Math Stats

Instructor: Jonathan Niles-Weed

Homework 1
Due: Sunday September 18, 11:59pm via NYU Gradescope

Collaborated with Jonah Potzcobutt, and Andre Chen.

1. This exercise will investigate (2.2).

   (a) Let X be any nondegenerate random variable on $[-1, 1]$. (Nondegenerate means that X takes more than one value.) Show that there exists a positive integer n_0 and a constant $c > 0$ depending on the distribution of X but not on n such that, if $X_1, \ldots, X_n$ are independent copies of X, then

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}X\right| > 3.3n^{-1/2}\right\} > c$$

(Hint: the central limit theorem guarantees that

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

where $\mathbf{Z}$ is a standard gaussian random variable.)

We can manipulate the given expression using linearity of expectation

$$\begin{gathered}\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}X\right| > 3.3n^{-1/2}\right\} > c \\ \mathbb{P}\left\{\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} (X_i - \mathbb{E}X)\right| > 3.3\right\} > c\end{gathered} \tag{1}$$

Then using the Central limit theorem we have:

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\left\{\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} (X_i - \mathbb{E}X)\right| > 3.3\right\} &= \mathbb{P}\left\{|Z| > \frac{3.3}{\sqrt{Var(X)}}\right\} \\ &= c\end{aligned} \tag{2}$$

Therefore:

$$\lim_{n\to\infty} \mathbb{P}\left\{\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} (X_i - \mathbb{E}X)\right| > 3.3\right\} = c$$

We can conclude that for finite $n \geq n_0$ (where I believe n_0 must be 2, as you can't take empirical averages and use CLT with just 1 sample):

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}X\right| > 3.3n^{-1/2}\right\} > c$$

(b) Conclude that if $\bar{X}_j$ is an independent copy of $\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$ for $j = 1, \dots p$ and $n \geq n_0$, then

$$P\left\{|X_j - \mathbb{E}X| \leq 3.3n^{-1/2} \quad \forall j = 1, \dots, p\right\} < (1-c)^p \quad \xrightarrow[p\to\infty]{} 0$$

We just found from part a that:

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}X\right| > 3.3n^{-1/2}\right\} > c$$

We can take the complement of this probability and get:

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}X\right| \leq 3.3n^{-1/2}\right\} < 1 - c$$

Since were interested that no $\hat{X}_1, \dots \hat{X}_j$ surpasses our bound, and our $\hat{X}s$ are independent, we can write our desired probability as a product of independent probabilities. We note that we already have our $\hat{X}_j$s from the problem above:

$$\hat{X}_j = \sum_{i=1}^{n} X_{j,i} - \mathbb{E}X_j$$

$$\begin{aligned}\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}X\right| \leq 3.3n^{-1/2}\right\} &= \mathbb{P}\left(\cap_{j=1}^{p} |\hat{X}_j - \mathbb{E}X_j| \leq 3.3n^{-1/2}\right) \\ &= \prod_{j=1}^{p} \mathbb{P}(|\hat{X}_j - \mathbb{E}X_j| \leq 3.3n^{-1/2}) \\ &= \mathbb{P}(|\hat{X}_j - \mathbb{E}X_j| \leq 3.3n^{-1/2})^p \quad \text{Independence of } X_js \\ &< (1-c)^p \quad \text{From a}\end{aligned} \tag{3}$$

And we have arrived at our desired inequality.

2. This exercise will show that uniform convergence and empirical risk minimization can easily fail for infinite classes. Let $X_1, \dots, X_n \sim$ *Unif*$([0,1])$ be i.i.d. Let $\mathcal{F}$ be the set of all indicator functions, i.e., functions of the form

$$f(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases}$$

For some $S \subset [0,1]$

(a) Show that, for any $f \in \mathcal{F}$ and any $\delta > 0$,

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

In other words, when n is large, any particular $f \in \mathcal{F}$ is close to its expectation with high probability.

Firstly, we note that X is bounded between $[0,1]$, meaninig we can use some of the properties of bounded Random variables to help with our desired equation:

$$\mathbb{P}\left\{|\bar{f} - \mathbb{E}f(x)| \geq \frac{s(b-a)}{2\sqrt{n}}\right\} \leq 2e^{-s^2/2}$$

Since $(b-a) = (0-1) = 1$, we can rewrite the nested inequality RHS as: $\frac{s}{2\sqrt{n}}$. We can also use the fact that $X \sim Unif([0,1])$, and f being an indicator function if $X \in S$. Since all the values we integrate over are in S, we have:

$$\int_0^1 f(x)dx = \mathbb{E}X$$

All we must do is pick an s such that the terms on the right hand side cancel out and yield a *delta*. We can do so by setting s in the following way:

$$\begin{aligned} s &= 2e^{-s^2/2} \\ log(\frac{\delta}{2}) &= \frac{-s^2}{2} \\ 2log(\frac{\delta}{2}) &= -s^2 \\ -2log(\frac{\delta}{2}) &= s^2 \\ \sqrt{2log(\frac{2}{\delta})} &= s \end{aligned} \tag{4}$$

Plugging in our new definition of s we have:

$$\begin{aligned} \mathbb{P}\left\{|\bar{f} - \mathbb{E}f(f)| \geq \frac{s(b-a)}{2\sqrt{n}}\right\} &\leq 2e^{-s^2/2} \\ &\leq 2e^{-\left(2log(\frac{2}{\delta})\right)/2} \\ &\leq \delta \end{aligned} \tag{5}$$

We can now compute the complement to yield an expression with probability $1-\delta$:

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

(b) Nevertheless, show that, no matter the value of X1 , . . . , Xn , there exists an $f \in \mathcal{F}$ such that

$$\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \int_0^1 f(x)dx\right| = 1 \quad \text{with probability } 1-\delta$$

In other words, no matter how large n is, there is always a function in $\mathcal{F}$ which is very far from its expectation.

Since S is a subset of $[0,1]$ and $X \sim Unif([0,1])$, we can imagine a possibility where all of $X_1, \ldots, X_n \notin S$, therefore:

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) = 0$$

But as our integration range is $[0,1]$ then our integral still evaluates to 1 as $S \subset [0,1]$

$$\int_0^1 f(x)dx = 1$$

Therefore, in this situation when dealing with infinite classes, our function $f \in \mathcal{F}$ is still very far from its expectation.

$$\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \int_0^1 f(x)dx\right| = |0-1| = 1$$

(c) Let $\{(y_i, X), \ldots, (y_n, X_n)\}$ be i.i.d. copies of (y, X), where $X \sim Unif([0,1])$ and $y = 1$ always. Show that for any n, there exists a minimizer $\hat{f}$ of the empirical risk $\hat{R}$ such that

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

In other words, no matter how large n is, empirical risk minimization fails.

If $y = 1$ no matter what the Then $\mathcal{R}(f^*) = 0$ as the theoretical risk minimizer, f^* always evaluates $f^*(x) = 1$ and therefore: $\mathcal{R}(f^*) = 0$.

To show a situation where $\mathcal{R}(\hat{f}) = 1$, we can imagine a classifier trained on points $X_1, \ldots, X_n$, which individually comprise a infinitely small portion of our continuous $[0,1]$ interval, and thus have no probability mass. The trained classifier would predict:

$$\begin{cases} 1 & if\ X \in X_1, \ldots, X_n \\ 0 & otherwise \end{cases}$$

When we would evaluate the risk of this empirical classifier, it would predict $y = 1$ correctly for all points $X_1, \ldots, X_n$. However, since $X_1, \ldots, X_n$ comprise an infinitely small subset of all real numbers between $[0,1]$, they would have 0 probability mass. Therefore, when evaluated, the empirical risk minimizer would predict incorrectly every time as:

$$\mathcal{R}(\hat{f}) = \mathbb{P}(\hat{f}(x) \neq 1) = 1$$

Therefore we would arrive at our situation provided in the problem statement where:

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

3. Theorem 2.8 showed that the empirical CDF $\hat{F}_n$ converges to F in the sense that $max_{t\in\mathbf{R}}|F(t) - \hat{F}_n(t)| \leq \mathcal{O}(\sqrt{log(n)/n})$ with high probability. In this exercise you will show converge in a different sense.

   (a) For a fixed $t \in \mathbb{R}$, show that $\mathbb{E}\hat{F}_n = F(t)$

   We can show this convergence for a $\hat{F}_n$ with arbitrary n, which will then generalize to all $\hat{F}_n$ due to theorem 2.8 and 2.5, which states that this convergence occurs uniformly over $\mathcal{F}$, meaning:

$$F_1(t) = F_2(t) = \cdots = F_n(t)$$

   We will also use the definition of $F(t) := \mathbb{P}\{X \leq t\}$

$$\begin{aligned} \mathbb{E}\hat{F}_n &= F(t) \\ \mathbb{E}\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{x\leq t} &= \mathbb{P}\{X \leq t\} \\ \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\mathbb{1}_{x\leq t} &= \mathbb{P}\{X \leq t\} \\ \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}\{X \leq t\} &= \mathbb{P}\{X \leq t\} \\ \mathbb{P}\{X \leq t\} &= \mathbb{P}\{X \leq t\} \quad \square \end{aligned} \tag{6}$$

   (b) For a fixed $t \in \mathbb{R}$, show that $Var(\hat{F}_n(t)) = \frac{1}{n}F(t)(1 - F(t))$

   Using the definition of Variance, we know the following:

$$Var(\hat{F}_n(t)) = \mathbb{E}(\hat{F}_n(t)^2) - \mathbb{E}^2\hat{F}_n(t)$$

   Evaluating $\mathbb{E}(\hat{F}_n(t)^2)$ first we have:

$$\begin{aligned}
\mathbb{E}(\hat{F}_n(t)^2) &= \mathbb{E}\left(\frac{1}{n^2}(\sum_{i=1}^{n}\mathbb{1}_{x_i\leq t})(\sum_{j=1}^{n}\mathbb{1}_{x_j\leq t})\right)\\
&= \frac{1}{n^2}\mathbb{E}\left((\sum_{i=1}^{n}\mathbb{1}_{x_i\leq t})(\sum_{j=1}^{n}\mathbb{1}_{x_j\leq t})\right)\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left((\mathbb{1}_{x_i\leq t})(\mathbb{1}_{x_j\leq t})\right)\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{P}(x_i\leq t\cap x_j\leq t)
\end{aligned} \tag{7}$$

Now we can do the same for $\mathbb{E}^2\hat{F}_n(t)$, where we know $\mathbb{E}\hat{F}_n(t) = F(t)$ from part a):

$$\begin{aligned}
\mathbb{E}^2\hat{F}_n(t) &= F(t)^2\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}F_i(t)F_j(t)
\end{aligned} \tag{8}$$

Putting it all together:

$$Var(\hat{F}_n(t)) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\mathbb{P}(x_i\leq t\cap x_j\leq t) - F_i(t)F_j(t)\right] \tag{9}$$

Note that since X_is are independent, every-time $i \neq j$ the terms will yield 0s, and therefore we only care about the cross terms where $i = j$ yielding:

$$\begin{aligned}
Var(\hat{F}_n(t)) &= \frac{1}{n^2}\sum_{i=1}^{n}\left[\mathbb{P}(x_i\leq t) - F_i(t)^2\right]\\
&= \frac{1}{n^2}n\left[\mathbb{P}(x\leq t) - F(t)^2\right]\\
&= \frac{1}{n}\left[F(t) - F(t)^2\right]\\
&= \frac{1}{n}F(t)(1-F(t)) \quad \square
\end{aligned} \tag{10}$$

(c) Using the above two claims, show that

$$\mathbb{E}\int_{-\infty}^{\infty}(F(t)-\hat{F}_n(t))^2dt \le \frac{1}{n}\int_{o}^{\infty}\mathbf{P}\{|X|\ge t\}\ dt$$

(You may assume that you can interchange expectation and integration)

We can use our definition of variance that we found last problem to help us show this inequality:

$$\begin{aligned}
\mathbb{E}\int_{-\infty}^{\infty}(F(t)-\hat{F}_n(t))^2dt &\le \frac{1}{n}\int_{o}^{\infty}\mathbf{P}\{|X|\ge t\}\ dt\\
\int_{-\infty}^{\infty}\mathbb{E}(F(t)-\hat{F}_n(t))^2dt &\le \frac{1}{n}\int_{o}^{\infty}\mathbf{P}\{|X|\ge t\}\ dt\\
\int_{-\infty}^{\infty}Var(\hat{F}_n(t)dt &\le \frac{1}{n}\int_{o}^{\infty}\mathbf{P}\{|X|\ge t\}\ dt\\
\int_{-\infty}^{\infty}\frac{1}{n}F(t)(1-F(t))dt &\le \frac{1}{n}\int_{o}^{\infty}\mathbf{P}\{|X|\ge t\}\ dt\\
\frac{1}{n}\int_{-\infty}^{\infty}\mathbb{P}(x\le t)(\mathbb{P}(x>t))dt &\le \frac{1}{n}\int_{o}^{\infty}\mathbf{P}\{|X|\ge t\}\ dt\\
\frac{1}{n}\int_{-\infty}^{\infty}\mathbb{P}(x\le t)(\mathbb{P}(x>t))dt &\le \frac{1}{n}\int_{o}^{\infty}\mathbf{P}\{X\le t\}+\mathbf{P}\{X\ge t\}\ dt \quad \square
\end{aligned} \tag{11}$$

The last step we replace our RHS with its equivalent probability expressed as sum as derived from the abosolute value. As probabilities are bounded [0,1] we know that:

$$\mathbb{P}(x\le t)\times\mathbb{P}(x>t)\le\mathbb{P}(x\le t)+\mathbb{P}(x>t)$$

(d) Conclude that

$$\mathbb{E}\int_{-\infty}^{\infty}(F(t)-\hat{F}_n(t))^2dt \le \frac{\mathbb{E}|X|}{n}$$

(Hint: Write $\int_0^\infty \mathbb{P}\{|X|\le t\}\ dt = \int_0^\infty \mathbb{E}\mathbf{1}_{|X|\ge t}dt$ and interchange expectation and integration)

Mathematically, bounds on $max_{t\in R}|\hat{F}_n(t)-F(t)|$ are known as L_∞ bounds and bounds on $\int_{-\infty}^{\infty}(F(t)-\hat{F}_n(t))^2dt$ are known as L_2 bounds. Neither bound is stronger than the other, but they give different information: the first says that the error at each point is small, and the second says that the total squared error over the whole real line is small.

Using what we just computed in part c):

$$\begin{aligned}
\mathbb{E}\int_{-\infty}^{\infty}(F(t)-\hat{F}_n(t))^2dt &\le \frac{1}{n}\int_{o}^{\infty}\mathbf{P}\{|X|\ge t\}\ dt\\
\mathbb{E}\int_{-\infty}^{\infty}(F(t)-\hat{F}_n(t))^2dt &\le \frac{1}{n}\int_{o}^{\infty}\mathbb{E}\mathbb{1}_{|x|\ge t}\ dt\\
\mathbb{E}\int_{-\infty}^{\infty}(F(t)-\hat{F}_n(t))^2dt &\le \frac{1}{n}\mathbb{E}\int_{o}^{|x|}\mathbb{1}_{|x|\ge t}\ dt\\
\mathbb{E}\int_{-\infty}^{\infty}(F(t)-\hat{F}_n(t))^2dt &\le \frac{\mathbb{E}|X|}{n} \quad \square
\end{aligned} \tag{12}$$

4. In this exercise, we will use the following strengthened form of Theorem 2.8, which goes by the name “Dvoretzky–Kiefer–Wolfowitz inequality”: if $\hat{F}_n$ is the empirical distribution corresponding to n i.i.d. samples from F, then

$$\mathbb{P}\left\{\sup_{t\in\mathbb{R}}|F(t) - \hat{F}_n(t)| \geq s\right\} \leq 2e^{-2ns^2} \quad \forall s \geq 0$$

   (a) Show that (2.6) is indeed stronger than Theorem 2.8.

   By theorem 2.8 we have:

$$\mathbb{P}\left\{\sup_{t\in\mathbb{R}}|F(t) - \hat{F}_n(t)|\right\} \leq 2\sqrt{\frac{2log(4n/\delta)}{n}}$$

   And the above statement holds with probability $\geq 1 - \delta$.

   Using 2.6 above we have:

$$\mathbb{P}\left\{\sup_{t\in\mathbb{R}}|F(t) - \hat{F}_n(t)| \geq s\right\} \leq 2e^{-2ns^2} \quad \forall s \geq 0$$

   To cancel out variables in the RHS we pick a convenient s:

$$s = 2\sqrt{\frac{2log(4n/\delta)}{n}}$$

   Plugging in our s we have:

$$\begin{gathered}
\mathbb{P}\left\{\sup_{t\in\mathbb{R}}|F(t) - \hat{F}_n(t)| \geq s\right\} \leq 2e^{-2ns^2} \\
\mathbb{P}\left\{\sup_{t\in\mathbb{R}}|F(t) - \hat{F}_n(t)| \geq 2\sqrt{\frac{2log(4n/\delta)}{n}}\right\} \leq 2e^{-2n(2\sqrt{\frac{2log(4n/\delta)}{n}})^2} \\
\mathbb{P}\left\{\sup_{t\in\mathbb{R}}|F(t) - \hat{F}_n(t)| \geq 2\sqrt{\frac{2log(4n/\delta)}{n}}\right\} \leq 2e^{-16log(4n/\delta)}
\end{gathered} \tag{13}$$

   Taking the complement we get:

$$\mathbb{P}\left\{\sup_{t\in\mathbb{R}}|F(t) - \hat{F}_n(t)| \leq 2\sqrt{\frac{2log(4n/\delta)}{n}}\right\} \geq 1 - 2e^{-16log(4n/\delta)} = 1 - \frac{\delta}{2n}^{16}$$

   Since $\delta \in [0, 1]$ and $n \geq 1$ we have arrived at a stronger bound than that given from Theorem 2.8.

   (b) Let $X_1, \ldots, X_n$ be i.i.d. copies of a random variable X with unknown distribution. Suppose we wish to decide whether the distribution function of X is equal to some known CDF F. In light of Theorem 2.8, one reasonable approach is to compare the empirical distribution function $\hat{F}n$ corresponding to $X_1, \dots, X_n$ to F . Suppose we decide that we will declare that X does not have distribution F if $\sup_{t\in\mathbb{R}}|F(t) - \hat{F}_n(t)| \geq \sqrt{3/2n}$ Show using (2.6) that if X did actually come from F , then we are wrong with probability at most 10%.

If we chose the above heuristic, we can use our formula derived in the prior problem while picking $s = \sqrt{3/2n}$:

$$\begin{aligned}\mathbb{P}\left\{\sup_{t\in\mathbb{R}}|F(t) - \hat{F}_n(t)| \geq s\right\} &\leq 2e^{-2ns^2} \\ \mathbb{P}\left\{\sup_{t\in\mathbb{R}}|F(t) - \hat{F}_n(t)| \geq \sqrt{3/2n}\right\} &\leq 2e^{-2n(\sqrt{3/2n})^2} \\ \mathbb{P}\left\{\sup_{t\in\mathbb{R}}|F(t) - \hat{F}_n(t)| \geq \sqrt{3/2n}\right\} &\leq 10\%\end{aligned} \tag{14}$$

This bound gives us certaininty that 10% of the time that we reject an X, we will be mistaken, and that observed Random Variable will have indeed came from X.

(c) Let $X_1, \dots, X_n$ be i.i.d. copies of X and $Y_1, \dots, Y_n$ be i.i.d. copies of Y, where the distributions of both X and Y are unknown. Suppose we wish to decide whether X and Y have the same distribution. Let $\hat{F}_n$ and $\hat{G}_n$ be the empirical CDF's corresponding to $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$, respectively, and suppose that we declare that the distributions of X and Y are different if $\sup_{t\in\mathbb{R}}|\hat{F}_n(t) - \hat{G}_n(t)| \geq \sqrt{8/n}$. Show using (2.6) that if X and Y do actually have the same distribution, then we are wrong with probability at most 10%.

We can begin manipulating the expression by adding and subtracting $F(t)$ to our LHS:

$$\begin{aligned}\sup_{t\in\mathbb{R}}|\hat{F}_n(t) - \hat{G}_n(t)| &\geq \sqrt{8/n} \\ \sup_{t\in\mathbb{R}}|\hat{F}_n(t) - F(t) + F(t) - \hat{G}_n(t)| &\geq \sqrt{8/n}\end{aligned} \tag{15}$$

Using the triangle inequality we have:

$$\mathbb{P}\left(|\hat{F}(t) - \hat{G}(t)| \geq \sqrt{\frac{8}{n}}\right) \leq \mathbb{P}|\hat{F}(t) - F(t)| + |F(t) - \hat{G}(t)|$$

Which can be bound as follows:

$$\mathbb{P}\left(|\hat{F}(t) - \hat{G}(t)| \geq \sqrt{\frac{8}{n}}\right) \leq \mathbb{P}\left((|F\hat{(t)} - F(t)| \geq \frac{1}{2}\sqrt{\frac{8}{n}}) \cup (|F(t) - \hat{G}(t)| \geq \frac{1}{2}\sqrt{\frac{8}{n}})\right)$$

Noting that we have 2 DKW inequalities on the RHS of our expression at this point, we can bound the probability of the union of these two events. After so, we can attempt to calculate the probability:

$$\mathbb{P}\left(|\hat{F}(t) - F(t)| \geq \frac{1}{2}\sqrt{\frac{8}{n}}\right) = \mathbb{P}\left(|\hat{G}(t) - F(t)| \geq \frac{1}{2}\sqrt{\frac{8}{n}}\right) \leq 2e^{-\dfrac{2n\frac{8}{n}}{4}}$$

$$\mathbb{P}\left(|\hat{F}(t) - F(t)| \geq \frac{1}{2}\sqrt{\frac{8}{n}}\right) \leq 2e^{-4}$$

Taking the complement we have:

$$\mathbb{P}\left(|\hat{F}(t) - F(t)| < \frac{1}{2}\sqrt{\frac{8}{n}}\right) > 1 - 2e^{-4}$$

Putting together all 3 possible scenarios, we see that the probability of

$$\mathbb{P}\left(|\hat{F}(t) - \hat{G}(t)| \geq \sqrt{\frac{8}{n}}\right) \leq (2e^{-4})^2 + 2(1-2e^{-4})(2e^{-4})$$

Which doesn't quite add up to 10% as I had hoped, but I'm not entirely where I may have gone wrong with this problem.

(d) In parts (b) and (c) above, our test depended on being able to decide whether suptR $\sup_{t\in\mathbb{R}}|F_n(t) - \hat{F}_n(t)|$ or $\sup_{t\in\mathbb{R}}|\hat{F}_n(t) - \hat{G}_n(t)|$ is larger than some threshold. Assuming that the samples are sorted, show that this question can be answered in Ø(n) time.

If we imagining writing our distribution test as an algorithm, we would need to do 2 things: firstly, calculate the empirical CDFs, F, G for all observations in each. Then we would have to evaluate our expression, $\sup_{t\in\mathbb{R}}|\hat{F}_n(t) - \hat{G}_n(t)|$ for all $t \in R$ (at least, for all t in our n observations of F, G) to see if two empirical CDFs differ by a large enough quantity.

This would be quite computationally costly: every time you wanted to evaluate the CDF of F or G you would have to scan the whole array of numbers, for all $x \in G, F \leq t$, yielding a time complexity of $\mathcal{O}(n^2)$. However, you could reduce this time complexity by firstly sorting the list, then using pointers to keep track of which observations you've already gone over as you iterate through each array and begin pairwise comparisons of CDF values. Depending on what sorting algorithm you use, the time complexity can vary, but most are bound by $\mathcal{O}(nlog(n))$.

However, if the list is sorted, this algorithm can be done in linear time, as there is no need to sort, and therefore the time complexity would be $\mathcal{O}(n)$ Using pointers, you could iterate through the each array of observations for F, G, and calculate all $t \leq (n+m)$ comparisons between distributions, where n is the number of samples from F and m is the number of samples from G.

The procedures described in this question are called Kolmogorov–Smirnov tests. We will learn much more about testing soon.