Assignment 3

(1) Consider the space of random variables $\mathcal{H}$ with finite second moment on a probability space $(\Omega, \mathcal{F}, P)$, with inner product $\langle X, Y\rangle = E[XY]$. Show for a given sub sigma algebra $\mathcal{G}$ of $\mathcal{F}$ that the mapping $P : \mathcal{H} \to \mathcal{H}$ defined by

$$PX = E[X|\mathcal{G}]$$

satisfies the two properties of an orthogonal projection:

(a) $P^2 = P$

(b) P is symmetric, that is, $\langle PX, Y\rangle = \langle X, PY\rangle$.

Onto what space is P projecting?

(2) Suppose that X is supported on $[0, \infty)$ and satisfies

$$P(X > s + t|X > s) = P(X > t) \quad \forall s > 0, t > 0.$$

Prove that X has the exponential distribution for some $\lambda > 0$.

(3) Let $X \in \mathbb{R}^{n\times p}$, where $p > n$, and suppose that X has full row rank. Assume that

$$y = X\beta,$$

so that there is no observational noise. Define

$$\widehat{\beta} = X^{\mathsf{T}}(XX^{\mathsf{T}})^{-1}y, \qquad P_X = X^{\mathsf{T}}(XX^{\mathsf{T}})^{-1}X.$$

(a) Show that

$$\widehat{\beta} = P_X\beta \qquad \text{and} \qquad X\widehat{\beta} = X\beta.$$

(b) Show that

$$\widehat{\beta} - \beta = -(I_p - P_X)\beta,$$

and that $\widehat{\beta}$ and $\widehat{\beta} - \beta$ are orthogonal.

(c) Let $x_0 \sim N(0, I_p)$ be independent of X. Show that

$$\mathbb{E}_{x_0}\left[\left(x_0^{\mathsf{T}}\widehat{\beta} - x_0^{\mathsf{T}}\beta\right)^2 \,\middle|\, X\right] = \|(I_p - P_X)\beta\|_2^2 .$$

(d) Now suppose that the entries of X are independent standard normal random variables. Using

$$\mathbb{E}P_X = \frac{n}{p}I_p,$$

show that

$$\mathbb{E}\left(x_0^\intercal \widehat{\beta} - x_0^\intercal \beta\right)^2 = \left(1 - \frac{n}{p}\right) \|\beta\|_2^2.$$

(e) Explain how $\widehat{\beta}$ can have zero training error while having positive prediction error.

(4) Let $X_1, \ldots, X_n$ be independent with distribution $\mathcal{N}(\mu, \sigma^2)$, with $\mu \in \mathbb{R}$ unknown, and $\sigma^2 > 0$ known. We want to test the (null) hypothesis H_0, that the true value of the mean is μ_0. Let

$$T = \frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma}$$

(a) Show that under H_0 the statistic $T \sim \mathcal{N}(0, 1)$.

(b) Suppose we decide to perform our test in the following way. We will set a threshold $t > 0$ and if $|T| \geq t$ ($\overline{X}$ deviates from the hypothesized mean by "too much" relative to its normal variation on the scale of σ) we will reject H_0, and otherwise not reject. If we want to have our false alarm probability (the probability of declaring the null invalid) at some small level $\alpha \in (0, 1)$, how should you set t?

(c) Suppose now we test as with t determined in the previous part. Now we want to know with what probability the test can detect when the actual value of the mean is not given by the null, but by some other value μ. This function of μ is called the "power" of the test, and can be written out

$$\beta(\mu) = P_\mu(|T| \geq t).$$

Determine the power function of the test for all $\mu \in \mathbb{R}$ and make a plot of it.

(d) Explore the behavior of the test as α varies, and as σ varies.

(5) (*) Under our usual assumptions for the linear model with multivariate normal errors, for testing the hypothesis $H_0 : D\beta = 0$ with $D \in \mathbb{R}^{q \times p}, r(D) = q \leq p$, let

$$F = \frac{\|\widehat{Y} - \widehat{Y}_{M_0}\|^2 / q}{\|Y - \widehat{Y}\|^2 / (n - p)}$$

where $\widehat{Y}_{M_0} = P_{M_0}Y$, with P_{M_0} the orthogonal projection onto the subspace $M_0 = \{X\beta : D\beta = 0\}$. Show that when H_0 is true then $F \sim F_{q,n-p}$.

(6) Show that the F statistic in the previous problem may be written as

$$F = \frac{(D\widehat{\beta} - c)^T[D(X^TX)^{-1}D^T]^{-1}(D\widehat{\beta} - c)}{qS^2},$$

with $c = 0$. It may be helpful to define

$$G = X(X^\top X)^{-1}D^\top \quad \text{and show} \quad \mathcal{R}(G) = \mathcal{R}(X) \cap M_0^\top$$

and hence that $G(G^\top G)^{-1}G^\top$ is the orthogonal projection onto the subspace W.

(7) Compute, the F statistic for testing the hypothesis that $H_0 : \beta_2 = 0$ in the simple linear regression setting: $Y = \beta_1 + \beta_2 x_2 + \epsilon$. Interpret the resulting formula, and give intuition as to how the data needs to behave in order that this F be large.

(8) Same as the problem above, for the test of $\beta_1 = \beta_2$ for the model $Y = \beta_1 x_1 + \beta_2 x_2 + \epsilon$.

(9) Reduce the problem of least squares estimation under the constraint $D\beta = c$ to one under the constraint $D\beta = 0$, and use that reduction to give the least squares estimate of the β parameter in general.

(10) Use the results of the previous problem to show (6).

(11) Show how the independent two sample t test with possibly unequal sample sizes can be derived as a special instance of the F test. Can one generalize this result to the natural k sample extension? (No need to work any computations to their endpoints. If an extension exists, show how one effects the generalization, and if not explain why it cannot be done.)

(12) Having made n_1, n_2 and n_3 measurements on the three angles of a triangle, each estimate unbiased, and having errors of equal variance, find the least squares estimates of the three angles subject to the constraint that they sum to π.