Stats 210A, Fall 2023
Homework 7

Due date:Thursday, Oct. 19

Instructions: The same standing instructions are in effect as in previous weeks.

1. MLR and location families

(a) Assume $X \sim p_\theta(x) = p_0(x - \theta)$, a location family with p_0 continuous and strictly positive. Show that the family has MLR in x if and only if $\log p_0$ is concave.

Note: For full credit, you should not assume that p_0 is differentiable.

Hint 1: It may help to recall that $f(x)$ is convex if and only if

$$R(x_1, x_2) = \frac{f(x_1) - f(x_2)}{x_1 - x_2}$$

is non-decreasing in x_1 and x_2.

Hint 2: It may also help to recall that a continuous function f is convex if and only if it is *midpoint convex* meaning

$$f\left(\frac{x_1 + x_2}{2}\right) \leq \frac{f(x_1) + f(x_2)}{2}, \qquad \text{for all } x_1, x_2.$$

(b) Consider testing in the Cauchy location family:

$$p_\theta(x) = \frac{1}{\pi(1 + (x - \theta)^2)}.$$

Let θ_0, θ_1 be any two real numbers with $\theta_1 > \theta_0$ and consider the LRT for testing $H_0 : \ \theta = \theta_0$ vs $H_1 : \ \theta = \theta_1$ at some level $\alpha \in (0, 1)$. Show that for some $\alpha^*(\theta_0, \theta_1)$, the rejection region for any $\alpha < \alpha^*$ is a bounded interval, and the rejection region for any $\alpha > \alpha^*$ is a union of two half intervals. Find α^*.

Hint: recall that $\frac{d}{dx} \arctan(x) = \frac{1}{1+x^2}$.

(c) In the Cauchy location family, prove that, for any $\alpha \in (0, 1)$, there exists no UMP level-α test of $H_0 : \ \theta = 0$ vs. $H_1 : \ \theta > 0$.

(d) Consider testing $H_0 : \theta = 0$ vs. $H_1 : \ \theta = 6$ in the Cauchy location family at level $\alpha = 0.01$. Numerically calculate the rejection region and the power for the LRT, and also for the one-tailed test that rejects for large values of X.

(e) **Optional:** (not graded, no extra points) In words, can you explain why the optimal LRT rejection regions for the Cauchy distribution take this odd form? Think about how you would explain to a scientific collaborator why you are proposing such an odd test, beyond "it fell out of an optimization problem."

Moral: When we think carefully about how to design rejection regions, we can get surprising results. In particular, for location families with heavy tails, extreme values are not that informative for distinguishing between two smaller values of the location parameter. Concretely, $X = 10^6$ doesn't help us distinguish between $\theta_1 = 1$ vs. $\theta_0 = 0$. By contrast, if the tails are lighter ($\log p_0$ concave implies the density shrinks at least exponentially) then more extreme X values always give stronger evidence for distinguishing between any two parameter values; this is what MLR means.

2. Some UMP tests

Numerically find the UMP test for the following hypothesis testing problems at level $\alpha = 0.05$. For each problem,

(i) derive the appropriate test on paper,

(ii) numerically compute the cutoff value c (and γ if necessary), and

(iii) plot the power function of the level-α test for an appropriate range of parameter values.

(a) $X_i \overset{\text{ind.}}{\sim} \text{Pois}(a_i\lambda)$ for $i = 1, \ldots, n$, where $a_1, \ldots, a_n$ are known positive constants and $\lambda > 0$ is unknown. Test $H_0 : \lambda = 1$ vs. $H_1 : \lambda > 1$, with $n = 5$ and $a_i = i$.

(b) $X_i \overset{\text{ind.}}{\sim} N(\theta, \sigma_i^2)$ for $i = 1, \ldots, n$, where σ_i^2 are known positive constants and $\theta \in \mathbb{R}$ is unknown. Test $H_0 : \theta = 0$ vs. $H_1 : \theta > 0$, with $n = 20$ and $\sigma_i^2 = i$. On your power plot, also plot the power function of the (sub-optimal) test that rejects for large $\sum_i X_i$.

(c) $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Pareto}(\theta) = \theta x^{-(1+\theta)}$, for $\theta > 0$ and $x > 1$ (also called a power law distribution). Test $H_0 : \theta = 1$ vs. $H_1 : \theta < 1$, for $n = 100$. On your power plot, also plot the power function of the (sub-optimal) test that rejects for large $\sum_i X_i$.

Moral: Once again, when we use the right test we often can deliver noticeably better power than if we chose an *ad hoc* test.

3. Uniform UMP test

We usually can't get a UMP two-sided test, but this problem gives an amusing counterexample where it is possible. Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Unif}[0, \theta]$ for $\theta > 0$.

(a) Consider the problem of testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$. Show that any test ϕ for which $\phi(x) = 1$ when $x_{(n)} = \max\{x_1, \ldots, x_n\} > \theta_0$ is UMP at level $\alpha = \mathbb{E}_{\theta_0}[\phi(X)]$.

(b) Now consider the problem of testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$. Show that a unique UMP level-α test exists, and is given by

$$\phi(x) = 1\left\{x_{(n)} > \theta_0 \text{ or } x_{(n)} < \theta_0\alpha^{1/n}\right\}$$

4. Bayesian hypothesis testing

Consider a univariate Gaussian problem with $X \mid \theta \sim N(\theta, 1)$, where $\theta = 0$ under the null hypothesis and $\theta \sim \Lambda_1$ under the alternative hypothesis (assume $\Lambda_1(\{0\}) = 0$). In addition let π_0 denote the *a priori* probability that the null hypothesis is true; therefore the full prior is a mixture between a point mass at 0 and Λ_1.

(a) Compute the posterior probability that the null hypothesis is true, i.e.

$$\pi_{\text{post}}(x; \Lambda_1, \pi_0) = \mathbb{P}(\theta = 0 \mid X = x).$$

(b) Assume $\pi_0 = 0.5$ (we are initially agnostic between the null and the alternative), and find

$$\pi^*_{\text{post}}(x) = \min_{\Lambda_1} \pi_{\text{post}}(x; \Lambda_1, 0.5),$$

as a function of x, for $x > 0$. Give the minimizing prior Λ_1, which also depends on x.

Note: This is not an optimization problem the analyst is going to solve: it is definitely not allowed to choose the prior after seeing the data. Instead, think of a large and diverse population of analysts who all have different priors before seeing the data, and therefore different posteriors after seeing the data (but with the constraint that none of them are initially "biased" against the null). Then we as theoreticians are calculating a lower bound π^*_{post} for any of these analysts' conditional belief in the null: all of their posterior credences in the null will be at least π^*_{post}. So everyone has their own prior but the only way someone could be really convinced that the null is false (more convinced than $1 - \pi^*_{\text{post}}$) is if they already thought it was probably false before seeing the data.

(c) Now restrict $\Lambda_1 = N(0, \tau^2)$ for $\tau > 0$, a class of more "realistic" priors. Compute π_{post} as a function of τ^2 and x. Find

$$\pi^*_{\text{post},N}(x) = \min_{\tau^2 > 0} \pi_{\text{post}}(x; N(0, \tau^2), 0.5),$$

and give the minimizing value of τ^2, both as functions of x, for $x > 1$.

(d) Now assume we observe a value of X such that the two-sided p-value $p(X)$ (i.e., $p(x) = \mathbb{P}_0(|X| > |x|)$) takes the values $0.05, 0.01, 0.005$, or 0.001. Numerically compute π^*_{post} and $\pi^*_{\text{post},N}$ for each value and make a small table. In words, interpret the results.

Moral: p-values are commonly misinterpreted as representing "the probability that the null hypothesis is true, given the data." This is an Bayesian statement and it depends on our prior beliefs. In fact, as this problem shows, even in a Bayesian setting, the p-value is generally not a good approximation for the posterior probability that the null is true.

5. Mean estimation

(a) Suppose $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} N_d(\theta, I_d)$ and consider estimating $\theta \in \mathbb{R}^d$. Show that $\overline{X} = \frac{1}{n}\sum_i X_i$ is the minimax estimator of θ under squared error loss.

Hint: Find a least favorable sequence of priors.

(b) Suppose $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} P$ where P is any distribution over the real numbers such that $\text{Var}_P(X) \leq 1$. Show that $\overline{X} = \frac{1}{n}\sum_i X_i$ is minimax for estimating $\theta(P) = \mathbb{E}_P X$ under the squared error loss.

Hint: Try to relate this problem to the Gaussian problem with $d = 1$.

(c) Assume $X \sim N(\theta, 1)$ with the constraint that $|\theta| \leq 1$. Show that the minimax estimator for squared error loss is

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

Plot its risk function.

Hint: Plot the risk function first. For this problem if you need to show that a function is maximized or minimized somewhere, you may do it numerically or by inspecting a graph if it is obvious enough.