2011 年度秋学期　統計学　第１２回

χ^2 分布と適合度検定

今回は，各グループに属するデータの数の比率が理論通りかどうかを検定する「適合度検定」を紹介します。この検定は，よく用いられる手法ではありますが，いくぶん問題をはらんだ手法でもあります。そのことを今回の講義で理解してください。

適合度検定

次のような問題を考えます。

> あるエンドウマメの交配実験の結果は，メンデルの法則によれば「黄色・丸」「黄色・しわ」「緑色・丸」「緑色・しわ」の４種類の形質のマメが $9:3:3:1$ の割合で現れるはずだという。実際にこの実験を行ったところ，それぞれの形質をもったマメの数は 447, 131, 152, 38 であった。この実験においてメンデルの法則が成り立っているかどうかを有意水準 5% で検定せよ。

この問題で，もしも各形質をもつマメの個数が，正確に理論通りに $9:3:3:1$ の比率になっているとすれば，得られたマメの総数は 768 ですから，それぞれの個数は 432, 144, 144, 48 になるはずです。

しかし，実際には，得られたマメは標本でしかありません。たとえ，世界全体で，各形質のマメが理論通りに $9:3:3:1$ の比率で実っているとしても，この実験で得られたマメは，偶然のために，$9:3:3:1$ にはならないかもしれません。

そこで，実際に観測されたデータをカテゴリに分類したときの比率と，今仮定している理論（モデル）から予想される比率とのずれが，偶然起こりうる程度のものか，それともモデルの方が間違っていると考えたほうがいいのかを検定するのが，**適合度検定**です。ここでは，よく用いられる χ^2（カイ２乗）適合度検定を紹介します。

上の問題を一般的に考えると，次のようになります：母集団が k 個のグループに分けられるとしましょう。ただし，各グループは，同時に２つ以上のグループに属する個体はない (「排反」といいます）とします。実験によって得た n 個の標本のうち，グループ 1〜k に属するものの数が，おのおの $X_1, X_2, ..., X_k$ であったとき，「各グループに属する個体数（例題でいえば豆の数）の比率が $p_1, p_2, ..., p_k$ になる」というモデルが正しいかどうかを検定します。

グループ名	1	2	$\cdots$	k	合計
実際に観測された個体数	X_1	X_2	$\cdots$	X_k	n
理論で予想される比率	p_1	p_2	$\cdots$	p_k	1
理論で予想される個体数	np_1	np_2	$\cdots$	np_k	n

表 1: 適合度検定の考え方。$X_i = np_i$ とみなしてよいかどうかを検定する。

仮定したモデルによれば，n 個からなる標本に対してはグループ 1〜k に属する個体数は $np_1, np_2, ..., np_k$ になるはずです（表を参照）。ここで，各グループについて「現実に観測された個体数 X_i と，モデルか

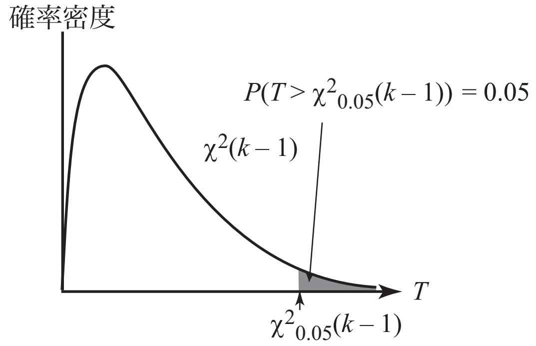

図 1: 自由度 $k-1$ の χ^2 分布の確率密度関数と，上側 5 パーセント点 $\chi^2_{0.05}(k-1)$

ら予想される値 np_i との差」の 2 乗の比率，すなわち $(X_i - np_i)^2/np_i$ を全部のグループに対して合計したもの，つまり

$$T = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i} \tag{1}$$

を考えます。T は，観測値と理論値の差が大きいほど大きくなります。

T は，n が大きいとき，**自由度 $k-1$ の χ^2 分布**という確率分布モデルに近似的にしたがいます。その証明は複雑なのでここでは示しませんが，グループが 2 つ，つまり $k=2$ の場合について，付録に説明を載せておきます。

さて，ここで行う検定では，帰無仮説 H_0 は「グループ $1,2,\ldots,k$ に属する個体数の比率は $p_1,p_2,\ldots,p_k$ である」で，対立仮説 H_1 は「グループ $1,2,\ldots,k$ に属する個体数の比率は $p_1,p_2,\ldots,p_k$ ではない」です。ここで，T の値が大きいほど，理論値と観測値とのずれが大きい，つまり帰無仮説の成立が疑わしい，ということなので，「T がそんな大きな値になる確率は，有意水準以下」のとき帰無仮説を棄却します。

自由度 $k-1$ の χ^2 分布の確率密度関数のグラフ（ヒストグラム）は，図 1 のような形をしています。このとき，自由度 $k-1$ の χ^2 分布にしたがう確率変数 T が，$\chi^2_{0.05}(k-1)$ より大きい確率が 5%であるとします（t 分布の場合と同様，「上側 5 パーセント点」といいます）。

したがって，有意水準が 5%のときは，帰無仮説が正しいとしたときに T の値を計算して，それが $\chi^2_{0.05}(k-1)$ よりも大きければ，帰無仮説は棄却されます。

結論は何なのか？ ー適合度検定の問題点

上の例では，帰無仮説 H_0:「4 つのグループに属する個体数の比率は 9/16, 3/16, 3/16, 1/16 である」，対立仮説 H_1:「4 つのグループに属する個体数の比率は 9/16, 3/16, 3/16, 1/16 でない」となります。そこで (1) 式の T の値を求めると $T = (447-432)^2/432 + (131-144)^2/144 + (152-144)^2/144 + (38-48)^2/48 = 4.22$ となります。4 つのグループに分かれるので自由度は 3 で，有意水準が 5%のとき，$T > \chi^2_{0.05}(3)$ のとき，H_0 は棄却されます。数表より，$\chi^2_{0.05}(3) = 7.8147$ ですから，$T > \chi^2_{0.05}(3)$ ではなく，H_0 は棄却されません。

これまでに説明した検定の考え方では，帰無仮説が「モデルは実測データに適合している」となっていますから，帰無仮説が棄却されたときは「モデルは実測データに適合していないと言える」と結論さ

れ，棄却されないときは「モデルは実測データに適合していないとまでは言えない」という結論が得られるはずです。したがって，今回の例の結論は「メンデルの法則が成り立っていないとまでは言えない」ということになるはずです。

ところが，適合度検定では中心極限定理を用いています（付録参照）から，標本サイズが大きくないと理論の正当性が保たれません。しかし一方で，標本サイズが大きいと棄却域が広くなり，帰無仮説が棄却されやすくなります[1]。例えば，母平均の両側検定を考えてみましょう。標本サイズが大きくなると，母平均の信頼区間は狭くなります。両側検定では，帰無仮説で述べた母平均の値が信頼区間に入っていなければ棄却するわけですから，棄却域は広くなります。したがって，標本サイズが大きくなると，理論比率と実際に観察された比率のほんのわずかの違いによっても帰無仮説が棄却され，「モデルは適合しないと言える」という結論が出てしまいます。

現実には，たとえ理論通りの現象が起こっていても，現象には誤差があり，理論通りの比率が実際に観察されることはまずありません。したがって，標本が多ければ，わずかの誤差を検出して帰無仮説が棄却され，どんな場合でも「モデルは適合しない」という結論が出てしまいます。これでは，この検定は役に立ちません。

そこで，適合度検定では，通常の検定とは逆に「標本サイズが大きく，検出力が十分に高いのに帰無仮説が棄却されないときは，『帰無仮説は正しい。モデルは実測データに適合している』あるいは『実測データはモデルと矛盾しない』と言える」という結論を導く，という考え方があります。しかし，標本が十分多くないときには，棄却されないからといって間違った結論を出してしまうことがあります。また，そんなに標本が多ければ，検定などしなくても，モデルによる比率と実測データにおける比率を直接比較するだけでいいはずです。

第１０回の講義で説明した，検定のそもそもの考え方を思い出してください。それは

> 「標本として何回かだけ温度を測定した時，t 統計量が極端に大きな（あるいは小さな）値になる確率は，（有意水準より）小さな確率でしか起こらない」
> →「『真の沸点が，t 統計量が極端な値になるような温度である』などという仮説を無理に信じるよりも，『そんな仮説はウソ』と考えるほうが自然」

というものでした。このような推測をしなければならないのは，実際に測定できる回数が少ないからです。もし，無限に多くの測定ができるのなら，真の沸点はほぼ一目瞭然です。

つまり，標本サイズが大きくないと成り立たず，しかも「帰無仮説が棄却されないとき，積極的な結論が得られる」という検定は，上のような検定の考え方や有意水準の意味を逸脱したものです。したがって，統計学者の間では適合度検定の有用性を否定する見解もあり，使用には注意が必要です。

付録：χ^2 分布と適合度検定について

χ^2 分布は，標準正規分布から派生する確率分布のひとつです。確率変数 $Z_1, Z_2, \ldots, Z_n$ が互いに独立で，それぞれが標準正規分布 $N(0,1)$ にしたがうとします。このとき，

$$\chi^2 = Z_1^2 + Z_2^2 + \cdots + Z_n^2 \tag{A1}$$

がしたがう確率分布を「自由度 n の χ^2 分布」といい，$\chi^2(n)$ で表します。

[1] 帰無仮説が棄却される確率を「検出力」といいます。教科書 110 ページ，および私の講義・2010 年度後期（広島大）「情報統計学」第１１回を参照してください。

さて，適合度検定において，グループが２つ，つまり $k=2$ の場合を考えます。このとき，本文 (1) 式の T がどうなるかをみてみましょう。グループが２つしかないので，各グループの個体数 X_1, X_2 の合計は全個体数 n で，また各グループの理論比率 p_1, p_2 の合計は１です。このとき，T は

$$\begin{aligned} T &= \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_2 - np_2)^2}{np_2} \qquad \text{(A2)} \\ &= \frac{(X_1 - np_1)^2}{np_1} + \frac{((n - X_1) - n(1-p_1))^2}{n(1-p_1)} \\ &= \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_1 - np_1)^2}{n(1-p_1)} \\ &= (X_1 - np_1)^2 \left\{ \frac{(1-p_1)+p_1}{np_1(1-p_1)} \right\} = \left\{ \frac{X_1 - np_1}{\sqrt{np_1(1-p_1)}} \right\}^2 \end{aligned}$$

となります。

ここで，p_1 は「ある個体がグループ１に属する確率」と考えられるので，「n 個の個体のうち，グループ１に属している数」を表す X_1 は２項分布 $B(n, p_1)$ にしたがいます。$B(n, p_1)$ の期待値は np_1，分散は $np_1(1-p_1)$ ですから，中心極限定理により，n が大きいとき，(A3) 式の $\{\}$ 内は標準正規分布にしたがいます。よって，T は自由度１（すなわち $k-1$）の χ^2 分布にしたがいます。

χ^2 分布は，母分散の推定・検定など，母集団分布が正規分布であるときの，さまざまな統計的推測に現れる確率分布です。また，第９回で説明した t 分布は，χ^2 分布からさらに派生する確率分布モデルです。くわしくは，私の講義 (広島大)・2010 年度後期「情報統計学」第１４，１５回を参照してください。