Models for time to failure

With the hazard function defined below,
$H(t) = \frac{f(t)}{1 - F(t)}$,
$F(t)$ and $f(t)$ are the CDF and PDF respectively.

If the exponential distribution is used for the model, i.e.,
$f(t) = \lambda exp(-\lambda t)$ and $F(t) = 1 - exp(-\lambda t)$.
The modeled harzard function will be
$H(t) = \lambda$,
which is a constant function and is therefore unreasonable for describing
$H(t) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \mathrm{P}(\text{Get disease during time } t \text{ to } t + \epsilon \mid \text{Never had disease up tp time } t)$.

So, one alternative is to model with *Weibull* distribution, given the PDF and CDF below,
$g_{\alpha,\beta}(t) = \frac{\beta}{\alpha}(\frac{t}{\alpha})^{\beta-1} exp(-(\frac{t}{\alpha})^{\beta})$ and $G_{\alpha,\beta}(t) = 1 - exp(-(\frac{t}{\alpha})^{\beta})$.
We will have the harzard function,
$H(t) = \frac{\beta}{\alpha}(\frac{t}{\alpha})^{\beta-1}$.
This is a more useful model compared to the one generated by exponential distribution. In this case, with appropriate parameters $\alpha > 0, \beta > 0$, we may choose different scenarios for different studies. For example, $\alpha = 2, \beta = 2$ will give us positive correlation in *getting disease at time t* and *no disease up till time t*, which is a plausible assumption.

Survival curves

A new surgical procedure effectiveness test is set up as the following:
$N = 1000$ individuals divided equally into surgical treament group X and control group Y;
times to death are *i.i.d* with $X_i \sim G_{3,2}$ and $Y_i \sim G_{2,2}$.
We would like to estimate the two survival functions $S_X(t) = P(X_i > t)$ and $S_Y(t) = P(Y_i > t)$.

1. Random generation for the Weibull distribution

The R function that I write for drawing *Weibull* random variables employs the inverse CDF sampling method. From the CDF, we derive that the inverser CDF is $G^{-1}_{\alpha,\beta}(x) = \alpha(-\log(1-x))^{1/\beta}$.

```
my.rweibull <- function(n, shape, scale = 1) {
    if (n <= 0 || shape <= 0 || scale <= 0) {
        warning("Invalid arguments! All arguments should be positive.")
        return("NaN")
    }
    u <- runif(n)
    x <- scale * (-log(1 - u))^(1/shape)
    return(x)
}
```

2. Kaplan-Meier survival-function estimate

With the two input vectors, *times* of event times and *censor* of censoring indicators, we can produce the Kaplan-Meier survivial-function estimate from
$S(t_i) = \dfrac{N(t_i) - E(t_i)}{N(t_i)} \times S(t_{i-1})$, with $S(0) = 1$
where vector t contains valid estimation points (excluding censored times), N denotes the number of subjects at risk and E represents the number of events(deaths) at the evaluated time. The output of my function contains the estimation points t vector, survival rates *surv* vector and the lookup table function for the Kaplan-Meier estimates.

```
my.KMsurv <- function(times, censor) {
    if (length(times) <= 0 || length(censor) <= 0)
        stop("Invalid input vectors!")
    if (length(times) != length(censor))
        stop("Mismatch of dimension in input vectors!")
    t <- c(0, sort(unique(times[censor])))  # prepare t estimates points
    s <- rep(0, length(t))  # create empty survival time estimates vector
    s[1] <- 1  # initialize
    for (i in 2:length(t)) {
        n <- length(times[times >= t[i]])  # num of subjects at risk at time t
        e <- length(times[times == t[i]])  # num of events at time t
        s[i] <- (n - e)/n * s[i - 1]
    }
    # estimated survival function S(t)
    my.surv <- function(eval) {
        return(s[t == eval])
    }
    # return a list of related KM objects
    km <- list(t = t, surv = s, s.func = my.surv)
    return(km)
}
```

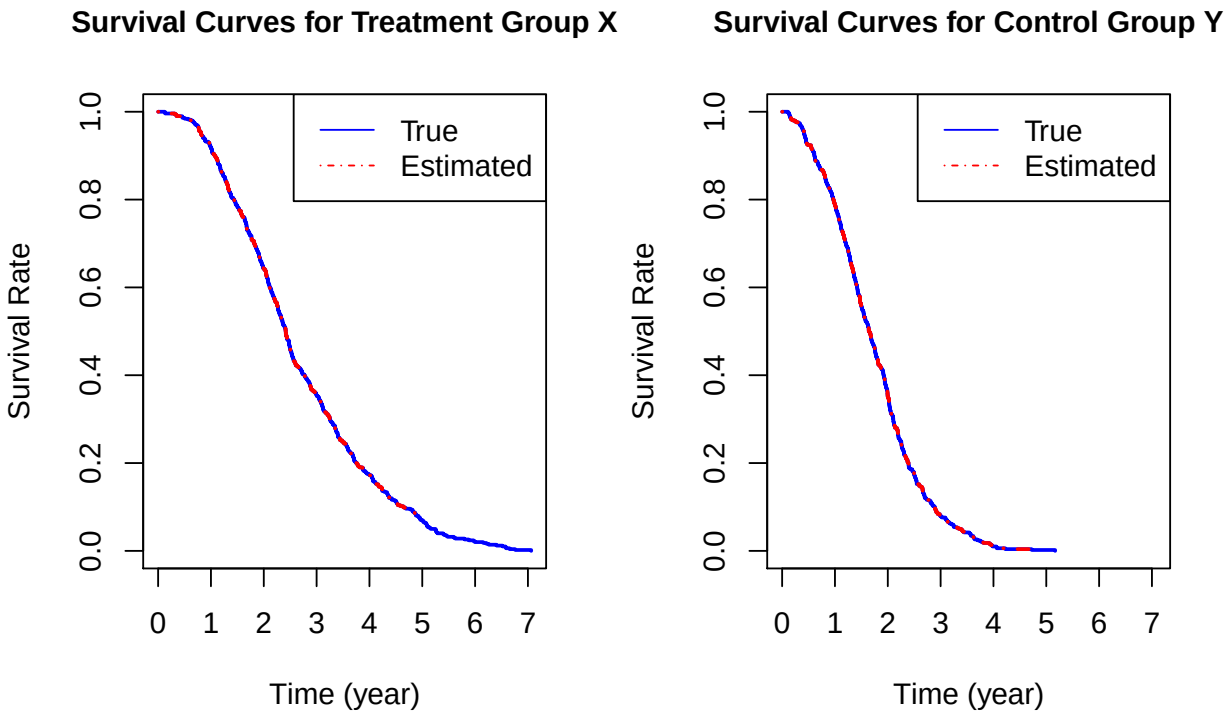

Figure 1: Visual comparisons for constant cut-off simulations

3. Simple simulations with constant cut-off

Since the study has a length of 5 years, we set the censoring indicator for the data by cut-off at 5. We then simulate the performace of the clinic trial with the setup described above. The graphical comparison for both treament and control groups between the true and estimated survival curves are shown in Figure 1.

We can see that the estimated curves are both very close to the real curves before the cut-off point, but beyond that point (in this case, 5), we have no estimation at all. Moreover, the treatment group shows higher survival rate from the plots indicating effectiveness of this surgical procedure.

4. Advanced simulations with independent random cut-off

We try to inject randomness into the censoring for estimation beyond the clinical cut-off point. Here, we assume $i.i.d\ Z_i \sim\ Exp(\lambda = 1/10)$, that gives the time at which i will be censored. Z being Independent of X and Y, our simulation results of the trial setup are shown graphically in Figure 2.

From the plot, we can see that we do have more estimations beyond the clinical trial cut-off point, which makes the estimated curve more complete. However, we have seen more error in this plot. The estimated curve for the treatment group seems to overestimate the survival rates. This is probably caused by the fact that we are adding more *effective* patients into the counts, making the estimations overly optimistic.

5. Advanced simulations with dependent random cut-off

A little tweak to 4, here we assume $Z_I|X_I < 2 \sim Exp(\lambda = 1/10)$ and $Z_i|X_i \geq 2 \sim Exp(\lambda = 1/5)$. This creates dependence of Z on X, similarly for Z and Y, which coule arise in a study where the sicker patients are more likely to remain under the care of their doctors and have longer expected follow-up time. The results of this run of simulation are shown in Figure 3.

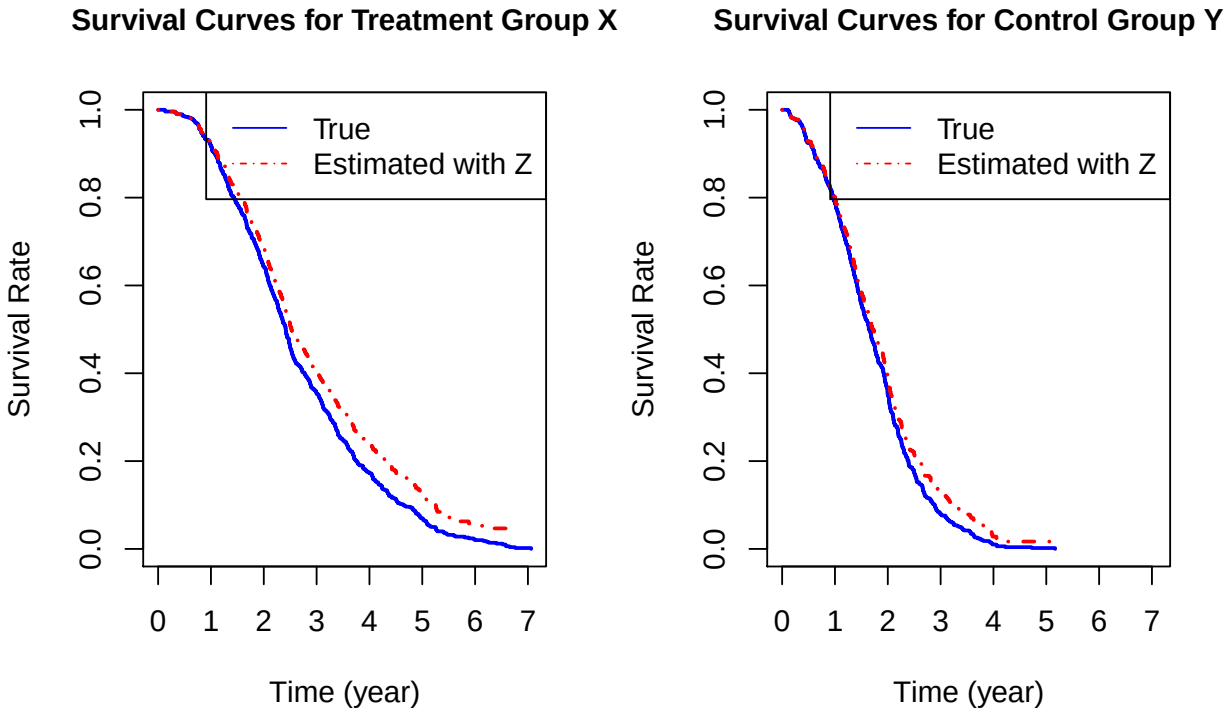

Figure 2: Visual comparisons for independent random cut-off simulations

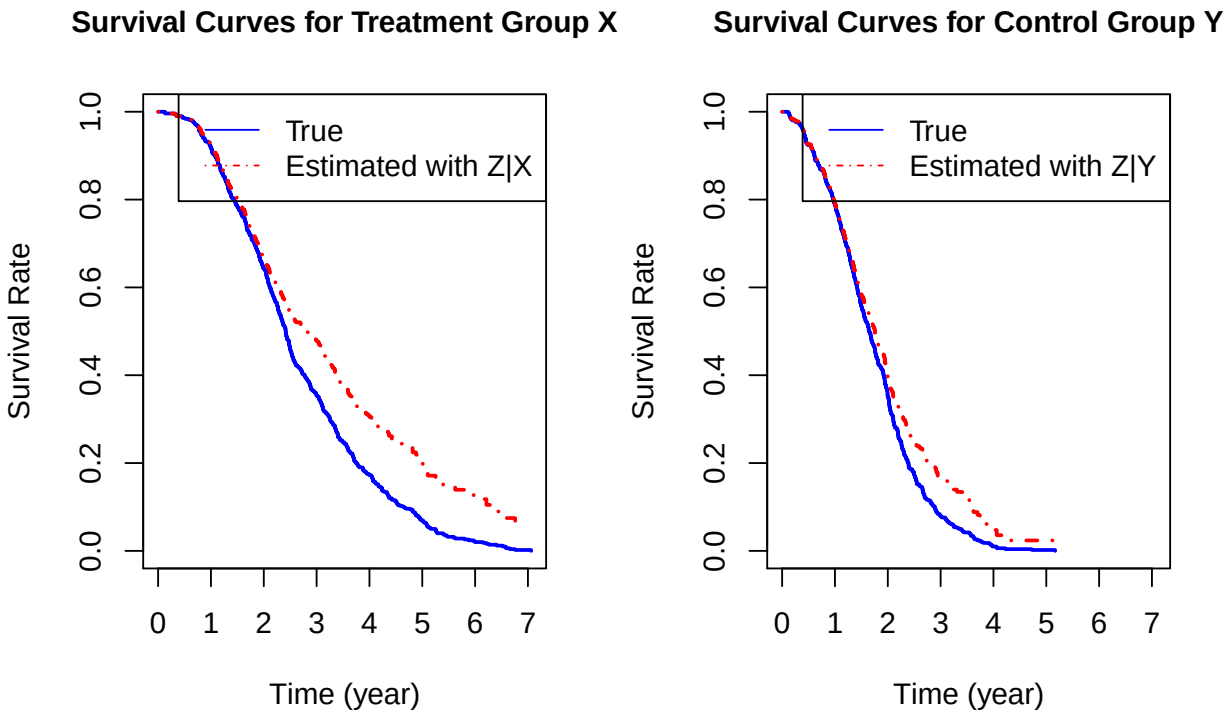

Figure 3: Visual comparisons for dependent random cut-off simulations

Similar to Figure 2, we can see a even more complete estimated curve but also more optimistically biased curve, especially for the treament group. This censoring scenario amplifies the effectiveness of the treament on the sicker patients by prolonging their follow-up time threshold and adding more effectively treated patients into the counts. This seems overly optimistic for the estimation of the survival rates.