Assignment 5

(1) Consider an i.i.d. sample $X_1, \dots, X_n$ from the Cauchy location family $\mathcal{C}(\mu, 1)$, where each observation has density

$$p(x;\mu) = \frac{1}{\pi}\frac{1}{1+(x-\mu)^2}, \qquad \mu \in \mathbb{R}.$$

(a) Let $M_n = X_{(k_n)}$, where $k_n = \lceil n/2 \rceil$, be the sample median. Derive its asymptotic distribution. You may use the following order-statistic limit: if F has density f, $\xi_q = F^{-1}(q)$, f is continuous and strictly positive at ξ_q, and

$$\frac{k_n}{n} \longrightarrow q, \qquad \sqrt{n}\left(\frac{k_n}{n} - q\right) \longrightarrow 0,$$

then

$$\sqrt{n}\big(X_{(k_n)} - \xi_q\big) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{q(1-q)}{f(\xi_q)^2}\right).$$

(b) Let $\widehat{\mu}_n$ denote the maximum likelihood estimator. Apply Cramér's theorem to derive the limiting distribution of

$$\sqrt{n}(\widehat{\mu}_n - \mu),$$

including the computation of the Fisher information. You may assume that all regularity conditions required by Cramér's theorem are satisfied.

(c) Define the asymptotic relative efficiency of the median relative to the MLE by

$$\mathrm{ARE}(M_n, \widehat{\mu}_n) = \frac{\text{asymptotic variance of } \widehat{\mu}_n}{\text{asymptotic variance of } M_n}.$$

Compute this quantity.

(d) If statistician A has n observations and uses the MLE to estimate μ, while statistician B uses the median, how large a sample size must statistician B use to obtain the same asymptotic variance as statistician A? State one statistical advantage of using the median.

(e) Make the same comparison between the MLE and the median for the parameter μ in the family $\mathcal{N}(\mu, \sigma^2)$, taking σ^2 to be known if desired.

(2) Let $Y \in \mathbb{R}^n$ and $X \in \mathbb{R}^{n\times p}$, and let $\Gamma \in \mathbb{R}^{n\times n}$ and $K \in \mathbb{R}^{p\times p}$ be symmetric positive-definite matrices. Define

$$\|u\|_\Gamma^2 = u^\top \Gamma u, \qquad \|\beta\|_K^2 = \beta^\top K \beta,$$

and let $\Gamma^{1/2}$ and $K^{1/2}$ denote the symmetric positive-definite square roots of Γ and K, respectively.

Recall that for an ordinary least-squares problem of minimizing $\|Z - W\beta\|^2$, the minimizing residual satisfies the orthogonality relation

$$\langle Z - W\widehat{\beta}, Wb\rangle = 0 \qquad \text{for every } b \in \mathbb{R}^p,$$

and, when $W^\top W$ is invertible,

$$\widehat{\beta} = (W^\top W)^{-1} W^\top Z.$$

(a) For $\lambda > 0$, consider the generalized ridge-regression problem

$$\underset{\beta\in\mathbb{R}^p}{\text{minimize}} \left\{ \|Y - X\beta\|_\Gamma^2 + \lambda\|\beta\|_K^2 \right\}.$$

As shown in class for ridge regression, this problem can be solved as an ordinary least-squares problem by finding an augmented response vector Z and an augmented design matrix W. Find Z and W, and verify that

$$\|Z - W\beta\|^2 = \|Y - X\beta\|_\Gamma^2 + \lambda\|\beta\|_K^2.$$

(b) Apply the ordinary least-squares formula above to find the minimizer $\widehat{\beta}$. Explain why the matrix that must be inverted is nonsingular.

(3) Let $\mathcal{H}$ be an RKHS of functions on χ. Show that its reproducing kernel is unique: if K and $\widetilde{K}$ both have the reproducing property for point evaluation in $\mathcal{H}$, then $K = \widetilde{K}$.

(4) Let $\mathcal{H}$ be an RKHS of functions on χ with reproducing kernel K. Let P be a probability measure on χ, and suppose that

$$\int K(x,x)\, dP(x) < \infty.$$

For $f \in \mathcal{H}$, define

$$T_P f = \int f(x) K(x,\cdot)\, dP(x).$$

(a) Show that T_P is a bounded linear operator from $\mathcal{H}$ into itself. Also show that, for all $f, g \in \mathcal{H}$,

$$\langle T_P f, g\rangle_{\mathcal{H}} = \int f(x) g(x)\, dP(x).$$

Deduce that T_P is self-adjoint and nonnegative.

(b) Let (X, Y) be defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, suppose that the distribution of X is P, and assume that $\mathbb{E}[Y^2] < \infty$. Define

$$g_P = \mathbb{E}\big[YK(X, \cdot)\big].$$

For $\lambda > 0$, we wish to sove the (true population) regularized least-squares problem

$$\underset{f \in \mathcal{H}}{\text{minimize}} \left\{\mathbb{E}\big[(Y - f(X))^2\big] + \lambda\|f\|_{\mathcal{H}}^2\right\}.$$

As in ridge regression, find an augmented response Z and a linear operator W such that, with respect to a natural norm on the left hand side,

$$\|Z - Wf\|^2 = \mathbb{E}\big[(Y - f(X))^2\big] + \lambda\|f\|_{\mathcal{H}}^2.$$

Then use an extension of the linear model least-squares orthogonality relation $\langle Y - X\widehat{\beta}, X\beta\rangle = 0$ for all $\beta \in \mathbb{R}^p$, in the form

$$\langle Z - Wf_\lambda, Wh\rangle = 0 \quad \text{for every } h \in H$$

to deduce that any minimizer f_λ must satisfy the equation

$$(T_P + \lambda I)f_\lambda = g_P.$$

(5) Let $\mathcal{H}$ be an RKHS of functions on χ. Show that convergence in norm implies pointwise convergence: if $f_n, f \in \mathcal{H}$ and

$$\|f_n - f\|_{\mathcal{H}} \longrightarrow 0,$$

then

$$f_n(x) \longrightarrow f(x) \qquad \text{for every } x \in \chi.$$

(Hint: Use the reproducing property together with the Cauchy–Schwarz inequality.)

(6) Let $\mathcal{H}$ be an RKHS of functions on χ with kernel K. For $n \geq 1$ and $\boldsymbol{x} = (x_1, \ldots, x_n)$, define

$$\mathbb{L}_{n,\boldsymbol{x}} = \left\{\sum_{i=1}^{n} \alpha_i K(x_i, \cdot) : \alpha_1, \ldots, \alpha_n \in \mathbb{R}\right\}.$$

Suppose that $f, g \in \mathbb{L}_{n,\boldsymbol{x}}$ and

$$f(x_i) = g(x_i), \qquad i = 1, \ldots, n.$$

Prove that $f = g$ in $\mathcal{H}$.

(7) Consider the Cauchy location family

$$p(x;\theta) = \frac{1}{\pi\big(1 + (x-\theta)^2\big)}, \qquad \theta \in \mathbb{R}.$$

Construct the Neyman–Pearson test of $H_0 : \theta = 0$ versus $H_1 : \theta = 1$ based on one observation X. Show that, for some values of the significance level α, the rejection region is disconnected and, in particular, is not of the form $[t, \infty)$.

(8) Construct Neyman–Pearson tests for simple-versus-simple hypotheses in classical families such as the Poisson, Gamma, Beta, Uniform, and Binomial families. For a multiparameter family, you may either vary one parameter while holding the others fixed, or specify all parameters under both hypotheses. For each example you consider, compute the power function of the test—that is, the probability of rejecting H_0 when the alternative is true—as a function of the null and alternative parameters, the sample size, and the Type I error level α.

(9) Compute the Type I error and power of the randomized Neyman–Pearson test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, where $\theta_0 < \theta_1$, based on an i.i.d. sample $X_1, \ldots, X_n$ from $\mathcal{U}[0,\theta]$, $\theta > 0$. The test rejects H_0 with probability 1 on the event $\{\theta_0 < X_{(n)} \le \theta_1\}$ and with probability α on the event $\{X_{(n)} \le \theta_0\}$.