

# Chapter04\_likelihood

Joey Votto's 출루율(OBP) = 0.46

주어진 출루율로 전체 시행(게임에 등판한 수) 대비 해당 조건이 충족되어 성공한 시행의 빈도수 비율을 확률로 정의할 수 있다. 각 타석에서 베르누이 분포가 발생하며, 다섯 번의 베르누이 분포로 구성된 이항확률분포를 수학적으로 구하는 방법과 시뮬레이션으로 발생시키는 두 가지 방법이 있다.

시뮬레이션 접근법은 모수로서 시뮬레이션을 돌릴 횟수, 타석수, 그리고 출루율을 지정해줘야 한다.

10,000 simulated results

```
a = rbinom(10000, 5, 0.46)
table(a) / 10000
```

```
## a
##      0      1      2      3      4      5
## 0.0427 0.1910 0.3363 0.2902 0.1194 0.0204
```

1,000 simulated result

```
a = rbinom(1000, 5, 0.46)
table(a) / 1000
```

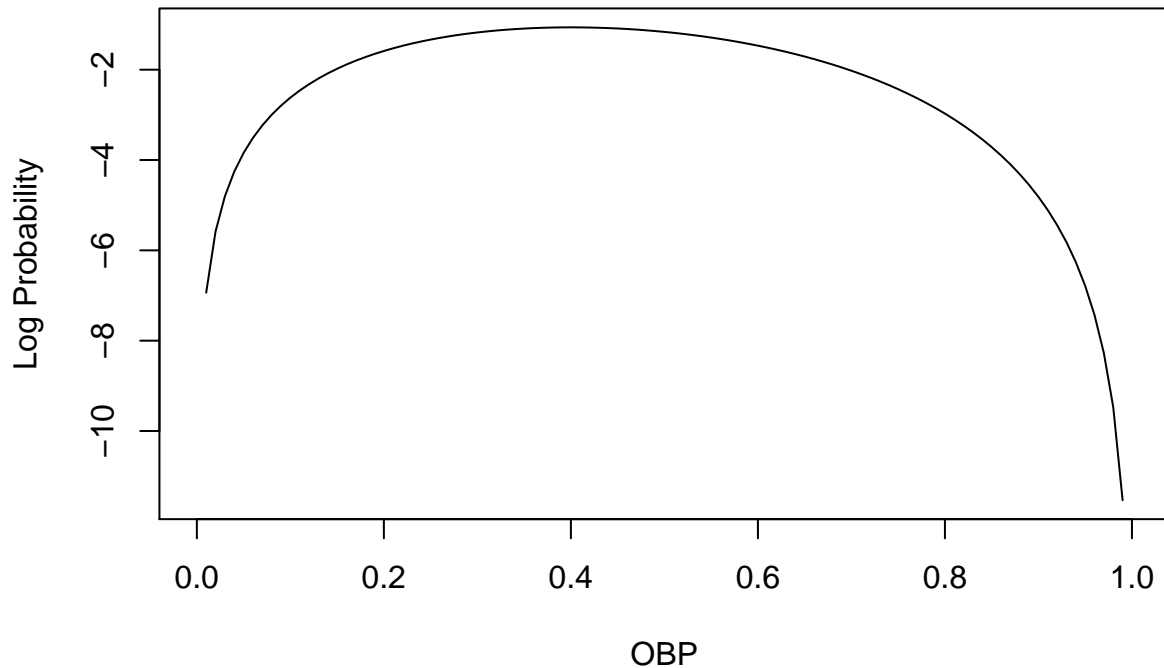
```
## a
##      0      1      2      3      4      5
## 0.046 0.189 0.326 0.286 0.135 0.018
```

## 최대 우도 추정법(Maximum Likelihood Estimation)

다섯번의 타석에서 두번의 출루를 할 가능성이 가장 높은 조건인 출루율 모수 (parameter) 를 찾기 위해서 이항확률밀도함수를 역으로 이용하는 방법을 최대우도추정이라고 한다. 보트 선수의 출루율이 알려져 있지 않은 경우, 그가 다섯 번의 타석에서 두 번 출루하는 것을 옆에서 지켜보고서, 앞으로도 다섯 번의 타석에서 두 번 출루할 수 있는 확률을 극대화하는 데 필요한 출루율을 4할 이라고 역추적했다면, 4할은 다섯 번의 타석에서 두 번의 출루 확률을 극대화하는 우도(likelihood)가 된다.

1. derivatives

```
curve(log(10*x^2*(1-x)^3), 0, 1, ylab='Log Probability', xlab='OBP')
```



## 2. 6 cases given 0.4 OBP

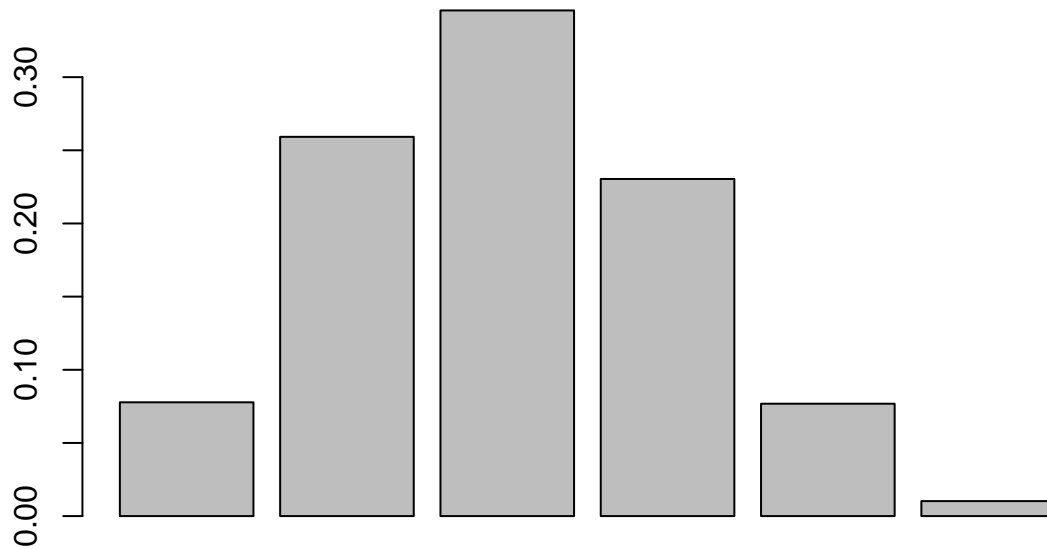
타석에 다섯 번 섰다는 건 다섯 번의 시행을 가졌다는 의미이며, 출루율 4할인 선수가 다섯 번의 타석에서 한 번이라도 출루를 만들어내지 못할 케이스를 포함해서 총 여섯 가지 각각의 시나리오마다 다른 확률을 갖게 된다. 다섯 번의 타석에서 두 번 출루할 확률이 34.5%로 가장 높은 것으로 확인됐다.

```
OBP = 0.4
base = 0:5
P = OBP^base*(1-OBP)^(5-base)
case = choose(5, base)
EV = P*case
EV
```

```
## [1] 0.07776 0.25920 0.34560 0.23040 0.07680 0.01024
```

여섯 가지 시나리오를 확률분포로 그려보면 세 번째 막대가 가장 높고 우측으로 약간 기울어진 분포가 나타난다. 다섯 번의 타석에서 출루를 두 번 할 확률이 34.6%로 가장 높게 만든 우도는 0.4였다. 전 타석에서 모두 출루할 확률은 1% 남짓 되며, 야구팬의 눈으로 볼 때 아무리 출루율 4할 선수라도 전 타석 출루를 한다는 것이 얼마나 힘든 일인지 잘 알고 있기 때문에, 출루할 것이냐 못할 것이냐의 이항확률분포가 합리적으로 만들어졌음이 눈으로 확인된다.

```
barplot(EV)
```



다른 출루율을 갖는 타자들이 다섯 번의 타석에서 출루를 두 번 할 수 있는 확률이 34.6% 보다 낮은 것으로 확인되면, 최대우도추정방식으로 모수를 정확하게 추정한 것으로 볼 수 있다.

출루율이 다음과 같을 때 4개 막대그래프를 한 눈에 나타내 보자. [OBP = 0.2 / 0.3 / 0.4 / 0.5]

```
par(mfrow=c(2,2))

OBP = 0.2
base = 0:5
P = OBP^base*(1-OBP)^(5-base)
case = choose(5, base)
EV = P*case
barplot(EV, main='OBP=0.2 (20.48%)', ylab='possibility')

OBP = 0.3
base = 0:5
P = OBP^base*(1-OBP)^(5-base)
case = choose(5, base)
EV = P*case
barplot(EV, main='OBP=0.3 (30.87%)', ylab='possibility')

OBP = 0.4
base = 0:5
P = OBP^base*(1-OBP)^(5-base)
case = choose(5, base)
EV = P*case
barplot(EV, main='OBP=0.4 (34.56%)', ylab='possibility')
```

```

OBP = 0.5
base = 0:5
P = OBP^base*(1-OBP)^(5-base)
case = choose(5, base)
EV = P*case
barplot(EV, main='OBP=0.5 (31.25%)', ylab='possibility')

```

