STATS 200: Homework 4

Due Friday, October 28, at 5PM

1. **Permutation tests for paired samples.** Let $D_1, \ldots, D_n \overset{IID}{\sim} f$ for a probability density function f on $\mathbb{R}$, and consider a test of the null hypothesis

$$H_0 : f \text{ is symmetric about } 0$$

(against some alternative, say $H_1 : f$ is symmetric about a value $\mu > 0$) that rejects for large values of a test statistic $T = T(D_1, \ldots, D_n)$.

(a) Describe the distribution of T conditional on $|D_1|, \ldots, |D_n|$, under H_0. (What values can T take conditional on $|D_1|, \ldots, |D_n|$, and with what probabilities? You may assume no value of D_i is exactly equal to 0.)

(b) Explain how computer simulation can be used to approximate the conditional distribution of T in part (a) (even if n is very large), and hence to perform a level-α test of H_0 based on T.

(c) If each D_i is the difference $D_i = X_i - Y_i$ of values from two paired samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, explain how your test in part (b) may be interpreted as a "permutation" test. Generalize your procedure to the following setting: Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ be random paired samples of "objects" represented in some data space $\mathcal{X}$, and consider the null hypothesis H_0 that $(X_1, Y_1), \ldots, (X_n, Y_n)$ are IID pairs such that (X_i, Y_i) has the same (joint) distribution as (Y_i, X_i). For a test statistic $T = T(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$, how can you use simulation to determine the rejection threshold of a test of H_0 based on T?

2. **Local power of the sign test.** Let $X_1, \ldots, X_n \overset{IID}{\sim} f$ for a PDF f. For the problem of testing

$$H_0 : f \text{ has median } 0$$
$$H_1 : f \text{ has median greater than } 0,$$

recall from Homework 3 that the sign statistic S is the number of positive X_i's, and the asymptotic sign test rejects H_0 when $S > \frac{n}{2} + \sqrt{\frac{n}{4}} z(\alpha)$.

In this problem, we'll study the power of this test against the specific alternative $\mathcal{N}(\frac{h}{\sqrt{n}}, 1)$, for a fixed constant $h > 0$ (say $h = 1$ or $h = 2$) and large n.

(a) If $X \sim \mathcal{N}(\frac{h}{\sqrt{n}}, 1)$, show that

$$\mathbb{P}[X > 0] = \Phi\left(\frac{h}{\sqrt{n}}\right)$$

where Φ is the CDF of the standard normal distribution $\mathcal{N}(0,1)$. Applying a first-order Taylor expansion of Φ around 0, show that for large n

$$\mathbb{P}[X > 0] \approx \frac{1}{2} + \frac{h}{\sqrt{2\pi n}}.$$

(b) Let $X_1, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(\frac{h}{\sqrt{n}}, 1)$. In this case, show that $\sqrt{\frac{4}{n}}(S - \frac{n}{2})$ has an approximate normal distribution that does not depend on n (but depends on h)—what is the mean and variance of this normal distribution? (A heuristic argument using the CLT here is fine—don't worry about formalizing convergence in distribution.) Using this result, derive an approximate formula for the power of the sign test against this alternative $\mathcal{N}(\frac{h}{\sqrt{n}}, 1)$, in terms of $z(\alpha)$, h, and the CDF Φ.

(c) Recall the simulations from Problem 5 of Homework 3, where you computed the power of the sign test at level $\alpha = 0.05$ against the alternatives $\mathcal{N}(\mu, 1)$ for $n = 100$ and $\mu = 0.1, 0.2, 0.3, 0.4$. Compute the values of h corresponding to these four values of μ, and evaluate your power formula in part (b) for these values of h. How closely does your power formula match your simulated powers from Homework 3? (If you did not do Homework 3, please use the simulated power values obtained by a classmate or on the Homework 3 solutions on the course webpage.)

(d) Suppose you wish to design an experiment such that you want the power of this sign test (at level $\alpha = 0.05$) against the alternative $\mathcal{N}(0.2, 1)$ to be 0.9. Using your formula from part (b), what is the approximate sample size n that you would need?

3. **Effect of a confounding factor.** To study the effectiveness of a drug that claims to lower blood cholesterol level, we design a simple experiment with n subjects in a control group and n (different) subjects in a treatment group. We administer the drug to the treatment group and a placebo to the control group, measure the cholesterol levels of all subjects at the end of the study, and look at whether cholesterol levels are lower in the treatment group than in the control. Let $X_1, \ldots, X_n$ be the cholesterol levels in the control group and $Y_1, \ldots, Y_n$ be those in the treatment group, and let

$$T = \frac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{n} + \frac{1}{n}}}$$

be the standard two-sample t-statistic where S_p^2 is the pooled variance.

Assume, throughout this problem, that the drug in fact *is not* effective and has the exact same effect as the placebo. However, suppose there are two types of subjects, high-risk and low-risk. (Approximately half of the human population is high-risk and half is low-risk; assume that we cannot directly observe whether a person is high-risk or low-risk.) The cholesterol level for high-risk subjects is distributed as $\mathcal{N}(\mu_H, \sigma^2)$, and for low-risk subjects as $\mathcal{N}(\mu_L, \sigma^2)$.

(a) A carefully-designed study randomly selects subjects for the two groups so that each subject selected for either group is (independently) with probability 1/2 high-risk and probability 1/2 low-risk. Explain why, in this case, $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ are IID from a common distribution. What are $\mathbb{E}[X_i]$ and $\mathrm{Var}[X_i]$?

(b) Explain (using the CLT and Slutsky's lemma) why, when n is large, T is approximately distributed as $\mathcal{N}(0, 1)$, and hence a test that rejects for $T > z(\alpha)$ is approximately level-α for large n.

(c) A poorly-designed study fails to properly randomize the treatment and control groups, so that each subject selected for the control group is with probability p high-risk and probability $1-p$ low-risk, and each subject selected for the treatment group is with probability q high-risk and probability $1-q$ low-risk. In this case, what are $\mathbb{E}[X_i]$, $\mathrm{Var}[X_i]$, $\mathbb{E}[Y_i]$, and $\mathrm{Var}[Y_i]$?

(d) In the setting of part (c), show that S_p^2 converges in probability to a constant $c \in \mathbb{R}$ as $n \to \infty$, and provide a formula for c. Show that T is approximately normally distributed, and provide formulas for the mean and variance of this normal. Is the rejection probability $\mathbb{P}[T > z(\alpha)]$ necessarily close to α? Discuss briefly how this probability depends on the values μ_H, μ_L, σ^2, p and q.

4. **Improving upon Bonferroni for independent tests.**
(a) Let $P_1, \ldots, P_n$ be the p-values from n different hypothesis tests. Suppose that the tests are performed using independent sets of data, and in fact all of the null hypotheses are true, so $P_1, \ldots, P_n \overset{IID}{\sim} \text{Uniform}(0, 1)$. Show that for any $t \in (0, 1)$,

$$\mathbb{P}\left[\min_{i=1}^{n} P_i \leq t\right] = 1 - (1-t)^n.$$

(b) Under the setting of part (a), if we perform all tests at significance level $1 - (1-\alpha)^{1/n}$ (that is, we reject a null hypothesis if its p-value is less than this level), show that the probability of (falsely) rejecting any of the n null hypotheses is exactly α. Is this procedure more or less powerful than the Bonferroni procedure (of performing all tests at level α/n)?

(c) Suppose, now, that all of the p-values $P_1, \ldots, P_n$ are still independent, but not necessarily all of the null hypotheses are true. (So the p-values corresponding to the *true* null hypotheses are still IID and distributed as Uniform$(0, 1)$.) If we perform all tests at significance level $1 - (1-\alpha)^{1/n}$, does this procedure control the familywise error rate (FWER) at level α? (Explain why, or show a counterexample.)