What to do today (Oct 28, 2020)?

1. *Introduction*

2. *Probability and Distribution (Chp 1-3)*

3. Essential Topics in Mathematical Statistics

3.1 Elementary Statistical Inferences (Chp 4)

- *3.1.1 Sampling and Statistics*
- *3.1.2 Confidence Interval*
- *3.1.3 Order Statistics*
- *3.1.4 Hypothesis Testing*
- **3.1.5 Statistical Simulation and Bootstrap**

3.2 Consistency and Limiting Distributions (Chp 5)

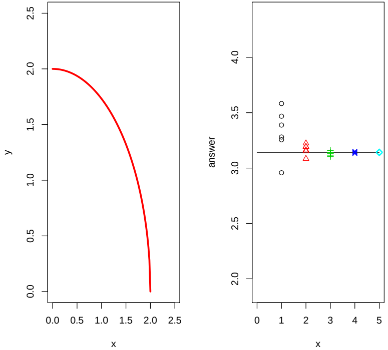
3.3 Maximum Likelihood Methods (Chp 6)

3.1.5 Statistical Simulation and Bootstrap: Monte Carlo Integration

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

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

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



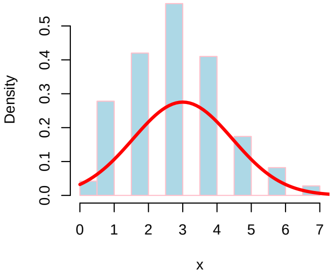
(a) y=(4−x^2)^.5
(b) true answer=pi

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

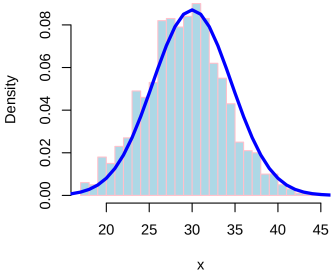
3.1.5 Statistical Simulation and Bootstrap: Simulation Example 1

To verify the normal approximation to binomial distn:

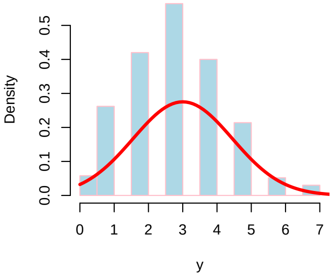
```
xtmp=matrix(ifelse(runif(1000*10)<.3,1,0),ncol=10)
x=apply(xtmp,1,sum)
hist(x, freq=FALSE, ...)
y=rbinom(n=1000,size=10,prob=0.3)
hist(y, freq=FALSE, breaks=11, ...)

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, col=4,
      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, col=4,
      lwd=3, add=TRUE)
```



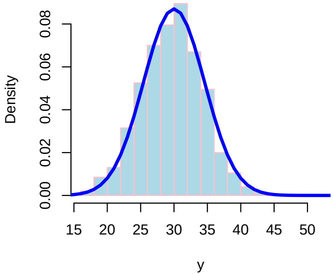
(a1). X ~ B(10,0.3)



(a2). X ~ B(100,0.3)



(b1). Y ~ B(10,0.3)

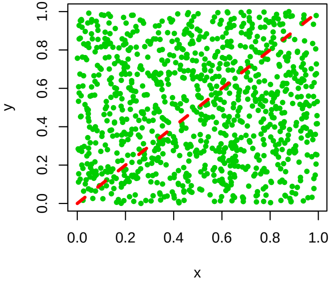


(b2). Y ~ B(100,0.3)

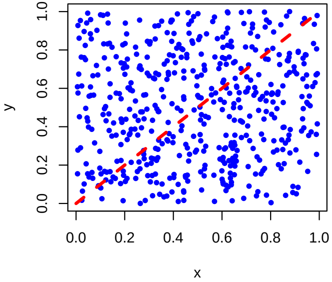
3.1.5 Statistical Simulation and Bootstrap: Simulation Example 2

What can data mssing result in?

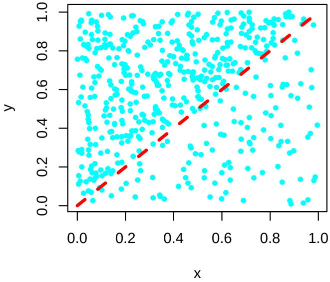
```
x=runif(1000);
y=runif(1000);
w=ifelse(x<y,1,0); sum(w)/1000

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);
```



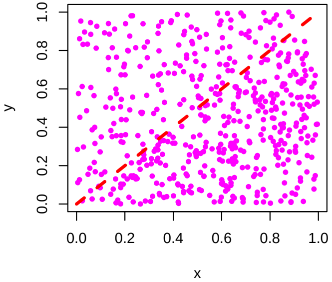
(a).all obs: rate of 'x<y'=.495



(b).random missing 50%: rate of 'x<y'=.486



(c).nonrandom missing 50%: rate of 'x<y'=.789



(d).nonrandom missing 50%: rate of 'x<y'=.337

3.1.5 Statistical Simulation and Bootstrap: Preparation for Bootstrap

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

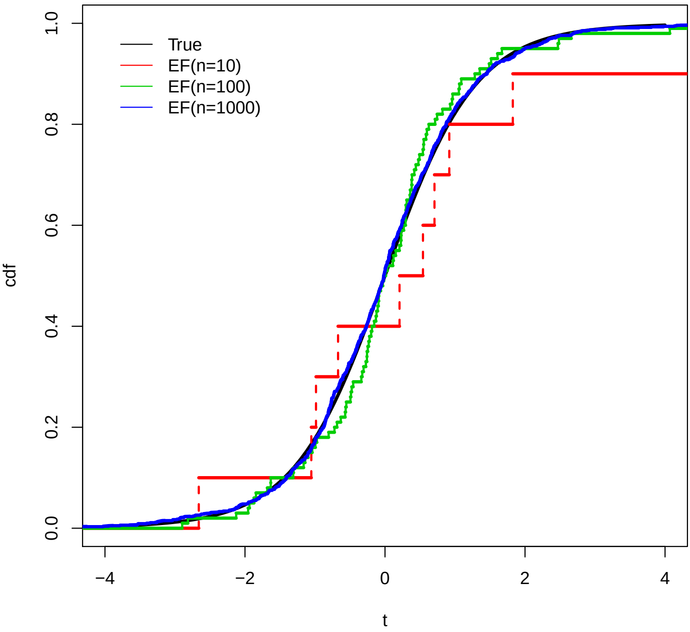
Definition. The following is the **empirical function** with the random sample:

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

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 $F(\cdot)$.



$T \sim t(6)$.

3.1.5 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 $\left\{\widehat{\theta}_b^* : b = 1, \ldots, B\right\}$.
- Calculate
  $$s_{\widehat{\theta}^*}^2 = \sum_{b=1}^{B} (\widehat{\theta}_b^* - \bar{\widehat{\theta}}^*)^2 / (B - 1)$$
  with $\bar{\widehat{\theta}}^* = \sum_{b=1}^{B} \widehat{\theta}_b^* / B$.
- Use $s_{\widehat{\theta}^*}^2$ to estimate $Var(\widehat{\theta})$.

3.1.5 Statistical Simulation and Bootstrap: Bootstrap

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

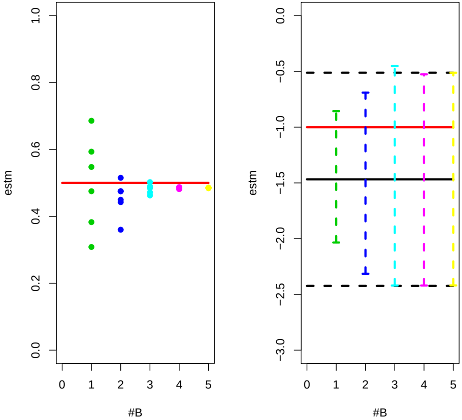
- How to obtain an interval estimator of a population parameter θ based on a point estmator $\widehat{\theta}(X_1, \ldots, X_n)$?

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 $k = 1, \ldots, 5$.

What will we study next class?

1. *Introduction*

2. *Probability and Distribution (Chp 1-3)*

3. **Essential Topics in Mathematical Statistics (Chp 4-6)**

   - *3.1 Elementary Statistical Inferences (Chp 4)*
   - **3.2 Consistency and Limiting Distributions (Chp 5)**
     - **3.2.1 Convergence in Probability**
     - **3.2.2 Convergence in Distribution**
   - *3.3 Maximum Likelihood Methods (Chp 6)*

4. *Further Topics, Selected from Chp 7-11*