Stats 210A, Fall 2018
Homework 13

Due date: Thursday, Dec. 6

You may disregard measure-theoretic niceties about conditioning on measure-zero sets, almost-sure equality vs. actual equality, "all functions" vs. "all measurable functions," etc. (unless the problem is explicitly asking about such issues).

1. Score test with nuisance parameters

Consider a testing problem with $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} p_{\theta,\zeta}(x)$ 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:

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

Let $J(\theta, \zeta)$ denote the Fisher Information (i.e. the full-sample information for all n observations), 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).

2. Pearson's χ^2 Test

Suppose that $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} p(x)$, a density with respect to the counting measure on $\mathcal{X} = \{1, \ldots, d\}$. Let $N_j = \sum_{i=1}^n 1\{X_i = j\}$ denote the counts (so $(N_1, \ldots, N_d)$ comprise a complete sufficient statistic for the sample X).

Assume $p_0(x)$ is a hypothesized distribution, which is strictly positive. The *Pearson* χ^2 *test statistic* for goodness-of-fit is defined as

$$S(X) = \sum_{j=1}^d \frac{(N_j - np_0(j))^2}{np_0(j)}$$

(a) Show that, if $p = p_0$, then $S(X) \Rightarrow \chi^2_{d-1}$ as $n \to \infty$.
(**Hint**: recall the multivariate CLT. If $Y_1, Y_2, \ldots$ are i.i.d. random vectors with mean $\mu \in \mathbb{R}^k$ and variance-covariance matrix $\Sigma \in \mathbb{R}^{k \times k}$, then $\frac{1}{\sqrt{n}} \sum_i^n (Y_i - \mu) \Rightarrow N(0, \Sigma)$.)

(b) Consider testing $H_0 : p = p_0$ vs. $H_1 : p \neq p_0$. Show that the test that rejects for large $S(X)$ is equivalent to the score test from class (Hint: note that there are really $d-1$, not d, free parameters in this problem).

3. Poisson score test

Suppose we observe covariate $x_i \in \mathbb{R}$ (fixed and known) and Poisson response $Y_i \sim \text{Pois}(\lambda_i)$ for $i = 1, \ldots, n$. We assume that $\lambda_i = \alpha + \beta x_i$, with the restriction that $\min_i \lambda_i \geq 0$, but $\alpha, \beta \in \mathbb{R}$ otherwise unknown. Assume there are at least 3 distinct values represented among $x_1, \ldots, x_n$.

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

(b) Derive the score test for the null hypothesis $H_0 : \beta = 0$ vs. $H_1 : \beta > 0$ (one-sided alternative). Give the test statistic and asymptotic rejection cutoff (this is not an i.i.d. sample but base your answer on the full-sample score and Fisher information; you do not need to justify the asymptotic approximation).

4. Trio of likelihood-based tests

Consider the three likelihood-based confidence intervals for a model with a single real parameter θ: the Wald, score, and generalized likelihood ratio intervals, which we can define respectively as $C_1^\theta(X)$, $C_2^\theta(X)$, and $C_3^\theta(X)$.

Define a new parameterization $\eta = f(\theta)$ where $f'(\theta) > 0$ for all $\theta \in \mathbb{R}$, and let $C_i^\eta(X)$ denote the corresponding confidence interval constructed based on the new parameterization. For which $i \in \{1, 2, 3\}$ are we guaranteed to have $C_i^\eta(X) = f(C_i^\theta(X))$ (i.e., which are invariant to parameterization)?

5. 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:

$$\theta^* = \operatorname*{arg\,min}_{\theta \in \Theta} D_{\text{KL}}(p \parallel p_\theta) = \operatorname*{arg\,max}_{\theta \in \Theta} \mathbb{E}_p \left[\ell(\theta; X)\right].$$

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:

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

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

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

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

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

(b) Is it true in general that

$$-\mathbb{E}_p[\ell''(\theta^*; X)] = \text{Var}_p(\ell'(\theta^*; X)) \text{ ?}$$

Prove or give a counterexample.

(c) Find the limiting distribution of $\hat{\theta}_n$ as $n \to \infty$. Will the Wald confidence interval cover θ^*?