機率與統計：從入門到精通

附錄 A　分布與恆等式速查

Distributions and Identities Reference

難度：參考手冊

本附錄是全書符號、常見分布與核心恆等式的速查手冊。建議印出來放在手邊：讀教科書、寫作業、面試前複習、或在鍵盤前調模型時，它都能讓你少翻十次書。每一條目都註明對應章節，方便回頭查閱完整推導。

A.1 符號總表

符號	意義
$\Omega,\ A,\ A^c$	樣本空間、事件、補集（第 1 章）
$\mathbb{P}(A),\ \mathbb{P}(A \mid B)$	機率、條件機率（第 1、2 章）
$X,\ p(x),\ f(x),\ F(x)$	隨機變數、PMF、PDF、CDF（第 3 章）
$\mathbb{E}[X],\ \mathrm{Var}[X],\ \sigma$	期望值、變異數、標準差（第 3 章）
$\mathrm{Cov}(X,Y),\ \mathrm{Corr}(X,Y)$	共變異數、相關係數（第 5 章）
$\boldsymbol{\mu},\ \Sigma$	均值向量、共變異數矩陣（第 5 章）
$\mathcal{N}(\mu,\sigma^2)$	常態（高斯）分布（第 4 章）
$\hat{\theta},\ L(\theta)$	估計量、概似函數（第 7 章）
$p(\theta \mid D)$	後驗分布（第 9 章）
$H(p),\ \mathrm{KL}(p\,\|\,q),\ I(X;Y)$	熵、KL 散度、互資訊（第 12 章）
$X \perp\!\!\!\perp Y$	獨立（第 2、5 章）
$X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim} p$	獨立且同分布（i.i.d.）的樣本
$F,\ \mathrm{SSB},\ \mathrm{SSW}$	F 統計量、組間 / 組內平方和（第 14 章）

表 A.1: 全書符號一覽。

A.2 機率核心恆等式

事件運算（第 1、2 章）

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B), \qquad \mathbb{P}(A^c) = 1 - \mathbb{P}(A),$$

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \qquad \mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\,\mathbb{P}(B).$$

貝氏定理與全機率定理：

$$\mathbb{P}(H \mid E) = \frac{\mathbb{P}(E \mid H)\,\mathbb{P}(H)}{\mathbb{P}(E)}, \qquad \mathbb{P}(E) = \sum_i \mathbb{P}(E \mid H_i)\,\mathbb{P}(H_i).$$

鏈式法則（自迴歸模型的骨架）：

$$\mathbb{P}(x_1, \ldots, x_n) = \prod_{t=1}^{n} \mathbb{P}(x_t \mid x_1, \ldots, x_{t-1}).$$

期望值與變異數（第 3、5 章）

$$\begin{aligned}
\mathbb{E}[aX + bY + c] &= a\,\mathbb{E}[X] + b\,\mathbb{E}[Y] + c \quad \text{(相依亦成立)},\\
\operatorname{Var}[X] &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2, \qquad \operatorname{Var}[aX + b] = a^2 \operatorname{Var}[X],\\
\operatorname{Cov}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y],\\
\operatorname{Var}[X + Y] &= \operatorname{Var}[X] + \operatorname{Var}[Y] + 2\operatorname{Cov}(X, Y) \quad \text{(獨立時交叉項為0)}.
\end{aligned}$$

極限定理（第 6 章）

$$\text{馬可夫：} \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}, \qquad \text{柴比雪夫：} \mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2},$$

$$\text{大數法則：} \bar{X}_n \to \mu, \qquad \text{中央極限定理：} \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \to \mathcal{N}(0, 1),$$

$$\text{Jensen（凸}\varphi\text{）：} \varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)].$$

A.3 常見分布速查

分布之間的關係（第 4 章）：

- n 個獨立 $\mathrm{Ber}(p)$ 之和 $= \mathrm{Bin}(n, p)$。
- $\mathrm{Bin}(n, p)$ 在 $n \to \infty$、$p \to 0$、$np = \lambda$ 固定時 $\to \mathrm{Poi}(\lambda)$；在 n 大時由 CLT $\approx \mathcal{N}(np,\, np(1-p))$。
- 指數分布是幾何分布的連續版本；兩者皆無記憶。
- Beta 是白努利 / 二項成功率的共軛先驗；k 個獨立標準常態的平方和 $\sim \chi^2_k$。

常態分布經驗法則（68–95–99.7）：$\mathbb{P}(|X-\mu| \leq \sigma) \approx 0.683$、$\mathbb{P}(|X-\mu| \leq 2\sigma) \approx 0.954$、$\mathbb{P}(|X-\mu| \leq 3\sigma) \approx 0.997$。常用臨界值：90% 對應 $z = 1.645$、95% 對應 $z = 1.960$、99% 對應 $z = 2.576$。

A.4 統計推論速查

估計（第 7 章）

$$\hat{\theta}_{\mathrm{MLE}} = \arg\max_\theta \sum_i \log p(x_i \mid \theta), \qquad \hat{\theta}_{\mathrm{MAP}} = \arg\max_\theta \Big[\sum_i \log p(x_i \mid \theta) + \log p(\theta)\Big],$$

分布	PMF / PDF	均值	變異數	典型用途
白努利 $\mathrm{Ber}(p)$	$p^x(1-p)^{1-x}$	p	$p(1-p)$	二元標籤
二項 $\mathrm{Bin}(n,p)$	$\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$	成功次數
幾何 $\mathrm{Geom}(p)$	$(1-p)^{k-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	首次成功等待
卜瓦松 $\mathrm{Poi}(\lambda)$	$\frac{\lambda^k e^{-\lambda}}{k!}$	λ	λ	罕見事件計數
均勻 $\mathrm{U}(a,b)$	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	亂數、先驗
指數 $\mathrm{Exp}(\lambda)$	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	等待時間
常態 $\mathcal{N}(\mu,\sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	雜訊、CLT
$\mathrm{Beta}(\alpha,\beta)$	$\propto x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	比率的先驗

表 A.2: 常見分布一覽（詳見第 4 章）。幾何分布此處採「第 k 次試驗首次成功」版本。

$$\mathrm{MSE}(\hat\theta) = \mathrm{Bias}(\hat\theta)^2 + \mathrm{Var}(\hat\theta), \qquad \text{樣本變異數（不偏）：} S^2 = \frac{1}{n-1}\sum_i (x_i - \bar x)^2.$$

區間與檢定（第 8 章）

$$\mathrm{SE}(\bar X) = \frac{s}{\sqrt{n}}, \qquad 95\%\ \text{信賴區間（均值）：}\ \bar X \pm 1.96\,\frac{s}{\sqrt{n}},$$

$$z = \frac{\bar X - \mu_0}{s/\sqrt{n}}, \qquad p\ \text{值} = \mathbb{P}(\text{資料至少此極端} \mid H_0).$$

型一錯誤率 α（誤拒真 H_0）、型二錯誤率 β（漏接真效應）、檢定力 $= 1-\beta$。

變異數分析（第 14 章）

$$\mathrm{SST} = \mathrm{SSB} + \mathrm{SSW}, \qquad F = \frac{\mathrm{MSB}}{\mathrm{MSW}} = \frac{\mathrm{SSB}/(k-1)}{\mathrm{SSW}/(N-k)} \sim F_{k-1,\,N-k}\ (H_0),$$

$$\eta^2 = \frac{\mathrm{SSB}}{\mathrm{SST}}, \qquad k=2 \Rightarrow F = t^2, \qquad \text{Tukey：}\ |\bar y_i - \bar y_j| > q_{\alpha;k,N-k}\sqrt{\mathrm{MSW}/n}.$$

雙因子：$\mathrm{SST} = \mathrm{SS}_A + \mathrm{SS}_B + \mathrm{SS}_{AB} + \mathrm{SSE}$；先檢定交互作用再解讀主效果。

貝氏推論（第 9 章）

$$p(\theta \mid D) \propto p(D \mid \theta)\,p(\theta), \qquad p(x_{\text{new}} \mid D) = \int p(x_{\text{new}} \mid \theta)\,p(\theta \mid D)\,\mathrm{d}\theta.$$

共軛配對：$\mathrm{Beta}(\alpha,\beta)$ + 白努利觀察(k 正, $n-k$ 反) $\Rightarrow \mathrm{Beta}(\alpha+k,\ \beta+n-k)$；高斯先驗 + 高斯概似 $\Rightarrow$ 高斯後驗。

正則化即先驗：L2（權重衰減、ridge）≡ 高斯先驗；L1（lasso）≡ Laplace 先驗。

A.5 資訊理論速查（第 12 章）

$$H(p) = -\sum_x p(x)\log p(x), \qquad H(p,q) = -\sum_x p(x)\log q(x),$$

$$\mathrm{KL}(p\,\|\,q) = H(p,q) - H(p) \geq 0 \quad (p = q \text{ 時取等號}),$$

$$I(X;Y) = \mathrm{KL}\big(p(x,y)\,\|\,p(x)\,p(y)\big) = H(X) - H(X\mid Y) \geq 0.$$

最小化交叉熵 $\iff$ 最大化概似 $\iff$ 最小化 $\mathrm{KL}(p\,\|\,q)$。

A.6 NumPy ／ SciPy 最小參考

任務	呼叫方式
抽樣 / 密度 / 累積機率	`scipy.stats.<dist>.rvs / pdf / cdf`
均值、變異數、標準差	`np.mean`、`np.var`、`np.std`
共變異數 / 相關係數	`np.cov`、`np.corrcoef`
多元常態抽樣	`np.random.multivariate_normal`
t 檢定 / 卡方 / ANOVA	`ttest_ind`、`chi2_contingency`、`f_oneway`、`tukey_hsd`
MLE（數值最佳化 NLL）	`scipy.optimize.minimize`
自助法信賴區間	`np.random.choice(..., replace=True)` 重抽

表 A.3: 最常用的 NumPy / SciPy 呼叫。

```python
import numpy as np
from scipy import stats

# 速查示範：常態分布的三件常用事
x = stats.norm(loc=0, scale=1)
print(x.cdf(1.96) - x.cdf(-1.96))   # 約 0.95：95% 在 ±1.96 內
print(x.ppf(0.975))                 # 約 1.96：反查臨界值
print(x.rvs(size=5, random_state=0))  # 抽 5 個樣本

# Beta-白努利共軛更新（第 9 章）
alpha, beta, k, n = 1, 1, 7, 10
post = stats.beta(alpha + k, beta + n - k)
print(post.mean())                  # (7+1)/(10+2) = 0.667
```

全書完。機率告訴你該期待什麼；統計告訴你該相信什麼。