

Stats 210A, Fall 2024

Homework 10

Due on: Friday, Nov. 22

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

Problem 1 (Some Maximum Likelihood Estimators). Find the MLE for each model below, and find its asymptotic distribution. You do not need to check the conditions for convergence theorems; just calculate assuming they are in force.

(a) Laplace: $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \frac{1}{2}e^{-|x-\theta|}$.

Note: although the log-likelihood is non-differentiable at one point, we can still use the Fisher information as defined by $J_1(\theta) = \text{Var}_\theta[\dot{\ell}_1(\theta; X_i)]$ to get the asymptotic distribution; you may assume this without proof. You may assume n is odd.

(b) Binomial: $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Binom}(m, \theta)$. Find the MLE for θ and for the canonical parameter $\eta = \log \frac{\theta}{1-\theta}$.

(c) Gaussian: $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\theta, \sigma^2)$. Find (i) the MLE for θ if σ^2 is known, (ii) the MLE for σ^2 if θ is known, and (iii) the MLE for (θ, σ^2) if neither is known.

Problem 2 (Limiting distribution of U -statistics). Suppose $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} P$ in some sample space \mathcal{X} . $U_n = U_n(X_1, \dots, X_n)$ is called a rank-2 U -statistic if

$$U_n = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n h(X_i, X_j)$$

where h is a symmetric function, i.e. $h(x_1, x_2) = h(x_2, x_1)$ for any $x_1, x_2 \in \mathcal{X}$.

In this problem, we denote $\theta = \mathbb{E}h(X_1, X_2)$ and assume that $\mathbb{E}h(X_1, X_2)^2 < \infty$. Note that U_n is the nonparametric UMVU estimator of θ .

Perhaps surprisingly, we can derive the asymptotic distribution of U_n in a relatively small number of steps using a technique called *Hájek projection* where we approximate it by an additive function of the independent X_i variables. We walk through the proof below.

(a) Define $g(x) = \mathbb{E}h(x, X_2) - \theta = \int h(x, u) dP(u) - \theta$. Show that, for all i ,

$$\mathbb{E}g(X_i) = 0, \quad \text{and} \quad \text{Var}(g(X_i)) < \infty.$$

(Note: g is a specific function from \mathcal{X} to \mathbb{R} . It is not a rule for naively substituting symbols into expressions. In particular, note that $g(X_i)$, a random variable, is not the same as the deterministic expression $\mathbb{E}h(X_i, X_2) - \theta$.)

(b) Define $\widehat{U}_n = \theta + \frac{2}{n} \sum_{i=1}^n g(X_i)$. Show that $\mathbb{E}[(U_n - \widehat{U}_n)f(X_i)] = 0$ for any i and any measurable function $f(X_i)$ with $\mathbb{E}[f(X_i)^2] < \infty$.

(Hint: Condition on X_i)

- (c) Show that $\sqrt{n}(U_n - \hat{U}_n) \xrightarrow{P} 0$ as $n \rightarrow \infty$. (Hint: show that U_n and \hat{U}_n have the same asymptotic variance, and then apply part (b)).
- (d) Conclude that $\sqrt{n}(U_n - \theta) \Rightarrow N(0, 4\zeta_1)$, where $\zeta_1 = \text{Var}(g(X_1))$.
- (e) Assume that $\mathcal{X} = \mathbb{R}$ with $\mathbb{E}X_i^4 < \infty$. Express the sample variance $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ as a rank-2 U-statistic and use the above results to derive its asymptotic distribution.

(**Note:** a similar result holds in general for rank- r U-statistics if we set $\hat{U}_n = \theta + \frac{r}{n} \sum_i g(X_i)$ where $g(x) = \mathbb{E}[h(x, X_2, \dots, X_r)] - \theta$.)

Moral: If P^n is the distribution of (X_1, \dots, X_n) then it is easy to check that the set of all square-integrable random variables of the form $f(X_1, \dots, X_n)$ (where $f : \mathcal{X}^n \rightarrow \mathbb{R}$ is measurable) forms a vector space over \mathbb{R} , which we call $L^2(P^n)$, where we can define an inner product as

$$\langle f(X), g(X) \rangle_{L^2} = \mathbb{E}[f(X)g(X)] \leq \sqrt{\mathbb{E}[f(X)^2]\mathbb{E}[g(X)^2]} < \infty.$$

Moreover, the subset of those random variables that can be written as $\sum_i f_i(X_i)$, where each f_i is measurable, forms a subspace. Part (b) establishes that the simpler random variable \hat{U}_n is the *projection* of U_n onto this subspace, and part (c) establishes that U_n is asymptotically very close to its projection.

Problem 3 (Score test with nuisance parameters). Consider a testing problem with $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} p_{\theta, \zeta}(x)$ 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 $[\frac{\partial}{\partial \theta} \ell(\theta, \zeta)]|_{\theta_0, \hat{\zeta}_0}$, and **not** $\frac{d}{d\theta_0} [\ell(\theta_0, \hat{\zeta}_0(\theta_0))]$).

- (b) Using part (a), conclude that

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

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 4 (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

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

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 .

- Show that this model is a curved exponential family.
- Derive the score test statistic for H_0 vs H_1 . Give the test statistic and asymptotic rejection cutoff.
- 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, \dots 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

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

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.

- 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).

Problem 5 (Super-Efficient Estimator). Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\theta, 1)$ and consider estimating θ via:

$$\delta_n(X) = \bar{X}_n 1\{|\bar{X}_n| > a_n\},$$

where $a_n \rightarrow 0$ but $a_n \sqrt{n} \rightarrow \infty$ as $n \rightarrow \infty$ (for example, $a_n = n^{-1/4}$).

- Show that δ_n has the same asymptotic distribution as \bar{X}_n when $\theta \neq 0$, but that $\sqrt{n}(\delta_n - \theta) \xrightarrow{P} 0$ if $\theta = 0$.
- Show that, pointwise in θ , as $n \rightarrow \infty$,

$$n \text{MSE}(\delta_n; \theta) \rightarrow 1\{\theta \neq 0\},$$

but that the convergence is not uniform in θ ; in fact,

$$\sup_{\theta \in \mathbb{R}} n \text{MSE}(\delta_n; \theta) \rightarrow \infty.$$

(**Note:** this is an example of a situation where it is incorrect to exchange a limit with a supremum.)

Moral: The sense in which asymptotically efficient estimators are “optimal” is not easy to define, and it isn't obvious how we should compare the asymptotic behavior of different estimators. In this example it would appear initially that the super-efficient estimator renders the sample mean inadmissible. But this is only true if we look at the pointwise limit for fixed θ ; at any n there are some values of θ for which the estimator is performing very badly, and this gets worse and worse as n gets larger.