文章编号:1000-5862(2016)04-0358-05

峰度与偏度系数的近似经验贝叶斯估计

章　溢[1],吕凤虎[2]
(1. 江西师范大学计算机信息工程学院,江西 南昌　330022;2. 南昌工程学院理学院,江西 南昌　330099)

摘要:建立了单样本数据的贝叶斯模型,给出了偏度系数和峰度系数的线性贝叶斯估计及近似信度估计.进而,将模型推广到多样本数据模型下,并讨论了近似信度估计的统计性质,比较了贝叶斯估计、线性贝叶斯估计及近似信度估计的均方误差.最后,给出了超参数的估计,得到了近似信度估计的经验贝叶斯估计,使该估计可直接运用于实际问题.

关键词:峰度系数;偏度系数;线性贝叶斯估计;近似信度估计;超参数;经验贝叶斯估计
中图分类号:O 211.9　　文献标志码:A　　DOI:10.16357/j.cnki.issn1000-5862.2016.04.06

0　引言

峰度系数与偏度系数是概率统计中度量随机变量密度曲线的重要特征量[1-2].关于这 2 个特征量的研究不仅在数理统计中得到广泛的关注,而且被运用到金融风险管理与决策、保险精算、时间序列预测等方面[3-7].

设随机变量 X 的分布函数为 $F(x)$,则该随机变量的峰度系数和偏度系数被定义为

$$\kappa=\frac{E(X-EX)^4}{[E(X-EX)^2]^2},\gamma=\frac{E(X-EX)^3}{[E(X-EX)^2]^{3/2}}.\quad(1)$$

注意到(1)式中的峰度系数 κ 和偏度系数 γ 涉及到分布的 3 阶矩和 4 阶矩,因此对这 2 个特征量的估计有一定的难度[8].

在对峰度系数和偏度系数进行估计过程中,一般有 2 类信息可以使用:(i) 对总体进行观测得到的样本数据信息 $\{X_1,X_2,\cdots,X_n\}$;(ii) 根据总体已有的历史资料或经验形成的先验信息.因此对它们的估计就落入了贝叶斯框架.设随机变量 X 的分布为 $F(x,\theta)$,而 θ 本身也是随机变量,其分布称为先验分布,记为 $\pi(\theta)$.在贝叶斯框架下,样本 $X_1,X_2,\cdots,X_n$ 可以看成在先验参数 θ 给定下的 n 次独立观测.此时峰度系数与偏度系数都与先验参数 θ 有关,记为 $\kappa(\theta)$ 和 $\gamma(\theta)$.根据贝叶斯定理,所有的统计推断都在后验分布上进行.

根据贝叶斯决策准则,若取平方损失函数,则峰度系数和偏度系数的最优估计分别为它们各自的后验均值.但是,由于后验均值不仅依赖于样本分布的具体信息,而且还依赖于先验分布的具体形式.而这些信息,特别是先验分布的具体形式,在实际问题中很难获取,在许多情况下带有主观因素[9-10].此时,解决这个问题的办法之一就是建立线性贝叶斯模型,将峰度系数和偏度系数的估计限定在某种线性函数类中,得到这种函数估计类中的最优估计.这种方法来源于精算学中的信度理论[11-14].若有多样本数据,则可结合经验贝叶斯方法对估计中的超参数进行估计,最后得到峰度系数和偏度系数的经验贝叶斯估计[15-17],这些估计可以直接运用于实际问题.

本文首先建立单样本数据的贝叶斯模型,研究峰度系数和偏度系数的贝叶斯估计及近似信度估计,然后将该结论推广到多样本数据的贝叶斯模型,并利用经验贝叶斯方法估计超参数,进而得到峰度系数和偏度系数的经验贝叶斯估计.

1　单样本数据下的贝叶斯模型

假设随机变量 X 的分布函数为 $F(x,\theta)$,而 θ 为随机变量,具有某个先验分布 $\pi(\theta)$.设在给定参数 θ 下对总体 X 进行了 n 次观测得到样本 $X_1,X_2,\cdots,X_n$.若记 $\mu_k(\theta)=E(X^k|\theta)$,则在这种单样本数据

收稿日期:2016-02-19
基金项目:国家自然科学基金(71361015),教育部人文社会科学基金(15YJC910010)和江西师范大学研究生创新基金(2014010654)资助项目.
作者简介:章　溢(1985-),女,江西南昌人,讲师,主要从事统计学与精算学方面的研究.

下,总体的峰度系数和偏度系数分别为

$$\kappa(\theta) = E\{[X-E(X|\theta)]^4|\theta\}/\{E[(X-E(X|\theta))^2|\theta]\}^2 = (\mu_4(\theta) - 4\mu_4(\theta)\mu_1(\theta) + 6\mu_2(\theta)[\mu_1(\theta)]^2 - 3[\mu_1(\theta)]^3)/[\mu_2(\theta) - (\mu_1(\theta))^2]^2,$$

$$\gamma(\theta) = E\{[X-E(X|\theta)]^3|\theta\}/\{E[(X-E(X|\theta))^2|\theta]\}^{3/2} = (\mu_3(\theta) - 3\mu_2(\theta)\mu_1(\theta) + 2[\mu_1(\theta)]^3)/[\mu_2(\theta) - (\mu_1(\theta))^2]^{3/2}.$$

为了对总体的峰度系数和偏度系数进行合适的估计,定义样本峰度和样本偏度为

$$\kappa_n = n\sum_{i=1}^{n}(X_i - \bar{X})^4/[\sum_{i=1}^{n}(X_i - X)^2]^2,$$

$$\gamma_n = \sqrt{n}\sum_{i=1}^{n}(X_i - \bar{X})^3/[\sum_{i=1}^{n}(X_i - \bar{X})^2]^{3/2}.$$

且引入记号

$\gamma_0 = E[\gamma(\theta)]$, $\gamma_n(\theta) = E(\gamma_n|\theta)$, $\gamma = E[\gamma_n(\theta)]$, $\delta_1 = \mathrm{Cov}(\gamma_n(\theta), \gamma(\theta))$, $\lambda_1 = E[\mathrm{Var}(\gamma_n|\theta)]$, $\rho_1 = \mathrm{Var}(\gamma_n(\theta))$.

类似地,对峰度系数,引入记号

$\kappa_0 = E[\kappa(\theta)]$, $\kappa_n(\theta) = E(\kappa_n|\theta)$, $\kappa = E[\kappa_n(\theta)]$, $\delta_2 = \mathrm{Cov}(\kappa_n(\theta), \kappa(\theta))$, $\lambda_2 = E[\mathrm{Var}(\kappa_n|\theta)]$, $\rho_2 = \mathrm{Var}(\kappa_n(\theta))$.

目标是利用样本信息 $X_1, X_2, \cdots, X_n$ 和先验信息 $\pi(\theta)$ 对峰度系数 $\kappa(\theta)$ 和偏度系数 $\gamma(\theta)$ 进行合适的估计. 若不考虑先验信息,则样本峰度 κ_n 和样本偏度 γ_n 分别是峰度系数 $\kappa(\theta)$ 和偏度系数 $\gamma(\theta)$ 的合适估计,且当 $n \to \infty$ 时有 $\kappa_n \to \kappa$ a. s., $\gamma_n \to \gamma$ a. s. [18]. 若引入先验信息,则得到的估计应充分利用样本信息和先验信息. 根据贝叶斯定理,容易得到如下结论,证明思路参见文献[18].

定理 1　在平方损失函数下,$\kappa(\theta)$ 与 $\gamma(\theta)$ 的最优估计分别为其后验均值 $\widehat{\kappa(\theta)} = E[\kappa(\theta)|X_1, X_2\cdots, X_n]$ 和 $\widehat{\gamma(\theta)} = E[\gamma(\theta)|X_1, X_2, \cdots, X_n]$,即

$$E(\kappa(\theta)|\underline{X_n}) = \arg\min_{g\in M} E\{[\kappa(\theta) - g)^2]\},$$

$$E(\gamma(\theta)|\underline{X_n}) = \arg\min_{g\in M} E\{[\gamma(\theta) - g)^2]\},$$

其中 $M = \{g: g$ 是 $X_1, X_2, \cdots, X_n$ 的可测函数$\}$.

称后验均值 $E(\kappa(\theta)|\underline{X_n})$ 和 $E(\gamma(\theta)|\underline{X_n})$ 分别为峰度系数 $\kappa(\theta)$ 与偏度系数 $\gamma(\theta)$ 的贝叶斯估计,记为 $\widehat{\kappa(\theta)}^B$ 及 $\widehat{\gamma(\theta)}^B$. 从上面的分析及结论可看出峰度系数 $\kappa(\theta)$ 和偏度系数 $\gamma(\theta)$ 的形式有很高的相似度,估计的方法也基本相同,本文仅对峰度系数 $\kappa(\theta)$ 进行估计(峰度系数涉及到 4 阶矩,而偏度系数只涉及到 3 阶矩),类似的结果可适用于偏度系数 $\gamma(\theta)$.

显然,后验均值

$$\widehat{\kappa(\theta)}^B = E(\kappa(\theta)|\underline{X_n}) = \int_{-\infty}^{+\infty}\kappa(\theta)\pi(\theta)\prod_{i=1}^{n}f(x_i|\theta)\,\mathrm{d}\theta \Big/ \int_{-\infty}^{+\infty}\pi(\theta)\prod_{i=1}^{n}f(x_i|\theta)\,\mathrm{d}\theta$$

不仅依赖于样本的具体分布,而且依赖于先验分布的具体形式. 而这些信息,特别是先验分布的具体形式,在实际中很难获取,并且即使对先验分布有合理的假设,结合样本分布也很难得到 $\widehat{\kappa(\theta)}^B$ 的显示表达式.

为此,求解最小化问题

$$\min_{a,b\in\mathbf{R}} E\{[\kappa(\theta) - a - b\kappa_n)^2]\}, \tag{2}$$

得到定理 2.

定理 2　在贝叶斯模型下,求解最小化问题(2)式,得到峰度系数 $\kappa(\theta)$ 的最优线性贝叶斯估计

$$\widehat{\kappa(\theta)}^{LB} = \kappa_0 + \delta_2(\kappa_n - \kappa)/(\lambda_2 + \rho_2).$$

证　记 $\psi = E\{[\kappa(\theta) - a - b\kappa_n)^2]\}$. 由于 $E[\kappa(\theta)] = \kappa_0$, $E[\kappa_n(\theta)] = \kappa$,因此对 ψ 关于 a 求导,并令导数为 0,有 $E[\kappa(\theta)] - a - bE(\kappa_n) = 0$,即

$$a = \kappa_0 - b\kappa. \tag{3}$$

将(3)式代入 ψ,得到 $\psi = E\{[(\kappa(\theta) - E(\kappa(\theta))) - b(\kappa_n - E(\kappa_n))]^2\}$,再对 b 求导并令导数为 0,得到

$$b = \mathrm{Cov}(\kappa(\theta), \kappa_n(\theta))/\mathrm{Cov}(\kappa_n, \kappa_n) = \delta_2/(\lambda_2 + \rho_2),$$

所以 $\widehat{\kappa(\theta)}^{LB} = \hat{a} + \hat{b}\kappa_n = \kappa_0 + \delta_2(\kappa_n - \kappa)/(\lambda_2 + \rho_2)$.

注 1　同理求解 $\min_{c,d\in\mathbf{R}} E\{[\gamma(\theta) - c - d\gamma_n)^2]\}$ 可以得到偏度系数 $\gamma(\theta)$ 的线性贝叶斯估计为 $\widehat{\gamma(\theta)}^{LB} = \gamma_0 + \delta_1(\gamma_n - \gamma)/(\lambda_1 + \rho_1)$.

在定理 2 与注 1 中得到的线性贝叶斯估计不依赖于样本分布和先验分布的具体形式. 而仅仅依赖于一些超参数,这些参数在多样本数据下是可以估计的.

在定理 2 中得到的线性贝叶斯估计并不能表示为"信度"加权的形式,一般地,"信度"加权形式需要条件 $\gamma_n(\theta) = \gamma(\theta)$ 或 $\kappa_n(\theta) = \kappa(\theta)$.

定理 3　若 $\kappa_n(\theta) = \kappa(\theta)$,则得到 $\kappa(\theta)$ 的近似信度估计为

$$\widehat{\kappa(\theta)}^{MB} = Z_2\kappa_n + (1 - Z_2)\kappa_0,$$

其中 $Z_2 = \rho_2/(\rho_2 + \lambda_2)$.

证　在 $\kappa_n(\theta) = \kappa(\theta)$ 条件下,有 $\kappa = \kappa_0$ 以及 $\rho_2 = \delta_2 = \mathrm{Var}(\kappa(\theta))$,因此有

$$b = \mathrm{Cov}(\kappa(\theta), \kappa_n(\theta))/\mathrm{Cov}(\kappa_n, \kappa_n) = \rho_2/(\lambda_2 + \rho_2) = Z_2,$$

则 $\kappa(\theta)$ 的近似信度估计为

$\widehat{\kappa(\theta)}^{MB} = \hat{a} + \hat{b}\gamma_n = Z_2\kappa_n + (1 - Z_2)\kappa_0$.

注 2 类似地,当 $\gamma_n(\theta) = \gamma(\theta)$ 时,得到 $\gamma(\theta)$ 的近似信度估计为 $\widehat{\gamma(\theta)}^{MB} = Z_1\gamma_n + (1 - Z_1)\gamma_0$.

2 多样本数据下的近似信度估计

已经得到偏度系数 $\gamma(\theta)$ 和峰度系数 $\kappa(\theta)$ 的贝叶斯估计,线性贝叶斯估计及近似信度估计. 然而,这些估计仍然包含许多未知参数,在贝叶斯统计中称这些参数为超参数. 为了估计这些超参数,需要有多个 θ 值下的样本信息.

设 θ 有 K 个观测值 $\theta_1,\theta_2,\cdots,\theta_K$,在 θ_i 下有容量为 n 的样本 $X_{i1},X_{i2},\cdots,X_{in}$. 即假设 θ_i 给定条件下 $X_{i1},X_{i2},\cdots,X_{in}$ 相互独立并服从相同的分布 $F(x,\theta_i)$,而 $\theta_1,\theta_2,\cdots,\theta_K$ 相互独立且服从相同的先验分布 $\pi(\theta)$,记 $E(\kappa_{in}|\theta_i) = \kappa_n(\theta_i)$,$E(\gamma_{in}|\theta_i) = \gamma_n(\theta_i)$.

对每个 i,有样本峰度和样本偏度

$$\kappa_{in} = \frac{n\sum_{j=1}^{n}(X_{ij} - \overline{X_i})^4}{[\sum_{j=1}^{n}(X_{ij} - \overline{X_i})^2]^2}, \gamma_{in} = \frac{\sqrt{n}\sum_{j=1}^{n}(X_{ij} - \overline{X_i})^3}{[\sum_{j=1}^{n}(X_{ij} - \overline{X_i})^2]^{3/2}},$$

注意到 $\{\theta_i,X_{i1},X_{i2},\cdots,X_{in}\}$ 对 $i = 1,2,\cdots,K$ 相互独立,则类似于定理 2,可以得到 $\kappa(\theta_i)$ 的线性贝叶斯估计及近似信度估计.

定理 4 求解最优化问题

$$\min_{b_i\in\mathbf{R}} E[\kappa(\theta_i) - b_0 - \sum_{i=1}^{k} b_i\kappa_{in})^2]$$

可得到 $\kappa(\theta_i)$ 的线性贝叶斯估计为

$$\widehat{\kappa(\theta_i)}^{LB} = \kappa_0 + \delta_2(\kappa_{in} - \kappa)/(\lambda_2 + \rho_2).$$

进一步地,若 $\kappa_n(\theta) = \kappa(\theta)$,则 $\kappa(\theta_i)$ 的近似信度估计为

$$\widehat{\kappa(\theta_i)}^{MB} = Z_2\kappa_{in} + (1 - Z_2)\kappa_{i0}.$$

注3 求解 $\min_{b_i\in\mathbf{R}} E[\gamma(\theta_i) - b_0 - \sum_{i=1}^{k} b_i\gamma_{in})^2]$ 可以得到 $\gamma(\theta_i)$ 的线性贝叶斯估计为 $\widehat{\gamma(\theta_i)}^{LB} = \gamma_0 + \delta_1(\gamma_{in} - \gamma)/(\lambda_1 + \rho_1)$. 进一步地,若 $\gamma_n(\theta) = \gamma(\theta)$,则 $\gamma(\theta_i)$ 的近似信度估计为 $\widehat{\gamma(\theta_i)}^{MB} = Z_1\gamma_{in} + (1 - Z_1)\gamma_{i0}$.

显然,近似信度估计 $\widehat{\kappa(\theta_i)}^{MB}$ 比线性贝叶斯估计 $\widehat{\kappa(\theta_i)}^{LB}$ 有更少的结构参数,并且形式较为简单,容易理解和使用. 因此,讨论近似信度估计 $\widehat{\kappa(\theta_i)}^{MB}$ 的统计性质.

定理 5 当 $\kappa_n(\theta_i) = \kappa(\theta_i)$ 时,有 $E[\widehat{\kappa(\theta_i)}^{MB}] = E[\kappa(\theta_i)] = \kappa_0$. 进一步地,若存在随机变量 W 使得 $|\kappa_n(\theta_i)| \leqslant W$,且 $EW < \infty$,则近似信度估计 $\widehat{\kappa(\theta_i)}^{MB}$ 是 $\kappa(\theta_i)$ 的相合估计,这里 $i = 1,2,\cdots,K$.

证 当 $\kappa_n(\theta_i) = \kappa(\theta_i)$ 时,容易得到 $E[\widehat{\kappa(\theta_i)}^{MB}] = Z_2E(\kappa_{in}) + (1 - Z_2)\kappa_0 = \kappa_0$. 另一方面,由于 $|\kappa_n(\theta_i)| \leqslant W$,且 $\kappa_{in} \to \kappa(\theta_i)$ a. s. . 由控制收敛定理知 $\kappa_n(\theta_i) = E(\kappa_{in}|\theta_i) \to \kappa(\theta_i)$ a. s. ,且 $E(\kappa_{in}^2|\theta_i) \to [\kappa(\theta_i)]^2$ a. s. ,则有

$$\lambda_2 = E[\mathrm{Var}(\kappa_{in}|\theta_i)] = E[E(\kappa_{in}^2|\theta_i) - (E(\kappa_{in}|\theta_i))^2] \to 0 \text{ a. s. },$$

因此 $Z_2 = \rho_2/(\rho_2 + \lambda_2) \to 1$,则得到

$$\widehat{\kappa(\theta_i)}^{MB} = Z_2\kappa_{in} + (1 - Z_2)\kappa_0 \to \kappa(\theta_i) \text{ a. s. }.$$

注 4 类似地,若 $\gamma_n(\theta_i) = \gamma(\theta_i)$ 时,有 $E[\widehat{\gamma(\theta_i)}^{MB}] = E[\gamma(\theta_i)] = \gamma_0$. 进一步地,若存在随机变量 V 使得 $|\gamma_n(\theta_i)| \leqslant V$,且 $EV < \infty$,则近似信度估计 $\widehat{\gamma(\theta_i)}^{MB}$ 是 $\gamma(\theta_i)$ 的相合估计,其中 $i = 1,2,\cdots,K$.

从上面的分析中,得到了峰度系数 $\kappa(\theta_i)$ 的贝叶斯估计 $\widehat{\kappa(\theta_i)}^{B} = E(\kappa(\theta_i)|X_{i1},X_{i2},\cdots,X_{in})$,线性贝叶斯估计 $\widehat{\kappa(\theta_i)}^{LB}$ 以及近似信度估计 $\widehat{\kappa(\theta_i)}^{MB}$. 然而,这些估计的优劣如何?下面将比较这些估计的均方误差.

定理 6 估计 $\widehat{\kappa(\theta_i)}^{B}$,$\widehat{\kappa(\theta_i)}^{LB}$ 和 $\widehat{\kappa(\theta_i)}^{MB}$ 关于 $\kappa(\theta_i)$ 的均方误差大小关系为 $MSE(\widehat{\kappa(\theta_i)}^{B}) \leqslant MSE(\widehat{\kappa(\theta_i)}^{LB}) \leqslant MSE(\widehat{\kappa(\theta_i)}^{MB})$.

证 根据贝叶斯估计的定义,$\widehat{\kappa(\theta_i)}^{B}$ 是使贝叶斯风险达到最小的估计,即

$$MSE(\widehat{\kappa(\theta_i)}^{B}) = \min_{\widehat{\kappa(\theta_i)}\in N} E[\widehat{\kappa(\theta_i)} - \kappa(\theta_i))^2],$$

其中 $N = \{g: g \text{ 是 } X_{i1},X_{i2},\cdots,X_{in} \text{ 的可测函数},i = 1,2,\cdots,K\}$,因此有

$$MSE(\widehat{\kappa(\theta_i)}^{B}) \leqslant MSE(\widehat{\kappa(\theta_i)}^{LB}).$$

同理,根据线性贝叶斯估计的定义,$\widehat{\kappa(\theta_i)}^{LB}$ 是估计类 $H = \{b_0 + \sum_{i=1}^{k} b_i\kappa_{in}\}$ 中均方误差达到最小的估计,而 $\widehat{\kappa(\theta_i)}^{MB} \in H$,因此 $MSE(\widehat{\kappa(\theta_i)}^{LB}) \leqslant MSE(\widehat{\kappa(\theta_i)}^{MB})$.

注5 同理可证 $MSE(\widehat{\gamma(\theta_i)}^{B}) \leqslant MSE(\widehat{\gamma(\theta_i)}^{LB}) \leqslant MSE(\widehat{\gamma(\theta_i)}^{MB})$,$i = 1,2,\cdots,K$.

3　峰度系数和偏度系数的经验贝叶斯估计

在实际运用中,近似信度估计不仅形式简单,而且不依赖于具体的先验分布,因此它是最方便使用的估计. 但是,近似信度估计仍然包含一些未知的超参数,如 γ_0,κ_0,λ_i 和 $\rho_i(i=1,2)$,这些超参数需要根据多样本数据来估计.

首先,考虑 $\lambda_1 = E[\mathrm{Var}(\gamma_n|\theta)]$ 和 $\lambda_2 = E[\mathrm{Var}(\kappa_n|\theta)]$ 的估计,可采用重抽样技术[9]. 从样本 $X_{i1},X_{i2},\cdots,X_{in}$ 的经验分布 $F_{in}(x) = \frac{1}{n}\sum_{j=1}^{n} I(X_{ij} \leqslant x)$ 中独立重复抽取 B 次容量为 n 的样本,得到样本 $(X_{i1}^b,X_{i2}^b,\cdots,X_{in}^b)$,$b=1,2,\cdots,B$. 因此 λ_1 的 1 个重抽样估计为 $\hat{\lambda}_1^* = \frac{1}{K}\sum_{i=1}^{K} S_i^2$,其中 $S_i^2 = \frac{1}{B-1}\cdot\sum_{b=1}^{B}\left(\gamma_{in}^b - \frac{1}{B}\sum_{m=1}^{B}\gamma_{in}^m\right)^2$. 同理,$\lambda_2$ 的 1 个重抽样估计为 $\hat{\lambda}_2^* = \frac{1}{K}\sum_{i=1}^{K} T_i^2$,其中

$$T_i^2 = \frac{1}{B-1}\sum_{b=1}^{B}\left(\kappa_{in}^b - \frac{1}{B}\sum_{m=1}^{B}\kappa_{in}^m\right)^2.$$

进而考虑 γ_0,κ_0,ρ_1,ρ_2 的估计. 注意到 $\kappa_0 = E[\kappa(\theta_i)]$,$\rho_2 = \mathrm{Var}(\kappa(\theta_i))$,由矩估计方法容易得到 κ_0 和 ρ_2 的矩估计为

$$\hat{\kappa}_0 = \frac{1}{K}\sum_{i=1}^{K}\kappa_{in},\hat{\rho}_2 = \frac{1}{K-1}\sum_{i=1}^{K}\left(\kappa_{in} - \overline{\kappa_n}\right)^2 - \hat{\lambda}_2,$$

这里 $\overline{\kappa_n} = \frac{1}{K}\sum_{i=1}^{K}\kappa_{in}$.

定理 7　当条件 $\kappa_n(\theta_i) = \kappa(\theta_i)$ 时,估计 $\hat{\kappa}_0$ 与 $\hat{\rho}_2$ 关于超参数 κ_0 与 ρ_2 是无偏的,即 $E[\hat{\kappa}_0] = \kappa_0$,$E[\hat{\rho}_2] = \rho_2$.

证　根据已知条件得

$$E[\hat{\kappa}_0] = \frac{1}{K}\sum_{i=1}^{K}E(\kappa_{in}) = \frac{1}{K}\sum_{i=1}^{K}E[E(\kappa_{in}|\theta_i)] = E(\kappa_n(\theta_i)) = E(\kappa(\theta_i)) = \kappa_0.$$

另一方面,由于 $\hat{\rho}_2 = \frac{K}{K-1}\left(\frac{1}{K}\sum_{i=1}^{K}\kappa_{in}^2 - \overline{\kappa_n^2}\right)$,在条件 $\kappa_n(\theta_i) = \kappa(\theta_i)$ 下,κ_{in} 对 $i=1,2,\cdots,K$ 独立同分布,且 $E(\kappa_{in}) = \kappa_0$,$\mathrm{Var}(\kappa_{in}) = \mathrm{Var}[E(\kappa_{in}|\theta_i)] + E[\mathrm{Var}(\kappa_{in}|\theta_i)] = \lambda_2 + \rho_2$,因此由矩估计的无偏性得

$$E[\hat{\rho}_2] = E\left[\frac{1}{K-1}\sum_{i=1}^{K}\left(\kappa_{in} - \overline{\kappa_n}\right)^2\right] - E[\hat{\lambda}_2] = \mathrm{Var}(\kappa_{in}) - \lambda_2 = \rho_2.$$

注 6　同理可得到 γ_0 和 ρ_1 的矩估计为 $\hat{\gamma}_0 = \frac{1}{K}\sum_{i=1}^{K}\gamma_{in},\hat{\rho}_1 = \frac{1}{K-1}\sum_{i=1}^{K}\left(\gamma_{in} - \overline{\gamma_n}\right)^2$,其中 $\overline{\gamma_n} = \frac{1}{K}\sum_{i=1}^{K}{}^{k}\gamma_{in}$. 且可以证明当 $\kappa_n(\theta_i) = \kappa(\theta_i)$ 时,估计 $\hat{\gamma}_0$ 与 $\hat{\rho}_1$ 关于 γ_0 和 ρ_1 是无偏的.

最后将超参数 $\kappa_0,\gamma_0,\lambda_i,\rho_i$ 的估计 $\hat{\kappa}_0,\hat{\gamma}_0,\hat{\lambda}_i,\hat{\rho}_i$ 分别代入,得到的估计记为 $\widehat{\gamma(\theta_i)}^{MB} = \hat{Z}_1\gamma_{in} + (1-\hat{Z}_1)\hat{\gamma}_0$,$\widehat{\kappa(\theta_i)}^{MB} = \hat{Z}_2\kappa_{in} + (1-\hat{Z}_2)\hat{\kappa}_0$,其中 $\hat{Z}_1 = \hat{\rho}_1/(\hat{\rho}_1 + \hat{\lambda}_1^*)$,$\hat{Z}_2 = \hat{\rho}_2/(\hat{\rho}_2 + \hat{\lambda}_2^*)$. 称之为偏度系数和峰度系数的经验贝叶斯估计,在实际问题中可以直接运用.

4　结论

本文讨论了随机变量的峰度系数及偏度系数的估计问题. 首先建立了单样本数据的贝叶斯模型,得到了峰度系数及偏度系数的贝叶斯估计. 进而,利用信度理论的线性化方法得到了线性贝叶斯估计及近似信度估计. 显然,近似信度估计有较少的结构参数及较简单的形式,其大样本性质也能保证,因此在实际问题中更方便使用. 然后,将模型推广到多样本数据,得到了峰度系数与偏度系数的近似信度估计,并给出了超参数的估计,证明了这些估计的统计性质. 最后,得到了峰度系数与偏度系数的经验贝叶斯估计. 本文的方法结合了贝叶斯统计方法、信度理论方法以及经验贝叶斯方法,得到的结论可以适用于保险精算、生存分析以及其他的统计领域. 注意到,在近似信度估计中,一般要求满足条件 $\kappa_n(\theta_i) = \kappa(\theta_i)$. 在实际使用中,由于 $\kappa_n(\theta_i) = E\left[n\sum_{j=1}^{n}(X_{ij} - \overline{X_i})^4/\left(\sum_{j=1}^{n}(X_{ij} - \overline{X_i})^2\right)^2 \mid \theta_i\right]$ 有非常复杂的形式,与 $\kappa(\theta_i) = E\{[X_{ij} - E(X_{ij}|\theta_i))|\theta_i]^4/(E(X_{ij} - E(X_{ij}|\theta)|\theta_i)^2)^2$ 在一般情况下是不相等的. 然而,由于 $\kappa_n(\theta_i) \to \kappa(\theta_i)$ a. s. 因此,当样本容量较大时,两者近似相等.

5　参考文献

[1]　王学民. 偏度和峰度概念的认识误区[J]. 统计与决策,2008(12):145-146.

[2]　邵建平,邓兆卉. 分配公平性的分布偏度与峰度描述研究[J]. 统计与决策,2008(3):144-147

[3] 余婧. 均值-方差-近似偏度投资组合模型和实证分析[J]. 运筹学学报,2010,14(1):106-114.
[4] 傅俊辉,张卫国,陆倩,等. 考虑偏度风险和峰度风险的非线性期货套期保值模型[J]. 系统工程,2009,27(10):44-48.
[5] 王鹏,王建琼,魏宇. 自回归条件方差-偏度-峰度:一个新的模型[J]. 管理科学学报,2009,12(5):121-129.
[6] Conrad J,Dittmar R F,Ghysels E. Ex ante skewness and expected stock returns[J]. The Journal of Finance,2013,68(1):85-124.
[7] Grigoletto M,Lisi F. Looking for skewness in financial time series[J]. The Econometrics Journal,2009,12(2):310-323.
[8] Yevjevich V,Obeysekera J T B. Estimation of skewness of hydrologic variables[J]. Water Resources Research,1984,20(7):935-943.
[9] Huang Y,Getachew. A Bayesian approach to joint mixed-effects models with a skew-normal distribution and measurement errors in covariates[J]. Biometrics,2011,67(1):260-269.
[10] Cabras S,Racugno W,Castellanos M E,et al. A matching prior for the shape parameter of the skew-normal distribution[J]. Scandinavian Journal of Statistics,2012,39(2):236-247.
[11] 温利民,邹思思,吕凤虎. 偏度系数和峰度系数的信度估计[J]. 统计与决策,2015(3):24-25.
[12] Buhlmann H,Gisler A. A course in credibility theory and its applications[M]. Netherlands: Springer,2005.
[13] 郑丹,章溢,温利民. 具有时间变化效应的信度模型[J]. 江西师范大学学报:自然科学版,2012,36(3):249-252.
[14] 方婧,章溢,温利民. 聚合风险模型下的信度估计[J]. 江西师范大学学报:自然科学版,2012,36(6):607-611.
[15] Robbins H. An empirical Bayes approach to statistics[C]//Proceedings of the Third Berkeley Symposium on Mathematics,Statistics and Probability,1956:157-164.
[16] Robbins H. The empirical Bayes approach to statistical decision problems[J]. Annals of Mathematics Statistics,1964,35(1):1-20.
[17] 李乃医. 随机删失下伽玛分布族参数的经验 Bayes 双边检验[J]. 系统科学与数学,2011,31(4):458-465.
[18] Ferguson T S. A course in large-sample theory[M]. New York: Chapman and Hall,1996.
[19] Efron B,Tibshirani R. An introduction to the bootstrap[M]. New York: Chapman and Hall,1993.

The Approximate Empirical Bayesian Estimation of Kurtosis and Skewness Coefficient

ZHANG Yi[1],LYU Fenghu[2]
(1. College of Computer Information Engineering,Jiangxi Normal University,Nanchang Jiangxi 330022,China;
2. College of Science,Nanchang Institute of Technology,Nanchang Jiangxi 330099,China)

Abstract: A Bayesian model of single sample data is established,and the Bayesian estimation,linear Bayesian estimation and approximate credibility estimation of skewness and kurtosis coefficient are given. Furthermore,the model is extended to multitude data model. In this model,the statistical properties of approximate credibility estimation are discussed,the mean square errors of Bayesian estimation,linear approximation and approximate credibility estimation are compared. Finally,the estimation of supper-parameters are given,thus the empirical Bayes estimation of approximate credibility estimation is derived,and it can be directly applied to practice.

Key words: kurtosis coefficient; skewness coefficient; linear Bayesian estimation; approximate credibility estimation; supper-parameter; empirical Bayes estimation

(责任编辑:曾剑锋)