1.频率学派
vs.
贝叶斯学派统计学

戴俊毅

研究员/长聘副教授

频率学派统计学基本概念

- p值：
- α水平：
- 置信区间：
- 统计效力：

频率学派统计学基本概念

- p值：在零假设为真的前提下，观测到实际数据或者更极端数据的概率
- α水平：
- 置信区间：
- 统计效力：

频率学派统计学基本概念

- p值：在零假设为真的前提下，观测到实际数据或者更极端数据的概率
- α水平：犯第1类统计推断错误的概率
- 置信区间：
- 统计效力：

频率学派统计学基本概念

- p值：在零假设为真的前提下，观测到实际数据或者更极端数据的概率
- α水平：犯第1类统计推断错误的概率
- 置信区间：有可能包含参数真值的（随机）区间
- 统计效力：

频率学派统计学基本概念

- p值：在零假设为真的前提下，观测到实际数据或者更极端数据的概率
- α水平：犯第1类统计推断错误的概率
- 置信区间：有可能包含参数真值的（随机）区间
- 统计效力：在备择假设为真的前提下，得到统计意义上显著结果的概率

对p值的（错误）解读

1. p值代表零假设为真的概率 ✖
2. p值代表观测结果完全由随机性导致的概率 ✖
3. $p<0.05$意味着零假设是错误的，应该被拒绝✖
4. $p>0.05$意味着零假设是正确的，应该被接受✖
5. p值较大意味着存在支持零假设的证据 ✖
6. p值代表在零假设为真的前提下，出现观测数据的概率 ✖
7. ……

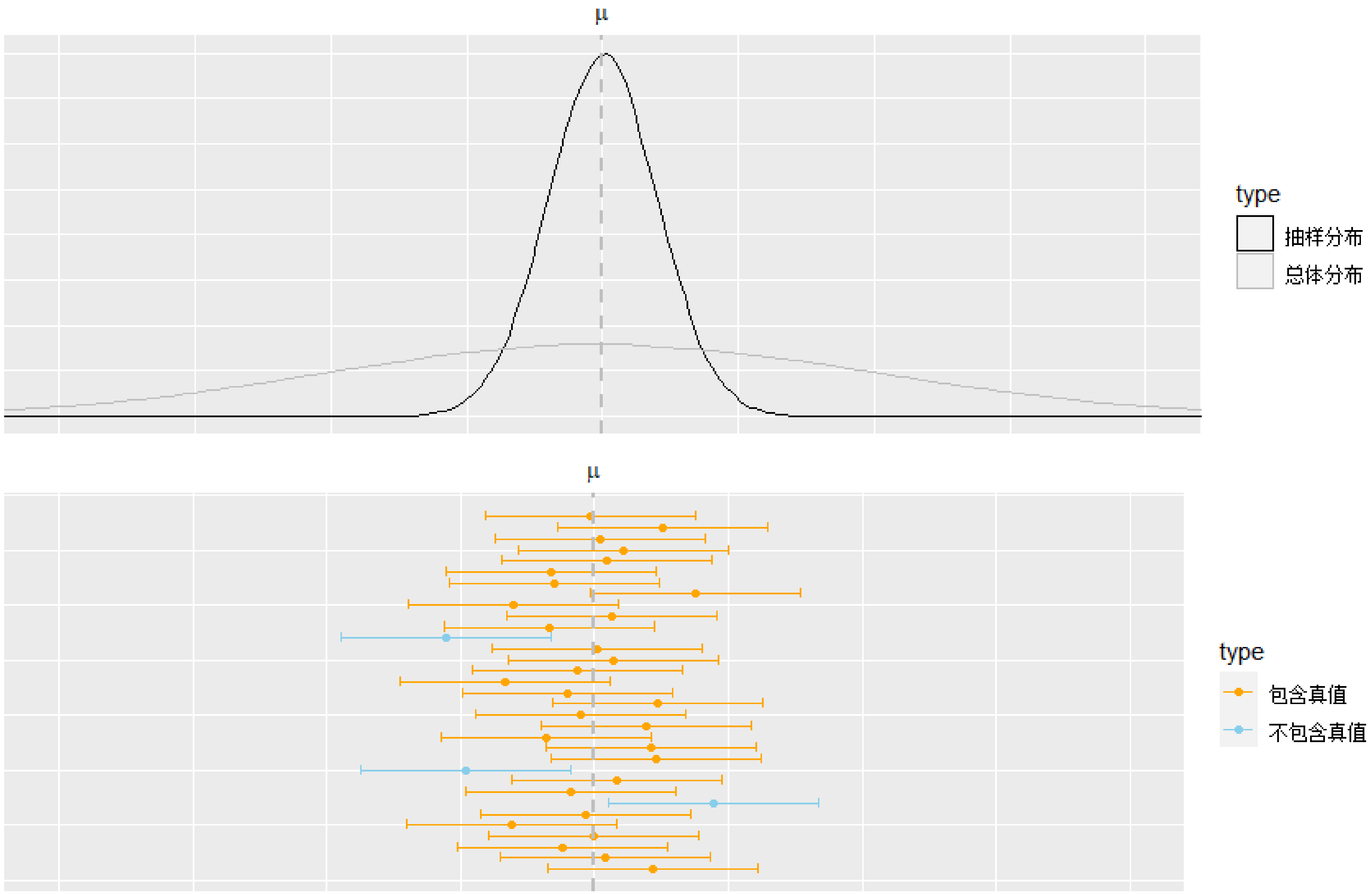
对置信区间的（错误）解读

1. 由一项研究的数据计算得到的95%置信区间有95%的概率包含参数真值 ✖
2. 未来研究所得的参数估计值有95%的概率会落入当前研究得出的95%置信区间 ✖
3. ……

Greenland et al. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretation. European Journal of Epidemiology, 31, 337—350.

置信区间示例

- 假定某一总体满足正态分布，其数学期望（即总体平均数）为μ（未知），（总体）标准差为5（已知）
- 那么，所有大小为25的随机样本的样本平均数组成的抽样分布满足正态分布，其数学期望为μ，标准差为1（=5/sqrt(25)）
- 每一个样本对应的总体平均数的95%置信区间为$m \pm 1.96$，其中m为样本平均数
- 当样本平均数m离μ较远时，对应的95%置信区间将不包括μ，反之，则包括μ


μ
type
抽样分布
总体分布
μ
type
包含真值
不包含真值

对置信区间的（错误）解读

1. 由一项研究的数据计算得到的95%置信区间有95%的概率包含参数真值 ✖
2. 未来研究所得的参数估计值有95%的概率会落入当前研究得出的95%置信区间 ✖
3. ……

Greenland et al. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretation. European Journal of Epidemiology, 31, 337—350.

概率

频率学派

- 各种事件或者结果在无穷多次试验中出现的相对频次（频率）
- 仅适用于可以在相同条件下重复进行的试验
- 相对客观

贝叶斯学派

- 对于各种事件或者结果的可能性大小的主观信念
- 既适用于可以在相同条件下重复进行的试验，也适用于仅发生若干次或者一次性的试验
- 相对主观

数据

频率学派

- 实证研究中观测到的数据，是众多可能观察到的数据中的一种
- 既要考虑实际观测到的数据，也要考虑实际未观测到，但是可能出现的数据

贝叶斯学派

- 实证研究中观测到的数据，是用来更新信念的唯一信息来源
- 实际上未观测到的数据，不应作为统计分析的依据

模型参数

频率学派

- 统计模型的参数表征的是客观事实
- 客观事实是确定的，所以参数的真实值是唯一的

贝叶斯学派

- 统计模型的参数表征的是主观上对于客观事实的认识
- 在通常情况下，由于信息的不完备，参数可以取各种不同的值，且被视为一个随机变量
- 在模型假设合理的前提下，随着数据的累积，主观信念强度将向参数真实值逼近

假设检验

频率学派

- 由于模型参数真值的唯一性，假设检验中的零假设通常以点假设形式出现
- 以点假设形式出现的零假设是数据分析的基础
- 假设检验的目的，通常是做出接受备择假设的统计推断
- 统计假设要么为真，要么为假

贝叶斯学派

- 由于模型参数可以取各种不同值，假设检验中的零假设也可以是区间假设
- 在统计假设检验中点假设和区间假设具有同等的地位
- 假设检验的目的，是做出接受某一假设的统计推断，既可以是零假设，也可以是备择假设
- 统计假设有一定的概率为真，一定的概率为假

发展简史

频率学派

- Gauss和Legendre: 误差分析、正态分布和最小二乘法
- K. Pearson: p值，统计假设检验理论，统计决策理论
- R. A. Fisher: 零假设显著性检验，方差分析，极大似然估计……
- Neyman: 置信区间
- Neyman & E. Pearson: 基于似然比的假设检验

贝叶斯学派

- Bayes, T. (1763)提出贝叶斯统计学派的核心问题和思想，即如何根据观测数据来推断特定情况发生的可能性
- Laplace, P.-S. (1774)独立提出了同样的思想
- Laplace, P.-S. (1812)提出了贝叶斯公式
- Jeffreys, H. (1939)确立了现代贝叶斯学派
- Ramsey, de Finitti, Savage, Lindley完善了对于主观概率的科学解释
- 1960年以来算法和计算能力的改善使得贝叶斯数据分析的实际应用成为可能

举例：有关流行病发病率的统计分析

- 模型参数Θ：在所有人口中该流行病的发病率
- 数据D: 随机抽取的10个人的结果如下，YNNYNYNNYN
  Y = 患病、N = 未患病

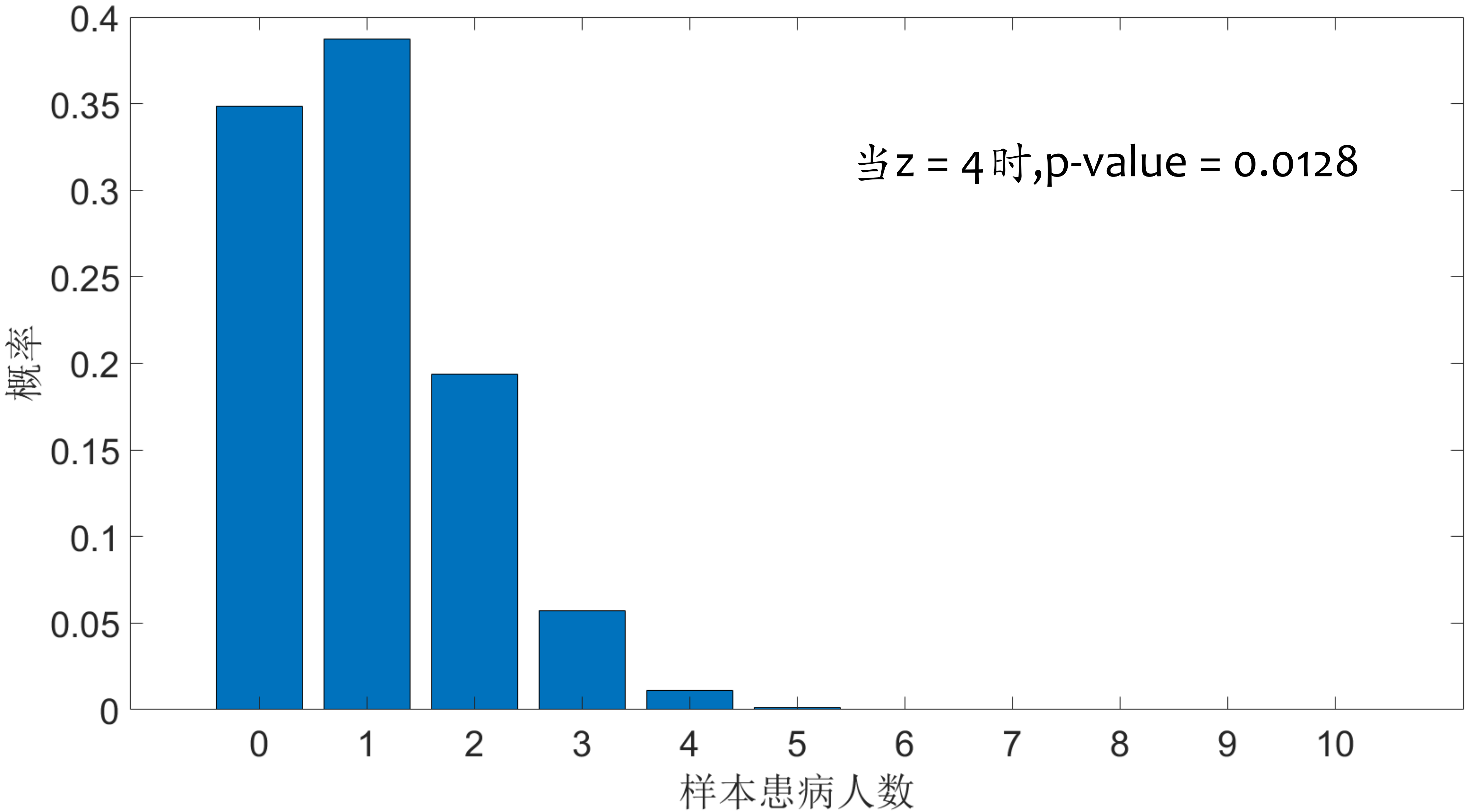
频率学派统计分析

- Θ是一个客观量，有唯一的真值
- 观察到的数据，是随机抽样的结果，需要考虑数据的抽样分布
- 由于Θ有唯一的真值，所以统计假设一般为点假设，比如Θ =0.1

抽样分布

- 抽样分布（频率学派统计学）是指在无穷多次抽样后样本统计量的概率分布
- p值等于在抽样分布中和实际观测结果（用样本统计量表示）相当或者更加极端的结果的比例

本示例对应的抽样分布的分布律：$\Pr(z|\theta,n)=\binom{n}{z}\theta^{z}(1-\theta)^{n-z}$，其中n=10代表抽样人数，z代表样本患病人数。当Θ =0.1时，抽样分布为



问题

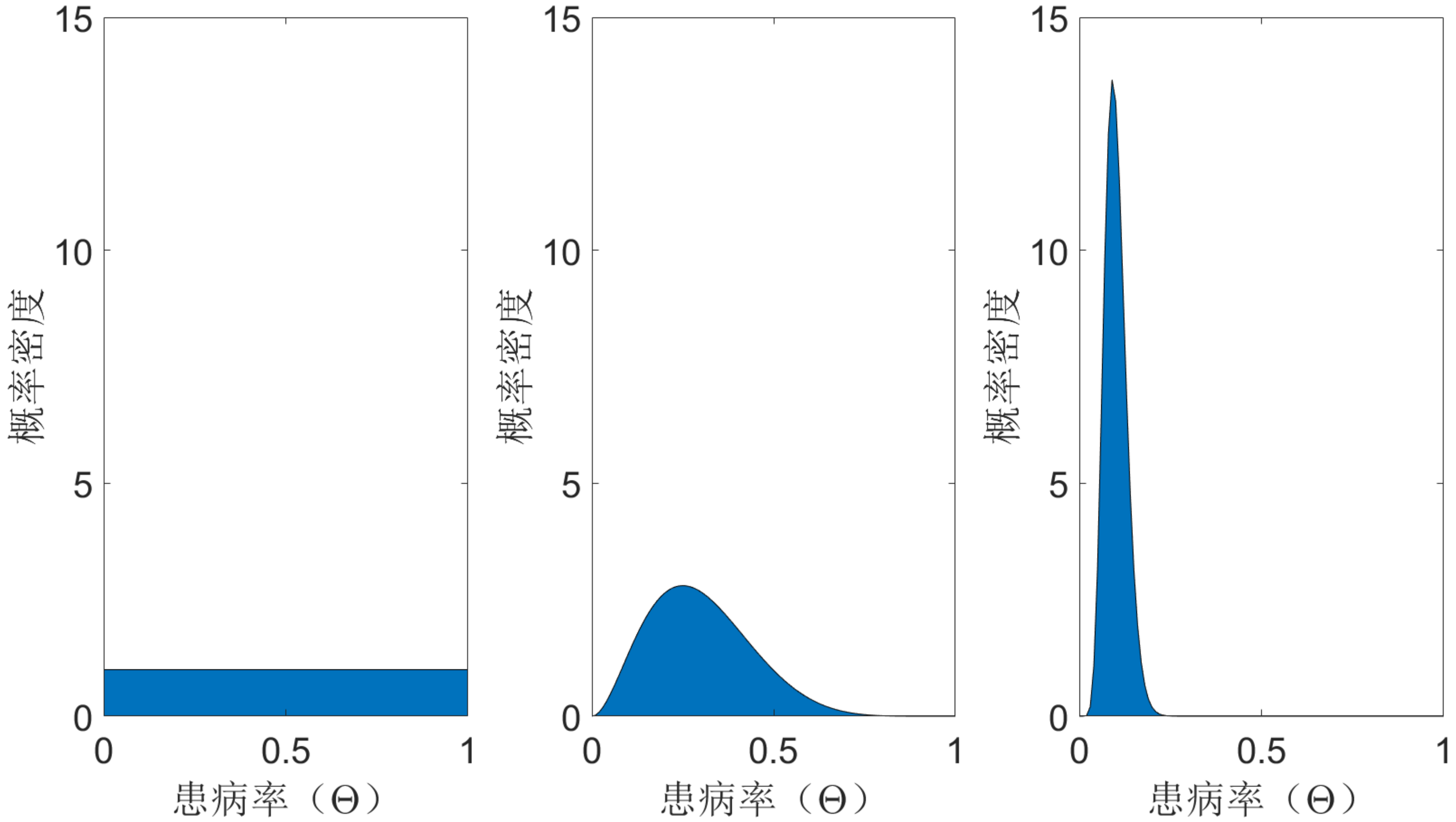
- 在频率学派统计学下，我们是否可以拒绝Θ =0.1这一假设?
- p值小，就意味着对应假设成立的可能性小吗?
- 在频率学派统计学下，怎样的数据可以让我们接受这一假设?
- 如果我们想知道，Θ <0.1的可能性有多大，应该如何处理?

贝叶斯学派统计分析

- Θ代表我们对于客观比率的主观认识，是一个随机变量
- 观察到的数据，是用来更新信念的唯一信息来源
- 未观察到的数据，不应作为统计推断的依据（条件观点）
- 由于Θ代表主观认识，所以有关的统计假设可以是区间假设，比如Θ <0.1

先验信念

- （贝叶斯统计学下的）先验信念是指我们在考虑数据之前，对于客观世界的看法，可以用先验概率分布表示
- 此处的概率分布，反映的是主观信念意义上的各种可能情况的可能性大小


15
10
5
0
概率密度
0
0.5
1
患病率（Θ）
15
10
5
0
概率密度
0
0.5
1
患病率（Θ）
15
10
5
0
概率密度
0
0.5
1
患病率（Θ）

后验信念

- 在考虑了数据之后，我们可以对先验信念进行调整，获得后验信念
- 调整的依据，是所谓的似然函数，即在给定数据的条件下，获得观测数据的可能性随参数值变化的函数
- 具体调整的方式，由贝叶斯公式决定

$$\underbrace{\Pr(\theta|D)}_{\text{后验信念}} \propto \underbrace{\Pr(\theta)}_{\text{先验信念}} \times \underbrace{\Pr(D|\theta)}_{\text{似然函数(数据信息)}}$$

贝叶斯定理（离散事件形式）

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

其中，A, B代表两个事件且$P(B) \neq 0$

$P(A|B)$：在B发生的前提下A发生的条件概率

$P(B|A)$：在A发生的前提下B发生的条件概率

$P(A)$：在不指定B是否发生的情况下A发生的边缘概率$(= \mathrm{P}(\mathrm{A}\&\mathrm{B}) + \mathrm{P}(\mathrm{A}\&\bar{B}))$

$P(B)$：在不指定A是否发生的情况下B发生的边缘概率$(= \mathrm{P}(\mathrm{A}\&\mathrm{B}) + \mathrm{P}(\bar{A}\&B))$

示例

- 假设A代表检测者患有艾滋病这一事件
- 假设B代表艾滋病检测结果为阳性这一事件
- 假定

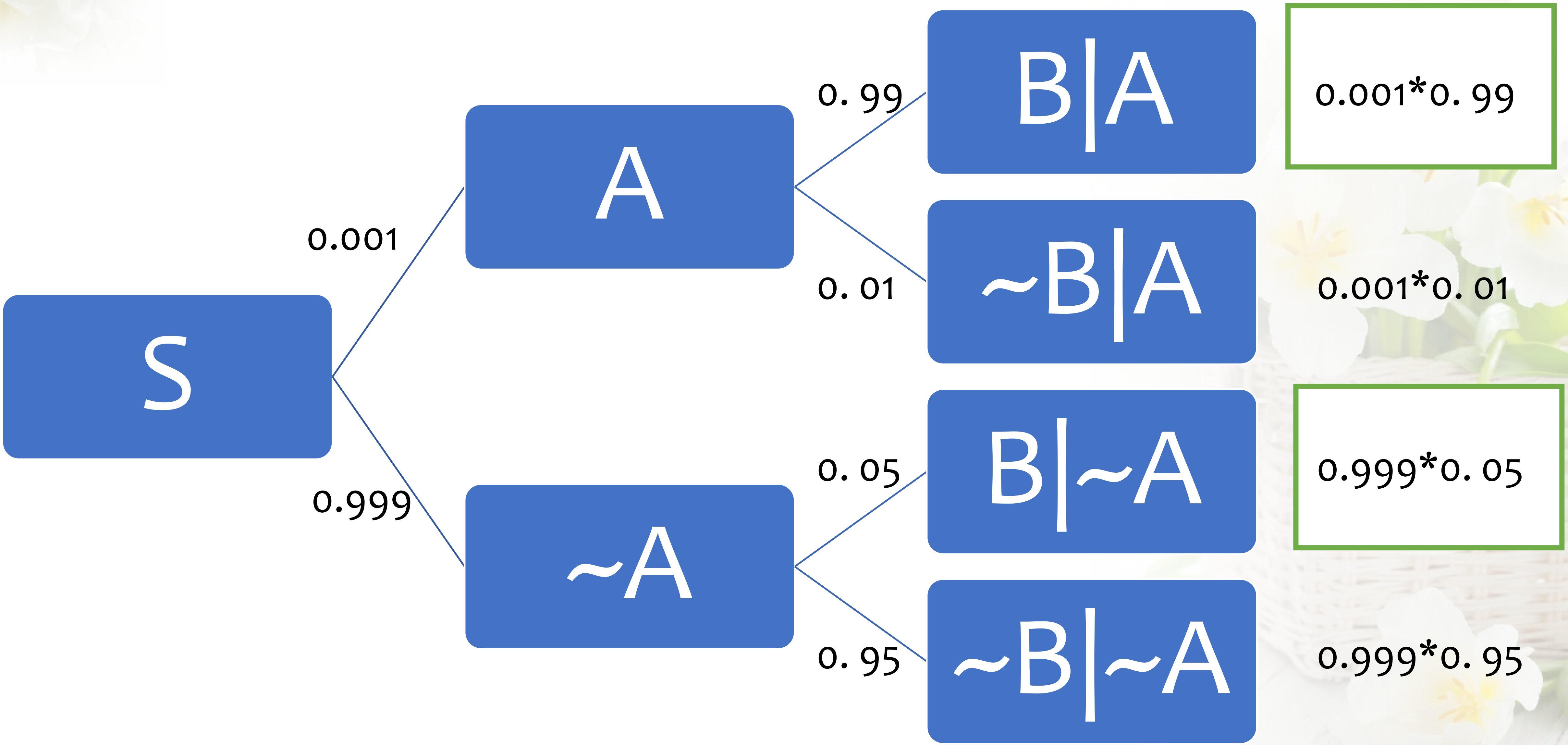
1. 艾滋病的总体患病率为1/1000
2. 在检测者患有艾滋病的条件下，检测结果为阳性的概率，即检出率为0.99
3. 检测者实际未得病，但检验结果为阳性的概率，即假阳性率为0.05

- 那么P(A) = 1/1000
- P(B|A) = 0.99
- $P(B|\bar{A})$ = 0.05
- P(B)代表检测结果为阳性的概率，有两种出现的可能性

  受测者的确患有艾滋病且检测结果正确,即$P(A)P(B|A)$

  受测者未患艾滋病且检测结果错误，即$P(\bar{A})P(B|\bar{A})$
- 因此

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}=\frac{0.99\times0.001}{0.99\times0.001+0.05*0.999}=1.94\%$$


S
0.001
0.999
A
~A
0. 99
0. 01
0. 05
0. 95
B|A
~B|A
B|~A
~B|~A
0.001*0. 99
0.001*0. 01
0.999*0. 05
0.999*0. 95

参数估计应用

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

其中，D代表观测数据, θ代表参数可能取值 (A = θ, B = D)

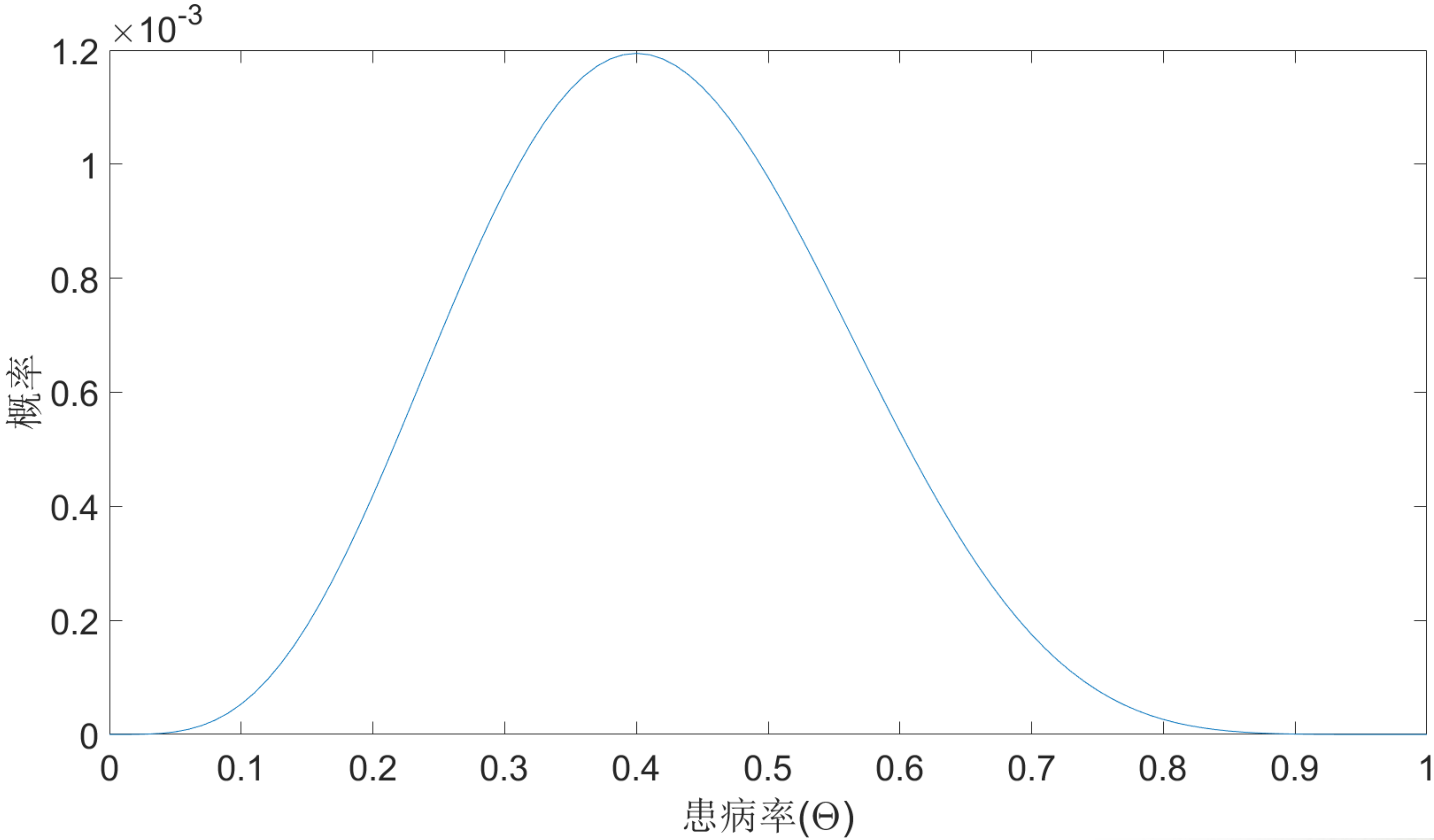
$P(\theta|\mathrm{D})$代表在出现数据D的前提下参数取值为θ的概率(如果Θ 离散)或概率密度(如果Θ连续)，即后验概率（或概率密度）

$P(\theta)$代表参数取值为θ的概率（或概率密度），即先验概率（或概率密度）

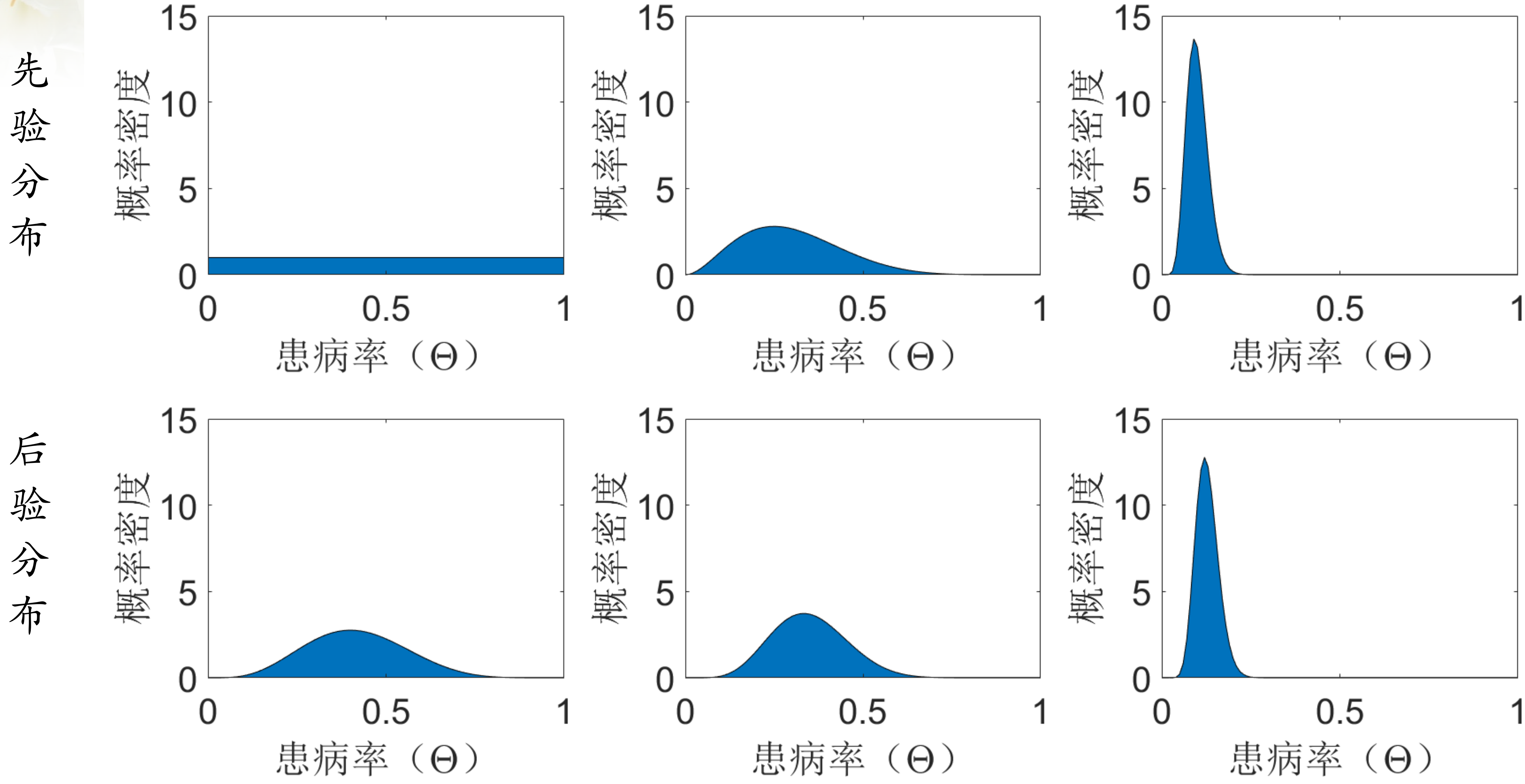
$P(\mathrm{D}|\theta)$代表在参数取值为θ的条件下出现数据D的概率(如果D离散)或概率密度(如果D连续)，即θ对应的似然值

$P(\mathrm{D})$代表出现数据D的边缘概率(如果D离散)或概率密度 (如果D连续)

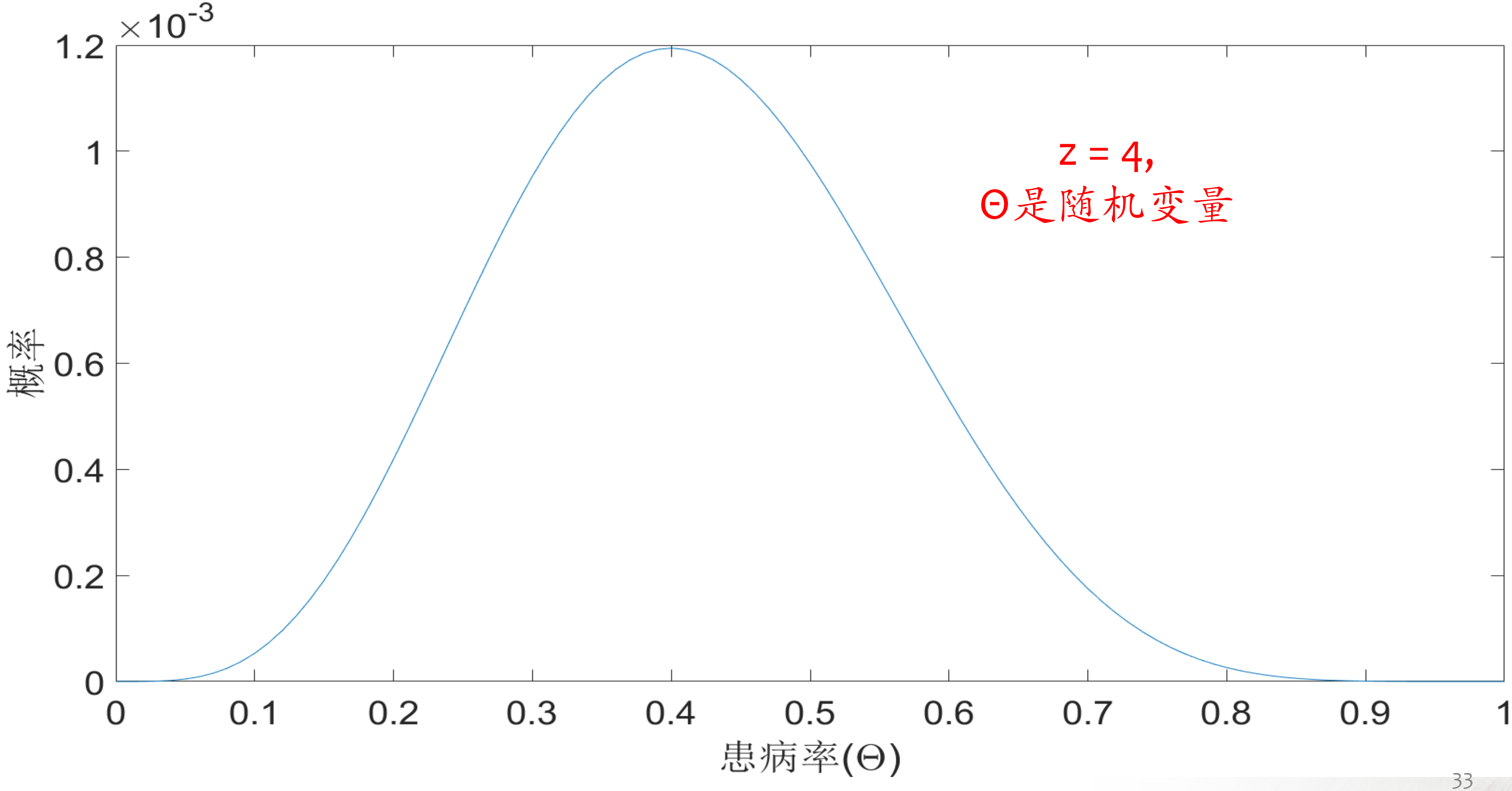
当Θ连续时， $P(\mathrm{D}) = \int P(D|\theta)P(\theta)d\theta$ ， 当Θ离散时， $P(\mathrm{D}) = \sum P(D|\theta)P(\theta)$


×10-3
1.2
1
0.8
0.6
0.4
0.2
0
概率
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
患病率(Θ)

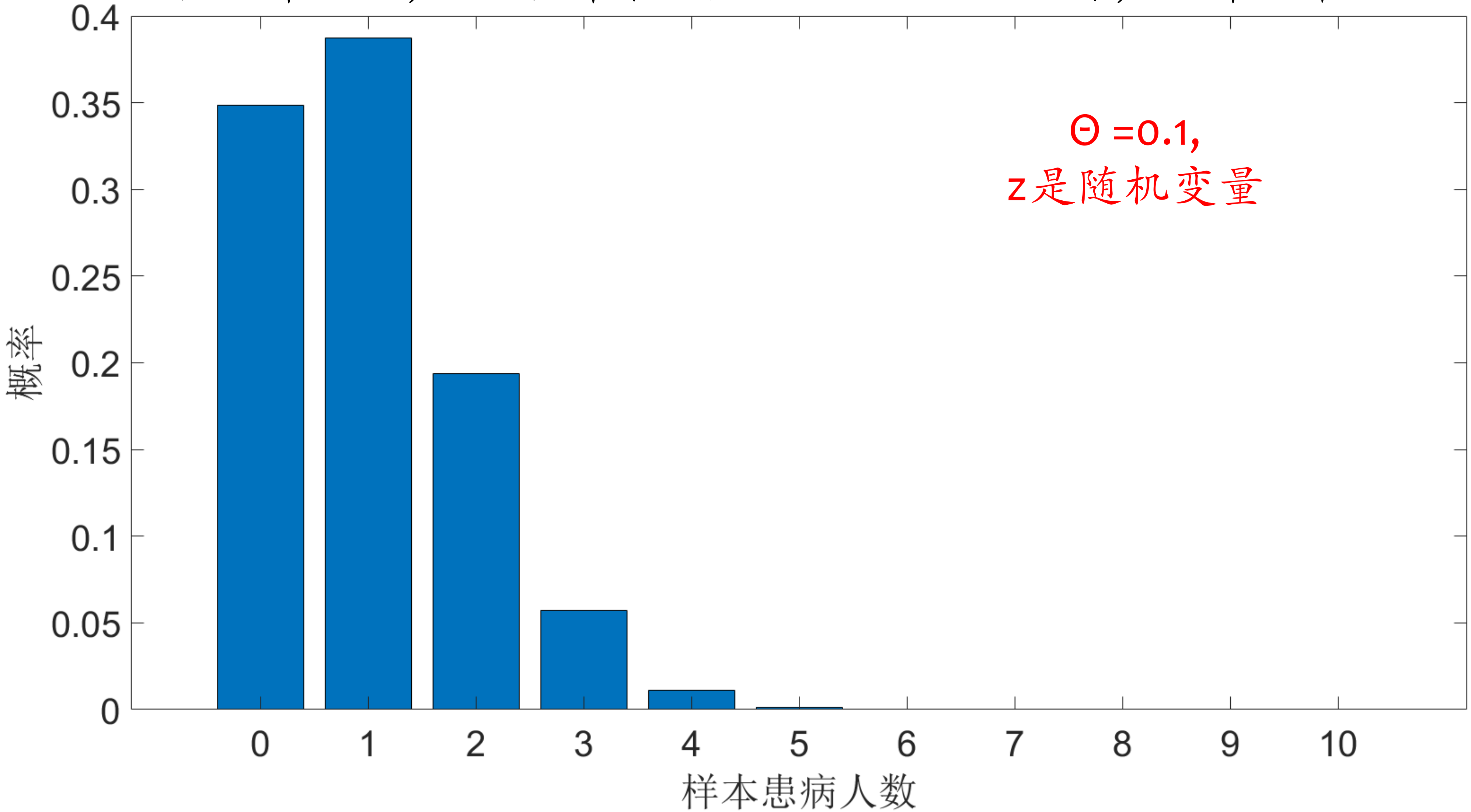
Θ是随机变量



对于本示例而言，似然函数$L(\theta;n,z) = \Pr(z|\theta,n) = \theta^z(1-\theta)^{n-z}$



本示例对应的抽样分布的分布律：$\Pr(z|\theta,n) = C_n^z\theta^z(1-\theta)^{n-z}$，其中n=10代表抽样人数，z代表样本患病人数。当Θ =0.1时，抽样分布为



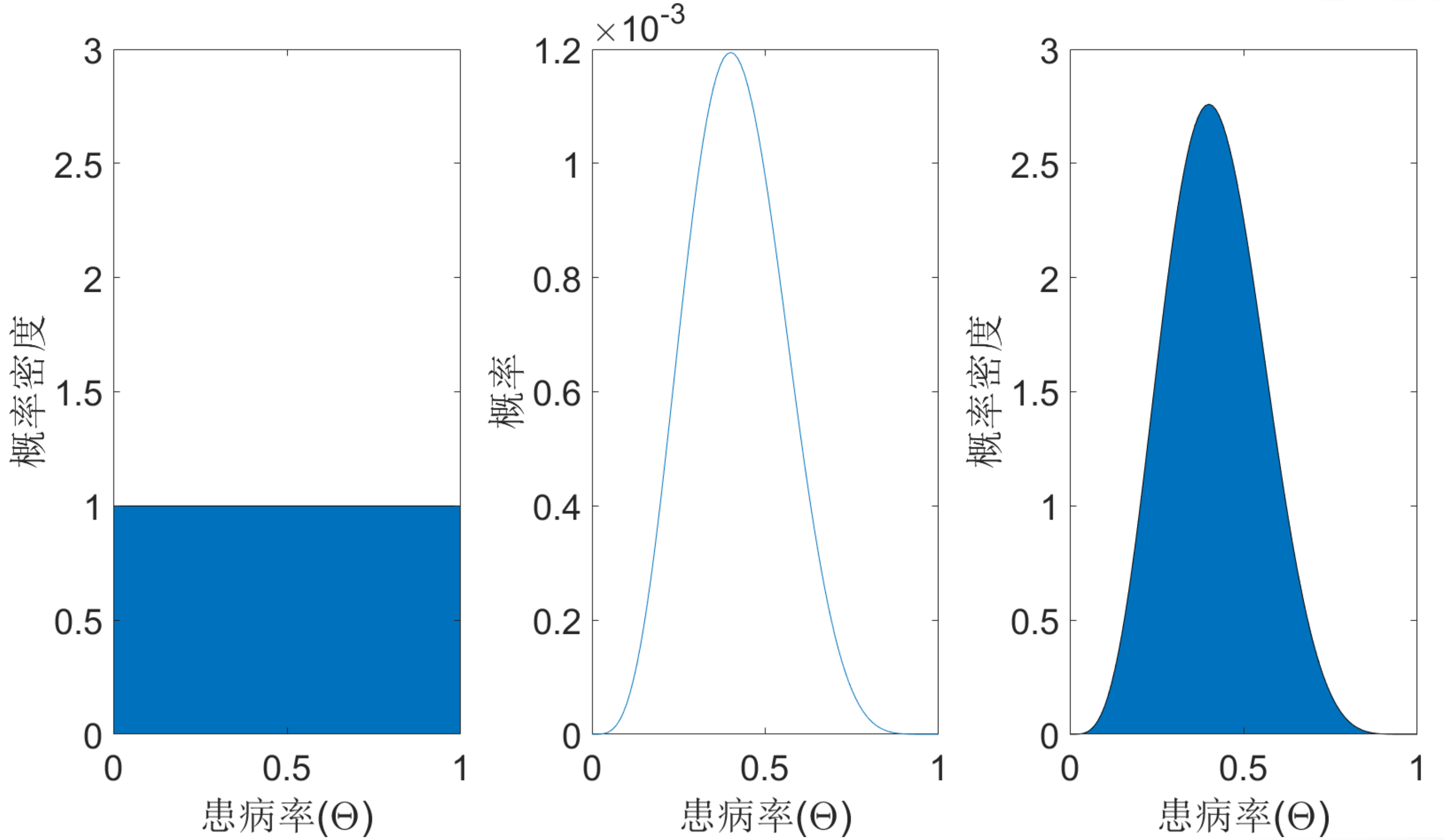
贝叶斯统计推断

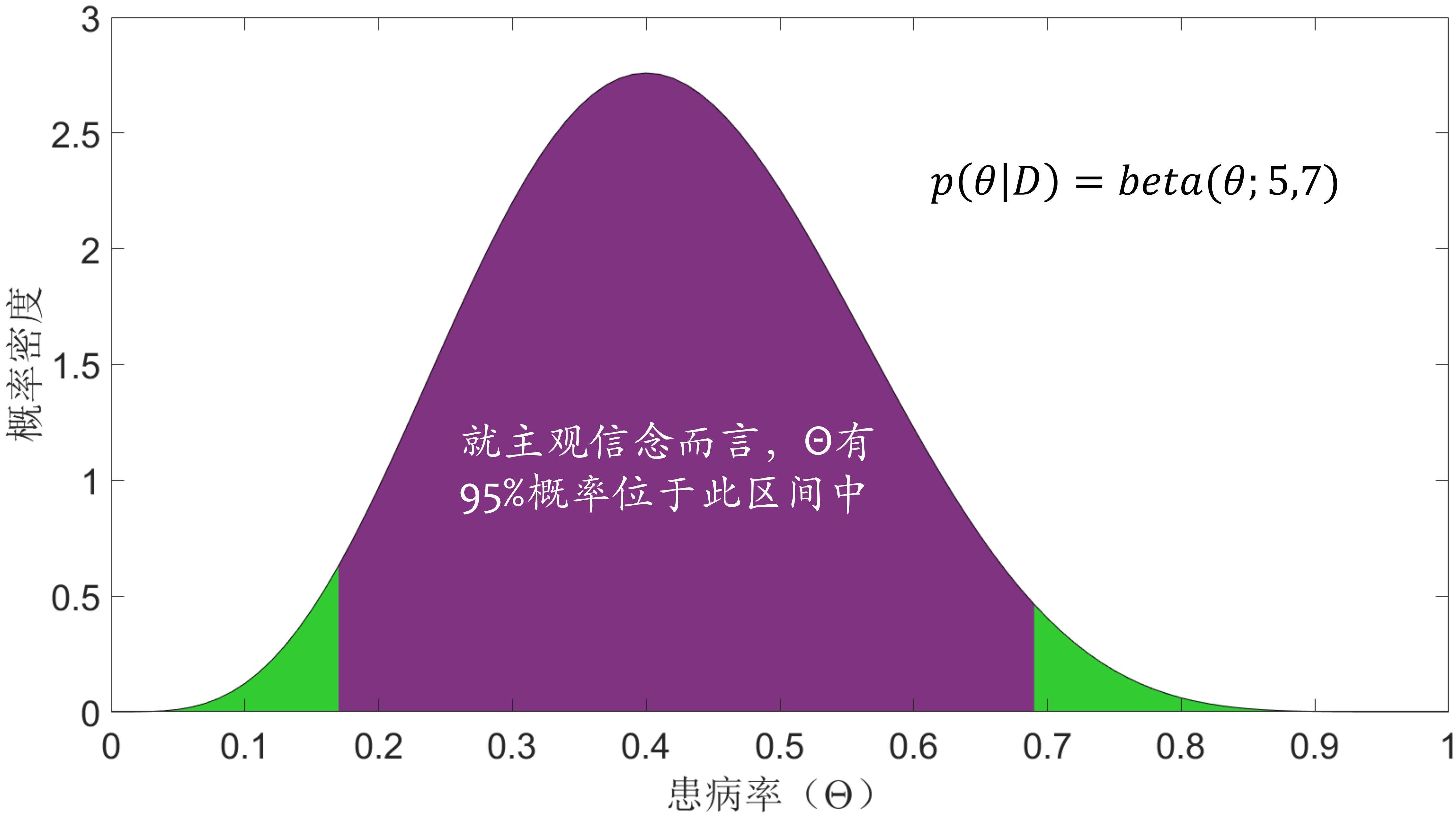
- 在贝叶斯统计学中，对假设进行统计推断主要有两种方式

1. 根据假设与后验分布之间的匹配程度
2. 根据不同假设间的证据强度之比（贝叶斯因子）

贝叶斯统计推断方法1

- 假定我们在考察数据之前，认为患病率可以取任意可能值，且可能性相等，那么我们的先验分布、似然函数以及后验分布如下




3
2.5
2
1.5
1
0.5
0
概率密度
$p(\theta|D) = beta(\theta; 5,7)$
就主观信念而言，Θ有
95%概率位于此区间中
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
患病率（Θ）

贝叶斯统计推断方法1

- 由后验概率分布可知，Θ的取值有95%的概率位于[0.1675,0.6921]这一区间里，且小于0.1675或者大于0.6921的概率都是2.5%
- 此例中的[0.1675,0.6921]被称为Θ 的95%可信区间(credible interval)
- 可信区间这一概念所涉及的概率，是贝叶斯意义上的概率，其含义和常识意义上的含义相一致
- 由后验概率分布可知， Θ <0.1的概率为0.0028
- 由于Θ <0.1这一区域和95%可信区间完全不重合，该假设可以被拒绝

贝叶斯统计推断方法2

- 在贝叶斯框架下，也可以根据数据支持不同假设的证据强度之比，做出接受或者拒绝统计假设的推断
- 比如，针对Θ =0.1这一假设(H_0)，可以通过对比支持它和支持Θ ~uniform(0,1)(H_1)的证据强度，来进行统计推断
- 在本例中，$P(D|H_0) = 0.000053$，$P(D|H_1) = 0.00043$，所以贝叶斯因子$(\text{Bayes Factor})_{10} = 0.00043/0.000053 = 8.15$，代表相比于零假设，有中等程度的证据支持备择假设
- 在此情况下，频率学派统计学的p-value = 0.0128

贝叶斯统计推断方法2

- 当H_0为Θ =0.4时， $P(D|H_0) = 0.0012$，$BF_{01} = 2.76$，代表相比于备择假设，有一定的证据支持零假设，但证据较弱
- 在此情况下，频率学派的p-value = 1 （双侧）
- 假定研究人员又随机抽取了10个人，他们的结果如下，YNNYNYNYYN, 那么总样本的结果为9名患者，11名非患者。此时
- 在频率学派统计学下，p-value = 0.8203（双侧）
- 在贝叶斯学派统计学下，$BF_{01} = 3.3545$（在贝叶斯框架下可以接受零假设）

贝叶斯因子的解读方式

Jeffreys(1961)建议用以下方法来解读BF_{12}

贝叶斯因子	解读
<1	(不同程度)支持模型2
1到$10^{0.5}$(≈3.16)	对模型1的支持程度弱到不值一提
$10^{0.5}$到10	较为（中等程度）支持模型1
10到$10^{3/2}$(≈31.6)	强烈支持模型1
$10^{3/2}$到100	极强地支持模型1
大于100	决定性地支持模型1

问题

- 针对Θ <0.1这一区间零假设，会有怎样的结果?

贝叶斯统计学的优势

- 更接近常识理解
- 客观和主观相匹配
- 适用范围更大
- 一致性程度更高