Caltech Department of Mathematics

Ma 3/103
Introduction to Probability and Statistics

KC Border
Winter 2020

Assignment 6: Exercises on Maximum Likelihood Estimation; Confidence Intervals

Due Tuesday, February 25 by 4:00 p.m.
in the dropbox in the lobby of Linde Hall.

Instructions:

When asked for a probability or an expectation, give both a formula and an explanation for why you used that formula, and also give a numerical value when available.

When asked to plot something, use informative labels (even if handwritten), so the TA knows what you are plotting, attach a copy of the plot, and, if appropriate, the commands that produced it.

No collaboration is allowed on optional exercises.

Exercise 1 This exercise has been dropped.

□

Exercise 2 (25 pts) [1, Problem 7.4.9, pp. 399–400]:

Creativity, as any number of studies have shown, is very much a province of the the young. Whether the focus is music, literature, science, or mathematics an individual's best work seldom occurs late in life. Einstein, for example, made his most profound discoveries at the age of twenty-six; Newton, at the age of twenty-three. Robert Wood [2] compiled the following list of twelve scientific breakthroughs dating from the middle of the sixteenth century to to the early years of the twentieth century. All represented high-water marks in the careers of the scientists involved.

Discovery	Discoverer	Year	Age, y
Earth orbits sun	Copernicus	1543	40
Telescope, basic laws of astronomy	Galileo	1600	34
Principles of motion, gravity, calculus	Newton	1665	23
Nature of electricity	Franklin	1746	40
Burning is oxidation	Lavoisier	1774	31
Earth evolved by gradual processes	Lyell	1839	33
Natural selection and evolution	Darwin	1858	49
Field equations for light	Maxwell	1864	33
Radioactivity	Curie	1896	34
Quantum theory	Planck	1901	43
Special relativity, $e = mc^2$	Einstein	1905	26
Quantum wave equation	Schrödinger	1926	39

1. (10 pts) What are the sample average and standard deviation? What can be inferred from these data about the *true* age at which scientists do their best work? Answer the question by constructing a 95% confidence interval. Hint: To do this you need to assume the ages are normally distributed, and you will need to use the t-cutoffs, not the z-cutoffs.

2. (5 pts) Before constructing a confidence interval for a set of observations over a long time period, we should be convinced the y_i's exhibit no biases or trends. If, for example, the age at which scientists made major discoveries decreased from century to century, the the parameter μ would no longer be a constant, and the confidence interval would be meaningless. Plot the age versus date for these twelve discoveries. (Put the date on the abscissa.) Does the variability of the y_i's appear to be random with respect to time? (Later on we will come up with a more rigorous test.)

3. (10 pts) Give a good reason for questioning how much light these data shed on the age of greatest scientific creativity. Hint: How do you think the sample was constructed? How would you have picked your sample? □

Exercise 3 (20 pts)

I love hardware stores. One of the things you can now find at a decent hardware store is a laser "tape measure." It is battery powered and has a digital readout of the length. The length is measured with an error ε that has mean zero and small variance σ^2, which is independent of the length being measured.

Suppose you wish to measure two boards, of (unknown) lengths A and B, where A is obviously greater than B. You can make only two measurements before the battery in your laser tape measure dies. Errors on different measurements are stochastically independent.

You could (i) measure each board separately. Or you could (ii) align the boards on one end and measure the difference $A - B$ of their lengths, and then put them end-to-end and measure the sum $A + B$ of their lengths.

1. (5 pts) Using procedure (i), what is the variance of the measurement of A? What is the variance of the measurement of B?

2. (12 pts) Using procedure (ii), let D be the measurement of the difference $A - B$ in lengths, and let S be the measurement of the sum $A + B$ of lengths. Note that due to measurement error, D and S are random variables. Find the estimate a of the length A and the estimate b of the length B, in terms of S and D. What is the variance of the estimate a? What is the variance of the estimate b?

3. (3 pts) What is the better method for ascertaining the lengths? □

Exercise 4 (10 pts) How much total time did you spend on the previous exercises? **Please put the answer to this exercise on the *front page* of your answers and identify it as such.** □

Exercise 5 (Optional Exercise) (50 pts)

Make repeated independent draws from a uniform distribution on the interval $[0, 1]$ until the last draw is bigger than the previous draw. Let X be the count of draws (including the last draw). What is $\boldsymbol{E}\, X$?

Hint: This is a discrete-time Markov process with a continuous state space and an absorbing state, call it STOP.

Hint: The answer is very pretty. □

References

[1] R. J. Larsen and M. L. Marx. 2012. *An introduction to mathematical statistics and its applications*, fifth ed. Boston: Prentice Hall.

[2] R. M. Wood. 1970. Giant discoveries of future science. *Virginia Journal of Science* 21:169–177.