基础数理统计（研究生公共课）

肖柳青 教授 博士主讲

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

主要内容

1. 点估计的基本概念
2. 两种基本的点估计方法
3. 点估计的优良标准（无偏性及有效估计和C-R下界）
4. 置信区间估计
5. 最佳点估计（充分统计量）
6. 贝叶斯估计

2.6 贝叶斯估计（Bayes Estimation）

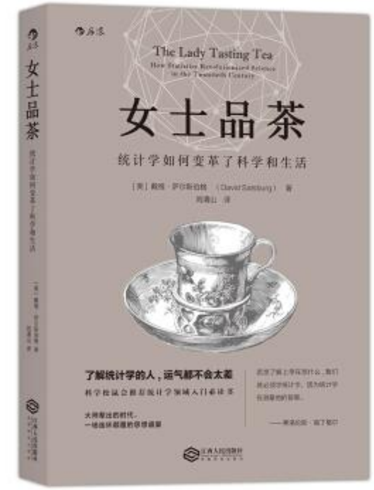
- 经典学派的观点：统计推断是根据样本信息对总体分布或总体的特征数进行推断，这里用到两种信息：总体信息和样本信息；
- 贝叶斯学派的观点：除了上述两种信息以外，统计推断还应该使用第三种信息：先验信息。

（1）**总体信息**:总体分布提供的信息。

（2）**样本信息**:抽取样本所得观测值提供的信息。

（3）**先验信息**:人们在试验之前对要做的问题在经验上和资料上总是有所了解的，这些信息对统计推断是有益的。先验信息即是抽样（试验）之前有关统计问题的一些信息。一般说来，先验信息来源于经验和历史资料。先验信息在日常生活和工作中是很重要的。

举例：



某人打靶，打了5枪，枪枪中靶，

问：此人枪法如何？

某人打靶，打了500枪，枪枪中靶，

问：此人枪法如何？

经典方法：极大似然估计：100%

但是：……

举例

英国统计学家Savage, L. J. 曾考察了如下两个统计试验:

(1) 一位常饮牛奶加茶的妇女声称, 她能分辨出先倒进杯子里的是茶还是牛奶.对此做了十次试验, 她都正确地说出来.

(2) 一位音乐家声称, 他能从一页乐谱辨别出是海顿(Haydn) 还是莫扎特(Mozart) 的作品. 在十次这样的试验中, 他都分辨正确.

在这两个统计试验中, 假如认为被试验者是在猜测, 每次成功概率为0.5, 那么十次都猜中的概率为$2^{-10} = 0.0009766$. 这是一个小概率事件,是几乎不可能发生的. 因此不能认为是猜测,而是他们的经验帮了他们的忙.

贝叶斯统计学

基于上述三种信息进行统计推断的统计学称为贝叶斯统计学。它与经典统计学的差别就在于是否利用先验信息。

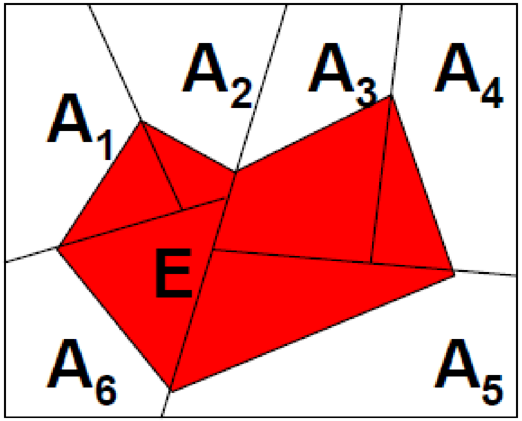
贝叶斯统计在重视使用总体信息和样本信息的同时，还注意先验信息的收集、挖掘和加工，使它数量化，形成先验分布，参加到统计推断中来，以提高统计推断的质量。忽视先验信息的利用，有时是一种浪费，有时还会导出不合理的结论。

贝叶斯学派的基本观点

贝叶斯学派的基本观点：任一未知量θ 都可看作随机变量，可用一个概率分布去描述，这个分布称为先验分布；在获得样本之后，总体分布、样本与先验分布通过贝叶斯公式结合起来得到一个关于未知量θ 新的分布—后验分布；任何关于θ 的统计推断都应该基于θ 的后验分布进行。

贝叶斯统计模型

$$p(A_i \mid B)=\frac{p(B \mid A_i)p(A_i)}{p(B)}=\frac{p(B \mid A_i)p(A_i)}{\sum_i p(B \mid A_i)p(A_i)}$$



i=1…n

- 设事件A_1， A_2， …， A_n构成互不相容的事件组，贝叶斯公式如上给出，先验信息以{$P[A_i]$, **i=1…n**}给出，即先验分布。由于事件B的发生，可以对A_1，A_2， …， A_n发生的概率重新估计。
- 贝叶斯公式综合了先验信息与试验提供的新信息，获得了后验信息，以后验概率{$P(A_i|B)$, **i=1…n**}体现出来，**贝叶斯公式反映了先验分布向后验分布的转化**。

复习：条件概率

1. 定义　设E为随机试验，Ω为其样本空间，A、B为任意两个事件，　若 $P(A)>0$

则称

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

为事件A出现的情况下，事件B的*条件概率*，或简称事件B关于事件A的*条件概率*。

2. 基本公式

定理1（乘法公式）

假设 $A_1, A_2, \cdots, A_n$ 为任意**n**个事件（ $n \geq 2$），

若 $$P(A_1A_2\cdots A_n) > 0$$

则

$$P(A_1A_2\cdots A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1A_2)\cdots P(A_n \mid A_1A_2\cdots A_{n-1})$$

定理2（全概率公式与贝叶斯公式）

设事件 $B_1, B_2, \cdots, B_n$ 两两互不相容，$\bigcup_{i=1}^{n} B_i = \Omega$

$P(B_i) > 0 \quad i = 1,2,\cdots,n$

则（1）对任意事件A，有

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i)$$

（2）对任意事件A，若 $P(A) > 0$，有

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}$$

3．分布密度

连续型随机变量

如果对于随机变量*X*的分布函数为*F*（*x*），存在非负的函数*f*（*x*）,使对任意的实数*x*有

$$F(x)=\int_{-\infty}^{x} f(t)dt$$

则称*X*为连续型随机变量，*f*（*x*）称为*X*的概率密度，且满足

$$f(x)\geq 0 \qquad \int_{-\infty}^{+\infty} f(x)dx=1$$

二、随机变量的联合分布

1. 联合分布函数

设 $X_1, X_2, \cdots, X_n$ 是样本空间Ω的***n***个随机变量，$x_1, x_2, \cdots, x_n$ 为任意实数，则称

$$F(x_1, x_2, \cdots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \cdots, X_n \le x_n)$$

为随机变量的***n***维联合分布函数 。

特别地 $$F(x, y) = P(X \le x, Y \le y)$$

即是***X***，***Y***的二维联合分布函数。

2．二维分布密度

连续型

如果存在一个非负的二元函数**f（x，y）**,使对任意的实数x，y有

$$F(x, y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dudv$$

则称（X，Y）为二维连续型随机变量，f（x，y）称为（X，Y）的概率密度，满足：

$$f(x,y)\geq 0 \qquad \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy=1$$

2．二维分布密度

离散型

设（*X*，*Y*）所有可能的取值为 (x_i, y_j)，而 p_{ij} 是（*X*，*Y*）取值 为 (x_i, y_j) 的概率，即

$$P(X = x_i, Y = y_j) = p_{ij}(i = 1, 2, \cdots \ j = 1, 2, \cdots)$$

则称上式为二维离散型随机向量（*X*，*Y*）的*联合分布律*。它满足

$$p_{ij} \geq 0 \qquad \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} p_{ij} = 1$$

离散型　若随机变量（*X*，*Y*）的联合分布律

$$P(X = x_i, Y = y_j) = p_{ij}$$

则

$$p_{i\bullet} = P(X = x_i) = \sum_{j=1}^{\infty} p_{ij} \quad (i = 1, 2, \cdots$$

$$p_{\bullet j} = P(Y = y_j) = \sum_{i=1}^{\infty} p_{ij} \qquad j = 1, 2, \cdots)$$

分别称为（*X*，*Y*）关于*X*和*Y*的边缘分布律。

*X和Y相互独立*的充要条件是　$p_{ij} = p_{i\bullet} \cdot p_{\bullet j}$

3．边缘分布及独立性

边缘分布

设（*X*，*Y*）的分布函数为 $F(x，y)$，则*X*，*Y*的分布函数$F_X(x)$、$F_Y(y)$，依次称为关于*X*和关于*Y*的边缘分布函数，且有

$$F_X(x)=F(x,+\infty) \qquad F_Y(y)=F(+\infty,y)$$

独 立 性

$$F(x，y)=F_X(x)\cdot F_Y(y)$$

则称随机变量*X*和*Y*是相互独立的。

连续型

若随机变量（*X*，*Y*）的概率密度为 $f(x,y)$

则

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx$$

分别称为（*X*，*Y*）关于*X*和*Y*边缘概率密度。

*X和Y相互独立*的充要条件是 $f(x,y) = f_X(x)\ f_Y(y)$

条件期望

一、条件期望的定义

离散型

$$E(X \mid Y=y_j) = \sum_{i=1}^{\infty} x_i P(X=x_i \mid Y=y_j)$$

其中 $$P(X=x_i \mid Y=y_j) = \frac{P(X=x_i, Y=y_j)}{P(Y=y_j)}$$

连续型

$$E(X \mid Y=y) = \int_{-\infty}^{+\infty} x f(x \mid y) dx$$

其中 $f(x \mid y)$ 条件概率密度

全数学期望公式

$E(X|Y)$ 是随机变量**Y**的函数，当 $Y=y$ 时取值 $E(X|Y=y)$ 因而它也是随机变量。

定理1 对一切随机变量**X**和**Y**，有

$$E(X)=E[E(X|Y)]$$

离散型

$$E(X)=\sum_{j=1}^{\infty}E(X|Y=y_j)P(Y=y_j)$$

连续型

$$E(X)=\int_{-\infty}^{+\infty}E(X|Y=y)f_Y(y)dy$$

定理1证明: $\mathrm{E(X)} = E[E(X|Y)]$

- 证：设（X, Y）的密度函数为$f(x,y)$，Y的边缘密度为 $f_Y(y)$
- 记 $g(y)=E(X|Y=y), g(Y)=E(X|Y)$
- 注意到 $g(y)=E(X|Y=y)=\int_{-\infty}^{+\infty} xf(x|y)dx$
- 由于 $f(x,y)=f_Y(y)f(x|y)$

$$EX=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf(x,y)dxdy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf_Y(y)f(x|y)dxdy$$

$$=\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} xf(x|y)dx\right)f_Y(y)dy=\int_{-\infty}^{+\infty} g(y)f_Y(y)dy$$

$$=E(g(Y))=E(E(X|Y))$$

*方差分解公式

定理2 对一切随机变量**X**和**Y**，有

$$Var(X)=Var(E(X|Y))+E(Var(X|Y))$$

- 首先给出条件方差的直观理解：
- 对于随机变量**X**，记 $g(Y)=E(X|Y)$，它是**Y**的函数，
- 定义 $\varepsilon=X-g(Y)=X-E(X|Y)$
- 由定理1 $\because E(g(Y))=EX$

$$\therefore E\varepsilon=EX-E(g(Y))=EX-E(E(X|Y))=0$$

- 故 $\varepsilon=X-g(Y)$ 可写成回归形式

$$X=g(Y)+\varepsilon=E(X|Y)+\varepsilon$$

定理2证明：$Var(X)=Var(E(X|Y))+E(Var(X|Y))$

- 证：下面来计算**X**的方差

$$Var(X)=Var(g(Y)+\varepsilon)=Var(g(Y))+Var(\varepsilon)+2cov(g(Y),\varepsilon)$$

$cov(g(Y),\varepsilon)=0$

$$\because cov(g(Y),\varepsilon)=E(g(Y)-E(g(Y))(\varepsilon-E\varepsilon))$$

$$=E(g(Y)\varepsilon)-E(g(Y))E\varepsilon-E\varepsilon E(g(Y))+E(g(Y))E\varepsilon$$

$$=E(g(Y)\varepsilon)-E(g(Y))E\varepsilon=E(g(Y)\varepsilon)$$

$$=E(g(Y)|Y)E(\varepsilon|Y)=E(g(Y))E(\varepsilon)=0$$

$$\therefore Var(X)=Var(g(Y))+Var(\varepsilon)=Var(E(X|Y))+Var(\varepsilon)$$

续

$$\therefore Var(\varepsilon)=E(X-g(Y))^2$$

$$=E(X^2+g(Y)^2-2Xg(Y))$$

$$=E(E(X^2|Y))-g(Y)^2$$

$$=E(Var(X|Y))$$

- 注【1】

$$EX=E(E(X|Y))$$

$$EX^2=E(E(X^2|Y))$$

$$EX=Eg(Y)$$

$$EX^2=Eg(Y)^2$$

- 注【2】

$$Var(X|Y)=E((X-E(X|Y))^2|Y)$$

$$=E(X^2|Y)-(E(X|Y))^2$$

贝叶斯估计的思路：

- 1、未知参数视为随机变量： θ
  - 数据的不可设计性与经验的不能穷尽性？
- 2、取样本$x_1 \ldots x_n$，求联合分布密度
  - $f(x_1,x_2,..x_n\,;\,\theta)$， θ是参数
- 3、联合分布密度—>条件分布密度
  - $f(x_1,x_2,..x_n \mid \theta)$， θ是随机变量
- 4、确定θ的先验分布$\pi(\theta)$
- 5、利用贝叶斯公式求后验分布密度
- 6、使用后验分布做推断（参数估计、假设检验）

贝叶斯公式的密度函数形式

- 总体依赖于参数$\boldsymbol{\theta}$的概率函数在贝叶斯统计中记为$f(x\,|\,\boldsymbol{\theta})$，它表示在随机变量$\theta$取某个给定值时总体的条件概率函数；
- 根据参数$\boldsymbol{\theta}$的先验信息可确定先验分布$\pi(\boldsymbol{\theta})$；
- 从贝叶斯观点看，样本 $x_1, x_2, \ldots, x_n$ 的产生分两步进行：首先从先验分布$\pi(\boldsymbol{\theta})$产生一个样本$\boldsymbol{\theta}_0$，然后从$f(x\,|\boldsymbol{\theta}_0)$中产生一组样本。这时样本的联合条件概率函数
- $$f(x_1,\cdots,x_n|\theta_0)=\prod_{i=1}^{n} f(x_i|\theta_0)$$
- 这个分布综合了总体信息和样本信息；

- $\boldsymbol{\theta}_0$ 是未知的，它是按先验分布$\pi(\theta)$产生的。为把先验信息综合进去，不能只考虑$\boldsymbol{\theta}_0$，对θ的其它值发生的可能性也要加以考虑，故要用$\pi(\boldsymbol{\theta})$进行综合。这样一来，样本$x_1, \ldots, x_n$和参数$\boldsymbol{\theta}$的联合分布为:

$$f(x_1, x_2, \ldots, x_n, \boldsymbol{\theta}) = f(x_1, x_2, \ldots, x_n | \boldsymbol{\theta})\pi(\boldsymbol{\theta}),$$

这个联合分布把**总体信息**、**样本信息**和**先验信息**三种可用信息都综合进去了；

➢在没有样本信息时，人们只能依据先验分布对θ作出推断。在有了样本观察值$x_1, x_2, \ldots, x_n$之后，则应依据$f(x_1, x_2, \ldots, x_n, \theta)$对$\theta$作出推断。由于

$$f(x_1,x_2,\ldots,x_n,\theta)=\pi(\theta\,|\,x_1,x_2,\ldots,x_n)\,f(x_1,x_2,\ldots,x_n),$$

其中

$$f(x_1,\cdots,x_n)=\int_\Theta f(x_1,\cdots,x_n,\theta)d\theta=\int_\Theta f(x_1,\cdots,x_n|\theta)\pi(\theta)d\theta$$

是$x_1, x_2, \ldots, x_n$的边际概率函数，它与θ无关，不含θ的任何信息。因此能用来对θ作出推断的仅是条件分布$\pi(\theta\,|\,x_1, x_2, \ldots, x_n)$，它的计算公式是

$$\pi(\theta|x_1,\cdots,x_n)=\frac{f(x_1,\cdots,x_n,\theta)}{f(x_1,\cdots,x_n)}=\frac{f(x_1,\cdots,x_n|\theta)\pi(\theta)}{\int_\Theta f(x_1,\cdots,x_n|\theta)\pi(\theta)d\theta}$$

$$\pi(\theta|x_1,\cdots,x_n)=\frac{f(x_1,\cdots,x_n,\theta)}{f(x_1,\cdots,x_n)}=\frac{f(x_1,\cdots,x_n|\theta)\pi(\theta)}{\int_\Theta f(x_1,\cdots,x_n|\theta)\pi(\theta)d\theta}$$

这个条件分布称为θ 的**后验分布，**它集中了总体、样本和先验中有关θ 的一切信息。

后验分布$\pi(\theta|\ x_1,\ x_2,\ \cdots,\ x_n)$的计算公式就是用密度函数表示的**贝叶斯公式**。

它是用总体和样本对先验分布$\pi(\theta)$作调整的结果，贝叶斯统计的一切推断都基于后验分布进行。

Bayes统计的三个基本假设:

假设I: 总体X有一个概率分布(pdf或pmf)$p(x;\theta)$, 其中θ是参数, 不同的θ对应着不同的分布. 在Bayes统计中, $p(x;\theta)$是给定θ后的一个条件概率分布, 记为$p(x|\theta)$. $p(x|\theta)$提供的关于θ的信息就是总体信息.

假设II: 当给定θ后,从总体$p(x|\theta)$中随机抽取一个样本$\widetilde{X} = (X_1, X_2, \cdots, X_n)$, 该样本中含有的有关$\theta$的信息,就是样本信息.

当给定θ后, 样本的联合条件概率分布(pdf或pmf)为

$$p(\widetilde{x}|\theta) = \prod_{i=1}^{n} p(x_i|\theta).$$

这个条件概率分布综合了总体信息和样本信息.

假设III: 未知参数θ是一个随机变量, 它有一个概率分布(pdf或pmf)$\pi(\theta)$(称为先验分布). 先验分布是已知的. 先验分布提供的信息就是先验信息.

这样, 样本$\widetilde{X}=(X_1, X_2, \cdots, X_n)$和参数的$\theta$的联合概率分布是

$$p(\widetilde{x}, \theta)=p(\widetilde{x}|\theta)\pi(\theta).$$

这个联合概率分布综合了总体信息、样本信息和先验信息.

给定样本$\widetilde{X}=\widetilde{x}$, 由$\widetilde{X}$和参数的$\theta$的联合概率分布求得参数$\theta$的条件概率分布

$$\begin{aligned}\pi(\theta|\widetilde{x}) &= \frac{p(\widetilde{x},\theta)}{p(\widetilde{x})} \\ &= \frac{p(\widetilde{x}|\theta)\pi(\theta)}{\int p(\widetilde{x}|\theta)\pi(\theta)d\theta} \text{ ———Bayes公式.}\end{aligned}$$

$\pi(\theta|\widetilde{x})$是有了试验结果后,得到的参数$\theta$的条件分布, 称为后验分布(posterior distribution). 而$p(\widetilde{x})$是样本的边际分布.

$$\widetilde{X}=\widetilde{x}$$
$$\vdots$$
$$\pi(\theta) \longrightarrow \pi(\theta|\widetilde{x})$$

Bayes方法对参数θ的统计推断是建立在后验分布$\pi(\theta|\widetilde{x})$的基础上的.

在Bayes公式中，分子和分母上同乘上一个与θ无关的量$h(\tilde{x})$不改变后验分布. 特别地如果$T = T(\tilde{X})$是充分统计量，

$$p(\tilde{x}|\theta) = p(\tilde{x}|T = t)p_T(t|\theta) \propto p_T(t|\theta).$$

其中$t = T(\tilde{x})$. 从而

$$\pi(\theta|\tilde{x}) = \frac{p_T(t|\theta)\pi(\theta)}{\int p_T(t|\theta)\pi(\theta)d\theta} = \pi(\theta|t).$$

用充分统计量代替样本所得的后验分布是一样的.

例1 （后验分布的计算）

$$p(x\mid\theta)=\frac{1}{\sqrt{2\pi}}e^{-(x-\theta)^2/2}$$

$$\pi(\theta_0)=\pi(6)=0.01 \qquad \pi(\theta_1)=\pi(10)=0.99$$

于是有

$$p(x)=\sum_{i=0}^{1}h(x,\theta_i)=\sum_{i=0}^{1}\pi(\theta_i)p(x\mid\theta_i)=\pi(6)p(x\mid 6)+\pi(10)p(x\mid 10)$$

$$p(8.30)=\pi(6)p(8.30\mid 6)+\pi(10)p(8.30\mid 10)=0.01\times e^{-2.30^2/2}+0.99\times e^{-1.70^2/2}$$

$$P(\theta=6\mid x=8.30)=\pi(6\mid 8.30)$$

$$=\frac{\pi(6)p(8.30\mid 6)}{p(8.30)}$$

$$= \frac{0.01 \times e^{-2.30^2/2}}{0.01 \times e^{-2.30^2/2} + 0.99 \times e^{-1.70^2/2}} = 3.033 \times 10^{-3}$$

$$P(\theta = 10 \mid x = 8.30) = \pi(10 \mid 8.30)$$

$$= \frac{\pi(10) p(8.30 \mid 10)}{p(8.30)}$$

$$= \frac{0.99 \times e^{-1.70^2/2}}{0.01 \times e^{-2.30^2/2} + 0.99 \times e^{-1.70^2/2}} = 0.997$$

例2 设X 服从参数为θ 的指数分布,$\theta \sim \Gamma(\alpha,\beta)$.

有容量为1的样本 x .则

$$p(x\,|\,\theta)=\theta e^{-\theta x}I_{(0,+\infty)}(x)$$

$$\pi(\theta)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}I_{(0,+\infty)}(\theta)$$

$$h(x,\theta)=\pi(\theta)p(x\,|\,\theta)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha}e^{-(x+\beta)\theta}I_{(0,+\infty)}(x)I_{(0,+\infty)}(\theta)$$

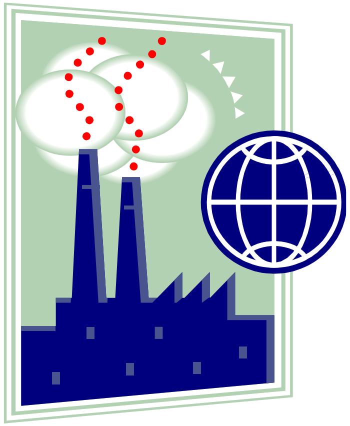
$$p(x)=\int_{-\infty}^{+\infty}h(x,\theta)d\theta=I_{(0,+\infty)}(x)\int_{0}^{+\infty}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha}e^{-(x+\beta)\theta}d\theta=\frac{\beta^{\alpha}(x+\beta)^{-(\alpha+1)}\Gamma(\alpha+1)}{\Gamma(\alpha)}I_{(0,+\infty)}(x)$$

故对 $x>0$ ，

$$\pi(\theta \mid x)=\frac{h(x,\theta)}{p(x)}=\frac{(x+\beta)^{\alpha+1}}{\Gamma(\alpha+1)}\theta^{\alpha}e^{-(x+\beta)\theta}I_{(0,+\infty)}(\theta)$$

即

$$\pi(\theta \mid x)\sim\Gamma(\alpha+1,x+\beta)$$



例3 设总体 $X\sim N(\theta,\sigma^2)$, 其中 θ未知, σ^2已知 $\theta\sim N(\mu,\tau^2)$, 其中 μ和 τ^2都已知. 又设 $x_1,\cdots,x_n$ 为样本值. 则

$$\pi(\theta)=\frac{1}{\sqrt{2\pi\tau}}\exp\left(-\frac{(\theta-\mu)^2}{2\tau^2}\right)$$

$$p(x_1,\cdots,x_n\mid\theta)=\left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\theta)^2\right)$$

由此可以算出。

$$\pi(\theta \mid x_1, \cdots, x_n) = \frac{1}{\sqrt{2\pi\tau_1}} \exp\left(-\frac{(\theta - \mu_1)^2}{2\tau_1^2} \right) \sim N(\mu_1, \tau_1^2)$$

其中

$$\mu_1 = \frac{\bar{x}\tau^2 + \mu\sigma^2 / n}{\tau^2 + \sigma^2 / n}$$

$$\frac{1}{\tau_1^2} = \frac{1}{\sigma^2 / n} + \frac{1}{\tau^2}$$

推导见书上或者后面的例子

3、贝叶斯估计

基于后验分布$\pi(\theta|x_1, x_2, ..., x_n)$对$\theta$所作的贝叶斯估计有多种，常用有如下三种：

- 使用后验分布的密度函数最大值作为θ的点估计，称为最大后验估计；
- 使用后验分布的中位数作为θ的点估计，称为后验中位数估计；
- 使用后验分布的均值作为θ的点估计，称为后验期望估计。

用得最多的是后验期望估计，它一般也简称为贝叶斯估计，记为 $\hat{\theta}_B$。

$$\widehat{\theta}_B(\widetilde{x}) = \mathsf{E}(\theta|\widetilde{x}) = \int \theta\pi(\theta|\widetilde{x})d\theta.$$

定义1 设$a=a(x)$为决策，$L(\theta,a)$为损失函数. 称

$$c(x,a)=\int_{\Theta} L(\theta,a)\pi(\theta\mid x)d\theta$$

为决策a 的后验期望损失. 称对所有满足$p(x)>0$的$\mathbf{x}$，使得$c(x,a)$有最小值的决策$a(x)$，或等价地，使得

$$\int_{\Theta} L(\theta,a)p(x\mid\theta)\pi(\theta)d\theta$$

有最小值的决策$a(x)$（因为$\pi(x)p(x\mid\theta)=p(x)\pi(\theta\mid x)$），为该统计问题的贝叶斯决策（函数）。在估计问题中，它又称为贝叶斯估计。

因为贝叶斯决策使得后验期望损失达到最小,所以从经济的角度考虑,贝叶斯决策是一个好的决策.

定理2.5.1 1) 在平方损失函数$L(\theta,a)=(a-\theta)^2$下，θ的贝叶斯估计为后验期望值

$$E(\theta \mid x)=\int_{\Theta}\theta\pi(\theta \mid x)d\theta$$

2) 在平方损失函数的加权函数 $L(\theta,a)=\lambda(\theta)(a-\theta)^2$ 下，θ 的贝叶斯估计为

$$\frac{E[\lambda(\theta)\theta \mid x]}{E[\lambda(\theta) \mid x]}$$

证　仅证1)，用类似的方法可以证明2)．由于

$$c(x,a)=\int_{\Theta}L(\theta,a)\pi(\theta\,|\,x)d\theta=\int_{\Theta}(a-\theta)^2\pi(\theta\,|\,x)d\theta$$

$$=a^2\int_{\Theta}\pi(\theta\,|\,x)d\theta-2a\int_{\Theta}\theta\pi(\theta\,|\,x)d\theta+\int_{\Theta}\theta^2\pi(\theta\,|\,x)d\theta$$

$$=a^2-2a\int_{\Theta}\theta\pi(\theta\,|\,x)d\theta+\int_{\Theta}\theta^2\pi(\theta\,|\,x)d\theta$$

是关于a的二次三项式，不难知道，当

$a=\int_{\Theta}\theta\pi(\theta\,|\,x)d\theta$ 时 $c(a,x)$有最小值

例. 设某事件A在一次试验中发生的概率为θ，为估计θ，对试验进行了n次独立观测，其中事件A发生了X次，显然 $X|\theta \sim B(n,\theta)$，即

$$f(X=x|\theta)=\binom{n}{x}\theta^{x}(1-\theta)^{n-x},\quad x=0,1,\cdots,n$$

θ的极大似然估计值为

$$\hat{\theta}_{MLE}=\frac{x}{n}$$ ————-经典统计学派的估计.

假若我们在试验前对事件A没有什么了解，从而对其发生的概率θ也没有任何信息。在这种场合，贝叶斯本人建议采用“同等无知”的原则使用区间（0,1）上的均匀分布$U(0,1)$作为θ的先验分布，因为它取（0,1）上的每一点的机会均等。

贝叶斯的这个建议被后人称为贝叶斯假设。

由此即可利用贝叶斯公式求出θ的后验分布。具体如下：先写出X和θ的联合分布

$$f(x,\theta)=\binom{n}{x}\theta^{x}(1-\theta)^{n-x},\quad x=0,1,\cdots,n,\quad 0<\theta<1$$

然后求X的边际分布

$$f(x_1,\cdots,x_n)=\int_{\Theta}f(x_1,\cdots,x_n,\theta)d\theta=\int_{\Theta}\binom{n}{x}\theta^{x}(1-\theta)^{n-x}d\theta=\binom{n}{x}\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}$$

最后求出θ的后验分布

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

最后的结果说明$\theta|X\sim Be(x+1,n-x+1)$，其后验期望估计为

$$\hat{\theta}_B=E(\theta|x)=\frac{x+1}{n+2}$$

$$f(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1},\quad 0\le x\le 1$$

$$\beta(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}=\int_0^1 t^{a-1}(1-t)^{b-1}\mathrm{d}t$$

$$E(X)=\frac{a}{a+b};\quad Var(X)=\frac{ab}{(a+b)^2(a+b+1)}$$

某些场合，贝叶斯估计要比极大似然估计更合理一点。比如："抽检3个全是合格品"与"抽检10个全是合格品"，后者的质量比前者更信得过。这种差别在不合格品率的极大似然估计中反映不出来（两者都为0），而用贝叶斯估计两者分别是0.2 和 0.083。

由此可以看到，在这些极端情况下，贝叶斯估计比极大似然估计更符合人们的理念。

另一种更简洁做法:

为求Bayes估计, 先给定一个先验分布$\pi(\theta)$, 然后求后验分布

$$\pi(\theta|x) = \frac{\binom{n}{x}\theta^x(1-\theta)^{n-x}\pi(\theta)}{\int \binom{n}{x}\theta^x(1-x)^{n-x}\pi(\theta)d\theta} \propto_\theta \theta^x(1-\theta)^{n-x}\pi(\theta).$$

然后求后验期望就得到Bayes估计

$$\widehat{\theta}_B = \int \theta\pi(\theta|x)d\theta = \frac{\int \theta\theta^x(1-\theta)^{n-x}\pi(\theta)d\theta}{\int \theta^x(1-x)^{n-x}\pi(\theta)d\theta}.$$

先验分布的选取:

1. ”同等无知”

在没有先验信息的情况下，对未知参数θ的所有可能取值同等对待. 现θ取值范围为$[0,1]$，故在没有先验信息的情况下，取均匀分布$U[0,1]$作为先验分布:

$$\pi(\theta)=1;\quad 0\le\theta\le 1.$$

这时Bayes估计为

$$\widehat{\theta}_B=\frac{\int\theta\theta^x(1-\theta)^{n-x}d\theta}{\int\theta^x(1-x)^{n-x}d\theta}=\frac{x+1}{n+2}.$$

No.	n	x	$\widehat{\theta}_{MLE} = \frac{x}{n}$	$\widehat{\theta}_{B} = \frac{x+1}{n+2}$
1	5	5	1	0.867
2	20	20	1	0.955
3	5	0	0	0.143
4	20	0	0	0.045

2. 共轭分布法 我们希望先验分布和后验分布是同类型的分布, 即它们在同一个分布族中. 这样的分布族称为样本(或统计量)的分布族的共轭分布族(conjugate family).

现在

$$\pi(\theta|x) \propto_\theta \theta^x(1-\theta)^{n-x}\pi(\theta).$$

为使得先验分布和后验分布有同类型的分布, 我们有

$$\pi(\theta) \propto_\theta \theta^{a-1}(1-\theta)^{b-1}.$$

因此取beta分布作为先验分布, beta分布族作为先验分布族.

$$\pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}; \quad 0 \le \theta \le 1.$$

这时

$$\pi(\theta|x) \propto_\theta \theta^{x+a-1}(1-\theta)^{n-x+b-1}.$$

Bayes估计为

$$\widehat{\theta}_B = \mathsf{E}[beta(x+a, n-x+b)] = \frac{x+a}{n+a+b}.$$

$\widehat{\theta}_B$可写成

$$\widehat{\theta}_B = \frac{n}{a+b+n} \cdot \widehat{\theta}_{MLE} + \frac{a+b}{a+b+n} \cdot \frac{a}{a+b}.$$

它是先验期望$a/(a+b)$和MLE$\widehat{\theta}_{MLE}$的加权平均.

n 小时，先验信息在估计中占主要地位; n大时，试验信息在估计中占主要地位.

参数a, b不同,对应的先验分布也不同，得到的Bayes估计也不同. a, b可以通过先验信息估计出来.

参数a, b不同,对应的先验分布也不同, 得到的Bayes估计也不同. a, b可以通过先验信息估计出来.

例如: 1. 如果知道θ的均值$\overline{\theta}$和方差s_θ^2, 则可由下列方程解出a, b:

$$\begin{cases} \frac{a}{a+b} = \overline{\theta}, \\ \frac{ab}{(a+b)^2(a+b+1)} = s_\theta^2. \end{cases}$$

2. 如果知道θ的分位数, 则也可解出a, b; 如

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

3. 如果根据先验信息只能获得先验均值θ. 可令

$$\frac{a}{a+b}=\overline{\theta}.$$

但从一个方程不能唯一确定两个未知参数.

方差

$$\mathsf{Var}(\theta)=\frac{ab}{(a+b)^2(a+b+1)}=\frac{\overline{\theta}(1-\overline{\theta})}{a+b+1}$$

随$a+b$的增大而减少, 方差减少意味着概率向均值$\mathsf{E}(\theta)=\overline{\theta}$集中, 从而提高了$\mathsf{E}(\theta)=\overline{\theta}$的确信程度. 这样以来, 选择$a+b$的问题转化为决策人对$\mathsf{E}(\theta)=\overline{\theta}$的确信程度大小的问题. 如果对$\mathsf{E}(\theta)=\overline{\theta}$很确信, 那么$a+b$可选得大一些, 否则就选得小一些.

Beta 分布中参数与方差的关系

Beta分布	a	$a+b$	$\mathsf{E}(\theta)$	$\mathsf{Var}(\theta)$
beta(2,3)	2	5	0.4	0.0400
beta(4,6)	4	10	0.4	0.0218
beta(8,12)	8	20	0.4	0.0114
beta(10,15)	10	25	0.4	0.0092
beta(14,21)	14	35	0.4	0.0067

例. 设$x_1, x_2, \cdots, x_n$是来自正态分布$N(\mu,\sigma_0^{\ 2})$的一个样本，其中$\sigma_0^{\ 2}$已知，μ未知，假设μ的先验分布亦为正态分布$N(\theta,\tau^2)$，其中先验均值θ和先验方差τ^2均已知，试求μ的贝叶斯估计。

解：样本X的分布和μ的先验分布分别为

$$f(x|\mu)=(2\pi\sigma_0^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

$$\pi(\mu)=(2\pi\tau^2)^{-1/2}\exp\left\{-\frac{1}{2\tau^2}(\mu-\theta)^2\right\}$$

由此可以写出x与μ的联合分布

$$f(x,\mu) = k_1 \cdot \exp\left\{-\frac{1}{2}\left[\frac{n\mu^2 - 2n\mu\bar{x} + \sum_{i=1}^{n} x_i^2}{\sigma_0^2} + \frac{\mu^2 - 2\theta\mu + \theta^2}{\tau^2}\right]\right\}$$

其中 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ ， $k_1 = (2\pi)^{-(n+1)/2}\tau^{-1}\sigma_0^{-n}$ 若记

$$A = \frac{n}{\sigma_0^2} + \frac{1}{\tau^2}, \quad B = \frac{n\bar{x}}{\sigma_0^2} + \frac{\theta}{\tau^2}, \quad C = \frac{\sum_{i=1}^{n} x_i^2}{\sigma_0^2} + \frac{\theta^2}{\tau^2}$$

则有

$$f(x,\mu) = k_1 \exp\{-\frac{1}{2}[A\mu^2 - 2B\mu + C]\}$$

$$= k_1 \exp\{-\frac{(\mu - B/A)^2}{2/A} - \frac{1}{2}(C - B^2/A)\}$$

注意到A,B,C均与μ无关，由此容易算得样本的边际密度函数

$$f(x)=\int_{-\infty}^{\infty} f(x,\mu)d\mu = k_1 \exp\left\{-\frac{1}{2}(C-B^2/A)\right\}(2\pi/A)^{1/2}$$

应用贝叶斯公式即可得到后验分布

$$\pi(\mu|x)=\frac{f(x,\mu)}{f(x)}=(2\pi/A)^{1/2}\exp\left\{-\frac{1}{2/A}(\mu-B/A)^2\right\}$$

这说明在样本给定后，μ的后验分布为

$N(B/A,1/A)$，即

$$\mu|x\sim N\left(\frac{n\bar{x}\sigma_0^{-2}+\theta\tau^{-2}}{n\sigma_0^{-2}+\tau^{-2}},\frac{1}{n\sigma_0^{-2}+\tau^{-2}}\right)$$

后验均值即为其贝叶斯估计:

$$\hat{\mu} = \frac{n/\sigma_0^2}{n/\sigma_0^2 + 1/\tau^2}\bar{x} + \frac{1/\tau^2}{n/\sigma_0^2 + 1/\tau^2}\theta$$

它是样本均值 $\bar{x}$ 与先验均值 θ 的加权平均。

Example

在儿童智商测验中，儿童的智商$X \sim N(\theta, 100)$，而$\theta \sim N(100, 225)$. 一儿童在一次智商测验中得$x = 115$分，求θ的Bayes估计.

$$\widehat{\theta}_B = \frac{x(100)^{-1} + 100(225)^{-1}}{(100)^{-1} + (225)^{-1}} = \frac{400 + 9x}{13} = 110.38.$$

共轭先验分布

设θ是某分布的一个参数，$\pi(\theta)$是其先验分布. 假如由抽样信息算得的后验分布$\pi(\theta|\tilde{x})$与$\pi(\theta)$是属于一个分布族，则称$\pi(\theta)$ 是θ 的共轭先验分布.

常用的共轭先验分布

总体分布	参数	共轭先验分布
二项分布	成功概率	beta分布
泊松分布	均值	gamma分布
指数分布	均值	倒gamma分布
指数分布	均值倒数	gamma分布
正态分布(方差已知)	均值	正态分布
正态分布(均值已知)	方差	倒gamma分布

例.指数分布

彩电的寿命服从指数分布，它的密度函数为

$$p(t|\theta) = \frac{1}{\theta} e^{-t/\theta}, \quad t > 0.$$

其中，$\theta > 0$是彩电的平均寿命. 现从一批彩电中随机抽取n台进行寿命试验. 试验到第r台失效时为止. 记录失效时间为$t_{(1)} \le t_{(2)} \le \cdots \le t_{(r)}$. 另外，$n-r$台直到试验停止时还没失效. 求$\theta$的Bayes估计.

设$(T_1, T_2, \cdots, T_n)$为抽取的n台彩电的寿命, $t_{(1)}, t_{(2)}, \cdots, t_{(r)}$是观察到的前$r$个次序统计量$T_{(1)}, T_{(2)}, \cdots, T_{(r)}$的观察值. $(T_{(1)}, T_{(2)}, \cdots, T_{(r)})$ 是截断样本, 这个截断样本的密度函数为

$$p(t_{(1)}, \cdots, t_{(r)}|\theta) \propto_\theta \prod_{i=1}^{r}(\theta^{-1}e^{-t_{(i)}/\theta}) \cdot [e^{-t_{(r)}/\theta}]^{n-r}$$

$$=\theta^{-r}\exp\{-s/\theta\}.$$

其中$s = t_{(1)} + \cdots + t_{(r)} + (n-r)t_{(r)}$为总试验时间.

若设$\pi(\theta)$为先验分布，则后验分布为

$$\pi(\theta|t_{(1)},\cdots,t_{(r)}) \propto_\theta \theta^{-r}\exp\{-s/\theta\}\pi(\theta).$$

为使得$\pi(\theta)$和$\pi(\theta|t_{(1)},\cdots,t_{(r)})$有同样的形式，应取

$$\pi(\theta) \propto_\theta \theta^{-(a+1)}\exp\{-b/\theta\}.$$

即

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

这样的密度函数是倒gamma分布的密度函数，记为$\Gamma^{-1}(a,b)$.

易得

$$\mathsf{E}[\Gamma^{-1}(a,b)] = \frac{b}{a-1}.$$

现在后验密度函数为

$$\pi(\theta|t_{(1)},\cdots,t_{(r)}) \propto_\theta \theta^{-(a+r+1)} \exp\{-(b+s)/\theta\}, \quad \theta > 0.$$

即$\pi(\theta|t_{(1)},\cdots,t_{(r)}) \sim \Gamma^{-1}(a+r, b+s)$. Bayes估计为

$$\widehat{\theta}_B = \mathsf{E}[\Gamma^{-1}(a+r,b+s)] = \frac{b+s}{a+r-1}.$$

根据已掌握的资料, 我国已做了大量的彩电寿命试验,由这些试验数据可估算出彩电的平均寿命不低于30000小时, 10%分位数$\theta_{0.1}$大约为11250小时. 这样解方程

$$\begin{cases} \frac{b}{a-1} = 30000 \\ \int_0^{\theta_{0.1}} \pi(\theta)d\theta = 0.1. \end{cases}$$

得

$$a = 1.956, \qquad b = 2868.$$

从而

$$\widehat{\theta}_B = \frac{2868 + s}{0.956 + r}.$$

加权平均均方误差与Bayes 估计

设θ为待估参数, $\widetilde{X}$为一个样本.

证明: 在均方误差意义下, θ的最优估计量不存在. 即, 不存在这样的估计量$\delta(\widetilde{X})$使得

$$\mathsf{E}_{\theta}\{\theta-\delta(\widetilde{X})\}^{2} \quad \text{关于}\theta\text{一致地最小.}$$

而将寻找范围缩小到无偏估计类时, 这样的最优估计量在一定条件下存在且唯一;

下面我们换一个角度思考最优估计的问题. 我们不需要

$$R(\theta,\delta)=\mathsf{E}_\theta\{\theta-\delta(\widetilde{X})\}^2 \text{ 关于}\theta\text{一致地最小.}$$

而考察使得均方误差$R(\theta,\delta)$关于θ的加权平均最小的问题:

$$\min_{\delta}\int R(\theta,\delta)\pi(\theta)d\theta.$$

$\pi(\theta)\geq 0$是某个给定的权函数. 我们不妨设

$$\int \pi(\theta)d\theta=1.$$

这样$\pi(\theta)$可以看做一个概率密度. 记样本的概率密度(pdf)或分布列(pmf)为$p(\widetilde{x}|\theta)$.

则

$$\int R(\theta,\delta)\pi(\theta)d\theta = \int\int \{\theta-\delta(\tilde{x})\}^2 p(\tilde{x}|\theta)\pi(\theta)d\tilde{x}d\theta.$$

这恰好是将$\theta \sim \pi(\theta)$看作随机变量时，在Bayes意义下，$(\tilde{X},\theta)$之函数$\{\theta-\delta(\tilde{X})\}^2$ 的期望，即

$$\int R(\theta,\delta)\pi(\theta)d\theta = \mathsf{E}\{\theta-\delta(\tilde{X})\}^2.$$

$\mathsf{E}\{\theta-\delta(\tilde{X})\}^2$称为Bayes风险.

如果写$p(\tilde{x}|\theta)\pi(\theta)=\pi(\theta|\tilde{x})p(\tilde{x})$，其中$\pi(\theta|\tilde{x})$ 是θ的后验分布，$p(\tilde{x})$是$\widetilde{X}$的边际分布，那么

$$\begin{aligned}\int R(\theta,\delta)\pi(\theta)d\theta &= \mathsf{E}\{\theta-\delta(\widetilde{X})\}^2 \\ &= \int\left[\int\{\theta-\delta(\tilde{x})\}^2\pi(\theta|\tilde{x})d\theta\right]p(\tilde{x})d\tilde{x} \\ &= \int \mathsf{E}\left[\{\theta-\delta(\tilde{x})\}^2\middle|\widetilde{X}=\tilde{x}\right]p(\tilde{x})d\tilde{x}.\end{aligned}$$

$\mathsf{E}\left[\{\theta-\delta(\tilde{x})\}^2\middle|\widetilde{X}=\tilde{x}\right]$称为后验期望损失(posterior expected loss).

为$\int R(\theta,\delta)\pi(\theta)d\theta$最小，只要

$$\min_{\delta(\widetilde{x})} \mathsf{E}\left[\{\theta-\delta(\widetilde{x})\}^2\middle|\widetilde{X}=\widetilde{x}\right].$$

而

$$\arg\min_{\delta(\widetilde{x})} \mathsf{E}\left[\{\theta-\delta(\widetilde{x})\}^2\middle|\widetilde{X}=\widetilde{x}\right]=\mathsf{E}[\theta|\widetilde{X}=\widetilde{x}]=\int\theta\pi(\theta|\widetilde{x})d\theta$$

即为后验期望，这就是Bayes 估计.

所以

$$\mathsf{E}[\theta|\widetilde{X}=\widetilde{x}]=\arg\min_{\delta(\widetilde{x})}\int R(\theta,\delta)\pi(\theta)d\theta.$$

即

$$\mathsf{E}[\theta|\widetilde{X}]=\arg\min_{\delta(\widetilde{X})}\int \mathsf{E}_\theta\{\theta-\delta(\widetilde{X})\}^2\pi(\theta)d\theta.$$

结论 对给定的权函数$\pi(\theta)$, 在加权均方误差(或称Bayes均方风险意义)下, 期望型Bayes估计是最优估计.

即

$$
\begin{aligned}
&\arg\min_{\delta(\tilde{x})} \int \mathsf{E}_\theta(\theta-\delta(\tilde{X})^2\pi(\theta)d\theta \\
=&\arg\min_{\delta(\tilde{x})} \mathsf{E}\left[(\theta-\delta(\tilde{X}))^2 \middle| \tilde{X}=\tilde{x}\right] \\
=&\mathsf{E}[\theta|\tilde{x}] = \text{mean of } \ \pi(\theta|\tilde{x}).
\end{aligned}
$$

例 1 某厂生产电子元件，从过去的每批产品得到次品率的概率分布如下表

次品率 p	0.02	0.05	0.10	0.15	0.20
概率 P	0.4	0.3	0.15	0.10	0.05

现有一批产品需要出厂，厂方没有作 100%的检验，只是抽样 20 件，检查结果其中有一件为次品。假定损失函数为误差的平方，即 $L(p,d)=(p-d)^2$，试求次品率的贝叶斯估计。

解：总体 X 为一件产品中的次品数，则总体 $X\sim b(1,p)$，其中 $p\in\Theta=\{0.02,0.05,0.10,0.15,0.20\}$。

样本 $(X_1,\cdots,X_n)$，样本观测值 $(x_1,\cdots,x_n)$， $n=20$

参数 p 的后验分布：

$$P(\theta=t|X_1=x_1,\cdots;X_n=x_n)=\frac{P(\theta=t)\prod_{i=1}^{n}P(X_i=x_i|\theta=t)}{\sum_{t\in\Theta}P(\theta=t)\prod_{i=1}^{n}P(X_i=x_i|\theta=t)}=\frac{P(\theta=t)\prod_{i=1}^{n}t^{x_i}(1-t)^{1-x_i}}{\sum_{t\in\Theta}P(\theta=t)\prod_{i=1}^{n}t^{x_i}(1-t)^{1-x_i}}=\frac{P(\theta=t)t^{\sum_{i=1}^{n}x_i}(1-t)^{n-\sum_{i=1}^{n}x_i}}{\sum_{t\in\Theta}P(\theta=t)t^{\sum_{i=1}^{n}x_i}(1-t)^{n-\sum_{i=1}^{n}x_i}}$$

参数 p 的贝叶斯估计：

$$p^*(x_1,\cdots,x_n)=E(p|x_1,\cdots,x_n)=\sum_{t\in\Theta}tP(p=t|X_1=x_1,\cdots,X_n=x_n)=\frac{\sum_{t\in\Theta}tP(p=t)t^{\sum_{i=1}^{n}x_i}(1-t)^{n-\sum_{i=1}^{n}x_i}}{\sum_{t\in\Theta}P(p=t)t^{\sum_{i=1}^{n}x_i}(1-t)^{n-\sum_{i=1}^{n}x_i}}$$

将 $\sum_{i=1}^{n}x_i=1$， $n=20$ 及 θ 的先验分布代入上式可得次品率的贝叶斯估计 $\hat{p}=0.052$ 。

例 2 总体 $X \sim b(1,p)$，参数 p 未知且 $p \sim U[0,1]$，样本 $(X_1,\cdots,X_n)$，损失函数是平方损失函数 $L(p,d)=(p-d)^2$，求参数 p 的贝叶斯估计和贝叶斯风险。

解：$f(x|p)=p^x(1-p)^{1-x}, x=0,1$

p 的后验分布 $$h(p|x_1,\cdots,x_n)=\frac{\pi(p)\prod_{i=1}^{n}P(X_i=x_i|p)}{\int_{\theta\in\Theta}\pi(p)\prod_{i=1}^{n}P(X_i=x_i|p)dp},\quad 0\le p\le 1$$

$$\pi(p)=\begin{cases}1, & 0\le p\le 1\\ 0, & otherwise\end{cases},\quad \prod_{i=1}^{n}P(X_i=x_i|p)=\prod_{i=1}^{n}p^{x_i}(1-p)^{1-x_i}=p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i}$$

$$\int_{\theta\in\Theta}\pi(p)\prod_{i=1}^{n}P(X_i=x_i|p)dp=\int_{-\infty}^{+\infty}\pi(p)q(x_1,\cdots,x_n|p)dp=\int_0^1 p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i}dp$$

$$=B(\sum_{i=1}^{n}x_i+1,n-\sum_{i=1}^{n}x_i+1)=\frac{\Gamma(\sum_{i=1}^{n}x_i+1)\Gamma(n-\sum_{i=1}^{n}x_i+1)}{\Gamma(n+2)}$$

p 的后验分布：

$$h(p|x_1,\cdots,x_n)=\frac{p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i}}{\dfrac{\Gamma(\sum_{i=1}^{n}x_i+1)\Gamma(n-\sum_{i=1}^{n}x_i+1)}{\Gamma(n+2)}}=\frac{\Gamma(n+2)}{\Gamma(\sum_{i=1}^{n}x_i+1)\Gamma(n-\sum_{i=1}^{n}x_i+1)}p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i},\quad 0\le p\le 1$$

即 $p|(x_1,\cdots,x_n)\sim Be(\sum_{i=1}^{n}x_i+1,n+1-\sum_{i=1}^{n}x_i)$

p 的贝叶斯估计： $p^*(x_1,\cdots,x_n)=E(p|x_1,\cdots,x_n)=\dfrac{\sum_{i=1}^{n}x_i+1}{n+2}=\dfrac{n\bar{x}+1}{n+2}$。

贝叶斯风险 $R_B(p^*)=E\left[R(p,p^*)\right]=E\left[E\left(L(p,p^*)|p\right)\right]$

$$E\left(L(p,p^*)|p\right)=E\left((p-p^*)^2|p\right)=E\left((p-\frac{n\bar{X}+1}{n+2})^2\Big|p\right)=E\left(p^2-2p\left(\frac{n\bar{X}+1}{n+2}\right)+\left(\frac{n\bar{X}+1}{n+2}\right)^2\Big|p\right)$$

$$=p^2-2pE\left(\frac{n\bar{X}+1}{n+2}\right)+E\left(\frac{n\bar{X}+1}{n+2}\right)^2=p^2-2p\left(\frac{nE\bar{X}+1}{n+2}\right)+\frac{\left(n^2E\bar{X}^2+2nE\bar{X}+1\right)}{(n+2)^2}$$

$$=p^2-2p\left(\frac{np+1}{n+2}\right)+\frac{\left(n^2(\frac{p(1-p)}{n}+p^2)+2np+1\right)}{(n+2)^2}=\frac{-(n-4)p^2+(n-4)p+1}{(n+2)^2}$$

$$R_B(p^*)=E\left[E\left(L(p,p^*)|p\right)\right]=E\left[\frac{-(n-4)p^2+(n-4)p+1}{(n+2)^2}\right]=\frac{-(n-4)Ep^2+(n-4)Ep+1}{(n+2)^2}$$

$$=\frac{-(n-4)\left[\frac{1}{12}+\left(\frac{1}{2}\right)^2\right]+\frac{1}{2}(n-4)+1}{(n+2)^2}=\frac{\frac{1}{6}(n-4)+1}{(n+2)^2}=\frac{1}{6(n+2)}$$

注：（1）经典估计 $\hat{p}=\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ 。在 $\sum_{i=1}^{n}X_i=0$，即样本均是废品时，估计正品率 $\hat{p}=0$；

在 $\sum_{i=1}^{n}X_i=n$，即样本中无废品时，估计正品率 $\hat{p}=1$。这样的估计似乎太极端了，有点不切实际，但当样本容量 n 较小时，这样的情形难以避免。而贝叶斯估计 $\hat{p}=\frac{1}{n+2}$ 或 $\hat{p}=\frac{n+1}{n+2}$，这样的估计留有余地，给人以比较可靠之感。

（2）经典估计 $\hat{p}=\bar{X}$，其贝叶斯风险

$$R_B(\hat{p})=E[R(p,\hat{p})]=E\left[E\left(L(p,\hat{p})\middle|p\right)\right]=E\left[E\left((p-\hat{p})^2\middle|p\right)\right]=E\left[E\left((p-\bar{X})^2\middle|p\right)\right]$$

$$=E\left[E\left((p^2-2p\bar{X}+\bar{X}^2)\middle|p\right)\right]=E\left[p^2-2pE\bar{X}+E\bar{X}^2\right]=E\left[p^2-2p\cdot p+\left(\frac{p(1-p)}{n}+p^2\right)\right]$$

$$=E\left[\frac{p(1-p)}{n}\right]=\frac{E\left[p(1-p)\right]}{n}=\frac{Ep-Ep^2}{n}=\frac{\frac{1}{2}-\left(\frac{1}{12}+(\frac{1}{2})^2\right)}{n}=\frac{1}{6n}$$

例 3 假设总体 $X\sim N(\mu,\sigma^2)$，σ^2 已知，$\mu\sim N(\tau,v^2)$，样本 $(X_1,\cdots,X_n)$，损失函数 $L(\mu,d)=(\mu-d)^2$，求参数 μ 的贝叶斯估计和贝叶斯风险。

解：$f(x|\mu)=\dfrac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<+\infty$

μ 的后验分布 $h(\mu|x_1,\cdots,x_n)=\dfrac{\pi(\mu)\prod_{i=1}^{n}p(x_i|\mu)}{\int_{-\infty}^{+\infty}\pi(\mu)\prod_{i=1}^{n}p(x_i|\mu)d\mu}$，$-\infty<\mu<+\infty$

$\pi(\mu)=\dfrac{1}{\sqrt{2\pi v^2}}e^{-\frac{(\mu-\tau)^2}{2v^2}},-\infty<\mu<+\infty$，$\prod_{i=1}^{n}p(x_i|\mu)=\prod_{i=1}^{n}\dfrac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}=(2\pi\sigma^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$

$$\int_{-\infty}^{+\infty}\pi(\mu)\prod_{i=1}^{n}p(x_i|\mu)d\mu=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi v^2}}e^{-\frac{(\mu-\tau)^2}{2v^2}}(2\pi\sigma^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}d\mu$$

$$=\frac{1}{\sqrt{2\pi v^2}}(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{\tau^2}{2v^2}-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2+\frac{\left(\tau\sigma^2+n\bar{x}v^2\right)^2}{2v^2\sigma^2(nv^2+\sigma^2)}\right\}$$

$$\times\int_{-\infty}^{+\infty}\exp\left\{-\frac{(nv^2+\sigma^2)}{2v^2\sigma^2}\left(\mu-\frac{\tau\sigma^2+n\bar{x}v^2}{nv^2+\sigma^2}\right)^2\right\}d\mu$$

$$=(2\pi\sigma^2)^{-\frac{n}{2}}\left(\frac{nv^2+\sigma^2}{\sigma^2}\right)^{-\frac{1}{2}}\exp\left\{-\frac{\tau^2}{2v^2}-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2+\frac{\left(\tau\sigma^2+n\bar{x}v^2\right)^2}{2v^2\sigma^2(nv^2+\sigma^2)}\right\}$$

μ 的后验分布：

$$h(\mu|x_1,\cdots,x_n)=\frac{1}{\sqrt{\dfrac{2\pi v^2\sigma^2}{nv^2+\sigma^2}}}\exp\left\{-\frac{\left(\mu-\dfrac{\tau\sigma^2+n\bar{x}v^2}{nv^2+\sigma^2}\right)^2}{\dfrac{2v^2\sigma^2}{nv^2+\sigma^2}}\right\},\quad \mu|(x_1,\cdots,x_n)\sim N\left(\frac{\tau\sigma^2+nv^2\bar{x}}{nv^2+\sigma^2},\frac{v^2\sigma^2}{nv^2+\sigma^2}\right)$$

μ 的贝叶斯估计： $\mu^*(x_1,\cdots,x_n)=E(\mu|x_1,\cdots,x_n)=\dfrac{\tau\sigma^2+nv^2\bar{x}}{nv^2+\sigma^2}$

贝叶斯风险 $R_B(\mu^*)=E\left[R(\mu,\mu^*)\right]=E\left[E\left(L(\mu,\mu^*)\middle|\mu\right)\right]$

$$E\left(L(\mu,\mu^*)\middle|\mu\right)=E\left((\mu-\mu^*)^2\middle|\mu\right)=E\left(\left(\mu-\frac{\tau\sigma^2+nv^2\bar{X}}{nv^2+\sigma^2}\right)^2\middle|\mu\right)=E\left(\mu^2-2\mu\left(\frac{\tau\sigma^2+nv^2\bar{X}}{nv^2+\sigma^2}\right)+\left(\frac{\tau\sigma^2+nv^2\bar{X}}{nv^2+\sigma^2}\right)^2\middle|\mu\right)$$

$$=\mu^2-2\mu E\left(\frac{\tau\sigma^2+nv^2\bar{X}}{nv^2+\sigma^2}\right)+E\left(\frac{\tau\sigma^2+nv^2\bar{X}}{nv^2+\sigma^2}\right)^2=\mu^2-2\mu\left(\frac{\tau\sigma^2+nv^2E\bar{X}}{nv^2+\sigma^2}\right)+\frac{E(\tau\sigma^2+nv^2\bar{X})^2}{(nv^2+\sigma^2)^2}$$

$$=\mu^2-2\mu\left(\frac{\tau\sigma^2+nv^2\mu}{nv^2+\sigma^2}\right)+\frac{(\tau\sigma^2)^2+2(\tau\sigma^2)(nv^2)\mu+(nv^2)^2\left(\dfrac{\sigma^2}{n}+\mu^2\right)}{(nv^2+\sigma^2)^2}$$

$$=\left[\frac{\sigma^2}{nv^2+\sigma^2}\right]^2(\mu-\tau)^2+\frac{n(v^2)^2\sigma^2}{(nv^2+\sigma^2)^2}$$

$$R_B(\mu^*) = E\left[E\left(L(\mu,\mu^*)\middle|\mu\right)\right] = E\left[\left[\frac{\sigma^2}{nv^2+\sigma^2}\right]^2(\mu-\tau)^2 + \frac{n(v^2)^2\sigma^2}{(nv^2+\sigma^2)^2}\right]$$

$$= \left[\frac{\sigma^2}{nv^2+\sigma^2}\right]^2 D\mu + \frac{n(v^2)^2\sigma^2}{(nv^2+\sigma^2)^2} = \left[\frac{\sigma^2}{nv^2+\sigma^2}\right]^2 v^2 + \frac{n(v^2)^2\sigma^2}{(nv^2+\sigma^2)^2}$$

$$= \frac{\sigma^2 v^2}{nv^2+\sigma^2}$$

注：若 $X \sim N(\mu,1)$， $\mu \sim N(0,1)$，则 μ 的贝叶斯估计 $\mu^*(x_1,\cdots,x_n) = \dfrac{n\overline{x}}{n+1} = \dfrac{1}{n+1}\sum_{i=1}^{n} x_i$，

贝叶斯风险 $R_B(\mu^*) = \dfrac{1}{n+1}$。

例 4 彩电寿命服从指数分布，其密度函数为

$$p(t|\theta)=\frac{1}{\theta}e^{-t/\theta},\quad t>0$$

其中 $\theta>0$ 是彩电的平均寿命。现从一批彩电中随机抽取 n 台进行寿命试验。试验到第 r 台失效为止，其失效时间为 $t_1\le t_2\le\cdots\le t_r$，另外 $n-r$ 台彩电直到试验停止时（t_r）还未失效。这种试验称为截尾寿命试验，所得样本 $(t_1,\cdots,t_r)$ 为截尾样本。试求彩电平均寿命 θ 的贝叶斯估计。

解：截尾样本的联合分布为

$$p(t_1,\cdots,t_r|\theta)=\frac{n!}{(n-r)!}\prod_{i=1}^{r}p(t_i|\theta)\left[1-F(t_r|\theta)\right]^{n-r}=\frac{n!}{(n-r)!}\prod_{i=1}^{r}\left(\frac{1}{\theta}e^{-t_i/\theta}\right)\left[e^{-t_r/\theta}\right]^{n-r}$$

$$=\frac{n!}{(n-r)!}\frac{1}{\theta^r}e^{-\frac{1}{\theta}\left(\sum_{i=1}^{r}t_i+(n-r)t_r\right)}$$

选用倒Γ分布 $IGa(\alpha,\lambda)$ 作为 θ 的先验分布，其密度函数为

$$\pi(\theta)=\frac{\lambda^\alpha}{\Gamma(\alpha)}\theta^{-(\alpha+1)}e^{-\lambda/\theta},\quad \theta>0$$

其中 $\alpha>0,\lambda>0$ 是两个待定参数，其数学期望 $E\theta=\dfrac{\lambda}{\alpha-1}$。

θ 的后验分布

$$h(\theta|t_1,\cdots,t_r)=\frac{\pi(\theta)p(t_1,\cdots,t_r|\theta)}{\int_0^{+\infty}\pi(\theta)p(t_1,\cdots,t_r|\theta)d\theta}=\frac{\theta^{-(\alpha+1)}e^{-\lambda/\theta}\frac{1}{\theta^r}e^{-\frac{1}{\theta}\left(\sum_{i=1}^r t_i+(n-r)t_r\right)}}{\int_0^{+\infty}\theta^{-(\alpha+1)}e^{-\lambda/\theta}\frac{1}{\theta^r}e^{-\frac{1}{\theta}\left(\sum_{i=1}^r t_i+(n-r)t_r\right)}d\theta}=\frac{\theta^{-(\alpha+r+1)}e^{-\frac{1}{\theta}\left(\lambda+\sum_{i=1}^r t_i+(n-r)t_r\right)}}{\int_0^{+\infty}\theta^{-(\alpha+r+1)}e^{-\frac{1}{\theta}\left(\lambda+\sum_{i=1}^r t_i+(n-r)t_r\right)}d\theta}$$

$$=\frac{\left(\lambda+\sum_{i=1}^r t_i+(n-r)t_r\right)^{\alpha+r}}{\Gamma(\alpha+r)}\theta^{-(\alpha+r+1)}e^{-\frac{1}{\theta}\left(\lambda+\sum_{i=1}^r t_i+(n-r)t_r\right)},\qquad \theta>0$$

即 $\theta|t_1,\cdots,t_r \sim IGa(\alpha+r,\lambda+\sum_{i=1}^r t_i+(n-r)t_r)$，$\theta$ 的贝叶斯估计为

$$\theta^*(t_1,\cdots,t_r)=E(\theta|t_1,\cdots,t_r)=\frac{\lambda+\sum_{i=1}^r t_i+(n-r)t_r}{\alpha+r-1}$$

注：若随机变量 $X\sim Ga(\alpha,\lambda)$，则 $\frac{1}{X}$ 的分布称为倒伽玛分布，记为 $IGa(\alpha,\lambda)$。

贝叶斯估计的误差

定义 5 参数 θ 的后验分布为 $h(\theta|x_1,\cdots,x_n)$， θ 的贝叶斯估计为 $\hat{\theta}$， 则

$$MSE(\hat{\theta}|x_1,\cdots,x_n)=E_{\theta|(x_1,\cdots,x_n)}(\hat{\theta}-\theta)^2$$

称为 $\hat{\theta}$ 的后验均方误差， $\sqrt{MSE(\hat{\theta}|x_1,\cdots,x_n)}$ 称为 $\hat{\theta}$ 的后验标准误差。

估计量 $\hat{\theta}$ 的后验均方误差越小，贝叶斯估计的误差越小。当 $\hat{\theta}=E\left(\theta|x_1,\cdots,x_n\right)=\hat{\theta}_E$ 时，

$$MSE(\hat{\theta}_E|x_1,\cdots,x_n)=E_{\theta|(x_1,\cdots,x_n)}(\hat{\theta}_E-\theta)^2=D\left[\theta|(x_1,\cdots,x_n)\right]$$

称为后验方差，称 $\sqrt{D\left[\theta|(x_1,\cdots,x_n)\right]}$ 为后验标准差。

$$\begin{aligned}MSE(\hat{\theta}|x_1,\cdots,x_n)&=E_{\theta|(x_1,\cdots,x_n)}(\hat{\theta}-\theta)^2=E_{\theta|(x_1,\cdots,x_n)}\left[(\hat{\theta}-\hat{\theta}_E)+(\hat{\theta}_E-\theta)\right]^2\\&=E_{\theta|(x_1,\cdots,x_n)}(\hat{\theta}-\hat{\theta}_E)^2+E_{\theta|(x_1,\cdots,x_n)}(\hat{\theta}_E-\theta)^2\\&=(\hat{\theta}-\hat{\theta}_E)^2+D(\theta|x_1,\cdots,x_n)\end{aligned}$$

当 $\hat{\theta}_E=E\left(\theta|x_1,\cdots,x_n\right)$ 时，后验均方误差最小，所以实际中常常取后验均值作为 θ 的贝叶斯估计值。

贝叶斯区间估计

定义 6 参数 θ 的后验分布为 $h(\theta|x_1,\cdots,x_n)$，对给定的样本 $(X_1,\cdots,X_n)$ 和概率 $1-\alpha(0<\alpha<1)$，

若存在两个统计量 $\hat{\theta}_L=\hat{\theta}_L(X_1,\cdots,X_n)$ 和 $\hat{\theta}_U=\hat{\theta}_U(X_1,\cdots,X_n)$，使得

$$P(\hat{\theta}_L\le\theta\le\hat{\theta}_U\big|x_1,\cdots,x_n)\ge 1-\alpha$$

称区间 $\left[\hat{\theta}_L,\hat{\theta}_U\right]$ 为参数 θ 的置信度 $1-\alpha$ 的贝叶斯估计区间，简称 θ 的 $1-\alpha$ 置信区间。

$$P(\theta\ge\hat{\theta}_L\big|x_1,\cdots,x_n)\ge 1-\alpha\text{，}\quad P(\theta\le\hat{\theta}_U\big|x_1,\cdots,x_n)\ge 1-\alpha$$

称 $\hat{\theta}_L,\hat{\theta}_U$ 为 θ 的 $1-\alpha$ （单侧）置信下限和（单侧）置信上限。

寻求贝叶斯置信区间只需要 θ 的后验分布

例 5 设$(X_1,\cdots,X_n)$是来自正态总体$N(\mu,\sigma^2)$的一个样本，σ^2已知，μ的先验分布$N(\tau,v^2)$，求μ的$1-\alpha$贝叶斯置信区间。

解：μ的后验分布：

$$\mu|(x_1,\cdots,x_n)\sim N(\mu_1,\sigma_1^2)\text{，}\ \mu_1=\frac{\tau\sigma^2+nv^2\overline{x}}{nv^2+\sigma^2},\sigma_1^2=\frac{v^2\sigma^2}{nv^2+\sigma^2}\text{，}\ \frac{\mu|(x_1,\cdots,x_n)-\mu_1}{\sigma_1}\sim N(0,1)$$

$$P\left(\left|\frac{\mu|(x_1,\cdots,x_n)-\mu_1}{\sigma_1}\right|\geq u_{1-\frac{\alpha}{2}}\right)=1-\alpha\text{，即 }P\left(\mu_1-\sigma_1u_{1-\frac{\alpha}{2}}\leq\mu|(x_1,\cdots,x_n)\leq\mu_1+\sigma_1u_{1-\frac{\alpha}{2}}\right)=1-\alpha$$

μ的$1-\alpha$贝叶斯置信区间为：$\left[\mu_1-\sigma_1u_{1-\frac{\alpha}{2}},\mu_1+\sigma_1u_{1-\frac{\alpha}{2}}\right]$，区间长度为$2\sigma_1u_{1-\frac{\alpha}{2}}$。

注：经典估计中μ的$1-\alpha$置信区间$\left[\mu-\frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}},\mu+\frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\right]$，区间长度为$\frac{2\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}$；

$$\frac{2\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}-2\sigma_1u_{1-\frac{\alpha}{2}}=2u_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}-\sigma_1\right)=\frac{2\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\left(1-\sqrt{\frac{v^2}{v^2+\sigma^2/n}}\right)>0$$

，即贝叶斯区间估计的精度高于经典区间估计。

第8次作业:

- 孙p.62 习题二
- 50、51.

谢 谢！