STAT 391
Homework 3
Out April 12, 2018
Due April 19, 2018

©Marina Meilă
mmp@cs.washington.edu

[Problem 1 – CDF's and densities – Not graded]

Let

$$F(x) = \begin{cases} 0, & x \leq 0 \\ x^2, & 0 < x \leq 1 \\ 1, & 1 < x \end{cases} \tag{1}$$

and

$$G(x) = \begin{cases} 0, & x \leq 0 \\ 2x^2, & 0 < x \leq 0.5 \\ 1 - 2(1-x)^2 & 0.5 < x \leq 1 \\ 1, & 1 < x \end{cases} \tag{2}$$

be two cumulative distribution functions.

1. – **Not graded** Plot F, G (OK to do by hand, but make a neat drawing, labelling all the important coordinates).

2. Compute their corresponding densities f and g. Plot them on a graph (OK to do by hand, but make a neat drawing, labelling all the important coordinates).

3. Denote by P_F and P_G the probability distributions defined by F, G. Find a, a' such that $P_F(0, a) = P_F(a, 1)$ and $P_G(0, a') = P_G(a', 1)$. a, a' i.e. the *medians* of the distributions P_F, P_G. Represent x, x' on the graphs in questions **1., 2.**

4. – **Not graded** Find the probabilities of the following intervals $[0, 0.25]$, $[1, 1.75]$ under P_F, P_G.

5. – **Not graded** Find the shortest interval $[a_F, b_F]$ that has probability 0.1 under F. Find the shortest interval $[a_G, b_G]$ that has probability 0.1 under G. Motivate your answer. Show the intervals $[a_F, b_F]$, $[a_G, b_G]$ on the graphs of f, g respectively.

6. Calculate the means of f, g, denoted by $E_f[X]$, $E_g[X]$.

Problem 2

The exponential distribution with parameter $\gamma > 0$ is defined by the density $f_\gamma(x) = \gamma e^{-\gamma x}$ on $S = [0, \infty)$.

1. Denote by p_n the probability of the interval $[n-1, n)$ under the exponential distribution, i.e. $p_n = Pr[\, x \in [n-1, n)\,]$ for $n = 1, 2, \ldots$. What is the expression of p_n as a function of γ and n? What is this expression if $\gamma = \ln 2$?
2. What is the expression of $\frac{p_n}{p_{n+1}}$ as a function of γ and n? What is this expression if $\gamma = \ln 2$?
3. Plot on the same graph the densities $f_\gamma(x)$ for $\gamma = \ln 2, \ln 3, \ln 4$.
4. Let $g(x) = \frac{1}{Z} e^{-\gamma(x+3)}, \;\; x \in S = [-3, \infty)$. Evaluate the normalization constant Z as a function of γ. Evaluate the expression of the CDF G of this distribution.

[Problem 3 - Not graded]

A rabbit can jump a distance x, which is random an uniformly distributed between 1 and 3 ft. A fox can jump a distance y which is uniformly distributed between 0.75 and 1.5ft, and is independent of the distance jumped by the rabbit (if any rabbit is present).

a. Make a neatly labeled sketch of the densities f_X, f_Y of the rabbit's and fox's jump lenghts.

b. What are the CDF's of x and y? Write their expressions F_X, F_Y graph them.

c. What is the probability that a rabbit jumps more than 2.5ft? What is the probability that a fox jumps less than 1 ft?

d. A fox is $d = 1$ft away from an unsuspecting rabbit. What is the probability that the fox will catch the rabbit, if the fox jumps once directly towards the rabbit and the rabbit is too surprised to move? To catch a rabbit, this fox must land within 0.2ft of the rabbit.

e. The same question, assuming now $d = 1.4$ft.

f. The fox is now only 0.5 ft from the rabbit, but the rabbit also takes a jump away from the fox. What is the probability that the fox will catch the rabbit, assuming that the fox jumped $y = 1ft$?

g. Make a plot of the probability that the fox catches the rabbit under the conditions in **f.** ($d = 0.5$ft, rabbit jumps)q, as a function of y.

Problem 4 – Rayleigh distribution

The Rayleigh parametrized family of distributions is described by

$$f(r; a) = \frac{r}{a^2} e^{-\frac{r^2}{2a^2}} \quad r \geq 0 \tag{3}$$

If we shoot at a target centered at $(0, 0)$ and our bullets hit at (x, y), where each of x, y is normally distributed with mean $\mu = 0$, then the distance $r = \sqrt{x^2 + y^2}$ from the target center is distributed according to a Rayleigh distribution.

Assume you are given a data set $\mathcal{D} = \{r_1, r_2, \ldots r_n\}$ drawn independently from a Rayleigh distribution. The task is to determine the formula for the ML estimate of the parameter a as a function of the data.

1. Write the formula of the likelihood of a, $L(a)$.

2. Take the logarithm of $L(a)$ to obtain the log-likelihood $l(a) = \log L(a)$. Then compute the derivative

$$\frac{\partial l}{\partial a}$$

3. Now solve the equation $\frac{\partial l}{\partial a} = 0$ to obtain a formula for a^{ML} as a function of the data.

4. Does this problem have sufficient statistics? What are they (is it)?

Problem 4 – Two random samples

James and Yali are two statistics students who want to estimate the distribution of running times of a machine learning program they have just written. James runs the program n_1 times, and obtains times $\{t_1, \ldots t_{n_1}\}$. Then Yali runs the same program n_2 times, and obtains times $\{t_{n_1+1}, \ldots t_{n_1+n_2}\}$. The each run time is independent of all the others, and is sampled from the same exponential distribution with unknown parameter γ.

1. Write the expression of $l(\gamma)$, the log-likelihood of the data obtained by the two students. Motivate your answer. Then maximize $l(\gamma)$ w.r.t γ to obtain the expression of the Maximum Likelihood estimate γ^{ML}.

2. Let $n_1 = 2, n_2 = 5$, James' times be $3.5, 0.8$ seconds, and Yali's times be $4.2, 0.5, 1.1, 2.0, 0.3$ seconds. Calculate the Maximum Likelihood of γ for these data, using the formula obtained in **1.**.

3. Suppose now that $n_1 = 2,000,000$, $n_2 = 5,000,000$ samples. James has left town to attend a conference after performing his part of the experiment, and has neglijently taken with him all his data $t_1, \ldots t_{n_1}$ on his laptop. Yali needs to perform the estimation as in **1.** but has no information about the experiment

James performed, except that he took some samples from the same f_γ as she. How many numbers must James send her, so that she can correctly estimate γ^{ML}? What are these numbers and how should Yali use them? *Prove your answer.*

Problem 5 – Least Squares

Let $x_1, x_2, \ldots x_n$ be real numbers, and define by $g(z)$ the function

$$g(z) \;=\; \sum_{i=1}^{n}(x_i - z)^2$$

Show that the minimum of g is attained for

$$z^* \;=\; \frac{1}{n}\sum_{i=1}^{n} x_i$$

What is the value $g(z^*)$?

[Hint: Take the derivative of g w.r.t. z and solve the equation $g'(z) = 0$.]