Stats 210A, Fall 2020
Homework 1

Due date: Wednesday, Sep. 9

You may disregard measure-theoretic niceties about conditioning on measure-zero sets, almost-sure equality vs. actual equality, “all functions” vs. “all measurable functions,” etc. (unless the problem is explicitly asking about such issues).

1. Risk of a shrinkage estimator
Let $\theta$ denote the proportion of people working in a company who are left-handed, and suppose we are in charge of ordering equipment and need to estimate $\theta$. Let $X$ denote the number of left-handers in a sample of size $n$ from the company (for simplicity, assume we sample with replacement).

It is known that 10% of the U.S. population is left-handed. Instead of using the “obvious” estimator $\hat{\theta}_0(X) = X/n$, we could “shrink” the estimator toward 10% by using:

$$\hat{\theta}_1(X) = 0.2 \cdot 10\% + 0.8 \cdot \frac{X}{n},$$

Let $\text{MSE}_n(\theta, \hat{\theta}_i)$ denote the mean squared error of an estimator $\hat{\theta}_i$, as a function of the sample size $n$ and true parameter $\theta$.

(a) Find $\text{MSE}_n(\theta, \hat{\theta}_i)$ for $i = 0, 1$.

(b) For what values of $\theta$ is $\text{MSE}_n(\theta, \hat{\theta}_1) < \text{MSE}_n(\theta, \hat{\theta}_0)$? Give the answer as a function of $n$. What happens as $n \to \infty$?

2. Convexity of $A(\eta)$ and $\Xi_1$
Let $\mathcal{P} = \{p_\eta: \eta \in \Xi_1\}$ denote an $s$-parameter exponential family in canonical form

$$p_\eta(x) = e^{\eta' T(x) - A(\eta) h(x)}, \quad A(\eta) = \log \int_{\mathcal{X}} e^{\eta' T(x) h(x)} d\mu(x),$$

where $\Xi_1 = \{\eta: A(\eta) < \infty\}$ is the natural parameter space.

Recall Hölder’s inequality: if $q_1, q_2 \geq 1$ with $q_1^{-1} + q_2^{-1} = 1$, and $f_1$ and $f_2$ are ($\mu$-measurable) functions from $\mathcal{X}$ to $\mathbb{R}$, then

$$\|f_1 f_2\|_{L^1(\mu)} \leq \|f_1\|_{L^{q_1}(\mu)} \|f_2\|_{L^{q_2}(\mu)}, \quad \text{where} \quad \|f\|_{L^q(\mu)} = \left(\int_{\mathcal{X}} |f(x)|^q d\mu(x)\right)^{1/q}.$$

(Note that $q_1 = q_2 = 2$ reduces to Cauchy-Schwarz).

(a) Show that $A(\eta): \mathbb{R}^s \to [0, \infty]$ is a convex function: that is, for any $\eta_1, \eta_2 \in \mathbb{R}^s$ (not just in $\Xi_1$), and $c \in [0, 1]$ then

$$A(c\eta_1 + (1-c)\eta_2) \leq cA(\eta_1) + (1-c)A(\eta_2).$$

(Hint: try $q_1 = q_2 = 2$ and $f_1(x)^{1/c} = e^{\eta_1' T(x) h(x)}$.)

(b) Conclude that $\Xi_1 \subseteq \mathbb{R}^s$ is convex.

3. Expectation of an increasing function
4. Exponential families maximize entropy

The entropy (with respect to \( \mu \)) of a random variable \( X \) with density \( p \), is defined by

\[
h(p) = \mathbb{E}_{p}(-\log p(X)) = -\int_{\{x : p(x) > 0\}} \log(p(x))p(x) \, d\mu(x).
\]

This quantity arises naturally in information theory as a minimal expected code length.

Now consider the problem of maximizing \( h(p) \) subject to the constraints that \( p \) is a probability density with \( \mathbb{E}_{p}[T(X)] = \alpha \), for some \( T : \mathcal{X} \rightarrow \mathbb{R}^s \) and \( \alpha \) in the interior of the convex hull of \( T(\mathcal{X}) = \{ T(x) : x \in \mathcal{X} \} \).

That is, \( p(x) > 0, \int p(x) \, d\mu = 1, \) and \( \int p(x)T_j(x) \, d\mu(x) = \alpha_j \), for \( 1 \leq j \leq s \).

(a) If \( \mathcal{X} \) is a finite set with \( \mu(\{ x \}) > 0 \) for all \( x \in \mathcal{X} \), show that the optimal \( p^* \) is a member the \( s \)-parameter exponential family

\[
p_{\eta^*}(x) = e^{\eta^* T(x) - A(\eta^*)},
\]

with parameter \( \eta^* \in \mathbb{R}^s \) chosen so that \( p_{\eta^*} \) satisfies the constraints. (**Hint**: use Lagrange multipliers).

(b) Blithely\(^1\) applying the result of (a) to \( \mathcal{X} = \mathbb{R} \), find the distribution that maximizes entropy with respect to the Lebesgue measure, subject to the constraint that \( \mathbb{E}(X) = \mu, \text{Var}(X) = \sigma^2 \).

5. Gamma family

The gamma family is a two-parameter family of distributions on \( \mathbb{R}_+ = [0, \infty) \), with density

\[
p_{k, \theta}(x) = \frac{x^{k-1}e^{-x/\theta}}{\Gamma(k)\theta^k}
\]

with respect to the Lebesgue measure on \( \mathbb{R}_+ \). \( k > 0 \) and \( \theta > 0 \) are respectively called the shape and scale parameters, and \( \Gamma(k) \) is the gamma function, defined as

\[
\Gamma(k) = \int_{0}^{\infty} x^{k-1}e^{-x} \, dx.
\]

The gamma distribution generalizes the exponential distribution (\( \text{Exp}(\theta) = \theta^{-1}e^{-x/\theta} = \Gamma(1, \theta) \)) and the chi-squared distribution (\( \chi^2_d = \frac{d}{2}^{-1/2}e^{-x/2} = \Gamma(d/2, 2) \)).

\(^1\)Meaning naively, without any concern that anything new might go wrong in a continuous space.
(a) Show that the Gamma is a 2-parameter exponential family by putting it into its canonical form. Find the natural parameter, sufficient statistic, carrier density, and log-partition function (Note: there are multiple valid ways of doing this).

(b) Find the mean and variance of $X \sim \Gamma(k, \theta)$.

(c) Find the moment generating function of $X \sim \Gamma(k, \theta)$:

$$M_X(u) = \mathbb{E}_{k, \theta}[e^{uX}],$$

and use it to find the distribution of $X_+ = \sum_{i=1}^{n} X_i$ where $X_1, \ldots, X_n$ are mutually independent with $X_i \sim \Gamma(k_i, \theta)$.

You may use without proof the following uniqueness result about MGFs: If $Y$ and $Z$ are two random variables whose MGFs coincide in a neighborhood of 0 ($\exists \delta > 0$ for which $M_Y(u) = M_Z(u) < \infty$ for all $u \in [-\delta, \delta]$), then $Y$ and $Z$ have the same distribution.