1. **Minimax risk for Bernoulli model of coin flips.** Consider the statistical decision theory simple setting with $\Theta \triangleq [0,1]$, observation space $\mathcal{X} \triangleq \{0,1\}$, *i.i.d.* observation sample $X : \Omega \to \mathcal{X}^m$ under unknown Bernoulli probability distribution with parameter $\theta \in \Theta$, and quadratic loss function $L : (\theta, \hat{\theta}) \mapsto (\theta - \hat{\theta})^2$. We denote the minimax risk for sample of size m by R_m^*.

   (a) Compute the risk $R_\theta(\hat{\theta}_{\text{emp}})$ for the empirical estimate $\hat{\theta}_{\text{emp}} \triangleq \bar{X} \triangleq \frac{1}{m}\sum_{i=1}^m X_i$, and show that $R_m^* \leqslant \frac{1}{4m}$.

   (b) Compute the Fisher information of $P^\theta \triangleq \text{Ber}(\theta)^{\otimes m}$ and $Q^\theta \triangleq \text{Bin}(m,\theta)$. Explain why they are equal.

   (c) Invoke the Bayesian Cramér-Rao lower bound theorem to show that $R_m^* = \frac{1+o(1)}{4m}$.

   (d) Notice that the risk of $\hat{\theta}$ is maximized at $\theta = 1/2$ (fair coin), which suggests that it might be possible to hedge against this situation by the following randomized estimator
   $$\hat{\theta}(X,U) \triangleq \hat{\theta}_{\text{emp}} \mathbb{1}_{\{U \leqslant \delta\}} + \frac{1}{2}\mathbb{1}_{\{U > \delta\}},$$
   for an external uniform randomness $U : \Omega \to [0,1]$ independent of observations. Find the worst-case risk of $\hat{\theta}(X,U)$ as a function of δ. Optimizing over δ, show the improved upper bound of $R_m^* \leqslant \frac{1}{4(m+1)}$.

   (e) Recall that a randomized estimator can always be improved if the loss is convex; so we should average out the randomness in $\hat{\theta}(X,U)$ by considering the estimator
   $$\hat{\theta}^* \triangleq \mathbb{E}[\hat{\theta}(X,U) \mid X] = \bar{X}\delta + \frac{1}{2}(1-\delta).$$
   Optimizing over δ to minimize the worst-case risk, find the resulting estimator $\hat{\theta}^*$ and its risk, show that it is constant independent of θ, and conclude $R_m^* \leqslant \frac{1}{4(1+\sqrt{m})^2}$.

   (f) Show that $\hat{\theta}^*$ found in part (e) is exactly minimax and hence $R_m^* = \frac{1}{4(1+\sqrt{m})^2}$. Recall that $\Gamma(a) \triangleq \int_{\mathbb{R}_+} x^{a-1}e^{-x}dx$. Consider the following prior Beta(a,b) with the following density for all $\theta \in \Theta$,
   $$\pi(\theta) \triangleq \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1},$$
   Show that if $a = b = \frac{\sqrt{m}}{2}$, then estimator $\hat{\theta}^*$ coincides with the Bayes estimator for this prior, which is therefore least favorable.
   (**Hint:** work with the sufficient statistic $S \triangleq \sum_{i=1}^m X_i$.)

   (g) Show that the least favorable prior is not unique; in fact, there is a continuum of them. (**Hint:** consider the Bayes estimator $\mathbb{E}[\theta \mid X]$ and show that it only depends on the first $m+1$ moments of π.)

   (h) ***k*-ary alphabet.** Consider k-ary input space $\mathcal{X} \triangleq [k]$, such that *i.i.d.* observation sample $X : \Omega \to \mathcal{X}^m$ has a common distribution $P \in \mathcal{M}(\mathcal{X})$. Show that for any k, m, the minimax squared risk of estimating P is exactly
   $$R^*(k,m) \triangleq \inf_{\hat{P}} \sup_{P \in \mathcal{P}_k} \mathbb{E}\left\|\hat{P} - P\right\|_2^2 = \frac{k-1}{k(1+\sqrt{m})^2},$$
   achieved by add $\frac{\sqrt{m}}{k}$ estimator.
   (**Hint:** For the lower bound, show that the Bayes estimator for the squared loss and the KL loss coincide.)

2. **Maximum of ratio.** Show that for vector $x, y \in \mathbb{R}^d$ and a positive definite matrix $\Sigma \in \mathbb{R}^{d\times d}$, we have $\sup_{x\neq 0} \frac{\langle x,y\rangle^2}{x^\top \Sigma x} = y^\top \Sigma^{-1} y$, where the maxima is achieved at $x^* = \Sigma^{-1} y$.

3. **Chernoff-Rubin-Stein lower bound.** Consider statistical decision theory simple setting with $\Theta \triangleq [-a, a]$ and $\mathcal{Y}, \mathcal{Y}' \subseteq \mathbb{R}$. Consider an *i.i.d.* observation sample $X : \Omega \to \mathcal{X}^m$ with common distribution $P_\theta \in \mathcal{M}(\mathcal{X})$.

   (a) State the appropriate regularity conditions and prove the following minimax lower bound,
   $$\inf_{\hat\theta} \sup_{\theta\in\Theta} \mathbb{E}_\theta(\theta - \hat\theta)^2 \geqslant \min_{\epsilon\in(0,1)} \max\left\{\epsilon^2 a^2, \frac{(1-\epsilon)^2}{m\bar{J}_F}\right\}.$$
   where $\bar{J}_F \triangleq \frac{1}{|\Theta|}\int_\Theta J_F(\theta)d\theta$ is the average Fisher information. (Hint: Consider the uniform prior on Θ and proceed as in the proof of Theorem 29.2 by applying integration by parts.)

   (b) Simplify the above bound and show that
   $$\inf_{\hat\theta} \sup_{\theta\in\Theta} \mathbb{E}_\theta(\theta - \hat\theta)^2 \geqslant \frac{1}{(a^{-1} + \sqrt{m\bar{J}_F})^2}. \quad (1)$$

   (c) Assuming the continuity of map $\theta \mapsto J_F(\theta)$, show that the above result also leads to the optimal local minimax lower bound in Theorem 29.4 obtained from Bayesian Cramér-Rao lower bound, i.e.
   $$\inf_{\hat\theta} \sup_{\theta-\theta_0\in[-m^{1/4}, m^{1/4}]} \mathbb{E}_\theta(\theta - \hat\theta)^2 \geqslant \frac{1+o(1)}{mJ_F(\theta_0)}.$$

4. **Prior with smallest Fisher information.** Show that the optimal density with a compact support is a squared cosine density $g : [-1, 1] \to \mathbb{R}$ defined as $g(u) \triangleq \cos^2 \frac{\pi u}{2}$ for all $u \in [-1, 1]$. Further, show that the minimum Fisher information is
   $$\min_{g \text{ on } [-1,1]} J(g) = \pi^2.$$

5. **Data processing inequality for f-divergence.** For any Markov chain $X \to Y \to Z$, a pair of measures $P_{X,Y,Z}$ and $Q_{X,Y,Z}$ with common Markov kernel $P_{Z|Y} = Q_{Z|Y}$, a convex map $f : (0, \infty) \to \mathbb{R}_+$, and arbitrary function $g : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}$, we have
   $$D_f(P_{X,Y} \| Q_{X,Y}) \geqslant D_f(P_{X,Z} \| Q_{X,Z}) \geqslant D_f(P_{g(X,Z)} \| Q_{g(X,Z)}).$$

6. **HCR for GLM.** Consider an *i.i.d.* random vector $X : \Omega \to \mathcal{X}^m$ with the common probability distribution $P_\theta \triangleq \mathcal{N}(\theta, 1)$ with unknown mean $\theta \in \mathbb{R}$ and known variance 1. Consider the sample mean estimator $\hat\theta(X) \triangleq \frac{1}{m}\sum_{i=1}^m X_i$. Using the Hammersley-Chapman-Robbins (HCR) lower bound, compute a lower bound for the variance of any unbiased estimator of θ, assuming you are comparing θ and $\theta' \triangleq \theta + \delta$ where $\delta \neq 0$.

   (a) Given that P_θ and $P_{\theta'}$ are Gaussian distributions, calculate $\chi^2(P_{\theta'} \| P_\theta)$ in terms of δ.

   (b) Use the result from the χ^2-divergence to provide the HCR lower bound for the variance of any unbiased estimator of θ.

   (c) Based on your results, if $n = 10$ and $\delta = 0.1$, what is the numerical lower bound for the variance of the estimator $\hat\theta$?