**「データ解析」（下平英寿）　講義資料２**

# 確率分布と確率変数

- 目標 :「確率」に関する復習
  1. 分布関数と密度関数を復習する.
  2. R による擬似乱数生成を通して「確率変数」を理解する.
  3. 一様分布，正規分布，ガンマ分布などの性質.
  4. 大数 (たいすう) の法則，中心極限定理など確率論の基本的な定理.

## 1　分布関数と密度関数

- 確率変数 $X$
- １変量の確率変数には，実数連続値を取るもの，離散値（たとえば整数値）だけを取るもの，さらにそれらの混合などがあるが，ここでは連続値確率変数だけを考える.
- 実現値 $x$
- 確率分布関数 $F_X(x) = \Pr\{X \leq x\}$
- 確率密度関数 $f_X(x) = \frac{dF_X}{dx}$
- 逆に $F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy$
- 期待値 $E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx$
- 分散 $V(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$
- ２変量 $X, Y$ に関しては，共分散 $\mathrm{cov}(X,Y) = E\{(X - E(X))(Y - E(Y))\}$. 特に $\mathrm{cov}(X,X) = V(X)$.
- 相関係数 $\rho(X,Y) = \mathrm{cov}(X,Y)/\sqrt{V(X)V(Y)}$.

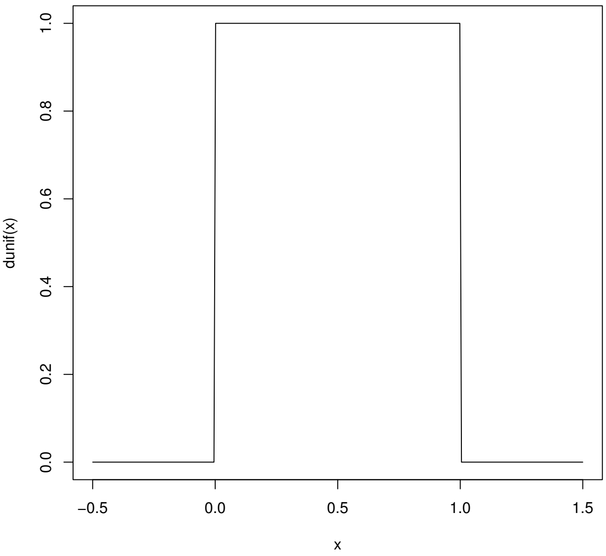
## 2　一様分布

### 2.1　密度関数，分布関数

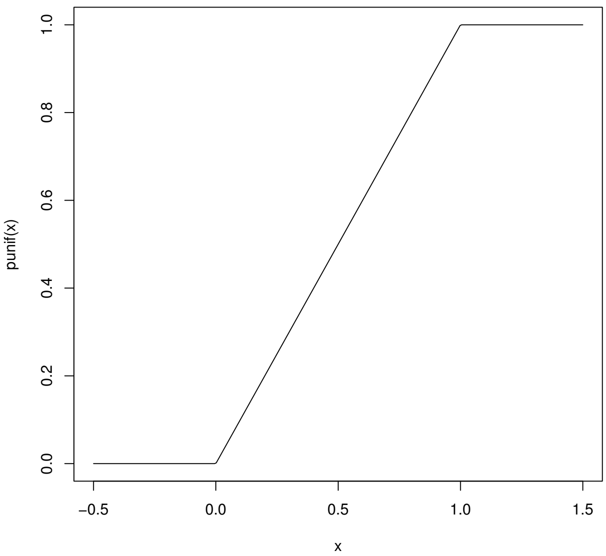
- 一様分布 $U(0,1)$ の密度関数 $f_X(x)$ と分布関数 $F_X(x)$ をプロットする.

```
# run0006.R
# 一様分布 U(0,1) 密度関数，分布関数
```

```
x <- seq(-0.5,1.5,length=300) # (-0.5,1.5)の区間に300個の点を取る
plot(x,dunif(x),type="l") # 密度関数のプロット. "l"は"line"の意味で, 点をプロットする代わりに, 線分でつなぐ.
dev.copy2eps(file="run0006-d.eps")
plot(x,punif(x),type="l") # 分布関数のプロット
dev.copy2eps(file="run0006-p.eps")
rm(x) # オブジェクトの削除
```




run0006-d




run0006-p

- 区間 $(0,1)$ の一様分布: $U(0,1)$

$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & x \leq 0, \text{or } x \geq 1 \end{cases}$$

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

$$E(X) = \frac{1}{2}, \quad V(X) = \frac{1}{12}$$

- 区間 $(a,b)$ の一様分布: $U(a,b)$

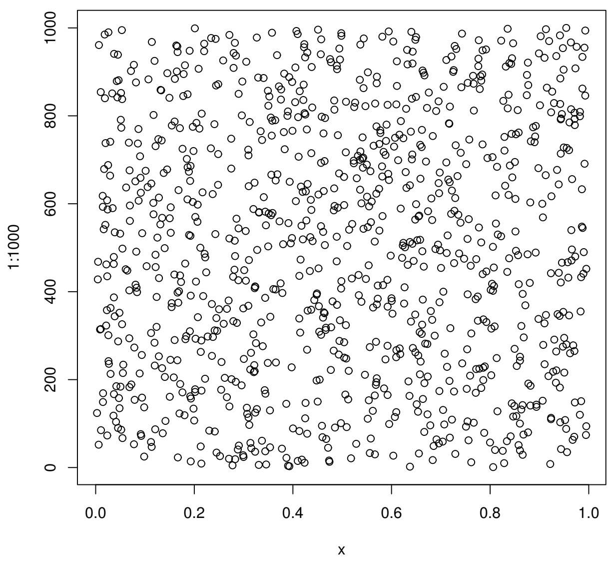
$$f_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & x \leq a, \text{or } x \geq b \end{cases}$$

$$F_X(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$

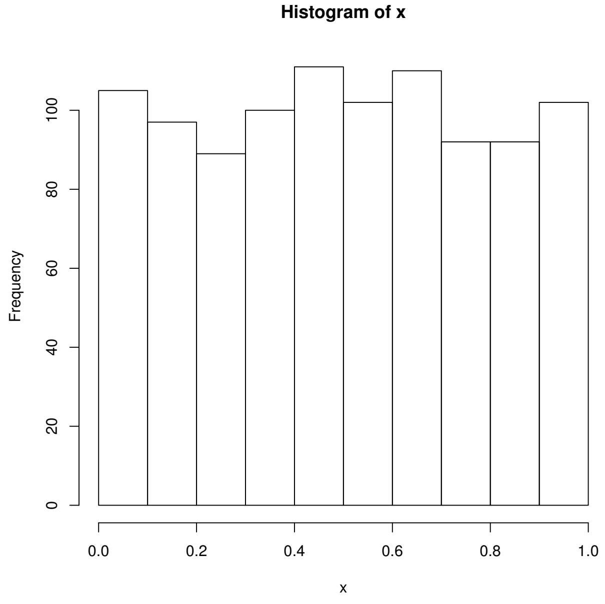
## 2.2 擬似乱数の生成

- 一様分布 $U(0,1)$ に従う擬似乱数を多数（1000 個）生成し，３通りのプロットをする．
- 擬似乱数：十分にランダムな系列を生成するアルゴリズム．実質的に確率変数とみなしてよい．最近の主流はメルセンヌ・ツイスター法.

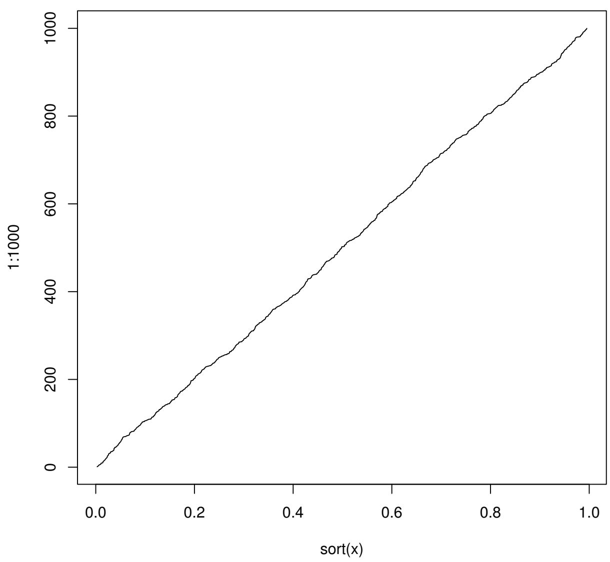
```
# run0014.R
# 一様分布 U(0,1) 擬似乱数の生成

x <- runif(1000) # (0,1)区間の一様乱数を1000個生成
plot(x,1:1000) # 横軸に乱数，縦軸に番号1,...,1000をプロット
dev.copy2eps(file="run0014-r.eps")
hist(x) # ヒストグラム
dev.copy2eps(file="run0014-h.eps")
plot(sort(x),1:1000,type="l") # 横軸に乱数をソートしたものをプロット
dev.copy2eps(file="run0014-s.eps")
boxplot(x) # 箱ヒゲ図(Boxプロット)
dev.copy2eps(file="run0014-b.eps")
rm(x) # オブジェクトの削除
```



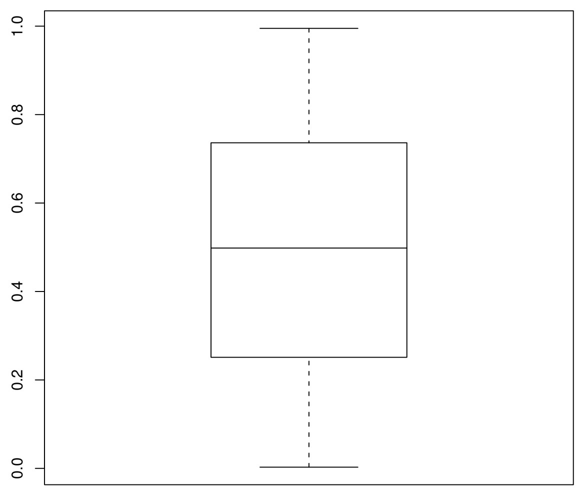

run0014-r




run0014-h




run0014-s



run0014-b

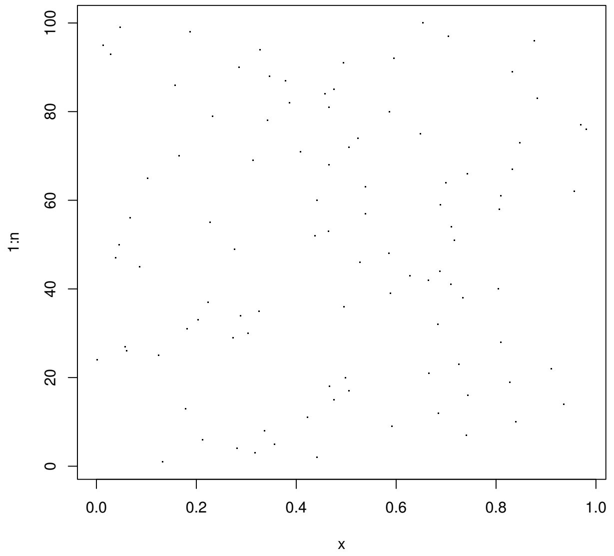
- run0014-h.eps は横軸=x，縦軸=幅 0.1 の各領域に入った回数．この縦軸を 1000*0.1=100 で割ったグラフ (ヒストグラム) は，縦軸が確率密度になる．シミュレーション回数がここでは $n = 1000$ だが，もし $n \to \infty$ とすれば，$f_X(x)$ に収束する．(run0006-d.eps 参照)
- run0014-s.eps は横軸=sort(x)，縦軸=1:1000．この縦軸を 1000 で割ったグラフ (経験分布関数) は，縦軸が確率になる．$n \to \infty$ とすれば，$F_X(x)$ に収束する．(run0006-p.eps 参照)
- 箱ヒゲ図 (Box プロット) の箱の中心=メディアン (50%のパーセンタイル点)，箱の両端 (hinges)=ほぼ 25%と 75%のパーセンタイル点．一般に任意の $0 \le \alpha \le 1$ に対して $100\alpha$ パーセントのパーセンタイル点というは，データを小さい順に並べて $n\alpha$ 番目の数. (もう少し正確な定義は「講義資料 3」を参照)

## 2.3 擬似乱数（その２）

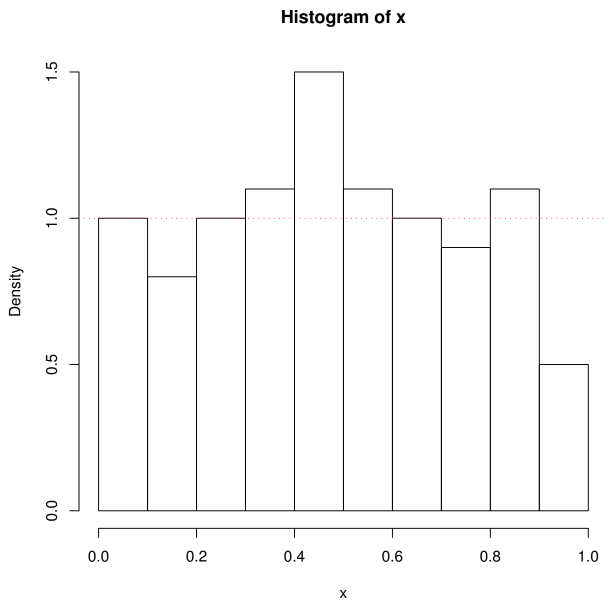
- ヒストグラムと経験分布関数のプロットを作成する．シミュレーション回数を $n = 100, 1000, 10000$ と増加させると，ヒストグラムは確率密度関数に，経験分布関数は確率分布関数に収束していく様子がわかる．

```
# run0007.R
# 一様分布 U(0,1) n=10^2,10^3,10^4

for(i in 2:4) {
  n <- 10^i  # n を変化させて実行
  x <- runif(n)
  plot(x,1:n,pch=".")  # 小さい点に変更
  dev.copy2eps(file=paste("run0007-r",i,".eps",sep=""))
```

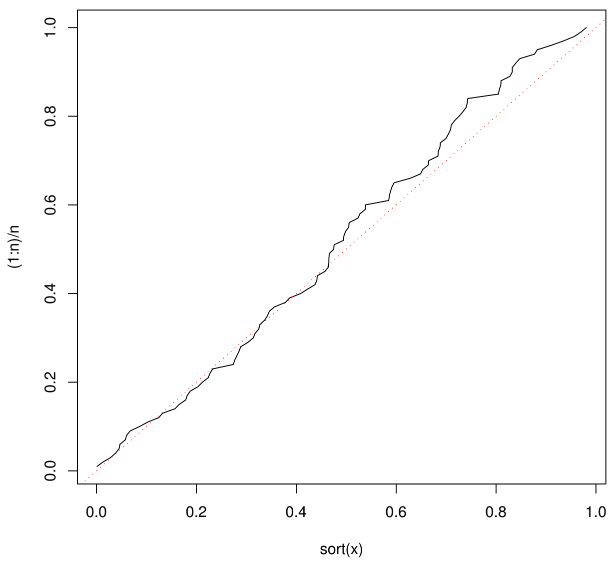
```
  hist(x,freq=F) # 縦軸=確率密度 f_x(x)
  abline(h=1,lty=3,col=2) # 極限値
  dev.copy2eps(file=paste("run0007-h",i,".eps",sep=""))
  plot(sort(x),(1:n)/n,type="l") # 縦軸=確率 F_X(x)
  abline(0,1,lty=3,col=2) # 極限値
  dev.copy2eps(file=paste("run0007-s",i,".eps",sep=""))
  boxplot(x) # 箱ヒゲ図(Box プロット)
  dev.copy2eps(file=paste("run0007-b",i,".eps",sep=""))
}
rm(i,n,x) # オブジェクトの削除
```



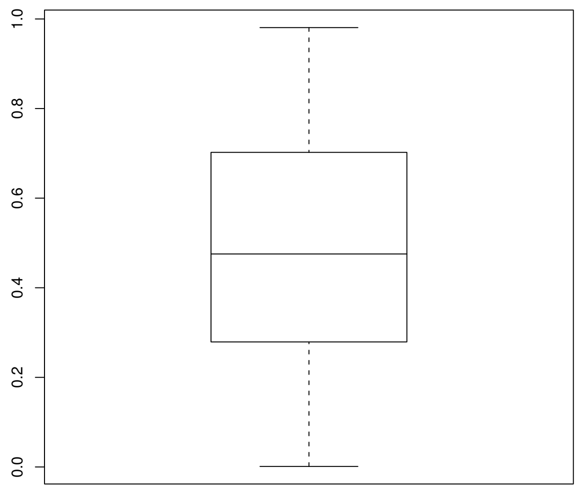

run0007-r2



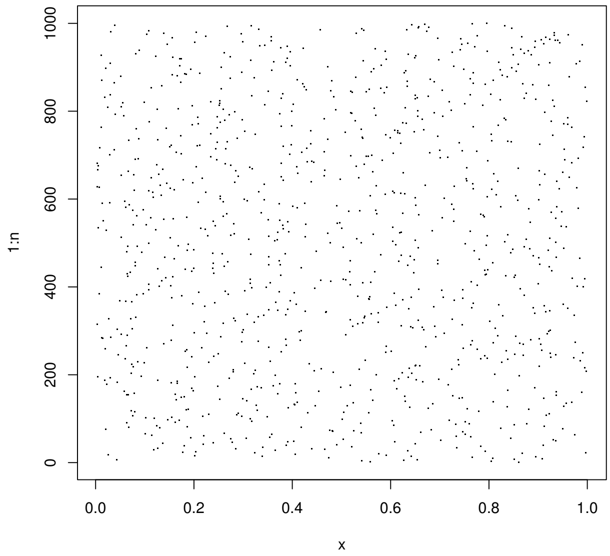

run0007-h2



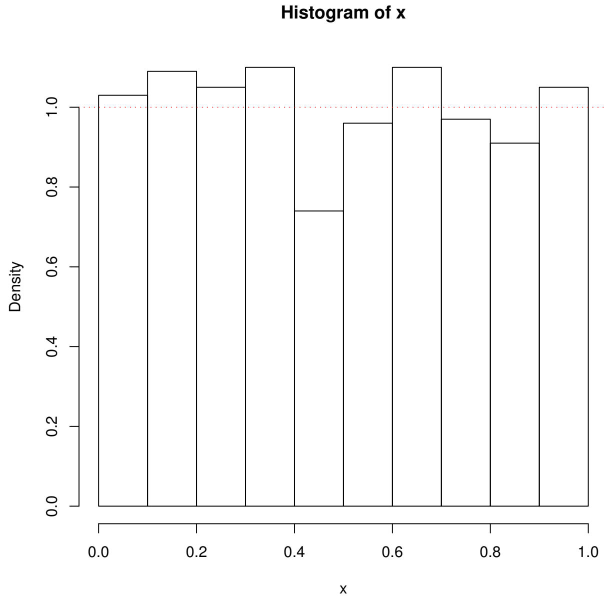

run0007-s2



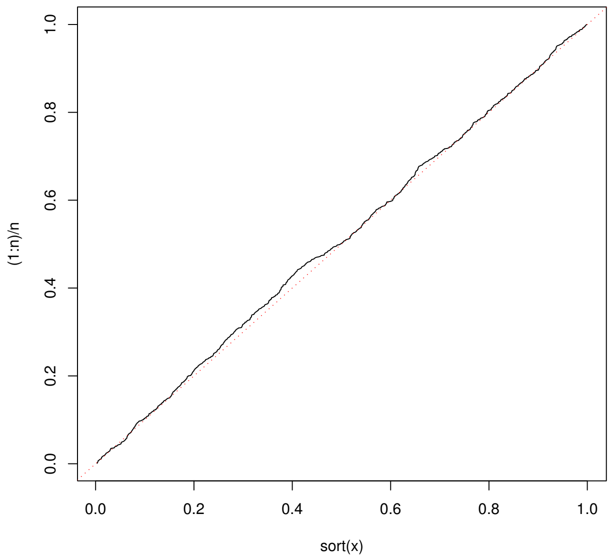
run0007-b2



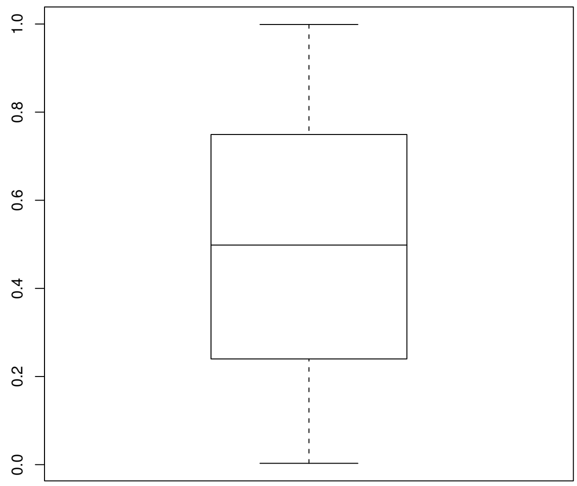

run0007-r3



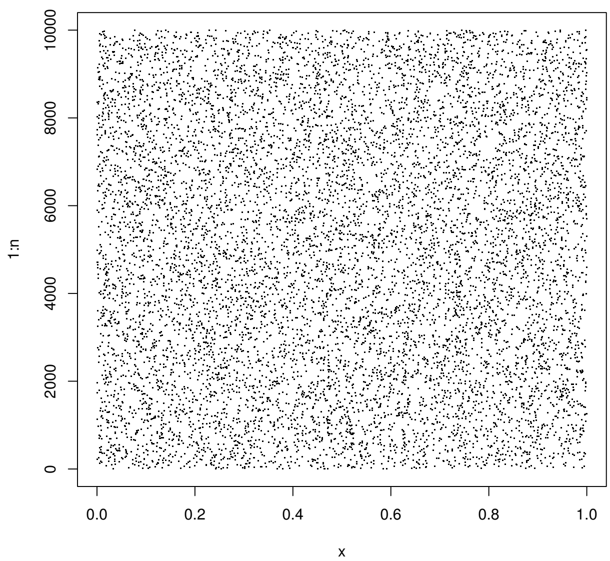

run0007-h3



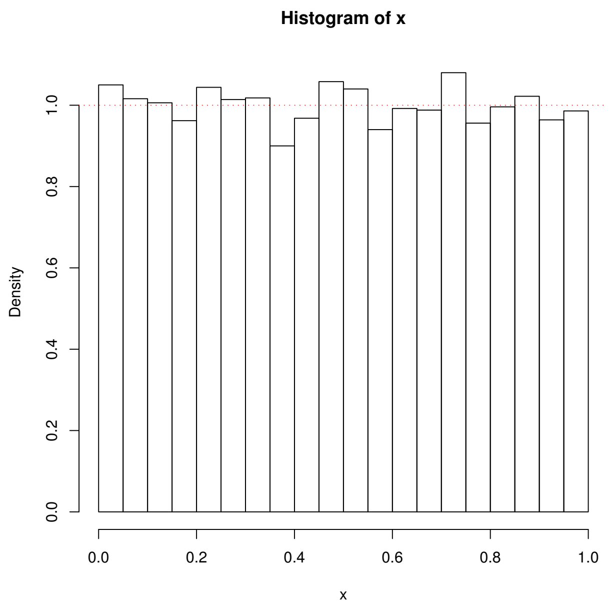

run0007-s3



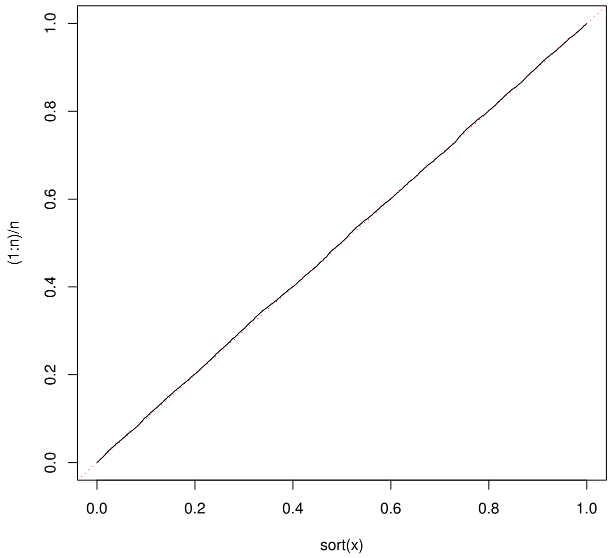
run0007-b3



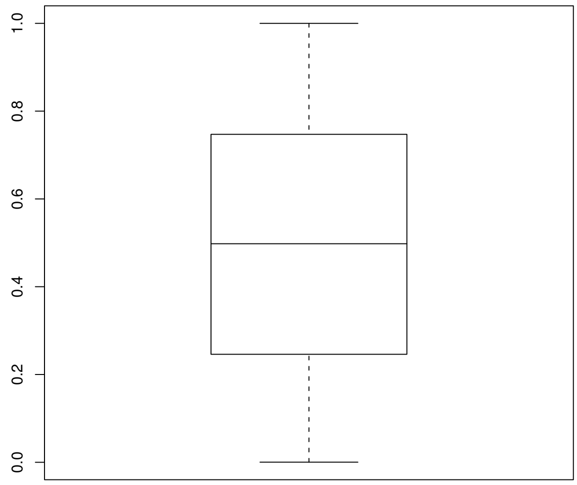

run0007-r4




run0007-h4




run0007-s4



run0007-b4

## 2.4 大数の法則

- 平均 $\mu$，分散 $\sigma^2$ の確率変数の列（各 $X_i$ は互いに独立）

$$X_1, X_2, \ldots$$

$$E(X_i) = \mu, \quad V(X_i) = \sigma^2$$

- $m$ 個の和

$$S = X_1 + \cdots + X_m$$

$$E(S) = m\mu, \quad V(S) = m\sigma^2$$

(説明) 一般に系列が独立でない場合は,

$$V(S) = V(X_1 + \cdots + X_m) = \sum_{i=1}^{m} \sum_{j=1}^{m} \mathrm{cov}(X_i, X_j)$$
$$= V(X_1) + \cdots + V(X_m) + 2 \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \mathrm{cov}(X_i, X_j)$$

いま系列が独立と仮定しているので，$\mathrm{cov}(X_i, X_j) = 0$ $(i \neq j)$ である.

- $m$ 個の平均

$$Y = \frac{X_1 + \cdots + X_m}{m}$$

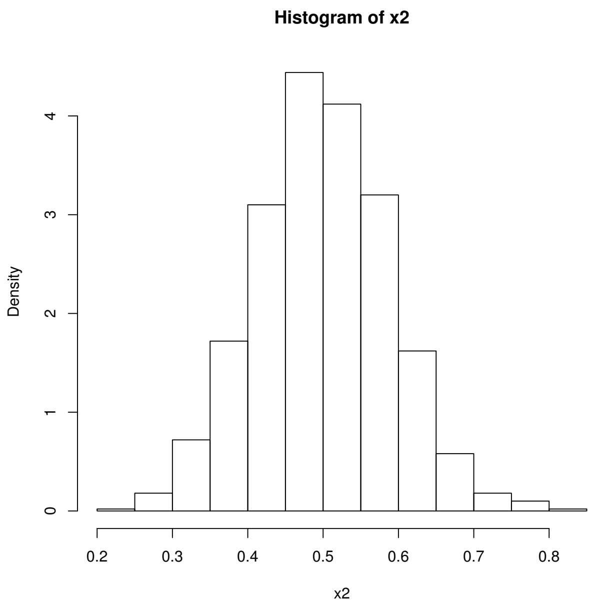
$$E(Y) = \mu, \quad V(Y) = \frac{\sigma^2}{m}$$

(説明) $Y = S/m$ であるから,

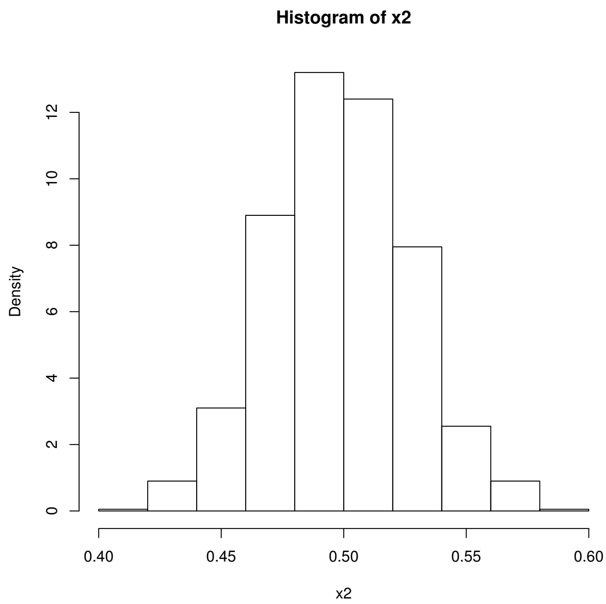
$$V(Y) = V\left(\frac{S}{m}\right) = \frac{1}{m^2} V(S) = \frac{\sigma^2}{m}$$

- $m \to \infty$ とすれば，$Y \to \mu$ に「収束」する．(収束することは，$V(Y) \to 0$ からも理解できる．)

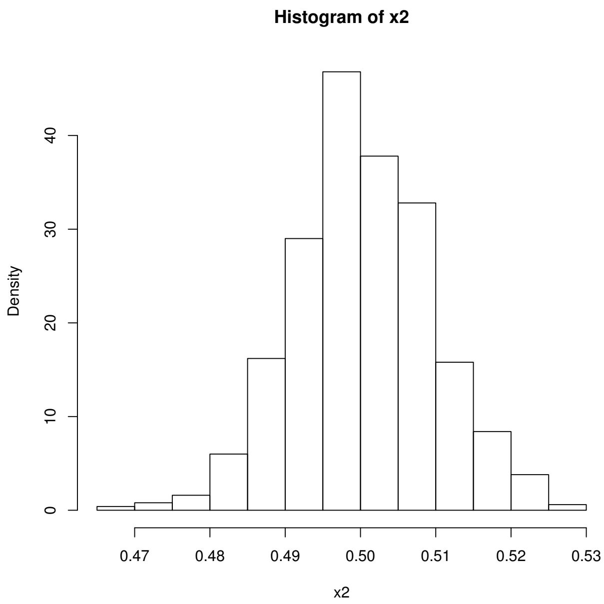
```
# run0013.R
# 一様乱数の和（大数の法則）

n <- 1000 # シミュレーションの回数
for(m in c(10,100,1000,10000)) {  # m=和を取る個数
  x1 <- matrix(runif(n*m),m,n)  # 一様乱数を要素とするサイズ m*n の行列
  x2 <- apply(x1,2,mean) # 各列の m 個の平均を取り，長さ n のベクトルを返す
  hist(x2,freq=F) # m 個の平均のヒストグラム
  dev.copy2eps(file=paste("run0013-h",m,".eps",sep=""))
}
rm(n,m,x1,x2)
```



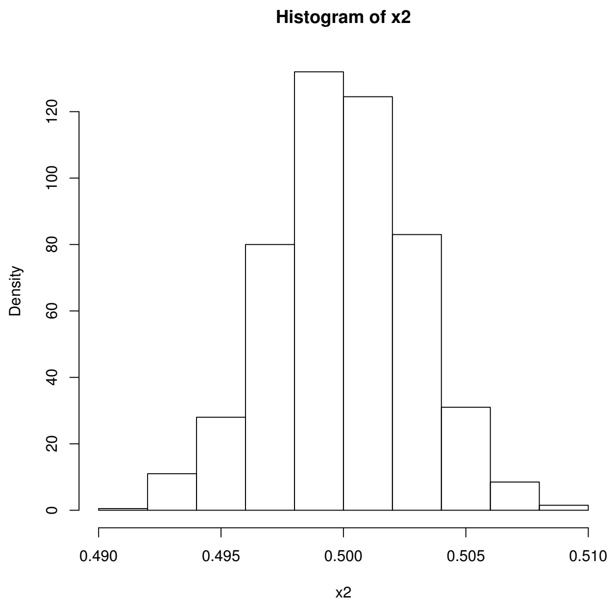

run0013-h10




run0013-h100




run0013-h1000




run0013-h10000

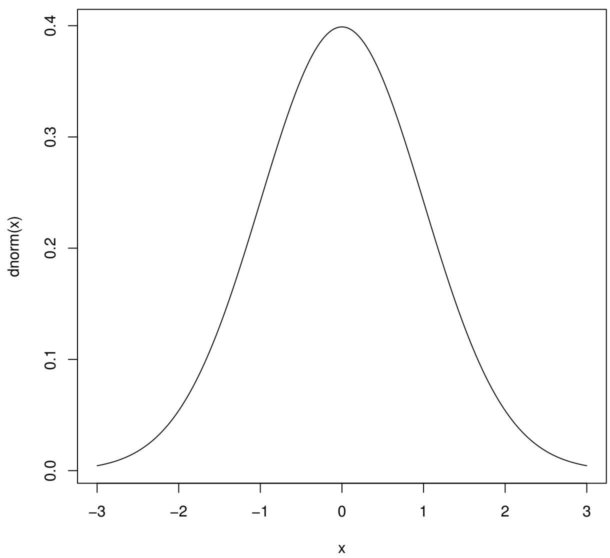
- $\mu = 0.5$ を中心にバラツキがある．横軸の数値を見ると分かるように，バラツキの大きさ（標準偏差=$\sqrt{\text{分散}}$）が，$n$ の増加とともに小さくなる．
- ヒストグラムの形状が釣鐘型になっていることに注意（後述の「中心極限定理」参照）．

# 3 正規分布

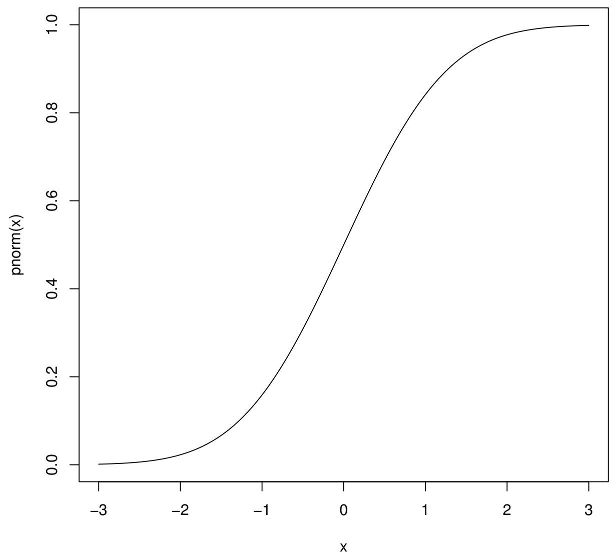
## 3.1 密度関数，分布関数

- 平均０，分散１の正規分布の密度関数と分布関数をプロットする．

```
# run0008.R
# 正規分布 N(0,1) 密度関数，分布関数

x <- seq(-3,3,length=300) # (-3,3) の区間に 300 個の点を取る
plot(x,dnorm(x),type="l") # 密度関数のプロット
dev.copy2eps(file="run0008-d.eps")
plot(x,pnorm(x),type="l") # 分布関数のプロット
dev.copy2eps(file="run0008-p.eps")
rm(x) # オブジェクトの削除
```




run0008-d




run0008-p

- 平均 $\mu$ と分散 $\sigma^2$ をパラメタとする正規分布を $N(\mu, \sigma^2)$ と書く.
- 平均 0，分散 1 の正規分布（標準正規分布）： $N(0, 1)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

$$F_X(x) = \int_{-\infty}^{x} f_X(s)\, ds$$

- $X \sim N(0, 1)$ から $Y = \sigma X + \mu$ と定義すれば，$Y \sim N(\mu, \sigma^2)$ である．一般に

$$X \sim N(\mu, \sigma^2) \quad \Rightarrow \quad aX + b \sim N(a\mu + b, a^2\sigma^2)$$

- 平均 $\mu$，分散 $\sigma^2$ (標準偏差 $\sigma$) の正規分布: $N(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

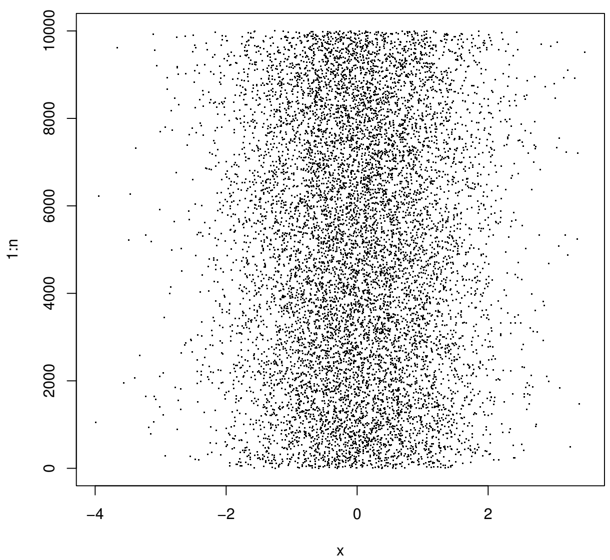
$$E(X) = a, \quad V(X) = b$$

## 3.2 擬似乱数の生成

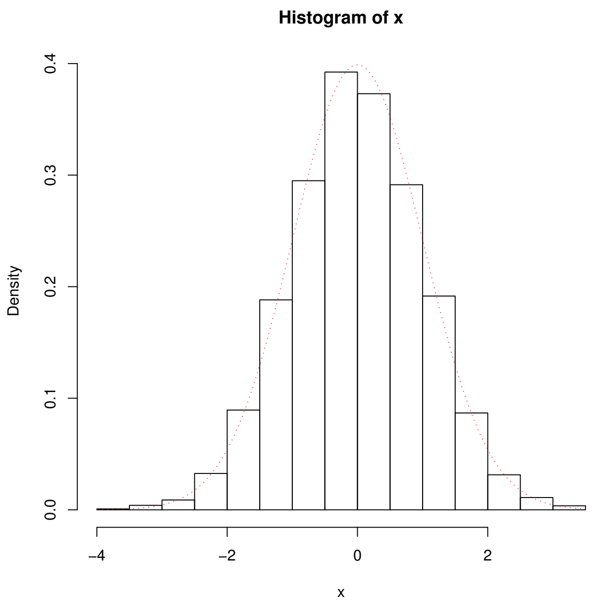
- 正規分布 $N(0, 1)$ に従う擬似乱数を多数（10000 個）生成し，３通りのプロットをする.

```
# run0015.R
# 正規分布 N(0,1) 擬似乱数の生成

n <- 10000 # シミュレーション回数
x <- rnorm(n) # (平均=0, 標準偏差=1) の正規乱数を n 個生成
plot(x,1:n,pch=".") # 横軸に乱数, 縦軸に番号 1,...,n をプロット
```

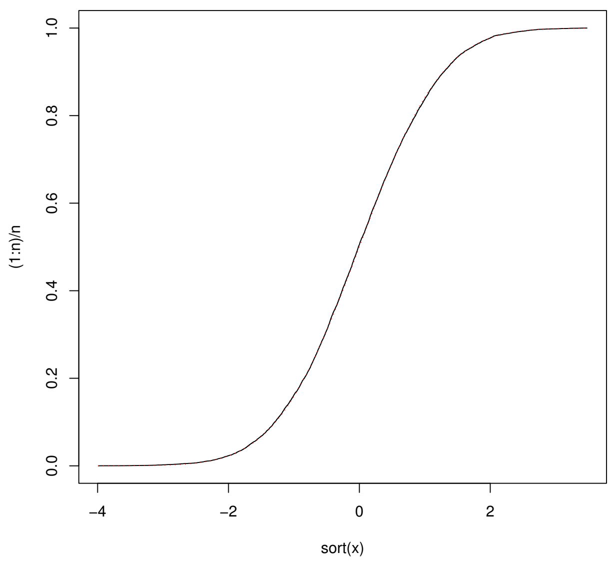
```
dev.copy2eps(file="run0015-r.eps")

hist(x,freq=F) # ヒストグラム，縦軸＝確率密度
x0 <- seq(min(x),max(x),length=300) # x の範囲に等間隔で 300 個の値を取る
lines(x0,dnorm(x0),col=2,lty=3) # 確率密度関数の曲線を赤の点線で描く
dev.copy2eps(file="run0015-h.eps")
plot(sort(x),(1:n)/n,type="l") # 横軸に乱数をソートしたものをプロット
lines(x0,pnorm(x0),col=2,lty=3) # 分布関数
dev.copy2eps(file="run0015-s.eps")

rm(n,x,x0) # オブジェクトの削除
```




run0015-r




run0015-h




run0015-s

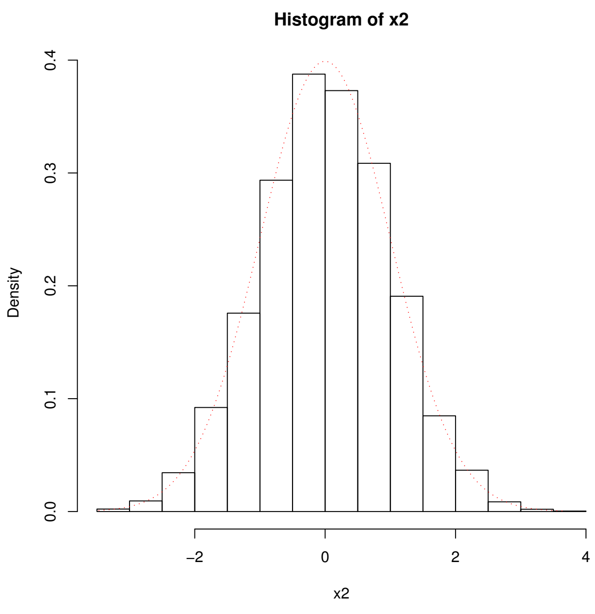
## 3.3 確率変数の変換

- 一様分布に従う確率変数を変換して，正規分布に従う確率変数を与える．

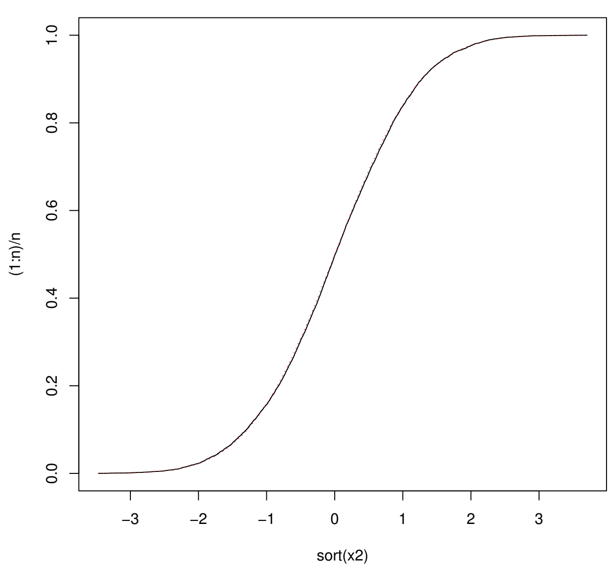
```
# run0009.R
# 一様分布から正規分布へ変換

n <- 10000
x1 <- runif(n) # (0,1)区間の一様乱数を10000個生成
x2 <- qnorm(x1) # 正規分布の分布関数の逆関数
hist(x2,freq=F) # ヒストグラム(縦軸は確率密度)
x0 <- seq(min(x2),max(x2),length=300) # x2の範囲に等間隔で300個の値を取る
lines(x0,dnorm(x0),col=2,lty=3) # 確率密度関数の曲線を赤の点線で描く
dev.copy2eps(file="run0009-h.eps")
plot(sort(x2),(1:n)/n,type="l") # 横軸に乱数をソートしたものをプロット
lines(x0,pnorm(x0),col=2,lty=3) # 分布関数
dev.copy2eps(file="run0009-s.eps")
rm(n,x0,x1,x2)
```

- 一様分布 $U(0,1)$ に従う確率変数 $U$ (`runif` で生成)
- 正規分布 $N(0,1)$ の分布関数 $F_X(x)$ (`pnorm` 関数)
- その逆関数 $F_X^{-1}(u)$ (`qnorm` 関数)
- $X = F_X^{-1}(U)$ は正規分布 $N(0,1)$ に従う

注意：上記の考え方は，他の確率分布にも適用できる．ただし，R における `rnorm` の実装では，計算コストの大きな `qnorm` 関数を使わずに，もっと計算効率のよく精度も高い rejection sampling 法の一種が用いられている (`src/nmath/snorm.c` 参照)． rejection sampling 法については，ここでは説明しない．




run0009-h




run0009-s

## 3.4 中心極限定理

- 平均 $\mu$, 分散 $\sigma^2$ の確率変数の列（各 $X_i$ は互いに独立）

$$X_1, X_2, \ldots, X_m$$

$$E(X_i) = \mu, \quad V(X_i) = \sigma^2$$

- $m$ 個の平均

$$Y = \frac{X_1 + \cdots + X_m}{m}$$

$$E(Y) = \mu, \quad V(Y) = \frac{\sigma^2}{m}$$

- $m$ が十分に大きければ，近似的に $Y$ は正規分布 $N(\mu, \sigma^2/m)$ に従う.
- $Z$ を平均 0，分散 1 にする（「標準化」という）

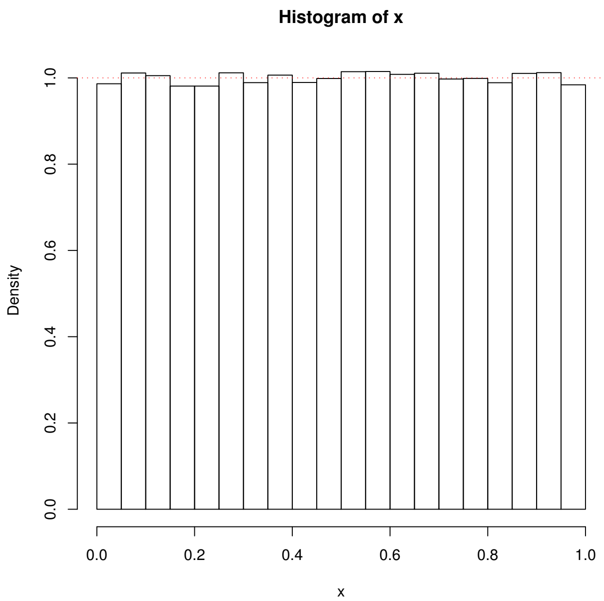
$$Z = \frac{\sqrt{m}(Y - \mu)}{\sigma}$$

$$E(Z) = 0, \quad V(Z) = 1$$

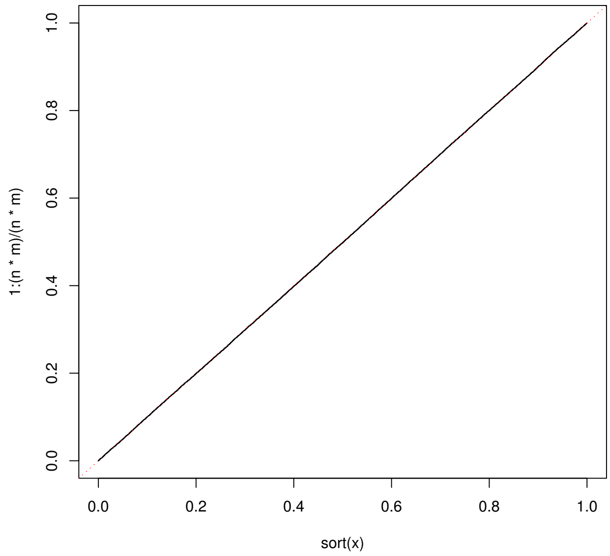
- 中心極限定理：$m \to \infty$ の極限で $Z$ の従う分布は正規分布 $N(0, 1)$ に収束する.

```
# run0010.R
# 一様乱数の和（中心極限定理）

n <- 10000 # シミュレーションの回数
m <- 10 # 平均の計算において和を取る個数

x <- matrix(runif(n*m),m,n)  # 一様乱数を要素とするサイズ m*n の行列
y <- apply(x,2,mean) # 各列の m 個の平均を取り，長さ n のベクトルを返す

hist(x,freq=F) # 一様乱数のヒストグラムを確認
abline(h=1,col=2,lty=3)
dev.copy2eps(file="run0010-hx.eps")
plot(sort(x),1:(n*m)/(n*m),type="l") # 経験分布関数を確認
abline(0,1,col=2,lty=3)
dev.copy2eps(file="run0010-sx.eps")

hist(y,freq=F) # m 個の平均のヒストグラム
heikin <- 1/2; hensa <- sqrt((1/12)/m) # Y の平均と標準偏差の計算
y0 <- seq(min(y),max(y),length=300) # y の範囲に等間隔に 300 個の値
lines(y0,dnorm(y0,mean=heikin,sd=hensa),col=2,lty=3) # 正規分布の密度関数
dev.copy2eps(file="run0010-h.eps")
plot(sort(y),(1:n)/n,type="l") # 経験分布関数
```

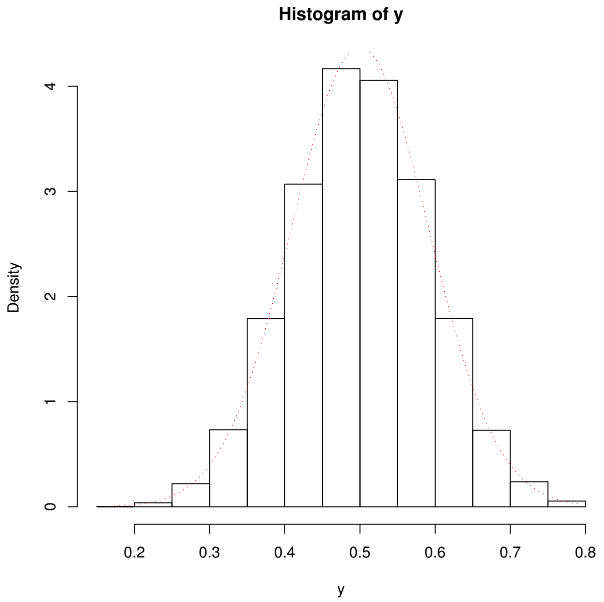
```
lines(y0,pnorm(y0,mean=heikin,sd=hensa),col=2,lty=3) # 分布関数
dev.copy2eps(file="run0010-s.eps")

rm(n,m,x,y,y0,heikin,hensa)
```



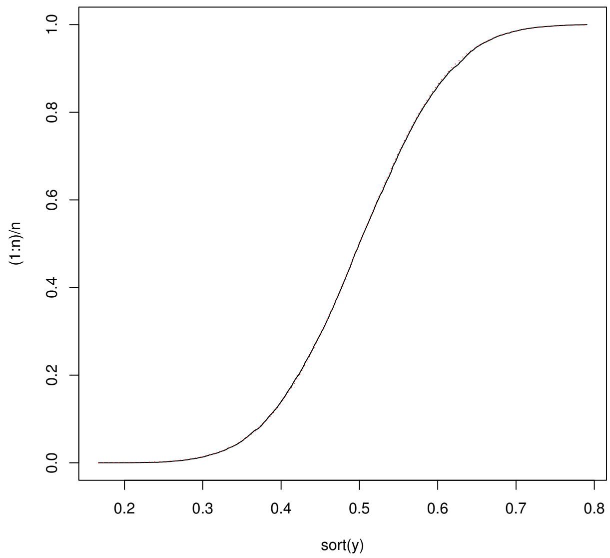
run0010-hx



run0010-sx



run0010-h



run0010-s

# 4　ガンマ分布

## 4.1　密度関数，分布関数，擬似乱数

- ガンマ分布のパラメタは shape=$a$ と scale=$s$ の 2 個．これを $Ga(a,s)$ と書く．shape は形状を変える．scale は倍率を変える．

- shape $= a$，scale=1 のガンマ分布：$Ga(a,1)$

$$f_X(x) = \frac{1}{\Gamma(a)} x^{a-1} e^{-x}, \quad x > 0$$

ただしガンマ関数は

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

- $X \sim Ga(a,1)$ から $Y = sX$ とすれば，$Y \sim Ga(a,s)$ となる．一般に

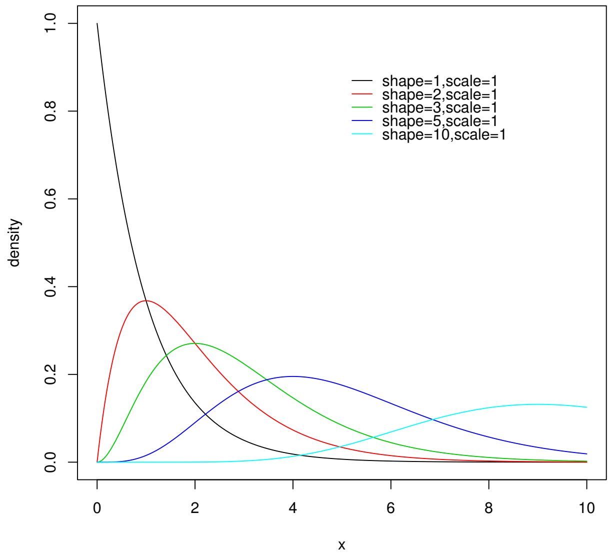
$$X \sim Ga(a,s) \quad \Rightarrow \quad kX \sim Ga(a,ks)$$

- shape $a$，scale $s$ のガンマ分布: $Ga(a,s)$

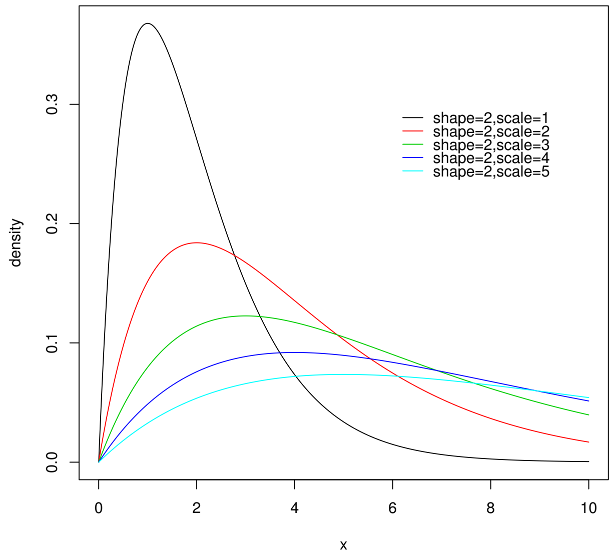
$$f_X(x) = \frac{1}{s\Gamma(a)} \left(\frac{x}{s}\right)^{a-1} \exp\left(-\frac{x}{s}\right), \quad x > 0$$

$$E(X) = as, \quad V(X) = as^2$$

```
# run0011.R
# ガンマ分布 Ga(a,s) a=shape,s=scale

x <- seq(0,10,length=500) # (0,10) の区間に 500 個の点を取る

i <- c(1,2,3,5,10) # shape を変化させる
j <- 1 # scale 固定
y <- sapply(i,function(a) dgamma(x,shape=a,scale=j)) # サイズ 500*5 の行列
matplot(x,y,type="l",lty=1,col=1:5,ylab="density") # まとめてプロット
legend(5,0.9,paste("shape=",i,",scale=",j,sep=""),
       col=1:5,lty=1,bty="n") # 凡例
dev.copy2eps(file="run0011-d1.eps")

i <- 2 # shape 固定
j <- c(1,2,3,4,5) # scale を変化させる
y <- sapply(j,function(s) dgamma(x,shape=i,scale=s))
matplot(x,y,type="l",lty=1,col=1:5,ylab="density")
legend(6,0.3,paste("shape=",i,",scale=",j,sep=""),col=1:5,lty=1,bty="n")
dev.copy2eps(file="run0011-d2.eps")

rm(i,j,x,y)
```




run0011-d1




run0011-d2

## 4.2 指数分布，カイ二乗分布

- 指数分布もカイ二乗分布も，ガンマ分布の特殊な場合
- $Ga(1,s)$ つまり，shape $a=1$ と固定し，scale=$s$ とおけば，平均 $s$ の指数分布

$$f_X(x)=\frac{1}{s}\exp\left(-\frac{x}{s}\right),\quad x>0$$

- $Ga(n/2,2)$ つまり，shape $a=m/2$ とし，scale=2 と固定すれば，自由度 $m$ のカイ二乗分布．ただし，奇数の $m$ に対して，ガンマ関数は

$$\Gamma(m/2)=\frac{(m-2)!!\sqrt{\pi}}{2^{(m-1)/2}},\quad (m-2)!!=(m-2)\times(m-4)\times\cdots\times 1$$

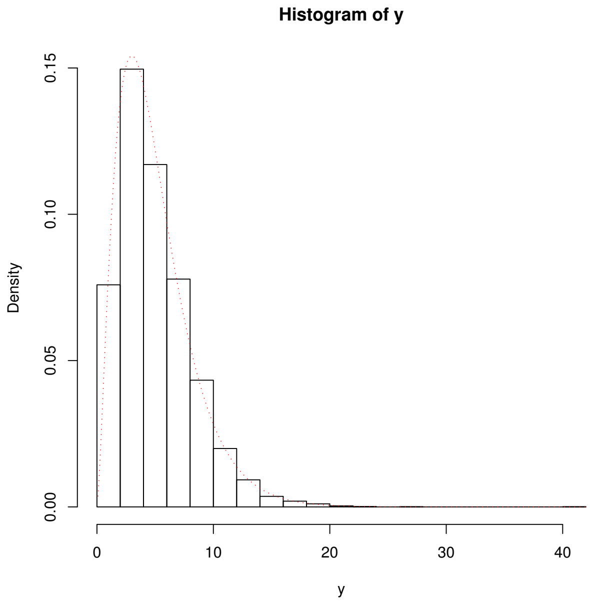
$$\Gamma(\tfrac{1}{2})=\sqrt{\pi},\ \Gamma(\tfrac{3}{2})=\tfrac{1}{2}\sqrt{\pi},\ \Gamma(\tfrac{5}{2})=\tfrac{3}{4}\sqrt{\pi}$$

- 標準正規分布に従う $m$ 個の確率変数の 2 乗和は，自由度 $m$ のカイ二乗分布に従う．

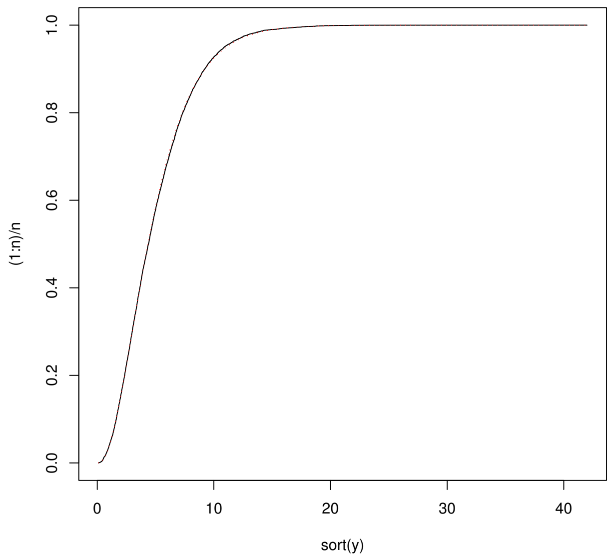
$$X_1,\ldots,X_m\sim N(0,1)\quad\Rightarrow\quad X_1^2+\cdots+X_m^2\sim\chi_m^2$$

```
# run0018.R
# 正規確率変数の二乗和 （カイ二乗分布）

n <- 10000 # シミュレーションの回数
m <- 5 # 和の計算において和を取る個数

x <- matrix(rnorm(n*m),m,n)  # N(0,1) を要素とするサイズ m*n の行列
y <- apply(x,2,function(v) sum(v^2)) # 各列の 2 乗和を計算
```

```
hist(y,freq=F) # m個の平均のヒストグラム
y0 <- seq(min(y),max(y),length=300) # yの範囲に等間隔に300個の値
lines(y0,dchisq(y0,df=m),col=2,lty=3) # カイ二乗分布の密度関数
dev.copy2eps(file="run0018-h.eps")
plot(sort(y),(1:n)/n,type="l") # 経験分布関数
lines(y0,pchisq(y0,df=m),col=2,lty=3) # 分布関数
dev.copy2eps(file="run0018-s.eps")

rm(n,m,x,y,y0)
```




run0018-h




run0018-s

## 4.3 リパラメトリゼーション

- パラメタ $(a, s)$ を変数変換する．平均 $\mu$ と分散 $\sigma^2$(またはその平方根 $\sigma$) を新たにパラメタとする．パラメタ $(\mu, \sigma)$ を用いて，ガンマ分布を表現する．

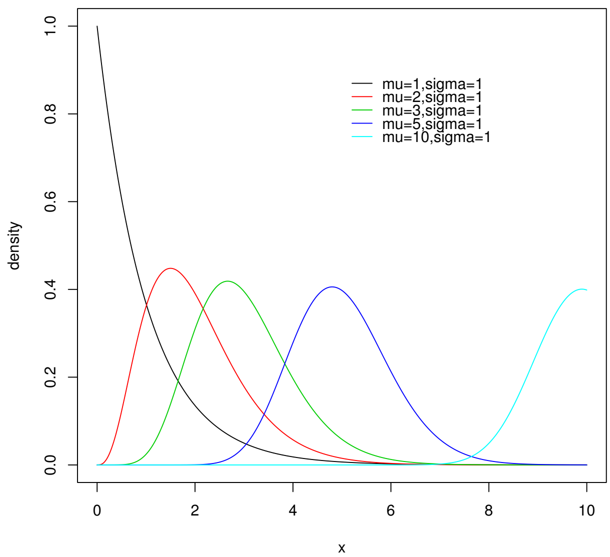
$$\mu = as, \quad \sigma^2 = as^2 \quad \Leftrightarrow \quad a = \mu^2/\sigma^2, \quad s = \sigma^2/\mu$$

$$E(X) = \mu, \quad V(X) = \sigma^2$$

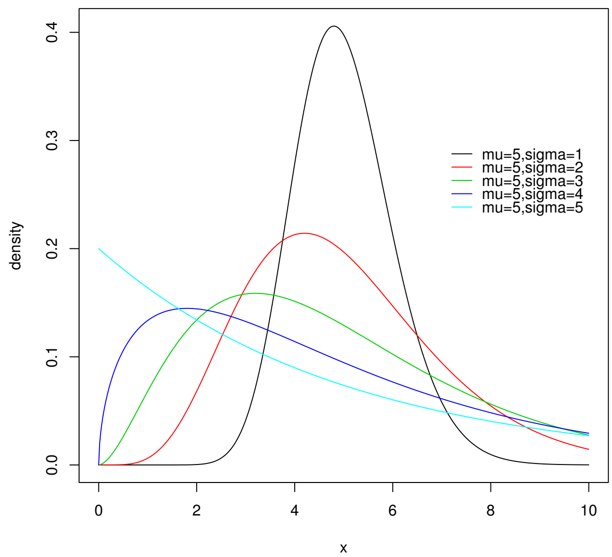
```
# run0012.R
# ガンマ分布 Ga(a,s) muとsigmaで表現

gshape <- function(mu,sigma) mu^2/sigma^2 # shapeの計算
gscale <- function(mu,sigma) sigma^2/mu # scaleの計算

x <- seq(0,10,length=500) # (0,10)の区間に500個の点を取る
```

```
i <- c(1,2,3,5,10) # mu を変化させる
j <- 1 # sigma 固定
print(rbind(gshape(i,j),gscale(i,j))) # shape と scale の表示
y <- sapply(i,function(a) dgamma(x,shape=gshape(a,j),scale=gscale(a,j)))
matplot(x,y,type="l",lty=1,col=1:5,ylab="density")
legend(5,0.9,paste("mu=",i,",sigma=",j,sep=""),col=1:5,lty=1,bty="n")
dev.copy2eps(file="run0012-d1.eps")

i <- 5 # mu 固定
j <- c(1,2,3,4,5) # sigma を変化させる
print(rbind(gshape(i,j),gscale(i,j))) # shape と scale の表示
y <- sapply(j,function(a) dgamma(x,shape=gshape(i,a),scale=gscale(i,a)))
matplot(x,y,type="l",lty=1,col=1:5,ylab="density")
legend(7,0.3,paste("mu=",i,",sigma=",j,sep=""),col=1:5,lty=1,bty="n")
dev.copy2eps(file="run0012-d2.eps")

rm(i,j,x,y)
```

```
> source("run0012.R")
     [,1] [,2]      [,3] [,4]  [,5]
[1,]    1  4.0 9.0000000 25.0 100.0
[2,]    1  0.5 0.3333333  0.2   0.1
     [,1] [,2]     [,3]   [,4] [,5]
[1,] 25.0 6.25 2.777778 1.5625    1
[2,]  0.2 0.80 1.800000 3.2000    5
```



run0012-d1



run0012-d2

### 4.4 再生性

- ガンマ分布に従う確率変数の系列（shape は変化，scale は固定)

$$X_1 \sim Ga(a_1, s),\, X_2 \sim Ga(a_2, s), \ldots,\, X_m \sim Ga(a_m, s)$$

- 和もガンマ分布

$$Y = X_1 + X_2 + \ldots + X_m$$

$$Y \sim Ga(a_1 + a_2 + \cdots + a_m, s)$$

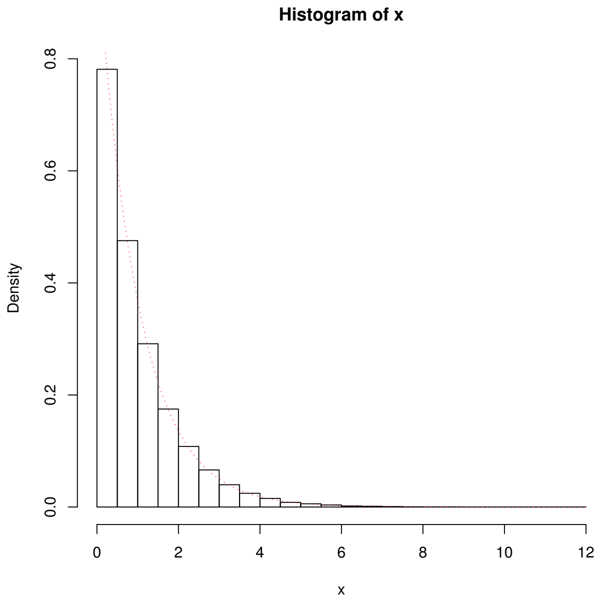
- したがって，平均もガンマ分布

$$Y = \frac{X_1 + X_2 + \ldots + X_m}{m}$$

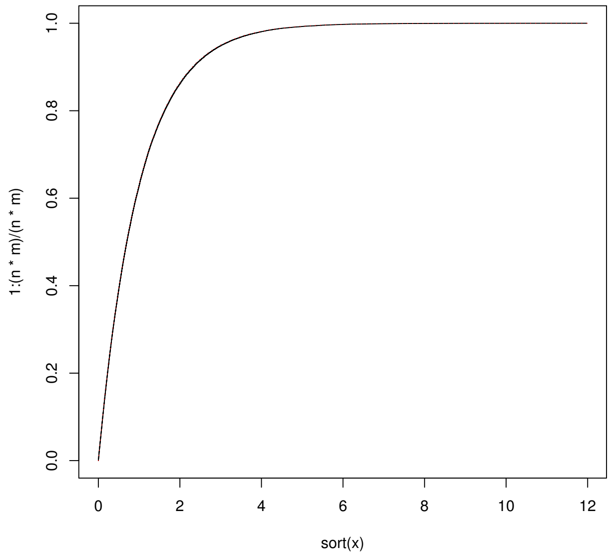
$$Y \sim Ga(a_1 + a_2 + \cdots + a_m, s/m)$$

```
# run0016.R
# ガンマ分布 Ga(a,1) 擬似乱数の生成と再生性

n <- 10000 # シミュレーションの回数
m <- 10 # 平均の計算において和を取る個数
a <- 1 # shape パラメタ

x <- matrix(rgamma(n*m,shape=a,scale=1),
            m,n)  # Ga(a,1) を要素とするサイズ m*n の行列
y <- apply(x,2,mean) # 各列の m 個の平均を取り，長さ n のベクトルを返す


hist(x,freq=F) # Ga(a,1) のヒストグラムを確認
x0 <- seq(min(x),max(x),length=300) # x の範囲に等間隔に 300 個の値
lines(x0,dgamma(x0,shape=a,scale=1),col=2,lty=3)
dev.copy2eps(file="run0016-hx.eps")
plot(sort(x),1:(n*m)/(n*m),type="l") # 経験分布関数を確認
lines(x0,pgamma(x0,shape=a,scale=1),col=2,lty=3)
dev.copy2eps(file="run0016-sx.eps")


hist(y,freq=F) # m 個の平均のヒストグラム
y0 <- seq(min(y),max(y),length=300) # y の範囲に等間隔に 300 個の値
lines(y0,dgamma(y0,shape=a*m,scale=1/m),col=2,lty=3) # Ga(a*m,1/m)
dev.copy2eps(file="run0016-h.eps")
plot(sort(y),(1:n)/n,type="l") # 横軸に乱数をソートしたものをプロット
lines(y0,pgamma(y0,shape=a*m,scale=1/m),col=2,lty=3) # Ga(a*m,1/m)
```

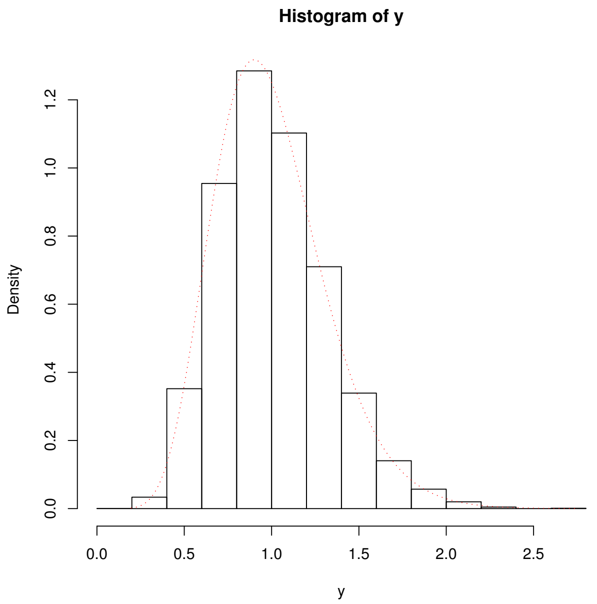
```
dev.copy2eps(file="run0016-s.eps")

rm(n,m,a,x,y,x0,y0)
```



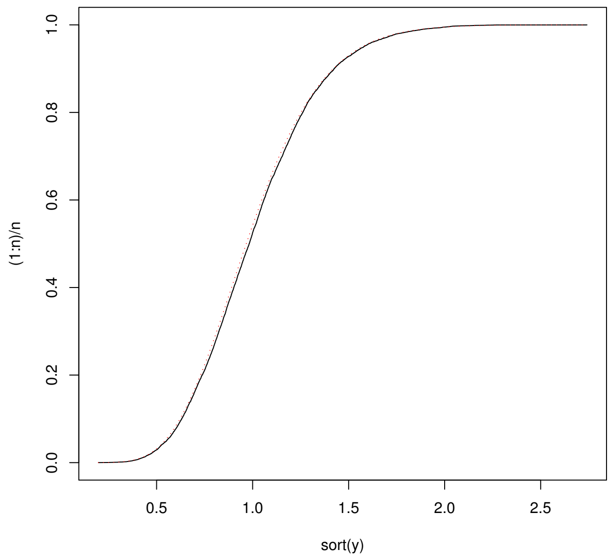

run0016-hx




run0016-sx




run0016-h




run0016-s

# 5　課題

## 5.1　課題 2-1

中心極限定理の収束の様子を以下のように確かめよ．run0010.R において，$m$ を変化させ，$m$ がどのくらい大きければ正規分布へ十分に収束したといえるか判断せよ (ヒストグラム run0010-h.eps および経験分布関数 run0010-s.eps のプロットを参考に判断する)．ただし run0010.R において run0010-hx.eps と run0010-sx をプロットする部分は削除してよい．

### 5.2　課題 2-2

上の課題 2-1 で一様分布 $U(0,1)$ をガンマ分布 $Ga(1,1)$(すなわち平均 1 の指数分布) に置き換えた場合，$m$ がどのくらい大きい必要があるか判断せよ．ただし，`run0010.R` において，`heikin` と `hensa` の定義を適切に変更しないと，赤い点線による極限値のプロットが正常に行われないので注意すること．

### 5.3　課題 2-3*

上の課題 2-2 で，`run0010.R` を利用せず，擬似乱数生成をまったく行わないで解答せよ．(ガンマ分布の再生性に注意する)．

### 5.4　課題 2-4*

課題 2-1 において一様分布 $U(0,1)$ に従う確率変数を，確率 $0<p<1$ で値 1，確率 $1-p$ で値 0 となる (離散値) 確率変数に置き換える．$m$ がどのくらい大きい必要があるかを判断せよ．ただし，$p=0.5$ とする．もし，$p=1/m$ のように $m$ に依存して $p$ を変化させたらどうなるか考察せよ．