Zhangsheng Lai
1002554

Statistics: Homework 2

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

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

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

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

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

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

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

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

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

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

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

8.7 (a)

$$\mathbb{P}(\hat{\theta} \leq k) = \mathbb{P}(\max\{X_1, \ldots, X_n\} \leq k)$$
$$= \prod_{i=1}^{n} \mathbb{P}(X_i \leq k)$$
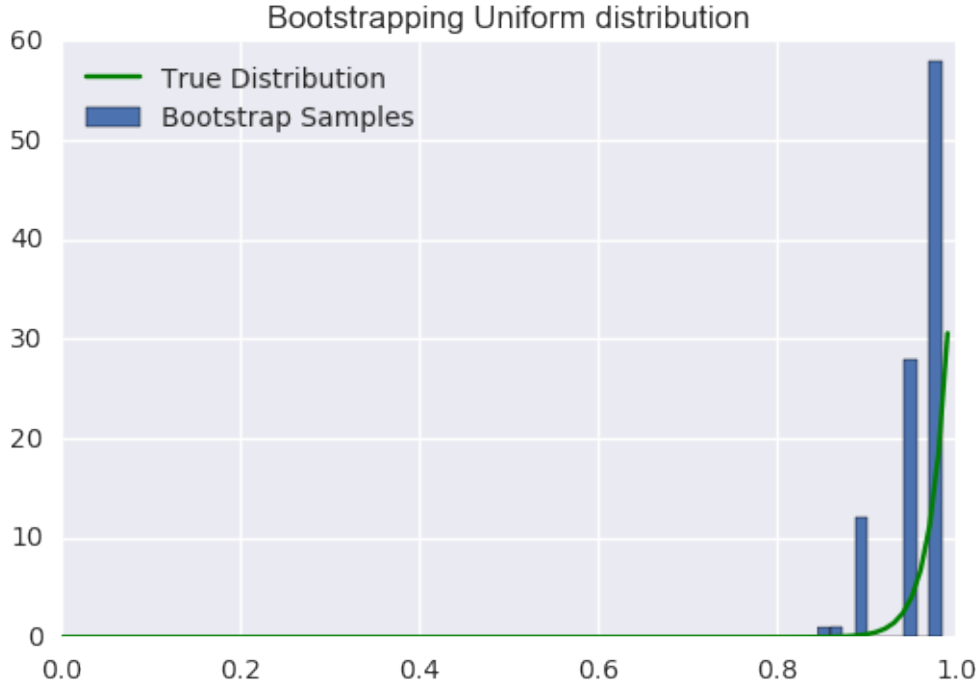$$= \left(\frac{k}{\theta}\right)^n$$

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

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

$$\mathbb{P}(\hat{\theta}^* = \hat{\theta}) = 1 - \mathbb{P}(\hat{\theta}^* \neq \hat{\theta})$$
$$= 1 - \left(1 - \frac{1}{n}\right)^n$$

The second equality holds as $\mathbb{P}(\hat{\theta}^* \neq \hat{\theta})$ 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

9.6