[X^2検定]
『このレポートには間違いはない』
→99％の信頼度で棄却される

M1 高山正輝

平成22年12月14日

1　X^2 の取扱説明書

X^2 検定の解説を書く前に、X^2 統計量とはいかなる量なのかを書いておく。X^2 分布がいかなる量かはその後紹介する。X^2 にはいくつかの表記が存在するがしばしば出てくるものをいくつか紹介する。

1.1　何回「パネェ！」って言うの？

まず一つ目は、

$$\mathrm{X}^2 = \sum_{i=1}^{N} \frac{(N_i - n_{i,model})^2}{n_{i,model}} \tag{1}$$

である。ここで、N_i は i 番目のビンに入った度数、N はビンの総数、n_{model} はある仮定されたモデルから想定される i 番目のビンに入る期待値、を意味している。N_i は整数でなくてはならないが、$n_{i,model}$ は必ずしも整数でなくともよい。

ところで、X^2 分布を考えるとき自由度という概念を正しく理解していなければならない。かくいう小生自身が十分理解しているとは言いがたいのだが、理解している範囲で解説したい。

まず上の例で自由度が最大になる、すなわち自由度 N になる場合を取り上げる。これは、観測量とモデルとが独立 (という言い方は正しくないかもしれないが) である場合である。つまり N_i と $n_{i,model}$ の間には何ら関係式が存在しないことを指している。例を挙げるなら、『T 君が「パネェ！」と発言する曜日毎の発言回数と、理論的に導かれた発言回数との比較』などである。この場合、観測量あるいはモデルの期待値は互いの値がいくらであろうとそんなの関係ねぇ、である。

	日曜日	月曜日	火曜日	水曜日	木曜日	金曜日	土曜日
観測値 [回]	18	74	3	5	93	24	50
モデル (期待値)[回]	15.7	57.2	8.8	2.1	103.2	36.9	49.5

表 1: ある週の T 君の「パネェ！」の発言回数の曜日毎の分布

ところが、問題が『T 君が一回の「パネェ！」発言に対して、周囲の人 100 人に聞いた T 君の好感度の 5 段階評価の分布』となると、前問と異なり N_i と $n_{i,model}$ の間には $\sum_{i=1}^{N} N_i = \sum_{i=1}^{N} n_{i,model} = 100$ という関係を要求する。つまり、観測の総量 (この場合「100 人」) によってモデルの期待値を規格化したために自由度が減り、この場合の自由度は $N-1$ となる。

	ステキ	好き	KY(カッコよい)	JK(実はキテる)	IR(居ない方が楽)
観測値 [人]	30	38	21	9	2
モデル (期待値)[人]	31.4	39.7	19.3	5.7	3.9

表 2: T 君の「パネェ！」発言の好感度の分布

以上のように観測量とモデルの間に関係式が増えるほど、すなわち束縛が厳しくなるほど、自由度は奪われていく。

こうして得られた X^2 はその自由度に対応した X^2 分布に従う (と勝手に思う) のである。厳密には上の X^2 が X^2 分布に従うかは怪しい。その理由は、各ビンでの $\sqrt{\frac{(N_i-n_{i,model})^2}{n_{i,model}}}$ という値が正規分布しているとは限らないためである。勘違いしないでもらいたいのは、i について正規分布しているか否か、を議論しているわけではないということである。先ほどの曜日毎の「パネェ！」の発言回数を例にすると、137 週間観測したときの日曜日の $\sqrt{\frac{(N-n_{model})^2}{n_{model}}}$ の値が正規分布か否かを今は議論しているのであって、一週間の分布が正規分布しているか否かを議論しているのではない (水曜日に最大値を取り、標準偏差 2 日とかだったらマジ「パネェ！」)、ということである。しかしながら、ビンの数 N が十分大きい、あるいは各ビンの度数 N_i が十分大きければ帰無仮説 (観測量はモデルから予想される分布に従う) が成り立つ場合に X^2 の分布は X^2 分布で近似できる (らしい)。従って、「$\sqrt{\frac{(N_i-n_{i,model})^2}{n_{i,model}}}$ がどういう分布なのか知らないけれど、正規分布であると勝手に思って検定しました。その時 90 %の信頼度でモデルと一致していました。」ということを前半部分については目をつむって行っている、、、と私は解釈した。違うかもしれない。ちゃんと知りたい人は numerical recipes in C の 13 章 5 節を読んで勉強してほしい。

1.2　T 君が「パネェ！」と言うと、L さんが「へっ？」と言う？

前のセクションで紹介した事柄を応用すると、二つの現象に相関があるかどうかの検定も可能である。例によって例題を使って説明する。

例題『曜日毎に T 君が「パネェ！」と言う回数と、L さんが「へっ？」と言う回数には何か相関がある』ことを示せ。

ここで帰無仮説 (X^2 の値が小さいことに対応する) は「何か相関がある」とする。つまり期待値を求めるモデルには「相関がある」と思って作ったモデルを起用する。

	日曜日	月曜日	火曜日	水曜日	木曜日	金曜日	土曜日	total[回]
T 君 [回]	18	74	3	5	93	24	50	267
L さん [回]	5	64	21	3	17	694	94	898
total[回]	23	138	24	8	110	718	144	1165
T 君のモデル [回]	$\frac{23*267}{1165}$	$\frac{138*267}{1165}$	$\frac{24*267}{1165}$	$\frac{8*267}{1165}$	$\frac{110*267}{1165}$	$\frac{718*267}{1165}$	$\frac{144*267}{1165}$	267
L さんのモデル [回]	$\frac{23*898}{1165}$	$\frac{138*898}{1165}$	$\frac{24*898}{1165}$	$\frac{8*898}{1165}$	$\frac{110*898}{1165}$	$\frac{718*898}{1165}$	$\frac{144*898}{1165}$	898

表 3: ある週の曜日毎の T 君の「パネェ！」の発言回数と L さんの「へっ？」の発言回数の分布と、その期待値

モデルの値 $n_{ij,model}$($\rightarrow$ i さんの j 曜日の期待値) を観測量 N_{ij} から求めるわけだが、どのようにして求めたかというと、個々人の分布に相関があると思っているので、j 曜日の度数の total の内 i さんの割合は、一週間の合計 (全体) の度数の内の i さんの取り分の割合に等しい、と考えるのである。

$$\frac{n_{ij,model}}{N_{,j}} = \frac{N_{i,}}{N} \tag{2}$$

$$N_{,j} = \sum_i N_{ij} \tag{3}$$

$$N_{i,} = \sum_j N_{ij} \tag{4}$$

$$N = \sum_i N_{i,} = \sum_j N_{,j} \tag{5}$$

また X^2 は以下のように書かれる。

$$\mathrm{X}^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{(N_{ij} - n_{ij,model})^2}{n_{ij,model}} \tag{6}$$

これは自由度 $(2-1)(7-1) = 6$ の X^2 分布に従う。なぜなら各曜日の total が決まっていて (T 君が求まれば、自動的に L さんの度数が解ってしまう)、一週間の total も知っている (T 君の日曜から金曜まで求まれば、自動的に土曜日の度数が解ってしまう) ためである。くどいようだが今帰無仮説として「相関がある」を選択したので X^2 が小さいほど相関がある可能性が高い、と言える。

1.3 ご冗談でしょう、Y さん！

ここまでは観測量は各ビンに入ったカウント数 (度数) を扱ってきた。しかし実際の観測量は度数だけではないはずである。そのような場合でも X^2 を定義することは可能か？答えはイエスである。今観測量 f_i が母平均 μ、母分散 v の正規分布に従う (こういう分布を $N(\mu, v)$ と書く) とき、$\frac{f_i - \mu}{\sqrt{v}}$ は $N(0,1)$ の正規分布に従う確率変数である。従って、

$$\mathrm{X}^2 = \frac{1}{v} \sum_{i=1}^{N} (f_i - \mu)^2 \tag{7}$$

で書かれる X^2 は自由度 N の X^2 分布である。また母平均 μ がわからないときに μ の代わりに標本平均

$$\overline{f} = \frac{1}{N} \sum_{i=1}^{N} f_i \tag{8}$$

を用いて、

$$\mathrm{X}^2 = \frac{1}{v} \sum_{i=1}^{N} (f_i - \overline{f})^2 \tag{9}$$

とすることもある。このときは「f_i は自由に取ってきてもいいよ。でも平均値は $\overline{f}$ になるようにね」と言う制限が一つ付くので、自由度は $N-1$ となるのである。こうして観測量から $\sum_{i=1}^{N}(f_i - \overline{f})^2$ が具体的に求まるので、「90 %の信頼度で v はこれこれの範囲に収まる」といった議論が出来るのである。最後に検定に移る前座としてゼミでも取り上げた「Y の悲劇」を例に挙げたい。

あるところに、周囲に「私は銀河を照らす巨星だ、いや超巨星だ」などと戯言をのたまう大学院生、Y さんがいました。Y さんは自分の身長を聞かれると決まって「197cm！141cm なんて測定誤差の内だしぃ」などととち狂った発言を繰り返していました。しかし周囲の人間にはどう見ても 140cm 代にしか見えません。そこで数理統計が得意な Y さんの担当教官である M 准教授は、その真偽を確かめるためにゼミの時こっそりとその身長を 5 回観測しました。以下がその結果

	1 回目	2 回目	3 回目	4 回目	5 回目
観測値 [cm]	142.5	143.0	141.9	142.2	142.4

表 4: Y さんの身長の観測値

ここから、

$$\overline{f} = \frac{1}{N}\sum_{i=1}^{N} f_i = 142.4 \tag{10}$$

を得る。したがって、

$$\mathrm{X}^2 = \frac{0.66}{v} \tag{11}$$

となる。ここで X^2 検定に用いる適当な数表を持ってくると

α	0.99	0.95	0.05	0.01
自由度 (4)	0.30	0.71	9.49	13.28

表 5: 自由度 4 の X^2 の数表

である。ここで、M 准教授はもっとも信頼できる母分散 v の値の範囲を推定してみることにした。信頼できるとは「何%の信頼度でこの範囲に入っている」の「何%」の値の大きなものである。数表を見ると α $= 0.99$(これより大きいところに 99 %の信頼度で入っている) と、α $= 0.01$(これより小さいところに 99 %の信頼度で入っている) とがある。したがって、M 准教授は「98 %の信頼度で母分散 v の範囲を推定」した。

$$0.30 < \frac{0.66}{v} < 13.28 \tag{12}$$

$$0.05 < v < 2.2 \tag{13}$$

$$0.22 < \sqrt{v} < 1.48 \tag{14}$$

結果から言えることは、『測定誤差は 98 %の信頼度で、0.22cm より大きいが、1.48cm よりは小さいだろう』ということである。この結果をうけて M 准教授は見事長年謎とされていた Y さんの身長に対し「197cm？たわごと。せいぜい $142.4 + 1.48cm$ だし！」という結論を出すことに成功したのである。かわいそうな Y さんは自分の身長が数理統計的に暴かれてしまい、以後 197cm などと見栄を張れなくなってしまったとさ。めでたしめでたし！

2 X^2 分布と X^2 検定

2.1 X^2 分布

ここまで X^2 にいくつかのバリエーションがあることを見てきたが、この節ではそもそも X^2 分布とはいかなる量か、その定義は何か、に触れていきたい。今まで書いてきた X^2 を思い出してもらいたい。いくつか表記の仕方があったが

どれも $\sum_{i=1}^{N} X_i$ と言う形で書かれ、確率変数 X_i が $N(0,1)$ の正規分布にしたがう (と勝手に思う) という表記になっている。実はこれは X^2 分布の定義そのもので、数理統計の教科書などでは、

定義
『標準正規分布 $N(0,1)$ に従う確率変数 X_i(この内 $N-M$ 個が独立) の二乗和 $\sum_{i=1}^{N} X_i^2$ の従う分布を自由度 $N-M$ の X^2 分布といい、X^2_{N-M} と書く』

『自由度 n の X^2 分布は平均 $m=n$、分散 $v=2n$ の非負実関数上の連続分布で、密度関数は

$$f_n(z) = \frac{1}{2\ \Gamma\ (\frac{n}{2})}(\frac{z}{2})^{\frac{n}{2}-1}e^{-\frac{z}{2}} \tag{15}$$

$$\Gamma\ (s) = \int_0^\infty x^{s-1}e^{-x}dx,\quad s>0 \tag{16}$$

で与えられる。』

ここで、X^2 の値の意味は、帰無仮説 ($\mathrm{X}^2=0$ となる説) が、

$$\alpha\ = \int_{\mathrm{X}^2}^\infty f_n(z)dz \tag{17}$$

の信頼度で棄却される、あるいは $1-\ \alpha$ の信頼度で採択される、と言う意味である。

2.2　X^2 検定

　X^2 の持つ意味がわかったところで実際の検定の考え方に触れていこうと思う。と言っても既に答えに近いヒントはたくさん出ているのでここではそれをまとめるだけなのだが。
　前節で信頼度について触れたが、これは例えば次のようにも言い換えることもできる。

『

$$\alpha\ = \int_c^\infty f_n(z)dz \tag{18}$$

となる c について X^2 が

$$\mathrm{X}^2 > c \tag{19}$$

を満たすとき、帰無仮説は (少なくとも)$1-\ \alpha$ の信頼度で棄却される。』

『

$$1-\ \alpha\ = \int_0^c f_n(z)dz \tag{20}$$

となる c について X^2

$$\mathrm{X}^2 < c \tag{21}$$

を満たすとき、帰無仮説は (少なくとも) α の信頼度で採択される。』

というのが X^2 検定である。α を自分の求める制度に設定し、それに対応する c の値と求められた X^2 の大きさを比較し議論するのである。この際 c の値というのは数表として出ているので適当な文献等で確認して欲しい。

3 天文的 X^2fit

天文で使われる表記としてよく以下のようなものを目にする。

$$\mathrm{X}^2 = \sum_{i=1}^{N} \frac{(f_i - f_{i,model})^2}{\sigma_i^2} \tag{22}$$

今までと何が違うかというと、個々の i について標準偏差 (→分散) が異なる点である。これは、今までの表記では個々の f_i の従う分布は $N(\mu, v)$ の正規分布であったのに対し、上の表記は個々の f_i の従う分布は $N(\mu_i, \sigma_i^2)$ の正規分布であることを意味しており、すなわち個々の index i について母集団が異なる場合の X^2 の表記であることを示している。個々の母集団が異なるとはどういうことかというと、変光星の光度 (f_i) の時間変化 (t_i) などがこれに相当する。各時刻 t_i での光度が μ_i、観測のエラーバーが σ_i となるので、この時の観測量 f_i の従う母集団は $N(\mu_i, \sigma_i^2)$ の正規分布に従う (と我々は思い込んでいる) のである。繰り返しになるが、そのような f_i を $\frac{f_i - f_{i,model}}{\sigma_i}$ と言う形に焼き直すことで母分布を $N(0,1)$ の標準正規分布に規格化しているので、その二乗和は X^2 分布になるのである。よって先と同様な検定が行える。

補足だが、次のような表記を目安にするのもいいかもしれない。

$$\frac{\mathrm{X}^2}{N} = \frac{1}{N} \sum_{i=1}^{N} \frac{(f_i - f_{i,model})^2}{\sigma_i^2} \tag{23}$$

これは個々の i での観測量とモデル値がエラーバーに対しどれだけズレているか、の平均を見ている。従って大雑把にいって $\mathrm{X}^2 < 1$ であればモデルは観測のエラーバーの中にはいっているので (一見) 良さそうだといえる。

最後にモデルのパラメーターの信頼度についてほんの少しだけ触れておく。

A さんと B さんが測った観測量から求めたモデルのパラメーター (X^2 を最小 (極小) にする) は一般に二人の観測結果が異なるので一致することはない。そこで、自分の得た観測結果から導かれたモデルのパラメーター (これは真のパラメーターの推定値にすぎない) をもとに真の値がどの範囲になら何%の信頼度で存在するか、を知るのが狙いである。概念だけの説明になってしまうが、例えばモデルとして

$$f_{i,model} = ax_i + b \tag{24}$$

を考える。この場合モデルのパラメーターの数は二つ (制限が二つ) なので自由度は $N-2$ となる。次にこのモデルが真の値であると思って、これから予想される擬似的観測量 $f_i^{(s)}$ $(s = 1, 2, 3, , ,)$ を $N(\mu_{i,model}, \sigma_i^2)$ の正規分布に従う乱数としていくつか $(s >> 1)$ 発生させ、これらから擬似的 X^2 を作り同じく擬似的パラメーター $a^{(s)}, b^{(s)}$ を得る。擬似的パラメーターの分布はある確率密度で書かれる (我々がそれを具体的な関数の形で書けるか否かによらない) ので、例えば

発生させた乱数の内 68 %が入る範囲を囲めばめでたく最初に求めたパラメーターからこの範囲内に真のパラメーターが 68 %の信頼度でありそう、等といえるのである。残念ながら小生が未だに十分理解できていないので、ここまでしか紹介することができない。詳しいことは numerical recipes in C の 14 章 5 節を見てほしい。

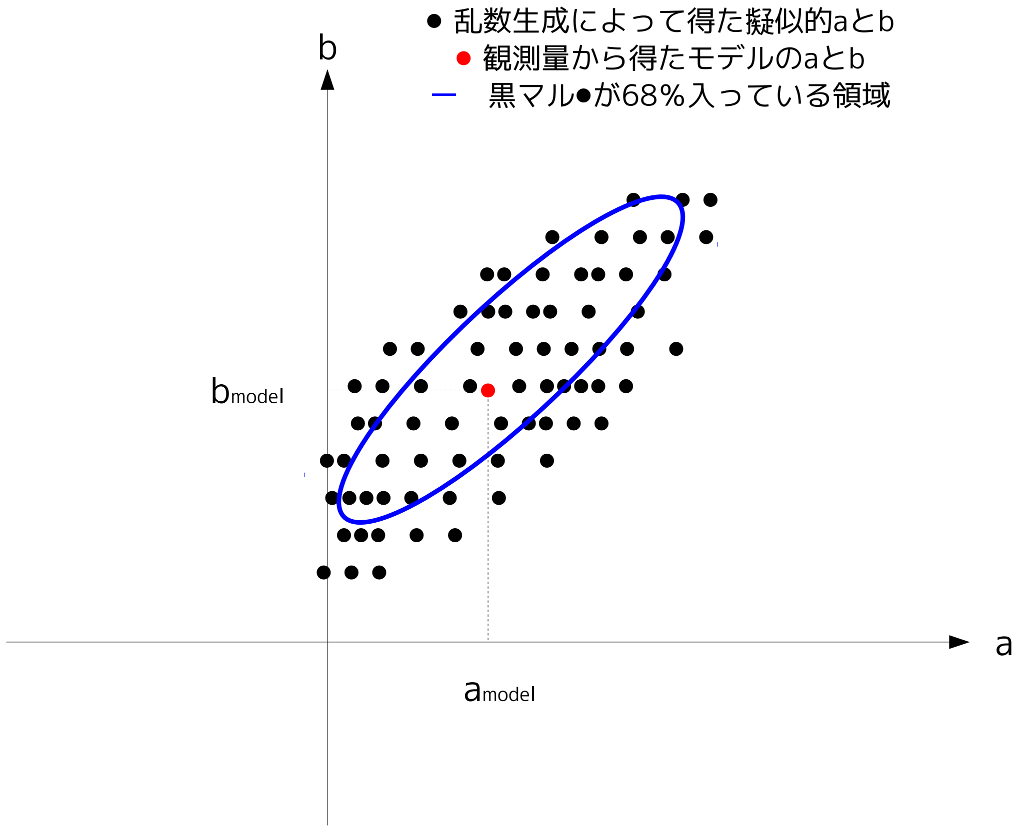

図 1: パラメーター推定のイメージ

最後に一言。X^2 検定はある信頼度でモデルが観測量に fit している、といっているだけで、モデルがある信頼度で (真の) 正しい答えであるといっているわけではない。よく fit したからと言って自分のモデルが物理的に正しいというのは早計である。以上、X^2 検定の紹介でした。