为什么要将样本方差除以 N-1?

Vincent Spruyt

2014/03

Abstract

在本文中，我们解释了在计算正态分布数据的样本方差时，为什么以及何时被 N 或被 $N-1$ 除。我们推导了用极大似然参数估计法计算平均值和方差的公式。

TL;DR

在本文中，我们将导出计算正态分布数据的均值和方差的著名公式，以回答文章标题中的问题。然而，对于那些对这个问题的“为什么”不感兴趣而只对“何时”感兴趣的读者来说，答案很简单：

如果必须同时估计数据的平均值和方差 (通常是这种情况)，则除以 $N-1$，得到的方差为：

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^2$$

另一方面，如果已知真实总体的平均值，因此只需要估计方差，则除以 N，从而得到方差：

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$$

前者是你通常需要的，后者的一个例子是高斯白噪声的传播估计。由于已知高斯白噪声的均值为零，在这种情况下只需估计方差。

Contents

1 序言

如果数据是正态分布的，我们可以用它的均值 μ 和方差 σ^2 来完全描述它。方差是表示每个数据点与平均值的平均偏差的标准偏差的平方。换句话说，方差表示数据的扩散。对于正态分布数据，68.3% 的观测值将在 $\mu-\sigma$ 和 $\mu+\sigma$ 之间。下图说明了这一点，它显示了平均值 $\mu = 10$ 和方差 $\sigma^2 = 3^2 = 9$ 的高斯密度函数：

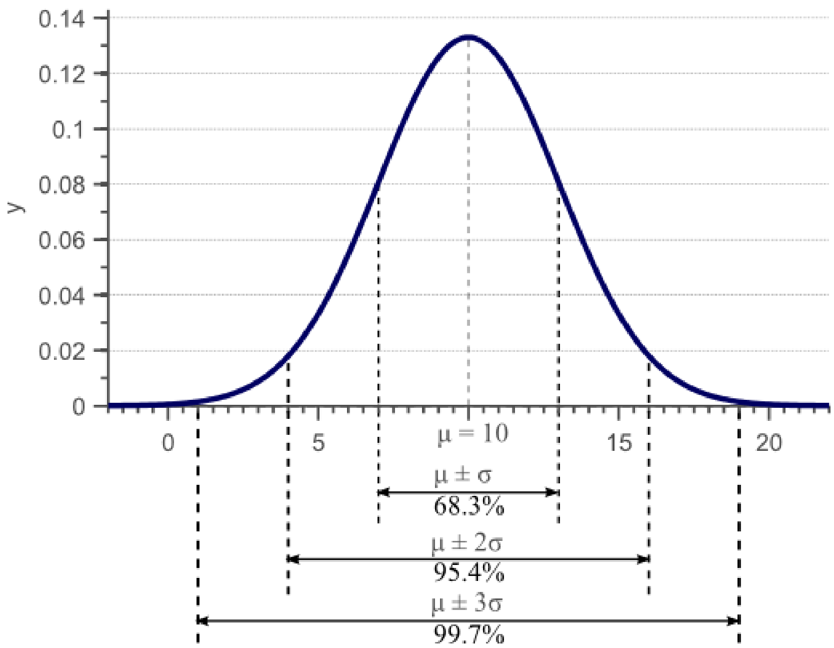

图 1：高斯密度函数。对于正态分布的数据，68% 的样本落在均值加上和减去标准差定义的区间内。

通常我们无法访问完整的数据。在上面的例子中，我们通常会有一些观察结果可供我们使用，但我们无法访问定义图的 x 轴的所有可能的观察结果。例如，我们可能有以下一组观察结果：

Table 1

Observation ID	Observed Value
Observation 1	10
Observation 2	12
Observation 3	7
Observation 4	5
Observation 5	11

如果我们现在计算经验平均值，将所有值相加，除以观察值的数量，我们得到：

$$\mu = \frac{10 + 12 + 7 + 5 + 11}{5} = 9. \tag{1}$$

通常我们假设经验平均值接近分布的实际未知平均值，因此假设观测数据是从平均值为 $\mu = 9$ 的高斯分布中取样的。在这个例子中，分布的实际平均值是 10，所以经验平均值确实接近实际平均值。

数据方差计算如下：

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2 = \frac{(10-9)^2 + (12-9)^2 + (7-9)^2 + (5-9)^2 + (11-9)^2}{4}) = 8.5 \tag{2}$$

同样，我们通常假设这个经验方差接近于潜在分布的真实和未知方差。在这个例子中，实际方差是 9，所以实际上经验方差接近实际方差。

现在的问题是，为什么用来计算经验平均数和经验方差的公式是正确的。实际上，另一个常用的计算方差的公式定义如下：

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 = \frac{(10-9)^2 + (12-9)^2 + (7-9)^2 + (5-9)^2 + (11-9)^2}{5}) = 6.8 \tag{3}$$

方程 (2) 和 (3) 的唯一区别是前者除以 $N-1$，而后者除以 N。两个公式实际上都是正确的，但何时使用哪一个取决于具体情况。

在下面几节中，我们将完全推导出最佳逼近正态分布的未知方差和均值的公式，并给出该分布的几个样本。我们将展示在哪些情况下将方差除以 N，在哪些情况下将方差归一化为 $N-1$。

近似一个参数 (均值或方差) 的公式称为估计值。在下面，我们将用 $\hat{\mu}$ 和 $\hat{\sigma}^2$ 表示分布的实参数和未知参数。估计值，例如经验平均值和经验方差，用 μ 和 σ^2 表示。

为了找到最优估计值，我们首先需要一个观察特定数据点 x_i 的可能性的解析表达式，假设总体是正态分布的，具有给定的平均值 μ 和标准偏差 σ。具有已知参数的正态分布通常表示为 $N(\mu, \sigma^2)$。似然函数为：

$$x_i \sim N(\mu, \sigma^2) \Rightarrow P(x_i; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}. \tag{4}$$

为了计算均值和方差，我们显然需要从这个分布中提取多个样本。在下面，让向量 $\vec{x} = (x_1, x_2, ...x_N)$ 是包含所有可用样本的向量 (例如，表 1 中示例中的所有值)。如果所有这些样本在统计上是独立的，我们可以把它们的联合似然函数写为所有个体可能性的总和：

$$\begin{aligned} P(\vec{x}; \mu, \sigma^2) &= P(x_1, x_2, ..., x_N; \mu, \sigma^2) \\ &= P(x_1; \mu, \sigma^2) P(x_2; \mu, \sigma^2) ... P(x_N; \mu, \sigma^2) \\ &= \textstyle\prod_{i=1}^N P(x_i; \mu, \sigma^2) \end{aligned} \tag{5}$$

将方程 (4) 代入方程 (5)，然后得出该联合概率密度函数的解析表达式：

$$\begin{aligned} P(\vec{x}; \mu, \sigma^2) &= \textstyle\prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^N (x_i-\mu)^2} \end{aligned} \tag{6}$$

方程 (6) 在下一节中将很重要，并将用于导出高斯分布的均值和方差的估计值的众所周知的表达式。

2 最小方差，无偏估计

为了确定一个估计值是否是一个“好”估计值，我们首先需要定义什么是“好”估计值。估计值的优度取决于两个度量，即其偏差和方差 (是的，我们将讨论均值估计值的方差和方差估计值的方差)。这两项措施将在本节中简要讨论。

2.1 参数偏差

想象一下，我们可以得到不同的 (不相交的) 子集的完整人口。与前面的例子类似，想象一下，除了表 1 中的数据，我们还有一个表 2 和一个表 3，其中有不同的观测值。那么一个好的均值估计值，就是一个平均值等于实际均值的估计值。尽管我们可以接受这样一种观点，即一个子集数据的经

验平均值不等于我们示例中的实际平均值，但一个好的估计员应该确保所有子集中估计平均值的平均值等于实际平均值。这个约束在数学上表示为估计值的期望值应等于实际参数值：

$$\begin{aligned} E[\mu] &= \hat{\mu} \\ E[\sigma^2] &= \hat{\sigma^2} \end{aligned} \tag{7}$$

如果上述条件成立，则估计值称为“无偏估计值”。如果条件不成立，估计被称为“有偏”，因为平均而言，他们要么低估或高估了参数的真实值。

2.2 参数方差

无偏估计保证了它们平均得到的估计值等于实参数。然而，这并不意味着每一个估计都是一个好的估计。例如，如果实际均值是 10，则无偏估计值可以估计一个总体子集的均值为 50，另一个子集的均值为 -30。估计值的期望值实际上是 10，这等于实际参数，但估计值的质量显然也取决于每个估计值的分布。一个对 5 个不同的人口子集产生估计 (10，15，5，12，8) 的估计值是无偏的，就像一个产生估计 (50，-30，100，-90，10) 的估计值一样。然而，第一个估计值的所有估计都比第二个估计值的估计更接近真值。

因此，一个好的估计值不仅有一个低偏差，而且产生一个低方差。该方差表示为估计值的均方误差：

$$\begin{aligned} Var(\mu) &= E[(\hat{\mu}-\mu)^2] \\ Var(\sigma^2) &= E[(\hat{\sigma}-\sigma)^2] \end{aligned}$$

因此，一个好的估计是一个低偏差，低方差估计。最优估计值，如果存在这样的估计值，则是一个没有偏差和方差低于任何其他可能的估计值。这种估计称为最小方差无偏 (minimum variance, unbiased, MVU) 估计。在下一节中，我们将导出高斯分布的均值和方差估计的解析表达式。我们将证明正态分布方差的 MVU 估计要求我们在某些假设下将方差除以 N，如果这些假设不成立，则要求我们除以 $N-1$。

3 最大似然估计

尽管可以使用许多技术来获得基于总体数据子集的参数估计值，但最简单的方法可能是最大似然 (maximum likelihood) 法。

方程 (6) 将 $\vec{x}$ 的观测概率定义为 $P(\vec{x};\mu,\sigma^2)$。如果我们在这个函数中固定 μ 和 σ^2，同时让 $\vec{x}$ 变化，我们就得到了如图 1 所示的高斯分布。但是，

我们也可以选择一个固定的 $\vec{x}$，让 μ 和/或 σ 变化。例如，我们可以选择 $\vec{x} = (10, 12, 7, 5, 11)$，就像前面的例子一样。我们也选择一个固定的 $\mu = 10$，让 σ^2 变化。图 2 显示了使用建议的固定的向量 $\vec{x}$ 和 μ，σ^2 的每个不同值的分布的曲线图：

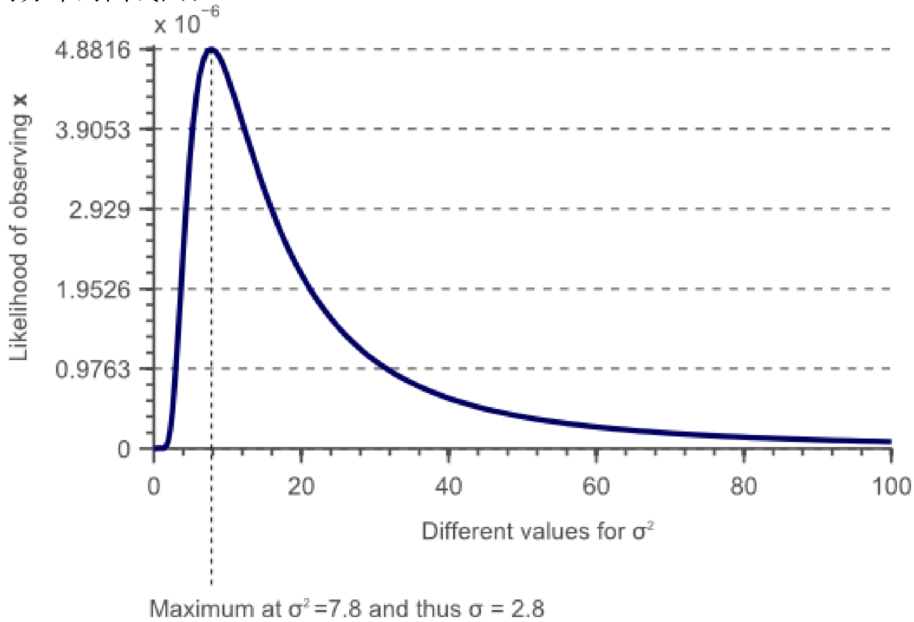

图 2：此图显示了观察固定数据 $\vec{x}$ 的可能性，如果数据是正态分布的，选择固定的 $\mu = 10$，与变化的 σ^2 的各种值作图。

在上图中，我们通过改变固定的 $\mu = 10$ 的 σ^2 来计算可能性 $P(\vec{x}; \sigma^2)$。结果曲线中的每个点表示观测 $\vec{x}$ 是参数为 σ^2 的高斯分布样本的可能性。对应于最高可能性的参数值很可能就是定义数据来源的分布的参数。因此，我们可以通过在这个似然曲线中找到最大值来确定最佳的 σ^2。在本例中，最大值为 $\sigma^2 = 7.8$，因此标准偏差为 $\sqrt{(\sigma^2)} = 2.8$。实际上，如果我们用传统的方法计算方差，给定的 $\mu = 10$，我们会发现它等于 7.8：

$$\frac{(10-10)^2 + (12-10)^2 + (7-10)^2 + (5-10)^2 + (11-10)^2}{5} = 7.8$$

因此，通过求最大似然函数的峰值，简单地导出了基于样本数据的方差计算公式。此外，我们不固定 μ，而是让 μ 和 σ^2 同时变化。找到这两个估计值对应于在二维似然函数中找到最大值。

为了求函数的最大值，我们只需把它的导数设为零。如果我们想求一个有两个变量的函数的最大值，我们计算对每个变量的偏导数，并把它们都设为零。在下面，让 $\hat{\mu}_{ML}$ 作为使用最大似然方法获得的总体平均数的最优估计值，并且让 $\hat{\sigma}^2_{ML}$ 作为方差的最优估计值。为了使似然函数最大化，我们

只需计算其 (偏) 导数并将其设置为零，如下所示：

$$\hat{\mu}_{ML} = \arg\max_{\mu} P(\vec{x};\mu,\sigma^2) \Rightarrow \frac{\partial P(\vec{x};\mu,\sigma^2)}{\partial \mu} = 0$$

并且

$$\hat{\sigma}^2_{ML} = \arg\max_{\sigma^2} P(\vec{x};\mu,\sigma^2) \Rightarrow \frac{\partial P(\vec{x};\mu,\sigma^2)}{\partial \sigma^2} = 0$$

在下面的段落中，我们将使用这种技术来获得 $\hat{\mu}$ 和 $\hat{\sigma}$ 的 MVU 估计值。我们考虑两种情况：

第一种情况假设分布的真实平均值是已知的。因此，我们只需要估计方差，然后问题就对应于在一维似然函数中求最大值，用 σ^2 参数化。虽然这种情况在实践中并不经常发生，但它确实有实际应用价值。例如，如果我们知道一个信号 (例如图像中像素的颜色值) 应该有一个特定的值，但是该信号已经被白噪声 (均值为零的高斯噪声) 污染，那么分布的均值是已知的，我们只需要估计方差。

第二种情况涉及真实均值和真实方差都未知的情况。这是您遇到最多的情况，您将根据样本数据获得均值和方差的估计值。

在接下来的段落中，我们将说明每种情况都会产生不同的 MVU 估计值。更具体地说，第一种情况要求方差估计值被 N 归一化为 MVU，而第二种情况要求被 $N-1$ 除为 MVU。

4 如果均值已知求估计方差

4.1 参数估计

如果分布的真实平均值已知，则似然函数仅在 σ^2 上参数化。然后，获得最大似然估计值对应于求解：

$$\hat{\sigma}^2_{ML} = \arg\max_{\sigma^2} P(\vec{x};\sigma^2) \tag{8}$$

然而，由于函数中的指数，计算由方程 (6) 定义的 $P(\vec{x};\sigma^2)$ 的导数相当复杂。实际上，最大化对数似然函数要比最大化似然函数本身容易得多。因为对数是单调函数，所以最大值是相同的。因此，我们解决以下问题代替：

$$\hat{\sigma}^2_{ML} = \arg\max_{\sigma^2} \log(P(\vec{x};\sigma^2)). \tag{9}$$

下面我们设置 $s=\sigma^2$ 以获得一个更简单的表示法。为了求对数似然函数的最大值，我们只需计算方程 (6) 对数的导数并将其设为零：

$$
\begin{aligned}
&\frac{\partial \log(P(\vec{x};\sigma^2))}{\partial \sigma^2} = 0\\
&\Leftrightarrow \frac{\partial \log(P(\vec{x};s))}{\partial s} = 0\\
&\Leftrightarrow \frac{\partial}{\partial s}\log\left(\frac{1}{(2\pi s)^{\frac{N}{2}}}e^{-\frac{1}{2s}\sum_{i=1}^{N}(x_i-\mu)^2}\right) = 0\\
&\Leftrightarrow \frac{\partial}{\partial s}\log\left(\frac{1}{(2\pi)^{\frac{N}{2}}}\right) + \frac{\partial}{\partial s}\log\left(\frac{1}{\sqrt{(s)}^{\frac{N}{2}}}\right) + \frac{\partial}{\partial s}\log\left(e^{-\frac{1}{2s}\sum_{i=1}^{N}(x_i-\mu)^2}\right) = 0\\
&\Leftrightarrow \frac{\partial}{\partial s}\log\left((s)^{-\frac{N}{2}}\right) + \frac{\partial}{\partial s}\left(-\frac{1}{2s}\sum_{i=1}^{N}(x_i-\mu)^2\right) = 0\\
&\Leftrightarrow -\frac{N}{2}\frac{\partial}{\partial s}\log(s) - \frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^2\frac{\partial}{\partial s}\left(\frac{1}{s}\right) = 0\\
&\Leftrightarrow -\frac{N}{2s} + \frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^2\left(\frac{1}{s^2}\right) = 0\\
&\Leftrightarrow \frac{N}{2s^2}\left(-s + \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2\right) = 0\\
&\Leftrightarrow \frac{N}{2s^2}\left(\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2 - s\right) = 0
\end{aligned}
$$

很明显，如果 $N > 0$，则上述问题的唯一可能解决方案是：

$$s = \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2. \tag{10}$$

注意，这个 $\hat{\sigma}$ 的最大似然估计确实是计算正态数据方差的传统公式。规范化因子是 $\frac{1}{N}$。

然而，最大似然法不能保证提供无偏估计。另一方面，如果得到的估计值是无偏的，那么最大似然方法确实保证了估计值也是最小方差，因此 MVU。因此，我们需要检查方程 (10) 中的估计值是否无偏。

4.2 效果评估

为了检验方程 (10) 定义的估计值是否无偏，我们需要检验方程 (7) 的条件是否成立，从而

$$E[s] = \hat{s}.$$

为此，我们将方程 (10) 插入 $E[s]$，并写为：

$$
\begin{aligned}
E[s] = E\left[\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2\right] &= \frac{1}{N}\sum_{i=1}^{N}E\left[(x_i-\mu)^2\right] = \frac{1}{N}\sum_{i=1}^{N}E\left[x_i^2-2x_i\mu+\mu^2\right] \\
&= \frac{1}{N}\left(NE[x_i^2]-2N\mu E[x_i]+N\mu^2\right) \\
&= \frac{1}{N}\left(NE[x_i^2]-2N\mu^2+N\mu^2\right) \\
&= \frac{1}{N}\left(NE[x_i^2]-N\mu^2\right)
\end{aligned}
$$

此外,方差的一个重要性质是,真实的方差 $\hat{s}$ 可以写为 $\hat{s}=E[x_i^2]-E[x_i]^2$ 使得满足 $E[x_i^2]=\hat{s}+E[x_i]^2=\hat{s}+\mu^2$ 的条件。在上述公式中使用此属性可得出：

$$
\begin{aligned}
E[s] &= \frac{1}{N}\left(NE[x_i^2]-N\mu^2\right) \\
&= \frac{1}{N}\left(N\hat{s}+N\mu^2-N\mu^2\right) \\
&= \frac{1}{N}\left(N\hat{s}\right) \\
&= \hat{s}
\end{aligned}
$$

由于 $E[s]=\hat{s}$，方程 (7) 所示的条件成立，因此所获得的数据方差 $\hat{s}$ 的估计值是无偏的。此外，由于最大似然方法保证了无偏估计也是最小方差 (MVU)，这意味着没有其它估计值比这里得到的更好。

因此，在计算正态分布数据的方差时，如果已知基础分布的真实平均值，则必须除以 N 而不是 $N-1$。

5 如果均值未知求估计方差

5.1 参数估计

在上一节中，分布的真实平均值是已知的，因此我们只需找到数据方差的估计值。然而，如果真平均数未知，则也必须找到平均数的估计值。此外，方差估计使用这个均值估计。因此，我们将证明先前得到的方差估计不再是无偏的。此外，我们将证明，在这种情况下，我们可以通过除以 $N-1$ 而不是除以 N 来“无偏”估计值，这稍微增加了估计值的方差。

和以前一样，我们使用极大似然方法来获得基于对数似然函数的估计值。我们首先找到 $\hat{\mu}$ 的 ML 估计值：

$$\begin{aligned}
&\frac{\partial \log(P(\vec{x}; s, \mu))}{\partial \mu} = 0 \\
&\Leftrightarrow \frac{\partial}{\partial \mu} \log\left(\frac{1}{(2\pi s)^{\frac{N}{2}}} e^{-\frac{1}{2s}\sum_{i=1}^{N}(x_i-\mu)^2}\right) = 0 \\
&\Leftrightarrow \frac{\partial}{\partial \mu} \log\left(\frac{1}{(2\pi)^{\frac{N}{2}}}\right) + \frac{\partial}{\partial \mu} \log\left(e^{-\frac{1}{2s}\sum_{i=1}^{N}(x_i-\mu)^2}\right) = 0 \\
&\Leftrightarrow \frac{\partial}{\partial \mu}\left(-\frac{1}{2s}\sum_{i=1}^{N}(x_i-\mu)^2\right) = 0 \\
&\Leftrightarrow -\frac{1}{2s}\frac{\partial}{\partial \mu}\left(\sum_{i=1}^{N}(x_i-\mu)^2\right) = 0 \\
&\Leftrightarrow -\frac{1}{2s}\left(\sum_{i=1}^{N}-2(x_i-\mu)\right) = 0 \\
&\Leftrightarrow \frac{1}{s}\left(\sum_{i=1}^{N}(x_i-\mu)\right) = 0 \\
&\Leftrightarrow \frac{N}{s}\left(\frac{1}{N}\sum_{i=1}^{N}(x_i)-\mu\right) = 0
\end{aligned}$$

如果 $N > 0$，那么很明显，上述方程只有在以下情况下才有解：

$$\mu = \frac{1}{N}\sum_{i=1}^{N}(x_i). \tag{11}$$

请注意，这确实是计算分布平均值的众所周知的公式。虽然我们都知道这个公式，但现在我们证明了它是正态分布的真均值和未知均值 $\hat{\mu}$ 的极大似然估计。现在，我们只假设我们在前面找到的由方程 (10) 定义的方差 $\hat{s}$ 的估计值仍然是 MVU 方差估计值。然而，在下一节中，我们将证明这个估计现在不再是无偏的。

5.2 效果评估

为了检查 $\hat{\mu}$ 的真均值的估计值 μ 是否无偏，我们必须确保方程 (7) 的条件成立：

$$E[\mu] = E\left[\frac{1}{N}\sum_{i=1}^{N}(x_i)\right] = \frac{1}{N}\sum_{i=1}^{N}E[x_i] = \frac{1}{N}NE[x_i] = \frac{1}{N}N\hat{\mu} = \hat{\mu}.$$

由于 $E[\mu] = \hat{\mu}$，这意味着所得到的分布均值的估计是无偏的。由于极大似然方法保证了在估计值无偏的情况下给出最小方差估计值，我们证明了 μ 是均值的 MVU 估计值。

为了检查方差 $\hat{s}$ 的早期发现的估计值 s 是否仍然是无偏的，如果它是基于经验平均值 μ 而不是真实平均值 $\hat{\mu}$，我们只需将获得的估计值 μ 插入方程 (10) 的早期导出估计值 s 中：

$$
\begin{aligned}
s = \sigma^2 &= \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \frac{1}{N}\sum_{i=1}^{N}(x_i)\right)^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[x_i^2 - 2x_i\frac{1}{N}\sum_{i=1}^{N}(x_i) + \left(\frac{1}{N}\sum_{i=1}^{N}(x_i)\right)^2\right] \\
&= \frac{\sum_{i=1}^{N} x_i^2}{N} - \frac{2\sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i}{N^2} + \left(\frac{\sum_{i=1}^{N} x_i}{N}\right)^2 \\
&= \frac{\sum_{i=1}^{N} x_i^2}{N} - \frac{2\sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i}{N^2} + \left(\frac{\sum_{i=1}^{N} x_i}{N}\right)^2 \\
&= \frac{\sum_{i=1}^{N} x_i^2}{N} - \left(\frac{\sum_{i=1}^{N} x_i}{N}\right)^2
\end{aligned}
$$

为了检查估计值是否仍然无偏，我们现在需要再次检查方程 (7) 的条件是否成立：

$$
\begin{aligned}
E[s] &= E\left[\frac{\sum_{i=1}^{N} x_i^2}{N} - \left(\frac{\sum_{i=1}^{N} x_i}{N}\right)^2\right] \\
&= \frac{\sum_{i=1}^{N} E[x_i^2]}{N} - \frac{E[(\sum_{i=1}^{N} x_i)^2]}{N^2}
\end{aligned}
$$

如前一节所述，方差的一个重要性质是真实的方差 $\hat{s}$ 可以写成 $\hat{s} = E[x_i^2] - E[x_i]^2$ 以满足 $E[x_i^2] = \hat{s} + E[x_i]^2 = \hat{s} + \mu^2$ 的条件。在上述公式中使用此性质可得出：

$$
\begin{aligned}
E[s] &= \frac{\sum_{i=1}^{N} E[x_i^2]}{N} - \frac{E[(\sum_{i=1}^{N} x_i)^2]}{N^2} \\
&= s + \mu^2 - \frac{E[(\sum_{i=1}^{N} x_i)^2]}{N^2} \\
&= s + \mu^2 - \frac{E[\sum_{i=1}^{N} x_i^2 + \sum_{i}^{N} \sum_{j \neq i}^{N} x_i x_j]}{N^2} \\
&= s + \mu^2 - \frac{E[N(s + \mu^2) + \sum_{i}^{N} \sum_{j \neq i}^{N} x_i x_j]}{N^2} \\
&= s + \mu^2 - \frac{N(s + \mu^2) + \sum_{i}^{N} \sum_{j \neq i}^{N} E[x_i]E[x_j]}{N^2} \\
&= s + \mu^2 - \frac{N(s + \mu^2) + N(N-1)\mu^2}{N^2} \\
&= s + \mu^2 - \frac{N(s + \mu^2) + N^2\mu^2 - N\mu^2}{N^2} \\
&= s + \mu^2 - \frac{s + \mu^2 + N\mu^2 - \mu^2}{N} \\
&= s + \mu^2 - \frac{s}{N} - \frac{\mu^2}{N} - \mu^2 + \frac{\mu^2}{N} \\
&= s - \frac{s}{N} \\
&= s\left(1 - \frac{1}{N}\right) \\
&= s\left(\frac{N-1}{N}\right)
\end{aligned}
$$

由于明显的 $E[s] \neq \hat{s}$,这表明分布方差的估计不再是无偏的。事实上,这个估计值平均低估了因子 $\frac{N-1}{N}$ 的真实方差。当采样数接近无穷大 ($N \to \infty$) 时，该偏差收敛为零。然而，对于小样本集，偏差是有意义的，应该消除。

5.3 固定偏差

由于偏差只是一个因素，我们可以通过将方程 (10) 中定义的偏差估计值 s 按偏差的倒数缩放来消除偏差。因此，我们定义一个新的无偏估计 $s\prime$ 如下：

$$s\prime = \left(\frac{N-1}{N}\right)^{-1} s$$

$$s\prime = \left(\frac{N-1}{N}\right)^{-1} \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$$

$$s\prime = \left(\frac{N}{N-1}\right) \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$$

$$s\prime = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^2$$

这个估计值现在是无偏的，确实类似于传统的计算方差的公式，在这里我们除以 $N-1$ 而不是 N。但是，请注意，得到的估计值不再是最小方差估计值，而是所有无偏估计值中方差最小的估计值。如果除以 N，则估计值是有偏的，如果除以 $N-1$，则估计值不是最小方差估计值。然而，一般来说，有偏估计要比有略高的方差估计差得多。因此，如果总体平均数未知，则应使用 $N-1$ 除法，而不是 N 除法。

6 结论

在本文中，我们展示了计算正态分布数据的均值和方差的常用公式的来源。此外，我们还证明了方差估计公式中的归一化因子在总体的真实均值已知的情况下应该是 $\frac{1}{N}$，而在均值本身也需要估计的情况下应该是 $\frac{1}{N-1}$。

If you’re new to this blog, don’t forget to subscribe, or follow me on twitter!