Midterm: ARE 210

Fall 2015

Due: Wednesday, October 28, 2015

This exam is due by 5pm on Wednesday, October 28 at the start of lecture. No late exams will be accepted.

You may use *Introduction to Mathematical Statisitics* by Hogg, Craig, and McKean, DeGroot and Schervish's *Probability and Statistics*, Alder's *R in a Nutshell*, my lecture slides, this semester's problem sets and solutions, your own notes, and `R`'s built-in help system. You may not use any other resources, including notes or other materials from previous editions of this course. You may not discuss this exam with anyone until after the due date. Any violations of these policies will result in a 0 for this exam and could result in failure of the course and further disciplinary action.

If questions are unclear, you can e-mail me. Any clarification(s) that I may offer will be distributed via the class-wide e-mail list.

You must show your work for full credit. Any results given or proven in lecture or on a problem set can be used directly; otherwise, you should prove it for full credit. If you are not able to prove a result, partial credit will be given for logical explanations.

1 (13 pts) Transformation relations

Let $X_i \sim \text{Uniform}[0, 1], i = 1, \ldots, N$. Let $Y_i = \exp(X_i)$.

1. (4 pts) Derive the correlation between X and Y.

Consider the vector (x_i, y_i) paired to (x_j, y_j). The pair is *concordant* if $x_i < x_j$ and $y_i < y_j$ or $x_i > x_j$ and $y_i > y_j$. The pair is *discordant* if $x_i < x_j$ and $y_i > y_j$ or $x_i > x_j$ and $y_i < y_j$. A pair is neither if $x_i = x_j$ or $y_i = y_j$.

2. (2 pts) How many unique pairs are possible?

Consider the ratio

$$\frac{\text{number of concordant pairs} - \text{number of discordant pairs}}{\text{total number of pairs}}$$

3. (1 pt) What is the range of this ratio?

4. (2 pts) What is the value of this ratio for X_i and Y_i?

5. (2 pts) Under what circumstances does this ratio take on its minimum and maximum values?

6. (2 pts) When does (the standard measure of) correlation take on its minimum and maximum values?

2 (20 pts) Contamination implications

Suppose that $X_1, \ldots, X_N$ are i.i.d. with mean μ and variance σ^2. Suppose that X_{N+1} is contaminated such that $X_{N+1} = \mu + k$, both μ and k unknown, but not random.

Define:

$$
\begin{aligned}
\bar{X}_N &= \frac{1}{N}\sum_{i=1}^{N} X_i \\
\bar{X}_{N+1} &= \frac{1}{N+1}\sum_{i=1}^{N+1} X_i \\
\hat{\sigma}_N^2 &= \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \bar{X}_N\right)^2 \\
\hat{\sigma}_{N+1}^2 &= \frac{1}{N}\sum_{i=1}^{N+1}\left(X_i - \bar{X}_{N+1}\right)^2.
\end{aligned}
$$

1. (4 pts) Write $\bar{X}_{N+1}$ as a function of $\bar{X}_N$.

2. (2 pts) Find the expected value of $\bar{X}_{N+1}$ as a function of μ.

3. (4 pts) Write $\hat{\sigma}_{N+1}^2$ as a function of $\hat{\sigma}_N^2$.

4. (2 pts) Find the expected value of $\hat{\sigma}_{N+1}^2$ as a function of σ^2.

In practice, the researcher does not know which observation is contaminated (*i.e.*, he doesn't know which observation is the $N+1$st) or even whether a value is contaminated. Instead, he uses an outlier detection procedure. Specifically, he classifies as an outlier any observation that is more than c standard deviations away from the mean; *i.e.*, observation i is an outlier if

$$\left|X_i - \bar{X}_{N+1}\right| > c\sqrt{\hat{\sigma}_{N+1}^2}.$$

5. (8 pts) Find the values of k that would result in an observation being rejected as a function of $\hat{\sigma}_N^2$ and c.

3 (17 pts) Mixture MLEs

Let $Y_1 \sim N(\mu_1, \sigma^2)$ and $Y_2 \sim N(\mu_2, \sigma^2)$. We know that an observation X_i, $i = 1, \ldots, N$, comes from distribution 1 with probability p.

Using `R`, generate 100 observations with $\mu_1 = 0$, $\mu_2 = 5$, $\sigma = 1$, and $p = 0.5$.

1. (4 pts) Derive the likelihood for the simulated data.
2. (4 pts) Code the likelihood function using `R`.
3. (1 pt) Find the MLE values by setting the starting values for the means to 0 and the starting values for the other parameters to their true values.
4. (1 pt) Find the MLE values by setting the starting values for p to near 0 and the starting values for the other parameters to their true values.
5. (1 pt) Find the MLE values by setting the starting values to the true values of the parameters.
6. (6 pts) Using the results that maximize the likelihood, find the variance of the MLE.

4 (25 pts) Distribution of income

Vilfredo Pareto claimed that the distribution of household wealth could be described using what has come to be known as the Pareto distribution. In this question, we apply this distribution to household income.

For an income of $\$x$, the Pareto distribution implies:

$$F(x) = 1 - \left(\frac{1}{x}\right)^{\alpha}.$$

According to the U.S. Census Bureau, the distribution of income in 2014 is given in Table 1. These results are based on a survey of 99,461 households.

1. (8 pts) Derive (perhaps to a constant factor) the log likelihood using these data.
2. (3 pts) Derive the score using these data.

Table 1: Distribution of income, 2014

Income range	Percent in range
$0 - \$21,432$	20
$\$21,432 - \$41,186$	20
$\$41,186 - \$68,212$	20
$\$68,212 - \$112,262$	20
$\$112,262 - \$206,568$	15
$\$206,268+$	5

3. (4 pts) Code the log likelihood and score functions using `R`.
4. (2 pts) Find the MLE for α.
5. (4 pts) Using each of the six income buckets in Table 1 separately, give a MOM estimator for α.
6. (4 pts) How do all these estimates compare to one another? What conclusion do you draw?

5 (25 pts) EL Chupacabra

A disadvantage of maximum likelihood estimation is that a parametric density must be specified. Consider the empirical density for X_i, ($i = 1, \ldots, N$ unique):

$$f(x_i) = \pi_i$$

for $0 \leq \pi_i \leq 1$ and $\sum_i \pi_i = 1$.

1. (3 pts) What is the empirical likelihood (*i.e.*, the likelihood based upon the empirical density)?
2. (5 pts) Derive the MLE for $\{\pi_i\}$.

Now, consider maximizing the log empirical likelihood subject to the moment conditions:

$$\sum_{i=1}^{N} \pi_i m(x_i; \theta) = 0,$$

where $\theta \in \mathbb{R}^K$ and $m(\cdot;\cdot) \in \mathbb{R}^R$, $R \geq K$.

3. First, maximize over $\{\pi_i\}$ and the Lagrange multipliers, holding θ fixed.

(a) (5 pts) What are the MLE for $\{\pi_i\}$?

(b) (4 pts) What are the equations for the Lagrange multipliers?

(c) (4 pts) How do the MLE for $\{\pi_i\}$ under the constrained optimization compare to the unconstrained MLEs? Provide an intuitive explanation.

(d) (2 pts) What is the purpose of adding the moment constraints?

4. (2 pts) Now, set up an unconstrained maximization problem for θ using a likelihood based on the MLEs for $\{\pi_i\}$ that you found in 3(a). (You do not need to solve it.)