Calculus for Statistical Genetics 統計遺伝学のための微分積分学

Introduction まえがき

Before we start this module, we check the role of this module in the the Joint Degree (JD) course .

このモジュールを始める前に、ジョイントディグリー(JD)専攻におけるこのモジュールの役割を確認する。

Read Data science in JD course https://github.com/ryamada22/JD_lectures/blob/6a58e1a078b449f41b937ec3b3a77aab2302f0d9/cells/Data_science_in_JD.ipynb (https://github.com/ryamada22/JD_lectures/blob/6a58e1a078b449f41b937ec3b3a77aab2302f0d9/cells/Data_science_in_JD.ipynb) .

JD専攻におけるデータサイエンスを読め。

Basics of Calculs 微分積分学の基礎 Read the first paragraph of Calculus document of Wikipedia https://en.wikipedia.org/wiki/Calculus (https://en.wikipedia.org/wiki/Calculus) .

Wikipediaの微分積分学 の記事 https://en.wikipedia.org/wiki/Calculus (https://en.wikipedia.org/wiki/Calculus) の最初のパラグラフを読め。

Assignment 課題

Write a short description sentence on calculus, derivative, and integral, respectively.

微分積分学、微分、積分のそれぞれを説明する単文を書け。

Expected value and basics of differentiation and integration 期待値と微分・積分の基礎

Sample average 標本平均

Assignment 課題

Generate random value series of positive integers and calculate their sample average.

正の整数の乱数列を発生させ、その標本平均を計算せよ。

```
n <- 10
x <- sample(1:3,n,replace=TRUE)
x
```

```
##  [1] 2 1 3 2 3 2 1 3 2 3
```

```
mean(x)
```

```
## [1] 2.2
```

```
1/n * sum(x)
```

```
## [1] 2.2
```

平均 Average

$$m = \frac{1}{n}\sum_{i=1}^{n} x_i$$

```
n <- 10
x <- sample(1:3,n,replace=TRUE)
x
```

```
##  [1] 2 3 3 2 1 3 1 2 1 3
```

```
mean(x)
```

```
## [1] 2.1
```

```
1/n * sum(x)
```

```
## [1] 2.1
```

重み付き平均 Weighted average

$$m_w = \sum_{j=1}^{k} v_j \times Pr(j)$$

```
x
```

```
##  [1] 2 3 3 2 1 3 1 2 1 3
```

```
tabulate(x)
```

```
## [1] 3 3 4
```

```
w <- tabulate(x)/n
w
```

```
## [1] 0.3 0.3 0.4
```

```
v <- sort(unique(x))
v
```

```
## [1] 1 2 3
```

```
mw <- sum(v * w)
mw
```

```
## [1] 2.1
```

期待値 Expected value

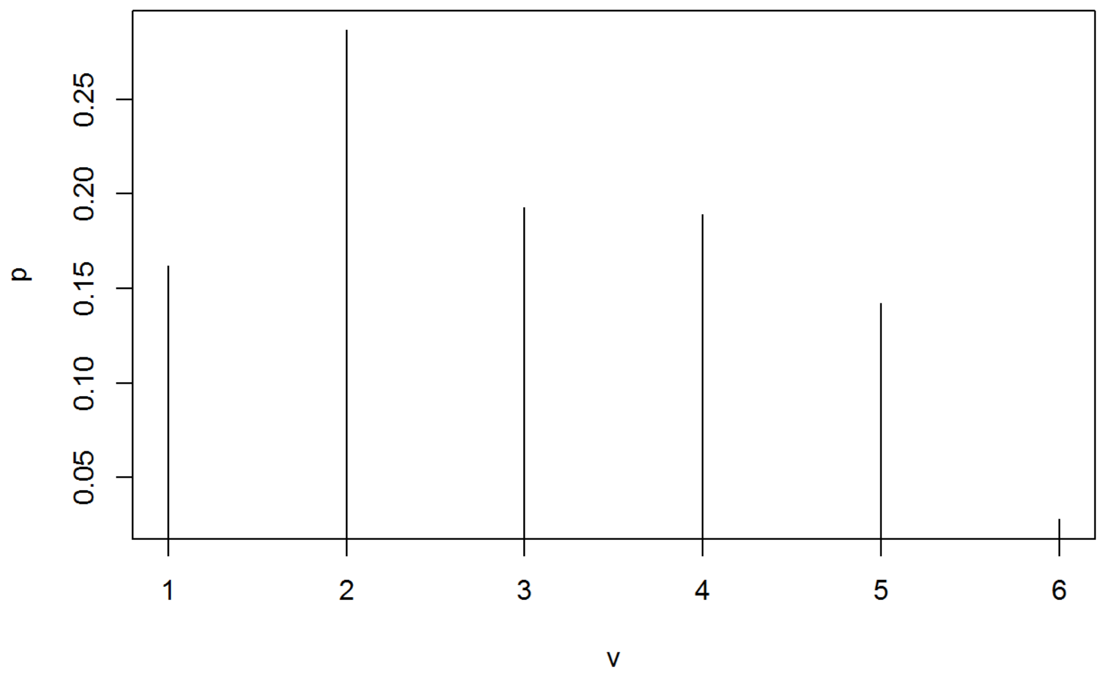
サイコロの目の数の期待値 Expected value of dice

```
v <- 1:6
p <- rep(1/6,6)
sum(v*p)
```

```
## [1] 3.5
```

出来の悪いサイコロの期待値 Dice in bad condition

```
v <- 1:6
p <- rep(1/6,6) + rnorm(6) * 0.1
p <- p/sum(p)
plot(v,p,type="h")
```




```
p
```

```
## [1] 0.16206033 0.28649117 0.19268056 0.18894725 0.14195306 0.02786763
```

```
sum(p)
```

```
## [1] 1
```

```
sum(v*p)
```

```
## [1] 2.945844
```

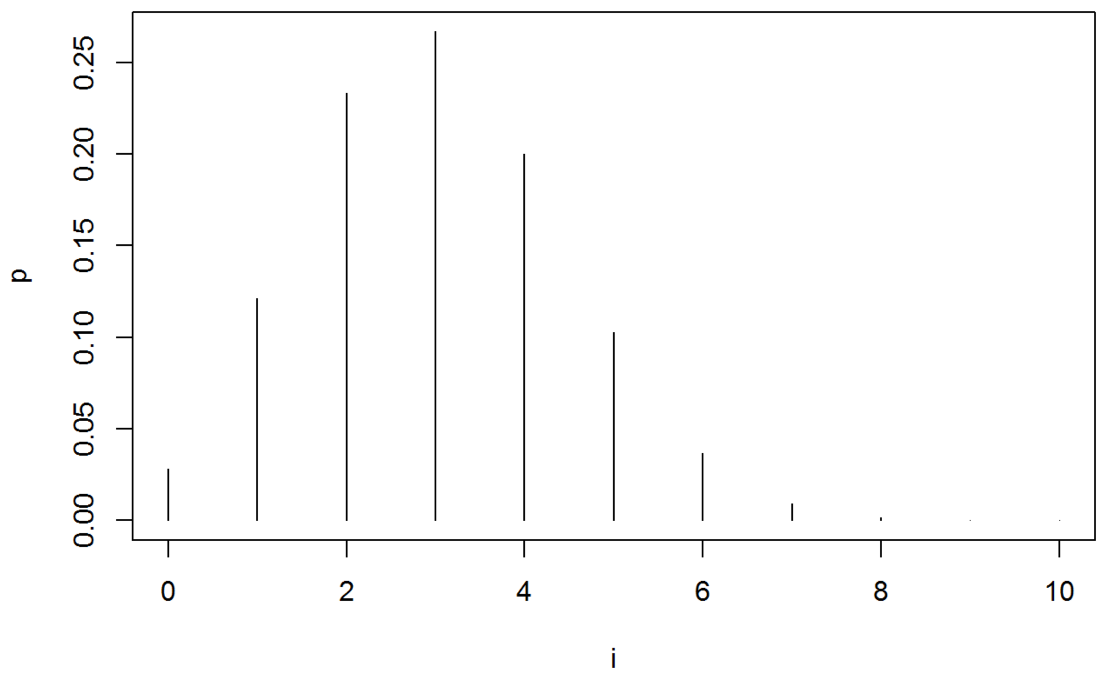
二項分布の期待値は np Expected value of binomial distribution : np

$$(p+(1-p))^n = 1^n = 1 = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^i$$

```
p0 <- 0.3
n <- 10
i <- 0:n
i.inv <- i[(n+1):1]
choose(n,i)
```

```
## [1]   1  10  45 120 210 252 210 120  45  10   1
```

```
p <- choose(n,i) * p0^i * (1-p0)^i.inv
plot(i,p,type="h")
```




```
sum(p)
```

```
## [1] 1
```

```
sum(p*i)
```

```
## [1] 3
```

```
n * p0
```

```
## [1] 3
```

ベータ分布

ベータ分布の正規化 Normalization of beta distribution

成功・失敗が、n回とm回だったとき、成功率がpである尤度は　n successes and m failures. Likelihood of success rate p is proportional to ;

$$p^n(1-p)^m$$

に比例する。

With $h(n,m)$ below,

$$h(n,m) = \int_0^1 p^n(1-p)^m dp$$

とおけば、the following equaion follows;

$$\int_0^1 \frac{1}{h(n,m)} p^n(1-p)^m dp = 1$$

となるから

The following is the likelihood function of success rate p when n successes and m failures.

$$\frac{1}{h(n,m)} p^n(1-p)^m$$

が成功n回、失敗m回のときの成功率pの尤度関数。

$p^n(1-p)^m$が関数の形を決め、$h(n,m)$は積分が1となるように正規化しているので、$h(n,m)$によって(尤度)関数を正規化する、と言う。

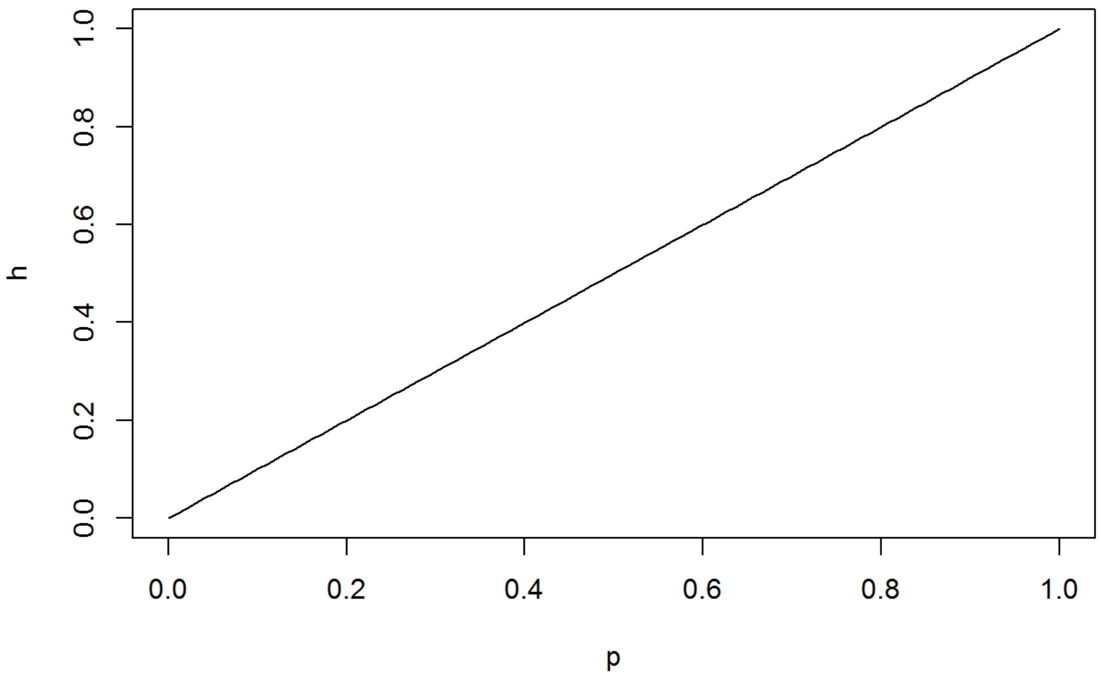
$p^n(1-p)^m$ determines its shape and $h(n,m)$ normalizes its integration to 1.

$h(n,m) = \int_0^1 p^n(1-p)^m dp$を計算してみる Calculation of $h(n,m)$

n=1, m=0

$$h(1,0) = \int_0^1 pdp$$

```
p <- seq(from=0,to=1,length=100)
h <- p
plot(p,h,type="l")
```



$h(1,0)$は面積として計算できる。 Area of $h(1,0)$ is given geometrically.

$$\frac{1}{2}\times 1\times 1=\frac{1}{2}$$

積分するなら Integration;

$$\frac{d}{dx}x^2=2x$$
$$\frac{d}{dx}\frac{1}{2}x^2=x$$

を使って、

$$x^2+C=\int 2xdx$$
$$\frac{1}{2}(x^2+C)=\int xdx$$

から、

$$\int_0^1 pdp=[\frac{1}{2}x^2]_0^1=\frac{1}{2}(1^2-0^2)=\frac{1}{2}$$

となる。

結局、n=1,m=0のときの尤度関数は The likelihood function when n=1 and m=0;

$$\frac{1}{h(1,0)}p^1(1-p)^0=\frac{1}{\frac{1}{2}}p=2p$$

ベータ分布の期待値 Expected value of beta distribution

$n=1,m=0$のときのベータ分布 When$n=1,m=0$, beta distribution is

$$2p$$

その期待値は Its expected value is

$$\int_0^1 (2p) \times pdp = \int_0^1 2p^2 dp = \frac{2}{3}[p^3]_0^1 = \frac{2}{3}$$

Exercise 1

Exercise 1-1

n=1, m=1の場合、二項観察の尤度関数は Likelihood function for binomial observation n=1 and m=1

$$\frac{1}{h(1,1)}p(1-p) = \frac{1}{h(1,1)}(p-p^2)$$

$$h(1,1) = \int_0^1 p - p^2 dp$$

を求めたい。

$f(x) = x^1(1-x)^1$ のグラフを描け。Draw $f(x) = x^1(1-x)^1$.

Exercise 1-2

[0,1]区間を、k等分してその小区間ごとの面積を近似的に計算し、その和を[0,1]の範囲の$p-p^2$の面積とみなすこととする。第i小区間の面積を、長方形の面積とみなして、計算し、kを、1,2,…,100と変化させ、その様子をプロットせよ。ただし、長方形は幅$\frac{1}{k}$、高さはその小区間の両端の$p-p^2$の値の平均値とせよ。

Divide the interval [0,1] into k evenly. Calculate subintervals' area approximately and sum them which is approcimation of the area under the curve. The area of the i-th subinterval should be considered a rectangle whose width is $/frac1k$ and its hight is the average of the hights of the both ends of the rectangle. Calculate and plot for k=1,2,…,100.

Exercise 1-3

$\frac{d}{dx}x^2 = 2x, \frac{d}{dx}x^3 = 3x^2$を使って$h(1,1)$を求め、近似で求めた値と比較せよ。

Integrate the function and compare the value with the approximation above.

Exercise 1-4

期待値を重み付き平均$ _0^1 p Pr(p) dp$ の積分を解くことで求めよ。 Answer its expected value by integrating $_0^1 p Pr(p) dp.

Exercise 1-5

n=2,m=3の場合の$p^n(1-p)^3$を展開し、n=1,m=1の場合と同様のことをせよ

Do the same for n=2 and m=3.

Exercise 1-6

指数分布の期待値は $\frac{1}{\lambda}$であると言う。このことを、離散的な計算をすることで確認せよ。

The expected value of exponential distribution is $\frac{1}{\lambda}$. Calculate its expected value discretely.

$$Pr(x) = \lambda \times e^{-\lambda x}$$

Exercise 1-6

微分積分の基礎技術 Basic skills of calculus

Go through the every item in the page https://en.wikibooks.org/wiki/Calculus/Differentiation/Basics_of_Differentiation/Solutions#Find_The_Derivative_By_Definition (https://en.wikibooks.org/wiki/Calculus/Differentiation/Basics_of_Differentiation/Solutions#Find_The_Derivative_By_Definition)