

Survival analysis_HW2

김민국

2019-10-17

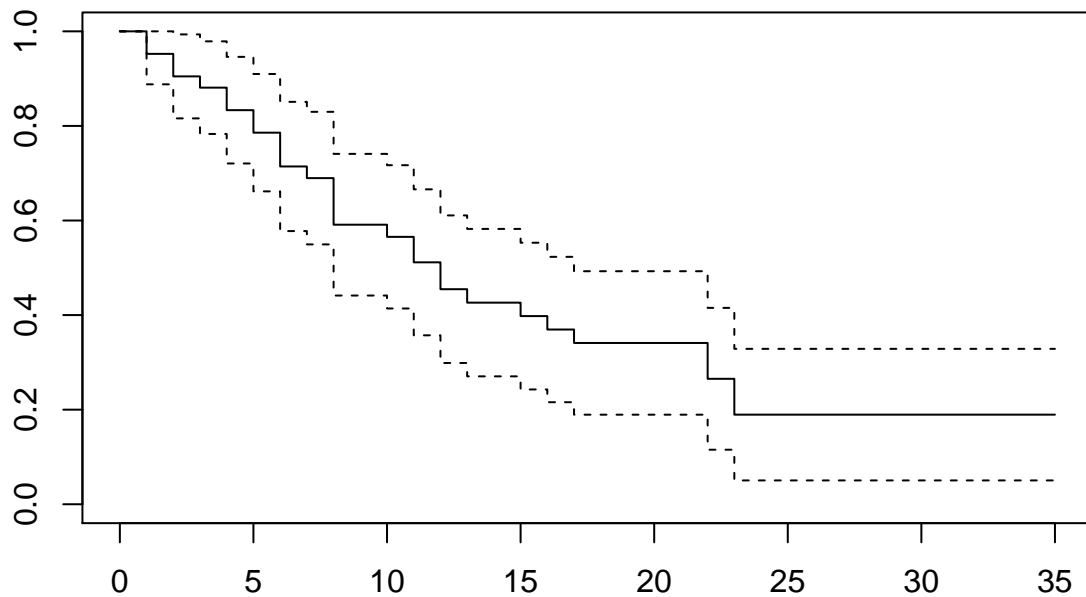
```
library(survival)
library(KMsurv)
```

```
data(drug6mp)
mp <- drug6mp
head(mp)
```

```
##   pair remstat t1 t2 relapse
## 1    1      1  1 10        1
## 2    2      2 22  7        1
## 3    3      2  3 32        0
## 4    4      2 12 23        1
## 5    5      2  8 22        1
## 6    6      1 17  6        1
```

-> help를 통해 drug6mp 변수들의 설명을 보면 t1은 placebo 환자들의 time, t2는 6-mp 환자들의 time이다.
-> placebo 환자들은 모두 fail 됐으며, 6-mp 환자들의 경우 fail 됐으면 relapse 변수가 1의 값을 갖게 된다.

```
mp_time <- c(mp$t1, mp$t2)
mp_event <- c(rep(1, length(mp$t1)), mp$relapse)
surv_x_1 <- Surv(time = mp_time, event = mp_event)
surv_x_1_fit <- survfit(formula = surv_x_1 ~ 1, data = surv_x_1, conf.type = "plain")
plot(surv_x_1_fit)
```

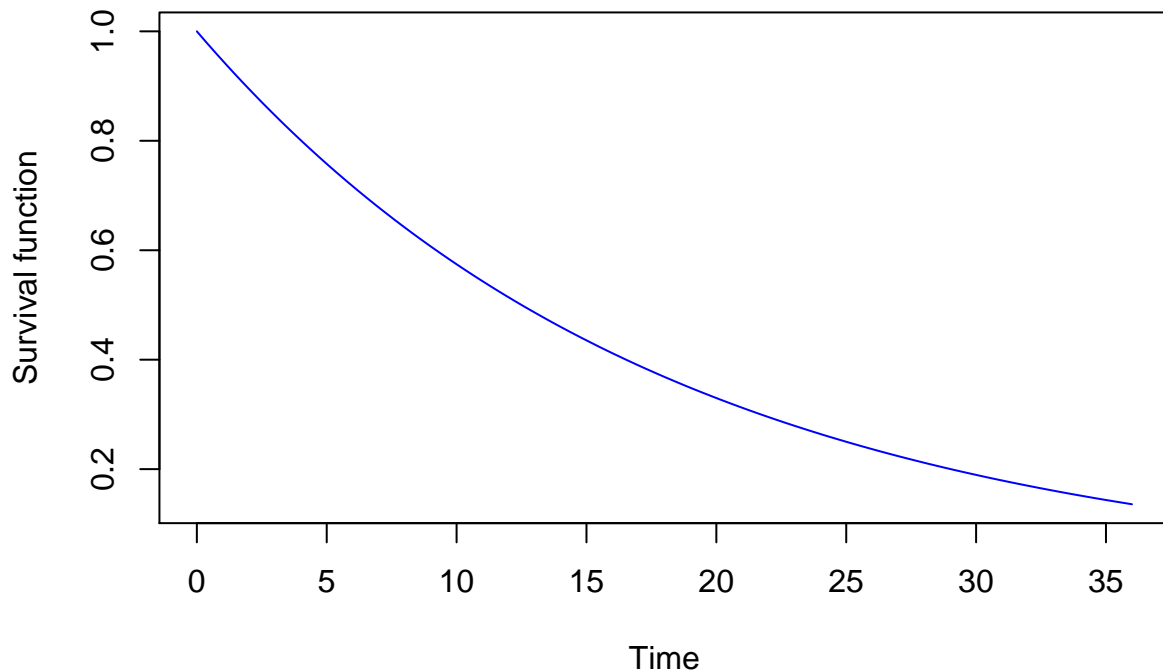


-> KM estimator를 그려주기 위해 mp_time 이라는 time들을 모아 놓은 벡터와 mp_event 라는 indicator 벡터를 만들어준다.

-> KM estimator를 그려준 결과이다.

```
lamda_hat_1 <- sum(mp_event)/sum(mp_time) # exponential ditribution's mle
curve(exp(-lamda_hat_1*x), from = 0, to = max(mp_time) + 1, col = "blue",
      xlab = "Time", ylab = "Survival function", main = "Using exponetial ditribution")
```

Using exponential ditribution

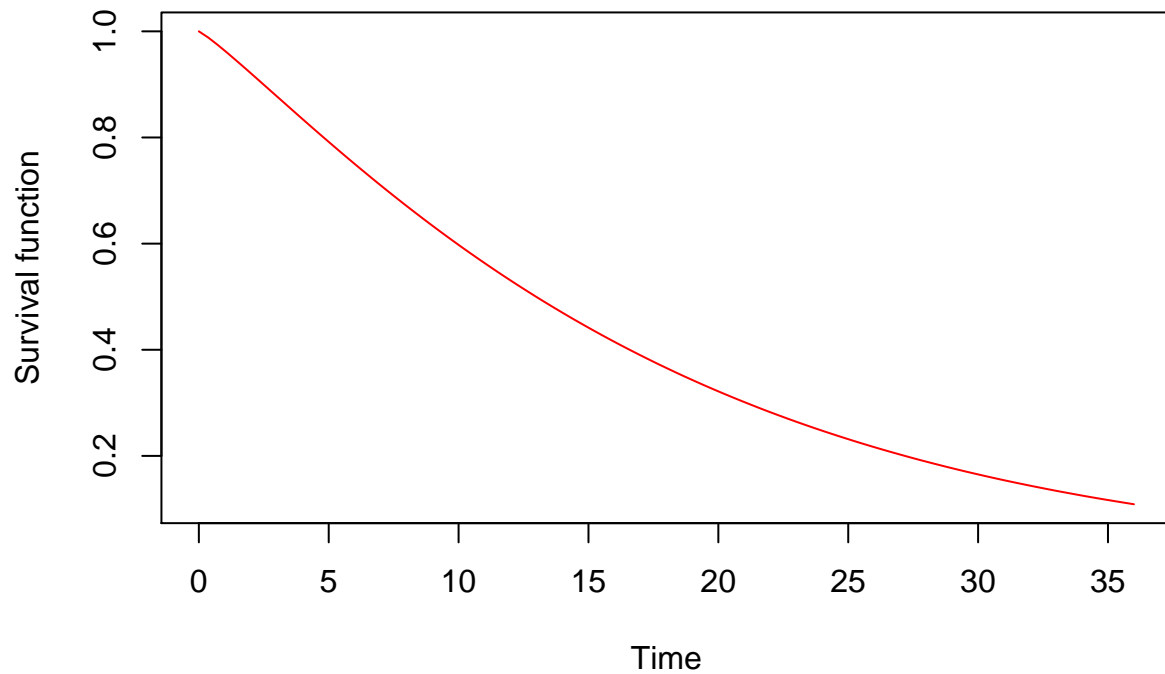


-> 지수분포를 통해 $S(x)$ 를 추정해본다. 이 때 모수의 값을 모르기 때문에 MLE를 이용한다.

-> 지수분포의 MLE값의 식을 구하고 구한 MLE값으로 지수분포를 그려준다.

```
alpha_1 <- 1.14 # weibull ditribution's mle
lamda_1 <- sum(mp_event) / sum((mp_time)^alpha_1) # weibull ditribution's mle
ln_1 <- log(alpha_1*lamda_1)*sum(mp_event) + (alpha_1 - 1)*sum(log(mp_time)*mp_event) -
  lamda_1 * sum((mp_time)^alpha_1)
curve(exp(-lamda_1*(x^alpha_1)), from = 0, to = max(mp_time) + 1, col = "red",
      xlab = "Time", ylab = "Survival function", main = "Using weibull ditribution")
```

Using weibull ditribution

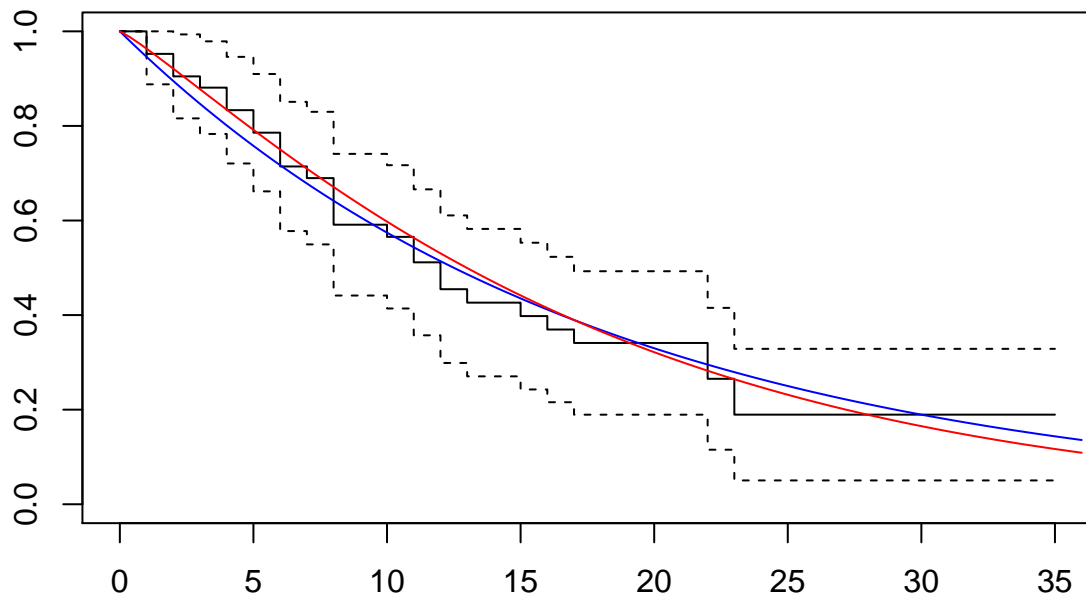


-> 와이불분포를 통해 $S(x)$ 를 추정해본다. 이 역시 MLE를 통해 모수의 값을 추정해야 한다.

-> λ 의 경우 α 의 함수로 표현이 가능하지만 α 는 \ln_1 값이 최대가 되도록 경험적으로 값을 찾아준다.

-> \ln_1 식을 먼저 구한 후 편미분을 통해 λ 식을 도출 후 α 는 값을 대입하며 찾아준다.

```
plot(surv_x_1_fit)
curve(exp(-lamda_hat_1*x), from = 0, to = max(mp_time) + 1, col = "blue", add = T)
curve(exp(-lamda_1*(x^alpha_1)), from = 0, to = max(mp_time) + 1, col = "red", add = T)
```



- > KM estimator와 지수분포, 와이불분포를 이용했을 때 추정한 $S(x)$ 를 한 번에 그린 그림이다.
- > 세가지 모두 Time이 커질수록 $S(x)$ 값이 작아지고 있는 것을 확인할 수 있다.
- > KM estimator와 두가지 경우를 비교하자면 KM estimator가 초반에는 $S(x)$ 를 크게 추정하지만 시간이 흐를수록 추정값이 작아지게 된다.
- > $S(x)$ 의 경향성을 보았을 때 parametric 함수들의 MLE과정이 번거로울 경우 non-parametric 방법인 KM estimator 방법을 이용하는 것도 괜찮을 것 같다.