1. Gaussian mean estimation.

1.1. Optimal shrinkage factor [10pts]. Let \( x_1, x_2, \ldots, x_n \in \mathbb{R}^d \) be i.i.d. multivariate Gaussian random vectors, i.e., \( x_i \sim \mathcal{N}(\mu, \sigma^2 I) \). Denoting the sample mean estimator with \( \hat{\mu} \triangleq \frac{1}{n} \sum_{i=1}^n x_i \), consider an estimator of the form \( \hat{\mu}^s = \left( 1 - \frac{\tau}{\|\hat{\mu}\|^2} \right) \hat{\mu} \). Find the optimal \( \tau \) that minimizes the risk \( R(\hat{\mu}^s, \mu) = \mathbb{E}[\|\hat{\mu}^s - \mu\|^2] \).

1.2. Generalizing Stein’s lemma [10pts]. Assume that \( X \sim p_\eta(x) \) and \( g_\eta : \mathbb{R}^d \to \mathbb{R}^d \) where \( p_\eta \) and \( g_\eta \) are differentiable w.r.t \( \eta \) and \( x \), and \( \mathbb{E}[g_\eta(X)] = \xi(\eta) \) for some function \( \xi \). Show that
- (a) \( \mathbb{E}[\nabla_x \log p_\eta(X) g_\eta(X)^\top] + \mathbb{E}[\nabla_x g_\eta(X)] = 0 \).
- (b) \( \mathbb{E}[\nabla_\eta \log p_\eta(X) g_\eta(X)^\top] + \mathbb{E}[\nabla_\eta g_\eta(X)] = \nabla_\eta \xi(\eta) \).

1.3. Generalizing SURE [10pts]. Let \( x \sim \mathcal{N}(\mu, \Sigma) \) where \( \mu \in \mathbb{R}^d \) and \( \Sigma \in \mathbb{R}^{d \times d} \). If \( \hat{\mu}(x) \in \mathbb{R}^d \) is an estimator of the form \( \hat{\mu}(x) = g(x) \) where \( g : \mathbb{R}^d \to \mathbb{R}^d \) is differentiable. Define the functional
\[
S(x, \hat{\mu}) = \text{Tr}(\Sigma) + 2 \text{Tr}(\Sigma \nabla_x g(x)) + \|g(x)\|_2^2.
\]
Then show that \( S(x, \hat{\mu}) \) is an unbiased estimator of the risk, i.e., \( \mathbb{E}[\|\hat{\mu}(x) - \mu\|^2] = \mathbb{E}[S(x, \hat{\mu})] \).

2. Exponential families.

2.1. Second moment [10pts]. For a density of the form \( p_\eta(x) = \exp(\langle \eta, \phi(x) \rangle - \psi(\eta)) \), let \( \mathbb{E}[\phi(x)] = \mu \). For a vector \( \xi \in \mathbb{R}^d \), find \( \text{Tr}(\mathbb{E}[\phi(x) - \phi(\xi)](\phi(x) - \xi)^\top]) \).

2.2. Score function [10pts]. Assume that \( x \sim p_\eta(x) \) where \( p_\eta \) is not necessarily in the exponential family form. Denote the log-likelihood by \( \ell_\eta(x) = \log p_\eta(x) \), show that
- (a) \( \mathbb{E}[\nabla_\eta \ell_\eta(x)] = 0 \).
- (b) \( \mathbb{E}[\nabla_\eta \ell_\eta(x) \nabla_\eta \ell_\eta(x)^\top] = -\mathbb{E}[\nabla_\eta^2 \ell_\eta(x)] \) (Problem 1.2 may be helpful).

2.3. Maximum entropy principle [Bonus 2pts]. Assume that \( p(x) \) is a probability mass function of a discrete random variable taking values from a finite set \( \mathcal{X} \). Entropy of \( p \) is defined as \( H(p) = -\sum_{x \in \mathcal{X}} p(x) \log p(x) \). For \( \phi : \mathcal{X} \to \mathbb{R}^d \), show that the maximum entropy density satisfying \( \mathbb{E}_p[\phi(X)] = \mu \in \mathbb{R}^d \) is a member of exponential family. That is, show that the solution to
\[
\text{maximize}_p H(p) \text{ subject to: } \mathbb{E}_p[\phi(X)] = \mu,
\]
is an exponential family. (Hint: Write the Lagrangian associated with the above optimization problem. Since \( \mathcal{X} \) is finite, think of \( p(x) \) as a vector and maximize over it.)