

機率和統計回顧

Afshine AMIDI 和 Shervine AMIDI

December 15, 2019

翻譯: kevingo. 審閱: 徐承漢.

幾率與組合數學介紹

□ **樣本空間** – 一個實驗的所有可能結果的集合稱之為這個實驗的樣本空間，記做 S

□ **事件** – 樣本空間的任何子集合 E 被稱之為一個事件。也就是說，一個事件是實驗的可能結果的集合。如果該實驗的結果包含 E ，我們稱我們稱 E 發生

□ **機率公理** – 對於每個事件 E ，我們用 $P(E)$ 表示事件 E 發生的機率

$$(1) \quad 0 \leq P(E) \leq 1 \quad (2) \quad P(S) = 1 \quad (3) \quad P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i)$$

□ **排列** – 排列指的是從 n 個相異的物件中，取出 r 個物件按照固定順序重新安排，這樣安排的數量用 $P(n, r)$ 來表示，定義為：

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

□ **組合** – 組合指的是從 n 個物件中，取出 r 個物件，但不考慮他的順序。這樣組合要考慮的數量用 $C(n, r)$ 來表示，定義為：

$$C(n, r) = \frac{P(n, r)}{r!} = \frac{n!}{r!(n-r)!}$$

注意：對於 $0 \leq r \leq n$ ，我們會有 $P(n, r) \geq C(n, r)$

條件機率

□ **貝氏定理** – 對於事件 A 和 B 滿足 $P(B) > 0$ 時，我們定義如下：

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

注意： $P(A \cap B) = P(A)P(B|A) = P(A|B)P(B)$

□ **分割** – 令 $\{A_i, i \in [1, n]\}$ 對所有的 i ， $A_i \neq \emptyset$ ，我們說 $\{A_i\}$ 是一個分割，當底下成立時：

$$\forall i \neq j, A_i \cap A_j = \emptyset \quad \text{和} \quad \bigcup_{i=1}^n A_i = S$$

注意：對於任何在樣本空間的事件 B 來說， $P(B) = \sum_{i=1}^n P(B|A_i)P(A_i)$

□ **貝氏定理的擴展** – 令 $\{A_i, i \in [1, n]\}$ 為樣本空間的一個分割，我們定義：

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^n P(B|A_i)P(A_i)}$$

□ **獨立** – 當以下條件滿足時，兩個事件 A 和 B 為獨立事件：

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

隨機變數

□ **隨機變數** – 一個隨機變數 X ，它是一個將樣本空間中的每個元素映射到實數域的函數

□ **累積分佈函數(CDF)** – 累積分佈函數 F 是單調遞增的函數，其 $\lim_{x \rightarrow -\infty} F(x) = 0$ 且 $\lim_{x \rightarrow +\infty} F(x) = 1$ ，定義如下：

$$F(x) = P(X \leq x)$$

注意： $P(a < X \leq b) = F(b) - F(a)$

□ **機率密度函數** – 機率密度函數 f 是隨機變數 X 在兩個相鄰的實數值附近取值的機率

□ **機率密度函數和累積分佈函數的關係** – 底下是一些關於離散(D) 和連續(C) 的情況下的重要屬性

| 情況 | 累積分佈函數 F | 機率密度函數 f | 機率密度函數的屬性 |
|-----|---------------------------------------|------------------------|---|
| (D) | $F(x) = \sum_{x_i \leq x} P(X = x_i)$ | $f(x_j) = P(X = x_j)$ | $0 \leq f(x_j) \leq 1$ 和 $\sum_j f(x_j) = 1$ |
| (C) | $F(x) = \int_{-\infty}^x f(y)dy$ | $f(x) = \frac{dF}{dx}$ | $f(x) \geq 0$ 和 $\int_{-\infty}^{+\infty} f(x)dx = 1$ |

□ **變異數** – 隨機變數的變異數通常表示為 $\text{Var}(X)$ 或 σ^2 ，用來衡量一個分佈離散程度的指標。其表示如下：

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

□ **標準差** – 一個隨機變數的標準差通常表示為 σ ，用來衡量一個分佈離散程度的指標，其單位和實際的隨機變數相容，表示如下：

$$\sigma = \sqrt{\text{Var}(X)}$$

□ **分佈的期望值和動差** – 底下是期望值 $E[X]$ 、一般期望值 $E[g(X)]$ 、第 k 個動差和特徵函數 $\psi(\omega)$ 在離散和連續的情況下的表示式：

| 情況 | $E[X]$ | $E[g(X)]$ | $E[X^k]$ | $\psi(\omega)$ |
|-----|--------------------------------------|---|--|--|
| (D) | $\sum_{i=1}^n x_i f(x_i)$ | $\sum_{i=1}^n g(x_i) f(x_i)$ | $\sum_{i=1}^n x_i^k f(x_i)$ | $\sum_{i=1}^n f(x_i) e^{i\omega x_i}$ |
| (C) | $\int_{-\infty}^{+\infty} x f(x) dx$ | $\int_{-\infty}^{+\infty} g(x) f(x) dx$ | $\int_{-\infty}^{+\infty} x^k f(x) dx$ | $\int_{-\infty}^{+\infty} f(x) e^{i\omega x} dx$ |

□ **隨機變數的轉換** – 令變數 X 和 Y 由某個函式連結在一起。我們定義 f_X 和 f_Y 是 X 和 Y 的分佈函式，可以得到：

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|$$

□ **萊布尼茲積分法則** – 令 g 為 x 和 c 的函數， a 和 b 是依賴於 c 的的邊界，我們得到：

$$\frac{\partial}{\partial c} \left(\int_a^b g(x) dx \right) = \frac{\partial b}{\partial c} \cdot g(b) - \frac{\partial a}{\partial c} \cdot g(a) + \int_a^b \frac{\partial g}{\partial c}(x) dx$$

□ **柴比雪夫不等式** – 令 X 是一隨機變數，期望值為 μ 。對於 $k, \sigma > 0$ ，我們有以下不等式：

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

聯合分佈隨機變數

□ **條件密度** – X 對於 Y 的條件密度，通常用 $f_{X|Y}$ 表示如下：

$$f_{X|Y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

□ **獨立** – 當滿足以下條件時，我們稱隨機變數 X 和 Y 互相獨立：

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

□ **邊緣密度和累積分佈** – 從聯合密度機率函數 f_{XY} 中我們可以得到：

| 種類 | 邊緣密度函數 | 累積函數 |
|-----|---|---|
| (D) | $f_X(x_i) = \sum_j f_{XY}(x_i, y_j)$ | $F_{XY}(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} f_{XY}(x_i, y_j)$ |
| (C) | $f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy$ | $F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(x', y') dx' dy'$ |

□ **獨立** – 當滿足以下條件時，我們稱隨機變數 X 和 Y 互相獨立：

$$\psi_{X+Y}(\omega) = \psi_X(\omega) \times \psi_Y(\omega)$$

□ **共變異數** – 我們定義隨機變數 X 和 Y 的共變異數為 σ_{XY}^2 或 $\text{Cov}(X, Y)$ 如下：

$$\text{Cov}(X, Y) \triangleq \sigma_{XY}^2 = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y$$

□ **相關性** – 我們定義 σ_X 、 σ_Y 為 X 和 Y 的標準差，而 X 和 Y 的相關係數 ρ_{XY} 定義如下：

$$\rho_{XY} = \frac{\sigma_{XY}^2}{\sigma_X \sigma_Y}$$

注意一：對於任何隨機變數 X 和 Y 來說， $\rho_{XY} \in [-1, 1]$ 成立

注意二：當 X 和 Y 獨立時， $\rho_{XY} = 0$

□ **主要的分佈** – 底下是我們需要熟悉的幾個主要的不等式：

| 種類 | 分佈 | PDF | $\psi(\omega)$ | $E[X]$ | $\text{Var}(X)$ |
|-----|---|--|--|---------------------|-----------------------|
| (D) | $X \sim \mathcal{B}(n, p)$ Binomial | $P(X = x) = \binom{n}{x} p^x q^{n-x}$ $x \in \llbracket 0, n \rrbracket$ | $(pe^{i\omega} + q)^n$ | np | npq |
| | $X \sim \text{Po}(\mu)$ Poisson | $P(X = x) = \frac{\mu^x}{x!} e^{-\mu}$ $x \in \mathbb{N}$ | $e^{\mu(e^{i\omega} - 1)}$ | μ | μ |
| (C) | $X \sim \mathcal{U}(a, b)$ Uniform | $f(x) = \frac{1}{b-a}$ $x \in [a, b]$ | $\frac{e^{i\omega b} - e^{i\omega a}}{(b-a)i\omega}$ | $\frac{a+b}{2}$ | $\frac{(b-a)^2}{12}$ |
| | $X \sim \mathcal{N}(\mu, \sigma)$ Gaussian | $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ $x \in \mathbb{R}$ | $e^{i\omega\mu - \frac{1}{2}\omega^2\sigma^2}$ | μ | σ^2 |
| | $X \sim \text{Exp}(\lambda)$ Exponential | $f(x) = \lambda e^{-\lambda x}$ $x \in \mathbb{R}_+$ | $\frac{1}{1 - \frac{i\omega}{\lambda}}$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |

參數估計

□ **隨機抽樣** – 隨機抽樣指的是 n 個隨機變數 X_1, \dots, X_n 和 X 獨立且同分佈的集合

□ **估計量** – 估計量是一個資料的函數，用來推斷在統計模型中未知參數的值

□ **偏差** – 一個估計量的偏差 $\hat{\theta}$ 定義為 $\hat{\theta}$ 分佈期望值和真實值之間的差距：

$$\text{Bias}(\hat{\theta}) = E[\hat{\theta}] - \theta$$

注意：當 $E[\hat{\theta}] = \theta$ 時，我們稱為不偏估計量

□ **樣本平均** – 一個隨機樣本的樣本平均是用來預估一個分佈的真實平均 μ ，通常我們用 \bar{X} 來表示，定義如下：

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

注意：當 $E[\bar{X}] = \mu$ 時，則為不偏樣本平均

□ **樣本變異數** – 一個隨機樣本的樣本變異數是用來估計一個分佈的真實變異數 σ^2 ，通常使用 s^2 或 $\hat{\sigma}^2$ 來表示，定義如下：

$$s^2 = \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

注意：當 $E[s^2] = \sigma^2$ 時，稱之為不偏樣本變異數

□ **中央極限定理** – 當我們有一個隨機樣本 X_1, \dots, X_n 滿足一個給定的分佈，其平均數為 μ ，變異數為 σ^2 ，我們有：

$$\bar{X} \underset{n \rightarrow +\infty}{\sim} \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$