1. **Bayes risk for square loss function.** Consider the statistical decision theory simple setting with unconstrained parameter set $\Theta \triangleq \mathbb{R}^d$, input space $\mathcal{X} = \Theta$, estimate $\hat{\theta} \triangleq \hat{T}(X,U)$ for observation X and external randomness $U : \Omega \to [0,1]$, a prior $\pi \in \mathcal{M}(\Theta)$, and the quadratic loss $L : (\theta,\hat{\theta}) \mapsto \left\|\theta - \hat{\theta}\right\|^2$.
   (a) Show that the best Bayes estimator is deterministic for any loss function. Consequently, it suffices to focus on deterministic estimators $\hat{T}(X)$.
   (b) Show that for any estimator $\hat{T}(X)$, we have $\mathbb{E}[(\theta - \mathbb{E}[\theta \mid X])\hat{T}(X)] = 0$.
   (c) Show that the Bayes estimator for quadratic loss is $\hat{T}_B(X) \triangleq \mathbb{E}[\theta \mid X]$.
   (d) Show that the Bayes risk is $\mathbb{E}[\mathrm{tr}(\mathrm{cov}(\theta \mid X))]$.

2. **Bayes risk for quadratic GLM.** Consider Consider the statistical decision theory simple setting with unconstrained parameter space $\Theta \triangleq \mathbb{R}^d$, input space $\mathcal{X} = \Theta$, and estimate $\hat{\theta} \triangleq \hat{T}(X,U)$ for observation X and external randomness $U : \Omega \to [0,1]$. For GLM, the observation $X \triangleq \theta + Z$, where Z is independent of θ and has a zero-mean Gaussian distribution $\mathcal{N}(0,\sigma^2 I_d)$. Consider a Gaussian prior $\pi \in \mathcal{M}(\mathcal{X})$ with zero mean and covariance matrix sI_d.
   (a) Given the observation X, derive the posterior distribution $P_{\theta|X}$ and identify the parameters of this distribution in terms of s and σ^2.
   (b) Find the Bayes estimator and Bayes risk for quadratic loss function $L : \theta \times \hat{\theta} \mapsto \left\|\theta - \hat{\theta}\right\|^2$.

3. **Minimax for quadratic GLM with constrained parameter space.** Consider the statistical decision theory simple setting for Gaussian location model with constrained parameter space $\Theta \triangleq \mathbb{R}_+$, input space $\mathcal{X} = \mathbb{R}$, observation $X \sim \mathcal{N}(\theta,\sigma^2)$, and quadratic loss function $L : (\theta,\hat{\theta}) \mapsto \left\|\hat{\theta} - \theta\right\|^2$.
   (a) Show that the minimax quadratic risk of the GLM $X \sim \mathcal{N}(\theta,\sigma^2)$ with constrained parameter space $\Theta = \mathbb{R}_+$ is the same as the unconstrained case $\Theta = \mathbb{R}$.
   (b) Show that the thresholded estimator $X_+ = X \vee 0$ achieves a better risk compared to maximum likelihood estimator, pointwise at every $\theta \in \mathbb{R}_+$.

4. **Bayes risk for Cauchy prior and exponential family.** Let $\mathcal{X} \triangleq \mathbb{R}_+$ and consider *i.i.d.* random vector $X : \Omega \to \mathcal{X}^n$ with common distribution $P_\theta \in \mathcal{M}(\mathcal{X})$ where $P_\theta(0,x] \triangleq 1 - e^{-\theta x}$ for all $x \in \mathbb{R}_+$. Let $\Theta \triangleq \mathbb{R}$ and the prior distribution $\pi \in \mathcal{M}(\Theta)$ follows the Cauchy distribution with parameter s, such that
$$\pi(\theta) \triangleq \int_{-\infty}^{\theta} \frac{1}{\pi s(1 + \frac{x^2}{s^2})} dx.$$
   Show that the Bayes risk $R^*_\pi \triangleq \inf_{\hat{\theta}} \mathbb{E}\left(\hat{\theta}(X) - \theta\right)^2$ satisfies the inequality $R^*_\pi \geqslant \frac{2s^2}{2ns^2+1}$.

5. **Minimax risk for multiple observations under Bernoulli family.** Consider statistical decision theory simple setting with $\Theta \triangleq [0,1]$, input space $\mathcal{X} \triangleq \{0,1\}$, observation sample $X : \Omega \to \mathcal{X}^m$ *i.i.d.* with common Bernoulli 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 m-sized sample by R^*_m.
   (a) Compute the risk $R_\theta(\hat{\theta}_{\text{emp}})$ for the empirical estimate $\hat{\theta}_{\text{emp}} \triangleq \bar{X} \triangleq \frac{1}{m}\sum_{i=1}^m X_i$. Show that $R^*_m \leqslant \frac{1}{4m}$.
   (b) Compute the Fisher information of distribution $P_\theta \triangleq \mathrm{Ber}(\theta)^{\otimes m}$ and $Q_\theta \triangleq \mathrm{Bin}(m,\theta)$. Explain why they are equal.
   (c) 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).

6. **Minimax risk for multiple observations under Gaussian family.** Consider statistical decision theory simple setting for $\Theta \triangleq \left\{\theta \in \mathbb{R}^{d\times d} : \|\theta\|_F \triangleq \sum_{i,j=1}^d \theta_{ij}^2 \leqslant r\right\}$, observation space $\mathcal{X} \triangleq \mathbb{R}^d$, *i.i.d.* observation sample $X : \Omega \to \mathcal{X}^m$ with common distribution $\mathcal{N}(0,\Sigma)$, and quadratic loss function $L : \Sigma \times \hat{\Sigma} \mapsto \left\|\Sigma - \hat{\Sigma}\right\|_F^2$. Show that the minimax quadratic risk satisfies
$$R^* \geqslant \left(\frac{d}{m} \wedge 1\right) r^2.$$
To show this, one may have to find matching upper and lower bound on minimax risk.
   (a) To show the upper bound, consider the sample covariance matrix defined as $\hat{\Sigma} \triangleq \frac{1}{m}\sum_{i=1}^m X_i X_i^\top$, and show that
   $$\mathbb{E}\left\|\Sigma - \hat{\Sigma}\right\|_F^2 = \frac{1}{m}.$$
   (b) To show the matching lower bound, show that for any positive semidefinite (PSD) matrices $\Sigma_0, \Sigma_1 \succeq 0$ and identity matrix I_d, the KL divergence satisfies
   $$D\Big(\mathcal{N}(0, I_d + \Sigma_0) \| \mathcal{N}(0, I_d + \Sigma_1)\Big) \leqslant \frac{1}{2}\left\|\Sigma_0^{\frac{1}{2}} - \Sigma_1^{\frac{1}{2}}\right\|_F^2.$$

7. **Sample complexity as a function of dimensions.** Consider the matrix case $\Theta \triangleq \mathbb{R}^{d\times d}$ with m independent observations in zero mean unit variance Gaussian noise, and let ϵ be a small constant. Then we have
   (a) For quadratic loss, namely, $\left\|\theta - \hat{\theta}\right\|_F^2$, we have $R_m^* = \frac{d^2}{m}$ and hence $m^*(\epsilon) = \Theta(d^2)$.
   (b) If the loss function is $\left\|\theta - \hat{\theta}\right\|_{\text{op}}^2$ then $R_m^* \asymp \frac{d}{m}$ and hence $m^*(\epsilon) = \Theta(d)$.
   (c) If $T(\theta) \triangleq \max_{i\in[d]} \theta_i$, then $m^*(\epsilon) = \Theta(\sqrt{\ln d})$.

8. **Minimax loss for bowl shaped functions.** Consider the statistical decision theory simple setting where $\Theta = \mathcal{Y} = \mathcal{Y}'$, and the loss function $L : \theta \times \hat{\theta} \mapsto L(\theta, \hat{\theta}) \triangleq \rho(\theta - \hat{\theta})$ is defined in terms of bowl-shaped loss functions $\rho : \mathbb{R}^d \times \mathbb{R}_+$ for all $x \in \mathbb{R}^d$. Show the following statements are true.
   (a) For $\Theta \subseteq \mathbb{R}^d$ and $\rho(x) \triangleq \|x\|_2^2$, the minimax risk is $R^* \asymp \frac{1}{m}\mathbb{E}\|Z\|^2 = \frac{d}{m}$.
   (b) For $\Theta \subseteq \mathbb{R}^d$ and $\rho(x) \triangleq \|x\|_\infty$, we have $\mathbb{E}\|Z\|_\infty \asymp \sqrt{\ln d}$ and the minimax risk is $R^* = \sqrt{\frac{d}{m}}$.
   (c) For $\Theta \subseteq \mathbb{R}^{d\times d}$ and $\rho(\theta) = \|\theta\|_{\text{op}}$ denote the operator norm that is the maximum singular value. In this case, $\mathbb{E}\|Z\|_{\text{op}} \asymp \sqrt{d}$ and so minimax risk is $R^* = \sqrt{\frac{d}{m}}$.
   (d) For $\Theta \subseteq \mathbb{R}^{d\times d}$ and $\rho(\theta) = \|\theta\|_F$, the minimax risk $R^* \asymp \frac{d}{\sqrt{m}}$.