確率統計と情報処理・演習（2009 年度後期）

確率分布

2009 年 11 月 06 日

日本女子大学理学部数物科学科　　今野 良彦

October 23, 2009

今日の講義の目的と概要

- 標本分布
  - 標本分布とは
  - 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）
    - t 分布と F 分布（たぶん次週になる）

母集団

```
> x<-1:100
> x
  [1]   1   2   3   4   5   6   7   8   9  10  11  12  13  14
 [15]  15  16  17  18  19  20  21  22  23  24  25  26  27  28
 [29]  29  30  31  32  33  34  35  36  37  38  39  40  41  42
 [43]  43  44  45  46  47  48  49  50  51  52  53  54  55  56
 [57]  57  58  59  60  61  62  63  64  65  66  67  68  69  70
 [71]  71  72  73  74  75  76  77  78  79  80  81  82  83  84
 [85]  85  86  87  88  89  90  91  92  93  94  95  96  97  98
 [99]  99 100
> mean(x)
[1] 50.5
> (sum(x**2)-100*mean(x)**2)/100
[1] 833.25
> v<-var(x)*99/100
> v
[1] 833.25
> sqrt(v)
[1] 28.86607
```


sample 標本抽出

```
> y<-sample(x,replace=T,5)
> y
[1]  44 63 48 38 23
> mean(y)
[1] 43.3
> y<-sample(x,replace=T,5)
> y
[1] 96 10 99 22  6
> mean(y)
[1] 46.6
>
```

⋆ 母集団から n 個の標本を抽出することを繰り返すことによって得られる標本平均や標本分散の分布はどうなるだろうか？

標本を $X_1, X_2, \ldots, X_n$ とする．このとき，

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

例　$n = 5$ とする．前のスライドの 1 回目の標本抽出では，$X_1, X_2, \ldots, X_n$ の実現値は

$$X_1 = 47,\ X_2 = 59,\ X_3 = 45,\ X_4 = 91,\ X_5 = 100, \bar{X}_5 = 68.4$$

2 回目の標本抽出では，

$$X_1 = 96,\ X_2 = 10,\ X_3 = 99,\ X_4 = 22,\ X_5 = 6, \bar{X}_5 = 46.4$$

n = 5 の標本抽出を 50 回繰り返した結果

```
      [,1] [,2] [,3] [,4] [,5] [,6]
 [1,]   56   64   72   75   33 60.0
 [2,]   57   79   96   77   22 66.2
 [3,]   58   48   51   93   29 55.8
 [4,]   72   48   16   68   45 49.8
 [5,]   22   59   55   52   72 52.0
 #  省略
[48,]   58   44   51   64   97 62.8
[49,]   70   48   82   92   87 75.8
[50,]   27   60   63   92   23 53.0
```

n = 5 の標本抽出を 50 回繰り返した結果

```
 [1] 60.0 66.2 55.8 49.8 52.0 52.0 64.2 34.4 46.8 67.6
[11] 48.0 47.0 56.4 48.0 52.4 61.2 55.0 62.4 67.0 71.8
[21] 30.8 45.6 47.4 48.8 52.0 46.4 43.2 46.0 60.8 64.4
[31] 48.0 63.8 58.0 49.0 57.0 52.4 31.4 68.2 55.0 41.4
[41] 35.2 53.8 65.8 48.6 65.8 56.2 53.0 62.8 75.8 53.0
>
```

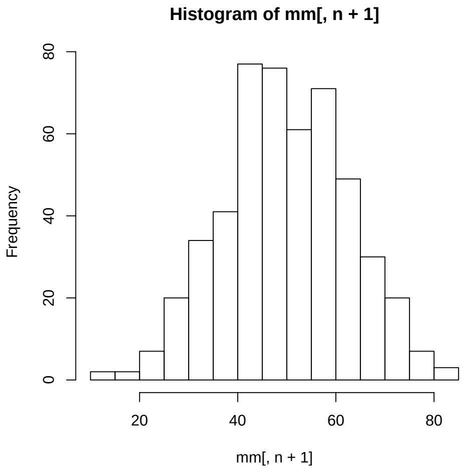



Figure 1: $n = 5$ のときの標本平均のヒストグラム（平均 = 51.0424, 標準偏差 = 12.30351）

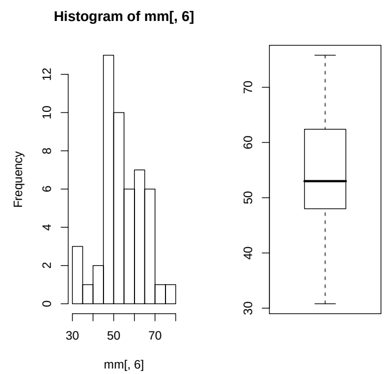


Figure 2: $n = 5$ のときの標本平均のヒストグラム（平均 = 51.0424, 標準偏差 = 12.30351）

標本の大きさ n の標本抽出を rep 回繰り返すプログラム

```
> rep<-500      # 実験回数
> n<-5     # 標本の大きさ
> mm<-matrix(,rep,n+2)     # 各実験の標本の値，標本平均，標本分
散を入れるオブジェクト
> for (i in 1:rep){
+ y<-sample(x,replace=T,n)       # 標本抽出
+ m<-mean(y)             # 標本平均の計算
+ v<-(sum(y**2)-n*mean(y)**2)/(n-1)      # 標本分散の計算
+ mm[i,]<-c(y,m,v)
+ }
> op<-par(mfrow=c(1,2))
> hist(mm[,n+1])      # 標本平均のヒストグラム
> mean(mm[,n+1])     # 標本平均の分布の平均
[1]  51.0424
> sqrt(var(mm[,n+1])*(rep-1)/rep)    # 標本平均の分布の分散
[1] 12.30351
> boxplot(mm[,n+1])
>
```

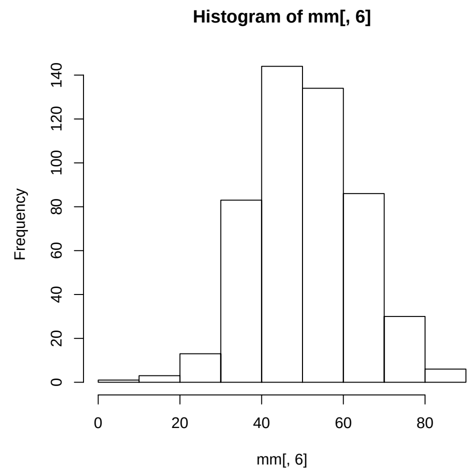



Figure 3: $n = 5$ のときの標本分散のヒストグラム（平均 = 50.8884, 標準偏差 = 12.78582）

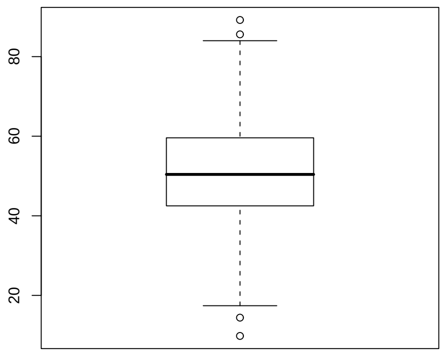


Figure 4: $n = 5$ のときの標本分散の箱ひげ図（平均 = 859.4496）

標本分散のヒストグラムの出力

```
>
> hist(mm[,n+2])
> mean(mm[,n+2])
[1] 859.4496
> boxplot(mm[,n+2])
>
```

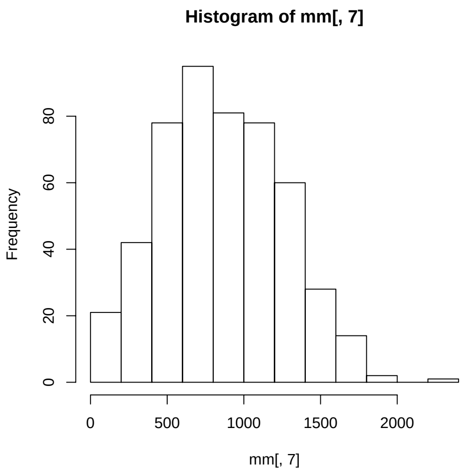


Figure 5: $n=5$ のときの標本分散のヒストグラム（平均 = 859.4496）

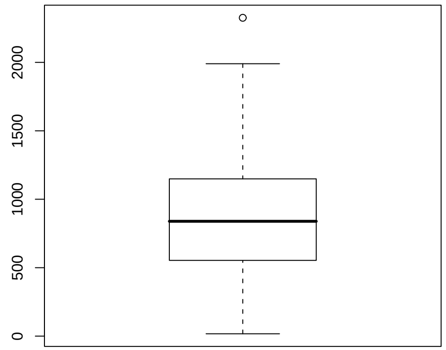


Figure 6: $n=5$ のときの標本分散の箱ひげ図（平均 = 859.4496）

標本の大きさ n の標本抽出を rep 回繰り返すプログラムための関数の作成

```
> x<-1:1000
> rep<-500
> sampledis<-function(n,x){
+ mm<-matrix(,rep,n+2)
+ for(i in 1:rep){
+ y<-sample(x,replace=T,n)
+ m<-mean(y)
+ v<-(sum(y**2)-n*mean(y)**2)/(n-1)
+ mm[i,]<-c(y,m,v)
+ }
+ return(mm)
+ }
> mm<-sampledis(10,x)
> hist(mm[,11])
> hist(mm[,12])
>
```


ここまでのまとめ

- 標本分布を説明しました.
  - 有限母集団を設定して,
  - そこから標本の大きさが n の標本を抽出し,
  - 標本平均や標本分散の分布をヒストグラムで見ました.
  - 母集団の母平均と標本平均の分布の平均とを比較しました.

以上の事項についての定義や性質について説明しました.

問題 1

- 有限母集団を以下のように作成よせ．

有限母集団の成績

```
> x<-1:100*(誕生月)
> x
```

(1) オブジェクト x のヒストグラム（20816***-目白花子-091120-poplation.pdf）を作成せよ．

(2) 母集団（オブジェクト x）の母平均と母分散を求めよ．

(3) 母集団から標本の大きさ 6 の標本を 500 回抽出し，標本平均の標本分布のヒストグラム（20816***-目白花子-091120-samplemean.pdf）を作成せよ．さらに，標本平均の標本分布の平均を計算せよ．

(4) 母集団から標本の大きさ 6 の標本を 500 回抽出し，標本分散の標本分布のヒストグラム（20816***-目白花子-091120-samplemvar.pdf）を作成せよ．さらに，標本分散の標本分布の平均を計算せよ．

- 締め切りは 2009 年 11 月 13 日（金）13 時
- 20816***-目白花子-091120.txt
- 提出はメール：mtoukei(at)mp[dot]jwu[dot]ac[dot]jp

これからの講義の目的と概要

- 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）
  - t 分布と F 分布

	R での名前（コマンド）			
分布名	名前	確率密度関数	累積分布関数	擬似乱数
正規分布	norm	dnorm(x, 平均, 標準偏差)	pnorm(x,0,1)	rnorm(個数,0,1)
t 分布	t	dt(x, 自由度)	pt(x,1)	rt(個数,1)
χ^2 分布[1]	chisq	dchisq(x, 自由度)	pchisq(x,1)	rchisq(個数,1)
F 分布	f	df(x, 自由度, 自由度)	pf(x,1,2)	rf(個数,1,2)

χ^2 分布（カイ自乗分布とよむ）

χ^2 分布（カイ自乗分布）

確率変数 Z が標準正規分布 $N(0,1)$ に従うとする．すわわち，

$$\mathbb{P}(a < Z \le b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz, \qquad (\text{任意の } -\infty < a < b < \infty)$$

このとき，

$$Y = Z^2$$

の分布を**自由度 1 の χ^2 分布**という．

⋆ 定義から Y は非負の値をとる確率変数である．

⋆ Y は連続型確率変数である．

⋆ Y の分布はどうなるか？

すわなち，確率はどうもとめればよいか？

そのために，Y の累積分布関数を求める．

それを微分すれば，確率密度関数を導出できる．

Y の累積分布関数 $F_Y(y)$ の導出

Y は非負値確率関数なので，$y \leq 0$ に対して，$F_Y(y) = \mathbb{P}(Y \leq y) = 0$ となる．

つぎに，$y > 0$ に対して，

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(Z^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq Z \leq \sqrt{y}) \\ &= \mathbb{P}(-\sqrt{y} < Z \leq \sqrt{y}) + \mathbb{P}(Z = -\sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \end{aligned}$$

注意：3 番目の等号は

$$\{Z^2 \leq y\} \Longleftrightarrow \{-\sqrt{y} \leq Z \leq \sqrt{y}\}$$

なので，事象 $\{Z^2 \leq y\}$ と $\{-\sqrt{y} \leq Z \leq \sqrt{y}\}$ の確率が等しいことがわかる．

Z は連続型確率変数なので，$\mathbb{P}(Z = \sqrt{y}) = 0$ に注意せよ．

Y の確率密度関数 $f_Y(y)$ を求めるために，$F_Y(y)$ を y について微分する：

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}F_Y(y) = \frac{d}{dy}\int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \\ &= \frac{d}{dy}\left\{\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz - \int_0^{-\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz\right\} \\ &= \frac{1}{\sqrt{2\pi}} e^{-y/2} \frac{d}{dy}\sqrt{y} - \frac{1}{\sqrt{2\pi}} e^{-y/2} \frac{d}{dy}(-\sqrt{y}) \\ &= \frac{1}{\sqrt{2\pi}} e^{-y/2} \frac{1}{2\sqrt{y}} - \frac{1}{\sqrt{2\pi}} e^{-y/2} \frac{-1}{2\sqrt{y}} \\ &= \frac{1}{\sqrt{2}\sqrt{\pi}} \frac{1}{\sqrt{y}} e^{-y/2} \\ &= \frac{1}{2^{1/2}\Gamma(1/2)} y^{\frac{1}{2}-1} e^{-y/2} \end{aligned}$$

注意：

$\star$ 2 番目の等号は定積分の性質を使っている！

$$\begin{aligned} \int_a^b f(x)\, dx &= \int_a^c f(x)\, dx + \int_c^b f(x)\, dx, \\ \int_a^b f(x)\, dx &= -\int_b^a f(x)\, dx \end{aligned}$$

$\star$ 3 番目の等号は微積分の基本的定理

$$\frac{d}{dx}\int_0^x g(z)\, dz = g(x)$$

と合成関数の微分を使っている！

$\star$ $\Gamma(a)\,(a>0)$ はガンマ関数で

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

とくに,

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

自由度 1 の χ^2 分布の確率密度関数

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

ただし,

$$\mathbb{1}_{(0,\infty)}(y)=\begin{cases}1 & y>0\\ 0 & y\le 0\end{cases}$$

自由度 1 の χ^2 分布を一般化するために, $Z_1, Z_2, \ldots, Z_m$ を独立同一に標準正規分布 $N(0,1)$ に従う確率変数を考える.

すわわち, 各 $Z_i\,(i=1,2,\ldots,m)$ は標準正規分布に従い, $Z_1, Z_2, \ldots, Z_m$ は独立である.

任意の $-\infty<a_i<b_i<\infty\,(i=1,2,\ldots,m)$ に対して,

$$\begin{aligned}&\mathbb{P}(a_1<Z_1\le b_1,\,a_2<Z_2\le b_2,\,\cdots,\,a_m<Z_m\le b_m)\\ &\quad=\Pi_{i=1}^m\mathbb{P}(a_i<Z_i\le b_i)\\ &\quad=\mathbb{P}(a_1<Z_1\le b_1)\times\mathbb{P}(a_2<Z_2\le b_2)\times\cdots\times\mathbb{P}(a_m<Z_m\le b_m)\end{aligned}$$

このとき,

$$Y_m=Z_1^2+Z_2^2+\cdots+Z_m^2$$

とおく. Y_m の確率分布を**自由度 m の χ^2 分布**という.

自由度 m の χ^2 分布の確率密度関数

$Y_m=Z_1^2+Z_2^2+\cdots+Z_m^2$ の確率密度関数は

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

ただし,

$$\mathbb{1}_{(0,\infty)}(y)=\begin{cases}1 & y>0\\ 0 & y\le 0\end{cases}$$

χ^2 分布の再帰性

- Y_{m_1} が自由度 m_1 の χ^2 分布に従い，
- Y_{m_2} が自由度 m_2 の χ^2 分布に従って，
- 互いに独立とする．

このとき，$Y_{m_1}+Y_{m_2}$ は自由度 m_1+m_2 の χ^2 分布に従う．

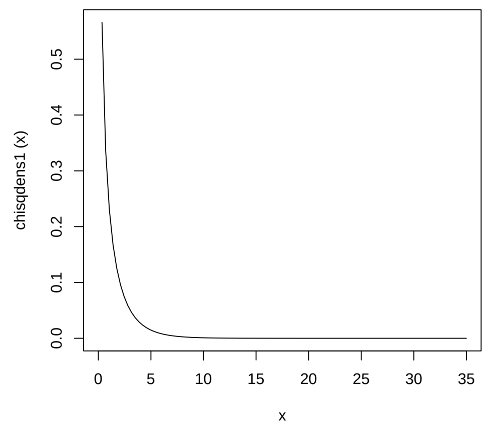


Figure 7: 自由度 1 の χ^2 分布の確率密度関数のグラフ

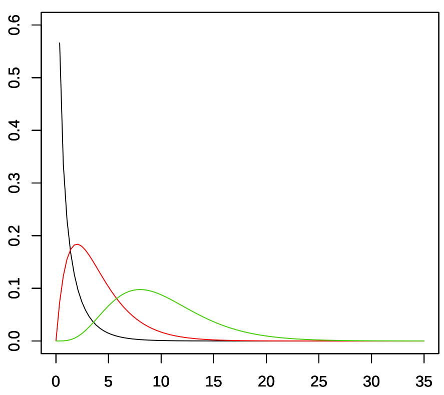


Figure 8: 自由度 1 の χ^2 分布の確率密度関数のグラフ（赤=4, 黄緑=10）

自由度 m の χ^2 分布の確率密度関数のグラフの作図プログラム

```
> chisqdens1<-function(x){dchisq(x,1)}
> # 自由度 1 の$ \chi ^2 $ 分布の確率密度関数
> chisqdens4<-function(x){dchisq(x,4)}
> # 自由度 4 の$ \chi ^2 $ 分布の確率密度関数
> chisqdens10<-function(x){dchisq(x,10)}
> # 自由度 10 の$ \chi ^2 $ 分布の確率密度関数
>
> curve(chisqdens1,0,35,ylim=c(0,0.6),ylab="",xlab="")
> # 自由度 1 の$ \chi ^2 $ 分布の確率密度関数の出力
> par(new=T)
> # グラフの重ね書きのコマンド
> curve(chisqdens4,0,35,ylim=c(0,0.6),ylab="",xlab="",col=2)
> # 自由度 4 の$ \chi ^2 $ 分布の確率密度関数の出力
> par(new=T)
> # グラフの重ね書きのコマンド
>
> curve(chisqdens10,0,35,ylim=c(0,0.6),ylab="",xlab="",col=3)
> # 自由度 10 の$ \chi ^2 $ 分布の確率密度関数の出力
```

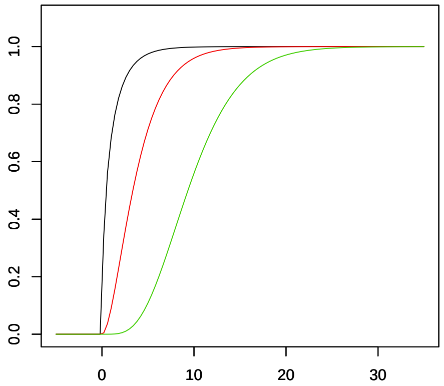


Figure 9: 自由度 m の χ^2 分布の累積分布関数のグラフ (m=4(赤), m=10(黄緑))

自由度 m の χ^2 分布の累積分布関数のグラフの作図プログラム

```
> chisqdis1<-function(x){pchisq(x,1)}
> # 自由度 1 の$ \chi ^2 $ 分布の累積分布関数
> chisqdis4<-function(x){pchisq(x,4)}
> # 自由度 4 の$ \chi ^2 $ 分布の累積分布関数
> chisqdis10<-function(x){pchisq(x,10)}
> # 自由度 10 の$ \chi ^2 $ 分布の累積分布関数
>
> curve(chisqdis1,-5,35,ylim=c(0,1.1),ylab="",xlab="",col=1)
> # 自由度 1 の$ \chi ^2 $ 分布の累積分布関数の出力
> par(new=T)
> # グラフの重ね書きのコマンド
> curve(chisqdis4,-5,35,ylim=c(0,1.1),ylab="",xlab="",col=2)
> # 自由度 4 の$ \chi ^2 $ 分布の累積分布関数の出力
> par(new=T)
> # グラフの重ね書きのコマンド
>
> curve(chisqdis10,-5,35,ylim=c(0,1.1),ylab="",xlab="",col=3)
> # 自由度 4 の$ \chi ^2 $ 分布の累積分布関数の出力
```


χ^2 分布の再帰性の証明

χ^2 分布の再帰性

- Y_{m_1} が自由度 m_1 の χ^2 分布に従い，
- Y_{m_2} が自由度 m_2 の χ^2 分布に従って，
- 互いに独立とする．

このとき，$Y_{m_1}+Y_{m_2}$ は自由度 m_1+m_2 の χ^2 分布に従う．

χ^2 分布の再帰性の証明のはじまり

Y_{m_1} と Y_{m_2} は独立なので，任意の $y_1>0$ と $y_2>0$ に対して，

$$\mathbb{P}(Y_{m_1}\le y_1,\,Y_{m_2}\le y_2)=\mathbb{P}(Y_{m_1}\le y_1)\mathbb{P}(Y_{m_2}\le y_2)$$

したがって，Y_{m_1} と Y_{m_2} の同時確率密度関数 $f_{Y_{m_1},\,Y_{m_2}}$ は

$$\begin{aligned} f_{Y_{m_1},\,Y_{m_2}}(y_1,\,y_2) &= \frac{\partial^2}{\partial y_1\partial y_2}\mathbb{P}(Y_{m_1}\le y_1,\,Y_{m_2}\le y_2) \\ &= \frac{\partial}{\partial y_1}\mathbb{P}(Y_{m_1}\le y_1)\frac{\partial}{\partial y_2}\mathbb{P}(Y_{m_2}\le y_2) \\ &= f_{Y_{m_1}}(y_1)f_{Y_{m_2}}(y_2) \end{aligned}$$

ただし，$f_{Y_{m_1}}(y_1)$ と $f_{Y_{m_2}}(y_2)$ は自由度 m_1 と m_2 の χ^2 分布の確率密度関数である．

すわなち，Y_{m_1} と Y_{m_2} の同時確率密度関数は

$$\begin{aligned} f_{Y_{m_1},Y_{m_2}}(y_1,\,y_2) &= f_{Y_{m_1}}(y_1)f_{Y_{m_2}}(y_2) \\ &= \frac{1}{2^{m_1/2}\Gamma(m_1/2)}y^{\frac{m_1}{2}-1}e^{-y_1/2}\frac{1}{2^{m_2/2}\Gamma(m_2/2)}y^{\frac{m_2}{2}-1}e^{-y_2/2} \end{aligned}$$

ただし，$\Gamma(a)\,(a>0)$ はガンマ関数で

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

$W=Y_{m_1}+Y_{m_2}$ とおく．W の確率密度関数 f_W を求めるために，

$$\mathbb{P}(W\le w)\,(w\ge 0)$$

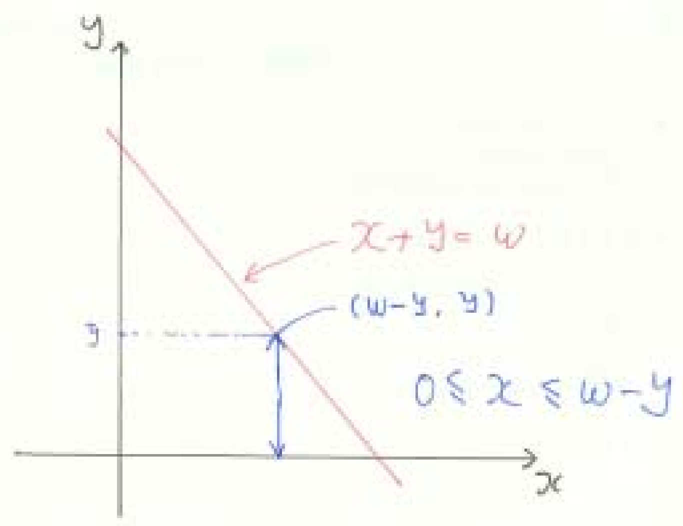
を計算する：そのために，

$$\{W\le w\}\Longleftrightarrow\{Y_{m_1}+Y_{m_2}\le w\}$$

なので，

$$\{(x,\,y)\in\mathbb{R}^2:\,x+y\le w,\,x\ge 0,\,y\ge 0\}=\{(x,\,y)\in\mathbb{R}^2:\,0\le y\le w,\,0\le x\le w-y\}$$

となる領域に注意する．



$$\begin{aligned} F_W(w) &= \mathbb{P}(Y_{m_1}+Y_{m_2}\le w)=\int\int_{x+y\le w,\,x\ge 0,\,y\ge 0} f_{Y_{m_1},Y_{m_2}}(x,\,y)\,dx\,dy \\ &= \int_0^w\left\{\int_0^{w-y} f_{Y_{m_1}}(x)f_{Y_{m_2}}(y)\,dx\right\}dy \end{aligned}$$

この両辺を w で微分する（右辺の微分は微積分の基本定理と合成関数の偏微分）と

$$f_W(w)=\frac{d}{dw}F_W(w)=\int_0^w f_{Y_{m_1}}(w-y)f_{Y_{m_2}}(y)\,dy \tag{1}$$

この計算はやさしくない！

(1) の右辺をさらに評価する：

$$
\begin{aligned}
f_W(w) &= \int_0^w f_{Y_{m_1}}(w-y) f_{Y_{m_2}}(y)\,dy \\
&= \int_0^w \frac{1}{2^{m_1/2}\Gamma(m_1/2)}(w-y)^{\frac{m_1}{2}-1}e^{-(w-y)/2}\frac{1}{2^{m_2/2}\Gamma(m_2/2)}y^{\frac{m_2}{2}-1}e^{-y/2}\,dy \\
&= \int_0^w \frac{1}{2^{(m_1+m_2)/2}\Gamma(m_1/2)\Gamma(m_2/2)}(w-y)^{\frac{m_1}{2}-1}y^{\frac{m_2}{2}-1}e^{-w/2}\,dy
\end{aligned}
$$

ここで，$y = wt$ と変換すると

$$
\begin{aligned}
f_W(w) &= \int_0^w \frac{1}{2^{(m_1+m_2)/2}\Gamma(m_1/2)\Gamma(m_2/2)}(w-y)^{\frac{m_1}{2}-1}y^{\frac{m_2}{2}-1}e^{-w/2}\,dy \\
&= \int_0^1 \frac{1}{2^{(m_1+m_2)/2}\Gamma(m_1/2)\Gamma(m_2/2)}(w-wt)^{\frac{m_1}{2}-1}(wt)^{\frac{m_2}{2}-1}e^{-w/2}w\,dt \\
&= \frac{1}{2^{(m_1+m_2)/2}\Gamma(m_1/2)\Gamma(m_2/2)}w^{\frac{m_1+m_2}{2}-1}e^{-w/2}\int_0^1 (1-t)^{\frac{m_1}{2}-1}t^{\frac{m_2}{2}-1}\,dt
\end{aligned}
$$

最後に，ベータ関数とガンマ関数の関係式

$$
\int_0^1 (1-t)^{\frac{m_1}{2}-1}t^{\frac{m_2}{2}-1}w\,dt = B(m_1/2,\, m_2/2) = \frac{\Gamma(m_1/2)\Gamma(m_2/2)}{\Gamma((m_1+m_2)/2)}
$$

をつかうと

$$
\begin{aligned}
f_W(w) &= \frac{1}{2^{(m_1+m_2)/2}\Gamma(m_1/2)\Gamma(m_2/2)}w^{\frac{m_1+m_2}{2}-1}e^{-w/2}\int_0^1 (1-t)^{\frac{m_1}{2}-1}t^{\frac{m_2}{2}-1}w\,dt \\
&= \frac{1}{2^{(m_1+m_2)/2}\Gamma((m_1+m_2)/2)}w^{\frac{m_1+m_2}{2}-1}e^{-w/2}
\end{aligned}
$$

となり，自由度 m_1+m_2 の χ^2 分布の確率密度関数を得る！ □

(1) の説明

$$
\begin{aligned}
h(s,\,t) &= \int_0^{s-t} f_{Y_{m_1}}(x) f_{Y_{m_2}}(t)\,dx, \\
H(s,\,w) - H(s,\,0) &= H(s,\,t)\Big|_{t=w} - H(s,\,t)\Big|_{t=0} = \int_0^w h(s,\,z)\,dz
\end{aligned}
$$

とおく．このとき，

$$
h(w,\,w) = 0
$$

と

$$
\int_0^w \frac{\partial}{\partial s}h(s,\,z)\Big|_{s=w}dz = \int_0^w f_{Y_{m_1}}(w-z) f_{Y_{m_2}}(z)\,dz
$$

になることに注意．

なぜならば,

$$
\begin{aligned}
\int_0^w \frac{\partial}{\partial s} h(s,\, z)\Big|_{s=w} dz &= \int_0^w \frac{\partial}{\partial s} \int_0^{s-z} f_{Y_{m_1}}(x) f_{Y_{m_2}}(z)\Big|_{s=w} dz \\
&= \int_0^w \frac{\partial}{\partial s} \left\{ \int_0^{s-z} f_{Y_{m_1}}(x) f_{Y_{m_2}}(z)\, dx \right\} \Big|_{s=w} dz \\
&= \int_0^w f_{Y_{m_1}}(s-z) f_{Y_{m_2}}(z)\, dx \Big|_{s=w} dz \\
&= \int_0^w f_{Y_{m_1}}(w-z) f_{Y_{m_2}}(z)\, dz
\end{aligned}
$$

なる.

$$
\begin{aligned}
\frac{d}{dw} \int_0^w h(s,\, t)\Big|_{s=w} ds &= \frac{d}{dw} H(s,\, t)\Big|_{s=w,\, t=w} - H(s,\, t)\Big|_{s=w,\, t=0} \\
&= \frac{\partial}{\partial s} H(s,\, t)\Big|_{s=w,\, t=w} + \frac{\partial}{\partial t} H(s,\, t)\Big|_{s=w,\, t=w} - \frac{\partial}{\partial s} H(s,\, t)\Big|_{s=w,\, t=0} \\
&= \frac{\partial}{\partial s} \int_0^t h(s,\, z)\, dz\Big|_{s=w,\, t=w} + \frac{\partial}{\partial t} \int_0^t h(s,\, z)\, dz\Big|_{s=w,\, t=w} \\
&\qquad - \frac{\partial}{\partial s} \int_0^t h(s,\, z)\, dz\Big|_{s=w,\, t=0} \\
&= \int_0^t \frac{\partial}{\partial s} h(s,\, z)\, dz\Big|_{s=w,\, t=w} + h(s,\, t)\Big|_{s=w,\, t=w} \\
&\qquad - \int_0^t \frac{\partial}{\partial s} h(s,\, z)\, dz\Big|_{s=w,\, t=0}
\end{aligned}
$$

よって,

$$
\begin{aligned}
\frac{d}{dw} \int_0^w h(s,\, t)\Big|_{s=w} ds &= \int_0^w \frac{\partial}{\partial s} h(s,\, z)\Big|_{s=w} dz + h(w,\, w) - \int_0^0 \frac{\partial}{\partial s} h(w,\, z)\, dz \\
&= \int_0^w f_{Y_{m_1}}(w-z) f_{Y_{m_2}}(z)\, dz
\end{aligned}
$$

□

乱数とシミュレーション：χ^2 分布（カイ自乗分布）

標準正規分布に従う乱数を発生させて，それを 2 上したものの分布を描いてみよう.

標準正規乱数の生成プログラム

```
> rnorm(10,0,1)    #  10 個の標準正規分布に従う擬似乱数を生成
 [1]  0.33597940 -0.70646612  0.08368782 -0.41461020
 [5]  0.94410755 -0.84296657  0.26558547  1.03757889
 [8] -3.57854450  0.16011053
>
> x<-rnorm(100000,0,1)    #  100000 個の標準正規分布に従う擬似
乱数を生成
>
> hist(x,freq=F,main=paste("Histgram of standart normal"))
>    # オブジェクト x のヒストグラムを作図
> curve(dnorm,-4,4,add=T,col=2)
>    ヒストグラムに標準正規分布の確率密度関数を重ね書き.
```

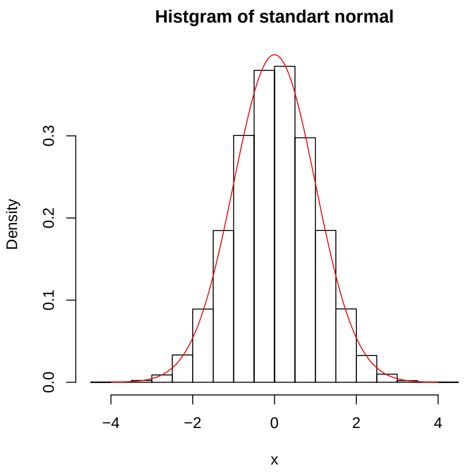


Figure 10: 100000 個の擬似標準正規分布乱数のヒストグラムと標準正規分布の確率密度関数のグラフ

自由度 1 の χ^2 分布乱数の生成プログラム

```
> x<-rnorm(4,0,1)
> x
[1]  0.2655685 -0.8015436 -0.4070102 -1.7365083
> x**2    # オブジェクト x の各成分を 2 乗
[1] 0.07052663 0.64247207 0.16565729 3.01546108
>
> x<-rnorm(100000,0,1)     #  100000 個の標準正規分布に従う擬似乱数を生成
>
> chi<-x**2  #  100000 個の標準正規分布に従う擬似乱数を 2 乗
> hist(chi,freq=F,main=paste("Histgram of chi-squared"))
>   # オブジェクト chi のヒストグラムを作図
> chidens1<-function(x){dchisq(x,1)}
>  # 自由度 1 のカイ自乗分布の確率密度関数をすくった.
> curve(chidens1,0,25,add=T,col=2)
> # 自由度 1 のカイ自乗分布のグラフを重ね書き
>
```

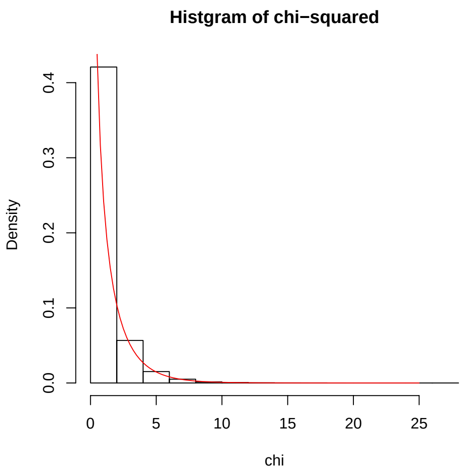



Figure 11: 100000 個の自由度 1 のカイ自乗分布擬似乱数のヒストグラムと自由度 1 のカイ自乗分布の確率密度関数のグラフ

自由度 2 の χ^2 分布乱数の生成プログラム

```
> x1<-rnorm(4,0,1)
> x2<-rnorm(4,0,1)
> x1**2
[1] 0.3031748 2.5525505 0.1713847 0.1230666
> x1
[1] 0.5506132 1.5976703 0.4139864 0.3508085
> x2
[1] -0.5799540 -0.5375283  0.1570062  1.0189586
> x1**2
[1] 0.3031748 2.5525505 0.1713847 0.1230666
> x2**2
[1] 0.33634666 0.28893663 0.02465094 1.03827668
> x1**2+x2**2
[1] 0.6395215 2.8414871 0.1960356 1.1613433
>
>
> x2<-rnorm(100000,0,1)
> x1<-rnorm(100000,0,1)
> chi2<-x1**2+x2**2
> chidens2<-function(x){dchisq(x,2)}
> hist(chi2,freq=F,main=paste("Histgram of chi-squared"))
> curve(chidens2,0,20,add=T,col=2)
>
```

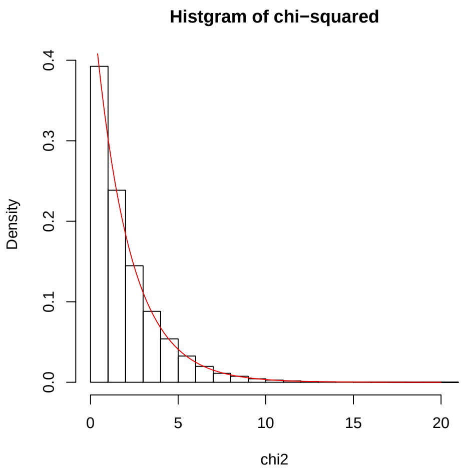


Figure 12: 100000 個の自由度 2 のカイ自乗分布擬似乱数のヒストグラムと自由度 2 のカイ自乗分布の確率密度関数のグラフ

自由度 3 の χ^2 分布乱数の生成プログラム

```
>
> x1<-rnorm(100000,0,1)
> x2<-rnorm(100000,0,1)
> x3<-rnorm(100000,0,1)
> chi3<-x1**2+x2**2+x3**2
>
> hist(chi3,freq=F,main=paste("Histgram of chi-squared"),
+ ylim=c(0,0.25))
> chidens3<-function(x){dchisq(x,3)}
> curve(chidens3,0,25,add=T,col=2)
```

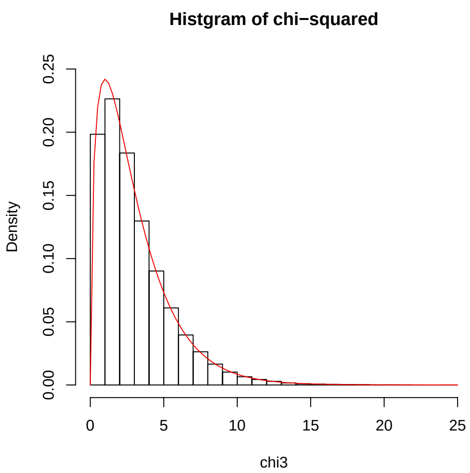



Figure 13: 100000 個の自由度 3 のカイ自乗分布擬似乱数のヒストグラムと自由度 2 のカイ自乗分布の確率密度関数のグラフ

標本分散の乱数とシミュレーション：χ^2 分布（カイ自乗分布）

⋆ 平均 μ, 分散 1 の正規分布に従う n 個の独立な確率変数を $X_1, X_2, \ldots, X_n$ とする.

⋆ $X_i - \mu$ は標準正規分布に従うので,

$$\sum_{i=1}^{n}(X_i-\mu)^2,$$

は自由度 n の χ^2 分布に従う.

⋆ 上の式で μ を $\bar{X}_n$ で置き換えた

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

の分布をシミュレーションで調べよう.

$n=4, \mu=2$ とする.

$\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ を乱数をひとつ生成するプログラム

```
> x<-rnorm(4,2,1)
> x
[1] 2.304172 1.843616 1.861465 3.018427
> sum(x)
[1] 9.02768
> mean(x)
[1] 2.25692
> x
[1] 2.304172 1.843616 1.861465 3.018427
> mean(x)
[1] 2.25692
> x-mean(x)
[1]  0.04725206 -0.41330438 -0.39545464  0.76150696
> (x-mean(x))
[1]  0.04725206 -0.41330438 -0.39545464  0.76150696
> (x-mean(x))**2
[1] 0.002232757 0.170820511 0.156384375 0.579892855
> sum((x-mean(x))**2)
[1] 0.9093305
>
```

$n = 4$, $\mu = 2$ とする.

$\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ を乱数を 100000 個生成するプログラム

```
> x<-rnorm(4,2,1)
> x
[1] 2.304172 1.843616 1.861465 3.018427
> sum(x)
[1] 9.02768
> mean(x)
[1] 2.25692
> mean(x)
[1] 2.25692
> x-mean(x)
[1]  0.04725206 -0.41330438 -0.39545464  0.76150696
> (x-mean(x))
[1]  0.04725206 -0.41330438 -0.39545464  0.76150696
> (x-mean(x))**2
[1] 0.002232757 0.170820511 0.156384375 0.579892855
> sum((x-mean(x))**2)
[1] 0.9093305
>
```

$n = 4$, $\mu = 2$ とする.

$\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ を乱数を 100000 個生成するプログラム

```
> repp<-100000
>  求める乱数の個数を指定
> svar<-rep(0,repp)
>   求めた乱数を収納するオブジェクトを作成. 100000 個の乱数を収納
> for (i in 1:repp){      # 繰り返し文の始まり
+ x<-rnorm(4,2,1)
+ svar[i]<-sum((x-mean(x))**2)
+ }                      # 繰り返し文の終わり
>
> hist(svar,freq=F,main=paste("Histgram of chi-squared"),
+ ylim=c(0,0.25))
>  # オブジェクト svar のヒストグラムを作図
> chidens3<-function(x){dchisq(x,3)}
> curve(chidens3,0,12,add=T,col=2)
>
```

$n = 4$, $\mu = 2$ とする.

$\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ を乱数を 100000 個生成するプログラムための関数

```
repp<-1000
chidis<-function(x,me,sig){
svar<-rep(0,repp)
for (i in 1:repp){
x<-rnorm(x,me,sqrt(sig))
svar[i]<-sum((x-mean(x))**2)
}
return(svar)
}
hist(chidis(5,0,1),freq=F,ylim=c(0,0.25))
```

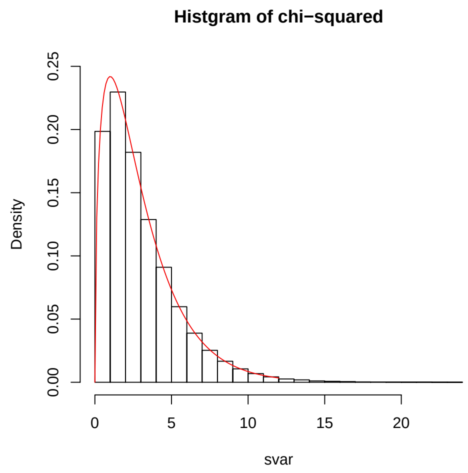


Figure 14: 100000 個の自由度 3 のカイ自乗分布擬似乱数のヒストグラムと自由度 2 のカイ自乗分布の確率密度関数のグラフ

ここまでのまとめ

- χ^2 分布の定義と性質を説明しました．
  - 自由度 m の χ^2 分布の確率密度関数，
  - χ^2 分布の再帰性，
  - 標準正規分布の 2 乗は自由度 1 の χ^2 分布に従うこと，
  - χ^2 分布の再帰性と標準正規分布の 2 乗は自由度 1 の χ^2 分布に従うことをシミュレーションで確認，
  - 標本分散と χ^2 分布の関係．

以上の事項についての定義や性質について説明しました．

問題 2

(1) m を

自由度を決定

```
> m<-round(runif(1,3.5,6.5))
> m
[1] 4
>
```

で定めよ．

(2) 定めた自由度だけの標準正規分布の擬似乱数を 2 乗してそれを足し合わせることにより，自由度 m の χ^2 分布乱数を作成し，ただしい自由度の χ^2 分布の確率密度関数を上書き（20816***-目白花子-091120-chisq1.pdf）せよ．

- 締め切りは 2009 年 11 月 13 日（金）13 時

- 20816***-目白花子-091120.txt
- 提出はメール：mtoukei(at)mp[dot]jwu[dot]ac[dot]jp

ヒント

```
>
> x1<-rnorm(100000,0,1)
> x2<-rnorm(100000,0,1)
> x3<-rnorm(100000,0,1)
> chi<-x1**2+x2**2+x3**2
>
> hist(chi,freq=F,main=paste("Histgram of chi-squared"),
+ ylim=c(0,0.25))
> chidens<-function(x){dchisq(x,??)}
> curve(chidens,0,25,add=T,col=2)
```


ヒント


```
>
> x1<-rnorm(100000,0,1)
> x2<-rnorm(100000,0,1)
> x3<-rnorm(100000,0,1)
> chi<-x1**2+x2**2+x3**2
>
> hist(chi,freq=F,main=paste("Histgram of chi-squared"),
+ ylim=c(0,0.25))
> chidens<-function(x){dchisq(x,??)}
> curve(chidens,0,25,add=T,col=2)
```


← ここを指定され自由度に従いかきかえる！