Stats 210A, Fall 2025
Homework 9

Due on: Wednesday, Nov. 5

Instructions: See the standing homework instructions on the course web page

Problem 1 (Fisher's exact test)**.** Suppose $X_i \sim \text{Binom}(n_i, \pi_i)$ independently for $i = 0, 1$ and $\pi_0, \pi_1 \in (0, 1)$. Consider testing $H_0 : \ \pi_1 \leq \pi_0$ vs. $H_1 : \ \pi_1 > \pi_0$.

(a) A natural object of inference in this model is the *odds ratio*:

$$\rho = \frac{\pi_1/(1-\pi_1)}{\pi_0/(1-\pi_0)}.$$

Write the model in exponential family form with $\theta = \log \rho$ as one of the natural parameters, and reframe H_0 as an equivalent hypothesis about θ.

(b) The hypergeometric distribution $\text{Hypergeom}(N, K, n)$ describes the probability distribution for sampling n binary values without replacement from a finite population of N binary values, of which K are equal to 1 and the other $N - K$ are equal to 0. If X is the number of successes in the subsample, its probability mass function is

$$p_{N,K,n}(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}, \quad \text{for } x = \max\{0, n+K-N\}, \ldots, \min\{K, n\}.$$

Find the UMPU level-α test of H_0 in part (a), show that the test statistic has a hypergeometric distribution for appropriate N, K, n, and describe how to find the cutoffs $c(u), \gamma(u)$.

(c) Find the conditional distribution of your test statistic for general θ. Note you do not need to find a closed-form expression for the normalizing constant.

(d) Suppose $n_0 = n_1 = 40$, $X_0 = 18$ and $X_1 = 7$. Give a 95% confidence interval for the odds ratio ρ by numerically inverting the two-sided, equal-tailed, conditional test of $H_0 : \ \rho = \rho_0$ vs. $H_1 : \ \rho \neq \rho_0$. Don't randomize the interval, just return the conservative non-randomized interval. (Hint: it is equivalent to set up the problem in terms of θ, and may be a little easier to think about that way.)

Moral: Fisher's exact test is almost certainly the most important non-Gaussian example of a UMPU test with nuisance parameters, and has been used in countless clinical trials and observational studies. For example, we might give n_1 cardiac disease patients a new drug and give n_0 a placebo, then observe how many patients in each group suffer a heart attack within the next 5 years. It can be derived directly from the simple tools we are learning in class.

Problem 2 (One-sample t-interval)**.** If $Z \sim N(0, 1)$ and $V \sim \chi^2_d$ with Z, V independent, we say that $T = Z/\sqrt{V/d}$ follows a *Student's t distribution* with d degrees of freedom, denoted by $T \sim t_d$. Note that $T^2 \sim F_{1,d}$ but T preserves sign information in case we want to do one-sided tests.

Now suppose $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$ with $\sigma^2 > 0$ unknown and consider testing $H_0 : \ \mu = \mu_0$ vs. $H_1 : \ \mu \neq \mu_0$.

We showed in class that the one-sided UMPU test for $H_0 : \ \mu \leq 0$ vs. $H_1 : \ \mu > 0$ rejects for large values of $T_X = \frac{\overline{X}\sqrt{n}}{\sqrt{S_X^2}}$, where S_X^2 is defined as in Problem 2.

(a) Show that $T_X \sim t_{n-1}$ if $\mu = 0$ (see hint for previous problem).

(b) To test $H_0 : \ \mu = 0$ vs. $H_1 : \ \mu \neq 0$, show that the UMPU test rejects for large values of $|T_X|$ (Hint: the simplest way is to use symmetry).

(c) Find a UMPU test of $H_0 : \ \mu = \mu_0$ for a generic $\mu_0 \in \mathbb{R}$, and invert to find a confidence interval for μ in terms of $\overline{X}$, S_X^2, quantiles of the t_{n-1} distribution, and the desired level α (Hint: consider the distribution of $X_i - \mu_0$).

Moral: The t-test of $\mu = 0$ that we are familiar with can be derived from the theory of UMPU conditional testing, but we need to additionally use the location family structure to get a conditional interval.

Problem 3 (McNemar's test)**.** Suppose we have paired binary data: for $i = 1, \ldots, n$ we observe $(X_i, Y_i) \in \{0, 1\}^2$. The pairs are i.i.d. with

$$\mathbb{P}\left[(X_i, Y_i) = (a, b)\right] = \pi_{a,b} \quad a, b \in \{0, 1\}.$$

This model could describe the performance of two prediction models on a test set, where X_i and Y_i represent respectively whether each model gets the ith prediction right. Or it could represent binary outcomes in a matched-pairs clinical trial, where similar patients are matched into pairs and then within each pair a coin is flipped to see who gets the treatment and who gets the placebo.

Write $\pi_X = \mathbb{P}(X_i = 1) = \pi_{1,0} + \pi_{1,1}$ and $\pi_Y = \mathbb{P}(Y_i = 1) = \pi_{0,1} + \pi_{1,1}$, and let $N_{a,b} = \sum_{i=1}^n 1\{X_i = a, Y_i = b\}$.

(a) Find the UMPU test of $H_0 : \ \pi_X \leq \pi_Y$ vs. $H_1 : \ \pi_X > \pi_Y$, giving the cutoffs $c(u), \gamma(u)$ in terms of solutions to integral equalities for a binomial distribution. (Hint: it may help to first reframe the hypothesis in terms of the $\pi_{a,b}$ parameters.)

(b) Suppose $N_{0,0} = N_{1,1} = 1000$, $N_{0,1} = 5$ and $N_{1,0} = 25$. Compute 95% confidence intervals for π_X and π_Y (invert the two-sided equal-tailed test but without randomizing). Then compute a p-value for $H_0 : \ \pi_X \leq \pi_Y$ (do not randomize). Does anything about the respective answers surprise you?

(Note: This test is called McNemar's test; it is very useful for clinical trials with matched pairs of subjects, and also for comparing the performance of different classifiers on a held-out sample.)

Moral: When we have paired data, we can often make much more precise comparisons between two distributions; even more precise than our ability to infer things about either of the distributions individually. This is often worth taking into account if we are designing an experiment: for example, if we match patients into pairs on demographic characteristics and then randomize a treatment/placebo assignment within each pair, we may get a very good inference about whether the treatment is better than the placebo, much better than we would get if we randomly assigned all $2n$ subjects independently of each other.

Problem 4 (Nonparametric tests)**.** In this problem you will design tests for two nonparametric hypothesis testing problems. There is necessarily some wiggle room in how you choose the test statistic, and it will probably not be possible to determine the cutoff explicitly. Just choose a reasonable one, define the cutoff in terms of a quantile of a well-defined distribution, and show that your test has significance level α.

(a) Suppose $X_1, \ldots, X_n \in \mathbb{R}$ are independent random variables with $X_i \sim P_i$. Consider testing the null hypothesis $H_0 : P_1 = P_2 = \cdots = P_n$ (i.e., the observations are i.i.d.) against the alternative that there is a systematic trend toward larger values of X_i as i increases (this is sometimes called a *test of trend*). Design a level-α test.

(b) Suppose $(X_1, Y_1), \ldots, (X_n, Y_n) \overset{\text{i.i.d.}}{\sim} P$ where P is an unknown joint distribution on $\mathbb{R}^2$. Consider testing the null hypothesis that X_i and Y_i are independent within each pair (i.e., $P = P_X \times P_Y$, with P_X and P_Y unknown and not necessarily the same) versus the alternative that (X_i, Y_i) are positively correlated within each pair. Design a level-α test.

Note that the alternative is defined a little vaguely in each part above. If that troubles you, we could formally take the alternative be "P_i are arbitrary but not all equal" in part (a), or "$P \neq P_X \times P_Y$" in part (b). The alternative hypotheses as I've defined them informally are meant to suggest which alternatives to prioritize when you design your test.

Moral: We can often design our own nonparametric tests by conditioning on an appropriate sufficient statistic for the null distribution.