確率統計と情報処理・演習（2007 年度後期）

確率分布

2007 年 11 月 30 日

日本女子大学理学部数物科学科　　今野 良彦

November 30, 2007

今日の講義の目的と概要

- 標本分布
  - 標本分布とは（先週説明）
  - 正規母集団からの標本に関連した標本分布
    - χ^2 分布（カイ自乗分布）（先週説明）
    - t 分布と F 分布

t 分布

m を自然数とする．確率変数 T が自由度 m の t 分布に従うとは，T の確率密度関数が

$$f_m(x) = \frac{\Gamma((m+1)/2)}{\sqrt{m\pi}\Gamma(m/2)}\left(1+\frac{x^2}{m}\right)^{-(m+1)/2}, \qquad -\infty < x < \infty$$

で与えられるときをいう．

ただし，$\Gamma(a)\,(a>0)$ はガンマ関数で

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

とくに，

$$\begin{aligned} \Gamma(1) &= 1, \\ \Gamma(1/2) &= \sqrt{\pi}, \\ \Gamma(a+1) &= a\,\Gamma(a), \\ \Gamma(n) &= (n-1)! \qquad (n \text{ は自然数}) \end{aligned}$$

t 分布のグラフの出力

```
> # 自由度 10 の t 分布の確率密度関数
> tdens10<-function(x){
+ dt(x,10)
+ }
>
> # 自由度 10 の t 分布の確率密度関数のグラフの出力
> curve(tdens10,-4,4)
>
> # 自由度 2 の t 分布
> tdens2<-function(x){
+ dt(x,2)
+ }
>
>
> curve(tdens10,-6,6)
> # 自由度 2 の t 分布の確率密度関数のグラフの出力を追加
> curve(tdens2,-6,6,col=2,add=T)
> # 標準正規分布の確率密度関数のグラフの出力の追加
> curve(dnorm,-6,6,col=3,add=T)
>
```


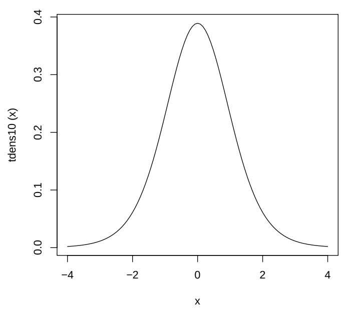

Figure 1: 自由度 10 のと t 分布の確率密度関数のグラフ

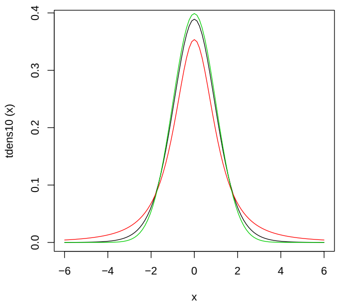

Figure 2: 自由度 10 の t 分布 (black)，自由度 2 の t 分布 (red) と標準正規分布（green）の確率密度関数のグラフ

t 分布の確率密度関数の導出

$\star$ X を標準正規分布に従う確率変数，すわわち，

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$\star$ Y を自由度 m の χ^2 に従う確率変数，すなわち

$$f_Y(y) = \begin{cases} \frac{1}{2^{m/2}\Gamma(m/2)} y^{\frac{m}{2}-1} e^{-y/2} & (y > 0), \\ 0 & (y \le 0). \end{cases}$$

$\star$ X と Y は独立

このとき，

$$Z = \frac{X}{\sqrt{Y/m}}$$

は自由度 m の t 分布に従う．

このことを示すために，任意の $-\infty < a_1 < a_2 < \infty$ と $0 < b_1 < b_2 < \infty$ に対して，

$$\mathbb{P}(a_1 < X \le a_2,\, b_1 < Y \le b_2) = \int_{a_1}^{a_1} \int_{b_1}^{b_2} f_{X,Y}(x,\, y)\, dx\, dy$$

を満たす非負の関数 $f_{X,Y}$ を求める．これを X **と** Y **の同時確率密度関数**という．

$-\infty < x < \infty,\, y > 0$ に対して，

$$\begin{aligned} \mathbb{P}(X \le x,\, Y \le y) &= \mathbb{P}(X \le x)\mathbb{P}(Y \le y) \\ &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du \int_0^y \frac{1}{2^{m/2}\Gamma(m/2)} v^{\frac{m}{2}-1} e^{-v/2}\, dv \end{aligned}$$

となるので，微積分の基本定理を用いると

$$\begin{aligned} f_{X,Y}(x,\, y) &= \frac{\partial^2}{\partial x \partial y} \mathbb{P}(X \le x,\, Y \le y) \\ &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{2^{m/2}\Gamma(m/2)} y^{\frac{m}{2}-1} e^{-y/2} \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{2^{m/2}\Gamma(m/2)} e^{-x^2/2} y^{\frac{m}{2}-1} e^{-y/2} \end{aligned}$$

任意の z に対して，

$$\begin{aligned} \mathbb{P}(Z \le z) &= \int\int_{(x/\sqrt{y/m}) \le z,\, y > 0} \frac{1}{\sqrt{2\pi}} \frac{1}{2^{m/2}\Gamma(m/2)} e^{-x^2/2} y^{\frac{m}{2}-1} e^{-y/2}\, dx\, dy \\ &= \int_0^\infty \left\{ \int_{-\infty}^{\sqrt{y/m}z} \frac{1}{\sqrt{2\pi}} \frac{1}{2^{m/2}\Gamma(m/2)} y^{\frac{m}{2}-1} e^{-(x^2+y)/2}\, dx \right\} dy \end{aligned}$$

$w = \sqrt{y/m}z$ とおき，この両辺を z で部分する．積分記号と微分記号を入れ替えを行う．

$$\begin{aligned} f_Z(z) &= \frac{d}{dz} \int_0^\infty \left\{ \int_{-\infty}^{\sqrt{y/m}z} \frac{1}{\sqrt{2\pi}} \frac{1}{2^{m/2}\Gamma(m/2)} y^{\frac{m}{2}-1} e^{-(x^2+y)/2}\, dx \right\} dy \\ &= \int_0^\infty \frac{d}{dz} \left\{ \int_{-\infty}^{\sqrt{y/m}z} \frac{1}{\sqrt{2\pi}} \frac{1}{2^{m/2}\Gamma(m/2)} y^{\frac{m}{2}-1} e^{-(x^2+y)/2}\, dx \right\} dy \\ &= \int_0^\infty \frac{dw}{dz} \frac{d}{dw} \left\{ \int_{-\infty}^{w} \frac{1}{\sqrt{2\pi}} \frac{1}{2^{m/2}\Gamma(m/2)} y^{\frac{m}{2}-1} e^{-(x^2+y)/2}\, dx \right\} dy \\ &= \int_0^\infty \sqrt{\frac{y}{m}} \frac{1}{\sqrt{2\pi}} \frac{1}{2^{m/2}\Gamma(m/2)} y^{\frac{m}{2}-1} e^{-(w^2+y)/2}\, dy \\ &= \frac{1}{\sqrt{\pi}\Gamma(m/2)} \frac{1}{\sqrt{m}} \int_0^\infty \frac{1}{2^{(m+1)/2}} y^{(m-1)/2} \exp\left[-\frac{(1+z^2/m)y}{2} \right] dy \end{aligned}$$

Insert 2006-11-17-zu1.pdf

ここで

$$u = \frac{(1+z^2/m)y}{2}, \qquad du = \frac{(1+z^2/m)}{2}\,dy$$

から

$$\begin{aligned}
f_Z(z) &= \frac{1}{\sqrt{\pi}\Gamma(m/2)}\frac{1}{\sqrt{m}}\int_0^\infty \frac{1}{2^{(m+1)/2}}\left(u\Big/\frac{(1+z^2/m)}{2}\right)^{(m-1)/2} \\
&\quad \times e^{-u}\Big/\frac{(1+z^2/m)}{2}\,du \\
&= \frac{1}{\sqrt{\pi}\Gamma(m/2)}\frac{1}{\sqrt{m}}\left(1+z^2/m\right)^{-(m+1)/2}\int_0^\infty u^{(m+1)/2-1}e^{-u}\,du \\
&= \frac{\Gamma((m+1)/2)}{\sqrt{\pi}\Gamma(m/2)}\frac{1}{\sqrt{m}}\left(1+\frac{z^2}{m}\right)^{-(m+1)/2}
\end{aligned}$$

t 分布のシミュレーション

目的　1000 個の標準正規乱数と自由度 m の χ^2 分布乱数を生成して，自由度 m の t 分布乱数を 1000 個作成して，その分布が t 分布の確率密度関数に近いことを確認する．

chi^2 分布のグラフの出力

```
> # 標準正規乱数 1000 個の生成
> nrdata<-rnorm(1000,0,1)
> # 自由度 2 のカイ自乗分布乱数 1000 個の生成
> chisqdata2<-rchisq(1000,2)
>  カイ自乗分布乱数 1000 個の平均
> mean(chisqdata2)
[1] 2.005096
>  カイ自乗分布乱数 1000 個の標準偏差
> sd(chisqdata2)
[1] 2.028897
>  # >  カイ自乗分布乱数 1000 個のヒストグラムの作成
> hist(chisqdata2,freq="F")
>  # 自由度 2 のカイ自乗分布の確率密度関数のグラフを上書き
> chisq2<-function(x){dchisq(x,2)}
> curve(chisq2,0,15,add=T)
>
```


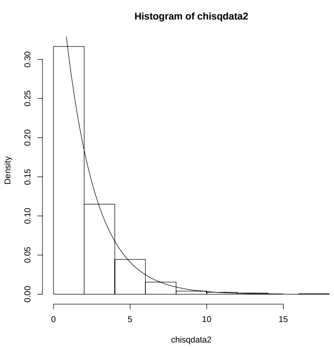

Figure 3: カイ自乗分布乱数 1000 個のヒストグラムと自由度 2 のカイ自乗分布の確率密度関数のグラフ

t 分布乱数のグラフの出力

```
> nrdata<-rnorm(5000,0,1)
> chisqdata<-rchisq(5000,6)
> # t 乱数の作成
> t6data<-nrdata/(sqrt(chisqdata/6))
>
> # 自由度 6 の t 乱数 5000 個のヒストグラム
> hist(t6data,freq=F,ylim=c(0,0.4))
> # 自由度 6 の t の確率密度関数の上書き
> t6<-function(x){dt(x,6)}
> curve(t6,-10,10,add=T,col=2)
>
```

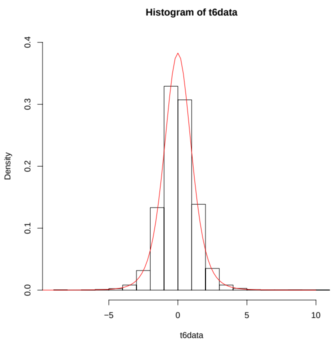

Figure 4: t 分布乱数 5000 個のヒストグラムと自由度 2 のカイ自乗分布の確率密度関数のグラフ

正規分布と t 分布の関係

t 分布が使われるのは正規分布から標本の場合である．すなわち，

$$X_1,\, X_2,\, \dots,\, X_n \sim N(\mu,\, \sigma^2)$$

ただし，$n \geq 2$，$-\infty < \mu < \infty,\, 0 < \sigma < \infty$ である．すなわち，

⋆ $X_1,\, X_2,\, \dots,\, X_n$ は互いに独立$-\infty < a_{i1} < a_{i2} < \infty\, (i = 1, 2, \dots, n)$ に対して，

$$\mathbb{P}(a_{11} < X_1 \leq a_{12},\, a_{21} < X_2 \leq a_{22},\, \dots,\, a_{n1} < X_n \leq a_{n2}) = \Pi_{i=1}^n \mathbb{P}(a_{i1} < X_i \leq a_{i2})$$

$$\mathbb{P}(a_{i1} < X_i \leq a_{i2}) = \int_{a_{i1}}^{a_{i2}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}\, dx$$

⋆

⋆ 標本平均 $\bar{X}_n = \dfrac{1}{n} \sum_{i=1}^n X_i$

⋆ 標本分散 $S_n^2 = \dfrac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$

このとき，

$$Z_n = \frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} \sim N(0,\, 1)$$

すわわち，$-\infty < a < b < \infty$ に対して，

$$\mathbb{P}(a < Z_n \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$

さらに，

$$\frac{(n-1)S_n^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma} \right)^2$$

は自由度 $n-1$ の χ^2 分布に従う．

⋆ $0 < a < b < \infty$ に対して，

$$\mathbb{P}(a < \frac{(n-1)S_n^2}{\sigma^2} \leq b) = \int_a^b \frac{1}{2^{(n-1)/2}\Gamma((n-1)/2)} y^{\frac{n-1}{2}-1} e^{-y/2}\, dy$$

ただし，$\Gamma(a)\, (a > 0)$ はガンマ関数で

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

⋆ $\bar{X}_n$ と S_n^2 は独立である．すわわち，$-\infty < a < b < \infty,\, 0 < c < d < \infty$ に対して，

$$\mathbb{P}(a < \bar{X}_n \leq b,\, c < S_n^2 \leq d) = \mathbb{P}(a < \bar{X}_n \leq b)\mathbb{P}(c < S_n^2 \leq d)$$

このことに注意して，

$$T = \frac{(\bar{X}_n - \mu)/\sqrt{\sigma^2/n}}{\sqrt{\sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma} \right)^2 /(n-1)}}$$

とおく．

U	標準正規分布	V	自由度 m の χ^2 分布
U と V は独立			
$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$	標準正規分布	$\sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma} \right)^2$	自由度 $n-1$ の χ^2 分布
Z_n と $\sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma} \right)^2$ は独立			

このとき，

$$T = \frac{U}{\sqrt{V/m}} \iff T = \frac{(\bar{X}_n - \mu) \Big/ \sqrt{\frac{\sigma^2}{n}}}{\sqrt{\sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma} \right)^2 \Big/ (n-1)}}$$

は自由度 m の t 分布に従う．

$$T \;=\; \frac{(\bar{X}_n-\mu)\Big/\sqrt{\frac{\sigma^2}{n}}}{\sqrt{\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}_n}{\sigma}\right)^2\Big/(n-1)}} = \frac{\bar{X}_n-\mu}{\sqrt{S_n^2/n}}$$

は自由度 $n-1$ の t 分布に従う．すなわち，$-\infty < a < b < \infty$ に対して，

$$\mathbb{P}\left(a < \frac{\bar{X}_n-\mu}{\sqrt{S_n^2/n}} \le b\right) = \int_a^b \frac{\Gamma(n/2)}{\sqrt{\pi}\Gamma((n-1)/2)}\frac{1}{\sqrt{n-1}}\left(1+\frac{t^2}{n-1}\right)^{-n/2} dt$$

ただし，

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}_n}{\sigma}\right)^2, \qquad \Gamma(a) = \int_0^\infty x^{a-1}e^{-x}\,dx, \qquad a>0$$

$T = (\bar{X}_n-\mu)/\sqrt{S_n^2/n}$ の分布をシミュレーションで確かめるアルゴリズム

(1) 平均が μ ($\mu = 0$ でなくともよい) の正規乱数を n 個発生される．$x_1, x_2, \ldots, x_n$ とする．

(2) このデータの標本平均の実現値を $\bar{x}_n$，標本標準偏差の実現値を

$$u_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2}$$

(3) T の実現値

$$t = \frac{\bar{x}_n-\mu}{u_n/\sqrt{n}}$$

(4) (1) ～ (3) を 1000 回繰り返して T の実現値を 1000 個計算する．

t の値を計算する関数

```
> tcalc<-function(x,y){
+ barx<-mean(x)
+ sdx<-sd(x)
+ tval<-(barx-y)/(sdx/sqrt(length(x)))
+ return(tval)
+ }
```

1000 個の t の値を計算し，グラフに出力

```
> samt<-rep(0,1000)              # ベクトル samt の成分は $0$
> for (i in 1:1000){              # 繰り返しのプログラム
+ samt[i]<-tcalc(rnorm(10,2,4),2)        # samt の $i$ 番 目
に t の値を入れる．
+ }
> hist(samt,nclass=20,freq=F)
> tdens9<-function(x)(dt(x,9))     # 自由度 $9$ の $t$ 分布の
密度関数
> curve(tdens9,-4,4,col=2,add=T)
>
```


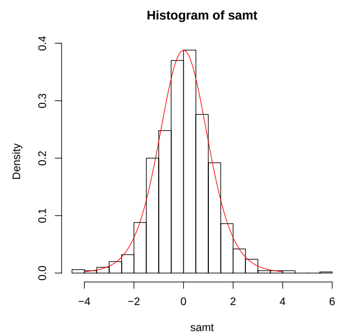

Figure 5: t 分布乱数 1000 個のヒストグラムと自由度 9 の t 分布の確率密度関数のグラフ

ここまでのまとめ

- t 分布の確率密度関数の形はわかりました？
- t 分布は独立な標準正規分布と χ^2 分布から導出されることはわかりましたか？
- 正規分布と t 分布の関係はかわりましたか？

問題１

(1) 適当な自由度が t の確率密度関数と標準正規分布の確率密度関数のグラフを書け．(meriro-hanaka-071207-t1.pdf)

(2) 1000 個の標準分布の乱数と自由度 m の χ^2 分布の乱数を 1000 個つくり 1000 個の t 分布乱数を作成し，そのグラムと自由度 m の t の分布の確率密度関数を書き入れたものを作成せよ．ただし，自由度は誕生月プラス 3 とする．(meriro-hanaka-071207-t2.pdf)

(3) (a) 平均が μ の正規乱数を n 個発生される．$x_1, x_2, \ldots, x_n$ とする．

(b) このデータの標本平均の実現値を $\bar{x}_n$，標本標準偏差の実現値を

$$u_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}$$

(c) T の実現値

$$t = \frac{\bar{x}_n - \mu}{u_n/\sqrt{n}}$$

(d) (a) ∼ (c) を 1000 回繰り返して T の実現値を 1000 個計算する．

のアルゴリズムに基づいて t 分布の乱数を作成し，1000 個の t 分布の乱数のヒストグラムと対応する自由度の t 分布の確率密度関数のグラフを書け．(meriro-hanaka-071207-t3.pdf) ただし，標本数 n は 16-(誕生月) とする．

- それぞれの問題の解説を mejiro-hanako-071214.txt に書け．
- 締め切りは 2007 年 12 月 14 日（金）13 時

F 分布の確率密度関数

m, n は正の整数とする．自由度 (m, n) の F 分布の確率密度関数は

$$\begin{aligned} f_{(m,n)}(x) &= \begin{cases} \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}\left(\frac{m}{n}\right)^{m/2} x^{m/2-1}\left(1+\frac{m}{n}x\right)^{-(m+n)/2}, & x > 0 \\ 0 & x \le 0 \end{cases} \\ &= \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}\left(\frac{m}{n}\right)^{m/2} x^{m/2-1}\left(1+\frac{m}{n}x\right)^{-(m+n)/2} \mathbb{1}_{(0,\infty)}(x) \end{aligned}$$

ただし，

$$\begin{aligned} \mathbb{1}_{(0,\infty)}(x) &= \begin{cases} 1 & x \in (0, \infty) \\ 0 & x \notin (0, \infty) \end{cases} \\ &= \begin{cases} 1 & (x > 0), \\ 0 & (x \le 0), \end{cases} \end{aligned}$$

自由度 $(1, 10)$, $(2, 10)$, $(3, 10)$, $(8, 10)$, $(8, 20)$ の F の分布の確率密度関数のグラフの作図

F のグラフの出力

```
> fdens1.10<-function(x)(df(x,1,10))
> fdens2.10<-function(x)(df(x,2,10))
> fdens3.10<-function(x)(df(x,3,10))
> fdens8.10<-function(x)(df(x,8,10))
> fdens8.20<-function(x)(df(x,8,20))
> curve(fdens1.10,0.1,5,ylim=c(0,1.5))
> curve(fdens2.10,0.00000001,5,col=2,add=T)
> curve(fdens3.10,0,5,col=3,add=T)
> curve(fdens8.10,0,5,col=4,add=T)
> curve(fdens8.20,0,5,col=5,add=T)
```


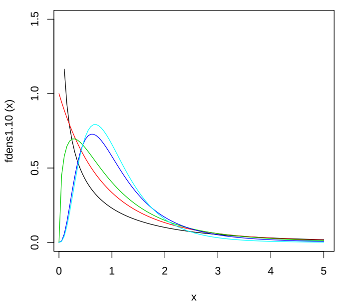

Figure 6: 自由度 $(1, 10)$(黒), $(2, 10)$(赤), $(3, 10)$(緑), $(8, 10)$(青), $(8, 20)$ の F の分布の確率密度関数のグラフ

X が自由度 m の χ^2 分布に従い，Y が自由度 n の χ^2 分布に従い，独立であれば，

$$Z = \frac{X/m}{Y/n}$$

は自由度 (m, n) の F 分布に従う．

X の確率密度関数は

$$\begin{aligned} f_X(x) &= \begin{cases} \frac{1}{2^{m/2}\Gamma(m/2)} x^{\frac{m}{2}-1} e^{-x/2} & (x > 0), \\ 0 & (x \le 0). \end{cases} \\ &= \frac{1}{2^{m/2}\Gamma(m/2)} x^{\frac{m}{2}-1} e^{-x/2}\, \mathbb{1}_{(0,\infty)}(x) \end{aligned}$$

同様に，Y の確率密度関数は

$$\begin{aligned} f_Y(y) &= \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} y^{\frac{n}{2}-1} e^{-y/2} & (y > 0), \\ 0 & (y \le 0). \end{cases} \\ &= \frac{1}{2^{n/2}\Gamma(n/2)} y^{\frac{n}{2}-1} e^{-y/2}\, \mathbb{1}_{(0,\infty)}(y) \end{aligned}$$

ただし，$a > 0$ に対して，

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

X と Y が独立であることから，任意の $0 < a_1 < a_2 < \infty$, $0 < b_1 < b_2 < \infty$ に対して，

$$\mathbb{P}(a_1 < X \le a_2,\, b_1 < Y \le b_2) = \int_{a_1}^{a_2}\int_{b_1}^{b_2} f_X(x) f_Y(y)\,dx\,dy$$

もっと一般的に $D \subset (0,\,\infty)\times(0,\,\infty)$ としてとき，

$$\mathbb{P}((X,\,Y)\in D) = \int\int_D f_X(x) f_Y(y)\,dx\,dy$$

が成立する．

このことに注意して，Z と分布関数を求めるために，任意の正の実数 z に対して，

$$D = \left\{(x,\,y)\in(0,\,\infty)\times(0,\,\infty) : \frac{x/m}{y/n} \le z\right\}$$

$z > 0$ に対して，X と Y は正値確率変数なので，

$$G_Z(z) = \mathbb{P}(Z \le z) = \mathbb{P}\left(\frac{X/m}{Y/n} \le z\right) = 0$$

また，任意の正の実数 z に対して，

$$\begin{aligned} G_Z(z) &= \mathbb{P}(Z \le z) = \mathbb{P}\left(\frac{X/m}{Y/n} \le z\right) \\ &= \int\int_D f_X(x) f_Y(y)\,dx\,dy \\ &= \int_0^\infty \left(\int_0^{\frac{m}{n}yz} f_X(x)\,dx\right) f_Y(y)\,dy \end{aligned}$$

両辺を z で微分(積分記号と微分記号の交換可能を仮定)する：$z > 0$ に対して，

$$\begin{aligned} g_Z(z) &= \frac{d}{dz} G_Z(z) \\ &= \int_0^\infty \frac{d}{dz}\left(\int_0^{\frac{m}{n}yz} f_X(x)\,dx\right) f_Y(y)\,dy \\ &= \int_0^\infty f_X(\frac{m}{n}yz)\frac{m}{n}y f_Y(y)\,dy \\ &= \int_0^\infty \frac{1}{2^{m/2}\Gamma(m/2)}\left(\frac{m}{n}yz\right)^{\frac{m}{2}-1} e^{-\frac{1}{2}\frac{m}{n}yz}\frac{m}{n}y \\ &\quad \times \mathbb{1}_{(0,\,\infty)}(\frac{m}{n}yz)\frac{1}{2^{n/2}\Gamma(n/2)} y^{\frac{n}{2}-1} e^{-y/2}\,\mathbb{1}_{(0,\,\infty)}(y)\,dy \end{aligned}$$

$z \le 0$ に対して，$g_Z(z) = \dfrac{d}{dz}G_Z(z) = 0$

$z > 0$ のとき，これを整理すると

$$\begin{aligned} g_Z(z) &= \int_0^\infty \frac{1}{2^{m/2}\Gamma(m/2)}\left(\frac{m}{n}yz\right)^{\frac{m}{2}-1} e^{-\frac{1}{2}\frac{m}{n}yz}\frac{m}{n}y \\ &\quad \times \mathbb{1}_{(0,\,\infty)}(\frac{m}{n}yz)\frac{1}{2^{n/2}\Gamma(n/2)} y^{\frac{n}{2}-1} e^{-y/2}\,\mathbb{1}_{(0,\,\infty)}(y)\,dy \\ &= \frac{\left(\frac{m}{n}\right)^{m/2} z^{m/2-1}}{\Gamma(m/2)\Gamma(n/2)2^{(m+n)/2}}\int_0^\infty e^{-\frac{y}{2}\left(1+\frac{m}{n}x\right)} y^{(m+n)/2-1}\,dy \end{aligned}$$

あとはつぎの積分を評価すればよい：

$$\int_0^\infty e^{-\frac{y}{2}\left(1+\frac{m}{n}x\right)} y^{(m+n)/2-1}\,dy$$

そのために

$$v = \frac{y}{2}\left(1+\frac{m}{n}z\right)$$

と置換すると

$$y\Big|_0^\infty = \frac{2v}{1+\frac{m}{n}z}\Big|_0^\infty, \qquad dy = \frac{2}{1+\frac{m}{n}z}dv$$

から

$$\begin{aligned} \int_0^\infty e^{-\frac{y}{2}\left(1+\frac{m}{n}z\right)} y^{(m+n)/2-1}\,dy &= \int_0^\infty e^{-v}\frac{2^{(m+n)/2-1}v^{(m+n)/2-1}}{(1+\frac{m}{n}x)^{(m+n)/2-1}}\frac{2}{1+\frac{m}{n}z}\,dv \\ &= 2^{(m+n)/2-1}\left(1+\frac{m}{n}z\right)^{-(m+n)/2}\int_0^\infty e^{-v} v^{(m+n)/2-1}\,dv \\ &= 2^{(m+n)/2-1}\left(1+\frac{m}{n}z\right)^{-(m+n)/2}\Gamma\left(\frac{m+n}{2}\right) \end{aligned}$$

これより，$z > 0$ のとき，

$$\begin{aligned} g_Z(z) &= \frac{\left(\frac{m}{n}\right)^{m/2} z^{m/2-1}}{\Gamma(m/2)\Gamma(n/2)2^{(m+n)/2}}\int_0^\infty e^{-\frac{y}{2}\left(1+\frac{m}{n}x\right)} y^{(m+n)/2-1}\,dy \\ &= \frac{\left(\frac{m}{n}\right)^{m/2} z^{m/2-1}}{\Gamma(m/2)\Gamma(n/2)2^{(m+n)/2}} 2^{(m+n)/2-1}\left(1+\frac{m}{n}z\right)^{-(m+n)/2}\Gamma\left(\frac{m+n}{2}\right) \\ &= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}\right)^{m/2} z^{m/2-1}\left(1+\frac{m}{n}z\right)^{-(m+n)/2} \end{aligned}$$

したがって，

$$g_Z(z) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}\right)^{m/2} z^{m/2-1}\left(1+\frac{m}{n}z\right)^{-(m+n)/2}\mathbb{1}_{(0,\,\infty)}(z)$$

独立な 1000 個の χ^2 から F 分布の値を 1000 個作成するプログラム

```
> c8rand<-rchisq(1000,8)    # 自由度 8 のカイ自乗分布乱数を 1000 個生成
> c10rand<-rchisq(1000,10)    # 自由度 10 のカイ自乗分布乱数を 1000 個生成
> fprop<-(c8rand/8)/(c10rand/10)    # 自由度 10 の F 分布乱数を 1000 個生成
> hist(fprop)
> hist(fprop)$count
 [1] 501 342 100  35  13   4   2   1   1   1
> hist(fprop,nclass=20,freq=F,ylim=c(0,0.8))
> fdens8.10<-function(x)(df(x,8,10))    # 自由度 (8, 10) の F 分布確率密度関数を設定
> curve(fdens8.10,0,5,col=4,add=T)    # 密度関数をヒストグラムに上書き
```


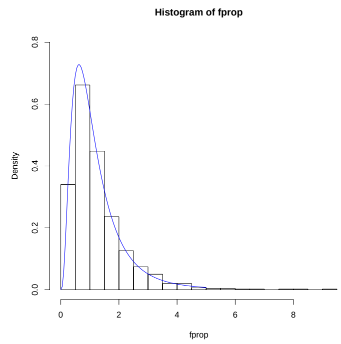

Figure 7: 自由度 $(8, 10)$ の F 分布乱数とその確率密度関数のグラフ

独立な 1000 個の χ^2 から F 分布の値を 1000 個作成するプログラム

```
> c5rand<-rchisq(1000,5)
> c8rand<-rchisq(1000,8)
> fdens5.8<-function(x)(df(x,5,8))
> fprop<-(c5rand/5)/(c8rand/8)
> hist(fprop)$count
 [1] 396 359 150  59  21   7   1   3   0   1   1   1   1
>
> hist(fprop,nclass=20,freq=F,ylim=c(0,0.7))
> curve(fdens5.8,0,10,col=2,add=T)
>
```


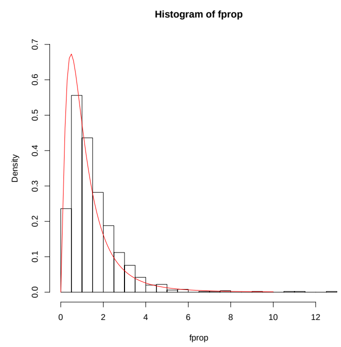

Figure 8: 自由度 $(5, 8)$ の F 分布乱数とその確率密度関数のグラフ

ここまでのまとめ

- F 分布の確率密度関数の形はわかりました？
- F 分布は独立なふたつの χ^2 分布から導出されることはわかりましたか？

問題２

(1) 適当な自由度が F の確率密度関数のグラフを 5 つ書け.（meriro-hanaka-071207-F1.pdf）色を変えてグラフを書き，各色のグラフに対応する自由度を mejiro-hanako-071207.txt に書け.

(2) 自由度 m と n の χ^2 分布の乱数を 1000 個つくり 1000 個の F 分布乱数を作成し，そのグラムと自由度 (m, n) の F の分布の確率密度関数を書き入れたものを作成せよ．ただし，自由度は m と n は適当なもの（例にないもの）を選択せよ.（meriro-hanaka-021207-F2.pdf）

- それぞれの問題の解説を mejiro-hanako-071207.txt に書け.
- 締め切りは 2007 年 12 月 07 日（金）13 時