

Stats 210A, Fall 2025

Homework 11

Due on: Wednesday, Nov. 19

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

Problem 1 (Precision-weighted average). Suppose that we observe two independent samples $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} (\mu, \sigma^2)$ and $Y_1, \dots, Y_m \stackrel{\text{i.i.d.}}{\sim} (\mu, \tau^2)$, with $n, m > 1$. The notation means that the expectation of a single X_i or Y_i is $\mu \in \mathbb{R}$, and the variance is $\sigma^2 > 0$ for a single X_i and $\tau^2 > 0$ for a single Y_i . All three parameters are unknown, but we are primarily interested in estimating the common expectation μ .

A natural estimator is to take a convex combination of the sample averages:

$$\delta_\gamma(X, Y) = \gamma \bar{X} + (1 - \gamma) \bar{Y},$$

for $\gamma \in [0, 1]$.

(a) Show that the optimal (variance-minimizing) choice of γ is

$$\gamma^* = \frac{n\sigma^{-2}}{n\sigma^{-2} + m\tau^{-2}} = \frac{1}{1 + \rho m/n},$$

where $\rho = \sigma^2/\tau^2$. δ_{γ^*} is called the *precision-weighted average* because $n\sigma^{-2}$ and $m\tau^{-2}$ are the precisions (inverse variances) of $\bar{X}$ and $\bar{Y}$, respectively. Give the variance of $\delta_{\gamma^*}(X, Y)$.

(b) Since σ^2 and τ^2 are unknown, we must estimate them. Let S_X^2 and S_Y^2 denote the usual sample variances for the two samples. Show that $\hat{\rho} = S_X^2/S_Y^2$ is a consistent estimator for ρ as $m, n \rightarrow \infty$.

Hint: It may help to recall the identity $(n-1)S_X^2 = \sum_i X_i^2 - n\bar{X}^2$.

Note: If you are wondering what it means for both m and n to go to ∞ , you may assume that we have a sequence of problems indexed by $k = 1, 2, \dots$ and $\min\{m_k, n_k\} \rightarrow \infty$ as $k \rightarrow \infty$. You should feel free to work more informally than this.

(c) Let $\hat{\gamma} = 1/(1 + \hat{\rho}m/n)$ and assume that $m, n \rightarrow \infty$ with $m/n \rightarrow c \in (0, \infty)$. Show that the adaptive estimator

$$\delta_{\hat{\gamma}}(X, Y) = \hat{\gamma} \bar{X} + (1 - \hat{\gamma}) \bar{Y}$$

has an asymptotic normal distribution as $n, m \rightarrow \infty$, and give its asymptotic distribution after appropriately centering and scaling it. Compare the asymptotic distribution of the adaptive estimator $\delta_{\hat{\gamma}}(X, Y)$ to the asymptotic distribution of the oracle estimator $\delta_{\gamma^*}(X, Y)$.

Hint: Start by considering the asymptotic distribution of $(\bar{X}, \bar{Y})$. You may use without proof the result that if $Z_n \Rightarrow P$ and $W_n \Rightarrow Q$, and Z_n and W_n are independent for each n , then $(Z_n, W_n) \rightarrow P \times Q$ (meaning the product measure between the distributions P and Q).

Note: Again, if we want to set up a formal sequence of problems in which the distribution converges, we could assume the ratio $c_k = m_k/n_k$ is converging to $c \in (0, \infty)$, in addition to our previous assumption that $\min\{m_k, n_k\} \rightarrow \infty$. As before, you can also work more informally.

Moral: This is an example of an estimator that's adaptive to an unknown problem parameter, in this case the ratio $\rho = \sigma^2/\tau^2$ which tells us how to choose the optimal value of some tuning parameter $\gamma^*(\rho, m, n)$. A tried and true method in statistics is to get a consistent estimator for the unknown parameter and just plug it in. Asymptotically, this very often works just as well as knowing the value of the nuisance parameter (it's not necessarily that we don't pay a price for not knowing ρ , it's just that the price we pay might be lower order than the source of error that matters).

Problem 2 (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.

Moral: Our general formula for the asymptotic distribution of the MLE allows us to quickly get good asymptotic approximations to their distributions.

Problem 3 (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 $\hat{U}_n = \theta + \frac{2}{n} \sum_{i=1}^n g(X_i)$. Show that $\mathbb{E}[(U_n - \hat{U}_n)f(X_i)] = 0$ 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 4 (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, $a_n = n^{-1/4}$).

(a) 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 $\theta = 0$.

(b) 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.