# 有关无偏统计量的解释

此pdf用以解释无偏统计量的概念，大部分选课的同学还没有学过概率统计，所以此pdf从比较基础的部分开始讲起，也会涉及到Lecture4中有关期望和方差的数学推导。同时也会证明为什么样本方差的计算方法是无偏的，但是df=n-1的样本标准差的计算方法是有偏的。

## 有关期望的理解

### 期望的定义

期望是针对随机变量而言的一个量，可以理解成是站在"上帝视角"的值，它是一种概率论的概念，是一个数学的特征，首先我们给出其定义公式：

离散型随机变量$X$的取值为$X_1, X_2, X_3, \ldots, X_n$，$X$取对应值的概率$P(X_1), P(X_2), P(X_3), P(X_n)$，

$$E(X) = X_1 * P(X_1) + X_2 * P(X_2) + \ldots + X_n * P(X_n) = \sum_{k=1}^{n} X_k * P(X_k) \tag{1}$$

### 期望与均值之间的联系

均值，其实是针对实验观察到的特征而言，比如实验结果得到了$x_1$,$x_2$,$x_3$,...,$x_n$这n个值，那么均值则是$m = 1/n * (x_1 + \ldots + x_n)$。比如我们进行掷骰子，掷了六次，点数分别为1，1，1，3，3，3，这六次的观察就是我们的**样本**，于是我们可以说 **均值**为(1+1+1+3+3+3)/6=2，但这并不意味着掷骰子，每次掷骰子点数期望值为2。因为在这里并未考虑到掷一次骰子得到不同点数的概率。所以**均值**仅仅是针对观测值而言。

但这并不意味着期望与均值毫无联系。当我们掷骰子的次数逐渐增多，会发现我们观察到的样本里，不同点数出现的次数(频率,frequency)比较接近(假设这个骰子是均匀的，那么每一面朝上的概率都是相等的)，样本量越大，频率可能约接近概率(这一点也叫作大数定理)。在此时计算均值时，会发现样本的均值和真实的数学期望比较接近。比如1,2,3,4,5,6朝上的次数分别为1000,1010,999,995,992,1004，此时**均值**是:

$$m = (1 * 1000 + 2 * 1010 + 3 * 999 + 4 * 995 + 5 * 992 + 6 * 1004)/6000 = 3.497$$

而掷骰子的数学期望则是:

$$E(X) = 1 * 1/6 + 2 * 1/6 + \ldots + 6 * 1 * 6 = 3.5$$

二者之间的差值已经很小，随着样本数目的增大，样本均值会比较接近随机变量的数学期望。所以可以认为**数学期望是样本均值在样本量趋于无穷大的极限**

## 有关方差的理解

在概率论的视角，假设一个随机变量X的期望值为$\mu$，则方差的定义式为：

$$Var(X) = E(X-u)^2 = E(X^2) - [E(X)]^2 \tag{2}$$

有关第二个等号的证明过程，大家可以看一下Lecture4中相关的证明：

$$\boxed{Var(X) = E(X-\mu)^2 = E(X^2) - [E(X)]^2}$$

### Derivation

$$Var(X) = \sum_{i=1}^{k}(X_i-\mu)^2 \Pr(X=x_i) = \sum_{i=1}^{k}(x_i^2 - 2\mu x_i + \mu^2)\Pr(X=x_i)$$

$$= \sum_{i=1}^{n} x_i^2 \Pr(X=x_i) - 2\mu\sum_{i=1}^{n} x_i \Pr(X=x_i) + \sum_{i=1}^{n}\mu^2 \Pr(X=x_i)$$

$$= \sum_{i=1}^{n} x_i^2 \Pr(X=x_i) - 2\mu^2 + \mu^2 = \sum_{i=1}^{n} x_i^2 \Pr(X=x_i) - \mu^2$$

只不过，在实际抽样问题时，我们对于一个样本倾向于计算方差减去的是样本的均值。

## 对于无偏统计量的理解

有了以上对于期望和方差一些数学理解，我们来介绍什么是无偏估计，并利用数学推导证明以下三个结论：

- 样本均值是总体均值的无偏估计量
- 样本方差(df=n-1)是总体方差的无偏估计量
- 利用df=n-1算得的样本标准差是总体标准差的有偏估计量

**什么是无偏估计**：无偏估计量是用样本统计量来估计总体参数时的一种无偏推断。**样本估计量的数学期望**等于**被估计参数的真实值**，则成此估计量为被估计参数的无偏估计。具体可见：无偏估计_百度百科 (baidu.com)，如果样本统计量的数学期望不等于被估计参数的真实值，则我们认为其实有偏的。下面我们分别证明上述提到的三个结论：

假设对于总体，其随机变量X的数学期望为$\mu$,方差为$\sigma^2$，我们从中抽样，$X_1, X_2, .., X_n$分别为样本中不同数据点的值，彼此之间独立(注意这里的$X_i$和期望定义的$X_i$指的不是一个内容，期望定义时，$X_i$表示的是可以取的不同值，而此时定义的$X_i$表示的是每个观测点的值，也可以认为我们有n个随机变量X。

**证明1：样本均值是总体均值的无偏估计量**
对于一个样本，我们有：

$$m = \frac{\sum_{k=1}^{n} X_i}{n} \tag{3}$$

我们想要说明m是无偏估计量，只需证明：

$$E(m) = \mu \tag{4}$$

下面即为证明：

$$E(m) = E(\frac{\sum_{k=1}^{n} X_i}{n}) = \frac{1}{n} * (E(X_1) + ... + E(X_n)) \tag{5}$$

又因为对于每一个$X_i$，都有$E(X_i) = E(X) = \mu$，所以上式转化为：

$$E(m) = \frac{1}{n} * (n * \mu) = \mu \tag{6}$$

所以我们认为样本均值是总体均值的无偏估计

**证明2：样本方差是总体方差的无偏估计量**
对于一个样本，假设样本均值为$m$，我们有：

$$s^2 = \frac{\sum_{k=1}^{n} (X_i - m)^2}{n-1} \tag{7}$$

我们想要说明$s^2$是无偏估计量，只需证明：

$$E(s^2) = \sigma \tag{8}$$

下面即为证明:

$$
\begin{aligned}
E(s^2) &= E(\frac{\sum_{k=1}^{n}(X_i-m)^2}{n-1}) \\
&= \frac{1}{n-1}E(\sum_{k=1}^{n}(X_i^2-2mX_i+m^2)) \\
&= \frac{1}{n-1}E(\sum_{k=1}^{n}X_i^2-nm^2)(\text{因为}\sum X_i = n*m) \\
&= \frac{1}{n-1}(\sum_{k=1}^{n}E(X_i^2)-nE(m^2))
\end{aligned} \tag{9}
$$

而由上述(2)，$Var(X)=E(X-u)^2=E(X^2)-[E(X)]^2$，我们有：

$$
\begin{aligned}
E(X_i^2) &= Var(X_i)+[E(X_i)]^2=\sigma^2+\mu^2 \\
E(m^2) &= Var(m)+[E(m)]^2=\frac{\sigma^2}{n}+\mu^2 \\
Var(m) &= \frac{\sigma^2}{n}\text{证明见}(11)
\end{aligned} \tag{10}
$$

带入(9)，我们有：

$$
\begin{aligned}
E(s^2) &= \frac{1}{n-1}(\sum_{k=1}^{n}E(X_i^2)-nE(m^2)) \\
&= \frac{1}{n-1}(n*\sigma^2+n*\mu^2-n*(\frac{\sigma^2}{n}+\mu^2)) \\
&= \sigma^2
\end{aligned} \tag{10}
$$

所以我们证明了，有关样本方差的确是总体方差的无偏统计量。下面补充(11)式:

$$
\begin{aligned}
Var(m) &= Var(\frac{\sum_{k=1}^{n}X_i}{n}) \\
&= \frac{1}{n}Var(X) \\
&= \frac{\sigma^2}{n}
\end{aligned} \tag{11}
$$

**证明3：利用df=n-1算得的样本标准差是总体标准差的有偏估计量**

由(2)(10)知，

$$
E(s^2)=Var(s^2)+[E(s)]^2=\sigma^2 \tag{12}
$$

由此我们有，

$$
E(s)=\sqrt{\sigma^2-Var(s^2)} \tag{13}
$$

因为$Var(s^2)\geq 0$(每次抽样时会存在一定的误差，如果数据点不全都一致的话，显然大于0)，

我们有:

$$E(s) \leq \sqrt{\sigma^2} = \sigma \tag{14}$$

所以我们使用df=n-1时进行计算，得到的结果总是会系统性的偏小，因此df=n-1方法计算得到的标准差是总体标准差的有偏估计量。