

要求：利用 ROOT 生成数量为 N 的随机数，服从均值为 50，标准差为 10 的高斯分布，并拟合，比较不同 N 、不同 bin 大小对拟合结果的影响。

一.部分概念

1 高斯分布

高斯分布（正态分布）的概率密度函数为：

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

其中 μ 为均值， σ 为标准差。本次实验设定真值为 $\mu=50$ ， $\sigma=10$ 。

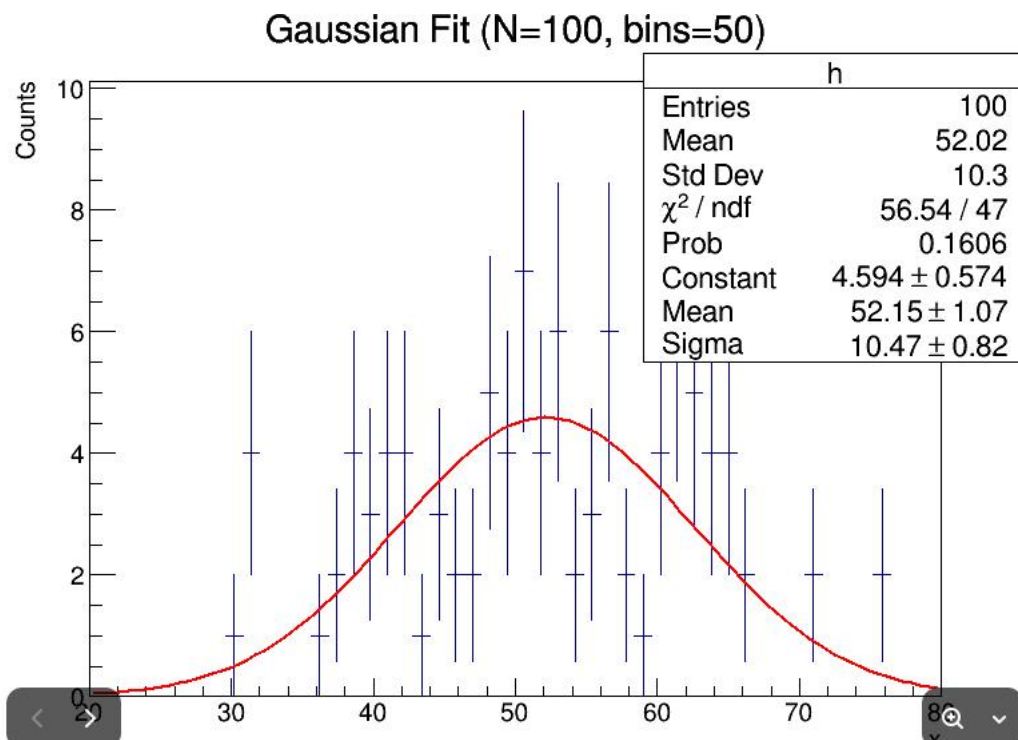
2 直方图与 bin

直方图将数据范围划分为若干个等宽区间，每个区间称为一个 bin。落在每个 bin 内的数据点个数即为该 bin 的计数。选择合适的 bin 数目需要在“形状精度”和“统计可靠性”之间权衡。

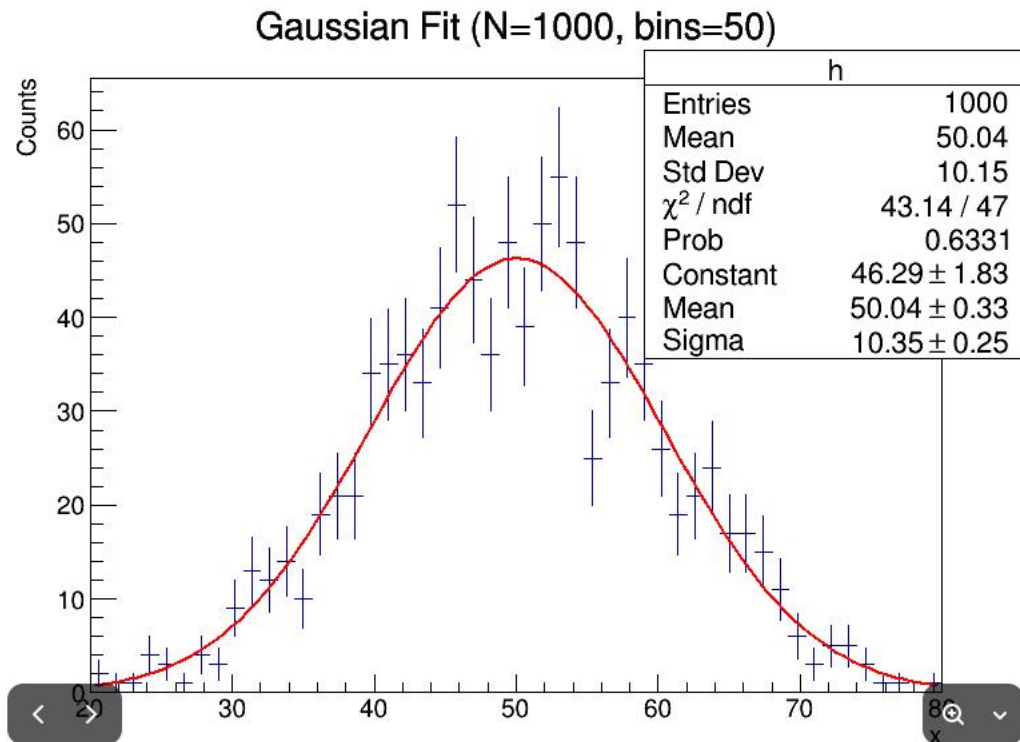
二.数据绘图

I 固定 bins=50，改变 N 值

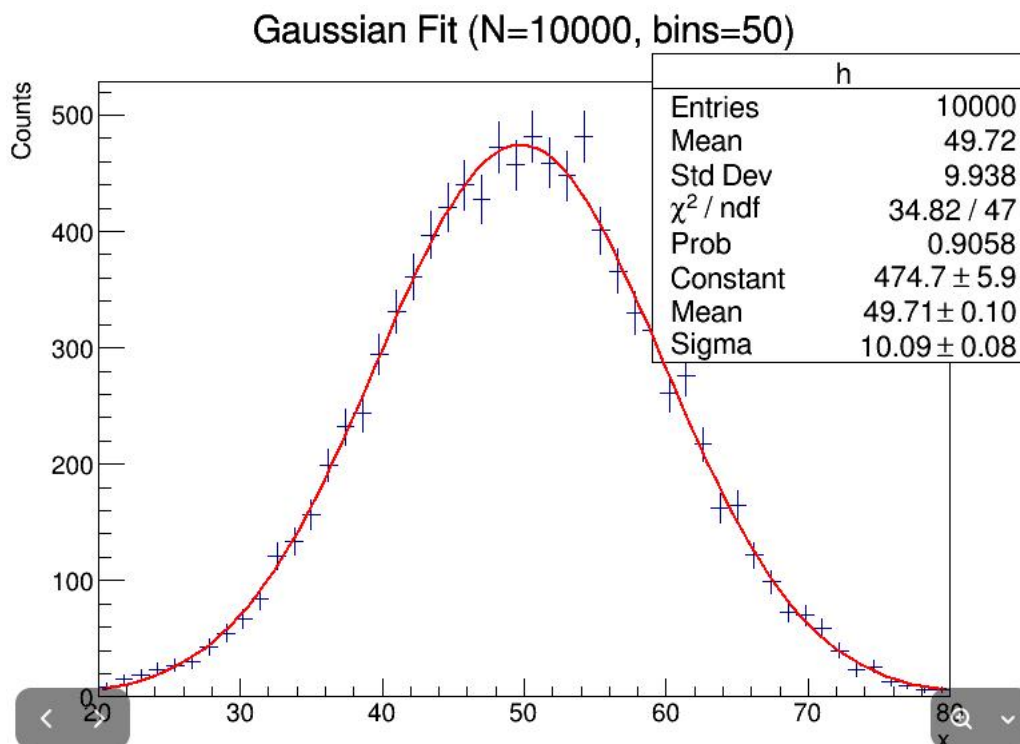
★ $N=100$



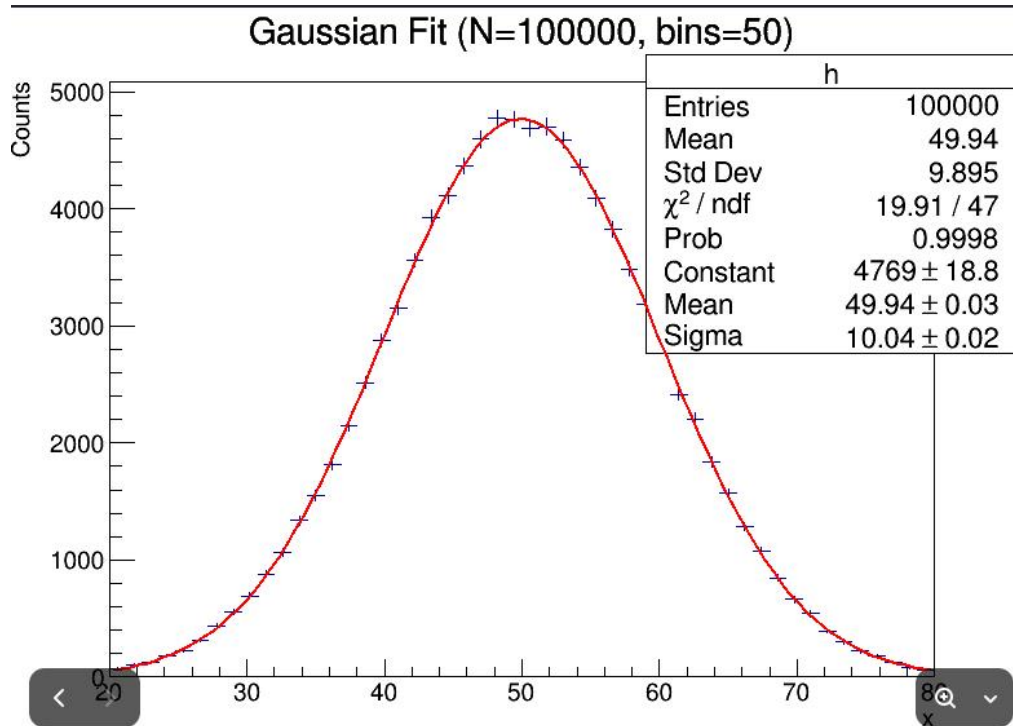
★ N=1000



★ N=10000

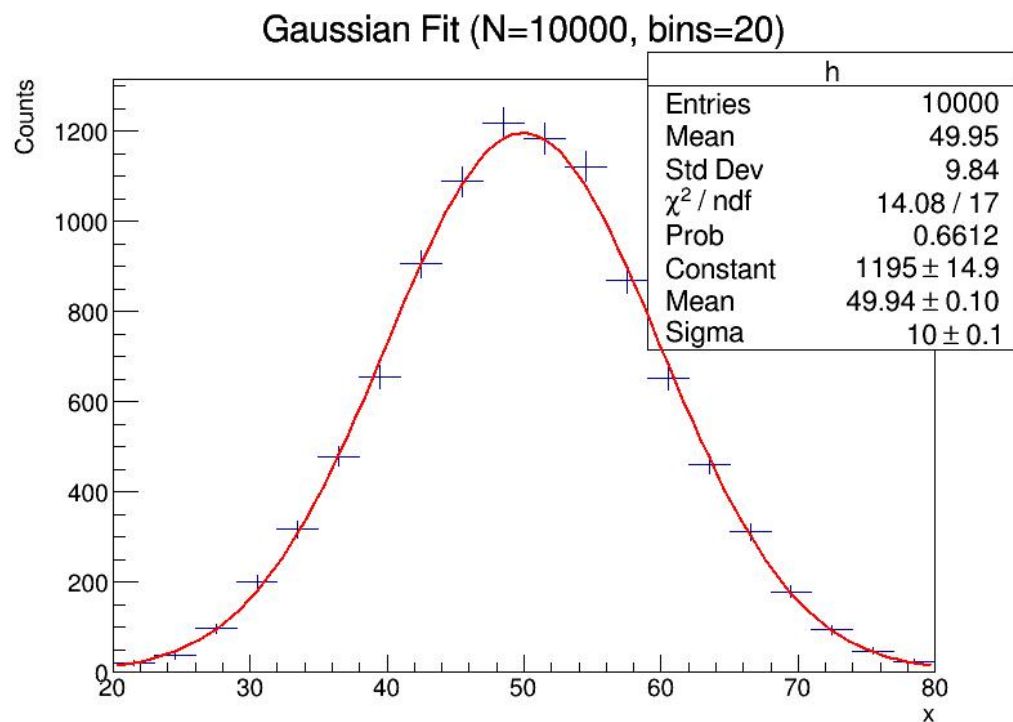


★ N=100000

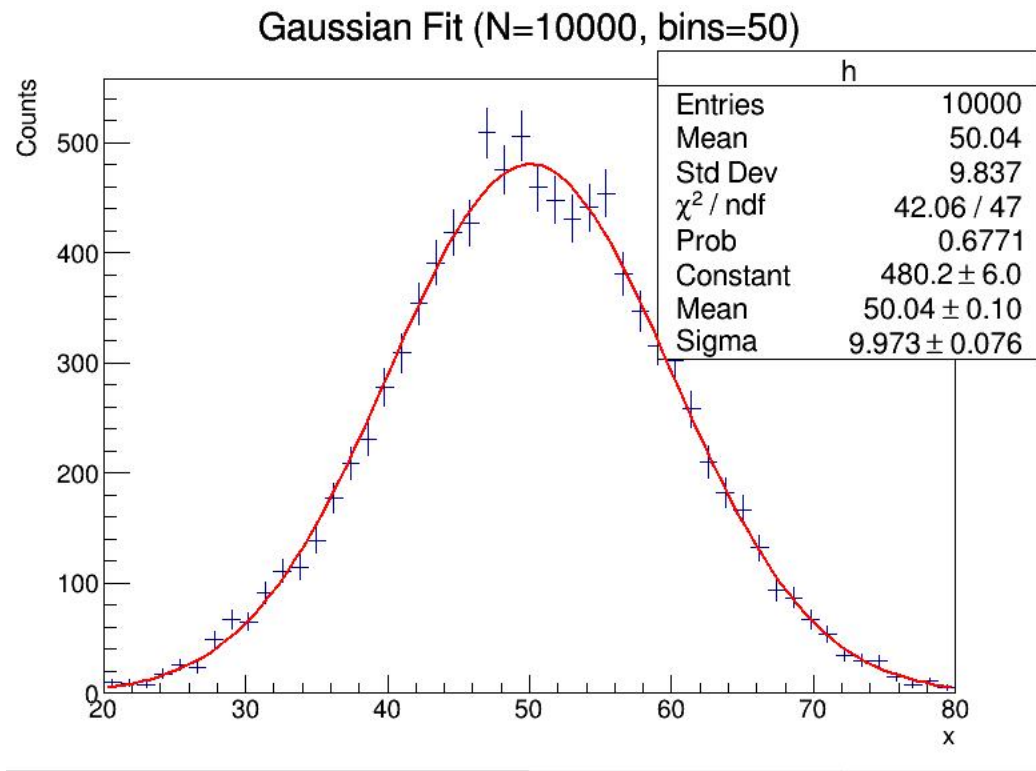


II 固定 N=10000, 改变 bins 值

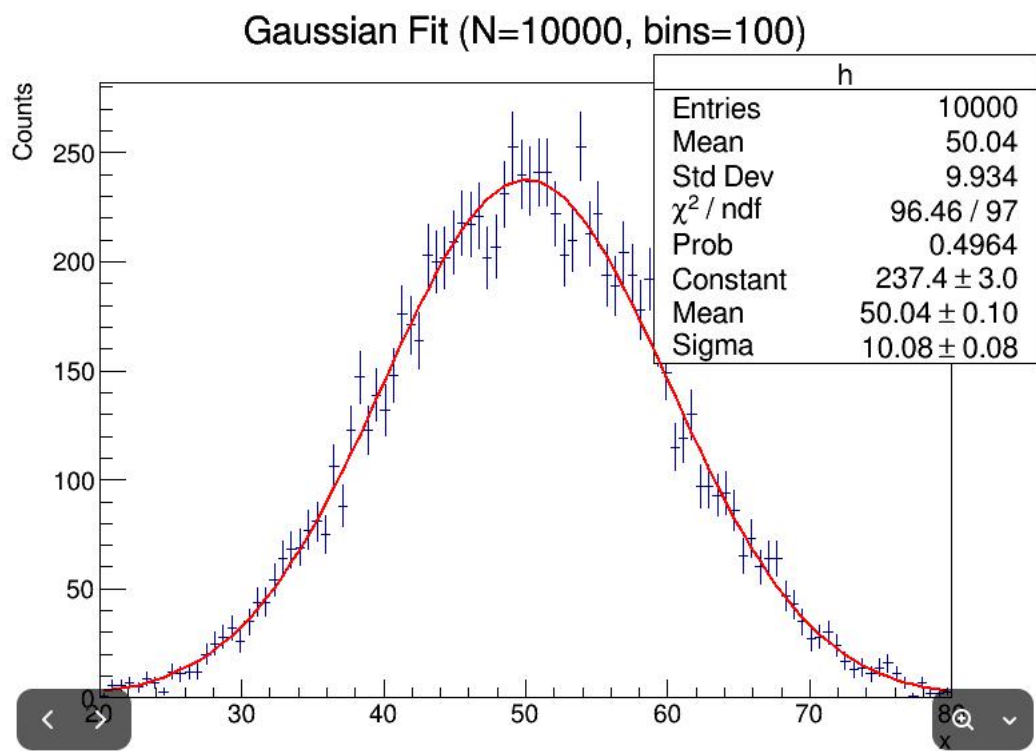
★ bins=20



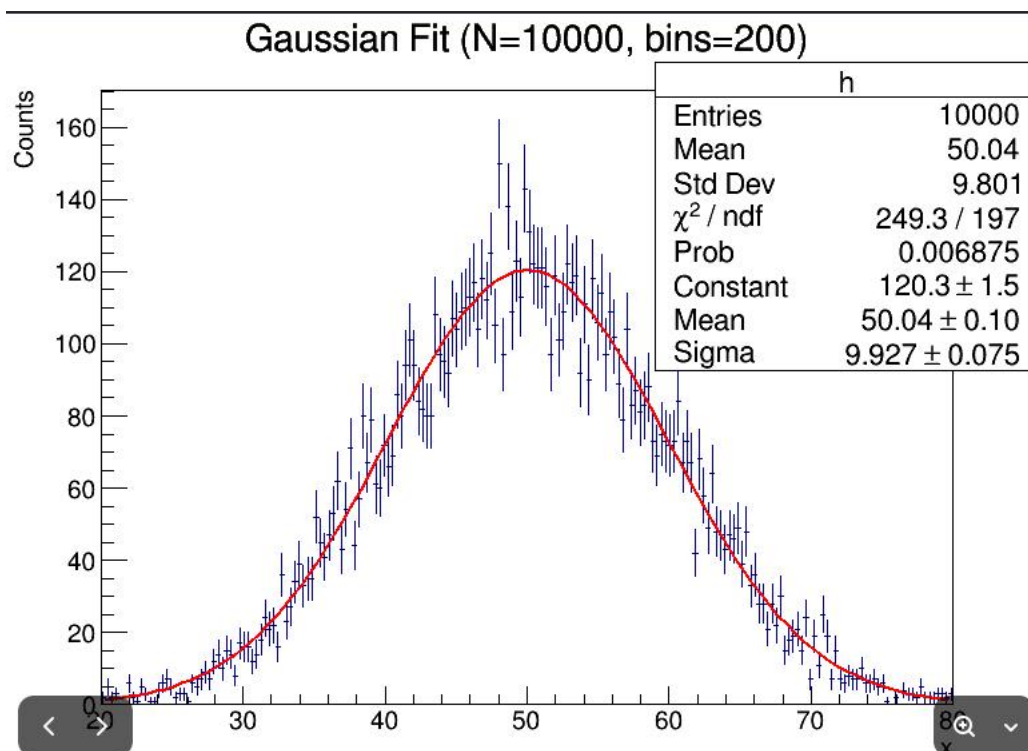
★ bins=50



★ bins=100



★ bins=200



三.实验分析

1.随着 N 增大，直方图越来越平滑，拟合结果趋近真值(50,10)，参数误差按 $1/\sqrt{N}$ 的规律减小。

2. bin 数太少会平滑掉分布特征，降低参数分辨率；bin 数太多会导致单个 bin 统计量不足，拟合不稳定。实际分析中需根据数据量选择合适 bin 数。最优 bin 数需要保证每个 bin 有足够的计数（一般 ≥ 5 ），同时尽可能保留分布形状。

比如在此实验中，bin=100 时，形状保留较好，统计涨落可控，拟合结果最准。但在 bin=200 时，部分 bin 计数太少，直方图出现毛刺， χ^2 明显增大，拟合变得不稳定。