编程实训作业

陆博福

2025 年 9 月

1 查漏补缺

以下是一些在课堂学习过程中提及的、较为陌生或不熟悉的词汇，在进一步查阅资料学习后记录了笔记。

1.1 最小点覆盖与最大边匹配

边匹配是一个边集合 $M \subseteq E$，其中任意两条边不共享同一个顶点。即：对于任意 $e_1, e_2 \in M$，$e_1 \neq e_2$，且 e_1 和 e_2 没有公共顶点。

点覆盖是一个顶点集合 $C \subseteq V$，使得每一条边 $(u, v) \in E$ 至少有一个端点属于 C。即：对任意边 $(u, v) \in E$，有 $u \in C$ 或 $v \in C$。

在一个二分图中，最大匹配的大小等于最小点覆盖的大小。假设图 $G = (V, E)$ 是一个二分图，其中 M 是一个边匹配，C 是一个点覆盖。

则：最大匹配的大小为 $|M_{\max}|$，最小点覆盖的大小为 $|C_{\min}|$

根据 König 定理：

$$|M_{\max}| = |C_{\min}|$$

应用场景：任务分配、网络流中的最大流与最小割、计算机视觉中的图像分割、电路布线与逻辑门设计、生物信息学中的基因调控网络、社交网络中的影响力传播、在线广告投放策略等。

1.2 匈牙利算法

匈牙利算法（Hungarian Algorithm）是一种用于解决分配问题（Assignment Problem）的算法。它主要用于在一个成本矩阵中找到一种最优的分配方式，使得总成本最小（或最大）。例如，在一个任务分配问题中，每个工人只能分配一个任务，每个任务也只能由一个工人完成，而目标是使总成本最低。

算法的核心是通过一系列操作来寻找**最小权完美匹配**，其步骤大致如下：

构造代价矩阵：将问题转化为一个矩阵，其中元素表示某人做某事的成本或时间。

行减法和列减法：对每一行和每一列减去该行或该列的最小值，使得至少有一行或一列出现 0。

尝试覆盖所有 0：用最少的直线覆盖所有 0，判断是否可以找到一个完美匹配。

调整矩阵：如果不能找到完美匹配，则调整矩阵，重复上述过程，直到找到一个完美匹配为止。

Listing 1: 匈牙利算法伪代码

```
for each row:
    subtract min(row)
for each column:
    subtract min(column)
while lines < n:
    cover zeros with minimum lines
    if lines < n:
        find min uncovered, adjust matrix
find optimal assignment
```

1.3 阿列夫零

阿列夫零（Aleph-null，符号为 $\aleph_0$）是数学中用来表示最小的无限基数的符号，它代表的是可数无限集合的大小。例如，自然数集合 $\{0, 1, 2, 3, ...\}$ 的基数就是阿列夫零。

在集合论中，阿列夫零是第一个阿列夫数，用于描述无限集合的“大小”。与之相对的是更大的无限基数，如阿列夫一（$\aleph_1$）、阿列夫二（$\aleph_2$）等，它们分别对应不可数无限集合的大小。

简单来说，任何与自然数集合具有相同基数的无限集合都是可数无限的，而像实数集合这样的无限集合则是不可数无限的，其基数大于阿列夫零。

1.4 共轭先验

共轭先验（Conjugate Prior）是贝叶斯统计中的一个重要概念，用于简化后验分布的计算。它的基本思想是：**选择一个与似然函数具有相同形式的先验分布**，这样在计算后验分布时，可以保持相同的分布形式，从而避免复杂的积分运算。

如果一个先验分布 $p(\theta)$ 与似然函数 $p(x|\theta)$ 的乘积（即联合分布）在归一化之后，仍然是一个已知的、容易处理的分布形式，那么这个先验分布就被称为**共轭先验**。

换句话说，若：

$$p(\theta|x) \propto p(x|\theta) \cdot p(\theta)$$

且 $p(\theta|x)$ 是某种已知分布（如正态分布、贝塔分布等），那么 $p(\theta)$ 就是 $p(x|\theta)$ 的共轭先验。

计算简便 共轭先验与似然函数属于同一概率分布族，因此后验分布的形式与先验分布相同。这使得后验分布的计算变得简单，无需复杂的数值积分或近似方法。

易于更新 当获得新的数据时，只需对先验参数进行更新，而不需要重新计算整个后验分布。这种特性使得共轭先验非常适合在线学习和动态模型更新。

适合解析求解 在许多情况下，共轭先验允许直接解析地得到后验分布，而不是依赖于蒙特卡洛模拟或其他数值方法。

似然函数	共轭先验	后验分布
$\mathcal{N}(\mu)$	$\mathcal{N}(\mu_0, \sigma_0^2)$	$\mathcal{N}(\mu_n, \sigma_n^2)$
$\mathcal{N}(\sigma^2)$	$\text{Inv-Gamma}(\alpha_0, \beta_0)$	$\text{Inv-Gamma}(\alpha_n, \beta_n)$
$\text{Bernoulli}(p)$	$\text{Beta}(\alpha_0, \beta_0)$	$\text{Beta}(\alpha_n, \beta_n)$
$\text{Binomial}(p)$	$\text{Beta}(\alpha_0, \beta_0)$	$\text{Beta}(\alpha_n, \beta_n)$
$\text{Poisson}(\lambda)$	$\text{Gamma}(\alpha_0, \beta_0)$	$\text{Gamma}(\alpha_n, \beta_n)$
$\text{Gamma}(\alpha)$	$\text{Inv-Gamma}(\alpha_0, \beta_0)$	$\text{Inv-Gamma}(\alpha_n, \beta_n)$

表 1: 常见分布及其共轭先验与后验分布

2 二项分布与泊松二项复合分布的比较分析

现有两种模型，一是二项分布，固定投 100 次硬币，三次实验正面朝上的次数分别为 49、50、51 次，二是先泊松分布再二项分布，投硬币固定计时 2 分钟，平均投 100 次，实验得到的正面朝上次数也是 49、50、51。下面比较两种模型的差异：

2.1 两种模型设定及其期望与方差

2.1.1 单独二项分布

$X \sim B(n,p)$ X 表示在 n 次独立重复试验中成功次数，每次成功的概率为 p。

- 期望：$E(X) = np$
- 方差：$Var(X) = np(1-p)$

2.1.2 先泊松后二项

$X \sim B(N,p), N \sim Poisson(\lambda)$ 设 N 表示在可能进行试验的次数，X 表示 N 次试验成功的次数，每次成功的概率为 p

- 边缘期望：

  $$E(X) = E(E(X|N)) = E(Np) = pE(N) = \lambda p$$

  即 $E(X) = \lambda p$

- 边缘方差：

  $$Var(X) = E(Var(X|N)) + Var(E(X|N))$$

  其中：$E(Var(X|N)) = E(Np(1-p)) = p(1-p)E(N) = p(1-p)\lambda$ $Var(E(X|N)) = Var(Np) = p^2Var(N) = p^2\lambda$

  因此：

  $$Var(X) = p(1-p)\lambda + p^2\lambda = \lambda p$$

即

表 2: 两种模型设定及其期望与方差

模型设定	**分布**	**期望** $E(X)$	**方差 Var**(X)
单独二项分布	$X \sim B(n,p)$	np	$np(1-p)$
先泊松后二项	$X \mid N \sim B(N,p), N \sim \text{Poisson}(\lambda)$	λp	λp

2.2 矩估计

矩估计的核心思想：用样本的矩（如均值、方差等）去估计总体的矩。

2.2.1 二项分布

总体期望（一阶矩）：$E(X_1) = np_1$
总体方差（二阶矩）：$\mathrm{Var}(X_1) = np_1(1-p_1)$

计算样本均值

$$\bar{X}_1 = \frac{1}{n}\sum_{i=1}^{n} X_1 i$$

用样本均值估计总体均值

$$E(X_1) = np_1 = \bar{X}_1$$

解得：

$$\hat{p}_1 = \frac{\bar{X}_1}{n}$$

2.2.2 先泊松再二项

计算样本均值

$$E(X_2) = \sum_{n=0}^{\infty}\sum_{x=0}^{n} x \cdot f_{X_2|N_2}(x|n) \cdot f_{N_2}(n)$$

其中：
$f_{X_2|N_2}(x|n) = \binom{n}{x} p_2^x (1-p_2)^{n-x} \quad (x \le n)$
$f_{N_2}(n) = \frac{\lambda^n e^{-\lambda}}{n!}$
因此：

$$E(X_2) = \sum_{n=0}^{\infty} E(X_2 \mid N_2 = n) \cdot f_{N_2}(n) = \sum_{n=0}^{\infty} np_2 \cdot f_{N_2}(n)$$

$$= p_2 \sum_{n=0}^{\infty} n f_{N_2}(n) = p_2 \cdot E(N_2)$$

因为 $N_2 \sim \mathrm{Poisson}(\lambda)$，所以：

$$E(N_2) = \lambda$$

因此：

$$E(X_2) = p_2 \cdot \lambda$$

用样本均值估计总体均值

$$\bar{X}_2 = p_2 \cdot \lambda$$

解得:

$$\hat{p_2} = \frac{\bar{X}_2}{\lambda}$$

2.3 矩估计的无偏性与有效性

2.3.1 单独二项分布

无偏性

$$E[\hat{p}_1] = E\left[\frac{x_1}{n}\right] = \frac{1}{n}E[x_1] = \frac{1}{n}(np_1) = p_1$$

——无偏（期望等于自身）

有效性

$$\mathrm{Var}(\hat{p}_1) = \mathrm{Var}\left(\frac{x_1}{n}\right) = \frac{1}{n^2} \cdot \mathrm{Var}(x_1)$$

由于 $\mathrm{Var}(X_1) = np_1(1-p_1)$ 因此:

$$\mathrm{Var}(\hat{p}_1) = \frac{p_1(1-p_1)}{n}$$

Fisher 信息量:

$$I(p_1) = \frac{n}{p_1(1-p_1)}$$

Cramér-Rao 下界:

$$\frac{1}{nI(p_1)} = \frac{p_1(1-p_1)}{n}$$

——有效（在 Cramér-Rao 下界下）

2.3.2 先泊松再二项

无偏性

$$E[\hat{p}_2] = E\left[\frac{x_2}{\lambda}\right] = \frac{1}{\lambda}E[x_2] = \frac{1}{\lambda}(p_2\lambda) = p_2$$

——无偏（期望等于自身）

有效性

$$\mathrm{Var}(\hat{p}_2) = \frac{1}{\lambda^2} \cdot \mathrm{Var}(x_2)$$

因为 $\mathrm{Var}(x_2) = p_2\lambda$，所以：

$$\mathrm{Var}(\hat{p}_2) = \frac{p_2\lambda}{\lambda^2} = \frac{p_2}{\lambda}$$

方差比直接对二项分布估计要大 $\frac{n}{(1-p)\lambda}$ 倍（若假设 $\lambda = n$ 则为 $\frac{1}{1-p}$），但仍是有效的

2.4 最大似然估计

最大似然估计的核心思想：找到一组参数值，使得在这组参数下，观测到当前数据的概率（即似然函数）最大。

数学表达 假设我们有一组独立同分布（i.i.d.）的观测数据 $x_1, x_2, \ldots, x_n$，这些数据来自一个概率分布，该分布的概率密度函数（PDF）或概率质量函数（PMF）为 $f(x \mid \theta)$，其中 θ 是我们要估计的未知参数。

似然函数定义为：

$$L(\theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

为了方便计算，通常取对数似然函数：

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i \mid \theta)$$

最大似然估计的目标是找到使对数似然函数最大的参数值 $\hat{\theta}$：

$$\hat{\theta}_{\mathrm{MLE}} = \arg\max_{\theta} \ell(\theta)$$

2.4.1 二项分布

概率质量函数

$$P(X_1 = x) = \binom{n}{x} p_1^x (1-p_1)^{n-x}, \quad x = 0, 1, \ldots, n$$

对数似然函数 假设观测数据为 x_1，则似然函数为：

$$L(p_1) = \binom{n}{x_1} p_1^{x_1} (1-p_1)^{n-x_1}$$

对数似然函数：

$$\ell(p_1) = \log L(p_1) = \log \binom{n}{x_1} + x_1 \log p_1 + (n - x_1) \log(1 - p_1)$$

最大化对数似然函数 对 p_1 求导并令其为零：

$$\frac{d\ell}{dp_1} = \frac{x_1}{p_1} - \frac{n - x_1}{1 - p_1} = 0$$

解得：

$$\frac{x_1}{p_1} = \frac{n - x_1}{1 - p_1} \Rightarrow x_1(1 - p_1) = p_1(n - x_1)$$

$$x_1 - x_1 p_1 = n p_1 - x_1 p_1 \Rightarrow x_1 = n p_1 \Rightarrow p_1 = \frac{x_1}{n}$$

最大似然估计量为：

$$\hat{p}_1 = \frac{x_1}{n}$$

2.4.2 先泊松再二项

联合分布与边缘分布

- 联合分布

$$P(X_2 = x, N = n) = P(N = n) \cdot P(X_2 = x | N = n)$$

$$= e^{-\lambda} \frac{\lambda^n}{n!} \binom{n}{x} p_2^x (1 - p_2)^{n-x}$$

- 边缘分布

$$P(X_2 = x) = \sum_{n=x}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \binom{n}{x} p_2^x (1 - p_2)^{n-x}$$

这个求和可以简化为：

$$P(X_2 = x) = e^{-\lambda} \frac{(\lambda p_2)^x}{x!} \sum_{n=x}^{\infty} \frac{(\lambda(1 - p_2))^{n-x}}{(n - x)!}$$

$$= e^{-\lambda} \frac{(\lambda p_2)^x}{x!} e^{\lambda(1-p_2)} = \frac{(\lambda p_2)^x e^{-\lambda p_2}}{x!}$$

也就是说，尽管我们从泊松分布再加一个二项分布，得到的边缘分布仍然是一个**泊松分布**

似然函数 设观测数据为 x_2，且 $X_2 \sim \text{Poisson}(\lambda p_2)$，那么：

$$P(X_2 = x_2) = \frac{(\lambda p_2)^{x_2} e^{-\lambda p_2}}{x_2!}$$

- 似然函数

$$L(p_2, \lambda) = \prod_{i=1}^{n} \frac{(\lambda p_2)^{x_i} e^{-\lambda p_2}}{x_i!}$$

- 对数似然函数

$$\ell(p_2, \lambda) = \sum_{i=1}^{n} [x_i \log(\lambda p_2) - \lambda p_2 - \log(x_i!)]$$

$$= \sum_{i=1}^{n} x_i \log(\lambda p_2) - n\lambda p_2 - \sum_{i=1}^{n} \log(x_i!)$$

最大化对数似然 将 λ 和 p_2 视为独立参数，分别对它们求偏导。

- 对 λ 求偏导：

$$\frac{\partial \ell}{\partial \lambda} = \sum_{i=1}^{n} \frac{x_i}{\lambda p_2} - np_2 = 0$$

$$\Rightarrow \frac{1}{\lambda p_2} \sum x_i = np_2 \Rightarrow \frac{\bar{x}}{p_2} = np_2 \Rightarrow \bar{x} = np_2^2$$

- 对 p_2 求偏导：

$$\frac{\partial \ell}{\partial p_2} = \sum_{i=1}^{n} \frac{x_i}{p_2} - n\lambda = 0 \Rightarrow \frac{\bar{x}}{p_2} = n\lambda \Rightarrow \bar{x} = n\lambda p_2$$

- 联立两个方程：

  $\bar{x} = np_2^2$ （来自 $\partial/\partial\lambda$）

  $\bar{x} = n\lambda p_2$ （来自 $\partial/\partial p_2$）

$$n\lambda p_2 = np_2^2 \Rightarrow \lambda = p_2$$

  最大似然估计量：

$$\hat{p}_2 = \sqrt{\frac{\bar{x}}{n}}, \quad \text{当}\lambda = p_2$$

2.5 最大似然估计的无偏性与有效性

2.5.1 单独二项分布

$$\hat{p}_1 = \frac{x_1}{n}$$

无偏性

$$E\left[\frac{X_1}{n}\right] = \frac{1}{n}E[X_1] = \frac{1}{n}(np_1) = p_1$$

单独二项分布的最大似然估计是无偏的。

有效性

- 方差计算：
$$\text{Var}\left(\frac{X_1}{n}\right) = \frac{1}{n^2}\cdot\text{Var}(X_1) = \frac{1}{n^2}\cdot np_1(1-p_1) = \frac{p_1(1-p_1)}{n}$$
- Fisher 信息量：对于 $X \sim \text{Binomial}(n,p)$，Fisher 信息量为：
$$I(p) = \frac{n}{p(1-p)}$$
- Cramér-Rao 下界：
$$\frac{1}{nI(p)} = \frac{p(1-p)}{n}$$
  与方差一致。达到 Cramér-Rao 下界，独立二项分布的最大似然估计是有效的估计量。

2.5.2 先泊松再二项

$$\hat{p}_2 = \sqrt{\frac{\bar{x}}{n}}, \quad 假设\lambda = p_2$$

无偏性

$$E[\hat{p}_2] = E\left[\sqrt{\frac{\bar{x}}{n}}\right]$$

由于 $\bar{x} \sim \text{Poisson}(\lambda p_2)$，且 $\hat{p}_2 = \sqrt{\frac{\bar{x}}{n}}$ 是非线性变换，因此：

$$E[\hat{p}_2] \neq \sqrt{\frac{E[\bar{x}]}{n}} = \sqrt{\frac{\lambda p_2}{n}}$$

而真实值是 p_2，显然不等于上式。

结论：先泊松再二项的最大似然估计不是无偏的。

有效性

- Cramér-Rao 下界对于参数 p_2，Fisher 信息量为：

$$I(p_2) = \frac{n}{p_2(1-p_2)}$$

- 实际方差

$$\hat{p}_2^{\mathrm{MLE}} = \sqrt{\frac{\bar{x}}{n}}, \quad \bar{x} \sim \mathrm{Poisson}(\lambda p_2)$$

  用泰勒展开近似期望和方差：

$$E[\hat{p}_2^{\mathrm{MLE}}] \approx \sqrt{\frac{\lambda p_2}{n}} + \frac{1}{2} \cdot \frac{1}{\sqrt{\frac{\lambda p_2}{n}}} \cdot \frac{\lambda p_2}{n} = \sqrt{\frac{\lambda p_2}{n}} + \frac{1}{2}\sqrt{\frac{\lambda p_2}{n}} = \frac{3}{2}\sqrt{\frac{\lambda p_2}{n}}$$

$$\mathrm{Var}(\hat{p}_2^{\mathrm{MLE}}) \approx \frac{1}{4} \cdot \frac{\mathrm{Var}(\bar{x})}{E[\bar{x}]} = \frac{1}{4} \cdot \frac{\lambda p_2}{\lambda p_2/n} = \frac{n}{4}$$

  这与期望不符，说明估计量的方差也偏离了理论下界。

先泊松再二项的最大似然估计在一般情况下不是有效的。

2.6 两种模型的差异比较

- 参数结构差异（固定试验 vs 随机试验次数）：单独二项分布中，试验次数 n 是固定的，参数为成功概率 p；而在“先泊松再二项”模型中，试验次数 $N \sim \mathrm{Poisson}(\lambda)$ 是随机的，因此需要同时估计两个参数 p 和 λ。
- 期望与方差形式差异（如复合模型方差更大）：单独二项分布的期望为 $E[X] = np$，方差为 $\mathrm{Var}(X) = np(1-p)$；在“先泊松再二项”模型中，边缘分布为 $X \sim \mathrm{Poisson}(\lambda p)$，其期望为 $E[X] = \lambda p$，方差为 $\mathrm{Var}(X) = \lambda p$。因此，在相同期望下，复合模型的方差通常更大。
- 应用场景差异（稀有事件计数 vs 抽样调查）：单独二项分布适用于抽样调查、质量检测等固定试验次数的场景；“先泊松再二项”模型更适合描述稀有事件或随机发生次数的场景，例如网络流量、顾客到达、罕见疾病病例等。