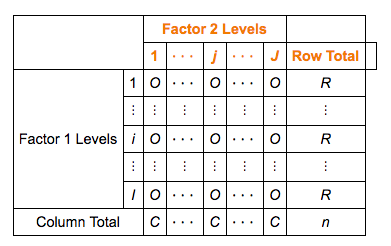
卡方检验：

1. 卡方分布
2. 卡方检验

这里讨论一般情况下两个因子的卡方检验，预设：

我们的样本有两个属性（X，Y），预测变量X和预测目标Y，其中：预测变量X有I种取值，预测目标Y有J种取值，我们定义如下方式的表格：



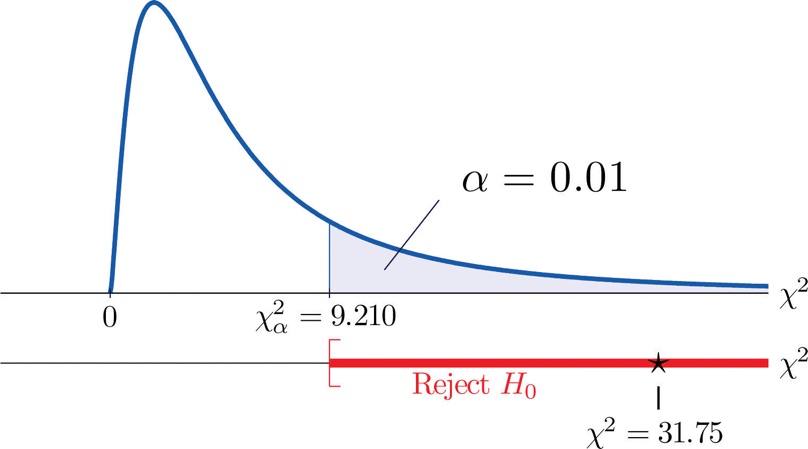
General Contingency Table

记：O是一个IxJ的矩阵，C是1xI的行矩阵，R是Ix1的列矩阵，n是样本数。

E是一个IxJ的矩阵，它是C与R的直积，计算方式如下：

卡方值的计算公式如下：

满足一个自由度为的卡方分布。



上图中，alpha值为0.01，其拒绝域是[9.210,\infty]，表示如计算出的卡方值大于9.21，那么我们只有0.99的概率（0.01的出错率）拒绝零假设（两个因子独立），即：99%的可能性两因子独立。

1. Python实现卡方检验
2. 定义函数chi\_test(X,Y,I,J)

输入预测变量X，和预测目标Y，及两个变量的取值自由度（I，J）。

1. 计算Chi值
2. 找到Chi对应的拒绝域