STA 732: Homework 10

April 1, 2026

Please hand in solutions for Problems 1 and 2, as well as any 2 of Problems 3, 4, and 5 (so select 2 out of 3 for the last three problems).

Problem 1: DQM models have $\sqrt{n}$-uniform contraction rates

Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} P_{\theta_0}$ where $\{P_\theta : \theta \in \Theta\}$, $\Theta \subseteq \mathbb{R}$, is DQM at θ_0 with Fisher information $I(\theta_0) > 0$.

(a) Show that for δ small, $\rho(P_{\theta_0}, P_{\theta_0+\delta}) = 1 - \frac{1}{8} I(\theta_0)\delta^2 + o(\delta^2)$.

(b) Conclude that $\rho(P^n_{\theta_0}, P^n_{\theta_0+h/\sqrt{n}}) \to e^{-h^2 I(\theta_0)/8}$ for every fixed $h \in \mathbb{R}$.

(c) Apply the constraint risk inequality of Chapter 2 with $f = \theta_0$ and $g = \theta_0 + h/\sqrt{n}$ to show that for any estimator $\hat{\theta}_n$,

$$\liminf_{n\to\infty} \sup_{\theta \in \{\theta_0,\, \theta_0+h/\sqrt{n}\}} n\, \mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2] \ge \frac{h^2}{4}\, e^{-h^2 I(\theta_0)/4}.$$

(d) Optimize over h to conclude that any estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ satisfies

$$\liminf_{n\to\infty} \sup_{\theta\in\Theta_0} n\, \mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2] \ge \frac{1}{e\, I(\theta_0)}.$$

where Θ_0 is any neighborhood of θ_0.

(e) Explain why part (d) suggests that the rescaled quantity $\sqrt{n}(\hat{\theta}_n - \theta_0)$ and the local experiments $\{P^n_{\theta_0+h/\sqrt{n}} : h \in \mathbb{R}\}$ are the natural objects for studying asymptotic optimality of estimators at θ_0.

Problem 2: Super-efficiency comes at a price

Consider the normal location model in which we observe $X_1, \dots, X_n \overset{\text{iid}}{\sim} N(\theta, 1)$ and we are interested in estimating the parameter $\theta \in \mathbb{R}$ in squared error loss. Consider the Hodges estimator

$$S_n = \begin{cases} \bar{X}_n & \text{if } |\bar{X}_n| \ge n^{-1/4}, \\ 0 & \text{if } |\bar{X}_n| < n^{-1/4}. \end{cases}$$

Write $R_n(\theta)$ for the risk corresponding to the estimator: $\mathbb{E}_\theta[(S_n - \theta)^2]$.

(a) Show that for every fixed $\theta \neq 0$, $\sqrt{n}(S_n - \theta) \xrightarrow{d} N(0,1)$.

(b) Show that $r_n := n\, R_n(0) \leq e^{-c\sqrt{n}}$ for some constant $c > 0$ and n large.

(c) Let $\hat{\theta}_n$ be *any* estimator with $n\, \mathbb{E}_0[\hat{\theta}_n^2] = r_n \to 0$. Use the constraint risk inequality via the likelihood ratio (Lemma 2.57) to show that for any sequence h_n (possibly depending on n),

$$\sqrt{n\, \mathbb{E}_{h_n/\sqrt{n}}[(\hat{\theta}_n - h_n/\sqrt{n})^2]} \geq |h_n| - \sqrt{r_n}\, e^{h_n^2/2}.$$

(*Hint:* First show that $\mathbb{E}_0[(dP^n_{h/\sqrt{n}}/dP^n_0)^2] = e^{h^2}$.)

(d) Conclude that if $h_n \to \infty$ with $h_n^2 \leq (1-\epsilon)\log(1/r_n)$ for some $\epsilon > 0$, then

$$n\, \mathbb{E}_{h_n/\sqrt{n}}[(\hat{\theta}_n - h_n/\sqrt{n})^2] \gtrsim (1 - o(1))|h_n|^2 \to \infty$$

(e) For the Hodges estimator S_n, use the bound from part (b) to show that $n\, R_n(\theta_n) \to \infty$ for every sequence $\theta_n = h_n/\sqrt{n}$ with $h_n \to \infty$ and $h_n = o(n^{1/4})$.

(f) Compare to the MLE estimator – $\bar{X}_n$ – which satisfies $\sqrt{n}(\bar{X}_n - \theta) \sim N(0,1)$ for all $\theta \in \mathbb{R}$. In light of parts (a) and (b), the Hodges estimator appears to dominate the MLE at every fixed θ. Explain why this does not contradict your findings in part (e).

Problem 3: Illustrating the Bernstein-von Mises theorem

Consider the Laplace location model $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \text{Laplace}(\theta, 1)$ with density $f_\theta(y) = \frac{1}{2}e^{-|y-\theta|}$.

(a) Show that the model is DQM at every $\theta_0 \in \mathbb{R}$ with score (a version of) $S_{\theta_0}(y) = \text{sign}(y - \theta_0)$ and Fisher information $I(\theta_0) = 1$.

(b) Choose a voluntary value $\theta_0 \in \mathbb{R}$ and a proper prior Π on $\mathbb{R}$ whose Lebesgue density is continuous and positive at θ_0. For a single simulated dataset $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \text{Laplace}(\theta_0, 1)$ and each $n \in \{10, 30, 100, 300\}$, plot on the same axes:

   (i) the posterior density of θ (use whichever sampling method you prefer),

   (ii) the density of the 'oracle posterior' $N(\tilde{\theta}_n,\, n^{-1})$, where

   $$\tilde{\theta}_n = \theta_0 + \frac{1}{n}\sum_{i=1}^{n} \text{sign}(Y_i - \theta_0)$$

   is the *efficient oracle estimator*.

   (iii) vertical lines at the posterior mean, the efficient oracle estimator, and at θ_0.

(c) Report your findings. In particular, comment on:

- the quality of the Gaussian 'oracle posterior' comparison as n grows;
- the shape of the posterior for small n and why it is non-Gaussian;
- where the posterior mean falls relative to $\tilde{\theta}_n$ and θ_0;
- whether (and try to think of a reason why) the posterior centers at $\tilde{\theta}_n$ rather than the unknown θ_0.

Problem 4: Frequentist coverage of symmetric credible intervals

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} P_{\theta_0}$ for a scalar parameter $\theta_0 \in \Theta \subseteq \mathbb{R}$, and suppose that a Bernstein-von Mises theorem holds:

$$d_{\text{TV}}\Big(\Pi\Big(\sqrt{n}(\theta - \theta_0) \in \cdot \mid X^{(n)}\Big),\; N\Big(\sqrt{n}(\tilde{\theta}_n - \theta_0),\; I(\theta_0)^{-1}\Big)\Big) \xrightarrow{P_{\theta_0}} 0,$$

where $\tilde{\theta}_n$ is the *efficient oracle estimator*

$$\tilde{\theta}_n = \theta_0 + I(\theta_0)^{-1}\Delta_{n,\theta_0}/\sqrt{n}.$$

$\Delta_{n,\theta_0} = n^{-1/2}\sum_{i=1}^n S_{\theta_0}(X_i)$, $\mathbb{E}_{\theta_0} S_{\theta_0}(X_i) = 0$ and $I(\theta_0) = \mathbb{E}_{\theta_0}[S_{\theta_0}(X_1)^2] \in (0, \infty)$.

Let $L_n \equiv L_n(X^{(n)})$ and $U_n \equiv U_n(X^{(n)})$ be such that $[L_n, U_n]$ is a symmetric $(1-\alpha)$ posterior credible interval, e.g.

$$\Pi(\theta < L_n \mid X^{(n)}) = \Pi(\theta > U_n \mid X^{(n)}) = \alpha/2.$$

(a) *(Quantile convergence from TV convergence.)* Let $F_n(\cdot \mid X^{(n)})$ denote the posterior distribution function of θ and let Φ_n denote the distribution function of the BvM approximation $N(\tilde{\theta}_n,\, (nI(\theta_0))^{-1})$. Show that the convergence in total variation implies

$$\sup_{y \in \mathbb{R}} \big| F_n(y \mid X^{(n)}) - \Phi_n(y) \big| \xrightarrow{P_{\theta_0}} 0.$$

Conclude, using the results of Exercises 6.3 and 6.4 (or by a direct argument), that for every $p \in (0, 1)$,

$$\sqrt{n}\left(F_n^{-1}(p \mid X^{(n)}) - \Phi_n^{-1}(p)\right) \xrightarrow{P_{\theta_0}} 0.$$

(b) *(Oracle interval.)* Let $\tilde{L}_n = \tilde{\theta}_n + z_{\alpha/2}(nI(\theta_0))^{-1/2}$ and $\tilde{U}_n = \tilde{\theta}_n - z_{\alpha/2}(nI(\theta_0))^{-1/2}$ be the symmetric $(1-\alpha)$ interval from the BvM normal approximation. Show that

$$P_{\theta_0}\Big(\theta_0 \in [\tilde{L}_n, \tilde{U}_n]\Big) \to 1 - \alpha.$$

(*Hint*: Express $\theta_0 \in [\tilde{L}_n, \tilde{U}_n]$ in terms of Δ_{n,θ_0} and apply the central limit theorem.)

(c) *(Coverage of the credible interval.)* Combine parts (a) and (b) to show that

$$P_{\theta_0}(\theta_0 \in [L_n, U_n]) \to 1 - \alpha.$$

(*Hint*: Show that the events $\{\theta_0 \in [L_n, U_n]\}$ and $\{\theta_0 \in [\tilde{L}_n, \tilde{U}_n]\}$ differ on a set whose probability vanishes.)

Problem 5: Bayes estimators are asymptotically efficient under a Bernstein-von Mises theorem

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} P_\theta$ for a model that is DQM at a parameter value $\theta_0 \in \Theta \subseteq \mathbb{R}$. Suppose a BvM theorem holds; the rescaled posterior measure $\Pi_n^*(\cdot) := \Pi(\sqrt{n}(\theta - \theta_0) \in \cdot \mid X^{(n)})$ satisfies

$$d_{\text{TV}}\big(\Pi_n^*,\; N(\mu_n,\, I_0^{-1})\big) \xrightarrow{P_{\theta_0}} 0,$$

where $I_0 = I(\theta_0)$ and $\mu_n = I_0^{-1}\Delta_{n,\theta_0} \xrightarrow{d} N(0, I_0^{-1})$ as $n \to \infty$.

Let $L_0 : [0,\infty) \to [0,\infty)$ be nondecreasing, continuous, nonconstant, and bounded, and define the loss function

$$L(\theta, d) = L_0\big(\sqrt{n}\,|d-\theta|\big)\,.$$

Consider the corresponding risk function $R_n(\delta,\theta) = \mathbb{E}_\theta[L(\theta,\delta)]$.

(a) *(Localizing.)* Define

$$\delta^* \in \arg\min_{s\in\mathbb{R}} \int L_0(|s-h|)\, d\Pi_n^*(h).$$

Show that δ_Π is a Bayes estimator under loss $L(\theta,d) = L_0(\sqrt{n}\,|d-\theta|)$ if and only if $\sqrt{n}(\delta_\Pi - \theta_0) = \delta^*$.

(b) *(Bayes estimator in the normal experiment.)* Argue that if we had access to the oracle posterior $N(\theta_0 + \mu_n/\sqrt{n}, (nI_0)^{-1})$, our corresponding Bayes estimator would be the oracle estimator $\theta_0 + I_0^{-1}\Delta_{n,\theta_0}/\sqrt{n}$.

Hint: Suppose $H \sim N(\mu, I_0^{-1})$. Show that the map $s \mapsto \mathbb{E}[L_0(|s-H|)]$ is minimized at $s = \mu$.

(c) *(Argmin convergence.)* Define the rescaled posterior risk $R_n^*(s) = \int L_0(|s-h|)\, d\Pi_n^*(h)$ and the oracle posterior risk $\tilde{R}_n^*(s) = \int L_0(|s-h|)\, dN(\mu_n, I_0^{-1})(h) = g(s-\mu_n)$.

   (i) Show that $\sup_s |R_n^*(s) - \tilde{R}_n^*(s)| \le 2\|L_0\|_\infty\, d_{\mathrm{TV}}\big(\Pi_n^*,\ N(\mu_n,\ I_0^{-1})\big)$.

   (ii) Use the earlier questions to show that $\delta^* - \mu_n \xrightarrow{P_{\theta_0}} 0$.

   (iii) Conclude that $\sqrt{n}(\delta_\Pi - \theta_0) \xrightarrow{d} N(0, I_0^{-1})$ under P_{θ_0}.

(d) *(The posterior mean.)* The squared error loss $L(\theta,d) = n(d-\theta)^2$ (equivalently, $L_0(u) = u^2$) is unbounded, so part (b) does not apply directly. Let $\bar{\delta}^* = \int h\, d\Pi_n^*(h)$ denote the rescaled posterior mean. Assuming that the rescaled posterior second moment satisfies $\int h^2\, d\Pi_n^*(h) = O_{P_{\theta_0}}(1)$, show that $\bar{\delta}^* - \mu_n \xrightarrow{P_{\theta_0}} 0$ and hence that the posterior mean is also asymptotically efficient. (*Hint*: For $M > 0$, split the integral at $|h| = M$. Use the bound $|\int f\, d(Q_1 - Q_2)| \le 2\|f\|_\infty\, d_{\mathrm{TV}}(Q_1, Q_2)$ for the bounded part, and Cauchy–Schwarz for the tail.)