Survival Models: Week 3

Estimation - Parametric

- Often we will assume that the random survival times of a population follow a particular *parametric* distribution defined by $f(t)$. For example, last week we saw an example where $f(t) = \lambda \exp(-\lambda t)$.
- Once we have assumed a particular distribution we need a means of estimating the parameters that define the distribution.
- For the moment we will ignore censoring, that is, our data are complete.

Method of Moments (MOM)

Method of moments is a relatively simple form of estimation that involves equating sample moments with population moments. The population moments for a random variable X are $E(X^j)$ and the corresponding sample moments are $n^{-1}\sum_{i=1}^{n} x_i^j$. The method is best illustrated via an example.

$$E(X^j) = \frac{\sum_{i=1}^{n} x_i^j}{n}$$

The future lifetimes of a population follow an exponential distribution ($f(t) = \lambda \exp(-\lambda t)$). Based on a sample of survival times $x_1, x_2, ..., x_n$, compute the method of moments estimator of λ.

For the exponential distribution we know that $E(X) = \frac{1}{\lambda}$ and that the first sample moment is $n^{-1}\sum_{i=1}^{n} x_i$. Equating the sample and population moments gives $\frac{1}{\hat{\lambda}} = n^{-1}\sum_{i}^{n} x_i$ and $\hat{\lambda} = (n^{-1}\sum_{i=1}^{n} x_i)^{-1}$.

MOM: method of moments

normal distribution

$X \sim N(\mu, \sigma^2)$

$\hat{\mu}_{MOM}$ $\hat{\sigma}^2_{MOM}$

① First moment

$$E(X) = \frac{1}{n}\sum_{i=1}^{n} x_i$$ → first sample moment

↓

first population moment

$$\hat{\mu}_{MOM} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

② Second moment

$$E(X^2) = \frac{1}{n}\sum x_i^2$$

↓

$$\mu^2 + \sigma^2 = \frac{1}{n}\sum x_i^2$$

$$\sigma^2 = \frac{1}{n}\sum x_i^2 - \boxed{\mu^2} \rightarrow \boxed{\hat{\mu}^2_{MOM}}$$

$$\hat{\sigma}_{MOM} = \frac{1}{n}\sum x_i^2 - \boxed{\left(\frac{1}{n}\sum x_i\right)^2}$$

The standard way to find more than one estimated parameters:

plug in ⇒ For n parameters we need n equations

i.e up to ~~the~~ n-th sample moment.

Maximum Likelihood Estimation (MLE)

Maximum likelihood estimation involves maximizing the likelihood of the observed data, viewed as a function of the unknown parameters.

The future lifetimes of a population follow an exponential distribution. Based on a sample of survival times $x_1, x_2, ..., x_n$, compute the maximum likelihood estimate of λ.

1. Write the likelihood: $L(\lambda | x_1, ..., x_n) = \prod_{i=1}^{n} \lambda \exp(-\lambda x_i)$.
2. Typically easier to work with log-likelihood $l(\lambda | x_1, ..., x_n) = log L(\lambda | x_1, ..., x_n) = n log(\lambda) - \sum_{i=1}^{n} \lambda x_i$.
3. Maximize the likelihood w.r.t λ. $\frac{dl(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$.
4. Setting the derivative to zero and solving gives $\hat{\lambda} = (n^{-1} \sum_{i=1}^{n} x_i)^{-1}$.

MLE with Exponential Distribution

$x \sim \exp(\lambda)$

$f(x) = \lambda e^{-\lambda x}$

$$L(\lambda | x_1, \dots, x_n) = \lambda \cdot e^{-\lambda x_1} \cdot \dots \lambda e^{-\lambda x_n}$$

$$= \lambda^n e^{-\lambda(x_1 + \dots + x_n)}$$

$$= \lambda^n e^{-\lambda \sum_{i=1}^n x_i}$$

$$= \prod_{i=1}^n \lambda e^{-\lambda x_i}$$

LL

$$l(\lambda | x_1, \dots x_n) = n \cdot \log \lambda - \lambda \cdot \sum_{i=1}^n x_i$$

$$l'(\lambda | x_1, \dots, x_n) = \frac{n}{\lambda} - \sum_{i=1}^n x_i$$

w.r.t. λ

$$\Rightarrow \hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i}$$

$$l'' = -\frac{n}{\lambda^2} < 0$$

so $\hat{\lambda}$ is a maximizer.

Maximum Likelihood Estimation (MLE)

- CALCULATE VARIANCE OF MLE ESTIMATION

Asymptotically (as $n \to \infty$) maximum likelihood estimators have the following properties:

1. Consistent: $\hat{\theta} \to \theta$.
2. Normally distributed with variance $\frac{1}{I(\theta)}$, where $I(\theta)$ is the Fisher Information.
3. $I(\theta) = -E(\frac{d^2}{d\theta^2} \log L(\theta))$. check if $I(\theta)$ is positive.

Note: The above results mean that for large enough n, our MLE $\hat{\theta}$ is approximately $N(\theta, \frac{1}{I(\theta)})$.

$Var(\hat{\lambda}) = \frac{1}{I(\hat{\lambda})}$, $\hat{\lambda}$ is random

Note: Find expected second derivative first, then plug in $\hat{\lambda}$.

$I(\lambda) = -E_X(l''(\lambda)) = \frac{n}{\lambda^2}$

$I(\hat{\lambda}) = \frac{n}{\hat{\lambda}^2} \Rightarrow Var(\hat{\lambda}) = \frac{\hat{\lambda}^2}{n} = \frac{n}{(\sum_{i=1}^{n} x_i)^2}$

Maximum Likelihood Estimation (MLE)

The number of deaths in a particular town (each week) follows a Poisson distribution ($f(X, \theta) = P(X = k) = \frac{\theta^k \exp(-\theta)}{k!}$). Over a period of 11 weeks the following death counts were observed $6, 6, 0, 3, 5, 4, 9, 5, 3, 7, 6$. Compute the MLE of θ and provide a variance for your estimate.

Poisson MLE

$$f(x_i, \theta) = \frac{\theta^{x_i} \exp(-\theta)}{x_i!}$$

$$L(\theta | x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{\theta^{x_i} \exp(-\theta)}{\cancel{}}$$

$$\propto \theta^{\sum_{i=1}^{n} x_i} \exp(-n\theta)$$

$$l(\theta | x_1, \dots, x_n) = \sum_{i=1}^{n} x_i \ln \theta - n \cdot \theta$$

$$l'(\theta | x_1, \dots, x_n) = \frac{\sum_{i=1}^{n} x_i}{\theta} - n = 0$$

$$\hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{54}{11}$$

Compute variance.

$$l''(\theta | x_1, \dots, x_n) = l''(\theta) = -\frac{\sum_{i=1}^{n} x_i}{\theta^2}$$

$$-E(l''(\theta)) = \frac{E(\sum_{i=1}^{n} x_i)}{\theta^2} = \frac{\sum_{i=1}^{n} E(x_i)}{\theta^2}$$

$$= \frac{n\theta}{\theta^2} = \frac{n}{\theta}$$

$$Var(\hat{\theta}) = \frac{1}{I(\hat{\theta})} = \left(\frac{n}{\hat{\theta}}\right)^{-1} = \frac{\hat{\theta}}{n} = \frac{54}{121}$$

R Example

```
poisson.L<-function(mu,x) {

n<-length(x)
logL<-sum(x)*log(mu)-n*mu
return(-logL)

}

optim(par=1,fn=poisson.L,method="BFGS",x=c(6,6,0,3,5,4,9,5,3,7,6))
```

R Example

```
normal.L<-function(theta,x) {

mu<-theta[1]
sigma2<-theta[2]
n<-length(x)
logL<- -.5*n*log(2*pi) -.5*n*log(sigma2) -
(1/(2*sigma2))*sum((x-mu)**2)
return(-logL)

}
xdata<-rnorm(20,2,sd=2)
optim(par=c(1,1),fn=normal.L,method="BFGS",x=xdata)
mean(xdata)
var(xdata) #not exactly the mle of sigma2
```

R Example

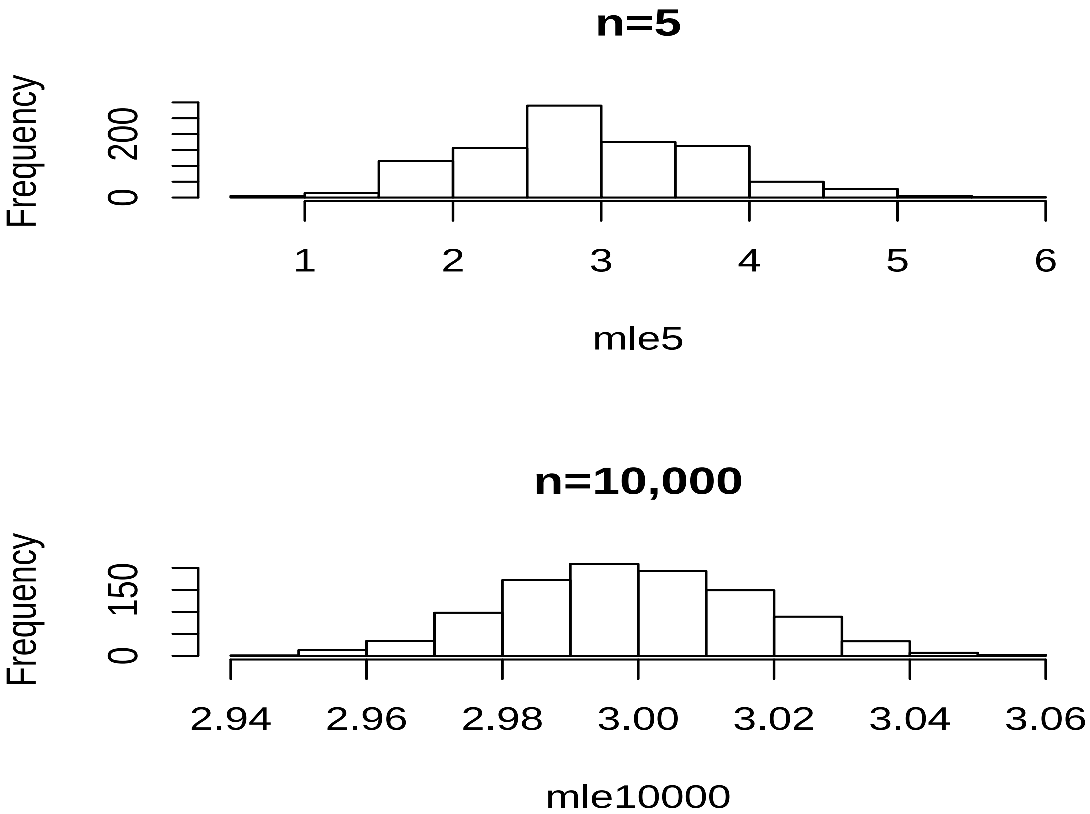


Figure 1: Histograms of MLE for n=5 and n=10,000 and a Poisson distribution

A Non-parametric Approach

If we observe N lifetimes a non-parametric estimate of the cumulative distribution function at time t is:

$$\hat{F}(t) = \frac{d(t)}{N},$$

where $d(t)$ is the number of the N observed lifetimes that are $\leq t$.

- An alternative to specifying a particular parametric distribution is to use a "distribution free" or non-parametric approach.
- One advantage of a non-parametric approach is that we do not need to specify a particular distribution - can be more robust.
- One disadvantage of a non-parametric approach is that it can be less "efficient" than a suitable parametric approach.

really bad at prediction b/c it's discrete CDF

A Non-parametric Approach

$$\hat{F}(t) = \frac{d(t)}{N}$$

where $d(t)$ is the number of the N observed lifetimes that are $\leq t$.

- $d(t)$ is binomial$(N, F(t))$.
- $E(\hat{F}(t)) = E(\frac{d(t)}{N}) = N^{-1} \times N \times F(t) = F(t)$. (unbiased).
- $V(\hat{F}(t)) = \frac{1}{N} F(t)(1 - F(t))$

Non parametric approach

$d(t) = \sum_{i=1}^{N} X_i$ $\quad X_i = \begin{cases} 1 & \text{ith dead by } t \\ 0 & \text{ith alive by } t \end{cases}$

$X_i \overset{iid}{\sim} \text{Bernoulli}(F(t))$

$\downarrow$ $P(T < t)$

$d(t) \sim \text{Bin}(N, F(t))$

$$\hat{F}(t) = \frac{d(t)}{N}$$

$E(\hat{F}(t)) = \frac{1}{N} E(d(t)) = \frac{1}{N} \cdot N \cdot F(t) = F(t)$ $\Rightarrow$ $\hat{F}(t)$ unbiased

$$\begin{aligned} Var(\hat{F}(t)) &= \frac{1}{N^2} Var(d(t)) \\ &= \frac{1}{N^2} \cdot N \cdot F(t) \cdot (1 - F(t)) \\ &= \frac{\hat{F}(t)(1 - \hat{F}(t))}{N} \quad \text{(plug in)} \end{aligned}$$

R Example

```
#example of estimating F(t) using parametric
#and non-parametric approach.

observed<-rexp(20,0.25)
lammle<-1/mean(observed)
Fparam<-1-exp(-lammle*seq(0,max(observed),by=0.01))
Fnonparam<-rep(0,length(seq(0,max(observed),by=0.01)))
j<-1
for(i in seq(0,max(observed),by=0.01)) {
Fnonparam[j]<-sum(observed<=i)/length(observed)
j<-j+1
}
plot(seq(0,max(observed),by=0.01),Fparam,type="l",main=
```

```
"Estimation of CDF",xlab="Time",ylab="Probability")
lines(seq(0,max(observed),by=0.01),Fnonparam,col="red")
```

R Example

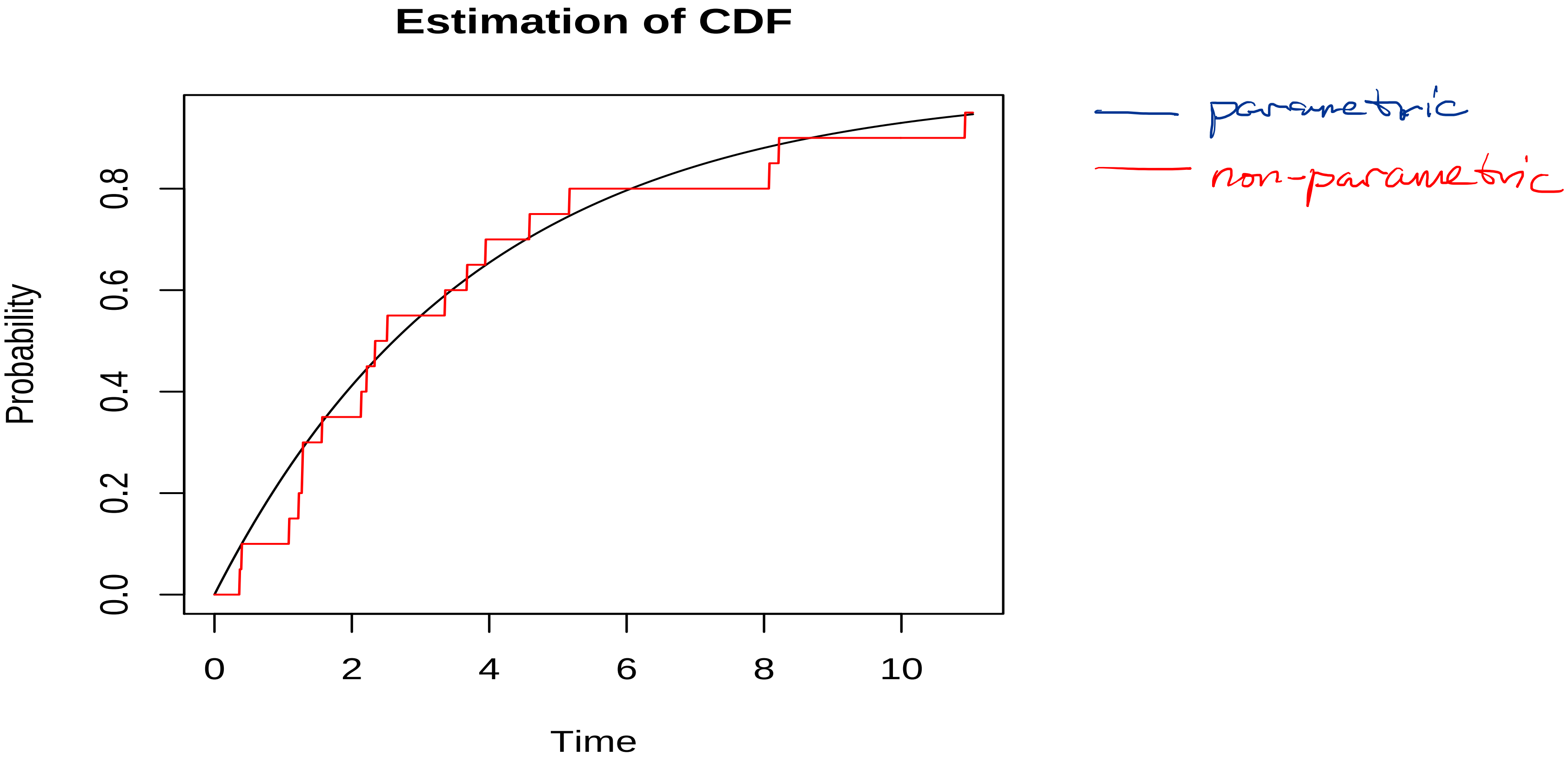


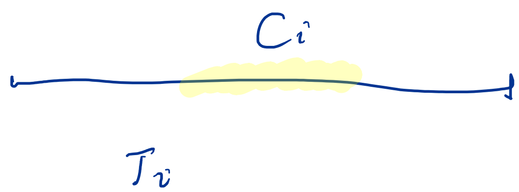
Figure 2: Comparison of parametric and non-parametric estimation

Censoring

≠ missing data.

- Censoring refers to the situation where we only know that an observation (survival time) falls in a particular interval - we do not know the exact value of the observation.
- *Right censoring* - only know that survival time is equal to or larger than a particular value. Example: subject moves interstate or investigation ends.
- *Left censoring* - do not know when condition of interest started. Example: survival once contract a particular disease.
- *Interval censoring* - only know survival time falls in a particular interval. Both left and right censoring are forms of interval censoring.

Censoring



- *Random censoring* - let C_i represent the time of censoring of the *ith* life and T_i the random lifetime of the *ith* life. Life is censored if $C_i < T_i$.
  1. Definition 1: Censoring is random if C_i is a random variable.
  2. Definition 2: Censoring is random if C_i and T_i are independent random variables. (This is the most common definition)
- *Informative/non-informative censoring* - Censoring is non-informative if it provides no information about the future lifetime (T_i). Definition 2 of random censoring implies non-informative censoring, definition 1 does not.

The methods we will speak about in this course will rely on the fact that the censoring is non-informative.