Stats 210A, Fall 2025
Homework 7

Due on: Wednesday, Oct. 22

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

Problem 1 (Minimax estimation for the exponential)**.** Assume we observe $X \sim \text{Exp}(\theta)$ and want to
In this problem we'll consider minimax estimation of θ in the exponential scale family

$$X \sim \text{Exp}(\theta) = \frac{1}{\theta} e^{-x/\theta}, \quad x > 0,$$

under the relative squared error loss $L(\theta, d) = \left(\frac{d-\theta}{\theta}\right)^2$. The mean and variance of an $\text{Exp}(\theta)$ random variable are θ and θ^2, respectively.

(a) Consider linear scaling estimators of the form $\delta(x) = cx$, for $c \in [0, 1]$. Show that one of these estimators dominates all others, find the optimal c and give its risk function.

(b) A conjugate prior for this problem is the inverse-Gamma prior, which has parameters $\alpha, \beta > 0$ and density

$$\theta \sim \text{Inv-Gamma}(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{-\alpha-1} e^{-\beta/\theta}.$$

Note $\theta \sim \text{Inv-Gamma}(\alpha, \beta)$ when $\theta^{-1} \sim \text{Gamma}(\alpha, 1/\beta)$. Find the posterior distribution and Bayes estimator.

Note: Since we are not using squared error loss, the Bayes estimator is not just the posterior mean.

Hint: It may help to recall that if $Y \sim \text{Gamma}(k, \sigma)$ then $\mathbb{E}Y = k\sigma$ and $\text{Var}(Y) = k\sigma^2$.

(c) Find the Bayes risk for the inverse-Gamma prior with parameters α, β.

(d) Use the previous results to show that your estimator from part (a) is minimax.

(e) Now find the objective Bayes estimator, using the Jeffreys prior, and give its risk function. Can you relate the result to the results in the previous parts?

Hint: It may help to recall the general result you showed on a previous homework regarding the Jeffreys prior for scale families.

Moral: In some sense, the Jeffreys prior in scale families plays the same role that the flat prior plays in location families. In this problem, since we chose a scale-invariant loss based on relative error, the problem is equally hard everywhere in the parameter space, so the problem exhibits a kind of scale symmetry that is analogous to the translational symmetry that we see when we have a location family and a translationally invariant loss.

Problem 2 (Monotone likelihood ratio and location families)**.** A one-parameter family $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, with $\Theta \subseteq \mathbb{R}$, has *monotone likelihood ratios* in a statistic $T(X)$ if

$$\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} \text{ is a non-decreasing function of } T(X), \text{ for all } \theta_0 \leq \theta_1.$$

(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

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

(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) 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.

Problem 3 (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.

Problem 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) If $\pi_0 = 0.5$ and $X = x$, how small could the posterior null probability be?

That is, 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. Any given analyst will use their own actual prior to calculate their own posterior probability. We are just getting a lower bound on how small the analyst's posterior null probability could be.

(c) Now restrict $\Lambda_1 = N(0, \tau^2)$ for $\tau > 0$, a subclass of "realistic" priors an analyst might use if they were initially unsure about what alternative value to focus on. Compute π_{post} as a function of τ^2 and x.

Continuing to assume the analyst puts 0.5 prior on the null and the alternative, now how small could the analyst's posterior probability be? That is, 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.