Mathematical Statistics Solutions: HW 2

October 19, 2022

Exercise 1

Part (a)

By the Central Limit Theorem (note that $\mathsf{Var}(x) > 0$ since X is nondegenerate),

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mathbb{E}X)\right| > 3.3\right] = \mathbb{P}\left[|Z| > 3.3\mathsf{Var}(X)^{-1/2}\right]$$

where $Z \sim \mathcal{N}(0,1)$. Therefore, for any $\epsilon > 0$, there exists some $n_0(\epsilon) > 0$, such that for all $n > n_0(\epsilon)$, we have

$$\mathbb{P}\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mathbb{E}X)\right| > 3.3\right] > \mathbb{P}\left[|Z| > 3.3\mathsf{Var}(X)^{-1/2}\right] - \epsilon$$

Choosing $\epsilon = 0.5\mathbb{P}\left[|Z| > 3.3\mathsf{Var}(X)^{-1/2}\right]$ and $n_0 = n_0(\epsilon)$, we know that for any $n \geq n_0$,

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}X)\right| > 3.3n^{-1/2}\right] > 0.5\mathbb{P}\left[|Z| > 3.3\mathsf{Var}(X)^{-1/2}\right] =: c > 0.$$

Part (b)

From Part (a) we know that

$$\mathbb{P}\left[\left|\bar{X} - \mathbb{E}X\right| \leq 3.3n^{-1/2}\right] \leq 1 - c < 1.$$

Since $\bar{X}_j$'s are independent copies of $\bar{X}$, we have

$$\mathbb{P}\left[\left|\bar{X}_j - \mathbb{E}X\right| \leq 3.3n^{-1/2} \quad \forall j\right] = \prod_{j=1}^{p}\mathbb{P}\left[\left|\bar{X}_j - \mathbb{E}X\right| \leq 3.3n^{-1/2}\right] \leq (1-c)^p \to 0.$$

Exercise 2

Part (a)

For any $f \in \mathcal{F}$, the random variable $f(X)$ where $X \sim \text{Unif}([0,1])$ is bounded in $[-1,1]$, with $\mathbb{E}f(X) = \int_0^1 f(x)dx$. Hoeffding's inequality gives

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \int_0^1 f(x)dx\right| \geq \sqrt{\frac{2\log(2/\delta)}{n}}\right] \leq \delta;$$

therefore,

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \int_0^1 f(x)dx\right| < \sqrt{\frac{2\log(2/\delta)}{n}}\right] \geq 1 - \delta.$$

Part (b)

Let $S = \{X_1, X_2, \dots, X_n\}$, and

$$f(x) = \begin{cases} 1, & x \in S \\ -1, & x \notin S. \end{cases}$$

We compute

$$\left| \frac{1}{n} \sum_{i=1}^{n} f(X_i) - \int_0^1 f(x)dx \right| = |1 - (-1)| = 2.$$

Part (c)

Note that $f^*(x) \equiv 1$ is a minimizer of the population risk:

$$\mathcal{R}(f^*) = \mathbb{P}[f^*(x) \neq y] = 0.$$

Meanwhile, for any fixed n, consider

$$\hat{f}(x) = \begin{cases} 1, & x \in \{X_1, X_2, \dots, X_n\} \\ -1, & \text{otherwise}, \end{cases}$$

$\hat{f}$ is a minimizer of the empirical risk $\hat{R}$, since

$$\hat{R}(\hat{f}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{\hat{f}(X_i) \neq y_i\} = 0.$$

Now we compute the population risk of $\hat{f}$:

$$\mathcal{R}(\hat{f}) = \mathbb{P}[\hat{f}(X) \neq y] = 1,$$

therefore we get

$$\mathcal{R}(\hat{f}) - \mathcal{R}(f^*) = 1 - 0 = 1.$$

Exercise 3

Part (a)

We compute using linearity of expectation and that the X_i are i.i.d.:

$$\mathbb{E}\hat{F}_n(t) = \mathbb{E}\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \leq t\} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}[X_i \leq t] = \mathbb{P}[X_1 \leq t] = F(t).$$

Part (b)

Since the X_i are independent, we use the additivity and scaling of variance to compute:

$$\mathsf{Var}\hat{F}_n(t) = \frac{1}{n^2} \sum_{i=1}^{n} \mathsf{Var}[\mathbb{1}\{X_i \leq t\}] = \frac{1}{n}\mathsf{Var}[\mathbb{1}\{X_i \leq t\}] = \frac{1}{n}\mathbb{P}[X_i \leq t](1 - \mathbb{P}[X_i \leq t]) = \frac{1}{n}F(t)(1 - F(t)),$$

where at the end we have used that $\mathbb{1}\{X_i \leq t\}$ is a Bernoulli random variable with success probability $p = \mathbb{P}[X_i \leq t]$, whose variance is $p(1-p)$.

Part (c)

We begin by substituting in the previous parts:

$$\begin{aligned}
\mathbb{E}\int_{-\infty}^{\infty}(F(t)-\hat{F}_n(t))^2dt &= \mathbb{E}\int_{-\infty}^{\infty}(\hat{F}_n(t)-\mathbb{E}\hat{F}_n(t))^2dt && \text{(Part (a))}\\
&= \int_{-\infty}^{\infty}\mathbb{E}(\hat{F}_n(t)-\mathbb{E}\hat{F}_n(t))^2dt && \text{(exchange expectation and integral)}\\
&= \int_{-\infty}^{\infty}\mathsf{Var}[\hat{F}_n(t)]dt && \text{(definition of variance)}\\
&= \frac{1}{n}\int_{-\infty}^{\infty}F(t)(1-F(t))dt && \text{(Part (b))}\\
&\le \frac{1}{n}\int_{-\infty}^{\infty}\mathbb{P}[X\le t]\cdot\mathbb{P}[X\ge t]dt
\end{aligned}$$

and now divide into positive and negative t and pick out the smaller term in each case to conclude:

$$\begin{aligned}
&= \frac{1}{n}\left(\int_0^{\infty}\mathbb{P}[X\le t]\cdot\mathbb{P}[X\ge t]dt+\int_0^{\infty}\mathbb{P}[X\le -t]\cdot\mathbb{P}[X\ge -t]dt\right)\\
&\le \frac{1}{n}\left(\int_0^{\infty}\mathbb{P}[X\ge t]dt+\int_0^{\infty}\mathbb{P}[X\le -t]dt\right)\\
&= \frac{1}{n}\int_0^{\infty}\mathbb{P}[|X|\ge t]dt.
\end{aligned}$$

Part (d)

Following the hint, we have

$$\int_0^{\infty}\mathbb{P}[|X|\ge t]dt=\int_0^{\infty}\mathbb{E}\mathbb{1}\{|X|\ge t\}dt=\mathbb{E}\int_0^{\infty}\mathbb{1}\{|X|\ge t\}dt=\mathbb{E}|X|,$$

and substituting gives the result.

Exercise 4

Part (a)

Recall that Theorem 2.7 states that

$$\mathbb{P}\left[\sup_t|\hat{F}_n(t)-F(t)|\ge 2\sqrt{\frac{2\log(4n/\delta)}{n}}\right]\le\delta.$$

If $s=2\sqrt{\frac{2\log(4n/\delta)}{n}}$ then $\delta=4n\exp(-ns^2/8)$, so this is equivalent to

$$\mathbb{P}\left[\sup_t|\hat{F}_n(t)-F(t)|\ge s\right]\le 4n\exp(-s^2/8).$$

Since $4n\exp(-ns^2/8)\ge 2\exp(-2ns^2)$ for all $n\ge 1$ and $s\ge 0$, the DKW inequality is indeed stronger than Theorem 2.7.

Part (b)

Our error probability is bounded by evaluating the inequality with $s=\sqrt{3/n}$, which gives $2\exp(-2n\cdot 3/n)=2/e^6\approx 0.005$, indeed (much) smaller than 0.1.

Part (c)

We use the bound

$$\begin{aligned}
&\mathbb{P}\left[\text{there exists } t \text{ with } |\hat{F}_n(t) - \hat{G}_n(t)| \geq s\right] \\
&\leq \mathbb{P}\left[\text{there exists } t \text{ with } |\hat{F}_n(t) - F(t)| \geq \frac{s}{2} \text{ or } |\hat{G}_n(t) - F(t)| \geq \frac{s}{2}\right] \\
&\leq \mathbb{P}\left[\text{there exists } t \text{ with } |\hat{F}_n(t) - F(t)| \geq \frac{s}{2}\right] + \mathbb{P}\left[\text{there exists } t \text{ with } |\hat{G}_n(t) - F(t)| \geq \frac{s}{2}\right] \\
&\leq 4\exp(-ns^2/2),
\end{aligned}$$

the last part following if the X_i and Y_i are indeed drawn from the same distribution, using the DKW inequality twice.

Now, our error probability is bounded by evaluating this with $s = \sqrt{8/n}$, which gives $4\exp(-n/2 \cdot 8/n) = 4/e^4 \approx 0.07 < 0.1$.

Part (d)

Suppose that the X_i and Y_i are sorted in ascending order in the index i. Let us write $X_{n+1} = Y_{n+1} = \infty$ and $X_0 = Y_0 = -\infty$ for the sake of convenience, where we take this to mean that for any distribution function A, we have $A(X_0) = A(Y_0) = 0$ and $A(X_{n+1}) = A(Y_{n+1}) = 1$.

Suppose A is any distribution function, and $|A(t) - \hat{F}_n(t)| > \epsilon$ for some t. Let i be the greatest index such that $X_i \leq t$. Then, $\hat{F}_n(t) = \hat{F}_n(X_i)$, while $A(X_i) \leq A(t) \leq A(X_{i+1})$. Therefore, either $|A(X_i) - \hat{F}_n(X_i)| > \epsilon$ or $|A(X_{i+1}) - \hat{F}_n(X_i)| > \epsilon$. Taking $A = F$ or $A = \hat{G}_n$, we see that in either case it suffices to check the distances between the values of A and $\hat{F}_n$ at the $\leq 2(n+2)$ pairs of values (X_i, X_i) and (X_{i+1}, X_i).