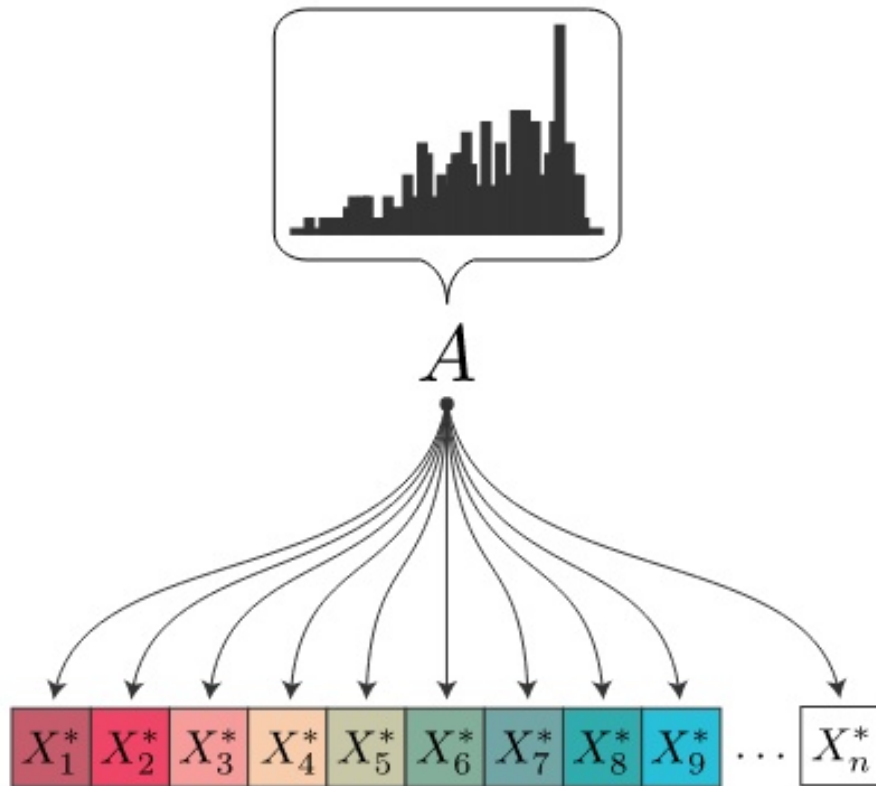


知乎

首发于  
川流不息

## 用 Bootstrap 进行参数估计大有可为



石川

量化交易 话题的优秀回答者

已关注

68 人赞同了该文章

### 1 从 t 分布说起

在量化投资领域，有大量需要进行参数估计（parameter estimation）的场景。比如在按照马科维茨的均值方差框架配置资产时，就必须计算投资品的收益率均值和协方差矩阵。很多时候，对于需要的统计量，仅有点估计（point estimate）是不够的，我们更感兴趣的是从样本数据得到的点估计和该统计量在未知总体中的真实值之间的误差。在这方面，区间估计——即计算出目标统计量的置信区间（confidence interval）——可以提供我们需要的信息。

谈到置信区间，人们最熟悉的当属计算总体均值（population mean）的置信区间。这是因为在中心极限定理（Central Limit Theorem）和正态分布假设（Normal distribution）下，总体均值的置信区间存在一个优雅的解析表达。利用样本均值和其 standard error 求出的 test statistic 满足 t 分布（Student's t-distribution），通过查表找到置信区间两边各自对应的 t 统计量的临界值（critical value）便可以方便的求出置信区间。由于 t 分布是对称的，因此总体均值的置信区间是关于样本均值对称的。



假设总体满足正态分布，而我们想计算均值的置信区间。如果总体的标准差  $\sigma$  已知，则可以使用正态分布计算均值的置信区间；如果  $\sigma$  未知，则使用样本的标准差  $s$  代替，并且利用 t 分布来代替正态分布计算均值的计算区间。这就是 t 分布被提出来的初衷。因此，**使用 t 分布计算均值的置信区间隐含着总体分布满足正态分布这个假设。**

但是，对于实际中的问题，总体并不满足正态分布，因此看起来我们不能使用 t 分布计算均值的置信区间。好消息是，我们还有另外一个“大招”：中心极限定理。中心极限定理告诉我们，不管总体的分布是什么样，总体的均值近似满足正态分布，因此我们仍然可以使用 t 分布计算置信区间。

中心极限定理是概率论中的一组定理。中心极限定理说明，大量相互独立的随机变量，其均值的分布以正态分布为极限。这组定理是数理统计学和误差分析的理论基础，指出了大量随机变量之和近似服从正态分布的条件。

可见，对于一个未知分布总体均值的推断，我们必须倚赖中心极限定理和正态分布的假设。**如果未知分布非常不规则或样本数不足，则中心极限定理指出的均值近似为正态分布便难以成立，而基于 t 分布计算出来的均值置信区间也不够准确。**

除了均值外，对于人们关心的许多其他统计量，比如中位数、分位数、标准差、或者相关系数，它们与均值不同，无法从 Normal Theory 中可以得到优雅的解析表达式来计算其置信区间，因此上述传统方法无能为力。

从上面的分析可知，仅仅掌握传统的 Normal Theory 方法局限性很大，使得我们在求解置信区间的很多问题面前举步维艰。因此，今天就给大家介绍一个利器——**Bootstrap 方法**。它在计算统计量的置信区间时大有可为。

## 2 Bootstrap 的由来和原则

*The bootstrap is a computer-based method for assigning measures of accuracy to statistical estimates. -- Efron & Tibshirani, An introduction to the bootstrap, 1993*  
译：Bootstrap 是一个基于计算机的方法，它可以计算统计估计的准确性。

自 1979 年以来，Bootstrap 方法得到了广泛的推广，其始作俑者是 Bradley Efron（Bootstrap 这个词也是他发明的）。它的核心思想是通过使用数据本身，从而估计从该数据中计算出来的统计数据的变化。现代计算机强大的计算能力使得该方法的实现非常简单。

有些费解？别着急，马上解释 Bootstrap 的意思和它的核心思想。

Bootstrap 一词出自英文习语“pull yourself up by your bootstraps”，它的直译是“通过拉你自己靴子的鞋带把你从地面上拉起来”。它的隐含意是“improve your situation by you”



放到参数估计的上下文中，**Bootstrap** 意味着我们仅仅通过使用手头上的样本数据（样本数据“自力更生”）而不对总体的分布做任何假设（比如传统方法中的正态分布假设），来计算样本统计量在估计总体统计量时的误差。

*The central idea is that it may sometimes be better to draw conclusions about the characteristics of a population strictly from the sample at hand, rather than by making perhaps unrealistic assumptions about the population. -- Mooney & Duval, Bootstrapping, 1993*

译：Bootstrap 的核心思想是，通过手头的样本数据本身对总体统计量进行推论，而非基于对于总体分布做出不切实际的假设。

目标够伟大（样本数据自力更生），但具体要怎么做呢？**如何仅仅通过（反复的）使用手头的数据来对同样从这些数据中得到的统计量进行误差估计呢？**

这里面要用到一个非常重要的技巧：**可置换的重采样（resampling with replacement）**。在这个定义中，“可置换”是核心。什么是“可置换”呢？举个例子。假设袋子里有标号 1 到 10 的小球。我们“可置换”地不断地从袋子里随机抽出小球。第一次抽出了 3 号小球；“可置换”是说在下次抽取之前把 3 号小球重新放回到袋子里；即在第二次抽取的时候，我们仍然有可能再次抽到 3 号小球（它和其他 9 个球被抽到的概率是一样的），这便是可置换的含义。作为对比，生活中更多的是“无置换的抽取”，比如体彩 36 中 7 或者世界杯抽签，抽出的小球都不会再放回池子中。

下面就来看看 Bootstrap 的原则。假设我们有如下设定：

1. 令  $v$  代表我们感兴趣的一个总体统计量（比如均值、中位数、标准差等），它来自未知的总体分布  $F$ 。
2. 令  $x_1, x_2, \dots, x_n$  为来自总体的一组样本数据，它们称为原始样本数据。
3. 令  $u$  代表从该样本中计算出的统计量。
4. 以原始样本数据中的数据作为“总体”，进行可置换的重采样，得到一个重采样样本（又称为 Bootstrap 样本），记为  $x_1^*, x_2^*, \dots, x_n^*$ （重采样样本中的数据个数必须和原始样本数据中的数据个数相同）。
5. 令  $u^*$  代表利用上述 Bootstrap 样本数据计算出来统计量。

**Bootstrap 原则指出：**“Bootstrap 样本统计量  $u^*$  围绕原始样本统计量  $u$  的变化（简称为  $u^*$  的变化）”是“原始样本统计量  $u$  围绕总体统计量  $v$  的变化（简称为  $u$  的变化）”的一个很好的近似。

为了计算  $u^*$  的变化，我们只需要对原始样本数据进行大量的可置换重采样（为此需要使用计算机的计算能力，在没有计算机的年代，手动进行大量重采样的工作量可想而知），得到许多 Bootstrap 样本，并从每个样本中计算出统计量  $u^*$  的一个取值，这些取值便构成  $u^*$  的分

Bootstrap 样本的大小和原始样本大小相同。

根据 Bootstrap 原则，使用经验 Bootstrap 方法（empirical Bootstrap method）就可以计算任何总体统计量的置信区间。

### 3 经验 Bootstrap 方法

我们以计算某未知分布均值的置信区间为例说明经验 Bootstrap 方法。假设我们从某未知分布的总体中得到下面 10 个样本数据：30, 37, 36, 43, 42, 48, 43, 46, 41, 42。

我们的问题有两个：（1）估计总体的均值（点估计），（2）计算置信水平为 80% 的 Bootstrap 置信区间。

第一个问题很容易回答，样本均值 40.8 就是总体均值  $\mu$  的点估计。对于第二个问题，由于样本点太少（仅有 10 个）且总体分布未知（无法做正态分布假设），因此我们摒弃传统的方法，而采用经验 Bootstrap 方法计算其置信区间。

计算  $\mu$  的置信区间的本质是回答这样一个问题：样本均值  $\bar{x}$  的分布是如何围绕总体均值  $\mu$  变化的。换句话说，我们想知道  $\delta = \bar{x} - \mu$  的分布。 $\delta$  就是当我们使用  $\bar{x}$  来估计  $\mu$  时的误差。

如果我们知道  $\delta$  的分布，则可以找到待求置信区间左右两端的临界值。在本例中，因为我们关心的是置信水平为 80% 的置信区间，因此  $\delta$  的临界值是 10% 和 90% 分位对应的  $\delta_{0.9}$  和  $\delta_{0.1}$ 。由此计算出  $\mu$  置信区间为：

$$[\bar{x} - \delta_{0.1}, \bar{x} - \delta_{0.9}]$$

这是因为：

$$\begin{aligned} \text{prob}(\delta_{0.9} \leq \bar{x} - \mu \leq \delta_{0.1} | \mu) &= 0.8 \\ \Rightarrow \text{prob}(\bar{x} - \delta_{0.1} \leq \mu \leq \bar{x} - \delta_{0.9}) &= 0.8 \end{aligned}$$

值得一提的是，上面的概率是**条件概率**，它表示假设总体均值为  $\mu$  的条件下，样本均值  $\bar{x}$  围绕总体均值  $\mu$  的变化在  $\delta_{0.1}$  和  $\delta_{0.9}$  之间的概率。

不幸的是，由于来自总体的样本只有一个（上面的 10 个数）且  $\mu$  的真实值未知，我们并不知道  $\delta$  的分布（因此也就不知道  $\delta_{0.9}$  和  $\delta_{0.1}$ ）。但是我们仍然利器在手，那就是 Bootstrap 原则。它指出虽然我们不知道  $\bar{x}$  如何围绕  $\mu$  变化（即  $\delta$  的分布），但是它可以由  $\bar{x}^*$  如何围绕  $\bar{x}$  变化（即  $\delta^*$  的分布）来近似，这里  $\delta^*$  是利用 Bootstrap 样本计算的均值与原始样本均值的差：

知乎

首发于  
川流不息

通过进行多次有置换的重采样，得到多个 Bootstrap 样本，每一个样本中都可以计算出一个均值。使用每一个 Bootstrap 样本均值减去原始样本均值 (40.8) 就得到  $\delta^*$  的一个取值。利用计算机，很容易产生足够多的 Bootstrap 样本，即足够多的  $\delta^*$  的取值。**根据大数定理 (law of large numbers)**，当样本个数足够多时， $\delta^*$  的分布是  $\delta$  的分布好的近似。

有了  $\delta^*$  的分布，就可以找到  $\delta_{0.9}^*$  和  $\delta_{0.1}^*$ ，并用它们作为  $\delta_{0.9}$  和  $\delta_{0.1}$  的估计，从而计算出  $\mu$  的置信区间：

$$[\bar{x} - \delta_{0.1}^*, \bar{x} - \delta_{0.9}^*]$$

上述思路就是经验 Bootstrap 方法的强大所在。

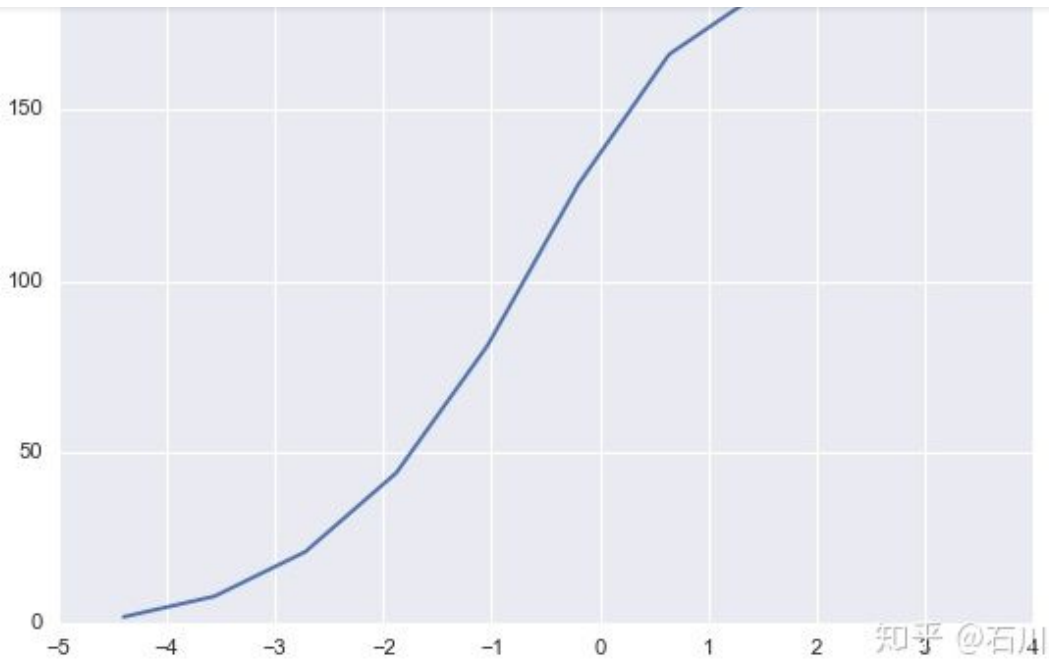
回到上面这个例子中。利用计算机产生 200 个 Bootstrap 样本（下图显示了前 10 个 Bootstrap 样本，每列一个）。

30	43	43	30	37	41	37	41	43	42
42	42	43	41	30	36	46	46	46	48
42	36	43	37	43	42	36	48	30	42
42	46	43	48	46	46	46	43	41	41
43	41	36	30	36	42	37	46	42	43
37	41	30	41	41	36	42	36	36	46
36	42	46	42	43	46	37	43	37	43
41	37	30	30	30	41	43	43	37	46
41	42	43	48	37	36	36	42	36	42
42	48	41	48	37	43	48	知乎 @石川		

由这 200 个 Bootstrap 样本计算出 200 个  $\delta^*$ ，它们的取值范围在 -4.4 到 4.0 之间， $\delta^*$  的累积密度函数如下图所示。



知乎

首发于  
川流不息

知乎 @石川

接下来，从这 200 个  $\delta^*$  中找出  $\delta_{0.9}^*$  和  $\delta_{0.1}^*$ 。由于  $\delta_{0.9}^*$  对应的是 10% 分位数，而  $\delta_{0.1}^*$  对应的是 90% 分位数，我们将 200 个  $\delta^*$  从小到大排序，其中第 20 个和第 181 个就是我们需要的数值： $\delta_{0.9}^* = -1.9$  以及  $\delta_{0.1}^* = 2.2$ 。由于原始样本均值为 40.8，因此求出  $\mu$  的 80% 的置信区间为：

$$[\bar{x} - \delta_{0.1}^*, \bar{x} - \delta_{0.9}^*] = [40.8 - 2.2, 40.8 + 1.9] = [38.6, 42.7]$$

## 4 Bootstrap 百分位法

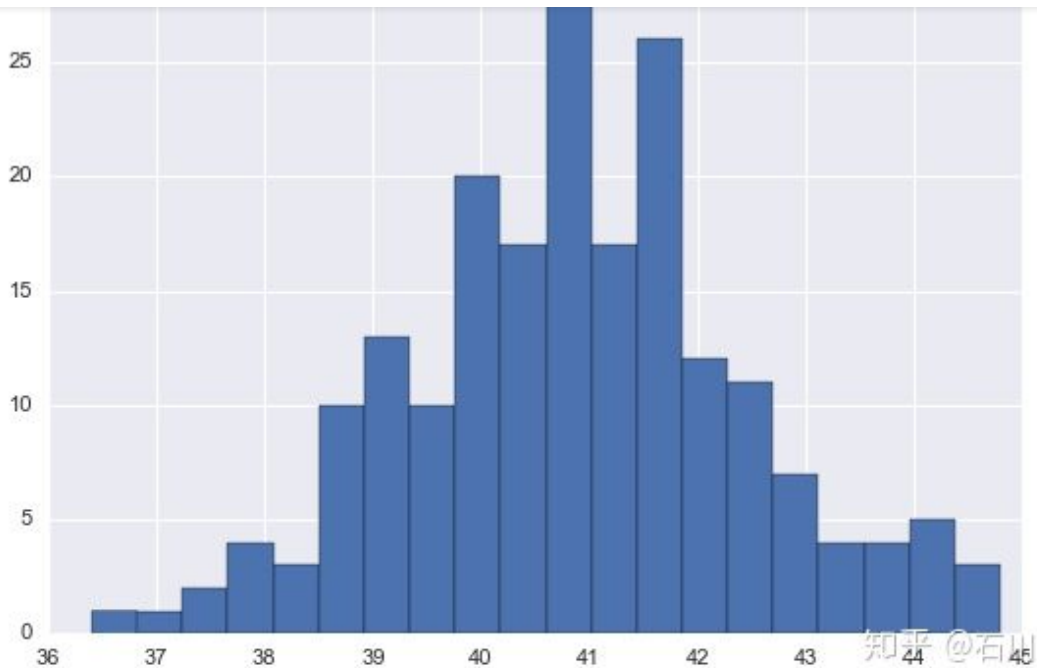
让我们来看看另外一种方法：**Bootstrap 百分位法 (Bootstrap percentile method)**。它与经验 Bootstrap 方法的不同之处在于，它不是用  $\delta^*$  的分布去近似  $\delta$  的分布，而是直接使用来自 Bootstrap 样本的统计量的分布作为原始样本统计量的分布。

让我们仍然用上一节中的例子来说明这种方法。

在那个例子中，我们对原始样本数据进行有置换的重采样，得到了 200 个 Bootstrap 样本。对于每个样本，计算出样本均值，因此一共有 200 个均值，它们构成了 Bootstrap 样本统计量  $\bar{x}^*$  的分布（下图）。







Bootstrap 百分位法使用来自 Bootstrap 样本统计量  $\bar{x}^*$  的分布作为原始样本统计量  $\bar{x}$  的分布的一个近似。因此，在这种方法下，我们只需要找到  $\bar{x}^*$  分布中 10% 分位和 90% 分位对应的  $\bar{x}^*$  的取值，它们就构成了  $\mu$  的置信区间。在本例中，这两个分位对应的  $\bar{x}^*$  的取值分别为 38.9 和 43，因此按这种方法得到的  $\mu$  的置信区间为：[38.9, 43]。

不难发现，上述两种方法得到的置信区间并不相同。它们是各有千秋还是说其中一个更准确呢？

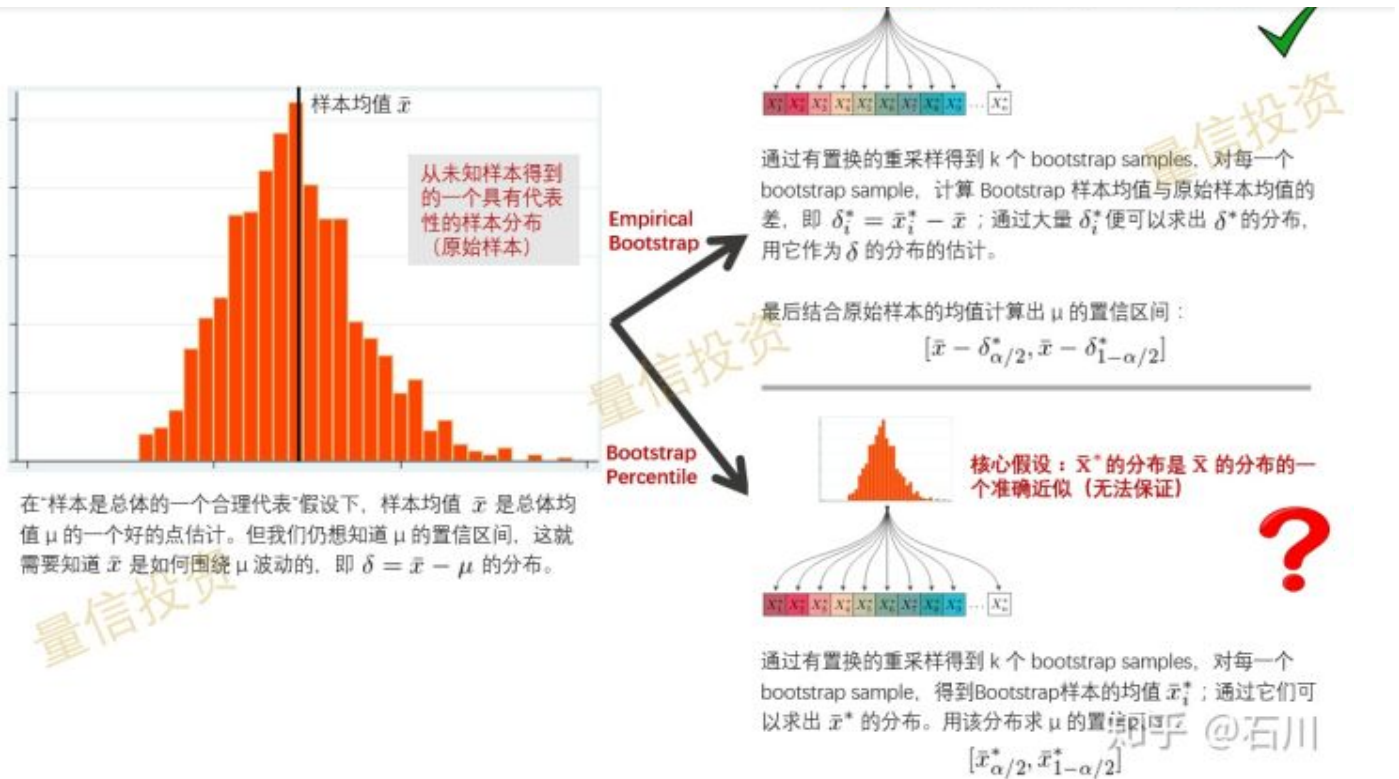
经验 Bootstrap 法和 Bootstrap 百分位法的区别如下：

- 经验 Bootstrap 法用  $\delta^*$  的分布去近似  $\delta$  的分布；之后再把误差加到原始样本均值的两侧，该置信区间是以样本均值  $\bar{x}$  为中心的。
- Bootstrap 百分位法直接用  $\bar{x}^*$  的分布来近似  $\bar{x}$  的分布（由于我们只有一个来自于总体的样本，因此我们没有  $\bar{x}$  的分布，而这种方法说我们可以是使用  $\bar{x}^*$  的分布代替）；它直接从  $\bar{x}^*$  的分布找到的置信区间作为总体均值的置信区间。这里一个很强的假设是  $\bar{x}^*$  的分布是  $\bar{x}$  分布的一个很好的近似。然而在现实中这是无法保证的，因此这种方法不好，它的准确性存疑。

Bootstrap 原则传达的是这样一个意思：样本统计量  $\bar{x}$  是以总体统计量  $\mu$  为中心围绕其波动；Bootstrap 样本统计量  $\bar{x}^*$  是以原始样本统计量  $\bar{x}$  为中心围绕其波动。如果  $\bar{x}$  和  $\mu$  有较大的差异，则  $\bar{x}$  和  $\bar{x}^*$  的分布也会不同（即 Bootstrap 百分位法的假设不成立）。反观  $\delta$  和  $\delta^*$ ，它们的分布各自描述  $\bar{x}$  如何围绕  $\mu$  波动以及  $\bar{x}^*$  如何围绕  $\bar{x}$  波动。Bootstrap 原则指出即使  $\bar{x}$  和  $\bar{x}^*$  分布不同， $\delta^*$  的分布仍然是  $\delta$  的分布的一个很好的近似，因此以原始样本均值  $\bar{x}$  为中心，以  $\delta^*$  的分布计算出误差，最终得到的  $\mu$  的置信区间是比较准确的。由此可知，经验 Bootstrap 方法优于 Bootstrap 百分位法。在实践中，应该使用前者。

下图概括了上文中对二者的比较。

知乎

首发于  
川流不息

## 5 Bootstrapped-t 方法

除了上面介绍的两种方法外, 最后我还想再提另一种方法: **Bootstrapped-t 方法**。

这种方法和第一节中介绍的传统方法十分接近。在传统方法中, 基于 Normal Theory 的假设, 我们只需要知道  $t$  统计量的临界值就可以计算均值的置信区间。**传统方法假设待估计的统计量的分布是对称的。**

然而在现实问题中, 这个假设可能无法满足, 所以假设对称并通过查表找出  $t$  statistic 的临界值会有问题 (因为得到的置信区间是对称的)。由此提出了 Bootstrapped-t 方法。

这种方法的核心思想是将每个 Bootstrap 样本中计算的统计量转化为一个对应的  $t$  statistic。这样, 有多少个 Bootstrap 样本我们就有多少个 Bootstrapped  $t$  statistic。由此, 可以计算出 Bootstrapped  $t$  statistic 的分布。用这个分布代替查表来找到计算置信区间时所需的  $t$  statistic 的临界值, 从而计算置信区间:

$$[\bar{x} - t_{\alpha/2}^* s_{\bar{x}}, \bar{x} - t_{1-\alpha/2}^* s_{\bar{x}}]$$

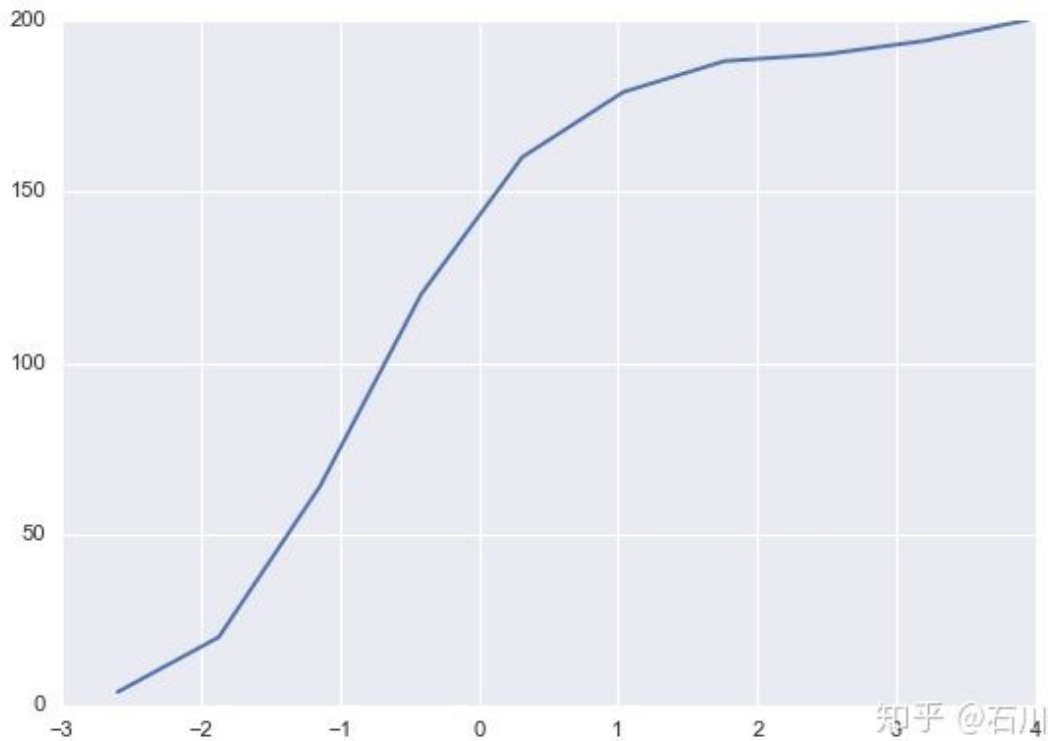
其中  $s_{\bar{x}}$  是  $\bar{x}$  的 standard error。

以均值为例, 可以通过下面的关系式将每个 Bootstrap 样本的均值转化为对应的 Bootstrapped  $t$  statistic (注: 如果研究的对象不是均值, 则 Bootstrapped  $t$  statistic 会出现不存在解析式的情况):



其中,  $\bar{x}_i^*$  和  $s_i^*$  分别为第  $i$  个 Bootstrap 样本的均值和标准差;  $n$  为样本大小。

仍以前面的例子说明这种方法如何计算  $\mu$  的置信区间。对于每个 Bootstrap 样本, 计算其 Bootstrapped t 统计量, 它们的累积密度函数为:



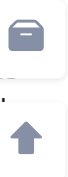
通过 Bootstrapped t 统计量很容易找到临界值 -1.17 和 1.81。因此,  $\mu$  的置信区间为: [31.82, 46.62]。在本例中, 本方法计算出的置信区间的范围大于前面两种方法。介绍这种方法的目的是为了给读者开拓思路。在实践中仍然推荐使用经验 Bootstrap 方法。

## 6 远远不止均值

到目前未知, 本文的例子中均已均值作为目标统计量, 这便于将不同的 Bootstrap 方法得到的置信区间进行比较。然而, Bootstrap 方法在计算置信区间时可以考虑各种传统方法无能为力的统计量。

下面就来看看中位数的例子。仍然以第三节中的十个数 (30, 37, 36, 43, 42, 48, 43, 46, 41, 42) 作为来自某个未知总体的一组样本。采用经验 Bootstrap 方法, 我们来计算中位数的 95% 的置信区间。

使用之前用到的 200 个 Bootstrap 样本, 可以得到中位数误差的临界值。由于考虑的是 95% 置信区间, 因此临界值为 2.5% 和 97.5% 分位对应的误差: -5.0 和 2.5。从原始数据易知, 样的中位数是 42。因此, 中位数的 95% 的置信区间为: [39.5, 47]。



知乎

首发于  
川流不息

本文介绍了如何使用 Bootstrap 技术计算参数估计的误差。Bootstrap 方法对总体分布不做假设，且可以被应用于我们感兴趣的各種统计量，这些特点使得它非常强大。当然，需要说明的是 **Bootstrap 中的重采样并不能够帮助我们改进点估计 (point estimate)**。以均值为例，原始样本均值  $\bar{x}$  就是总体均值  $\mu$  的点估计。我们使用重采样得到很多 Bootstrap 样本，并且得到很多 Bootstrap 样本均值  $\bar{x}^*$ ，则这些  $\bar{x}^*$  的平均值将会非常接近  $\bar{x}$ （事实上，可以证明  $E[\bar{x}^*]$  ——  $\bar{x}^*$  的期望 —— 就是  $\bar{x}$ ）。换句话说，对于点估计，Bootstrap 样本均值并不能比  $\bar{x}$  提供任何新的信息。但是，这些  $\bar{x}^*$  的取值对于估计  $\bar{x}$  如何围绕  $\mu$  变化非常有效，这便是我们在全文中反复强调的 Bootstrap 的核心。

在量化投资领域，Bootstrap 也有广泛的应用。例如，Bootstrap 可以用来对参数估计的偏差进行修正，比如投资品收益率之间的相关系数。投资品的历史收益率数据就是我们仅有的样本，通过重采样并利用经验 Bootstrap 方法，可以求出各种统计量的估计误差，这无疑有助于我们更好的构建投资策略，进行风险防控。又比如，简单的分类算法（比如分类树）可以用来进行选股，但是它对样本数据比较敏感，预测的方差较大。在这方面可以采用 Bootstrap 技巧作为元算法技术用于一般分类算法（比如结合 Bootstrap 和分类树得到的装袋算法），这可以明显地降低分类算法的方差，从而提高预测的准确性。在多因子模型中，Barra 在协方差矩阵的 Eigenfactor 调整时也用了 Bootstrap 思想。

最后，本文介绍的几种方法都属于无参数 Bootstrap 方法，即对总体分布不做任何假设。在一些应用中，如果能够明确总体分布的类型，也可以使用 Bootstrap 方法进行参数估计，这称之为参数化 Bootstrap 方法。比如，我们已知总体分布满足指数分布，但是不知道其参数  $\lambda$ 。这时，可以利用参数化 Bootstrap 方法计算出 Bootstrap 样本中  $\lambda^*$  的误差的分布，用它来估计  $\lambda$  的置信区间。由于空间有限，本文不再展开介绍。

(全文完)

**免责声明：**文章内容不可视为投资意见。市场有风险，入市需谨慎。

原创不易，请保护版权。如需转载，请联系获得授权，并注明出处，谢谢。已委托“维权骑士”（[维权骑士\\_免费版权监测/版权保护/版权分发](#)）为进行维权行动。

编辑于 2019-07-25

[Bootstrap\(statistics\)](#)   [参数估计](#)   [统计学](#)

▲ 赞同 68   ▼   ● 12 条评论   ➤ 分享   ★ 收藏   ...

文章被以下专栏收录

知乎

首发于  
川流不息

北京量信投资管理有限公司是一家在中国基金业协会备案登记的专业私募基金管理人...

八上卷一

## 推荐阅读

## 如何理解非参数检验？

数据描述的三个角度：集中趋势，离散程度和分布形态。而常用统计推断检验方法分为两大类：参数检验和非参数检验。它们两者有什么联系呢？参数检验通常是假设总体服从正态分布，样本统计量...

胡保强

## 参数估计中的Frequentist与Bayesian

近期在读MIT的概率论【1】，书上讲得深入浅出，觉得有点收获，于是在知乎尝试写出自己对参数估计的一些理解（其实就是总结书上的东西），个人实在是水平有限，加上是第一次尝试创作，有欠妥...

风清云

## 参数估计与

统计推断的分布，参数估计二、与假设检验估计1.估计指用样本统

社会小学生

## 12 条评论

切换为时间排序

写下你的评论...



Wu Kevin

1 年前

感谢分享，很有启发。我读的比较细致，有个细节：文中有一句“根据大数定理 (law of large numbers)，随着样本个数的增加， $\delta^{\star}$  的分布也越来越精确。”多抽样肯定没坏处，但是bootstrap存在这种渐进性质吗？I think it remains doubt

👍 2



石川 (作者) 回复 Wu Kevin

1 年前

非常感谢。我觉得你说的是对的。这句话不够准确，应该改为：根据大数定理，当有足够多样本数据的时候， $\delta^{\star}$  的分布会是  $\delta$  的分布的好的近似。

👍 1



balancex 回复 石川 (作者)

2 个月前

当有足够多样本时， $\delta^{\star}$  分布的方差会趋近于零吧，同时  $\delta^{\star}$  分布收敛到  $\delta$  的分布？是这个意思吗

👍 赞



杨柏伟

1 年前



知乎

首发于  
川流不息

赞



石川 (作者) 回复 杨柏伟

1 年前

这是个好问题，我们也采用了很多想法来评价策略的表现。不过还没试过你说的这个思路。:)

赞



杨柏伟 回复 石川 (作者)

1 年前

但单纯对收益率重采样，收益率序列之间的相关关系应该会被破坏掉了，所以对时序数据用bootstrap应该要考虑如何最大限度保留原有的相关性信息，对此，有什么好方法可以解决？

赞

展开其他 3 条回复



18856857963

1 年前

您好，请问你们做量化投资策略是做证券还是期货的？有固定合作的期货公司吗？

赞



石川 (作者) 回复 18856857963

1 年前

您好。证券和商品期货都做。有合作的期货公司。谢谢。

赞



萧嫣

10 个月前

我想问问，使用bootstrap可以模拟出原分布函数吗

赞

