第 3 章损失分布的统计推断方法

【内容要点】

1. 经验分布函数

（1） 经验分布函数的定义

以频率代替概率，Fn（x）实际上就是 n 个观测值落在区间（-∞，x］的频率，注意分布函数在 x（k）≤x<x（k+1）这一左闭右开区间为一恒值。

变量观测值不多时，此时即为离散型变量的分布函数； 变量观测值较多时，将其人为分组，此时即为连续型变量的分布函数，同时为了消除组距不同对频率的影响，可以计算频率密度：频率密度=频率组距。

（2） 频率直方图和曲线图

频率密度直方图以损失随机变量的观测值分组为横坐标，以各组的频率密度为纵坐标，将频率密度直方图的各个矩形上底中点用线段连接起来，就是频率密度折线图。把折线图修匀（光滑化），就成为频率密度曲线图。

（3） 损失次数的经验分布

即以损失次数为离散随机变量的分布函数。

（4） 损失额的经验分布

即为连续型变量的经验分布，如果直接以损失额为因变量的修匀分布函数不够光滑，可试用损失额的其他函数作为因变量，如损失额对数形式等。

2. 损失分布参数的估计、假设检验和损失分布函数的拟合检验

完整的拟合过程为：

（1） 损失分布的选择

根据损失观测数据的频率密度的直方图或曲线图形态（如连续还是离散，单峰还是多峰，正态还是偏态，左偏还是右偏等），从与其形似的理论分布族中，选择适当的分布律或概率密度函数曲线。

特别地，根据损失次数经验分布离散、右偏的特点，一般常选择的理论分布有：泊松（Poisson）分布、二项分布、负二项（Pascal）分布等，其中泊松分布是单参数的，二项分布和负二项分布都是双参数的分布。

根据损失额经验分布连续、非负、右偏和长尾的特点，一般常选择对数正态分布、帕雷托（pareto）分布、伽玛（Gamma）分布和韦伯

（Weibull）分布等作为理论分布，它们都是双参数的分布。有的还用三参数的广义帕雷托分布和布尔（Burr）分布作为损失额的理论分布。

（2） 利用观测数据对理论分布的参数进行估计（点估计含矩法估计、极大似然估计、分位数估计、最小二乘估计、区间估计）和假设检验。

① 点估计

a. 矩法估计

即以样本原点矩作为随机变量 X 的同类同阶原点矩，一般的，待估参数有几个就建立几个矩

等方程。样本原点矩是根据样本数据计算的，随机变量的原点矩是根据其分布函数计算的，是随机变量的数字特征，其中含有未知参数。

b. 极大似然估计

是使得联合概率函数（称为似然函数）的 L（x1，x2，…，xn； θ）= ∏ni=1f（xi； θ）达到最大的参数 θ。这一点很好理解，因为似然函数是概率的乘积，显然使得该乘积最大的 θ 是较好的估计。

至于其求解就是最大化 L 函数，一般求使其一阶倒数为 0 即可，有时为了简化计算，极大化 L 的对数 lnL。

值得注意的是均匀分布的唯一参数 θ（均匀分布的区间长度）的极大似然估计为样本的最大值。

c. 分位数估计

就是应用样本或者总体的分布函数，找到使得 P（X≤xi）＝分位百分比的 xi，一般的有几个未知参数就建立几个等式。

要切记的是，在样本中查找分位数对应的分位点注意是要使得 P（X≤xi）＝分位百分比，而非 xi 的下一个或上一个样本。另外，首先要将样本按从小到大排序。

d. 最小二乘估计

随机变量 X 的先验分布函数为 F（x； θ），根据观测值 x1，x2，…，xn 建立的经验分布函数为 Fn（x），所谓未知参数 θ 的最小二乘估计，就是使得先验分布函数与经验分布函数的距离 D（θ）=∑ni=1［F（xi； θ）-Fn（xi）］2 达到最小的 θ。

在实际应用时为了提高最小二乘估计的效果，常将 D（θ）加权调整为 D（θ）=∑ni=1W（xi）F（xi； θ）-nn+1Fn（xi）2，其中权 W（xi）可以有不同的取法，一般可取

W（xi）=1F（xi； θ）［1-F（xi； θ）］

根据概率统计理论，有三种性质常用来评价估计的好坏： 无偏性、有效性和一致性。具体视实际问题的要求在这三个标准中权衡。

② 区间估计

是使得参数 θ 处于某一置信区间的概率等于置信度的估计，当然一般要求事先知道样本或者总体的分布。

③ 假设检验

先对随机变量的参数作出假设，然后选择检验统计量，确定检验规则和拒绝域，再根据观测到的数据计算检验统计量的值，从而作出拒绝还是接受原假设的判断。

（3） 对理论分布作拟合优度检验（注意： 与上面的参数假设检验不同，参数假设检验是先已知样本或者总体分布，假设参数的某些数字特征，然后来检验假设的正确性与否； 而这里的拟合优度检验是对上述前两项工作即损失分布的选择和参数估计或检验的评估）。

一般选用 χ2 检验。 χ2 拟合优度检验的原假设是 H0： 损失数据服从某个理论分布。为了检验 H0，先把观测到的数据分成 n 组。选用检验统计量 χ2=∑ni=1（Qi-Ei）2Ei，式中的 Qi 是第 i 组中观测到的实际频数，Ei 是根据假设的理论分布计算出来的理论频数。在假设 H0 成立的前提下，统计量 χ2 服从自由度为 n-k-1 的 χ2 分布，其中 k 为理论分布中用估计方法得到的参数的个数。

值得注意的是，在分组时，一定要包含样本的所有可能取值范围，切不可因为在某一区间没有样本，就丢掉这一区间，因为随机变量的分布在整个区间是完整的。比如，一索赔额分布如下，索赔额在 0~1000 有 200 次，1000~2000 有 300 次，2000 以上 0 次，一定要将索赔额分为三段，0~1000，1000~2000 和 2000 以上。

3.部分保险数据损失分布的条件分布问题

主要研究免赔和再保险业务风险的有关问题（详见下面的重难点解析）。

4. 多风险损失的分布函数的拟合

已知两种风险（如火灾和盗窃）损失数据，如何据此求解总损失的分布函数。

第一种方法是把两类损失分布函数的加权平均作为总损失的分布函数。若 F1（x）和 F2（x）分别为两类损失的分布函数，p 是在已知发生损失的条件下发生的是盗窃损失的概率，是根据观测数据来估计的。那么总损失的分布可以确定为

F（x）=pF1（x）+（1-p）F2（x）

第二种方法是将两类损失的观测值合并在一起，相当于一种风险进行拟合。

5. 贝叶斯估计

步骤：

（1） 选择先验分布，注意是关于待估参数 θ 的分布，而不是关于 x 的分布；

（2） 确定条件密度函数，是在假设 θ 已知的情况下，观测值 x 的密度，并求该密度的乘积即联合密度函数；

（3） 求出 θ 的后验分布，实际就是一个条件密度函数，π（θ|x）= f（x|θ）π（θ）∫f（x|θ）π（θ）dθ。

由于其分母是关于 θ 的定积分，是关于 θ 的常数，与 θ 的分布无关，因此，可以不求解其结果。即 π（θ|x）∝f（x|θ）π（θ）（∝表示成比例的意思）

常见的先验分布和后验分布

损失随机变量的分布观测值为 x1，x2，…，xn

未知参数

参数的先验分布

参数的后验分布

泊松分布 P（λ）

伽玛分布 Γ（α，λ）

λ>0

Γ（β，γ）

Γ∑xi+β，n+γ

Γnα+β，∑xi+γ

泊松分布 P（λ）

伽玛分布Γ（α，λ）
λ>0
U（0，+∞）
Γ∑xi+1，n

Γn+α，∑xi
二项分布 B（m，p）

负二项分布 NB（k，p）
0<p<1
Beta（α，β）
Beta∑xi+α，nm-∑xi+β

Betank+α，∑xi+β

正态分布 N（μ，σ2）

-∞<μ<+∞
N（θ，τ2）

N∑xiσ2+θτ2nσ2+1τ2，1nσ2+1τ2

正态分布 N（μ，σ2）
-∞<μ<+∞

U（-∞，+∞）
N∑xin，σ2n

（4） 选择损失函数 Loss（θ^，θ）
所谓损失函数是指真实θ与估计值θ^的误差的评价函数，常用的有三种损失函数： 平方损失函数（θ^-θ）2、绝对值损失函数|θ^-θ|、0　1 损失函数 L=1，|θ^-θ|>ε

0，|θ^-θ|≤ε （ε>0，ε极小）。
（5） 不同损失函数下的θ的估计
① 平方损失函数下的贝叶斯估计是后验分布的均值

E（Loss［θ^，θ）］=∫Loss（θ^，θ）π（θ|x）dθ＝∫（θ^-θ）2π（θ|x）dθ

ddθ^E［Loss（θ^，θ）］＝-2∫（θ^-θ）π（θ|x）dθ＝0

所以 θ^∫ π （θ|x）dθ=∫ θ π （θ|x）dθ
而∫ π （θ|x）dθ=1，
所以 θ^=E（θ|x）=∫ θ π （θ|x）dθ
② 绝对误差函数下的贝叶斯估计是后验分布的中位数

E［Loss（θ^，θ）］=∫Loss（θ^，θ） π
（θ|x）dθ=∫ θ^-θ π （θ|x）dθ

=∫ θ^-∞（θ^-θ） π （θ|x）dθ+∫+∞ θ^（θ-θ^） π （θ|x）dθ

ddθ^E（Loss（θ^，θ））=θ^π （θ^|x）-∫ θ^-∞
π （θ|x）dθ-θ^π （θ^|x）-θ^π （θ^|x）

+∫+∞ θ^π （θ|x）dθ+θ^π （θ^|x）

=0

所以∫ θ^-∞ π （θ|x）dθ=∫+∞ θ^π （θ|x）dθ

所以 θ^为 π （θ|x）的中位数。
③ 绝对误差函数下的贝叶斯估计是后验分布的众数

E［Loss（θ^，θ）］=∫Loss（θ^，θ） π （θ|x）dθ=1-∫ θ^+ε θ^-ε π （θ|x）dθ

ε 极小，故只有 π （θ|x）最大时上式最小，
所以 θ^为 π （θ|x）的众数。
6. 信度理论
（1） 信度理论概述
非寿险保费的估算可以根据两类数据：一类是通过观察得到的本险种一组保单的近期数据；另一类是同险种保单早期损失数据或者类似险种保单的同期损失数据。根据前一类数据确定的保费称为经验保费
PMo，根据后一类数据得到的保费称为先验信息保费 PMe。所谓信度理论，是如何合理将这两者加权（(1-z）PMo+z PMe）以作为后验保费的估计，其中 z 称为信度或者信度因子。
（2） 有限扰动信度
① 原理和思想：如果 C 的估计量 C^是由先验信息数据 M 和近期观察值 T 加权获得，即 C^=（1-z）M+zT=（1-z）M+zET+z（T-ET），z（T-ET）就是 C^和真值 C 距离的扰动，有限扰动信度是要该扰动的相对值不超过一定限度 γ 的概率足够达到 1-α，有

Pzx--μ μ≤ɣ
≥1-α

② 完全可信性
a. 完全可信性的条件
完全可信性即指完全信度经验，z＝1，因此，
Px--μ μ≤ɣ
≥1-α，其逆定理同样成立，即完全可信性成立的条件是
Px--μ μ≤ɣ
≥1-α
b. 意义和应用： 对样本量的要求
运用中心极限定理，假设 X 的方差为 Var（X）=σ2，则 n（x--μ）σ 的渐近分布为标准正态分布 N（0，1）。根据上式，我们有

Pn（x--μ）σ≤
nɣ μ σ ≥1-α

因而

nɣ μ σ ≥U1-α2

式中的 U1-α2 是标准正态分布 N（0，1）的 1-α2 分位点。于是，完全可信性条件可以简化为

n≥U1-α2ɣ2
σ μ2=U1-α2ɣ2Var（X）[E（X）]2
Δn0

③ 平方根法则
在上面所说的同样假设（运用中心极限定理，假设 X 的方差为 Var（X）=σ2，则 n（x--μ）σ 的渐近分布为标准正态分布 N（0，1））下

z=nγμσU1-α2=nλ0σμ

所以 z=nn0，这就是所谓的平方根法则。
④ 值得注意的一点是，z
的最大值，n≥n0 时，不可根据平方根法则 z=nn0

z=min nn0，1=min μ σ n λ 0，1=min E（X）Var（X）n λ 0，1

（3） Buhlmann 信度
① 背景知识
由于贝叶斯估计方法需要事先知道参数 θ 的先验分布 π（θ）条件密度 f（x| θ），而且计算也较困难。因此，在非寿险精算中，贝叶斯信度并不实用。于是，瑞士精算学家 Buhlmann 提出了用观测值的线性函数作为损失随机变量 X 或者其参数的可信性估计，并由此来推算保费的信度模型。
② 模型结构参数及其无偏估计
该模型同样假设损失随机变量 X 的分布参数 θ 也是随机变量，非寿险精算中常把 θ 的不同取值理解为异质风险的不同风险水平，因而也叫风险参数。记 μ（θ）=E（X| θ），ν（θ）=Var（X| θ），并称 μ（θ）为假设均值，称 ν（θ）为过程方差（process variance），它们都是随机变量 θ 的函数。记

μ=E（X）=E［E（X| θ）］=E［μ（θ）］

（先对 θ i 下的 x 求得期望值 μ（θ i），再对不同 θ i 下的 μ（θ i）求得均值，所以 μ 称为 X 的总均值。因此，可以引入估计量

μ^（θ i）=x-i=∑nj=1
xijn，μ^=∑ri=1x-
ir=∑ri=1∑nj=1xijrn

容易证得这些估计都是无偏估计。

ν^=E［Var（X| θ）］=E［ν（θ）］

（先对θi 下的 x 求得方差 v^（θi），是在同一风险水平下的内在差异，故称为同质方差，再对不同θi 下的 v^（θi）求得均值，因此，可以引入估计量

v^（θi）=∑nj=1（xij-x-i）2n-1， v^=∑ri=1
v^（θi）r=∑ri=1∑nj=1（xij-x-i）2r（n-1）

容易证得这些估计都是无偏估计。

a=Var［E（X|θ）］=Var［v（θ）］

（先对θi 下的 x 求得期望值μ（θi），再对这些期望值求得方差，描述不同风险水平下的 x 的均值之间的差异，故称为异质方差。据此考虑引入估计量∑ri=1（x-i-x-）2r-1，由于 E∑ri=1（x-i-x-）2r-1
=a+ vn，即∑ri=1（x-i-x-）2r-1 不是 a 的无偏估计，所以修改后的
a^=∑ri=1（x-i-x-）2r-1- v^n 为 a 的无偏估计。）
③ 两个定理
Var（X）=E［Var（X|θ）］+Var［E（X|θ）］= v+a（可以用来简便计算）
若两个损失随机变量 X1 和 X2 的风险参数相等，则 Cov（X1，X2）=a。
④ Buhlmann 方法
基本思想是： 把损失随机变量 X 的观测值 x1，x2，…，xn 的线性函数β+β1x1+β2x2+…+βnxn 作为将来损失的可信性估计，并使均方损失达到最小。正由于这个原因，Buhlmann 信度也叫最小平方信度。
这里的βi 是观测值 xi 在线性函数中的权重，i=1，2，…，n。由于观测值 x1，x2，…，xn 的风险特征相同，所以可令β1=β2=…=βn，这时β+β1x1+β2x2+…+βnxn=β+nβ1x-，若记 z=nβ1，则将来损失的可信性估计为β+zx-。
式中的β和 z 可以通过求均方损失函数

L（β，z）=E{［β+zx--μ（θ）］2}

的极小值点得到。经过简单的计算可知，当

β=（1-z）μ

z=nn+ va 时，L（β，z）达到极小。

用这样计算得到β+zx-=（1-z）μ+zx-作为将来损失的估计而厘定的保费，称为 Buhlmann 信

度保费（buhlmann credibility premium），简称信度保费。
⑤ Buhlmann 信度保费与平方损失条件下的贝叶斯估计得到的保费是一致的
这一点非常有用，可以通过平方损失条件下贝叶斯估计求得 Buhlmann 信度保费的信度，同时可以通过 Buhlmann 信度保费的信度求得平方损失条件下贝叶斯估计。
（4） Buhlmann Straub 信度
实际就是加权的 Buhlmann 信度，具体来说是对同质方差的加权，可以参考韩天雄主编的教材，此处不再赘述。
【重难点解析】
【例 3.1】某公司在 2002 年中共发生了 100 个理赔个案，并按损失大小整理出下列频率密度直方图。但在整理过程中，数字 x 和 y 遗失了，仅知道有 60 个理赔在 x 和 y 之间发生。试根据直方图信息求 x 的值。

解频率密度直方图面积为 1，即 ∫3000100f（x）dx＝1

0.0006×（y-x）=0.6

0.0009×（x-100）+0.0001×（3000-y）=0.4

得 x=362.5
【例 3.2】已知索赔记录如下
表所示。

索 赔 次 数
平均索赔金额
2001 年
100
1000
2002 年
200
1250

每年的通货膨胀率为 10%，现用 Pareto（3，λ）分布作为 2003 年内平均索赔金额的模型。求 λ 的矩估计值。
解
E（X）＝λ α-1＝λ 3-1＝
x-＝100300×1000×1.12＋200300×1250×1.1＝1320

λ ＝2640

部分保险数据损失分布的条件分布问题，是容易让读者疑惑的问题，在这里给予重点说明如下。
存在免赔额时保险人承担的风险 Y 和保险标的风险 X 的关系为

Y=XX＞d

Y 未定义，其他

此时

FY（y）=P（Y≤y）=0，y<d

P（Y≤y|X>d）=Fx（y）-Fx（d）1-Fx（d），y≥d

fY（y）=0，y<d

fx（y）1-Fx（d），y≥d

E（Y）＝∫∞dxfx（x）dx

超额分保再保险（自留额为 d）的再保险人承担的保险责任与保险标的风险的关系为

Y=X-d，X＞d

Y 未定义，其他

此时

FY（y）=P（Y≤y）=0，y<0

P（Y≤y|X>d）=Fx（y+d）-Fx（d）1-Fx（d）
，y≥0

fY（y）=0，y<0

fx（y+d）1-Fx（d），y≥0

E（Y）＝∫∞d（x-d）fx（x）dx

【例 3.3】设原保险人与再保险人签订超额分保合同，自留额为 d，无最高限额，设保险标的风险 x 服从参数为（u，σ2）的对数正态分布，求此时再保险人承担的保险责任 y 的期望值。

解
lnx～N（u，σ2），

y=x-d，x＞d

y 未定义，其他

E（Y）＝∫∞d（x-r）fx（x）dx＝
∫∞d（x-d）1σ2π
e-12lnx-uσ2xdx

＝expu+σ22Φ
ln d-u-σ2σ2+
1-Φln d-uσ

（此积分较为复杂，本例宜记住）
同时，E（Y）＝E（X）-∑d-11
［1-Fx（x）],

E（yd+1）=E（yd）-［1-Fx（d）］

【例 3.4】
F（x）＝1-qe-px（x≥0），求其矩母函数。
解这是一个典型的混合分布函数，
由 F（x）＝1-qe-px＝p×1＋q×（1-e-px）知其是
p(x＝0)＝1 和参数为 p 的指数分布的混合分布，前者的概率是 p，后者的概率是 q(＝1-p)，
F（x）＝pqe-px（x≥0），且 F（x=0）=p（x=0）=p，故矩母函数

MX（t）=E（etx）=p×e0＋∫∞0etxpe-pxdx＝p＋qpp-t

实际上就是 p（x＝0）＝1 和参数为 p 的指数分布的混合分布的矩母函数的线性组合。
贝叶斯估计是重点内容之一，该估计可应用于多个领域，并且程式化较强。
贝叶斯估计主观性的来源： 贝叶斯方法主观性主要来源于先验分布的选择和对损失函数的选择。
【例 3.5】被保险人每年的索赔次数服从泊松分布，50%的被保险人每年平均索赔次数为 2，另外 50%的被保险人每年平均索赔次数为 4，随机选择一位被保险人，发现其在开始两年的索赔次数都是 4 次，求该被保险人在第三年内索赔次数的贝叶斯估计值。
解
先验分布 p（λ＝2）＝p（λ＝4）＝0.5

p（x1＝x2＝4|λ=2）＝24e-24!2＝0.00814

同理，p（x1＝x2＝4|λ=4）＝0.03817

p（λ=2|x1＝x2＝4）

＝p（x1=x2=4|λ=2）p（λ=2）

p（x1=x2=4|λ=2）p（λ=2）+p（x1=x2=4|λ=4）p（λ=4）

＝0.18213

同理，p（λ=4|x1＝x2＝4）＝0.81787

E（λ|x1＝x2＝4）＝2×0.18213＋4×0.81787＝3.6357

【例 3.6】设
x1，x2，…，xn 是来自总体分布的伯努利分布的样本。
伯努利分布的参数为 p，其先验分布为均匀分布 U（0，1）。设后验分布的均值为 15。试求 n（n≤5）和∑xi。
解伯努利分布函数为
f（x|p）=pxq1-x，L（x|p）=pxiq
(1-xi)＝pxiqn-xi，其先验分布为均匀分布即π（p）=1，则后验分布π（p|x）=f（x|p）•π（p）

∫10f(x|p)π(p)dp
为 Beta 分布 BetaΣxi+1，n-Σxi+1，均值为
Σxi+1n+2＝15，所以

n＝3，Σxi＝0

【例 3.7】已知索赔额变量的先验分布是 Pareto（α，10），参数α等可能地取 1，2，3。现观察到某风险的索赔额为 20。计算该风险下次索赔额大于 30 的概率。
Pareto（α，λ）的函数密度为 f（x）＝αλα（λ+x）α+1，x＞0
解λ＝10，f（x＝20|α=1）
＝αλα（λ+x）α+1
＝1×10α（10+20）α+1＝
0.011111
同理可得
f（x＝20|α=2）＝0.007407，f（x＝20|α=3）＝0.003704
先验分布
P（α＝1）＝P（α＝2）＝P（α＝3）＝13
后验分布
π（θ|x）=
f（x|θ）π（θ）∫f（x|θ）π（θ）dθ
P（α＝1|x=20）＝0.011111×130.011111×13+0.007407×13+0.03704×13＝0.5
同理 P（α＝2|x=20）＝0.33，P（α＝3|x=20）＝0.17
索赔额大于30的概率应该是在α取各可能值下索赔额大于30的概率乘以α取各可能值的后验概率，其中 P（x＞30|α=n）=∫∞30αλα（λ+x）α+1dx=λα（λ+α）-α（λ＝10），具体的将α＝1，2，3 代入略，因此，p（x＞30）＝Σ3n=1P（x＞30|α=n）P（α=n|n=20）＝0.15。
贝叶斯估计和极大似然估计的区别与联系是两者都需要求得联合密度函数。区别在于贝叶斯估计需要已知θ的先验分布和求解θ的后验分布，极大似然估计不需要； 贝叶斯估计是以极小化估计值和真值的误差为目标，极大似然估计以极大化联合密度函数为目标。
信度理论是一个难点，其涉及的统计知识已经超出正常的数理统计范围，我们具体例说如下。
（1） 有限扰动信度
【例 3.8】
保险公司有 8000 份保单，当年总赔付成本是 450000 元，而总风险保费是 520000 元。已知满足完全可信性条件是保单达到 14833 份。试问保险公司在下年应收取多少风险保费。

解
z=minnn0，1＝min800014833，1＝0.735

p^＝0.735×450000＋（1-0.735）×520000＝468550

因此，下一年每份保单的风险保费为 4685508000＝58.6 元

（2） Buhlmann 信度

【例 3.9】现有 4 类风险共 7 年损失记录。Xij 表示第 i 类风险在第 j 年内的损失。已知：

∑4i=1∑7j=1

（xij-x-i）＝33.6

∑4i=1（x-i-x-）＝3.3

按最小平方信度方法，求信度因子。

解 v^=∑ri=1

v^（θi）r=∑ri=1∑nj=1（xij-x-i）2r（n-1）

a^=∑ri=1（x-i-x-）2r-1- v^n

∴ v^＝33.64×（7-1）＝1.4

a^＝3.34-1-1.47＝0.9

∴ z=nn+ v a＝77+1.40.9＝0.82

（3） Buhlmann 信度与平方损失条件下的贝叶斯估计

【例 3.10】某类保险产品索赔次数 x 服从参数为 θ 的泊松分布，而参数 θ 服从参数为 α，β 的 Γ 分布，它的密度函数为

f（θ）=β Γ（α）e-β θ（β θ）

α-1（θ>0）

先观察 n 份保单，在责任期内索赔次数分别为 Xi，试求 θ 的贝叶斯估计。

解联合密度

f（x1，x2，…，xn|θ）=∏e-θ

θ xx!=e-nθ

θ xi∏（xi!）

∝e-（β+n）θ θ xi+α-1

这是以∑xi+α，β+n 为参数的 Γ 分布，在平方损失条件下 θ 的贝叶斯估计为

∑xi+α β+n=β β+nα β+nβ+n∑xin

【例 3.11】对上例采用 Buhlmann 信度求其信度及其信度估计。

解由于索赔次数 x 服从参数为 θ 的泊松分布，因此

E（x|θ）=Var（x|θ）=θ，又由于参数 θ 服从参数为 α，β 的 Γ 分布所以

E（Var（x|θ））=E（θ）=α β，
Var［(E（x|θ))］=Var（θ）=α β2，

ν a=E（Var（x））Var（E（x））=β

z=nn+va=nn+β

μ=E（θ）=α β

所以（1-z）μ+zx-=β β+nα β+nβ+n∑xin，和例 3.7 结果一致，其中可以挖掘的其他知识不再赘述。
由于贝叶斯估计的重要性，我们再给出几个例子来加深读者的认识。
【例 3.12】设某运输车队每年的事故发生数服从泊松分布，参数λ可取 1.0 或 1.5，又设λ的先验分布为
P（λ=1.0）=0.4，P（λ=1.5）=0.6，假如某一年该车队发生了三次事故，求λ的后验分布，并在平方损失函数下求参数的贝叶斯估计。
解 P（x=3|λ）=λ33!e-λ

P（x=3|λ=1）=13！e-1

P（x=3|λ=1.5）=1.533！e-1.5

P（λ=1|x=3）=e-1×0.43!1.533!e-1.5×0.6+13!e-1×0.4=0.2457

P（λ=1.5|x=3）=e-1.5×0.6×1.533!1.533!e-1.5×0.6+13!e-1×0.4=0.7543

所以λ^=E（λ）=P（λ=1|x=3）×1+P（λ=1.5|x=3）×1.5

=0.2457+0.7543×1.5=1.3772
【例 3.13】设θ是某险种的索赔频率，抽取 8 份保单，发现在有效期内有三次索赔，假如先验分布为:
（1） θ~U（0，1）;
（2） θ~f（θ）=2（1-θ），0<θ<1

0，其他
分别求θ的后验分布，并在平方损失函数下求参数的贝叶斯估计。

解（1） P（x=3|θ）=C38θ3（1-θ）5
P（θ|x=3）∝C38θ4-1（1-θ）6-1~Beta（4，6）
所以参数的贝叶斯估计为

θ^=44+6=25

（2） P（θ|x=3）∝C38θ4-1（1-θ）6-12（1-θ）

∝C38θ4-1（1-θ）7-1

~Beta（4，7）

所以参数的贝叶斯估计为

θ^=44+7=411

【例 3.14】从一组有效保单中抽取 100 份，发现有三个索赔，假如该险种的索赔频率θ的先验分布为
Beta（2，200），求θ的后验分布. 并在平方损失函数下求参数的贝叶斯估计。

解 P（θ|x=3）∝C3100θ3（1-θ）97Cθ2-1（1-θ）200-1
∝C×θ5-1（1-θ）297-1~Beta（5，297）

所以参数的贝叶斯估计为

θ^=55+297=5302

注:上两题中，在 0 1 误差函数下的贝叶斯估计，该是多少呢?请读者思考。
【例 3.15】关于参数θ的贝叶斯估计，下列选项哪一项是正确的?
① 在二次损失函数下，θ的估计是后验分布的中位数。
② 在二次损失函数下，θ的估计是后验分布的众数。
③ 在 0 1 误差函数下，θ的估计是后验分布的均值。
④ 在 0 1 误差函数下，θ的估计是后验分布的众数。
A. 仅①正确 B. 仅②正确 C. 仅③正确

D. 仅④正确 E. 全都不正确

解选 D。贝叶斯估计有三种损失函数，它们对应的参数的后验分布的估计要牢记。

【例 3.16】设 p 的先验分布为（0，1）上的均匀分布，已知 x1，x2…，xn 是来自总体分布为二点分布的样本，二点分布的参数为 p，并且已知后验分布的均值为 14，问以下结论哪一个是正确的?

A. ∑xi=1，n=2B. ∑xi=1，n=4C.∑xi=0，n=2

D. ∑xi=0，n=4E.∑xi=0，n=6

解根据贝叶斯公式，p 的后验分布

f（p|x1，x2，…，xn）∝f（p）f（x1，x2，…，xn）

∝px1（1-p）1-x1px2（1-p）1-x2…pxn（1-p）1-xn

∝pni=1xi+1-1（1-p）

n-∑ni=1xi+1-1

~Beta∑xi+1，n-∑xi+1

所以其均值为

∑xi+1∑xi+1+n-∑xi+1=∑xi+1n+2=14

选 C，∑xi=0，n=2 可以满足要求。

【例 3.17】设定某种疾病发病次数服从泊松分布，大约一半的人每年的发病次数为 1 次，另一半的人每年发病次数为 2 次，随机选取一人，发现其在前两年的发病次数均为 1 次，求该人在第三年内的索赔 1 次的贝叶斯估计值。

A.22+e2B. 44+e2C. 4+e28+e2

D. 8+e24+e2 E. 2+e24+e2

解涉及贝叶斯估计，必须明确以下几个问题:

（1） 总体分布是什么；

（2） 参数是哪一个；

（3） 参数的先验分布；

（4） 样本值；

（5） 损失函数是哪一个。

对本题而言，设 X 表示发病次数，则 X 为总体，X~Possion（λ）。λ是参数，由“大约一半的人每年的发病次数为 1 次，另一半的人每年发病次数为 2 次”，则λ的先验分布为 P（λ=1）=P（λ=2）=0.5； 那样本值是多少? “随机选取一人，发现其在前两年的发病次数均为 1 次”，说明样本容量为 2，且 x1=x2=1。这里的损失函数没作说明，但这是离散型随机变量，就取平方损失函数即可。

λ的后验分布

P（λ=1|x1=1，x2=1）

=P（λ=1）P（x1=x2=1|λ=1）P（λ=1）P（x1=x2=1|λ=1）+P（λ=2）P（x1=x2=1|λ=2）

=11+4e-2

同理，

P（λ=2|x1=1，x2=1）

=P（λ=2）P（x1=x2=1|λ=2）P（λ=1）P（x1=x2=1|λ=1）+P（λ=2）P（x1=x2=1|λ=2）

=4e-21+4e-2

所以第三年内索赔次数估计的贝叶斯估计值为（也就是后验分布的期望）

λ^=E（λ|x1=1，x2=1）=1×11+4e-2+2×4e-21+4e-2=8+e24+e2

选 D。

【例 3.18】有关贝叶斯方法的陈述，下列选项中正确的是哪一项?

① 在 0~1 损失函数下，贝叶斯方法得到的信度因子的估计与最小平方信度是一致的；

② 在估计非线性问题时，贝叶斯方法比最小平方信度更有优越性；

③ 贝叶斯方法含有主观成分，此主观成分主要表现在对先验分布及损失函数的选取上。

A. 仅①正确 B.仅②正确 C.仅③正确

D. ②③正确 E.全部正确

解②③正确，选 D。

在平方损失函数下，贝叶斯估计值才与最小平方信度一致。

最小平方信度方法实际上更倾向于一个线性模型，而贝叶斯方法没有此限制。

【例 3.19】若 X 服从参数为 p 的几何分布，p 为随机变量且 p 服从参数为（α，β）的贝塔分布，那么 p 的后验分布是什么?

A.贝塔分布 B.几何分布 C.均匀分布

D.参数为（α+1，x+β）的贝塔分布

E.参数为（x+β，α+1）的贝塔分布

解 f（p|x）∝f（p）f（x|p）

∝Γ（α+β）Γ（α）Γ（β）pα-1（1-p）β-1（1-p）β-1（1-p）xp

∝pα+1-1（1-p）β+x-1~Beta（α+1，β+x）

选 AD。

【例 3.20】设 40 张同类保单，用 Xi 表示第 i 张保单的索赔次数，并设 Xi~P（λ），i=1，2，…，40，又设参数 λ 为随机变量，且服从均值为 0.6，方差为 0.02 的 Γ（α，β）分布，分布密度为 f（x）=βαΓ（α）xα-1e-βx，x>0，已知观察到 40 张保单共有 18 次索赔，试计算在平方损失函数下 λ 的贝叶斯估计。
解 λ 的后验分布

f（λ|x1，x2，…，xn）∝f（λ）•f（x1，x2，…，xn|λ）

∝λα-1e-βλ•e-λλx1x1!•e-λλx2x2!•…e-λλx40x40!

∝λα-1e-βλe-40λλ40i=1xi

∝λα+xi-1e-（β+40）λ

~Γα+∑40i=1xi，β+40

由条件有αβ=0.6，αβ2=0.02，可解得

α=18，β=30

在平方损失函数下，λ 的贝叶斯估计为后验分布的均值。
所以 λ 的贝叶斯估计为

λ^=α+∑40i=1xiβ+40=18+1830+40=3670=0.51428

【例 3.21】评估损失分布的贝叶斯方法是一种主观方法，其主观性的主要表现有几个?
A.1 个 B.2 个 C.3 个
D.4 个 E.以上都不对
解选 B。
【例 3.22】（多选）在用贝叶斯方法估计损失分布中，其主观性表现在何处?
A. 选择先验分布 B.确定似然函数 C.确定参数 θ 的后验分布

D.选择损失函数 E.估计参数
解选 AD。本题是上一题的详细考察。
【例 3.23】（多选）以下对未知参数的贝叶斯估计中，分别选择二次损失函数、绝对误差函数、0　1 误
差函数，得到结果相同的有哪几项?

A. 总体服从 N（μ， σ2），未知参数μ的先验分布为 N（μ′， σ′2），求未知参数μ的贝叶斯估计

B. 总体服从对数正态 log N（μ， σ2），未知参数μ的先验分布为 U（0，1），求未知参数μ的贝叶斯估计

C. 总体服从 Exp（λ），未知参数λ的先验分布为 U（0，∞），求未知参数λ的贝叶斯估计

D. 总体服从 Exp（λ），未知参数λ的先验分布为Γ（α′， λ′），求未知参数λ的贝叶斯估计

E. 总体服从Γ（α， λ），未知参数λ的先验分布为Γ（α′， λ′），求未知参数λ的贝叶斯估计

解在贝叶斯估计中，二次损失函数下的贝叶斯估计为后验分布的均值，绝对误差损失函数下为中位数，0　1 误差损失函数下为众数。要想结果相同，也就是要求后验分布的均值，同时它也是众数，也是中位数，即均值=众数=中位数。这就要求后验分布为对称分布才行，如正态分布，t 分布等。

A 中，参数μ的后验分布是正态分布；

B 中，参数μ的后验分布是正态分布；

C 中，参数λ的后验分布是伽玛分布；

D 中，参数λ的后验分布也是伽玛分布；

E 中，参数λ的后验分布也是伽玛分布；

故选 AB。

【习题解答】

1. 从一组有效保单中抽取 100 份，发现有三个索赔，假如该险种的索赔频率θ的先验分布为贝塔（2，200），求θ的后验分布均值。

解 P（θ|X=3）=C3100θ3

（1-θ）97Cθ2-1（1-θ）200-1

∫10C3100θ3（1-θ）97Cθ2-1

（1-θ）200-1dx

=Cθ5-1（1-θ）297-1

∴ E（θ）=55+297=5302

2. 设 40 张同类保单，用 Xi 表示第 I 张保单的索赔次数，并设 Xi～P（λ），I=1，2，…，40，又设参数λ为随机变量，且服从均值为 0.6，方差为 0.02 的伽玛（α， β）分布，并且已知观察到 40 张保单共有 18 次索赔，试计算在平方损失函数下λ的贝叶斯估计。

解由已知条件可知 40 张报单的联合概率密度函数为

f（x1，x2，…，x40|λ）

=∏40i=1λxiXi!e-40λ

=λxie-40λx1!x2!…x40!

对于Γ（α， β）分布，均值为αβ，方差为αβ2。

由已知有αβ=0.6，αβ2=0.02

得α=18，β=30

f（λ|x1，x2，…，x40）

=∏40i=11Xi!λxie-40λαβλα-1e-βλ∫+∞0∏40i=11Xi!λxie-40λαβλα-1e-βλdλ

=（40+β）α+xiΓα+∑xiλα+xi-1e-（40+β）λ

∴ 在平方损失函数下，λ的估计为

λ^=α+∑xi40+β=0.5143

3. 在对某工业产品的保险研究中，需要知道该产品在一年内损坏的概率。已知同为用过两年的该产品 500 件，在一年观察期中观察到 15 件损坏，求该产品在一年内损坏概率的极大似然估计。

解设该产品在一年内损坏的概率为 q，那么一年内损坏的件数 N 服从参数 500 与 q 的二项分布。似然函数 L（q）=P（N=k）=Ck500qk（1-q）500-k

ln ［L（q）］=ln Ck500+kln q+（500-k）ln（1-q）

令

d［ln L（q）］dq=kq+k-5001-q=0

得到 q^=k500=15500=0.03

4. 试求一枚匀称硬币投掷 4 次所得正面朝上次数的分布函数经验分布函数 F（x）。

解正面 0 次朝上的概率为： 116；

正面 1 次朝上的概率为： 416；

正面 2 次朝上的概率为： 616；

正面 3 次朝上的概率为： 416；

正面 4 次朝上的概率为： 116。

∴ F（x）图形如下。

F（x）=0，x<0

116，0≤x<1

516，1≤x<2

1116，2≤x<3

1516，3≤x<4

1，x≥4

5. 下表显示了索赔情况调查所揭示的索赔分布情况，画出观察到的分布函数的图形，并用它估计在 150~200 美元之间发生索赔的概率。

索赔额频数索赔额频数
0~10045500~10008
100~200101000 以上 0
200~50023

解设 q 为所估计的索赔频率，于是似然函数

L（q）=（1-q）100-2λe-0.2λ λe-0.6λ
=e-98λλ2e-0.8λ

ln L=-98.8+2λ

令 d(ln L)dλ=-98.8+2λ=0
∴ λ^=298.8
所求索赔频率的极大似然估计为

q^=1-e-λ^=0.02

6. 在某一特定年度中，1000 份保险单中发生赔款 140 起。假定某一保单在一定时期遇到赔款的次数服从泊松分布。估测某特定保单持有人在 9 个月中不遇到损失的概率。
解理赔次数服从泊松分布，因此，我们假定某一保单在三个月内遇到的索赔次数是一个以 q 为均值的泊松变量，并且假定每一时间区段赔款相互独立。从而一份保单在 12 个月内总索赔次数是以 4q 为均值的泊松变量。对于 1000 份保单，其均值为 4000q。通过赔款次数 140，我们估计 q 为 1404000=0.035，
那么一保单在 9 个月中遇到索赔次数是以 0.105（0.035×3）为均值的泊松变量，则由泊松分布可知，其无理赔的概率为

（0.105）0e-0.1050!=e-0.105=0.90

7. 下表概括了某保险公司 100 个赔款样本的赔款额状况。若该分布适用反对数正态分布模型，求其参数 μ 和 σ 的矩估计值，并估计一笔赔款规模超过 4000 美元的概率。

索赔额频数索赔额频数
0~40022400~28003
400~800242800~32001
800~1200323200~36001
1200~1600213600 以上 0
1600~200010 总数 100
2000~24006

解平均索赔额=200×2100+600×24100+…+3400×1100
方差=2002×2100+6002×24100+…+34002×1100-12162=362944
而对数正态分布的均值与方差分别为 eμ+12σ2 和 e2μ+σ2（eσ2-1）

eμ+12σ2=1216

e2μ+σ2（eσ2-1）=362944

解得 μ=6.993 σ=0.469
某一特定索赔 X 超过 4000 美元的概率等于 lnX 大于 8.29 的概率为

1-Φ8.294-6.9930.469=1-Φ（2.77）=0.00280

8. 下表给出了 4000 份保单的赔款经历情况，其中每一保单承担风险期限为一年。假定每保单经历赔款的次数服从泊松分布，用最大似然估计原理估计泊松分布的参数。

赔 款 次 数被观察的保单数赔 款 次 数被观察的保单数
0328834
1642 总和 4000
266

解未发生索赔的概率为 e-qq00!
发生一次概率为 e-qq11!

发生二次概率为 e-qq22!
发生三次概率为 e-qq33!
∴ 似然函数

L=e-qq00!3288
e-qq11!642
e-qq22!66
e-qq33!4

=e-4000qq7989.56×1022

L=-4000q+786ln q-52.91

令 d(ln L)dq=-4000+786q
∴ q^=786/4000=0.196
9. 求上题中数据所代表的风险种类的赔款频率的置信度为 95％的置信区间。
解由中心极限定理的正态逼进结论可知，
索赔总次数是以 4000q 为均值。4000q 为方差的正态变量。

∴ N-4000q4000q～N（0，1）

令 N=486 并使 N-4000q4000q 等于标准正态分布上、下 2.5%点，即

786-4000q4000q=±1.96

∴ 16000000q2-6303366q+617796=0

∴ q=0.183 或 0.211

即 q 的 95%置信区间为 0.183 至 0.211
10. 某保险公司在某种业务中有 100 项索赔的样本，已在题 7 中给出。其索赔频率约为 0.015，计算当对此索赔经历设定完全可信性（k＝0.05，P＝0.9）时未满期责任的最小容量。
解由第 7 题可知，索赔额均值 m 为 1216，方差 σ 2 为 362944。因此 σ /m=0.50，
将此数字及对应 k=0.05，p=0.9 的值（见下表）
代入公式 nF=n0［1+（σ /m）2］，
可得 nF 为 1353，而已知索赔频率为 0.015，所以可以推算具有完全可信的最小未满期责任业务量约 90000 张保单。

k
p
0.30.20.10.050.01

0.93068271108227060
0.954396384153738416
0.9970166663265466358
0.9991202711083433110827 4

注： 对应于 k 与 p 的可信性参数 n0。

11. 在上题中的保险公司在其未满期业务中有 19307 份保险单。其总的风险保费为 366833 元，一年内索赔总额为 340575 元。试求保险公司在下一年对每份保险合同应收取的风险保费的可信性估计值。

解由上题，完全可信性的最小期望索赔次数为 1353，因此，对于索赔频率为 0.015 而未满期责任为 19307 份保单的期望索赔次数为 290。

由 z=nznF=0.46

对 19303 份保单应收取的总风险保费的可信性估计

按公式 C^z=（1-z）C^C+zC

算出 C^C=366833z=0.46C=340575

∴ C^z=354754

∴ 保险公司在下一年对每一张保险合同应收取的风险保费可信性估计为 35475419307=18.37 元，而目前每一份收取的保费为 19.00 元。

12. 在城市机动车辆保险保单持有人的报案中，90％的事故发生在市内，10％则发生在市外。而非城市保单持有人的报案中，15％的事故发生在市内，85％发生在市外。已知该保险公司的保单持有人 80％居住在市内。现得到一市外发生交通事故的报告，问该保单持有人系市内居住者的概率为多少?

解以 D 表示居住地，

L 表示事故发生地，D 分为城市（u）和乡村（r），L 分为 u 和 r。

那么

P（L=u|D=u）=0.9

P（L=r|D=u）=0.1

P（L=u|D=r）=015

P（L=r|D=r）=0.85

P（D=u）=0.8

P（D=r）=0.2

∴ P（D=u|L=r）

=P（D=u）P（L=r|D=u）P（D=u）P（L=r|D=u）+P（D=r）P（L=r|D=r）

=0.8×0.10.8×0.1+0.2×0.85

=0.32

13. 设下表中的理赔记录用韦伯分布来拟合，试用 0.2 和 0.7 分位点估计参数 r。

0.1
0.2
0.3
0.4
0.5
0.6
0.8
0.95
0.98
1.0

解 0.2 分位点为 0.25
0.7 分位点为 0.875
分别令 0.2=1-e-cxr
0.7=1-e-cxr
得 x=-ln 0.8c1r 和 x=-ln 0.3c1r
将 0.25 和 0.875 代入上面两式有

0.25=-ln 0.8c1r

0.875=-ln 0.3c1r

∴ r=1.35

14. 已知总体分布为Γ（α， λ），其中参数λ的先验分布为 exp（λ′），试问其后验分布是什么?

解 f（λ|x1，x2，…，xn）∝f（x1，x2，…，xn|λ）f（λ）

∝∏ni=1（λ αe-λxi）• e-λ′ λ=λnαe-xiλ • e-λ′ λ

=λnαe-（xi+λ′） λ

即其分布为Γnα， ∑xi+λ′ +1

15. 设某保险人经营某种车辆险，对过去所发生的 1000 次理赔情况作了记录，平均理赔为 2200，又按赔付金额分为 5 档，各档中的记录次数如下。

理赔额	次数
0～1000	200
1000～2000	300
2000～3000	250
3000～4000	150
4000～5000	100
5000 以上	0

试用χ2 分布检验判断能否用指数分布模拟个别理赔额的分布（假设置信水平为 99.5%）。

解先假设个别理赔额 X~exp（λ）

用矩方法或最大似然法都得到λ^=1x-=12200

为了计算 Ei，先计算个别理赔额落入每一个档次内的概率，

比如 2000～3000 内概率为

∫30002000λe-λxdx=e

-2000λ-e-3000λ=0.1472

在 2000～3000 内平均次数 E2=147.2

类似地算出 E1=365.3，E3=231.8，E4=93.4，E5=59.3，E6=103

χ2 统计量的值为

χ2=（200-365.3）2365.3+（300-231.8）2231.8+…+（0-103）22=331.89

由于只有一个参数，并且记录完整。χ2 分布的自由度为 6-1-1=4，查 χ2 分布表在置信水平 99.5%下值为 14.86，远低于观察值。
故应拒绝原假设，即选择指数分布不恰当。
16. 根据上题所给出的损失记录，画出频率直方图和频率折线图。问：
（1） 这两条曲线和概率论中什么曲线相似?
（2） 根据频率直方图和频率折线图，你准备用什么分布来模拟个别理赔额分布？
解频率直方图如下。

频率折线图如下。

由于这两条曲线具有尾部较厚的特点，可以考虑用广义帕雷托分布模拟。
17. 对于承保某一类风险，某保险公司将其保费计算基于平均赔款额为 1200 美元的假定之上。1980 年中，这类保险项目发生了 1243 次赔款，平均额为 1283. 70 美元，其观察值的标准差等于 1497.31 美元。是否有证据能说明该公司在计算保险费率时所假定的平均赔款额太低了呢?
解根据中心极限定理及独立随机变量之和的性质，我们假设这 1243 起索赔支出数的分布近似于均值为 1243μ，方差为 1243σ2 的正态分布
即均值为 1243×1200=1491600
标准差近似为 1243×（1497.31）2=52789
1980 年实际总支出额为 1283×1243=159639。在原假设条件下，真正的平均索赔额为 1200，则 1243 起赔款中支出为 1595639 美元，或更多数目的概率为

1-Φ1595639-149160052789=1-Φ（1.97）=0.024

显然偏低，即平均索赔额假设太低了。
18. 试证明采用绝对误差函数时，参数 θ 的贝叶斯估计是后验分布的中位数。

证： E［Loss（θ^， θ）］=∫+∞-∞Loss（θ^， θ）f（θ|x）dθ
=∫+∞-∞|θ^-θ|f（θ|x）dθ
=∫ θ^-∞ （θ^-θ）f（θ|x）dθ+
∫+∞ θ^（θ^-θ）f（θ|x）dθ
=∫ θ^-∞ θ^f（θ|x）dθ-∫ θ^-∞ θf（θ|x）dθ
+∫+∞ θ^θ^f（θ|x）dθ-
∫+∞ θ^θf（θ|x）dθ
=θ^-2θ^∫+∞ θ^f（θ|x）dθ-∫ θ^-∞ θf（θ|x）dθ+∫+∞ θ^θf（θ|x）dθ
∴ θ^E［Loss(θ^,θ)］=1-2∫+∞ θ^f（θ|x）dθ+2θ^f（θ^|x）-θ^f（θ^|x）-θ^f

（θ^|x）
=1-2∫+∞θ^f（θ|x）dθ

令　　θ^E［Loss(θ^,θ)］=0 有
∫+∞θ^f（θ|x）dθ=12

即θ^为后验分布的中位数。
19. 设θ是某险种的索赔频率，抽取 8 份保单，发现在有效期内有三次索赔，假如先验分布为
（1）　θ～U（0，1）；
（2）　θ～f（θ）=2（1-θ），0<θ<1。
分别求θ的后验分布。
解（1）　P（X=3|θ）=C38θ3（1-θ）5

P（θ|X=3）=C38θ4-1（1-θ）6-12（1-θ）
∫10C38θ4-1（1-θ）6-12（1-θ）dθ
=C38θ4-1（1-θ）6-1∫10C38θ4-1（1-θ）6-1dθ

∴　θ后验分布为 Beta（4，6）
（2）　P（θ|X=3）=C38θ4-1（1-θ）6-12（1-θ）∫10C38θ4-1（1-θ）6-12（1-θ）dθ

=C•θ4-1（1-θ）7-1
其中 C 为常数，故θ后验分布为 Beta（4，7）
20. 假设某一工业新产品的寿命服从均匀分布，参数为 e，其密度函数为 f（t）=1/e。现对 7 件这样的产品进行观察，得到如下的产品余命为 3，4，5，7，8，10，12。试求 e 的最小二乘估计。
解由 f（t）=1e　可知

S（t）=e-te

SS=∑7i=1S（ti）-12［S^（ti）+S^（ti-1）］2

=∑7i=1
1-tie-12
［S^（ti）+S^（ti-1）］2

dSSde=2∑7i=1
1-tie-12［
S^（ti）+S^（ti-1）］tie2

令 dSSde=0
可得到 e 的最小二乘估计表达式

e^=∑7i=1（ti）2
∑7i=1ti-∑7i=1ti12（S^（ti）+S^（ti-1））

具体计算如下表所示。

i ti t2iS^（ti） ti12
［S^（ti）+S^（ti-1）］

1390.92.85
24160.83.4
35250.73.75
47490.54.2
58640.43.6
6101000.12.5
7121440.00.6
∑4940720.90

故有 e^=40749-20.9=14.48