Problem Set 4
Estimation and Decision Theory

For the Exercise Session on Oct 29 — **Due: Tue, November 4, 10am, on Moodle**

1 Problem for Class

Problem 1: Tweedie's Formula

For the special case where $X = D + N$, where N is Gaussian noise of mean zero and variance σ^2, *Tweedie's formula* says that the conditional mean (that is, the MMSE estimator) can be expressed as

$$\mathbb{E}\left[D \mid X = x\right] = x + \sigma^2 \ell'(x), \tag{1}$$

where

$$\ell'(x) = \frac{d}{dx} \log f_X(x), \tag{2}$$

where $f_X(x)$ denotes the marginal PDF of X. In this exercise, we derive this formula.

(a) Assume that $f_{X|D}(x|d) = e^{\alpha dx - \psi(d)} f_0(x)$ for some functions $\psi(d)$ and $f_0(x)$ and some constant α (such that $f_{X|D}(x|d)$ is a valid PDF for every value of d). Define

$$\lambda(x) = \log \frac{f_X(x)}{f_0(x)}, \tag{3}$$

where $f_X(x)$ is the marginal PDF of X, i.e., $f_X(x) = \int f_{X|D}(x|\delta) f_D(\delta) d\delta$. With this, establish that

$$\mathbb{E}\left[D \mid X = x\right] = \frac{1}{\alpha} \frac{d}{dx} \lambda(x). \tag{4}$$

(b) Show that the case where $X = D + N$, where N is Gaussian noise of mean zero and variance σ^2, is indeed of the form required in Part *(a)* by finding the corresponding $\psi(d), f_0(x)$, and α. Show that in this case, we have

$$\frac{f_0'(x)}{f_0(x)} = -\frac{x}{\sigma^2}, \tag{5}$$

and use this fact in combination with Part *(a)* to establish Tweedie's formula.

2 The Homework

Problem 2: Bernoulli Data with Beta Prior

Let $S \in [0, 1]$ be distributed with a Beta distribution with parameters $(1/2, 1/2)$, which, as we have seen in class, is $p(s) = \frac{1}{\pi} s^{-\frac{1}{2}} (1 - s)^{-\frac{1}{2}}$. We make n observations $X_1, X_2, \ldots, X_n$ that are (conditionally) independent Bernoulli(S) random variables.

a) Calculate the conditional distribution $p(s|x_1, x_2, \dots, x_n)$. Express it in terms of the integer t, which is the number of '1's in the sample $(x_1, x_2, \dots, x_n)$.

Hint: For $a, b \in \mathbb{R}^+$, we have $\int_0^1 y^{a-1}(1-y)^{b-1} dy = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, where $\Gamma(\cdot)$ denotes the Gamma function.

b) We would like to estimate S from $X_1, X_2, \dots, X_n$ such as to minimize the mean-squared error $\mathbb{E}[(S - \hat{S}(X_1, X_2, \dots, X_n)^2]$. Find the optimum estimate $\hat{S}(X_1, X_2, \dots, X_n)$. Simplify your result as much as possible.

Hint: The Gamma function satisfies the property, for $c \in \mathbb{R}^+$, that $\Gamma(c+1) = c\Gamma(c)$.

3 Additional Problems

Problem 3: Parameter Estimation and Fisher Information

Find the Fisher information for the following families:

(a) $f_\theta(x) = N(0, \theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{x^2}{2\theta}}$

(b) $f_\theta(x) = \theta e^{-\theta x}, x \geq 0$

(c) What is the Cramèr Rao lower bound on $\mathbb{E}_\theta(\hat{\theta}(X) - \theta)^2$, where $\hat{\theta}(X)$ is an unbiased estimator of θ for (a) and (b)?

Problem 4: Wiener Filter and Irrelevant Data

As we have seen in class, the (FIR) Wiener filter is given by

$$\mathbf{w} \quad = \quad R_x^{-1} \mathbf{r}_{dx}, \tag{6}$$

where R_x is the autocorrelation matrix of the data that's being used, and $\mathbf{r}_{dx}$ is the cross-correlation between the data and the desired output. For this to be well defined, R_x should be full rank. In this problem, we study this question in more detail.

(a) In many applications, the signal acquisition process is noisy. That is, the data $x[n] = s[n] + w[n]$, where $s[n]$ is an *arbitrary* signal, and $w[n]$ is white noise. Prove that in this case, the $p-$dimensional autocorrelation matrix R_x is full rank (i.e., invertible) for any p. (Note: Be careful not to make *any* assumptions about the signal $s[n]$.)

(b) In some other cases, R_x could be rank-deficient. To study this, prove first that if the (FIR) Wiener filter based on the data $\mathbf{x} = \{x[n]\}_{n=0}^{p-1}$ is $\mathbf{w}$, then the (FIR) Wiener filter based on the modified data $A\mathbf{x}$ (where A is an invertible matrix) is $A^{-H}\mathbf{w}$, (where we use the relatively common notation $A^{-H} = (A^{-1})^H$).

(c) Explain how to find the (FIR) Wiener filter when R_x is rank-deficient. Discuss existence and uniqueness. *Hint:* Use Part *(b)* to transform your data to a more convenient basis.

Problem 5: Fisher Information and Divergence

Suppose we are given a family of probability distributions $\{p(\cdot\,;\theta) : \theta \in \mathbb{R}\}$ on a set $\mathcal{X}$, parametrized by a real valued parameter θ. (Equivalently, a random variable X whose distribution depends on θ.) Assume that the parametrization is smooth, in the sense that

$$p'(x;\theta) := \frac{\partial}{\partial\theta} p(x;\theta) \quad \text{and} \quad p''(x;\theta) := \frac{\partial^2}{\partial\theta^2} p(x;\theta)$$

exist. (Note that the derivatives are with respect to the parameter θ, not with respect to x.) We will use the notation $\mathbb{E}_{\theta_0}[\cdot]$ to denote expectations when the parameter is equal to a particular value θ_0, i.e., $\mathbb{E}_\theta[g(X)] = \sum_x p(x;\theta)g(x)$.

Define the function $K(\theta,\theta') := D\big(p(\,\cdot\,;\theta)\|p(\,\cdot\,;\theta')\big)$.

(a) Show that for any θ_0, $\dfrac{\partial}{\partial\theta}K(\theta,\theta_0) = \displaystyle\sum_x p'(x;\theta)\log\frac{p(x;\theta)}{p(x;\theta_0)}$.

(b) Show that $\dfrac{\partial^2}{\partial\theta^2}K(\theta,\theta_0) = \displaystyle\sum_x p''(x;\theta)\log\frac{p(x;\theta)}{p(x;\theta_0)} + J(X;\theta)$ with

$$J(X;\theta) := \mathbb{E}_\theta\Big[\big(p'(X;\theta)/p(X;\theta)\big)^2\Big].$$

(c) Show that when θ is close to θ_0

$$K(\theta,\theta_0) = \tfrac{1}{2}J(X;\theta_0)(\theta-\theta_0)^2 + o((\theta-\theta_0)^2)$$

(d) Show that $J(X;\theta) = -\mathbb{E}_\theta\big[\frac{\partial^2}{\partial\theta^2}\log p(X;\theta)\big]$.