What to do this week (March 7 and 9, 2023)?

Part I. Introduction
Part II. Epidemiologic Concepts and Designs
Part III. Clinical Trials

Part IV. Modern Biostatistical Approaches

Part IV.1 Incomplete Data Analysis

Part IV.2 Some Other Important Topics (Chp 8 - 18, Koepsell and Weiss, 2003)

IV.2.1 Measures of Risks
IV.2.2 Measurement Error Revisit
IV.2.3 Confounding and Its Control
IV.2.4 Causation vs Association

Part IV.3 Selected Widely-Used Algorithms

IV.3.1 Statistical Simulation and Bootstrap
IV.3.2 EM Algorithm and Related

Part IV.4 Lifetime Data Analysis

Part IV.3.1 Statistical Simulation and Bootstrap: Monte Carlo Methods

What do we care about a random variable X?

Its **distribution**: its pattern of taking different values, that is, what values X takes and how often it takes a particular value.

How do we find out X's distribution from its observations (**data**: $x_1, \ldots, x_n$)?

- (i) by descriptive analysis: plotting/tabulating the data; summarizing the data with statistics
- (ii) by **making inference** on X's disnt $F(\cdot)$

  (iia) to approximate (estimate) $F(\cdot)$ by point/interval estimation;
  (iib) to choose between (test on) two contradictory claims about $F(\cdot)$ by hypothesis testing

  How to verify a conclusion? How to assess performance of an inference procedure?

Part IV.3.1 Statistical Simulation and Bootstrap: Monte Carlo Methods

- **Monte Carlo** refers to an area of Monaco, where the *Monte Carlo Casino* is located.
- **Monte Carlo methods** (or Monte Carlo experiments) are a class of computational algorithms that obtain numerical results by repeated random sampling.
- Monte Carlo methods are especially useful for simulating phenomena with significant uncertainty in inputs and random systems.

How does an Monte Carlo method work?

- How to simulate a particular system?
- After quantifying the system by a rv, how to simulate the rv?

Part IV.3.1 Statistical Simulation and Bootstrap: Monte Carlo Generation

Uniform generator. eg the one in the software package *R*: "runif(n,min,max)"

```
x=runif(100);
hist(x, freq=FALSE, ... ...)
curve(dunif(x), col = 2, lty = 2, lwd = 3, add = TRUE)
y=runif(100);
plot(x,y, xlab=x, ylab=y, pch=18, col=4, sub=(a2). n=100)
x=runif(1000);
hist(x, freq=FALSE, ... ...)
curve(dunif(x), col = 2, lty = 2, lwd = 3, add = TRUE)
y=runif(1000);
plot(x,y, xlab=x, ylab=y, pch=18, col=4, sub=(b2). n=1000)
```

Part IV.3.1 Statistical Simulation and Bootstrap: Monte Carlo Generation

How to generate random variables?

- *R* has generators of most commonly used rvs: eg. "rnorm(n,mean,sd)"
- Use transformations of commonly used rvs: for example,

```
x=runif(1000);
w=3*x-1;
z=rnorm(1000);
v=3*z+5;
hist(x, freq=FALSE, ...)
curve(dunif(x), col = 2, lty = 2, lwd = 3, add = TRUE)
hist(w, freq=FALSE, ...)
lines(w,rep(1/3,1000), col = 3, lty = 2, lwd = 3)
hist(z, freq=FALSE, ...)
curve(exp(-x^2/2)/(2*pi)^.5, min(z),max(z),
                col = 4, lty = 2, lwd = 3, add = TRUE)
hist(v, freq=FALSE, ...)
curve(exp(-(x-5)^2/(2*9))/(2*pi*9)^.5, min(v),max(v),
                 col = 5, lty = 2, lwd = 3, add = TRUE)
```

Part IV.3.1 Statistical Simulation and Bootstrap: Monte Carlo Generation

How to generate random variables?

- Use transformations of commonly used rvs: for example,
  - If given a cdf $F(\cdot)$, $Y = F^{-1}(U)$ with $U \sim U(0,1)$ has $Y \sim F(\cdot)$

```
ztmp=matrix(rnorm(4000),ncol=4)
b=apply(ztmp^2,1,sum)
x=runif(1000);
t=-log(1-x)/2
hist(b, freq=FALSE, ...)
lines(density(b), col = 2, lty = 2, lwd = 3)
b2=rchisq(1000,df=4)
hist(b2, freq=FALSE, ...)
lines(density(b2), col = 3, lty = 2, lwd = 3)
hist(t, freq=FALSE, ...)
lines(density(t), col = 4, lty = 2, lwd = 3)
t2=rexp(1000,2)
hist(t2, freq=FALSE, ...)
lines(density(t2), col = 5, lty = 2, lwd = 3)
```

Part IV.3.1 Statistical Simulation and Bootstrap: Monte Carlo Generation

How to generate random variables?

- **Accept-Reject Algorithm.** If $f(\cdot)$ is a pdf and $f(x) \leq Mg(x)$ with M a constant and $g(\cdot)$ the *instrumental* pdf.

  Step 1. Generate $Y \sim g(\cdot)$ and $U \sim U(0,1)$ indptly.

  Step 2. If $U \leq \frac{f(Y)}{[Mg(Y)]}$, take $X = Y$ and go to Step 3; otherwise, return to Step 1.

  Step 3. Obtain X, which follows $f(\cdot)$.

To prove it? (p298, the textbook by Hogg et al)

To prove $P(X \leq x) = \int_{-\infty}^{x} f(z)dz$.

Proof.

$$
\begin{aligned}
P(X \leq x) &= P\Big(Y \leq x \Big| U \leq \frac{f(Y)}{Mg(Y)}\Big) \\
&= \frac{P\Big(Y \leq x, U \leq \frac{f(Y)}{Mg(Y)}\Big)}{P\Big(U \leq \frac{f(Y)}{Mg(Y)}\Big)} \\
&= \frac{\int_{-\infty}^{x} \big[\int_0^{f(y)/Mg(y)} 1du\big] g(y)dy}{\int_{-\infty}^{\infty} \big[\int_0^{f(y)/Mg(y)} 1du\big] g(y)dy} \\
&= \frac{\int_{-\infty}^{x} \big[\frac{f(y)}{Mg(y)}\big] g(y)dy}{\int_{-\infty}^{\infty} \big[\frac{f(y)}{Mg(y)}\big] g(y)dy} \\
&= \int_{-\infty}^{x} f(z)dz
\end{aligned}
$$

Remark. The algorithm's probability of an acceptance is $1/M$:

$$P\Big(U \leq \frac{f(Y)}{Mg(Y)}\Big) = \frac{1}{M}$$

Example. Suppose $X \sim N(0,1)$ with pdf $f(x) = (2\pi)^{-1/2}\exp(-x^2/2)$, and $Y \sim \mathit{Cauchy}(0,1)$ with pdf $g(y) = \pi^{-1}(1+x^2)^{-1}$.
Note that

$$\frac{f(x)}{g(x)} = \frac{\pi}{\sqrt{2\pi}}(1+x^2)\exp(-x^2/2).$$

Because $(1+x^2)\exp(-x^2/2) \leq 2\exp(-1/2)$, $f(x) \leq Mg(x)$ with $M = \frac{\pi}{\sqrt{2\pi}}(1.213) = 1.520$.

Use the Accept-Reject Algorithm to generate observations from N(0,1):

```
x<-rep(0,1000)
for(i in 1:1000){
        y<-rcauchy(1, location = 0, scale = 1);
        u<-runif(1, min=0, max=1)
        while(u>(exp(-y^2/2)/(2*pi)^.5/1.520*pi*(1+y^2))){
        y<-rcauchy(1, location = 0, scale = 1);
        u<-runif(1, min=0, max=1)
}
        x[i]<-y
}
```

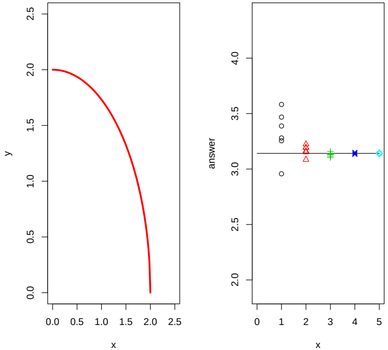
Monte Carlo Integration

How to calculate $\int_a^b g(x)dx$?

Example.

$$\int_0^2 \sqrt{4-x^2}dx = 2\int_0^2 \sqrt{4-x^2}(\frac{1}{2})dx = 2E\{\sqrt{4-X^2}\} \quad (\pi)$$

provided that $X \sim U(0,2)$.



(a) y=(4−x^2)^.5 (b) true answer=pi

(b) Approximates to the integral by generating n observations from $U(0,2)$, with $n = 10^k$ for $k = 1, \ldots, 5$.

Part IV.3.1 Statistical Simulation and Bootstrap: Simulation Example 1

To verify the normal approximation to binomial distn:

```
xtmp=matrix(ifelse(runif(1000*10)<.3,1,0),ncol=10)
x=apply(xtmp,1,sum)
hist(x, freq=FALSE, ...)
y=rbinom(n=1000,size=10,prob=0.3)
hist(y, freq=FALSE, breaks=11, ...)

xtmp=matrix(ifelse(runif(1000*100)<.3,1,0),ncol=100)
x=apply(xtmp,1,sum)
hist(x, freq=FALSE,breaks=20, ...)
curve(exp(-(x-30)^2/2/(30*.7))/(2*pi*21)^.5, 0, 100, lty=1, c
        lwd=3, add=TRUE)
y=rbinom(n=1000,size=100,prob=0.3)
hist(y, freq=FALSE,breaks=20, ...)
curve(exp(-(x-30)^2/2/(30*.7))/(2*pi*21)^.5, 0, 100, lty=1, c
        lwd=3, add=TRUE)
```

Part IV.3.1 Statistical Simulation and Bootstrap: Simulation Example 2

What can data missing result in?

```
x=runif(1000);
y=runif(1000);
w=ifelse(x<y,1,0);  sum(w)/1000

r0=ifelse(runif(1000)<.5,1,0)
x0=x[r0==1];y0=y[r0==1];
w0=ifelse(x0<y0,1,0);sum(w0)/sum(r0);

r1=rep(0,1000)
r1[x<y]=rbinom(length(x[x<y]),size=1,prob=.8)
r1[x>=y]=rbinom(length(x[x>=y]),size=1,prob=.2)
x1=x[r1==1];y1=y[r1==1];
w1=ifelse(x1<y1,1,0);sum(w1)/sum(r1);

r2=rep(0,1000)
r2[x<y]=rbinom(length(x[x<y]),size=1,prob=.3)
r2[x>=y]=rbinom(length(x[x>=y]),size=1,prob=.7)
x2=x[r2==1];y2=y[r2==1];
w2=ifelse(x2<y2,1,0);sum(w2)/sum(r2);
```

Part IV.3.1 Statistical Simulation and Bootstrap: Preparation for Bootstrap

Consider rv $X \sim F(\cdot)$: iid observations $X_1, \ldots, X_n$

Definition. The following is the **empirical function** with the random sample:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}(X_i \leq x) \quad \text{for} \quad -\infty < x < \infty.$$

For all $x \in (-\infty, \infty)$,

- $E\{F_n(x)\} = F(x)$ and $Var\{F_n(x)\} = F(x)\big[1 - F(x)\big]/n$.
- more ... ...

$F_n(\cdot)$ is a very good estimator of $F(\cdot)$.

Part IV.3.1 Statistical Simulation and Bootstrap: Bootstrap

Consider rv $X \sim F(\cdot)$: iid observations $X_1, \ldots, X_n$

- When to use a point estmator $\widehat{\theta}(X_1, \ldots, X_n)$ of a population parameter θ, how to estimate its variance $Var(\widehat{\theta})$?

Thinking ... ...

- If we could have a random sample $\widehat{\theta}_b$ for $b = 1, \ldots, B$ from the same population as $\widehat{\theta}$, we can estimate the variance with

$$s^2_{\widehat{\theta}} = \sum_{b=1}^{B} (\widehat{\theta}_b - \bar{\widehat{\theta}})^2 / (B-1)$$

with $\bar{\widehat{\theta}} = \sum_{b=1}^{B} \widehat{\theta}_b / B$.

- That can be achieved if there are $X_{1b}, \ldots, X_{nb}$ iid from $F(\cdot)$. However, $F(\cdot)$ is unknown. How to overcome it?

Part IV.3.1 Statistical Simulation and Bootstrap: Bootstrap

Consider rv $X \sim F(\cdot)$: iid observations $X_1, \ldots, X_n$

- When to use a point estmator $\widehat{\theta}(X_1, \ldots, X_n)$ of a population parameter θ, how to estimate its variance $Var(\widehat{\theta})$?

Bootstrap variance estimation:

- Step 1. Generate $X_{1b}^*, \ldots, X_{nb}^*$ iid from the empirical function $F_n(\cdot)$.
  (*Resample with size n from* $X_1, \ldots, X_n$ *with replacement.*)
- Step 2. Calculate $\widehat{\theta}(X_{1b}^*, \ldots, X_{nb}^*)$, denoted by $\widehat{\theta}_b^*$.
- Repeat Steps 1. and 2. B times and obtain $\{\widehat{\theta}_b^* : b = 1, \ldots, B\}$.
- With $\bar{\widehat{\theta}}^* = \sum_{b=1}^B \widehat{\theta}_b^*/B$, calculate $s_{\widehat{\theta}^*}^2 = \sum_{b=1}^B (\widehat{\theta}_b^* - \bar{\widehat{\theta}}^*)^2/(B-1)$.
- Use $s_{\widehat{\theta}^*}^2$ to estimate $Var(\widehat{\theta})$.

Resampling methods: Jackknife (J.W. Tukey, 1958); Bootstrap (Bradley Efron, 1979)

Part IV.3.1 Statistical Simulation and Bootstrap: Bootstrap

Consider rv $X \sim F(\cdot)$: iid observations $X_1, \ldots, X_n$

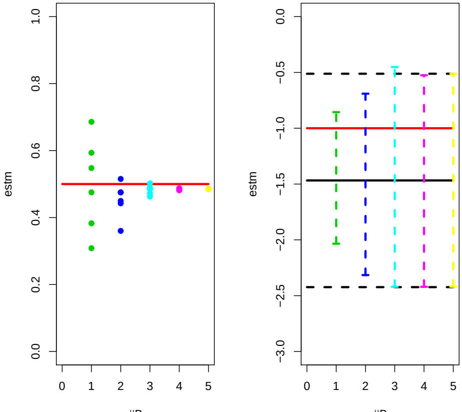
- How to obtain an interval estimator of a population parameter θ based on a point estmator $\widehat{\theta}(X_1, \ldots, X_n)$?

Bootstrap confidence interval:

- Step 1. Generate $X^*_{1b}, \ldots, X^*_{nb}$ iid from the empirical function $F_n(\cdot)$.
  (*Resample with size n from* $X_1, \ldots, X_n$ *with replacement.*)
- Step 2. Calculate $\widehat{\theta}(X^*_{1b}, \ldots, X^*_{nb})$, denoted by $\widehat{\theta}^*_b$.
- Repeat Steps 1. and 2. B times and obtain $\left\{\widehat{\theta}^*_b : b = 1, \ldots, B\right\}$.
- Sort the sequence as $\widehat{\theta}^*_{(1)} \leq \ldots \leq \widehat{\theta}^*_{(B)}$, and obtain bootstrap percentiles: $\widehat{\theta}^*_{((\alpha/2)100)}$ and $\widehat{\theta}^*_{((1-\alpha/2)100)}$.
- Use $(\widehat{\theta}^*_{((\alpha/2)100)}, \widehat{\theta}^*_{((1-\alpha/2)100)})$ as a $(1-\alpha)100\%$ CI for θ.

Bootstrap example

Consider $X \sim F(\cdot)$ with $\mu = E(X)$ and iid obs $X_1, \ldots, X_n$.



(a) Bootstrap Estm for SD of Xbar

(b) Bootstrap CI Estm for Population Mean

$X \sim N(-1, 5^2)$ with $n = 100$ and $B = 10^k$ for $k = 1, \ldots, 5$.

Part IV.3.2 EM Algorithm and Related: Maximum Likelihood Estimation (MLE)

- **Likelihood Function:** $X_1, \ldots, X_n$ are iid observations (a random sample) from the population with distribution $f(x; \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Omega$. If the observed values are $x_1, \ldots, x_n$, then $L(\boldsymbol{\theta} \mid \text{ data }) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta})$.

- **Maximum Likelihood Estimator (MLE):** The **MLE** $\widehat{\boldsymbol{\theta}}$ is maximizes $L(\boldsymbol{\theta})$: $\widehat{\boldsymbol{\theta}} = argmax_{\boldsymbol{\theta} \in \Omega} L(\boldsymbol{\theta} \mid \text{ data })$.

  The MLE $\widehat{\boldsymbol{\theta}}$ gives the parameter value that agrees most closely with the observed sample (the data).

- **Asymptotic Properties** Provided some regularity conditions,

  1. $\widehat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}$.
  2. $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{D} MN(\mathbf{0}, \mathrm{FI}(\boldsymbol{\theta})^{-1})$.

  For $j = 1, \ldots, K$, the jth component of $\widehat{\boldsymbol{\theta}}_n$ satisfies

$$\sqrt{n}(\widehat{\theta}_{n,j} - \theta_j) \xrightarrow{D} N(0, [\mathrm{FI}(\boldsymbol{\theta})^{-1}]_{jj}).$$

Part IV.3.2 EM Algorithm and Related: Likelihood-Based Tests:

Consider $H_0 : \boldsymbol{\theta} \in \Omega_0$ vs $H_1 : \boldsymbol{\theta} \in \Omega_1$ using a random sample $X_1, \ldots, X_n \sim f(x; \boldsymbol{\theta})$ for $\theta \in \Omega \subseteq \mathcal{R}^K$ with $\Omega_0 \bigcap \Omega_1 = \emptyset$.

- **Wald Test:** If $\Omega_0 = \{\boldsymbol{\theta}_0\}$ and $\Omega_1 = \Omega \setminus \{\boldsymbol{\theta}_0\}$ and $n >> 1$, $W = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \left[n\mathrm{FI}(\boldsymbol{\theta}_0 \right] (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \underset{H_0}{\sim} \chi^2(K)$ approximately.

  $\Longrightarrow$ Rejection region with level of α: $\{\mathbf{x} : W > c\}$, the critical value c is $\chi^2(K)_{1-\alpha}$ of the $\chi^2(K)$-distn.

- **Likelihood Ratio Test:** If $\Omega_1 = \Omega \setminus \Omega_0$ and $n >> 1$, $-2\log(\Lambda) \underset{H_0}{\sim} \chi^2(K-q)$ approximately. Here $\Lambda = \frac{\max_{\theta \in \Omega_0} L(\theta)}{\max_{\theta \in \Omega} L(\theta)}$ and if q the number of indpt parameters under H_0 is q.

  $\Longrightarrow$ Rejection region with level of α: $\{\mathbf{x} : -2\log(\Lambda) > c\}$, the critical values c is $\chi^2(K-q)_{1-\alpha}$ of the $\chi^2(K-q)$-distn.

Part IV.3.2 EM Algorithm and Related: Expectation-Maximization (EM) Algorithm

What does it do?
An iterative procedure (algorithm) to calculate MLEs of the population parameters $\boldsymbol{\theta}$ when it is hard to maximize the likelihood function of θ with the available data.

Original References:

- Dempster, Laird, and Rubin (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm". *Journal of the Royal Statistical Society, Series B*. 39 (1): 1–38.
- C.F. Jeff Wu (1983). "On the Convergence Properties of the EM Algorithm". *Annals of Statistics*. 11 (1): 95–103.
- B.W. Turbull (1976) "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data". *Journal of the Royal Statistical Society, Series B*. 38 (3): 290–295.

Part IV.3.2 EM Algorithm and Related: Expectation-Maximization (EM) Algorithm

Goal: Consider to maximize $L(\boldsymbol{\theta}|\mathbf{X})$, the (observed) likelihood function with the available data $\mathbf{X}$ to obtain the MLE of population parameter $\boldsymbol{\theta}$.

Suppose that it is relatively easier to maximize $L_C(\boldsymbol{\theta}|\mathbf{X},\mathbf{Z})$, the (complete) likelihood function of $\boldsymbol{\theta}$ with the "augmented" data $(\mathbf{X},\mathbf{Z})$.

Define $Q(\boldsymbol{\theta}|\boldsymbol{\theta}_0,\mathbf{x}) = E_{\boldsymbol{\theta}_0}\big[\log L_C(\boldsymbol{\theta}|\mathbf{X},\mathbf{Z})|\boldsymbol{\theta}_0,\mathbf{X}=\mathbf{x}\big]$.

EM Algorithm. Let $\hat{\boldsymbol{\theta}}^{(m)}$ be the estimate on the mth step with $m \geq 0$. The following compute the estimate on the $(m+1)$th step:

- *Expectation-Step.* Compute

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)},\mathbf{x}) = E_{\hat{\boldsymbol{\theta}}^{(m)}}\big[\log L_C(\boldsymbol{\theta}|\mathbf{X},\mathbf{Z})|\hat{\boldsymbol{\theta}}^{(m)},\mathbf{X}=\mathbf{x}\big].$$

- *Maximization-Step.* Let the updated estimate be

$$\hat{\boldsymbol{\theta}}^{(m+1)} = argmax_{\boldsymbol{\theta}\in\Omega} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)},\mathbf{x}).$$

Part IV.3.2 EM Algorithm and Related: Expectation-Maximization (EM) Algorithm

- Starting with an initial estimate (guess) of $\boldsymbol{\theta}$, say, $\hat{\boldsymbol{\theta}}^{(0)}$, repeat the *E-Step* and *M-Step* and record the sequence of estimates $\{\hat{\boldsymbol{\theta}}^{(m)} : m = 1, 2, \dots\}$.

- Under some assumptions, the sequence converges in probability to the MLE $\widehat{\boldsymbol{\theta}}$ as $m \to \infty$.

- The EM algorithm works to improve $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)}, \mathbf{x})$. This implies improvements to $L(\boldsymbol{\theta}|\mathbf{x})$: $L(\hat{\boldsymbol{\theta}}^{(m+1)}|\mathbf{x}) \geq L(\hat{\boldsymbol{\theta}}^{(m)}|\mathbf{x})$.

EM algorithm has a broad range of applications!

... and a lot of variants, eg MCEM Algorithm, ES Algorithm, ...

Example. Consider a mixture of normal distributions: $X = (1 - W)Y_1 + WY_2$ with $Y_j \sim N(\mu_j, \sigma_j^2)$ for $j = 1, 2$ and $W \sim B(1, \epsilon)$. Suppose the observations on a random sample $\mathbf{X}' = (X_1, \ldots, X_n)$ from the mixture distn are available.

- Apply the EM algorithm to estimate the parameter $\boldsymbol{\theta}' = (\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \epsilon)$
- Similar to Ex6.6.8, generate data with $n = 1000$ and $\boldsymbol{\theta}' = (100, 15^2, 200, 10^2, .8)$, and then evaluate the MLE of $\boldsymbol{\theta}'$ by EM algorithm.

Solution. Let $f_j(y)$ be the pdf of Y_j. The log-likelihood functn with the observed data is

$$l(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^{n} \log\left[(1-\epsilon)f_1(x_i) + \epsilon f_2(x_i)\right].$$

With observations $\mathbf{w}'$ on a random sample $W_1, \ldots, W_n$ from $W \sim B(1, \epsilon)$, the log-likelihood functn with the complete data is

$$l_C(\boldsymbol{\theta}|\mathbf{x}, \mathbf{w}) = \sum_{i=1}^{n} \left\{\left[(1-w_i)\log f_1(x_i) + w_i \log f_2(x_i)\right] + \left[w_i \log \epsilon + (1-w_i)\log(1-\epsilon)\right]\right\}.$$

Provided the estimate at the mth stage is $\boldsymbol{\theta}^{(m)}$,

E-Step. $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}, \mathbf{x}) = l_C(\boldsymbol{\theta}|\mathbf{x}, \mathbf{w})\Big|_{\mathbf{w}=\widehat{\mathbf{w}}^{(m)}}$ with

$$\hat{w}_i^{(m)} = E(W_i|\boldsymbol{\theta}^{(m)}, \mathbf{x}) = \frac{\epsilon^{(m)} f_2(x_i)^{(m)}}{(1-\epsilon^{(m)}) f_1(x_i)^{(m)} + \epsilon^{(m)} f_2(x_i)^{(m)}}.$$

M-Step. Maximizing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}, \mathbf{x})$ wrt $\boldsymbol{\theta}$ yields

$$\mu_1^{(m+1)} = \frac{\sum_{i=1}^n (1-\hat{w}_i^{(m)}) x_i}{\sum_{i=1}^n (1-\hat{w}_i^{(m)})}, \quad \sigma_1^{2\,(m+1)} = \frac{\sum_{i=1}^n (1-\hat{w}_i^{(m)})(x_i - \mu_1^{(m+1)})^2}{\sum_{i=1}^n (1-\hat{w}_i^{(m)})},$$

$$\mu_2^{(m+1)} = \frac{\sum_{i=1}^n \hat{w}_i^{(m)} x_i}{\sum_{i=1}^n \hat{w}_i^{(m)}}, \quad \sigma_2^{2\,(m+1)} = \frac{\sum_{i=1}^n \hat{w}_i^{(m)} (x_i - \mu_2^{(m+1)})^2}{\sum_{i=1}^n \hat{w}_i^{(m)}},$$

and $\epsilon^{(m+1)} = \sum_{i=1}^n \hat{w}_i^{(m)} / n$.

What to study next?

Part I. Introduction
Part II. Epidemiologic Concepts and Designs
Part III. Clinical Trials

Part IV. Analytic Epidemiology

Part IV.1 Incomplete Data Analysis (*supplementary*)

Part IV.2 Some Other Important Topics (Chp 8 - 18, Koepsell and Weiss, 2003)

Part IV.3 Selected Widely-Used Algorithms
IV.3.1 Bootstrap and Related
IV.3.2 EM Algorithm and Related

Part IV.4 Lifetime Data Analysis
IV.4.1 Introduction
IV.4.2 Parametric Inference
IV.4.3 Nonparametric Inference: Estimation
IV.4.4 Nonparametric Inference: Testing
IV.4.5 Semiparametric Inference