绪论

- 贝叶斯学派的最基本的观点是:**任一个未知量θ都可看作一个随机变量,应该用一个概率分布去描述对θ的未知状况。**这个概率分布是在抽样前就有的关于θ的先验信息的概率称述。
- **似然函数**属于联合密度函数，综合了总体信息和样本信息

$$L(\theta')=p(X|\theta')=\prod_{i=1}^{n}p(x_i|\theta') \tag{1}$$

- 贝叶斯公式的密度函数形式与离散形式，其中θ的条件分布称为θ的后验分布，集中了总体、样本和先验等三种信息中有关θ的一切信息，排除了与之无关的信息。一般先验分布$\pi(\theta)$反映人们抽样前的认识，通过抽样信息（总体信息和样本信息）对先验进行调整形成后验分布。

$$\pi(\theta|\boldsymbol{x})=\frac{p(\boldsymbol{x}|\theta)\pi(\theta)}{h(\boldsymbol{x},\theta)}=\frac{p(\boldsymbol{x}|\theta)\pi(\theta)}{\int_{\Theta}p(\boldsymbol{x}|\theta)\pi(\theta)\mathrm{d}\theta} \tag{2}$$

$$\pi(\theta_i|x)=\frac{p(x|\theta_i)\pi(\theta)}{\sum_j p(x|\theta_j)\pi(\theta_j)} \tag{3}$$

- 贝叶斯假设，对无信息时，可认为θ在区间(0,1)的均匀分布

$$\pi(\theta)=\begin{cases}1,0<\theta<1\\0,\text{其他场合}\end{cases} \tag{4}$$

- 重要分布
  - 二项分布B(n, p)： 重复n次独立的伯努利试验，每次试验的成功概率为p，当试验次数为1时，二项分布服从0-1分布，其分布为：$P(X=k)=C_n^k p^k(1-p)^{n-k}$, 常用于观察单位只能具有相互对立的一种结果的猜测活动。
  - 指数分布： 描述泊松过程中的事件之间的时间的概率分布 ，即事件以恒定平均速率连续且独立地发生的过程， 具有无记忆的关键性质。常用于描述对发生的缺陷数或系统故障数的测量结果，但不能作为机械零件功能参数的分布规律。密度函数为：$f(x)=\lambda e^{-\lambda x};x>0$
  - 泊松分布$P(\lambda)$： 适合于描述单位时间内随机事件发生的次数。 概率函数为：$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$;k=0,1.... 当二项分布的n很大而p很小时，泊松分布可作为二项分布的近似，其中λ为np。
  - **贝塔分布**，也称B分布，定义在(0,1) 区间的连续概率分布，其概率密度函数为：$f(x;\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}=\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$，其中**贝塔函数**
    $B(\alpha,\beta)=\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$，Γ为伽马函数
    $\Gamma(x)=\int_0^{+\infty}t^{x-1}e^{-t}dt;(x>0)$，贝塔分布的核为$\theta^{\alpha-1}(1-\theta)^{\beta-1}$(注意区分二项分布的核$\theta^x(1-\theta)^{n-x}$中x为变量，贝塔分布中$\theta$是变量)
  - **伽马分布**$Ga(\alpha,\lambda)$，其中α>0为形状参数，$\lambda>0$为尺度参数，其密度函数为
    $p(x|\alpha,\lambda)=\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}$，通过此可以得到$Y=X^{-1}$的密度函数：
    $p(y|\alpha,\lambda)=\frac{\lambda^\alpha}{\Gamma(\alpha)}\frac{1}{y}^{\alpha+1}e^{\frac{-\lambda}{y}}$，称为**倒伽马分布**记为$IGa(\alpha,\lambda)$
  - 正态分布$N(\mu,\sigma^2)$：其概率函数为$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- 指数分布簇
  - 形如 $f_X(x|\theta)=h(x)\,g(\theta)\exp[\,\eta(\theta)\cdot T(x)\,]$

- 包含如正态分布、多项式分布、泊松分布、伽马分布、指数分布、贝塔分布和 $Dirichlet$分布等

共轭先验

- 设θ是<u>总体分布</u>中的参数(或参数向量)，$\pi(\theta)$是θ的先验密度函数,假如由抽样信息算得的后验密度函数与$\pi(\theta)$有**相同的函数形式**,则称$\pi(\theta)$是θ的(自然)**共轭先验分布**。通过这种方式计算得到的后验分布的一些参数可以很好解释。共轭先验分布的选区是由似然函数所含的θ因式所决定，即选与似然函数(θ的函数)具有相同核的分布作为先验分布。
  - 正态**均值**(方差已知)的共轭先验分布是正态分布。可以理解为：后验均值是在先验均值与样本均值间采取折衷方案，在处理正态分布时,方差的倒数发挥着重要作用,并称其为**精度**，则*后验分布的精度是样本均值分布的精度与先验分布精度之和*，<u>增加样本量n或减少先验分布方差都有利于提高后验分布的精度</u>。

$$\begin{gathered} \text{先验知识}\theta \sim N(\mu,\tau^2)\text{总体分布}x \sim N(\theta,\sigma^2)\text{样本}\overline{x},\sigma_0^2 = \frac{\sigma^2}{n} \qquad (5) \\ \text{后验知识}\pi(\theta|\boldsymbol{x}) \sim N(\mu_1,\tau_1^2) \end{gathered}$$

$$\begin{gathered} \mu_1 = \frac{\frac{\mu}{\tau^2}+\frac{\overline{x}}{\sigma_0^2}}{\frac{1}{\tau^2}+\frac{1}{\sigma_0^2}} \qquad (6) \\ \frac{1}{\tau_1^2} = \frac{1}{\tau^2}+\frac{1}{\sigma_0^2} \end{gathered}$$

  - 二项分布的成功概率θ的共轭先验分布是贝塔分布

$$\begin{gathered} \text{先验}\theta \sim Be(\alpha,\beta)\text{总体}X \sim b(n,\theta) \qquad (7) \\ \text{后验}\pi(\theta|\boldsymbol{x}) \sim Be(\alpha+x,\beta+n-x) \end{gathered}$$

$$\begin{gathered} E(\theta|x) = \frac{\alpha+x}{\alpha+\beta+n} = \frac{n}{\alpha+\beta+n}\frac{x}{n} + \frac{\alpha+\beta}{\alpha+\beta+n}\frac{\alpha}{\alpha+\beta} \qquad (8) \\ Var(\theta|x) \approx \frac{1}{n}\frac{x}{n}(1-\frac{x}{n}) \end{gathered}$$

  - 常用共轭先验分布

表 1.3.1 常用共轭先验分布

总体分布	参数	共轭先验分布
二项分布	成功概率	贝塔分布 $Be(\alpha,\beta)$
泊松分布	均值	伽玛分布 $Ga(\alpha,\lambda)$
指数分布	均值的倒数	伽玛分布 $Ga(\alpha,\lambda)$
正态分布(方差已知)	均值	正态分布 $N(\mu,\tau^2)$
正态分布(均值已知)	方差	倒伽玛分布 $IGa(\alpha,\lambda)$

  - 在单参数指数族场合,使用共轭先验分布得后验均值一定值于先验均值与样本均值(或样本方差等)之间。
- **后验分布的计算**：由于$m(x)$不依赖于θ，在计算时仅起到正则化因子的作用，$\pi(\theta|\boldsymbol{x}) \propto p(\boldsymbol{x}|\theta)\pi(\theta)$，其中各因子提取出仅与$\theta$有关的称为核。计算时可以略去与$\theta$无关的因子。
- **先验分布的选取**，应以合理性作为首要原则

确定先验信息

> 超参数：先验分布中所含的未知参数称为超参数。无信息先验分布一般不含超参数。

- 确定超参数的估计值

    - 利用先验矩（根据历史若干个估计值，进行加工整理，得到相关值，估计值来源一般为专家经验）
    - 利用先验分位数（确定两个分位数，得到方程式，解得相关值）
    - 利用先验矩和先验分位数
- 多参数模型（实际问题中常有多个未知参数，而一般不关注的参数称为讨厌参数）
    - 正态均值与正态方差的(联合)共轭先验分布为**正态-逆伽马分布**记为$N-IGa(v_n, \mu_n \sigma_n^2)$
- 充分统计量
    - 设**x**是来自分布函数$F(x|\theta)$的一个样本，$T=T(x)$是统计量，假如在给定T(x)的条件下，x的条件分布与θ无关的话,则称该统计量为θ的充分统计量。
    - 设x为密度函数$p(x|\theta)$的一个样本，$T(x)$为θ的充分统计量的充要条件是，用样本分布$p(x|\theta)$算得的后验分布与统计量$T(x)$算得的后验分布是相同的。如二维统计量$T=(\overline{x}, Q)$恰好是量(μ, σ^2)的充分统计量。
    - 使用充分统计量可以简化数据、降低样本维数，从而简化后验分布的计算。

贝叶斯估计

- 条件方法

  后验分布是在样本x给定下θ的条件分布,基于后验分布的统计推断就意味着只考虑已出现的数据(样本观察值),而认为<u>未出现的数据与推断无关</u>，这一重要的观点被称为“条件观点“,基于这种观点提出的统计推断方法被称为**条件方法**。

贝叶斯估计

- 从后验分布中选用某个特征量作为θ的估计。使后验密度达到最大的值θ_{MG}称为最大后验估计;后验分布的中位数θ_{Me}称为θ的后验中位数估计;后验分布的期望值θ_E称为θ的后验期望估计,这三个估计也都称为θ的**贝叶斯估计**,记为θ_B,在不引起混乱时也记为θ_0。实际中，一般采用后验期望估计作为贝叶斯估计。
    - 估计的误差。取后验均值可使后验均方差达到最小。

定义 2.2.2 设参数θ的后验分布$\pi(\theta|\boldsymbol{x})$,贝叶斯估计为$\hat{\theta}$,则$(\theta-\hat{\theta})^2$的后验期望

$$MSE(\hat{\theta}|\boldsymbol{x})=E^{\theta|x}(\theta-\hat{\theta})^2$$

称为$\hat{\theta}$的后验均方差,而其平方根$[MSE(\hat{\theta}|\boldsymbol{x})]^{\frac{1}{2}}$称为$\hat{\theta}$的后验标准误,其中符号$E^{\theta|x}$表示用条件分布$\pi(\theta|\boldsymbol{x})$求期望,当$\hat{\theta}$为$\theta$的后验期望$\hat{\theta}_E=E(\theta|\boldsymbol{x})$时,则

$$MSE(\hat{\theta}_E|\boldsymbol{x})=E^{\theta|x}(\theta-\hat{\theta}_E)^2=\mathrm{Var}(\theta|\boldsymbol{x})$$

称为后验方差,其平方根$[\mathrm{Var}(\theta|\boldsymbol{x})]^{\frac{1}{2}}$称为后验标准差。

- 柯西分布 期望不存在

区间估计

> 对于区间估计问题,贝叶斯方法具有处理方便和含义清晰的优点,而经典方法寻求的置信区间常受到批评。

- 可信区间:

设参数θ的后验分布为$\pi(\theta|x)$，给定样本x和概率α (0<α<1)，若存在这样两个统计量θ_U θ_L，使得$P(\theta_L \leq \theta \leq \theta_U|x) > 1-\alpha$，则称区间[$\theta_U$,θ_L]为θ的可信水平为$1-\alpha$的贝叶斯可信区间，即参数θ的$1-\alpha$的可信区间。仿照经典方法，可以得到$1-\alpha$的单侧可信下限和$1-\alpha$的单侧可信上限。

 - 贝叶斯方法可信区间的寻求，较经典统计方法更简单。
 - 经典统计求得的是置信区间，而贝叶斯得到的是可信区间，可信区间更符合理解和解释。

- 最大后验密度(HPD)可信区间

> 区间长度最短，并把具有最大后验密度的点都包含在区间内，而区间外的点上的后验密度函数值不超过区间内的后验密度函数值

 - 若后验密度函数是单峰对称的，则$(1-\alpha)HPD$可信空间为等尾可信区间，单峰不对称需要计算机器辅助计算；多峰则可能出现可信区间不连续的情况。
 - PS： 当后验密度函数出现多峰时，常常是由于先验信息与抽样信息不一致引起的，而共轭先验分布大多是单峰的，这必导致后验分布也是单峰的，它可能会掩盖这种不一致信息，故而要慎重对待和使用共轭先验分布。

假设检验

获得后验分布后，计算两个假设H0与H1的后验概率，然后比较两者的大小，即观察**后验概率比**α_0/α_1，从中选择最大概率的一方；但当两者相接近时需要进一步抽样或搜集信息。此种方法可推广到三个及以上的假设状况。

- 贝叶斯因子，既依赖于样本数据x，还依赖于先验分布π，这会减弱先验的影响，突出数据的影响；贝叶斯因子体现了数据支持某假设的程度。贝叶斯因子对样本信息变化的反应是灵敏的，而对先验信息变化的反应是迟钝的。

$$B^\pi(x) = \frac{\text{后验机会比}}{\text{先验机会比}} = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1} = \frac{\alpha_0\pi_1}{\pi_0\alpha_1} \tag{9}$$

 - 简单对简单（参数假设为特定值）

$$B^\pi(x) = \frac{\alpha_0\pi_1}{\pi_0\alpha_1} = \frac{p(x|\theta_0)}{p(x|\theta_1)} \tag{10}$$

 - 复杂对复杂（参数假设为特定区间，使用g(θ)约束θ的范围表示θ的分布情况，特别的取两个区间θ的极大似然估计代替g(θ)的加权结果可以得到经典统计的似然比统计量）

$$B^\pi(\boldsymbol{x}) = \frac{\alpha_0\pi_1}{\alpha_1\pi_0} = \frac{\int_{\Theta_0} p(\boldsymbol{x} \mid \theta) g_0(\theta) d\theta}{\int_{\Theta_1} p(\boldsymbol{x} \mid \theta) g_1(\theta) d\theta} = \frac{m_0(\boldsymbol{x})}{m_1(\boldsymbol{x})}$$

 - 简单对复杂（综合前两种情况的思维，将特定值转化为以特定值附近区间）

$$B^\pi(\boldsymbol{x}) = \frac{\alpha_0\pi_1}{\alpha_1\pi_0} = \frac{p(\boldsymbol{x}|\theta_0)}{m_1(\boldsymbol{x})}$$

由于此类情况的贝叶斯因子计算简单，可以使用其计算得到θ的后验分布：

$$\pi(\Theta_0 \mid \boldsymbol{x}) = \left[1 + \frac{1-\pi_0}{\pi_0}\frac{1}{B^\pi(\boldsymbol{x})}\right]^{-1}$$

- 以上的三种可以拓展到多重假设问题，PS: 针对现实问题，需要根据已知的信息和分布特定，设定总体分布和先验函数。
- 预测（对随机变量未来观察值做出统计推断，一般先获得变量分布，再取期望、中位数、众数、一定区间等作为预测值）预测值的方差一般大于实测值的方差。
  - 如果无样本观察数据，则使用先验分布获得随机变量 x 的边缘分布m(x)。
  - 如果有样本观察数据，则使用先验分布求得后验分布，再计算随机变量 x 的后验预测分布m(x|**x**)。

$$m(x\mid\boldsymbol{x})=\int_{\Theta}p(x\mid\theta)\pi(\theta\mid\boldsymbol{x})d\theta$$

  - 如果有样本观察数据，并估计同参数的另一个随机变量，则使用先验分布获得随机变量 z 的后验预测分布m(z|**x**)。

$$m(z\mid\boldsymbol{x})=\int_{\Theta}g(z\mid\theta)\pi(\theta\mid\boldsymbol{x})d\theta$$

- 似然原理 当x的样本值给出时，似然函数为$L(\theta)=p(x|\theta)=\prod_{i=1}^{n}p(x_i|\theta)$ 这是一个关于θ的函数，使似然函数在参数空间取最值的$\hat{\theta}$称为最大似然估计。
  - 有了观测值后，似然函数L(θ)包含了所有与试验有关的θ的信息；
  - 如果两个似然函数成比例，比例函数与θ无关，则两者包含θ的信息相同

先验分布的确定

- 主观概率(人们根据经验对一个事件发生可能性的个人信念，对取值范围是离散时更有效)
  - 对立事件比较
  - 专家意见（询问专家时需要设计好问题，并对专家有一定的了解便于修正形成自己的主观概率，或者向多个专家咨询综合修正）
  - 历史资料
- 利用先验信息（参数空间连续）
  - 等分区间统计各区间的频率，绘制直方图
  - 选定先验密度后再估计超参数
  - 定分度与变分度
- 利用边缘分布m(x)
  - 边缘分布可以看作是混合分布（多个总体加权平均）的推广，如果p(x|θ)已知，则m(x)可以反映先验函数的合理性；
  - 把m^{π}作为先验函数π的似然函数，通过极大似然法选取π，这种方法称为二型极大似然先验。如果先验密度函数形式已知，则求解先验函数中的超参数即可。
  - 矩方法(先验函数形式已知时，利用先验矩和边缘分布矩的关系建立方程寻求超参数的估计值)
- 无信息先验与广义先验分布

贝叶斯决策

- 决策三要素：状态集合、行动集、收益函数Q
- 行动的容许性：行动集中只存在容许的行动（有选择地可能，有存在地必要）
- 决策准则：悲观准则(max min)、乐观准则(max max)、折中准则（乐观系数）

- 损失函数L = max(Q) - Q "该赚却没赚到的钱"。损失函数包含了较多的信息，使用其做决策将更为合理
- 先验期望准则：以收益函数在先验信息下得到的先验期望收益，取最大处为最优行动（与收益函数的原点和单位无关）； 或以损失函数在先验信息下得到的先验损失，取最大处为最优行动。两种方式只用到了先验信息，故只能使用正常的先验分布，而不能使用广义先验分布。
- 把损失函数引入贝叶斯统计推断，就构成了贝叶斯决策问题。
- 后验风险准则：损失函数对后验分布的期望称为后验风险R，以后验风险最小处为最优行动（和样本有关，故是一个决策函数），此时的决策函数为贝叶斯解。
- 决策函数（从样本到决策的映射）与决策函数类

贝叶斯小结

- 认识到贝叶斯学派的最基本的观点是:**任一个未知量θ都可看作一个随机变量,应该用一个概率分布去描述对θ的未知状况。**
- 牢记贝叶斯原理的公式，分清离散与连续的区别和使用情景。

$$\pi(\theta|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\theta)\pi(\theta)}{h(\boldsymbol{x},\theta)} = \frac{p(\boldsymbol{x}|\theta)\pi(\theta)}{\int_\Theta p(\boldsymbol{x}|\theta)\pi(\theta)\mathrm{d}\theta} \tag{11}$$

- 针对现实问题，需要根据已知的信息和分布特征，把握总体分布和先验函数形式。
- 牢记各种分布的函数的分布函数，核的形式，共轭先验分布，期望与方差，掌握利用似然原理计算后验的超参数，了解区间(0,1)上的均匀分布$U(0,1)$是贝塔分布$Be(1,1)$，从$U(0,1)$中样本的第k个顺序量$x_{(k)} \sim Be(k,n-k+1)$；

常见分布	概率密度或概率函数	数学期望	方差	共轭先验分布
均匀分布	$f(x) = 1/(b-a), a \le x \le b,$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	a=0时，b—pareto分布
二项分布	$P(X=k) = C_n^k p^k (1-p)^{n-k}$	np	$np(1-p)$	p—贝塔分布
泊松分布	$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$	λ	λ	λ—伽马分布
指数分布	$f(x) = \lambda e^{-\lambda x}; x > 0$，即$Ga(1, \lambda)$	$1/\lambda$	$1/\lambda^2$	λ—伽马分布，$1/\lambda$—逆伽马分布
正态分布	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	μ—正态分布；σ^2—逆伽马分布
贝塔分布	$f(x; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$	$\alpha/(\alpha+\beta)$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	
伽马分布	$p(x \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x > 0$	α/β	α/β^2	
逆伽马分布	$p(y \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} (1/y)^{\alpha+1} e^{-\beta/y}, y > 0$	$\beta/(\alpha-1)$	$\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$	
卡方分布	即$Ga(n/2, 1/2)$，n为自由度	n	$2n$	

- 正态分布作正态分布均值的共轭先验的性质（方差的倒数——精度可以看作是先验与样本精度的和）

$$
\begin{gathered}
\text{先验}\,\theta \sim N(\mu, \tau^2)\text{总体分布}\,x \sim N(\theta, \sigma^2)\text{样本}\,\bar{x}, \sigma_0^2 = \frac{\sigma^2}{n} \qquad (12) \\
\text{后验}\,\pi(\theta|\boldsymbol{x}) \sim N(\mu_1, \tau_1^2) \\
\mu_1 = \frac{\frac{\mu}{\tau^2} + \frac{\bar{x}}{\sigma_0^2}}{\frac{1}{\tau^2} + \frac{1}{\sigma_0^2}} \\
\frac{1}{\tau_1^2} = \frac{1}{\tau^2} + \frac{1}{\sigma_0^2}
\end{gathered}
$$

- 贝塔分布作二项分布成功概率的共轭先验的性质（α β 分别可以看作实验成功次数与不成功次数的累积）

$$
\begin{gathered}
\text{先验}:\theta \sim Be(\alpha, \beta)\text{总体}:X \sim b(n, \theta)\text{样本}:x \qquad (13) \\
\text{后验}\,\pi(\theta|\boldsymbol{x}) \sim Be(\alpha + x, \beta + n - x)
\end{gathered}
$$

- 伽马分布对泊松分布强度的共轭先验的性质（α β 可以看作泊松分布样本信息和与样本量的累计）

$$\text{先验：}\theta \sim Ga(\alpha,\beta)\text{总体：}P(x=x)=\frac{\theta^x}{x!}e^{-\theta}\text{样本：}\boldsymbol{x} \qquad (14)$$
$$\text{后验：}\theta \sim Ga(n\bar{x}+\alpha, n+\beta)$$

- 伽马分布对指数分布的共轭先验的性质（β α 可以看作泊松分布样本信息和与样本量的累计）

$$\text{先验：}\theta \sim Ga(\alpha,\beta)\text{总体：}X \sim Exp(\lambda)\text{样本：}\boldsymbol{x} \qquad (15)$$
$$\text{后验：}\theta \sim Ga(\alpha+n, \beta+n\bar{x})$$

- 认识到后验函数是样本信息与先验信息的综合，在后验的期望和方差中可以体现为加权或整合的形式，理解方差倒数作为精度的思维。
- 掌握常用的先验函数中超参数的求法：ML二型先验法和矩方法（利用边缘函数m(x)极值或矩关系）；
- 掌握常用的贝叶斯估计：后验期望估计（后验函数的期望）和最大后验估计（得到后验函数，求导取极值点）的求法。
- 区分样本均值的方差和样本的方差的不同，区贝叶斯估计中后验均方差与后验方差、后验标准误与后验标准差的不同。
- 理解假设检验中贝叶斯因子可以体现样本数据对原假设的支持程度

假设的形式	Bayes因子	说　明
简单假设对简单假设	$B=\dfrac{\alpha_0\pi_1}{\alpha_1\pi_0}=\dfrac{p(x\mid\theta_0)}{p(x\mid\theta_1)}$	B的大小表示样本支持θ_0的程度。
复杂假设对复杂假设	$B=\dfrac{\alpha_0\pi_1}{\alpha_1\pi_0}=\dfrac{m_0(x)}{m_1(x)}$	B可看作Θ_0与Θ_1上的加权似然比，它部分地消除了先验分布的影响，而强调了样本观察值的作用。
简单假设对复杂假设	$B=\dfrac{\alpha_0\pi_1}{\alpha_1\pi_0}=\dfrac{p(x\mid\theta_0)}{m_1(x)}$	在实际应用中，可以先计算B，再计算后验概率：$\alpha_0=\left[1+\dfrac{1-\pi_0}{\pi_0}\dfrac{1}{B}\right]^{-1}$

- 理解利用贝叶斯决策的方法来实现假设检验时，引入了贴合实际情况和目标的损失函数，从而提供了一个显著性水平的合理选取方法。
- 掌握收益函数$Q(\theta,a)$和损失函数$L(\theta,a)$的计算方式，区分 先验期望损失 与后验风险$R(a\mid\boldsymbol{x})$（损失函数$L(\theta,a)$对后验分布$\pi(\theta\mid\boldsymbol{x})$的期望）与完全信息期望EVPI（理想）
- 区分估计中的贝叶斯估计和决策中的贝叶斯估计的异同，在贝叶斯推断中定义了贝叶斯估计的概念，并没有结合实际的场景提出选取的方法而一般选取后验期望估计，决策中引入决策函数结合了具体的场景，并在行动集空间与参数空间同为某个实数集时，提出选取估计的依据为后验风险准则。
- 掌握计算贝叶斯决策中常用损失函数下的贝叶斯估计的方法：平方损失下的（均值或其变形）、绝对值损失下的（分位点）。$\frac{dR}{d\delta}=\frac{dE(L(\theta,\delta)|x)}{d\delta}=0$

参考书籍：《贝叶斯统计》

部分练习习题：1.2、1.4、1.8、**1.11**、1.13、1.19、**2.2**、2.5、**2.8**、**2.10**、3.12、5.12、例1.3.1

参考答案：https://wenku.baidu.com/view/4643d9c180eb6294dd886c66?fromShare=1

贝叶斯网络

- 贝叶斯网络是用来表示变量间连接概率的图形模式，能表示复杂联合概率分布的紧凑表示形式，它提供了一种自然的表示因果信息的方法，用来发现数据间的潜在关系。在这个网络中，用节点表示变量，有向边表示变量的依赖关系， 并使用条件概率表（CPT）来描述联合概率分布。

利用贝叶斯方法学习权重

贝叶斯方法考虑的权重空间（即权重的整个解空间上许多解）上权重的概率分布， 通过先验和观测的新的数据，不断更新后验概率，使得权重逐渐集中到权重空间的局部上。

- 先验选择：一般选择性质较好的高斯分布
- 贝叶斯巧妙的将最大似然技术（误差平方和最小）和正则化（权重衰减）技术结合起来。 当数据集较大，随着训练进行，上式中样本信息会随着增大并成为主导，同时先验的重要性逐渐减弱。当数据集较小的时候，先验在决定解就会起很重要作用。
- 很多情况下，积分是很难进行的，就需要通过近似方法来求解。 如使用泰勒展开保留2阶项来近似计算。

贝叶斯的简单运用

- 拼写检查
- 邮件分类