Stats 210A, Fall 2021
Homework 12

Due date: Friday, Dec. 3

Instructions: See the Standing Homework Instructions pinned post on Piazza.

1. Maximum Likelihood for Uniform

(a) Let $X_1, \ldots, X_n$ i.i.d. $\sim \text{Unif}[0, \theta]$. Find the maximum likelihood estimator $\hat{\theta}_n$ for $\theta$ and show that its asymptotic distribution is given by:

$$n(\theta - \hat{\theta}_n) \Rightarrow \text{Exp}(\theta) = \frac{1}{\theta} e^{-x/\theta} \mathbb{1}\{x > 0\}.$$ 

Here the error is of order $\frac{1}{n}$ instead of $\frac{1}{\sqrt{n}}$ as we usually expect, and as is predicted by the theorem from class. Why doesn’t the $\frac{1}{n}$ rate of convergence contradict the theorem from class on the asymptotic distribution of the MLE?

(Note: a factor of $n$ appears in the display equation above where we would normally see a factor of $\sqrt{n}$ instead. This is not a typo.)

(b) We showed previously that the UMVU estimator is $\delta_n = \frac{n+1}{n} X_{(n)}$, where $X_{(n)} = \max\{X_1, \ldots, X_n\}$. Find the asymptotic distribution of $\delta_n$. Which asymptotic distribution is better?

2. Bootstrap bias correction

Download the dataset in hw11-sample.csv, which consists of a $1000 \times 20$ matrix of integers. The data matrix consists of $1000$ random vectors $X_1, \ldots, X_n \sim P$ for a certain distribution with finite variance.

Consider estimating $\theta(P) = \lambda_{\max}(\text{Var}_P(X))$.

(a) Compute the plugin estimator of $\theta(P)$ and use the bootstrap to estimate its standard error.

(b) Estimate the bias and compute a bias-corrected estimator.

(c) Use the bootstrap to estimate the standard error and MSE of the bias-corrected estimator. Does it seem like the bias correction will be worth it, assuming our loss is measured in terms of MSE?

(d) Download the dataset in hw11-validation.csv, which consists of a $200,000 \times 20$ matrix of integers giving 200 more independent data sets of the same size, sampled from the true population, so that $X^{(k)}$ runs from row $1000(k - 1) + 1$ up to row $1000k$. Compute both estimators for each fresh data set and make a plot showing both histograms. Add a vertical line at 750, which is the true value of $\theta(P)$. Estimate the “ground truth” MSE for each estimator using your 200 estimates.

3. Estimation in misspecified models

Assume we observe a sample $X_1, \ldots, X_n \overset{i.i.d.}{\sim} p(x)$, and we perform maximum likelihood estimation for a real parameter $\theta$ 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 $\theta^*$ denote the parameter value that minimizes KL divergence:

$$\theta^* = \arg\min_{\theta \in \Theta} D_{\text{KL}}(p \parallel p_\theta) = \arg\max_{\theta \in \Theta} \mathbb{E}_{p} [\ell_n(\theta; X)].$$

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

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

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

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

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

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

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

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

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

(c) Find the limiting distribution of $\hat{\theta}_n$ as $n \to \infty$. Will the Wald confidence interval for the misspecified model asymptotically achieve the correct coverage of $\theta^*$?