

参数估计

导读

点估计

矩估计

极大似然估计

估计量的评选标准

区间估计

基本思路

导读

参数估计导读来自 [Junyi Hou](#).

点估计就是用一个数据（data）的函数（通常称为估计统计量，estimator）来给出一个未知参数的估计值。

即使是固定的参数真值（虽然我们不知道这个值），由于数据的随机性，不同的数据代入这个函数往往会得出不同的估计值（estimation）。所以我们往往在点估计的基础上包裹上一个邻域，即得到一个**区间估计**。

那么点估计周围的这个邻域的大小是怎么确定的呢？一个最直接的答案就是：确定一个百分比， $p\%$ ，使得给定任意数据集，参数的估计值（estimation）落在这个邻域内的概率为 $p\%$ 。那么，确定邻域大小的问题就变成了确定**参数估计量**（estimator）的分布的问题了。

首先，如果我们假设总体服从**正态分布**。那么可以证明，统计量作为随机变量的函数，往往会服从从正态分布中推导出来的一系列分布（如 t 分布，chi-square分布和F分布），那么通过统计量（estimator）的分布，我们可以很轻松的得到所求邻域的大小。

但是，在日常生活中，数据并不一定服从正态分布的。如果总体不是正态分布的，那么估计统计量（estimator）很可能也不服从正态分布， t 分布，chi-square 分布和 F 分布这些我们已知的分布。如果我们不知道统计量的分布，就无法确定应该给这个点估计包裹一个多大的邻域。

于是我们退而求其次，由于在满足一定正则条件的情况下，很多数据的分布都会在数据量趋近于无穷的情况下趋近于正态分布。比如，**中心极限定理**（Central Limit Theorem，以下简称 CLT）说的是样本均值的极限分布为正态分布。而估计量一般可以表示成样本均值的函数（e.g. OLS, GMM），所以知道了样本均值的极限（正态）分布也就知道了这些估计量的极限分布。于是我们就可以计算区间估计中的置信区间了。

最后，如果正则条件不满足，CLT 无法适用。数据分布即使在数据量趋于无穷的情况下仍然不是正态分布，这时候，采用传统方法得到区间估计的办法就行不通了。需要采用更加先进的方法（比如 bootstrapping 寻找区间估计；比如彻底抛弃 parametric model 转用 semi- non-parametric model 等等）。

其实 CLT 不单单在找区间估计的时候用到。很多**假设检验**的问题都依赖于统计量（或者数据等）的分布是正态分布这一假设。所以如果假设统计量本身就是正态的，那么当然可以以这些统计量为基础进行假设检验。但是如果分布不是正态的，那很有可能就需要 CLT 来帮助（至少建立在极限状态下的正态性）证明假设检验（包括区间估计）的正当性：因为如果统计量不是正态的，那么得出来的东西根本对不上号，假设检验也就没啥大意义了。

点估计

矩估计

在统计学中，矩估计（英语：method of moments）是估计总体参数的方法。首先推导涉及感兴趣的参数的总体矩（即所考虑的随机变量的幂的期望值）的方程。然后取出一个样本并从这个样本估计总体矩。接着使用样本矩取代（未知的）总体矩，解出感兴趣的参数。从而得到那些参数的估计。矩估计是英国统计学家卡尔·皮尔逊于1894年提出的。

假设问题是要估计表征随机变量 W 的分布 $f_W(w; \theta)$ 的 k 个未知参数 $\theta_1, \theta_2, \dots, \theta_k$ 。如果真实分布（“总体矩”）的前 k 阶矩可以表示成这些 θ 的函数：

$$\mu_1 \equiv E[W] = g_1(\theta_1, \theta_2, \dots, \theta_k),$$

$$\mu_2 \equiv E[W^2] = g_2(\theta_1, \theta_2, \dots, \theta_k),$$

\vdots

$$\mu_k \equiv E[W^k] = g_k(\theta_1, \theta_2, \dots, \theta_k).$$

设取出一大小为 n 的样本，得到 w_1, \dots, w_n 。对于 $j = 1, \dots, k$ ，令

$$\hat{\mu}_j = \frac{1}{n} \sum_{i=1}^n w_i^j$$

为 j 阶样本矩，是 μ_j 的估计。 $\theta_1, \theta_2, \dots, \theta_k$ 的矩估计量记为 $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ ，由这些方程的解（如果存在）定义：

$$\begin{aligned}\hat{\mu}_1 &= g_1(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k), \\ \hat{\mu}_2 &= g_2(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k), \\ \hat{\mu}_k &= g_k(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k).\end{aligned}$$

极大似然估计

在统计学中，最大似然估计（英语：maximum likelihood estimation，缩写为MLE），也称极大似然估计、最大概似估计，是用来估计一个概率模型的参数的一种方法。

给定一个概率分布 D ，已知其概率密度函数（连续分布）或概率质量函数（离散分布）为 f_D ，以及一个分布参数 θ ，我们可以从这个分布中抽出一个具有 n 个值的采样 X_1, X_2, \dots, X_n ，利用 f_D 计算出其似然函数：

$$L(\theta | x_1, \dots, x_n) = f_{\theta}(x_1, \dots, x_n).$$

若 D 是离散分布， f_{θ} 即是在参数为 θ 时观测到这一采样的概率。若其是连续分布， f_{θ} 则为 X_1, X_2, \dots, X_n 联合分布的概率密度函数在观测值处的取值。一旦我们获得 X_1, X_2, \dots, X_n ，我们就能求得一个关于 θ 的估计。最大似然估计会寻找关于 θ 的最可能的值（即，在所有可能的 θ 取值中，寻找一个值使这个采样的“可能性”最大化）。从数学上来说，我们可以在 θ 的所有可能取值中寻找一个值使得似然函数取到最大值。这个使可能性最大的 $\hat{\theta}$ 值即称为 θ 的最大似然估计。由定义，最大似然估计是样本的函数。

估计量的评选标准

无偏性，有效性，相合性。

区间估计

基本思路

1. 如果已知总体分布为正态分布，那么从抽样分布一节我们知道，其某些样本统计量满足正态分布，卡方分布， t 分布和 F 分布。

2. 知道样本统计量的分布对我们来说是非常有帮助的，即使在我们不知道分布参数的情况下。比如我们知道了总体 $X \sim N(\mu, \sigma^2)$ ，那么样本均值 $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ ，如果我们知道 σ 和 n ，那么只有一个未知参数 μ 不知道，这个 μ 也是我们要估计的未知参数。

3. 由 3σ 原则我们知道，如果我们得到一个样本，并求出样本均值的观察值 \bar{X}_0 ，那么

$$P\left(\mu - 2\frac{\sigma}{\sqrt{n}} < \bar{X}_0 < \mu + 2\frac{\sigma}{\sqrt{n}}\right) \approx 95\%$$

$$P\left(\bar{X}_0 - 2\frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_0 + 2\frac{\sigma}{\sqrt{n}}\right) \approx 95\%$$

这个式子的意思就是：我们要估计的未知参数 μ 在 $\left(\bar{X}_0 - 2\frac{\sigma}{\sqrt{n}}, \bar{X}_0 + 2\frac{\sigma}{\sqrt{n}}\right)$ 这个区间内的概率约等于 95%。

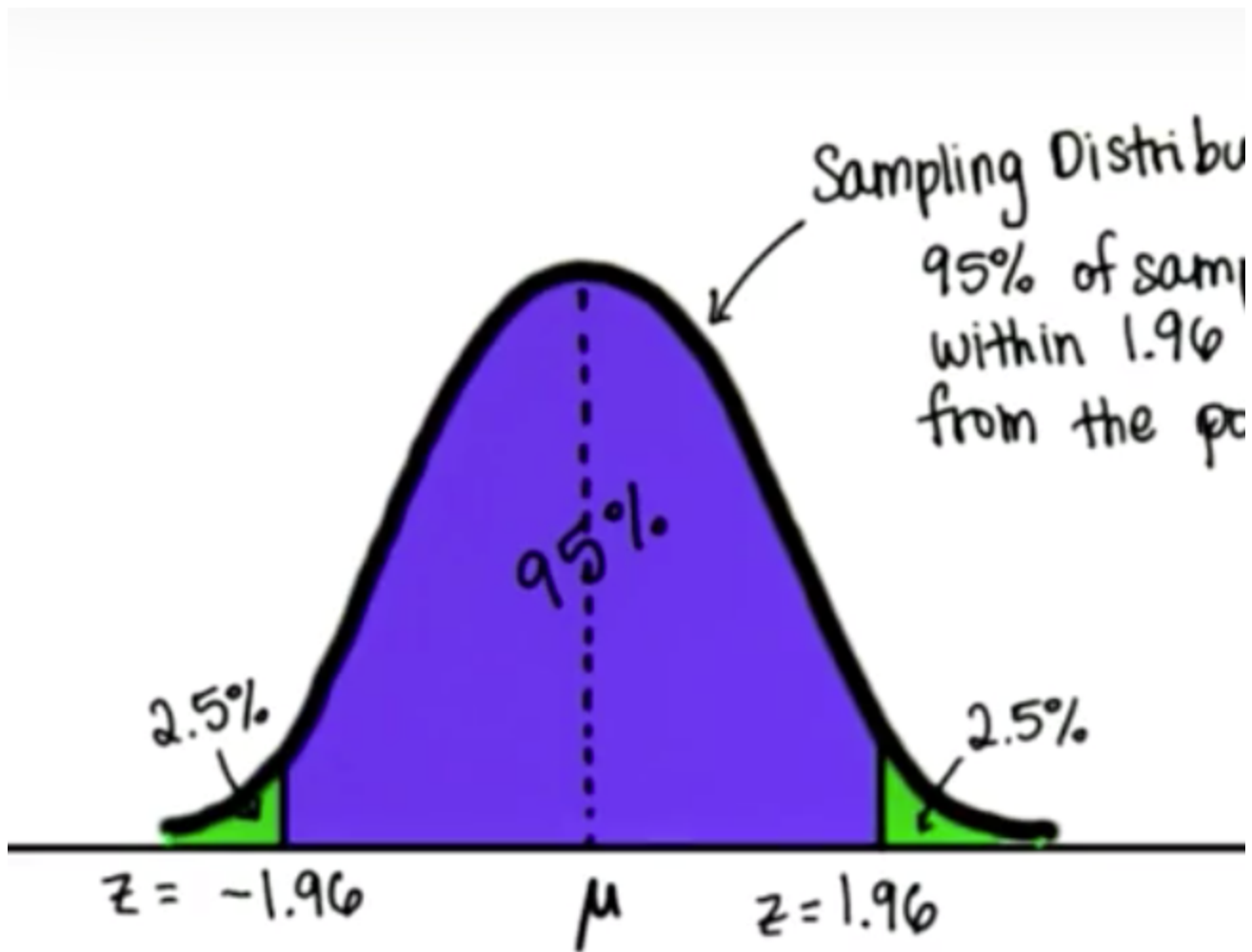
4. 3 中的结论，我们可以换一种等价的说法或思考的角度，也就是说：如果我们要求未知参数的真值在某个区间内的概率为 95%，那么这个区间就得是 $\left(\bar{X}_0 - 2\frac{\sigma}{\sqrt{n}}, \bar{X}_0 + 2\frac{\sigma}{\sqrt{n}}\right)$ 。这个 95% 叫做置信水平 $1 - \alpha$ ， α 为弃真错误，得到的区间成为置信区间（Confidence interval），区间的上下限称为置信上限和置信下限。

5. 但是 3σ 原则毕竟有局限性，我们也不能每次估计时都要求是 95% 的置信水平，所以为了更精确地估计，我们需要知道置信水平和置信区间的具体关系。

6. 我们回到 3 中，当我们知道 $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ ， \bar{X} 是一个统计量，我们可以进一步知道

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1), \quad \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

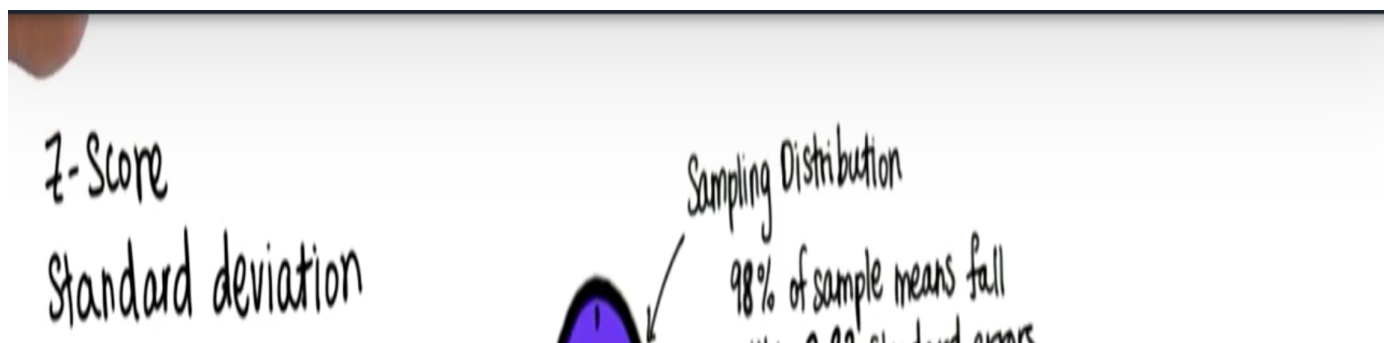
是一个我们构造的统计量 z ，因为其分布是更加简单的单位正态分布，所以我们用它来分析。其概率密度图如下：



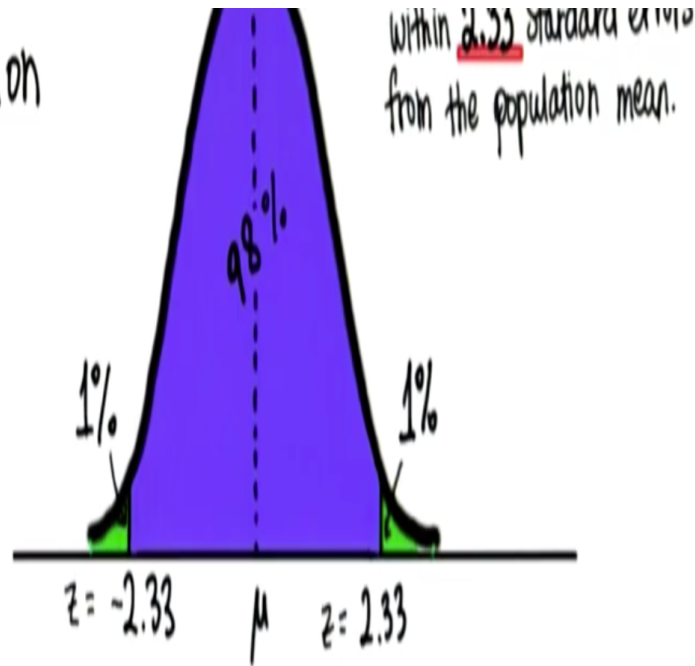
7. 那么当我们有了置信水平 α 以及 z 的定义，我们可以得到一个类似的等式：

$P\left(-z_{\frac{\alpha}{2}} < z < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$ ，这个等式就将 α 和 z 联系起来，将 $z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ 代入该式，并做等价变换可得 $P\left(\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$ ，这样在 α, \bar{X}, σ ，和 n 都已知的情况下，用这四个量表示出了未知参数 μ 的置信区间。

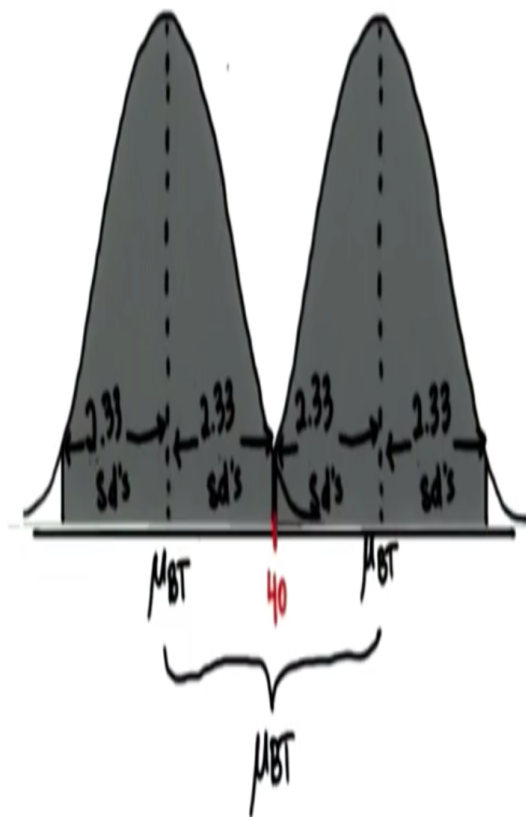
8. 整理思路，我们如果知道总体分布为正态分布，并且知道其标准差 σ 和样本容量 n ，当我们给出置信区间 $1 - \alpha$ 情况下，利用得到的一个样本计算出其样本均值 \bar{X} ，我们通过构造统计量 z 结合其分布性质，就可以得到未知参数 μ 的置信区间，也就是对 μ 进行了区间估计：



sampling distribution
standard error



± 2.33 are the
CRITICAL VALUES OF
Z for 98%
Confidence.



± 1.96 are the
CRITICAL VALUES OF
Z for 95%
confidence.

9. 以上是我们对特定总体情况的区间估计，这个特定情况是总体服从正态分布且已知方差，但是，实际中可能我们并不知道方差，或者我们知道均值想估计方差，亦或是均值和方差都不知道，甚至总体可能都

不服从正态分布，这个时候我们就不能通过样本均值这个统计量来进行参数估计了，就需要通过其他一些统计量，比如 χ^2, t 等等。但我们有[区间估计与假设检验表](#)，通过这个表我们可以快速判断自己应该使用哪个统计量以及估计或统计方法，这是十分方便的。

10. 样本容量越大，置信区间越窄；置信度越高，置信区间越宽。