

统计学习：第一章

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。

答：

	极大似然估计	贝叶斯估计
模型	$\{P P(0) = p, P(1) = 1 - p, p \in [0, 1]\}$	$\{P P(0) = p, P(1) = 1 - p, p \in [0, 1]\}$
策略	$\operatorname{argmax}_p p^{(n-k)}(1-p)^k$	$P(D p)P(p)$
算法	求导 $p = (n - k)/n$	计算与积分

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。

答：经验风险为：

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

模型是条件概率分布 $P_\theta(Y|X)$ ，当损失函数是对数损失函数 $L(Y|X) = -P(Y|X)$ 时，有

$$R_{\text{emp}} = \min_{\theta} \frac{1}{N} \sum_{i=1}^n -\log P_\theta(y_i|x_i) \quad (2)$$

$$= \max_{\theta} \frac{1}{N} \sum_{i=1}^n \log P_\theta(y_i|x_i) \quad (3)$$

$$= \max_{\theta} \log \prod_{i=1}^n P_\theta(y_i|x_i) \quad (4)$$

$$= \max_{\theta} \prod_{i=1}^n P_\theta(y_i|x_i) \quad (5)$$

。