χ^2 分布, t 分布, F 分布：資料

芳賀昭弘 *

1 χ^2 分布

1. 好きか嫌いか、などのような２つもしくはそれ以上の質的変数が独立かどうかを調べるために使われる（独立性を調べるために使われる）。

2. 観測度数の分布が理論的な度数と一致しているかどうかを調べるために使われる（適合性を調べるために使われる）。

1 の例：治療 A と治療 B の効果を調べたところ、次のような結果が得られた。

	治療 A	治療 B	計
有効（P）	8	13	21
無効（N）	2	7	9
計	10	20	30

治療 A と治療 B に差があると言えるか？（答えは。。。演習問題で！）

2 の例：メンデルは 556 個のえんどう豆を標本にとって、その色と形を同時に調べて次の結果を得た。

円形の黄色のえんどう豆	315
円形の緑色のえんどう豆	108
尖った黄色のえんどう豆	101
尖った緑色のえんどう豆	32

この割合が 9 : 3 : 3 : 1 に従っていると言えるか？（答えは。。。演習問題で！）

さて、χ^2 分布がどんな分布か示していません。もしかしたら、知らなくても統計解析できてしまうかもしれません。なぜなら、χ^2 分布をどのような問題の時にどのように使えば良いのかさえ知っておけば、あとはコンピュータが計算してくれる時代だからです。これは他の分布にも言えることです。実は研究者と言われている人の中でさえ、分布の導出を見たことも聞いたこともなく使用している人がほとんどかもしれません（言い過ぎかな？）。私はそれを良いこととは思っていません。少なくとも、どういう前提があって χ^2 分布が導入さ

* Electronic address: haga@tokushima-u.ac.jp

れなければならないのか、だけは最低限押さえておく必要があります（さもなくば、前提に合わない問題に適応してしまう可能性があります）。そして、さらに大事なことは、今日利用されている統計理論がしっかりとした数理的土台の上に成り立っていることを感じることにあると思っています。それが大学において統計学という学問を学ぶ本質的な意義だと考えています。

余談はさておき、この資料に立ち帰れば導出も適応も判断がつけられる、という意味を込めて、以下に χ^2 分布の導出を示しておきます。試験には出しませんのでご安心を。ただ、いずれ皆さんは研究室に配属され、卒業研究をすることになり、そしてデータ解析の中で統計学をもう一度勉強し直す人がいると思います。その時に、いくら勉強しても中々理解できず、上述のように「使えればいい／解析できればいい」と妥協するかもしれません。それでも構いません（研究者と言われる人たちでさえ、そうなので！）。が、「理解がフワフワして気持ちが悪い／きちんと腑に落ちたい」という人はこの資料がとっても役に立つと思います（χ^2 分布、t-分布、F-分布をここまで整理した資料は見つからないと思います）。以下の導出があって初めて私たちは安心して χ^2 分布を観測データの理解に使用することが出来ていることを時々思い出してください。

1.1 χ^2 分布の導出

1. 自由度 1 の χ^2 分布は、 確率変数 x が標準正規分布 $\mathcal{N}(x|0,1)$ に従う時の、x^2 の分布のことである（これが定義）。どのような分布になるかを以下で考えてみよう。まず標準正規分布 $\mathcal{N}(x|0,1)$ は、

$$\mathcal{N}(x|0,1) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2} \tag{1}$$

であるので、$y = x^2$ と変数変換して y の確率分布を求めれば良い。注意すべきことは、y が確率分布となるためには y の定義された空間で積分すると 1 になることである。つまり、$dx = \frac{1}{2\sqrt{y}}dy$ となることを踏まえて、

$$\begin{aligned}\chi^2(y) =& 2\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-y/2}\frac{1}{2\sqrt{y}} \\ =& \frac{1}{2^{1/2}\Gamma(\frac{1}{2})}y^{-1/2}\mathrm{e}^{-y/2}\end{aligned} \tag{2}$$

となる（2 倍されているのは積分範囲が $[-\infty,\infty]$ から $[0,\infty]$ と半分になるから）。これで、y に対して範囲 $[0,\infty]$ で積分を実行すると 1 となる。この分布を自由度 1 の χ^2 分布という。上式の 2 段目は、次に示す自由度 n の χ^2 分布

$$\chi^2(x) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})}x^{n/2-1}\mathrm{e}^{-x/2} \tag{3}$$

に繋げるために変形を施した。なお、$\Gamma(z)$ はガンマ関数と言われる関数であり、

$$\Gamma(z) = \int_0^\infty \mathrm{e}^{-t}t^{z-1}dt \tag{4}$$

である。特に $\Gamma(1/2) = \sqrt{\pi}$ となる。

2. 自由度 n の χ^2 分布は、確率変数 $x_1^2+x_2^2+\cdots+x_n^2$（ここで x_i は独立に標準正規分布 $\mathcal{N}(x_i|0,1)$ に従っている、つまり x_i^2 はそれぞれ単独では自由度 1 の χ^2 分布に従っている）の分布のことである（これが定義）。この命題を念頭に入れて、式（3）を帰納法より求めてみよう。すなわち、「確率変数 $x_1^2+x_2^2+\cdots+x_{n-1}^2$ が自由度 $n-1$ の χ^2 分布に従う時、$x_1^2+x_2^2+\cdots+x_n^2$ は自由度 n の χ^2 分布に従う」ことを示す。これは次の関係を示せば良い。

$$\chi_n^2(x) = \int_0^x \chi_{n-1}^2(t)\chi_1^2(x-t)dt \tag{5}$$

なぜなら、独立な事象の確率は積（$\chi^2_{n-1}(t)\chi^2_1(t')$）で表され、且つそれぞれの出る目 t, t' の組み合わせ全てを考えなければならないからである（積分の中の変数が t と $x-t$ になって、全ての t の値を考慮するために t で積分)。式（5）の右辺は、

$$\int_0^x \chi^2_{n-1}(t)\chi^2_1(x-t)dt = \frac{\mathrm{e}^{-x/2}}{2^{n/2}\Gamma(\frac{n-1}{2})\Gamma(\frac{1}{2})}\int_0^x t^{(n-3)/2}(x-t)^{-1/2}dt \tag{6}$$

$x = tu$ と置くと、$dt = -t/udu$ であり積分範囲は $[0, 1]$ となる；

$$\frac{\mathrm{e}^{-x/2}}{2^{n/2}\Gamma(\frac{n-1}{2})\Gamma(\frac{1}{2})}\int_0^x t^{(n-3)/2}(x-t)^{-1/2}dt = \frac{\mathrm{e}^{-x/2}x^{(n-3)/2}x^{-1/2}x}{2^{n/2}\Gamma(\frac{n-1}{2})\Gamma(\frac{1}{2})}\int_0^1 u^{(n-3)/2}(1-u)^{-1/2}du \tag{7}$$

この最後の積分はよく知られた結果を与え、

$$\int_0^1 u^{(n-3)/2}(1-u)^{-1/2}du = \frac{\Gamma(\frac{n-1}{2})\Gamma(\frac{1}{2})}{\Gamma(\frac{n}{2})} \tag{8}$$

これはベータ分布 $B(\frac{n-1}{2}, \frac{1}{2})$ と呼ばれている。これで整理すると、確かに式（3）が成り立つことがわかる（証明終了)。

なお、x_i は独立に標準正規分布 $\mathcal{N}(x_i|0, 1)$ に従っているという仮定であったが、x_i は平均 μ_i で標準偏差が σ_i の正規分布 $\mathcal{N}(x_i|\mu_i, \sigma_i^2)$ に従う場合でももちろんよく、その場合は確率変数

$$\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} + \cdots + \frac{(x_n-\mu_n)^2}{\sigma_n^2} \tag{9}$$

が自由度 n の χ^2 分布となる。独立同分布（母集団の平均、標準偏差が同じ）の場合は、

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \tag{10}$$

が自由度 n の χ^2 分布となる。

母集団の平均がわからない場合、サンプルされたデータの平均（標本平均 $\bar{x}$）を使わざるを得ないが、その場合の

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2 \tag{11}$$

は自由度 $n-1$ の χ^2 分布となる。その理由は次に示す。

1.2 χ^2 分布において、なぜ母集団の平均 μ ではなく、標本の平均 $\bar{x}$ を使うと自由度が n から $n-1$ となるのか？

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}x_i^2 - \frac{1}{\sigma^2}n\bar{x}^2 \tag{12}$$

となることを考えると、$(x_1^2, x_2^2, \cdots, x_n^2)$ のうち一つが最後の項によって打ち消されて良いように思われるであろう。実際にベクトル $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ を定義し、直交変換 $\boldsymbol{y} = C\boldsymbol{x}$ を考える。C は変換行列であり、$|C| = 1, C^TC = I$ を満たすとする。$\boldsymbol{y}$ の第一成分として $y_1 = \sqrt{n}\bar{x}$ となるような変換を定義しておけば、

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}y_i^2 - \frac{1}{\sigma^2}y_1^2 \tag{13}$$

となるので、自由度が $n-1$ となる（証明終了)。

1.3 適合性についての補足

k 個のカテゴリーがあり、それぞれのカテゴリーに入る標本の数を数えたところ、
観測数 $x_1, x_2, \cdots, x_k$ 　合計 n
であった。一方、ある理論では、それぞれのカテゴリーに入る確率 p_i が与えられており、
観測数 $np_1, np_2, \cdots, np_k$ 　合計 n
となるはずであった。両者は一致したかどうかを調べるにはどうすれば良いのでしょう？これは、カテゴリー i となる標本数 x_i を試行回数と考え、i となる確率が p_i であると考えると、x_i は二項分布となり、

$$\chi^2_{k-1} = \sum_{i=1}^{k} \frac{(x_i - np_i)^2}{np_i} \tag{14}$$

という関係が成り立つ。つまり、$\sum_{i=1}^{k} \frac{(x_i-np_i)^2}{np_i}$ は自由度 $k-1$ の χ^2 分布となる。

簡単のため、$k=2$ のケースで確かめてみましょう。その場合、$p_2 = 1 - p_1$（p_1 と p_2 は確率なので、足すと 1 になる）及び $x_2 = n - x_1$ なので

$$\begin{aligned}
\chi^2 =& \frac{(x_1 - np_1)^2}{np_1} + \frac{(x_2 - np_2)^2}{np_2} \\
=& \frac{(x_1 - np_1)^2}{np_1} + \frac{(x_1 - np_1)^2}{n(1-p_1)} \\
=& \frac{(x_1 - np_1)^2}{np_1(1-p_1)} \\
=& \frac{(x_1 - np_1)^2}{\sigma^2}
\end{aligned} \tag{15}$$

ここで、最後の変形は二項分布の分散 σ^2 が $np_1(1-p_1)$ となることを使った（証明終了）。
よって観測データと理論との適合性を観るために、自由度 $k-1$ の χ^2 分布が使える。式 (14) の x_i を観測値 O_i、np_i を理論値（期待値）E_i に置き換えて、

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{16}$$

とすることもできる。これは観測値を理論曲線でフィッティングするときにも使え、この値を最小にするように理論曲線のパラメータを決めることを<u>χ^2 フィッティング</u>という（その場合、パラメータの数を m とすると自由度は $k-m$ となる。ただしフィッティングでは上記が最小になれば良いので自由度を意識しなければならない場面はほとんどない）。

1.4 χ^2 分布のまとめ

1. 好きか嫌いか、などのような 2 つもしくはそれ以上の質的変数が独立かどうかを調べるために使われる（独立性を調べるために使われる）。分割表が $m \times n$ の時、自由度は $(m-1) \times (n-1)$ となる。

   例：A クラス（10 人）, B クラス（20 人）, C クラス（20 人）に数学が好きかどうかを調べたところ、次のような結果が得られた。

	A クラス	B クラス	C クラス	計
好き	6	11	8	25
普通	2	5	8	15
嫌い	2	4	4	10
計	10	20	20	50

A クラス, B クラス, C クラスの結果がどの程度異なっているか調べるには、自由度 $2 \times 2 = 4$ の χ^2 分布を使う。

(注意) : 上記のような表を分割表という。分割表を使って統計解析を行うものとしては、分散分析というものもある。χ^2 検定と分散分析の違いは、χ^2 検定では「質的変数」を用いた度数のみを使った分割表しか扱えないということ。

2. 観測度数の分布が理論的な度数と一致しているかどうかを調べるために使われる（適合性を調べるために使われる)。

   例：メンデルは 556 個のえんどう豆を標本にとって、その色と形を同時に調べて次の結果を得た。

円形の黄色のえんどう豆	315
円形の緑色のえんどう豆	108
尖った黄色のえんどう豆	101
尖った緑色のえんどう豆	32

   この割合が 9 : 3 : 3 : 1 と比べて異なっているかどうかを調べるには自由度 $4 - 1 = 3$ の χ^2 分布を使う。

2 t 分布

1. 正規分布に従う母集団で分散がわからない場合に使用する。

2. 正規分布に従う 2 つの母集団の平均に差があるかどうかを調べる場合に使用する。

3. 相関係数の比較に使用する。

2.1 t 分布の導出

次の変数 t に従う分布を自由度 (n-1) の t 分布という。

$$t = \frac{\bar{x} - \mu}{\sqrt{s^2/(n-1)}} \tag{17}$$

ここで $\bar{x}$ は標本平均、s^2 は標本分散、μ は母集団の平均である（母集団は正規分布）。これが定義。まずこの式がどういう意味を持つか考えてみよう。母集団の分散 σ^2 がわかっている場合は、n 個の標本の平均の分布は正規分布であり、変数 $z = \frac{\bar{x}-\mu}{\sqrt{\sigma^2/n}}$ は標準正規分布となる（中心極限定理）。t 分布と見比べると、分散の部分が $\sigma^2/n \to s^2/(n-1)$ という違いがあることがわかる。つまり、母集団の分散がわからないので、計算できる標本分散に変えた、という意図が伝わる。ところで、

$$n\frac{s^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{18}$$

は、自由度 n-1 の χ^2 分布に従うということを前章でみた。ということは、標本で得られた分散は自由度 n-1 の χ^2 分布となる。標本の平均は正規分布に従うのであるが、その分散は未知である。しかし、分散の分布は自由度 n-1 の χ^2 分布であるというのが t 分布の素顔のようである。もっと平たく言うと、χ^2 分布に従うあらゆる分散をもつ正規分布の重ね合わせが t 分布ということ。標本の数が n のとき、自由度は n-1 となる。

さて、上記の考察をもとに t 分布の導出を行おう。正規分布 $\mathcal{N}(x|\mu, \tau^{-1})$ において $\tau = 1/\sigma^2$ が自由度 $n-1$ の χ^2 分布に従っているということなので、

$$\mathrm{Gam}(\tau|a,b) = \frac{1}{\Gamma(a)}b^a\tau^{a-1}\mathrm{e}^{-b\tau} \tag{19}$$

$$\tag{20}$$

という関数を導入しておく（これはガンマ分布と呼ばれる）。$a = (n-1)/2$、$b = 1/2$ とおくと、これは $n-1$ の χ^2 分布に一致することは容易に確かめることができる。t 分布の変数は、

$$t = \frac{\bar{x}-\mu}{\sqrt{s^2/(n-1)}} = \frac{(\bar{x}-\mu)/\sigma/\sqrt{n}}{\sqrt{s^2n/(n-1)/\sigma^2}} \tag{21}$$

この分子

$$y = \frac{(\bar{x}-\mu)}{\sigma/\sqrt{n}} \tag{22}$$

は標準正規分布 $\mathcal{N}(y|0,1)$ に従い、分母の

$$\sqrt{z/(n-1)} = \sqrt{n\frac{s^2}{\sigma^2}/(n-1)} \tag{23}$$

の z は自由度 $n-1$ の χ^2 分布に従う。今、$n-1=k$ と置いておく。$y = t\sqrt{z/k}$ より

$$\mathcal{N}(y|0,1) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}y^2} = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}t^2\frac{z}{k}} \to \mathcal{N}(t|0,1) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}t^2\frac{z}{k}}\sqrt{\frac{z}{k}} \tag{24}$$

ここで y の確率分布から t の確率分布とみなすことにしたので、全範囲で積分して 1 となる条件から $\sqrt{\frac{z}{k}}$ が掛けられている（ヤコビアン）。自由度 k の χ^2 分布、

$$\mathrm{Gam}(z|\frac{k}{2},\frac{1}{2}) = \frac{1}{\Gamma(\frac{k}{2})}(\frac{1}{2})^{\frac{k}{2}}z^{\frac{k}{2}-1}\mathrm{e}^{-z/2} \tag{25}$$

$$\tag{26}$$

の下、あらゆる z について重ね合わせる。$z/k=\tau$ と置き、$dz=kd\tau$ に注意すると、あらゆる τ に対する正規分布の重ね合わせは、次の式で求められる。

$$\begin{aligned}
p(t) &= \int_0^\infty \mathcal{N}(t|0,\tau)\mathrm{Gam}(\tau|\frac{k}{2},\frac{k}{2})d\tau \\
&= \int_0^\infty \sqrt{\frac{\tau}{2\pi}}\mathrm{e}^{-\frac{\tau}{2}t^2}\frac{1}{\Gamma(\frac{k}{2})}(\frac{k}{2})^{\frac{k}{2}}\tau^{\frac{k}{2}-1}\mathrm{e}^{\frac{k}{2}\tau}d\tau \\
&= \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})}\left(\frac{1}{\pi k}\right)^{1/2}\left[1+\frac{t^2}{k}\right]^{-(k+1)/2}
\end{aligned} \tag{27}$$

これが自由度 k の t-分布の具体的な形である。（式 (21) をみて正規分布と χ^2 分布の除算じゃないか、なぜここで積をとっているんだ、と思う人がいるかもしれませんが、標本抽出は独立なので確率は積で表されます。除算となっていることは (t,y) への変数変換を通して考慮されています。次の F 分布でも同じことを行ないます）

2.2　相関係数と t 分布

$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$ の二変数データに関して、相関係数

$$r=\frac{s_{xy}}{\sqrt{s_xs_y}}=\frac{\sum_k(x_k-\bar{x})(y_k-\bar{y})}{\sqrt{\sum_i(x_i-\bar{x})^2\sum_j(y_j-\bar{y})^2}} \tag{28}$$

に対して

$$t=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{29}$$

は自由度 $n-2$ の t 分布に従う。
（証明）
変数 $X\in x_1,\cdots,x_n$ を固定し、Y のみの分布を考える。

$$\begin{aligned}
t=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} &= \frac{s_{xy}\sqrt{n-2}}{\sqrt{s_xs_y\left(1-\frac{s_{xy}^2}{s_xs_y}\right)}} \\
&= \frac{s_{xy}\sqrt{n-2}}{\sqrt{s_xs_y-s_{xy}^2}} \\
&= \frac{\sqrt{n-2}\sum_k(x_k-\bar{x})(y_k-\bar{y})}{\sqrt{\sum_i(x_i-\bar{x})^2\sum_j(y_j-\bar{y})^2-\sum_k(x_k-\bar{x})(y_k-\bar{y})}}
\end{aligned} \tag{30}$$

ここで $v_i=(x_i-\bar{x})/\sqrt{\sum_i(x_i-\bar{x})^2}$ と置くと、

$$t=\frac{\sqrt{n-2}\sum_k v_ky_k}{\sqrt{\sum_j(y_j-\bar{y})^2-(\sum_k v_ky_k)^2}} \tag{31}$$

分母のルートの中身は

$$\begin{aligned}
\sum_j(y_j-\bar{y})^2-(\sum_k v_ky_k)^2 &= \sum_j(y_j)^2-2\sum_j y_j\bar{y}+n\bar{y}^2-(\sum_k v_ky_k)^2 \\
&= \sum_j(y_j)^2-n\bar{y}^2-(\sum_k v_ky_k)^2
\end{aligned} \tag{32}$$

変数 y_i を次の直交行列により変換（$\boldsymbol{z} = C\boldsymbol{y}$）する:

$$C = \begin{pmatrix} \sqrt{n} & \sqrt{n} & \cdots & \sqrt{n} \\ v_1 & v_2 & \cdots & v_n \end{pmatrix} \tag{33}$$

すると、

$$\begin{aligned} \sum_j (y_j)^2 - n\bar{y}^2 - (\sum_k v_k y_k)^2 &= \sum_j (z_j)^2 - z_1^2 - z_2^2 \\ &= \sum_{j=3}^{n} z_j^2 \end{aligned} \tag{34}$$

これは自由度 $n-2$ の χ^2 分布：χ^2_{n-2} である。書き直すと、

$$t = \frac{z_2}{\sqrt{\chi^2_{n-2}/(n-2)}} \tag{35}$$

と書ける。これは自由度 $n-2$ をもつ t 分布である（証明終了）。

なお、「母相関 ρ が 0 かどうか」を調べる場合、
帰無仮説 H_0: $\rho = 0$
対立仮説 H_1: $\rho \neq 0$
として上記の自由度 $n-2$ の t 分布を使う。

なお、「母相関 ρ が任意の値 c かどうか」を調べる場合、
帰無仮説 H_0: $\rho = c$
対立仮説 H_1: $\rho \neq c$
として z 変換 $z = \frac{1}{2}\ln\frac{1+r}{1-r}$ を使って $u = z\sqrt{n-3}$ が近似的に正規分布に従うことを使う。

2.3 t 分布のまとめ

1. 正規分布に従う母集団で分散がわからない場合に使用する。n 個の標本では自由度 (n-1) の t-分布を使う。２つの母集団（同分散）からの標本 m, n に対しては自由度 (m+n-2) の t-分布を使う。

2. ２つの母集団の平均に差があるかどうかを調べる場合、それぞれの標本に対応がある場合は自由度 (n-1) の t-分布、対応がない場合は自由度 ν の t-分布（Welch 法）を使う。ν の出し方は前節参照のこと。

3. n 個の標本を２つ用意し、その相関係数を調べて標本間に相関があるかどうかを調べる場合、自由度 (n-2) の t-分布を使う。

3 F 分布

1. ２つの標本グループが正規分布に従う同じ母集団からサンプルされたかどうかを調べる（等分散性を調べる）場合に使用する。

2. 3つ以上の母集団について平均値に差があるかどうかを調べる（分散分析）場合に使用する。

3.1 F 分布の導出

正規分布 $\mathcal{N}(\mu, \sigma_x^2)$ である母集団からの標本を $(x_1, x_2, \cdots, x_{n_1})$、正規分布 $\mathcal{N}(\nu, \sigma_y^2)$ である母集団からの標本を $(y_1, y_2, \cdots, y_{n_2})$ とする時、

$$F = \frac{\chi_x^2/n_1}{\chi_y^2/n_2} \tag{36}$$

の標本分布は自由度 (n_1, n_2) の F 分布に従う。この分布の具体的な表式を求めてみよう。まず、

$$\frac{\chi_x^2/n_1}{\chi_y^2/n_2} = \frac{z_1/n_1}{z_2/n_2} \tag{37}$$

とおいて、変数 (z_1, z_2) の確率密度関数 $p(z_1, z_2)$ を考える（確率密度関数ということは $p(z_1, z_2)$ を z_1 と z_2 で積分すると 1 とならなければならない）。これを

$$\frac{z_1/n_1}{z_2/n_2} = y, \tag{38}$$

$$z_1 = yz/n_2, \tag{39}$$

$$z_2 = z/n_1. \tag{40}$$

となる新たな変数 (z, y) により変換する。式 (39) より確率密度関数 $p(z_1, z_2)$ を y 単独で表したものが F 分布となる。そのためには、$p(z_1, z_2)$ を変数 (z, y) で表した確率密度関数を求め、それを z で積分することが必要となる。確率密度関数は、変数で積分すると 1 にならなければならないので、

$$\int p(z_1, z_2) dz_1 dz_2 = \int p(z, y) dz dy \tag{41}$$

という関係が成り立つ。このためには、

$$\int p(z, y) dz dy = \int p(z_1, z_2) |J(z_1, z_2/z, y)| dz dy \tag{42}$$

となるヤコビアン J を元の確率分布関数 $p(z_1, z_2)$ に掛ければ良い（t 分布では式 (24) でヤコビアンが掛けられている）。

$$|J(z_1, z_2/z, y)| = \begin{vmatrix} \partial z_1/\partial z & \partial z_1/\partial y \\ \partial z_2/\partial z & \partial z_2/\partial y \end{vmatrix} = \begin{vmatrix} y/n_2 & z/n_2 \\ 1/n_1 & 0 \end{vmatrix} = \frac{z}{n_1 n_2} \tag{43}$$

$p(z_1, z_2)$ は z_1 と z_2 の変数を持つ χ^2 分布の積（式 (37) をみて「除算じゃないか」と思うかもしれませんが、標本抽出は独立なので確率は積で表されます。除算となっていることは (z, y) への変数変換を通して考慮されています。t 分布でも同じことを行なっています）である。これに上記のヤコビアンを掛けて z で積分した結果が F 分布である；

$$\begin{aligned} p(y) &= \int \frac{1}{2^{n_1/2}\Gamma(n_1/2)} z_1^{n_1/2-1} \mathrm{e}^{-z_1/2} \frac{1}{2^{n_2/2}\Gamma(n_2/2)} z_2^{n_2/2-1} \mathrm{e}^{-z_2/2} \frac{z}{n_1 n_2} \\ &= \int \frac{y^{n_1/2-1}}{2^{(n_1+n_2)/2} n_1^{n_2/2} n_2^{n_1/2} \Gamma(n_1/2)\Gamma(n_2/2)} z^{(n_1+n_2)/2-1} \mathrm{e}^{-(1/n_1+y/n_2)z/2} \end{aligned} \tag{44}$$

ここで $(1/n_1 + y/n_2)z/2 = \omega$ と変数変換してから積分を実行する。結果は、

$$p(y) = \frac{n_1^{n_1/2} n_2^{n_2/2} y^{n_1/2-1}}{\Gamma(n_1/2)\Gamma(n_2/2)(n_2 + yn_1)^{(n_1+n_2)/2}} \Gamma((n_1 + n_2)/2) \tag{45}$$

となる。式 (45) が F 分布と呼ばれる（この表式における確率変数は y）。