

本次作业测试的具体数据表格如下所示：

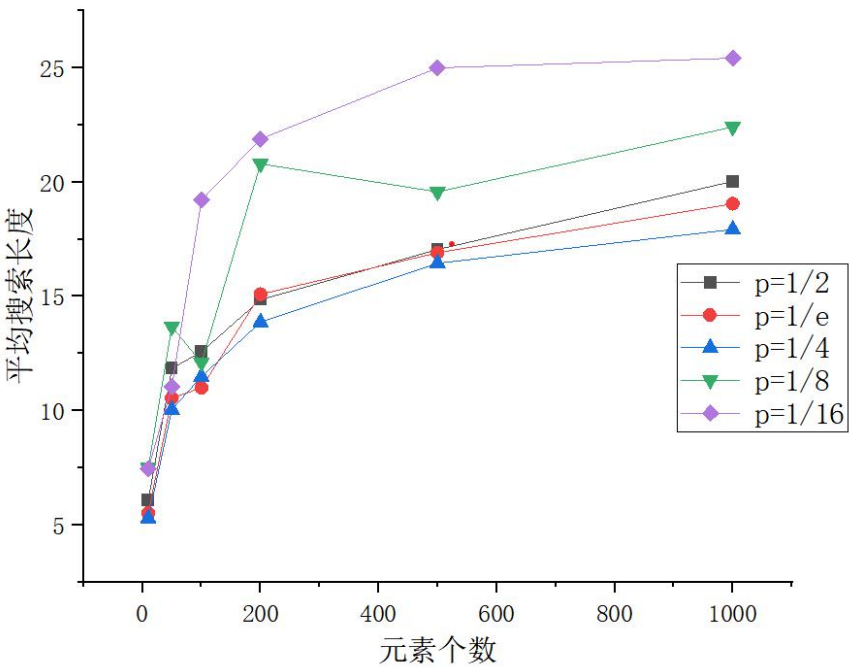
	10	50	100	200	500	1000
$p=1/2$	6.0828	11.8525	12.5706	14.85	17.0505	20.0302
$p=1/e$	5.5129	10.5534	11.0054	15.0947	16.9078	19.0502
$p=1/4$	5.2931	10.0465	11.4881	13.8734	16.4491	17.9264
$p=1/8$	7.4926	13.6759	12.1026	20.8043	19.5711	22.4166
$p=1/16$	7.4641	11.0651	19.2114	21.8817	24.9939	25.4222

理论数值如下表所示：

$p=1/2$	8.6438	13.2877	15.2877	17.2877	19.9316	21.9316
$p=1/e$	7.8391	12.2139	14.0981	15.9823	18.4731	20.3572
$p=1/4$	7.9772	12.621	14.6211	16.6211	19.2649	21.2649
$p=1/8$	10.0013	16.1931	18.8598	21.5265	25.0516	27.7183
$p=1/16$	14.3544	23.6421	27.6421	31.6421	36.9298	40.9298

一．对于同一个概率，改变其元素个数时其平均搜索长度变化曲线如下所示：

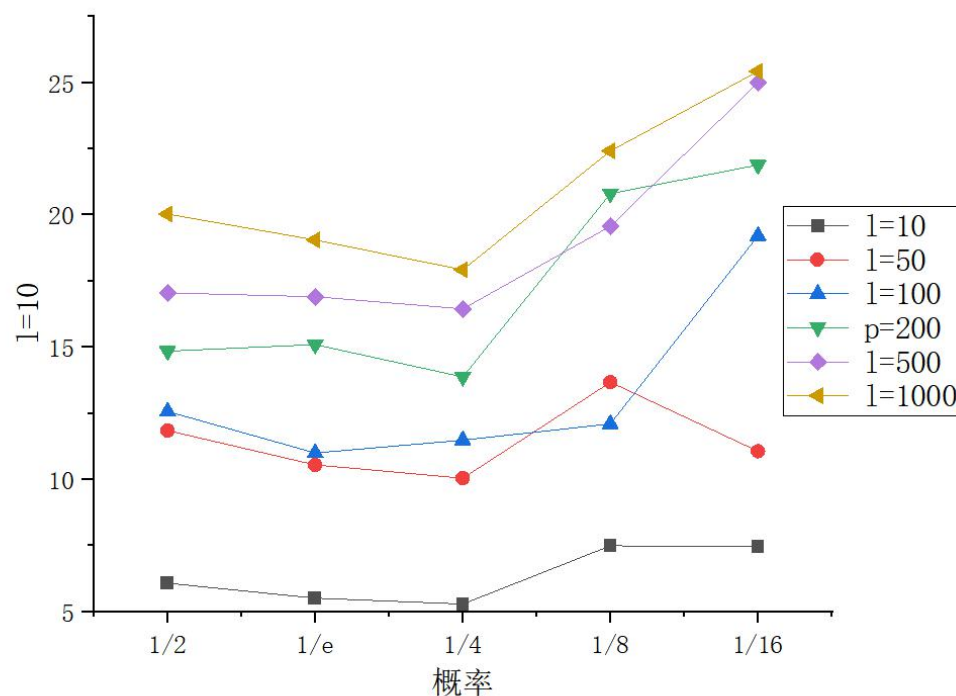
1



由上图可以知道,对于同一个概率而言,其平均搜索长度会随着元素个数的增加而逐渐增大。同时根据理论知识可以知道,跳表在理想状况下的平均搜索长度具有 $O(\log n)$ 的平均复杂度,而从上图以及上面两个表格的数据对比可以知道,对于所有实际测得数值而言,其数值均小于理论数值,这是符合理论情况的;而对于曲线的拟合程度而言,当 p 为 $1/2$ $1/e$ $1/4$ 时曲线与理论值的 $\log n$ 曲线符合的很好,而对于 $p=1/8$ 以及 $p=1/16$ 两组曲线,其个别数据与理论值之间的偏差较大, ($p=1/8$ 组的 50 200 两个数据点) 分析可能的原因如下:

1. 在生成 search 的目标值时, 使用了时间作为随机种子来生成伪随机数, 这可能导致在较短时间内的调用结果并不是那么理想的均匀, 这可能使得个别点的数据与理论值之间的偏差过大。
2. 每个数据都是有相同情况下运行五次取结果的平均值得到的, 出现一定的偏差可能是由于实验次数过少而导致的偶然误差。

二. 对于相同的元素个数, 改变对应的概率时平均搜索长度的变化如下所示:



根据上图可以知道,对于同一个元素个数的跳表,降低其插入概率时总体有平均搜索长度上升的趋势。

同时根据上面的理论数值可以知道, $p=1/2$ $p=1/e$ $p=1/4$ 的时候各个数据点的值几乎相等,而这点在上图中体现的很好;同时通过上面的理论值表格可以知道,对于每个长度的跳表而言, $p=1/e$ 时候的数据基本都是最小的,而对于上面绘制的折线图而言, $p=1/4$ 的时候才能够具有最小的搜索长度,这与理论值的分析具有一定的偏差,但是实际上差距并不是很大,分析原因如下:

1. 个人认为主要的误差原因应该就是随机数的生成并不能真正的做到所谓的平均取值, 而使得部分数据与理论排序之间存在一定的误差。
2. 在进行实验数据的取得时, 我对于 MAX_LEVEL 的计算采用的是 $\log n / \log(1/p)$ 并对结果进行取整的方式来进行的, 这会导致计算时的最大层数与理论最大层数之间存在一定的误差, 而跳表不同层之间的跳转也会对搜索长度产生较大的影响, 这也可能导致结果中不同概率之

间的排序并不完全符合理论的排序。