

第六章

贝叶斯估计

第一节 贝叶斯推断方法

第二节 贝叶斯决策方法

第一节 贝叶斯推断方法

一、统计推断中可用的三种信息

美籍波兰统计学家耐曼 (E. L. Lehmann 1894—1981) 高度概括了在统计推断中可用的三种信息：

1. **总体信息**，即总体分布或所属分布族给我们的信息。譬如“总体视察指数分布”或“总体是正态分布”在统计推断中都发挥重要作用，只要有总体信息，就要想方设法在统计推断中使用
2. **样本信息**，即样本提供我们的信息，这是任何一种统计推断中都需要

3. 先验信息，即在抽样之前有关统计推断的一些信息。譬如，在估计某产品的不合格率时，假如工厂保存了过去抽检这种产品质量的资料，这些资料（包括历史数据）有时估计该产品的不合格率是有好处的。这些资料所提供的信息就是一种先验信息。又如某工程师根据自己多年积累的经验对正在设计的某种彩电的平均寿命所提供的估计也是一种先验信息。由于这种信息是在“试验之前”就已有的，故称为先验信息。

以前所讨论的点估计只使用前两种信息，没有使用先验信息。假如能把收集到的先验信息也利用起来，那对我们进行统计推断是有好处的。只用前两种信息的统计学称为经典统计学，三种信息都用的统计学称为贝叶斯统计学。本节将简要介绍贝叶斯统计学中的点估计方法。

二、贝叶斯公式的密度函数形式

贝叶斯统计学的基础是著名的贝叶斯公式，它是英国学者贝叶斯（T. R. Bayes 1702~1761）在他死后二年发表的一篇论文《论归纳推理的一种方法》中提出的。经过二百年的研究与应用，贝叶斯的统计思想得到很大的发展，目前已形成一个统计学派——贝叶斯学派。为了纪念他，英国历史最悠久的统计杂志《Biometrika》在1958年又全文刊登贝叶斯的这篇论文。

初等概率论中的贝叶斯公式是用事件的概率形式给出的。可在贝叶斯统计学中应用更多的是贝叶斯公式的密度函数形式。下面结合贝叶斯统计学的基本观点来引出其密度函数形式。贝叶斯统计学的基本观点可以用下面三个观点归纳出来。

假设 I 随机变量 X 有一个密度函数 $p(x; \theta)$ ，其中 θ 是一个参数，不同的 θ 对应不同的密度函数，故从贝叶斯观点看， $p(x; \theta)$ 是在给定后 θ 是个条件密度函数，因此记为 $p(x | \theta)$ 更恰当一些。这个条件密度能提供我们的有关的 θ 信息就是总体信息。

假设II 当给定 θ 后，从总体 $p(x | \theta)$ 中随机抽取一个样本 X_1, \dots, X_n ，该样本中含有 θ 的有关信息。这种信息就是样本信息。

假设III 我们对参数 θ 已经积累了很多资料，经过分析、整理和加工，可以获得一些有关 θ 的有用信息，这种信息就是先验信息。参数 θ 不是永远固定在一个值上，而是一个事先不能确定的量。从贝叶斯观点来看，未知参数 θ 是一个随机变量。而描述这个随机变量的分布可从先验信息中归纳出来，这个分布称为先验分布，其密度函数用 $\pi(\theta)$ 表示。

1 先验分布

定义3.1 将总体中的未知参数 $\theta \in \Theta$ 看成一取值于 Θ 的随机变量，它有一概率分布，记为 $\pi(\theta)$ ，称为参数 θ 的先验分布。

2 后验分布

在贝叶斯统计学中，把以上的三种信息归纳起来的最好形式是在总体分布基础上获得的样本 X_1, \dots, X_n ，和参数的联合密度函数

$$p(x_1, \dots, x_n, \theta) = p(x_1, \dots, x_n | \theta) \pi(\theta)$$

在这个联合密度函数中。当样本 X_1, \dots, X_n 给定之后，未知的仅是参数 θ 了，我们关心的是样本给定后， θ 的条件密度函数，依据密度的计算公式，容易获得这个条件密度函数

$$\begin{aligned}\pi(\theta|x_1, \dots, x_n) &= \frac{p(x_1, \dots, x_n, \theta)}{p(x_1, \dots, x_n)} \\ &= \frac{p(x_1, \dots, x_n|\theta)\pi(\theta)}{\int p(x_1, \dots, x_n|\theta)\pi(\theta)d\theta}\end{aligned}$$

这就是贝叶斯公式的密度函数形式，其中 $\pi(\theta|x_1, \dots, x_n)$ 称为 θ 的后验密度函数，或后验分布。而

$$p(x_1, \dots, x_n) = \int p(x_1, \dots, x_n|\theta)\pi(\theta)d\theta$$

是样本的边际分布，或称样本 X_1, \dots, X_n 的无条件分布，它的积分区域就是参数 θ 的取值范围，随具体情况而定。

前面的分析总结如下：人们根据先验信息对参数 θ 已有一个认识，这个认识就是先验分布 $\pi(\theta)$ 。通过试验，获得样本。从而对 θ 的先验分布进行调整，调整的方法就是使用上面的贝叶斯公式，调整的结果就是后验分布 $\pi(\theta|x_1, \dots, x_n)$ 。后验分布是三种信息的综合。获得后验分布使人们对 θ 的认识又更进一步，可看出，获得样本的效果是把我们对 θ 的认识由 $\pi(\theta)$ 调整到 $\pi(\theta|x_1, \dots, x_n)$ 。所以对 θ 的统计推断就应建立在后验分布 $\pi(\theta|x_1, \dots, x_n)$ 的基础上。

例1 设事件A的概率为 θ ，即 $\pi(A) = \theta$ 。为了估计 θ 而作 n 次独立观察，其中事件出现次数为 X ，则有 X 服从二项分布 $b(n, \theta)$

$$\text{即 } P(X = x | \theta) = C_n^x \theta^x (1 - \theta)^{n-x}, x = 0, 1, \dots, n.$$

如果此时我们对事件A的发生没有任何了解, θ 对 θ 的大小也没有任何信息。在这种情况下，贝叶斯建议用区间 $(0, 1)$ 上的均匀分布作为先验分布。因为它在 $(0, 1)$ 上每一点都是机会均等的。这个建议被后人称为贝叶斯假设。

$$\pi(\theta) = \begin{cases} 1, & 0 < \theta < 1 \\ 0, & \text{others} \end{cases}$$

样本 \mathbf{X} 与参数的联合分布为

$$h(x, \theta) = C_n^x \theta^x (1 - \theta)^{n-x}, x = 0, 1, \dots, n, 0 < \theta < 1$$

此式在定义域上与二项分布有区别。再计算 \mathbf{X} 的边缘密度为

$$m(x) = \int_0^1 h(x, \theta) d\theta = C_n^x \frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}, x = 0, 1, n$$

$$\pi(\theta|x) = \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)} \theta^x (1-\theta)^{n-x}, 0 < \theta < 1$$

即 $\theta|X \sim Be(x+1, n-x+1)$

拉普拉斯计算过这个概率,研究男婴的诞生比例是否大于0.5?如抽了251527个男婴,女婴241945个

贝叶斯统计学首先要想方设法先去寻求 θ 的先验分布。先验分布的确定大致可分以下几步：

第一步，选一个适应面较广的分布族作先验分布族，使它在数学处理上方便一些，这里我们选用 β 分布族

$$\pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1}, 0 \leq \theta \leq 1, a > 0, b > 0$$

注： $\Gamma(s) = \int_0^{\infty} x^{s-1} e^{-x} dx, s > 0, \Gamma(n+1) = n!$

$$B(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx, p > 0, q > 0$$

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}, p > 0, q > 0$$

作为 θ 的先验分布族是恰当的，从以下几方面考虑：

1 参数 θ 是废品率，它仅在 $(0, 1)$ 上取值。因此，必需用区间 $(0, 1)$ 上的一个分布去拟合先验信息。 β 分布正是这样一个分布。

2 β 分布含有两个参数 a 与 b ，不同的 a 与 b 就对应不同的先验分布，因此这种分布的适应面较大

3 样本 X 的分布为二项分布 $b(n, \theta)$ 时，假如 θ 的先验分布为 β 分布，则用贝叶斯估计算得的后验分布仍然是 β 分布，只是其中的参数不同。这样的先验分布（ β 分布）称为参数 θ 的共轭先验分布。选择共轭先验分布在处理数学问题上带来不少方便。

4 国内外不少人使用 β 分布获得成功。

第二步，根据先验信息在先验分布族中选一个分布作为先验分布，使它与先验信息符合较好。利用 θ 的先验信息去确定 β 分布中的两个参数 a 与 b 。从文献来看，确定 a 与 b 的方法很多。例如，如果能从先验信息中较为准确地算得 θ 先验平均和先验方差，则可令其分别等于 β 分布的期望与方差最后解出 a 与 b 。

$$\frac{a}{a+b} = \bar{\theta}$$

$$\frac{ab}{(a+b)^2(a+b+1)} = S_{\theta}^2$$

$$\Rightarrow a = \frac{(1-\bar{\theta})\bar{\theta}^2}{S_{\theta}^2} - \bar{\theta}$$

$$b = \frac{a(1-\bar{\theta})}{\bar{\theta}}$$

如果从先验信息获得 $\bar{\theta} = 0.2, S_{\theta}^2 = 0.01$ ，则可解得 $a=3, b=12$ 这意味着 θ 的先验分布是参数 $a=3, b=12$ 的 β 分布。

假如我们能从先验信息中较为准确地把握 θ 的两个分位数，如确定 θ 确定的10%分位数 $\theta(0.1)$ 和50%的中位数 $\theta(0.5)$ ，那可以通过如下两个方程来确定 a 与 b 。

$$\begin{cases} \int_0^{\theta_{0.1}} \pi(\theta) d\theta = 0.1, \\ \int_0^{\theta_{0.5}} \pi(\theta) d\theta = 0.5. \end{cases}$$

假如的信息较为丰富，譬如对此产品经常进行抽样检查，每次都对废品率作出一个估计，把这些估计值看作的一些观察值，再经过整理，可用一个分布去拟合它。

假如关于的信息较少，甚至没有什么有用的先验信息，那可以用区间 $(0, 1)$ 上的均匀分布 ($a=b=1$ 情况)。用均匀分布意味着我们对的各种取值是“同等对待的”，是“机会均等的”。贝叶斯本人认为，当你对参数 θ 的认识除了在有限区间 (c, d) 之外，其它毫无所知时，就可用区间 (c, d) 上的均匀分布作为 θ 的先验分布。这个看法被后人称之为“贝叶斯假设”。

确定了先验分布后，就可计算出后验分布，过程如下

$$\begin{aligned}
 p(x, \theta) &= p(X = x | \theta) \pi(\theta) \\
 &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \binom{n}{x} \theta^{a+x-1} (1-\theta)^{b+n-x-1} \quad x=0, 1, \dots, n, 0 < \theta < 1
 \end{aligned}$$

于是X的边际分布为

$$p(x) = \int_0^1 p(x, \theta) d\theta = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+x)\Gamma(b+n-x)}{\Gamma(a+b+n)} \binom{n}{x}, x = 0, 1, \dots, n.$$

最后在给出 $X=x$ 的条件下， θ 的后验密度为

$$\pi(\theta|x) = \frac{p(x, \theta)}{p(x)} = \frac{\Gamma(a+b+n)}{\Gamma(a+x)\Gamma(b+n-x)} \theta^{a+x-1} (1-\theta)^{b+n-x-1}, 0 < x < 1$$

显然这个后验分布仍然是 β 分布，它的两个参数分别是 $a+x$ 和 $b+n-x$ 。我们选后验期望作为的贝叶斯估计，则 θ 的贝叶斯估计为

$$\hat{\theta}_B = \int_0^1 \theta \pi(\theta|x) d\theta = \frac{a+x}{a+b+n}$$

与前面的极大似然估计是不同的。

$$X \sim \beta(a, b), E(X) = \frac{a}{a+b}$$

如果用 $(0, 1)$ 上的均匀分布作为 θ 的先验分布, 则 θ 的贝叶斯估计为

$$\hat{\theta}_B = \frac{x+1}{n+2}$$

计算如下:

$$\begin{aligned} p(x, \theta) &= p(X = x | \theta) \pi(\theta) \\ &= C_n^x \theta^x (1-\theta)^{n-x}, x = 0, 1, \dots, n, 0 < \theta < 1 \end{aligned}$$

$$p(x) = \int_0^1 C_n^x \theta^x (1-\theta)^{n-x} d\theta = C_n^x \frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}$$

$$\pi(\theta|x) = \frac{p(x, \theta)}{p(x)} = \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)} \theta^x (1-\theta)^{n-x}, 0 < x < 1$$

后验分布为 $\beta(x+1, n-x+1)$

三、常用的一些共轭先验分布

对于一些常用的指数分布族，如果仅对其中的参数 θ 感兴趣，下表列出了它们的共轭先验分布及后验期望。

分 布	共 轭 先 验 分 布	后 验 分 布
正态分布 $N(\theta, \sigma^2)$	正态分布 $N(\mu, \tau^2)$	$\frac{\tau^2 x + \mu \sigma^2}{\tau^2 + \sigma^2}$
二项分布 $b(n, p)$	β 分布 $\beta(a, b)$	$\frac{a + x}{a + b + x + n}$
Poisson分布 $\pi(\theta)$	Γ 分布 Γ (a , b)	$\frac{a + x}{b + 1}$

EX1 设 θ 是一批产品的不合格率，已知它不是0.1就是0.2，且其先验分布为

$$\pi(0.1) = 0.7, \pi(0.2) = 0.3$$

假如从这批产品中随机取8个进行检查，发现有2个不合格，求 θ 的后验分布。

解： $P(X = 2|\theta) = C_8^2 \theta^2 (1-\theta)^6$

θ	0.1	0.2
P	0.7	0.3

$\theta X = 2$	0.1	0.2
P	0.542	0.458

$$\begin{aligned} P(\theta = 0.1|X = 2) &= \frac{P(X = 2|\theta = 0.1)P(\theta = 0.1)}{P(X = 2|\theta = 0.1)P(\theta = 0.1) + P(X = 2|\theta = 0.2)P(\theta = 0.2)} \\ &= \frac{1}{1 + \frac{C_8^2 0.2^2 0.8^6}{C_8^2 0.1^2 0.9^6} \cdot \frac{0.3}{0.7}} \\ &\approx 0.542 \end{aligned}$$

$$P(\theta = 0.2|X = 2) = 1 - P(\theta = 0.1|X = 2) = 0.458$$

EX2 设一卷磁带上的缺陷数服从泊松分布 $P(\lambda)$ 其中 λ 可取 1.0 和 1.5 中的一个, 又设 λ 的先验分布为

$$\pi(1.0) = 0.4 \quad \pi(1.5) = 0.6$$

假如检查一卷磁带发现了 3 个缺陷, 求 λ 的后验分布。

四、贝叶斯推断（估计）

I 条件方法

由于未知参数的后验分布是集三种信息（总体、样本和后验）于一身，它包含了所有可供利用的信息。故有关的参数估计和假设检验等统计推断都按一定方式从后验分布提取信息，其提取方法与经典统计推断相比要简单明确得多。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）而认为未出现的数据与推断无关，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方法被称为条件方法。

例如经典统计学认为参数的无偏估计应满足：
其中平均是对样本空间中所有可能出现的样本而求的，可实际中样本空间中绝大多数样本尚未出现过，而多数从未出现的样本也要参与平均是实际工作者难以理解的。故在贝叶斯推断中不用无偏性，而条件方法是容易被实际工作者理解和接受的。

II 估计

1. 贝叶斯估计

定义3.2 使后验密度 $\pi(\theta|x)$ 达到最大的值 θ_{MD} 称为最大后验估计；后验分布的中位数 $\hat{\theta}_{Me}$ 称为后验中位数估计；后验分布的期望值 $\hat{\theta}_E$ 称为 θ 的后验期望值估计，这三个估计都称为贝叶斯估计，记为 $\hat{\theta}_B$ 。

例1 为估计不合格率 θ ，今从一批产品中随机抽取 n 件，其中不合格品数 X 服从 $B(n, p)$ ，一般选取 $Be(\alpha, \beta)$ 为 θ 的先验分布，设 α, β 已知，由共轭先验分布可知， θ 的后验分布为

$$Be(\alpha + x, \beta + n - x)$$

可计算得：
$$\hat{\theta}_{MD} = \frac{\alpha + x - 1}{\alpha + \beta + n - 2}, \quad \hat{\theta}_E = \frac{\alpha + x}{\alpha + \beta + n}$$

我们选用贝叶斯假设 $\alpha = \beta = 1$ 则

$$\hat{\theta}_{MD} = \frac{x}{n}, \quad \hat{\theta}_E = \frac{x+1}{n+2}$$

第一、在二项分布时， θ 的最大后验估计就是经典统计中的极大似然估计，即 θ 的极大似然估计就是取特定的先验分布下的贝叶斯估计。

第二、 θ 的后验期望值估计 $\hat{\theta}_E$ 要比最大后验估计 θ_{MD} 更合适一些。

第三、 θ 的后验期望值估计要比最大后验估计更合适一些。表2.1列出四个实验结果,在试验1与试验2中,“抽检3个产品没有一件不合格”与抽检10个产品没有一件是不合格”这两件事在人们心目中留下的印象是不同的。后者的质量要比前者的质量更信得过。

表3.1 不合格率 θ 的二种贝叶斯估计的比较

试验号	样本量n	不合格数x	$\hat{\theta}_{MD} = \frac{x}{n}$	$\hat{\theta}_E = \frac{x+1}{n+2}$
1	3	0	0	0.200
2	10	0	0	0.083
3	3	3	1	0.800
4	10	10	1	0.917

在试验3和誓言4中，“抽检3个产品全部不合格”与抽检“10个产品全部不合格”也是有差别的。在实际中，人们经常选用后验期望估计作为贝叶斯估计。

2. 贝叶斯估计的误差

设 $\hat{\theta}$ 是 θ 的一个贝叶斯估计，在样本给定后， $\hat{\theta}$ 是一个数，在综合各种信息后， θ 是按 $\pi(\theta|x)$ 取值，所以评价一个贝叶斯估计的误差的最好而又简单的方式是用 θ 对 $\hat{\theta}$ 的后验均方差或平方根来度量，定义如下：

定义3.2 设参数 θ 的后验分布为 $\pi(\theta|x)$, 贝叶斯估计为 $\hat{\theta}$, 则 $\hat{\theta}$ 的后验期望

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

称为 $\hat{\theta}$ 的后验均方差, 而其平方根称为后验标准误.

当 $\hat{\theta}_E = E(\theta|x)$ 时, 则 $MSE(\hat{\theta}_E|x) = E^{\theta|x}(\theta - \hat{\theta}_E)^2 = Var(\theta|x)$, 称为后验均方差.

后验均方差与后验方差有如下关系:

$$\begin{aligned}MSE(\hat{\theta}|x) &= E^{\theta|x}(\theta - \hat{\theta})^2 \\&= E^{\theta|x}[(\theta - \hat{\theta}_E) + (\hat{\theta}_E - \hat{\theta})]^2 \\&= \text{Var}(\theta|x) + (\hat{\theta}_E - \hat{\theta})^2\end{aligned}$$

这表明,当 $\hat{\theta} = \hat{\theta}_E$ 时,可使后验均方差达到最小,实际中常取后验均值作为 θ 的贝叶斯估计值.

例2 设一批产品的不合格率为 θ ,检查是一个一个进行,直到发现第一个不合格品为止,若 X 为发现第一个不合格品时已检查的产品数,则 X 服从几何分布,其分布列为

$$P(X = x|\theta) = \theta(1-\theta)^{x-1}, x = 1, 2, \dots$$

设 θ 的先验分布为 $P(\theta = \frac{i}{4}) = \frac{1}{3}, i = 1, 2, 3$, 如今只获得一个样本观察值 $x=3$, 求 θ 的最大后验估计, 后验期望估计, 并计算它的误差. 故联合分布为

$$P(X = 3, \theta = \frac{i}{4}) = \frac{1}{3} \cdot \frac{i}{4} \cdot (1 - \frac{i}{4})^2$$

$X=3$ 的无条件概率为(利用全概率公式)

$$P(X = 3) = \frac{1}{3} \left[\frac{1}{4} \left(\frac{3}{4}\right)^2 + \frac{2}{4} \left(\frac{2}{4}\right)^2 + \frac{3}{4} \left(\frac{1}{4}\right)^2 \right] = \frac{5}{48}$$

故
$$P(\theta = i/4 | X = 3) = \frac{P(X = 3, \theta = i/4)}{P(X = 3)} = \frac{4i}{5} \left(1 - \frac{i}{4}\right)^2, i = 1, 2, 3$$

θ	1/4	2/4	3/4
$P(\theta = i/4 X = 3)$	9/20	8/20	3/20

可看出, 的最大后验估计 $\hat{\theta}_{MD} = 1/4$

$$\hat{\theta}_E = E(\theta | X = 3) = 17/40$$

$$\text{Var}(\theta | x) = E(\theta^2 | x) - E^2(\theta | x) = 17/80 - (17/40)^2 = 51/1600$$

$\hat{\theta}_{MD}$ 的后验方差为

$$\text{MSE}(\hat{\theta} | x) = \text{Var}(\theta | x) + (\hat{\theta}_E - \hat{\theta})^2 = 51/1600 + (1/4 - 17/40)^2 = \frac{1}{16}$$

3.区间估计(可信区间)

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,而经典方法求置信区间常受到批评.

定义3.3 参数 θ 的后验分布为 $\pi(\theta|x)$,对给定的样本 x 和概率 $1-\alpha(0 < \alpha < 1)$,若存在这样的二个统计量 $\hat{\theta}_L = \hat{\theta}_L(x)$ 与 $\hat{\theta}_U = \hat{\theta}_U(x)$ 使得

$$P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U | x) \geq 1 - \alpha$$

则称区间 $[\hat{\theta}_L, \hat{\theta}_U]$ 为参数的可信水平为 $1-\alpha$ 贝叶斯可信区间,或简称为 θ 的 $1-\alpha$ 可信区间.而满足

$P(\theta \geq \hat{\theta}_L | x) \geq 1-\alpha$ 的 $\hat{\theta}_L$ 称为 θ 的 $1-\alpha$ (单侧)可信下限.

满足 $P(\theta \leq \hat{\theta}_U | x) \geq 1-\alpha$ 的 $\hat{\theta}_U$ 称为 θ 的 $1-\alpha$ (单侧)可信上限.

这里的可信水平和可信区间与经典统计中的置信水平与置信区间虽是同类的概念,但两者还是有本质的差别,主要表现在下面二点:

1. 在条件方法下,对给定的样本和可信水平,通过后验分布可求得具体的可信区间,譬如,的可信水平为0.9的可信区间是,这时我们可以写出

$$P(1.5 \leq \theta \leq 2.6 | x) = 0.9$$

2.在经典统计中寻求置信区间有时是困难的,因为它要设法构造一个枢轴量,使它的分布不含未知参数,这是一项技术性很强的工作.相比之下可信区间只要利用后验分布,不需要再去寻求另外的分布,可信区间的寻求要简单得多.

例3 设 x_1, x_2, \dots, x_n 是来自正态总体 $N(\theta, \sigma^2)$ 的一个样本观察值,其中 σ^2 已知,若正态均值的先验分布取为 $N(\mu, \tau^2)$, 其中与 τ 已知,则可求得 θ 的后验分布为 $N(\mu_1, \sigma_1^2)$, 由此很容易获得 θ 的 $1-\alpha$ 可信区间

$$P(\mu_1 - \sigma_1 \mu_{1-\alpha/2} \leq \theta \leq \mu_1 + \sigma_1 \mu_{1-\alpha/2}) = 1 - \alpha$$

EX1 设随机变量 X 的密度函数为 $p(x|\theta) = \frac{2x}{\theta^2}, 0 < x < \theta < 1$

(1) 假如 θ 的先验分布为 $U(0, 1)$, 求 θ 的后验分布.

(2) 假如 θ 的先验分布为

$$\pi(\theta) = 3\theta^2, 0 < \theta < 1$$

求 θ 的后验分布及后验期望估计

$$(1) \pi(\theta|x) = \frac{x}{(1-x)\theta^2}, x < \theta < 1, \hat{\theta}_E = \frac{x \ln x}{x-1}$$

$$(2) \pi(\theta|x) = \frac{1}{(1-x)}, x < \theta < 1, \hat{\theta}_E = \frac{1+x}{2}$$

EX2 对正态分布 $N(0,1)$ 观察,获得三个观察值

$$x_1 = 2, x_2 = 3, x_3 = 4$$

若 θ 的先验分布为 $N(3,1)$,求 θ 的**0.95**可信区间

第二节 贝叶斯决策方法

决策就是对一件事要作决定.它与推断的差别在于是否涉及后果.统计学家在作推断时是按统计理论进行的,很少考虑结论在使用后的损失.可决策者在使用推断结果时必需与得失联系在一起,能带来利润的就会用,使他遭受损失的就不会被采用,度量得失的尺度就是损失函数.它是著名的统计学家A.Wald(1902-1950)在40年代引入的一个概念.从实际归纳出损失函数是决策的关键.

贝叶斯决策:把损失函数加入贝叶斯推断就形成贝叶斯决策论,损失函数被称为贝叶斯统计中的第四种信息.

一、决策的基本概念

例1 设甲乙二人进行一种游戏,甲手中有三张牌,分别标以 $\theta_1, \theta_2, \theta_3$. 乙手中也有三张牌, 分别标以 a_1, a_2, a_3 . 游戏的规则是双方各自独立地出牌,按下表计算甲的得分与乙的得分.

甲的得分矩阵(乙的失分矩阵)

	3	-2	0
	1	4	-3
	-4	-1	2

这是一个典型的双人博弈(赌博)问题.不少实际问题可归结为双人博弈问题.把上例中的乙方改为自然或社会,就形成人与自然(或社会)的博弈问题.

例2 某农作物有两个品种:产量高但抗旱能力弱的品种 a_1 和抗旱能力强但产量低的品种 a_2 .在明年雨量不知的情况下,农民应选播哪个品种可使每亩平均收益最大?这是人与自然界的博弈.以明年60mm雨量为界来区分雨量充足 θ_1 和雨量不充足 θ_2 .写出收益矩阵(单位:元)

	a_1	a_2
θ_1	1000	200
θ_2	-200	400

例3一位投资者有一笔资金要投资.有以下几个投资向供他选择:

a_1 :购买股票,根据市场情况,可净赚5000元,但也可能亏损10000元;

a_2 :存入银行,不管市场情况如何总可净赚1000元.

这位投资者在与金融市场博弈.未来的金融市场也有二种情况:看涨 θ_1 与看跌 θ_2 可写出投资者的收益矩阵

	a_1	a_2
θ_1	5000	1000
θ_2	-10000	1000

二、决策的三要素

1. 状态集 $\Theta = \{\theta\}$, 其中每个元素 θ 表示自然界(或社会)可能出现的一种状态, 所有可能状态的全体组成状态集.

2. 行动集 $A = \{a\}$, 其中 a 表示人对自然界可能采取的一个行动.

一般行动集有两个以上的行动供选择. 若有两个行动无论对自然界的哪一个状态出现, a_1 总比 a_2 收益高, 则 a_2 就没有存在的必要, 可把它从行动集中去掉, 使留在行动集中的行动总有可取之处.

3. 收益函数 $Q(\theta, a)$. 函数值 $Q(\theta_i, a_j) = Q_{ij}$ 表示当自然界处于状态 θ_i , 而人们选取行动 a_j 时所得到的收益大小.

收益函数的值可正可负, 其正表示赢利, 负表示亏损, 单位常用货币单位。收益函数的建立不是件容易的事, 要对所研究的问题有全面的了解才能建立起来。收益矩阵

$$Q = \begin{pmatrix} Q_{11} & Q_{12} & \cdots & Q_{1m} \\ Q_{21} & Q_{22} & \cdots & Q_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{n1} & Q_{n2} & \cdots & Q_{nm} \end{pmatrix}$$

三、损失函数

1. 从收益到损失

为了以后的统一处理，在决策中常用一个更为有效的概念：损失函数。在状态集和行动集都为有限时用损失矩阵。

这里的损失函数不是负的收益，也不是亏损。例如，某商店一个月的经营收益为-1000元，即亏1000元。这是对成本而言。我们不称为损失，而称其为亏损。我们讲的损失是指“该赚而没有赚到的钱”，例如该商店本可以赚2000元，但由于某种原因亏了1000元，那我们说该商店损失了3000元。用这种观点认识损失对提高决策意识是有好处的。

按上述观点从收益函数可以很容易获得损失函数。

例4 某公司购进某种货物可分大批、中批和小批三种行动，记为 a_1, a_2, a_3 .未来市场需求量可分为高、中、低三种状态，记为 $\theta_1, \theta_2, \theta_3$.三个行动在不同市场的利润如下：

$$Q = \begin{pmatrix} 10 & 6 & 2 \\ 3 & 4 & 2 \\ -2.7 & -0.8 & 1 \end{pmatrix}$$

这是一个收益矩阵,我们把它改写成损失矩阵如下:

$$L = \begin{pmatrix} 0 & 4 & 8 \\ 1 & 0 & 2 \\ 3.7 & 1.8 & 0 \end{pmatrix}$$

2. 损失函数

构成决策问题的三要素: $\Theta = \{\theta\}$ $A = \{a\}$ $L(\theta, a)$

由收益函数容易获得损失函数

$$L(\theta, a) = \max_{a \in A} Q(\theta, a) - Q(\theta, a)$$

例5 某公司购进一批货物投放市场,若购进数量 a 低于市场需求量 θ ,每吨可赚15万元,若购进数量 a 超过市场需求量 θ ,超过部分每吨反而要亏35万元.由此可写出收益函数

$$Q(\theta, a) = \begin{cases} 15a, & a \leq \theta \\ 15\theta - 35(a - \theta), & a > \theta \end{cases}$$

显然,当购进数量 a 等于市场需求量 θ 时,收益达到最大为
15 · θ

$$L(\theta, a) = \begin{cases} 15(\theta - a), & a \leq \theta \\ 35(a - \theta), & a > \theta \end{cases}$$

3. 损失函数下的悲观准则

第一步,对每个行动 a ,选出最大损失值,记为 $\max_{\theta \in \Theta} L(\theta, a), a \in A$

a' a'

$$\min_{a \in A} \max_{\theta \in \Theta} L(\theta, a) = \max_{\theta \in \Theta} L(\theta, a')$$

则称 a' 为悲观准则下的最优行动.这是一种保守策略.不求零损失,但愿少损失.

例6某公司购进某种货物可分大批、中批和小批三种行动，记为， θ_1 θ_2 θ_3 未来市场需求量可分为高、中、低三种状态，记为， a_1 a_2 a_3 ，三个行动在不同市场的利润如下：

$$Q = \begin{pmatrix} 10 & 6 & 2 \\ 3 & 4 & 2 \\ -2.7 & -0.8 & 1 \end{pmatrix}$$

这是一个收益矩阵，我们把它改写成损失矩阵如下：

$$L = \begin{matrix} & \begin{matrix} a_1 & a_2 & a_3 \end{matrix} \\ \begin{pmatrix} 0 & 4 & 8 \\ 1 & 0 & 2 \\ 3.7 & 1.8 & 0 \end{pmatrix} \end{matrix}$$

在悲观准则下,第一步:行动 a_1, a_2, a_3

的最大损失值依此为3.7,4,8.

第二步,在上面三个最大损失值中最小值为3.7,而对应的行动为 a_1

4. 常用损失函数

(1) 平方损失函数 $L(\theta, a) = (a - \theta)^2$ $L(\theta, a) = \lambda(\theta)(a - \theta)^2$

这是在统计决策中用得最多的损失函数.

$$L(\theta, a) = \begin{cases} k_0(\theta - a), & a \leq \theta \\ k_1(a - \theta), & a \geq \theta \end{cases}$$

(3) 0-1损失函数

$$L(\theta, a) = \begin{cases} 0, & |\theta - a| \leq \varepsilon \\ 1, & |\theta - a| \geq \varepsilon \end{cases}$$

(4) 多元二次损失函数

$$L(\theta, a) = \sum_{i=1}^p \omega_i (a_i - \theta_i)^2$$

四、贝叶斯决策问题

先验信息和抽样信息都用的决策问题称为贝叶斯决策问题。若以下条件已知，则我们认为一个贝叶斯决策问题给定了。

$$(1) \quad X \sim p(x|\theta), \theta \in \Theta \quad (2) \quad \theta \sim \pi(\theta)$$

$$(3) \quad A = \{a\}$$

(4) 定义在 $\Theta \times A$ 上的二元函数 $L(\theta, a)$ 称为损失函数

1. 后验风险函数

我们把损失函数 $L(\theta, a)$ 对后验分布 $\pi(\theta|x)$ 的期望称为后验风险, 记 $R(a|x)$, 即

$$R(a|x) = E^{\theta|x} [L(\theta, a)] = \begin{cases} \sum_i L(\theta_i, a) \pi(\theta_i|x) \\ \int_{\Theta} L(\theta, a) \pi(\theta|x) d\theta \end{cases}$$

后验风险就是用后验分布计算的平均损失.

2.决策函数

定义3.5 在给定的贝叶斯决策问题中,从样本空间 $\mathcal{X} = \{x = (x_1, x_2, \dots, x_n)\}$ 到行动集A上的一个映照 $\delta(x)$ 称为该决策问题的一个决策函数, $D = \{\delta(x)\}$ 表示所有样本空间到A上的决策函数组成的类称为决策函数类.

在贝叶斯决策中我们面临的是决策函数类D,要在D中选择决策函数 $\delta(x)$,使其风险最小.

3. 后验风险准则

定义 在给定的贝叶斯决策问题中 $D = \{\delta(x)\}$ 是其决策函数称

$$R(\delta|x) = E^{\theta|x}[L(\theta, \delta(x))], x \in \mathcal{X}, \theta \in \Theta$$

为决策函数 $\delta = \delta(x)$ 的后验风险. 假如在决策函数类中存在这样的决策函数, 它在 D 中有最小的风险, 即

$$R(\delta'|x) = \min_{\delta \in D} R(\delta|x)$$

则称 $\delta' = \delta'(x)$ 为后验风险准则下的最优决策函数, 或称贝叶斯决策, 或贝叶斯解.

4. 平方损失函数下的贝叶斯估计

定理3.1 在平方损失函数 $L(\theta, \delta) = (\delta - \theta)^2$ 下, θ 的贝叶斯估计为后验均值, 即

$$\delta_B(x) = E(\theta|x)$$

[Pr] 在平方损失函数下, 任何一个决策函数 $\delta = \delta(x)$ 的后验风险为

$$E[(\delta - \theta)^2|x] = \delta^2 - 2\delta E(\theta|x) + E(\theta^2|x)$$

$$2\delta - 2E(\theta|x) = 0$$

$$\delta = E(\theta|x)$$