Math 541b: Introduction to Mathematical Statistics
Homework 2 (due on Wed 11/16)

Please submit electronically directly to Blackboard in a PDF file.

1^o (Neyman-Pearson lemma for discrete distributions via randomization). Assume that we would like to test $H_0 : X \sim P$ against $H_1 : X \sim Q$, where P and Q are two discrete distributions on $\mathbb{R}$ with common support $\mathcal{X}$ (for simplicity assume that there is a finite number m of possible outcomes). Let p and q be the corresponding p.m.f.'s. Let $r(x) := \frac{q(x)}{p(x)}$ on $\mathcal{X}$, and consider the following test parametrized by $0 \le k \le \infty$ and $0 \le \gamma \le 1$:

$$\phi(x) = \begin{cases} 1, & r(x) > k, \\ \gamma, & r(x) = k, \\ 0, & r(x) < k. \end{cases}$$

(i) Show that for any $\alpha \in (0,1)$ one can select $k = k(\alpha)$ and $\gamma = \gamma(\alpha)$ such that $\mathbb{E}_P[\phi(x)] = \alpha$ (i.e., the test has size α).

(ii) Show that any such selection results in a test $\phi^*(x)$ which has the maximal power among all tests of level α. (Note that in our setup the power of a test $\varphi(\cdot)$ is $\beta_\varphi = \mathbb{E}_Q[\varphi(x)]$.)

(iii) Assuming $k \ge 1$, show that $\phi^*(x)$ is unbiased: $\beta_{\phi^*} \ge \alpha$, and moreover $\beta_{\phi^*} > \alpha$ unless $P = Q$.

2^o (Monotone likelihood ratio). Assume there is a set with N items, of which D are defective. One randomly selects $n < N$ items (i.e., they are sampled from N items without replacement) and shows them to you. You know N and n but not D. Let X be the number of defective items among those observed, and let $p_D(x)$ be the corresponding family of p.m.f.'s parametrized by D (and supported on $\mathcal{X} = \{0, 1, ..., \min\{n, D\}\}$). Find $p_D(x)$ and show that this is an MLR family with respect to $T(X) = X$.

3^o (Simple null against a composite alternative in an exponential family). Let $X_{1:n} := (X_1, ..., X_n)$ be an i.i.d. sample with $X \sim \mathcal{N}(0, \sigma^2)$ for each $i \in [n] := \{1, 2, ..., n\}$. Given $\sigma_0^2 > 0$, find the (asymptotically) UMP unbiased test of size α for testing $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 \ne \sigma_0^2$. Use the normal approximation of χ_n^2 (Chi-squared distribution) via CLT and a characterization of the UMP unbiased test in an exponential family given in the class (you are not required to prove it).

Hint 1: *Express $\mathcal{N}(0, \sigma^2)$ as an exponential family with a* canonical parametrization.

Hint 2: *Recall that the test must satisfy two conditions: its size is α, and $\mathbb{E}_0[T\phi(T)] = \mathbb{E}_0[T]\alpha$. where $\mathbb{E}_0[\cdot]$ its expectation under the null. Try to find a form of the test for which the second condition holds "automatically" in the CLT limit $n \to \infty$, regardless of α. The key word is "symmetry."*

4^o (Bayes-optimal test). Let $\mathcal{P} := \{P_\theta, \theta \in \Theta \subseteq \mathbb{R}^k\}$ be a family of distributions supported on $\mathcal{X} \subseteq \mathbb{R}^d$, and consider the general problem of testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$, where Θ_0 and Θ_1 are disjoint and $\Theta = \Theta_0 \cup \Theta_1$. Let P_θ have the density $p_\theta(x)$ with respect to the Lebesgue measure on $\mathbb{R}^d$, i.e., a p.d.f. We are also given a *prior* Π on Θ – a distribution supported on Θ with density $\pi(\theta)$, which reflects our "a priori guess" about the actual P_θ generating X.

Definition 1. *The* average risk *of a test $\phi(x)$ with respect to Π is defined as the sum of errors of both types averaged over $\theta \sim \Pi$:*

$$R_\Pi(\phi) := \int_{\Theta_0} \int_{\mathcal{X}} \phi(x) p_\theta(x) dx \cdot \pi(\theta) d\theta + \int_{\Theta_1} \int_{\mathcal{X}} (1 - \phi(x)) p_\theta(x) dx \cdot \pi(\theta) d\theta.$$

Any test minimizing R_Π is called Bayes-optimal*, and its average risk is called* the Bayes risk *(for Π).*

(i) <u>Show that the following test is Bayes-optimal:</u>

$$\phi_\Pi(x) = \begin{cases} 1, & \Pi(\Theta_1 | X = x) \ge \Pi(\Theta_0 | X = x), \\ 0, & \Pi(\Theta_1 | X = x) < \Pi(\Theta_0 | X = x), \end{cases}$$

where $\Pi(\Theta_1 | X = x)$ is the posterior probability of Θ_1 if $X = x$ is observed (similarly for Θ_0).[1]

(ii) <u>Show that for any choice of prior Π and test $\phi(\cdot)$, $R_\Pi(\phi)$ lower-bounds the *worst-case risk* of ϕ:</u>

$$\bar{R}(\phi) = \sup_{\theta \in \Theta} \left\{ \mathbb{1}[\theta \in \Theta_0] \cdot \int_{\mathcal{X}} \phi(x) p_\theta(x) dx \quad + \quad \mathbb{1}[\theta \in \Theta_1] \cdot \int_{\mathcal{X}} (1 - \phi(x)) p_\theta(x) dx \right\}.$$

This is useful because in composite testing problems (i.e., when Θ_0, Θ_1 are not singletons), worst-case risk is usually hard to evaluate precisely; however, if the prior Π is chosen reasonably, the Bayes risk will be close to $\bar{R}(\phi)$. <u>*How would you choose a prior?*</u> (This is not evaluated.)

(iii) Let $\bar{\phi}$ be any test minimizing $\bar{R}(\phi)$. Such test is called *minimax (or worst-case) optimal*, and $\bar{R}(\bar{\phi})$ is called the *minimax risk.*

<u>Using the results of (ii), conclude that the minimax risk is lower-bounded by the Bayes risk.</u>

5^o (Confidence-boosting via voting). Let $X_1, ..., X_n$ be an i.i.d. sample from $\mathbb{P}_\theta$, $\theta \in \Theta$. Assume also that $n = mk$ for some $m, k \in \mathbb{N}$, and there is a deterministic test $\phi(x_{1:k})$ that, using k observations, distinguishes between the two hypotheses H_0, H_1[2] with confidence $2/3$, that is

$$\max \left\{ \sup_{\theta \in \Theta_0} \mathbb{E}_\theta[\phi(X_{1:k})], \ \sup_{\theta \in \Theta_1} \mathbb{E}_\theta[1 - \phi(X_{1:k})] \right\} \le \frac{1}{3}.$$

Now, consider the following simple procedure:

1. Split $X_{1:n}$ into m batches $X^{(1)}, ..., X^{(m)}$ of k observations each, i.e. $X^{(j)} := X_{k(j-1)+1\,:\,k(j-1)+k}$.

[1]Note that, in fact, we have $\Pi(\Theta_0 | X = x) + \Pi(\Theta_1 | X = x) = \Pi(\Theta) = 1$, if Π is a probability measure; however, the results of this exercise are preserved even when Π is improper (that is $\Pi(\Theta) \neq 1$), and in particular when $\Pi(\Theta) = +\infty$.

[2]Corresponding to some partition $\Theta = \Theta_0 \sqcup \Theta_1$, but this is not important in the context of this problem.

2. Let $Z_j := \phi(X^{(j)})$, and consider the test

$$\varphi = \varphi(X_{1:n}) = \mathbb{1}\left[\sum_{j \in [m]} Z_j \ge \frac{m}{2}\right]$$

—in other words, accept/reject H_0 by aggregating the "basic" tests via the majority-vote rule.

(i) Working with the normal approximation for the binomial distribution (neglecting the error of this approximation), and using that $\mathbb{P}[U \ge u] \le e^{\frac{-u^2}{2}}$ where $U \sim \mathcal{N}(0,1)$, show the following:

$$\text{Risk}_m(\varphi) \lesssim e^{-cm}.$$

Here $c > 0$ is a constant; $\text{Risk}_m(\varphi)$ is the worst-case error (of either type) for test $\varphi(X_{1:n})$ with $n = km$; finally, $\lesssim$ "hides" the CLT approximation in the following sense: have an actual inequality, with "$\le$", if the distribution of the appropriate asymptotically normal statistic (converging to to $\mathcal{N}(0,1)$ by CLT) is *replaced with* $\mathcal{N}(0,1)$.[3]

(ii) Your next task is to show that the above bound (perhaps with some other $c > 0$) holds in finite sample and with $\le$ instead of $\lesssim$. To this end, assuming w.l.o.g. that m is even, note that

$$\text{Risk}_m(\varphi) \le \sum_{j=m/2}^{m} \binom{m}{j}\left(\frac{1}{3}\right)^j \left(\frac{2}{3}\right)^{m-j} < \left(\frac{m}{2}+1\right)\binom{m}{m/2}\left(\frac{1}{3}\right)^{m/2}\left(\frac{2}{3}\right)^{m/2}$$

and upper-bound the right-hand side with e^{-cm}.

(iii) Finally, show that we will get the same result in *(ii)* if, instead of 1/3, the "basic" test makes an error with probability at most $\delta < 1/2$; the only difference is that c will now depend on δ.

[3] Note that this does not suffice to argue that, say, $\text{Risk}_m(\varphi)e^{cm} \to \text{const}$ as $n \to \infty$. Could you explain why?