Calculus2 Calculus for Likelihood Function and Maximum Likelihood Estimation

ryamada

2017年1月22日

1 ベータ分布

1.1 尤度関数

(n,m)成否観察の尤度関数

$$L(p|n,m) = \frac{\Gamma(n+m+2)}{\Gamma(n+1)\Gamma(m+1)} p^n (1-p)^m$$

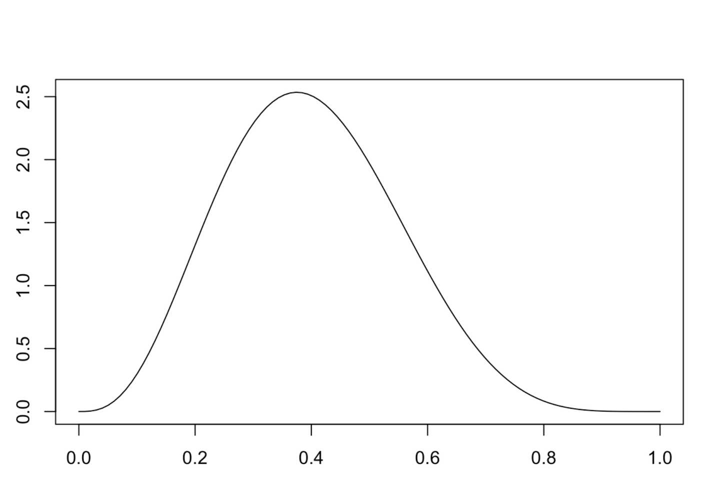
```
n <- 3
m <- 5
p <- 0.2
gamma(n+m+2)/(gamma(n+1)*gamma(m+1))*p^n*(1-p)^m
```

```
## [1] 1.321206
```

```
dbeta(p,n+1,m+1)
```

```
## [1] 1.321206
```

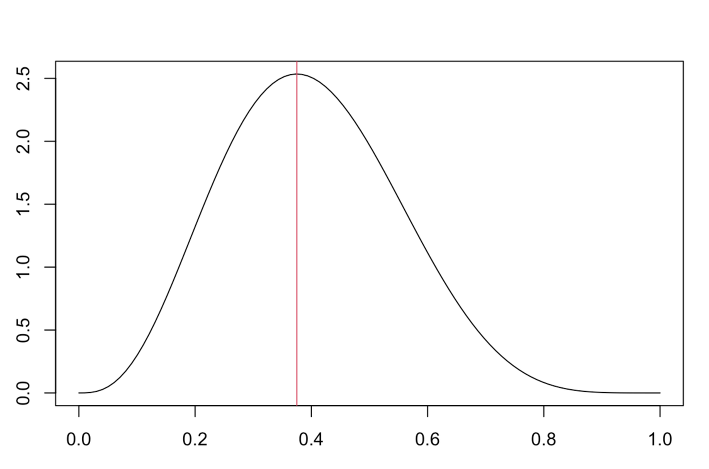
```
p <- seq(from=0,to=1,length=100)
d <- dbeta(p,n+1,m+1)
plot(p,d,type="l")
```



最尤推定値p_{MLE}は

$$p_{MLE} = \frac{n}{n+m}$$

```
plot(p,d,type="l")
abline(v=n/(n+m),col=2)
```



微分を使って求める。

$$L(p|n,m) = \frac{\Gamma(n+m+2)}{\Gamma(n+1)\Gamma(m+1)} p^n (1-p)^m$$

$$\frac{d}{dp} L(p_{MLE}|n,m) = 0$$

$$\begin{aligned} \frac{d}{dp} L(p|n,m) &= \frac{d}{dp} \frac{\Gamma(n+m+2)}{\Gamma(n+1)\Gamma(m+1)} p^n (1-p)^m \\ &= \frac{\Gamma(n+m+2)}{\Gamma(n+1)\Gamma(m+1)} \frac{d}{dp} (p^n (1-p)^m) \end{aligned}$$

$$\frac{d}{dx}(f(x) \times g(x)) = (\frac{d}{dx} f(x)) \times g(x) + f(x) \times (\frac{d}{dx} g(x))$$

$$\frac{d}{dp} p^n (1-p)^m = (\frac{d}{dp} p^n) \times (1-p)^m + p^n \times (\frac{d}{dp}(1-p)^m)$$

$$\frac{d}{dp} p^n = np^{n-1}$$

$$\frac{d}{dp}(1-p)^m = \frac{d}{dq} q^m \frac{dq}{dp}$$

$$q = (1-p)$$

$$\frac{d}{dq} q^m = mq^{m-1}$$

$$\frac{dq}{dp} = \frac{d}{dp}(1-p) = -1$$

$$\frac{d}{dp}(1-p)^m = mq^{m-1} \times (-1) = -m(1-p)^{m-1}$$

$$\begin{aligned} \frac{d}{dp} p^n (1-p)^m &= (\frac{d}{dp} p^n) \times (1-p)^m + p^n \times (\frac{d}{dp}(1-p)^m) \\ &= np^{n-1}(1-p)^m + p^n \times (-m(1-p)^{m-1}) \\ &= p^{n-1}(1-p)^{m-1}(n(1-p) - mp) \\ &= p^{n-1}(1-p)^{m-1}(n - (n+m)p) \\ &= p^{n-1}(1-p)^{m-1}(n+m)(\frac{n}{n+m} - p) \end{aligned}$$

1.2 Exercise 1

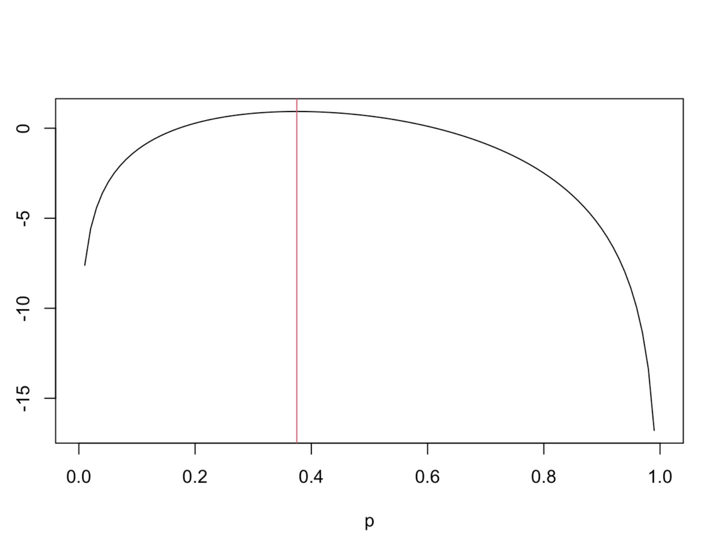
1.2.1 Exercise 1-1

p_{MLE}では$\frac{d}{dp} L(p_{MLE}|n,m) = 0$であり、かつ、$\frac{d^2}{dp^2} L(p_{MLE}|n,m) = \frac{d}{dp}(\frac{d}{dp} L(p_{MLE}|n,m)) < 0$である。 これを示せ。

1.3 対数尤度関数

$$\begin{aligned} \log L(p|n,m) &= \log(\frac{\Gamma(n+m+2)}{\Gamma(n+1)\Gamma(m+1)} p^n (1-p)^m) \\ &= \log(\frac{\Gamma(n+m+2)}{\Gamma(n+1)\Gamma(m+1)}) + n \log p + m \log(1-p) \\ &= C + n \log p + m \log(1-p) \end{aligned}$$

```
p <- seq(from=0,to=1,length=100)
d <- dbeta(p,n+1,m+1)
logd <- log(d)
plot(p,logd,type="l")
abline(v=n/(n+m),col=2)
```



最尤推定のために微分するなら対数尤度関数を微分してもよい。

$$\begin{aligned} \frac{d}{dp}(\log(L(p|n,m))) &= \frac{d}{dp}(C + n \log p + m \log(1-p)) \\ &= 0 + n \frac{d}{dp} \log p + m \frac{d}{dp} \log(1-p) \\ &= n \frac{1}{p} + m \frac{1}{1-p} \frac{d}{dp}(1-p) \\ &= n \frac{1}{p} + m \frac{1}{1-p} \times (-1) \\ &= \frac{1}{p(1-p)}(n(1-p) - mp) \\ &= \frac{n+m}{p(1-p)}(\frac{n}{n+m} - p) \end{aligned}$$

2つの大事なこと。

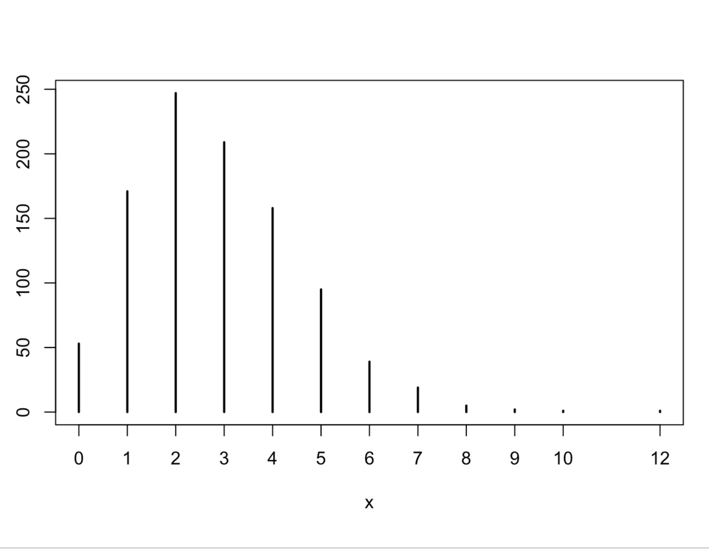
- 尤度関数のうち、pの関数ではない成分($\frac{\Gamma(n+m+2)}{\Gamma(n+1)\Gamma(m+1)}$))は最尤推定には不要
- 尤度関数の積は対数尤度関数とすることで和とすると楽になる

2 ポアソン分布

単位時間当たり平均λ回起きる現象がある。

単位時間あたりの生起回数はポアソン分布に従う。

```
lambda <- 2.8
n <- 1000
x <- rpois(n,lambda)
plot(table(x))
```



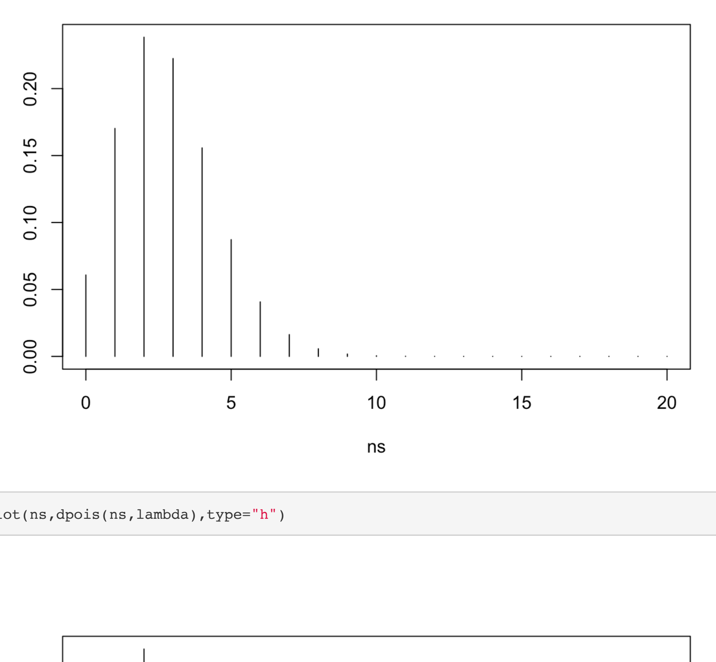

```
mean(x)
```

```
## [1] 2.846
```

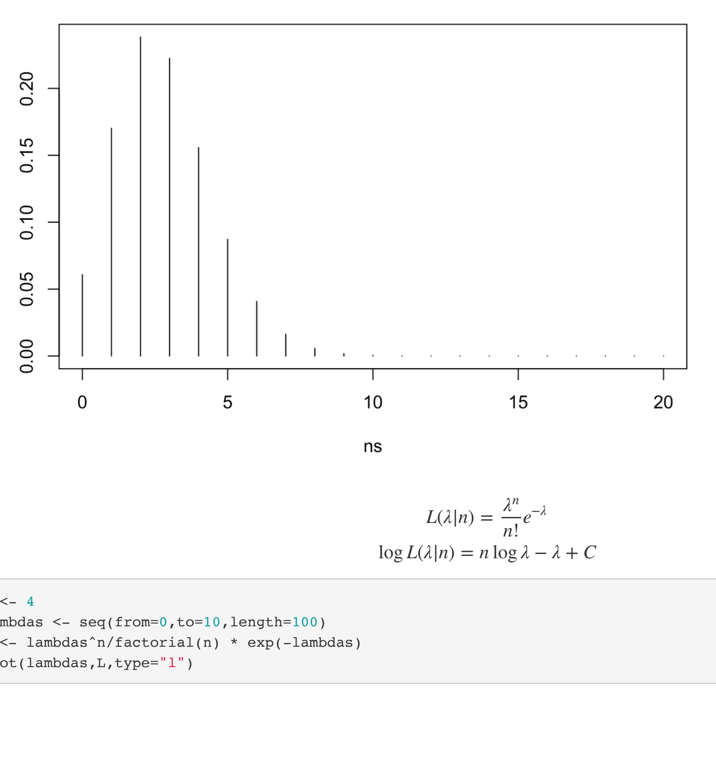
2.1 確率質量関数

$$P(n|\lambda) = \frac{\lambda^n}{n!} e^{-\lambda}$$

```
ns <- 0:20
p <- lambda^ns/factorial(ns) * exp(-lambda)
plot(ns,p,type="h")
```




```
plot(ns,dpois(ns,lambda),type="h")
```

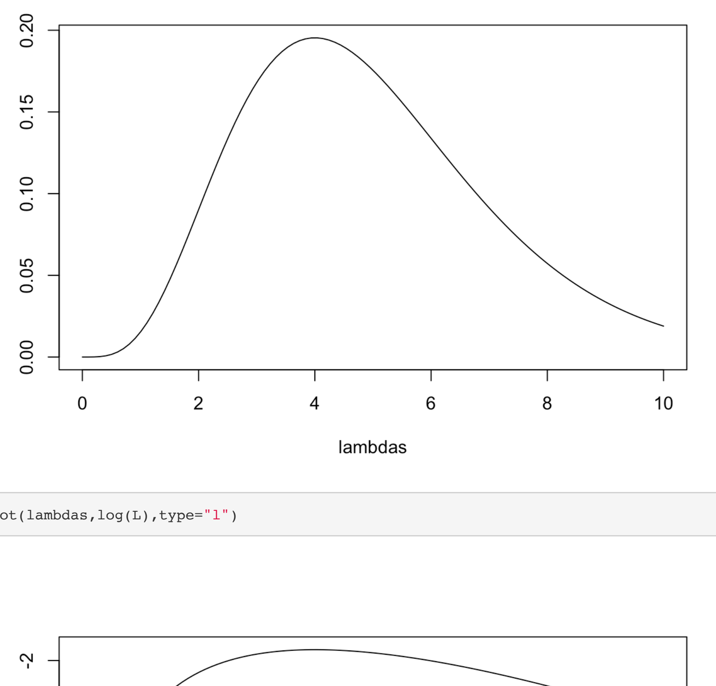


尤度関数・対数尤度関数

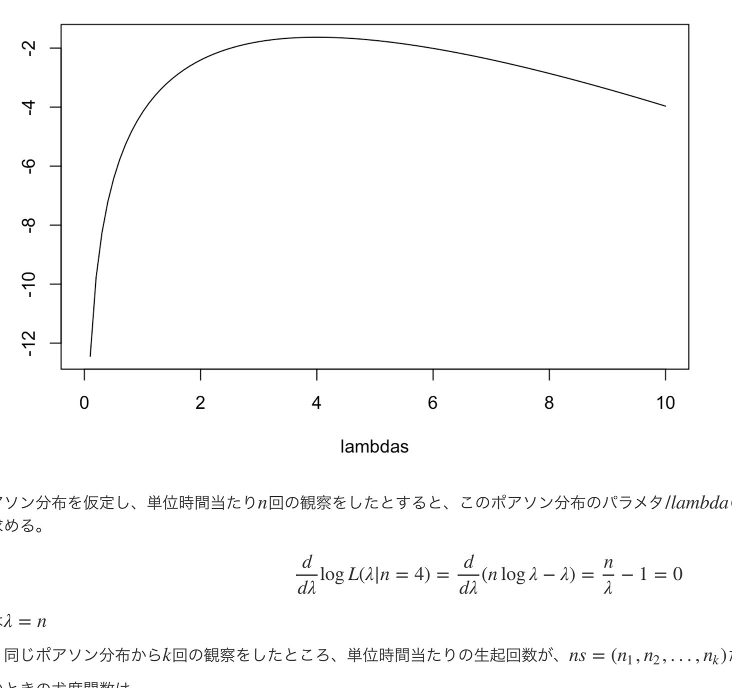
$$L(\lambda|n) = \frac{\lambda^n}{n!} e^{-\lambda}$$

$$\log L(\lambda|n) = n \log \lambda - \lambda + C$$

```
n <- 4
lambdas <- seq(from=0,to=10,length=100)
L <- lambdas^n/factorial(n) * exp(-lambdas)
plot(lambdas,L,type="l")
```




```
plot(lambdas,log(L),type="l")
```



ポアソン分布を仮定し、単位時間当たりn回の観察をしたとすると、このポアソン分布のパラメタ$llambda$の最尤推定値はいくつなのかを微分して求める。

$$\frac{d}{d\lambda} \log L(\lambda|n=4) = \frac{d}{d\lambda}(n \log \lambda - \lambda) = \frac{n}{\lambda} - 1 = 0$$

解は$\lambda = n$

今、同じポアソン分布からk回の観察をしたところ、単位時間当たりの生起回数が、$ns = (n_1, n_2, \ldots, n_k)$だったという。

このときの尤度関数は

$$L(ns|\lambda) = \prod_{i=1}^{k} L(n_i|\lambda)$$

対数尤度関数は

$$\log L(ns|\lambda) = \sum_{i=1}^{k} \log L(n_i|\lambda)$$

微分する

$$\begin{aligned} \frac{d}{d\lambda}(\sum_{i=1}^{k} \log L(n_i|\lambda)) &= \sum_{i=1}^{k} \frac{d}{d\lambda} \log L(n_i|\lambda) \\ &= \sum_{i=1}^{k} (\frac{n_i}{\lambda} - 1) \\ &= \frac{1}{\lambda}(\sum_{i=1}^{k} n_i) - k\lambda \end{aligned}$$

$\frac{d}{d\lambda} \log L(ns|\lambda) = 0$の解は$\lambda = \frac{\sum_{i=1}^{k} n_i}{k}$

3 Exercise 2

3.1 Exercise 2-1

k個の正の実数値$x = (x_1, \ldots, x_k)$が観察された。指数分布からの独立な標本とみなし、指数分布のパラメタに関する(対数)尤度関数を作れ。また、微分を使って、パラメタの最尤推定値を求めよ。

3.2 Exercise 2-2

k個の0付近の実数値が観察された。平均0、標準偏差sの正規分布からの独立な標本とみなし、sに関する(対数)尤度関数を作れ。また、微分を使って、sの最尤推定値を求めよ。

3.3 Exercise 2-3

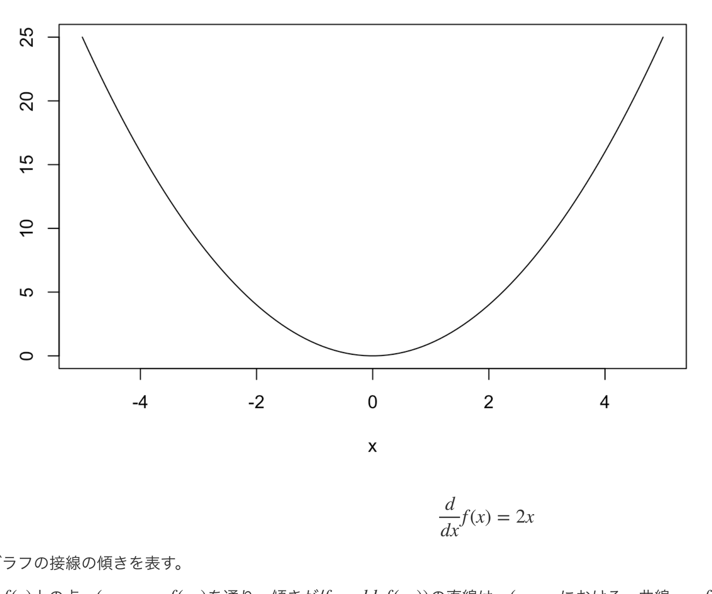
同じ標本が標準偏差が1で平均が未知の正規分布からの独立な標本であるとみなし、平均mに関する(対数)尤度関数を作れ。また、微分を使って、mの最尤推定値を求めよ。

4 傾き、接線

$$y = f(x) = x^2$$

のグラフは

```
x <- seq(from=-5,to=5,length=100)
y <- x^2
plot(x,y,type="l")
```



$$\frac{d}{dx} f(x) = 2x$$

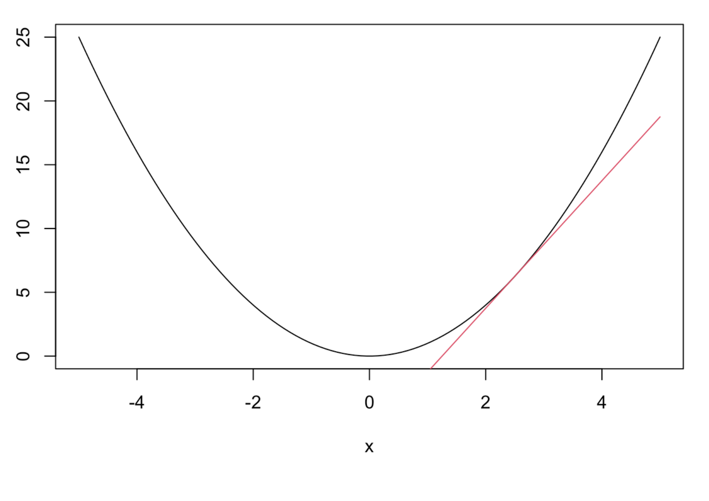
はグラフの接線の傾きを表す。

$y = f(x)$上の点、$(x_0, y_0 = f(x_0)$を通り、傾きが$/fracddxf(x_0))$の直線は、$(x_0, y_0$における、曲線$y = f(x))$の接線である。

接線上の点は

$$y = y_0 + \frac{d}{dx} f(x_0) \times (x - x_0)$$

```
x <- seq(from=-5,to=5,length=100)
y <- x^2
plot(x,y,type="l")
x0 <- 2.5
y0 <- x0^2
dfdx <- 2*x0
y.tangent <- y0 + dfdx*(x-x0)
points(x,y.tangent,type="l",col=2)
```



5 Exercise 3

5.1 Exercise 3-1

$y = x^2$ のグラフを描き、その上の多数の点の接線を同一のプロットに重ねて描け。

5.2 Exercise 3-2

$y = sin(x)$のグラフを描き、多数の接線を重ねて描け。

5.3 Exercise 3-3

同様に正規分布の確率密度関数の接線を描け。