2.1.0.

对伯努利分布.

$$P(x|\mu) = \mu^x (1-\mu)^{1-x}$$

$E[x] = \mu \quad var(x) = \mu(1-\mu)$

对于二项分布 $P(m|N,\mu) = \binom{N}{m}\mu^m (1-\mu)^{N-m}$

$E(m) = N\mu \quad var(m) = N(1-\mu)\mu.$

伯努利分布中对 $\mu$ 的最大似然估计.

$$\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

∴假设 扔一个硬币3次且均正面朝上.

则 $N=m=3$ ∴ $\mu_{ML}=1$ 最大似然的结果将预测所有未来的观测值.都是正面向上，这是最大似然中过拟合现象的一个例子.

2.1.1

共轭性，选择与概率密度函数形式相似的先验概率分布，那么后验概率分布（正比于似然函数和先验的乘积）就会有着与先验分布 ~~同~~ 相同的函数形式

对 ~~$P(x|\mu) = \mu^x(1-\mu)^{1-x}$~~ 取先验概率分布函数.

$P(m|N,\mu) = \binom{N}{m}\mu^m(1-\mu)^{N-m}$

$$Beta(\mu|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}.$$

其中 $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du.$ 保证归一化.

$E(\mu) = \frac{a}{a+b} \quad var(\mu) = \frac{ab}{(a+b)^2(a+b+1)}$

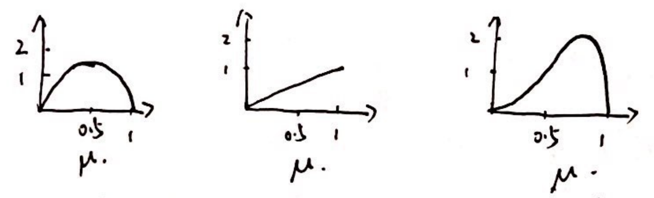
此时 后验概率分布 $= Beta(\mu|a,b)\cdot\binom{N}{m}\mu^m(1-\mu)^{N-m}$

$$P(\mu|m,l,a,b) = \frac{\Gamma(m+a+l+b)}{\Gamma(m+a)\Gamma(l+b)}\mu^{m+a-1}(1-\mu)^{l+b-1}.$$

∴ $a$ 和 $b$ 分别为 $x=1$ 和 $x=0$ 的观测数.（不一定是整数）.

每次取得一个观测值，然后在每次观测值后更新当前的后验分布.（更新方法是让当前的后验分布与新观测值的似然函数相乘，然后归一化，获得新的修正后的后验分布.，即将每次得到的后验分布作为下一次观测的先验分布。

顺序方法学习，可以被用于实时学习场景中.，在实时学习的场景中，输入为一个稳定的持续的数据流，模型必须在观测所有数据之前就进行预测.




$a=2\ b=2$ $N=m=1$ $a=3\ b=2$

贝叶斯的结果和最大似然的结果在 $N\to\infty$ 会统一到一起，对于有限规模的数据集，μ的后验均值总是位于先验均值和μ的最大似然估计之间，随着观测数据增多，后验概率的不确定性会持续下降.