Homework 7: Mathematical Statistics (MATH-UA 234)

Due 12/08 at the beginning of class on Gradescope. The quiz will still be 12/06, and will cover content from problems 1-3 (i.e. Bayesian inference). No solutions will be posted prior to the quiz.

Reminder. Remember than the project presentations are on December 14th!

Problem 1. *Suppose* $X_1, \dots, X_n \sim \mathrm{Ber}(p)$ *(with* 1 *representing heads and zero representing tails) and that we use the prior distribution* $p \sim \mathrm{Beta}(\alpha, \beta)$.

(a) *Compute the posteriori distribution for* $p | X_1 = x_1, \dots, X_n = x_n$.

(b) *For each of the coins below, find values of* α *and* β *so that your prior distribution represents your belief about the parameter p of the coin. Plot and label these 6 prior distributions. Note that the head side is the side marked with the number.*

(c) *Suppose you flipped coin zero and got 53 heads and 47 tails. Make a plot showing the prior and posterior densities for p.*

(d) *Suppose you flipped coin 4 and got 39 heads and 61 tails. Make a plot showing the prior and the posterior densities for p.*

(e) *Suppose you flipped coin 6 and got 0 heads and 100 tails. Make a plot showing the prior and the posterior densities for p.*

(f) *For the coin 6 example, is the probability that* $p = 0$ *under your posterior 100%? Does this make sense? Why or why not?*

This image was taken from this site: `https://izbicki.me/blog/how-to-create-an-unfair-coin-and-prove-it-with-math.html`

Problem 2 (Wasserman 11.1). *Suppose* $X_1, \dots, X_n \sim N(\theta, \sigma^2)$, *and that we use the prior distribution* $\theta \sim N(a, b^2)$. *Show that* $\theta | X_1 = x_1, \dots, X_n = x_n \sim N(\bar{\theta}, \tau^2)$ *where*

$$\bar{\theta} = w \frac{x_1 + \cdots + x_n}{n} + (1 - w)a, \qquad w = \frac{1/\mathrm{se}^2}{1/\mathrm{se}^2 + 1/b^2}, \qquad \tau = 1/\sqrt{1/\mathrm{se}^2 + 1/b^2}, \qquad \mathrm{se} = \sigma/\sqrt{n}.$$

problems with a textbook reference are based on, but not identical to, the given reference

Problem 3 (Wasserman 11.2). *Let* $X_1, \ldots, X_n \sim N(\mu, 1)$.

(a) *Simulate a dataset (using* $\mu = 5$*) consisting of* $n = 100$ *observations*

(b) *Take* $f(\mu) = 1$ *as the prior density, and find the posterior density given the observed data. Plot this density*

Problem 4. *Consider a model of the form* $f(x) = \hat{\beta}_0 + \hat{\beta}_1 y$ *and, given data*

$$(X_1, Y_1), \ldots, (X_n, Y_n),$$

define the loss function

$$L(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} (Y_i - f(X_i))^2.$$

(a) *Compute the partial derivatives* $\partial L(\hat{\beta}_0, \hat{\beta}_1)/\partial \hat{\beta}_0$ *and* $\partial L(\hat{\beta}_0, \hat{\beta}_1)/\partial \hat{\beta}_1$

(b) *Find the minimizers* $\hat{\beta}_0$ *and* $\hat{\beta}_1$ *for* $L(\hat{\beta}_0, \hat{\beta}_1)$.

(c) *Show that you can write the loss function in the form* $\|\vec{b} - \vec{A}\vec{x}\|_2^2$, *where* $\vec{b}$ *is a particular vector of length n,* $\vec{A}$ *is a* $n \times 2$ *matrix, and* $\vec{x}$ *is a length 2 vector.*

Problem 5. *Consider the following four data sets:*

```
x1 = [10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5]
y1 = [8.04, 6.95, 7.58, 8.81, 8.33, 9.96, 7.24, 4.26, 10.84, 4.82, 5.68]

x2 = [10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5]
y2 = [9.14, 8.14, 8.74, 8.77, 9.26, 8.10, 6.13, 3.10, 9.13, 7.26, 4.74]

x3 = [10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5]
y3 = [7.46, 6.77, 12.74, 7.11, 7.81, 8.84, 6.08, 5.39, 8.15, 6.42, 5.73]

x4 = [8, 8, 8, 8, 8, 8, 8, 19, 8, 8, 8]
y4 = [6.58, 5.76, 7.71, 8.84, 8.47, 7.04, 5.25, 12.50, 5.56, 7.91, 6.89]
```

(a) *Find the sample mean and sample variance of each datasets' X and Y values. Compute the sample correlation between the datasets.*

(b) *Find the linear regression line and compute the* R^2 *value for each dataset.*

(c) *Now, plot the datasets and the linear regression lines. Explain what happened.*

Problem 6 (Wasserman 13.2). *Suppose* $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, *where* $\mathbb{E}[\epsilon_i | X_i] = 0$ *and* $\mathbb{V}[\epsilon_i | X_i] = \sigma^2$.

Let $\hat{\beta}_0$ *and* $\hat{\beta}_1$ *be the least squares estimates given in Theorem 13.4. Show that* $\mathbb{E}[\hat{\beta}_0 | X_1, \ldots, X_n] = \beta_0$ *and* $\mathbb{E}[\hat{\beta}_1 | X_1, \ldots, X_n] = \beta_1$. *You should regard* $X_1, \ldots, X_n$ *as constant.*

Problem 7. *Pick at least one of the following articles to read. Provide a one paragraph summary of what you think the most important points of the article were. Discuss how this is relevant to what we are learning in class.*

- *Why algorithms can be racist and sexist*
- *All the Ways Hiring Algorithms Can Introduce Bias*
- *Racial Discrimination in Face Recognition Technology*