Zhangsheng Lai
1002554

Statistics: Homework 2

6.3 Given $\hat{\theta} = 2\overline{X}_n$ and $X_1, \dots, X_n \sim \text{Uniform}(0, \theta)$,

$$\begin{aligned}
\mathsf{bias}(\hat{\theta}) &= \mathbb{E}(2\overline{X}_n) - \theta \\
&= 2n^{-1}\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) - \theta \\
&= 2n^{-1}\sum_{i=1}^{n}\mathbb{E}\left(X_i\right) - \theta \\
&= 2n^{-1}\frac{n\theta}{2} - \theta = 0 \\
\mathsf{se}(\hat{\theta})^2 &= \mathbb{V}(2\overline{X}_n) \\
&= 4\mathbb{V}(\overline{X}_n) \\
&= 4n^{-2}\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) \\
&= 4n^{-2}\sum_{i=1}^{n}\mathbb{V}\left(X_i\right) \\
&= 4n^{-2}\frac{n\theta^2}{12} = \frac{\theta^2}{3n} \\
\mathsf{MSE}(\hat{\theta}) &= \text{bias}(\hat{\theta})^2 + \text{se}(\hat{\theta})^2 = \frac{\theta^2}{3n}
\end{aligned}$$

7.2 For $X_1, \dots, X_n \sim \text{Bernoulli}(p)$ plug-in estimator for p is

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and the estimated standard error is given by

$$\hat{\mathsf{se}}_p = \sqrt{\mathbb{V}(\hat{p})} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

As the X_i's are iid, by Central Limit Theorem, $\hat{p}$ is asymptotically normal with mean p and variance $\hat{\mathsf{se}}_p^2$. Thus an approximate 90% confidence interval for p is $(\hat{p} - 1.645\mathsf{se}, \hat{p} + 1.645\mathsf{se})$.

For $X_1, \dots, X_n \sim \text{Bernoulli}(p)$ and $Y_1, \dots, Y_n \sim \text{Bernoulli}(q)$ plug-in estimator for $p - q$ is

$$\hat{p} - \hat{q} = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{m}\sum_{i=1}^{m} Y_i$$

with estimated standard error

$$\hat{\mathsf{se}}_{p-q} = \sqrt{\mathbb{V}(\hat{p} - \hat{q})} = \sqrt{\mathbb{V}(\hat{p}) + \mathbb{V}(\hat{q})} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{\hat{q}(1-\hat{q})}{m}}$$

Since the Y_i's are iid, by Central Limit Theorem $\hat{q}$ is asymptotically normal with mean q and variance $\hat{\mathsf{se}}_q^2$. The difference of two asymptotically normal random variables is asymptotically normal, thus $\widehat{p - q}$ is asymptotically normal with mean $p - q$ and variance se_{p-q}^2. An approximate 90% confidence interval is

$$(\hat{p} - \hat{q} - 1.645\mathsf{se}_{p-q}, \hat{p} - \hat{q} + 1.645\mathsf{se}_{p-q})$$

7.9 An estimate for $p_1 - p_2$ is $0.9 - 0.85 = 0.05$ with standard error

$$\sqrt{\frac{0.9(1-0.9)}{100} + \frac{0.85(1-0.85)}{100}} = 0.0466368953$$

with 80% and 90% confidence intervals given by

$$80\%: \quad (\hat{p} - \hat{q} - 1.282\mathsf{se}_{p-q}, \hat{p} - \hat{q} + 1.282\mathsf{se}_{p-q}) = (-0.0097885, 0.1097885)$$
$$90\%: \quad (\hat{p} - \hat{q} - 1.96\mathsf{se}_{p-q}, \hat{p} - \hat{q} + 1.96\mathsf{se}_{p-q}) = (-0.041408315, 0.141408315)$$

8.7 (a)

$$\begin{aligned}\mathbb{P}(\hat{\theta} \le k) &= \mathbb{P}(\max\{X_1, \dots, X_n\} \le k) \\ &= \prod_{i=1}^{n} \mathbb{P}(X_i \le k) \\ &= \left(\frac{k}{\theta}\right)^n\end{aligned}$$

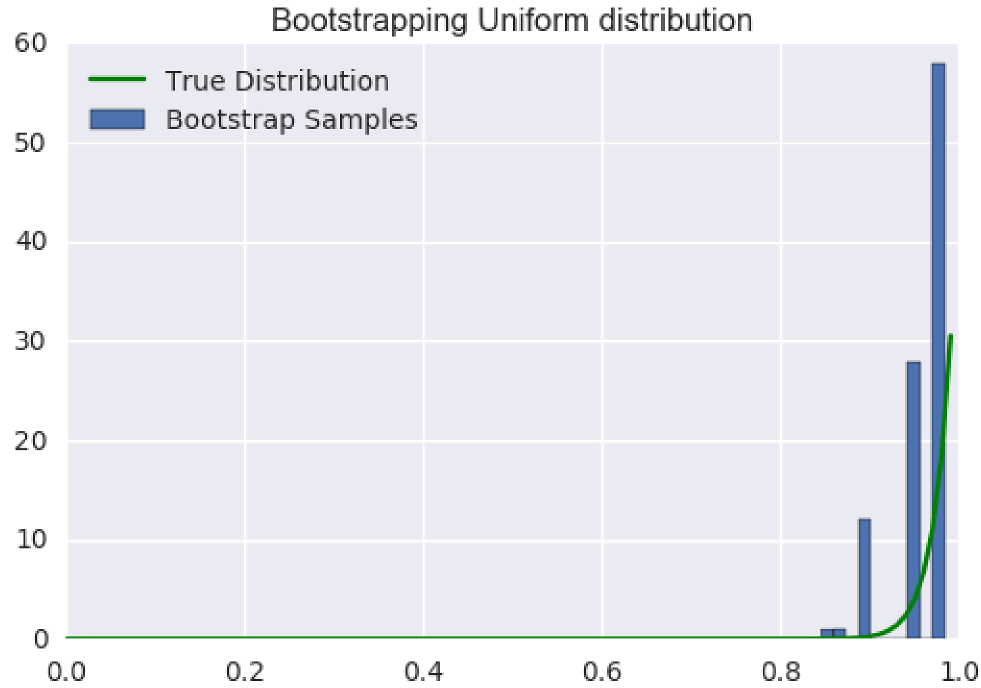

Figure 1: Comparison of the true distribution $\hat{\theta}$ to histrograms from bootstrap

(b) Let $\hat{\theta} = X_{max} = \max\{X_1, \dots, X_n\}$. Then

$$\begin{aligned}\mathbb{P}(\hat{\theta}^* = \hat{\theta}) &= 1 - \mathbb{P}(\hat{\theta}^* \ne \hat{\theta}) \\ &= 1 - \left(1 - \frac{1}{n}\right)^n\end{aligned}$$

The second equality holds as $\mathbb{P}(\hat{\theta}^* \ne \hat{\theta})$ denotes the probability that any random sampling with replacement of the n samples drawn has probability of $1 - 1/n$ not being x_{max} (which is fixed since a random sample of n has been drawn). As each sampling process is iid due to replacement, probability of them all not being x_{max} is $(1 - 1/n)^n$. Thus we have $\mathbb{P}(\hat{\theta}^* = \hat{\theta}) \to 0$ as $n \to \infty$ and $\mathbb{P}(\hat{\theta}^* = \hat{\theta}) = .632$ for $n = 50$.

9.2 (a) For $X_1, \dots, X_n \sim \text{Uniform}(a, b)$

$$\hat{\mu} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{\hat{b} + \hat{a}}{2}$$
$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\overline{X}_n - X_i)^2 = \frac{(\hat{b} - \hat{a})^2}{12}$$

thus

$$\begin{aligned}(\hat{b}-\hat{a})+(\hat{b}+\hat{a}) &= +\sqrt{12\hat{\sigma}^2}+2\hat{\mu}\\ \hat{b} &= \frac{1}{2}\left(\sqrt{12\hat{\sigma}^2}+2\hat{\mu}\right)\\ \hat{a} &= 2\hat{\mu}-\hat{b}\end{aligned}$$

the positive root is taken as $b-a>0$.

(b) Let $X_1,\ldots,X_n \sim \text{Uniform}(a,b)$, with $X_{max}=\max\{X_1,\ldots,X_n\}$. If $b<X_{max}$, then $f(X_j;a,b)=0$ for some j. Thus if $b\geq X_{max}$, then $f(X_i;a,b)=1/b-a$ for all i. In a similar fashion, letting $X_{min}=\{X_1,\ldots,X_n\}$, if $X_{min}<a$ we also have $f(X_j;a,b)=0$ for some j and $f(X_i;a,b)=1/b-a$ for all i if $X_{min}\geq a$. Therefore,

$$\mathcal{L}_n(a,b) := \begin{cases}0, & X_{min}<a \text{ or } X_{max}>b\\ \left(\frac{1}{b-a}\right)^n, & \text{otherwise}\end{cases}$$

$\mathcal{L}(a,b)$ strictly decreasing over $(-\infty, X_{min}]$ and $[X_{max},\infty)$, thus the maximum likelihood estimators $\hat{a}=X_{min}$ and $\hat{b}=X_{max}$.

(c) Let $\tau=\int x\,dF(x)$ be given, then from (b) we know that MLE's $\hat{a}$ and $\hat{b}$ are given by X_{min} and X_{max} respectively. Then the MLE of τ follows from MLE's $\hat{a}$ and $\hat{b}$. Thus MLE of τ is $(X_{min}+X_{max})/2$.

(d) By simulation, the MSE of $\hat{\tau}\approx 0.015$. Analytically, for the MSE of the nonparametric plugin estimator $\tilde{\tau}$ we have

$$\begin{aligned}\mathbb{E}(\hat{\theta}-\theta)^2 &= \mathbb{E}(\hat{\theta}^2-2\hat{\theta}\theta+\theta^2)\\ &= \mathbb{E}(\hat{\theta}^2)-2\theta\mathbb{E}(\hat{\theta})+\mathbb{E}(\theta^2)\\ &= \mathbb{E}(\hat{\theta}^2)-2\theta\mathbb{E}(\hat{\theta})+\mathbb{E}(\theta^2)\\ &= \mathbb{E}(\hat{\theta}^2)-4\end{aligned}$$

$$\begin{aligned}\mathbb{E}(\hat{\theta})^2 &= n^{-2}\left[\mathbb{E}\left(\sum_{i=1}^n X_i^2\right)+2\mathbb{E}\left(\sum_{i\neq j}X_iX_j\right)\right]\\ &= n^{-2}\left[n\mathbb{E}\left(X^2\right)+n(n-1)\mathbb{E}\left(X_iX_j\right)\right]\\ &= n^{-2}\left[n\mathbb{E}\left(X^2\right)+n(n-1)\mathbb{E}\left(X\right)^2\right]\\ &= 121/30\end{aligned}$$

using the substitution $\mathbb{E}(X^2)=2$, $\mathbb{E}(X)=13/2$ and $n=10$. The expectations are computed with $a=1, b=2$. Thus we have MSE to be $1/30$.

9.6 (a) The log-likelihood function is given to be

$$\begin{aligned}\ell_n(\theta) &= n\log\frac{1}{\sqrt{2\pi}}-\sum_{i=1}^n\frac{(X_i-\theta)^2}{2}\\ &= n\log\frac{1}{\sqrt{2\pi}}-\frac{n}{2}S^2-\frac{n}{2}(\overline{X}_n-\theta)^2\\ \text{with } \frac{\partial\ell_n}{\partial\theta} &= n(\overline{X}_n-\theta)\end{aligned}$$

thus the maximum of the log-likelihood function is when $\theta=\overline{X}_n$. Now, $X_1\sim N(\overline{X}_n,1)$

$$\begin{aligned}\psi=\mathbb{P}(X_1>0) &= \mathbb{P}\left(X_1-\overline{X}_n>-\overline{X}_n\right)\\ &= \mathbb{P}\left(Z>-\overline{X}_n\right)\end{aligned}$$

where the Z refers to the standard normal distribution.

(b) The distribution of $\hat{\psi}$ is normally distributed with $\hat{se}=\sqrt{1/I_n(\hat{\theta})}$

(c) Let $\tilde{\psi}=(1/n)\sum_i Y_i$, to show $\tilde{\psi}$ is consistent, we need to show that $\mathbb{P}(|\tilde{\psi}-\psi|>\epsilon)\to 0$ as $n\to\infty$.

$$\mathbb{P}(|\tilde{\psi}-\psi|>\epsilon)$$

(d)