DOI:10.13656/j.cnki.gxkx.2012.02.021

广西科学 Guangxi Sciences 2012,19(2):125~128

复合 LINEX 对称损失下 Poisson 分布参数的 Bayes 估计和 E-Bayes 估计*

Bayesian Estimation and E-Bayesian Estimation for Parameters of Poisson Distribution in Composite LINEX Loss of Symmetry

韦 师,莫达隆,苏玉华
WEI Shi,MO Da-long,SU Yu-hua

(贺州学院数学系,广西贺州 542800)
(Department of Mathematics' Hezhou University,Hezhou,Guangxi,542800,China)

摘要: 在复合 LINEX 对称损失函数下,给出 Poisson 分布参数 λ 的 Bayes 估计及 E-Bayes 估计,并通过数值模拟验证 Bayes 估计及 E-Bayes 估计的合理性及优越性.

关键词: Poisson 分布 复合 LINEX 对称损失函数 Bayes 估计 E-Bayes 估计
中图法分类号: O212.0 **文献标识码:** A **文章编号:** 1005-9164(2012)02-0125-04

Abstract: The Bayesian estimation and E-Bayesian estimation for parameters of Poisson distribution were studied mainly in the composite LINEX loss of symmetry, and their rationality and superiority were verified through numerical analysis.

Key words: Poisson distribution, composition LINEX loss of symmetry, Bayesian estimation, E-Bayesian estimation

Poisson 分布是概率论中一种重要离散分布,适合于描述单位时间(或空间)内随机事件发生的次数,在实际生活中有着广泛的应用.例如,如某一服务设施在一定时间内到达的人数、电话交换机接到呼叫的次数、汽车站台的候客人数、机器出现的故障数、自然灾害发生的次数、一块产品上的缺陷数、显微镜下单位分区内的细菌分布数等等都可以用 Poisson 分布来描述.对 Poisson 分布的研究不仅可以提高工作效率,而且可以提高经济效益.国内外对 Poisson 分布已有一定的研究,李建波等[1]经数据模拟和正态分布比较后,验证了 Lenth 方法对服从 Poisson 分布的响应变量在某些情况下是适用的.范洪福[2]建立一种服从 Poisson 分布的随机变量的数学模型,并进行推广.王德辉等[3]研究熵损失下 Poisson 分布参数倒数的估计.徐宝等[4]研究 Poisson 分布参数倒数在一种对称损失下的 Bayes 估计.韦莹莹等[5]讨论 Q-对称熵损失下 Poisson 分布参数倒数的估计问题.最近,韩明[6]提出了一种新的参数估计方法——E-Bayes 估计法,并在只有一个失效数据下,研究失效率的 E-Bayes 估计.而在复合 LINEX 对称损失函数下张睿[7]研究了正态分布及指数分布的参数估计并证明其可容许性.

本文也在复合 LINEX 对称损失函数下,给出 Poisson 分布参数 λ 的 Bayes 估计及 E-Bayes 估计,并通过数值分析验证该 Bayes 估计及 E-Bayes 估计的合理性及优良性.

1 参数 λ 的 Bayes 估计

设随机变量 X 服从参数 λ 的 Poisson 分布,则其分布律为

$$P(X=x_i)=\frac{e^{-\lambda}\lambda^{x_i}}{x_i!},\lambda>0. \qquad (1)$$

当随机变量 X 服从(1)式的 Poisson 分布时,设 $(X_1,X_2,\cdots,X_n)$ 为来自总体 X 的容量 n 的随机 i.i.d 样本,$(x_1,x_2,\cdots,x_n)$ 为其观测值,则其联合密度函数

收稿日期:2012-02-20
作者简介:韦 师(1984-),男,助教,硕士,主要从事概率论与数理统计教学与研究。
* 广西教育厅科研项目(201106LX643,200106YB141),贺州学院科研项目(2011ZRKY13)资助。

为

$$f(x_1,x_2,\cdots,x_n \mid \lambda)=\frac{e^{-n\lambda}\lambda^{T}}{\prod_{i=1}^{n}x_i!}, \qquad (2)$$

其中 $T=\sum_{i=1}^{n}x_i$.

文献[7]提出的复合 LINEX 对称损失函数，其表达形式为

$$L(\theta,\delta)=L_a(\theta,\delta)+L_{-a}(\theta,\delta)=e^{-a(\theta-\delta)}+e^{a(\theta-\delta)}-2,a>0. \qquad (3)$$

事实上由于 $L(\theta,\delta)=e^{-a(\theta-\delta)}+e^{a(\theta-\delta)}-2\geqslant 2\sqrt{(e^{-a(\theta-\delta)})(e^{a(\theta-\delta)})}-2=0$，所以损失函数的函数值是非负的. 又在(3)式中对 δ 求偏导得 $L'(\theta,\delta)=ae^{-a(\theta-\delta)}-ae^{a(\theta-\delta)}$. 任取 $0<\delta_1<\delta_2$，那么就有

$$L'(\theta,\delta_1)-L'(\theta,\delta_2)=a(e^{-a(\theta-\delta_1)}-e^{-a(\theta-\delta_2)}+e^{a(\theta-\delta_2)}-e^{a(\theta-\delta_1)})<0. \qquad (4)$$

即 $L'(\theta,\delta_1)-L'(\theta,\delta_2)<0$，所以对称损失函数 $L(\theta,\delta)$ 关于 δ 是严格凸函数.

定理 1.1 记 $X=(X_1,X_2,\cdots,X_n)$，在复合 LINEX 对称损失函数(3)下，对任何先验分布 $\pi(\lambda)$，λ 的 Bayes 估计为

$$\delta(x)=\frac{1}{2a}\ln(E(e^{a\lambda}\mid X)/E(e^{-a\lambda}\mid X)). \qquad (5)$$

证明 设 $\delta(x)$ 为 λ 的任一估计，在损失函数(3)下，由定义得 $\delta(x)$ 的 Bayes 风险为

$$E[L(\lambda,\delta)]=E\{E[L(\lambda,\delta)]\mid X\}=E\{E[e^{-a(\lambda-\delta)}+e^{a(\lambda-\delta)}-2]\mid X\}. \qquad (6)$$

(6)式左端 $E[L(\lambda,\delta)]$ 表示关于 λ 与样本 $X_1,X_2,\cdots,X_n$ 的联合分布取期望，所以要求 λ 的 Bayes 解，只要求极小化 $E\{[e^{-a(\lambda-\delta)}+e^{a(\lambda-\delta)}-2]\mid X\}$ 即可. 令

$$g(\delta)=E\{[e^{-a(\lambda-\delta)}+e^{a(\lambda-\delta)}-2]\mid X\}=e^{a\delta}E(e^{-a\lambda}\mid X)+e^{-a\delta}E(e^{a\lambda}\mid X)-2. \qquad (7)$$

对(7)式关于 δ 求导并令其等于0，即

$$g'(\delta)=ae^{a\delta}E(e^{-a\lambda}\mid X)-ae^{-a\delta}E(e^{a\lambda}\mid X)=0. \qquad (8)$$

解得 $\delta(x)=\frac{1}{2a}\ln(E(e^{a\lambda}\mid X)/E(e^{-a\lambda}\mid X))$.

事实上由函数 $g(\delta)$ 自身的性质知，$g(\delta)$ 是严格凸函数，因为对任取的 $0<\delta_1<\delta_2$ 有

$$g'(\delta_1)-g'(\delta_2)=aE(e^{-a\lambda}\mid X)(e^{a\delta_1}-e^{a\delta_2})-aE(e^{a\lambda}\mid X)(e^{-a\delta_1}-e^{-a\delta_2})<0. \qquad (9)$$

所以函数 $g(\delta)$ 是关于 δ 的严格凸函数. 从而 $\delta(x)$ 是函数 $g(\delta)$ 唯一的极小值点，所以 λ 的 Bayes 解为

$$\delta(x)=\frac{1}{2a}\ln(E(e^{a\lambda}\mid X)/E(e^{-a\lambda}\mid X)).$$

选取 $\Gamma(\alpha,\beta)$ 为 Poisson 分布参数 λ 的先验分布，则先验分布的密度函数为

$$\pi(\lambda;\alpha,\beta)=\Gamma(\alpha,\beta)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda},\alpha>0,\beta>0,\lambda>0. \qquad (10)$$

由 Bayes 公式得参数 λ 的后验密度为

$$H(\lambda\mid x)=\frac{f(x_1,x_2,\cdots,x_n\mid\lambda)\pi(\lambda;\alpha,\beta)}{\int_0^{\infty}f(x_1,x_2,\cdots,x_n\mid\lambda)\pi(\lambda;\alpha,\beta)\mathrm{d}\lambda}=\frac{e^{-(n+\beta)\lambda}\lambda^{T+\alpha-1}}{\int_0^{\infty}e^{-(n+\beta)\lambda}\lambda^{T+\alpha-1}\mathrm{d}\lambda}=\frac{e^{-(n+\beta)\lambda}\lambda^{T+\alpha-1}(n+\beta)^{T+\alpha}}{\int_0^{\infty}e^{-(n+\beta)\lambda}[\lambda(n+\beta)]^{T+\alpha-1}\mathrm{d}[\lambda(n+\beta)]}=\frac{(n+\beta)^{T+\alpha}e^{-(n+\beta)\lambda}\lambda^{T+\alpha-1}}{\Gamma(T+\alpha)}. \qquad (11)$$

定理 1.2 在复合 LINEX 对称损失函数(3)下，对于先验分布为 $\Gamma(\alpha,\beta)$，Poisson 分布参数 λ 的 Bayes 估计为

$$\delta_B(x)=\frac{T+\alpha}{2a}\ln\left(\frac{n+\beta+a}{n+\beta-a}\right).$$

证明 由(11)式可得 $e^{-a\lambda}$ 及 $e^{a\lambda}$ 的条件期望分别为

$$E(e^{-a\lambda}\mid X)=\int_0^{\infty}e^{-a\lambda}H(\lambda\mid x)\mathrm{d}\lambda=\frac{(n+\beta)^{T+\alpha}}{\Gamma(T+\alpha)}\int_0^{\infty}\lambda^{T+\alpha-1}e^{-(n+\beta+a)\lambda}\mathrm{d}\lambda=\frac{(n+\beta)^{T+\alpha}}{\Gamma(T+\alpha)(n+\beta+a)^{T+\alpha}}\int_0^{\infty}[\lambda(n+\beta+a)]^{T+\alpha-1}e^{-(n+\beta+a)\lambda}\mathrm{d}[\lambda(n+\beta+a)]=\frac{(n+\beta)^{T+\alpha}}{\Gamma(T+\alpha)}\frac{\Gamma(T+\alpha)}{(n+\beta+a)^{T+\alpha}}=\left(\frac{n+\beta}{n+\beta+a}\right)^{T+\alpha}, \qquad (12)$$

$$E(e^{a\lambda}\mid X)=\int_0^{\infty}e^{a\lambda}H(\lambda\mid x)\mathrm{d}\lambda=\frac{(n+\beta)^{T+\alpha}}{\Gamma(T+\alpha)}\int_0^{\infty}\lambda^{T+\alpha-1}e^{-(n+\beta-a)\lambda}\mathrm{d}\lambda=\frac{(n+\beta)^{T+\alpha}}{\Gamma(T+\alpha)(n+\beta-a)^{T+\alpha}}\int_0^{\infty}[\lambda(n+\beta-a)]^{T+\alpha-1}e^{-(n+\beta-a)\lambda}\mathrm{d}[\lambda(n+\beta-a)]=\frac{(n+\beta)^{T+\alpha}}{\Gamma(T+\alpha)}\frac{\Gamma(T+\alpha)}{(n+\beta-a)^{T+\alpha}}=\left(\frac{n+\beta}{n+\beta-a}\right)^{T+\alpha}. \qquad (13)$$

故由定理 1.1 得 Poisson 分布参数 λ 的 Bayes 估计为

$$\delta_B(x)=\frac{T+\alpha}{2a}\ln\left(\frac{n+\beta+a}{n+\beta-a}\right).$$

由于复合 LINEX 对称损失函数是严格的凸函数，所以在此损失函数下的 Bayes 估计也是唯一的，则由文献[8]可知该 Bayes 估计是可容许的.

2 参数 λ 的 E-Bayes 估计

定义 2.1[6,9] 若 $\delta(\alpha,\beta)$ 是参数 λ 的 Bayes 估

计，含有超参数 α,β，并且 $\delta(\alpha,\beta)$ 是连续的，则称

$$\delta_E(x)=\iint_D\delta(\alpha,\beta)\pi(\alpha,\beta)\mathrm{d}\alpha\mathrm{d}\beta \qquad (14)$$

为参数 λ 的 E-Bayes 估计，其中 $\iint_D\delta(\alpha,\beta)\pi(\alpha,\beta)\mathrm{d}\alpha\mathrm{d}\beta<\infty$，$D$ 是 α 和 β 的可能取值范围组成的集合，$\pi(\alpha,\beta)$ 是 α 和 β 在 D 上的密度函数.

定理 1.2 求出参数 λ 的 Bayes 估计 $\delta_B(x)$ 的表达式中含有超参数 α,β，在没有历史经验的情况下，参数 α,β 还是未知的变量. 在此通常把超参数 α,β 看成是随机变量，再根据具体的情况选择适当的先验分布，我们根据 Poisson 分布的性质，采用减函数法[10]来确定超参数数 α,β 的先验分布.

根据 $\Gamma(\alpha,\beta)$ 函数的性质并结合 $\Gamma(\alpha,\beta)$ 函数的图像（图 1）可知，$\Gamma(\alpha,\beta)$ 函数一般有两种情况:

（i）当 $\beta>0,\alpha>1$ 时，$\Gamma(\alpha,\beta)$ 函数图像是单峰曲线;

（ii）当 $\beta>0,0<\alpha<1$ 时，$\Gamma(\alpha,\beta)$ 函数图像是单调递减;

由此可知，当 $0<\alpha<1,\beta>0$ 时，此函数关于参数 θ 为减函数. 所以先验分布可以采取均匀分布，即

$$\pi(\alpha)=U(0,1),\pi(\beta)=U(0,c). \qquad (15)$$

从稳健性的角度考虑，为了保证 Bayes 估计的稳健性，这里的 c 应该有一个界限.

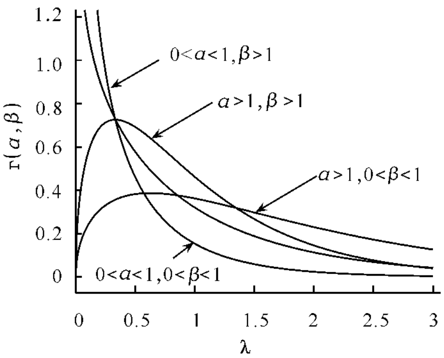

图 1 $\Gamma(\alpha,\beta)$ 分布的密度函数

Fig. 1 Density function image of the $\Gamma(\alpha,\beta)$

在确定好超参数 α,β 的先验分布后，就可以运用参数 E-Bayes 估计的定义来求出 Poisson 分布参数 λ 的 E-Bayes 估计.

定理 2.1 对于 Poisson 分布，若参数 λ 的先验密度函数 $\pi(\lambda;\alpha,\beta)$ 为 Γ 分布 $\Gamma(\alpha,\beta)=\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}$，超参数 α,β 的先验分布为 D 上的均匀分布，则 λ 的 E-Bayes 估计为

$$\delta_E(x)=\frac{2T+1}{4ac}(f(n,a,c)-g(n,a)-f_1(n,a,c)+g_1(n,a)).$$

其中 $f(n,a,c)=(n+a+c)\ln(n+a+c)$，$f_1(n,a,c)=(n+c-a)\ln(n+c-a)$，$g(n,a)=(n+a)\ln(n+a)$，$g_1(n,a)=(n-a)\ln(n-a)$.

证明 由 E-Bayes 的定义得

$$\delta_E(x)=\iint_D\delta_B(x)\pi(\alpha,\beta)\mathrm{d}\alpha\mathrm{d}\beta=$$

$$\int_0^1\int_0^c\frac{1}{c}\frac{T+\alpha}{2a}\ln\left(\frac{n+\beta+a}{n+\beta-a}\right)\mathrm{d}\alpha\mathrm{d}\beta=$$

$$\frac{1}{2ac}\int_0^c\ln\left(\frac{n+\beta+a}{n+\beta-a}\right)\int_0^1(T+\alpha)\mathrm{d}\alpha\mathrm{d}\beta=$$

$$\frac{2T+1}{4ac}\int_0^c\ln\left(\frac{n+\beta+a}{n+\beta-a}\right)\mathrm{d}\beta=\frac{2T+1}{4ac}\int_0^c\ln(n+\beta+a)-\ln(T+\beta-a)\mathrm{d}\beta=$$

$$\frac{2T+1}{4ac}\left[\int_{n+a}^{n+a+c}\ln x\mathrm{d}x-\int_{n-a}^{n-a+c}\ln x\mathrm{d}x\right]=\frac{2T+1}{4ac}\left[(n+a+c)\ln(n+a+c)-(n+a)\ln(n+a)-(n+c-a)\ln(n+c-a)+(n-a)\ln(n-a)\right]=$$

$$\frac{2T+1}{4ac}(f(n,a,c)-g(n,a)-f_1(n,a,c)+g_1(n,a)).$$

其中 $f(n,a,c)=(n+a+c)\ln(n+a+c)$，$f_1(n,a,c)=(n+c-a)\ln(n+c-a)$，$g(n,a)=(n+a)\ln(n+a)$，$g_1(n,a)=(n-a)\ln(n-a)$. 即参数 λ 的 E-Bayes 估计为

$$\delta_E(x)=\frac{2T+1}{4ac}(f(n,a,c)-g(n,a)-f_1(n,a,c)+g_1(n,a)).$$

3 数值模拟

已知某细胞单位所含白血球的个数服从 Poisson 分布，对 1008 个细胞单位进行观察，用 k 表示细胞单位含白血球的个数，n_k 表示 1008 个观测单位中，含 k 个白血球的细胞单位个数. 白血球分布数据为 $n_0=64$，$n_1=171$，$n_2=239$，$n_3=220$，$n_4=155$，$n_5=83$，$n_6=46$，$n_7=20$，$n_8=6$，$n_9=3$，$n_{10}=0$，$n_{11}=1$.

根据这些数据用第 1,2 节所给出的参数 λ 的 Bayes 估计及 E-Bayes 估计，估计出白血球所服从的 Poisson 分布的参数 λ. 取 $a=2$，计算结果见表 1 和表 2. 从表 1 和表 2 可以看出参数 λ 的 Bayes 估计与 E-Bayes 估计比较接近.

为了进行参数 λ 的 Bayes 估计与 E-Bayes 估计精确度的比较，用 R 软件进行模拟. 对 $n=50$ 产生 5000 个 Poisson 分布的随机次序样本，分别求出参数 λ 的 Bayes 估计与 E-Bayes 估计. 从偏差大小的角度去比较参数 λ 的 Bayes 估计与 E-Bayes 估计. 参数真值取 $\lambda=1.36$. 模拟结果见表 3 和表 4:

表1 参数λ的Bayes估计

Table 1 Bayesian estimation of poisson distribution parameters

α	$\delta_B(x)$								$\Delta\delta_B$
	$\beta=0.5$	$\beta=1$	$\beta=2$	$\beta=3.5$	$\beta=4$	$\beta=5.5$	$\beta=7$	$\beta=8$	
0.1	2.8221	2.8207	2.8179	2.8137	2.8124	2.8082	2.8040	2.8013	0.0208
0.25	2.8223	2.8209	2.8181	2.8139	2.8125	2.8083	2.8042	2.8014	0.0209
0.4	2.8224	2.8210	2.8182	2.8140	2.8127	2.8085	2.8043	2.8016	0.0208
0.55	2.8226	2.8212	2.8184	2.8142	2.8128	2.8086	2.8045	2.8017	0.0209
0.7	2.8227	2.8213	2.8185	2.8143	2.8129	2.8088	2.8046	2.8019	0.0208
0.8	2.8228	2.8214	2.8186	2.8144	2.8130	2.8089	2.8047	2.8020	0.0208
0.95	2.8230	2.8216	2.8188	2.8146	2.8132	2.8090	2.8049	2.8021	0.0209
$\Delta\delta_B$	0.0009	0.0009	0.0009	0.0009	0.0008	0.0008	0.0009	0.0008	0.0001

表2 参数λ的E-Bayes估计

Table 2 E-Bayesian estimation of poisson distribution parameters

$\delta_E(x)$									$\Delta\delta_E(x)$
$c=0.5$	$c=1$	$c=2$	$c=3$	$c=4$	$c=5$	$c=6$	$c=7$	$c=8$	
2.8232	2.8225	2.8221	2.8197	2.8183	2.8169	2.8155	2.8142	2.8128	0.0097

表3 模拟后参数λ的Bayes估计

Table 3 Simulation of Bayesian estimation of poisson distribution parameters

α	$\delta_B(x)$					$\Delta\delta_B$
	$\beta=0.5$	$\beta=1$	$\beta=3.5$	$\beta=4.5$	$\beta=6$	
0.1	1.4312	1.4200	1.3667	1.3467	1.3178	0.1134
0.25	1.4343	1.4230	1.3697	1.3495	1.3206	0.1137
0.4	1.4373	1.4261	1.3726	1.3524	1.3234	0.1139
0.55	1.4404	1.4291	1.3755	1.3553	1.3262	0.1142
0.7	1.4435	1.4321	1.3784	1.3582	1.3290	0.1145
0.8	1.4455	1.4342	1.3804	1.3601	1.3309	0.1146
$\Delta\delta_B$	0.0143	0.0142	0.0137	0.0134	0.0131	0.0012

表4 模拟后参数λ的E-Bayes估计

Table 4 Simulation of E-Bayesian estimation of poisson distribution parameters

$\delta_E(x)$								$\Delta\delta_E$
$c=0.5$	$c=1$	$c=2$	$c=3$	$c=4$	$c=5$	$c=6$	$c=7$	
1.4037	1.3968	1.3832	1.3700	1.3571	1.3445	1.3322	1.3203	0.0734

从表3可以看出，参数λ的Bayes估计$\delta_B(x)$的偏差$\Delta\delta_B=|\delta_B(x)-\delta_0|$，$\delta_0=1.36$，$\Delta\delta_B\in[0.0422,0.0855]$。而从表4分析结果看，参数$\lambda$的E-Bayes估计$\delta_E(x)$的偏差$\Delta\delta_E=|\delta_E(x)-\delta_0|$，$\delta_0=1.36$，$\Delta\delta_B\in[0.0397,0.0437]$。由于估计中超参数对估计值有一定的影响，所以，只有在一定的条件下，参数λ的E-Bayes估计比参数λ的Bayes估计精确度更高。

综合表1到表4得出结论：

(1)参数λ的Bayes估计很稳健，横向极差最大只有0.0209(表1)，而纵向极差最大也仅有0.0009。而模拟后参数λ的Bayes估计的估计值的极差最大也只是0.1164(表3)。所以λ的Bayes估计都符合统计决策的稳健性。

(2)参数λ的E-Bayes估计的极差分别为0.0097(表2)和0.0734(表4)。从统计决策的稳健性的角度分析，参数λ的E-Bayes估计很稳健。

总之，本文给出的参数λ的Bayes估计及E-Bayes估计是合理可行的。此外在一定的条件下，参数λ的E-Bayes估计比Bayes估计稳健性更强，精确度更高，即在一定条件下，参数λ的E-Bayes估计比Bayes估计优良性更好。

参考文献:

[1] 李建波，张日全. Lenth方法对Poisson分布有效性的研究[J]. 应用概率统计，2011(3):331-336.

[2] 范洪福. 一类Poisson分布的数学模型[J]. 大学数学，2011，27(4):150-151.

[3] 王德辉，赖民，宋立新. 熵损失下Poisson分布参数倒数的估计[J]. 吉林大学学报:自然科学版，2000(4):17-22.

[4] 徐宝，于春艳，孙宪军. 一种对称损失下Poisson分布参数倒数的Bayes估计[J]. 吉林师范大学学报，2006(3):53-54.

[5] 韦莹莹，韦程东，薛婷婷. Q-对称熵损失下Poisson分布参数倒数的估计[J]. 广西师范学院学报:自然科学版，2007，24(2):33-38.

[6] 韩明. 只有一个失效数据情形失效率的E-Bayes估计[J]. 数学物理学报，2011，31A(2):577-583.

[7] 张睿. 复合LINEX损失下的参数估计[D]. 大连:大连理工大学硕士学位论文，2007.

[8] 范金城，吴可法. 统计推断导引[M]. 北京:科学出版社，2001.

[9] 韩明. 可靠度的一种新估计方法[J]. 兵工学报，2004，25(1):60-64.

[10] 韩明. 多层先验分布的构造及其应用[J]. 运筹与管理，1997，6(3):31-40.

(责任编辑:尹 闯)