贝叶斯方法在审计子总体稀少项目估计中的应用

胡桂华(博士)

(广西财经学院　南宁　530003)

【摘要】在抽样审计过程中有时需要估计一个相对于整个总体来说单位数非常稀少的子总体的均值、总值或是单位数目。这时,从总体所抽取的样本中可能只包含数目很少的子总体单位,甚至可能一个子总体的单位都没有。这时,传统的子总体估计技术便遇到了困难。本文通过对Y. Liu等人的研究进行解读和再加工,介绍了他们所设计的方法技术。

【关键词】子总体估计　贝叶斯统计　抽样审计

一、引言

假设审计对象是由N张发票组成的一个总体。每张发票的票面金额是已知的,记作x_i(i=1,2,…,N)。发票总体分为两个子总体:错误发票(稀少项目,后同)子总体和非错误发票子总体。我们的审计目标是估计错误发票子总体(记为"j域")错误发票票面金额的总值。

传统的做法是应用子总体估计技术。即首先定义第i(i=1,2,…,N)总体单位的下列标志值:

$$y_i=\begin{cases}x_i & (若发票为错误发票)\\ 0 & (若发票为非错误发票)\end{cases}$$

于是,用来自整个总体的概率样本来估计y_i的总体总值,也就得到了对错误发票子总体(j域)总值$X_{(j)}$的估计。如果是样本量为n的简单随机样本,则估计量为:

$$\hat{X}_{(j)}=\hat{Y}=N\frac{1}{n}\sum_{i=1}^{n}y_i=N\frac{1}{n}\sum_{i=1}^{n(j)}x_{(j)k}$$

从上式可看出,估计的精度除了与样本量n有关以外,还与样本中错误发票的数目n(j)有关。

事实上,在大多数被审计的发票总体中,错误发票张数所占的比例很小。在这种情况下,样本中可能会只有很少的几张错误发票,甚至一张也没有,这样,用上面的方法来估计$X_{(j)}$就遇到了困难。

Y. Liu等(2005)设计一种巧妙解决此类问题的方法。在这个方法中,首先应用贝叶斯公式来估计总体中错误发票的张数,然后在这个估计的基础上把样本内外有关错误发票的信息综合起来得到错误发票票面金额总体总值的估计。

二、总体中错误发票张数的估计

设发票总体中发票的总张数为N,第i张发票票面金额记为x_i(i=1,2,…,N);总体中,错误发票的张数为M(未知),第k张错误发票票面金额记为$x_k^{\otimes}$(k=1,2,…,M);总体中,非错误发票的张数为N−M(未知),第l张非错误发票票面金额记做$x_l^{\oplus}$(l=1,2,…,N−M)。

Y. Liu等(2005)用贝叶斯公式来估计M。在贝叶斯方法中,M被看做随机变量,理论上来看,它可以取从0到N这样N+1个整数值,至于它的概率分布(先验分布)Pr(M),在没有任何可用信息的情况下,假定M的N+1个整数值全都等可能。为了收集关于M的信息,从发票总体中不放还地抽取简单随机样本,样本量为n,对所抽得的样本观察结果为:样本中错误发票的张数为m(m是一个实际观察到的确定的数值,$M\in[0,n]$)。由于样本中出现了m张错误发票,根据这个信息可知,总体中错误发票的数目至少应该有m张,于是,M的先验分布Pr(M)应当修改为:M取0到m−1诸整数值的概率都是0,M取m到N诸整数值的概率都是1/(N−m)。另外,M的似然函数Pr(m|M)(总体错误发票的张数为M时样本中出现m张错误发票的概率)应当用超几何分布写出。这样,便可以用贝叶斯公式写出M的后验分布Pr(M|m)为:

$$\Pr(M|m)=\frac{\frac{1}{N-m}\Pr(m|M)}{\sum_{M=m}^{N}\frac{1}{N-m}\Pr(m|M)}=\frac{\Pr(m|M)}{\sum_{M=m}^{N}\Pr(m|M)}\qquad(1)$$

其中:

$$\Pr(m|M)=\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}\qquad(2)$$

一旦求得M的后验分布,便可以把这个分布的位置特征数(数学期望值或是众数)作为M的一个估计($\hat{M}$)。下面分别写出M的后验分布数学期望和后验分布众数。

M的后验分布数学期望是在已知容量为n的样本中观察到的错误发票数目为m的条件下的期望值。即:

$$E(M|m)=\sum_{M=m}^{N}M\Pr(M|m)=\sum_{M=m}^{N}M\frac{\Pr(m|M)}{\sum_{M=m}^{N}\Pr(m|M)}$$

$$=\sum_{M=m}^{N}M\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}\Bigg/\sum_{M=m}^{N}\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}\qquad(3)$$

在式(3)中,依次代入M=m,M=m+1,…,M=m+N各个数值,然后求和算出结果。

M的后验分布数学期望众数是后验分布函数的极大值。

考虑：

$$Pr(M|m)=\frac{Pr(m|M)}{\sum_{M=m}^{N}Pr(m|M)}=\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}\Bigg/\sum_{M=m}^{N}\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}=\frac{C_M^m C_{N-M}^{n-m}}{C_N^n} \qquad (4)$$

用二项分布来描述式(4)的超几何分布：

$$Pr(M|m)=C_n^m(M/N)^m[(N-M)/N]^{n-m} \qquad (5)$$

对式(5)求解关于变量M的极大值点，在$M\neq0, M\neq N$的条件下可得：

$$\hat{M}_o=N(m/n) \qquad (6)$$

这就是M的后验分布众数。

Y. Liu等(2005)指出，当样本中观察到的错误发票数目m非常小时，众数可能不是一个好的估计值。特别是当样本中没有错误发票时，无法用众数充当估计量。总之，比较M的后验数学期望和后验众数，选用后验期望值$E(M|a)$充当估计量较为适宜。当错误发票为总体中的稀少项目时，M的条件概率分布（以观察到m为条件）呈正偏，此时，平均数估计量比众数估计量保守一些。因此，当希望获得较为保守的估计量时，应考虑用后验数学期望样本中的平均数作为M的估计量。

三、总体中错误发票票面总金额的点估计

为了估计错误发票票面金额的总体总值，Y. Liu、M. Batcher、Scheuren F(2005)给出了一个很有用的假设：在一定的条件下，可以用所有发票（包括错误的和非错误的）票面金额的总体均值$\bar{X}$代替错误发票票面金额总体均值$\bar{X}^{\otimes}$。下面我们来说明这个假设。假定根据$x_i(i=1,2,\cdots,N)$对发票总体做了组距很小的分层，并且假定各层中错误发票的数目（张数）差不多。这时，次数分布直方图近似成为连续型随机变量的概率密度函数曲线。由于各层中错误发票的数目（张数）差不多，因此错误发票面值$x^{\otimes}$服从均匀分布，令$x^{\otimes}$在区间[a，b]上取值，于是，其密度函数为：

$$p(x^{\otimes})=\begin{cases}1/(b-a) & (a\leq x^{\otimes}\leq b)\\ 0 & (x^{\otimes}<a, b<x^{\otimes})\end{cases} \qquad (7)$$

由式(7)给出的均匀分布随机变量$x^{\otimes}$的期望值是$E(x^{\otimes})=(a+b)/2$。另外，把非错误发票面值$x^{\oplus}$概率分布的密度函数记作$f(x^{\oplus})$。在上述假定下，发票总体（含错误发票和非错误发票）票面金额x的平均值是：

$$E(x)=\int_a^b x\frac{Mp(x^{\otimes})+(N-M)f(x^{\oplus})}{M+(N-M)}dx=\frac{Mp(x^{\otimes})+(N-M)f(x^{\oplus})}{N}\int_a^b\frac{1}{2}dx^2=\frac{Mp(x^{\otimes})+(N-M)f(x^{\oplus})}{N}\ \frac{(b-a)(b+a)}{2} \qquad (8)$$

当$f(x^{\oplus})$也在区间[a，b]上均匀分布时，式(8)的值为$(a+b)/2$。可见，在错误发票票面金额服从均匀分布的假定下，若非错误发票票面金额的分布也接近相同区间上的均匀分布，可以用整个总体的平均数代替错误发票总体平均数，非错误发票票面金额的分布距离均匀分布越远，这种代替的近似性越大。

Y. Liu等(2005)指出，在上述事实下有下面的关系成立（本文使用了与该文献不同的公式记号，下文不再说明）：

$$X^{\otimes}=\bar{X}M \qquad (9)$$

其中，$X^{\otimes}$是错误发票票面金额的总体总值，$\bar{X}$是所有发票（包括错误的和非错误的）票面金额的总体均值，M是总体中错误发票的张数。

基于式(9)，用样本构造式(9)等号右边M的估计量，将它与$\bar{X}$相乘便得到了$X^{\otimes}$的估计。但是，这样做丢弃了样本中m张错误发票票面金额的信息。怎样把这些信息纳入估计量呢？Y. Liu等(2005)为式(9)写了一个脚注，给出了式(9)的修正形式：

$$\hat{X}^{\otimes}=X_n^{\otimes}+\bar{X}_{(N-n)}(\hat{M}-m) \qquad (10)$$

式(10)中，$X_n^{\otimes}$是样本中错误发票票面金额之和，$\bar{X}_{(N-n)}(\hat{M}-m)$则是未进入样本的那些发票中错误发票票面金额总值的估计量，它是根据式(9)写出的。

四、总体中错误发票张数和票面金额总值的区间估计

式(1)给出了在观察到m的条件下总体中错误发票张数M的后验分布。

根据式(1)和式(9)可得：

$$Pr(X_m^{\otimes}|m;M)=Pr(M|m) \qquad (11)$$

式(11)给出了在观察到m的条件下总体中错误发票票面金额总值$X^{\otimes}$的后验分布（即M取m，m+1，…，N等各个不同值时相应的错误发票票面金额总值$X_m^{\otimes}$与相应概率的分布列）。

对于区间估计，Y. Liu等（2005）所构造的是由Bain和Engelhard(1991)所提出的最小期望长度可信区间。最小期望长度可信区间又叫做最大后验密度可信区间，它应符合这样的条件：这个区间只包含那些具有最大后验密度的点，也就是说，这个区间外部的任意一个点的后验密度与区间内不论哪一个点的后验密度相比都要小。

为了寻找最小期望长度可信区间，首先要把M的后验分布具体地写出来。这就是，先把M的各个可能值(m，m+1，…，N)逐一代入式(1)，计算出各个M值相应的后验概率，将它们编制成M后验概率分布列；然后以此为基础编制M的累积后验概率分布列。

所需要的最小期望长度可信区间要用逐步逼近法来寻找。操作方法是：

1. 在众数的右边任意选择变量轴上的一个点（变量M的某一个值）作为初始可信区间的上限（记做$M_U^{(1)}$），将该值代入式(1)算出相应的初始可信区间的上限概率密度函数值（记做$k^{(1)}$）。

2. 用概率密度函数值在众数的左边用计算机搜索，找出该值（或与该值近似相等的概率密度函数值）所对应的自变量M的值，作为初始可信区间的下限（记做$M_L^{(1)}$）。

3. 应用M的累积后验概率分布列计算随机变量M在$M_L^{(1)}$与$M_U^{(1)}$之间取值的概率。

基于价值网的战略成本管理框架研究

盛 革

（肇庆学院 广东肇庆 526061）

【摘要】网络经济时代，价值网已替代价值链成为新的价值创造模式。文章在解析价值网系统构造的基础上，初步提出了一个基于价值网的战略成本管理框架。

【关键词】价值网 价值链 战略成本管理

战略成本管理是以战略的眼光从成本的源头识别成本驱动因素，对价值链进行成本管理，为战略管理的每一个关键步骤提供战略性成本信息，以形成企业竞争优势，从而有效适应企业外部环境的变化。价值链分析作为一种战略性分析工具，是战略成本管理的核心和出发点。企业通过对内部价值链和外部价值链进行分析实现对企业内外环境因素的分析，将成本管理从企业内部延伸至企业所在的整个价值链。然而，随着网络经济的到来，价值创造的基本逻辑与理念已发生深刻变化。哈佛商学院的Applegate在2000年就明确指出，在网络经济时代，应该以“价值网”而不是“价值链”来分析企业的商业模式。现代企业竞争，已不再是单一企业或线性价值链的竞争，而是企业与其协力者所共同营造的价值网的竞争（汪涛、李威，2003）。价值网被定义为“由顾客、供应商、合作企业和它们之间的信息流构成的动态网络”，其通过定制化服务与顾客保持一致，纳入由供应商、客户甚至竞争对手构成的唯一增值网络，并具有敏捷生产与分销和快捷的市场响应特征。特别是近年来，由于新一代使能技术在企业管理及商务领域的应用扩展，以及对象管理组织（OMG）和企业应用集成（EAI）等关键技术的驱动，进一步加速了企业价值模块的整合、价值链的解构和重建，促使企业业务模式被整体纳入价值网范畴，结成了企业与合作者之间的战略联盟以及一体化的“竞合”商务关系。现代企业的管理重心已由沟通、合作、协调向以网络、协同、共享为主题的网链管理转变。由于企业经营环境的变化和价值网概念的扩展，理论上需要对以价值链为基础的战略成本管理框架进行拓展。本文基于战略成本管理理论及价值网理论，初步提出了一个基于价值网的战略成本管理基本框架，以期为价值网内企业战略成本管理实践提供参考。

一、相关理论简介

1. 价值链理论的局限性。迈克尔·波特在上世纪80年代首先提出了价值链概念，认为企业所进行的一切活动均为价值活动，因而企业的一系列相关业务（或彼此关联的增值活动）被恰当地描述为一个价值链。引入价值链分析不仅能较好地控制经营成本与改善运作效率，同样也能有效降低管理成本以体现管理价值，并能够为客户传递价值和创造价值，即提供差别化的价值或服务。价值链作为竞争优势来源的重要分析工具已被企业界广泛采纳。然而，在网络经济环境下，价值链理论凸现出明显的局限性（傅代国、田小刚，2008）。

（1）线性、静态思维模式。在价值链理论中，企业经营战略的一个重要问题是构建企业价值链和产业价值链，企业经营的核心问题是“在价值链上定位”和“将战略建立在独特的经

---

4. 用这个概率（初始可信区间的可信概率）与事先所要求的可信概率（如90%）比较。若初始可信区间的可信概率小于90%，说明需要将初始可信区间扩大；若初始可信区间的可信概率大于90%，说明需要将初始可信区间缩小。

5. 构建第二轮可信区间。若是需要将初始可信区间扩大，则在$M_U^{(1)}$的右侧适当选一个新的M值，重复前面的过程；若是需要将初始可信区间缩小，则在$M_U^{(1)}$的左侧适当选一个新的M值，重复前面的过程。

6. 若是第二轮可信区间的可信概率与90%仍有距离，便继续构建第三轮可信区间。如此进行下去，直到所构建的可信区间的可信概率恰好等于（或近似等于）90%为止。

这样，便建立了总体中错误发票张数M的90%可信概率的最小期望长度可信区间。将这个区间的上限和下限分别代入式（10）算出结果，便得到了总体中错误发票票面金额总值$X^{\otimes}$的90%可信概率的最小期望长度可信区间。

上面的计算过程中，把众数作为参照，分别在众数的两侧寻找可信区间的上限和下限，这是为了减少寻找的工作量。有的时候无法计算众数，在一开始，只好任意选择一个适当的数值作为初始可信区间的上限。

主要参考文献

1. Yan Liu, Mary Batcher, Scheuren F. Efficient Sampling Design in Audit Data. Journal of Data Science, 2005; 3

2. Yan Liu, Mary Batcher, Wendy Rotz. Application of The Hypergeometric Distribution in a Special Case of Rare Events, Proceedings of The Annual Meeting of The American Statistical Association, 2001