1.1 说明伯努利模型的极大似然估计以及贝叶斯估计中的统计学习方法三要素。伯努利模型是定义在取值为 0 与 1 的随机变量上的概率分布。假设观测到伯努利模型 n 次独立的数据生成结果，其中 k 次的结果为 1，这时可以用极大似然估计或贝叶斯估计来估计结果为 1 的概率。

解：设随机变量 X 服从参数为 P 的伯努利分布，P 为待估参数 $(0<P<1)$。随机变量 X 的分布列为：

$$P\{X=x_i\}=P^{x}(1-P)^{1-x},\quad x=0\text{或}1$$

设 $(X_1,X_2,\ldots,X_n)$ 为从总体中取出的容量为 n 的样本，其观测值为：$(x_1,x_2,\ldots,x_n)$，因为 n 次试验中有 k 次的结果为 1，所以 $\sum_{i=1}^{n}x_i=k$。

I：伯努利模型的极大似然估计：

1：模型：伯努利模型

2：策略：经验风险最小化

3：算法：

似然函数为：

$$L(x_1,x_2,\ldots,x_n;P)=P^{k}(1-P)^{n-k}$$

取对数后：

$$\log L=k\log P+(n-k)\log(1-P)$$

对 P 求偏导数：

$$\frac{\partial \log L}{\partial p}=\frac{k-np}{p(1-p)}$$

令 $\frac{\partial \ln L}{\partial p}=0$，解得 $P=\frac{k}{n}$，

因此，伯努利模型参数 P 的极大似然估计值为：

$$P=\frac{k}{n}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

Ⅱ: 伯努力模型的贝叶斯估计

1: 模型: 伯努利模型

2: 策略: 结构风险最小化

3: 算法:

假设参数p服从a和b的Beta分布(Beta分布作为伯努利模型的先验分布), 即: $p \sim Be(a,b)$, 其概率密度函数为:

$$f(x;a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$$

样本的联合分布为:

$$P(D|p) = \prod_{i=1}^{n} P(X_i|p) = p^k(1-p)^{n-k}$$

根据贝叶斯公式:

$$P(p|D) = \frac{P(p)P(D|p)}{P(D)} \propto P(p)P(D|p)$$

所以, ~~参数~~参数p的最大后验概率估计为:

$$\hat{p} = \arg\max_p P(D|p)P(p) = \arg\max_p p^k \cdot (1-p)^{n-k} \cdot p^{a-1} \cdot (1-p)^{b-1}$$

$$= \arg\max_p p^{k+a-1}(1-p)^{n+b-k-1}$$

令: $f(p) = p^{a+k-1}(1-p)^{n+b-k-1}$, $0<p<1$,

则: 当$f'(p)=0$ 时, $f(p)$取得极大值,

解得: 参数p的贝叶斯估计值为:

$$p = \frac{a+k-1}{n+a+b-2}$$

1.2 通过经验风险最小化推导极大似然估计。证明模型是条件概率分布，当损失函数是对数损失函数时，经验风险最小化等价于极大似然估计。

证明:

经验风险最小化即求解:

$$\min_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^{N} L(y_i, f(x_i))$$

当模型是条件概率分布, 损失函数是对数损失函数时, 上式等

价于: $\min_{f \in F} \frac{1}{N} \sum_{i=1}^{N} -\log P(Y|X)$

又$\because N$为常数,

$\therefore$ 上述等价于: $\max_{f \in F} \sum_{i=1}^{N} \log P(Y|X) \Leftrightarrow \max_{f \in F} \log \prod_{i=1}^{N} P(Y|X)$

证毕