

卡方分布（chi square distribution）即 $\chi^2$ 分布。

假设随机变量X服从正态分布，记作 $X \sim N(0,1)$ ，均值是0，方差是1。我们再定义一个变量Q，它是从X的正态分布中抽取样本,然后对得到的值进行平方，也就是它等于随机变量X的平方：

$$Q = X^2$$

卡方分布就是这样一些数量的随机变量平方之和

这个Q就是一个 $\chi^2$ 分布的随机变量。由于它只是一个独立的标准正态分布随机变量平方之和，说它只有一个自由度。表示为：

$$Q \sim \chi_1^2$$

再假设 $Q_2$ 是两个独立的标准正态分布随机变量平方之和：

$$Q_2 = X_1^2 + X_2^2$$

则

$$Q_2 \sim \chi_2^2.$$

我们来看下卡方分布表的使用：

自由度k \ P value （概率）	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88

假设自由度为2，临界值为2.41的概率为0.30。这就是说在密度函数下，临界值以上所占的总面积为30%：

The Chisq Density Distribution

