適合度検定，独立性の検定の例題

母集団の属性が k 個の互いに排反なクラス $C_1, C_2, \cdots, C_k$ に分かれているものとする。そのクラスに入る確率が理論的に $p_1, p_2, \cdots, p_k$ （ただし $\sum_i p_i = 1$） とする。実際に母集団からの抽出した標本の値が $n_1, n_2, \cdots, n_k$ （ただし $\sum_i n_i = n$） は，その理論値 $np_1, np_2, \cdots, np_k$ とは，必ずしも合致していないから, ずれが生じる。このずれに関して適当な尺度を考えて，理論と実際とが一致しているかどうかを検定する。

例題（確率分布の適合度検定）人間の血液型は，4 種類で，その構成比率は　0.16, 0.48, 0.20, 0.16 という。いまある職業をもつ 770 人について調べたら，　180, 360, 132, 98　であった。これに対して，もとの構成比率のそれを比べて検定せよ。

（解）

血液型	C_1	C_2	C_3	C_4	計
観測度数（標本値）	180	360	132	98	770
期待度数（理論値）	$770 * 0.16$	$770 * 0.48$	$770 * 0.20$	$770 * 0.16$	770
	$= 123.2$	$= 369.6$	$= 154.0$	$= 123.2$	$= 770.0$

検定統計量は

$$
\begin{aligned}
\chi^2 \quad &:= \textstyle\sum_{i=1}^{k} \dfrac{(n_i - np_i)^2}{np_i} \\
&= \frac{(180-123.2)^2}{123.2} + \frac{(360-369.6)^2}{369.6} + \frac{(132-154.0)^2}{154.0} + \frac{(98-123.2)^2}{123.2} = 34.73
\end{aligned}
$$

を用いる。有意水準は $\alpha = 0.05$，自由度 $\nu = 4-1 = 3$. χ^2 分布表より，$\chi^2_{0.05}(3) = 7.81$ よって棄却域は $\{\chi^2 | \chi^2 > 7.81\}$ χ^2 の値は棄却域に入っているから，仮説は棄却される。したがってこの血液型の分布は仮説が与える構成比率とは異なるものである。(終)

つぎに属性が 2 つの場合 A, B を考える。一方の属性 A を縦 (行) ともう一方 B を横 (列) に並べれば，行列の形に枠が書ける。これを**分割表**とよぶ。つまり 2 つの分類基準の間に関してこれらの関連性の有無を調べる。「関連がない」ということは，それぞれの枠に属する/属さないことが「独立である」(**独立性の検定**)。これを実際の標本値と理論値と比較する。確率の理論値はそれぞれの周辺度数から独立性の仮定のもとで計算できる。

例題（独立性の検定）つぎの表はある科目の成績評価を学部別にしたものである。これについて学部間に成績の優劣が認められるかどうかを有意水準 1% で検定せよ。

	A 学部	B 学部	C 学部
優	8	12	6
良	25	32	10
可	10	9	14
不可	7	5	12

（解）周辺和から各マスの期待度数を $\frac{r_1 c_1}{n} = 26 * 50/150 = 8.67, \frac{r_1 c_2}{n} = 26 * 58/150 = 10.05, \cdots$ と計算してもとめると

8.67	10.05	7.28
22.33	25.91	18.76
11.00	12.76	9.24
8.00	9.28	6.72

$$
\chi^2 = \sum_i \sum_j \frac{\left(\left| n_{ij} - \frac{r_i c_j}{n} \right| \right)^2}{\frac{r_i c_j}{n}} = \frac{(8-8.67)^2}{8.67} + \cdots + \frac{(12-6.72)^2}{6.72} = 16.39
$$

有意水準と自由度 $\alpha = 0.05, \nu = 3 \times 2 = 6$ のカイ 2 乗分布表から $\chi^2_{0.05}(6) = 12.59$ よって棄却されない（有意である）。すなわち学部間に優劣が認められる。(終)