Stats 210A, Fall 2025
Homework 12

Due on: Friday, Dec. 5

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

Problem 1 (Score test with nuisance parameters)**.** Consider a testing problem with $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} p_{\theta,\zeta}(x)$ with parameter of interest $\theta \in \mathbb{R}$ and nuisance parameter $\zeta \in \mathbb{R}$. That is, we are testing $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$, and ζ is unknown; let ζ_0 denote its true value. Then there is a version of the score test where we plug in an estimator for ζ, but we must use a corrected version of the variance.

Let $\hat{\zeta}_0$ denote the maximum likelihood estimator of ζ under the null:

$$\hat{\zeta}_0(\theta_0) = \arg\max_{\zeta \in \mathbb{R}} \; \ell(\theta_0, \zeta; X).$$

Assume $\hat{\zeta}_0$ is consistent under the null hypothesis.

Let $J(\theta, \zeta)$ denote the full-sample Fisher Information (omitting the usual n subscript), and assume it is continuous and positive-definite everywhere.

(a) Use Taylor expansions informally to show that, for large n,

$$\frac{\partial}{\partial\theta}\ell(\theta_0, \hat{\zeta}_0) \approx \frac{\partial}{\partial\theta}\ell(\theta_0, \zeta_0) - \frac{\frac{\partial^2}{\partial\theta\partial\zeta}\ell(\theta_0, \zeta_0)}{\frac{\partial^2}{\partial\zeta^2}\ell(\theta_0, \zeta_0)} \frac{\partial}{\partial\zeta}\ell(\theta_0, \zeta_0).$$

(Note: the LHS should be read as $[\frac{\partial}{\partial\theta}\ell(\theta, \zeta)]\big|_{\theta_0, \hat{\zeta}_0}$, and **not** $\frac{d}{d\theta_0}[\ell(\theta_0, \hat{\zeta}_0(\theta_0))]$).

(b) Using part (a), conclude that

$$\left(J_{11} - \frac{J_{12}^2}{J_{22}}\right)^{-1/2} \frac{\partial}{\partial\theta}\ell(\theta_0, \hat{\zeta}_0) \Rightarrow N(0, 1) \quad \text{ as } n \to \infty$$

where $J = J(\theta_0, \hat{\zeta}_0)$. Compare this to the score test statistic we would use if ζ_0 were known rather than estimated. (Note: you may assume without proof that the approximation error in part (a) is negligible; i.e. you may take the "$\approx$" as an exact equality).

Moral: The score test can be carried out with nuisance parameters, but the fact that we estimate the nuisance parameter affects the distribution of the test statistic in a way that we need to take into account.

Problem 2 (Poisson score test)**.** Suppose that for $i = 1, \dots, x_n$ we observe a real covariate $x_i \in \mathbb{R}$ (fixed and known) and a Poisson response $Y_i \sim \text{Pois}(\lambda_i)$. We assume that $\lambda_i = \alpha + \beta x_i$, with the restriction that $\lambda_i \geq 0$ for all i, but with $\alpha, \beta \in \mathbb{R}$ otherwise unrestricted. Assume that

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n |x_i - \bar{x}_n|^3}{\left(\sum_{i=1}^n (x_i - \bar{x}_n)^2\right)^{3/2}} = 0,$$

where $\bar{x}_n = n^{-1}\sum_{i=1}^n x_i$. We observe the first n pairs (x_i, y_i) and our goal is to test the hypothesis $H_0 : \beta = 0$ vs. $H_1 : \beta > 0$. Assume that there are at least 3 distinct values represented among $x_1, \dots, x_n$.

(a) Show that this model is a curved exponential family.

(b) Derive the score test statistic for H_0 vs H_1. Give the test statistic and asymptotic rejection cutoff.

(c) Show that your test statistic is indeed asymptotically normally distributed, and find an asymptotically valid rejection cutoff.

Hint: It may help to use the *Lyapunov CLT*, which applies to sums of independent random variables that are not necessarily identically distributed: Suppose $Z_1, Z_2, \ldots$ is a sequence of random variables with $Z_i \sim (\mu_i, \sigma_i^2)$, for $\sigma_i^2 < \infty$. Define $s_n^2 = \sum_{i=1}^n \sigma_i^2$. If for some $\delta > 0$, we have

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}\left[|Z_i - \mu_i|^{2+\delta}\right] = 0,$$

then $s_n^{-1} \sum_{i=1}^n (Z_i - \mu_i) \Rightarrow N(0, 1)$.

Hint: It may also help to start by assuming $\bar{x} = 0$, and then generalize your result.

(d) Suppose n is small, so we don't want to rely on the asymptotic normality. Explain how we could find a finite-sample exact conditional cutoff for the score test from part (b) (it is not necessary to give a closed form for the test, or to prove any optimality property).

Moral: Likelihood-based methods give us good options for estimation in models where exact inference could be difficult.

Problem 3 (Estimation in misspecified models)**.** Assume we observe a sample $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} p(x)$, and we perform maximum likelihood estimation for a real parameter θ using a dominated family $\mathcal{P} = \{p_\theta(x) : \theta \in \Theta \subseteq \mathbb{R}\}$, where $p \notin \mathcal{P}$. Let $\hat{\theta}_n$ denote the maximum likelihood estimator, and let θ^* denote the parameter value that minimizes KL divergence:

$$\theta^* = \underset{\theta\in\Theta}{\operatorname{argmin}}\ D_{\text{KL}}(p \parallel p_\theta) = \underset{\theta\in\Theta}{\operatorname{argmax}}\ \mathbb{E}_p\left[\ell_n(\theta; X)\right].$$

Since there is no true value of θ, we might still hope to "fail gracefully" by estimating θ^*, which (in some sense) best approximates the true distribution p.

Assume θ^* is unique, that the parameter space Θ is compact, that θ^* is in its interior, and that the supremum log-likelihood ratio between any pair of parameter values is bounded in expectation:

$$\mathbb{E}_p\left[\sup_{\theta_1,\theta_2\in\Theta} |\ell_1(\theta_1; X) - \ell_1(\theta_2; X)|\right] < B.$$

In addition, assume that the log-likelihood is twice continuosly differentiable, and that for all $\theta \in \Theta$,

$$\text{Var}_p(\dot{\ell}_1(\theta; X)) \in (0, \infty), \quad \mathbb{E}_p\left[\ddot{\ell}_1(\theta; X)\right] \in (-\infty, 0).$$

Note that by dominated convergence we can bring derivatives inside the integral; you do not need to justify this. Finally, assume $\mathbb{E}_p\left[\sup_{\theta\in\Theta} |\ddot{\ell}_1(\theta; X)|\right] < \infty$.

(a) Show that the maximum likelihood estimator converges in probability to θ^*.

(b) Give a counterexample to show why, even for smooth models, we should not expect under misspecification to have

$$-\mathbb{E}_p[\ddot{\ell}_n(\theta^*; X)] = \text{Var}_p(\dot{\ell}_n(\theta^*; X)).$$

(c) Find the limiting distribution of $\hat{\theta}_n$ as $n \to \infty$. We can think of a confidence interval for θ more generally as a confidence interval for the best-fitting parameter value θ^*. Can we expect the Wald confidence interval for the misspecified model to asymptotically achieve the correct coverage of θ^*?

Moral: There is a reasonable fallback interpretation for parameter estimates in misspecified models, but we need to be careful about standard errors and likelihood-based intervals.