RANCAS算法参数K的确定

当我们从估计模型参数时，用p表示一些迭代过程中从数据集内随机选取出的点均为局内点的概率；此时，结果模型很可能有用，因此p也表征了算法产生有用结果的概率。用w表示每次从数据集中选取一个局内点的概率，如下式所示：

w = 局内点的数目 / 数据集的数目

通常情况下，我们事先并不知道w的值，但是可以给出一些鲁棒的值。假设估计模型需要选定n个点，wn是所有n个点均为局内点的概率；1−wn是n个点中至少有一个点为局外点的概率，此时表明我们从数据集中估计出了一个不好的模型。(1−wn)k表示算法永远都不会选择到n个点均为局内点的概率，它和1-p相同。因此，1−p=(1−wn)k。我们对上式的两边取对数，得出:k=log(1−p)/log(1−wn)

值得注意的是，这个结果假设n个点都是独立选择的；也就是说，某个点被选定之后，它可能会被后续的迭代过程重复选定到。这种方法通常都不合理，由此推导出的k值被看作是选取不重复点的上限。例如，要寻找适合的直线，RANSAC算法通常在每次迭代时选取2个点，计算通过这两点的直线可能模型，要求这两点必须唯一。

为了得到更可信的参数，标准偏差或它的乘积可以被加到k上。k的标准偏差定义为：

SD(k)= √（1−wn）/wn

源自：<https://blog.csdn.net/robinhjwy/article/details/79174914>