4.基于共轭分布的贝叶斯数据分析

戴俊毅

研究员/长聘副教授

伯努利分布和二项分布

- Bern(p)

$$Pr\{X=z\}=p^z(1-p)^{1-z}, z=0\text{或}1$$

- Bino(n,p)

$$Pr\{X=z\}=\binom{n}{z}p^z(1-p)^{n-z}, z=0,1,\ldots \mathrm{n}$$

- 伯努利分布是二项分布当n=1时的特例
- 在二项分布中，p是参数

贝塔先验分布

- X~beta(α,β)

$$f(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1},0\le x\le 1,\alpha>0,\beta>0$$

- 当关于参数p的先验信念用 $beta(\alpha,\beta)$表达时，对应的概率密度函数为

$$f(p)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1},0\le p\le 1$$

基于贝叶斯定理的信念更新

- 贝叶斯定理

$$P(\theta|D)=\frac{P(\theta)P(D|\theta)}{P(D)}=\frac{P(\theta)P(D|\theta)}{\int P(D|\theta)P(\theta)d\theta}$$

- 在本例中，

$$\theta=p,$$

$$P(\theta)=f(p)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}$$

$$P(D|\theta)=p^{z}(1-p)^{n-z}$$

因此，

$$P(\theta|D) = P(\theta)P(D|\theta)/P(D)$$
$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \times p^{\mathrm{z}}(1-p)^{n-\mathrm{z}}/P(D)$$
$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha+\mathrm{z}-1}(1-p)^{\beta+n-\mathrm{z}-1}/P(D)$$

因为$P(\theta|D)$代表给定D条件下的Θ的概率密度函数，所以

$$1=\int_{-\infty}^{\infty}P(\theta|D)d\theta$$

$$=\frac{1}{P(D)}\int_0^1\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha+z-1}(1-p)^{\beta+n-z-1}dp$$

$$=\frac{1}{P(D)}\frac{\Gamma(\alpha+z)\Gamma(\beta+n-z)}{\Gamma(\alpha+\beta+n)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1\frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+z)\Gamma(\beta+n-z)}p^{\alpha+z-1}(1-p)^{\beta+n-z-1}dp$$

$$=\frac{1}{P(D)}\frac{\Gamma(\alpha+z)\Gamma(\beta+n-z)}{\Gamma(\alpha+\beta+n)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

- 因此，

$$\frac{1}{P(D)}\frac{\Gamma(\alpha+z)\Gamma(\beta+n-z)}{\Gamma(\alpha+\beta+n)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}=1$$

$$P(D)=\frac{\Gamma(\alpha+z)\Gamma(\beta+n-z)}{\Gamma(\alpha+\beta+n)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

$$P(\theta|D)=\frac{1}{P(D)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha+z-1}(1-p)^{\beta+n-z-1}$$

$$=\frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+z)\Gamma(\beta+n-z)}p^{\alpha+z-1}(1-p)^{\beta+n-z-1}$$

也就是说，当p(=θ)的先验分布为beta(α, β)时，它的后验分布也是beta分布，调整后的分布参数为α+z和β+n-z

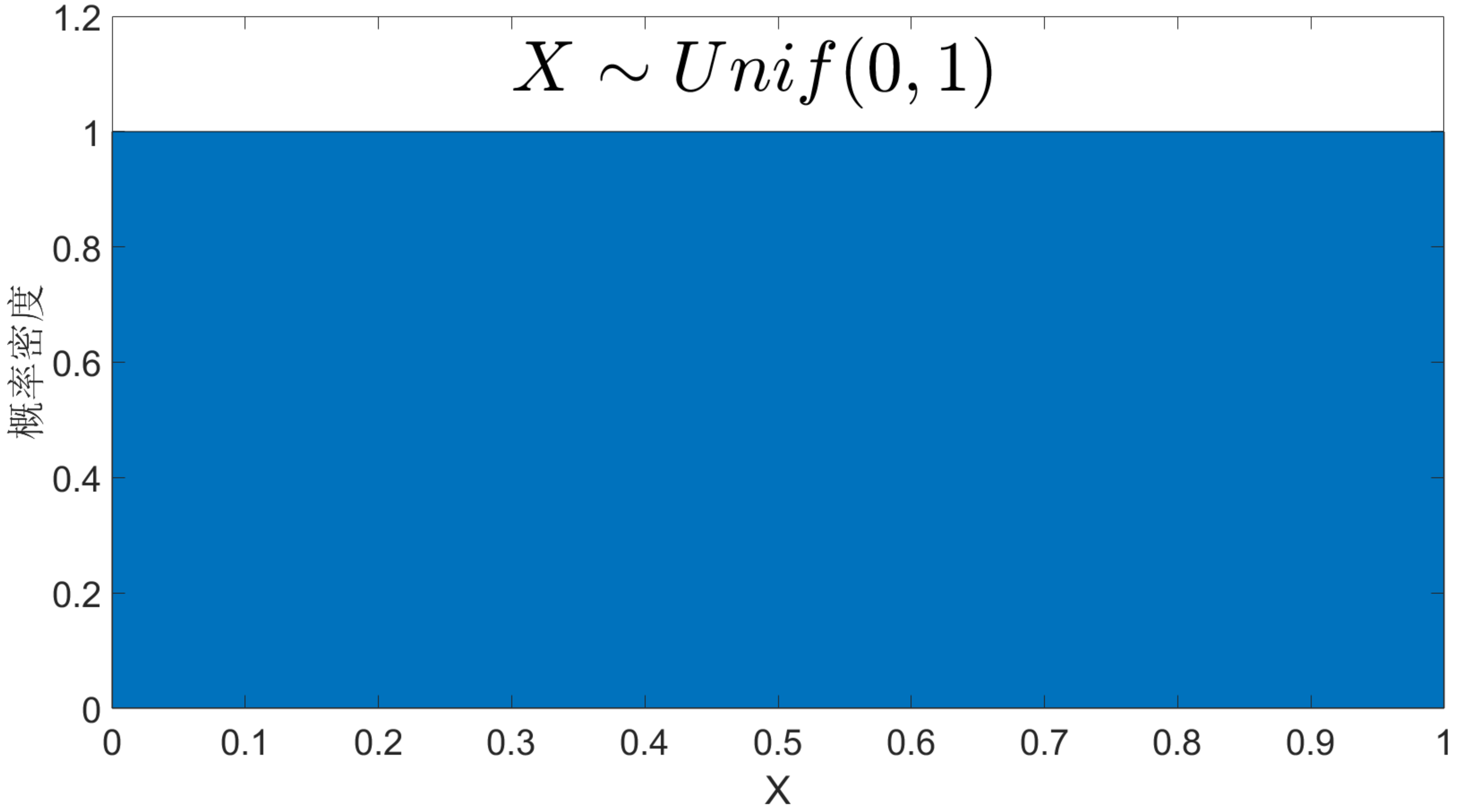
共轭分布

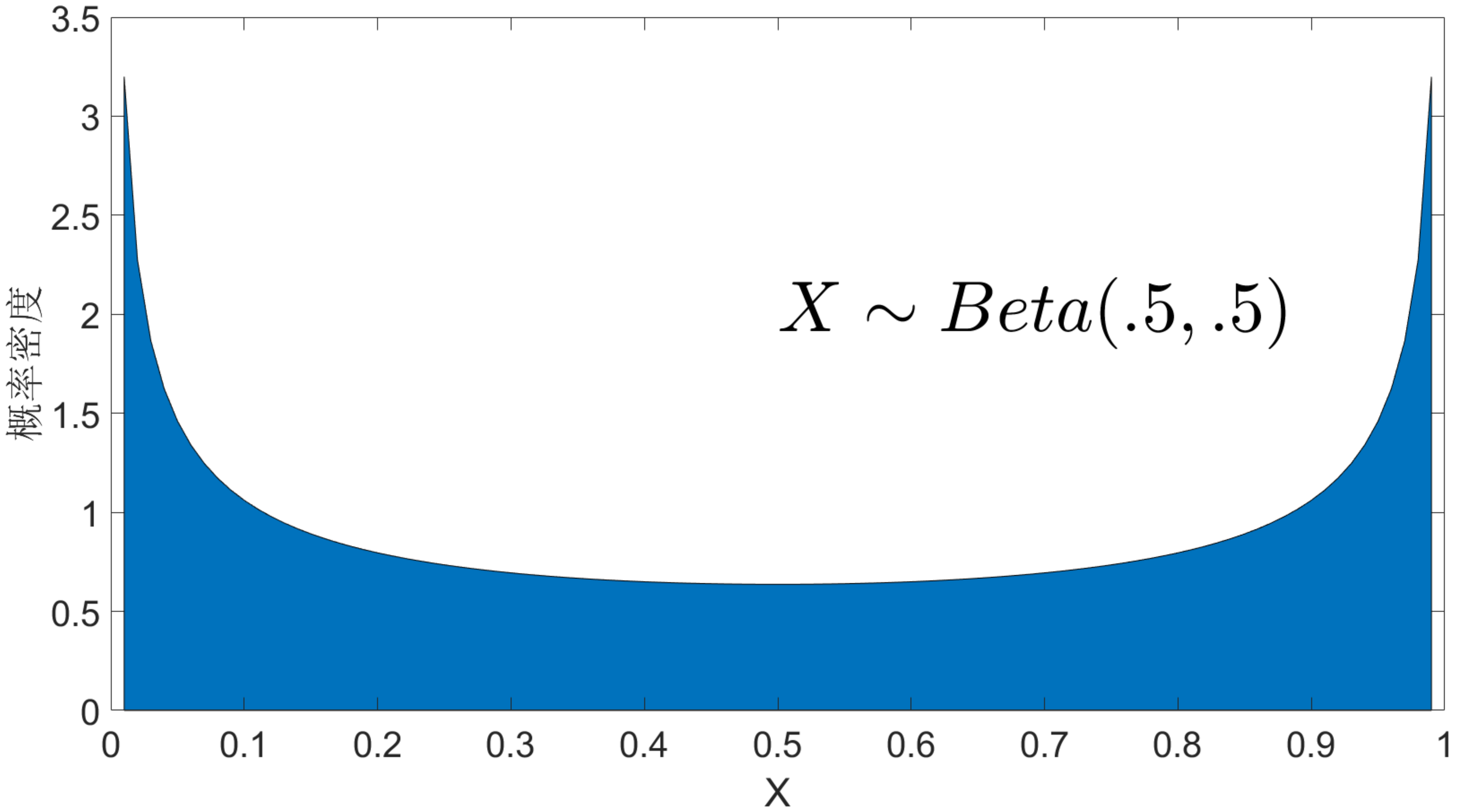
- 当使用贝塔分布作为二项分布参数p的先验分布时，后验分布仍然为贝塔分布，因此，我们称贝塔分布为针对二项分布参数p的共轭分布
- 当先验分布采用对应的共轭分布时，这样的先验分布称为共轭先验分布，对于后验分布亦是如此

无信息共轭先验分布

- 有两种针对二项分布参数的无信息共轭先验

1. 0-1之间的均匀分布，即beta(1,1)
2. Jeffreys先验，即beta(1/2,1/2)


$X \sim Unif(0,1)$
1.2
1
0.8
0.6
0.4
0.2
0
概率密度
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
X



3.5
3
2.5
2
1.5
1
0.5
0
概率密度
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
X
$X \sim Beta(.5, .5)$

共轭后验贝塔分布的调整方式

- 当使用贝塔分布作为二项分布参数p的先验分布且先验参数为α和β时，后验分布的对应参数为α+z和β+n-z。其中p代表出现对应于X=1的结果的概率，n代表总的试验次数，z代表出现对应于X=1的结果的次数。

参数估计方法

- 由于使用贝塔先验分布时，后验分布仍为贝塔分布且参数已知，我们很容易得出参数p的点估计和区间估计
- 例如，当采用均匀先验分布即beta(1,1)，且10次试验中X=1的对应结果出现了3次时，后验分布为beta(4,8)，其点估计为0.3(后验众数), 1/3（后验数学期望）或者0.324(后验中位数)
- 其95%对称可信区间为[0.1093,0.6097]（两侧各去除2.5%）
- 其95%最高概率密度可信区间 (highest density interval, HDI)为[0.0934,0.5880]

模型比较方法

- 当使用贝塔分布作为参数p的先验分布时，容易求得P(D|H)，此时，H可以用beta(α,β)来表达，而P(D|H) = $\frac{\Gamma(\alpha+z)\Gamma(\beta+n-z)}{\Gamma(\alpha+\beta+n)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$
- 在之前的例子中，对于beta(1,1)这样的先验分布和n=10,z=3这样的数据，

$$P(D|H) = \frac{\Gamma(\alpha+z)\Gamma(\beta+n-z)}{\Gamma(\alpha+\beta+n)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

$$= \frac{\Gamma(4)\Gamma(8)}{\Gamma(12)}\frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} = 0.00076$$

模型比较方法

- 对于p=0.5这样的点假设，容易求得$P(D|H)=0.5^3(1-0.5)^7=0.00098$
- 故针对H0: p = 0.5, H1: p ~ beta(1,1)而言，BF10 = 0.00076/0.00098 = 0.78
- 如果我们先验地认为P(H0) = 0.5, P(H1) = 0.5, 那么

$$\frac{P(H_1|D)}{P(H_0|D)}=\frac{P(D|H_1)}{P(D|H_0)}\times\frac{P(H_1)}{P(H_0)}=BF\times 1=0.78$$

- 也就是说，根据所得数据，两个模型的概率比从先验的1调整为了后验的0.78

问题

- 先验地认为P(H0) = 0.5, P(H1) = 0.5,应该如何理解?
- 所谓模型的概率比，这里面的模型，到底指的是什么意思?

数据预测方法

- 在得到了n=10, z = 3的观测数据，并且先验地假定p~beta(1,1)的情况下，我们可以进一步预测之后观测的结果
- 此时，p的后验分布为beta(4,8), 它反映了我们在得到n=10, z=3的数据之后对于p的看法
- 针对每个p的可能值，可以计算得到一批特定的新数据的概率，那么得到一批特定的新数据的总概率（即边缘概率），就是在p的后验分布下，似然函数的数学期望

数据预测方法

- 例如，假如我们再做n次相同的试验，那么得到X=1对应结果的次数为z的概率等于

$$\Pr(z|n,\text{Posterior})$$

$$= \int_0^1 C_n^z p^z (1-p)^{n-z} \frac{\Gamma(12)}{\Gamma(4)\Gamma(8)} p^3 (1-p)^7 dp$$

$$= C_n^z \frac{\Gamma(12)}{\Gamma(4)\Gamma(8)} \frac{\Gamma(z+4)\Gamma(n-z+8)}{\Gamma(n+12)} \int_0^1 p^{z+3}(1-p)^{n-z+7} \frac{\Gamma(n+12)}{\Gamma(z+4)\Gamma(n-z+8)} dp$$

$$= C_n^z \frac{\Gamma(12)}{\Gamma(4)\Gamma(8)} \frac{\Gamma(z+4)\Gamma(n-z+8)}{\Gamma(n+12)}$$

问题

- 频率学派统计学能否进行类似的数据预测?

数学期望的调整方式

- 上例中先验分布的数学期望为

$$\mu_{prior} = \frac{\alpha}{\alpha + \beta}$$

后验分布的数学期望为

$$\mu_{posterior} = \frac{\alpha + z}{\alpha + \beta + n}$$
$$= \frac{\alpha + \beta}{\alpha + \beta + n}\frac{\alpha}{\alpha + \beta} + \frac{n}{\alpha + \beta + n}\frac{z}{n}$$
$$= \frac{\alpha + \beta}{\alpha + \beta + n}\mu_{prior} + \frac{n}{\alpha + \beta + n}M_{sample}$$

即后验数学期望是先验数学期望和样本平均数的加权平均，权重和$\alpha + \beta$以及n成正比

正态分布参数的共轭分布

- 对于连续概率分布以及对应的似然函数，也可能存在共轭分布，例如，对于正态分布的两个参数μ和σ，存在如下共轭分布

1. 当μ未知而σ^2已知时，μ的共轭分布为正态分布
2. 当μ已知而σ^2未知时，σ^2的共轭分布为倒伽马分布
3. 当μ和σ^2都未知时，(μ,σ^2)的联合共轭分布为正态-倒伽马分布

当使用共轭先验分布时，有关正态分布参数的统计分析和推断也会变得容易许多。

参见DBDA(ed2), Ch16; 贝叶斯统计，例1.5.1

注意事项

- 虽然共轭分布使我们的统计分析大大简化了，但是并非所有情况下都可以使用共轭先验分布

1. 有时候共轭分布并无法准确刻画我们的先验信念
2. 有时候共轭先验分布的性质会导致一些错误的假定或者概念上的困难
3. 对于某些分布参数而言，并不存在共轭分布

- 在以上这些情况下，需要使用更加合理的非共轭先验分布，并且使用近似方法，进行贝叶斯统计分析