白鷗大学論集　第24巻　第２号

研究ノート

χ^2分布，t分布，F分布（１）

沖津　　直

THE χ^2, t and F Distributions

OKITSU Tadashi

はじめに
Ⅰ　χ^2分布
　1．χ^2分布の定義とその性質
　2．分散の区間推定
　　(1)母平均既知の場合
　　(2)母平均未知の場合
　3．分散の検定
　　(1)母平均既知の場合
　　(2)母平均未知の場合

沖 津 直

はじめに

推測統計において、統計量の標本分布が重要な役割を演ずる。最初の段階で学習する基本的な標本分布としては、標本平均の標本分布と標本比率の標本分布があるけれども、これからも新たな標本分布が必要となります。これらの分布がχ^2分布、t分布、F分布である。χ^2分布、t分布、F分布はいずれもいろいろな推定・検定において必要になる重要な標本分布である。これらの分布は推測統計のための手段として導入されたものであるけれども、これら分布の具体的な確率密度関数などをひとつひとつ記憶する必要はない。大切なのは、定義、グラフの形、それぞれの分布表から確率の見つけ方と使い方などをしっかり理解しておくことである。本稿では、χ^2分布とその分布表の使い方に主眼をおきながら、分散の推定・検定を通じてそれらがどのように利用されているかを説明していきたい。

I　χ^2分布

1．χ^2分布の定義と性質

標本平均や標本比率の２つの基本的な標本分布から、さらに一歩進んだ代表的な標本分布として、χ^2分布、t分布、F分布の３つがある。まず、χ^2分布であるが、平均０、分散１の標準正規母集団N（０，１）より抽出された大きさnの標本変量

$Z_1, Z_2, \cdots\cdots, Z_n$

をそれぞれ２乗して合計したものをχ^2とおくと、

$$\chi^2 = Z_1^2 + Z_2^2 + \cdots\cdots + Z_n^2 \qquad (1)$$

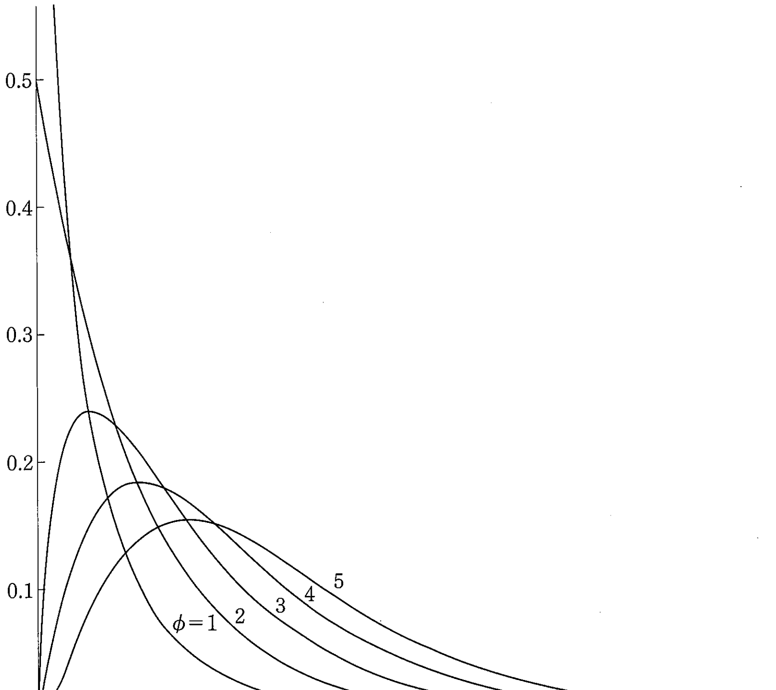
となる。この確率変数χ^2の確率密度関数は、証明は省略するが

$$\left.\begin{aligned} f(\chi^2) &= \frac{1}{2\Gamma(\frac{\phi}{2})}\left(\frac{\chi^2}{2}\right)^{\frac{\phi}{2}-1} . e^{\frac{-\chi^2}{2}} && \chi^2 > 0 \\ &= 0 && \chi^2 \leqq 0 \end{aligned}\right\} \qquad (2)$$

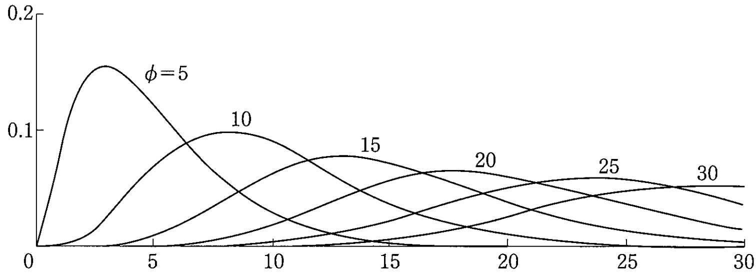
で与えられる。このとき、χ^2分布は自由度$\phi = n$のカイ二乗分布に従うという。ここに、Γ（ギリシャ文字でガンマと読む）は

$$\Gamma(m) = \int_0^\infty x^{m-1}.e^{-x}dx \, (m > 0)$$

で定義されるガンマ関数を表す記号である。(参考文献3 p158参照) χ^2分布は正規母集団を前提としているが、実際の場合、母集団分布が正規でないことも少なくない。幸いに、正規母集団の条件が正確に満たされなくとも、χ^2分布の形は大きな影響を受けない。このような性質をもつ標本分布を正規分布の条件に対して、頑丈な分布という。χ^2分布はかなり頑丈な分布である。後に説明するF分布も正規性の条件に対してかなり頑丈な分布である。

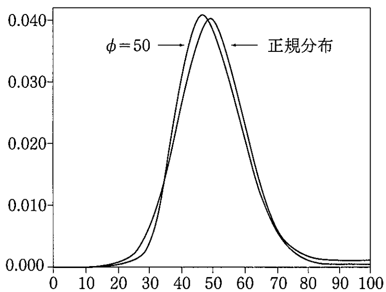


1図　自由度1，2，3，4，5のχ^2分布

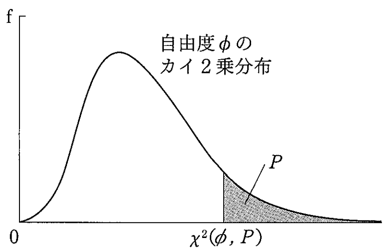


２図　自由度５，10，15，20，25，30のχ^2分布

χ^2分布の形は、自由度$\phi = n$の大きさによって決まる。つまり、自由度ϕはχ^2分布の母数である。χ^2分布の形は、一般に非対称であるが、自由度ϕが大となるにしたがって、同図からもわかるように次第に正規分布に接近していく。１図および２図はこの様子を表している。自由度１のχ^2分布は、０に近づくにつれて値が大きくなる。自由度３以上になると、右すそに長い分布になり、自由度が高くなるにつれて山の高さは低く、中心も右へ移動していく。山の頂上の位置や左右非対称の度合いも自由度によって変わる。そして、自由度が30～50ぐらいになると分布は正規分布にかなり一致してくる。３図は、自由度が50の場合のカイ二乗分布（平均50、分散100）とそれと同じ平均と分散をもつ正規分布の図が描かれているが、かなり重なりあっていることが視覚的にわかる。



３図　自由度50のカイ２乗分布と、それと同じ平均と分散をもつ正規分布のグラフ



４図　χ^2_P以上となる上側確率

χ^2分布の期待値と分散は次の式で与えられる。

E（χ^2）＝ϕ

V（χ^2）＝2ϕ

ここで、Eは期待値、Vは分散を表している。

したがって、χ^2分布は自由度ϕがそのまま分布の平均値となり、自由度を2倍したものが分布の分散となる。χ^2分布は自由度ϕによって形が変わるので、その数値は自由度ごとに計算する。4図に示すように、たとえば特定の数値χ^2に対して$\chi^2 > \chi^2_P$となる確率は

$$P(\chi^2 > \chi^2_P) = \int_{\chi^2_p}^{\infty} f(\chi^2) d\chi^2$$

で計算される。χ^2分布表によっては、実際によく使ういくつかの上側確率Pに対してχ^2の値を自由度ごとに計算したものを、通常使用している。χ^2（ϕ，P）を自由度ϕのカイ二乗分布の上側100％P点と呼んでその値は付表1（出所：参考文献1 p1081参照）に収録されている。一般に、上側というのは、分布に対面して右側のある任意の値よりも大きい領域の確率を示す言葉である。たとえば、上側2.5％点というのは、右側の領域の確率が2.5％となる横軸上における変数の値のことである。また、実際に必要となるのは、上側2.5％点である。左側の領域の面積に該当する確率が2.5％となる変数の値のことを上側97.5％点という。つまり、上側2.5％点と上側97.5％点の横軸上の変数の値は、推定や検定を行うときの重要な境界値になるのである。

ところで、χ^2分布は、t分布やF分布の統計量の理論的分布を導出するためにも用いられるという意味でも重要である。いま、正規母集団N（μ，σ^2）より、独立にそれぞれ取り出される大きさnの標本変量X_1, X_2,……，X_nより、標本分散という統計量をつくる。

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad (3)$$

ここで、$(X_1 - \overline{X})^2$，$(X_2 - \overline{X})^2$，・・・，$(X_n - \overline{X})^2$をそれぞれσ^2で割ったものの総和をχ^2とおくと、

$$\chi^2 = \left(\frac{X_1 - \overline{X}}{\sigma}\right)^2 + \left(\frac{X_2 - \overline{X}}{\sigma}\right)^2 + \cdots + \left(\frac{X_n - \overline{X}}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\sigma}\right)^2 \qquad (4)$$

となる。n 個の変数 $\left(\frac{X_i-\bar{X}}{\sigma}\right)^2$ は、それぞれ標準正規分布Ｎ（０，１）に従うので、（４）式の分子 $\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ が nS^2に等しいので、（４）式は

$$\chi^2=\frac{nS^2}{\sigma^2} \qquad (5)$$

とも書ける。しかし、母平均値が未知なので（５）式の自由度は $n-1$ の χ^2分布に従うことになる。したがって、nS^2/σ^2であるカイ二乗値は、（２）式の ϕ のかわりに $n-1$ を入れると、つぎの確率密度関数で与えられる。

$$\left.\begin{aligned} f(\chi^2) &= \frac{1}{2\Gamma(\frac{n-1}{2})}(\frac{\chi^2}{2})^{\frac{n-2}{2}}.e^{-\frac{\chi^2}{2}} && \chi^2>0 \\ &=0 && \chi^2 \leqq 0 \end{aligned}\right\} \qquad (6)$$

ここで、自由度がｎ－１になるというのは、

$$\sum_{i=1}^{n}(X_i-\bar{X})=0$$

なる関係が成立し、$\bar{X}$ が定数として与えられることになるために、n 個の変量のうちｎ－１個の変量の値が決まれば、残りは自ずとその値が決まるからである。１個の制約条件をもつことになり、自由度は n から１個だけ減じて、$n-1$ の自由度になるためである。

付表の χ^2分布表より、自由度 ϕ と右すその確率 P の組み合わせからいくつかの χ^2の値を求めてみよう。たとえば、自由度 $\phi=15$、$P=0.05$および $P=0.01$に対応する $\chi^2_{0.05}$と $\chi^2_{0.01}$の値は、それぞれ

$$\chi^2_{0.05}=25 \qquad \chi^2_{0.01}=30.6$$

である。χ^2の右下の添え字は上側確率を示している。あるいは、χ^2（15, 0.05）＝25、χ^2（15, 0.01）＝30.6とも表記する。また、$\phi=20$、Ｐ＝0.05および $\phi=20$、Ｐ＝0.95と対応する χ^2の値は、それぞれ

$$\chi^2_{0.05}=37.6 \qquad \chi^2_{0.95}=37.7$$

となる。ここで、通常、右すその確率を一般的に P とか α で表示する。自由度 ϕ と右すその確率 α に対応する χ^2の値は、χ^2（ϕ，α）、左側に α と同じ確率をとる χ^2値の左境界値は、χ^2（ϕ，１－α）と書くことがで

きる。χ^2（ϕ，α）を上側確率がαとなる境界値（前述の100％α点）と呼んでいるが、実際には上側確率と同じ大きさになる左側の確率に相当する境界値χ^2（ϕ，1－α）もほしいのである。また、χ^2分布の両すそ側からの確率が均等にα／2になるようなχ^2値は、χ^2（ϕ，α／2）の上側確率と左側からみた確率がα／2となる境界値χ^2（ϕ，1－α／2）もほしいのである。χ^2分布表の見方は、まず自由度ϕ、次に有意水準から決まる右すその確率Pを見つけ、その行と列の交差した数値がχ^2値であり、それは推定や検定において境界値あるいは臨界値となるのである。

2．分散の区間推定

（1）母平均既知の場合

さて、平均μ、分散σ^2の正規母集団から抽出された大きさnの無作為標本から計算される標本分散をS^2とし、母分散σ^2に関する信頼係数95％の信頼区間を考えよう。いま、n個の標本変量をX_1, X_2・・・X_nと表し、各標本変量から母平均を引いて標準偏差で割ったものをχ^2とおくと、

$$\chi^2=\left(\frac{X_1-\mu}{\sigma}\right)^2+\left(\frac{X_2-\mu}{\sigma}\right)^2+\cdots+\left(\frac{X_n-\mu}{\sigma}\right)^2=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \quad (7)$$

となる。これは、（1）式で示したように自由度$\phi=n$のχ^2分布に従った。さらに、（7）式は、$V^2=\sum_{i=1}^{n}(X_i-\mu)^2/n$とおけば、

$$\chi^2=\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}=\frac{nV^2}{\sigma^2} \quad (8)$$

とも書ける。この式の分子は、母平均μからの標本変量Xの偏差二乗和を表している。χ^2統計量には2つのパラメータμとσ^2が含まれるが、μは既知であるから、σ^2だけが未知である。そこで、母分散σ^2に関する信頼係数95％の信頼区間をつくるには、2図の斜線を施した部分が左右とも0.025となるような上側97.5％点となる$\chi^{2i}_{0.975}$値および上側2.5％点の$\chi^{2i}_{0.025}$値をχ^2分布表より求めればよい。すなわち、

$$\mathrm{P}\left(\chi^2_{0.975}<\chi^2<\chi^2_{0.025}\right)=1-0.05=0.95 \quad (9)$$

この式のカッコ内の不等式は、χ^2の値のかわりに（8）式にもどすと、

$$\chi^2_{0.975} < \frac{nV^2}{\sigma^2} < \chi^2_{0.025} \quad あるいは \quad \chi^2_{0.975} < \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} < \chi^2_{0.025}$$

となるから、つぎの等式と同等である。

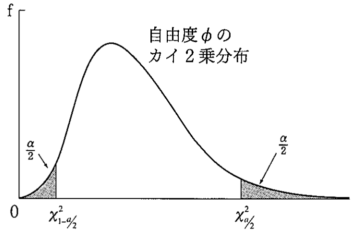
$$\frac{nV^2}{\chi^2_{0.025}} < \sigma^2 < \frac{nV^2}{\chi^2_{0.975}}$$

したがって、（9）式のカッコ内は、

$$P(\frac{nV^2}{\chi^2_{0.025}} < \sigma^2 < \frac{nV^2}{\chi^2_{0.975}}) = 0.95$$

と書くことができる。一般に

$$P(\frac{nV^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{nV^2}{\chi^2_{1-\alpha/2}}) = 1 - \alpha$$



５図　カイ２乗分布の２つの右と左の境界値

となるから、母分散σ^2に関する信頼係数（１－α）×100％の信頼区間は、

$$\frac{nV^2}{\chi^2_{0.975}} < \sigma^2 < \frac{nV^2}{\chi^2_{0.025}} \tag{10}$$

で与えられる。nV^2のかわりに$\sum_{i=1}^{n}(X_i - \mu)^2$で置き換えてもなんらかわらない。自由度は母平均$\mu$既知であるから$n$となる。

例題１　あるクラスから無作為にある科目の６人の学生の成績を得たとしよう。点数は正規分布に従っているものとして、母分散σ^2の信頼係数95％に対する信頼区間を求めてみよう。ただし、母平均μは過去の実績から経験的にわかっており、$\mu = 68$とする。

75、 70、 60、 75、 80、 60

自由度ϕ＝n＝６であるから、χ^2分布表より、上側2.5％点および上側97.5％点

$$\chi^2_{0.025} = 14.45 \qquad \chi^2_{0.975} = 1.237$$

を見つける。つぎに、nV^2は次のようにして求める。

$$nV^2 = \sum_{i=1}^{n}(X_i - \mu) = (75-68)^2 + (70-68)^2 + (60-68)^2 + (75-68)^2 + (80-68)^2 + (60-68)^2 = 374$$

したがって、母分散σ^2に関する信頼係数95％の信頼区間は、(10) 式にnV^2と$\chi^2_{\alpha/2}$と$\chi^2_{1-\alpha/2}$の値をそれぞれ代入してσ^2について整理すると

$$\frac{374}{14.45} < \sigma^2 < \frac{374}{1.237} \qquad \therefore 25.9 < \sigma^2 < 302.3$$

と求まる。χ^2の記号の右下の添え字に$\alpha/2$と$1-\alpha/2$、ある任意の確率α、つまり有意水準の大きさを左右均等にわけているからである。よって、標準偏差の信頼区間は、上の区間の正の平方根をとって、

$$5.09 < \sigma < 17.39$$

となる。

以上のように、未知母数σ^2の存在範囲が、不等式という形で示されるのである。ただ、ここでこの例題1のように、標本の大きさは、6人の学生の成績にすぎないので、このσ^2の信頼係数95％に対する信頼区間を求めると、σ^2の存在範囲が25.9から302.3とその範囲が広めになるのである。実際、統計家からみれば、その範囲が広すぎて漠然としすぎて不満に感ずるかもしれない。その理由は、明らかに標本の大きさが6しかなく、データから得られる情報量が乏しいからである。σ^2やσの存在範囲を絞り込みたいと思うなら、標本の数をもっと増やさなければならない。標本の数を大きくすれば、データの情報が増えるので、未知母数σ^2あるいはσの存在範囲は狭い不等式の区間で表示でき、絞ることができるのである。このような意味で、標本の大きさは推測統計の精度に合わせて標本の大きさを決めていかなければならないのである。また、例題の文章において、正規分布に従っているということは、言外に母集団の大きさが非常に大きいことを意味している。そして、母集団の大きさが小さければ小さいほど、母集団分布のヒストグラムは階段状のものになっており、曲線で描いた図と違って粗くなっているのである。

（2）母平均未知の場合

母平均μが未知の場合は、前述したように（7）式のμの代わりに$\overline{X}$を用いた前述した（4）式となる。

$$\chi^2 = (\frac{X_1 - \overline{X}}{\sigma})^2 + (\frac{X_2 - \overline{X}}{\sigma})^2 + \cdots\cdots + (\frac{X_n - \overline{X}}{\sigma})^2 = \frac{\sum_{i=1}^{n}\left(\overline{X}_i - \overline{X}\right)^2}{\sigma^2} \qquad (4)$$

さらに、ここで、標本分散の（３）式より、$\sum_{i=1}^{n}(X_i - \overline{X})^2 = nS^2$ であったから、母平均未知の場合、

$$\chi^2 = \frac{nS^2}{\sigma^2} \qquad (5)$$

とも書ける。このχ^2統計量は、前述したように自由度$\phi = n - 1$のχ^2分布に従った。自由度がnでなく$n - 1$になることに注意しておこう。したがって、母平均μが未知の場合、

$$P(\frac{nS^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{nS^2}{\chi^2_{1-\alpha/2}}) = 1 - \alpha \qquad (11)$$

と書くことができるから、一般に、σ^2未知の場合の信頼係数（$1 - \alpha$）×100％の信頼区間は、(11) 式のかっこ内の式で与えられる。ただし、ここではnS^2の代わりに $\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ として表現しておいてもよい。くれぐれも自由度は$n - 1$になることだけは注意しておこう。

例題２　例題１において、母平均μが未知であるとして、σ^2に関する信頼係数95％の信頼区間を求めてみよう。

$$\overline{X} = \frac{1}{6}(75 + 70 + 60 + 75 + 80 + 60) = 70$$

$$nS^2 = \sum_{i=1}^{n}(X - \overline{X})^2 = (75-70)^2 + (70-70)^2 + (60-70)^2 + (75-70)^2 + (80-70)^2 + (60-70)^2 = 350$$

一方、自由度$\phi = n - 1 = 6 - 1 = 5$であるから、χ^2分布表より、上側2.5％点および上側97.5％点となるχ^2の値は、それぞれ

$$\chi^{2i}_{0.025} = 12.83 \qquad \chi^{2i}_{0.975} = 0.834$$

である。これらを信頼区間の公式（11）式に代入して未知母数σ^2について整理すると、

$$350／12.83 < \sigma^2 < 350／0.834$$

$$27.28 < \sigma^2 < 420$$

という母分散σ^2に関する信頼係数95％の信頼区間が得られる。母平均μが既知の場合と比較して、未知分散の存在範囲が広くなっていることがわ

かる。母平均μがわからないので、そのぶん情報量が少なくなっているので当然である。母分散未知の場合、自由度が$n-1$になることに注意しておこう。また、標準偏差σの信頼区間は、上の式の正の平方根をとって、およそ

$$5.22 < \sigma < 20.49$$

となる。

なお、標本の大きさが大きい大標本の場合には、標準偏差の信頼区間はつぎの要領で比較的簡単に近似計算することができる。正規母集団より標本抽出する場合、標本の大きさnが$n>30$ないし50以上であるとき、標本標準偏差Sの分布は、平均がE(S)、分散V(S)のつぎの式

$$E(S)=\sigma \qquad V(S)=\sigma/2n$$

の正規分布で近似されることが証明されている。したがって、母標準偏差σに関する信頼係数95%の信頼区間はつぎのようになる。

$$S-1.96\frac{\sigma}{\sqrt{2n}}<\sigma<S+1.96\frac{\sigma}{\sqrt{2n}} \tag{12}$$

この不等式は、つぎの形に書き直すことができる。

$$\frac{S}{1+\frac{1.96}{\sqrt{2n}}}<\sigma<\frac{S}{1+\frac{1.96}{\sqrt{2n}}} \tag{13}$$

この式がσに関する信頼係数95%の信頼区間を定める。大標本の場合、σの信頼区間はこの式で近似的に計算することができる。

例題３　6章の新生児の体重の例（参考文献4 p127参照）において、母分散に関する信頼係数95%の信頼区間を求めてみよう。$n=1000$であるから、大標本の場合である。

（11）式を用いて、母分散σ^2に関する信頼係数95%の信頼区間を求めてみよう。$n=1000$、$S^2=0.269$、また、$\chi^2_{\alpha/2}$および$\chi^2_{1-\alpha/2}$は付表のχ^2分布表の欄外に示してあるように、$\phi>1000$のとき、

$$\chi^2_{0.025} = \frac{1}{2}(1.96 + \sqrt{(2\times 999)-1})^2 = 1088$$

$$\chi^2_{0.975} = \frac{1}{2}\left\{(-1.96) + \sqrt{(2\times 999)-1}\right\}^2 = 912.8$$

これらの値を（10）式に代入すると、

$$\frac{1000\times 0.269}{1088} < \sigma^2 < \frac{1000\times 0.269}{912.8}$$

$$\therefore\quad 0.247 < \sigma^2 < 0.295$$

したがって、母標準偏差σに関する信頼係数95％の信頼区間は正の平方根をとって、

$$\therefore\quad 0.497 < \sigma < 0.543$$

となる。あるいは、この場合、大標本であるから、直接（13）式に代入して求めると$S=0.519$であるから、

$$\frac{0.519}{1+\dfrac{1.96}{\sqrt{2\times 1000}}} < \sigma < \frac{0.519}{1-\dfrac{1.96}{\sqrt{2\times 1000}}}$$

$$\therefore\quad 0.497 < \sigma < 0.543$$

となって、大標本の場合には、（13）式を用いて計算するほうが速くて簡単であることがわかる。

3．分散の検定

（1）母平均既知の場合

正規母集団N(μ, σ^2)において、μは既知であるがσ^2が未知の場合に、次のような帰無仮設をたてる。

$$H_0:\ \sigma^2 = \sigma_0^2$$

σ_0^2は任意の定数である。この母集団より、抽出される大きさnの標本より、（7）式

$$\chi^2 = \left(\frac{X_1-\mu}{\sigma}\right)^2 + \left(\frac{X_2-\mu}{\sigma}\right)^2 + \cdots + \left(\frac{X_n-\mu}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \qquad (7)$$

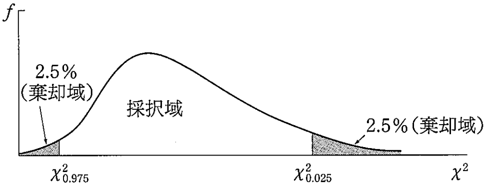
の標本統計量をつくると、これは自由度$\phi = n$のχ^2分布に従う。対立仮説H_1：$\sigma^2 \neq \sigma_0^2$をこのように設定する場合、検定はつぎの要領で行う。自由度ϕで有意水準αであるから、

$$P(\chi^2 \geq \chi^2_{\alpha/2}) = \frac{\alpha}{2} \qquad P(\chi^2 \leq \chi^2_{1-\alpha/2}) = \frac{\alpha}{2}$$

なる２つの境界点 $\chi^2_{\alpha/2}$、$\chi^2_{1-\alpha/2}$をχ^2分布表から見つけて検定すればよい。次の３つの場合がある。たとえば有意水準５％の両側検定の場合、６図に示すように、棄却域が図の両側に2.5％ずつとられる。したがって、一般的には、

$\chi^2 \leq \chi^2_{1-\alpha/2}$ または $\chi^2 \geq \chi^2_{\alpha/2}$ ならば、H_0を棄却する。

$\chi^2_{1-\alpha/2} < \chi^2 < \chi^2_{\alpha/2}$ならば、$H_0$を採択する。



６図　両側検定の採択地及び棄却域

ということになる。

つぎに、片側検定の場合には、次のような仮説の形をとる。

$(a)\ \begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases}$ (b) $\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{cases}$

（a）の右片側検定の場合

$\chi^2 \geq \chi^2_\alpha$ ならば、H_0を棄却する

$\chi^2 < \chi^2_\alpha$ ならば、H_0を採択する

また、（b）の左片側検定の場合

$\chi^2 \leq \chi^2_{1-\alpha}$ ならば、H_0を棄却する

$\chi^2 > \chi^2_{1-\alpha}$ ならば、H_0を採択する

の要領で検定すればよい。（a）と（b）のような片側検定の場合には、棄却域は右側だけ、あるいは左側だけにとられる。

例題４　同じ実験を５回繰り返して得られた液体中に含まれるエチルア

ルコールの含量を測定して、次の値を得た。(単位g)

23、27、20、21、24

このデータより、母分散σ^2について、H_0：$\sigma^2 = 5$を検定してみよう。ただし、母平均値μについては、従来の調査より$\mu = 25$であることがわかっている。有意水準を５％とする。まず、仮説を次のように立てる。

H_0：$\sigma^2 = 5$

H_1：$\sigma^2 \neq 5$

対立仮説については、σ^2が５よりも大きいのか小さいのか何の事前の情報もわかっていないから、両側検定とした。つぎに、検定統計量χ^2の値を求めると、

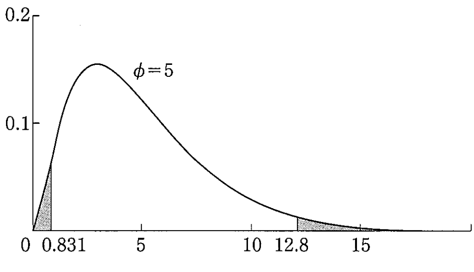
$$\chi^2 = \frac{1}{5}\left\{(23-25)^2 + (27-25)^2 + (20-25)^2 + (21-25)^2 + (24-25)^2\right\} = 10$$

となる。一方、自由度$\phi = n = 5$、有意水準５％であるから、χ^2分布の形はつぎの７図のごとくである。χ^2分布表より、検定の２つの境界点をそれぞれみつけると、

$$\chi^2_{0.975} = 0.831 \qquad \chi^2_{0.025} = 12.8$$

を得る。この場合、計算されるχ^2の値は10であるから、

$$0.831 < \chi^2 = 10 < 12.8$$



７図　自由度５の両側検定

となる。ゆえに、帰無仮説H_0を棄却するわけにはいかない。したがって、この液体のエチルアルコールの含量の分散が変わったとは考えられない、と結論づけることができる。

カイ二乗の仮説検定において、帰無仮説が正しいならば、およそ$V^2 \fallingdotseq \sigma^2$のはずであるから、$\chi^2 = nV^2/\sigma^2 \fallingdotseq n$である。したがって、$\chi^2$の値が$n$よりもはるかに大きくても小さくても仮説はおかしいことになる。よって、今までとは異なり、χ^2の値が７図の棄却域の部分に入っていたら仮説がおかしいと判定することになるのである。

（２）母平均未知の場合

正規母集団N（μ，σ^2）において、母平均μ未知の場合、

H_0：$\sigma^2 = \sigma_0^2$

H_1：$\sigma^2 \neq \sigma_0^2$

の仮説を検定しよう。母平均既知の場合のようにμを使うことができないから、

$$\chi^2 = \frac{1}{\sigma_0^2}\left\{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2\right\}$$

を使う。この検定統計量χ^2は、自由度$\phi = n - 1$のχ^2分布に従う。検定の仕方については母平均既知の場合と同様である。

例題５　あるブドウ栽培の果樹園において、今年の巨峰の10房の無作為標本の重さは、つぎのとおりであった。（単位ｇ）

350、380、360、330、350、340、370、350、360、350

平年作なら、分散は$\sigma^2 = 100$の正規分布にしたがっている。今年は平年作といってよいだろうか。有意水準５％で検定してみよう。

仮説はつぎのようになる。

H_0：$\sigma^2 = 100$

H_1：$\sigma^2 \neq 100$

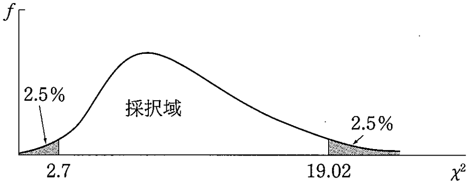
まず、$\overline{X}$を求めてから、χ^2の値を求めると、つぎのようになる。

$$\overline{X} = \frac{1}{10}(350 + 380 + 360 + 330 + 350 + 340 + 370 + 350 + 360 + 350) = 354$$

$$\chi^2 = \frac{1}{100}\left\{(350 - 354)^2 + (380 - 354)^2 + \cdots\cdots + (350 - 354)^2\right\} = 18.4$$

この検定統計量は、自由度 $\phi = n - 1$ の＝９の χ^2分布に従う。一方、有意水準 α は0.05で両側検定であるから、２つの境界点を χ^2分布表より求めると

$$\chi^2_{0.025} = 19.02 \qquad \chi^2_{0.975} = 2.7$$



８図　自由度９の両側検定

を得る。$\chi^2 = 18.4$であるかから、８図のように

$$2.7 < \chi^2 = 18.4 < 19.02$$

となる。ゆえに、帰無仮設 H_0を棄却するわけにはいかない。今年の巨峰の重さの散らばりは、平年とかわらない、と考えられる。

一方、この例題において、χ^2のところは、次のように計算しておいてもよい。

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n} = \frac{(350-354)^2 + (380-354)^2 + \cdots + (350-354)^2}{10} = \frac{1840}{10} = 184$$

$$nS^2 = 10 \times 184 = 1840$$

したがって、（５）式にそれぞれを代入すると、

$$\chi^2 = \frac{nS^2}{\sigma^2} = \frac{10 \times 184}{100} = 18.4$$

となって、同じ計算結果になることが確認できる。

以上のようになり、標本分散 S^2が100よりもかなり大きい値になるので帰無仮説 H_0が棄却されそうであるけれども意外と検定では帰無仮説は採択される結果となるのである。

また、例題５において、もし、ブドウの平年作における母平均 μ は350であると問題設定をしてみよう。この場合、（７）式から χ^2の値を求めると、

$$\chi^2=1/100[(350-350)^2+(380-350)^2+\cdots\cdots+(350-350)^2]=20$$

となる。したがって、この設定の場合、自由度は母平均が既知なので$\phi=n$となるし、自由度10のχ^2分布において、有意水準５％であるから、左右に均等に確率を2.5％ずつとると、χ^2分布表からの境界値はそれぞれ20.5と3.25であることがわかる。したがって、

$$\chi^2_{0.975}=3.25<\chi^2=20<\chi^2_{0.025}=20.5$$

となり、χ^2はほんのすこし右境界値よりも小さいのでかろうじて、採択域に入っている。したがって、帰無仮説はH_0を棄却するわけにはいかない。

さらに、この例題において、今年の重さの分散は、100よりも大きくなっているのではないか、ということに注目して対立仮説を100よりも大きいと設定してみよう。この場合、右片側検定となるので、対立仮説が$\sigma^2>100$となる。このように考えると、この問題は収穫物の巨峰の重さの分散が平年よりも大きくなったのではないかということを意味する。そして、右片側検定なので有意水準αを右側に５％分の確率を切り取るχ^2値の境界値を、自由度10のχ^2分布表よりみつけると、

$$\chi^2\ (9,\ 0.05)\ =16.92$$

であるから、

$$\chi^2=20>\chi^2\ (9,\ 0.05)\ =16.92$$

となるので、帰無仮説H_0を棄却する。すなわち、今年の巨峰の分散は平年よりも大きいといえる。

上記のように、問題設定の仕方、たとえば母平均既知なのか未知なのか、両側検定か片側検定か、などによって、検定の結論が変わってくることがある。特に、χ^2検定値と境界値がかなり接近している場合など、有意水準の大きさのとり方、棄却域の大きさや取り方によって、推定や検定の結論が変わってくるのである。

例題６． あるメーカーのある製品の重量の散らばりが0.01以下として管

理されているとしよう。製品の中から無作為に10個を検査したところ、分散が0.015であった。品質管理が適正に行われているかどうかを有意水準５％で検定してみよう。この場合、分散が規定の0.01よりも大きいのではないかに注目して、次のような右片側検定を設定したとする。

H_0：$\sigma^2 = 0.01$

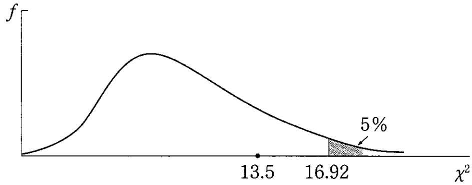
H_1：$\sigma^2 > 0.01$

（５）式の検定統計量は、自由度９のχ^2分布において棄却域を右片側にとるので

$\chi^2 = 9 \mathrm{x} 0.015 / 0.01 = 13.5$

自由度９の有意水準５％に対応する境界値は、16.92であるから、両者の大小を比較すると

$\chi^2 = 13.5 < \chi^2\ (9,\ 0.05) = 16.92$



９図　自由度９の右片側検定

となるので、帰無仮説H_0を棄却するわけにはいかない。したがって、有意水準５％で、品質管理がうまくいっていないということはない。

この例題６で、ある製品の代わりにいろいろな具体的な商品名を入れればそれらの製造過程での品質管理であれ、商品の規定重量が保証されているかどうかの消費者問題であれ、いろいろな分散の検定問題にも応用できるのである。

ちなみに、２組の母集団よりそれぞれ抽出した２組の標本について、それらが属する母集団の２つの分散の間に差異があるかどうかの検定についてはＦ検定の等分散の検定のところで扱われる。

以上のように、χ^2分布の定義やそれの利用、使い方について述べてき

たのであるが、χ^2分布を利用するに際して、注意しておくことは、まず推測統計における推定、検定において、母分散σ^2が既知の場合、（8）式の$\chi^2 = \frac{nV^2}{\sigma^2}$は自由度$n$の$\chi^2$分布に従う。ところが、母分散$\sigma^2$が未知の場合、（5）式の$\chi^2 = \frac{nS^2}{\sigma^2}$は自由度$n-1$の$\chi^2$分布に従うことをはっきり区別して認識しておくことである。

つぎに、カイ二乗分布を使って区間推定、検定の問題を扱う場合、データが例題4や例題5のように観測値のままで示されている場合、χ^2値の計算式は母平均μが既知の場合（8）式、母平均が未知の場合（5）式が使われる。それから、例題5や例題6のような検定において、母平均が既知か未知かによって、両側検定か片側検定かによって棄却域の取り方が異なってくるし、有意水準を大きくしたり小さくしたりすることによって、仮説検定の結果が変わってくることがあるので注意しておこう。また、母分散の区間推定をする場合、母平均が既知の場合は（10）式を、母平均が未知の場合は（11）式を用いる。

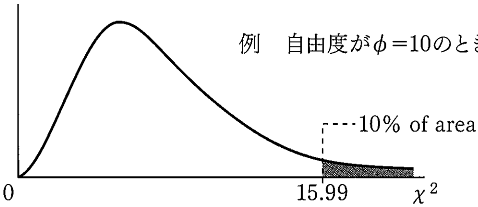
本稿はχ^2分布を使った母集団の母数の1つである分散を推測するパラメトリックの場合である。母集団の分布の型そのものを推測するノンパラメトリックの場合については、同論集の第23巻第2号のカイ二乗検定で扱われている。

参考文献

1．Statistics	taro Yamane	3ed. HARPER & ROW	1973
2．elementary statistics	P.G.Hoel	4ed. John Wily Sons, Inc.	1975
3．日常のなかの統計学	鷲尾泰俊著	岩波書店	1984
4．統計学入門	沖津　直著	八千代出版	1998

付表

χ^2 分布表



例　自由度がϕ=10のとき：$P\left[\chi^2>15.99\right]=.10$

$\phi \backslash P$	.995	.99	.975	.95	.90	.75	.50	.25	.10	.05	.025	.01	.005	P / ϕ
1	0.0^4393	0.0^3157	0.0^2982	0.0^23	0.0158	0.102	0.455	1.323	2.71	3.84	5.02	6.63	7.88	1
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.386	2.77	4.61	5.99	7.38	9.21	10.60	2
3	0.0717	0.115	0.216	0.352	0.584	1.213	2.37	4.11	6.25	7.81	9.35	11.34	12.84	3
4	0.207	0.297	0.484	0.711	1.064	1.923	3.36	5.39	7.78	9.49	11.14	13.28	14.86	4
5	0.412	0.554	0.831	1.145	1.610	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75	5
6	0.676	0.872	1.237	1.635	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55	6
7	0.989	1.239	1.690	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.3	7
8	1.344	1.646	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.1	22.0	8
9	1.735	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.7	23.6	9
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.5	23.2	25.2	10
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.9	24.7	26.8	11
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.0	23.3	26.2	28.3	12
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.4	24.7	27.7	29.8	13
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.1	23.7	26.1	29.1	31.3	14
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.3	25.0	27.5	30.6	32.8	15
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.5	26.3	28.8	32.0	34.3	16
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.5	24.8	27.6	30.2	33.4	35.7	17
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.6	26.0	28.9	31.5	34.8	37.2	18
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.7	27.2	30.1	32.9	36.2	38.6	19
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.8	28.4	31.4	34.2	37.6	40.0	20
21	8.03	8.90	10.28	11.59	13.24	16.34	20.3	24.9	29.6	32.7	35.5	38.9	41.4	21
22	8.64	9.54	10.98	12.34	14.04	17.24	21.3	26.0	30.8	33.9	36.8	40.3	42.8	22
23	9.26	10.20	11.69	13.09	14.85	18.14	22.3	27.1	32.0	35.2	38.1	41.6	44.2	23
24	9.89	10.86	12.40	13.85	15.66	19.04	23.3	28.2	33.2	36.4	39.4	43.0	45.6	24
25	10.52	11.52	13.12	14.61	16.47	19.94	24.3	29.3	34.4	37.7	40.6	44.3	46.9	25
26	11.16	12.20	13.84	15.38	17.29	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3	26
27	11.81	12.88	14.57	16.15	18.11	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6	27
28	12.46	13.56	15.31	16.93	18.94	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0	28
29	13.12	14.26	16.05	17.71	19.77	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3	29
30	13.79	14.95	16.79	18.49	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7	30
40	20.7	22.2	24.4	26.5	29.1	33.7	39.3	45.6	51.8	55.8	59.3	63.7	66.8	40
50	28.0	29.7	32.4	34.8	37.7	42.9	49.3	56.3	63.2	67.5	71.4	76.2	79.5	50
60	35.5	37.5	40.5	43.2	46.5	52.3	59.3	67.0	74.4	79.1	83.3	88.4	92.0	60
70	43.3	45.4	48.8	51.7	55.3	61.7	69.3	77.6	85.5	90.5	95.0	100.4	104.2	70
80	51.2	53.5	57.2	60.4	64.3	71.1	79.3	88.1	96.6	101.9	106.6	112.3	116.3	80
90	59.2	61.8	65.6	69.1	73.3	80.6	89.3	98.6	107.6	113.1	118.1	124.1	128.3	90
100	67.3	70.1	74.2	77.9	82.4	90.1	99.3	109.1	118.5	124.3	129.6	135.8	140.2	100
Z_α	−2.58	−2.33	−1.96	−1.64	−1.28	−0.674	0.000	0.674	1.282	1.645	1.960	2.33	2.58	Z_α

〔注〕$\phi>100$のときは、$\chi^2=\frac{1}{2}\left(Z_\alpha+\sqrt{2\phi-1}\right)^2$とする．$Z_\alpha$は、有意水準$\alpha$に対応する基準化された正規偏差であり、表の最下欄に示してある。

（本学経営学部教授）