数理統計学まとめ（その 7）：第 4 章 母集団と標本

4　正規分布から導かれる標本分布

4.1　カイ 2 乗分布（χ^2-分布）

母集団が標準正規分布 $N(0,1)$ に従うとき，大きさ n の無作為標本 $X_1,\ldots,X_n$ に対して

$$X=\sum_{j=1}^{n}X_i^2$$

が自由度 n のカイ 2 乗分布に従う．その分布密度関数は

$$f(x)=\frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}x^{n/2-1}e^{-x/2}\qquad(x>0)$$

となる．ただし，$m>0$ に対して

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

をガンマ関数という．$\Gamma(m)$ は

$$\Gamma(m+1)=m\Gamma(m)$$

を満たす．これを使うと自由度 ν のカイ 2 乗分布の平均と分散はそれぞれ ν と 2ν になることがわかる．

定理 4.1　（教科書 p.87 定理 4.5）
母集団が正規分布 $N(\mu,\sigma^2)$ に従うとき，無作為標本 $X_1,\ldots,X_n$ に対して，

$$Y=\frac{1}{\sigma^2}\left((X_1-\bar{X})^2+\cdots+(X_n-\bar{X})^2\right)$$

は自由度 $n-1$ のカイ 2 乗分布に従う．

4.2　ティー分布（t-分布）

確率変数 X と Y が独立で，X が $N(0,1)$ に従い，Y が自由度 ν のカイ 2 乗分布に従うとき，

$$T=\frac{X}{\sqrt{\frac{Y}{\nu}}}$$

は自由度 ν のティー分布（t-分布）t_ν に従う．このとき T の分布密度関数は

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \qquad (-\infty < t < \infty)$$

となる．

定理 4.2 （教科書 p.89 定理 4.6) 母集団が正規分布 $N(\mu, \sigma^2)$ に従うとき，無作為標本 $X_1, \ldots, X_n$ に対して，標本平均 $\bar{X}$ と

$$Y = \frac{1}{\sigma^2}\left((X_1-\bar{X})^2 + \cdots + (X_n-\bar{X})^2\right)$$

は独立であり，したがって

$$T = \frac{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}}{\sqrt{\frac{Y}{(n-1)}}}$$

は自由度 $n-1$ のティー分布（t-分布）t_{n-1} に従う．

$$U^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

を使うと[1]，

$$(n-1)U^2 = \sigma^2 Y$$

となるので

$$T = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2 Y}{n(n-1)}}} = \frac{\bar{X}-\mu}{\sqrt{\frac{U^2}{n}}}$$

となる．

$$E(T) = 0 \ \ (\nu \geq 2), \quad V(T) = \frac{\nu}{\nu-2} \ \ (\nu \geq 3)$$

となる．

[1] U^2 は不偏標本分散と呼ばれる

4.3 エフ分布 （F-分布）

確率変数 X_1, X_2 が独立で，それぞれ自由度 ν_1, ν_2 のカイ 2 乗分布に従うとき

$$F = \frac{\frac{X_1}{\nu_1}}{\frac{X_2}{\nu_2}}$$

は自由度 ν_1, ν_2 の F 分布 $F^{\nu_1}_{\nu_2}$ に従うと言う．その分布密度関数は

$$f(x) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} x^{\frac{\nu_1-2}{2}}\left(1+\frac{\nu_1}{\nu_2}x\right)^{-\frac{\nu_1+\nu_2}{2}} \qquad (x > 0)$$

となる．

$$E(F) = \frac{\nu_2}{\nu_2-2} \qquad (\nu_2 \geq 3)$$

$$V(F) = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)} \qquad (\nu_2 \geq 5)$$

定理 4.3 （教科書 p.91, 定理 4.7）

自由度 (ν_1, ν_2) の F-分布と (ν_2, ν_1) の F-分布との間には次の関係が成り立つ．

$$F^{\nu_1}_{\nu_2}(1-\alpha) = \frac{1}{F^{\nu_2}_{\nu_1}(\alpha)}$$

ただし，$F^{\nu_1}_{\nu_2}(\alpha)$ は F が自由度 (ν_1, ν_2) の F-分布に従うとき，

$$P\left(F \geq F^{\nu_1}_{\nu_2}(\alpha)\right) = \alpha$$

となる値として定義する．

証明 X を自由度 ν_1 のカイ 2 乗分布に従う確率変数とし，Y をこれと独立な自由度 ν_2 のカイ 2 乗分布に従う確率変数とする．このとき

$$F = \frac{\frac{Y}{\nu_2}}{\frac{X}{\nu_1}}$$

が自由度 (ν_2, ν_1) の F-分布に従うので，

$$\begin{aligned}
\alpha &= P(F \geq F^{\nu_2}_{\nu_1}(\alpha)) \\
&= P\left(\frac{Y}{X}\frac{\nu_1}{\nu_2} \geq F^{\nu_2}_{\nu_1}(\alpha)\right) \\
&= P\left(\frac{X}{Y}\frac{\nu_2}{\nu_1} \leq \frac{1}{F^{\nu_2}_{\nu_1}(\alpha)}\right) \\
&= 1 - P\left(\frac{1}{F} \geq \frac{1}{F^{\nu_2}_{\nu_1}(\alpha)}\right)
\end{aligned}$$

$\frac{1}{F}$ は自由度 (ν_1, ν_2) の F-分布だから，上の式は

$$F^{\nu_1}_{\nu_2}(1-\alpha) = \frac{1}{F^{\nu_2}_{\nu_1}(\alpha)}$$

を意味している．

例 4.1 （教科書 p.87 例題 4.6）　自由度が 10 のカイ 2 乗分布に従う確率変数 X に対して

$$P(X \geq a) = 0.05$$

を満たす a (上側 5 %点）を求める．教科書の巻末の付表 3 を見て左端の欄の自由度が 10 となっているところを横に見る．このうち，一番上の欄の α が .05 となっている列を下がり，自由度 10 の行と交わるところを見ると， $a = 18.307$ がわかる．

同じ調子で自由度 8 のカイ 2 乗分布に従う確率変数 X に対して上側 2.5 %点を求めると，17.535 となる．

表に載っていない上側 α 点 （ 100α %点）を求めるときは比例配分による．例えば，自由度 20 のカイ 2 乗分布に従う確率変数 X に対して

$$P(X \geq a) = 0.08$$

を求めるには，上側 10 %点が 28.412 で 5 % 点が 31.410 なので，比例配分して上側 8 %点は $28.412 + \frac{2}{5} \times (31.410 - 28.412) = 29.611$ と求めれば良い．表の数字は小数点以下 3 桁なので，結果も小数点以下 3 桁より詳しくは求められない．というのは，すでに表を作る時点で次の桁以下は丸められているので．