All of Statistics - Chapter 7 Solutions

Jan 20, 2020

1.

Note that

$$\mathbb{E}[\hat{F}_n(x)] = \mathbb{E}[I(X_1 \leq x)] = \mathbb{P}(X_1 \leq x) = F(x).$$

Moreover,

$$\mathbb{V}(\hat{F}_n(x)) = \mathbb{V}(I(X_1 \leq x))/n = F(x)(1 - F(x))/n.$$

By the bias-variance decomposition, the MSE converges to zero. Equivalently, we can say that $\hat{F}_n(x)$ converges to $F(x)$ in the L2 norm. Since Lp convergence implies convergence in probability, we are done.

Remark. For each x, $\hat{F}_n(x)$ is a random variable. The above proves only that each random variable $\hat{F}_n(x)$ converges in probability to the true value of the CDF $F(x)$. The Glivenko-Cantelli Theorem yields a much stronger result; it states that $\|\hat{F}_n - F\|_\infty$ converges almost surely (and hence in probability) to zero.

2.

Assumption. The Bernoulli random variables in the statement of the question are pairwise independent.

The plug-in estimator is $\hat{p} = \overline{X}_n$. The standard error is $\operatorname{se}(\hat{p})^2 = \mathbb{V}(X_1)/n = p(1-p)/n$. We can estimate the standard error by $\hat{\operatorname{se}}(\hat{p})^2 = \hat{p}(1-\hat{p})/n$. By the CLT,

$$\hat{p} \approx N(p, \operatorname{se}(\hat{p})^2) \approx N(\hat{p}, \hat{\operatorname{se}}(\hat{p})^2)$$

and hence an approximate 90% confidence interval is $\hat{p} \pm 1.64 \cdot \hat{\operatorname{se}}(\hat{p})$. The second part of this question is handled similarly.

3.

TODO (Computer Experiment)

4.

By the CLT

$$\sqrt{n}\left(\frac{\sum_i I(X_i \leq x)}{n} - \mathbb{E}\left[I(X_1 \leq x)\right]\right) \rightsquigarrow N(0, \mathbb{V}(I(X_1 \leq x))).$$

Equivalently,

$$\sqrt{n}\left(\hat{F}_n(x) - F(x)\right) \rightsquigarrow N(0, F(x)\left(1 - F(x)\right)).$$

Or, more conveniently,

$$\hat{F}_n(x) \approx N\left(F(x), \frac{F(x)\left(1 - F(x)\right)}{n}\right).$$

Remark. The closer (respectively, further) $F(x)$ is to 0.5, the more (respectively, less) variance there is in the empirical distribution evaluated at x.

5.

Without loss of generality, assume $x < y$. Then,

$$\begin{aligned}
\operatorname{Cov}(\hat{F}_n(x), \hat{F}_n(y)) &= \frac{1}{n^2}\operatorname{Cov}(\sum_i I(X_i \leq x), \sum_i I(X_i \leq y)) \\
&= \frac{1}{n^2}\sum_i \operatorname{Cov}(I(X_i \leq x), I(X_i \leq y)) = \frac{1}{n}\operatorname{Cov}(I(X_1 \leq x), I(X_1 \leq y)) \\
&= \frac{1}{n}(F(x) - F(x)F(y)) = \frac{1}{n}F(x)\left(1 - F(y)\right).
\end{aligned}$$

6.

By the results of the previous question,

$$\begin{aligned}
n \cdot \operatorname{se}(\hat{\theta})^2 &= n\mathbb{V}(\hat{F}_n(b) - \hat{F}_n(a)) \\
&= n\mathbb{V}(\hat{F}_n(b)) + n\mathbb{V}(\hat{F}_n(a)) - 2n\operatorname{Cov}(\hat{F}_n(b), \hat{F}_n(a)) \\
&= F(b)\left(1 - F(b)\right) + F(a)\left(1 - F(a)\right) - 2F(a)\left(1 - F(b)\right) \\
&= \left(F(b) - F(a)\right)\left[1 - \left(F(b) - F(a)\right)\right].
\end{aligned}$$

We can use the estimator

$$\hat{\operatorname{se}}(\hat{\theta})^2 = \frac{1}{n}\left(\hat{F}_n(b) - \hat{F}_n(a)\right)\left[1 - \left(\hat{F}_n(b) - \hat{F}_n(a)\right)\right].$$

An approximate $1 - \alpha$ confidence interval is $\hat{\theta} \pm z_{\alpha/2} \cdot \hat{\operatorname{se}}(\hat{\theta})$.

Remark. The closer $F(b) - F(a)$ is to zero or one, the smaller the standard error.

7.

TODO (Computer Experiment)

8.

TODO (Computer Experiment)

9.

This is an application of our findings in Question 2. In particular, we use the estimate $(90 - 85)/100 = 0.05$. A $1 - \alpha$ confidence interval for this estimate is $0.05 \pm z_{\alpha/2} \cdot \hat{\text{se}}$ where

$$\hat{\text{se}} = \sqrt{0.9\left(1 - 0.9\right)/100 + 0.85\left(1 - 0.85\right)/100} \approx 0.047.$$

The z-scores corresponding to 80% and 95% intervals are approximately 1.28 and 1.96.

10.

TODO (Computer Experiment)