確率統計と情報処理・演習（2007 年度後期）

確率分布

2007 年 12 月 07 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

- 中心極限定理
  - 一様分布と中心極点定理
  - さまざまな分布と中心極限定理
  - ２項分布と中心極限定理

一様分布の復習とたたみ込み

$X_1, X_2, \ldots, X_n$ が独立同一に区間 $[0, 1]$ 上の一様分布に従っているとする．すなわち，$0 \le a_i < b_i \le 1\,(i = 1, 2, \ldots, n)$ と任意の正の整数 $n \ge 2$ に対して

$$\begin{aligned}&\mathbb{P}(a_1 < X_1 \le b_1, a_2 < X_2 \le b_2, \ldots, a_n < X_n \le b_n)\\ &\quad = \Pi_{i=1}^{n} \mathbb{P}(a_i < X_i \le b_i)\\ &\quad = \mathbb{P}(a_1 < X_1 \le b_1)\mathbb{P}(a_2 < X_2 \le b_2) \times \cdots \times \mathbb{P}(a_n < X_n \le b_n)\end{aligned}$$

と

$$\mathbb{P}(a_1 < X_i \le b_1) = \int_{a_1}^{b_1} f_X(x)\,dx$$

が成立．ただし，

$$f_X(x) = \mathbb{1}_{[0,1]}(x) = \begin{cases} 1 & (0 \le x \le 1) \\ 0 & (\text{その他}) \end{cases}$$

$n = 1$ のとき

$$\begin{aligned}\mathbb{E}[X_1] &= \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^1 x\,dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}\\ \mathbb{E}[X_1^2] &= \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \int_0^1 x^2\,dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3}\\ \mathbb{VAR}[X_1] &= \mathbb{E}[X_1^2] - \{\mathbb{E}[X_1]\}^2 = \frac{1}{12}\end{aligned}$$

$0 < x < 1$ に対して，$F_X(x) = \mathbb{P}(X_1 \le x) = \int_0^x f_X(t)\,dt = \int_0^x dt = x$ より

$$F_X(x) = \begin{cases} 0 & (x < 0) \\ x & (0 \le x \le 1) \\ 1 & (x > 1) \end{cases}$$

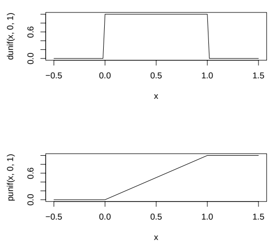

Figure 1: 一様分布の確率密度関数と累積分布関数のグラフ

$n = 2$ のとき

X_1 と X_2 は独立に閉区間 $[0, 1]$ 上の一様分布に独立同一に従っているとき，$Z = X_1 + X_2$ の確率密度関数はつぎで与えられある：

$$f_Z(z) = \begin{cases} z & (0 \le z \le 1) \\ 2 - z & (1 < z \le 2) \\ 0 & (\text{その他}) \end{cases}$$

平均は

$$\begin{aligned}\mathbb{E}[Z] &= \int_{-\infty}^{\infty} z f_Z(z)\,dz = \int_0^1 z^2\,dz + \int_1^2 z(2-z)\,dz = \left[\frac{z^3}{3}\right]_0^1 + \left[z^2 - \frac{z^3}{3}\right]_1^2\\ &= \frac{1}{3} + 4 - \frac{8}{3} - 1 + \frac{1}{3} = 1\end{aligned}$$

分散を求めるために

$$\mathbb{VAR}[Z] = \mathbb{E}[Z^2] - \{\mathbb{E}[Z]\}^2$$

に注意する．

$$\begin{aligned}\mathbb{E}[Z^2] &= \int_{-\infty}^{\infty} z^2 f_Z(z)\,dz = \int_0^1 z^3\,dz + \int_1^2 z^2(2-z)\,dz \\ &= \left[\frac{z^4}{4}\right]_0^1 + \left[\frac{2z^3}{3} - \frac{z^4}{4}\right]_1^2 = \frac{1}{4} + \frac{16}{3} - \frac{16}{4} - \frac{2}{3} + \frac{1}{4} = \frac{7}{6}\end{aligned}$$

したがって，

$$\mathbb{VAR}[Z] = \frac{7}{6} - 1 = \frac{1}{6}$$

$X_1 + X_2$ 分布のグラフの出力

```
> u2<-function(x){ # 確率密度関数の計算
+ switch(length(which(c((x>=0),(x>=1),(x>2))))+1,0,x,2-x,0)
+ }
> u2(-1)
[1] 0
> u2(0.3)
[1] 0.3
> u2(1.1)
[1] 0.9
> u2(2.1)
[1] 0
>
> x<-seq(-0.5,2.5,by=0.01)
> plot(x,sapply(x,u2),type="l")    # グラフの作図のコマンド
>
```


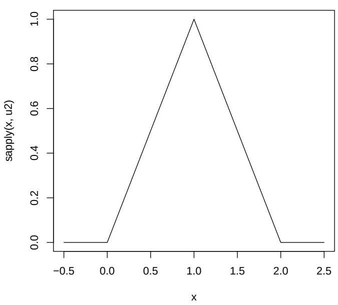

Figure 2: $X_1 + X_2$ の確率密度関数のグラフ

つぎに，Z を標準化する：

$$T = \frac{Z - \mathbb{E}[Z]}{\sqrt{\mathbb{VAR}[Z]}} = \frac{Z-1}{\sqrt{1/6}} = \sqrt{6}(Z-1)$$

T の確率密度関数 $f_T(t)$ を求めよう：$0 \le z \le 2$ のとき，$f_Z(z) > 0$ なので，$-\sqrt{6} \le t \le \sqrt{6}$ のとき，$f_T(t) > 0$ となる．

$-\sqrt{6} \le t \le 0\,(0 \le 1 + t/\sqrt{6} \le 1)$ のとき，

$$\mathbb{P}(T \le t) = \mathbb{P}(\sqrt{6}(Z-1) \le t) = \mathbb{P}(Z \le 1 + t/\sqrt{6}) = \int_{-\infty}^{1+t/\sqrt{6}} f_Z(z)\,dz$$

よって，

$$f_T(t) = \frac{d}{dt}\mathbb{P}(T \le t) = \left(1 + \frac{t}{\sqrt{6}}\right)\frac{1}{\sqrt{6}}$$

つぎに，$0 < t \le \sqrt{6}\,(1 < 1 + t/\sqrt{6} \le 2)$ のとき，

$$\mathbb{P}(T \le t) = \mathbb{P}(\sqrt{6}(Z-1) \le t) = \mathbb{P}(Z \le 1 + t/\sqrt{6}) = \int_{-\infty}^{1+t/\sqrt{6}} f_Z(z)\,dz$$

よって，

$$f_T(t) = \frac{d}{dt}\mathbb{P}(T \le t) = \left(2 - \left(1 + \frac{t}{\sqrt{6}}\right)\right)\frac{1}{\sqrt{6}}$$

したがって，

$$f_T(t) = \begin{cases} \left(1 + \frac{t}{\sqrt{6}}\right)\frac{1}{\sqrt{6}} & (-\sqrt{6} \le t \le 0) \\ \left(2 - \left(1 + \frac{t}{\sqrt{6}}\right)\right)\frac{1}{\sqrt{6}} & (0 < t \le \sqrt{6}) \\ 0 & (\text{その他}) \end{cases}$$

$X_1 + X_2$ 分布のグラフの出力

```
> u2s<-function(x){      # T の確率密度関数の計算
+ u2(x/sqrt(6)+1)/sqrt(6)
+ }
> u2s(0)
[1] 0.4082483
> u2s(1)
[1] 0.2415816
>
> x<-seq(-3,3,by=0.01)
> plot(x,sapply(x,u2s),type="l")      # T の確率密度関数の作図
> curve(dnorm,-3,3,add=T,col=2)      # 標準正規分布の確率密度関数の作図
>
```

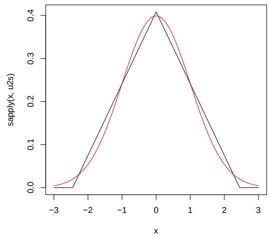

Figure 3: T と標準正規分布の確率密度関数のグラフ

$n = 3$ のとき

$X_1,\, X_2,\, X_3$ は閉区間 $[0,\,1]$ 上の一様分布に独立同一に従っているとき，$Z = X_1 + X_2 + X_3$ の確率密度関数はつぎで与えられる：

$$f_Z(z) = \begin{cases} \frac{z^2}{2} & (0 \le z \le 1) \\ -z^2 + 3z - \frac{3}{2} & (1 < z \le 2) \\ \frac{(z-3)^2}{2} & (2 < z \le 3) \\ 0 & (\text{その他}) \end{cases}$$

期待値と分散($X_1,\, X_2,\, X_3$ は独立であることを利用して)は

$$\begin{aligned} \mathbb{E}[Z] &= \mathbb{E}[X_1] + \mathbb{E}[X_2] + \mathbb{E}[X_3] = \frac{3}{2} \\ \mathbb{VAR}[Z] &= \mathbb{VAR}[X_1] + \mathbb{VAR}[X_2] + \mathbb{VAR}[X_3] = \frac{1}{4} \end{aligned}$$

確率変数は連続型として以下の議論を行う．X_1 と X_2 は独立のとき，同時確率密度関数は $f_{X_1,\,X_2}(x_1,\,x) = f_{X_1}(x_1) f_{X_2}(x_2)$ とかける．ただし，f_{X_1} と f_{X_2} は X_1 と X_2 の確率密度関数とした．

このとき，

$$\begin{aligned} \mathbb{E}[X_1 + X_2] &= \int\int (x_1 + x_2) f_{X_1}(x_1) f_{X_2}(x_2)\, dx_1\, dx_2 \\ &= \int\int x_1 f_{X_1}(x_1) f_{X_2}(x_2)\, dx_1\, dx_2 + \int\int x_2 f_{X_1}(x_1) f_{X_2}(x_2)\, dx_1\, dx_2 \\ &= \int f_{X_2}(x_2)\, dx_2 \int x_1 f_{X_1}(x_1)\, dx_1 + \int f_{X_1}(x_1)\, dx_1 \int x_2 f_{X_2}(x_2)\, dx_2 \\ &= \mathbb{E}[X_1] + \mathbb{E}[X_2] \end{aligned}$$

さらに，

$$\begin{aligned} \mathbb{VAR}[X_1 + X_2] &= \mathbb{E}[\{X_1 + X_2 - \mathbb{E}[X_1 + X_2]\}^2] \\ &= \mathbb{E}[\{(X_1 - \mathbb{E}[X_1]) + (X_2 - \mathbb{E}[X_2])\}^2] \\ &= \mathbb{E}[(X_1 - \mathbb{E}[X_1])^2 + (X_2 - \mathbb{E}[X_2])^2 + 2(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])] \\ &= \mathbb{E}[(X_1 - \mathbb{E}[X_1])^2] + \mathbb{E}[(X_2 - \mathbb{E}[X_2])^2] + 2\mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])] \\ &= \mathbb{VAR}[X_1] + \mathbb{VAR}[X_2] + 2\mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])] \end{aligned}$$

しかし，

$$\begin{aligned} &\mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])] \\ &\quad = \int\int (x_1 - \mathbb{E}[X_1])(x_2 - \mathbb{E}[X_2]) f_{X_1}(x_1) f_{X_2}(x_2)\, dx_1\, dx_2 \\ &\quad = \int (x_1 - \mathbb{E}[X_1]) f_{X_1}(x_1)\, dx_1 \int (x_2 - \mathbb{E}[X_2]) f_{X_2}(x_2)\, dx_2 = 0 \end{aligned}$$

$X_1 + X_2 + X_3$ 分布のグラフの出力

```
> u3<-function(x){
+ switch(length(which(c((x>=0),(x>=1),(x>=2),(x>3))))+1,
+ 0,x^2/2,-x^2+3*x-3/2,(x-3)^2/2,0)
+ }
> x<-seq(-0.5,3.5,by=0.01)
>
> plot(x,sapply(x,u3),type="l")
>
```


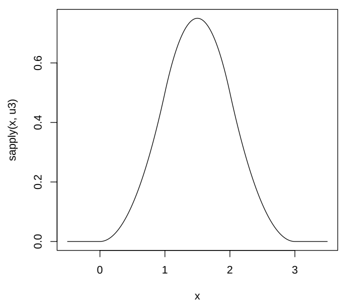

Figure 4: $X_1 + X_2 + X_3$ の確率密度関数のグラフ

つぎに，Z を標準化する：

$$T = \frac{Z - \mathbb{E}[Z]}{\sqrt{\mathbb{VAR}[Z]}} = \frac{Z - \frac{3}{2}}{\sqrt{1/4}} = 2\left(Z - \frac{3}{2}\right)$$

T の確率密度関数は

$$\begin{aligned} f_T(t) &= \frac{d}{dt}\mathbb{P}(T \le t) = \frac{d}{dt}\mathbb{P}\left(Z \le \frac{3}{2} + \frac{t}{2}\right) \\ &= \frac{d}{dt}\int_{-\infty}^{\frac{3}{2}+\frac{t}{2}} f_Z(z)\,dz = f_Z\left(\frac{3}{2} + \frac{t}{2}\right)\frac{1}{2} \end{aligned}$$

よって，

$$f_T(z) = \begin{cases} \frac{\left(\frac{3}{2}+\frac{t}{2}\right)^2}{2} & (-3 \le z \le -1) \\ -\left(\frac{3}{2}+\frac{t}{2}\right)^2 + 3\left(\frac{3}{2}+\frac{t}{2}\right) - \frac{3}{2} & (-1 < z \le 1) \\ \frac{\left(\left(\frac{3}{2}+\frac{t}{2}\right)-3\right)^2}{2} & (1 < z \le 3) \\ 0 & (\text{その他}) \end{cases}$$

$X_1 + X_2$ 分布のグラフの出力

```
> u3s<-function(x){
+ u3(1.5+(x*0.5))*0.5
+ }
> x<-seq(-3.5,3.5,by=0.01)
>
> plot(x,sapply(x,u3s),type="l",ylim=c(0,0.4))       # y 軸の長さを調整
> curve(dnorm,-4,4,add=T,col=2)
>
```


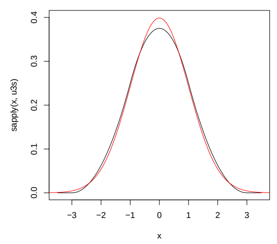

Figure 5: T と標準正規分布の確率密度関数のグラフ

一様乱数を用いた正規乱数の作成

$X_1, X_2, \ldots, X_{12}$ は独立に閉区間 $[0, 1]$ の一様分布に従っているとする．

$$Z = X_1 + X_2 + \cdots + X_{12}$$

とおく．このとき，

$$\begin{aligned} \mathbb{E}[Z] &= \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_{12}] = 6 \\ \mathbb{VAR}[Z] &= \mathbb{VAR}[X_1] + \mathbb{VAR}[X_2] + \cdots + \mathbb{VAR}[X_{12}] = 1 \end{aligned}$$

さらに

$$T = \frac{Z - \mathbb{E}[Z]}{\sqrt{\mathbb{VAR}[Z]}} = Z - 6$$

は標準正規分布に従うとみなしても（実用上は）差し支えない！

12 個の一様乱数から正規乱数を作成するためのプログラム

```
> repp=50000        # 50000 個の T を作成
> clt<-matrix(0,repp)
> for (i in 1:repp){
+ x<-runif(12,0,1)
+ clt[i]<-sum(x)-6
+ }
> hist(clt,nclass=40,freq=F)
> curve(dnorm,-4,4,add=T,col=2)
>
```


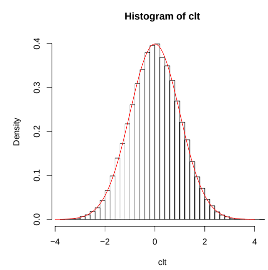

Figure 6: 12 個の一様乱数から 50000 の正規乱数のヒストグラムと標準正規分布の確率密度関数

中心極限定理

$X_1, X_2, \ldots, X_n$ は独立同一分布に従う確率変数として，$\mathbb{VAR}[X_1] < \infty$ とする．

$$Z = X_1 + X_2 + \cdots + X_n, \qquad \bar{X}_n = \frac{1}{n} Z$$

としたとき，

$$\begin{aligned}
\mathbb{E}[Z] &= \mathbb{E}[X_1 + X_2 + \cdots + X_n] \\
&= \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n] = n\mathbb{E}[X_1] \\
\mathbb{VAR}[Z] &= \mathbb{VAR}[X_1 + X_2 + \cdots + X_n] \\
&= \mathbb{VAR}[X_1] + \mathbb{VAR}[X_2] + \cdots + \mathbb{VAR}[X_n] = n\mathbb{VAR}[X_1]
\end{aligned}$$

ので，

$$T = \frac{Z - \mathbb{E}[Z]}{\sqrt{\mathbb{VAR}[Z]}} = \frac{n\bar{X}_n - n\mathbb{E}[X_1]}{\sqrt{n\mathbb{VAR}[X_1]}} = \frac{\sqrt{n}(\bar{X}_n - \mathbb{E}[X_1])}{\sqrt{\mathbb{VAR}[X_1]}}$$

中心極限定理

$n \to \infty$ のとき，

$$\mathbb{P}\left(\frac{\sqrt{n}(\bar{X}_n - \mathbb{E}[X_1])}{\sqrt{\mathbb{VAR}[X_1]}} \le x \right) \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt, \qquad {}^\forall x \in \mathbb{R}$$

例：自由度 1 の χ^2 分布の場合

$X_1, X_2, \ldots, X_n$ は独立同一に自由度 1 の χ^2 分布に従うとする．このとき，

$$\begin{aligned}
\mathbb{E}[X_1] &= 1 \\
\mathbb{VAR}[X_1] &= 2
\end{aligned}$$

となる．中心極限定理より

$$\frac{\sqrt{n}(\bar{X}_n - 1)}{\sqrt{2}}$$

は標準正規分布に近づくことがわかる．

自由度 1 の χ^2 分布の和と中心極限定理

```
> jiken<-function(x,y){
+ clt<-rep(0,y)
+ for (i in 1:y){
+ z<-rchisq(x,1)
+ clt[i]<-sqrt(x)*(mean(z)-1)/sqrt(2)
+ }
+ hist(clt,nclass=40,freq=F,xlim=c(-4,4))
+ curve(dnorm,-4,4,add=T,col=2)
+ }
> op<-par(mfrow=c(2,2))
> jiken(1,50000)
> jiken(2,50000)
> jiken(5,50000)
> jiken(10,50000)
>
> jiken(20,50000)
> jiken(30,50000)
> jiken(40,50000)
> jiken(60,50000)
>
```


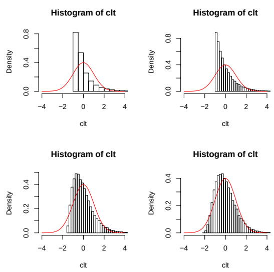

Figure 7: $n = 1, 2, 5, 10$ のときの T のヒストグラムと標準正規分布の確率密度関数

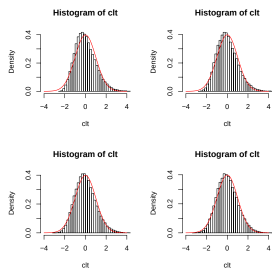

Figure 8: $n = 20, 30, 40, 60$ のときの T のヒストグラムと標準正規分布の確率密度関数

$\boxed{\text{問題 1}}$　$X_1, X_2, \ldots, X_n$ を閉区間 $[0, 1]$ 上の一様分布に独立に従うとし，

$$T = \frac{\sqrt{n}(\bar{X}_n - \frac{1}{2})}{\frac{1}{\sqrt{12}}}$$

とする．

- $n = 1, 4, 8, 12$ に対して，T 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を書き込んだものを作成（mejiro-hanako-071214-uniform.pdf）し，グラフから観察できること（簡単なことでよい）を述べよ．
- mejiro-hanako-071214.txt
- 2007 年 12 月 14 日(金）13　時締め切り．
- 提出先のメールアドレス:mtouke[at]mp[dot]jwu[dot]ac[dot]jp

ヒント

```
> jiken<-function(x,y){
+ clt<-rep(0,y)
+ for (i in 1:y){
+ z<-runif(x,0,1)
+ clt[i]<-sqrt(x)*(mean(z)-???)/(1/sqrt(12))
+ }
+ hist(clt,nclass=40,freq=F,xlim=c(-4,4),ylim=c(0,0.45))
+ curve(dnorm,-4,4,add=T,col=2)
+ }
```


$\boxed{\text{問題 2}}$　$X_1, X_2, \ldots, X_n$ は自由度 2 の χ^2 分布に独立に従うとし，

$$T = \frac{\sqrt{n}(\bar{X}_n - \mathbb{E}[X_1])}{\sqrt{\mathbb{VAR}[X_1]}}$$

とする．

- T のヒストグラムが標準正規分布の確率密度関に近づく様子がわかるようにいくつかの n を適当にったグラフを作成せよ（mejiro-hanako-071214-chi.pdf）し，グラフから観察できること（簡単なことでよい）を述べよ．
- mejiro-hanako-061208.txt
- 2007 年 12 月 14 日(金）13　時締め切り．
- 提出先のメールアドレス:mtouke[at]mp[dot]jwu[dot]ac[dot]jp

ヒント

$$\mathbb{E}[X_1] = 2, \qquad \mathbb{VAR}[X_1] = 4$$

```
> jiken<-function(x,y){
+ clt<-rep(0,y)
+ for (i in 1:y){
+ z<-rchisq(x,2)
+ clt[i]<-sqrt(x)*(mean(z)-???)/???
+ }
+ hist(clt,nclass=40,freq=F,xlim=c(-4,4),ylim=c(0,0.45))
+ curve(dnorm,-4,4,add=T,col=2)
+ }
```