Name: ______________________________

Instructions: This is a take-home exam. No collaboration with other people is allowed. You may consult books or online resources, but you must cite them. In particular, you might want to look up Chernoff's bound and exponential tilting. You must write all statistical algorithms yourself. You may use the built-in optimization routine `nlm` to check your work, but you must be the author of all routines used for the solutions you turn in. If you have a question about whether a built-in function is acceptable, email me. If you find an error or ambiguity in the exam, please email me.

Your final exam is due before 5pm on Friday December 16th.

To submit your final exam, send an electronic (PDF) copy of your exam to `forrest.crawford@yale.edu`, including derivations, code, output, and images. Put "BIS557A final exam" in the subject of your email.

Academic honesty: Sign on the line below when you are finished with the exam.

I have neither given nor received aid in this exam.

______________________________ ______________

Signature Date

Consider a randomized clinical trial of a surgical procedure to repair the medial meniscus (MM) after a sports injury. Patients were randomly assigned to the new surgical procedure, or given a "sham" incision but no actual treatment. Both groups received standard non-invasive physical therapy. Patients were unaware of their treatment status. Their outcome 6 months after the surgery is binary, based on pain level and mobility.

To formalize the inferential problem, consider a population of n units. The binary treatment of i is z_i. Suppose $z_i \sim \text{Bernoulli}(\pi_i)$ independently where $0 < \pi_i < 1$ is the probability of receiving treatment. Therefore the joint treatment probability (randomization distribution) is

$$f(\mathbf{z}|\boldsymbol{\pi}) = \prod_{i=1}^{n} \pi_i^{z_i}(1-\pi_i)^{1-z_i}$$

where $\mathbf{z} = (z_1, \dots, z_n)$ and $\boldsymbol{\pi} = (\pi_1, \dots, \pi_n)$. Let y_i be the binary outcome observed for subject i. Consider a test statistic that measures a weighted difference in oucomes for treated and untreated individuals,

$$T(\mathbf{z}; \mathbf{y}, \boldsymbol{\pi}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i z_i}{\pi_i} - \frac{y_i(1-z_i)}{1-\pi_i}\right) \tag{1}$$

where $\mathbf{y} = (y_1, \dots, y_n)$. Intuitively, if $T(\mathbf{z}; \mathbf{y}, \boldsymbol{\pi})$ is positive, we estimate that the treatment helps patients; if it is negative, we estimate that the treatment hurts.

To proceed, we need a notion of potential (or counterfactual) outcomes under different treatment assignments. Let y_{i1} be the potential outcome of subject i under treatment $z_i = 1$ and let y_{i0} be the potential outcome of subject i under no treatment $z_i = 0$. Then $y_i = z_i y_{i1} + (1-z_i)y_{i0}$ is the outcome actually observed in the trial for subject i. These potential outcomes are regarded as fixed; the only randomness in the trial comes from the Bernoulli-distributed treatment assignments.

After the trial has been conducted, researchers would like to know about the uncertainty in the estimate. One way to do this is to construct a model that links the observed outcomes y_i to the counterfactual outcomes y_{i0} and y_{i1}. Consider a model of counterfactual outcomes in which $\tau \in \mathbb{R}$ is an additive treatment effect: $y_{i1} = y_{i0} + \tau$. Given τ, we can reconstruct y_{i1} and y_{i0} from the realized (y_i, z_i) as follows. Let

$$y_{i1}(\tau) = \begin{cases} y_i + \tau & \text{if } z_i = 0 \\ y_i & \text{if } z_i = 1 \end{cases} \qquad y_{i0}(\tau) = \begin{cases} y_i & \text{if } z_i = 0 \\ y_i - \tau & \text{if } z_i = 1 \end{cases}$$

where z_i refers to the treatment assigned in the actual trial. This construction allows us to ask questions (conditional on τ) about what might have happened if the treatments were different from the realized assignments $\mathbf{z}$ in the trial. Let $\mathbf{w} = (w_1, \dots, w_n) \sim f(\mathbf{w}|\boldsymbol{\pi})$ be a vector of binary treatments possibly different from those assigned in the trial. Consider the counterfactual test statistic

$$T(\mathbf{w}; \mathbf{y}(\tau), \boldsymbol{\pi}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_{i1}(\tau) w_i}{\pi_i} - \frac{y_{i0}(\tau)(1-w_i)}{1-\pi_i}\right)$$

where $\mathbf{y}(\tau)$ is the collection of all counterfactual outcomes $y_{i1}(\tau)$ and $y_{i0}(\tau)$.

One of the first things that researchers would like to do is test a "sharp" hypothesis of the form $T(\mathbf{z}; \mathbf{y}, \boldsymbol{\pi}) = \tau$ for a particular value of τ. For example, setting $\tau = 0$ gives the null hypothesis of no treatment effect. To test this type of hypothesis, consider the one-sided p-value

$$p(\tau) = \Pr\big(T(\mathbf{w}; \mathbf{y}(\tau), \boldsymbol{\pi}) \geq t\big)$$

where $\mathbf{w} = (w_1, \dots, w_n) \sim f(\mathbf{w}|\boldsymbol{\pi})$, and $t = T(\mathbf{z}; \mathbf{y}, \boldsymbol{\pi})$ is the value of the test statistic (1) actually observed in the trial under the realized treatment assignments $\mathbf{z}$. Intuitively, $p(\tau)$ is the probability of observing a test statistic as or more extreme than the one actually observed in the trial, conditional on τ, under the randomization distribution $f(\mathbf{w}|\boldsymbol{\pi})$. Unfortunately $p(\tau)$ usually cannot be expressed in closed form and researchers must resort to approximations or Monte Carlo simulation methods to compute it.

Problem 1 Theory

a. Show that in general, $T(\mathbf{z}; \mathbf{y}, \boldsymbol{\pi})$ is an unbiased estimator of the average treatment effect

$$\text{ATE} = \frac{1}{n} \sum_{i=1}^{n} (y_{i1} - y_{i0}).$$

Under the additive model $y_{i1} = y_{i0} + \tau$, show that $T(\mathbf{z}; \mathbf{y}, \boldsymbol{\pi})$ is an unbiased estimator of τ.

b. Under the additive model with τ fixed, compute the mean and variance of $T(\mathbf{z}; \mathbf{y}, \boldsymbol{\pi})$ with respect to the randomization distribution $f(\mathbf{z}|\boldsymbol{\pi})$. Give an approximate expression for $p(\tau)$ using the normal CDF $\Phi(\cdot)$. Call this value $\hat{p}_0(\tau)$.

c. Find a and b_i so that $T(\mathbf{z}; \mathbf{y}(\tau), \boldsymbol{\pi})$ can be written as $a + \sum_{i=1}^{n} b_i z_i$. Suppose $\tau < t$ and use Chernoff's bound to show that for all $\lambda > 0$, we have the upper bound $p(\tau) \leq \tilde{p}(\tau|\lambda)$ where

$$\tilde{p}(\tau|\lambda) = e^{\lambda(a-t)} \prod_{i=1}^{n} (\pi_i e^{\lambda b_i} + 1 - \pi_i).$$

Note that a and $b_1, \dots, b_n$ are functions of τ. Write these expressions down; you will use them below.

d. To find the tightest upper bound for $p(\tau)$, devise an algorithm to find the value of λ that minimizes $\tilde{p}(\tau|\lambda)$.

e. Let M be a positive integer and consider the Monte Carlo estimator

$$\hat{p}_1(\tau) = \frac{1}{M} \sum_{j=1}^{M} \mathbb{1} \left\{ T(\mathbf{w}_j; \mathbf{y}(\tau), \boldsymbol{\pi}) \geq t \right\}$$

where $\mathbf{w}_j = (w_{j1}, \dots, w_{jn}) \sim f(\mathbf{w}|\boldsymbol{\pi})$. Show that $\hat{p}_1(\tau)$ is an unbiased estimator of $p(\tau)$. What is the variance of $\hat{p}_1(\tau)$?

f. Let $\lambda \in \mathbb{R}$ and define a "tilted" version of the treatment probability π_i for subject i as

$$\Pr(z_i = 1|\lambda) = \frac{e^{\lambda b_i} \pi_i}{e^{\lambda b_i} \pi_i + 1 - \pi_i}$$

where λb_i is called the "tilting parameter". Exponential tilting is a common method for shifting the properties of a density or probability mass function, while ensuring that the shifted distribution is still in the same parametric family as the original. Note that the denominator is the moment generating function of z_i. Show that the tilted randomization distribution becomes

$$f(\mathbf{z}|\boldsymbol{\pi}, \lambda) = \prod_{i=1}^{n} \frac{e^{\lambda b_i z_i} \pi_i^{z_i} (1 - \pi_i)^{1 - z_i}}{e^{\lambda b_i} \pi_i + 1 - \pi_i}.$$

g. Consider the Monte Carlo estimator

$$\hat{p}_2(\tau|\lambda) = \frac{e^{\lambda a} \prod_{i=1}^{n} (e^{\lambda b_i} \pi_i + 1 - \pi_i)}{M} \sum_{j=1}^{M} \mathbb{1} \left\{ T(\mathbf{w}_j; \mathbf{y}(\tau), \boldsymbol{\pi}) \geq t \right\} e^{-\lambda T(\mathbf{w}_j; \mathbf{y}(\tau), \boldsymbol{\pi})}$$

where $\mathbf{w}_j = (w_{j1}, \ldots, w_{jn}) \sim f(\mathbf{w}|\boldsymbol{\pi}, \lambda)$ is a draw from the tilted randomization distribution. Show that $\hat{p}_2(\tau|\lambda)$ is an unbiased estimator of $p(\tau)$.

h. Show that

$$\mathrm{Var}[\hat{p}_2(\tau|\lambda)] \leq \frac{e^{2\lambda(a-t)} \prod_{i=1}^{n} (e^{\lambda b_i} \pi_i + 1 - \pi_i)^2 - p^2}{M}$$

and devise an algorithm to find the value of λ that minimizes this variance bound.

i. Give a sufficient condition for $\mathrm{Var}[\hat{p}_2(\tau|\lambda)] < \mathrm{Var}[\hat{p}_1(\tau)]$. Now we have a recipe for finding the value of λ that gives a way of computing $p(\tau)$ that is more efficient than the naïve Monte Carlo estimator.

Problem 2 Application

These problems refer to the results of the MM randomized trial in the file `mmtear.csv`. Realized treatments $\mathbf{z}$ and observed outcomes $\mathbf{y}$ are given. In this trial $\pi_i = 1/2$ for all subjects i.

a. Compute the test statistic $T(\mathbf{z}; \mathbf{y}, \boldsymbol{\pi})$ for this trial.

b. Test the sharp null hypothesis that $\tau = 0$ under the additive model using the estimator $\hat{p}_2(\tau = 0|\lambda^*)$ where λ^* minimizes $\tilde{p}(\tau|\lambda)$. Report your computed estimate of $p(\tau = 0)$. Choose M large enough that the Monte Carlo error is small. Explain how you chose M.

c. Compute and plot $\hat{p}_0(\tau)$, $\hat{p}_1(\tau)$, $\tilde{p}(\tau|\lambda)$, and $\hat{p}_2(\tau|\lambda)$ as a function of $\tau \in [-5, 5]$ with $M = 1000$. For each value of τ, use the λ that minimizes $\tilde{p}(\tau|\lambda)$.

d. Let $0 < \alpha < 1$ and consider the confidence set

$$CI(\alpha) = \{\tau : \ \alpha/2 < \Pr(T(\mathbf{z}; \mathbf{y}(\tau), \boldsymbol{\pi}) \geq t) < 1 - \alpha/2\}$$

Use $\hat{p}_0(\tau)$, $\hat{p}_1(\tau)$, $\tilde{p}(\tau|\lambda)$, and $\hat{p}_2(\tau|\lambda)$ to compute approximate 50% and 95% confidence set for τ. Report your results in a table. What do you find?

Problem 3 Extra Credit

Randomized trials are often conducted by fixing the number of treated individuals at m of n total subjects, instead of using Bernoulli allocation as described above. Suppose the randomization distribution is uniform:

$$f(\mathbf{z}) = \mathbb{1} \left\{ \sum_{i=1}^{n} z_i = m \right\} \binom{n}{m}^{-1}.$$

This randomization distribution is more difficult to work with because the individual treatments z_i are not independent. Consider the test statistic

$$T(\mathbf{z}; \mathbf{y}) = \sum_{i=1}^{n} \left(\frac{y_i z_i}{m} - \frac{y_i (1 - z_i)}{n - m} \right).$$

Fix τ and let $y_{i0}(\tau)$ and $y_{i1}(\tau)$ be potential outcomes under the additive model. Consider the counterfactual test statistic

$$T(\mathbf{w}; \mathbf{y}(\tau)) = \sum_{i=1}^{n} \left(\frac{y_{i1}(\tau) w_i}{m} - \frac{y_{i0}(\tau)(1 - w_i)}{n - m} \right)$$

where $\mathbf{w} \sim f(\mathbf{w})$. Again we wish to compute

$$p(\tau) = \Pr\left(T(\mathbf{w}; \mathbf{y}(\tau)) \geq t\right)$$

where t is the value of the test statistic actually observed in the trial.

a. Derive a naïve monte Carlo estimator $\hat{p}_1(\tau)$ of $p(\tau)$ under the randomization distribution $f(\mathbf{z})$. Derive its expectation and variance.

b. Exhibit a family of *non-uniform* randomization distributions $g(\mathbf{z}|\boldsymbol{\phi})$ where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)$ is a vector of parameters that control how often each subject is assigned to treatment. Your randomization distribution must have the property that $g(\mathbf{z}|\boldsymbol{\phi}) = 0$ whenever $\sum_{i=1}^{n} z_i \neq m$. Explicitly state the probability mass function $g(\mathbf{z}|\boldsymbol{\phi})$ and describe how you would draw a realization $\mathbf{z} \sim g(\mathbf{z}|\boldsymbol{\phi})$.

c. Use your specification of $g(\mathbf{z}|\boldsymbol{\phi})$ to derive an estimator of $p(\tau)$ that has smaller variance than $\hat{p}_1(\tau)$. How do you choose $\boldsymbol{\phi}$ for a given τ?