

2016 威海暑期学校结业报告

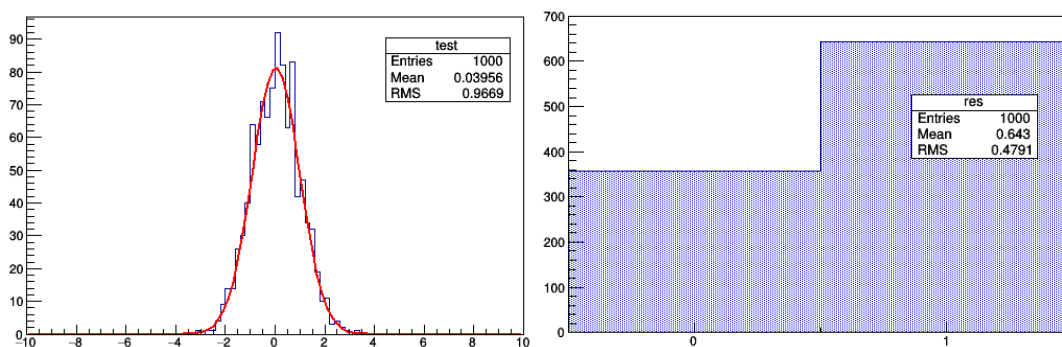
—两个统计问题的讨论

唐光毅, IHEP, 苑长征, 粒子物理与原子核物理, tanggy@ihep.ac.cn

1. 一倍标准差区间 $[\mu - \sigma, \mu + \sigma]$ 对应 68.4%的概率表示什么? 是真值落在这个区间里面的概率么?

Tommaso 反复强调的一点是, 一旦做出了一个区间估计, 真值要么落在其中, 要么没有落在其中, (非零即一), 所以 68.4%不能作为真值是否落在其中的概率。符合数学定义的一个说法也许是, 所估计的区间包含(覆盖)真值的概率是 68.4%。下面考察了一个小例子:

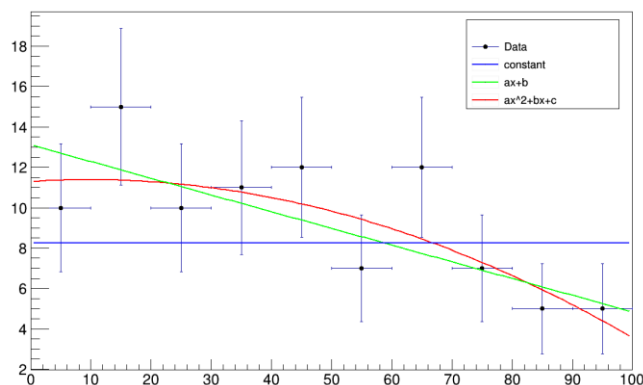
产生 1000 个标准正态分布随机数, 填入直方图, 并用高斯分布拟合该图, 可以得到拟合参数和误差。我们考察真值是否落在这个区间中: 如果在, 则记录 1; 不在, 记录 0。重复这个过程。



上两图分别是其中某一次拟合示例(左)和 1000 次重复之后的结果(右)。右图显示 64.3%的区间包含真值, 考虑到统计误差, 可以认为和 68.4%是一致的。(所以两种陈述差别就在于真值可不可以作为随机变量来考察它的分布? 或者说有没有后验的办法讨论真值的分布? 这里没有余力讨论这个问题。)

2.F 检验和 χ^2 检验显著性的比较。

利用 Tommaso 给出的例子, 这里尝试 χ^2 检验是不是可以给出同样的结论。



拟合如上图所示,常数、线性、二次拟合对应的 χ^2 值分别是 11.35、4.23、3.628。和 Tommaso 一样考虑 5%的置信度,线性和常数拟合的 p 值是 $0.0076 < 5\%$,说明线性拟合增加的参数作用“显著”。二次和常数拟合的 p 值是 $0.44 > 5\%$,说明二次拟合增加的参数作用不“显著”。结论仍然是线性拟合最佳,与 F 检验结论一致。

不过 F 检验和 χ^2 检验对应的 P 值不同,说明存在两种检验给出矛盾结果的可能(类似于选取不同的置信度的结果也可能出现矛盾)。值得注意的另一点是 F 检验仍然不能区分具有相同参数个数的模型的“好坏”。