接下来希望测试的两个方面

1.测试不同算法在不同数据集质量的情况下 的准确率。

如何定义数据集质量？能控制的因素：每一个attribute的候选值个数， 选到真值的概率

（候选值多的时候 选到真值的概率应该相应调低）

需要：一个能提供 输出文件名、候选值范围及选到真值概率的接口（希望也能提供source个数和attribute个数的接口）的函数，它要能输出到文件，在第一行输出每个attribute的ground truth，在接下来的 #source行输出每一个source对每一个attribute给出的值（暂时认为每一个source会给出所有attribute的值，后续可以考虑用0来代替没有给出值）格式为每一行代表每一个source，每一列代表每一个attribute。

2.测试不同算法在source accuracy普遍较高，但在某几个attribute上大家都犯错了，算法能否得到真值的能力。

前提：在某几个attribute上做对的source是整体比较可信的（即在其他attribute上也倾向于犯较少的错）

测试目标：不同的算法在满足上述条件的数据集上的效果。在特定attribute取到真值的频率达到多少时该算法能得到该attribute的真值？测出不同算法的这个threshold frequency