R で学ぶ確率統計

学生番号

氏名 手塚康久

2021/06/15

レポート

R にも ggplot2 にも不慣れで，確率分布の理解も “very poor”，復習メモしかできませんでした.

二項分布

起こる結果が 1 か 0 の 2 つしかないベルヌーイ試行では，その成功確率は 0≦p≦1 とすると，p(X=1)=P, P(X=0)=1-P

ベルヌーイ試行を n 回繰り返し k 回成功する確率 p は

dbinom(k,n,p)

例 試行回数 20，成功確率 1/3 のとき，

dbinom(3,20,1/3)= 0.0428538

3 回以下の確率

pbinom(3,20,1/3)=0.604

30%になる回数は

qbinom(0.3,20,1/3)=6

期待値 np，分散 np(1-p)

期待値は 20/3=6.666667,

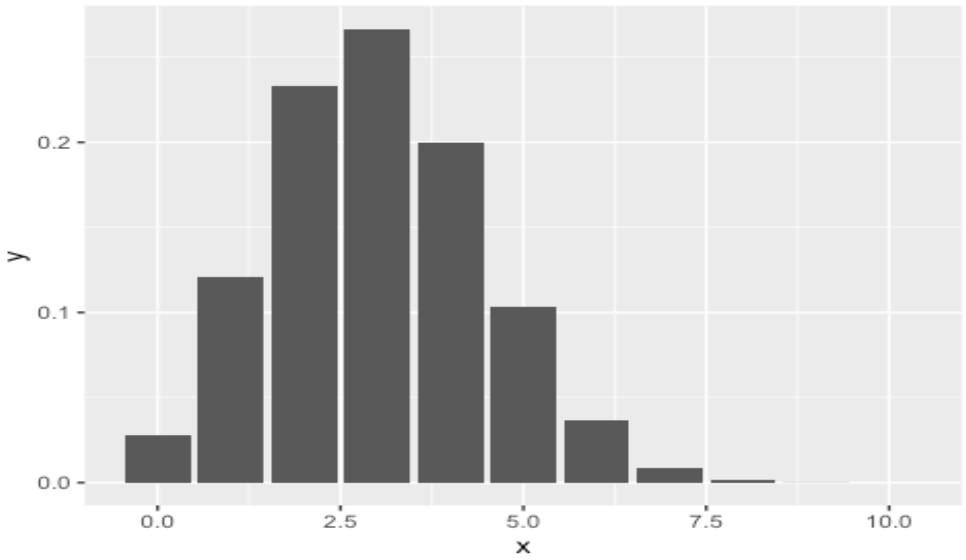
分散は 20/3×17/20=5.666667

図表：二項分布（離散型 p=0.3,0.5）

10 回のベルヌーイ試行，確率を 0.3（図表 1）から 0.5（図表 2）に上げると右に移り，0.3 の時は 3 の周辺の，0.5 の時は 5 の周辺に度数は集まる．試行を増やせば右に移り，正規分布に近づくことが推測される.

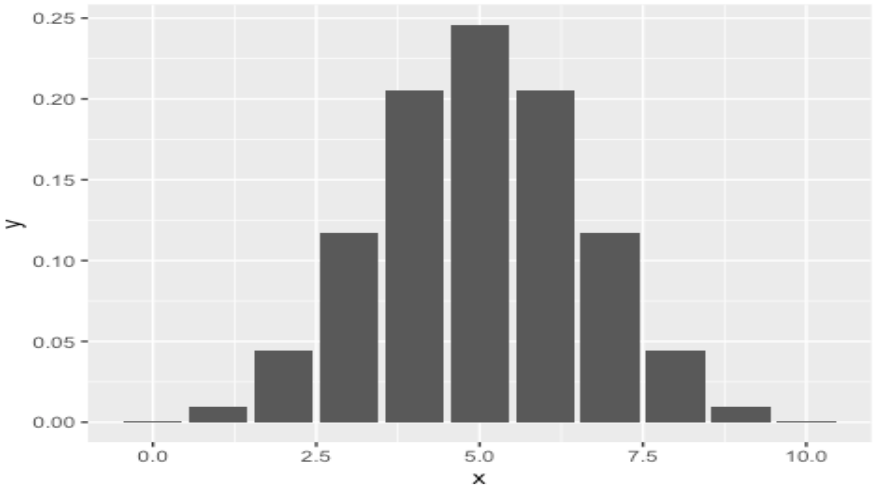
```
f2 <-tibble(x=0:10,y= dbinom(seq(0,10,1),10,0.3))
     ggplot(f2) + geom_bar(aes(x=x,y=y),stat="identity")
```

図表 1




```
f2 <-tibble(x=0:10,y= dbinom(seq(0,10,1),10,0.5))
      ggplot(f2) + geom_bar(aes(x=x,y=y),stat="identity")
```

図表 2



ポアソン分布

二項分布が基にあり，n が十分大きく，p が非常に小さい時「np＝一定」と考えられ，λとおく．期待値＝λ．n，p が分からなくてもλが分かっていれば求められる．

例 1 ：

1 年間自動車 1 万台に 1 件の割合で事故が発生する（1 日 2 件起きないとする），2 万台につき 3 件起こる確率は？

λ=20000*0.0001=2，dpois(3,2)=0.18 0

3 件以下の確率は ppois(3,2)=0.857

例 2 ：

ある工場で不良品が月 2 個出ている，月 1 個しかでない確率は

dpois(1,2) 0.2706706

月 2 個 dpois(2,2) 0.2706706

月 0 個 dpois(0,2) 0.1353353

月 2 個以下 ppois(2,2) 0.6766764

月 3 個以上出る確率

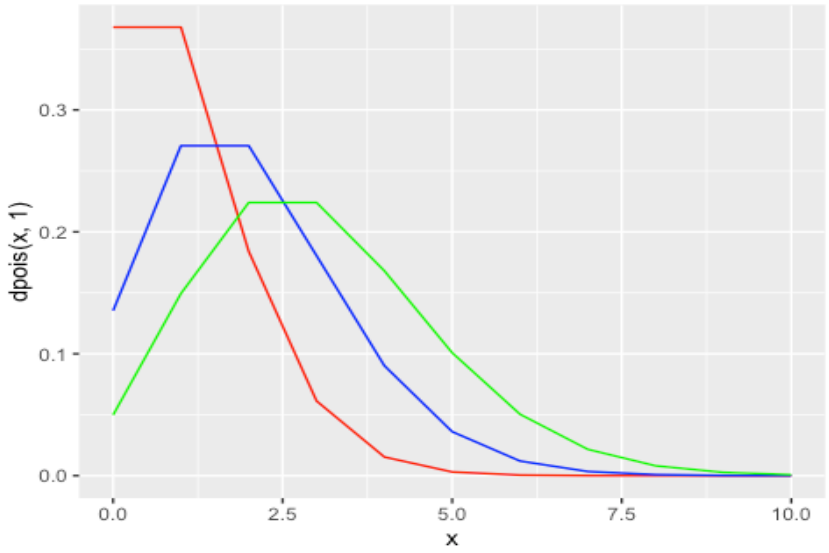
1- ppois(2,2),

期待値も分散も λ

```
ggplot(data.frame(x=c(0:10)), aes(x)) + geom_line(aes(y=dpois(x,1)),co
lour="red")+
geom_line(aes(y=dpois(x,2)),colour="blue") +
geom_line(aes(y=dpois(x,3)),colour="green")
```

λ が大きくなれば，右に移動して正規分布に近づくことが推測される．

図表 3



幾何分布（離散型,p=0.3,0.5）

ベルヌーイ試行で 1 回目が起こる確率の分布．

例：

確率 0.3 のくじで，3 回目ではじめて当てる確率

dgeon(2,0.3)=0.11574

1 回目から 10 回までに最初に当たる確率は図表 4.

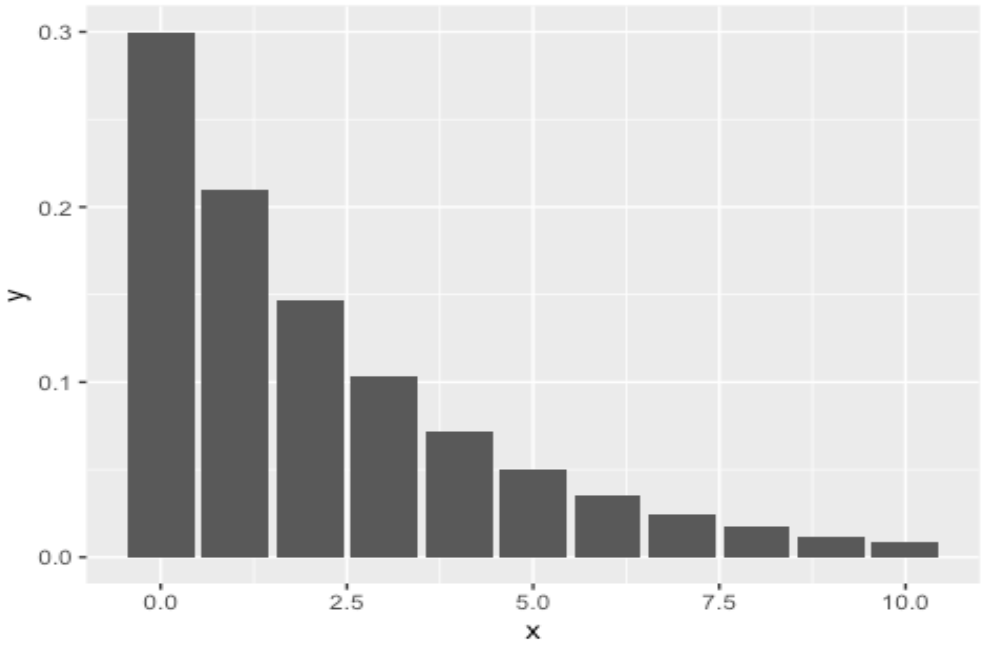
```
library(tidyverse)
## ─ Attaching packages ─────────────────── tidyverse 1.3.0 ─

## ✓ ggplot2 3.3.3     ✓ purrr   0.3.4
## ✓ tibble  3.0.4     ✓ dplyr   1.0.2
## ✓ tidyr   1.1.2     ✓ stringr 1.4.0
## ✓ readr   1.4.0     ✓ forcats 0.5.0

## ─ Conflicts ──────────────────── tidyverse_conflicts() ─
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

     f1 <-tibble(x=0:10,y=dgeom(seq(0,10,1),0.3))
      ggplot(f1) + geom_bar(aes(x=x,y=y),stat="identity")
```

図表 4



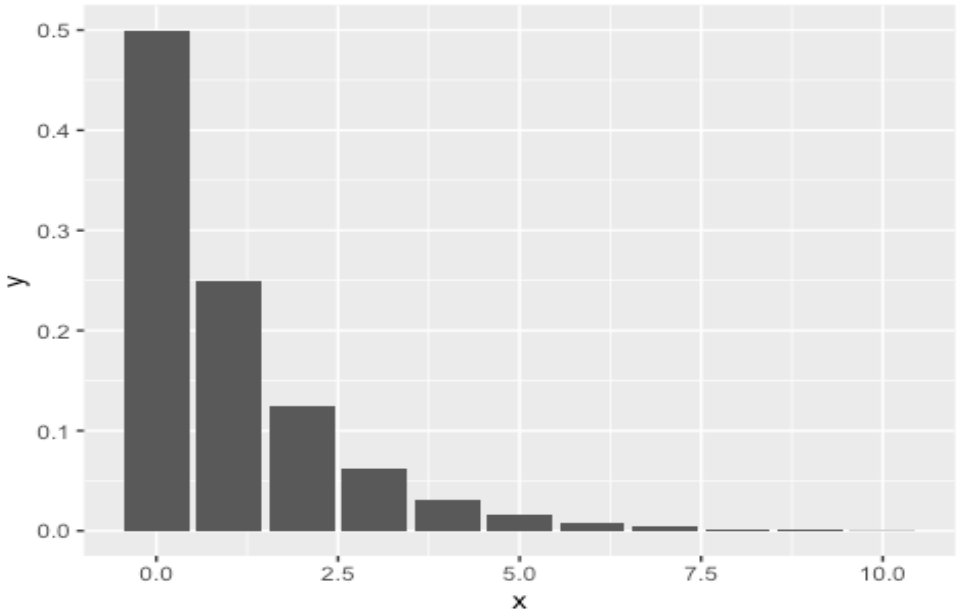
確率 0.5 なら

dgeom(2,0.5)=0.125

図表 4 は 1 回目から 10 回目に初めて当たる確率のヒストグラム，5 回までに当たる確率は pgeom(4,0.5)=0.969，　6 回目以降に当たるのは約 3%.

```
<ScaleContinuousPosition>
      f1 <-tibble(x=0:10,y=dgeom(seq(0,10,1),0.5))
      ggplot(f1) + geom_bar(aes(x=x,y=y),stat="identity")
```

図表 5



幾何分布，係数を 0.3 から 0.5 にすると，当然 1 回目に成功する可能性は 0.5 となる.

正規分布

平均を中心に左右対称の形をとる．平均を 50，標準偏差を 10（偏差値）を例に

pnorm(50,50,10)=0.5 平均点が真ん中に位置

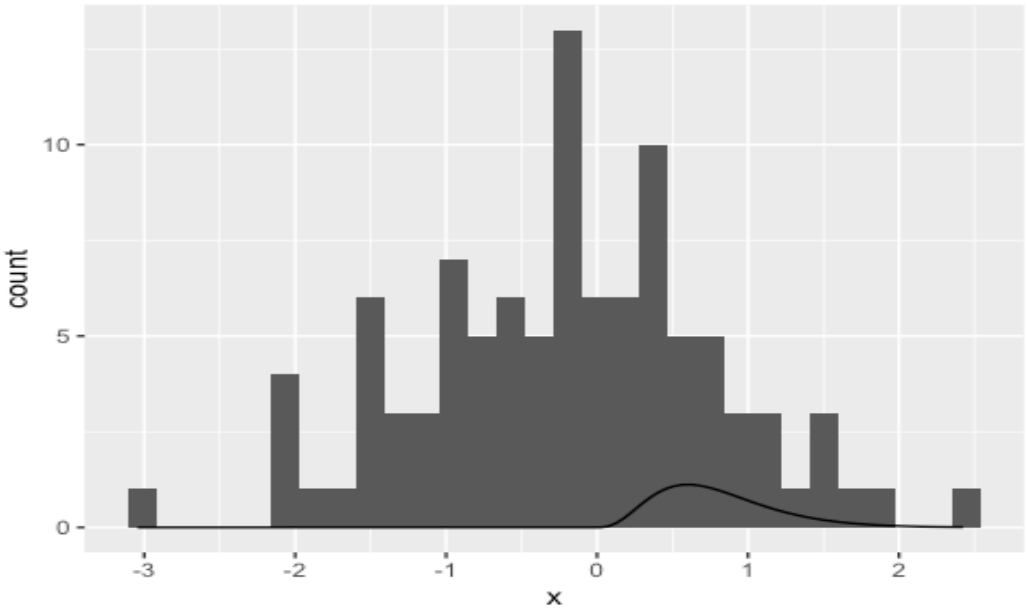
pnorm(60,50,10)=0.8413 偏差値 60（60 点）以下の人は全体の 84%

1-pnorm(70,50,10)=0.228,偏差値 70 以上の人は全体の 2.3%.

図表 6 が乱数を 100 まで発生させて shape=4,rate=5,bins=30)

```
ggplot(data.frame(x =rnorm(100,0,1)) )+
  geom_histogram(aes(x=x),bins=30)+
        stat_function(fun =dgamma,args=c(shape=4,rate=5))
```

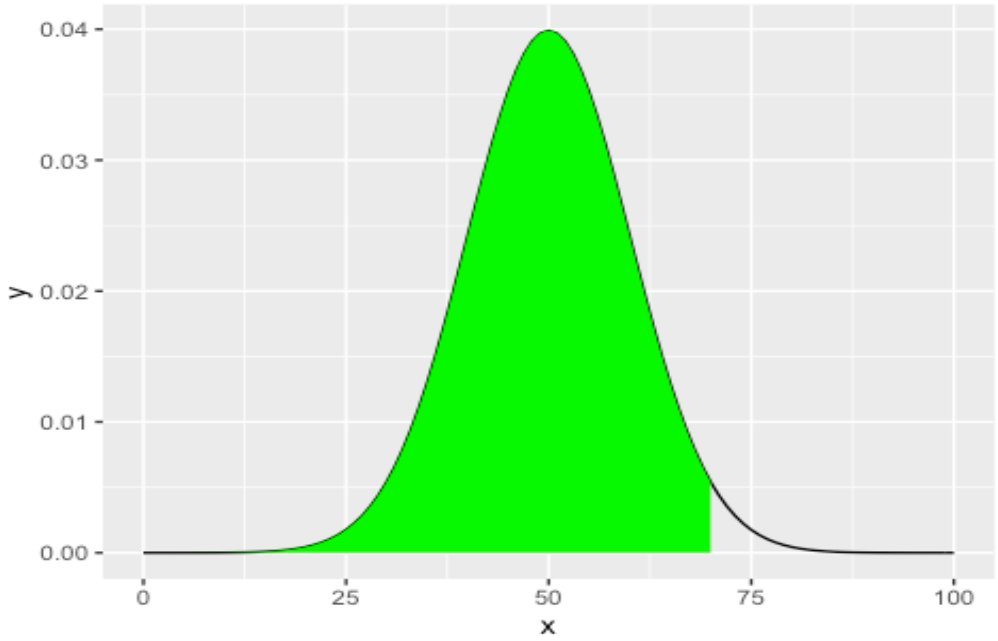
図表 6




```
mean <- 50; sd <- 10; x1 <- 70
y1 <- dnorm(x1,mean=mean,sd=sd)
ggplot(data=data.frame(x=c(0,100)), aes(x=x)) +
 stat_function(fun=dnorm, args=c(mean=mean,sd=sd) ) +
geom_area(stat = "function", fun = dnorm, args=c(mean=mean,sd=sd),fill
 = "green", xlim = c(0, x1))
```

図表 7 は 70 以下の 97.7%を緑色に配色.

図表 7



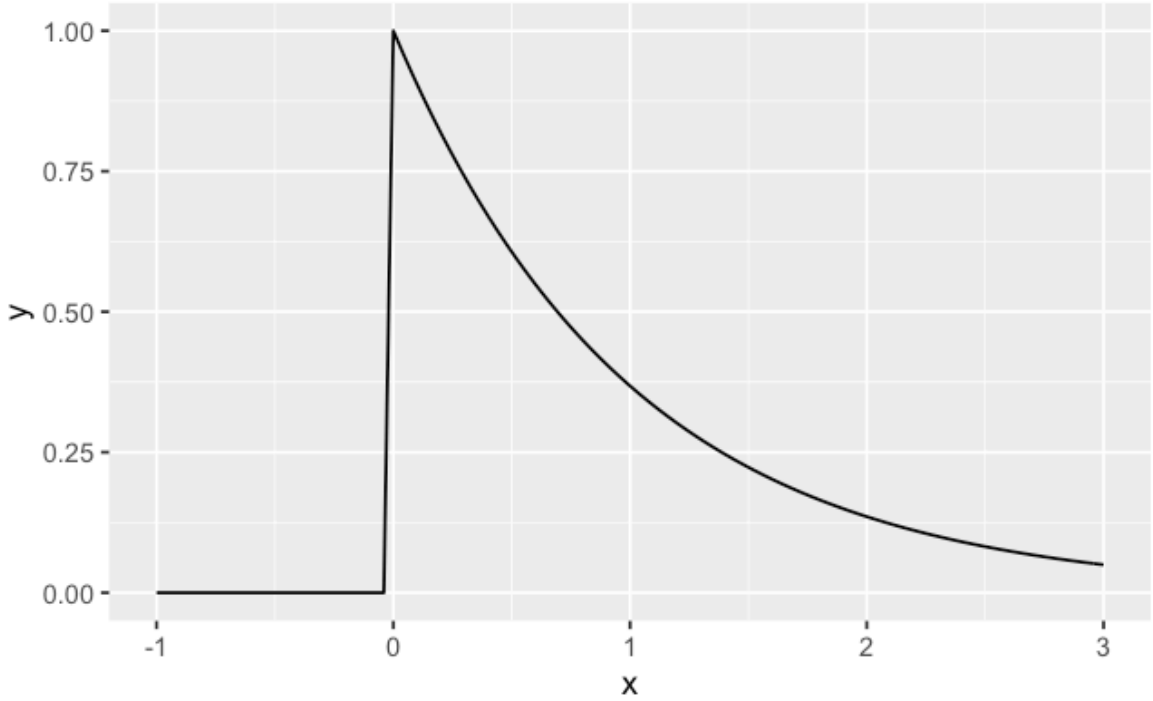
指数分布（連続型)

機械が故障してから次に故障するまでなど次に何かが起こるまでの期間が従う分布.
ある期間に平均してλ回起こるとすると，期待値は 1/λ，分散は 1/λ＾2

Dexp(0,1,0)=1, dexp(1,1,0)=0.36787,dexp(2,1,0)=0.13533

```
ggplot(data.frame(x =c(-1,3)),aes(x=x) )+
      stat_function(fun =dexp,args=c(rate=1))
```

図表 8



ガンマ分布

例： 待ち時間 λ=2 に従う指数分布，1 分に 3 人に出会う確率

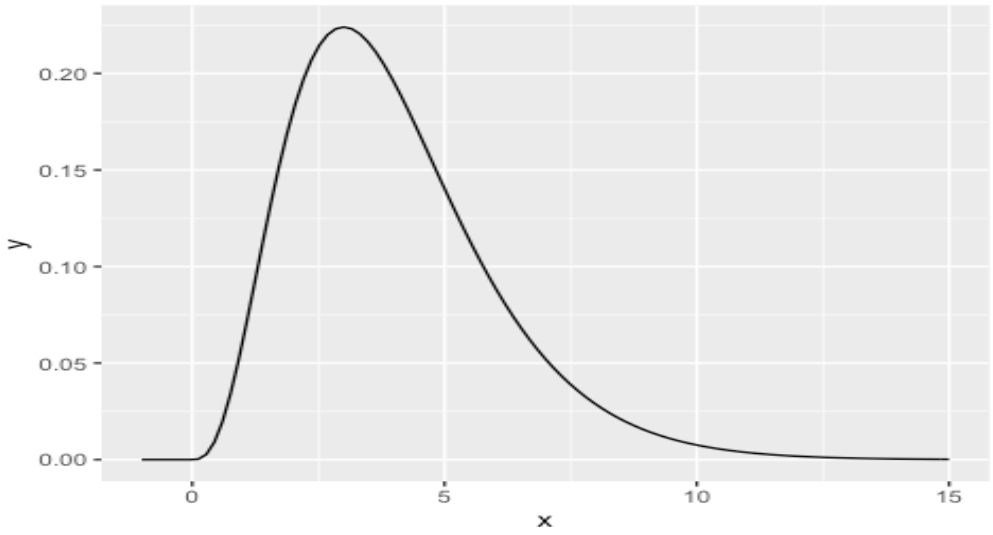
dgamma(1,rate=2,shape=3)=0.5413411

pgamma(1,rate=2,shape=3)=0.3233236

ガンマ分布(連続型，shape=4,2,1,rate:1)

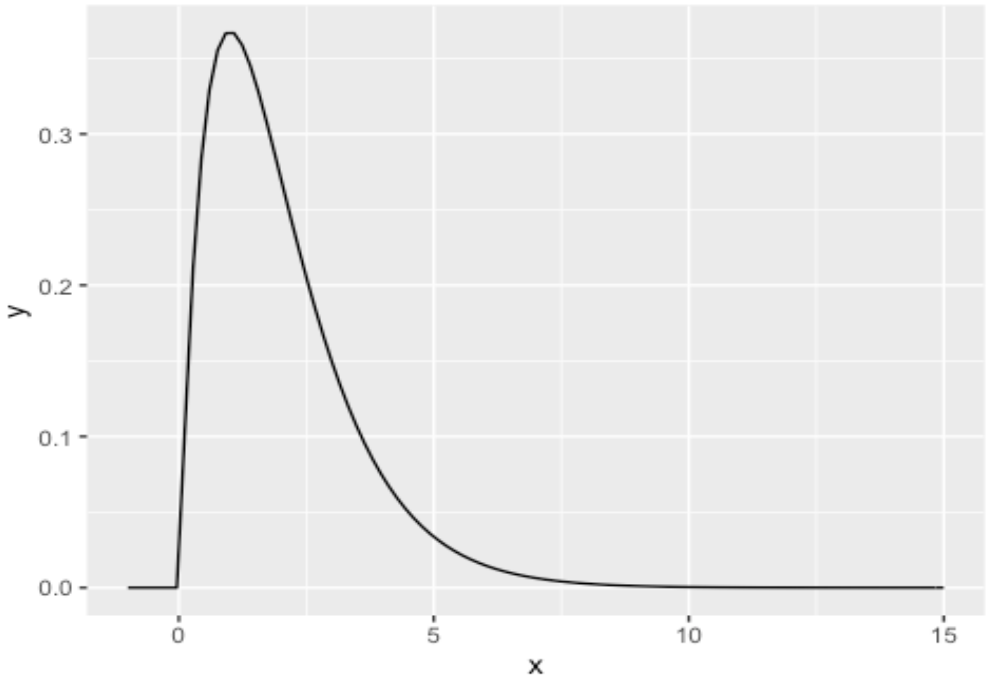
```
ggplot(data.frame(x =c(-1,15)),aes(x=x) )+
        stat_function(fun =dgamma,args =c(shape=4,rate=1))
```

図表 9



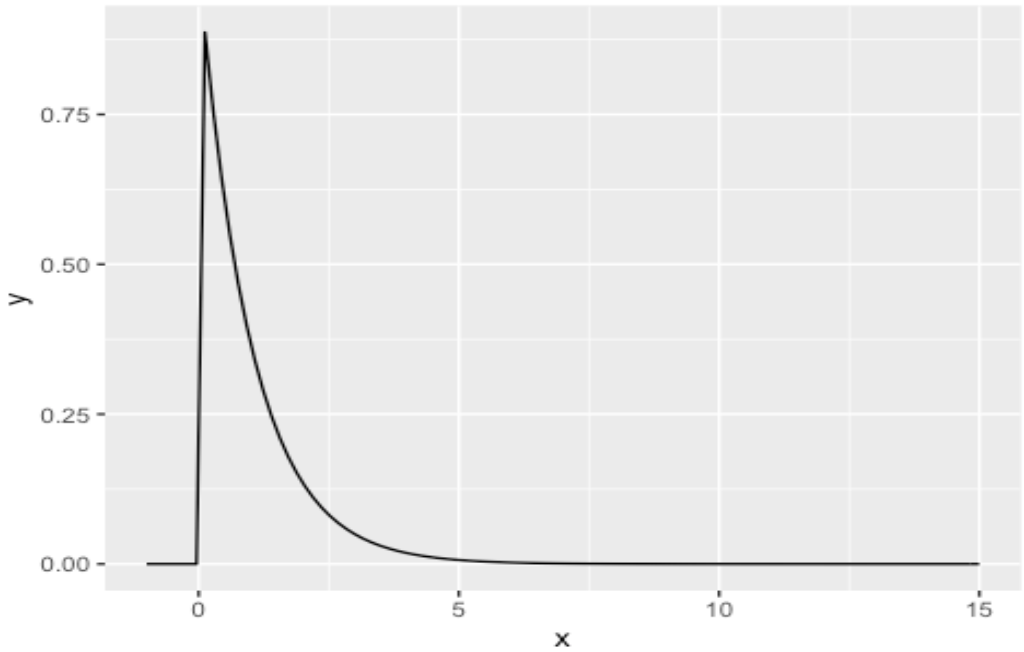
図表 10

```
ggplot(data.frame(x =c(-1,15)),aes(x=x) )+
        stat_function(fun =dgamma,args =c(shape=2,rate=1))
```




```
ggplot(data.frame(x =c(-1,15)),aes(x=x) )+
        stat_function(fun =dgamma,args =c(shape=1,rate=1))
```

図表 11



shape のパラメータが大きくなるとなだらかな正規分布に近づく.

結論

このメモを書くことにより，とりあえず，恐怖心はなくなった．しかし，解決すべき課題が多いことも認識した.

参考

BellCurve 統計 web　https://bellcurve.jp/