1. **[20 points] Constructing and interpreting confidence intervals:** Suppose that we have \( X_1, \ldots, X_n \sim \text{Ber}(p) \), where \( p \) is unknown.

- Use Hoeffding’s bound to construct a 95% confidence interval for \( p \).
- Use the CLT to construct an approximate 95% confidence interval for \( p \).
- Simulate the following setup: Repeat each of the following experiments 10000 times.
  Vary the sample size from 10 to 10000 on a logarithmic scale, and vary \( p \) from 0 to 1 on a linear scale. For each value of \((n, p)\) compute the coverage probability of the CLT interval and the Hoeffding interval, i.e. for each value of \((n, p)\) and for each of the 10000 experiments construct the interval and check if the true parameter is inside the constructed interval.
  Plot and summarize your findings. Particularly, do the two types of intervals have the advertised coverage and are they conservative?

2. **[15 points] Finding and Using Pivots:**

- Suppose that \( X_1, \ldots, X_n \sim \text{Exp}(\lambda) \), then show that

\[
Q(X_1, \ldots, X_n, \lambda) = \frac{2 \sum_{i=1}^{n} X_i}{\lambda}
\]

is a valid pivot. You can use the fact that the sum of exponentials has a Gamma distribution.

**Clarification:** As you might have guessed from the expression for \( Q \). We are using an unusual parametrization for the Exponential distribution here, in particular, for this question you can assume that the \( \text{Exp}(\lambda) \) distribution has pdf:

\[
p(x; \lambda) = \frac{1}{\lambda} \exp(-x/\lambda).
\]

Use this pivot to construct a \((1 - \alpha)\)-confidence interval for \( \lambda \).

- A location family is a collection of shifted (mean 0) distributions, i.e. for some fixed distribution \( F \) with mean 0, we define:

\[
P_\theta = \{ G : G(x) = F(x - \theta) \}.
\]

Show that \( \bar{X} - \theta \) is always a pivot for a location family.
3. **[20 points] Multiple Testing:** We conduct a multiple testing experiment and would like to test $d$ null hypotheses $H_{01}, \ldots, H_{0d}$. Throughout this question you should assume that we have access to the p-values for each hypothesis, and that these p-values satisfy the usual condition, i.e. under the null

$$P_0(p\text{-value} \leq t) \leq t,$$  

for $0 \leq t \leq 1$.

Our goal is to control the $k$-Family Wise Error Rate ($k$-FWER) – which is the probability of making $k$ or more false rejections (imagine $k > 1$).

In the exam you considered the following Bonferroni-type procedure:

- We reject each null for which the corresponding p-value $\leq k\alpha/d$.

and showed that it controls the $k$-FWER at $\alpha$.

The analogue of Holm’s procedure turns out to be the following:

- Sort the p-values so that $p(1) \leq p(2) \leq \ldots \leq p(d)$.
- For the first $k$ hypotheses we use the Bonferroni threshold, i.e. if $p(1) \leq k\alpha/d$ reject and move on. Else accept all hypotheses.
- If $p(2) \leq k\alpha/d$ reject and move on. Else accept $H_{0(2)}, \ldots, H_{0(d)}$.
  
  $\vdots$
  
- If $p(k) \leq k\alpha/d$ reject and move on. Else accept $H_{0(k)}, \ldots, H_{0(d)}$.
- After $k$ we start increasing the threshold. If $p(k+1) \leq k\alpha/(d-1)$, reject and move on. Else accept $H_{0(k+1)}, \ldots, H_{0(d)}$.
- If $p(k+2) \leq k\alpha/(d-2)$, reject and move on. Else accept $H_{0(k+2)}, \ldots, H_{0(d)}$.
  
  $\vdots$

Show that this procedure controls the $k$-FWER at $\alpha$. Compare this procedure to the Bonferroni controlling procedure above.

4. **[15 points] Bootstrap:** Suppose that we observe i.i.d. samples $X_1, \ldots, X_n$, and for an estimator $\hat{\theta}(X_1, \ldots, X_n)$ we consider using the empirical bootstrap to assess its variability. We have in lecture derived the mean and variance of the bootstrap distribution of $\hat{\theta^*}$ when $\hat{\theta}$ is the sample mean.

Suppose instead we considered a so-called “linear statistic”, where for some fixed, bounded function $f$:

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} f(X_i).$$

Compute the mean and variance of the bootstrap distribution of $\hat{\theta^*}$. 
5. **[15 points] Causal Inference 1:** Suppose we are in a randomized trial, i.e. we have \(n\) individuals sampled from some distribution and of these \(n\) individuals \(n/2\), (assume \(n\) is even for simplicity) are randomly assigned to receive treatment and the remaining \(n/2\) receive control. Assume that the outcomes are binary, and for each unit denote the observed outcomes as \(Y_{i}^{\text{obs}}\). Denote the units receiving treatment by \(T\) and those receiving control by \(C\). We want to estimate the average treatment effect:

\[
\tau = \mathbb{E}[Y(1) - Y(0)],
\]

and we use the natural estimator:

\[
\hat{\tau}_1 = \frac{2}{n} \sum_{i \in T} Y_{i}^{\text{obs}} - \frac{2}{n} \sum_{i \in C} Y_{i}^{\text{obs}}.
\]

Use Hoeffding’s inequality to give a bound on \(|\hat{\tau}_1 - \tau|\) that holds with probability at least \(1 - \delta\). Provide a reason as to why Hoeffding is a reasonable tool here.

6. **[15 points] Causal Inference 2:** Now we suppose that we are in an observational set-up, but we have no unmeasured confounding, i.e. for some covariates \(X\) we have that \(W \perp (Y(1), Y(0))|X\). We assume that in this setting we observe \(n\) i.i.d. samples of the form \((X_i, W_i, Y_{i}^{\text{obs}})\).

To begin with let us suppose that the propensity scores \(\mathbb{P}(W = 1|X)\) are known, i.e. we know the probability of receiving treatment for each unit in the population. This is not an unreasonable assumption in settings where say a doctor prescribes treatment say after conducting some tests (i.e. measuring some covariates). As we noted in lecture, when \(\pi(X)\) is not known we need to estimate it using regression.

To simplify things we will assume that there are constants \(\pi_{\text{min}} > 0\) and \(\pi_{\text{max}} < 1\) such that,

\[
\pi_{\text{min}} \leq \mathbb{P}(W = 1|X) \leq \pi_{\text{max}} \quad \forall \ x.
\]

A natural estimator then is the Horvitz-Thompson/Inverse-Propensity Score estimator:

\[
\hat{\tau}_2 = \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{Y_{i}^{\text{obs}}W_i}{\pi(X_i)} - \frac{Y_{i}^{\text{obs}}(1 - W_i)}{1 - \pi(X_i)} \right].
\]

Use Hoeffding’s inequality to give a bound on \(|\hat{\tau}_2 - \tau|\) that holds with probability at least \(1 - \delta\). You need to reason about why we needed the \(\pi_{\text{min}}, \pi_{\text{max}}\) assumption, as well as why Hoeffding is a reasonable tool here.