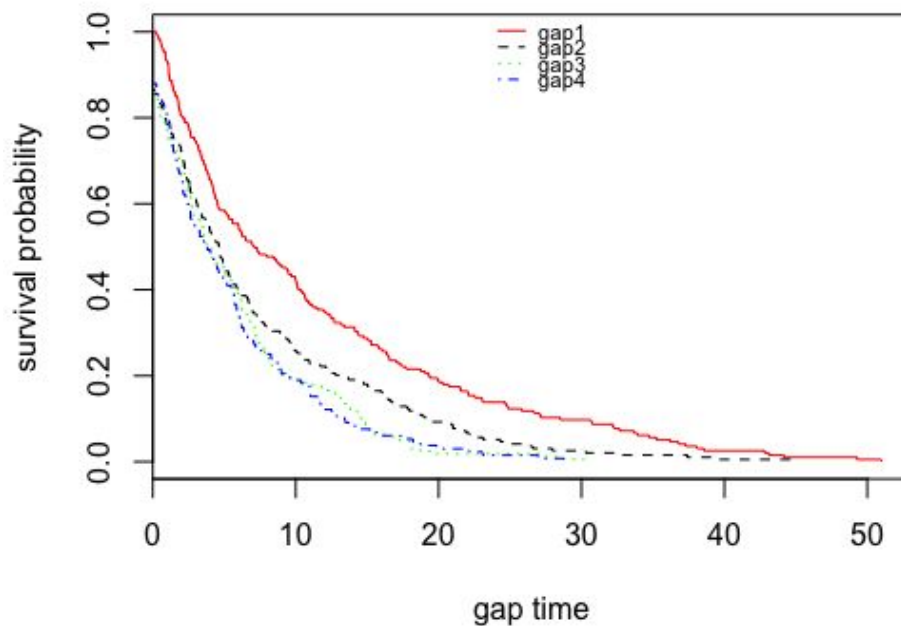


## Simulation 2:

(pg 138 Give an illustration of the effect of dependent censoring. )

Simulate a random sample of 200 individuals, where the log gap times,  $Y_{ij} = \log(W_{ij})$  ( $j=1,2,\dots$ ), were generated from a multivariate normal distribution in which each  $Y_{ij}$  was identically distributed as  $N(2,2)$ , and with equal correlations  $\text{corr}(Y_{ij}, Y_{ik}) = 0.5$ , for  $j \neq k$ . A common censoring time  $C_i = 52$  was imposed on each process.



## R code:

```
#figure 4.1 kaplan-meier estimates for the first four gap times, improperly ignoring  
dependent censoring
```

```
library(MASS)
```

```
# for individual i, use mvrnorm() function to generate random values
```

```
set.seed(101)
```

```
Yi_funct<- function(n){
```

```
  S<- matrix(1,nrow = n, ncol = n) #S is the covariance matrix of the log(Gap)
```

```
  for (i in 1:n){
```

```
    S[i,i]<- 2
```

```
  }
```

```
  Yi<-mvrnorm(n=1, rep(2,n),S)
```

```

Wi<-exp(Yi)
Ti<-sum(Wi)
return(c(Wi))
}

```

```

dataj<-matrix(0,nrow = 200,ncol = 50)
for (i in 1:200){
  dataj[i,]<-Yi_funct(50)
}
dataj
dataj[1,]
dim(dataj)

```

```

Gmat<-matrix(0,nrow = 200,ncol = 50)
for (i in 1:200){
  Gmat[i,]<-cumsum(dataj[i,])
}

```

```

#find the number of observations for each individual before or at censoring time
for (i in 1:200){
  for (j in 1:50){
    if (Gmat[i,j]>52){
      Gmat[i,j]=0
    }
  }
}

```

```

library(km.ci)
Nobs<-rowSums(Gmat!=0)
time1<-Gmat[,1]
status<- ifelse((time1!=0),1,0)
gap1<-data.frame(time1,status)
fit<-survfit(Surv(time1,status==1)~1)
plot(fit)
plot(fit, lty = c(1,0,0), col = "red", xlab = "gap time", ylab = "survival probability")
# c(a,b,c) the first a is the km-estimator , b/c are for the confidence interval

```

```

time2<-Gmat[,2]-Gmat[,1]
status<- ifelse((time2!=0),1,0)
gap2<-data.frame(time2,status)
fit2<-survfit(Surv(time2,status==1)~1)
lines(fit2, lty=c(2,0,0), col = "black")

```

```

time3<-Gmat[,3]-Gmat[,2]
status<- ifelse((time3!=0),1,0)
gap3<-data.frame(time3,status)
fit3<-survfit(Surv(time3,status==1)~1)

```

```
lines(fit3, lty=c(3,0,0), col = "green")
```

```
time4<-Gmat[,4]-Gmat[,3]  
status<- ifelse((time4!=0),1,0)  
gap4<-data.frame(time4,status)  
fit4<-survfit(Surv(time4,status==1)~1)  
lines(fit4, lty=c(4,0,0), col = "blue")
```

```
legend('top', legend=c("gap1","gap2","gap3","gap4"), bty = "n", cex=0.7, lty = 1:4, col =  
c("red","black","green","blue"))
```