Kyle McGraw

ACM 157 Set 2

1. $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{\#1}{N}$ $\quad x_i \in \{0,1\}$

a)

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(x_i^2 - 2x_i\mu + \mu^2\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_i - \frac{1}{N}\sum_{i=1}^{N} 2x_i\mu + \frac{1}{N}\sum_{i=1}^{N}\mu^2 \qquad x_i^2 = x_i$$

$$= \mu - 2\frac{1}{N}N\mu\sum_{i=1}^{N} x_i + \frac{1}{N}\mu^2\sum_{i=1}^{N} 1$$

$$= \mu - 2\mu^2 + \mu^2$$

$$= \mu - \mu^2$$

$$= \mu(1-\mu)$$

b)

$$S^2 = \hat{\sigma}_n^2\frac{Nn-n}{Nn-N} = \left(1-\frac{1}{N}\right)\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\overline{x}_n\right)^2$$

$$= \left(1-\frac{1}{N}\right)\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i^2 - 2x_i\overline{x}_n + \overline{x}_n^2\right)$$

$$= \left(1-\frac{1}{N}\right)\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i - 2\overline{x}_n\sum_{i=1}^{n} x_i + \overline{x}_n^2\sum_{i=1}^{n} 1\right)$$

$$= \left(1-\frac{1}{N}\right)\frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i - 2\overline{x}_n\frac{1}{n}\sum_{i=1}^{n} x_i + \overline{x}_n^2\frac{1}{n}n\right)$$

$$= \left(1-\frac{1}{N}\right)\frac{n}{n-1}\left(\overline{x}_n - 2\overline{x}_n^2 + \overline{x}_n^2\right)$$

$$= \left(1-\frac{1}{N}\right)\frac{n}{n-1}\left(\overline{x}_n - \overline{x}_n^2\right)$$

$$= \left(1-\frac{1}{N}\right)\frac{n}{n-1}\overline{x}_n\left(1-\overline{x}_n\right)$$

c) $N=300 \quad n=90 \quad \#1 = 70 \quad \overline{x}_n = 70/90 \quad 95\% = 100(1-\alpha)\% \quad \alpha = 0.05$

$$I = \overline{x}_n \pm z_{1-\frac{\alpha}{2}}\,se\left[\overline{x}_n\right]$$

$$= 70/90 \pm z_{1-\frac{0.05}{2}}\frac{s}{\sqrt{n}}\sqrt{\left(1-\frac{n-1}{N-1}\right)}$$

$$= 70/90 \pm z_{0.975}\sqrt{\left(1-\frac{1}{N}\right)\frac{n}{n-1}\overline{x}_n\left(1-\overline{x}_n\right)}\frac{1}{\sqrt{n}}\sqrt{\left(1-\frac{n-1}{N-1}\right)}$$

$$= 70/90 \pm 1.96\sqrt{\left(1-\frac{1}{300}\right)\frac{90}{89}\,70/90\left(1-70/90\right)}\frac{1}{\sqrt{90}}\sqrt{\left(1-\frac{89}{299}\right)}$$

$$= 70/90 \pm 0.0723$$

$$[0.706,\ 0.850]$$

2. 95% interval length ≤ 0.04

$2 \cdot z_{1-\frac{\alpha}{2}} se[\bar{X}_n] \leq 0.04$

$2 \cdot 1.96 \cdot \frac{S}{\sqrt{n}} \leq 0.04$ ignore population correction

$2 \cdot 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leq 0.04$ $S \approx \sigma$

$2 \cdot 1.96 \cdot \frac{\sqrt{0.5(1-0.5)}}{\sqrt{n}} \leq 0.04$ $\sigma = \sqrt{p(1-p)} = \sqrt{0.5(1-0.5)}$ worst-case variance

$n \geq 2401$

4. $X_1, \ldots, X_n$ uniform on $[0, \theta]$ $\mu = \frac{\theta}{2}$ $\sigma^2 = \frac{\theta^2}{12}$ $\mathbb{V}[\bar{X}_n] = \frac{\sigma^2}{n}$

a) $\hat{\theta} = 2\bar{X}_n$

$bias[\hat{\theta}] = \mathbb{E}[\hat{\theta}] - \theta = \mathbb{E}[2\bar{X}_n] - \theta = 2\mathbb{E}[\bar{X}_n] - \theta$
$= 2\frac{\theta}{2} - \theta = \theta - \theta = 0$

$se[\hat{\theta}] = \sqrt{\mathbb{V}[\hat{\theta}]} = \sqrt{\mathbb{V}[2\bar{X}_n]} = \sqrt{4\mathbb{V}(\bar{X}_n)} = \sqrt{4\frac{\sigma^2}{n}}$

$= \sqrt{4\frac{\theta^2}{12n}} = \theta\sqrt{\frac{1}{3n}}$

$MSE[\hat{\theta}] = bias[\theta]^2 + se[\hat{\theta}]^2 = \frac{\theta^2}{3n}$

b) $\hat{\theta} = X_{(n)} = \max\{X_1, \ldots, X_n\}$

CDF: $F(x) = P(X_{(n)} \leq x) = P(X_1 \leq x)P(X_2 \leq x) \cdot \ldots \cdot P(X_n \leq x)$
$= \frac{x}{\theta} \cdot \frac{x}{\theta} \cdot \ldots \cdot \frac{x}{\theta} = \left(\frac{x}{\theta}\right)^n$

$f(x) = \frac{nx^{n-1}}{\theta^n}$

$\mathbb{E}[\hat{\theta}] = \int_0^\theta x\, dF(x) = \int_0^\theta x f(x) dx = \int_0^\theta n\frac{x^n}{\theta^n} dx = \frac{n}{n+1}\frac{x^{n+1}}{\theta^n}\Big|_0^\theta$
$= \frac{n}{n+1}\frac{\theta^{n+1}}{\theta^n} = \frac{n}{n+1}\theta$

$bias[\hat{\theta}] = \mathbb{E}[\hat{\theta}] - \theta = \frac{n}{n+1}\theta - \theta = -\frac{1}{n+1}\theta$

$\mathbb{E}[\hat{\theta}^2] = \int_0^\theta x^2 dF(x) = \int_0^\theta x^2 f(x) dx = \int_0^\theta n\frac{x^{n+1}}{\theta^n} dx = \frac{n}{n+2}\frac{x^{n+2}}{\theta^n}\Big|_0^\theta$
$= \frac{n}{n+2}\frac{\theta^{n+2}}{\theta^n} = \frac{n}{n+2}\theta^2$

$\mathbb{V}[\hat{\theta}] = \mathbb{E}[\hat{\theta}^2] - \mathbb{E}[\hat{\theta}]^2 = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 = \frac{n(n+1)^2}{(n+2)(n+1)^2}\theta^2 - \frac{n^2(n+2)}{(n+2)(n+1)^2}\theta^2$
$= \frac{1}{(n+2)(n+1)^2}\theta^2\left(n^3 + 2n^2 + n - n^3 - 2n^2\right) = \frac{n}{(n+2)(n+1)^2}\theta^2$

$se[\hat{\theta}] = \sqrt{\mathbb{V}[\hat{\theta}]} = \theta\sqrt{\frac{n}{(n+2)(n+1)^2}}$

$MSE[\hat{\theta}] = bias[\theta]^2 + se[\hat{\theta}]^2 = \left(-\frac{1}{n+1}\theta\right)^2 + \frac{n}{(n+2)(n+1)^2}\theta^2$
$= \frac{1}{(n+1)^2}\theta^2 + \frac{n}{(n+2)(n+1)^2}\theta^2 = \frac{2n+2}{(n+2)(n+1)^2}\theta^2 = \frac{2}{(n+2)(n+1)}\theta^2$

c)

We can see that $X_{(n)}$ is more efficient at $n>2$ because:

$X_{(n)}$ MSE $<$ $2\bar{X}_n$ MSE

$$\frac{2}{(n+2)(n+1)}\theta^2 \leq \frac{\theta^2}{3n}$$

$$6n \leq (n+2)(n+1)$$

$$n^2+3n+2-6n \geq 0$$

$$n^2-3n+2 \geq 0$$

$$(n-2)(n-1) \geq 0$$

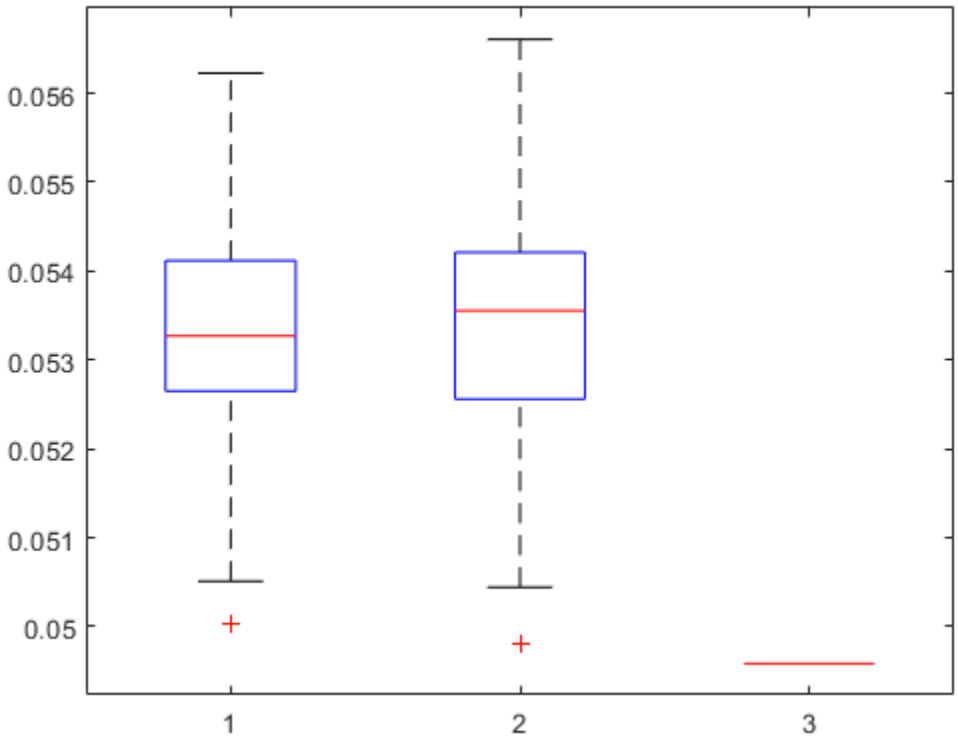
equal at $n=1,2$

greater at $n>2$

```
A = importdata("birth.txt");
A(A(:,1) == 999, :) = [];
bwt = A(:,1) * 0.0283495;
mu = mean(bwt);
n=100;
X = datasample(bwt,n,'Replace',false);
sampleMu = mean(X);
s = std(bwt);
N = length(bwt);
se = s/sqrt(n)*sqrt(1-(n-1)/(N-1));
disp([mu sampleMu se])
% (a) mu=3.3899 sampleMu=3.3750 se=3.3899

bootSe = bootB(X,N,n);
disp(bootSe)
% (b) se=0.0507

bootSeBF = bootC(X,N,n);
disp(bootSeBF)
% (c) se=0.0523

bootBs = zeros(100,1);
bootCs = zeros(100,1);
for i = 1:100
    bootBs(i) = bootB(X,N,n);
    bootCs(i) = bootC(X,N,n);
end
boxplot([bootBs,bootCs,repmat(se, 100, 1)])

function bootSe = bootB(X,N,n)
    B = 10^3;
    Pboot = repmat(X,round(N/n),1);
    sMu = zeros(B,1);
    for b = 1:B
        sb = datasample(Pboot,n,'Replace',false);
        sMu(b) = mean(sb);
    end
    aveMu = mean(sMu);
    bootSe =  sqrt((1/B)*sum((sMu-aveMu).^2));
end

function bootSeBF = bootC(X,N,n)
    B = 10^3;
    k = floor(N/n);
    r = N-n*k;
    Pboot1 = repmat(X,k,1);
    Pboot2 = repmat(X,k+1,1);
    p = (1-r/n)*(1-r/(N-1));
    sMuBF = zeros(B,1);
    for b = 1:B
        if rand() < p
            sbBF = datasample(Pboot1,n,'Replace',false);
```

```
        else
            sbBF = datasample(Pboot2,n,'Replace',false);
        end
        sMuBF(b) = mean(sbBF);
    end
    aveMuBF = mean(sMuBF);
    bootSeBF =  sqrt((1/B)*sum((sMuBF-aveMuBF).^2));
end
```

```
    3.3899    3.4413    0.0496

    0.0532

    0.0531
```



Published with MATLAB® R2021a