第 19 章　蒙特卡罗法与自助法

19.1　蒙特卡罗法的思想与用途

通过计算机模拟从总体抽取大量随机样本的计算方法统称为“蒙特卡罗法”(Monte Carlo Methods，简记 MC)。

例（计算圆周率π）：在边长为 1 的正方形中内接1/4单位圆。正方形面积为 1，1/4圆面积为π/4。如知道1/4单位圆占正方形面积的比例，就可计算π。

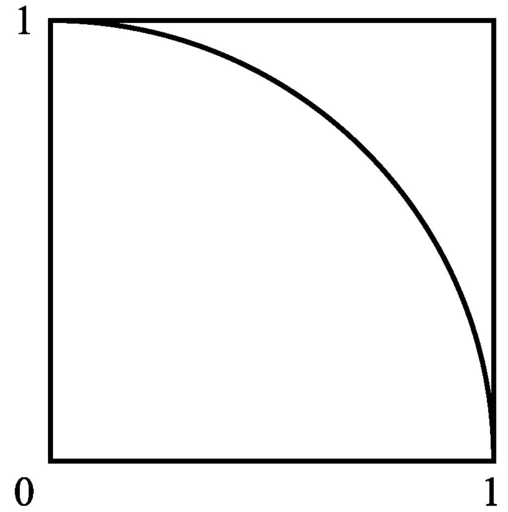

图 19.1　计算圆周率π的随机实验

向这个正方形随机地射箭，落点在正方形上服从二维均匀分布。

重复实验 n 次，其中有 m 次落在1/4圆内。

根据大数定律，$m/n \xrightarrow{p} \pi/4$，故$\pi \approx 4m/n$。

在计量中，常用 MC 来确定统计量的小样本性质。

【例】对于$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \ (i = 1, \cdots, n)$，对$H_0 : \boldsymbol{R\beta} = \boldsymbol{r}$进行显著性水平为 5%的大样本检验：

$$W \equiv n(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})' \left[\boldsymbol{R} \widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}) \boldsymbol{R}' \right]^{-1} (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}) \xrightarrow{d} \chi^2(m)$$

其中$\hat{\boldsymbol{\beta}}$为 OLS 估计量，m 为线性约束个数。

渐近χ^2分布只是真实分布的近似，故“5%”可能只是“名义显著性水平”(nominal size)，而非“真实显著性水平”(true or actual size)，二者之差称为“显著性水平扭曲”(size distortion)。

可用 MC 来确定“真实显著性水平”。

第一步，给定$\boldsymbol{\beta}$的具体取值，以及$\boldsymbol{x}$与ε的概率分布。

第二步，从$\boldsymbol{x}$与ε的分布中随机抽样，得到$\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$与$\{\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n\}$。

第三步，根据方程$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$计算$\{y_1, y_2, \cdots, y_n\}$。

第四步，对此样本进行 OLS 估计，计算统计量W，与$\chi^2(m)$的 5% 临界值比较，确定是否拒绝原假设$H_0: \boldsymbol{R\beta} = \boldsymbol{r}$。

第五步，大量重复第二至第四步，得到 M 个随机样本(比如，$M = 1000$)，进行 M 次检验，则拒绝原假设的比例就是真实显著性

水平。

19.2 蒙特卡罗法实例：模拟中心极限定理

19.3 蒙特卡罗法实例：服从卡方分布的扰动项

19.4 蒙特卡罗积分

MC 的另一用途是计算复杂或高维的积分，称为“蒙特卡罗积分”(Monte Carlo integration)。

考虑计算定积分$\int_a^b f(x)dx$，其中a,b为有限值。

通过变量替换，可将积分上下限变为1与0，故仅考虑$I \equiv \int_0^1 f(x)dx$。

假设x服从在$[0,1]$上的均匀分布，则随机变量函数$f(x)$的期望值

$$\mathrm{E}[f(x)] = \int_0^1 f(x)\cdot 1\,dx \equiv I$$

抽取随机变量x的样本容量为S的随机样本，记为$\{x_1,\cdots,x_s,\cdots,x_S\}$，则蒙特卡罗积分估计值为$f(x)$的样本均值：

$$\hat{I}_{\mathrm{MC}} = \frac{1}{S}\sum_{s=1}^{S} f(x_s)$$

根据大数定律，当$S \to \infty$时，样本均值$\hat{I}_{\mathrm{MC}} \xrightarrow{p} \mathrm{E}\left[f(x)\right] = I$。

如果积分上限a或下限b为无穷，可从某个适当的概率密度$g(x)$中抽取随机样本$\left\{x_1, \cdots, x_s, \cdots, x_S\right\}$。原积分总可写为

$$\int_a^b f(x)\,dx = \int_a^b \left[\frac{f(x)}{g(x)}\right] g(x)\,dx \equiv \int_a^b w(x) g(x)\,dx = \mathrm{E}\left[w(x)\right]$$

其中，$w(x) \equiv \dfrac{f(x)}{g(x)}$。蒙特卡罗积分估计值为

$$\hat{I}_{\mathrm{MC}} = \frac{1}{S} \sum_{s=1}^{S} w(x_s)$$

从密度函数$g(x)$中抽样的方法称为“重要性抽样”(importance

sampling)，因为函数$w(x)$决定了每个样本点的权重或重要性。

19.5 最大模拟似然法与模拟矩估计

使用 MLE 的前提是，能写出似然函数$f(\boldsymbol{y}\,|\,\boldsymbol{x},\boldsymbol{\theta})$。

有时，该似然函数可能包含无法求解的积分。

比如，在随机效应的非线性面板模型中，要将个体效应u_i积分掉(u_i不可观测)，才能写出似然函数。

记u_i的密度函数为$g(u_i)$，并假设第 i 个观测值的似然函数为

$$f(y_i \mid \boldsymbol{x}_i, \boldsymbol{\theta}) = \int h(y_i \mid \boldsymbol{x}_i, \boldsymbol{\theta}, u_i) g(u_i)\, du_i$$

如果积分无解析解，可使用蒙特卡罗积分进行估计。

从分布$g(u_i)$中随机抽取 S 个观测值，记为$\left\{u_i^1, \cdots, u_i^S\right\}$，则上式的估计值为

$$\hat{f}(y_i \mid \boldsymbol{x}_i, \boldsymbol{\theta}) = \frac{1}{S}\sum_{s=1}^{S} h(y_i \mid \boldsymbol{x}_i, \boldsymbol{\theta}, u_i^s)$$

假设样本为 iid，则整个样本的对数似然函数估计值为

$$\ln \hat{L}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln \hat{f}(y_i \mid \boldsymbol{x}_i, \boldsymbol{\theta})$$

其中，n 为样本容量。

最大化上式所得到的估计量$\hat{\boldsymbol{\theta}}_{\mathrm{MSL}}$称为“最大模拟似然估计量”(Maximum Simulated Likelihood Estimator，简记 MSL)。

在一定正则条件下，当模拟抽样的次数$S \to \infty$时，$\hat{f}$对 f 的近似程度越来越好，即$(\hat{f} - f) \xrightarrow{p} 0$，则 MSL 为一致估计量。

如果$\sqrt{n}/S \to 0$(即 S 的增长速度快于$\sqrt{n}$)，则 MSL 为渐近有效估计量(渐近等价于 MLE)，且服从渐近正态分布。

类似地，在进行矩估计时，如果矩条件中包含无解析解的积分，也可使用蒙特卡罗积分来估计此矩条件，然后进行矩估计。

此法称为“模拟矩估计”(Method of Simulated Moments)，简记 MSM。

19.6 自助法的思想与用途

MC 虽然威力大，但必须对总体模型做很具体的假定，所得结论不清楚在多大意义上能够推广。

Efron (1979)提出了对原始样本进行“再抽样”(resampling)的方法，即“自助法”或“自举法”(bootstrap)。

假设从总体抽得样本容量为 n 的随机样本。来自总体的样本带有总体的信息。

将此样本看作一个总体，进行“有放回”(with replacement)地抽样，样本容量仍然为 n。这种样本被称为“自助样本”(bootstrap sample)。

由于是有放回地抽样，原来的某些观测值可能不出现，而有些观测值则可能多次出现。

可通过计算机模拟获得许多自助样本，然后利用这些自助样本对总体进行统计推断。

假设$\{x_1, x_2, \cdots, x_n\}$是来自总体 F 的随机样本。

定义总体 F 的经验分布函数(empirical distribution function)F_n：

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{I}(x_i \le x), \quad -\infty < x < \infty$$

$\boldsymbol{I}(\cdot)$为示性函数，而$\sum_{i=1}^{n}\boldsymbol{I}(x_i \le x)$表示样本中小于或等于 x 的个数。

经验分布函数的形状为阶梯函数，在每个x_i处向上跳一个台阶。

可以证明，对任意 x，$F_n(x) \xrightarrow{p} F(x)$，这是自助法成立的前提。

自助法可看成是从经验分布函数中不断地抽样。

自助法的用途主要有两个方面。

首先，对于某些统计量(比如，样本中位数)，常规方法很难得到标准误。可使用自助法，计算每个自助样本的样本中位数，得到样本中位数的分布，并计算其标准误。

其次，可使用自助法得到更加渐近有效的估计量(asymptotic refinement)。

19.7 自助法的分类

(1) **非参数自助法**(nonparametric bootstrap)，也称“经验分布自助法”(empirical distribution function bootstrap)。将原始样本进行有放回地随机抽样。在回归模型中，意味着将$(y_i, \boldsymbol{x}_i)$成对抽样，故也称“成对自助法”(paired bootstrap)。

(2) **参数自助法**(parametric bootstrap)。

假设总体分布函数的形式已知，为$F(x,\theta)$，而θ未知。先得到θ的估计量$\hat{\theta}$(比如使用 MLE)，然后从总体$F(x,\hat{\theta})$中重复抽样。

此法的前提是对总体分布函数的形式比较确信。在此前提下，参数自助法比非参自助法更有效率。

在回归模型中，需先确定条件分布的具体形式，即$y\,|\,\boldsymbol{x} \sim F(\boldsymbol{x},\theta)$。

一种方法是，得到估计量$\hat{\theta}$后，从$F(\boldsymbol{x}_i,\hat{\theta})$中随机抽样得到对应的$y_i$。这相当于是“固定解释变量”(fixed regressors)的情形。

另一种方法是，先从$\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$中进行再抽样(resample)，得到

$\boldsymbol{x}_i^*$，然后再从$F(\boldsymbol{x}_i^*, \hat{\theta})$中随机抽样得到对应的$y_i$。这相当于“随机解释变量”(stochastic regressors)的情形。

(3) **残差自助法**(residual bootstrap)。

对于回归模型 $y_i = g(\boldsymbol{x}_i, \boldsymbol{\beta}) + \varepsilon_i$，首先通过估计得到残差 $\hat{\varepsilon}_i = y_i - g(\boldsymbol{x}_i, \hat{\boldsymbol{\beta}})$。

对残差$\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \cdots, \hat{\varepsilon}_n\}$使用自助法，得到残差的自助样本$\{\hat{\varepsilon}_1^*, \hat{\varepsilon}_2^*, \cdots, \hat{\varepsilon}_n^*\}$。

计算对应的 $y_i^* = g(\boldsymbol{x}_i, \hat{\boldsymbol{\beta}}) + \hat{\varepsilon}_i^*$，进而得到自助样本$\{(y_1^*, \boldsymbol{x}_1), \cdots, (y_n^*, \boldsymbol{x}_n)\}$。

19.8 使用自助法估计标准误

假设原始样本为$\{x_1, x_2, \cdots, x_n\}$。对于未知参数$\theta$的估计量$\hat{\theta} = \hat{\theta}(x_1, x_2, \cdots, x_n)$，需计算标准误$\sigma_{\hat{\theta}} \equiv \sqrt{\mathrm{Var}(\hat{\theta})}$，但有时无解析式。

如果从真实总体F获得样本容量为n的B个随机样本，对每个样本都可计算$\hat{\theta}$，得到B个估计值$\{\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_B\}$，则

$$s_{\hat{\theta}} \equiv \sqrt{\frac{1}{B-1}\sum_{i=1}^{B}(\hat{\theta}_i - \bar{\theta})^2}$$

其中$\bar{\theta} \equiv \dfrac{1}{B}\sum_{i=1}^{B}\hat{\theta}_i$。

但真实总体 F 的分布未知，而从总体多次抽样的成本可能很高。

以经验分布函数F_n来近似真实分布 F，并从F_n中大量抽取随机样本，即在原始样本$\{x_1, x_2, \cdots, x_n\}$中每次有放回地抽样，得到样本容量为 n 的自助样本$\{x_1^*, x_2^*, \cdots, x_n^*\}$，并计算$\hat{\theta}^* = \hat{\theta}(x_1^*, x_2^*, \cdots, x_n^*)$。

如此重复，共抽取 B 个自助样本，则得到θ的 B 个自助估计值$\{\hat{\theta}_1^*, \hat{\theta}_2^*, \cdots, \hat{\theta}_B^*\}$。可以定义标准误的自助估计为

$$s_{\hat{\theta}}^* \equiv \sqrt{\frac{1}{B-1}\sum_{i=1}^{B}(\hat{\theta}_i^* - \overline{\theta}^*)^2}$$

其中，$\overline{\theta}^* \equiv \dfrac{1}{B}\sum_{i=1}^{B}\hat{\theta}_i^*$。

19.9　使用自助法进行区间估计

(1) **百分位法**(percentile method)。

得到自助估计量$\hat{\theta}^*$的经验分布$\left\{\hat{\theta}_1^*, \hat{\theta}_2^*, \cdots, \hat{\theta}_B^*\right\}$。

将$\left\{\hat{\theta}_1^*, \hat{\theta}_2^*, \cdots, \hat{\theta}_B^*\right\}$按从小到大的顺序排列，并记其$\alpha/2$与$(1-\alpha/2)$上分位数(upper quantile)分别为$\hat{\theta}_{\alpha/2}^*$与$\hat{\theta}_{1-\alpha/2}^*$，则$\theta$的置信区间为

$$\left[\hat{\theta}_{1-\alpha/2}^*,\ \hat{\theta}_{\alpha/2}^*\right]$$

(2) **基于正态的置信区间**(normal-based confidence interval)。

使用标准正态分布来估计置信区间，即

$$\left[\hat{\theta}-1.96\times s_{\hat{\theta}}^{*},\ \hat{\theta}+1.96\times s_{\hat{\theta}}^{*}\right]$$

其中，$s_{\hat{\theta}}^{*}$是用自助法估计的标准误，并假定置信度为 95%。

(3) **百分位 t 法**(percentile-t method)。

根据每个自助样本计算对应的自助 t 统计量，

$$t_i^{*}\equiv\frac{\hat{\theta}_i^{*}-\hat{\theta}}{s_{\hat{\theta}_i^{*}}},\quad i=1,\cdots,B$$

其中，$\hat{\theta}$为根据原始样本计算的θ估计量，而$s_{\hat{\theta}_i^*}$是根据$\left\{\hat{\theta}_1^*, \hat{\theta}_2^*, \cdots, \hat{\theta}_B^*\right\}$计算的标准误。

得到自助t统计量的经验分布$\left\{t_1^*, t_2^*, \cdots, t_B^*\right\}$，并记其$\alpha/2$与$(1-\alpha/2)$上分位数(upper quantile)分别为$t_{\alpha/2}^* > 0$与$t_{1-\alpha/2}^* < 0$，则$\theta$的置信区间为

$$\left[\hat{\theta} + t_{1-\alpha/2}^* \times s_{\hat{\theta}},\ \hat{\theta} + t_{\alpha/2}^* \times s_{\hat{\theta}}\right]$$

19.10 使用自助法进行假设检验

考虑用自助法进行如下双边检验，$H_0: \theta = \theta_0 \;\; vs \;\; H_1: \theta \neq \theta_0$。

一种方法是，如果$\theta_0 \in \left[\hat{\theta}^*_{1-\alpha/2},\ \hat{\theta}^*_{\alpha/2}\right]$，则接受原假设$H_0$；反之，则拒绝。这就是“百分位法” (percentile method)。

另一方法是，在假设H_0成立的情况下，计算原始样本的 t 统计量，

$$t \equiv \frac{\hat{\theta} - \theta_0}{s_{\hat{\theta}}}$$

如果$t \in \left[t^*_{1-\alpha/2},\ t^*_{\alpha/2}\right]$，则接受原假设$H_0$；反之，则拒绝。其中，$t^*_{\alpha/2}$与$t^*_{1-\alpha/2}$的定义如上。这就是“百分位 t 法”。

19.11　自助法的一致性(选读)

19.11 自助法的一致性(选读)

自助法是一致估计量吗?考虑以下一般的自助法。随机样本$\{x_1,x_2,\cdots,x_n\}$来自总体$F(x,\theta)$。我们希望知道统计量$T_n=T_n(x_1,x_2,\cdots,x_n)$的概率分布$G_n(t,F)\equiv \mathrm{P}(T_n\leq t)$,其中$G_n(t,F)$的下标$n$表示该分布依赖于样本容量$n$。另外,$G_n(t,F)$也依赖于真实的总体分布$F$。

大样本理论(渐近理论)通过让$n\to\infty$,即用$G_\infty(t,F)$来近似$G_n(t,F)$;对于$G_n(t,F)$中未知的F,则使用其一致估计量$\hat{F}=F(x,\hat{\theta})$来替代,其中$\hat{\theta}$是$\theta$的一致估计量。自助法则不需要让$n\to\infty$,而是使用经验分布函数$F_n$来替代$F$,即用$G_n(t,F_n)$来近似$G_n(t,F)$。尽管$G_n(t,F_n)$一般没有解析解,但可以把样本$F_n$近似地看作总体$F$,并从$F_n$中进行再抽样,获得自助样本。从一个自助样本$\{x_1^*,x_2^*,\cdots,x_n^*\}$,就可以得到一个统计量$T_n^*=T_n(x_1^*,x_2^*,\cdots,x_n^*)$。独立地重复抽样$B$次,得到$B$个统计量$\{T_{n,1}^*,T_{n,2}^*,\cdots,T_{n,B}^*\}$,从中可以得到$T_n$分布函数的自助估计(bootstrap estimate)$\hat{G}_n(t,F_n)$,即$\{T_{n,1}^*,T_{n,2}^*,\cdots,T_{n,B}^*\}$的经验分布函数:

$$\hat{G}_n(t,F_n)\equiv\frac{1}{B}\sum_{i=1}^{B}\mathbf{1}(T_{n,i}^*\leq t) \tag{19.18}$$

其中,$\mathbf{1}(\cdot)$为示性函数。显然,当$B\to\infty$时,$\hat{G}_n(t,F_n)\xrightarrow{p}G_n(t,F_n)$,因为前者是后者的经验分布函数(参见本章附录)。然而,当$n\to\infty$时,是否$G_n(t,F_n)\xrightarrow{p}G_n(t,F)$?显然,这要求$F_n\xrightarrow{p}F$(即$F_n$是$F$的一致估计)。另外,这还要求总体分布函数$F(x,\theta)$与统计量$T_n$的分布函数$G_n(t,F)$都是光滑的。这些条件在实践中一般都成立,故自助估计量是一致的。

19.12 异方差情况下的自助法

由于异方差的存在不影响观测数据$(y_i, \boldsymbol{x}_i)$仍然为 iid，故仍可使用成对自助法(paired bootstrap)。

但残差自助法(residual bootstrap)却不成立，因为在条件异方差的情况下，扰动项不是 iid，故经验分布函数F_n不是总体分布函数 F 的一致估计，自助估计量也就不一致。

Wu(1986)与 Liu(1988)提出“野自助法”(wild bootstrap)，对残差$\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \cdots, \hat{\varepsilon}_n\}$先进行线性变换再进行抽样，以满足从异方差的扰动项进行抽样的要求。

定义具有两点分布的新残差为：

$$\hat{\varepsilon}_i^* = \begin{cases} \dfrac{1-\sqrt{5}}{2}\hat{\varepsilon}_i, & 以\dfrac{\sqrt{5}+1}{2\sqrt{5}}的概率 \\ \dfrac{1+\sqrt{5}}{2}\hat{\varepsilon}_i, & 以\dfrac{\sqrt{5}-1}{2\sqrt{5}}的概率 \end{cases}$$

可以证明，$\mathrm{E}(\hat{\varepsilon}_i^*) = 0$，$\mathrm{Var}(\hat{\varepsilon}_i^*) = \hat{\varepsilon}_i^2$。

对$\left\{\hat{\varepsilon}_1^*, \hat{\varepsilon}_2^*, \cdots, \hat{\varepsilon}_n^*\right\}$进行再抽样，就是从异方差的扰动项进行抽样。

19.13 面板数据与时间序列的自助法

可使用成对自助法对个体 i 进行再抽样，而不对时间 t 进行再抽样，即如果抽中个体 i，则个体 i 在所有时间的观测值都同时被抽中。

这种方法称为“分块自助法”(block bootstrap)，对于非线性面板数据或聚类数据(clustered data)也适用。

由于自助法假定样本为 iid，而时间序列数据通常存在自相关(故不是 iid)，因此针对时间序列的自助法更为复杂。

19.14 自助法的 Stata 命令

19.15 使用自助法进行稳健的豪斯曼检验