R-LAB实验报告

软件61 吴海旭 2016013223

1. 实验一

本实验的主要目标是验证以下内容:

$$if\ \ x \sim \mathbb{N}(\mu, \sigma^2)\ \ and\ \ x_1, x_2, \ldots, x_n \sim iid\, \mathbb{N}(\mu, \sigma^2)\quad s^2 = \frac{1}{n-1}\sum_i x_i$$

$$then\ \ E(\overline{x}) = \mu\quad E(s^2) = \sigma^2$$

1.1 实验设计

> 变量声明:
>
> μ 和 σ^2 为正态分布参数。
>
> n为每一组随机序列中随机数的个数，m为产生随机序列的组数。
>
> num为执行上述操作的次数。

我们设计实验如下:

- 固定 μ 和 σ^2 、m，调节n，进行模拟实验

1.2 实验结果

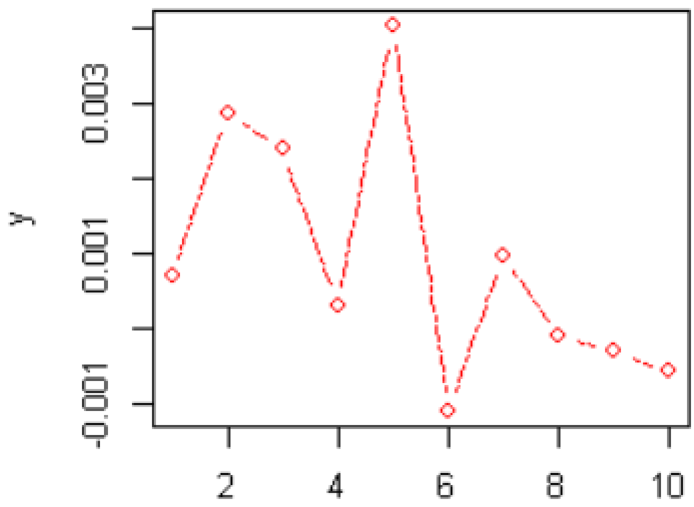
我们去操作次数num 为10，m为500，即我们将重复生成 500 组 n 个随机数的序列100次。

我们得到结果如下:

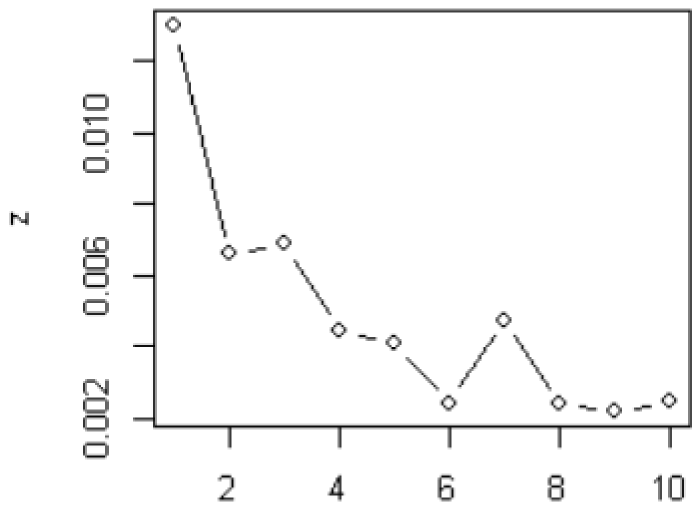
公式一

我们将 μ 和 σ^2 分别设置为0和1，调节 n 的值得到以下结果（n为从20到400以20为步长的等差数列）。

拟合图:



方差图:



上图中，横坐标表示第x组n值，红色线的纵坐标为$E(\overline{x})$的拟合值，黑色线为其方差值。

我们发现如下结论:

- 随着n的增加，$E(\overline{x})$逐渐波动下降趋向于0。
- 随着n的增加，$E(\overline{x})$的方差z逐渐趋向于0。

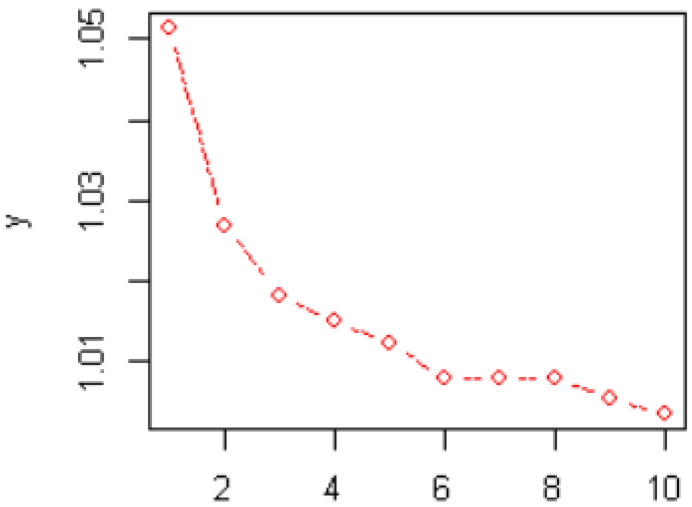
对于$E(\overline{x})$的波动下降，我认为因为此时数据量较小，所以抖动较大，但是也可以看出下降趋势。

而方差的值确实会随着n增大而减小，因为产生的随机独立变量由n个值来决定，n越大，数据越稳定，所以方差减小。

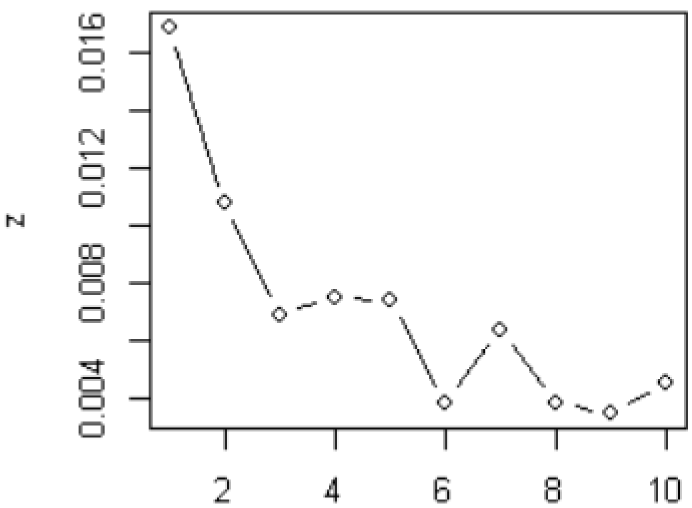
公式二

我们将 μ 和 σ^2 分别设置为0和1，调节 n 的值得到以下结果（n为从20到400以20为步长的等差数列）。

拟合图:



方差图:



上图中，横坐标表示第x组n值，红色线的纵坐标为$E(s^2)$的拟合值，黑色线为其方差值。

我们发现如下结论:

- 随着n的增加，$E(\sigma^2)$逐渐下降趋向于1。所以我们可以认为$E(s^2) = \sigma^2$在本实验中成立。
- 随着n的增加，$E(\sigma^2)$的方差z逐渐趋向于0。

2. 实验二

本实验的主要目标是验证以下内容:

$$if \;\; x \sim \mathbb{N}(\mu, \sigma^2) \;\; and \;\; x_1, x_2, \ldots, x_n \sim iid\, \mathbb{N}(\mu, \sigma^2)$$
$$\overline{x} \sim \mathbb{N}(\mu, \frac{\sigma^2}{n})$$

2.1 实验设计

> 变量声明:

> μ 和 σ^2 为正态分布参数。
>
> n为每一组随机序列中随机数的个数，m为产生随机序列的组数。

我们设计实验如下：

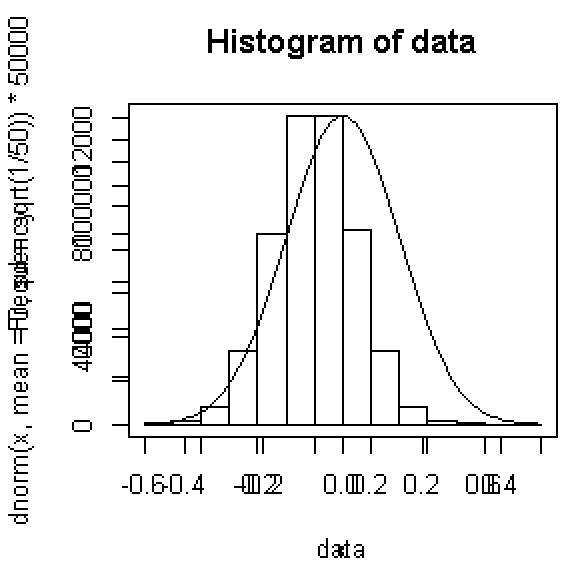
- 固定 μ 和 σ^2 ，调节 n ，进行模拟实验

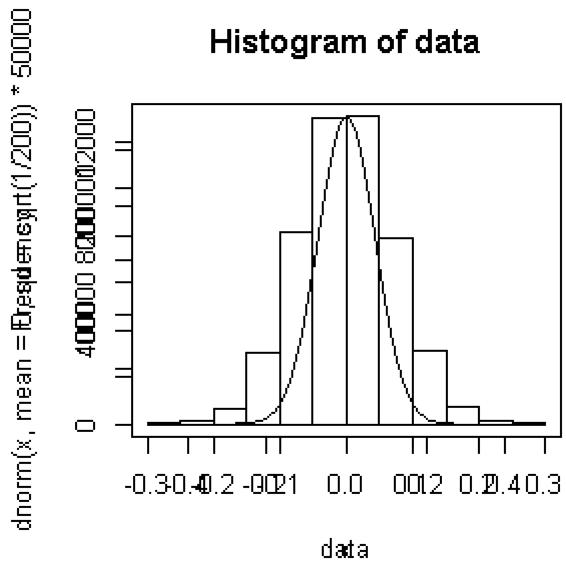
2.2 实验结果

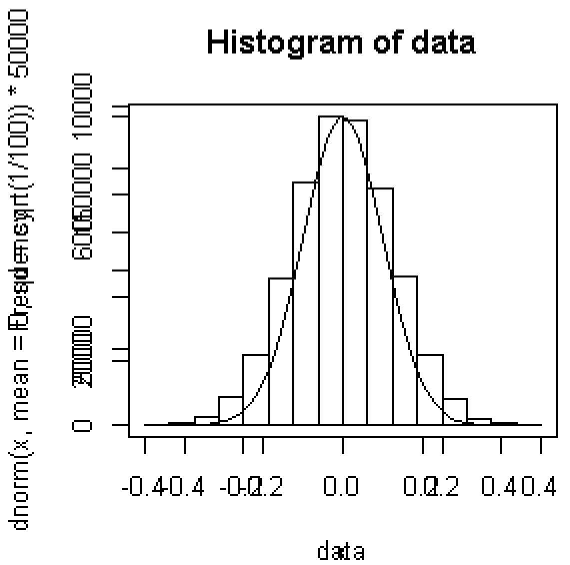
我们设定m即数据组数设置为50000，n值依次为50、100、200、400，μ 和 σ^2 分别设置为0和1。

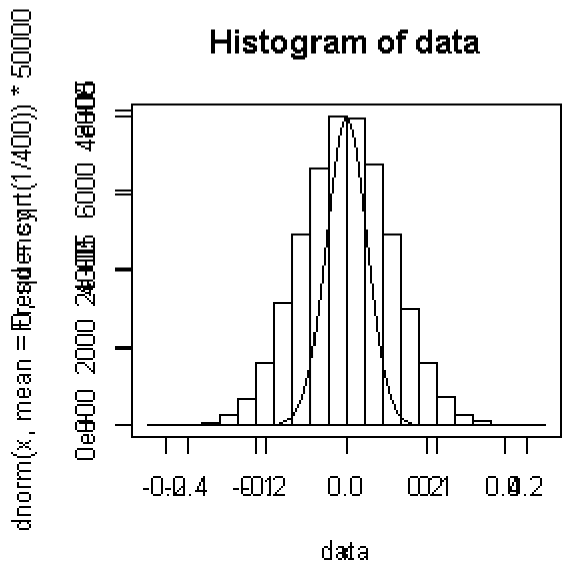
得到了以下数据结果。

> 图中曲线为对应的正态分布函数图像。因为两张图绘制在一起，所以有些凌乱。









根据上述结果，我们可以得到以下结论：

- 在本此实验中，符合下述结论。

$$\overline{x} \sim \mathbb{N}(\mu, \frac{\sigma^2}{n})$$

- 随着n值，即每一组数据量增大时，$\overline{x}$ 更接近理想分布。

3. 思考

- 数据越多是否拟合越好?

实际上，数据量的增大只是尽可能模拟蒙特卡洛的方式。所以，一般情况下，数据量大，数据拟合确实会变好。

- 如何衡量拟合结果?

对于上述结果，我们仅仅通过方差和目测是否符合曲线走势的方法来衡量，但是这样的方案是不完善的。比如，我们在做实验二的时候，可以使用插值法进行函数拟合，在和正态分布做比较。