慶應義塾大学ビジネス・スクール

統計学　ノート（5）

1　カイ2乗分布

今まで2つの平均または2つの割合が等しいかどうかの検定の問題を，正規曲線を用いる方法で解いてきた。しかし，3つ以上の変数がある問題をこの方法で解くことはできない。例えば，4種類の平均を調べるとき，1度に2種類ずつを比較するのは有効な解決にはならない。いくつかの割合を比較する問題を考え，いくつかの平均が等しいという仮説の検定法を与える。一般的な問題は次のように表される。

ある実験の可能な結果の数をkで表す。これらの可能な結果はk個のマスあるいは箱で表される。実験をn回行い，各マスに入った結果を実験全体の観測度数として表す。そのとき問題は，これらの度数が仮定したある理論から予想される度数に適合するかどうか決めることである。

このとき平均度数は期待度数とよばれ，これをe_iで表す。また観測度数はo_iで表す。

適合しているかどうか検定する一般的な方法は，観測度数と期待度数の一致の程度を測るある尺度を基にしている。この尺度はカイ2乗とよばれ，次の式で定義される。

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \tag{1}$$

ここで，o_iとe_iはそれぞれi番目のマスの観測度数と期待度数を表し，kはマスの数を表す。

公式(1)を見れば明らかなように，χ^2の値は観測度数と期待度数が完全に一致したとき0となり，その差が大きくなればχ^2は大きくなる。したがって，χ^2の値が大きくなればなるほど，実験と理論との一致はますます弱くなると考えられる。

1.1　カイ2乗分布

他の標本分布の場合と同様に，χ^2の理論分布を数学的に求めることは可能である。各マスの度数がとりうる値には限界があるので，χ^2がとりうる値にも限界がある。したがって，χ^2の理論分布は離散形分布となるに違いない。多数の値をもつ離散型分布は，膨大な計算を必要とするので，実際的な考えとしては，2項分布を正規分布で近似したように，離散型χ^2分布を連続型分布で近似させる必要がある。この近似として，カイ2乗分