(2)$。

示例：考慮（連續）二元 pdf $f(x,y) = 6xy^2, 0 \leq x \leq 1, 0 \leq y \leq 1$。經過一些工作（可以通過下一個定理避免），我們可以推導出

$$f_X(x) = \int_{\mathbb{R}} f(x,y)dy = 2x, 0 \leq x \leq 1; \quad \text{and}$$
$$f_Y(y) = \int_{\mathbb{R}} f(x,y)dx = 3y^2, 0 \leq y \leq 1$$

那麼 X 和 Y 是獨立的，因為 $f(x,y) = f_X(x) f_Y(y), \forall x, y$。這是判斷 X 和 Y 是否獨立（我們不會證明）的簡單方法，它避免了 $f_X(x)$ 和 $f_Y(y)$。

定理：對於某些函數 $a(x)$ 和 $b(y)$（不一定是pdf）。

因此，如果將 $f(x,y)$ 分解為 x 和 y 的單獨函數，則 X 和 Y 是獨立的。

但是，如果存在非矩形的有趣限制，這使得將邊際分解變得不可能。在這種情況下，X 和 Y 將是依賴的 - 注意！

特別是對於某些 (x,y)，有趣的限制可能允許 $f_X(x) > 0$ 和 $f_Y(y) > 0$，但是 $f(x, y) =0 $。這立即弄亂了分解，因為 $f_X(x) f_Y(y) \neq f(x,y)$。

示例：$f(x,y) = 6xy^2$，對於 $0 \leq x \leq 1, 0 \leq y \leq 1$。

取$a(x) = 6x, 0 \leq x \leq 1,$ 和$b(y) = y^2, 0 \leq y \leq 1$。因此，$X$ 和 Y 是獨立的（如上所述）。

示例：$f(x,y) = \frac{21}{4} x^2 y$，對於 $x^2 \leq y \leq 1$（回想一下 $f_X(x)$ 和 $f_Y(y)$ 很討厭）。

哎呀！有趣的限制意味著依賴！

示例：$f(x,y) = \frac{c}{x+y}$，對於 $1 \leq x \leq 2, 1 \leq y \leq 3$

不幸的是，您不能將 $f(x,y)$ 分別分解為 x 和 y 的函數。因此，X 和 Y 不是獨立的。

3.3.2 獨立的後果

現在我們可以確定 X 和 Y 是否是獨立的，那麼我們可以用這些知識做什麼呢？我們先介紹一個有用的工具。

定義/定理（無意識統計學家的二維定律）：設 $h(X,Y)$ 是隨機變量 X 和 Y 的函數。然後

$$\mathrm{E}[h(X,Y)] = \begin{cases} \sum_x \sum_y h(x,y) f(x,y) & \text{discrete case} \\ \int_{\mathbb{R}} \int_{\mathbb{R}} h(x,y) f(x,y) dx dy & \text{continuous case.} \end{cases}$$

示例： 2-D LOTUS 就像它的一維小弟一樣，可以幫助我們計算各種好東西，其中一些我們將在下面討論。以下是其在連續情況下的適用性的一些示例。

- $\mathrm{E}[X+Y] = \int_{\mathbb{R}} \int_{\mathbb{R}} (x+y) f(x,y) dx dy$
- $\mathrm{E}[XY] = \int_{\mathbb{R}} \int_{\mathbb{R}} xy f(x,y) dx dy$

- $\mathrm{E}\left[X^2\sin(Y)\right]=\int_{\mathbb{R}}\int_{\mathbb{R}}x^2\sin(y)f(x,y)dxdy$ 這個二維 LOTUS 最簡單的結果是預期值加起來的直觀事實。

定理：無論 X 和 Y 是否獨立，我們有

$$\mathrm{E}[X+Y]=\mathrm{E}[X]+\mathrm{E}[Y]$$

證明（連續案例）：

$$\begin{aligned}
\mathrm{E}[X+Y] &= \int_{\mathbb{R}}\int_{\mathbb{R}}(x+y)f(x,y)dxdy \quad (\text{two-dimensional LOTUS} \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}}xf(x,y)dxdy+\int_{\mathbb{R}}\int_{\mathbb{R}}yf(x,y)dxdy \\
&= \int_{\mathbb{R}}x\int_{\mathbb{R}}f(x,y)dydx+\int_{\mathbb{R}}y\int_{\mathbb{R}}f(x,y)dxdy \\
&= \int_{\mathbb{R}}xf_X(x)dx+\int_{\mathbb{R}}yf_Y(y)dy \\
&= \mathrm{E}[X]+\mathrm{E}[Y].
\end{aligned}$$

我們可以將此結果推廣到兩個以上的 RV。

推論：如果 $X_1,X_2,\ldots,X_n$ 是隨機變量（獨立與否），那麼

$$\mathrm{E}\left[\sum_{i=1}^{n}X_i\right]=\sum_{i=1}^{n}\mathrm{E}\left[X_i\right]$$

證明：歸納。

有時候，獨立確實很重要……

定理：如果X和Y是獨立的，則$\mathrm{E}[XY]=\mathrm{E}[X]\mathrm{E}[Y]$。

證明（連續案例）：

$$\begin{aligned}
\mathrm{E}[XY] &= \int_{\mathbb{R}}\int_{\mathbb{R}}xyf(x,y)dxdy \quad (\text{two-dimensional LOTUS}) \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}}xyf_X(x)f_Y(y)dxdy(X\text{ and }Y\text{ are independent}) \\
&= \left(\int_{\mathbb{R}}xf_X(x)dx\right)\left(\int_{\mathbb{R}}yf_Y(y)dy\right) \\
&= \mathrm{E}[X]\mathrm{E}[Y]
\end{aligned}$$

備註：如果 X 和 Y 依賴，上述定理不一定成立。請參閱下面 $3.5 中關於協方差的討論。

下一個定理是獨立性的一個非常重要的結果——它表明如果兩個 RV 是獨立的，則可以將方差相加。

定理：如果 X 和 Y 是獨立的，那麼

$$\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)$$

3.3.2.1 證明：

備註：獨立性的假設在這裡真的很重要。如果 X 和 Y 不是獨立的，那麼結果可能不成立！同樣，請參閱即將到來的關於協方差的討論。

我們可以概括求和結果的方差。.

推論：如果 $X_1, X_2, \ldots, X_n$ 是獨立隨機變量，那麼

$$\operatorname{Var}\left(\sum_{i=1}^{n} a_i X_i + b\right) = \sum_{i=1}^{n} a_i^2 \operatorname{Var}(X_i)$$

證明：歸納。

Corollary (of Corollary)：如果 $X_1, X_2, \ldots, X_n$ 是獨立的，那麼

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

到目前為止，本小節中的所有結果都著眼於獨立性如何影響（或不影響）隨機變量的某些函數（例如總和）的期望值和方差。我們現在提出一個有用的結果，在這個結果中，更具挑戰性的目的需要獨立性。

定理：如果 X 和 Y 是獨立的連續 RV，那麼（使用明顯的 pdf 和 cdf 符號）

$$\mathrm{P}(Y \leq X) = \int_{\mathbb{R}} F_Y(x) f_X(x) dx$$

證明：根據事件 $\{Y < X\}$ 的定義，我們有

$$\begin{aligned}
\mathrm{P}(Y \leq X) &= \iint_{y \leq x} f(x, y) dx dy \\
&= \int_{\mathbb{R}} \int_{-\infty}^{x} f_Y(y) f_X(x) dy dx \quad \text{(since } X, Y \text{ are independent)} \\
&= \int_{\mathbb{R}} \mathrm{P}(Y \leq x) f_X(x) dx.
\end{aligned}$$

$$\begin{aligned}
&\operatorname{Var}(X+Y) = \mathrm{E}\left[(X+Y)^2\right] - (\mathrm{E}[X+Y])^2 \\
&= \mathrm{E}\left[X^2 + 2XY + Y^2\right] - (\mathrm{E}[X] + \mathrm{E}[Y])^2 \\
&= \mathrm{E}\left[X^2\right] + 2\mathrm{E}[XY] + \mathrm{E}\left[Y^2\right] - \left\{(\mathrm{E}[X])^2 + 2\mathrm{E}[X]\mathrm{E}[Y] + (\mathrm{E}[Y])^2\right\} \\
&= \mathrm{E}\left[X^2\right] + 2\mathrm{E}[X]\mathrm{E}[Y] + \mathrm{E}\left[Y^2\right] - (\mathrm{E}[X])^2 - 2\mathrm{E}[X]\mathrm{E}[Y] - (\mathrm{E}[Y])^2 \\
&\text{(since } X \text{ and } Y \text{ are independent)} \\
&= \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 + \mathrm{E}\left[Y^2\right] - (\mathrm{E}[Y])^2.
\end{aligned}$$

示例：假設 $X \sim \operatorname{Exp}(\alpha)$ 是直到下一個男性司機出現在停車場的時間（以 $\alpha/$ 小時的速度），與此無關，$Y \sim \operatorname{Exp}(\beta)$ 是下一位女司機的時間（以 β / 小時計算）。然後

$$
\begin{aligned}
\mathrm{P}(Y \leq X) &= \int_{\mathbb{R}} F_Y(x) f_X(x) dx \\
&= \int_0^{\infty} \left(1 - e^{-\beta x}\right) \alpha e^{-\alpha x} dx \\
&= \frac{\beta}{\alpha + \beta}
\end{aligned}
$$

3.3.3 隨機樣本

隨機樣本的概念很容易理解並且在實踐中非常有用。

定義：如果 (i) X_i 都是獨立的，並且 (ii) 有相同的 pmf/pdf，例如 $f(x)$。

符號：$X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} f(x)$ (iid = “獨立同分佈”)。

示例/定理：假設 $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} f(x)$, with $\mathrm{E}\left[X_i\right] = \mu$，$\operatorname{Var}(X_i) = \sigma^2$。將樣本均值定義為 $\bar{X} \equiv \sum_{i=1}^n X_i / n$。然後

$$
\mathrm{E}[\bar{X}] = \mathrm{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n \mathrm{E}\left[X_i\right] = \frac{1}{n} \sum_{i=1}^n \mu = \frac{n\mu}{n} = \mu
$$

所以 $\bar{X}$ 的平均值與 X_i 的平均值相同。

同時，樣本均值的方差如何？

$$
\begin{aligned}
\operatorname{Var}(\bar{X}) &= \operatorname{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \operatorname{Var}\left(\sum_{i=1}^n X_i\right) \\
&= \frac{1}{n^2} \sum_{i=1}^n \operatorname{Var}(X_i) \quad (X_i \text{ 's are independent }) \\
&= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}
\end{aligned}
$$

所以 $\bar{X}$ 的均值與 X_i 的均值相同，但方差減小了！這使得 $\bar{X}$ 成為 μ 的一個很好的估計量（這在實踐中通常是未知的）。結果被稱為大數定律。請繼續關注第 4 章中有關此開始的更多信息

3.4 條件分佈的擴展

我們已經看到條件分佈如何在收到新信息後反映 pmf/pdf 的變化。我們現在將研究這個重要概念的擴展。以下是以下小節中的內容。$3.4.1 定義了條件期望並提供了介紹性示例。3.4.2 提出了“雙重期望”定理，這將使我們能夠間接計算出令人驚訝的各種期望值。$3.4.3 給出了幾個涉及條件分佈的榮譽問題。

3.4.1 條件期望

首先，讓我們回顧一下期望的通常定義。例如，某個人群（可能是任何身高）的男性的預期體重是多少？

$$\mathrm{E}[Y] = \begin{cases} \sum_y yf(y) & \text{discrete} \\ \int_{\mathbb{R}} yf(y)dy & \text{continuous} \end{cases}$$

現在，假設我們對條件期望值感興趣，例如，身高為 $6'$ 的男性的平均體重。有人會推測，高個子的預期體重往往比矮個子高。

在任何情況下，讓 $f(y \mid x)$ 成為 Y 的條件 pmf/pdf，給定 $X = x$。

定義：給定 $X = x$ 的 Y 的條件期望是

$$\mathrm{E}[Y \mid x] \equiv \mathrm{E}[Y \mid X = x] \equiv \begin{cases} \sum_y yf(y \mid x) & \text{discrete} \\ \int_{\mathbb{R}} yf(y \mid x)dy & \text{continuous.} \end{cases}$$

注意 $\mathrm{E}[Y \mid X = x]$ 是 x 的實函數。

離散示例：考慮以下聯合 pmf。

無條件期望是 $\mathrm{E}[Y] = \sum_y yf_Y(y) = 6.15$。但條件是，例如，$X = 3$，我們有

$$f(y \mid x = 3) = \frac{f(3, y)}{f_X(3)} = \frac{f(3, y)}{0.60} = \begin{cases} \frac{34}{60} & \text{if } y = 2 \\ \frac{5}{60} & \text{if } y = 5 \\ \frac{21}{60} & \text{if } y = 10. \end{cases}$$

那麼以 $X = 3$ 為條件的期望是

$$\mathrm{E}[Y \mid X = 3] = \sum_y yf(y \mid 3) = 2\left(\frac{34}{60}\right) + 5\left(\frac{5}{60}\right) + 10\left(\frac{21}{60}\right) = 5.05$$

這與無條件期望 $\mathrm{E}[Y] = 6.15$ 相比。因此，$X = 3$ 的信息將 Y 的條件期望值推低至 5.05。

舊連續示例：讓我們考慮一下我們的好朋友有有趣的極限，

$$f(x, y) = \frac{21}{4}x^2y, \quad \text{if } x^2 \le y \le 1$$

回顧

$$f_Y(y) = \frac{7}{2}y^{5/2}, 0 \le y \le 1, \quad \text{so that} \quad \mathrm{E}[Y] = \int_{\mathbb{R}} yf_Y(y)dy = \frac{7}{9}$$

進一步回憶

$$f(y \mid x) = \frac{2y}{1 - x^4} \quad \text{if } x^2 \le y \le 1$$

因此，

$$\mathrm{E}[Y \mid x] = \int_{\mathbb{R}} yf(y \mid x)dy = \frac{2}{1 - x^4}\int_{x^2}^1 y^2dy = \frac{2}{3} \cdot \frac{1 - x^6}{1 - x^4}, \quad -1 \le x \le 1$$

因此，例如，$\mathrm{E}[Y \mid X = 0.5] = \frac{2}{3} \cdot \frac{63}{64}/\frac{15}{16} = 0.70$ (略低於 $\mathrm{E}[Y] = 7/9$)。

3.4.2 雙重期望

現在我們已經到了本章中最有趣的定理......

3.4.2.1 定理（雙重期望）：

$$\mathrm{E}[\mathrm{E}(Y \mid X)] = \mathrm{E}[Y].$$

備註：這個看起來很有趣的結果有幾個分支。

- 簡單地說：條件期望值（$Y \mid X$）的期望值（所有 X 的平均值）是普通的舊期望值（Y）。
- 將外部期望值視為 $h(X) = \mathrm{E}(Y \mid X)$ 的期望值。然後 LOTUS 奇蹟般地給了我們 $\mathrm{E}[Y]$。
- 信不信由你，有時使用我們的雙重期望技巧間接計算 $\mathrm{E}[Y]$ 會更容易。證明（連續案例）：由無意識的統計學家，

$$\begin{aligned}
\mathrm{E}[\mathrm{E}(Y \mid X)] &= \int_{\mathbb{R}} \mathrm{E}(Y \mid x) f_X(x) dx \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} y f(y \mid x) dy \right) f_X(x) dx \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} y f(y \mid x) f_X(x) dx dy \\
&= \int_{\mathbb{R}} y \int_{\mathbb{R}} f(x, y) dx dy \\
&= \int_{\mathbb{R}} y f_Y(y) dy \\
&= \mathrm{E}_Y[Y]
\end{aligned}$$

舊示例：假設 $f(x, y) = \frac{21}{4} x^2 y$，如果 $x^2 \leq y \leq 1$。我們將以 two 的方式找到 $\mathrm{E}[Y]$。為此，我們注意到，通過前面的示例，

$$\begin{aligned}
f_X(x) &= \frac{21}{8} x^2 \left(1 - x^4\right), \quad \text{for } -1 \leq x \leq 1 \\
f_Y(y) &= \frac{7}{2} y^{5/2}, \quad \text{for } 0 \leq y \leq 1 \\
\mathrm{E}[Y \mid x] &= \frac{2}{3} \cdot \frac{1 - x^6}{1 - x^4}, \quad \text{for } -1 \leq x \leq 1
\end{aligned}$$

解決方案 # 1（舊的，無聊的方式）：

$$\mathrm{E}[Y] = \int_{\mathbb{R}} y f_Y(y) dy = \int_0^1 \frac{7}{2} y^{7/2} dy = \frac{7}{9}$$

解決方案#2（新的、令人興奮的方式）：

$$\begin{aligned}
\mathrm{E}[Y] &= \mathrm{E}[\mathrm{E}(Y \mid X)] \\
&= \int_{\mathbb{R}} \mathrm{E}(Y \mid x) f_X(x) dx \\
&= \int_{-1}^1 \left(\frac{2}{3} \cdot \frac{1 - x^6}{1 - x^4} \right) \left(\frac{21}{8} x^2 \left(1 - x^4\right) \right) dx = \frac{7}{9}
\end{aligned}$$

請注意，兩個答案都是相同的（哇！）。（一世）

3.4.3 榮譽申請

本小節討論了迄今為止提出的概念的一些有趣的應用。$3.4.3.1 與我們所說的“標準條件論證”有關——這是一個強大的工具，可用於得出各種有用的結果。3.4.3.2 介紹了雙重期望的應用，包括涉及所謂的第一步分析和隨機變量的隨機和的示例。

3.4.3.1 §3.4.3.1 標準條件參數應用

我們給出了幾個例子，它們調用了稱為標準條件參數的通用方法來計算各種有趣的概率。該方法直接遵循適用於二元隨機變量的總概率定律（參見 &1.8）。

示例/定理：假設 X 和 Y 是獨立的連續 RV，分別具有 pdf $f_X(\cdot)$ 和 cdf $F_Y(\cdot)$。那麼總和 $Z=X+Y$ 有 cdf

$$\mathrm{P}(Z \leq z)=\int_{\mathbb{R}} F_Y(z-x) f_X(x) d x$$

諸如上述 $\mathrm{P}(Z \leq z)$ 的表達式通常稱為卷積。

證明：Z 的 pdf 用 $f_Z(z)$ 表示，(X, Z) 的聯合 pdf 用 $f_{X, Z}(x, z)$ 表示，條件 pdf $Z \mid X=x$ 乘以 $f_{Z \mid X}(z \mid x)$。然後我們有

$$\begin{aligned}
\mathrm{P}(Z \leq z) & =\int_{-\infty}^z f_Z(t) d t \\
& =\int_{-\infty}^z \int_{\mathbb{R}} f_{X, Z}(x, t) d x d t \\
& =\int_{\mathbb{R}} \int_{-\infty}^z f_{Z \mid X}(t \mid x) f_X(x) d t d x \quad(\text { flip integrals }) \\
& =\int_{\mathbb{R}} \mathrm{P}(Z \leq z \mid X=x) f_X(x) d x \\
& =\int_{\mathbb{R}} \mathrm{P}(X+Y \leq z \mid X=x) f_X(x) d x \\
& =\int_{\mathbb{R}} \mathrm{P}(Y \leq z-x \mid X=x) f_X(x) d x \\
& =\int_{\mathbb{R}} \mathrm{P}(Y \leq z-x) f_X(x) d x \quad(X, Y \text { are independent })
\end{aligned}$$

示例：假設 $X, Y \stackrel{\text { iid }}{\sim} \operatorname{Exp}(\lambda)$，令 $Z=X+Y$。然後

$$\begin{aligned}
\mathrm{P}(Z \leq z) & =\int_{\mathbb{R}} F_Y(z-x) f_X(x) d x \\
& =\int_0^z F_Y(z-x) f_X(x) d x \quad(z-x \geq 0, x \geq 0 \Rightarrow 0 \leq x \leq z) \\
& =\int_0^z\left(1-e^{-\lambda(z-x)}\right) \lambda e^{-\lambda x} d x \\
& =1-e^{-\lambda z}-\lambda z e^{-\lambda z}
\end{aligned}$$

因此，Z 的 pdf 為

$$\frac{d}{dz}\mathrm{P}(Z \le z) = \lambda^2 z e^{-\lambda z}, \quad z \ge 0$$

這意味著 $X+Y \sim \text{Gamma}(2,\lambda)$，也就是 $\text{Erlang}(\lambda)$。可以使用離散的 RV 進行類似類型的捲積。我們在沒有證據的情況下陳述了以下結果（無論如何，這是直截了當的）。

示例/定理：假設 X 和 Y 是兩個獨立的整數值 RV，具有 pmf 的 $f_X(x)$ 和 $f_Y(y)$。那麼 $Z=X+Y$ 的 pmf 是

$$f_Z(z) = \mathrm{P}(Z=z) = \sum_{x=-\infty}^{\infty} f_X(x) f_Y(z-x)$$

示例：假設 X 和 Y 是 iid $\text{Bern}(p)$。此外，回想一下，對於任何集合 A 和元素 z，如果 $z \in A$ 和 $1_{A}(z)=0，則指示函數 $1_A(z)=1$ 否則。那麼 $Z=X+Y$ 的 pmf 是

$$\begin{aligned} f_Z(z) = &\sum_{x=-\infty}^{\infty} f_X(x) f_Y(z-x) \\ =& f_X(0) f_Y(z) + f_X(1) f_Y(z-1) \quad (X \text{ can only } = 0 \text{ or} \\ =& f_X(0) f_Y(z) 1_{\{0,1\}}(z) + f_X(1) f_Y(z-1) 1_{\{1,2\}}(z) \\ & \left(1_{\{\cdot\}}(z) \text{ functions indicate nonzero } f_Y(\cdot)\text{ 's}\right) \\ =& p^0 q^{1-0} p^z q^{1-z} 1_{\{0,1\}}(z) + p^1 q^{1-1} p^{z-1} q^{2-z} 1_{\{1,2\}}(z) \\ =& p^z q^{2-z} \left[1_{\{0,1\}}(z) + 1_{\{1,2\}}(z)\right] \\ =& \binom{2}{z} p^z q^{2-z}, \quad z = 0, 1_{1,2}. \end{aligned}$$

因此，$Z \sim \text{Bin}(2,p)$，一個來自過去的好消息！

3.4.3.2 $3.4.3.2 雙期望應用

雙重期望可以巧妙地用於獲得有趣的發現。

示例：我們將使用“第一步”方法來計算 $Y \sim \text{Geom}\ (p)$ 的平均值。我們將 Y 視為拋硬幣的次數，直到 H 第一次出現，其中 $\mathrm{P}(\mathrm{H})=p$ 來激發這個例子。

讓我們從第一個翻轉開始，就其本身而言。如果第一次翻轉是 H，則令 $X=1$；如果 T，則令 $X=1$。當然，如果$X=1$，那麼我們立即完成，在這種情況下$Y=1$。但是，如果 $X=0$，那麼就像我們已經浪費了第一次折騰，必須重新開始練習。換句話說，$\mathrm{E}[Y \mid X=1]=1$，但是 $\mathrm{E}[Y \mid X=0]=1+\mathrm{E}[Y]$。因此，基於第一步的結果$X$，我們有

$$\begin{aligned} \mathrm{E}[Y] &= \mathrm{E}[\mathrm{E}(Y \mid X)] \\ &= \sum_x \mathrm{E}[Y \mid x] f_X(x) \\ &= \mathrm{E}[Y \mid X=0]\mathrm{P}(X=0) + \mathrm{E}[Y \mid X=1]\mathrm{P}(X=1) \\ &= (1+\mathrm{E}[Y])(1-p) + (1)(p) \end{aligned}$$

求解，我們得到 $\mathrm{E}[Y]=1/p$（這確實是正確答案）！

示例：考慮一系列擲硬幣。在 HT 首次出現之前，預期翻轉 Y 的次數是多少？顯然，$Y=A+B$，其中 A 是直到第一次出現 H 之前的翻轉次數，B 是直到第一次出現 T 之前的後續翻轉次數在 H 的序列開始之後。例如，序列 TTTHHT 對應於 $Y=A+B=4+2=6$。

無論如何，很明顯 A 和 B 是 iid $\mathrm{Geom}(p=1/2)$，所以通過前面的例子，$\mathrm{E}[Y]=\mathrm{E}[A]+\mathrm{E}[B]=(1/p)+(1/p)=4$。

前面的示例不涉及第一步分析（除了使用幾何 RV 的期望值）。但是下一個相關示例將...

示例：再次考慮一系列擲硬幣。在 HH 第一次出現之前，預期的翻轉次數是多少 Y？例如，序列 TTHTTHH 對應於 $Y=7$ 次嘗試。使用增強的第一步分析，我們看到

$$\begin{aligned}
\mathrm{E}[Y]= & \mathrm{E}[Y \mid \mathrm{T}] \mathrm{P}(\mathrm{T})+\mathrm{E}[Y \mid \mathrm{H}] \mathrm{P}(\mathrm{H}) \\
= & \mathrm{E}[Y \mid \mathrm{T}] \mathrm{P}(\mathrm{T}) \\
& \quad+\{\mathrm{E}[Y \mid \mathrm{HH}] \mathrm{P}(\mathrm{HH} \mid \mathrm{H})+\mathrm{E}[Y \mid \mathrm{HT}] \mathrm{P}(\mathrm{HT} \mid \mathrm{H})\} \mathrm{P}(\mathrm{H}) \\
= & (1+\mathrm{E}[Y])(0.5)+\{(2)(0.5)+(2+\mathrm{E}[Y])(0.5)\}(0.5) \\
& \quad(\text { since we have to start over once we see a T}) \\
= & 1.5+0.75 \mathrm{E}[Y]
\end{aligned}$$

求解，我們得到 $\mathrm{E}[Y]=6$，考慮到前面例子的結果，這可能令人驚訝。

獎勵定理（隨機數 RV 之和的期望）：假設 X_1、X_2、$\ldots$ 是獨立的 RV，均具有相同的均值。還假設 N 是一個獨立於 X_i 的非負整數值 RV。然後

$$\mathrm{E}\left[\sum_{i=1}^N X_i\right]=\mathrm{E}[N] \mathrm{E}\left[X_1\right]$$

備註：小心！特別要注意 $N $，因為左邊是一個數字，而右邊是一個隨機變量。

證明（參見 Ross [6]）：通過雙重期望，以及 N 獨立於 X_i 的事實，我們有

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^N X_i\right] & =\mathrm{E}\left[\mathrm{E}\left(\sum_{i=1}^N X_i \mid N\right)\right] \\
& =\sum_{n=1}^{\infty} \mathrm{E}\left(\sum_{i=1}^N X_i \mid N=n\right) \mathrm{P}(N=n) \\
& =\sum_{n=1}^{\infty} \mathrm{E}\left(\sum_{i=1}^n X_i \mid N=n\right) \mathrm{P}(N=n) \\
& =\sum_{n=1}^{\infty} \mathrm{E}\left(\sum_{i=1}^n X_i\right) \mathrm{P}(N=n) \quad(N \text { and } X_i \text { 's independent }) \\
& =\sum_{n=1}^{\infty} n \mathrm{E}\left[X_1\right] \mathrm{P}(N=n) \\
& =\mathrm{E}\left[X_1\right] \sum_{n=1}^{\infty} n \mathrm{P}(N=n) . \quad \square
\end{aligned}$$

示例：假設我們擲骰子的次數是 $N \sim \operatorname{Pois}(10)$。如果 X_i 表示 i^{th} 拋擲的價值，那麼所有擲骰的預期總數為

$$\mathrm{E}\left[\sum_{i=1}^N X_i\right]=\mathrm{E}[N] \mathrm{E}\left[X_1\right]=10(3.5)=35$$

定理：在與之前相同的條件下，

$$\operatorname{Var}\left(\sum_{i=1}^{N} X_i\right) = \mathrm{E}[N] \operatorname{Var}(X_1) + (\mathrm{E}[X_1])^2 \operatorname{Var}(N)$$

證明：例如，參見羅斯 6.

3.5 協方差和相關性

協方差和相關性是用於定義 X 和 Y 之間的關聯程度的度量，如果它們碰巧不是獨立的。我們將在 3.5.1 $開始討論協方差和相關的基礎知識。$3.5.2 討論了相關性和因果關係之間的關係——兩個項目相關的事實是否意味著一個導致另一個？（答案是：不一定。）我們在 $3.5.3 中詳細研究了一些數值示例。最後，第 3.5.4 節闡述了一些將在後續文章中有用的定理。

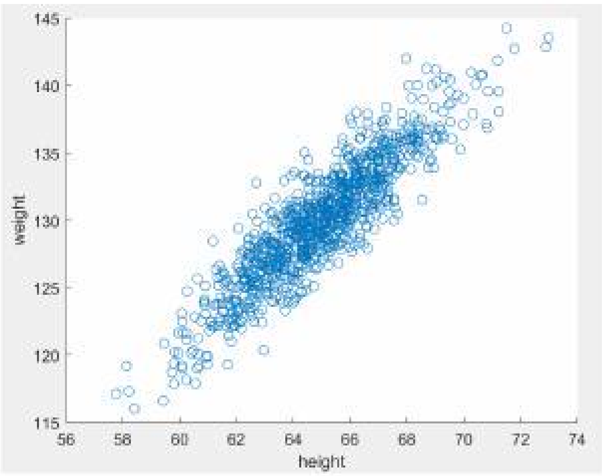
3.5.1 基礎知識

定義：X 和 Y 之間的協方差為

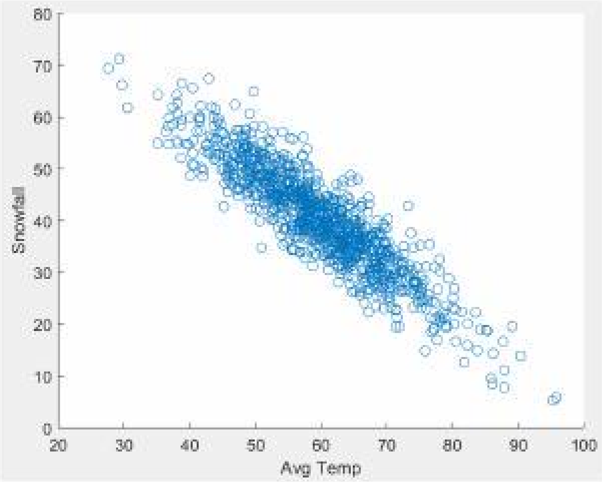
$$\operatorname{Cov}(X, Y) \equiv \sigma_{XY} \equiv \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$$

3.5.1.1 備註：

- $\operatorname{Cov}(X, X) = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = \operatorname{Var}(X)$
- 如果 X 和 Y 具有正協方差，那麼 X 和 Y 會"朝同一方向"移動。想想身高和體重。

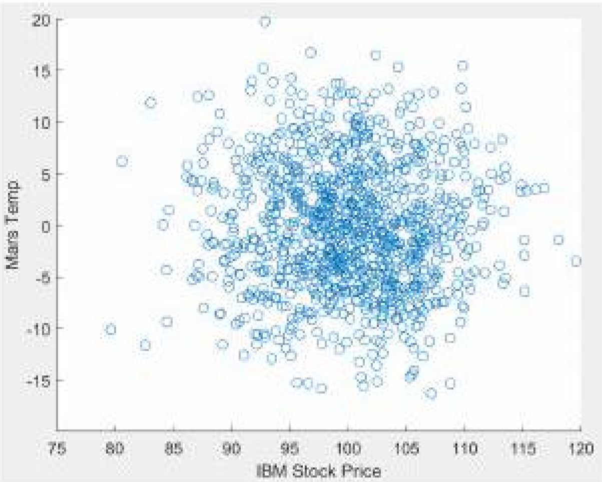


- 如果 X 和 Y 具有負協方差，那麼 X 和 Y 會"朝相反的方向"移動。想想降雪和溫度。



- 如果X 和Y 是獨立的，那麼它們之間當然沒有關聯。事實上，我們將在下面證明獨立性意味著協方差為 0（但不是相反）。

示例：IBM 股票價格與火星上的溫度是獨立的。（至少那是他們希望你相信的！）



定理（計算協方差的更簡單方法）：

$$\mathrm{Cov}(X,Y) = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y].$$

3.5.1.2 證明：

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{E}[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])] \\ &= \mathrm{E}[XY - X\mathrm{E}[Y] - Y\mathrm{E}[X] + \mathrm{E}[X]\mathrm{E}[Y]] \\ &= \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] - \mathrm{E}[Y]\mathrm{E}[X] + \mathrm{E}[X]\mathrm{E}[Y].\end{aligned}$$

定理：X 和 Y 獨立意味著 $\mathrm{Cov}(X,Y)=0$。

證明：根據 \$3.3.2 的一個定理，我們記得 X 和 Y 獨立意味著 $\mathrm{E}[XY]=\mathrm{E}[X]\mathrm{E}[Y]$。然後

$$\mathrm{Cov}(X,Y) = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] = \mathrm{E}[X]\mathrm{E}[Y] - \mathrm{E}[X]\mathrm{E}[Y].$$

危險威爾羅賓遜！$\mathrm{Cov}(X,Y)=0$ 並不意味著 X 和 Y 是獨立的！

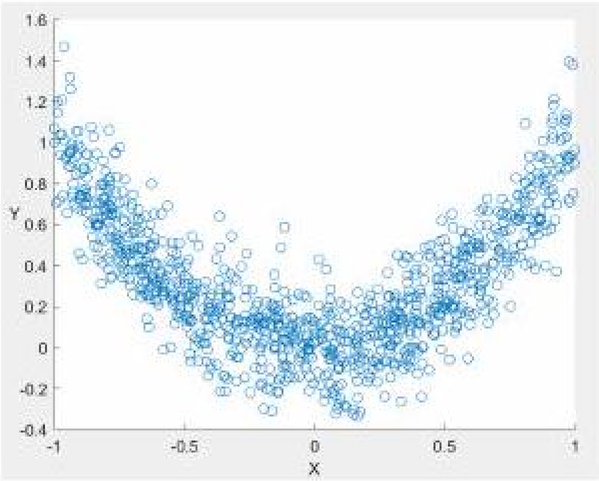
示例：假設 $X \sim \mathrm{Unif}(-1,1)$，並且 $Y = X^2$（因此 X 和 Y 顯然是依賴的）。但

$$\mathrm{E}[X]=\int_{-1}^{1} x \cdot \frac{1}{2} d x=0 \text { and }$$
$$\mathrm{E}[X Y]=\mathrm{E}\left[X^{3}\right]=\int_{-1}^{1} x^{3} \cdot \frac{1}{2} d x=0 .$$

所以

$$\operatorname{Cov}(X, Y)=\mathrm{E}[X Y]-\mathrm{E}[X] \mathrm{E}[Y]=0 . \quad ㄨ$$

這是這種零協方差依賴現象的圖形說明，我們實際上在 Y 中添加了一些正常噪聲以使其看起來更漂亮。



現在，關於協方差的兄弟。

定義：X 和 Y 之間的相關性為

$$\rho \equiv \operatorname{Corr}(X, Y) \equiv \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X) \operatorname{Var}(Y)}}=\frac{\sigma_{X Y}}{\sigma_{X} \sigma_{Y}} .$$

備註：協方差具有“平方”單位，而相關性是無單位的。事實上，相關性可以看作是協方差的“標準化”版本。

定理：可以證明 $-1 \leq \rho \leq 1$

（先前定理的）推論：X 和 Y 獨立意味著 $\rho=0$。

備註：$\rho \doteq 1$ 被視為“高”相關性；$\rho \doteq 0$ 是“低”；和 $\rho \doteq -1$ 是“高”負相關。

示例：身高與體重高度相關。火星溫度與 IBM 股價的相關性較低。

3.5.2 相關性和因果關係

筆記！相關並不一定意味著因果關係！這是數據分析和公共話語的許多領域中非常常見的陷阱。我們提供一些例子來說明相關問題。

相關性確實暗示因果關係的示例：如前所述，身高和體重呈正相關。在這種情況下，似乎更高的高度確實會導致更大的重量。

相關性不代表因果關係的例子：溫度與檸檬水銷量呈正相關，可以說溫度對檸檬水銷量有一定的因果影響。同樣，溫度和汽車過熱也與因果關係呈正相關。檸檬水銷量和汽車過熱也很可能正相關，但顯然沒有因果關係。

與因果關係的零相關關係示例！我們之前看到，兩個因隨機變量可能具有零相關性。

為了證明 X 導致 Y，必須證明：

- X 發生在 Y 之前（在某種意義上）。
- X 和Y 之間的關係並不完全是隨機的。
- 沒有其他原因可以解釋這種關係（在上面的檸檬水銷售/過熱汽車示例中違反了這一點）。

這些項目通常可以通過數學分析、適當數據的統計分析或諮詢適當的專家來確定。

關於相關性和因果關係之間的關係，上述三個例子似乎給出了相互矛盾的指導。我們如何才能以有意義的方式解釋這些發現？以下是要點： - 如果 X 和 Y 之間的相關性（顯著）非零，那麼這兩個項目之間存在某種類型的關係，這可能是因果關係，也可能不是因果關係，但這應該引起我們的好奇心。

- 如果 X 和 Y 之間的相關性為 0 ，那麼在依賴和因果關係方面我們還沒有完全脫離困境。為了明確排除 X 和 Y 之間的關係，始終強烈建議協議至少，
- 繪製來自 X 和 Y 的數據，以查看是否存在非線性關係，如在零相關但相關示例中。
- 諮詢適當的專家。

3.5.3 幾個工作的數值例子

離散示例：假設 X 是大學生的 GPA，Y 是學生每週在社交媒體上的平均小時數。這是聯合pmf。

我們可以很容易地計算

$$
\begin{gathered}
\mathrm{E}[X]=\sum_x x f_X(x)=2.7 \\
\mathrm{E}\left[X^2\right]=\sum_x x^2 f_X(x)=7.9, \quad \text{and} \\
\operatorname{Var}(X)=\mathrm{E}\left[X^2\right]-(\mathrm{E}[X])^2=0.61 .
\end{gathered}
$$

同樣，$\mathrm{E}[Y]=50$，$\mathrm{E}\left[Y^2\right]=2580$，$\operatorname{Var}(Y)=80$. 最後，

嗯……在這個例子中，更多的社交媒體似乎與較低的成績相關（儘管我們實際上還沒有證明因果關係）。

連續示例：假設 $f(x, y)=10 x^2 y$，對於 $0 \leq y \leq x \leq 1$。

$$
\begin{aligned}
& \mathrm{E}[X Y]=\sum_x \sum_y x y f(x, y) \\
& =2(40)(0.0)+3(40)(0.2)+\cdots+4(60)(0.0)=129, \\
& \operatorname{Cov}(X, Y)=\mathrm{E}[X Y]-\mathrm{E}[X] \mathrm{E}[Y]=-6.0, \text{ and } \\
& \rho=\frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X) \operatorname{Var}(Y)}}=-0.859
\end{aligned}
$$

請注意，有趣的限制通常會導致非零相關性。我們有

$$\begin{aligned}
f_X(x) &= \int_0^x 10x^2 y dy = 5x^4, 0 \leq x \leq 1 \\
\mathrm{E}[X] &= \int_0^1 5x^5 dx = 5/6 \\
\mathrm{E}\left[X^2\right] &= \int_0^1 5x^6 dx = 5/7, \quad \text{and} \\
\operatorname{Var}(X) &= \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = 0.01984
\end{aligned}$$

相似地，

$$\begin{aligned}
& f_Y(y) = \int_y^1 10x^2 y dx = \frac{10}{3} y\left(1-y^3\right), \quad 0 \leq y \leq 1 \\
& \mathrm{E}[Y] = 5/9, \quad \text{and } \operatorname{Var}(Y) = 0.04850.
\end{aligned}$$

最後，

3.5.4 涉及協方差的其他有用定理

在本小節中，我們將討論幾個稍後將非常有用的定理。

定理：X 和 Y 是否獨立，

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X, Y)$$

備註：如果 X 和 Y 是獨立的，那麼協方差項就消失了。但是如果 X 和 Y 是相關的，不要忘記協方差項！

證明：通過我們在之前的證明中所做的工作，

$$\begin{aligned}
\operatorname{Var}(X+Y) &= \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 + \mathrm{E}\left[Y^2\right] - (\mathrm{E}[Y])^2 + 2(\mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]) \\
&= \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X, Y).
\end{aligned}$$

這個結果可以概括...

3.5.4.1 定理：

$$\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{Var}(X_i) + 2\sum\sum_{i<j} \operatorname{Cov}\left(X_i, X_j\right)$$

$$\begin{aligned}
& \mathrm{E}[XY] = \int_{\mathbb{R}} \int_{\mathbb{R}} xy f(x, y) dx dy = \int_0^1 \int_0^x 10x^3 y^2 dy dx = 10/21, \\
& \operatorname{Cov}(X, Y) = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] = 0.01323, \quad \text{and} \\
& \rho = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} = 0.4265.
\end{aligned}$$

證明：歸納。

推論（恰好是來自 $3.3.2 的老朋友：如果所有 X_i 都是獨立的，那麼

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

現在我們得到了 X 和 Y 的線性泛函協方差的結果。

定理：$\mathrm{Cov}(aX, bY + c) = ab\,\mathrm{Cov}(X, Y)$。

證明：

$$\begin{aligned}\mathrm{Cov}(aX, bY + c) &= \mathrm{E}[aX \cdot (bY + c)] - \mathrm{E}[aX]\mathrm{E}[bY + c] \\ &= \mathrm{E}[abXY] + \mathrm{E}[acX] - \mathrm{E}[aX]\mathrm{E}[bY] - \mathrm{E}[aX]\mathrm{E}[c] \\ &= ab\mathrm{E}[XY] - ab\mathrm{E}[X]\mathrm{E}[Y] + ac\mathrm{E}[X] - ac\mathrm{E}[X] \\ &= ab\,\mathrm{Cov}(X, Y). \quad \square\end{aligned}$$

將上述兩個定理放在一起，我們得到了一個非常普遍的結果。

定理：

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i + c\right) = \sum_{i=1}^{n} a_i^2\,\mathrm{Var}(X_i) + 2\sum_{i<j} a_i a_j\,\mathrm{Cov}\big(X_i, X_j\big)$$

示例：這是一個有用的 $\mathrm{Var}(XY)$ 表達式，在統計和計算機模擬應用程序中經常出現。

$$\begin{aligned}\mathrm{Var}(X - Y) &= (1)^2\,\mathrm{Var}(X) + (-1)^2\,\mathrm{Var}(Y) + 2(1)(-1)\,\mathrm{Cov}(X, Y) \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\,\mathrm{Cov}(X, Y).\end{aligned}$$

示例：如果 $\mathrm{Var}(X) = \mathrm{Var}(Y) = \mathrm{Var}(Z) = 10, \mathrm{Cov}(X, Y) = 3, \mathrm{Cov}(X, Z) = -2$，$\mathrm{Cov}(Y, Z) = 0$，則

$$\begin{aligned}&\mathrm{Var}(X - 2Y + 3Z) \\ &= \quad (1)^2\,\mathrm{Var}(X) + (-2)^2\,\mathrm{Var}(Y) + (3)^2\,\mathrm{Var}(Z) \\ &\quad + 2(1)(-2)\,\mathrm{Cov}(X, Y) + 2(1)(3)\,\mathrm{Cov}(X, Z) + 2(-2)(3)\,\mathrm{Cov}(Y, Z) \\ &= 10 + 4(10) + 9(10) - 4(3) + 6(-2) - 12(0) = 116. \quad \square\end{aligned}$$

3.6 矩生成函數，再訪

我們的矩生成函數 (mgf) 汽車還剩一些里程。現在我們對聯合分佈有了一些了解，我們將進行另一個驅動。

舊定義：$M_X(t) \equiv \mathrm{E}\left[e^{tX}\right]$ 是 $\mathrm{RV}X$ 的 mgf。

舊示例：如果 $X \sim \mathrm{Bern}(p)$，那麼

$$M_X(t) = \mathrm{E}\left[e^{tX}\right] = \sum_x e^{tx} f(x) = e^{t \cdot 1} p + e^{t \cdot 0} q = pe^t + q$$

舊示例：如果 $X \sim \mathrm{Exp}(\lambda)$，那麼

$$M_X(t) = \mathrm{E}\left[e^{tX}\right] = \int_{\mathbb{R}} e^{tx} f(x) dx = \frac{\lambda}{\lambda - t} \quad \text{if } \lambda > t$$

舊定理（為什麼稱為 mgf）：在某些技術條件下，

$$\mathrm{E}\left[X^k\right]=\left.\frac{d^k}{dt^k}M_X(t)\right|_{t=0},\quad k=1,2,\ldots$$

新定理：假設 X_1、$\ldots$、X_n 是獨立的。令$Y=\sum_{i=1}^n X_i$。那麼總和的 mgf 是 mgf 的乘積，

$$M_Y(t)=\prod_{i=1}^n M_{X_i}(t)$$

證明：我們有

$$M_Y(t)=\mathrm{E}\left[e^{tY}\right]=\mathrm{E}\left[e^{t\sum_{i=1}^n X_i}\right]=\mathrm{E}\left[\prod_{i=1}^n e^{tX_i}\right]=\prod_{i=1}^n \mathrm{E}\left[e^{tX_i}\right]$$

最後一步是因為 X_I 是獨立的。

推論：如果X_1、$\ldots$、X_n是獨立同分佈的，並且$Y=\sum_{i=1}^n X_i$，那麼$M_Y(t)=\left[M_{X_1}(t)\right]^n$。

示例：假設 $X_1,\ldots,X_n \overset{\text{iid}}{\sim} \operatorname{Bern}(p)$. 然後，通過前面的示例，

$$M_Y(t)=\left[M_{X_1}(t)\right]^n=\left(pe^t+q\right)^n.$$

我們可以和我們的老朋友一起使用這樣的結果……

舊定理（識別分佈）：在本書中，每個分佈都有一個唯一的 mgf。

證明：不在這裡。

示例/定理：n iid $\operatorname{Bern}(p)$ 隨機變量的總和 Y 與 $\operatorname{Bin}(n,p)$ 隨機變量具有相同的分佈。

證明：通過前面的例子和唯一性，我們只需要證明 $Z\sim\operatorname{Bin}(n,p)$ 的 mgf 匹配 $M_Y(t)=\left(pe^t+q\right)^n$。為此，我們有

$$\begin{aligned}
M_Z(t)&=\mathrm{E}\left[e^{tZ}\right]\\
&=\sum_z e^{tz}\mathrm{P}(Z=z)\\
&=\sum_{z=0}^n e^{tz}\binom{n}{z}p^z q^{n-z}\\
&=\sum_{z=0}^n \binom{n}{z}\left(pe^t\right)^z q^{n-z}\\
&=\left(pe^t+q\right)^n\quad(\text{ by the Binomial Theorem})
\end{aligned}$$

示例：您可以通過其 mgf 來識別分佈。例如，$M_X(t)=\left(\frac{3}{4}e^t+\frac{1}{4}\right)^{15}$ 意味著 $ X (15,0.75)$。

舊定理（X 的線性函數的 mgf）：假設 X 有 mgf $M_X(t)$，並令 $Y=aX+b$。那麼 $M_Y(t)=e^{tb}M_X(at)$。

示例：假設 RV Y 具有以下 mgf，

$$M_Y(t)=e^{-2t}\left(\frac{3}{4}e^{3t}+\frac{1}{4}\right)^{15}=e^{bt}\left(pe^{at}+q\right)^n,$$

其中 $a=3, b=-2, p=3/4, q=1/4, n=15$。通過前面的例子和定理，我們立即得到

$$M_Y(t)=e^{bt}M_X(at),$$

其中 $X\sim\operatorname{Bin}(n,p)\sim\operatorname{Bin}(15,0.75)$。因此，$Y=aX+b=3X-2$，因此 Y 是一個“移位的”二項式 RV。

定理（二項式的加性性質）：如果 $X_1,\ldots,X_k$ 是獨立的，則 $X_i\sim\operatorname{Bin}(n_i,p)$（其中 p 對於所有 X_i 都是相同的），然後

$$Y\equiv\sum_{i=1}^k X_i\sim\operatorname{Bin}\left(\sum_{i=1}^k n_i,p\right).$$

證明：我們有

$$\begin{aligned}M_Y(t)&=\prod_{i=1}^k M_{X_i}(t)\quad(\text{ mgf of independent sum })\\&=\prod_{i=1}^k\left(pe^t+q\right)^{n_i}\quad(\operatorname{Bin}(n_i,p)\text{mgf})\\&=\left(pe^t+q\right)^{\sum_{i=1}^k n_i}.\end{aligned}$$

這是 $\operatorname{Bin}\left(\sum_{i=1}^k n_i,p\right)$ 的 mgf，所以我們完成了。

備註：您可以使用 mgf 技術來添加很多東西。以下是一些示例（在沒有證明或相關參數的情況下說明）。

- iid Geom (p) 隨機變量的總和是負二項式隨機變量（參見 \$4.1.32）。
- 獨立泊鬆的總和是另一個泊松（參見練習 3.821。
- iid Exp $(\lambda)'$ 的總和是 Erlang - 一種伽馬分佈（參見 \$4.2.2.2)
- 獨立法線的總和是另一個法線（見 4.3.1。
- 還有很多…

3.7 隨機變量的二元函數

3.7.1 導論和基本理論

到目前為止，我們已經研究了一個或多個隨機變量的函數的各種屬性：

- 單個變量的函數，例如，$h(X)$ 的期望值是多少？（這是 LOTUS，來自 §\$2.5.2 和 2.8.3 ）
- $h(X)$ 的分佈是什麼？（參見 \$2.8中關於RV功能的討論）

- 有時甚至是兩個（或更多）變量的函數。例如，如果 X_i 是獨立的，那麼 $\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right)$ 是什麼？（看看 \$3.3.3 ）
- 使用標準條件參數來獲得 $X+Y$ 的分佈。（見 \$3.4.3。）

目標：現在讓我們給出兩個隨機變量函數分佈的一般結果，其證明超出了本文的範圍。

榮譽定理：假設 X 和 Y 是具有聯合 pdf $f(x, y)$ 的連續隨機變量，並且 $V=h_1(X, Y)$ 和 $W=h_2(X, Y)$ 是 X 和 Y 的函數，其中

$$X=k_1(V, W) \text { and } Y=k_2(V, W),$$

對於適當選擇的反函數 k_1 和 k_2。那麼 V 和 W 的聯合 pdf 是

$$g(v, w)=f\left(k_1(v, w), k_2(v, w)\right)|J(v, w)|,$$

其中 $|J|$ 是變換的雅可比行列式（行列式）的絕對值，即

$$J(v, w)=\left|\begin{array}{ll}\frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w}\end{array}\right|=\frac{\partial x}{\partial v} \frac{\partial y}{\partial w}-\frac{\partial y}{\partial v} \frac{\partial x}{\partial w} .$$

哇，那是一口！

推論：如果X和Y是獨立的，那麼V和W的聯合pdf是

$$g(v, w)=f_X\left(k_1(v, w)\right) f_Y\left(k_2(v, w)\right)|J(v, w)| .$$

備註：這些結果概括了從\$2.8開始的一維方法實際上，你可以用這個方法找到各種很酷的東西，例如 $X+Y$的分佈，X / Y等.，以及 X 和 Y 的任何函數的聯合 pdf。

備註：雖然符號很討厭，但應用程序並沒有那麼糟糕。

3.7.2 例子

示例：假設 X 和 Y 是 iid $\operatorname{Exp}(\lambda)$。找到 $X+Y$ 的 pdf。

技巧：我們將設置 $V=X+Y$，以及“虛擬”隨機變量 $W=X$。這產生

$$X=W=k_1(V, W), \text { and } Y=V-W=k_2(V, W) .$$

為了得到雅可比項，我們計算 $\frac{\partial x}{\partial v}=0, \frac{\partial x}{\partial w}=1, \frac{\partial y}{\partial v}=1$ 和 $\frac{\partial y}{\partial w}=-1$，所以

$$|J(v, w)|=\left|\frac{\partial x}{\partial v} \frac{\partial y}{\partial w}-\frac{\partial y}{\partial v} \frac{\partial x}{\partial w}\right|=|0(-1)-1(1)|=1 .$$

這意味著 V 和 W 的聯合 pdf 是

$$\begin{aligned} g(v, w) & =f\left(k_1(v, w), k_2(v, w)\right)|J(v, w)| \\ & =f(w, v-w) \cdot 1 \\ & =f_X(w) f_Y(v-w) \quad(X \text { and } Y \text { independent }) \\ & =\lambda e^{-\lambda w} \cdot \lambda e^{-\lambda(v-w)}, \quad \text { for } w>0 \text { and } v-w>0 \\ & =\lambda^2 e^{-\lambda v}, \quad \text { for } 0<w<v .\end{aligned}$$

最後，我們獲得總和 V 的所需 pdf（在仔細注意積分區域之後），

$$g_V(v)=\int_{\mathbb{R}} g(v,w)dw=\int_0^v \lambda^2 e^{-\lambda v}dw=\lambda^2 v e^{-\lambda v},\quad \text{for } v>0.$$

這是 $\text{Gamma}(2,\lambda)$ pdf，它與我們在 \$3.4.3 中的答案相匹配

3.7.2.1 榮譽示例

假設 X 和 Y 是 iid $\text{Unif}(0,1)$。求 $V=X+Y$ 和 $W=X/Y$ 的聯合 pdf。經過一些代數，我們得到

$$X=\frac{VW}{W+1}=k_1(V,W),\quad \text{and}\quad Y=\frac{V}{W+1}=k_2(V,W).$$

經過更多的代數，我們計算

$$\frac{\partial x}{\partial v}=\frac{w}{w+1},\quad \frac{\partial x}{\partial w}=\frac{v}{(w+1)^2},\quad \text{and}\quad \frac{\partial y}{\partial v}=\frac{1}{w+1},\quad \frac{\partial y}{\partial w}=\frac{-v}{(w+1)^2},$$

所以在更多的代數之後，

$$|J|=\left|\frac{\partial x}{\partial v}\frac{\partial y}{\partial w}-\frac{\partial y}{\partial v}\frac{\partial x}{\partial w}\right|=\frac{v}{(w+1)^2}$$

這意味著 V 和 W 的聯合 pdf 是

$$\begin{aligned}
g(v,w)&=f\left(k_1(v,w),k_2(v,w)\right)|J(v,w)|\\
&=f\left(\frac{vw}{w+1},\frac{v}{w+1}\right)\cdot\frac{v}{(w+1)^2}\\
&=f_X\left(\frac{vw}{w+1}\right)f_Y\left(\frac{v}{w+1}\right)\frac{v}{(w+1)^2}\quad (X \text{ and } Y \text{ independent })\\
&=1\cdot 1\cdot\frac{v}{(w+1)^2},\ \text{for } 0<x,y<1\quad (\text{ since } X,Y\sim \text{Unif}(0,1))\\
&=\frac{v}{(w+1)^2},\ \text{for } 0<x=\frac{vw}{w+1}<1,\ \text{and } 0<y=\frac{v}{w+1}<1\\
&=\frac{v}{(w+1)^2},\ \text{for } 0<v<1+\min\left\{\frac{1}{w},w\right\},\ \text{and } w>0\ (\text{algebra!}).
\end{aligned}$$

請注意，您必須小心 v 和 w 的限制，但這個東西確實會雙重積分到 1 ！

我們還可以獲得邊際 pdf。首先，對於製服的比例，我們有

$$\begin{aligned}
g_W(w)&=\int_{\mathbb{R}} g(v,w)dv\\
&=\int_0^{1+\min\{1/w,w\}}\frac{v}{(w+1)^2}dv\\
&=\frac{(1+\min\{1/w,w\})^2}{2(w+1)^2}\\
&=\begin{cases}\frac{1}{2}, & \text{if } w\le 1\\ \frac{1}{2w^2}, & \text{if } w>1\end{cases}
\end{aligned}$$

這對我來說有點奇怪和出乎意料（對於 $w \leq 1$ 是平坦的，然後對於 $w > 1$ 很快減少到 0）。

對於製服總和的 pdf，我們必須計算 $g_V(v) = \int_{\mathbb{R}} g(v,w)dw$。但首先我們需要處理一些不等式約束，以便我們可以在適當的區域上進行整合，即，

$$0 \leq v \leq 1 + \min\{1/w, w\}, \quad 0 \leq v \leq 2, \quad \text{and} \quad w \geq 0.$$

稍加思考，我們看到如果 $0 \leq v \leq 1$，那麼除了它是正數之外，對 w 沒有任何約束。另一方面，如果 $1 < v \leq 2$，那麼你可以證明（這需要一點工作）$v - 1 \leq w \leq \frac{1}{v-1}$。因此，我們有

$$\begin{aligned} g_V(v) &= \begin{cases} \int_0^\infty g(v,w)dw, & \text{if } 0 \leq v \leq 1 \\ \int_{v-1}^{1/(v-1)} g(v,w)dw, & \text{if } 1 < v \leq 2 \end{cases} \\ &= \begin{cases} v, & \text{if } 0 \leq v \leq 1 \\ 2 - v, & \text{if } 1 < v \leq 2 \end{cases} \quad \text{(after algebra)} \end{aligned}$$

這是一個三角形 $(\mathbf{0}, \mathbf{1}, \mathbf{2})$ pdf。你能看出為什麼嗎？這個pdf有直觀的解釋嗎？

3.7.2.2 現在我們的讚助商說一句話...

我們終於完成了可能是文本中最困難的材料！祝賀和祝賀！從現在開始，事情會變得更容易！快樂的日子又來了！–(

3.8 練習

1. ($3.1) 假設 $f(x,y) = 6x$，對於 $0 \leq x \leq y \leq 1$。

a. 求 $\mathrm{P}(X < 1/2$ 和 $Y < 1/2)$。

b. 找到 X 的邊際 pdf $f_X(x)$。

2. ($3.1) 假設 X 和 Y 是具有以下聯合 pmf 的離散隨機變量，其中任何字母表示您需要計算的概率。

a. 用正確的字母值完成表格。

b. 求 $\mathrm{P}(X \leq 1)$

c. 求 $\mathrm{P}(X = 5 \mid Y = 27)$。

3. ($3.2) 假設 $f(x,y) = cxy^2$，對於 $0 \leq x \leq y^2 \leq 1$ 和 $0 \leq y \leq 1$。

a. 找到c。

b. 求 X的邊際pdf，$f_X(x)$。

c. 求 Y的邊際pdf，$f_Y(y)$。

d. 求 $\mathrm{E}[X]$。

e. 求 $\mathrm{E}[Y]$。

f. 給定$Y = y$，即$f(x \mid y)$，求X 的條件pdf。

4. (§3.2) 下表給出了兩個隨機變量的聯合 pmf $f(x, y)$：X（大學生的 GPA）和 Y（學生每週的平均小時數）在社交媒體上花費）。

a. 隨機學生在社交媒體上花費 50 小時的概率是多少？(b) 一個隨機學生每週在社交媒體上花費 50 小時的條件概率是多少，假設他的 $\mathrm{GPA} = 2$？

5.（第 3.3 節）考慮隨機變量 X 和 Y 具有現在熟悉的聯合 pmf

X 和 Y 是獨立的嗎？

6.（第 3.3 節）假設 $f(x, y) = c(x + y)$ 對於 $0 < x < y < 1$，對於適當的常數 c。X 和 Y 是獨立的嗎？

7. (§3.3) 如果 $\mathrm{E}[X] = 7, \mathrm{E}[Y] = -3, \mathrm{Var}(X) = 1, \mathrm{Var}(Y); = 9$，且$X$和$Y$是獨立的，求$\mathrm{Var}(X + Y)$。

8. (§3.3) 我們有兩個品牌的燈泡。品牌 X 的 $\mathrm{Exp}(\mu = 1)$ 壽命平均為 1 年。品牌 Y 的 $\mathrm{Exp}(\lambda = 1/2)$ 壽命平均為 2 年。假設 X 和 Y 是獨立的，那麼 X 將持續超過 Y 的機會是多少？

9. ($3.3) 假設 $X_1, \ldots, X_{\underline{n}} \stackrel{\text{iid}}{\sim} \text{Poisson}(5)$ 與 pmf $\mathrm{P}(X_i = x) = e^{-5} 5^x / x!, x = 0, 1, 2, \ldots$ 令 $\bar{X}_n = \sum_{i=1}^n X_i / n$ 表示 X_i 的樣本均值。什麼是 $\mathrm{E}[\bar{X}], \mathrm{Var}(\bar{X}), \lim_{n \to \infty} \mathrm{E}[\bar{X}]$ 和 $\lim_{n \to \infty} \mathrm{Var}(\bar{X})$？

10. ($3.4) $\mathrm{E}[Y \mid x]$ 的合法表達式是什麼？

a. E[Y]

b. $\int_{\mathbb{R}} y f(y \mid x) dy$

c. $\int_{\mathbb{R}} y f(y \mid x) dx$

d. $\frac{1}{fX(x)} \int_{\mathbb{R}} y f(x, y) dy$

e. b. 和 (d)

11. ($3.4) 假設

$$f(x, y) = 6x \quad \text{for } 0 \leq x \leq y \leq 1$$

您可能已經在某個地方看到 X 的邊際 pdf 原來是

$$f_X(x) = 6x(1 - x) \text{ for } 0 \leq x \leq 1$$

a. 給定 $X = x$，求 Y 的條件 pdf。

b. 求 $\mathrm{E}[Y \mid X = x]$。

c. 求 $\mathrm{E}[\mathrm{E}[Y \mid X]]$。12.（§3.4）回到問題 3

d. 求條件期望 $\mathrm{E}[X \mid y]$。

e. 求雙條件期望 $\mathrm{E}[\mathrm{E}[X \mid Y]]$。

13. (§3.4) 考慮問題 4 的設置，涉及兩個隨機變量的聯合 pmf $f(x, y)$：X（大學生的 GPA）和 Y（平均小時數）每週花在社交媒體上的時間），以及由此產生的邊緣。

什麼是 $\mathrm{E}[\mathrm{E}[Y \mid X]]$ ？

14. (§3.4) 考慮抛硬幣的序列——H 的概率是 $1/3$，T 是 $2/3$。在第一次出現“HT”之前，翻轉次數 Y 的期望值是多少？（例如，序列 TTHHT 對應於 $Y = 5$。）

15.（§3.4）我是一個好賭徒。當我玩遊戲時，我要么以 0.4 $的概率輸掉 10 $，要么以 0.6 $的概率輸掉 20 $。假設我在累了回家之前玩了 Poisson ($\lambda = 9.5$) 隨機數的遊戲。此外，假設一切都是獨立的（即所有遊戲以及遊戲數量）。我回家時預期的獎金是多少？

16. (§3.5) 假設 12 月降雪量與紐約州 Siberacuse 氣溫之間的相關性為 -0.5。進一步假設在 ^{2} 和 $\operatorname{Var}(T) = 25(\text{ degree } F)^2$ 中 $\operatorname{Var}(S) = 100$。找到 $\operatorname{Cov}(S, T)$（以度英寸為單位，不管是什麼）。

17.（§3.5）對還是錯？如果 X 和 Y 正相關，並且 Y 和 Z 正相關，那麼一定是 X 導致 Z，反之亦然。

18. (§3.5) 考慮以下兩個離散 RVsX 和 Y 的聯合 pmf。

找出 X 和 Y 之間的相關性。

19. (§3.5) 如果 X 和 Y 都具有均值 -7 和方差 4 ，並且 $\operatorname{Cov}(X, Y) = 1$，求 $\operatorname{Var}(3XY)$。

20. (§3.5) 令 $\operatorname{Var}(X) = \operatorname{Var}(Y) = 10, \operatorname{Var}(Z) = 20, \operatorname{Cov}(X, Y) = 2$、 $\operatorname{Cov}(X, Z) =-3$，$\operatorname{Cov}(Y, Z) = -4$。找到 $\operatorname{Corr}(X, Y)$ 和 $\operatorname{Var}(X - 2Y + 5Z)$。21. ($3.6) 假設 $X_1, X_2, \ldots, X_n$ 是具有 $X_i \sim \operatorname{Pois}(\text{ } lambda_i)$。使用 mgf 證明 $Y = \sum_{i=1}^n X_i \sim \operatorname{Pois}\left(\sum_{i=1}^n \lambda_i\right)$。

21. ($3.6) 識別隨機變量 Y 的概率分佈，如果它具有 mgf $M_Y(t) = \left(0.3e^t + 0.7\right)^4$

22. ($3.7) 假設$X \sim \operatorname{Exp}(\lambda)$，$Y \sim \operatorname{Exp}(\mu)$，並且$X$和$Y$是獨立的。找到 $X + Y$ 的 pdf。

ryNoSmBk

ryCh 4 【第 4 章】

4.0.0.1 分佈

本章將討論許多有趣的概率分佈。我們將提供結果概要，其中一些我們將證明，而其中一些我們已經看到。將特別強調正態分佈，因為它是如此重要且具有如此多的含義，包括中心極限定理。

4.1 - 離散分佈

4.2 - 連續分配

4.3 - 正態分佈和中心極限定理

4.4 - 正態分佈的擴展

4.5 $ - 計算機注意事項

4.1 離散分佈

4.1.1 伯努利分佈和二項分佈

我們將從最基本的發行版開始。

定義：伯努利 (p) 分佈由下式給出

$$X=\begin{cases}1 & \text{w.p. } p \quad (\text{" success" }) \\ 0 & \text{w.p. } q=1-p \quad (\text{"failure" })\end{cases}$$

回想一下：$\mathrm{E}[X]=p$，$\operatorname{Var}(X)=pq$，矩生成函數 (mgf) $M_{X}(t)=pe^{t}+q $

定義：參數為 n 和 p 的二項式分佈由概率質量函數 (pmf) 給出

$$\mathrm{P}(Y=k)=\binom{n}{k} p^{k} q^{n-k}, \quad k=0,1, \ldots, n$$

定理：$X_{1}, \ldots, X_{n} \stackrel{\text{iid}}{\sim} \operatorname{Bern}(p) \Rightarrow Y \equiv \sum_{i=1}^{n} X_{i} \sim \operatorname{Bin}(n, p)$。

證明：這種結果可以很容易地通過矩生成函數唯一性論證來證明，正如我們在 $3.6 中提到的

將 $\operatorname{Bin}(n, p)$ 視為 $n \operatorname{Bern}(p)$ 試驗的成功次數。

示例：擲兩個骰子並取和；重複實驗五次。讓 Y 是您看到的 7 的數量。每個實驗都可以看作是一次 $\operatorname{Bern}(p)$ 試驗，$p=\mathrm{P}(\text{Sum}=7)=1/6$，因此 $Y \sim \operatorname{Bin}(5,1/6)$。然後，例如，

$$\mathrm{P}(Y=4)=\binom{5}{4}\left(\frac{1}{6}\right)^{4}\left(\frac{5}{6}\right)^{5-4}.$$

定理：$Y \sim \operatorname{Bin}(n, p)$ 蘊含

$$\mathrm{E}[Y] = \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{E}\left[X_i\right] = np,$$

同樣，

$$\mathrm{Var}(Y) = npq.$$

我們在 \$3.6 中看到 $M_Y(t) = \left(pe^t + q\right)^n$。

定理：某些二項式相加：如果 $Y_1, \ldots, Y_k$ 是獨立的並且 $Y_i \sim \mathrm{Bin}(n_i, p)$，然後

$$\sum_{i=1}^{k} Y_i \sim \mathrm{Bin}\left(\sum_{i=1}^{k} n_i, p\right).$$

4.1.2 超幾何分佈

定義：假設您有類型 1 的 a 對象和類型 2 的 b 對象。從 $a+b$ 中選擇 n 個不替換的對象。令 X 為選擇的類型 1 的數量。然後 X 具有 pmf 的超幾何分佈

$$\mathrm{P}(X=k) = \frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}}, \quad k = \max\{0, n-b\}, \ldots, \min\{a, n\}.$$

pmf的支持看起來有點複雜，但是如果$n \leq \min\{a, b\}$, 那麼$k = 0, 1, \ldots, n$。

示例：假設一個盒子裡有 25 只襪子——15 隻紅色和 10 只藍色。選擇七個沒有替換。

$$\mathrm{P}(\text{ exactly three reds are picked }) = \frac{\binom{15}{3}\binom{10}{4}}{\binom{25}{7}}.$$

定理：經過一些代數，結果是

$$\mathrm{E}[X] = n\left(\frac{a}{a+b}\right), \quad \text{and } \mathrm{Var}(X) = n\left(\frac{a}{a+b}\right)\left(1 - \frac{a}{a+b}\right)\left(\frac{a+b-n}{a+b-1}\right).$$

備註：這裡，$\frac{a}{a+b}$ 在二項分佈中扮演 p 的角色。然後相應的 $Y \sim \mathrm{Bin}(n, p)$ 結果將是

$$\mathrm{E}[Y] = n\left(\frac{a}{a+b}\right), \quad \text{and} \quad \mathrm{Var}(Y) = n\left(\frac{a}{a+b}\right)\left(1 - \frac{a}{a+b}\right)$$

所以二項式具有與超幾何相同的均值，但方差稍大。

4.1.3 幾何和負二項分佈

本小節討論與伯努利和二項式密切相關的兩個分佈。

4.1.3.1 \$.1.3.1 幾何分佈

定義：假設我們考慮一個獨立的 Bern (p) 試驗的無限序列。讓 Z 等於試驗次數，直到獲得第一次成功。事件 $Z=k$ 對應於 $k-1$ 次失敗，然後是成功。例如，FFFFS 產生 $Z=5$。因此，

$$\mathrm{P}(Z=k) = q^{k-1}p, \quad k = 1, 2, \ldots$$

我們說Z 具有參數p 的幾何分佈。

符號：$X \sim \mathrm{Geom}(p)$。

我們將通過 mgf 獲得幾何的均值和方差...

定理：$\mathrm{Geom}(p)$ 的 mgf 為

$$M_Z(t)=\frac{pe^t}{1-qe^t}, \quad \text{for } t<\ln(1/q).$$

證明：我們有

$$\begin{aligned} M_Z(t) &= \mathrm{E}\left[e^{tZ}\right] \\ &= \sum_{k=1}^{\infty} e^{tk} q^{k-1} p \\ &= pe^t \sum_{k=0}^{\infty} \left(qe^t\right)^k \\ &= \frac{pe^t}{1-qe^t}, \text{ for } qe^t<1. \end{aligned}$$

推論：$\mathrm{E}[Z]=1/p$。證明：對期望值使用通常的 mgf 定理，

$$\begin{aligned} \mathrm{E}[Z]= & \left.\frac{d}{dt} M_Z(t)\right|_{t=0} \\ = & \left.\frac{\left(1-qe^t\right)\left(pe^t\right)-\left(-qe^t\right)\left(pe^t\right)}{\left(1-qe^t\right)^2}\right|_{t=0} \\ = & \left.\frac{pe^t}{\left(1-qe^t\right)^2}\right|_{t=0} \\ = & \frac{p}{(1-q)^2}=1/p \end{aligned}$$

備註：我們也可以直接從期望值的定義中證明這一點，如 \$2.5.1

同樣，經過很多代數，

$$\mathrm{E}\left[Z^2\right]=\left.\frac{d^2}{dt^2} M_Z(t)\right|_{t=0}=\frac{2-p}{p^2},$$

以便

$$\mathrm{Var}(Z)=\mathrm{E}\left[Z^2\right]-(\mathrm{E}[Z])^2=\frac{2-p}{p^2}-\frac{1}{p^2}=\frac{q}{p^2}$$

示例：反复擲骰子。我們在第八次投擲中第一次觀察到 3 的概率是多少？

答案：我們需要的投擲次數是 $Z \sim \mathrm{Geom}(1/6)$。因此，$\mathrm{P}(Z=8)=$$(5/6)^7(1/6)$

我們期望投擲多少次？

答案：$\mathrm{E}[Z]=1/p=6$ 投擲。

這是幾何的一個非常重要的屬性。

Theorem (Memoryless Property of the Geometric)：假設$Z \sim \text{Geom}\ (p)$。那麼對於正整數 s, t，我們有

$$\mathrm{P}(Z > s + t \mid Z > s) = \mathrm{P}(Z > t)$$

為什麼它被稱為無記憶屬性？如果一個事件在時間 s 之前還沒有發生，那麼它在額外的 t 時間單位之後發生的概率與它最初在時間 t 之後發生的（無條件）概率相同 - 它忘了它已經過去了s！

證明：首先，對於任意$t = 0, 1, 2, \ldots$，尾概率為

$$\mathrm{P}(Z > t) = \mathrm{P}(t\,\text{Bern}(p) \text{ failures in a row }) = q^t.$$

這立即意味著

$$\mathrm{P}(Z > s + t \mid Z > s) = \frac{\mathrm{P}(Z > s + t \cap Z > s)}{\mathrm{P}(Z > s)} = \frac{\mathrm{P}(Z > s + t)}{\mathrm{P}(Z > s)} = \frac{q^{s+t}}{q^s} = q^t$$

換句話說，$\mathrm{P}(Z > s + t \mid Z > s) = \mathrm{P}(Z > t)$

示例：讓我們擲骰子，直到第一次出現 5。假設我們已經拋了四次都沒有成功。在我們觀察到 5 之前，我們需要兩次以上的額外投擲的概率是多少？

令 Z 為所需的投擲次數。通過無記憶屬性（使用 $s = 4$ 和 $t = 2$ ），我們想要

$$\mathrm{P}(Z > 6 \mid Z > 4) = \mathrm{P}(Z > 2) = (5/6)^2$$

有趣的事實：Geom (p) 是唯一具有無記憶屬性的離散分佈。

不那麼有趣的事實：有些書將 Geom(p) 定義為 Bern(p) 失敗的次數，直到您觀察到成功；即，失敗次數 = 試驗次數 - 1。您應該意識到這種不一致，但現在不要擔心。

4.1.3.2 §4.1.3,2 負二項分佈

定義：假設我們考慮一個無限的獨立 Bern(p) 試驗序列。令 W 表示直到 r^{th} 成功為止的試驗次數，使得 $W = r, r + 1, \ldots$ 事件 $W = k$ 對應到 $k - 1$ 時恰好成功 $r - 1$，然後在 k 時成功 r^{th} 。我們說 W 具有負二項分佈（又名帕斯卡分佈），參數為 r 和 p。

符號：$W \sim \text{Neg Bin}(r, p)$。

示例：FFFFFSS 對應於 $W = 7$ 次試驗，直到 $r = 2^{\text{nd}}$ 成功。

備註：與 Geom (p) 一樣，NegBin 的確切定義取決於您正在閱讀的書。

讓我們獲得 W 的 pmf。請注意 $W = k$ 如果您在時間 $k - 1$ 前獲得了 $r - 1$ 次成功（和 kr 次失敗），然後在 r^{th} 次成功千元。所以，

$$\mathrm{P}(W = k) = \left[\binom{k-1}{r-1} p^{r-1} q^{k-r}\right] p = \binom{k-1}{r-1} p^r q^{k-r}, \quad k = r, r + 1, \ldots$$

示例：擲骰子直到第三次出現 5。我們需要進行 7 次拋擲的概率是多少？

令 W 為所需的投擲次數。顯然，$W \sim \text{NegBin}(3, 1/6)$。然後

$$\mathrm{P}(W = 7) = \binom{7-1}{3-1} (1/6)^3 (5/6)^{7-3}$$

定理：如果 $Z_1, \ldots, Z_r \overset{\text{iid}}{\sim}$ Geom(p)，則 $W = \sum_{i=1}^{r} Z_i \sim \text{NegBin}(r, p)$。換句話說，Geom $(p)'$ s 加起來就是一個 NegBin。證明：我們不會在這裡做，但你可以使用 mgf 技術。（或者，你甚至可以直覺地爭論。）

如果您將 Z_i 視為 $(i-1)^{\text{st}}$ 成功後的試驗次數，則該定理是有意義的，直到並包括 i^{th} 成功。

（NegBin 的）性質：由於 Z_i 是獨立同分佈的，上述定理給出：

$$\begin{gathered}\mathrm{E}[W] = r\mathrm{E}\left[Z_i\right] = r/p \\ \operatorname{Var}(W) = r \operatorname{Var}(Z_i) = rq/p^2 \\ M_W(t) = \left[M_{Z_i}(t)\right]^r = \left(\frac{pe^t}{1-qe^t}\right)^r\end{gathered}$$

4.1.3.3 二項式和 NegBin 有什麼關係？

如果 $X_1, \ldots, X_n \overset{\text{iid}}{\sim}$ Bern(p)，那麼

$$Y \equiv \sum_{i=1}^{n} X_i \sim \operatorname{Bin}(n, p), \text{ with } \mathrm{E}[Y] = np \text{ and } \operatorname{Var}(Y) = npq$$

如果 $Z_1, \ldots, Z_r \overset{\text{iid}}{\sim}$ Geom(p)，那麼

$$W \equiv \sum_{i=1}^{r} Z_i \sim \operatorname{Neg}\operatorname{Bin}(r, p), \text{ with } \mathrm{E}[W] = r/p \text{ and } \operatorname{Var}(W) = rq/p^2$$

4.1.4 泊松過程和泊松分佈

我們將首先討論泊松過程，然後討論它們如何導致泊松分佈，我們已經順便遇到過。

4.1.5 泊松過程

設$N(t)$ 是一個計數過程。也就是說，$N(t)$ 是某個過程在時間間隔 $[0, t]$內發生（或到達或事件）的次數。$N(t)$ 看起來像一個階躍函數。

示例：$N(t)$ 可以是以下任何一種：

a. 進入購物中心的汽車（在時間 t 之前）。

b. 電線上的缺陷（長度為 t ）。

c. 餅乾麵團中的葡萄乾（體積為 t ）。

令$\lambda > 0$ 為每單位時間（或長度或體積）的平均出現次數。在上面的例子中，我們可能有：(a) $\lambda = 10/\text{min}$, (b) $\lambda = 0.5/\text{ft}$, (c) \$ / \$^{3}。在繼續之前，讓我們享受一些有用的符號。

符號：$o(h)$ (“little-oh”) 是一個通用函數，使得 $o(h)/h \to 0$ 為 $h \to 0$，即 $o(h)$ 去零比 h 變為零“快”。例如 $f(h) = h^2 = o(h)$，因為$f(h)/h = h \to 0$; 但是 $g(h) = 2h$ 不是 o(h)，因為在這種情況下 $g(h)/h = 2$ 不會變為 0 。

泊松過程是一個特定的計數過程。

定義： 對於 $a < b$，增量 $N(b) - N(a)$ 是一個隨機變量，表示 $(a, b]$ 中的到達次數。泊松過程 (PP) 是一個計數過程滿足以下增量條件：

i. 有足夠短的時間間隔 h，使得對於所有 t，

$$\begin{aligned} &\mathrm{P}(N(t+h)-N(t)=0)=1-\lambda h+o(h) \\ &\mathrm{P}(N(t+h)-N(t)=1)=\lambda h+o(h) \\ &\mathrm{P}(N(t+h)-N(t)\geq 2)=o(h) \end{aligned}$$

ii. 增量$N(t+h)-N(t)$的分佈只取決於長度h

iii. 如果$a<b<c<d$，則兩個增量$N(d)-N(c)$ 和$N(b)-N(a)$ 是獨立隨機變量。

泊松過程假設的英文翻譯：

i. 到達幾乎一次發生一個，然後以 $\lambda/$ 單位時間的速率發生。

ii. 到達模式是固定的——它不會隨著時間而改變；特別是，速率 λ 是恆定的。

iii. 兩個不相交的時間間隔內的到達次數是獨立的。

泊松過程示例：中微子撞擊探測器。這種情況非常罕見，以至於它們確實一次只發生一次——你永遠不會得到中微子群的到達，1 此外，中微子的到達率不會隨時間而變化，並且所有到達都是相互獨立的。

反例：顧客到達餐廳。他們傾向於成群出現，而不是一次一個。價格在一天中有所不同（例如，晚餐時間更多）。此外，到達可能不是獨立的。這不是泊松過程。

4.1.6 泊松分佈

定義：令X為單位時間間隔內泊松(λ)過程的出現次數。那麼X 具有參數λ 的泊松分佈。

符號：$X\sim\mathrm{Pois}(\lambda)$

[1] 中微子是孤獨的生物。定理/定義：$X\sim\mathrm{Pois}(\lambda)\Rightarrow\mathrm{P}(X=k)=e^{-\lambda}\lambda^k/k!,k=0,1,2,\ldots$

證明：遵循泊松過程假設並涉及一些簡單的微分方程。

首先，讓我們定義$P_x(t)\equiv\mathrm{P}(N(t)=x)$，即在時間$t$ 之前恰好x 到達的概率。我們注意到在時間 $t+h$ 之前沒有任何到達的概率可以寫成在時間 t 之前沒有任何到達的概率。

$$\begin{aligned} &P_0(t+h) \\ &\quad=\mathrm{P}(N(t+h)=0) \\ &\quad=\mathrm{P}(\text{ no arrivals by time } t \text{ and then no arrivals by time } t+h) \\ &\quad=\mathrm{P}(\{N(t)=0\}\cap\{N(t+h)-N(t)=0\}) \\ &\quad=\mathrm{P}(N(t)=0)\mathrm{P}(N(t+h)-N(t)=0) \quad \text{(by independent increments (iii))} \\ &\quad=\mathrm{P}(N(t)=0)\mathrm{P}(N(h)=0) \quad \text{(by stationary increments (ii))} \\ &\quad\doteq P_0(t)(1-\lambda h) \quad \text{(by definition and (i)).} \end{aligned}$$

因此，

$$\frac{P_0(t+h)-P_0(t)}{h}\doteq-\lambda P_0(t)$$

取極限為 $h\to 0$，我們有

$$P_0'(t)=-\lambda P_0(t)$$

同樣，對於 $x>0$，我們有

$$
\begin{aligned}
P_x(t+h) = &\quad \mathrm{P}(N(t+h)=x) \\
= \quad \mathrm{P}(N(t+h)=x \text{ and no arrivals during } [t,t+h]) & \\
& + \mathrm{P}(N(t+h)=x \text{ and } \geq 1 \text{ arrival during } [t,t+h]) \\
& (\text{ Law of Total Probability }) \\
\doteq \quad \mathrm{P}(\{N(t)=x\} \cap \{N(t+h)-N(t)=0\}) & \\
& + \mathrm{P}(\{N(t)=x-1\} \cap \{N(t+h)-N(t)=1\}) \\
& (\text{ by (i), only consider case of one arrival in } [t,t+h] \\
= \quad \mathrm{P}(N(t)=x)\mathrm{P}(N(t+h)-N(t)=0) & \\
& + \mathrm{P}(N(t)=x-1)\mathrm{P}(N(t+h)-N(t)=1) \\
& (\text{ by independent increments (iii)}) \\
& P_x(t)(1-\lambda h)+P_{x-1}(t)\lambda h.
\end{aligned}
$$

取極限為 $h \to 0$，我們得到

$$P_x'(t)=\lambda\left[P_{x-1}(t)-P_x(t)\right], \quad x=1,2,\ldots$$

微分方程 4.1 和 4.2 的解很容易證明為

$$P_x(t)=\frac{(\lambda t)^x e^{-\lambda t}}{x!}, \quad x=0,1,2,\ldots$$

（如果您不相信我，只需插入並親自查看！）

注意 $\operatorname{Pois}(\lambda)$ 案例的 $t=1$ 最終完成了證明。

備註：λ 的值可以簡單地通過改變時間單位來改變。（由您來跟踪您正在使用的單位。）

4.1.6.1 示例：

$$
\begin{aligned}
X = &\text{ number of phone calls in one minute } \sim \operatorname{Pois}(\lambda=3/\min) \\
Y = &\text{ number of phone calls in five minutes } \sim \operatorname{Pois}(\lambda=15/5\min) \\
Z = &\text{ number of phone calls in 10 seconds } \sim \operatorname{Pois}(\lambda=0.5/10\mathrm{sec}).
\end{aligned}
$$

定理：$X \sim \operatorname{Pois}(\lambda) \Rightarrow$ mgf 是 $M_X(t)=e^{\lambda\left(e^t-1\right)}$。

4.1.6.2 證明：

$$M_X(t)=\mathrm{E}\left[e^{tX}\right]=\sum_{k=0}^{\infty} e^{tk}\left(\frac{e^{-\lambda}\lambda^k}{k!}\right)=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\left(\lambda e^t\right)^k}{k!}=e^{-\lambda}e^{\lambda e^t}$$

定理：$X \sim \operatorname{Pois}(\lambda) \Rightarrow \mathrm{E}[X]=\operatorname{Var}(X)=\lambda$。

證明（使用 mgf）：

$$\mathrm{E}[X]=\left.\frac{d}{dt}M_X(t)\right|_{t=0}=\left.\frac{d}{dt}e^{\lambda\left(e^t-1\right)}\right|_{t=0}=\left.\lambda e^t M_X(t)\right|_{t=0}=\lambda$$

相似地，

$$\begin{aligned}
\mathrm{E}\left[X^2\right] &= \left.\frac{d^2}{dt^2} M_X(t)\right|_{t=0} \\
&= \left.\frac{d}{dt}\left(\frac{d}{dt} M_X(t)\right)\right|_{t=0} \\
&= \left.\lambda \frac{d}{dt}\left(e^t M_X(t)\right)\right|_{t=0} \\
&= \left.\lambda\left[e^t M_X(t)+e^t \frac{d}{dt} M_X(t)\right]\right|_{t=0} \\
&= \left.\lambda e^t\left[M_X(t)+\lambda e^t M_X(t)\right]\right|_{t=0} \\
&= \lambda(1+\lambda).
\end{aligned}$$

因此，$\operatorname{Var}(X)=\mathrm{E}\left[X^2\right]-(\mathrm{E}[X])^2=\lambda(1+\ lambda)-\lambda^2=\lambda$。完畢。

示例：呼叫中心以每分鐘 3 次的 Poisson 過程到達。

讓 $X=$ 一分鐘內的通話次數。

所以 $X \sim \operatorname{Pois}(3), \mathrm{E}[X]=\operatorname{Var}(X)=3, \mathrm{P}(X \leq 4)=\sum_{k=0}^4 e^{-3} 3^k / k\,!$.

讓 $Y=40$ 秒內的通話次數。所以 $Y \sim \operatorname{Pois}(2), \mathrm{E}[Y]=\operatorname{Var}(Y)=2, \mathrm{P}(Y \leq 4)=\sum_{k=0}^4 e^{-2} 2^k / k!$

定理（泊鬆的加性性質）：假設 $X_1, \ldots, X_n$ 與 $X_i \sim \operatorname{Pois}(\lambda_i)$獨立，$i=1$，$\ldots$，$n$。然後

$$Y \equiv \sum_{i=1}^n X_i \sim \operatorname{Pois}\left(\sum_{i=1}^n \lambda_i\right)$$

證明：由於 X_i 是獨立的，來自 \$3.6 的定理意味著

$$M_Y(t)=\prod_{i=1}^n M_{X_i}(t)=\prod_{i=1}^n e^{\lambda_i\left(e^t-1\right)}=e^{\left(\sum_{i=1}^n \lambda_i\right)\left(e^t-1\right)}$$

這是 Pois $\left(\sum_{i=1}^n \lambda_i\right)$ 分佈的 mgf。

示例：男性 [女性] 駕駛的汽車按照泊松過程以 $\lambda_1=3 / \mathrm{hr}\left[\lambda_2=5 / 的速率到達停車場 hr\right]$。所有到達都是獨立的。在接下來的 30 分鐘內恰好有兩次到達的概率是多少？

答：到達的總數是$\operatorname{Pois}(\lambda_1+\lambda_2=8 / \mathrm{hr})$，所以接下來30分鐘的總數是 $X \sim \operatorname{Pois}(4)$。所以 $\mathrm{P}(X=2)=e^{-4} 4^2 / 2!$。

4.2 連續分佈

現在，我們繼續介紹我們的一些老朋友……

4.2.1 均勻分佈

定義：如果隨機變量 X 具有 pdf 和 cdf，則它具有邊界 a 和 b 的均勻分佈：

$$f(x)=\begin{cases}\frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise}\end{cases} \quad \text{and} \quad F(x)=\begin{cases}0, & x<a \\ \frac{x}{b-a}, & a \leq x \leq b \\ 1, & x \geq b .\end{cases}$$

符號：$X \sim \operatorname{Unif}(a, b)$。

以前的工作表明

$$\mathrm{E}[X]=\frac{a+b}{2} \text{ and } \operatorname{Var}(X)=\frac{(a-b)^2}{12}.$$

我們還可以推導出 mgf，

$$M_X(t)=\mathrm{E}\left[e^{tX}\right]=\int_a^b e^{tx}\frac{1}{b-a}dx=\frac{e^{tb}-e^{ta}}{t(b-a)}$$

4.2.2 指數、Erlang 和 Gamma 分佈

指數分佈有一個健壯的家譜，我們將在這裡討論其中的一些分支。

4.2.2.1 &.2.2.1 指數分佈

定義：帶有速率參數 λ 的指數分佈有 pdf

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & \text{otherwise}\end{cases}$$

符號：$X\sim \operatorname{Exp}(\lambda)$

以前的工作表明 cdf $F(x)=1-e^{-\lambda x}, \mathrm{E}[X]=1/\lambda$, 和 $\operatorname{Var}(X)$ =$1/\lambda^2$. 我們還推導出了 mgf，

$$M_X(t)=\mathrm{E}\left[e^{tX}\right]=\int_0^\infty e^{tx}f(x)dx=\frac{\lambda}{\lambda-t},\quad t<\lambda$$

指數分佈有一些非常酷的特性。

定理（指數的無記憶性質）：假設 $X\sim \operatorname{Exp}(\lambda)$。那麼對於正的 s,t，我們有

$$\mathrm{P}(X>s+t \mid X>s)=\mathrm{P}(X>t).$$

與離散幾何分佈類似，X 將在額外的 t 時間單位中存活的概率是它至少存活 t 的（無條件）概率 - 即，它忘記了它已經過了時間 $新元。它總是“像新的一樣！”

證明：我們有

$$\begin{aligned}\mathrm{P}(X>s+t \mid X>s)&=\frac{\mathrm{P}(X>s+t\cap X>s)}{\mathrm{P}(X>s)}=\frac{\mathrm{P}(X>s+t)}{\mathrm{P}(X>s)}\\&=\frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}=e^{-\lambda t}=\mathrm{P}(X>t).\end{aligned}$$

示例：假設燈泡的壽命是指數級的，平均為 10 個月。如果光能存活 20 個月，那麼它還能再存活 10 個月的概率是多少？

$$\mathrm{P}(X>30\mid X>20)=\mathrm{P}(X>10)=e^{-\lambda x}=e^{-(1/10)(10)}=e^{-1}$$

示例：如果到下一輛公共汽車的時間呈指數分佈，平均為 10 分鐘，而您已經等了 20 分鐘，那麼您預計還要再等 10 分鐘。

備註：指數是唯一具有無記憶性的連續分佈

備註：注意 $\mathrm{E}[X]$ 和 $\operatorname{Var}(X)$ 對於幾何分佈和指數分佈有多相似。（不是巧合！）定義：如果X是一個連續隨機變量，pdf $f(x)$和cdf $F(x)$，那麼它的失效率函數是

$$S(t) \equiv \frac{f(t)}{\mathrm{P}(X>t)}=\frac{f(t)}{1-F(t)}$$

可以粗略地認為是 X' 的瞬時死亡率，因為它到目前為止已經存活到時間 t。

示例：如果 $X \sim \operatorname{Exp}(\lambda)$，則 $S(t)=\lambda e^{-\lambda t} / e^{-\lambda t}=\lambda$。所以如果 X 是一個燈泡的指數壽命，那麼它的瞬時燒毀率總是 $\lambda-$ 總是和新的一樣！這顯然是無記憶屬性的結果。

指數（連續）也與泊松（離散）有關！

定理：令X 為速率為λ 的泊松過程第一次到達之前的時間量。然後 $X \sim \operatorname{Exp}(\lambda)$。

證明：$F(x)$ 的 cdf 為

$$\begin{aligned}
F(x) & =\mathrm{P}(X \leq x) \\
& =1-\mathrm{P}(\text{ no arrivals in }[0, x]) \\
& =1-\frac{e^{-\lambda x}(\lambda x)^0}{0!} \quad(\text{ the number of arrivals in }[0, x] \text{ is Pois }(\lambda x)) \\
& =1-e^{-\lambda x} .
\end{aligned}$$

定理：令人驚訝的是，它可以證明（經過大量工作）泊松過程的到達間隔時間都是 iid $\operatorname{Exp}(\lambda)$! 當你參加高級隨機過程課程時，你自己看看。

示例：假設到達購物中心的人來自泊松過程，速率為 $\lambda=20 / \mathrm{hr}$。13^{th} 和 14^{th} 客戶之間的時間至少為四分鐘的概率是多少？

答案：假設客戶 13 和 14 之間的時間為 X。因為我們有一個泊松過程，所以間隔是 iid $\operatorname{Exp}(\lambda=20 / \mathrm{hr})$，所以

$$\mathrm{P}(X>4 \min)=\mathrm{P}(X>1 / 15 \mathrm{hr})=e^{-\lambda t}=e^{-20 / 15} .$$

4.2.2.2 &4.2.2,2 Erlang 和 Gamma 分佈

有幾種分佈可以概括指數分佈。

定義：假設 $X_1, \ldots, X_k \stackrel{\text{iid}}{\sim} \operatorname{Exp}(\lambda)$，令 $Y=\sum_{i=1}^k X_i$。那麼 Y 具有帶有參數 k 和 λ 的 Erlang 分佈，記為 $Y \sim \operatorname{Erlang}(\lambda)$。

Erlang 只是 iid 指數的總和。例如，如果您購買了一包燈泡，那麼 Erlang 可能是對所有燈泡的整個生命週期進行建模的好方法。當然，很明顯 $\operatorname{Erlang}_1(\lambda) \sim \operatorname{Exp}(\lambda)$。定理：Erlang 的 pdf 和 cdf 是

$$\begin{gathered}
f(y)=\frac{\lambda^k e^{-\lambda y} y^{k-1}}{(k-1)!}, \quad y \geq 0, \quad \text{ and } \\
F(y)=1-\sum_{i=0}^{k-1} \frac{e^{-\lambda y}(\lambda y)^i}{i!}, \quad y \geq 0
\end{gathered}$$

請注意，cdf 是一堆泊松概率的總和。（雖然我們不會在這裡做，但這個觀察有助於 cdf 的推導。）

Erlang 的期望值、方差和 mgf：

$$
\begin{aligned}
\mathrm{E}[Y] &= \mathrm{E}\left[\sum_{i=1}^{k} X_i\right] = \sum_{i=1}^{k} \mathrm{E}\left[X_i\right] = k/\lambda \\
\mathrm{Var}(Y) &= k/\lambda^2 \\
M_Y(t) &= \left(\frac{\lambda}{\lambda - t}\right)^k, \quad t < \lambda
\end{aligned}
$$

示例：假設 X_1 和 X_2 是 iid Exp(3)。求 $\mathrm{P}\left(X_1 + X_2 < 1\right)$。由於 $X_1 + X_2 \sim$ Erlang $\mathrm{Eg}_{k=2}(\lambda = 3)$，我們有

$$
\mathrm{P}\left(X_1 + X_2 < 1\right) = 1 - \sum_{i=0}^{k-1} \frac{e^{-\lambda y}(\lambda y)^i}{i!} = 1 - \sum_{i=0}^{2-1} \frac{e^{-(3\cdot 1)}(3\cdot 1)^i}{i!} = 0.801
$$

我們的泛化節還沒有結束……

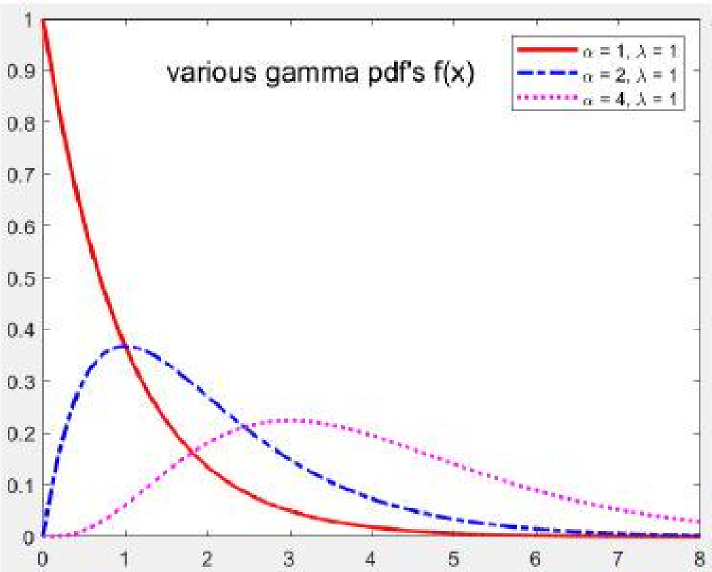
定義：X 具有參數 $\alpha > 0$ 和 $\lambda > 0$ 的 gamma 分佈，如果它有 pdf

$$
f(x) = \frac{\lambda^\alpha e^{-\lambda x} x^{\alpha - 1}}{\Gamma(\alpha)}, \quad x \ge 0
$$

在哪裡

$$
\Gamma(\alpha) \equiv \int_0^\infty s^{\alpha - 1} e^{-s} ds
$$

是伽馬函數。我們寫成$X \sim \mathrm{Gam}(\alpha, \lambda)$。



備註：伽瑪分佈概括了 Erlang 分佈（其中 α 必須是正整數）。它與 Erlang 具有相同的期望值、方差和 mgf，用 α 代替 k。（見練習 4.613）

備註：很容易證明$\Gamma(1) = 1$。此外，如果$\alpha > 0$，則$\Gamma(\alpha + 1) =\alpha\Gamma(\alpha)$。特別是，如果 α 是一個正整數，那麼 $\Gamma(\alpha + 1) = \alpha!$。這裡有一個讓你在聚會上受歡迎的花絮：$\Gamma(1/2) = \sqrt{\pi}$。（見練習 4.614）

4.2.3 其他連續分佈

我們將簡要提及其他一些值得了解的連續分佈，儘管我們會將正態分佈保存到它自己的特殊部分。

三角 (a, b, c) 分佈 - 適用於基於有限數據（最小值、模式、最大值）對 RV 進行建模。pdf和時刻是：

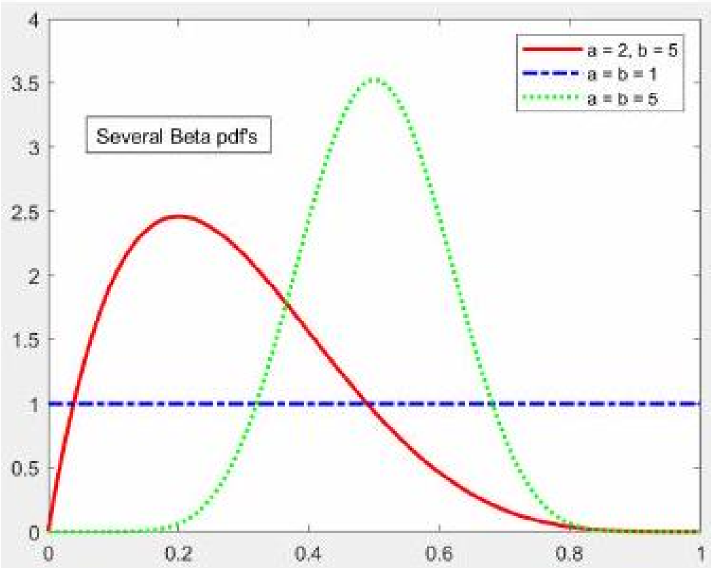
Beta(a, b) 分佈 - 適用於對受限於區間的 RV 進行建模。pdf和時刻是：

$$f(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}, 0<x<1,$$

$$\mathrm{E}[X]=\frac{a}{a+b} \quad \text{and } \operatorname{Var}(X)=\frac{ab}{(a+b)^2(a+b+1)} \quad \text{(see Exercise 4.615.}$$

此分佈的名稱來自 beta 函數，該函數定義為

$$\beta(a, b) \equiv \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}=\int_0^1 x^{a-1}(1-x)^{b-1} dx$$



$$f(x)=\begin{cases}\frac{2(x-a)}{(b-a)(c-a)}, & a<x \leq b \\ \frac{2(c-x)}{(c-b)(c-a)}, & b<x<c \\ 0, & \text{otherwise}\end{cases}$$

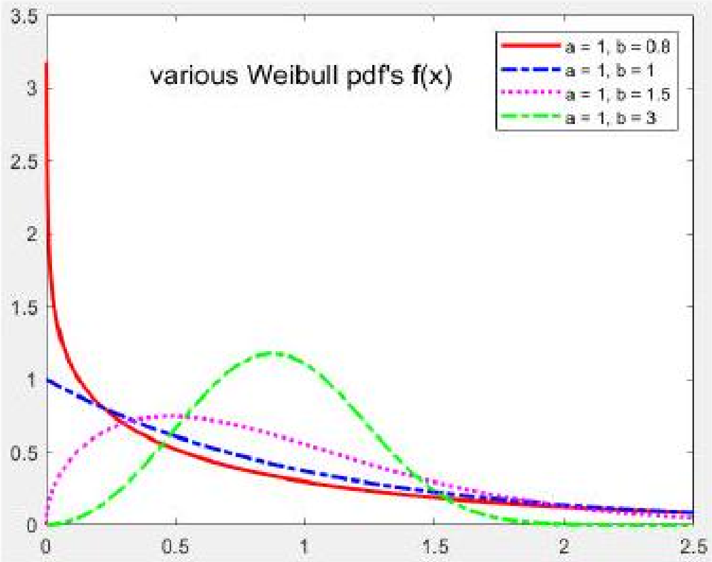
$$\mathrm{E}[X]=\frac{a+b+c}{3} \text{ and } \operatorname{Var}(X)= \text{ mess.}$$

備註：某些版本的統一 $(a=b=1)$ 和三角分佈是 beta 的特例。

Weibull (a, b) 分佈 - 適用於建模可靠性模型。常數 a 稱為"比例"參數，而 b 是"形狀"參數。pdf、cdf、期望值和方差是（見練習 4.616

$$f(x)=ab(ax)^{b-1}e^{-(ax)^b} \quad \text{and} \quad F(x)=1-\exp\left[-(ax)^b\right], x>0$$

$$\mathrm{E}[X]=\frac{1}{a}\Gamma\left(1+\frac{1}{b}\right) \quad \text{and} \quad \operatorname{Var}(X)=\frac{1}{a^2}\left[\Gamma\left(1+\frac{2}{b}\right)-\Gamma^2\left(1+\frac{1}{b}\right)\right]$$



備註：指數也是 Weibull 的一個特例。

示例：變送器的故障時間 T 具有 Weibull 分佈，其參數為 $a = 1/(200 \text{ hours })$ 和 $b = 1/3$。然後

$$\mathrm{E}[T] = 200\Gamma(1+3) = 1200 \text{ hours.}$$

它在 2000 小時之前失敗的概率是

$$F(2000) = 1 - \exp\left[-(2000/200)^{1/3}\right] = 0.884$$

柯西分佈 - 一個“肥尾”分佈，有利於反駁事物！

$$f(x) = \frac{1}{\pi\left(1+x^2\right)} \quad \text{and} \quad F(x) = \frac{1}{2} + \frac{\arctan(x)}{\pi}, \quad x \in \mathbb{R}$$

定理：柯西分佈具有未定義的均值和無限方差！

奇怪的事實：$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Cauchy} \Rightarrow \sum_{i=1}^n X_i/n \sim$ 柯西。即使你取一堆柯西的平均值，你也回到了你開始的地方！

4.2.3.1 其他分佈的字母湯

- χ^2 分佈 - 當我們談論統計時出現。- t 分配 - 即將推出。
- F 分佈 - 即將到來。
- 以老統計學家命名的分佈：Pareto、LaPlace、Rayleigh、Gumbel、Johnson。
- ETC…

4.3 正態分佈和中心極限定理

正態分佈非常重要，我們給它一個完整的部分！

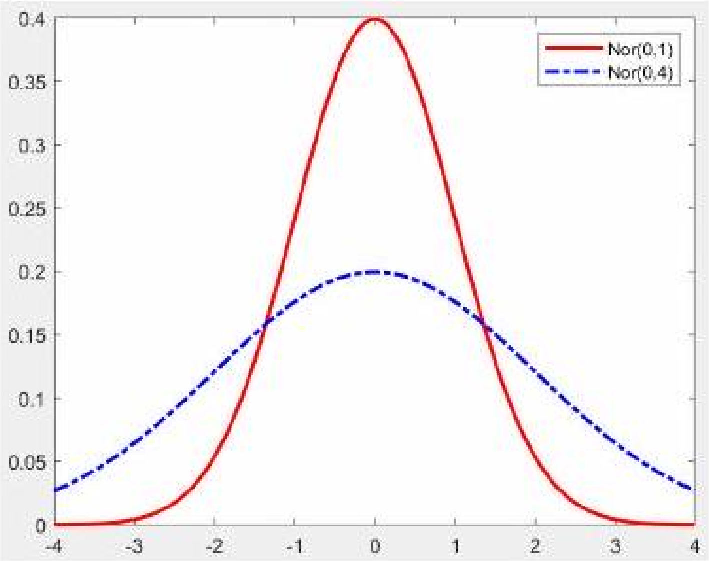
4.3.1 基礎知識

定義：隨機變量 X 具有參數 μ 和 σ^2 的正態分佈，如果它有 pdf

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right], \quad \forall x \in \mathbb{R}$$

$f(x)$ 是“鐘形”並且圍繞 $x=\mu$ 對稱，當你遠離 μ 時，尾巴會迅速脫落。

具有較小 σ^2 的法線對應於“又高又瘦”的鐘形曲線；大的 σ^2 與“短而肥”的鐘形曲線相關聯。



注：正態分佈也稱為高斯分佈。

示例：身高、體重、SAT 分數、作物產量和事物的平均值往往是正常的。

符號：$X \sim \mathrm{Nor}\left(\mu, \sigma^2\right)$。趣事 (1)：$\int_{\mathbb{R}} f(x)dx = 1$

證明：令 $z = (x-\mu)/\sigma$ 並註意

$$\int_{\mathbb{R}} f(x)dx = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] dx = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

現在的結果來自 § 1.2.2 6 中的極坐標示例。

趣事 (2)：cdf 是

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y-\mu)^2}{2\sigma^2}\right] dy = ??$$

備註：這個量沒有封閉形式的表達式，儘管通過數值方法獲得的非常精確的近似值被廣泛使用。敬請關注。

趣事 (3)：如果 X 是任何正常的 RV，那麼

$$\begin{aligned} \mathrm{P}(\mu-\sigma < X < \mu+\sigma) &\doteq 0.6827 \\ \mathrm{P}(\mu-2\sigma < X < \mu+2\sigma) &\doteq 0.9545 \\ \mathrm{P}(\mu-3\sigma < X < \mu+3\sigma) &\doteq 0.9973 \end{aligned}$$

所以幾乎所有的概率都包含在平均值的三個標準差內。（這就是豐田吹噓“六西格瑪”質量時所指的東西。）

趣事 (4)：mgf 為 $M_X(t) = \mathrm{E}\left[e^{tX}\right] = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$。

證明：通過 LOTUS，我們有

$$\begin{aligned}M_X(t) &= \int_{\mathbb{R}} e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] \\ &= \int_{\mathbb{R}} e^{t(\mu+\sigma z)} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2} \sigma dz \quad (\text{letting } z=(x-\mu)/\sigma) \\ &= \exp\left[\mu t+\frac{\sigma^2 t^2}{2}\right] \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(z-\sigma t)^2}{2}\right] dz\end{aligned}$$

我們完成了，因為被積函數只是 $\mathrm{Nor}(\sigma t, 1)$ 分佈的 pdf，因此積分為 1.

趣聞 (5) 和 (6)：$\mathrm{E}[X]=\mu$ 和 $\operatorname{Var}(X)=\sigma^2$。

證明：根據有趣的事實（4），鍊式法則，以及 $M_X(0)=1$ 的明顯結果，我們有

$$\mathrm{E}[X]= \left.\frac{d}{dt} M_X(t)\right|_{t=0}= \left.\left(\mu+\sigma^2 t\right) M_X(t)\right|_{t=0}=\mu M_X(0)=\mu .$$

類似地，但增加了乘積規則的使用，我們有

$$\begin{aligned}\mathrm{E}\left[X^2\right] &= \left.\frac{d^2}{dt^2} M_X(t)\right|_{t=0} \\ &= \left.\frac{d}{dt}\left[\left(\mu+\sigma^2 t\right) M_X(t)\right]\right|_{t=0} \\ &= \left.\left[\sigma^2 M_X(t)+\left(\mu+\sigma^2 t\right) M_X'(t)\right]\right|_{t=0} \\ &= \left.\left[\sigma^2 M_X(t)+\left(\mu+\sigma^2 t\right)^2 M_X(t)\right]\right|_{t=0} \\ &= \sigma^2+\mu^2\end{aligned}$$

因此，$\operatorname{Var}(X)=\mathrm{E}\left[X^2\right]-(\mathrm{E}[X])^2=\left(\sigma^2+\mu^2\right)-\mu^2=\sigma^2$。

以下非常一般的結果表明我們可以將獨立的法線相加並仍然得到法線。

定理（法線的加性性質）：如果 $X_1, \ldots, X_n$ 獨立於 $X_i \sim \mathrm{Nor}\left(\mu_i, \sigma_i^2\right), i=1, \ldots, n,$ 那麼

$$Y \equiv \sum_{i=1}^n a_i X_i+b \sim \mathrm{Nor}\left(\sum_{i=1}^n a_i \mu_i+b, \sum_{i=1}^n a_i^2 \sigma_i^2\right).$$

所以獨立法線的線性組合本身就是正常的。

證明：由於Y是一個線性函數，

我們是由 mgf 唯一性完成的。

備註：正態分佈的特徵完全在於其均值和方差。

通過以上，我們知道獨立法線的線性組合仍然是正常的。因此，當我們將獨立的法線相加時，我們所要做的就是找出均值和方差——總和的正態性是免費的。

例子：假設$X \sim \mathrm{Nor}(3,4)$，$Y \sim \mathrm{Nor}(4,6)$，並且X和Y是獨立的。求 $2X-3Y$ 的分佈。

解決方法：這是正常的

$$\mathrm{E}[2X-3Y]=2\mathrm{E}[X]-3\mathrm{E}[Y]=2(3)-3(4)=-6$$

$$\begin{aligned}
M_Y(t) &= M\sum_i a_iX_i + b(t) = e^{tb}M\sum_i a_iX_i(t) \\
&= e^{tb}\prod_{i=1}^{n} M_{a_iX_i}(t) \quad (X_i \text{ 's independent}) \\
&= e^{tb}\prod_{i=1}^{n} M_{X_i}(a_it) \ \text{(mgf of linear function)} \\
&= e^{tb}\prod_{i=1}^{n} \exp\left[\mu_i(a_it) + \frac{1}{2}\sigma_i^2(a_it)^2\right] \quad \text{(normal mgf)} \\
&= \exp\left[\left(\sum_{i=1}^{n}\mu_ia_i + b\right)t + \frac{1}{2}\left(\sum_{i=1}^{n}a_i^2\sigma_i^2\right)t^2\right]
\end{aligned}$$

和

$$\operatorname{Var}(2X-3Y) = 4\operatorname{Var}(X) + 9\operatorname{Var}(Y) = 4(4) + 9(6) = 70$$

因此，$2X-3Y \sim \operatorname{Nor}(-6,70)$。

推論（加性性質）：

$$X \sim \operatorname{Nor}(\mu,\sigma^2) \Rightarrow aX+b \sim \operatorname{Nor}(a\mu+b, a^2\sigma^2)$$

證明：立即從加性屬性，注意到 $\mathrm{E}[aX+b] = a\mu+b$ 和 $\operatorname{Var}(aX+b) = a^2\sigma^2$

這是前面推論的一個極其重要的推論。

4.3.1.1 推論（標準化法線）：

$$X \sim \operatorname{Nor}(\mu,\sigma^2) \Rightarrow Z \equiv \frac{X-\mu}{\sigma} \sim \operatorname{Nor}(0,1).$$

證明：使用上面的推論 $a = 1/\sigma$ 和 $b = -\mu/\sigma$

$\operatorname{Nor}(0,1)$ 分佈是特殊的，我們現在將看到。.

4.3.2 標準正態分佈

定義：$\operatorname{Nor}(0,1)$ 分佈稱為標準正態分佈，通常用 Z 表示。Nor $(0,1)$ 的 pdf 和 cdf 是

$$\phi(z) \equiv \frac{1}{\sqrt{2\pi}}e^{-z^2/2} \quad \text{and} \quad \Phi(z) \equiv \int_{-\infty}^{z}\phi(y)dy, \quad z \in \mathbb{R}$$

分別在 pdf / cdf 使用自己的希臘字母 ϕ/Φ 代替無聊的 ol’ f/F ！

備註：$\operatorname{Nor}(0,1)$ 很好，因為它的 cdf 有可用的表。例如，參見附錄中的表 B.1。您可以通過應用變換 $Z = (X-\mu)/\sigma$ 將任何正態隨機變量 X 標準化為標準正態。然後您可以使用 cdf 表。

更多備註：以下結果很容易通過對稱參數推導出來，應該記在內存中。

- $\mathrm{P}(Z \le a) = \Phi(a)$
- $\mathrm{P}(Z \ge b) = 1 - \Phi(b)$
- $\mathrm{P}(a \le Z \le b) = \Phi(b) - \Phi(a)$

- $\Phi(0) = \mathrm{P}(Z \le 0) = 1/2$
- $\Phi(-b) = \mathrm{P}(Z \le -b) = \mathrm{P}(Z \ge b) = 1 - \Phi(b)$ - $\mathrm{P}(-b \le Z \le b) = \Phi(b) - \Phi(-b) = 2\Phi(b) - 1$
- 更一般地說，

$$\mathrm{P}(\mu - k\sigma \le X \le \mu + k\sigma) = \mathrm{P}(-k \le Z \le k) = 2\Phi(k) - 1$$

因此，任何正常 RV 在其均值的 k 標準差內的概率不取決於均值或方差。

著名的 Nor $(0,1)$ 表值。很多人最終會記住 [2] 下表中的值。(i) 此外，表 B.1 焦急地等待著你的閱讀。或者您可以使用方便的軟件調用，例如 Excel 中的 normdist（計算任何正態分佈的 cdf）。

通過前面的"趣聞"和隨後的討論，對於各種 k 值，任何正常 RV 在其平均值的 k 標準差內的概率在下表中給出。

著名的逆 $\mathrm{Nor}(0,1)$ 表值。Z 的 p^{th} 分位數是 z 的值，使得 $\Phi(z) = \mathrm{P}(Z \le z) = p$; 它是用 $z_{1-p} \equiv \Phi^{-1}(p)$ 表示。

其中一些值 $\Phi^{-1}(p)$ 很容易記住。但總的來說，您始終可以使用表 B.1 或 Excel 的 norminv 函數等軟件，該函數實際上計算任何正態分佈的逆，而不僅僅是標準正態分佈。

[2] 也許"很多人"需要多出去走走。

示例：假設 $X \sim \mathrm{Nor}(21, 4)$。求 $\mathrm{P}(19 < X < 22.5)$。

答：標準化，我們得到

$$\begin{aligned}
\mathrm{P}(19 < X < 22.5) &= \mathrm{P}\left(\frac{19-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{22.5-\mu}{\sigma}\right) \\
&= \mathrm{P}\left(\frac{19-21}{2} < Z < \frac{22.5-21}{2}\right) \\
&= \mathrm{P}(-1 < Z < 0.75) \\
&= \Phi(0.75) - \Phi(-1) \\
&= \Phi(0.75) - [1 - \Phi(1)] \\
&= 0.7734 - [1 - 0.8413] \quad \text{(from Table B.1} \\
&= 0.6147. \square
\end{aligned}$$

示例：假設男性的身高是 $M \sim \mathrm{Nor}(68, 4)$，女性的身高是 $W \sim \mathrm{Nor}(65, 1)$。找出隨機女性比隨機男性高的概率。

答：請注意

$$\begin{aligned}
W - M &\sim \mathrm{Nor}(\mathrm{E}[W-M], \mathrm{Var}(W-M)) \\
&\sim \mathrm{Nor}(65-68, 1+4) \sim \mathrm{Nor}(-3, 5)
\end{aligned}$$

然後

$$\begin{aligned}
\mathrm{P}(W > M) &= \mathrm{P}(W - M > 0) \\
&= \mathrm{P}\left(Z > \frac{0+3}{\sqrt{5}}\right) \\
&= 1 - \Phi(3/\sqrt{5}) \\
&\doteq 1 - 0.910 = 0.090.
\end{aligned}$$

4.3.3 正態觀測的樣本均值

回想一下隨機變量 $X_1, \ldots, X_n$ 的樣本均值是 $\bar{X} \equiv \sum_{i=1}^{n} X_i / n$。

推論（舊定理）：$X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \operatorname{Nor}(\mu, \sigma^2) \Rightarrow \bar{X} \sim \operatorname{Nor}(\mu, \sigma^2/n)$。

證明：根據前人的工作，只要 $X_1, \ldots, X_n$ 是獨立同分佈的東西，我們有 $\mathrm{E}[\bar{X}] = \mu$ 和 $ ({X})=^{2} / n$。由於 $\bar{X}$ 是獨立法線的線性組合，因此也是正常的。完畢。備註：這個結果非常顯著！隨著觀察次數的增加，$\operatorname{Var}(\bar{X})$ 變小，而 $\mathrm{E}[\bar{X}] = \mathrm{E}[X_i] = \mu$ 保持不變。事實上，在下一個例子和定理中，我們將看到推論的一些結果。

首先，這是一個不錯的小應用程序，我們在其中確定了確保 $\bar{X}$ 很有可能接近 μ 所需的樣本量。

定理：假設 $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \operatorname{Nor}(\mu, \sigma^2)$。然後保證要求的樣本量$n$

$$\mathrm{P}(|\bar{X} - \mu| \le c) \ge \gamma$$

是

$$n \ge \frac{\sigma^2}{c^2}\left[\Phi^{-1}\left(\frac{1+\gamma}{2}\right)\right]^2$$

證明：注意$\bar{X} \sim \operatorname{Nor}(\mu, \sigma^2/n)$。然後

$$\begin{aligned}
\mathrm{P}(|\bar{X} - \mu| \le c) &= \mathrm{P}(-c \le \bar{X} - \mu \le c) \\
&= \mathrm{P}\left(\frac{-c}{\sigma/\sqrt{n}} \le \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le \frac{c}{\sigma/\sqrt{n}}\right) \\
&= \mathrm{P}\left(\frac{-c\sqrt{n}}{\sigma} \le Z \le \frac{c\sqrt{n}}{\sigma}\right) \\
&= 2\Phi(c\sqrt{n}/\sigma) - 1
\end{aligned}$$

現在我們必須找到 n 使得這個概率至少是 γ。為此，

$$\begin{aligned}
2\Phi(c\sqrt{n}/\sigma) - 1 \ge \gamma &\Leftrightarrow \Phi(c\sqrt{n}/\sigma) \ge \frac{1+\gamma}{2} \\
&\Leftrightarrow \frac{c\sqrt{n}}{\sigma} \ge \Phi^{-1}\left(\frac{1+\gamma}{2}\right),
\end{aligned}$$

平方後我們就完成了。

示例：假設 $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \operatorname{Nor}(\mu, 16)$。找到樣本大小 n 使得

$$\mathrm{P}(|\bar{X} - \mu| \le 1) \ge 0.95$$

解：根據上述結果，我們有

$$n \ge \frac{4^2}{1^2}\left[\Phi^{-1}\left(\frac{1+0.95}{2}\right)\right]^2 = 16\left[\Phi^{-1}(0.975)\right]^2 = 16(1.96)^2 = 61.47$$

因此，如果您取 62 次觀察的平均值，則 $\bar{X}$ 有 95% 的機會在真實（但未知）均值 μ 的 1 以內。

因此，該示例表明，隨著正態觀測值 n 的數量增加，樣本均值 $\bar{X}$ 趨向於在 μ 附近徘徊。

在這一點上，我們陳述（並同時證明）（弱）大數定律（LLN），它得出相同的結論，而不必假設基礎觀察是正常的！大數定律：假設 X_1、X_2、...、X_n 是具有有限均值 μ 和有限方差 σ^2 的任何事物。然後，

對於任何固定的 $c > 0$，樣本均值 $\bar{X}$ 在以下意義上收斂到 μ。

$$\mathrm{P}(|\bar{X}-\mu|>c) \leq \frac{\operatorname{Var}(\bar{X})}{c^2} = \frac{\sigma^2}{nc^2} \to 0, \quad \text{as } n \to \infty$$

證明：由於切比雪夫不等式，已經在結果聲明中完成了！

LLN 表明，隨著樣本大小 $n \to \infty$ ，樣本均值偏離真實均值超過任何固定量的概率變為零！或者用簡單的英語，

樣本均值 $\bar{X}$ 收斂到真實均值 μ 為 $n \to \infty$，即 $\bar{X} \to \mu.$

在課程即將到來的統計部分（例如，參見第 5 章，我們將了解到這使得樣本均值 $\bar{X}$ 成為均值 $\mathrm{E}[X_i] = \mu$，這在實踐中通常是未知的。你能想像強 LLN 能做什麼嗎？

4.3.4 中心極限定理

在我們開始討論好東西之前，讓我們先從一個有用的定義開始，這個定義是關於一系列 cdf 收斂到一個限制 cdf。

定義：隨機變量序列 $Y_1, Y_2, \ldots$ 與各自的 cdf 的 $F_{Y_1}(y), F_{Y_2}(y), \ldots$如果 $\lim_{n\to\infty} F_{Y_n}(y) = F_Y(y)$ 對屬於Y 的連續集的所有y，即$F_Y(y)$ 是連續的y 值的集合。

符號：$Y_n \xrightarrow{d} Y$。

思路：如果 $Y_n \xrightarrow{d} Y$ 和 n 很大，那麼你應該可以通過 $Y 的極限分佈來近似 Y_n 的分佈$。

示例：假設 $Y_n \sim \operatorname{Exp}\left(\lambda + \frac{1}{n}\right)$，對於 $n = 1, 2$， $\ldots$ 和 $Y \sim \operatorname{Exp}(\lambda)$。那麼 Y_n 的 cdf 為 $F_{Y_n}(y) = 1 - \exp\left[-\left(\lambda + \frac{1}{n}\right) y\right], y > 0, \text{and } Y'$ s 是 $F_Y(y) = 1 - e^{-\lambda y}, y > 0.$ 很明顯 $ F_{Y_{n}}(y) F_{Y}(y)$ 對於所有 $y > 0$，因此 $Y_n \xrightarrow{d} Y.$

我們現在為您提供概率和統計中最重要的定理，它表明隨著樣本量的增加，任何獨立同分佈觀察的樣本均值和總和趨於正態分佈。中心極限定理 (CLT)：假設 X_1、 $\ldots$ 、X_n 是具有 $\mathrm{E}[X_i] = \mu$ 和
的 任何事物$Var(X_i) = \sigma^2$。那麼作為$n \to \infty$，

$$Z_n \equiv \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - \mathrm{E}[\bar{X}]}{\sqrt{\operatorname{Var}(\bar{X})}} \xrightarrow{d} \operatorname{Nor}(0,1).$$

備註：對於這樣一個重要的定理，我們有很多話要說。

- 主要結論：如果 n 很大，那麼

$$\bar{X} \approx \operatorname{Nor}(\mathrm{E}[\bar{X}], \operatorname{Var}(\bar{X})) \sim \operatorname{Nor}(\mu, \sigma^2/n) \text{ and } \sum_{i=1}^{n} X_i \approx \operatorname{Nor}(n\mu, n\sigma^2).$$

- X_i 不必是正常的 CLT 才能工作！它甚至適用於離散分佈！
- 您通常需要 $n \geq 30$ 觀察值才能使近似值正常工作。（如果 X_i 來自對稱分佈，則需要更少的觀察。）
- 如果觀察是獨立同分佈的，你幾乎總是可以使用 CLT - 你只需要有限的 μ 和 σ^2。

- 事實上，CLT 實際上比這裡介紹的定理更通用！在某些情況下（此處未討論），它適用於適度相關和/或不來自同一分佈的 RV！

4.3.4.1 CLT 榮譽證明

假設 X_i 的 mgf $M_X(t)$ 存在並且滿足您不需要了解的某些技術條件。此外，在不失一般性的情況下（因為無論如何我們都在標準化）並且為了符號方便，我們將假設 $\mu=0$ 和 $\sigma^2=1$。如果我們能夠證明 Z_n 的 mgf 收斂到 $Z\sim\mathrm{Nor}(0,1)$ 的 mgf，我們就完成了. 也就是說，我們需要證明 $M_{Z_n}(t)\to e^{t^2/2}$ 為 $n\to\infty$。

為了讓事情順利進行，Z_n 的 mgf 是

$$\begin{aligned} M_{Z_n}(t) &= M_{\sum_{i=1}^n X_i/\sqrt{n}}(t) \\ &= M_{\sum_{i=1}^n X_i}(t/\sqrt{n}) \quad (\text{ mgf of a linear function of a RV}) \\ &= \left[M_X(t/\sqrt{n})\right]^n \quad (X_i \text{ 's are iid}) \end{aligned}$$

因此，記錄日誌，我們的目標是表明

$$\lim_{n\to\infty}\ln(M_{Z_n}(t)) = \lim_{n\to\infty} n\ell n\left(M_X(t/\sqrt{n})\right) = t^2/2$$

如果我們讓 $y=1/\sqrt{n}$，我們修改後的目標是證明

$$\lim_{y\to 0}\frac{\ln(M_X(ty))}{y^2} = t^2/2.$$

在繼續之前，請注意

$$\lim_{y\to 0}\ln(M_X(ty)) = \ell\mathrm{n}(M_X(0)) = \ln(1) = 0$$

和

$$\lim_{y\to 0} M_X'(ty) = M_X'(0) = \mathrm{E}[X] = \mu = 0$$

最後一個等式來自我們的標準化假設。

所以在所有這些積累之後，我們有

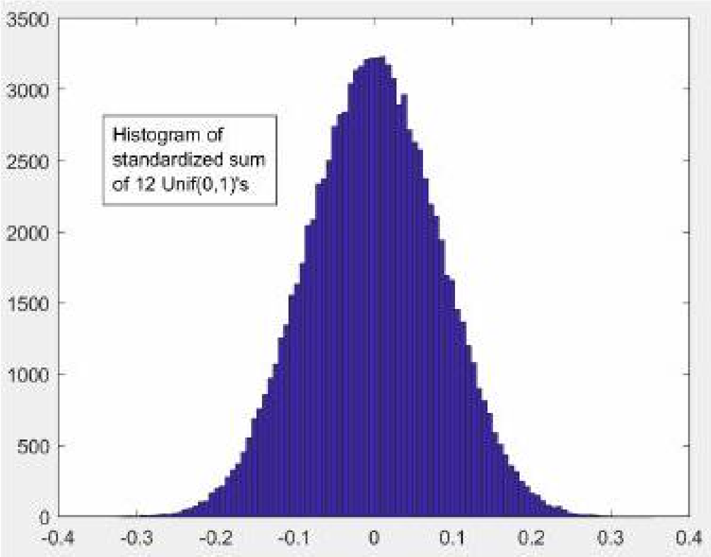
4.3.5 CLT 示例

為了證明 CLT 確實有效，我們將給出一些說明性示例。

示例：讓我們將一些 iid Unif $(0,1)$ 的、U_1、 ... 、U_n 相加。令 $S_n=\sum_{i=1}^n U_i$。注意 $\mathrm{E}\left[S_n\right]=n\mathrm{E}\left[U_i\right]=n/2$，並且 $\mathrm{Var}(S_n)=n\,\mathrm{Var}(U_i)=n/12$。因此，CLT 說，對於大的 n，

$$Z_n \equiv \frac{S_n-n/2}{\sqrt{n/12}} \approx \mathrm{Nor}(0,1)$$

下圖是使用 $Z_{12}=S_{12}-6$ for $n=12$ 的 100,000 次模擬實現編譯的直方圖。即使 $n=12$ 這個看似很小的值，生成的直方圖看起來也非常正常 - 幾乎可以肯定是由於 S_n 的組成部分的底層 Unif$(0,1)$ pdf 的對稱性。



$$
\begin{aligned}
&\lim_{y\to 0}\frac{\ln(M_X(ty))}{y^2}\\
&=\lim_{y\to 0}\frac{tM_X'(ty)}{2yM_X(ty)} \quad (\text{ by 4.3 et L'Hôpital to deal with } 0/0)\\
&=\lim_{y\to 0}\frac{t^2M_X''(ty)}{2M_X(ty)+2ytM_X'(ty)} \quad (\text{by 4.4 et L'Hôpital encore})\\
&=\frac{t^2M_X''(0)}{2M_X(0)+0}=\frac{t^2\mathrm{E}\left[X^2\right]}{2}=\frac{t^2}{2} \quad \left(\mathrm{E}\left[X^2\right]=\sigma^2+\mu^2=1\right).
\end{aligned}
$$

示例：假設 $X_1,\ldots,X_{100}\overset{\text{iid}}{\sim}\operatorname{Exp}(1/1000)$。求 $\mathrm{P}(950\le\bar{X}\le 1050)$。

解：回想一下，如果 $X_i\sim\operatorname{Exp}(\lambda)$，那麼 $\mathrm{E}\left[X_i\right]=1/\lambda$ 和 $\operatorname{Var}(X_i)=1/\lambda^2$。此外，由於 $\bar{X}$ 是基於 n 觀察的樣本均值，那麼

$$
\begin{gathered}
\mathrm{E}[\bar{X}]=\mathrm{E}\left[X_i\right]=1/\lambda \text{ and}\\
\operatorname{Var}(\bar{X})=\operatorname{Var}(X_i)/n=1/\left(n\lambda^2\right).
\end{gathered}
$$

對於我們的問題，$\lambda=1/1000$ 和 $n=100$，所以 $\mathrm{E}[\bar{X}]=1000$ 和 $\operatorname{Var}(\bar{X})=10000$。這些結果和 CLT 立即意味著

$$
\bar{X}\approx\operatorname{Nor}(\mathrm{E}[\bar{X}],\operatorname{Var}(\bar{X}))\sim\operatorname{Nor}(1000,10000).
$$

因此，

$$
\begin{aligned}
&\mathrm{P}(950\le\bar{X}\le 1050)\\
&=\mathrm{P}\left(\frac{950-\mathrm{E}[\bar{X}]}{\sqrt{\operatorname{Var}(\bar{X})}}\le\frac{\bar{X}-\mathrm{E}[\bar{X}]}{\sqrt{\operatorname{Var}(\bar{X})}}\le\frac{1050-\mathrm{E}[\bar{X}]}{\sqrt{\operatorname{Var}(\bar{X})}}\right)\\
&\doteq\mathrm{P}\left(\frac{950-1000}{100}\le Z\le\frac{1050-1000}{100}\right) \quad (\text{where } Z\sim\operatorname{Nor}(0,1))\\
&=\mathrm{P}(-0.5\le Z\le 0.5)=2\Phi(1/2)-1=0.383. \quad \square
\end{aligned}
$$

備註：由於 X_i 是 iid $\operatorname{Exp}(1/1000)$，我們知道

$$
100\bar{X}=\sum_{i=1}^{100}X_i\sim\operatorname{Erlang}_{100}(1/1000)\sim\operatorname{Gam}(100,1/1000).
$$

這表明這個問題可以完全解決 - 例如，Excel gamma cdf 函數 gammadist $(x, n, 1/\lambda, 1)$ 產生

$$\begin{aligned} \mathrm{P}(950 \leq \bar{X} \leq 1050) &= \mathrm{P}\left(95000 \leq \sum_{i=1}^{100} X_i \leq 105000\right) \\ &= \mathrm{gammadist}(105000, 100, 1000, 1) - \ \mathrm{gammadist}\ (95000, 100, 1000, 1). \end{aligned}$$

你知道嗎，你最終得到完全相同的答案 0.383，所以 CLT 正態近似真的很好！

示例：假設 $X_1, \ldots, X_{100}$ 是來自某個分佈的獨立同分佈，均值為 1000，標準差為 1000。求 $\mathrm{P}(950 \leq \bar{X} \leq 1050)$。

解決方案：通過與前面示例完全相同的操作，答案 $\doteq 0.383$。請注意，我們不關心數據是否來自指數分佈。我們只需要均值和方差。

二項式的正態逼近：假設 $Y \sim \mathrm{Bin}(n, p)$，其中 n 非常大。在這種情況下，我們通常通過適當的正態分佈來近似二項式。CLT 自 $Y = \sum_{i=1}^{n} X_i$ 起適用，其中 X_i 是 iid $\mathrm{Bern}(p)$。然後

$$\frac{Y - \mathrm{E}[Y]}{\sqrt{\mathrm{Var}(Y)}} = \frac{Y - np}{\sqrt{npq}} \approx \mathrm{Nor}(0, 1),$$

在這種情況下 $Y \approx \mathrm{Nor}(np, npq)$。

對於二項式的正態逼近，通常的經驗法則是只要 $np \geq 5$ 和 $nq \geq 5$ 就可以很好地工作。

為什麼我們需要這樣的近似值？

示例：假設 $Y \sim \mathrm{Bin}(100, 0.8)$，我們要計算

$$\mathrm{P}(Y \geq 84) = \sum_{i=84}^{100} \binom{100}{i} (0.8)^i (0.2)^{100-i}$$

祝你好運，二項式係數（它們太大）和要總結的項數（它會變得乏味）。一個小時後我會回來看你。

下一個示例顯示瞭如何使用近似值。請注意，它包含“連續性校正”以解釋二項式是離散的而法線是連續的這一事實。（但如果您不想使用它，請不要太擔心。）

示例：亞特蘭大勇士隊進行 100 場獨立棒球比賽，每場比賽的獲勝概率為 0.8。他們贏得至少 84 場比賽的概率是多少？

答案：$Y \sim \mathrm{Bin}(100, 0.8)$，我們想要 $\mathrm{P}(Y \geq 84)$（如上一個例子）。所以，

$$\begin{aligned} \mathrm{P}(Y \geq 84) &= \mathrm{P}(Y \geq 83.5) \quad (\text{"continuity correction," since } 84 \doteq [83.5, 84.5]) \\ &\doteq \mathrm{P}\left(Z \geq \frac{83.5 - np}{\sqrt{npq}}\right) \quad (\text{CLT, with } Z \sim \mathrm{Nor}(0, 1)) \\ &= \mathrm{P}\left(Z \geq \frac{83.5 - 80}{\sqrt{16}}\right) \\ &= \mathrm{P}(Z \geq 0.875) = 0.1908 \end{aligned}$$

實際答案（使用真正的 Bin(100, 0.8) 分佈）結果是 0.1923 $ - 非常接近！

4.4 正態分佈的擴展

正態分佈有大量的擴展/應用程序。我們在這裡介紹其中的一些。

4.4.1 二元正態分佈

定義：如果 (X, Y) 具有聯合 pdf，則它具有二元正態分佈

$$f(x,y) = C\exp\left\{\frac{-\left[z_X^2(x) - 2\rho z_X(x) z_Y(y) + z_Y^2(y)\right]}{2\left(1-\rho^2\right)}\right\}$$

其中 μ_X、μ_Y、σ_X^2 和 σ_Y^2 是 X 和 Y 的明顯邊際矩,

$$\rho \equiv \operatorname{Corr}(X,Y), \quad C \equiv \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$
$$z_X(x) \equiv \frac{x-\mu_X}{\sigma_X} \quad \text{and} \quad z_Y(y) \equiv \frac{y-\mu_Y}{\sigma_Y}$$

相當討厭的聯合pdf，嗯？

有趣的事實：邊際 $X \sim \operatorname{Nor}(\mu_X, \sigma_X^2)$ 和 $Y \sim \operatorname{Nor}(\ mu_Y, \sigma_Y^2)$

示例：(X, Y) 可以是一個人的（身高、體重）。這兩個量略微正常，但呈正相關。

如果要計算雙變量正態概率，則需要評估數量，例如

$$\mathrm{P}(a < X < b, c < Y < d) = \int_c^d \int_a^b f(x,y)\,dx\,dy$$

這可能需要數值積分技術。R 等統計數據包中還提供了易於使用的函數調用。

有趣的事實（在回歸研究中出現）：假設 $X = x$，Y 的條件分佈也是正常的。尤其是，

$$Y \mid X = x \sim \operatorname{Nor}\left(\mu_Y + \rho\left(\sigma_Y/\sigma_X\right)\left(x-\mu_X\right), \sigma_Y^2\left(1-\rho^2\right)\right).$$

因此，關於 X 的信息有助於更新 Y 的分佈。

示例：考慮一所艱難大學的學生。假設 X 是他們的綜合 SAT 成績（數學和語文），Y 是他們的新生 GPA（滿分 4 分）。假設一項研究表明

$$\mu_X = 1300, \quad \mu_Y = 2.3, \quad \sigma_X^2 = 6400, \quad \sigma_Y^2 = 0.25, \quad \text{and} \quad \rho = 0.6.$$

求 $\mathrm{P}(Y \ge 2 \mid X = 900)$。

解決方法：首先，

$$\begin{aligned}
\mathrm{E}[Y \mid X = 900] &= \mu_Y + \rho\left(\sigma_Y/\sigma_X\right)\left(x-\mu_X\right) \\
&= 2.3 + (0.6)(\sqrt{0.25/6400})(900-1300) \\
&= 0.8
\end{aligned}$$

表示一個 SAT 成績為 900 的孩子的預期 GPA 為 0.8 $。

第二，

$$\operatorname{Var}(Y \mid X = 900) = \sigma_Y^2\left(1-\rho^2\right) = 0.16$$

因此，

$$Y \mid X=900 \sim \operatorname{Nor}(0.8, 0.16)$$

現在我們可以計算

$$\mathrm{P}(Y \geq 2 \mid X=900)=\mathrm{P}\left(Z \geq \frac{2-0.8}{\sqrt{0.16}}\right)=1-\Phi(3)=0.0013.$$

這傢伙考GPA的機會不大。

雙變量正態分佈很容易推廣到多變量情況。

榮譽定義：隨機向量 $\boldsymbol{X}=\left(X_1, \ldots, X_k\right)^{\mathrm{T}}$ 具有多元正態分佈，平均向量 $ =({1}, , {k})^{}$ 和 $k \times k$ 協方差矩陣 $\boldsymbol{Sigma}=\left(\sigma_{ij}\right)$，如果它有多元 pdf

$$f(\boldsymbol{x})=\frac{1}{(2 \pi)^{k / 2}|\boldsymbol{\Sigma}|^{1 / 2}} \exp \left\{-\frac{(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{2}\right\}, \quad \boldsymbol{x} \in \mathbb{R}^k$$

其中 $|\boldsymbol{\Sigma}|$ 和 $\boldsymbol{\Sigma}^{-1}$ 分別是 $\boldsymbol{\Sigma}$ 的行列式和倒數。

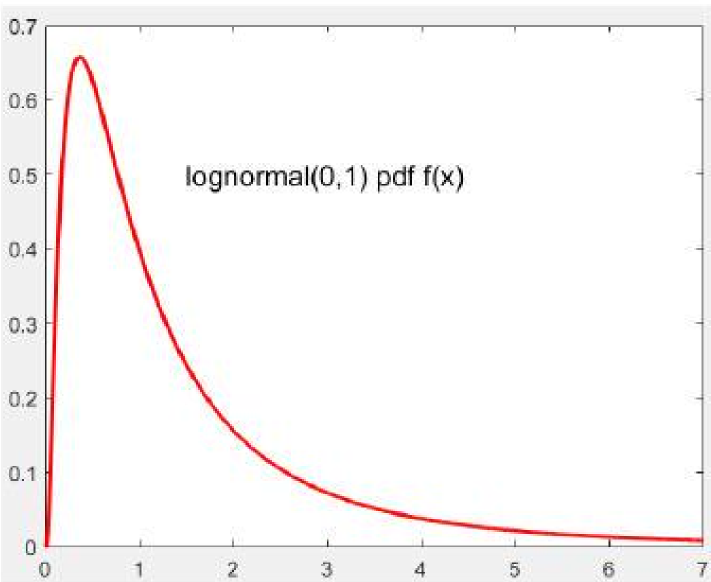
原來對於$i, j=1,2, \ldots, k,$

$$\mathrm{E}\left[X_i\right]=\mu_i, \quad \operatorname{Var}\left(X_i\right)=\sigma_{ii}, \quad \text{and } \operatorname{Cov}\left(X_i, X_j\right)=\sigma_{ij}.$$

符號：$\boldsymbol{X} \sim \operatorname{Nor}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

4.4.2 對數正態分佈

定義：如果$Y \sim \operatorname{Nor}\left(\nu, \tau^2\right)$，則$X \equiv e^Y$ 具有參數$\left(\nu, \tau^2\right)$。這種分佈具有巨大的用途，例如，在某些股票期權的定價中。



Fun Facts：讓 $\phi(x)$ 和 $\Phi(x)$ 分別表示我們的老朋友，標準正態分佈的 pdf 和 cdf。那麼對數正態的 pdf、cdf 和矩是

$$
\begin{aligned}
f(x) &= \frac{1}{\tau x}\phi\left(\frac{\ln(x)-\nu}{\tau}\right) = \frac{1}{x\tau\sqrt{2\pi}}\exp\left\{-\frac{[\ln(x)-\nu]^2}{2\tau^2}\right\}, x>0, \\
F(x) &= \Phi\left(\frac{\ln(x)-\nu}{\tau}\right), x>0, \quad \text{and} \\
E\left[X^k\right] &= \exp\left\{k\nu+\frac{k^2\tau^2}{2}\right\}, \quad k=1,2,\ldots
\end{aligned}
$$

尤其是，

$$
\mathrm{E}[X] = e^{\nu+\frac{\tau^2}{2}} \quad \text{and } \operatorname{Var}(X) = e^{2\nu+\tau^2}\left(e^{\tau^2}-1\right)
$$

示例：假設 $Y \sim \mathrm{Nor}(10,4)$，令 $X=e^Y$。然後

$$
\mathrm{P}(X \leq 1000) = \mathrm{P}\left(Z \leq \frac{\ln(1000)-10}{2}\right) = \Phi(-1.55) = 0.061
$$

和

$$
\mathrm{E}[X] = e^{10+\frac{4}{2}} = e^{12},
$$

這當然會引起很高的期望！

瘋狂的事實：雖然所有對數正態矩都存在，但 mgf 不存在！這是因為 $\mathrm{E}\left[e^{tX}\right]$ 對於 $t>0$ 是無限的。

4.4.2.1 榮譽示例：如何獲得諾貝爾獎

眾所周知，股票價格與對數正態分佈密切相關。實際上，對於固定時間 $t>0$ 的股票價格，通常使用以下模型

$$
S(t) = S(0)\exp\left\{\left(\mu-\frac{\sigma^2}{2}\right)t+\sigma\sqrt{t}Z\right\}
$$

其中 μ 與股票價格的“漂移”（即自然增長率）有關，σ 是其“波動性”（股票反彈的幅度），$S(0)$是初始價格，Z 是標準的普通房車。

金融的一個活躍領域是估計期權價格。例如，一個所謂的歐式看漲期權 C 允許其所有者在預先確定的到期日 $T 以預先商定的執行價格 k 購買該股票，該所有者為該特權支付了預付費用。$。

例如，假設 IBM 目前以每股 100 $的價格出售。如果我認為股票會升值，我可能想現在支付 $3/ 股，以獲得三個月後以 $105 的價格購買 IBM 的權利。

- 如果IBM 三個月後價值$120，我將能夠以$105 購買它，並且將獲得$120 − $105 − $3的利潤 = $12。
- 如果三個月後 IBM 以 107 $的價格出售，我仍然可以以 105 $的價格購買它，但會損失 1 $（從我最初的期權購買中收回 2 $）。
- 如果 IBM 以 $95 的價格出售，那麼我將不會行使我的選擇權，並且會夾著尾巴走開，因為我失去了最初的 $3。

那麼期權的價值是多少（我應該為此付出什麼）？它的期望值為

$$
\mathrm{E}[C] = e^{-rT}\mathrm{E}\left[(S(T)-k)^+\right]
$$

其中 $x^+ \equiv \max\{0,x\}$ 和 - r 是“無風險”利率，例如，您可以從美國國債中獲得的利率。這是用來代替漂移 μ。

- 術語e^{-rT} 表示貨幣的時間價值，即，如果我用我的錢購買一張國庫券，我可能獲得的利息對應的折舊項。

使用標準條件參數，我們可以計算

$$\begin{aligned}
\mathrm{E}[C] &= e^{-rT}\mathrm{E}\left[S(0)\exp\left\{\left(r-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}Z\right\}-k\right]^+ \\
&= e^{-rT}\int_{-\infty}^{\infty}\left[S(0)\exp\left\{\left(r-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}z\right\}-k\right]^+\phi(z)dz \\
&= S(0)\Phi(b+\sigma\sqrt{T})-ke^{-rT}\Phi(b) \quad \text{(after lots of algebra)}
\end{aligned}$$

其中 $\phi(\cdot)$ 和 $\Phi(\cdot)$ 是 $\mathrm{Nor}(0,1)$ pdf 和 cdf（和往常一樣），並且

$$b \equiv \frac{rT-\frac{\sigma^2 T}{2}-\ell n(k/S(0))}{\sigma\sqrt{T}}$$

布萊克和斯科爾斯在 1968 年進行了最初的計算（儘管方式不同）。1997 年，斯科爾斯和默頓因這項工作獲得了諾貝爾獎（可憐的布萊克已經到了他自己的有效期（ ））。這項工作有很多很多的概括用於實際金融應用，但這是一個很好的起點。同時，獲得前往挪威或瑞典或任何頒發諾貝爾經濟學獎的地方的門票！

4.5 計算機注意事項

4.5.1 評估 pmf’s / pdf’s 和 cdf’s

我們可以使用各種計算機軟件包（Excel、Minitab、R、SAS 等）來計算各種常見分佈的 pmf’s/pdf’s 和 cdf’s 值。例如，在 Excel 中，我們可以找到以下函數：

- binomdist = 二項分佈

-expondist = 指數

- negbinomdist = 負二項式
- normdist 和 normsdist = Normal 和 Standard Normal
- 泊松 = 泊松。

諸如 normsinv、tinv 等 Excel 函數可以計算標準正態、t 和其他分佈的倒數。此功能對於獲取分佈分位數很有用。它也可用於模擬各種 RV。

4.5.2 模擬隨機變量

如何模擬房車？這是一個重要的問題，因為模擬用於評估各種包含固有隨機性的現實世界過程（排隊系統、庫存系統、製造系統等）。雖然這不是模擬文本，但我們將快速提及一些關於 RV 生成的基礎知識。

首先，您可以使用 Excel 函數 rand 來模擬 $\mathrm{Unif}(0,1)$RV，例如 U。然後，基於逆變換定理 (§2.8.2，您可以模擬任何其他具有 cdf $F(x)$ 的連續 RVX，只需將 into $X=F^{-1}(U)$

當然，您必須手頭有反函數才能進行此計算，有時這很困難，因為反函數並不總是以封閉形式存在。標準統計數據包部分地為您解決了這個問題。例如，您可以使用 Excel 標準正態逆函數通過 normsinv (rand()) 生

成標準正態 RV。離散 RV 的逆變換方法類似，但您必須小心一點，因為對於離散情況，逆變換並不總是唯一的。

除了逆變換之外，還有許多 RV 生成方法。如果您感興趣，請查看 $2.9 中的練習 27 以獲得更多見解。

我們以一個非凡的 RV 生成示例結束本章，該示例使用了一種替代方法。.

定理（Box 和 Muller）：如果 $U_1, U_2 \stackrel{\text{iid}}{\sim} \text{Unif}(0,1)$，那麼

$$Z_1 = \sqrt{-2\ell n\left(U_1\right)}\cos(2\pi U_2) \quad \text{and} \quad Z_2 = \sqrt{-2\ell\ln(U_1)}\sin(2\pi U_2)$$

是 iid $\text{Nor}(0,1)$

示例：假設 $U_1 = 0.3$ 和 $U_2 = 0.8$ 是兩個 iid Unif $(0,1)'$ 的實現。使用 Box-Muller 方法生成兩個 iid 標準法線。

解決方案：我們有

$$\begin{aligned} Z_1 &= \sqrt{-2\ln(U_1)}\cos(2\pi U_2) = 0.480 \\ Z_2 &= \sqrt{-2\ln(U_1)}\sin(2\pi U_2) = -1.476. \end{aligned}$$

4.5.2.1 備註：

- 有很多方法可以生成 $\text{Nor}(0,1)'$ ，但這可能是最簡單的。
- 餘弦和正弦項以弧度計算，而不是度數。
- 要從 $Z \sim \text{Nor}(0,1)$ 得到 $X \sim \text{Nor}\left(\mu, \sigma^2\right)$，只需取 $X = \mu + \sigma Z$。
- 令人驚訝的是，它是“穆勒”，而不是“穆勒！”

[3] 一開始是一個很好的起點。

4.5.2.2 Box-Muller 的榮譽證明

我們遵循 {3.7.1 的主要推論給出的方法，即如果我們可以表示 $U_1 = k_1\left(Z_1, Z_2\right)$ 並且$U_2 = k_2\left(Z_1, Z_2\right)$ 對於一些函數 $k_1(\cdot,\cdot)$ 和 $k_2(\ cdot,\cdot)$，則 (Z_1, Z_2) 的聯合 pdf 由下式給出

$$\begin{aligned} g\left(z_1, z_2\right) &= f_{U_1}\left(k_1\left(z_1, z_2\right)\right) f_{U_2}\left(k_2\left(z_1, z_2\right)\right)\left|\frac{\partial u_1}{\partial z_1}\frac{\partial u_2}{\partial z_2} - \frac{\partial u_2}{\partial z_1}\frac{\partial u_1}{\partial z_2}\right| \\ &= \left|\frac{\partial u_1}{\partial z_1}\frac{\partial u_2}{\partial z_2} - \frac{\partial u_2}{\partial z_1}\frac{\partial u_1}{\partial z_2}\right| \quad (U_1 \text{ and } U_2 \text{ are iid Unif(0,1)}). \end{aligned}$$

為了獲得函數 $k_1\left(Z_1, Z_2\right)$ 和 $k_2\left(Z_1, Z_2\right)$，注意

$$Z_1^2 + Z_2^2 = -2\ln(U_1)\left[\cos^2(2\pi U_2) + \sin^2(2\pi U_2)\right] = -2\ln(U_1),$$

以便

$$U_1 = e^{-(Z_1^2+Z_2^2)/2}.$$

這立即意味著

$$\begin{aligned} Z_1^2 &= -2\ln(U_1)\cos^2(2\pi U_2) \\ &= -2\ln\left(e^{-\left(Z_1^2+Z_2^2\right)/2}\right)\cos^2(2\pi U_2) \\ &= \left(Z_1^2+Z_2^2\right)\cos^2(2\pi U_2), \end{aligned}$$

以便

$$U_2 = \frac{1}{2\pi}\arccos\left(\pm\sqrt{\frac{Z_1^2}{Z_1^2+Z_2^2}}\right) = \frac{1}{2\pi}\arccos\left(\sqrt{\frac{Z_1^2}{Z_1^2+Z_2^2}}\right),$$

我們在哪裡（非嚴格地）擺脫“ ± ”以平衡 $y = \arccos(x)$ 的範圍僅被視為 $0 \le y \le \pi$ 的事實（不是 $0 \le y \le 2\pi$ ）。

現在一些衍生的樂趣。讓我們先從簡單的開始。

$$\frac{\partial u_1}{\partial z_i} = \frac{\partial}{\partial z_i}e^{-\left(z_1^2+z_2^2\right)/2} = -z_i e^{-\left(z_1^2+z_2^2\right)/2}, i = 1, 2.$$

然後是更具挑戰性的傢伙。

$$\begin{aligned} \frac{\partial u_2}{\partial z_1} &= \frac{\partial}{\partial z_1}\frac{1}{2\pi}\arccos\left(\sqrt{\frac{z_1^2}{z_1^2+z_2^2}}\right) \\ &= \frac{1}{2\pi}\frac{-1}{\sqrt{1-\frac{z_1^2}{z_1^2+z_2^2}}}\frac{\partial}{\partial z_1}\sqrt{\frac{z_1^2}{z_1^2+z_2^2}} \quad \text{(chain rule)} \\ &= \frac{1}{2\pi}\frac{-1}{\sqrt{\frac{z_2^2}{z_1^2+z_2^2}}\frac{1}{2}}\left(\frac{z_1^2}{z_1^2+z_2^2}\right)^{-1/2}\frac{\partial}{\partial z_1}\frac{z_1^2}{z_1^2+z_2^2} \quad \text{(chain rule ag} \\ &= \frac{-\left(z_1^2+z_2^2\right)}{4\pi z_1 z_2}\frac{2z_1 z_2^2}{\left(z_1^2+z_2^2\right)^2} \\ \frac{\partial u_2}{\partial z_2} &= \frac{-z_2}{2\pi\left(z_1^2+z_2^2\right)}, \quad \text{and} \\ &= \frac{z_1}{2\pi\left(z_1^2+z_2^2\right)} \quad \text{(after similar algebra).} \end{aligned}$$

然後我們終於有了

$$\begin{aligned} g(z_1, z_2) &= \left|\frac{\partial u_1}{\partial z_1}\frac{\partial u_2}{\partial z_2} - \frac{\partial u_2}{\partial z_1}\frac{\partial u_1}{\partial z_2}\right| \\ &= \left|-z_1 e^{-\left(z_1^2+z_2^2\right)/2}\frac{z_1}{2\pi\left(z_1^2+z_2^2\right)} - \frac{z_2}{2\pi\left(z_1^2+z_2^2\right)}z_2 e^{-\left(z_1^2+z_2^2\right)/2}\right| \\ &= \frac{1}{2\pi}e^{-\left(z_1^2+z_2^2\right)/2} = \left(\frac{1}{\sqrt{2\pi}}e^{-z_1^2/2}\right)\left(\frac{1}{\sqrt{2\pi}}e^{-z_2^2/2}\right) \end{aligned}$$

這是兩個 iid $\mathrm{Nor}(0,1)$ pdf 的乘積，所以我們完成了！

4.6 練習

1. (4.1.1) 我的諮詢業務正在進行六個獨立的項目，每個項目的成功概率估計為 0.9 $。至少五個成功的概率是多少？

2. (4.1.1 使用矩生成函數求$\operatorname{Bin}(n, p)$ 分佈的均值和方差。

3. (4.1.2 假設我們有一盒襪子——四隻紅襪和五隻藍襪。讓我們在不放回的情況下對三隻襪進行抽樣。找到你得到的預期紅襪數量。

4. ($4.1.3 我可以在我的籃球罰球中賺到 80%。假設獨立投籃，第四次投球時我第一次投失的概率是多少？ 5. ($4.1.3) 如果本月生產中心的訂單是一個 Geom (0.7) 隨機變量，求我們最多有三個訂單的概率。

5. (4.1.3) 假設測試中的問題是獨立同分佈的，你能夠以 0.9 的概率正確回答任何問題。在您做出第二個錯誤答案之前，您將完成的預期問題數量是多少？

6. ($4.1.4) 顧客按照泊（10/小時）流程到達麵包店。麵包店可以在一小時內接待多達 15 位顧客，而不會超載。下一小時超載的概率是多少？

7. ($4.2.1) 假設 $U_1, U_2, \ldots$ 是 Unif $(0,1)$ 隨機變量的 iid 樣本。我正在進行一項實驗，在該實驗中，我按順序對 U_i 進行採樣，直到其中一個大於 0.75，此時實驗停止。讓我們用隨機變量 N 來表示停止時間。因此，例如，如果我觀察到 $U_1=0.62$，$U_2=0.17$，$U_3=0.84$，那麼我會在樣本 $N=3$ 處停止。求出預期的停止時間，$\mathrm{E}[N]$。

8. ($4.2.1) 回想一下 Unif$(a, b)$ 分佈的 mgf 是

$$M_X(t)=\mathrm{E}\left[e^{tX}\right]=\int_a^b e^{tx} \frac{1}{b-a} dx=\frac{e^{tb}-e^{ta}}{t(b-a)}$$

使用這個事實來找到 X 的期望值。

10. ($4.2.2) 空調故障的時間是指數級的，平均為四年。

a. 空調在四年前出現故障的概率是多少？

b. 假設空調已經使用了兩年。在六年大關之前失敗的概率是多少？

11. ($4.2.2) 根據泊松過程，孩子們以 13 / 小時的不幸率到達一所房子做萬聖節不給糖就搗蛋。15^{th} 和 16^{th} 到達之間的時間超過 4 分鐘的概率是多少？（提示：認為指數。）

12. ($4.2.2) 假設 X 和 Y 是具有速率 $\lambda=1/3$ 的 iid 指數隨機變量。求 $\mathrm{P}(1 \leq X+Y \leq 2)$

13. ($4.2.2) 如果$X \sim \operatorname{Gam}(\alpha, \lambda)$，求$M_X(t), \mathrm{E}[X]$, and $\operatorname{Var}(X)$。

14. ($4.2.2) (a) 證明如果$\alpha>0$，則$\Gamma(\alpha+1)=(\alpha) \Gamma(\alpha)$。(b) 當你在做的時候，證明$\Gamma(1/2)=\sqrt{\pi}$。

15. (4.2.3 如果$X \sim \operatorname{Beta}(a, b)$，導出$\mathrm{E}[X]$ 和$\operatorname{Var}(X)$。

16. ($4.2.3) 如果$X \sim$ Weibull (a, b)，導出$\mathrm{E}[X]$ 和$\operatorname{Var}(X)$。17. ($4.2.3) 如果隨機變量 X 具有 cdf，則稱其具有帕累托分佈，參數 $\lambda>0$ 和 $\beta>0$

$$F(x)=1-(\lambda/x)^\beta, \quad \text{for } x \geq \lambda$$

Pareto 是一種重尾分佈，在統計建模中有多種用途。求 $\mathrm{E}[X]$（但當 $\beta \leq 1$ 時要小心，因為那是重尾引起問題的時候）。

18. (4.3.1) 假設$X \sim \mathrm{Nor}(3,10)$，$Y \sim \mathrm{Nor}(-4,3)$，並且$X$和$Y$是獨立的. 求 $W = -2X + Y$ 的分佈。

19. ($4.3.2) 假設 Z 是標準正態。尋找

    a. $\mathrm{P}(-2 < Z < 0)$。

    b. $\mathrm{P}(-1 < Z < 1)$。

    c. $\mathrm{P}(Z > 1.65)$

    d. $\mathrm{P}(Z > -1.96)$。

    e. $\mathrm{P}(|Z| > 1.2)$

20. ($4.3.2) 求 z 使得 $\Phi(z) = 0.92$。

21. (4.3.2 如果 $\Phi(x)$ 是標準的普通 cdf，請使用你的武器庫中的任何方法找到 x 使得 $\Phi(x) = 2x$。我希望你找到至少保留兩位小數的答案。

22. ($4.3.2) 如果 X 具有均值 2 和方差 9 的正態分佈，求 $X \leq 5$ 的概率。

23. (4.3.2) 假設 ABC 大學學生的 SAT 數學成績可以近似為均值為 700 方差為 225 的正態分佈。找出學生至少得分 715 的概率。

24. (4.3.2) 如果 W, X, Y, Z 是 iid $\mathrm{Nor}(0,1)$RVs，求 $\mathrm{P}(W + X + Y + Z \leq 2)$.

25. (4.3.2) 如果 $Z \sim \mathrm{Nor}(0.1)$，找到 $\mathrm{E}\left[Z^{2k}\right]$ for $k=1,000. 2.3. 4 $。

26. ($4.3.3) 假設 $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \mathrm{Nor}(\mu, 9)$，其中$\mu$ 未知。用 $\bar{X} = \sum_{i=1}^{n} X_i/n$ 表示樣本均值。找到樣本大小 n 使得

$$\mathrm{P}(|\bar{X} - \mu| \leq 1) \geq 0.95.$$

換句話說，您應該進行多少次觀察才能使 $\bar{X}$ 有 95% 的機會接近 μ ？

27. ($4.3.4) 宇宙中最重要的定理是什麼？

    a. 東方極限定理

    b. 中心極限定理

    c. 中央極限血清

    d. Central Simit Theorem (simit 是一種美味的土耳其百吉餅) 28. ($4.3.5) 假設 $X_1, \ldots, X_{600}$ 是獨立同分佈的，值為 1,0 和 -1 ，每個概率為 1/3。（這就是所謂的隨機遊走。）求 sum $\sum_{i=1}^{600} X_i$ 最多為 40 的近似概率。

29. ($4.3.5 100 個大彈珠裝在一個盒子裡。每個平均重 2 盎司，標準差為 0.2 盎司。求一個盒子重量小於 202 盎司的近似概率。

30. (4.3.5 如果 $X_1, \ldots, X_{400}$ 來自某個分佈的獨立同分佈，均值為 1，方差為 400，求樣本均值 $\bar{X}$ 更大的近似概率超過 2 $。

31. ($4.3.5) 一個生產過程生產的項目，其中 6% 是有缺陷的。每天隨機抽取 200 件商品，統計有缺陷的商品數量 X。使用二項式的正態逼近，求 $\mathrm{P}(X \leq 10)$

32. (4.4.1 用(X, Y) 表示來自某個群體的隨機男性的配對(身高, 體重)。假設(X, Y) 與 μ_X是二元正態的 $= 70$ 英寸，$\sigma_X^2 = 100\text{in}^2$，$\mu_Y = 150$ 磅，$\sigma_Y^2 = 225\text{lb}^2$，$\rho = \text{Corr}(X, Y)$ =0.8。

求 $\mathrm{P}(Y \geq 165 \mid X = 75)$

33. ($4.4.2 正如我們在文中所討論的，您可以使用正態/對數正態分佈的屬性來估計股票的期權價格。我不會讓您嚴格地或通過模擬來這樣做，但我會給你一個快速查找任務。當我在 2020 年 7 月 10 日寫這篇文章時，IBM 目前以每股 118.35 $的價格出售。假設我有興趣保證我可以購買 IBM 的股票2020 年 10 月 16 日最多 $130。查找（可能使用互聯網上的 FaceTube 之類的東西）各種 IBM 期權價格。

34. 4.4.2 假設 $Y \sim \text{Nor}(1, 4)$，因此 $X = e^Y$ 是對數正態的。求 $\mathrm{P}(X > e)$。

35. (4.4.2) 假設隨機變量Y 具有Nor $(50, 25)$ 分佈。求 $X = e^Y$ 的均值和方差，然後求 $\mathrm{P}(X \leq \mathrm{E}[X])$。

36. (4.5) 假設我們有兩個 iid $\text{Unif}(0, 1)'\text{s}$，$U_1 = 0.6$ 和 $U_2 = 0.9$。

a. 使用 Box-Muller 方法生成兩個 iid Nor $(0, 1)$ 隨機變量 Z_1 和 Z_2。

b. $Z_1^2 + Z_2^2$的分佈是什麼。（提示：使用 Box-Muller 中的 Z_1 和 Z_2 的形式，看看你是否想出了任何有趣的東西。）

c. 使用前兩部分的結果生成一個 Exp(1/2) 隨機變量。

d. 我們都對“Box-Muller”沒有變音符號感到失望。你至少能找到一些有變音符號的消費品或重金屬樂隊嗎？

37.命名那個分佈！

a. 如果 $Z \sim \text{Nor}(0, 1)$，$3Z - 2$ 的分佈是什麼？(b) 如果 $Z \sim \text{Nor}(0, 1)$ 和 $\Phi(\cdot)$ 是標準正態 cdf，那麼討厭的隨機變量 $\Phi(Z)$ 的分佈是什麼？

b. 如果$U \sim \text{Unif}(0, 1)$，$-10 \ln(\sqrt{U})$ 的分佈是什麼？

c. 如果 U_1, U_2, U_3 是 iid $\text{Unif}(0, 1)$，那麼分佈是什麼

$$-3 \ln(U_1 (1 - U_2)(1 - U_3))?$$

e. 如果 U 和 V 是 iid $\text{Unif}(0, 1)$，那麼

$$-2 + \sqrt{-\ell n(U)} \cos(2\pi V) + \sqrt{-\ell n(U)} \sin(2\pi V)?$$

（提示： Think Box-Muller 起價 $4.5.2 ）

38. 數學獎勵：假設 a、k 和 e 都是非零的。用披頭士的歌詞來證明$m = t$。

ryNoSmBk

ryCh 5 【第 5 章】

5.0.0.1 描述性統計

現在進入正文的第二個主要部分——統計！到目前為止，我們已經使用了幾種適用於模擬現實世界現象的概率分佈，例如到達呼叫中心的客戶數量、電子元件的壽命、人的體重等。但是我們怎麼知道我們使用的具體模型實際上是否合理？在給定這些模型的情況下，我們如何做出適當的決定？

統計學使用觀察樣本數據來得出關於從中抽取樣本的總體的一般結論。我們將在 \$5.1 開始本章，通過回顧一些基本的描述性統計數據（直方圖、樣本均值等性能度量等），然後回顧適合使用某些分佈的情況。例如，您什麼時候想使用泊松分佈？然後，我們將在 \$5.2 中討論各種技術來估計與這些分佈相關的任何未知參數。例如，如果您認為零件修復時間可能是指數的，那麼您仍然需要估計費率 λ。

最後，我們將在 \$5.3 中做一個簡短的介紹，以了解幾個在統計學中特別重要的抽樣分佈。特別是，我們將結識一些新朋友——χ^2、Student t 和 F 分佈。該材料將為我們為第 6 章做準備，在那裡我們將推導出各種置信區間，這些置信區間可用於為任何未知分佈參數的值提供一定的下限和上限；然後是第 77 章，我們將在其中描述各種假設檢驗，以確定我們最終選擇的分佈（包括其參數）是否充分適合我們觀察到的數據集。

\$5.1 - 統計簡介

\$5.2 - 點估計

\$5.3 - 抽樣分佈

5.1 統計學概論

我們將通過對數據的高級討論以及一些可用作第一遍的簡單數據分析技術來輕鬆進入該主題。

5.1.1 什麼是統計數據？

統計數據構成了使用觀察或實驗數據進行決策的合理基礎。我們在面臨不確定性的情況下做出這些決定，而統計數據幫助我們回答以下問題：

- 一個群體（或系統）的分析。
- 兩個或多個群體的比較。哪個是最好的系統？

示例：統計無處不在！

- 選舉投票。
- 可口可樂與百事可樂。
- 吸煙對患癌症概率的影響。
- 新疫苗對感染肝炎概率的影響。
- 在特定時間段內最受歡迎的電視節目是什麼？
- 各種熱處理方法對鋼抗拉強度的影響。
- 哪些肥料可以提高作物產量？

想法（選舉投票示例）：我們不能投票給每個選民。因此，我們從（大量）選民中抽取（少量、有效、有代表性的）數據樣本，並嘗試根據樣本得出合理的結論。

博弈計劃：統計告訴我們如何進行抽樣（即要進行多少觀察，如何進行觀察等），然後如何從抽樣數據中得出結論。一般來說，統計研究大致按以下方式進行。

1. 收集和匯總數據以供分析（在本節中討論）。
2. 確定/估計基礎分佈（連同相關參數），例如，$\mathrm{Nor}\,(30, 8)$（在 $5.2 和第 60 章中討論。
3. 進行統計檢驗，看看你的分佈或一些相關假設是否“大致”正確（第 7 章 .
4. 必要時循環這些步驟，因為我們對其他問題感到好奇和/或遇到對更多數據的需求。

5.1.1.1 數據類型：

我們對數據進行採樣和分析的方式有時取決於可用數據的類型。

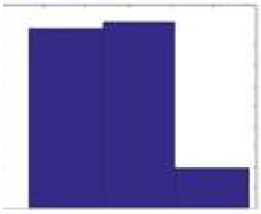
- 離散變量：此類數據只能取特定值，例如，本週工廠發生的事故數量，或可能擲出的一對骰子。
- 連續變量：可以在一定間隔內取任何實際值的數據。例如，燈泡的使用壽命，或新生兒的重量。
- 分類變量：這種類型的數據不是數字的。例如，在某個時間段內您最喜歡（廣播）的電視節目是什麼？

5.1.2 描述性統計

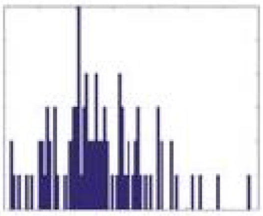
在本小節中，我們將討論如何匯總數據。

一張圖片值1000字。因此，如果只是為了識別諸如非標準分佈、缺失數據點、異常值等明顯問題，則應始終在做任何其他事情之前繪製數據。

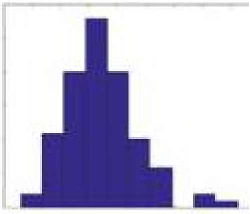
直方圖可以快速、簡潔地查看您正在處理的內容。可以證明（通過所謂的 Glivenko-Cantelli 定理），如果您進行足夠的觀察，直方圖最終將收斂到真實分佈，從而表明直方圖在描繪數據方面是可靠的。一個棘手的問題是確定對數據進行分區的單元格的適當數量（圖 5.1。幸運的是，許多軟件包（例如 Excel、R、Matlab）可以被哄騙自動完成這項工作。



不是 enuf 細胞



太多



正好！

圖 5.1：描繪相同數據的不同直方圖。

示例：測試成績（即原始數據）：

23	62	91	83	82	64	73	94	94	52
67	11	87	99	37	62	40	33	80	83
99	90	18	73	68	75	75	90	36	55

等級的莖葉圖。這個老朋友是寫下所有數據的簡單方法。它節省了一些空間，看起來像一個橫向直方圖。

$$
\begin{array}{r|l}
9 & 9944100 \\
8 & 73320 \\
7 & 5533 \\
6 & 87422 \\
5 & 52 \\
4 & 0 \\
3 & 763 \\
2 & 3 \\
1 & 81
\end{array}
$$

分組數據。有時，將數據分組到間隔桶中會提供信息，以及有關桶中項目的頻率和累積頻率的信息。

擁有大量數據真是太好了。但有時這太好了！所以我們經常需要用數字來簡潔地總結數據集。

匯總統計通過向我們提供有關數據特徵的信息（例如樣本量、樣本均值、樣本中位數和样本標準差），為我們描繪了一幅極其簡潔的數據圖景。

在運行數據集中，我們有。.

- $n=30$ 觀察。
- 如果 X_i 是 i^{th} 分數，那麼樣本均值是

$$
\bar{X} \equiv \sum_{i=1}^{n} X_i / n = 66.5
$$

樣本均值是集中趨勢的度量，它是真實但通常未知的均值 $\mathrm{E}\left[X_i\right]$ 的估計量。

- 樣本中位數是集中趨勢的另一種度量，當 X_i 以數字方式排列時，它只是"中間"觀察值。

示例：數據集 $16,7,83$ 給出的樣本中位數為 16 。當樣本量是偶數時，"合理"的方法是只使用兩個中間值的平均值，例如，集合 $16,7,83,20$ 給出 $=18 $。我們的運行數據集的樣本中位數是 73 。備註：樣本中位數比樣本均值更不容易受到"異常值"數據的影響。一個錯誤的數字可能會破壞樣本均值的一整天。

示例：$7,7,7,672,7$ 的樣本均值為 140，樣本中位數為 7.

- 樣本方差（真實但未知方差 $\operatorname{Var}(X_i)$) 的估計量是

$$
S^2 \equiv \frac{1}{n-1} \sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 = 630.6.
$$

方差是散佈或分散的量度。

- 其他此類散佈度量是樣本標準差 $S=+\sqrt{S^2}$ 和样本範圍 $\max_i X_i - \min_i X_i$。

備註：在您進行任何觀察之前，必須將 $\bar{X}$ 和 S^2 視為隨機變量，並分別稱為未知均值和方差的估計量。獲取數據後，$\bar{X}$ 和 S^2 成為常數，然後是未知均值和方差的估計值。

備註：假設數據取 p 不同的值 X_1、 ... 、X_p，頻率分別為 f_1、 ... 、f_p。要快速有效地計算 $\bar{X}$ 和 S^2，只需取

$$\bar{X} = \frac{1}{n}\sum_{j=1}^{p} f_j X_j \quad \text{and} \quad S^2 = \frac{\sum_{j=1}^{p} f_j X_j^2 - n\bar{X}^2}{n-1}.$$

示例：假設我們擲骰子 10 次。

那麼$\bar{X} = (2\cdot 1 + 1\cdot 2 + \cdots + 3\cdot 6)/10 = 3.7$，並且

$$S^2 = \frac{\left(2\cdot 1^2 + 1\cdot 2^2 + \cdots + 3\cdot 6^2\right) - 10(3.7)^2}{9} = 3.789$$

備註：如果個別觀察不能在頻率分佈中精確分類，您可以將觀察 □ 分解為 c 區間，並簡單地近似 $\bar{X}$ 和 S^2 .

示例：考慮下圖 $c = 3$，其中我們用 $m_{j} 表示 j^{th} 區間的中點，j=1,2,3$，並且總樣本量 $n = \sum_{j=1}^{c} f_j = 30$。

[1] 分手很難。然後我們有近似值

$$\begin{aligned}
\bar{X} &\doteq \frac{\sum_{j=1}^{c} f_j m_j}{n} = \frac{10(125) + 15(175) + 5(250)}{30} = 170.833 \\
S^2 &\doteq \frac{\sum_{j=1}^{c} f_j m_j^2 - n\bar{X}^2}{n-1} \\
&= \frac{\left[10(125)^2 + 15(175)^2 + 5(250)^2\right] - 30(170.833)^2}{29} \\
&= 1814.
\end{aligned}$$

請注意，一般來說，在上述 S^2 的計算中保持適當的數值精度是個好主意 - 主要是因為我們可能會從其他大數中減去大數，在這種情況下很容易丟失有效數字。

5.1.3 候選分佈

既然我們已經使用描述性統計來大致了解數據的性質，也許是時候對我們正在處理的概率分佈類型做出明智的猜測了。我們將在第 7 章中介紹更正式的分佈擬合方法

從某種意義上說，我們可能會很幸運，因為"標準"分佈（例如幾何、指數或正態）會導致完美的、直觀的數據擬合；或者我們可能不得不下注並使用非標準解決方案。但是我們如何開始決定呢？讓我們從思考以下問題開始。

- 數據是離散分佈、連續分佈還是混合分佈？
- 單變量/多變量？
- 有多少數據可用？
- 是否有專家詢問數據的性質？

- 如果我們沒有太多/任何數據怎麼辦 - 我們至少可以猜測一個好的分佈嗎？

如果分佈是離散隨機變量，那麼我們有許多熟悉的選擇可供選擇。

- $\mathrm{Bernoulli}(p)$（成功概率為 p ）
- $\mathrm{Binomial}(n, p)$（ $n\,\mathrm{Bern}(p)$ 試驗中的成功次數）
- 幾何 (p)（ $\mathrm{Bern}(p)$ 試驗次數，直到第一次成功）
- 負二項式（ $\mathrm{Bern}(p)$ 試驗次數，直到多次成功）
- Poisson (λ)（計算一段時間內到達的次數）
- 經驗（基於直方圖的通用“樣本”分佈）如果數據表明是連續分佈。.
- 統一（從數據中知道的不多，除了可能的最小值和最大值）
- 三角形（至少我們對最小值、最大值和“最可能”值有一個想法）
- 指數 (λ)（例如，泊松過程的到達間隔時間）
- 正常（身高、體重、智商、樣本均值等的良好模型）
- Beta（適用於指定有界數據）
- Gamma、Weibull、Gumbel、對數正態（可靠性數據）
- 經驗（我們的萬能朋友）

我們現在的遊戲計劃是： 選擇一個“合理的”分佈並估計 $5.2 中的相關參數，除了我們承諾在 $5.3 中採樣分佈之外。在第 6 章中進行額外的置信區間分析，然後通過在第 7 章中正式提出和評估我們的分佈選擇最終將所有內容放在一起

5.2 點估計

假設根據收集數據的直方圖以及過去的經驗，我們相信某些到達間隔時間遵循 $\mathrm{Exp}(\lambda)$。在第 7 章進行正式假設檢驗之前，我們需要估計指數分佈的未知參數值 λ。在本節中，我們將介紹估計的概念，然後介紹各種估計技術。

5.2.1 估計簡介

定義：統計量是觀測值 X_1、 $\ldots$ 、X_n 的函數，並且不明確依賴於任何未知參數。

（統計）示例：樣本均值和样本方差，通常分別表示為 $\bar{X} \equiv \sum_{i=1}^{n} X_i/n$ 和 $S^2 \equiv \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2/(n-1)$。

（參數的）示例：總體的真實但未知的期望值 μ 和總體方差 σ^2。因此，$(\bar{X} - \mu)/\sigma$ 不是一個統計量，因為該數量包含未知參數。

備註：統計數據是隨機變量。如果我們取兩個不同的樣本，我們’ d 期望得到兩個不同的統計值。但請注意，在觀察到 X_i 之後（此時它們不再是隨機的），您實際上可以計算統計數據 - 不涉及未知參

數。統計量通常用於從 X_i 的潛在概率分佈中估計一些未知參數。

令 $X_1, \ldots, X_n$ 為獨立同分佈 RV 並令函數 $T(\mathbf{X}) \equiv T\left(X_1, \ldots, X_n\right)$ 是基於 X_i 的統計數據。假設我們使用 $T(\mathbf{X})$ 來估計某個未知參數 θ。則 $T(\mathbf{X})$ 稱為 θ 的點估計器。

示例：樣本均值 $\bar{X}$ 通常是真實均值 $\mu = \mathrm{E}\left[X_i\right]$ 的點估計量；並且樣本方差 S^2 通常是真實方差 $\sigma^2 = \operatorname{Var}(X_i)$ 的點估計量

如果 $T(\mathbf{X})$ 具有某些理想的屬性就好了（我們稍後會詳細討論）。

- 它的期望值應該等於它試圖估計的參數，即$\mathrm{E}[T(\mathbf{X})] = \theta$，或者至少應該接近。
- 它應該具有低方差。

5.2.2 無偏估計

估計器的一個好特性是它“平均”是正確的。

定義：如果 $\mathrm{E}[T(\mathbf{X})] = \theta$，則 $T(\mathbf{X})$ 對 θ 是無偏的。

首先，我們將 $\bar{X}$ 的無偏性視為 μ 的估計量。

定理：假設 X_1、 ... 、X_n 是均值 μ 的任何事物。然後

$$\mathrm{E}[\bar{X}] = \frac{1}{n} \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left[X_i\right] = \mathrm{E}\left[X_1\right] = \mu$$

所以 $\bar{X}$ 對於 μ 總是無偏的。這就是為什麼 $\bar{X}$ 被稱為樣本均值的原因。

寶貝示例：特別是，假設 X_1、 ... 、X_n 是 iid $\operatorname{Exp}(\lambda)$。那麼對於 $\mu = \mathrm{E}\left[X_i\right] = 1/\lambda$，$\bar{X}$ 是無偏的。

備註：但要小心！事實證明（見習題 59，$1/\bar{X}$ 在這種指數情況下偏向於 λ，即 $\mathrm{E}[1/\bar{X}] \neq 1/\mathrm{E}[\bar{X}] = \lambda$。

現在我們將 S^2 的無偏性建立為 σ^2 的估計量。

定理：如果 X_1、 ... 、X_n 是均值 μ 和方差 σ^2 的任何事物，我們有

$$\mathrm{E}\left[S^2\right] = \frac{1}{n-1} \mathrm{E}\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right] = \operatorname{Var}(X_i) = \sigma^2$$

因此，在 iid 情況下，S^2 對於 σ^2 總是無偏的。這就是為什麼 S^2 被稱為樣本方差的原因（這也是我們除以 $(n-1)$ 而不是 n) 的原因。證明：首先，一些標準代數給出

$$\begin{aligned}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2&=\sum_{i=1}^{n}\left(X_i^2-2\bar{X}X_i+\bar{X}^2\right)\\&=\sum_{i=1}^{n}X_i^2-2\bar{X}\sum_{i=1}^{n}X_i+n\bar{X}^2\\&=\sum_{i=1}^{n}X_i^2-2n\bar{X}^2+n\bar{X}^2\\&=\sum_{i=1}^{n}X_i^2-n\bar{X}^2\end{aligned}$$

然後

$$\begin{aligned}\mathrm{E}\left[S^2\right]&=\frac{1}{n-1}\mathrm{E}\left[\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right]\quad\left(\text{by definition of } S^2 \text{ and Equation 5.1}]\right)\\&=\frac{1}{n-1}\left(\sum_{i=1}^{n}\mathrm{E}\left[X_i^2\right]-n\mathrm{E}\left[\bar{X}^2\right]\right)\\&=\frac{n}{n-1}\left(\mathrm{E}\left[X_1^2\right]-\mathrm{E}\left[\bar{X}^2\right]\right)\quad(\text{ since the } X_i \text{ 's are iid})\\&=\frac{n}{n-1}\left(\mathrm{Var}(X_1)+(\mathrm{E}\left[X_1\right])^2-\mathrm{Var}(\bar{X})-(\mathrm{E}[\bar{X}])^2\right)\\&=\frac{n}{n-1}\left(\sigma^2-\sigma^2/n\right)\quad\left(\text{since}\,\mathrm{E}\left[X_1\right]=\mathrm{E}[\bar{X}]\text{ and }\mathrm{Var}(\bar{X})=\sigma^2/n\right)\\&=\sigma^2\quad\text{Done.}\end{aligned}$$

小例子：假設 X_1 、 ... 、X_n 是獨立同分佈 $\mathrm{Exp}(\lambda)$。那麼 S^2 對於 $\mathrm{Var}(X_i)=1/\lambda^2$ 是無偏的

備註：小心！可以證明（參見練習 510，樣本標準偏差 S 對於標準偏差 σ 而言不是無偏的。

各個未知參數 μ 和 σ^2 的獨立無偏估計量 $\bar{X}$ 和 S^2 具有很好的直觀意義。然而，有時可能有多個合理的無偏估計量用於未知參數，在這種情況下，我們將不得不採取某些措施來確定哪個估計量在某種程度上是“最好的”。

大例子：假設 $X_1,X_2,\ldots,X_n\stackrel{\text{iid}}{\sim}\mathrm{Unif}(0,\theta)$，所以pdf 是 $f(x)=1/\theta,0<x<\theta$。動機是這樣的：我給你一堆介於 0 和 θ 之間的隨機數，你必須猜出 θ 是什麼。

我們將研究 θ 的三個無偏估計量——好的、更好的和醜陋的（稍後我們將解釋我們選擇的名稱）：

$$Y_1=2\bar{X},\quad Y_2=\frac{n+1}{n}\max_{1\le i\le n}X_i,\quad\text{and}\quad Y_3=\begin{cases}12\bar{X}&\text{w.p. }1/2\\-8\bar{X}&\text{w.p. }1/2.\end{cases}$$

我們將首先證明所有三個估計量確實是無偏的。首先考慮“好的”估計器，$Y_1=2\bar{X}$。

證明（Y_1 是無偏的）：$\mathrm{E}\left[Y_1\right]=2\mathrm{E}[\bar{X}]=2\mathrm{E}\left[X_i\right]=2(\theta/2)=\theta$

其次，讓我們研究“更好”的估計器，$Y_2=\frac{n+1}{n}M$，其中 $M\equiv\max_{1\le i\le n}X_i$。為什麼這個 θ 的估計量具有直觀意義？答案：因為它採取了迄今為止最大的觀測值，並給予它一點提升（因子 $(n+1)/n$）以試圖將它推近 θ。

證明（Y_2 是無偏的）：而不是證明 $\mathrm{E}\left[Y_2\right]=\frac{n+1}{n}\mathrm{E}[M]=\theta$，我們將等價地證明 $\mathrm{E}[M]=n\theta/(n+1)$。為此，讓我們從獲取 M 的 cdf 開始：

$$\begin{aligned}
\mathrm{P}(M \leq y) &= \mathrm{P}\left(X_1 \leq y \text{ and } X_2 \leq y \text{ and } \cdots \text{ and } X_n \leq y\right) \\
&= \mathrm{P}\left(X_1 \leq y\right) \mathrm{P}\left(X_2 \leq y\right) \cdots \mathrm{P}\left(X_n \leq y\right) \quad \left(X_i \text{ 's independent }\right) \\
&= \left[\mathrm{P}\left(X_1 \leq y\right)\right]^n \left(X_i \text{ 's identically distributed }\right) \\
&= \left[\int_0^y f_{X_1}(x) dx\right]^n \\
&= \left[\int_0^y (1/\theta) dx\right]^n \\
&= (y/\theta)^n .
\end{aligned}$$

這意味著 M 的 pdf 是

$$f_M(y) \equiv \frac{d}{dy}(y/\theta)^n = \frac{n y^{n-1}}{\theta^n}, \quad 0 < y < \theta$$

以便

$$\mathrm{E}[M] = \int_0^\theta y f_M(y) dy = \int_0^\theta \frac{n y^n}{\theta^n} dy = \frac{n\theta}{n+1}$$

哇！這最終表明 $Y_2 = \frac{n+1}{n} M$ 是 θ 的無偏估計量！

最後，讓我們看看 θ 的“醜陋”估計器，

$$Y_3 = \begin{cases} 12\bar{X} & \text{w.p. } 1/2 \\ -8\bar{X} & \text{w.p. } 1/2 \end{cases}$$

請注意，可能會得到 θ 的負估計，這很奇怪，因為 $\theta > 0$ ！儘管如此。.

證明（Y_3 是無偏的）：

$$\mathrm{E}\left[Y_3\right] = 12\mathrm{E}[\bar{X}] \cdot \frac{1}{2} - 8\mathrm{E}[\bar{X}] \cdot \frac{1}{2} = 2\mathrm{E}[\bar{X}] = \theta$$

如果多個估計器對 θ 都是無偏的，那麼哪個是最好的？當然，無偏估計器是好的，但“醜陋”的估計器 Y_3 表明無偏估計器有時會很愚蠢。我們可以通過查看估計器可能具有的其他屬性來打破無偏關係。例如，考慮估計量的方差。大例子（續）：再次假設 $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Unif}(0, \theta)$。回想一下“好的”$Y_1 = 2\bar{X}$、“更好的”$Y_2 = \frac{n+1}{n} M$ 和“醜陋的”Y_3 對於 θ 都是無偏的

讓我們找出它們的差異。首先，

$$\operatorname{Var}(Y_1) = 4 \operatorname{Var}(\bar{X}) = \frac{4}{n} \cdot \operatorname{Var}(X_i) = \frac{4}{n} \cdot \frac{\theta^2}{12} = \frac{\theta^2}{3n},$$

這不是太破舊，因為 $n \rightarrow \infty$ 的方差變為 0。

同時，

$$\begin{aligned}
\operatorname{Var}(Y_2) &= \mathrm{E}\left[Y_2^2\right] - \left(\mathrm{E}\left[Y_2\right]\right)^2 \\
&= \left(\frac{n+1}{n}\right)^2 \mathrm{E}\left[M^2\right] - \theta^2 \quad (\text{ since } Y_2 \text{ is unbiased}) \\
&= \left(\frac{n+1}{n}\right)^2 \int_0^\theta \frac{ny^{n+1}}{\theta^n} dy - \theta^2 \\
&= \theta^2 \cdot \frac{(n+1)^2}{n(n+2)} - \theta^2 = \frac{\theta^2}{n(n+2)}
\end{aligned}$$

它很快就變成了 0 - 比 Y_1 的速率要好！

最後，

$$\begin{aligned}
\operatorname{Var}(Y_3) &= \mathrm{E}\left[Y_3^2\right] - \left(\mathrm{E}\left[Y_3\right]\right)^2 \\
&= \frac{1}{2}\left(144\mathrm{E}\left[\bar{X}^2\right] + 64\mathrm{E}\left[\bar{X}^2\right]\right) - \theta^2 \quad (\text{since } Y_3 \text{ is unbiased}) \\
&= 104\mathrm{E}\left[\bar{X}^2\right] - \theta^2 \\
&= 104\left[\operatorname{Var}(\bar{X}) + (\mathrm{E}[\bar{X}])^2\right] - \theta^2 \\
&= 104\left[\frac{\theta^2}{12n} + \frac{\theta^2}{4}\right] - \theta^2 \quad (\text{by previous work}) \\
&= \left(\frac{26}{3n} + 25\right)\theta^2
\end{aligned}$$

而且這個爛攤子甚至不會像 $n \to \infty$ 那樣變成 0 ！

因此，Y_1、Y_2 和 Y_3 都是無偏的，但 Y_2 的方差遠低於 Y_1，而 $Y_{3} $具有瘋狂的高方差。現在你可以明白為什麼它們被稱為“好”、“更好”和“醜”了。在任何情況下，我們都可以通過選擇方差最小的 Y_2 來打破“無偏關係”。

5.2.3 均方誤差

我們在 $5.2.2 中看到，公正是一件好事，但它並不能說明全部情況；差異也必須進入對話。在評估估計器性能時，均方誤差 (MSE) 會同時考慮偏差和方差。

定義：θ 的估計量 $T(\mathbf{X})$ 的均方誤差為

$$\operatorname{MSE}(T(\mathbf{X})) \equiv \mathrm{E}\left[(T(\mathbf{X}) - \theta)^2\right]$$

在給出 MSE 的更簡單解釋之前，我們定義參數 θ 的估計器的偏差

$$\operatorname{Bias}(T(\mathbf{X})) \equiv \mathrm{E}[T(\mathbf{X})] - \theta$$

定理：更簡單的解釋：$\mathrm{MSE} = \mathrm{Bias}^2 + \mathrm{Var}$。

證明：根據定義，我們有

$$\begin{aligned}\operatorname{MSE}(T(\mathbf{X})) &= \mathrm{E}\left[(T(\mathbf{X})-\theta)^2\right] \\ &= \mathrm{E}\left[T^2\right]-2\theta\mathrm{E}[T]+\theta^2 \\ &= \mathrm{E}\left[T^2\right]-(\mathrm{E}[T])^2+(\mathrm{E}[T])^2-2\theta\mathrm{E}[T]+\theta^2 \\ &= \operatorname{Var}(T)+\underbrace{(\mathrm{E}[T]-\theta)}_{\text{Bias}}{}^2 \cdot \square\end{aligned}$$

因此，MSE 結合了估計量的偏差和方差。MSE 越低越好。如果 $T_1(\mathbf{X})$ 和 $T_2(\mathbf{X})$ 是 θ 的兩個估計量，我們通常更喜歡 MSE 較低的那個——甚至如果它恰好有更高的偏差。

定義：$T_2(\mathbf{X})$ 與 $T_1(\mathbf{X})$ 的相對效率為 $\operatorname{MSE}(T_1)/\operatorname{MSE}(T_2)$。如果這個數量是 <1，那麼我們更喜歡 T_1。

示例：假設估計器 A 有偏差 $=3$ 和方差 $=10$，而估計器 B 有偏差 $=-2$ 和方差 $=14$。哪個估計器 (A 或 B) 的均方誤差較小？

解：$\mathrm{MSE}=\mathrm{Bias}^2+\mathrm{Var}$，所以

$$\operatorname{MSE}(\mathrm{A})=9+10=19 \quad \text{and} \quad \operatorname{MSE}(\mathrm{B})=4+14=18$$

因此，B 的 MSE 較低。

示例：$X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \operatorname{Unif}(0, \theta)$。回想一下，我們檢查了 θ 的兩個相當不錯的無偏估計量，即 $Y_1=2\bar{X}$ 和 $Y_2=\frac{n+1}{n}\max_i X_i$。我們發現$\operatorname{Var}(Y_1)=\theta^2/(3n)$，而$\operatorname{Var}(Y_2)=\theta^2/\left(n^2+2n\right)$，所以

$$\frac{\operatorname{MSE}(Y_1)}{\operatorname{MSE}(Y_2)}=\frac{\theta^2/(3n)}{\theta^2/\left(n^2+2n\right)}=\frac{n+2}{3}>1, \quad \text{for } n>1$$

這表明 Y_2 是更好的估計量。

5.2.4 最大似然估計

最大似然估計器（MLE）作為無偏估計的替代方法（儘管 MLE 有時本身是無偏的）。MLE 具有許多理想的屬性，並且是第 7 章中將討論的擬合優度檢驗的重要組成部分

5.2.4.1 $5.2.4.1 MLE 和嬰兒示例介紹

定義：考慮一個 iid 隨機樣本 $X_1, X_2, \ldots, X_n$，其中每個 X_i 有 pmf/pdf$f(x)$。此外，假設 θ 是來自 X_i 的某個未知參數。似然函數是 $L(\theta) \equiv \prod_{i=1}^n f\left(x_i\right)$。

定義：θ 的最大似然估計是使$L(\theta)$ 最大化的θ 的值。MLE 是 X_i 的函數並且是一個隨機變量。

備註：我們可以非常非正式地將 MLE 視為 θ 的"最可能"估計。

示例：假設 $X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \operatorname{Exp}(\lambda)$。讓我們找到 λ 的 MLE。我們從似然函數開始，

$$L(\lambda)=\prod_{i=1}^n f\left(x_i\right)=\prod_{i=1}^n \lambda e^{-\lambda x i}=\lambda^n \exp \left(-\lambda \sum_{i=1}^n x_i\right).$$

現在的任務是相對於 λ 最大化 $L(\lambda)$。我們可以採用導數並以通常的方式遍歷所有可怕的代數，但這通常太乏味了。一個經常應用的有用技巧是簡單地對似然函數的自然對數執行最大化。由於自然的

$\log$ 函數是一對一的，很容易看出使$L(\lambda)$ 最大化的λ 也使$\ln(L(\lambda))$ 最大化。為此，我們有

$$\ln(L(\lambda)) = \ln\left(\lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right)\right) = n\ln(\lambda) - \lambda \sum_{i=1}^{n} x_i,$$

以便

$$\frac{d}{d\lambda}\ln(L(\lambda)) = \frac{d}{d\lambda}\left(n\ln(\lambda) - \lambda \sum_{i=1}^{n} x_i\right) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \equiv 0.$$

求解臨界點（並進行此處未說明的二階導數測試），我們發現 MLE 為 $\hat{\lambda} = 1/\bar{X}$。

5.2.4.2 備註：

- MLE $\hat{\lambda} = 1/\bar{X}$ 有意義，因為 $\mathrm{E}[X_1] = 1/\lambda$ 對於指數分佈。
- 也許令人驚訝，即使我們在 $5.2.2 中表明 $\bar{X} = 1/\hat{\lambda}$ 對於 $\mathrm{E}[X_1] = 1/\lambda$ 是無偏的 ，事實證明 MLE $\hat{\lambda}$ 對 λ 略有偏差（參見練習 599。這實際上是無偏估計的一個小缺點。
- 在我們的工作結束時，我們在 λ 上戴上一個小帽子，以表明這是 MLE。這就像一頂派對帽！- 最後，我們將所有的小 x_i 變成大 X_i 以表明這是一個隨機變量。我們的 MLE 都長大了！
- 您應該始終執行二階導數測試，但如果您不這樣做，也許我們不會責怪您。

MLE 也適用於離散分佈。

示例：假設 $X_1, X_2 \ldots, X_n \overset{\text{iid}}{\sim} \text{Bern}(p)$。我們將找到 p 的 MLE。由於 $\text{Bern}(p) \sim \text{Bin}(1, p)$，我們可以將 pmf 寫為

$$f(x) = p^x(1-p)^{1-x}, \quad x = 0, 1.$$

因此，似然函數是

$$L(p) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

以便

$$\ln(L(p)) = \sum_{i=1}^{n} x_i \ln(p) + \left(n - \sum_{i=1}^{n} x_i\right)\ln(1-p)$$

我們將關於 p 的導數設置為零，

$$\frac{d}{dp}\ln(L(p)) = \frac{\sum_i x_i}{p} - \frac{n - \sum_i x_i}{1-p} \equiv 0,$$

最後求解 MLE，$\hat{p} = \bar{X}$。這在直覺上很好，因為 $\mathrm{E}[X] = p$。

例如，假設我們觀察到以下 $30\,\text{Bern}(p)$ 試驗結果（1 是成功，0 是失敗）：

$$\begin{array}{ccccccccccccccc} 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 \end{array}$$

那麼$\hat{p}=\bar{X}=21/30=0.7$。

5.2.4.3 $5.2.4.2 青少年 MLE 示例

MLE 也可以用來解決更實質性的問題。首先，我們提供一個正態分佈示例，在該示例中，我們一次找到兩個參數的 MLE。

示例：假設 $X_1, X_2 \ldots, X_n \stackrel{\text{iid}}{\sim} \operatorname{Nor}\left(\mu, \sigma^2\right)$。我們將找到 μ 和 σ^2 的同時 MLE。似然函數是

$$\begin{aligned} L\left(\mu, \sigma^2\right) & =\prod_{i=1}^n f\left(x_i\right) \\ & =\prod_{i=1}^n \frac{1}{\sqrt{2 \pi \sigma^2}} \exp \left\{-\frac{1}{2} \frac{\left(x_i-\mu\right)^2}{\sigma^2}\right\} \\ & =\frac{1}{\left(2 \pi \sigma^2\right)^{n / 2}} \exp \left\{-\frac{1}{2 \sigma^2} \sum_{i=1}^n\left(x_i-\mu\right)^2\right\} . \end{aligned}$$

這意味著

$$\ell n\left(L\left(\mu, \sigma^2\right)\right)=-\frac{n}{2} \ell \mathrm{n}(2 \pi)-\frac{n}{2} \ell n\left(\sigma^2\right)-\frac{1}{2 \sigma^2} \sum_{i=1}^n\left(x_i-\mu\right)^2$$

因此，對 μ 進行偏導，我們得到

$$\frac{\partial}{\partial \mu} \ell \mathrm{n}\left(L\left(\mu, \sigma^2\right)\right)=\frac{1}{\sigma^2} \sum_{i=1}^n\left(x_i-\mu\right) \equiv 0$$

所以 MLE 是 $\hat{\mu}=\bar{X}$（這又是有道理的）。

現在對 σ^2 做同樣的事情，通過對 σ^2（不是 σ）取部分，

$$\frac{\partial}{\partial \sigma^2} \ell \mathrm{n}\left(L\left(\mu, \sigma^2\right)\right)=-\frac{n}{2 \sigma^2}+\frac{1}{2 \sigma^4} \sum_{i=1}^n\left(x_i-\hat{\mu}\right)^2 \equiv 0$$

（我們潛入 $\hat{\mu}$ for μ），最終得到

$$\widehat{\sigma^2}=\frac{1}{n} \sum_{i=1}^n\left(X_i-\bar{X}\right)^2$$

備註：請注意 $\widehat{\sigma^2}$ 與（無偏）樣本方差的相似程度，

$$S^2=\frac{1}{n-1} \sum_{i=1}^n\left(X_i-\bar{X}\right)^2=\frac{n}{n-1} \widehat{\sigma^2}$$

估計器 $\widehat{\sigma^2}$ 有一點偏差，但它的方差略小於 S^2。無論如何，隨著 n 變大，S^2 和 $\widehat{\sigma^2}$ 變得相同。

正態分佈的兩個參數 μ 和 σ^2 的 MLE 或多或少地解耦，使得各自的計算很容易。伽馬分佈的情況並非如此，它也有兩個參數。

示例：帶有參數 r 和 λ 的 gamma 分佈的 pdf 為

$$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, \quad x > 0$$

假設 $X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Gam}(r, \lambda)$。我們將找到 r 和 λ 的 MLE。首先，似然函數是

$$L(r, \lambda) = \prod_{i=1}^n f(x_i) = \frac{\lambda^{nr}}{[\Gamma(r)]^n} \left(\prod_{i=1}^n x_i \right)^{r-1} e^{-\lambda \sum_{i=1}^n x_i},$$

以便

$$\ln(L) = nr\ln(\lambda) - n\ln(\Gamma(r)) + (r-1)\ln\left(\prod_{i=1}^n x_i \right) - \lambda \sum_{i=1}^n x_i$$

取關於 λ 的偏導數並將其設置為 0 以獲得

$$\frac{\partial}{\partial\lambda} \ln(L) = \frac{nr}{\lambda} - \sum_{i=1}^n x_i \equiv 0$$

所以 λ 的 MLE 是 $\hat{\lambda} = \hat{r}/\bar{X}$。

River City (A pre-Hamilton rap) 的麻煩在於我們需要找到$\hat{r}$，這並不容易_但我們會盡力而為。與上述工作類似，我們得到

$$\frac{\partial}{\partial r} \ln(L) = n\ln(\lambda) - \frac{n}{\Gamma(r)} \frac{d}{dr} \Gamma(r) + \ell\,\mathrm{n}\left(\prod_{i=1}^n x_i \right) \equiv 0$$

其中 $\psi(r) \equiv \Gamma'(r)/\Gamma(r)$ 被稱為 digamma 函數。

此時，我們將$\hat{\lambda} = \hat{r}/\bar{X}$代入方程5.2，並使用計算機算法（如二分法、牛頓法等）搜索解決的 r 的值

$$n\ln(r/\bar{x}) - n\psi(r) + \ell n\left(\prod_{i=1}^n x_i \right) \equiv 0.$$

伽馬函數在任何合理的軟件包中都很容易獲得；但如果 digamma 函數恰好在您所在的城鎮不可用，您可以利用近似值

$$\Gamma'(r) \doteq \frac{\Gamma(r+h) - \Gamma(r)}{h} \quad \text{(for any small } h \text{ of your choosing).}$$

例如，選擇 $h = 0.01$，我們發現

$$\Gamma'(1.5) \doteq \frac{\Gamma(1.51) - \Gamma(1.5)}{0.01} = \frac{0.8865917 - 0.8862269}{0.01} = 0.03648$$

這與四捨五入為 0.03649 $的實際值（通過 Matlab 獲得）很好地比較。

請參閱 $7.5.3，我們在其中執行了一個涉及 gamma 分佈的擴展數值示例，包括搜索組件。

最後，一個有趣的例子沒有直接應用微積分。

示例：假設 $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Unif}(0, \theta)$。我們將找到 θ 的 MLE。首先，回想一下 pdf 是 $f(x) = 1/\theta, 0 < x < \theta$，你需要注意有趣的限制。似然函數是

$$L(\theta) = \prod_{i=1}^{n} f(x_i) = \begin{cases} 1/\theta^n & \text{if } 0 \leq x_i \leq \theta, \forall i \\ 0 & \text{otherwise.} \end{cases}$$

為了有 $L(\theta) > 0$，我們必須有 $0 \leq x_i \leq \theta$，對於所有 i。換句話說，我們必須有 $\theta \geq \max_i x_i$。受此約束，$L(\theta) = 1/\theta^n$ 在盡可能小的 θ 值處最大化，即 $\hat{\theta} = \max_i X_i$。鑑於我們之前在 $$ 中看到的類似（無偏）估計量 $Y_2 = \frac{n+1}{n} \max_i X_i$，這是有道理的5.2 .2$

備註：我們在這個例子中使用了很少的微積分！

5.2.4.4 $5.2.4.3 MLE 的不變性

到目前為止，我們在 MLE 上的所有工作都導致了不變性，這使我們能夠以極其一般的方式使用 MLE，尤其是我們將在第 7 章討論的擬合優度檢驗

定理（MLE 的不變性）：如果 $\hat{\theta}$ 是某個參數 θ 的 MLE，並且 $h(\cdot)$ 是任何合理的函數，那麼 $h(\hat{\theta})$ 是 $h(\theta)$ 的 MLE

備註：我們之前註意到，這樣的性質不具有無偏性。例如，雖然 $\mathrm{E}\left[S^2\right] = \sigma^2$，但通常情況下 $\mathrm{E}\left[\sqrt{S^2}\right] \neq \sigma$。

備註：當 $h(\cdot)$ 是一對一函數時，不變性的證明是“容易的”。當 $h(\cdot)$ 更糟糕時，這並不容易 - 但通常仍然如此。無論如何，我們不會在這裡進行證明。

示例：假設 $X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Nor}(\mu, \sigma^2)$

回想一下 σ^2 的 MLE 是 $\widehat{\sigma^2} = \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 / n$。如果我們考慮函數 $h(y) = +\sqrt{y}$，那麼不變性屬性說 σ 的 MLE 是

$$\hat{\sigma} = \sqrt{\widehat{\sigma^2}} = \left[\frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2\right]^{1/2}$$

示例：假設 $X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Bern}(p)$。

我們看到 p 的 MLE 是 $\hat{p} = \bar{X}$。然後不變性說 $\hat{p}(1-\hat{p}) = \bar{X}(1-\bar{X})$ 是 $\text{Var}(X_i) = p(1-p)$。

最後，這是一個在生存分析和精算科學中具有巨大應用的例子，我們對物品（或人類）是否會在一定時間內存活感興趣。

示例：假設 $X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Exp}(\lambda)$

回想一下 λ 的 MLE 是 $\hat{\lambda} = 1/\bar{X}$。同時，我們將生存函數定義為

$$\bar{F}(x) \equiv \mathrm{P}(X > x) = 1 - F(x) = e^{-\lambda x}$$

然後不變性說 $\bar{F}(x)$ 的 MLE 是

$$\widehat{\bar{F}(x)} = e^{-\hat{\lambda}x} = e^{-x/\bar{X}},$$

$\bar{F}(x)$ 上的那個東西更像是一個屋頂而不是一頂帽子！

考慮一下我從 $\operatorname{Exp}(\lambda)$ 分佈生成的以下 30 個電池壽命（以年為單位）。

4.082	1.575	4.031	2.008	4.491	4.003	3.041	0.898	0.401	2.005
0.084	0.396	3.761	2.407	3.776	5.296	3.672	0.620	2.528	5.255
3.930	2.750	3.931	1.502	1.741	0.218	1.727	2.923	0.714	0.290

一個簡單的計算表明樣本均值 $\bar{X} = 2.469$。因此，λ 的 MLE 是 $\hat{\lambda} = 1/\bar{X} = 0.405$。因此，特定部件將持續超過 3 年的概率的 MLE 為

$$\widehat{\bar{F}(x)} = e^{-x/\bar{X}} = e^{-3(0.405)} = 0.297$$

5.2.5 矩量法

另一類豐富的估計量來自矩量法（MoM）。我們已經看到，MLE 有時需要做大量工作才能提出（在我們達到 \$7.5.3 之前，甚至不要詢問 Weibull 分佈）。MoM 是 MLE 的替代方案，通常更易於應用。

首先，回想一下隨機變量 X 的 k^{th} 矩是

$$\mu_k \equiv \mathrm{E}\left[X^k\right] = \begin{cases} \sum_x x^k f(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} x^k f(x)dx & \text{if } X \text{ is continuous.} \end{cases}$$

定義：假設 $X_1, X_2, \ldots, X_n$ 是獨立同分佈的隨機變量。那麼 $\mu_k = \mathrm{E}\left[X^k\right]$ 的矩估計方法是樣本矩 $m_k \equiv \sum_{i=1}^n X_i^k/n$，對於 $k = 1, 2$， $\ldots$

備註：作為$n \to \infty$，大數定律意味著$\sum_{i=1}^n X_i^k/n \to \mathrm{E}\left[X^k\right]$，即 $m_k \to \mu_k$（所以這是一個很好的估計量）。

備註：你應該永遠愛你的媽媽！

5.2.5.1 示例：

- 真實均值 $\mu_1 = \mu = \mathrm{E}\left[X_i\right]$ 的 MoM 估計量是樣本均值 $m_1 = \bar{X} = \ sum_{i=1}^n X_i/n$
- $\mu_2 = \mathrm{E}\left[X_i^2\right]$ 的 MoM 估計量是 $m_2 = \sum_{i=1}^n X_i^2/n$。
- $\operatorname{Var}(X_i) = \mathrm{E}\left[X_i^2\right] - \left(\mathrm{E}的MoM估計量\left[X_i\right]\right)^2 = \mu_2 - \mu_1^2$ 是

$$m_2 - m_1^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{n-1}{n}S^2.$$

（對於較大的 n，也可以使用 S^2.）

一般博弈計劃：用真實矩 $\mu_k = \mathrm{E}\left[X^k\right]$ 表示感興趣的參數。然後代入樣本矩 $m_k = \sum_{i=1}^{n} X_i^k/n$。示例：假設 $X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \operatorname{Pois}(\lambda)$。回想一下 $\lambda = \mathrm{E}\left[X_i\right]$，所以 λ 的簡單 MoM 估計量是 $\bar{X}$。

但也要注意 $\lambda = \operatorname{Var}(X_i)$，所以 λ 的另一個 MoM 估計量是 $\frac{n-1}{n}S^2$（或普通的舊 S^2)

備註：如果可以選擇，請使用看起來更容易的 MoM 估計器。

示例：假設 $X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \operatorname{Nor}(\mu, \sigma^2)$。$\mu$ 和 σ^2 的 MoM 估計量是 $\bar{X}$ 和 $\frac{n-1}{n}S^2$（或 $S^{2} $ ），分別。所以對於這個例子，這些估計器與 MLE 的相同。

讓我們以一個不那麼瑣碎的例子結束。

示例：假設 $X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \operatorname{Beta}(a, b)$. pdf 是

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1$$

事實證明（經過大量代數）

$$\mathrm{E}[X] = \frac{a}{a+b} \quad \text{and} \quad \operatorname{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}.$$

讓我們通過 MoM 估計 a 和 b。為此，請注意

$$\mathrm{E}[X] = \frac{a}{a+b},$$

這意味著

$$a = \frac{b\mathrm{E}[X]}{1-\mathrm{E}[X]} \doteq \frac{b\bar{X}}{1-\bar{X}}$$

和

$$\operatorname{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)} = \frac{\mathrm{E}[X]b}{(a+b)(a+b+1)}.$$

獲取 b 的 MoM，將以下代入方程 5.4)：[X] 為 {X}，(X)$ 為 S^{2}$，和 $b\bar{X}/(1-\bar{X})$ 對於 a。這是我們在代數的一大堆之後得到的：

$$b \doteq \frac{(1-\bar{X})^2\bar{X}}{S^2} - 1 + \bar{X}$$

最後，插入方程 (5.3 以獲得 a 的 MoM 估計量。

示例：考慮以下數據集，其中包含我們從 beta 分佈中獲得的 $n = 10$ 觀察值。

$$0.86 \quad 0.77 \quad 0.84 \quad 0.38 \quad 0.83 \quad 0.54 \quad 0.77 \quad 0.94 \quad 0.37 \quad 0.40$$

我們立即有 $\bar{X} = 0.67$ 和 $S^2 = 0.04971$。MoM 估計量是

$$b \doteq \frac{(1-\bar{X})^2\bar{X}}{S^2} - 1 + \bar{X} = 1.1377$$

接著

$$a \doteq \frac{b\bar{X}}{1-\bar{X}} = 2.310$$

5.3 抽樣分佈

目標：我們將討論稍後在學習置信區間（第 6 章和假設檢驗（第 7 章）時將需要的一些分佈。特別是，我們將簡要介紹正態分佈（當然，，我們已經知道很多），以及一些新朋友 - χ^2、t 和 F 分佈。

定義：回想一下，統計只是來自隨機樣本的觀測值 X_1、X_2、 . . . 、X_n 的函數。該函數不明確依賴任何未知參數。

示例：$\bar{X}$、S^2 和 $\max\{X_1$、X_2、 . . . 、$X_n\}$ 是統計數據，但$(\bar{X}-\mu)/\sigma$ 不是。

由於統計數據是隨機變量，因此找出它們的分佈很有用。統計量的分佈稱為抽樣分佈。我們將研究幾個抽樣分佈：分別在 §$5.3.155.3.4 中的正態分佈、$\chi^2$、Student t 和 F。

5.3.1 正態分佈

主要外賣： 我們已經多次看到 $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Nor}(\mu, \sigma^2)$ 意味著樣本均值 $\bar{X} \sim \text{North}(\mu, \sigma^2/n)$.

此結果通常用於獲取 μ 的置信區間並進行假設檢驗。敬請關注！

我們現在將介紹一些其他重要的抽樣分佈……

5.4 2 χ^2 分佈

定義/定理：如果 $Z_1, \ldots, Z_k \overset{\text{iid}}{\sim} \text{Nor}(0,1)$，則 $Y \equiv \sum_{i=1}^{k} Z_i^2$ 具有 k 自由度 (df) 的卡方分佈，我們寫成 $Y \sim \chi^2(k)$.

備註：術語“df”非正式地對應於您擁有的“獨立信息片段”的數量。例如，假設 $\sum_{i=1}^{n} X_i = c$，其中 c 是您實際觀察到 X_i 後的已知常數。那麼你可能有 $n-1\text{df}$，因為知道 X_i 的任何 $n-1$ 會給你剩下的 X_i。

每次我們必須估計一個參數時，我們也會非正式地“失去”一個自由度。例如，如果我們可以訪問 n 觀察值，但必須估計兩個參數 μ 和 σ^2（例如，對於正態分佈），那麼我們可能只會得到 $n-2\text{df}$。

實際上，df 對應於某個數學空間的維數（本課程未涉及）！有趣的 $\chi^2(k)$ 事實：pdf 是

$$f_Y(y) = \frac{1}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} y^{\frac{k}{2}-1} e^{-y/2}, \quad y > 0.$$

我們還可以證明 $\text{E}[Y] = k$（參見習題 525 和 $\text{Var}(Y) = 2k$。

您可以輕鬆地比較 pdf 以了解指數分佈是一種特殊情況，即 $\chi^2(2) \sim \text{Exp}(1/2)$。更一般地說，如果 df 是偶數，我們會得到一點驚喜，即 $\chi^2(2k) \sim \text{Erlang}_k(1/2)$。最後，對於較大的 k，$\chi^2(k)$ 分佈近似

正態分佈（根據 CLT）。

定義：具有 cdf $F(x)$ 的（連續）隨機變量 X 的 $(1-\alpha)$ 分位數是 x_α 使得
$F\left(x_{alpha}\right)=P\left(X \leq x_\alpha\right)=1-\alpha$。注意 $x_\alpha=F^{-1}(1-\alpha)$，其中 $F^{-1}(\cdot)$ 是 X. 的逆 cdf

符號：如果 $Y \sim \chi^2(k)$，那麼我們用特殊符號 $\chi_{\alpha,k}^2$（表示 $(1-\alpha)$ 分位數 而不是 x_α）。也就是說，所謂的"尾"概率$\mathrm{P}\left(Y>\chi_{\alpha,k}^2\right)=\alpha$。您可以向上查看 $\chi_{\alpha,k}^2$，例如，在書後的表 B.3 中或通過 Excel 函數 chisq.inv $(1-\alpha, k)$.

示例：如果 $Y \sim \chi^2(10)$，則 $\mathrm{P}\left(Y>\chi_{0.05,10}^2\right)=0.05$，其中我們可以查$\chi_{0.05,10}^2=18.31$

定理：χ^2 相加。如果 $Y_1, Y_2, \ldots, Y_n$ 獨立於 $Y_i \sim \chi^2\left(d_i\right)$，對於所有$i$，然後 $\sum_{i=1}^n Y_i \sim \chi^2\left(\sum_{i=1}^n d_i\right)$。

證明：只需使用標準的 mgf 參數，雖然這裡沒有細節。

那麼 χ^2 分佈在統計中的什麼位置呢？當我們試圖估計 σ^2 時，通常會出現這種情況。見第 6 章和第 7 章

示例：如果 $X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \operatorname{Nor}\left(\mu, \sigma^2\right)$，那麼我們將在第 6 章中展示

$$S^2=\frac{1}{n-1} \sum_{i=1}^n\left(X_i-\bar{X}\right)^2 \sim \frac{\sigma^2 \chi^2(n-1)}{n-1}$$

5.4.1 學生 t 分佈

定義/定理：假設$Z \sim \operatorname{Nor}(0,1)$ 和$Y \sim \chi^2(k)$，並且Z 和Y 是獨立的。那麼 $T \equiv Z / \sqrt{Y / k}$ 具有自由度為 k 的 Student t 分佈，我們寫成 $T \sim t(k)$。

有趣的事實：pdf是

$$f_T(x)=\frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k} \Gamma\left(\frac{k}{2}\right)}\left(\frac{x^2}{k}+1\right)^{-(k+1)/2}, \quad x \in \mathbb{R}$$

$t(k)$ 看起來像 $\operatorname{Nor}(0,1)$，除了 t 有更粗的尾巴。$k=1$ 的情況給出了 Cauchy 分佈，它的尾巴非常肥大（正如我們之前所見）。隨著 df k 變大，$t(k) \rightarrow \operatorname{Nor}(0,1)$。

很容易證明 $\mathrm{E}[T]=0$ for $k>1$, $\operatorname{Var}(T)=\frac{k}{k-2}$ for $k >2 $。

符號：如果$T \sim t(k)$，那麼我們用$t_{\alpha,k}$ 來表示$(1-\alpha)$ 分位數。換句話說，$\mathrm{P}\left(T>t_{\alpha,k}\right)=\alpha$。見附錄表 B.2。

示例：如果 $T \sim t(10)$，則 $\mathrm{P}\left(T>t_{0.05,10}\right)=0.05$，其中 $t_{0.05,10}=1.812$ 在書的背面或通過 Excel 函數 t。 $\operatorname{inv}(1-\alpha, k)$。

那麼我們在統計中使用 t 分佈做什麼呢？當我們找到置信區間並對均值 μ 進行假設檢驗時，它就會出現，尤其是在方差未知的情況下。同樣，請繼續關注第 6 章和第 7 章。

名字背後的故事：順便問一下，為什麼我們最初稱它為 Student t 分佈？“Student”是William Gossett [2] 的筆名，他首先導出了這個分佈。Gossett 是愛爾蘭吉尼斯啤酒廠的一名統計學家，他不得不匿名進行研究，這樣吉尼斯就不會冒著商業機密的風險。

5.5 4 F 分佈

定義/定理：假設$X \sim \chi^2(n)$ 和$Y \sim \chi^2(m)$，並且X 和Y 是獨立的。那麼 $F \equiv \frac{X/n}{Y/m} = mX/(nY)$ 有 F 分佈，有 n 和 m df（是的，你必須指定兩個df 值）。它用 $F \sim F(n,m)$ 表示。

有趣的事實：令人討厭的 pdf 是

$$f_F(x) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \frac{\left(\frac{n}{m}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1}}{\left(\frac{n}{m}x+1\right)^{\frac{n+m}{2}}}, \quad x>0$$

$F(n,m)$ 通常有點偏右。

可以證明，對於 $m>2$，$\mathrm{E}[F] = m/(m-2)$，$\mathrm{Var}(F) = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$ 為 $m>4$。

t 分佈是一個特例——你能找出是哪一個嗎？

符號：如果 $F \sim F(n,m)$，那麼我們用 $F_{\alpha,n,m}$ 來表示 $(1-\alpha)$ 分位數。即$\mathrm{P}\left(F > F_{\alpha,n,m}\right) = \alpha$。表 B.4 B.7 可以在本書後面找到各種 α、n、m，或者您可以使用 Excel 函數 f。 $\mathrm{inv}(1-\alpha, n, m)$。

示例：如果 $F \sim F(5,10)$，則 $\mathrm{P}\left(F > F_{0.05,5,10}\right) = 0.05$，其中我們找到 $F_{0.05， 5,10}=3.326.

備註：F 分佈有一個有趣的性質 $F_{1-\alpha,m,n} = 1/F_{\alpha,n,m}$。如果您必須找到類似 $F_{0.95,10,5} = 1/F_{0.05,5,10} = 1/3.326$ 的東西，請使用這個事實，我們注意到 df 是交換的。此屬性減少了在表格中提供低於 0.5 的分位數的需要。

[2] 不要與奧斯卡獲獎演員 Louis Gossett, Jr. 混淆。那麼我們在統計中使用 F 分佈是什麼？當我們找到置信區間並對來自兩個不同總體的方差比率進行假設檢驗時使用它。例如，兩個總體的方差是否大致相同，或者它們是否“顯著”不同？稍後將在第 6 章和第 7 章中詳細介紹

5.6 5.4 練習

1. ($5.1.1) 我今天正在進行一項民意調查，詢問受訪者他們最喜歡的政治候選人是誰——史密斯、瓊斯或託馬斯。我正在收集什麼類型的數據？

a. 連續
b. 離散的
c. 分類的
d. 時間序列
e. 序數

2. ($5.1.2) 假設我們收集以下觀察結果：$7, -2, 1, 6$。什麼是樣本均值和样本方差？

3.（§5.1.2）這是我在佐治亞理工學院概率與統計課上學生的智商。我的學生真棒！

146	149	154	161
141	152	156	165
136	148	150	182
145	146	154	149
154	147	140	147
140	142	149	153
145	131	167	149
170	168	146	160
150	140	137	151
169	137	163	153

構建直方圖並根據該數據計算樣本均值和样本方差。

4. ($5.1.2) 這是我最近一次測試的分數的莖葉圖。求樣本均值、樣本中位數和样本標準差。

$$\begin{array}{r|l} 4 & 4 \\ 5 & 8\ 3 \\ 6 & 9\ 6\ 6\ 2 \\ 7 & 7\ 7\ 5\ 4\ 3\ 1\ 1\ 0 \\ 8 & 9\ 7\ 7\ 6\ 5\ 3\ 2\ 1\ 1 \\ 9 & 6\ 4\ 3\ 1 \\ 10 & 0\ 0 \end{array}$$

5. ($5.1.2) 讓我們通過 rand() 函數在 Excel 中生成 1000 個 Unif $(0, 1)$ 隨機變量（或者使用任何你喜歡的編程語言）。找出您的觀測值的樣本均值和方差。

6. ($5.1.2) c 的什麼值使加權平方和最小化 $\sum_{i=1}^{n} w_i\,(X_i - c)^2$, 其中 $w_1, \ldots, w_n$ 構成了一組給定的權重？

7. ($5.1.3) 我們對建模一個簡單的單服務器排隊系統感興趣，例如，一個單人理髮店。顧客隨機到達理髮店，但總是一次一個。此外，全天的總體到達率幾乎相同，並且已知不相交時間段的到達數量是獨立的。如果在他到達時，客戶看到至少有一個其他人在排隊（“理髮師隊列”(v)，那麼他有 25 $ %$ 的機會生氣並離開。否則，他加入這條線，以先進先出的方式處理。最終，輪到客戶由理髮師服務。現在，這是一個偉大的理髮師，他做了一百萬件小事，構成了整體服務時間，每一個都需要iid時間 - 洗頭，修剪，頭部按摩，談論運動，mani，pedi, pay up, 等等等等。在所有這些事情之後，客戶離開了。您如何對該系統的各種組件進行建模？

a. 到達之間的泊鬆時間，關於是否進入線路的伯努利決策過程，正常的整體服務時間

b. 到達之間的指數時間，關於是否進入線路的伯努利決策過程，正常的整體服務時間

c. 到站泊鬆時間、是否入線的伯努利決策過程、統一的整體服務時間

d. 到達之間的指數時間，關於是否進入線路的幾何決策過程，指數整體服務時間

8. ($5.2.1) 假設 $X_1, X_2, \ldots, X_n$ 是獨立同分佈的，均值未知 μ 和未知 σ^2。以下哪項可以被認為是統計數據？

a. 樣本均值，$\bar{X}$

b. 樣本標準差，S

c. 標準化樣本均值，$(\bar{X}-\mu)/\sqrt{\sigma^2/n}$

d. Both(a) 和 (b)

e. 以上都不是 9. (§5.2.2) 假設 $X_1, X_2, \ldots, X_n$ 是 iid $\operatorname{Exp}(\lambda)$。我們知道樣本均值 $\bar{X}$ 對於均值 $1/\lambda$ 是無偏的。但證明 $1/\bar{X}$ 對 λ 有一點偏差。

10. ($5.2.2) 假設 X_1 和 X_2 的兩個觀察值是獨立同分佈的。

   a. 證明對於這種只有兩個觀察的特殊情況，我們有代數恆等式 $S^2=\left(X_1-X_2\right)^2/2$。

   b. 假設，具體來說，X_1 和 X_2 的兩個觀測值是獨立同分佈 $\operatorname{Bern}(p)$。我們知道樣本方差 S^2 對於方差 $\sigma^2=pq$ 是無偏的，其中 $q=1-p$。但是證明$S=\left|X_1-X_2\right|/\sqrt{2}$ 偏向於 $\sigma=\sqrt{pq}$。

11. ($5.2.2) 假設 $X_1, X_2, \ldots, X_n$ 是獨立同分佈的 $\mathrm{P}\left(X_i=1\right)=1-p$，並且 $\mathrm{P}\left(X_i=2\right)=p($ 本質上是一個 $\operatorname{Bern}(p)+1$ 分佈 $)$。

   a. 證明估計量 $\hat{p}=\bar{X}-1$ 對於 p 是無偏的。

   b. 對於有限樣本大小 n，估計量 $\hat{p}^2=(\bar{X}-1)^2$ 對於 p^2 是否無偏？

12. ($5.2.3) 假設 $X_1, X_2, \ldots$ 是獨立同分佈的，均值 μ 和方差 σ^2。用 $\bar{X}_n=\sum_{i=1}^n X_i/n$ 表示基於 n 個觀測值的樣本均值，對於 $n \geq 1$。$\bar{X}_{10}$ 和 $\bar{X}_{20}$ 中哪一個是 μ的更好估計量？ 解釋你的選擇。

13. ($5.2.3) 考慮兩個估計器，$T_1$ 和 T_2，用於未知參數 θ。假設 $\operatorname{Bias}(T_1)=0, \operatorname{Bias}(T_2)=\theta, \operatorname{Var}(T_1)=4\theta^2$，$\operatorname{Var}(T_2)=\theta^2$。您可能決定使用哪個估算器，為什麼？

14. ($5.2.3) 假設 $\hat{\theta}_1$、$\hat{\theta}_2$ 和 $\hat{\theta}_3$ 是θ，我們知道

$$\mathrm{E}\left[\hat{\theta}_1\right]=\mathrm{E}\left[\hat{\theta}_2\right]=\theta, \quad \text{and} \quad \mathrm{E}\left[\hat{\theta}_3\right]=\theta+3,$$

和

$$\operatorname{Var}\left(\hat{\theta}_1\right)=12, \quad \operatorname{Var}\left(\hat{\theta}_2\right)=10, \quad \text{and} \quad \operatorname{Var}\left(\hat{\theta}_3\right)=4$$

這些估計器中哪個具有最低的 MSE？

15. ($5.2.4) 假設我們觀察到三個 iid $\operatorname{Exp}(\lambda)$ 客戶服務時間，分別為 2、4 和 9 分鐘。$的最大似然估計是多少？

16. ($5.2.4) 假設 $X_1, X_2, \ldots, X_n$ 是 iid $\operatorname{Pois}(\lambda)$。求 λ 的 MLE。

17. ($5.2.4) 考慮一個來自 $\operatorname{Unif}(0, \theta)$ 分佈的大小為 $n=4$ 的 iid 樣本，其中 $\theta>0$ 是未知的。如果 $U_1=3.7$、$U_2=16.3$、$U_3=1.6$ 和 $U_4=7.9$，求 MLE $\hat{\theta}$

18. ($5.2.4) 如果 $X_1, X_2, \ldots, X_n$ 是獨立同分佈 $\mathrm{Nor}\left(\mu, \sigma^2\right)$，其中 μ 和 σ^2 是未知的，σ^2 的 MLE 的期望值是多少？

19. ($5.2.4) 不擇手段地計算2.5點處的digamma函數，即計算$\psi(2.5) = \Gamma'(2.5)/\Gamma(2.5)$。

20. ($5.2.4) 假設$X_1, X_2, \ldots, X_n$是獨立同分佈$\mathrm{Exp}(\lambda)$，我們觀察樣本均值 $\bar{X} = 2$。$\mathrm{P}(X > 2)$的最大似然估計是多少？ 21. (§5.2.4) 假設 $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \mathrm{Geom}(p)$。

a. 找到 p 的最大似然估計量。

b. 假設您有 $n = 5$ 個觀測值：$1, 5, 1, 2, 3$。得到的 MLE 值是多少？

c. $2p^3$ 的 MLE 是多少？

22. (§5.2.5) 令 $X_1, \ldots, X_n$ 是參數 p 的伯努利隨機變量的獨立同分佈樣本。找到 p 的 MOM 估計量。是否有多個 MOM 估計器？

23. (§5.3) 一群男大學生的身高正態分佈，平均為 68 英寸，標準差為 3 英寸。隨機抽取 10 名學生作為樣本。指定樣本均值的抽樣分佈 $\bar{X}$。

24. (§5.3) 現在是分位數時間！

a. 假設 $Y \sim \chi^2(8)$。求 $\mathrm{P}(Y \leq 2.73)$。

b. 求 $\chi^2_{0.025,6}$，即 $\chi^2(6)$ 分佈的 97.5^{th} 分位數使得 $mathrmP\left(\chi^2(6) \leq \chi^2_{0.025,6}\right) = 0.975$。

c. 求 $t_{0.25,9}$，即 $t(9)$ 分佈的 75^{th} 分位數使得 $\mathrm{P}\left(t(9) \leq t_{0.25,9}\right) = 0.75$。

d. 假設 $T \sim t(1000)$。$\mathrm{P}(T > 1.96)$ 是什麼？

e. 求$F_{0.05,4,9}$，即分位數使得$(F(4,9) F_{0.05,4,9})= 0.95 $。

f. 求 $F_{0.975,4,5}$，即滿足 $(F(4,5) F_{0.975,4,5})= 的分位數0.025 $。

25. ($5.3) 求 $\chi^2(k)$ 分佈的期望值。

26. (§5.3)假設$X \sim \mathrm{Nor}(0,1)$和$Y \sim \chi^2(3)$，並且X和Y是獨立的。命名 $3X^2/Y$ 的分佈。（提示：回想一下 $[\mathrm{Nor}(0,1)]^2 \sim \chi^2(1).$）

ryNoSmBk

ryCh 6 【第 6 章】

6.0.0.1 置信區間

在本章中，我們終於開始探索令人興奮的置信區間 (CI) 主題！我們都聽說過置信區間，它提供了一種報告統計數據不確定性的有效方法。例如，假設根據一項新的民意調查，史密斯總統的受歡迎程度為 56 $，正負 3 $。這可以粗略地解釋為我們有 95 $ %$ 確定史密斯的真正受歡迎程度在 $[0.53, 0.59]$ 範圍內的某個陳述。

在本章中，我們將形式化這些概念以解決範圍廣泛的問題。$6.1 以對該主題的簡潔介紹開始。其餘部分展示了不同類型的 CI 的大雜燴。當我們碰巧知道潛在方差（這可能是也可能不是一個大假設）時，6.2 會查看 CI 的特殊情況以獲得正態分佈的均值。$6.3 考慮兩個具有已知方差的正態分佈，並給出它們均值差異的 CI。§$6.4 和 6.5 與它們的兩個前身一樣，但在更現實的情況下，潛在的方差是未知的。$6.666.8 給出正態分佈方差的 CI，兩個正態分佈的方差比，以及伯努利分佈的成功概率，

86.1 - 置信區間簡介

66.2 - 正態均值的置信區間（已知方差）

66.3 - 均值差的置信區間（已知方差）

6.4 - 正態均值的置信區間（方差未知）

86.5 - 均值差的置信區間（方差未知）

$6.6 - 正態方差的置信區間

66.7 - 正態方差比的置信區間

$6.8 - 伯努利成功概率的置信區間

6.1 置信區間簡介

思路：不要單獨通過點估計器來估計一個參數，而是以一定的概率找到一個包含未知參數的（隨機）區間。這為我們提供了一種表達我們對點估計器的信心的方法。

示例：樣本均值 $\bar{X}$ 是未知真實均值 μ 的點估計量。μ 的 95% 置信區間可能類似於

$$\mu \in \left[\bar{X} - z_{0.025}\sqrt{\sigma^2/n}, \bar{X} + z_{0.025}\sqrt{\sigma^2/n}\right]$$

其中 $z_{0.025} = 1.96$ 是 Nor $(0, 1)$ 分佈的 0.975 分位數（參見附錄中的表 B.1）。解釋是我們有信心0.95（即$1.0 - 2(0.025))\mu$ 位於區間內。

定義：未知參數 θ 的 $100(1-\alpha)\%$ 置信區間由兩個隨機變量 L 和 U 給出，滿足 $\mathrm{P}(L \leq \theta\, leqU) = 1 - \alpha$。

隨機變量 L 和 U 是置信下限和置信上限。

數量 $(1-\alpha)$ 是預先指定的置信係數。θ 實際上位於 L 和 U 之間的可能性為 $(1-\alpha)$。

由於 $L \leq \theta \leq U$，我們稱 $[L, U]$ 為 θ 的雙邊 CI

如果 L 滿足 $\mathrm{P}(L \leq \theta) = 1 - \alpha$，則 $[L, \infty)$ 是 $100(1-\alpha)\%$ 一- θ 的 CI 偏低。

類似地，如果 U 滿足 $\mathrm{P}(\theta \leq U) = 1 - \alpha$，則 $(-\infty, U]$ 是 \$100(1-) % \$$\theta$ 的單邊上 CI。

示例：我們以 95% 確定史密斯總統的受歡迎程度為 $56\% \pm 3\%$。

示例：表 6.1 說明了來自 10 個獨立樣本的置信區間的實現，每個樣本包含 100 個不同的觀察值。從每個樣本中，我們使用 100 個觀測值重新計算 L 和 U，以獲得 θ 的 95% 置信區間。在每種情況下，$[L, U]$ 要么覆蓋（包含）要么不覆蓋 θ 的未知真實值（為了討論的目的，假設為 2）。

表中的一些要點：

- 參數θ 是一個固定的未知常數（$\theta = 2$ 的真實值，但在實踐中我們不知道），因此它不會因樣本而異。
- 有些 CI 很瘦（例如，樣本 6），而有些則更圓（例如，樣本 7）。
- 有時 CI 因太低而錯過 θ，即 $\theta > U$ (sample 10)。
- 有時 CI 因太高而錯過 θ，即 $\theta < L$ (sample 3)。
- 我們看到十個 CI 中只有八個實際上覆蓋了 θ。這完全沒問題，即使對於 95 的% CI，因為…

表 6.1：十個樣本的十個置信區間。

- 隨著樣本數量的增加，覆蓋未知 θ 的 CI 的比例將接近 $1 - \alpha = 0.95$。所以，很高興，這一切最終都解決了（假設我們使用的是數學上有效的 CI）！
- 但是，請注意，在現實世界中，您通常只得到一個樣本，因此在 CI 上只有一個鏡頭 - 但您至少有 $1 - \alpha$ 的概率做對了！

置信區間無處不在！我們將查看許多用於正態分佈的均值和方差的 CI，以及用於 Bernoulli 分佈的成功概率的 CI。我們還將擴展這些結果以比較相互競爭的正態分佈（例如，兩個正態中的哪一個具有更大的均值？），以及相互競爭的伯努利分佈。

6.2 正態均值的置信區間（已知方差）

我們將從最簡單的情況開始。

目標：來自具有未知均值 μ 和已知方差 σ^2 的正態分佈的樣本。使用這些觀察來獲得 μ 的 CI。

備註：誠然，這是一個不現實的案例，因為如果我們在現實生活中不知道 μ，那麼我們可能也不知道 σ^2。但這是開始討論的好地方。

設置：假設 $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Nor}\big(\mu, \sigma^2\big)$，其中 σ^2 是已知的。使用 $\bar{X} = \sum_{i=1}^{n} X_i/n$ 作為 μ 的點估計量。（這是公正的！這是 MLE！這是我們的媽媽！）回想一下

$$\bar{X} \sim \text{Nor}\big(\mu, \sigma^2/n\big) \Rightarrow Z \equiv \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \sim \text{Nor}(0,1).$$

數量 Z 稱為樞軸。這對我們來說是一個“起點”。Z 的定義意味著

$$\begin{aligned}
1-\alpha &= \mathrm{P}\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) \\
&= \mathrm{P}\left(-z_{\alpha/2} \le \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \le z_{\alpha/2}\right) \\
&= \mathrm{P}(-z_{\alpha/2}\sqrt{\sigma^2/n} \le \bar{X}-\mu \le z^{\overbrace{\alpha/2}^{\sqrt{\sigma^2/n}}}_{\alpha/2\sqrt{\sigma^2/n})} \le \mu \le \underbrace{\bar{X}+z_{\alpha/2}\sqrt{\sigma^2/n})}_{L} \\
&= \mathrm{P}(\underbrace{\bar{X}-z_{\alpha}}_{U}) \\
&= \mathrm{P}(L \le \mu \le U)
\end{aligned}$$

因此，我們有

μ : 的 $100(1-\alpha)\%$ 雙面 CI

$$\bar{X} - z_{\alpha/2}\sqrt{\sigma^2/n} \le \mu \le \bar{X} + z_{\alpha/2}\sqrt{\sigma^2/n}$$

6.2.0.1 備註：

- 請注意我們如何使用樞軸和代數將 μ 全部單獨“隔離”到不等式的中間。
- 觀察X_1、 ... 、X_n後，可以計算出L和U。在那一點上，沒有什麼是未知的，因為 L 和 U 不涉及 μ。
- 有時我們會將 CI 寫成直觀的形式 $\mu \in \bar{X} \pm H$，其中半角（又名半長）是

$$H \equiv z_{\alpha/2}\sqrt{\sigma^2/n}.$$

示例：假設我們觀察大學橄欖球隊中 $n = 25$ 內線隊員的體重。假設這些是來自 Nor $\left(\mu, \sigma^2\right)$ 分佈的獨立同分佈觀察，我們以某種方式知道方差 $\sigma^2 = 324$。

262.4	274.6	245.0	307.7	281.6
320.5	261.7	283.5	273.5	308.5
257.7	296.2	294.8	279.1	242.1
299.3	254.5	281.1	297.4	281.8
315.8	284.1	286.2	251.8	258.6

該數據產生 $\bar{X} = 279.98$ 的樣本均值。由於 $z_{\alpha/2} = z_{0.025} = 1.96$，我們得到 μ 的以下兩側 $100(1-\alpha) = 95\%$ CI。

$$\mu \in \bar{X} \pm z_{\alpha/2}\sqrt{\sigma^2/n} = 280.0 \pm z_{0.025}\sqrt{324/25} = 280.0 \pm 7.1.$$

所以 μ 的 95% CI 是 $272.9 \leq \mu \leq 287.1$

樣本大小計算：如果我們進行更多的觀察，那麼 CI 會變得更短，因為 $H = z_{\alpha/2}\sqrt{\sigma^2/n}$。實際上，應該進行多少次觀察才能使半長（或“錯誤”）$\leq \epsilon$ ？我們很容易看到

$$z_{\alpha/2}\sqrt{\sigma^2/n} \leq \epsilon \quad \text{iff} \quad n \geq \sigma^2 z_{\alpha/2}^2/\epsilon^2$$

示例：假設在前面的示例中，我們希望半長為 ≤ 3，即 $\mu \in \bar{X} \pm 3$。n 應該是什麼？

$$n \geq \sigma^2 z_{\alpha/2}^2/\epsilon^2 = 324(1.96)^2/9 = 138.3$$

只是為了使 n 成為整數，向上舍入 $n = 139.$

單邊置信區間：我們可以類似地獲得 μ 的單邊 CI（如果我們只對一個方向的界限感興趣）：考慮相同的樞軸 $Z = (\bar{X}-\mu)/\sqrt{\sigma^2/n} \sim \mathrm{Nor}(0,1)$ 我們之前使用的。然後，根據 $(1-\alpha)$ 分位數 z_α 的定義，我們有

$$\begin{aligned} 1-\alpha &= \mathrm{P}\left(Z \leq z_\alpha\right) \\ &= \mathrm{P}\left(\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \leq z_\alpha\right) \\ &= \mathrm{P}\left(\bar{X}-\mu \leq z_\alpha\sqrt{\sigma^2/n}\right) \\ &= \mathrm{P}\left(\bar{X}-z_\alpha\sqrt{\sigma^2/n} \leq \mu\right) \end{aligned}$$

這立即為我們提供了 $100(1-\alpha)\%$ 低於 μ 的 CI，

$$\mu \geq \bar{X} - z_\alpha\sqrt{\sigma^2/n}$$

並且通過對稱性，μ 的 $100(1-\alpha)\%$ 上 CI 是

$$\mu \leq \bar{X} + z_\alpha\sqrt{\sigma^2/n}$$

請注意，我們使用 $1-\alpha$ 分位數 z_α 表示單邊 CI，而不是 $(1-\alpha/2)$ 分位數 $z_{\alpha/2}$。α 的這種“重新分配”使單邊邊界更接近 $\bar{X}$，因此信息量更大。

示例：再次使用我們的足球數據，讓我們得到這些大男孩平均體重的 95% 下限。由於 $z_{0.05} = 1.645$，我們有

$$\mu \geq \bar{X} - z_\alpha\sqrt{\sigma^2/n} = 280.0 - z_{0.05}\sqrt{324/25} = 280.0 - 5.9 = 274.1$$

6.3 正態均值差的置信區間（已知方差）

目標：從具有未知均值和已知方差的兩個正態分佈中採樣。使用這些觀察來獲得兩個均值差異的 CI。

備註：這將為我們提供有關哪個分佈比另一個“更好”的信息，至少就其手段而言。我們稍後會在 $6.5 中做更現實的未知方差案例

示例：為庫存策略 X 和 Y 產生的收入的平均差異給出一個 95% 的置信區間。

設置：假設我們有來自兩個競爭群體的大小為 n 和 m 的樣本：

$$\begin{array}{lcll} X_1, X_2, \ldots, X_n & \overset{\text{iid}}{\sim} & \text{Nor}\left(\mu_x, \sigma_x^2\right) & \text{(population 1)} \quad \text{and} \\ Y_1, Y_2, \ldots, Y_m & \overset{\text{iid}}{\sim} & \text{Nor}\left(\mu_y, \sigma_y^2\right) & \quad \text{(population 2))} \end{array}$$

其中 μ_x 和 μ_y 的平均值是未知的，而 σ_x^2 和 σ_y^2 是已知的。

還假設 X_i 獨立於 Y_i 。

讓我們找到均值差的 CI，$\mu_x - \mu_y$。為此，請從總體 1 和 2 中定義樣本均值，

$$\bar{X} \equiv \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } \bar{Y} \equiv \frac{1}{m}\sum_{i=1}^{m} Y_i$$

明顯地，

$$\bar{X} \sim \text{Nor}\left(\mu_x, \sigma_x^2/n\right) \quad \text{and} \quad \bar{Y} \sim \text{Nor}\left(\mu_y, \sigma_y^2/m\right)$$

以便

$$\bar{X} - \bar{Y} \sim \text{Nor}\left(\mu_x - \mu_y, \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}\right).$$

這給了我們支點

$$Z \equiv \frac{\bar{X} - \bar{Y} - \left(\mu_x - \mu_y\right)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \sim \text{Nor}(0, 1)$$

以便

$$\text{P}\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1 - \alpha$$

使用與單一種群情況相同的操作，我們立即獲得

$\mu_x - \mu_y$ 的 $100(1-\alpha)\%$ 雙面 CI：

$$\mu_x - \mu_y \in \bar{X} - \bar{Y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}.$$

單面 CI：類似地，我們有一個單面的上 CI，

$$\mu_x - \mu_y \leq \bar{X} - \bar{Y} + z_{\alpha}\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}$$

和一側較低的 CI，

$$\mu_x - \mu_y \geq \bar{X} - \bar{Y} - z_{\alpha}\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}.$$

示例：一位旅行教授對 XXX 大學學生 (X) 和 YYY 大學 (Y) 學生進行了相同的測試。她假設測試分數正態分佈，已知標準差分別為 $\sigma_x = 20$ 點和 $\sigma_y = 12$ 點。她隨機抽取 40XXX 分數和 24YYY 測試樣本，並分別觀察 $\bar{X} = 95$ 點和 $\bar{Y} = 60$ 的樣本均值。

讓我們找到 $\mu_x - \mu_y$ 的 90% 雙邊 CI。用 $z_{\alpha/2} = z_{0.05}$ 插入雙面版本，我們得到

$$\begin{aligned}\mu_x - \mu_y \in \bar{X} - \bar{Y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} \\ = 35 \pm 1.645\sqrt{\frac{400}{40} + \frac{144}{24}} \\ = 35 \pm 6.58,\end{aligned}$$

暗示 $28.42 \leq \mu_x - \mu_y \leq 41.58$。換句話說，我們 90% 確定 $\mu_x - \mu_y$ 位於這個區間內。這意味著我們可以非正式地得出結論，平均而言，XXX 學生的得分明顯高於 YYY 孩子。

備註：假設兩個樣本大小都等於 n。為了得到一個半長的 $\leq \epsilon$，我們使用類似於 $6.2 中的推理來發現我們需要

$$n \geq \frac{z_{\alpha/2}^2\left(\sigma_x^2 + \sigma_y^2\right)}{\epsilon^2}$$

下一節將研究更現實的情況，其中基礎觀察的方差是未知的。

6.4 正態均值的置信區間（方差未知）

在本節中，我們將研究更現實的情況，其中基礎正態隨機變量的方差是未知的。這需要更多的工作，但有更多的應用程序。

設置：$X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \operatorname{Nor}\left(\mu, \sigma^2\right)$, with σ^2 未知 事實：這是我們在接下來的討論中需要的三個結果。

(a) $\dfrac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \sim \mathrm{Nor}(0,1)$.

(b) $\bar{X}$ and S^2 are independent.

(c) $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 \sim \dfrac{\sigma^2\chi^2(n-1)}{n-1}$.

證明草圖：我們已經在 \$4.3.3 中證明了 Fact (a)。事實（b）的證明使用了所謂的科克倫定理，這有點超出了我們的範圍。為了推導出 Fact (c)，我們首先應用一些簡單的代數來建立恆等式

$$
\begin{aligned}
\sum_{i=1}^{n}\left(X_i-\mu\right)^2 &= \sum_{i=1}^{n}\left[\left(X_i-\bar{X}\right)+(\bar{X}-\mu)\right]^2 \\
&= \sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 + 2\sum_{i=1}^{n}\left(X_i-\bar{X}\right)(\bar{X}-\mu) + \sum_{i=1}^{n}(\bar{X}-\mu)^2 \\
&= \sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 + 2(\bar{X}-\mu)\sum_{i=1}^{n}\left(X_i-\bar{X}\right) + n(\bar{X}-\mu)^2 \\
&= \sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 + n(\bar{X}-\mu)^2 \left(\text{since} \sum_{i}\left(X_i-\bar{X}\right)=0\right)
\end{aligned}
$$

可以重寫為

$$
\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\mu\right)^2 = S^2 + \frac{n}{n-1}(\bar{X}-\mu)^2
$$

注意 $X_i \sim \mathrm{Nor}\left(\mu,\sigma^2\right)$ 以及 $[\mathrm{Nor}(0,1)]^2 \sim \chi^2(1)$ 暗示 $\left(X_i-\mu\right)^2 \sim \sigma^2\chi^2(1)$，對於所有 i。因此，獨立 χ^2 的加性屬性（來自 \$5.3.2 意味著

$$
\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\mu\right)^2 \sim \frac{\sigma^2\chi^2(n)}{n-1}
$$

同樣，事實 (a) 意味著

$$
\frac{n}{n-1}(\bar{X}-\mu)^2 \sim \frac{\sigma^2\chi^2(1)}{n-1}
$$

由於 $\bar{X}$ 和 S^2 是獨立的（根據事實 (b)），我們可以將方程 6.3 和 6.4 的結果代入方程 6.2，然後調用 χ^2加法再次獲得

$$
\frac{\sigma^2\chi^2(n)}{n-1} \sim S^2 + \frac{\sigma^2\chi^2(1)}{n-1} \Rightarrow S^2 \sim \frac{\sigma^2\chi^2(n-1)}{n-1}
$$

（現在差不多是 t.. 的時間了）使用事實 (a) - (c)，根據 t 分佈的定義，我們有，

$$
\frac{(\bar{X}-\mu)/\sqrt{\sigma^2/n}}{\sqrt{S^2/\sigma^2}} \sim \frac{\mathrm{Nor}(0,1)}{\sqrt{\chi^2(n-1)/(n-1)}} \sim t(n-1).
$$

換句話說，

$$\frac{\bar{X}-\mu}{\sqrt{S^2/n}} \sim t(n-1).$$

備註：此表達式不包含未知 σ^2。它已被 S^2 取代，我們可以計算出來。這個過程被稱為“標準化和學生化”。

現在，通過與已知方差情況相同的操作，我們可以獲得

6.4.0.1 $100(1-\alpha)\%$ μ **的雙面 CI :**

$$\mu \in \bar{X} \pm t_{\alpha/2,n-1}\sqrt{S^2/n}$$

其中 t 分位數可以在附錄的表 B.2 中或通過軟件找到。

單面 CI：相應的 $100(1-\alpha)\%$ 較低和較高 CI 分別是，

$$\mu \geq \bar{X} - t_{\alpha,n-1}\sqrt{S^2/n} \text{ and } \mu \leq \bar{X} + t_{\alpha,n-1}\sqrt{S^2/n}$$

備註：這裡我們使用 t 分佈分位數（而不是像 §6.2 的已知方差情況下的正常分位數。t 分位數往往大於相應的 $\text{Nor}(0,1)$ quantile，所以這些未知方差的 CI 往往比已知方差的 CI 長一點。較長的 CI 是由於我們缺乏關於方差的精確信息這一事實的結果。

示例：考慮我們來自 \$6.2 的數據集，其中列出了 $n=25$ 個大學橄欖球運動員的體重。回想一下，觀察的樣本均值是 $\bar{X}=279.98$。此外，這一次，假設我們不知道方差 σ^2，而是使用通常的樣本方差來估計它，結果是 \$S^{2}=479.65 \$

為了獲得 μ 的雙邊 95% 置信區間，我們可以使用 Excel 函數 t。 $\text{inv}(0.975, 24)$ 得到 $t_{\alpha/2,n-1}=t_{0.025,24}=2.064$。那麼CI的半長為

$$H = t_{\alpha/2,n-1}\sqrt{S^2/n} = 2.064\sqrt{479.65/25} = 9.04$$

因此，CI 為 $\mu \in \bar{X} \pm H = 279.98 \pm 9.04$，或 $270.94 \leq \mu \leq 289.02$，恰好比類似的已知方差 CI 稍寬在 §6.2

這是上面示例的 R 代碼：

$x <-$ 數據幀 $(x=c($

$262.4, 274.6, 245.0, 307.7, 281.6,$

```
320.5, 261.7, 283.5, 273.5, 308.5,
257.7, 296.2, 294.8, 279.1, 242.1,
299.3, 254.5, 281.1, 297.4, 281.8,
315.8, 284.1, 286.2, 251.8, 258.6))
print(confint(lm(x~1,x)))
  2.5 \%   97.5 \%
(Intercept) 270.94  289.02
```

6.5 正態均值差的置信區間（方差未知）

本節研究 CI 的兩個相互競爭的正常人群的平均值之間的差異，並且類似於在 §6.3 中開始的討論。然而，在這裡，我們假設潛在的總體方差是未知的，因此我們的 CI 將涉及 t 分佈分位數而不是正態分位數。

設置：假設我們有來自兩個總體的大小為 n 和 m 的樣本，

$$
\begin{array}{llll}
X_1, X_2, \ldots, X_n & \overset{\text{iid}}{\sim} \operatorname{Nor}\left(\mu_x, \sigma_x^2\right) & (\text{ population } 1), & \text{and} \\
Y_1, Y_2, \ldots, Y_m & \overset{\text{iid}}{\sim} & \operatorname{Nor}\left(\mu_y, \sigma_y^2\right) & (\text{ population } 2)
\end{array}
$$

我們假設均值 μ_x 和 μ_y 是未知的，方差 σ_x^2 和 σ_y^2 是也是未知數。

目標：找到均值差的 CI，$\mu_x-\mu_y$。

我們的討論將在隨後的小節中分為三個案例，每個案例都涉及不同的假設。

- (£6.5.1 隨機樣本 $X_1, X_2, \ldots, X_n$ 和 $Y_1, Y_2, \ldots, Y_m$ 是它們自己相互獨立，未知方差相等，即 $\sigma_x^2=\sigma_y^2=\sigma^2$。這是一個稍微不現實的情況，但當有證據表明方差至少大致相等時會很有用。
- (£6.5.2 隨機樣本 $X_1, X_2, \ldots, X_n$ 和 $Y_1, Y_2, \ldots, Y_m$ 是他們自己彼此獨立，但未知方差是任意的（因此可能不相等），即 $\sigma_x^2 \neq \sigma_y^2$。這可能是最- 逼真的場景，以及最流行的比較手段。
- ($6.5.3) 隨機樣本 $X_1, X_2, \ldots, X_n$ 和 $Y_1, Y_2, \ldots, Y_n$ 是大小相同，不一定相互獨立。特別是，任何特定對 (X_i, Y_i) 中的兩個分量可能是相關的，即 $\operatorname{Corr}(X_i, Y_i) \neq 0$。此外，未知方差是任意的。這種設置可以發揮很大的優勢 - 至少在每對之間具有顯著正相關的情況下。

在每種情況下，我們都將使用通常的樣本均值，

$$
\bar{X} \equiv \frac{1}{n} \sum_{i=1}^n X_i \quad \text{and } \bar{Y} \equiv \frac{1}{m} \sum_{i=1}^m Y_i
$$

現在，關於這三個案例……

6.5.1 方差未知但相等

設置：我們假設兩個樣本 $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \operatorname{Nor}\left(\mu_x, \sigma^2\right)$ 和 $Y_1, Y_2, \ldots, Y_m \overset{\text{iid}}{\sim} \operatorname{Nor}\left(\mu_y, \sigma^2\right)$ 相互獨立，但具有相同的未知方差 σ^2。

示例：我們比較了兩個工業體重秤的平均值，它們的變化相同，但可能以不同的點為中心。

已經定義了樣本均值，讓我們提醒自己樣本方差，

$$
\begin{aligned}
& S_x^2 \equiv \frac{1}{n-1} \sum_{i=1}^n\left(X_i-\bar{X}\right)^2 \sim \frac{\sigma^2 \chi^2(n-1)}{n-1} \text{ and } \\
& S_y^2 \equiv \frac{1}{m-1} \sum_{i=1}^m\left(Y_i-\bar{Y}\right)^2 \sim \frac{\sigma^2 \chi^2(m-1)}{m-1},
\end{aligned}
$$

其中分佈結果遵循公式 6.1p。S_x^2 和 S_y^2 都是公共方差 σ^2 的估計量。更好的估計器是 σ^2 的池化估計器，它對來自 S_x^2 和 S_y^2 的信息進行線性插值，

$$S_p^2 \equiv \frac{(n-1)S_x^2+(m-1)S_y^2}{n+m-2}.$$

6.5.1.1 定理：

$$S_p^2 \sim \frac{\sigma^2\chi^2(n+m-2)}{n+m-2}.$$

證明：結果直接來自於獨立的 χ^2 隨機變量相加的事實。

在一些常見的代數之後（以及 $\bar{X}-\bar{Y}$ 獨立於 S_p^2) 的鬼鬼祟祟的事實，我們看到樞軸

$$\frac{\bar{X}-\bar{Y}-\left(\mu_x-\mu_y\right)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} = \frac{\frac{\bar{X}-\bar{Y}-\left(\mu_x-\mu_y\right)}{\sqrt{\frac{\sigma^2}{n}+\frac{\sigma^2}{m}}}}{\sqrt{S_p^2/\sigma^2}} \sim \frac{\mathrm{Nor}(0,1)}{\sqrt{\frac{2^2(n+m-2)}{n+m-2}}} \sim t(n+m-2),$$

它比分別從 $\bar{X}$ 或 $\bar{Y}$ 的樞軸具有更多的自由度，更多的 df 是好的！

所以，在通常的附加代數之後，我們得到

$\mu_x-\mu_y$: 的 $100(1-\alpha)\%$ 雙邊池 CI

$$\mu_x-\mu_y \in \bar{X}-\bar{Y} \pm t_{\alpha/2,n+m-2}S_p\sqrt{\frac{1}{n}+\frac{1}{m}}$$

其中 $t_{\alpha/2,n+m-2}$ 是適當的 t 分佈分位數。

單面 CI 的：如果我們讓 $A \equiv t_{\alpha,n+m-2}S_p\sqrt{\frac{1}{n}+\frac{1}{m}}$，那麼一側下CI為$\mu_x-\mu_y \geq \bar{X}-\bar{Y}-A$，一側上CI為$\mu_x-\mu_y \leq \bar{X}-\bar{Y}+A$。

示例：讓我們考慮一下 Tech School of Technology 和 Clever University 學生的 IQ，我們將假設一個常見但未知的方差 σ^2。

TST 學生：$X_1,\ldots,X_{25} \stackrel{\text{iid}}{\sim} \mathrm{Nor}\left(\mu_x,\sigma^2\right)$

CU 學生：$Y_1,\ldots,Y_{36} \stackrel{\text{iid}}{\sim} \mathrm{Nor}\left(\mu_y,\sigma^2\right)$

假設事實證明

$$\bar{X}=120, \quad \bar{Y}=80, \quad S_x^2=100, \quad \text{and} \quad S_y^2=95$$

這兩個樣本方差非常接近，所以我們繼續對我們常見的 σ^2 假設感覺良好，我們將使用合併方差估計器，

$$S_p^2=\frac{(n-1)S_x^2+(m-1)S_y^2}{n+m-2}=\frac{(24)(100)+(35)(95)}{59}=97.03$$

因此，由於 $t_{0.025,59}=2.00$，我們看到 $\mu_x-\mu_y$ 的兩側 95% CI 是

$$\begin{aligned}\mu_x - \mu_y \in \bar{X} - \bar{Y} &\pm t_{\alpha/2, n+m-2} S_p \sqrt{\frac{1}{n} + \frac{1}{m}} \\ &= 120 - 80 \pm 2.00\sqrt{97.03}\sqrt{0.0678} \\ &= 40 \pm 5.13,\end{aligned}$$

可以改寫為 $34.87 \leq \mu_x - \mu_y \leq 45.13$。

上面的 CI 不包含 0（甚至不接近），這使得 TST 學生比 CU 學生更精明更加明顯。

6.5.2 方差未知且不等

我們將再次比較來自兩個相互競爭的正常總體的均值，但現在它們可能具有不相等的方差——這似乎是最現實的情況。

設置：這裡我們假設樣本 $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Nor}(\mu_x, \sigma_x^2)$ 和 $Y_1, Y_2, \ldots, Y_m \overset{\text{iid}}{\sim} \text{Nor}(\mu_y, \sigma_y^2)$ 是相互獨立的，但它們可能有不同的未知方差 $\sigma_x^2\ neq\sigma_y^2$

和以前一樣，首先計算樣本均值和方差 $\bar{X}$、$\bar{Y}$、S_x^2 和 S_y^2。由於方差可能不相等，我們不能使用池化估計器 S_p^2。相反，我們使用所謂的韋爾奇逼近法，

$$t^\star \equiv \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \approx t(\nu)$$

其中近似的自由度由強大的表達式給出

$$\nu \equiv \frac{\left(\frac{S_x^2}{n} + \frac{S_y^2}{m}\right)^2}{\frac{(S_x^2/n)^2}{n-1} + \frac{(S_y^2/m)^2}{m-1}}.$$

在通常的樞軸相關代數之後，我們得到以下在實踐中應用非常廣泛的CI。

$\mu_x - \mu_y$ 的近似 $100(1-\alpha)\%$ 雙面 Welch CI：

$$\mu_x - \mu_y \in \bar{X} - \bar{Y} \pm t_{\alpha/2, \nu} \sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}$$

單面 CI：定義數量 $B \equiv t_{\alpha,\nu}\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}$。那麼單邊下 CI 是 $\mu_x - \mu_y \geq \bar{X} - \bar{Y} - B$，單邊上 CI 是 $\mu_x - \mu_y \leq \bar{X} - \bar{Y} + B$。

示例：考慮兩個正常人群，我們將得到 $\mu_x - \mu_y$ 的雙邊 95% CI。假設事實證明

$$\begin{array}{lll} n = 25 & \bar{X} = 100 & S_x^2 = 400 \\ m = 16 & \bar{Y} = 80 & S_y^2 = 100 \end{array}$$

您可以從 S_x^2 和 S_y^2 看出這兩個方差不可能相等。所以我們必須使用近似方法。近似的df是

$$\nu \equiv \frac{\left(\frac{S_x^2}{n}+\frac{S_y^2}{m}\right)^2}{\frac{(S_x^2/n)^2}{n-1}+\frac{(S_y^2/m)^2}{m-1}} = \frac{\left(\frac{400}{25}+\frac{100}{16}\right)^2}{\frac{(400/25)^2}{24}+\frac{(100/16)^2}{15}} = 37.30 \doteq 37$$

在這種情況下，為了保守起見，我們將 df 向下舍入（即，CI 稍長）。由於置信係數為 0.95，我們有 $t_{\alpha/2,\nu} = t_{0.025,37} = 2.026$。這為我們提供了 $\mu_x - \mu_y$ 的以下 CI，

$$\begin{aligned} \mu_x - \mu_y \in \bar{X} - \bar{Y} \pm t_{\alpha/2,\nu}\sqrt{\frac{S_x^2}{n}+\frac{S_y^2}{m}} \\ = 20 \pm 2.026\sqrt{22.25} \\ = 20 \pm 9.56. \end{aligned}$$

示例：這裡有一些時間（以分鐘為單位）供人們解決某種類型的測試問題。X_i 表示未經培訓就試圖解決問題的人，而 Y_i' 表示接受培訓然後解決問題的人。X_i 和 Y_i 來自完全不同的人群（因此研究中總共使用了 $5+5=10$ 人），並且彼此獨立其他。

我們很容易計算以下樣本均值和样本方差。

$$\bar{X} = 24, \quad \bar{Y} = 15, \quad S_x^2 = 192.5, \quad S_y^2 = 142.5$$

然後我們有

$$\nu = \frac{\left(\frac{S_x^2}{n}+\frac{S_y^2}{n}\right)^2}{\frac{(S_x^2/n)^2}{n-1}+\frac{(S_y^2/n)^2}{n-1}} = \frac{4(192.5+142.5)^2}{(192.5)^2+(142.5)^2} = 7.83 \doteq 8$$

在這種情況下，我們已經四捨五入了，因為它很接近。這為我們提供了以下 $\mu_x - \mu_y$ 的 90% 近似 Welch CI，

$$\mu_x - \mu_y \in \bar{X} - \bar{Y} \pm t_{0.05,8}\sqrt{\frac{S_x^2}{n}+\frac{S_y^2}{n}} = 9 \pm 1.860\sqrt{67} = 9 \pm 15.22$$

因此，即使受過訓練的人（平均而言）減少了 9 分鐘，但置信區間非常寬 (± 15.22) 以至於它包含 0，這非正式地意味著我們不能真正確定培訓是否真的有幫助。

我想知道我們是否可以做得更好。..？

6.5.3 配對觀察

.. 是的，有時我們可以比韋爾奇做得更好！

設置：再次考慮兩個相互競爭的正常人群。假設現在我們成對收集兩個群體的觀察結果。不同對之間的隨機變量是獨立的。但是同一對中的兩個觀測值可能不是獨立的。

$$\text{independent}\begin{cases}\text{Pair 1:} & (X_1, Y_1)\\ \text{Pair 2 :} & (X_2, Y_2)\\ \quad\vdots & \quad\vdots\\ \text{Pair } n: & \underbrace{(X_n, Y_n)}\end{cases}$$

特別地，對 (X_i, Y_i) 對於所有 $i \neq j$ 都獨立於對 $\left(X_j, Y_j\right)$ ，但在任何對 i中，X_i 和 Y_i 可能不是獨立的。示例：想想成對的雙胞胎。一個雙胞胎服用一種新藥，另一個服用安慰劑。雙胞胎 Sal 和 Sally 的反應可能會有一些相關性，但它們將獨立於 Dennis 和 Denise 的反應。

想法：通過建立這樣的實驗，我們希望能夠更準確地捕捉兩個正常群體之間的差異，因為我們正在使用這些對來消除外來噪聲。

更多細節：取 n 對觀察，$(X_i, Y_i), i = 1, 2, \ldots, n$，這樣

$$X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \operatorname{Nor}\left(\mu_x, \sigma_x^2\right) \quad \text{and} \quad Y_1, Y_2, \ldots, Y_n \stackrel{\text{iid}}{\sim} \operatorname{Nor}\left(\mu_y, \sigma_y^2\right).$$

我們假設均值 μ_x 和 μ_y 是未知的，方差 σ_x^2 和 σ_y^2 是也是未知的，可能是不平等的。

此外，對 i 獨立於對 j（在對之間），但 X_i 可能不獨立於 Y_i（在一對內）。（另外一個技術要求是所有 (X_i, Y_i) 對都是二元正態的。）

目標：找到均值差的 CI，$\mu_x - \mu_y$。

為此，請定義成對差異，

$$D_i \equiv X_i - Y_i, \quad i = 1, 2, \ldots, n,$$

以便

$$D_1, D_2, \ldots, D_n \stackrel{\text{iid}}{\sim} \operatorname{Nor}\left(\mu_d, \sigma_d^2\right),$$

其中 $\mu_d \equiv \mu_x - \mu_y$（這是我們想要 CI 的目的），並且

$$\sigma_d^2 \equiv \sigma_x^2 + \sigma_y^2 - 2\operatorname{Cov}(X_i, Y_i)$$

現在問題簡化為舊的 $\operatorname{Nor}\left(\mu, \sigma^2\right)$ 情況，其中 μ 和 σ^2 未知。所以讓我們像以前一樣計算樣本均值和方差：

$$\begin{aligned}\bar{D} &\equiv \frac{1}{n}\sum_{i=1}^{n} D_i \sim \operatorname{Nor}\left(\mu_d, \sigma_d^2/n\right) \text{ and}\\ S_d^2 &\equiv \frac{1}{n-1}\sum_{i=1}^{n}\left(D_i - \bar{D}\right)^2 \sim \frac{\sigma_d^2\chi^2(n-1)}{n-1}.\end{aligned}$$

就像以前一樣，我們得到

$$\frac{\bar{D} - \mu_d}{\sqrt{S_d^2/n}} \sim t(n-1).$$

然後在通常的代數之後，我們得到

$\mu_x - \mu_y$: 的 $100(1-\alpha)\%$ 雙邊配對 CI

$$\mu_d = \mu_x - \mu_y \in \bar{D} \pm t_{\alpha/2,n-1}\sqrt{S_d^2/n}$$

單面 CI：定義數量 $C \equiv t_{\alpha,n-1}\sqrt{S_d^2/n}$。那麼單邊下 CI 是 $\mu_d \geq \bar{D} - C$，而單邊上 CI 是 $\mu_d \leq D + C$。

以下示例的目的是將paired-t 方法與$6.5.2 中的"通常" Welch CI 方法進行比較，當兩者都按預期使用時。

示例：讓我們重新審視 $6.5.2 中的示例，在該示例中，我們對沒有接受過培訓的人與接受過培訓的人的問題解決時間感興趣。現在我們將讓一個未經訓練的人解決問題，然後訓練那個人，然後讓同一個人在接受培訓後做類似的問題。因此，我們沒有使用兩組完全不同的五人（總共十人）進行實驗，而是在實驗的"之前"和"之後"部分使用相同的五個人。以下是該抽樣方案的結果。

儘管人們是獨立的，但同一個人解決測試問題的前後時間可能是相關的。（例如，一個敏銳的人可能在訓練前後相對較快，而一個遲鈍的人……，嗯，你知道的。）

讓我們假設所有時間都是正常的。我們想要 $\mu_d =$ ${b}-{a} 的 90% 雙面 CI。我們看到 $n = 5, \bar{D} = 9$, 和 $S_d^2 = 42.5$。

（請注意，我們在這裡"操縱"了數字，以便我們具有與 §6.5.2 的類似示例相同的差異樣本平均值，即 ${X}-{Y}=9 $ 在這兩種情況下。這將使我們能夠公平地比較 Welch 與成對的 t 方法。）

在任何情況下，我們發現 90% 雙邊配對 - t CI 是

$$\mu_d \in \bar{D} \pm t_{0.05,4}\sqrt{S_d^2/n} = 9 \pm 2.13\sqrt{42.5/5} = 9 \pm 6.21$$

6.5.3.1 備註：

- Welch 方法的近似df ν 大於paired-t 方法的df $n-1$。較大的 df 往往會使 CI 更短。這對韋爾奇有利。
- 成對的-t CI 傾向於通過有意地在 X_i 和 Y_i 之間引入正相關來去除一定數量的自然噪聲。這有利於成對的t。
- 在當前示例中，成對的 t 置信區間的寬度 (±6.21) 比 Welch CI 的寬度短得多，從 £6.5.2(±15.22) 開始。顯然，在這個例子中，自然噪聲的減少不僅抵消了兩個 df 的差異。
- 事實上，paired-t CI 不包含 0，這非正式地表明訓練實際上對解決問題的時間有顯著影響 - 這是通過 Welch CI 結果無法得出的結論。

6.5.3.2 道德：盡可能使用配對-t。

6.6 正態方差的置信區間

為了解決我們可以從某個系統中獲得多少可變性，我們現在將獲得正態分佈的方差 σ^2（而不是均值）的置信區間。

設置：$X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Nor}\big(\mu, \sigma^2\big)$。

回想一下，樣本方差的分佈是

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\big(X_i - \bar{X}\big)^2 \sim \frac{\sigma^2\chi^2(n-1)}{n-1}$$

這意味著

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

這樣一些代數和適當的 χ^2 分位數（見表 $B.3$

$$\begin{aligned}1-\alpha &= \mathrm{P}\left(\chi^2_{1-\frac{\alpha}{2},n-1} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\frac{\alpha}{2},n-1}\right)\\ &= \mathrm{P}\left(\frac{1}{\chi^2_{1-\frac{\alpha}{2},n-1}} \ge \frac{\sigma^2}{(n-1)S^2} \ge \frac{1}{\chi^2_{\frac{\alpha}{2},n-1}}\right)\\ &= \mathrm{P}\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right)\end{aligned}$$

因此，我們有

σ^2 的 $100(1-\alpha)\%$ 雙面 CI

$$\sigma^2 \in \left[\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right]$$

備註：σ^2 的 CI 與樣本方差 S^2 成正比。

備註：此 CI 不包含對未知 μ 的引用！單面 CI：σ^2 的 $100(1-\alpha)\%$ 較低和較高 CI 分別為

$$\sigma^2 \ge \frac{(n-1)S^2}{\chi^2_{\alpha,n-1}} \text{ and } \quad \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha,n-1}}.$$

示例：再次考慮我們從 86.2 開始的數據集，其中列出了 $n = 25$ 個足球運動員的體重。回想一下，樣本方差是 $S^2 = 479.65$。這是方差 σ^2 的 95% 上限 CI。

$$\sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha,n-1}} = \frac{(24)(479.65)}{\chi^2_{0.95,24}} = \frac{11512}{13.85} = 831.2.$$

6.7 正態方差比的置信區間

我們研究兩個正態分佈中的哪一個變量更少/更多，例如，兩種類型的工業規模。

設置：$X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Nor}\big(\mu_x, \sigma_x^2\big)$ 和 $Y_1, Y_2, \ldots, Y_m \overset{\text{iid}}{\sim} \text{Nor}\big(\mu_y, \sigma_y^2\big)$，其中所有 X 和 Y 都是獨立的，均值和方差未知。

為了比較兩個分佈的可變性，我們將得到比率 σ_x^2/σ_y^2 的 CI。

回想一下兩個樣本方差的分佈：

$$S_x^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\sim\frac{\sigma_x^2\chi^2(n-1)}{n-1},\text{ and}$$
$$S_y^2=\frac{1}{m-1}\sum_{i=1}^{m}\left(Y_i-\bar{Y}\right)^2\sim\frac{\sigma_y^2\chi^2(m-1)}{m-1}.$$

因此，

$$\frac{S_y^2/\sigma_y^2}{S_x^2/\sigma_x^2}\sim\frac{\chi^2(m-1)/(m-1)}{\chi^2(n-1)/(n-1)}\sim F(m-1,n-1).$$

使用將左尾 F 分位數轉換為右尾分位數的小技巧，然後是一些代數，我們得到

$$\begin{aligned}1-\alpha&=\mathrm{P}\left(F_{1-\frac{\alpha}{2},m-1,n-1}\leq\frac{S_y^2/\sigma_y^2}{S_x^2/\sigma_x^2}\leq F_{\frac{\alpha}{2},m-1,n-1}\right)\\&=\mathrm{P}\left(\frac{S_x^2}{S_y^2}\frac{1}{F_{\frac{\alpha}{2},n-1,m-1}}\leq\frac{\sigma_x^2}{\sigma_y^2}\leq\frac{S_x^2}{S_y^2}F_{\frac{\alpha}{2},m-1,n-1}\right).\end{aligned}$$

所以我們有 σ_x^2/σ_y^2 : 的 $100(1-\alpha)\%$ 雙邊 CI

$$\frac{\sigma_x^2}{\sigma_y^2}\in\left[\frac{S_x^2}{S_y^2}\frac{1}{F_{\frac{\alpha}{2},n-1,m-1}},\frac{S_x^2}{S_y^2}F_{\frac{\alpha}{2}},m-1,n-1\right]$$

備註：σ_x^2/σ_y^2 的 CI 與樣本方差的比率成正比，S_x^2/S_y^2，並且不包含對 μ_x 或 μ_y 的引用。

單面 CI：σ_x^2/σ_y^2 的 $100(1-\alpha)\%$ 較低和較高 CI 分別為

$$\frac{\sigma_x^2}{\sigma_y^2}\geq\frac{S_x^2}{S_y^2}\frac{1}{F_{\alpha,n-1,m-1}}\quad\text{and}\quad\frac{\sigma_x^2}{\sigma_y^2}\leq\frac{S_x^2}{S_y^2}F_{\alpha,m-1,n-1}$$

備註：如果您想要 σ_y^2/σ_x^2 的 CI，只需翻轉討論的各種 CI 中的所有 X 和 Y以上。

示例：假設 25 人參加考試 A，16 人參加考試 B。假設所有分數都是正常且獨立的。

如果 $S_{\mathrm{A}}^2=100$ 和 $S_{\mathrm{B}}^2=70$，讓我們為比率 $_ 找到 95% 的上置信區間{}^{2} / _{}^{2}$。查找 $F_{\alpha,n\mathrm{B}-1,n\mathrm{A}-1}=F_{0.05,15,24}$ 分位數（見表 B.4 B. 7、我們得到

$$\frac{\sigma_{\mathrm{A}}^2}{\sigma_{\mathrm{B}}^2}\leq\frac{S_{\mathrm{A}}^2}{S_{\mathrm{B}}^2}F_{\alpha,n\mathrm{B}-1,n\mathrm{A}-1}=\frac{100}{70}(2.11)=3.01$$

6.8 伯努利成功概率的置信區間

目標：獲得伯努利成功概率 p 的置信區間。

示例：有很多實際問題可以激發這種材料。

- 我們 95% 確定總統的受歡迎程度在 $53\% \pm 2.5\%$ 範圍內
- 我們有 90% 的信心，製造商生產的缺陷品的比例是 $2.3\% \pm 0.2\%$。

設置：假設 $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bern}(p)$

由於 $\sum_{i=1}^{n} X_i \sim \text{Bin}(n,p)$，我們知道 $\bar{X} \sim \frac{1}{n}\text{Bin}(n,p)$。讓我們假設 n 是"大的"，這樣我們就可以使用中心極限定理來近似正態分佈的二項式分佈 - 而不必擔心相關的"連續性校正"。我們所說的"大"是指 $np \geq 5$ 和 $n(1-p) \geq 5$，無論我們認為 p 可能是什麼——也許可以對 p 做一個初步的有根據的猜測。（如果 n 不大，那麼我們將不得不使用討厭的二項式表，我們不想在這里處理它！）請注意

$$\text{E}[\bar{X}] = \text{E}\left[X_i\right] = p \quad \text{and} \quad \text{Var}(\bar{X}) = \text{Var}(X_i)/n = pq/n.$$

那麼對於大的 n，CLT 意味著

$$\frac{\bar{X} - \text{E}[\bar{X}]}{\sqrt{\text{Var}(\bar{X})}} = \frac{\bar{X} - p}{\sqrt{pq/n}} \approx \text{Nor}(0,1)$$

現在讓我們做一些瘋狂的事情，並通過其最大似然估計器 $\bar{X}(1-\bar{X})$ 來估計 pq。這給

$$\frac{\bar{X} - p}{\sqrt{\bar{X}(1-\bar{X})/n}} \approx \text{Nor}(0,1)$$

那麼通常的代數意味著

$$\begin{aligned} 1-\alpha &\doteq \text{P}\left(-z_{\alpha/2} \leq \frac{\bar{X}-p}{\sqrt{\bar{X}(1-\bar{X})/n}} \leq z_{\alpha/2}\right) \\ &= \text{P}\left(\bar{X} - z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq p \leq \bar{X} + z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right) \end{aligned}$$

這導致

p 的近似 $100(1-\alpha)\%$ 雙面 CI：

$$p \in \bar{X} \pm z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$$

單面 CI：p 的大約 $100(1-\alpha)\%$ 較低和較高 CI 分別為

$$p \geq \bar{X} - z_{\alpha}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \text{ and } p \leq \bar{X} + z_{\alpha}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$$

示例：學生正確回答某個測試問題的概率是 p。假設 200 名學生的隨機樣本對問題產生 160 個正確答案。p 的 95% 雙邊 CI 由下式給出

$$p \in \bar{X} \pm z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} = 0.8 \pm 1.96\sqrt{\frac{(0.8)(0.2)}{200}} = 0.8 \pm 0.055$$

Sample-Size 計算：兩側 CI 的半寬為

$$z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$$

我們應該進行多少次觀察才能使半長為 $\leq \epsilon$ ？

$$z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq \epsilon \Longleftrightarrow n \geq \left(z_{\alpha/2}/\epsilon\right)^2 \bar{X}(1-\bar{X})$$

當然，$\bar{X}$ 在進行觀察之前是未知的。n 的保守選擇是通過最大化 $\bar{X}(1-\bar{X})=1/4$。然後我們有

$$n \geq z_{\alpha/2}^2/\left(4\epsilon^2\right).$$

另一方面，如果我們能夠以某種方式對 p 進行初步估計（例如，基於來自試點研究的初步樣本均值），我們可以使用

$$n \geq z_{\alpha/2}^2 \hat{p}(1-\hat{p})/\epsilon^2.$$

示例：假設我們想要獲取足夠多的樣本，使得 p 的 95% 雙邊 CI 不會大於 $\pm 3\%$。我們需要進行多少次觀察？

$$n \geq z_{\alpha/2}^2/\left(4\epsilon^2\right)=\frac{(1.96)^2}{4(0.03)^2}=1067.1$$

所以取 $n=1067$ 個樣本。

現在讓我們假設一項試點研究給出了 $\hat{p}=0.65$ 的估計值。n 應該如何修改？

$$n \geq z_{\alpha/2}^2 \hat{p}(1-\hat{p})/\epsilon^2=\frac{(1.96)^2}{(0.03)^2}(0.65)(0.35)=971.1$$

所以現在我們只需要$n=971$。

備註：我們還可以為兩個相互競爭的伯努利參數之間的差異推導出近似 CI。例如，假設 $X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \operatorname{Bern}(p_x)$ 和 $Y_1, Y_2, \ldots, Y_m \stackrel{\text{iid}}{\sim} \operatorname{Bern}\left(p_y\right)$ ，其中兩個樣本相互獨立。然後，使用本節中已經開發的技術，我們可以以簡單的方式獲得以下成功比例差異的 CI。

p_x-p_y 的近似 $100(1-\alpha)\%$ 雙邊 CI：

$$p_x-p_y \in \bar{X}-\bar{Y} \pm z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}+\frac{\bar{Y}(1-\bar{Y})}{m}}.$$

示例：Tech School of Technology 和 Universal College University 學生在 SAT 考試的數學部分至少獲得 700 分的概率分別為 $p_{\odot}$ 和 p_{Θ}。200 名 TST 學生的隨機樣本產生了 160 名學生，他們在測試中獲得了 ≥ 700。但是，遺憾的是，400 名 UCU 學生的樣本顯示只有 50 人成功。那麼成功概率差異的 95% CI 是

$$
\begin{aligned}
p_\Theta - p_\Theta \in \bar{X} - \bar{Y} \pm z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n} + \frac{\bar{Y}(1-\bar{Y})}{m}} \\
= 0.8 - 0.125 \pm 1.96\sqrt{\frac{(0.8)(0.2)}{200} + \frac{(0.125)(0.875)}{400}} \\
= 0.675 \pm 0.064.
\end{aligned}
$$

對UCU來說看起來不是很好。

6.9 6.9 練習

1. ($.1 考慮以下對於某個總體的未知均值的 95% 置信區間：

$$\mu \in \bar{X} \pm z_{0.025}\sqrt{\sigma^2/n} = -5 \pm 15$$

以下哪些陳述是正確的？

a. 我們 95% 確定 μ 介於 -20 和 10 之間。

b. 如果我們想要 μ 的 99% 置信區間，它會更寬。

c. 2μ 對應的95% 置信區間是-10 ± 30。

d. 如果我收集 100 組數據併計算相應的 100 個置信區間，那麼其中正好 95 個將包含 μ。

2. ($6.2) 從亞特蘭大開車到舊金山所需的時間大致呈正態分佈，標準差為 $\sigma = 25$ 小時。

a. 假設 20 個人每人進行一次 iid 旅行，他們所用時間的樣本均值為 $\bar{X} = 100$ 小時。為平均行程時間構建一個 95% 的雙邊置信區間。

b. 假設我們想要 95% 確信估計平均行程時間的誤差小於 2 小時。應該使用什麼樣本量？

3. ($6.3) 我喜歡玉米棒！讓我們研究兩種玉米的穗長——白色和黃色。1 假設我們（不知何故）知道兩種類型的長度具有大致相同的標準差，$\sigma_{\text{w}} = \sigma_{\text{y}} = 4$ cm。分別測試兩個大小為$n_{\text{w}} = 30$ 和$n_{\text{y}} = 20$的隨機樣本，樣本平均長度為${W}=22 $ 和 $\bar{Y} = 26$ cm。在長度的平均差上構建一個 95% 的置信區間。

4. ($6.4) 假設我們收集以下觀察值：$7, -2, 1, 6$。讓我們假設這些人是來自具有未知方差 σ^2 的正態分佈的獨立同分佈。給我一個平均 μ 的雙邊 95% 置信區間。

[1] 一個典型的玉米穗上有多少粒？800！這做了很多爆米花！5.（6.4 16 名亞特蘭大獵鷹隊足球運動員的體重如下（以磅為單位，而不是公斤）。

225	301	281	263
216	237	249	204
250	238	300	217
318	255	275	295

為平均權重構建一個 95% 的雙邊置信區間。

6.（6.4 考慮以下 $n = 15$ 觀察的數據集，方便地以數字順序呈現。

-123	-112	-100	-84	-83
-61	-7	17	20	21
26	33	33	43	80

為均值構建一個單邊 95% 上置信區間。

7. ($6.4) 我們喜歡在喬治亞山區的豪華小屋度假。這些年來，我們進行了 25 次 iid 旅行。平均而言，它需要 6.10 $加侖的汽油，樣本標準偏差為 0.16 $加侖。為每次旅行的平均加侖數找到一個低 90 $的置信區間。

8. ($6.4) 假設 $\mu \in [-30, 90]$ 是某個庫存策略產生的平均年成本的 90% 置信區間。進一步假設這個區間是基於對基礎庫存系統的四個獨立同分佈的觀察，我們假設這是正常的，但具有未知的方差。不幸的是，老闆決定她想要一個 95% 的置信區間。你能提供嗎？

9. ($6.4) 考慮方差未知時正態均值的通常雙邊置信區間，

$$\mu \in \bar{X} \pm t_{\alpha/2,n-1}\sqrt{S^2/n}$$

令 $H = t_{\alpha/2,n-1}\sqrt{S^2/n}$ 表示這個區間的半寬。可以證明

$$H \sim \sigma t_{\alpha/2,n-1}\frac{\chi(n-1)}{\sqrt{n(n-1)}} \quad \text{(the chi distribution!)}$$

然後，這反過來又導致了結果

$$\mathrm{E}[H] = \sigma t_{\alpha/2,n-1}\sqrt{\frac{2}{n(n-1)}}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)},$$

其中 $\Gamma(\cdot)$ 是 gamma 函數，並且具有眾所周知的性質 $\Gamma(a) = (a-1)\Gamma(a-1)$ for $a > 1$。特別是，經過一些代數（包括 $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}\Big)$ 的事實，結果證明如果 $n \geq 4$ 是偶數，那麼預期的半長與

$$\zeta(n) \equiv \frac{t_{\alpha/2,n-1}}{\sqrt{n(n-1)}}\frac{\left(\frac{n-2}{2}\right)!}{\left(\frac{n-3}{2}\right)\left(\frac{n-5}{2}\right)\cdots\frac{1}{2}}$$

使用 $\alpha = 0.05$ 的上述方程，確定 $n = 4, 6, 8$ 中的哪一個給出最小的預期寬度。

10. ($6.5.1) 大小為 20 的隨機樣本取自兩個獨立的正常人群。樣本均值和標準差分別為 $\bar{X} = 23.10$、$S_x^2 = 12.96$、$\bar{Y} = 20.00$ 和 $S_y^2 = 12.00$。假設 $\sigma_x^2 = \sigma_y^2$，求 $\mu_x - \mu_y$ 的 95% 雙邊置信區間。

11. (6.5.2 我們正在研究解決兩個初等數學問題所需的時間。假設我們讓四個學生嘗試問題 A，另外四個學生嘗試問題 B。假設一切都是獨立且正常的。結果如下所示（片刻之間）。

假設方差不相等，為問題 A 和 B 之間的均值差找到一個雙邊 95% 置信區間。

12. ($6.5.3 我們再次研究解決兩個初等數學問題所需的時間。假設我們要求四個學生同時嘗試問題 A 和問題 B。假設學生是獨立的並且所有結果都服從正態分佈，但請注意一個特定學生在這兩個問題上的時間很可能是正相關的。這些結果如下所示（以秒為單位）。

再次為 A 和 B 的均值差找到一個雙邊 95% CI。

13. (6.5.3) 假設我們想看看一種新藥是否能降低人們的膽固醇水平。我們給四個人服用這種新藥，我們獲得了以下關於他們膽固醇水平 (CL) 的“之前”和“之後”結果。

看起來這種藥物確實有助於降低 CL。讓我們假設上述所有數字都是正常的，並且四個測試對象彼此獨立。為“之前”和“之後”均值的差異找到一個兩側的 95% 置信區間。

14. ($6.6) 假設 $X_1, X_2, \ldots, X_6$ 是具有未知均值和未知方差 σ^2 的獨立同態正態分佈。進一步假設 $\bar{X} = 1000$ 和 $S^2 = 25$。求 σ^2 的 90% 雙邊置信區間。

15. ($6.6) 考慮具有未知均值 μ 和未知方差 σ^2 的 iid 正態觀察 X_1、X_2、 ... 、X_7。σ^2 的通常 90% 雙邊置信區間的預期寬度是多少？您可以以 σ^2 的形式保留您的答案

16. (6.7) 假設 $X_1, X_2, \ldots, X_{20}$ 是獨立同分佈 $\mathrm{Nor}\left(\mu_x, \sigma_x^2\right), Y_1, Y_2, \ldots, Y_{20}$ 是獨立同ID $\mathrm{Nor}\left(\mu_y, \sigma_y^2\right)$，並且所有 X_i 都獨立於所有 Y_i 。進一步假設$\bar{X} = 23$，$S_x^2 = 5184$，$\bar{Y} = -11$，$S_y^2 = 900$。求總體方差比的 95% 雙邊置信區間 σ_x^2/σ_y^2。

17. ($6.8) 一位民意測驗專家詢問了 2000 人的樣本，他們是否支持某項提議。正好有 1200 人回答“是”。為支持該提案的總體百分比找到一個雙邊 95% 置信區間。

18. (6.8 假設 $X_1, X_2, \ldots, X_n$ 是具有未知均值 p 的 iid Bernoulli，並且我們已經進行了初步試點調查，表明 $p \doteq 0.7$. n 必須有多大才能使兩側 95% 置信區間的半長為 0.1？（給出最小的數字。）

19. ($6.8) Tech School of Technology 和 Universal College University 的學生能成功積分 $\int_1^{\infty} e^{-x} dx = 1/e$ 的概率是 $p_{x} $ 和 p_y，分別。800 名 TST 學生的隨機樣本產生了 760 名學生，他們得到了正確的答案。但是，遺憾的是，800 名 UCU 學生的樣本顯示只有 100 人成功。找到成功概率差異的 95% CI。

ryNoSmBk

ryCh 7 【第 7 章】

7.0.0.1 假設檢驗

對於有興趣進行研究的學生和進行現實世界分析的從業者來說，假設檢驗都是一個重要的學科領域。這個想法是提出假設並以統計嚴格的方式測試其有效性。例如，“製造零件的平均壽命是否至少為兩年？” 或“新藥是否比目前市場上的藥物更有效？” 或“A 項的平均抗拉強度是否大於 B 項？”

我們在 $7.1 中介紹並激發了這個了不起且有用的主題，包括關於如何正確制定假設以及如何考慮在測試過程中可能發生的某些類型的錯誤的評論（例如，我們希望避免拒絕假設事實證明這是正確的）。$7.2 討論了與正態分佈均值相關的假設檢驗，在非常特殊（並且可能不現實）的情況下，正態分佈的方差以某種方式已知。$7.3 做同樣的事情，但現在當方差未知時 - 這種情況在實踐中具有巨大的適用性。$7.4 為其他參數提供了許多測試，例如正態分佈的方差和伯努利分佈的成功概率參數。$7.5 是關於擬合優度檢驗的擴展討論，

$7.1 - 假設檢驗簡介

87.2 - 正態均值的假設檢驗（已知方差）

$7.3 - 正態均值的假設檢驗（方差未知）

$7.4 - 其他參數的綜合測試

$7.5 - 擬合優度測試

7.1 假設檢驗簡介

本章的目標是通過收集數據來研究人口的屬性，然後使用這些數據得出關於人口的合理、統計上有效的結論。我們從假設檢驗的高級視圖開始。有關在特定分佈/參數場景中使用假設檢驗的詳細信息將在後續部分中介紹。

7.1.1 我們的一般方法

我們的遊戲計劃如下：

1. 陳述一個關於人口的假設。

2. 選擇一個檢驗統計量來檢驗假設是否為真。

3. 根據我們的觀察計算檢驗統計量。

4. 解釋結果——檢驗統計是否表明我們拒絕或未能拒絕假設？

7.1.1.1 &7.1.1 1 陳述一個假設

假設只是關於參數值或整個分佈的陳述或聲明。為了提供支持或反對聲明真實性的證據，我們執行假設檢驗，包括原假設 (H_0) 和替代假設 (H_1) 覆蓋整個參數空間。零假設有點代表“現狀”。這不一定是真的，但我們會不情願地堅持下去，直到“證明”不是這樣。

示例：我們目前認為一包雞肉的平均重量 μ 是 μ_0 盎司。（我們指定 μ_0。）但既然我們有懷疑，一個合理的測試是

$$H_0 : \mu = \mu_0 \quad \text{vs. } H_1 : \mu \neq \mu_0.$$

這是一個雙向測試。如果樣本均值 $\bar{X}$（μ 的估計量）“太大”或“太小”，我們將拒絕原假設 H_0 的信念。

示例：我們希望一個新品牌的輪胎能夠平均使用超過 μ_0 英里的平均 μ。（同樣，我們指定 μ_0。）但我們確實需要證據才能合理確定地陳述該主張。否則，我們將堅持使用舊品牌。

$$H_0 : \mu \leq \mu_0 \quad \text{vs. } H_1 : \mu > \mu_0.$$

這是一個單方面的測試。僅當 $\bar{X}$ “太大”時，我們才會拒絕 H_0。

示例：我們測試某種類型汽車的平均排放量是否小於 $\mu_0$ppm。但我們需要證據。

$$H_0 : \mu \geq \mu_0 \quad \text{vs. } H_1 : \mu < \mu_0.$$

這也是一種片面的考驗。如果 $\bar{X}$ “太小”，我們將拒絕原假設，然後才聲稱排放量很低。想法：H_0 是舊的、保守的“現狀”（認為它是“在被證明有罪之前是無辜的”），而 H_1 是新的、激進的假設，只是渴望看到光明的一天。儘管我們可能不太確定 H_0 的真實性，但除非我們看到支持 H_1 的大量證據，否則我們不會拒絕它而支持 H_1。如果沒有足夠的證據支持 H_1，那麼我們“無法拒絕”H_0。這大致意味著我們勉強接受了 H_0。

7.1.1.2 &.1.1 2 選擇一個檢驗統計量

測試統計量是我們用來測試 H_0 是否為真的隨機變量。

例如，如果我們進行涉及均值 μ 的假設檢驗，我們自然會將樣本均值 $\bar{X}$ 與假設均值 μ_0 進行比較。比較是使用具有已知抽樣分佈的檢驗統計量完成的，在這種情況下，

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \text{Nor}(0, 1) \quad \left(\text{if } \sigma^2 \text{ is known and } H_0 \text{ is true}\right)$$

或者

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1) \text{ (if } \sigma^2 \text{ is unknown and } H_0 \text{ is true).}$$

隨著本章的進行，我們將更詳細地討論這一點。

7.1.1.3 &7.1.1 **計算檢驗統計量**

在這裡，我們收集樣本數據，然後根據該數據計算檢驗統計量。

示例：考慮新藥代謝的時間。平均而言，當前藥物需要 $\mu_0 = 15$ 分鐘。新藥更快嗎？

聲明：新藥的預期時間 μ 是 < 15 min。所以明顯的測試是

$$H_0 : \mu \geq 15 \quad \text{vs.} \quad H_1 : \mu < 15$$

假設我們有以下數據：$n = 20$，$\bar{X} = 14.88$，$S = 0.333$。那麼檢驗統計量是

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = -1.61$$

什麼會導致我們拒絕 H_0 ？如果 $\bar{X} \ll \mu_0 = 15$，這將表明 H_0 可能是錯誤的。等效地，如果 T_0 是“顯著”$\ll 0$，我們將拒絕 H_0。所以現在的問題是：$T_0 = -1.61$ 小到足以拒絕 H_0 嗎？

7.1.1.4 \$7.1.14 **解釋檢驗統計量**

如果 H_0 實際上是事物的真實狀態，那麼 $\mu = \mu_0$，並且根據我們對未知方差情況下的正態平均置信區間 (\$6.4 的討論，我們有

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1).$$

考慮到這個分佈結果，我們將檢查檢驗統計量 T_0 的觀察值的合理性。

- 如果該值被認為不可信，那麼我們將拒絕 H_0 並選擇 H_1。
- 否則，如果合理性是合理的，那麼我們將無法拒絕 H_0。

示例（繼續上一個示例）：如果 H_0 為真，則 $T_0 \sim t(n-1) \sim t(19)$。那麼從 $t(19)$ 隨機變量中得到 $T_0 = -1.61$ 是否合理（或者至少不離譜）？

如果是，那麼我們將無法拒絕（“勉強接受”）H_0。如果否，那麼我們將拒絕 H_0 支持 H_1

讓我們看看… 從表 B.2 中。我們有

$$t_{0.95,19} = -t_{0.05,19} = -1.729 \quad \text{and} \quad t_{0.90,19} = -t_{0.10,19} = -1.328$$

換句話說，

$$\mathrm{P}(t(19) < -1.729) = 0.05 \text{ and } \mathrm{P}(t(19) < -1.328) = 0.10$$

這意味著

$$0.05 < \mathrm{P}(t(19) < \underbrace{-1.61}_{T_0}) < 0.10.$$

也就是說，如果 H_0 為真，那麼我們看到 $T_0 \leq -1.61$ 的機會只有 5% 到 10% 之間。這個概率不是很高，但也不算小。我們的拒絕/拒絕失敗決定將歸結為我們認為 T_0 的一個難以置信的值，它在尾部太遠了，無法合理地成為 $t(19)$ 的觀察值。

例如，我們會在“級別”0.10 拒絕 H_0，因為

$$T_0 = -1.61 < -t_{0.10,19} = -1.328.$$

但我們無法在“級別”0.05 拒絕 H_0，因為

$$T_0 = -1.61 > -t_{0.05,19} = -1.729.$$

更多關於這個即將到來的！

7.1.2 我們的方式的錯誤

哪裡會出錯？當我們進行假設檢驗時，我們通常會最終拒絕或未能拒絕原假設 H_0。有時我們會做出正確的決定，有時卻不會。事實上，可能會發生四件事——兩件好事，兩件壞事。

- 如果 H_0 確實是真的，並且我們認為它是真的，那就太好了！(五)
- 如果 H_0 實際上是假的並且我們得出結論它是假的 - 好！(- 如果 H_0 實際上是真的，但我們認為它是假的 - 糟糕！這稱為 I 類錯誤。

示例：我們錯誤地認為一種新的劣質藥物比目前市場上的藥物更好。（哎呀！）

- 如果 H_0 實際上是假的，但我們認為它是真的 - 糟糕！這稱為 II 型錯誤。

示例：我們錯誤地認為一種新的、優質的藥物比目前市場上的藥物更差。（哎呀！）

我們注意到 I 類錯誤通常被認為比 II 類錯誤“更糟”。

我們有責任為我們願意忍受的相應的 I 類和 II 類錯誤概率設置上限 α 和 β（儘管我們以後將這些界限視為等式），即，

$$\mathrm{P}(\text{ Type I error }) = \mathrm{P}(\text{ Reject } H_0 \mid H_0 \text{ true }) = \alpha$$

和

$$\mathrm{P}(\text{ Type II error }) = \mathrm{P}(\text{ Fail to Reject } H_0 \mid H_0 \text{ false }) = \beta.$$

我們通常會選擇非常小的數量 α 和 β（例如，0.01 或 0.05），並且肯定使得 $\alpha + \beta < 1$。但請注意，非常小的 α 和 β 值是有代價的，例如，非常小的 α 使得拒絕 H_0 變得更加困難，即使它恰好是錯誤的。

定義：I 類錯誤的概率，α，稱為檢驗的大小或顯著性水平。

定義：假設檢驗的功效是

$$1 - \beta = \mathrm{P}(\text{ Reject } H_0 \mid H_0 \text{ false}).$$

功率大就好，這種情況下我們說測試是“強大的”。

7.2 正態均值的假設檢驗（已知方差）

在本節中，我們將討論在各種情況下涉及正態分佈均值的假設檢驗，所有這些都涉及已知方差 (s)。在 {7.2.1 中，我們涵蓋了單個正態分佈均值的檢驗。$7.2.2 涉及此類測試的設計，以滿足對 I 類和 II 類錯誤的某些限制。$7.2.3 為我們提供了比較兩個相互競爭的正態分佈均值的檢驗。

7.2.1 一個樣本測試

假設 $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \operatorname{Nor}(\mu, \sigma^2)$，其中 σ^2 以某種方式已知（這不是很現實，但請逗我們一會兒）。

如 $7.1 中所述，雙面測試（也稱為簡單測試）檢查所討論的參數是否等於特定的給定值。例如，下面是均值的雙邊檢驗：

$$H_0: \mu = \mu_0 \quad \text{vS.} \quad H_1: \mu \neq \mu_0,$$

我們預先指定一個假設值 μ_0 作為平均值。

備註：當然，這個特別簡單的 H_0 永遠不可能精確到任意小數位。但這不是我們關心的問題——我們主要感興趣的是看看我們是否有足夠的統計顯著證據來斷然拒絕它。

無論如何，我們將使用 $\bar{X}$ 來估計 μ。如果 $\bar{X}$ 與 μ_0 “顯著不同”（瘋狂的高或瘋狂的低），那麼我們將拒絕 H_0。但是“顯著不同”有多少呢？為了確定“顯著不同”的含義，我們首先定義檢驗統計量，

$$Z_0 \equiv \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

如果 H_0 為真，則 $\mathrm{E}[\bar{X}] = \mu_0$ 和 $\operatorname{Var}(\bar{X}) = \sigma^2 / n$，因此 $Z_0 \sim \operatorname{Nor}(0,1)$。然後我們有

$$\mathrm{P}\left(-z_{\alpha/2} \leq Z_0 \leq z_{\alpha/2}\right) = 1 - \alpha.$$

因此，如果 H_0 為真，則區間 $\left[-z_{\alpha/2}, z_{\alpha/2}\right]$ 之外的 Z_0 值將極不可能，當然也很可疑。所以我們的假設檢驗決定如下。

μ 的雙面測試：

$$\text{Reject } H_0 \quad \text{iff} \quad |Z_0| > z_{\alpha/2}.$$

這向我們保證

$$\begin{aligned} \mathrm{P}(\text{ Type I error }) &= \mathrm{P}(\text{ Reject } H_0 \mid H_0 \text{ true }) \\ &= \mathrm{P}\left(|Z_0| > z_{\alpha/2} \mid Z_0 \sim \operatorname{Nor}(0,1)\right) \\ &= \alpha. \end{aligned}$$

如果 $|Z_0| > z_{\alpha/2}$，那麼我們處於測試的拒絕區域（也稱為臨界區域）。如果 $|Z_0| \leq z_{\alpha/2}$，那麼我們在接受區域。

單邊檢驗與參數趨向於某個方向的假設有關。例如，考慮均值最多為 μ_0 的原假設

$$H_0: \mu \leq \mu_0 \quad \text{vS.} \quad H_1: \mu > \mu_0.$$

我們實際上以一種模仿雙邊測試的方式來解決這個問題，但我們只是做出了一些不同的最終決定。事實上，我們考慮與之前相同的檢驗統計量 $Z_0 = \sqrt{n}\left(\bar{X} - \mu_0\right) / \sigma$。但是現在，如果 H_0 為真，那麼在單邊區間 $(-\infty, z_\alpha$] 之外的非常大的 Z_0 值是極不可能的. 因此，在這種情況下，我們

$$\text{Reject } H_0 \text{ iff } Z_0 > z_\alpha.$$

換句話說，如果$Z_0 > z_\alpha$，這表明$\mu > \mu_0$。

同樣，對於另一個單向測試，我們有

$$H_0 : \mu \geq \mu_0 \quad \text{vs.} \quad H_1 : \mu < \mu_0.$$

如果 H_0 為真，則 Z_0 的值超出區間 $[-z_\alpha, \infty)$ 的可能性極小。所以，

$$\text{Reject } H_0 \quad \text{iff} \quad Z_0 < -z_\alpha.$$

如果 $Z_0 < -z_\alpha$，這意味著 $\mu < \mu_0$。

示例：我們從 §6.2 中檢查 25 名大學橄欖球運動員的體重 假設我們以某種方式知道體重是正態分佈的 $\sigma^2 = 324$。在那個舊示例中，我們發現樣本均值 $\bar{X} = 280.0$。

我們將檢驗平均權重為 $\mu_0 = 272$ 的雙邊（簡單）假設，並且我們將保持 I 類型錯誤 $= 0.05$ 的概率。然後我們有

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0.$$

在這裡我們有

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{280 - 272}{\sqrt{324/25}} = 2.22.$$

由於 $|Z_0| = 2.22 > z_{\alpha/2} = z_{0.025} = 1.96$，我們在 $\alpha = 0.05$ 處拒絕 H_0。

請注意，較低的 α 會導致較高的 $z_{\alpha/2}$，在這種情況下，拒絕 H_0 會"更難"。例如，如果 $\alpha = 0.01$，則 $z_{0.005} = 2.58$，對於本例中給出的場景，我們將無法拒絕 H_0。

定義：測試的 p-value 是導致拒絕 H_0 的最小顯著性水平 α。

備註：研究人員經常報告他們進行的任何測試的 p-values。這樣做有助於防止不道德的 cad 在事後挑选和選擇某些 α 值，以操縱所需的拒絕/接受結論。

無論如何，讓我們推導出一個用於雙面測試的 p-value 的表達式

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0.$$

我們拒絕 H_0 iff

$$|Z_0| > z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2),$$

其中 $\Phi(x)$ 是我們的老夥伴，$\text{Nor}(0, 1)$ cdf。所以我們拒絕 H_0 iff

$$\Phi(|Z_0|) > 1 - \alpha/2 \quad \text{iff} \quad \alpha > 2(1 - \Phi(|Z_0|))$$

因此，對於已知方差情況下的雙邊正態均值檢驗，p-value 為 $p = 2(1 - \Phi(|Z_0|))$。

同樣，對於單面測試，

$$H_0 : \mu \leq \mu_0 \quad \text{vs.} \quad H_1 : \mu > \mu_0,$$

我們有 $p=1-\Phi\left(Z_0\right)$

而對於另一個單方面的測試，

$$H_0: \mu \geq \mu_0 \quad \text{vS.} \quad H_1: \mu<\mu_0,$$

很容易證明 $p=\Phi\left(Z_0\right)$

示例：對於前面的兩側數值示例，

$$p=2\left(1-\Phi\left(\left|Z_0\right|\right)\right)=2(1-\Phi(2.22))=0.0263$$

7.2.2 測試設計

目標：為已知方差正態均值檢驗設計一個雙邊檢驗，其中我們對 I 型和 II 型錯誤施加以下約束：

$$\mathrm{P}(\text{ Type I error }) \leq \alpha \quad \text{ and } \quad \mathrm{P}\left(\text{ Type II error } \mid \mu=\mu_1>\mu_0\right) \leq \beta$$

通過“設計”，我們僅僅意味著確定我們需要的觀察次數 n，我們需要這些觀察值來滿足 α 的 I 類錯誤界限和 β 的 II 類錯誤界限。

備註：綁定 β 是針對真均值 μ 恰好等於用戶指定值 $\mu=\mu_1>\mu_0$ 的特殊情況。換句話說，我們試圖“保護”自己免受 μ 實際上恰好等於 μ_1 的可能性。

如果我們更改“受保護的”μ_1，我們需要更改 n。一般來說，μ_1 與 μ_0 越接近，我們需要做的工作就越多（即 n 越高）——因為很難區分兩個相近的案例。

定理：假設實際均值與假設均值之差為

$$\delta \equiv \mu-\mu_0=\mu_1-\mu_0.$$

（不失一般性，我們可以假設$\mu_1>\mu_0$。）那麼α和β的設計要求可以通過取一個size的樣本來實現

$$n \doteq \sigma^2\left(z_{\alpha / 2}+z_\beta\right)^2 / \delta^2$$

備註：在接下來的證明中，我們將得到一個 β 的表達式，它涉及在包含 n 的混亂中評估的標準正態 cdf。然後我們將進行反演以獲得 n 的所需近似值。證明：讓我們首先看一下 β 的值。

$$\begin{aligned}
\beta & =\mathrm{P}\left(\text{ Type II error } \mid \mu=\mu_1>\mu_0\right) \\
& =\mathrm{P}\left(\text{ Fail to Reject } H_0 \mid H_0 \text{ false }\left(\mu=\mu_1>\mu_0\right)\right) \\
& =\mathrm{P}\left(\left|Z_0\right| \leq z_{\alpha / 2} \mid \mu=\mu_1\right) \\
& =\mathrm{P}\left(-z_{\alpha / 2} \leq Z_0 \leq z_{\alpha / 2} \mid \mu=\mu_1\right) \\
& =\mathrm{P}\left(-z_{\alpha / 2} \leq \frac{\bar{X}-\mu_0}{\sigma / \sqrt{n}} \leq z_{\alpha / 2} \mid \mu=\mu_1\right) \\
& =\mathrm{P}\left(-z_{\alpha / 2} \leq \frac{\bar{X}-\mu_1}{\sigma / \sqrt{n}}+\frac{\mu_1-\mu_0}{\sigma / \sqrt{n}} \leq z_{\alpha / 2} \mid \mu=\mu_1\right)
\end{aligned}$$

由於 $\mu=\mu_1$ 在這個表達式中，我們將定義

$$Z \equiv \frac{\bar{X}-\mu_1}{\sigma / \sqrt{n}} \sim \operatorname{Nor}(0,1)$$

以便

$$\begin{aligned}
\beta &= \mathrm{P}\left(-z_{\alpha/2} \leq Z + \frac{\sqrt{n}\delta}{\sigma} \leq z_{\alpha/2}\right) \quad (\text{ where } \delta = \mu_1 - \mu_0) \\
&= \mathrm{P}\left(-z_{\alpha/2} - \frac{\sqrt{n}\delta}{\sigma} \leq Z \leq z_{\alpha/2} - \frac{\sqrt{n}\delta}{\sigma}\right) \\
&= \Phi\left(z_{\alpha/2} - \frac{\sqrt{n}\delta}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\sqrt{n}\delta}{\sigma}\right)
\end{aligned}$$

現在，請注意 $-z_{\alpha/2} \ll 0$ 和 $-\sqrt{n}\delta/\sigma < 0(\text{ since } \delta > 0)$。這兩個事實意味著上面 β 表達式中的第二個 $\Phi(\cdot)$ 項是 $\doteq 0$，因此我們只需要使用該表達式中的第一個項。這導致

$$\beta \doteq \Phi\left(z_{\alpha/2} - \frac{\sqrt{n}\delta}{\sigma}\right)$$

當且當

$$\Phi^{-1}(\beta) = -z_\beta \doteq z_{\alpha/2} - \frac{\sqrt{n}\delta}{\sigma}$$

當且當

$$\frac{\sqrt{n}\delta}{\sigma} \doteq z_{\alpha/2} + z_\beta \quad \text{ iff } \quad n \doteq \sigma^2\left(z_{\alpha/2} + z_\beta\right)^2/\delta^2. \quad \text{ Done! Whew! }$$

回顧：如果您想測試 $H_0：\mu = \mu_0$ 與 $H_1：\mu \neq \mu_0$，以及

- 你知道σ^2，
- 你想要 $\mathrm{P}(\text{ Type I error }) = \alpha$，並且
- 你想要$\mathrm{P}(\text{Type II error}) = \beta$，當$\mu = \mu_1\ (\neq \mu_0)$，

那麼你必須採取 $n \doteq \sigma^2\left(z_{\alpha/2} + z_\beta\right)^2/\delta^2$ 觀察。

同樣，如果你在做單面測試，結果證明你需要取 $n \doteq \sigma^2\left(z_\alpha + z_\beta\right)^2/\delta^2$ 觀察。示例：足球運動員的體重正常，$\sigma^2 = 324$。我們希望測試 $H_0 : \mu = 272$ vs. $H_1 : \mu \neq 272$。假設我們想要防止 μ 恰好等於 $\mu_1 = 276$，並且我們想要 $\alpha = 0.05$ 和 $\beta = 0.05$。我們應該進行多少次觀察？

$$n \doteq \frac{\sigma^2}{\delta^2}\left(z_{\alpha/2} + z_\beta\right)^2 = \frac{324}{16}(1.96 + 1.645)^2 = 263.2$$

換句話說，我們需要大約 263 個觀測值。

7.2.3 兩個樣本測試

著眼於比較兩個相互競爭的正常總體的均值，假設我們有以下 X（來自總體 1）和 Y（來自總體 2）的樣本：

$$X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \operatorname{Nor}\left(\mu_x, \sigma_x^2\right) \quad \text{and} \quad Y_1, Y_2, \ldots, Y_m \overset{\text{iid}}{\sim} \operatorname{Nor}\left(\mu_y, \sigma_y^2\right),$$

其中樣本彼此獨立，並且 σ_x^2 和 σ_y^2 是已知的。

一個自然要問的問題是這兩個總體是否具有相同的均值？這是查看均值是否不同的雙邊測試：

$$H_0 : \mu_x = \mu_y \quad \text{vS.} \quad H_1 : \mu_x \neq \mu_y.$$

定義數量

$$Z_0 = \frac{\bar{X} - \bar{Y} - \left(\mu_x - \mu_y\right)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}.$$

如果 H_0 為真（即均值相等），那麼一切都是已知的或可以觀察到的，因此 Z_0 成為檢驗統計量，

$$Z_0 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \sim \operatorname{Nor}(0, 1).$$

因此，和以前一樣，

$\mu_x - \mu_y$ 的雙面檢驗：

$$\text{Reject } H_0 \quad \text{iff} \quad |Z_0| > z_{\alpha/2}.$$

使用更多與以前相同的推理，我們得到以下單邊測試：

$$\begin{gathered} H_0 : \mu_x \leq \mu_y \quad \text{vs.} \quad H_1 : \mu_x > \mu_y \\ \text{Reject } H_0 \quad \text{iff} \quad Z_0 > z_\alpha \\ H_0 : \mu_x \geq \mu_y \quad \text{vs.} \quad H_1 : \mu_x < \mu_y \\ \text{Reject } H_0 \quad \text{iff} \quad Z_0 < -z_\alpha \end{gathered}$$

It's so easy!! ☺

示例：現在是夏天，生活很容易！讓我們比較一下亞特蘭大和巴塞羅那每四天採集的 22 個所謂的 iid 正常溫度讀數。

首先，這是亞特蘭大的讀數。

83	89	82	78	80	94	83	92	76	89	83
90	80	91	88	88	88	79	87	83	86	84

這是巴塞羅那的。

88	82	84	92	95	103	106	88	73	84	91
84	102	91	95	100	96	101	96	81	92	89

我們可以很容易地計算出樣本均值，$\bar{A} = 85.14$ 和 $\bar{B} = 91.50$。此外，歷史證據允許我們假設亞特蘭大的已知方差為 $\sigma_a^2 = 36$，而巴塞羅那的已知方差為 $\sigma_b^2 = 64$。

假設我們要在 \$= 級別測試 $H_0 : \mu_a = \mu_b$ 與 $H_1 : \mu_a \neq \mu_b$ 0.05 \$。那麼檢驗統計量是

$$Z_0 = \frac{85.14 - 91.50}{\sqrt{\frac{36}{22} + \frac{64}{22}}} = -2.98$$

那麼$|Z_0| = 2.98 > z_{\alpha/2} = 1.96$，所以我們拒絕$H_0$。我們得出結論，城市有不同的平均溫度。

7.3 正態均值的假設檢驗（方差未知）

我們重複上一節的工作，除了現在我們處理未知方差的情況。具體來說，在 \$7.3.1。我們測試單個正態分佈的平均值。{7.3.2 的目標是在兩個方差都未知時比較兩個正態分佈的均值。

又到了 t 的時候了！

7.3.1 一個樣本測試

假設 $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Nor}(\mu, \sigma^2)$，其中 σ^2 是未知的 - 這是一個比 §7.2.1 更現實的場景。一個正常人群的雙邊（又名簡單）均值檢驗是

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0.$$

像往常一樣，我們將使用 $\bar{X}$ 來估計 μ。如果 $\bar{X}$ 與 μ_0 “顯著不同”，那麼我們將拒絕 H_0。為了確定“顯著不同”的含義，我們還需要估計 σ^2。為此，定義檢驗統計量

$$T_0 \equiv \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

其中 S^2 是我們心愛的老朋友，樣本方差，

$$S^2 \equiv \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \sim \frac{\sigma^2 \chi^2(n-1)}{n-1}$$

如果 H_0 為真，則

$$T_0 = \frac{\frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}}}{\sqrt{S^2/\sigma^2}} \sim \frac{\text{Nor}(0,1)}{\sqrt{\frac{\chi^2(n-1)}{n-1}}} \sim t(n-1)$$

所以我們有

μ 的雙面測試：

$$\text{Reject } H_0 \quad \text{iff} \quad |T_0| > t_{\alpha/2, n-1},$$

其中表 B.2 給出了分位數。

使用與上一節相同的推理，單邊測試是：

$$\begin{gathered} H_0 : \mu \leq \mu_0 \quad \text{vs.} \quad H_1 : \mu > \mu_0 \\ \text{Reject } H_0 \quad \text{iff} \quad T_0 > t_{\alpha, n-1} \\ H_0 : \mu \geq \mu_0 \quad \text{vs.} \quad H_1 : \mu < \mu_0 \\ \text{Reject } H_0 \quad \text{iff} \quad T_0 < -t_{\alpha, n-1} \end{gathered}$$

示例：考慮以下數據集，該數據集由 40 天內客戶在當地銀行的平均每日等待時間組成。由於各種原因，我們可以將這些每日平均值視為獨立同分佈的正常數據，儘管其均值和方差未知。

6.0	4.9	5.8	4.7	4.9	6.5	4.5	6.6	4.8	6.7
5.1	5.0	5.5	4.6	5.8	4.9	5.7	6.3	4.6	2.7
4.9	4.6	5.7	5.8	4.1	5.0	6.7	5.3	5.4	4.1
4.0	5.7	7.6	4.7	4.1	5.8	5.0	3.7	5.8	3.1

一個簡單的計算表明，樣本均值$\bar{X} = 5.168$，樣本方差為$S^2 = 1.003$。

假設我們要測試等待時間的平均值是否為 $\mu_0 = 5.5$。那麼檢驗統計量是

$$T_0 \equiv \frac{\bar{X} - \mu_0}{\sqrt{S^2/n}} = \frac{5.168 - 5.5}{\sqrt{1.003/40}} = -2.097.$$

讓我們在 $\alpha = 0.10$ 級別做一個雙邊測試。由於$|T_0| = 2.097 > t_{\alpha/2,n-1}$ =$t_{0.05,39} = 1.685$，我們拒絕H_0。因此，平均等待時間可能不等於 5.5

回想一下：測試的 p-value 是導致拒絕 H_0 的最小顯著性水平 α。我們可以計算這個具有未知方差的雙邊正態均值檢驗的 p-value。首先，我們拒絕 H_0 iff

$$|T_0| > t_{\alpha/2,n-1} = F_{n-1}^{-1}(1 - \alpha/2),$$

其中 $F_{n-1}(t)$ 是 $t(n-1)$ 分佈的 cdf（而 $F_{n-1}^{-1}(\cdot)$ 是逆分佈）。這種關係成立當且僅當

$$F_{n-1}\left(|T_0|\right) > 1 - \alpha/2 \text{ iff } \alpha > 2\left(1 - F_{n-1}\left(|T_0|\right)\right).$$

所以對於未知方差情況下的雙邊檢驗，我們有

$$p = 2\left(1 - F_{n-1}\left(|T_0|\right)\right).$$

示例：繼續銀行等待時間示例，回想一下我們有 $n = 40$ 個數據點，導致檢驗統計量 $T_0 = -2.097$。那麼p-value就是$2\left(1 - F_{39}(2.097)\right) = 0.0425$，其中Excel函數t。dist 可用於評估 cdf $F_{39}(2.097)$。由於 $p = 0.0425 < 0.05$，我們拒絕了 $\alpha = 0.05$ 級別的原假設

7.3.2 兩個樣本測試

假設我們有以下兩個樣本設置，

$$X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Nor}\left(\mu_x, \sigma_x^2\right) \quad \text{and} \quad Y_1, Y_2, \ldots, Y_m \overset{\text{iid}}{\sim} \text{Nor}\left(\mu_y, \sigma_y^2\right),$$

其中樣本相互獨立，σ_x^2 和 σ_y^2 是未知的。

自然會問：哪個總體的均值更大？在 §6.5 中類似的置信區間討論的同時，我們將研究三種情況，我們將以稍微不同的方式進行攻擊：

1. Pooled-t 檢驗：$\sigma_x^2 = \sigma_y^2 = \sigma^2$，比如說。

2. 近似（Welch）t 檢驗：$\sigma_x^2 \neq \sigma_y^2$。

3. 配對-t 測試：(X_i, Y_i) 觀察配對。

7.3.2.1 $7.3.2 Pooled-t 測試

讓我們從兩個未知方差恰好相等的偶然情況開始，即假設 ${x}^{2}={y}$^{\{2\}=}$\{2\}$ $，因此 σ^2 是常見的未知方差。

考慮雙邊檢驗以查看均值是否不同：

$$H_0 : \mu_x = \mu_y \quad \text{vs.} \quad H_1 : \mu_x \neq \mu_y.$$

為了進行測試，我們首先計算兩個總體的樣本均值和方差，

$$\bar{X} \equiv \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } \bar{Y} \equiv \frac{1}{m}\sum_{i=1}^{m} Y_i$$
$$S_x^2 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1} \text{ and } S_y^2 = \frac{\sum_{i=1}^{m}\left(Y_i - \bar{Y}\right)^2}{m-1}.$$

與第 6 章一樣，我們將兩個樣本的方差組合起來得到合併方差估計量，

$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}$$

它比單獨的任何一個成分都具有更多的自由度。事實上，如果 H_0 為真，則可以證明

$$S_p^2 \sim \frac{\sigma^2\chi^2(n+m-2)}{n+m-2}$$

然後是檢驗統計量

$$T_0 \equiv \frac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n+m-2)$$

因此，我們有一個匯集測試的集合。

$\mu_x - \mu_y$ 的匯總測試：

示例：戴爾的農夫聲稱紅色美味蘋果的平均重量小於金色美味蘋果的平均重量。因此，他想測試 $H_0 : \mu_{\mathrm{r}} \geq \mu_{\mathrm{g}}$ vs. $H_1 : \mu_{\mathrm{r}} < \mu_{\mathrm{g}}$。

假設他有以下數據（單位為克）：

$$\begin{aligned} n = 25, \quad \bar{R} = 106.5, \quad S_{\mathrm{r}}^2 = 6.2 \\ m = 25, \quad \bar{G} = 112.9, \quad S_{\mathrm{g}}^2 = 5.6 \end{aligned}$$

我們看到 S_{r}^2 非常接近 S_{g}^2，所以我們假設 $\sigma_{\mathrm{r}}^2 \doteq \sigma_{\mathrm{g}}^2$。這證明了使用合併方差估計器的合理性，

$$S_p^2 = \frac{(n-1)S_{\mathrm{r}}^2 + (m-1)S_{\mathrm{g}}^2}{n+m-2} = \frac{24(6.2) + 24(5.6)}{48} = 5.9$$

以便

$$T_0 = \frac{\bar{R} - \bar{G}}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{-6.4}{\sqrt{5.9}\sqrt{\frac{1}{25} + \frac{1}{25}}} = -9.32$$

讓我們在 $\alpha = 0.05$ 級別進行測試。那麼$t_{\alpha,n+m-2} = t_{0.05,48} = 1.677$，由於$T_0 < -t_{\alpha,n+m-2}$，我們拒絕 H_0（我們甚至不需要為此創建一個表格，因為 T_0 是如此消極）。換句話說，我們得出的結論是，金色的平均重量比紅色的重，正如農民所想的那樣。

7.3.2.2 \$7.3.2 近似 t 測試

所有的方差通常不是相等的，所以假設 $\sigma_x^2 \neq \sigma_y^2$，兩者都是未知的。與我們使用 Welch 假設檢驗的工作一樣，從 6.5.2 定義

$$T_0^\star \equiv \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \approx t(\nu) \quad (\text{ if } H_0 \text{ true })$$

其中近似自由度由下式給出

$$\nu \equiv \frac{\left(\frac{S_x^2}{n} + \frac{S_y^2}{m}\right)^2}{\frac{\left(S_x^2/n\right)^2}{n-1} + \frac{\left(S_y^2/m\right)^2}{m-1}}.$$

下表總結瞭如何進行各種雙面和單面測試。

7.3.2.3 $\mu_x - \mu_y$ 的近似（韋爾奇）測試：

two-sided	$H_0 : \mu_x = \mu_y$ $H_1 : \mu_x \neq \mu_y$	Reject H_0 iff $\lvert T_0^\star\rvert > t_{\alpha/2,\nu}$
one-sided	$H_0 : \mu_x \leq \mu_y$ $H_1 : \mu_x > \mu_y$	Reject H_0 iff $T_0^\star > t_{\alpha,\nu}$
one-sided	$H_0 : \mu_x \geq \mu_y$ $H_1 : \mu_x < \mu_y$	Reject H_0 iff $T_0^\star < -t_{\alpha,\nu}$

示例：假設我們正在模擬兩種不同的排隊策略。例如，考慮一條長線饋入多個並行服務器（如郵局）與幾條短線饋入特定服務器（如雜貨店）。我們模擬了第一個策略的 $n = 25$ 次重複，並獲得了 25 個近似 iid 的平均等待正常實現。同樣，我們模擬了第二個策略的 $m = 35$ iid 試驗。這是等待時間的摘要數據（以秒為單位）。

請注意，這些樣本方差表明真實方差不相等。讓我們在 $\alpha = 0.05$ 級別測試兩個排隊策略導致相同的預期等待時間的假設。

首先，我們需要計算近似的自由度，

$$\nu = \frac{\left(\frac{S_x^2}{n} + \frac{S_y^2}{m}\right)^2}{\frac{(S_x^2/n)^2}{n-1} + \frac{(S_y^2/m)^2}{m-1}} = \frac{\left(\frac{28.62}{25} + \frac{13.29}{35}\right)^2}{\frac{(28.62/25)^2}{24} + \frac{(13.29/35)^2}{34}} = 39.49 \to 39.$$

那麼檢驗統計量是

$$T_0^\star = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} = \frac{3.4}{\sqrt{\frac{28.62}{25} + \frac{13.29}{35}}} = 2.754.$$

由於$|T_0^\star| \le t_{\alpha/2,\nu} = t_{0.025,39} = 2.023$，我們拒絕 H_0。因此，我們得出結論，這兩個策略的均值不同——策略 2 的平均值較小。

7.3.2.4 $7.3.2 配對-$t$ 測試

再次考慮兩個相互競爭的正常人群。將 §6.5.3 中的置信區間討論配對，假設我們成對收集來自兩個總體的觀察結果。特別地，不同對之間的隨機變量是獨立的。同一對中的兩個觀察值可能不是獨立的——事實上，它們之間的正相關通常是好的（如 §6.5.3 中所述！

示例：考慮一項涉及成對雙胞胎的調查；對於每組，一個雙胞胎服用一種新藥，另一個服用安慰劑。

$$\text{independent} \begin{cases} \text{Pair 1:} & (X_1, Y_1) \\ \text{Pair 2:} & (X_2, Y_2) \\ \vdots & \vdots \\ \text{Pair } n: & \underbrace{(X_n, Y_n)}_{\text{not independent}} \end{cases}$$

在討論假設檢驗之前，我們將回顧一些符號。首先，定義成對差異，

$$D_i \equiv X_i - Y_i, \quad i = 1, 2, \ldots, n.$$

注意 $D_1, D_2, \ldots, D_n \stackrel{\text{iid}}{\sim} \text{Nor}\big(\mu_d, \sigma_d^2\big)$，其中

$$\mu_d \equiv \mu_x - \mu_y \quad \text{and} \quad \sigma_d^2 \equiv \sigma_x^2 + \sigma_y^2 - 2\,\text{Cov}(X_i, Y_i)$$

此外，定義差異的樣本均值和方差，

$$\bar{D} \equiv \sum_{i=1}^n D_i/n \sim \text{Nor}\big(\mu_d, \sigma_d^2/n\big)$$
$$S_d^2 \equiv \frac{1}{n-1}\sum_{i=1}^n \big(D_i - \bar{D}\big)^2 \sim \frac{\sigma_d^2 \chi^2(n-1)}{n-1}.$$

如果我們正在檢驗均值的相等性，即H_0：$\mu_d = \mu_x - \mu_y = 0$，那麼相關的檢驗統計量是

$$T_0 \equiv \frac{\bar{D}}{\sqrt{S_d^2/n}} \sim t(n-1).$$

使用與 $7.3.1 的未知方差的單樣本正態均值問題完全相同的操作，我們得到以下決策規則。

$\mu_x - \mu_y$ 的配對測試：

示例：考慮我們上次在 6.5.3 訪問的示例，在該示例中，我們對人們接受培訓前後的問題解決時間感興趣。

雖然這五個人是獨立的，但同一個人解決問題的前後時間幾乎可以肯定是正相關的。

讓我們假設所有時間都是正常的。我們將測試 H_0：$\mu_b = \mu_a$，級別為 $\alpha = 0.10$。我們看到 $n = 5$、$\bar{D} = 9$ 和 $S_d^2 = 42.5$。這給出了 $|T_0| = |\sqrt{n}\bar{D}/S_d| = 3.087$。同時，$t_{0.05,4} = 2.13$，所以我們拒絕$H_0$。

因此，我們得出結論 $\mu_b \neq \mu_a$（訓練可能有幫助）。

7.4 其他參數測試的雜燴

本節討論除均值之外的各種參數檢驗。

- 正態分佈的方差 σ^2 ($7.4.1.
- 兩個法線 ($7.4.2) 的方差比 σ_x^2/σ_y^2。
- 伯努利成功參數 p($7.4.3
- 成功參數的差異，$p_x - p_y$，來自兩個伯努利（7.4.4）。

7.4.1 正態方差檢驗

本小節討論對正態分佈的觀測方差的假設檢驗。為此，假設 $X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Nor}(\mu, \sigma^2)$，其中 μ 和 σ^2 是未知的。我們首先考慮雙邊檢驗（您指定假設的 σ_0^2)，

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{vs. } H_1 : \sigma^2 \neq \sigma_0^2.$$

回想一下（再次）樣本方差是

$$S^2 \equiv \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \sim \frac{\sigma^2 \chi^2(n-1)}{n-1}$$

我們將使用檢驗統計量

$$\chi_0^2 \equiv \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1) \quad (\text{if } H_0 \text{ is true})$$

因此，我們立即進行了以下一組測試。

測試 σ^2 ：

two-sided	$H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 \neq \sigma_0^2$	Reject H_0 iff $\chi_0^2 < \chi_{1-\alpha/2,n-1}^2$ or $\chi_0^2 > \chi_{\alpha/2,n-1}^2$
one-sided	$H_0 : \sigma^2 \leq \sigma_0^2$ $H_1 : \sigma^2 > \sigma_0^2$	Reject H_0 iff $\chi_0^2 > \chi_{\alpha,n-1}^2$
one-sided	$H_0 : \sigma^2 \geq \sigma_0^2$ $H_1 : \sigma^2 < \sigma_0^2$	Reject H_0 iff $\chi_0^2 < \chi_{1-\alpha,n-1}^2$

請注意，我們可以在表 B.3 中查找分位數

示例：假設我們有興趣減少我們在麵包店生產的產品的可變性 - 客戶喜歡一致性！出於這個原因，我們真的很想知道麵包大小的方差是否為 $\leq 4\ \mathrm{cm}^2$。因為我們想毫無疑問地證明可變性很低，所以我們將測試

$$H_0 : \sigma^2 \geq 4 \text{ vS. } H_1 : \sigma^2 < 4.$$

假設我們檢查 $n = 100$ 個麵包，我們發現 $\bar{X} = 32$ cm 和 $S^2 = 3.2\ \mathrm{cm}^2$. 然後檢驗統計量

$$\chi_0^2 = (n-1)S^2/\sigma_0^2 = (99)(3.2)/4 = 79.2$$

（並且不明確依賴於 $\bar{X}$ ）。此外，使用 Excel 例程 chisq. inv $(0.1, 99)$，我們計算 $\chi_{\alpha,n-1}^2 = \chi_{0.90,99}^2 = 81.4$。由於$\chi_0^2 < \chi_{1-\alpha,n-1}^2$，我們拒絕$\mathrm{H}_0$。看起來這種麵包的可變性很低——而且它真的聞起來令人垂涎欲滴！

7.4.2 等方差的兩樣本檢驗

現在我們將測試兩個總體是否具有相同的方差。我們從採樣開始

$$X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \operatorname{Nor}(\mu_x, \sigma_x^2) \quad \text{and} \quad Y_1, Y_2, \ldots, Y_m \stackrel{\text{iid}}{\sim} \operatorname{Nor}(\mu_y, \sigma_y^2),$$

我們假設所有的 X 和 Y 都是獨立的。當然，我們將通過樣本方差 S_x^2 和 S_y^2 來估計方差 σ_x^2 和 σ_y^2 。

為了進行雙面測試

$$H_0 : \sigma_x^2 = \sigma_y^2 \quad \text{vs.} \quad H_1 : \sigma_x^2 \neq \sigma_y^2$$

我們將使用檢驗統計量

$$F_0 \equiv \frac{S_x^2}{S_y^2} \sim F(n-1, m-1) \quad (\text{if } H_0 \text{ true})$$

因此，

測試 σ_x^2/σ_y^2

請注意，我們可以通過 Excel 函數 f 查找分位數。投資 但是如果我們在荒島上，我們可以使用表 B.4 B.7。儘管我們應該記得（參見 \$5.3.4，對於任何適當的 γ, a, b，$F_{1-\gamma,a,b} = 1/F_{\gamma,b,a}$.

示例：假設我們要在 $\alpha = 0.05$ 級別測試兩個過程是否具有相同的方差，即 $H_0 : \sigma_x^2 = \sigma_y^2$ 與 $H_1 : \sigma_x^2 \neq \sigma_y^2$。如果樣本方差的比率“太高”或“太低”，那麼我們將拒絕 H_0。

假設我們有以下數據（請注意，我們並不明確需要樣本均值）：$n = 7$ 個觀測值，其中 $S_x^2 = 17.78;$ $m = 16$，$S_y^2 = 12.04$。那麼 $F_0 = S_x^2/S_y^2 = 1.477$。同時，我們得到臨界點，

$$F_{1-\alpha/2,n-1,m-1} = 1/F_{\alpha/2,m-1,n-1} = 1/F_{0.025,15,6} = 1/5.269 = 0.190,$$

$F_{\alpha/2,n-1,m-1} = F_{0.025,6,15} = 3.415$。

由於 F_0 位於兩個臨界點之間，我們無法拒絕 H_0。

7.4.3 伯努利比例檢驗

我們對本小節的興趣在於檢驗有關伯努利成功參數 p 的假設

7.4.3.1 §7.4.3.1 近似伯努利檢驗

假設 $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bern}(p)$。雙面測試的形式為

$$H_0 : p = p_0 \quad \text{vS. } H_1 : p \neq p_0,$$

我們指定假設的 p_0。

為了進行測試，我們令 $Y = \sum_{i=1}^n X_i \sim \text{Bin}(n, p)$。我們將使用檢驗統計量

$$Z_0 \equiv \frac{Y - np_0}{\sqrt{np_0(1-p_0)}} = \frac{\bar{X} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

如果 H_0 為真（即 $p = p_0$ ），則中心極限定理 (CLT) 意味著

$$Z_0 \approx \text{Nor}(0, 1).$$

使用到目前為止我們已經多次應用的經過驗證的推理，我們得到了以下 p 的測試類別。

7.4.3.2 p : 的近似檢驗

0}$&& \end{tabular}

備註：為了讓 CLT 工作，我們需要 n 大（比如至少 30），並且 $np \geq 5$ 和 $nq \geq 5$ （這樣 p 不會太接近0 或 1)。如果 n 不是很大，我們可能不得不使用二項式表（而不是正常的近似值）。這有點乏味，所以我們不會在這裡討論！

示例：學生正確回答某個測試問題的概率是 p。假設隨機抽取 1000 名學生參加全國性考試，得到 820 個問題的正確答案。讓我們在 $\alpha = 0.01$ 級別測試假設 $H_0 : p \leq 0.8$ vs. $H_1 : p > 0.8$。（近似）檢驗統計量是

$$Z_0 \equiv \frac{Y - np_0}{\sqrt{np_0(1-p_0)}} = \frac{820 - 800}{\sqrt{1000(0.8)(0.2)}} = 1.58$$

由於 $Z_0 = 1.58 \leq z_\alpha = z_{0.01} = 2.33$，我們無法拒絕 H_0。

7.4.3.3 §7.4.3.2 伯努利比例檢驗的樣本量選擇

我們能否設計一個雙邊檢驗 $H_0 : p = p_0$ vs. $H_1 : p \neq p_0$，這樣

$$\mathrm{P}(\text{ Type I error }) = \alpha \text{ and } \mathrm{P}(\text{ Type II error } | p \neq p_0) = \beta\,?$$

也就是說，當我們需要 I 類錯誤界 α 和 II 類概率界 β 時，我們是否可以指定雙邊測試的樣本量？是的！我們現在將證明必要的樣本量是

$$n \doteq \left[\frac{z_{\alpha/2}\sqrt{p_0 q_0} + z_\beta \sqrt{pq}}{p - p_0}\right]^2,$$

其中，為了節省空間，我們讓 $q \equiv 1 - p$ 和 $q_0 \equiv 1 - p_0$。

請注意，n 是未知 p 的函數。在實踐中，我們會選擇一些 $p = p_1$ 並問：如果 p 恰好等於 p_1 而不是 p_0，我們應該進行多少次觀察？因此，我們要防止 p 實際上等於 p_1 的情況。

證明（類似於 \$7.2.2 的正常平均設計證明 : :

$$\begin{aligned}
\beta &= \mathrm{P}(\text{ Type II error }) \\
&= \mathrm{P}(\text{ Fail to Reject } H_0 \mid H_0 \text{ false }) \\
&\doteq \mathrm{P}\left(|Z_0| \leq z_{\alpha/2} \mid p \neq p_0\right) \quad (\text{ by the CLT }) \\
&= \mathrm{P}\left(-z_{\alpha/2} \leq Z_0 \leq z_{\alpha/2} \mid p \neq p_0\right) \\
&= \mathrm{P}\left(-z_{\alpha/2} \leq \frac{Y - np_0}{\sqrt{np_0(1-p_0)}} \leq z_{\alpha/2} \mid p \neq p_0\right) \\
&= \mathrm{P}\left(-z_{\alpha/2}\sqrt{\frac{p_0 q_0}{pq}} \leq \frac{Y - np_0}{\sqrt{npq}} \leq z_{\alpha/2}\sqrt{\frac{p_0 q_0}{pq} \mid p \neq p_0}\right) \\
&= \mathrm{P}\left(-c \leq \frac{Y - np}{\sqrt{npq}} + \frac{n(p - p_0)}{\sqrt{npq}} \leq c \mid p \neq p_0\right)
\end{aligned}$$

在哪裡

$$c \equiv z_{\alpha/2}\sqrt{\frac{p_0 q_0}{pq}}$$

現在註意到

$$Z \equiv \frac{Y - np}{\sqrt{npq}} \approx \mathrm{Nor}(0,1)$$

（因為 p 是真正的成功概率）。這給

$$\begin{aligned}
\beta &\doteq \mathrm{P}\left(-c \leq Z + \frac{n(p - p_0)}{\sqrt{npq}} \leq c\right) \\
&= \mathrm{P}\left(-c - \frac{\sqrt{n}(p - p_0)}{\sqrt{pq}} \leq Z \leq c - \frac{\sqrt{n}(p - p_0)}{\sqrt{pq}}\right) \\
&= \mathrm{P}(-c - d \leq Z \leq c - d) \\
&= \Phi(c - d) - \Phi(-c - d)
\end{aligned}$$

在哪裡

$$d \equiv \frac{\sqrt{n}\,(p-p_0)}{\sqrt{pq}}.$$

還要注意

$$-c-d=-z_{\alpha/2}\sqrt{\frac{p_0q_0}{pq}}-\frac{\sqrt{n}\,(p-p_0)}{\sqrt{pq}} \ll 0.$$

這意味著 $\Phi(-cd) \doteq 0$，因此 $\beta \doteq \Phi(cd)$。因此，

$$-z_\beta \equiv \Phi^{-1}(\beta) \doteq c-d = z_{\alpha/2}\sqrt{\frac{p_0q_0}{pq}}-\frac{\sqrt{n}\,(p-p_0)}{\sqrt{pq}}.$$

經過一點代數，我們終於（！）得到

$$n \doteq \left[\frac{z_{\alpha/2}\sqrt{p_0q_0}+z_\beta\sqrt{pq}}{p-p_0}\right]^2.$$

類似地，相應的單面檢驗的樣本量為

$$n \doteq \left[\frac{z_{\alpha}\sqrt{p_0q_0}+z_\beta\sqrt{pq}}{p-p_0}\right]^2. \text{ Whew!}$$

示例：假設我們正在進行一項研究以確定特定的抗過敏藥物是否有效。我們將假設該藥物對每個獨立受試者明顯有效或無效，因此我們將進行合法的伯努利試驗。我們的假設是 $H_0: p=p_0=0.8$ vs. $H_1: p \neq 0.8$。

為了設計我們的測試（即確定它的樣本量），讓我們將我們的類型 I 錯誤概率設置為 $\alpha=0.05$。我們想防止在性能不佳方面犯錯，所以讓我們將類型 II 設置為錯誤為 $\beta=0.10$，在 $p=p_1=0.7$ 的特殊情況下。然後

$$\begin{aligned} n &\doteq \left[\frac{z_{\alpha/2}\sqrt{p_0q_0}+z_\beta\sqrt{p_1q_1}}{p_1-p_0}\right]^2 \\ &= \left[\frac{1.96\sqrt{(0.8)(0.2)}+1.28\sqrt{(0.7)(0.3)}}{0.7-0.8}\right]^2 \\ &= 187.8 \to 188. \quad \Box \end{aligned}$$

7.4.4 等比例的兩個樣本檢驗

假設 $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bern}(p_x)$ 和$Y_1, Y_2, \ldots, Y_m \overset{\text{iid}}{\sim} \text{Bern}(p_y)$ 是獨立樣本來自兩個相互競爭的伯努利種群。現在我們有興趣檢驗關於成功參數差異的假設，即$p_x - p_y$。

雙面測試的形式為

$$H_0: p_x = p_y \quad \text{vs.} \quad H_1: p_x \neq p_y.$$

為了進行測試，用

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \frac{\mathrm{Bin}(n, p_x)}{n} \quad \text{and} \quad \bar{Y} \sim \frac{\mathrm{Bin}\left(m, p_y\right)}{m}$$

根據中心極限定理和我們從第 6 章開始的置信區間工作，我們知道對於大的 n 和 m，

$$\bar{X} \approx \mathrm{Nor}\left(p_x, \frac{p_x\,(1-p_x)}{n}\right) \quad \text{and} \quad \bar{Y} \approx \mathrm{Nor}\left(p_y, \frac{p_y\left(1-p_y\right)}{m}\right)$$

然後

$$\frac{\bar{X}-\bar{Y}-\left(p_x-p_y\right)}{\sqrt{\frac{px(1-px)}{n}+\frac{py(1-py)}{m}}} \approx \mathrm{Nor}(0,1).$$

此外，在原假設下，我們有 $p \equiv p_x = p_y$，在這種情況下

$$\frac{\bar{X}-\bar{Y}}{\sqrt{p(1-p)\left[\frac{1}{n}+\frac{1}{m}\right]}} \approx \mathrm{Nor}(0,1)$$

當然，上式左邊的p是未知的，但是在H_0下，我們假設$p = p_x = p_y$。因此，我們將通過合併估計器估計 p，

$$\hat{p} \equiv \frac{\sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j}{n+m}.$$

現在將其代入上一個方程，得到最後一個近似值——我們最終可以使用的檢驗統計量，

$$Z_0 \equiv \frac{\bar{X}-\bar{Y}}{\sqrt{\hat{p}(1-\hat{p})\left[\frac{1}{n}+\frac{1}{m}\right]}} \approx \mathrm{Nor}(0,1) \quad (\text{ under } H_0).$$

然後我們立即進行以下一組測試。

$p_x - p_y$ 的近似檢驗：

示例：讓我們根據客戶評論（“美味”或“討厭”）比較兩家餐廳。Burger Fil-A 在 $n = 950$ 的評論中獲得了 550 個美味的評分（+400 的討厭評分），而 McWendy 在 $m = 1025$ 的評論中獲得了 675 個 $（+$350 討厭的評分）的美味評分。我們將在 $\alpha = 0.05$ 級別測試假設

$$H_0 : p_b = p_m \quad \text{Vs.} \quad H_1 : p_b \neq p_m$$

首先，我們計算

$$\bar{B} = \frac{550}{950} = 0.5789, \quad \bar{M} = \frac{675}{1025} = 0.6585, \quad \text{and} \quad \hat{p} = \frac{550+675}{950+1025} = 0.6203.$$

這給了我們

$$Z_0 = \frac{\bar{B} - \bar{M}}{\sqrt{\hat{p}(1-\hat{p})\left[\frac{1}{n} + \frac{1}{m}\right]}} = \frac{0.5789 - 0.6585}{\sqrt{0.6203(0.3797)\left[\frac{1}{950} + \frac{1}{1025}\right]}} = -3.642$$

由於 $|Z_0| > z_{0.025} = 1.96$，我們很容易拒絕 H_0，並非正式地宣布 McWendy 獲勝。

7.5 擬合優度測試

此時，假設我們已經猜到了一個合理的分佈，然後估計了相關的參數。現在我們將進行正式測試，看看我們的辛勞有多成功——換句話說，我們假設的分佈+相關參數是否可以接受？

7.6 1 χ^2 擬合優度檢驗

特別是，我們將測試

$$H_0 : X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{pmf} / \text{pdf}\, f(x) \quad \text{vs.} \quad H_1 : \text{not}\, H_0,$$

在顯著水平

$$\alpha \equiv \mathrm{P}\left(\text{ Reject } H_0 \mid H_0 \text{ true }\right) = \mathrm{P}(\text{ Type I error })$$

這被稱為擬合優度檢驗。擬合優度測試程序具有以下主要步驟：

1. 將 $f(x)$ 的域劃分為 k 個集合，例如 A_1、A_2、 ... 、A_k（如果 X 是離散的或區間，則為不同的點如果 X 是連續的）。
2. 統計落在集合 $A_i, i = 1, 2, \ldots, k$ 中的實際觀測值 O_i 的數量。假設 H_0 為真（即 $f(x)$ 是實際的 pmf/pdf)，定義 $p_i \equiv \ mathrmP(X \in A_i)$, 在這種情況下 $O_i \sim \text{Bin}(n, p_i), i = 1, 2, \ldots, k$。
3. 確定如果 H_0 為真，將落入每個集合的預期觀測數，例如 $E_i = \mathrm{E}\left[O_i\right] = np_i, i = 1, 2, \ldots, k$。4. 根據 E_i 和 O_i 之間的差異計算檢驗統計量。卡方擬合優度檢驗統計量為 1

$$\chi_0^2 \equiv \sum_{i=1}^{k} \frac{\left(O_i - E_i\right)^2}{E_i}$$

5. χ_0^2 的大值表示不合適（所以只做一個單邊測試）。因此，如果 $\chi_0^2 > \chi_{\alpha,k-1-s}^2$，我們拒絕 H_0，其中：

- s 是必須估計的來自 $f(x)$ 的未知參數的數量。例如，如果 $X \sim \text{Nor}\left(\mu, \sigma^2\right)$，則 $s = 2$。
- $\chi_{\alpha,\nu}^2$ 是我們可以在表 B 中找到的 $\chi^2(\nu)$ 分佈的通常 $(1-\alpha)$ 分位數.3 即 $\mathrm{P}\left(\chi^2(\nu) < \chi_{\alpha,\nu}^2\right) = 1 - \alpha$。

如果 $\chi_0^2 \leq \chi_{\alpha,k-1-s}^2$，我們無法拒絕（因此我們勉強接受）$H_0$。

7.6.0.1 備註：

- 通常的建議：為了使 χ^2 擬合優度檢驗起作用，選擇 $n \geq 30$ 和 k 使得 $E_i \geq 5$ 對所有 我。

- 如果自由度 $\nu = k-1-s$ 恰好很大，那麼可以證明

$$\chi^2_{\alpha,\nu} \doteq \nu\left[1-\frac{2}{9\nu}+z_\alpha\sqrt{\frac{2}{9\nu}}\right]^3,$$

其中 z_α 是適當的標準正態分位數。

- 其他具有有趣名稱的擬合優度測試：Kolmogorov-Smirnov、Anderson-Darling、Shapiro-Wilk 等。

7.6.1 初學者示例

在本節中，我們將通過一些簡單的示例來說明 χ^2 擬合優度檢驗的使用。

寶貝示例：測試 H_0：X_i 是統一的 $(0,1)$。假設我們有 $n=10000$ 個觀察值（制服很便宜！），分成 $k=5$ 個等長的區間（所以這些是易於使用的“等概率”區間）。

“為什麼這個測試數據會讓你想起老麥克唐納的農場？我們可以輕鬆計算擬合優度統計量 $\chi^2_0 \equiv \sum_{i=1}^k (O_i-E_i)^2/E_i = 5.511$

現在我們取$\alpha=0.05$，注意沒有未知參數，所以$s=0$。

那麼$\chi^2_{\alpha,k-1-s}=\chi^2_{0.05,4}=9.49$。

由於$\chi^2_0 \le \chi^2_{\alpha,k-1-s}$，我們無法拒絕$H_0$。因此，我們將假設這些數字確實是統一的。

離散示例：要求一個籃球訓練營的 1000 名孩子罰球四次，並記錄成功投籃的次數，我們讓 X_i 表示孩子 i 的投籃次數。

我們將測試 H_0 ：觀察結果是 $\text{Bin}(4,p)$。讓我們從找到 p 的最大似然估計量開始。似然函數是

$$\begin{aligned}
L(p) &= \prod_{i=1}^n f(x_i) \\
&= \prod_{i=1}^n \binom{4}{x_i} p^{x_i}(1-p)^{4-x_i} \\
&= Cp\sum_{i=1}^n x_i(1-p)^{4n-\sum_{i=1}^n x_i}
\end{aligned}$$

where $C=\prod_{i=1}^n \binom{4}{x_i}$ is a constant with respect to p. Thus,

$$\ell\text{n}(L(p)) = \ell\text{n}(C) + \left(\sum_{i=1}^n x_i\right)\ell\text{n}(p) + \left(4n-\sum_{i=1}^n x_i\right)\ell\,\text{n}(1-p)$$

以便

$$\frac{d\ln(L(p))}{dp} = \frac{\sum_{i=1}^n x_i}{p} - \frac{4n-\sum_{i=1}^n x_i}{1-p} = 0$$

經過一點代數，我們得到了 MLE，

$$\hat{p}=\frac{1}{4n}\sum_{i=1}^{n}X_i=\frac{1}{4000}\sum_{i=1}^{1000}X_i=\frac{0(32)+1(131)+\cdots+4(142)}{4000}=0.61375$$

這是有道理的，因為它是所有 $n=1000$ 孩子拍攝的所有照片的總體平均值。

Our aim is to calculate the goodness-of-fit test statistic χ_0^2 associated with the $\operatorname{Bin}(4,\hat{p})$ distribution. To this end, we'll make a little table, assuming that $\hat{p}=0.61375$ is correct. By the Invariance Property of MLEs (this is why we learned it!), the expected number of kids making x shots is
$E_x=n\widehat{\mathrm{P}}(X=x)=n\binom{4}{x}\hat{p}^x(1-\hat{p})^{4-x}, x=0,1,\ldots,4($ assuming $\hat{p}$ is actually $p)$.

因此，檢驗統計量為

$$\chi_0^2=\sum_{x=0}^{4}\frac{(O_x-E_x)^2}{E_x}=\frac{(32-22.26)^2}{22.26}+\frac{(131-141.47)^2}{141.47}+\cdots=5.46$$

讓$k=5$ 表示單元格的數量，讓$s=1$ 表示我們必須估計的參數的數量。假設水平 $\alpha=0.05$。然後我們將 $\chi_0^2=5.46$ 與 $\chi_{\alpha,k-1-s}^2=\chi_{0.05,3}^2=7.81$ 進行比較。

由於$\chi_0^2<\chi_{\alpha,k-1-s}^2$，我們無法拒絕$H_0$。這意味著我們可以將孩子們投籃的次數視為 $\operatorname{Bin}(4,\hat{p})$。

連續分佈：對於連續情況，我們將表示區間 $A_i\equiv(a_{i-1},a_i]$, for $i=1,2,\ldots,k$. 為方便起見，我們選擇 a_i 以確保我們有等概率區間，即

$$p_i=\mathrm{P}\left(X\in A_i\right)=\mathrm{P}\left(a_{i-1}<X\leq a_i\right)=1/k,\text{ for all } i.$$

在這種情況下，對於所有 i，我們立即有 $E_i=n/k$，然後

$$\chi_0^2=\sum_{i=1}^{k}\frac{\left(O_i-(n/k)\right)^2}{n/k}.$$

問題是 a_i 可能取決於未知參數。

示例：假設我們有興趣將分佈擬合到一系列到達間隔時間。它們會是指數級的嗎？

$$H_0: X_1,X_2,\ldots,X_n\stackrel{\text{iid}}{\sim}\operatorname{Exp}(\lambda)$$

讓我們用等概率間隔進行 χ^2 擬合優度檢驗。這相當於選擇 a_i 使得 cdf

$$F\left(a_i\right)=\mathrm{P}\left(X\leq a_i\right)=1-e^{-\lambda a_i}=\frac{i}{k},\quad i=0,1,2,\ldots,k$$

經過一點代數，我們得到

$$a_i=-\frac{1}{\lambda}\ell n\left(1-\frac{i}{k}\right),\quad i=0,1,2,\ldots,k$$

很好，但是 λ 是未知的。所以我們需要估計 $s=1$ 參數。好消息是我們從 §5.2.4 知道 MLE 是 $\widehat{\lambda}=1/\bar{X}$。因此，根據不變性， a_i 的 MLE 是

$$\widehat{a}_i=-\frac{1}{\widehat{\lambda}}\ln\left(1-\frac{i}{k}\right)=-\bar{X}\ln\left(1-\frac{i}{k}\right),\quad i=0,1,2,\ldots,k.$$

示例（續）：考慮以下 $n = 100$ 觀察。

0.9448	0.8332	0.6811	0.9574	0.8351	1.4689	1.1664	0.6908	1.0641	1.0850
0.4271	0.6450	0.5775	0.7418	0.7728	0.8622	0.9768	1.4564	0.7689	1.2720
1.4364	0.3484	0.9015	0.1020	0.7372	0.5493	1.0788	0.8315	0.7200	1.6785
0.4792	1.2647	1.6250	0.5860	0.5293	1.0847	0.6903	0.3253	1.1239	1.2604
0.5390	0.4309	0.7868	0.3558	0.8677	1.0312	1.4873	1.2868	1.1287	0.5592
0.6776	1.3921	0.9592	0.4818	0.5250	1.2473	1.0830	0.4711	0.5856	0.9798
1.3020	0.3465	1.3818	0.6614	0.7619	1.1047	0.4816	0.4628	1.2492	0.7993
1.1733	1.0558	0.6630	0.7472	0.6149	1.2403	1.1319	0.9306	0.3116	0.7711
1.2537	0.8989	1.1378	0.5341	1.1170	0.5455	1.0143	0.9954	0.8152	0.8712
0.7990	1.6576	0.5394	1.0199	0.6693	0.9880	0.9220	0.7691	1.0232	0.5635

我們將使用 $k = 10$ 等概率區間對指數分佈進行擬合優度檢驗，因此對於所有 i，$E_i = n/k = 10$。事實證明，基於 100 次觀察的樣本均值是 $\bar{X} = 0.8778$。然後

$$\widehat{a}_i = -0.8778 \ln(1 - 0.1i), \quad i = 0, 1, 2, \ldots, 10$$

最後，相信我，我們已經確定了 100 個觀測值中的每一個屬於哪個區間，然後我們將它們匯總起來得到 O_i 的...

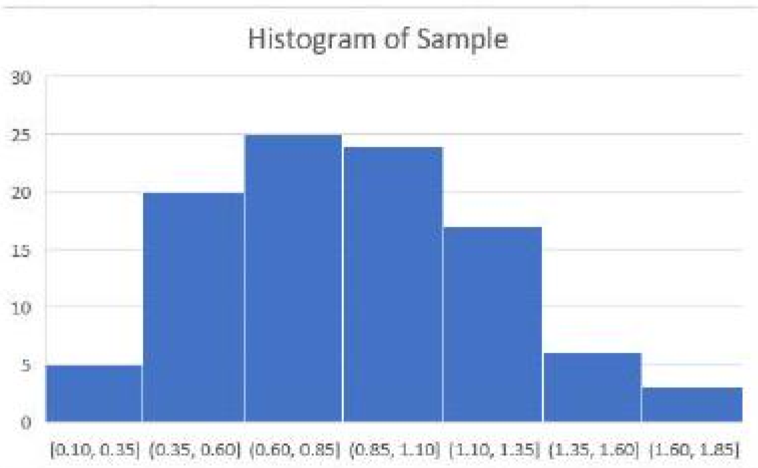
然後

$$\chi_0^2 = \sum_{i=1}^{k} (O_i - E_i)^2 / E_i = 92.2 \quad \text{and} \quad \chi^2_{\alpha,k-1-s} = \chi^2_{0.05,8} = 15.51.$$

所以我們拒絕H_0。這些觀察不是世博會。

7.6.2 小項目

讓我們通過擴展的示例/迷你項目讓事情變得更有趣。我們繼續考慮上一個示例中的 100 個獨立同分佈觀察的相同樣本，但現在有更多細節。該數據集的樣本均值（仍然）為 $X = 0.8778$，樣本標準差為 $S = 0.3347$。繪製數據集通常很有用，這就是這個小傢伙的樣子。.



在本節的其餘部分，我們將進行各種擬合優度檢驗以確定數據可能來自哪些分佈。

$$H_0 : X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} f(x),$$

我們將考慮以下可能性：

- 指數（上面拒絕）。
- Gamma（泛化指數）。
- Weibull（以不同的方式概括指數）。

在每種情況下，我們將數據劃分為 $k=10$ 個等概率區間，並在 $\alpha=0.05$ 級別執行 χ^2 擬合優度檢驗。在此過程中，我們會遇到一些需要處理的有趣問題，但最終都會解決。

7.6.3 指數擬合

帶回不好的記憶，我們在前面的例子中測試了零假設

$$H_0 : X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Exp}(\lambda)$$

回想一下，我們慘遭失敗。但是，回想起來，鑑於以下事實，這是有道理的：

- 該圖看起來並不接近指數。
- $\text{Exp}(\lambda)$ 隨機變量的期望值和標準差均為 $1/\lambda$，但樣本均值 $\bar{X}=0.8778$ 是樣本 $\gg$標準差 $S=0.3347$

這促使我們需要查看其他分佈以尋找良好的數據擬合。

7.6.4 伽瑪擬合

帶有參數 r 和 λ 的 gamma 分佈具有 pdf

$$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, \quad x > 0.$$

請注意，$r=1$ 產生 $\text{Exp}(\lambda)$ 作為特例，因此 gamma 具有更好的擬合數據集的潛力。我們將測試

$$H_0 : X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Gam}(r, \lambda).$$

回到過去的美好時光 ($5.2.4，我們找到了 r 和 λ 的 MLE：

$$\hat{\lambda} = \hat{r}/\bar{X},$$

$\hat{r}$ 解決的地方

$$g(r) \equiv n \ln(r/\bar{x}) - n\psi(r) + \ln\left(\prod_{i=1}^n x_i\right) = 0$$

其中 $\psi(r) \equiv \Gamma'(r)/\Gamma(r)$ 是 digamma 函數。

由於 digamma 有時在通常的軟件包中很難找到，我們將合併近似值

$$\Gamma'(r) \doteq \frac{\Gamma(r+h) - \Gamma(r)}{h} \text{ (for any small } h \text{ of your choosing).}$$

所以我們需要找到解決的$\hat{r}$

$$g(r) \doteq n\ln(r) - n\ln(\bar{x}) - \frac{n}{h}\left(\frac{\Gamma(r+h)}{\Gamma(r)} - 1\right) + \ell n\left(\prod_{i=1}^{n} x_i\right) = 0$$

回想一下$1.2.2，我們有很多方法可以解決這類問題。特別是，二分法是找到任何連續函數 $g(r)$ 的零點的簡單方法。它依賴於中間值定理，該定理指出如果 $g(\ell)g(u) < 0$，則 $r^\star \in [\ell, u]$ 為零。利用這個事實，很容易通過連續二等分來磨練零。

讓我們嘗試對我們的 $n = 100$ 觀察數據集進行二分，我們記得樣本均值是 $\bar{X} = 0.8778$。相信我 $\ell\mathrm{n}\left(\prod_{i=1}^{n} X_i\right) = -21.5623$。此外，我們將近似微分項設置為 $h = 0.01$。然後這是方程式 7.1 簡化為：

$$\begin{aligned} g(r) &\doteq 100\ln(r) - 100\ln(0.8778) - \frac{100}{0.01}\left(\frac{\Gamma(r+0.01)}{\Gamma(r)} - 1\right) - 21.5623 \\ &= 100\ln(r) - \frac{10000\Gamma(r+0.01)}{\Gamma(r)} + 9991.47 = 0. \end{aligned}$$

按照 $1.2.2 中概述的二分算法，我們首先需要通過注意 $g(5) = 0.5506$ 和 $g(7) = -3.0595$ 來初始化搜索。所以那裡有一個零，只是很想找到！該算法在表中得到了充分的描述。

我們看到算法最終給出$\hat{r} = r^\star = 5.2349$，然後$\hat{\lambda} = \hat{r}/\bar{X} = 5.9637.$

所以現在我們終於可以開始我們的 χ^2 擬合優度工作了，注意我們有 $s = 2$ 未知參數。我們將 $n = 100$ 個觀察值劃分為 $k = 10$ 個等概率區間，因此對於所有 i，$E_i = n/k = 10$。區間的（近似）端點由 $\widehat{F}(\hat{a}_i) = i/k, i = 0, 1, 2, \ldots, k$隱式給出，其中 $\widehat{F}(x)$ 是 $\mathrm{Gam}(\hat{r}, \hat{\lambda})$ 分佈的 cdf。

遺憾的是，伽馬分佈的 cdf 沒有封閉形式。但這就是為什麼我們周圍有 Excel（或其朋友），例如，

$$\hat{a}_i = \widehat{F}^{-1}(i/k) = \mathrm{gammainv}(i/k, \hat{r}, \hat{\lambda}),$$

結果如下表。

我們立即發現

$$\chi_0^2 = \sum_{i=1}^{k} (O_i - E_i)^2 / E_i = 6.2 \quad \text{and} \quad \chi^2_{\alpha, k-1-s} = \chi^2_{0.05, 7} = 14.07.$$

所以我們沒有拒絕H_0。這些可能確實是伽馬！

7.7 3,3 威布爾擬合

Weibull 分佈有 cdf $F(x) = 1 - \exp[-(\lambda x)^r]$，對於 $x \geq 0$。注意 $r = 1$ 產生 $\mathrm{Exp}(\lambda)$ 作為一個特例。（另請注意，這裡的 r 和 λ 與前面討論的 gamma 分佈不同。）

讓我們從獲取 $s = 2$ 未知參數 (r 和 λ) 的 MLE 開始。經過一點代數（涉及幾個鏈式規則），pdf是

$$f(x) = \lambda r(\lambda x)^{r-1} e^{-(\lambda x)^r}, \quad x \geq 0.$$

因此，大小為 n 的 iid 樣本的似然函數為

$$L(r, \lambda)=\prod_{i=1}^{n} f\left(x_{i}\right)=\lambda^{n r} r^{n}\left(\prod_{i=1}^{n} x_{i}\right)^{r-1} \exp \left[-\lambda^{r} \sum_{i=1}^{n} x_{i}^{r}\right]$$

以便

$$\ell \mathrm{n}(L)=n r \ln (\lambda)+n \ell \mathrm{n}(r)+(r-1) \ell \mathrm{n}\left(\prod_{i=1}^{n} x_{i}\right)-\lambda^{r} \sum_{i=1}^{n} x_{i}^{r}$$

此時，我們通過設置來最大化 r 和 λ

$$\frac{\partial}{\partial r} \ln (L)=0 \quad \text{and} \quad \frac{\partial}{\partial \lambda} \ln (L)=0.$$

經過更多代數 - 包括 $\frac{d}{dx} c^{x}=c^{x} \ell n(c)$ 的事實 - 我們得到聯立方程

$$\lambda=\left(\frac{1}{n} \sum_{i=1}^{n} x_{i}^{r}\right)^{-1 / r} \text{ and}$$

$$g(r)=\frac{n}{r}+\ln \left(\prod_{i=1}^{n} x_{i}\right)-\frac{n \sum_{i} x_{i}^{r} \ell \mathrm{n}\left(x_{i}\right)}{\sum_{i} x_{i}^{r}}=0$$

λ 的方程看起來很簡單，只要我們能解決 r ！

但我們可以！這次我們用牛頓法。它通常比二分法快得多。這是牛頓從 \$1.2.2 的合理實現

1.初始化$r_{0}=\bar{X} / S$，其中$\bar{X}$是樣本均值，S^{2}是樣本方差。設置 $j \leftarrow 0$

2. 更新$r_{j+1} \leftarrow r_{j}-g\left(r_{j}\right) / g^{\prime}\left(r_{j}\right)$。
3. 如果 $\left|g\left(r_{j+1}\right)\right|$ 或 $\left|r_{j+1}-r_{j}\right|$ 或者你的預算適當小，然後停止並設置 MLE $\hat{r} \leftarrow r_{j+1}$。否則，設 $j \leftarrow j+1$ 並轉到步驟 2

為了使用牛頓，我們需要（在更多代數之後）

$$g^{\prime}(r)=-\frac{n}{r^{2}}-\frac{n \sum_{i} x_{i}^{r}\left[\ell n\left(x_{i}\right)\right]^{2}}{\sum_{i} x_{i}^{r}}+\frac{n\left[\sum_{i} x_{i}^{r} \ell \mathrm{n}\left(x_{i}\right)\right]^{2}}{\left[\sum_{i} x_{i}^{r}\right]^{2}}$$

讓我們在 $n=100$ 觀察的數據集上嘗試 Newton，其中 $r_{0}=\bar{X} / S=0.8778 / 0.3347=2.6227$。這導致下表。

因此，$\hat{r}=r_{5}=2.8607$，因此，

$$\widehat{\lambda}=\left(\frac{1}{n} \sum_{i=1}^{n} X_{i}^{\hat{r}}\right)^{-1 / \hat{r}}=1.0148$$

我們現在可以進行 χ^{2} 等概率區間的擬合優度檢驗。為了得到端點，我們注意到 $F\left(a_{i}\right)=i / k$，然後一些代數樂趣 + MLE Invariance Property yield

$$\hat{a}_{i}=\frac{1}{\widehat{\lambda}}\left[-\ell n\left(1-\frac{i}{k}\right)\right]^{1 / \hat{r}}=0.9854[-\ell n(1-0.1 i)]^{0.3496}, \quad i=0,1,2, \ldots, 10$$

此外，事實證明（見下表）

$$\chi_0^2 = \sum_{i=1}^{k} (O_i - E_i)^2 / E_i = 5.0 \quad \text{and} \quad \chi^2_{\alpha,k-1-s} = \chi^2_{0.05,7} = 14.07$$

所以我們不能拒絕H_0，我們會勉強承認這些觀察是威布爾。

大啟示：我實際上是根據 Weibull 分佈生成觀察結果，參數 $r = 3$ 和 $\lambda = 1$。所以我們做得很好！

7.8 練習

1. ($7.1) 對還是錯？如果我們有統計上顯著的證據表明它是錯誤的，我們拒絕原假設。

2. ($7.1) 數量 α 被稱為

a. P（第一類錯誤）
b. P（II 型錯誤）
c. 重要程度
d. P (Reject H_0 | H_0 為真)

3. (7.1) $1-\beta$ 在典型的假設檢驗中代表什麼？

a. 權力
b. P（II 型錯誤）
c. P（第一類錯誤）
d. 1 - 置信水平
e. P (未能拒絕 H_1 | H_1 為真)

4. ($7.1) 假設一家藥品製造商發布了一種實際上不如其當前品牌有效的新藥。公司剛剛犯了什麼樣的錯誤——第一類還是第二類？

5. ($7.2.1 假設您進行假設檢驗，最終得到一個非常接近於零的 p-value。您會怎麼做？

a. 完全恐慌！
b. 拒絕 H_0。
c. 未能拒絕 H_0。
d. 獲取更多數據。
e. 適當增加/減少β。

6. ($7.2.1 TRUE 還是 FALSE？測試的 p-value 總是大於 Type I 錯誤 α。

7. ($7.2.1 亞特蘭大 8 月的平均氣溫正在研究中。過去五年導致了以下 8 月平均氣溫：81.6, $78.8, 80.8, 79.9, 81.3$。以前的經驗使我們能夠假設這些平均值是正常的，方差 $\sigma^2 = 5$ 度 ^{2}。有理由相信真正的月平均溫度低於 80° 嗎？也就是說，測試H_0：$\mu \geq 80$ 與 H_1：$\mu < 80$。使用 $\alpha = 0.025$。

8. ($7.2.1) 出於認證目的，某組學生的平均考試成績必須“可證明”大於 60 。為此，隨機測試了四名學生，他們的樣本平均測試分數為 61。我們假設分數是 iid normal，已知標準差為 3 。(a) 基於這個樣本，我們是否可以得出平均分數為 > 60，顯著性水平 $\alpha = 0.05$ 的結論？

b. 如果真實平均分是 65，拒絕 $H_0 : \mu \leq 60$ 的概率是多少？

9. ($7.2.2) 考慮一系列具有未知均值 μ 和已知方差 $^{2 的獨立同分佈正態觀測 $X_1, X_2, \ldots, X_n$ }=100 $。假設我們正在考慮單邊測試 $H_0 : \mu \geq 90$ vs. $H_1 : \mu < 90$，級別 $\alpha = 0.05$。如果我們希望當 μ 恰好等於 87 時 II 類錯誤的概率為 0.10，我們應該進行多少次觀察？

10. ($7.2.3) 喬和山姆做披薩！假設他們的比薩餅皮的厚度是正態分佈的，（已知的）標準偏差分別為 $\sigma_{\mathrm{J}} = 0.10$ 和 $\sigma_{\mathrm{S}} = 0.15$ 英寸. 以下是每位廚師製作的比薩餅厚度的隨機樣本（以英寸為單位）。

Joe 和 Sam 生產的比薩餅的平均厚度是否相同？使用 $\alpha = 0.05$。

11. ($7.3.1) 假設某個製造中心的每周利潤是 iid 正態分佈，均值未知 μ，方差未知 σ^2。在我們進行十次觀察後，我們發現樣本均值 $\bar{X} = 100$，樣本標準差 $S = 50$。測試看看平均值是否真的高於 90 ，即測試 $H_0 : \mu \leq 90$ vs. $H_1 : \mu > 90$。使用 $\alpha = 0.025$。

12. ($7.3.2 TRUE 或 FALSE？當您比較兩個總體時，如果您認為兩個總體的方差大致相等，則可以使用合併方差估計量。

13. ($7.3.2) 假設我們要比較兩個正態總體的均值，這兩個正態總體都有未知的方差。我們從第一個總體中獲取 $n = 6$ 的觀測值，發現樣本均值和样本方差分別為 ${X}=$50 和 $S_x^2 = 120$。我們從第二個總體中獲取 $m = 5$ 的觀測值，並獲得 $\bar{Y} = 75$ 和 $S_y^2 = 100$ 的樣本均值和样本方差。因為樣本方差非常接近，所以繼續使用合併方差估計器 S_p^2。用 $\alpha = 0.10$ 檢驗 $\mu_x = \mu_y$ 的假設。

14. ($7.3.2) 假設 $m = 10$ 男性和 $w = 10$ 女性參加某項績效考核，得到的樣本統計量 $\bar{M} = 105, S_{\mathrm{m}}^2 = 36$，$\bar{W} = 120$，$S_{\mathrm{w}}^2 = 100$。查看樣本方差，我們必須假設兩個總體的方差不相等。現在檢驗男性和女性在考試中表現大致相同的零假設。使用 $\alpha = 0.05$。

15. ($7.3.2 假設我們進行一個實驗來測試人們是否可以用右手或左手將球扔得更遠。我們讓 20 個人做這個實驗。每個人用右手扔一個球，用左手扔一個球（以隨機順序），然後我們測量距離。如果我們有興趣確定右手拋球的預期長度是否大於左手拋球的預期長度，我們可能會使用哪種類型的假設檢驗？（ a）z（正態）差異假設檢驗

b. pooled-t 差異假設檢驗

c. 配對-t 差異假設檢驗

d. χ^2 差異假設檢驗

e. F 差異假設檢驗

16. ($7.3.2 TRUE or FALSE？pooled-t 檢驗的自由度通常比成對的t 檢驗少。

17. ($7.3.2 我們正在研究解決兩個初等數學問題所需的時間。假設我們要求四個學生同時嘗試問題 A 和問題 B（以某種隨機順序）。假設所有分數都是正常的，並且孩子們是獨立的。結果如下所示（以秒為單位）。

讓我們看看我們是否可以毫無疑問地證明 B 比 A 更難（平均花費更多時間）。具體來說，測試 $H_0 : \mu_{\mathrm{A}} \geq \mu_{\mathrm{B}}$ vs. H_1 : $\mu_{\mathrm{A}} < \mu_{\mathrm{B}}$。使用 $\alpha = 0.05$

18. (7.4.1) 某個人口中的人的身高應該有一個2英寸的標準偏差。但是，我們剛剛隨機抽取了 $n=11$ 人，得到了 4.2 英寸的樣本標準差。這是怎麼回事？使用 $\alpha=0.01$ 來檢驗 $H_0:\sigma\leq 2$ 的說法是否合理。

19. (7.4.2) 假設我們要在0.05 級別測試兩個過程是否具有相同的方差。對於第一個系統，我們有 $n=12$ 個觀測值，$\bar{X}=-3.87$ 和 $S_x^2=67.78$；對於第二個系統，$m=8$，$\bar{Y}=32.16$ 和 $S_y^2=12.04$。那麼系統是否具有相等的方差？

20. ($7.4.2 考慮以下從兩個正常人群中抽取的隨機樣本，$20X'$ s 和 $10Y'$ s。

是否有證據表明 X 的方差大於 Y' 的方差？使用 $\alpha=0.05$。

21. ($7.4.3) 在特定人群中隨機選擇的 $n=1000$ 人中，$Y=$133 被發現具有 COVID-19 抗體。檢驗抗體的實際比率最多為10% 的零假設，即$H_0:p\leq 0.1$ vs. $H_1:p>0.1$。使用 $\alpha=0.01$。22. ($7.4.4 讓我們根據雇主對畢業生的評價（贏家或輸家）來比較兩所大學。在參與研究的 500 名科技學院學生中，475 人獲得了贏家評價，25 人獲得了輸家評價。關於另一方面，在賈斯汀比伯歌唱學院的 300 名學生中，只有 150 人獲得了成功的評價，而 150 人獲得了失敗的評價。（實際上，一些回复的雇主表示遺憾的是，有些 JBSA 學生沒有“完全失敗”的類別。) 儘管我們都知道結果如何，

$$H_0:p_{\mathrm{TST}}=p_{\mathrm{JB}}\quad \text{vS.}\quad H_1:p_{\mathrm{TST}}\neq p_{\mathrm{JB}}.$$

首先，我們計算樣本均值和合併概率估計，

$$\bar{T}=\frac{475}{500}=0.95,\quad \bar{J}=\frac{150}{300}=0.5,\quad \text{and}\quad \hat{p}=\frac{475+150}{500+300}=0.7813$$

這給了我們

$$Z_0=\frac{\bar{T}-\bar{J}}{\sqrt{\hat{p}(1-\hat{p})\left[\frac{1}{n}+\frac{1}{m}\right]}}=\frac{0.95-0.5}{\sqrt{0.7813(0.2187)\left[\frac{1}{500}+\frac{1}{300}\right]}}=14.91$$

由於 $|Z_0|>z_{0.025}=1.96$，我們很容易拒絕 H_0，並宣布 Tech 為贏家，JBSA 為輸家。

23. ($7.5.1) 假設我們觀察 10000 個數字得到以下數據。

進行 χ^2 擬合優度檢驗，看看這些數字是否近似於 $\mathrm{Unif}\,(0,1)$。使用顯著性水平$\alpha=0.05$。

24. ($7.5.2) 假設我們正在進行 χ^2 擬合優度檢驗以確定 $n=200$ 獨立同分佈觀察是否來自 Pearson 分佈，這是一個很酷的分佈具有必須估計的 $s=3$ 參數。假設我已經為您完成了一些工作，並將觀察結果劃分為 $k=5$ 等概率區間，並且已經計算了每個區間中下降的觀察值數量。

如果水平 $\alpha=0.05$，我們接受還是拒絕 Pearson 擬合？25. ($7.5.2) 在我最終上籃之前，我必須罰幾球？這是我從 70 次試驗中收集的數據。

換句話說，我第一次嘗試了 34 次，兩次嘗試了 18 次，等等。讓我們在 $\alpha=0.05$ 級別測試假設 H_0：所需的投擲次數是 $\mathrm{Geom}(p)$。

26. ($7.5.3 回想一下 $\psi(r)\equiv\Gamma'(r)/\Gamma(r)$ 是二伽馬函數。不擇手段地求解 $\psi(r)=0.$ 提示：實際上有無數個這樣的零——最容易找到的是區間$(-1,0)$，所以我建議你在那裡搜索。