

## STA 312f23 Assignment Three<sup>1</sup>

Question 3e of this assignment **will not be on Quiz 3**, because we did not cover the necessary lecture material in time. It will be repeated on Assignment 4. It's given here just so you can start work on it if you have time.

As usual, the paper and pencil parts of this assignment are not to be handed in. They are practice for Quiz 3 on September 29th. The R part of Question 2 may be handed in as part of the quiz. **Bring hard copy of your printout to the quiz.** The printout should show your complete input and output. Do not write anything on your printout in advance except possibly your name and student number.

1. In an earlier assignment, you proved that the log-normal distribution has density

$$f(t|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp - \left\{ \frac{(\log(t) - \mu)^2}{2\sigma^2} \right\} \frac{1}{t}$$

- (a) Derive formulas for the maximum likelihood estimates of  $\mu$  and  $\sigma^2$ . Start with  $\mu$ . Show your work. Don't bother with the Hessian or the second derivative test.
- (b) Based on the R output below, give numerical values of the MLE. This is something you could do with a calculator.

```
length(x); mean(log(x)); var(log(x))
[1] 17
[1] -1.263583
[1] 5.114263
```

2. Dead pixels are a big problem in manufacturing computer and cell phone screens. The physics of the manufacturing process dictates that dead pixels happen according to a spatial Poisson process, so that the numbers of dead pixels in cell phone screens are independent Poisson random variables with parameter  $\lambda$ , the expected number of dead pixels. Naturally,  $\lambda$  depends on details of how the screens are manufactured.

In an effort to reduce the expected number of dead pixels, six assembly lines were set up, each with a different version of the manufacturing process. A random sample of 50 phones was taken from each assembly line and sent to the lab for testing. Mysteriously, three phones from one assembly line disappeared in transit, and 15 phones from another assembly line disappeared. Sample sizes and sample mean numbers of dead pixels appear in the table below.

	Manufacturing Process					
	1	2	3	4	5	6
ybar	10.68	9.87234	9.56	8.52	10.48571	9.98
n	50	47	50	50	35	50

---

<sup>1</sup>This assignment was prepared by [Jerry Brunner](#), Department of Mathematical and Computational Sciences, University of Toronto. It is licensed under a [Creative Commons Attribution - ShareAlike 3.0 Unported License](#). Use any part of it as you like and share the result freely. The L<sup>A</sup>T<sub>E</sub>X source code is available from the course website: <http://www.utstat.toronto.edu/brunner/oldclass/312f23>

The first task is to carry out a large sample likelihood ratio test to see whether the expected numbers of dead pixels are different for the six manufacturing processes. Using R, calculate the test statistic and the  $p$ -value. Also report the degrees of freedom.

You are being asked for a computation, but *most of the task is thinking and working things out on paper*. I got away with only six lines of code: One line to enter the means, one line to enter the sample sizes, two lines to set up the computation of  $G^2$ , one line to compute  $G^2$ , and one line to compute the  $p$ -value. Here are some little questions to get you started.

- (a) Derive the maximum likelihood estimate of  $\lambda$  for a single random sample from a Poisson distribution. Don't bother with the second derivative test.
- (b) Denote the parameter vector by  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)^\top$ . What is  $p$ ?
- (c) What is the null hypothesis?
- (d) What is the distribution of a sum of independent Poisson random variables, all with parameter  $\lambda$ ? You will *not* be asked to prove this on the quiz.
- (e) What is the distribution of  $n_j \bar{Y}_j$ ?
- (f) What is the likelihood function? Write it down and simplify a bit. Denote the data values by  $y_{ij}$ , for  $j = 1, \dots, p$  and  $i = 1, \dots, n_j$ .
- (g) What is the log likelihood? Simplify.
- (h) What is the unrestricted MLE  $\hat{\boldsymbol{\lambda}}$ ? It's a vector. Just write it down, based on your answer to [2a](#).
- (i) What is the restricted MLE  $\hat{\boldsymbol{\lambda}}_0$ ? *It's a vector*. If you just think about it, you can write down the answer based on the MLE in [2a](#). As a hint, I used the following notation for the mean of all the data pooled (thrown together in one big pot).  $\bar{Y} = \frac{1}{n} \sum_{j=1}^p n_j \bar{Y}_j$ , where  $n = \sum_{j=1}^p n_j$ .
- (j) Now you are ready to write the test statistic. There are a lot of cancellations. Keep simplifying!
- (k) Now use R to compute the test statistic and  $p$ -value. For comparison, my  $p$ -value is 0.01169142. If you got this, we must be doing everything else the same too.
- (l) What do you conclude at  $\alpha = 0.05$ ? Does manufacturing process affect the expected number of dead pixels? Answer Yes or No. Note that directional conclusions are not possible here. For that, we would need pairwise follow-up tests.

3. The Weibull distribution, which is never mentioned in most statistics classes, will play an important role in this course. The Weibull is written (parameterized) in quite a few different ways, even in the same book. Here is the density of the Weibull from p. 16 of our text.

$$f(t|\alpha, \lambda) = \begin{cases} \alpha\lambda(\lambda t)^{\alpha-1} \exp\{-(\lambda t)^\alpha\} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases},$$

where the parameters  $\alpha$  and  $\lambda$  are both greater than zero.

- (a) Verify that  $f(t|\alpha, \lambda)$  above really is a density, by showing that it integrates to one.
- (b) The Weibull density reduces to an exponential for one particular value of  $\alpha$ . What is that value?
- (c) Prove that if  $T$  has a Weibull distribution,  $E(T^k) = \frac{\Gamma(1 + \frac{k}{\alpha})}{\lambda^k}$ . I used the fact that a gamma density must integrate to one.
- (d) The median of a continuous distribution is the point with 50% of the probability above and 50% below. Find the median of the Weibull distribution. Show your work.

- (e) The file <http://www.utstat.toronto.edu/brunner/data/legal/Weibull.data1.txt> contains a random sample from a Weibull distribution. Read it with `scan()`.
- Use R's `optim` function to find the MLE. The answer is a pair of numbers from your printout.
  - Calculate a 95% confidence interval for  $\alpha$ . The answer is a pair of numbers from your printout. My lower confidence limit is 1.780622.
  - Give a point estimate of the expected value for the Weibull data. The answer is a number that you calculate with R. It should appear on your printout. The answer should be fairly close to the sample mean. Why?
  - Calculate a 95% confidence interval for the expected value. Hint: to apply the delta method, you will need the derivative of the gamma function. See `help(digamma)`. Because of the Central Limit Theorem, the `t.test` function yields a confidence interval for  $\mu$  that is close to the right answer.
  - Give a point estimate of the median for the Weibull data. The answer is a number that you calculate with R. It should appear on your printout. The answer should be fairly close to the sample median.
  - Give a 95% confidence interval for the median. The answer is a pair of numbers on your printout, the lower confidence limit and the upper confidence limit. My lower confidence limit is 3.182938.
  - This is the last question, and you should do it last. Using the usual 0.05 significance level, test whether the expected value equals the median. What do you conclude? There is more than one right way to do this. I did it several ways. You only need to do it one way. Using the delta method, I got  $Z = 9.92$ . Another approach is to observe that the mean equals the median if and only if  $\alpha$  has a particular value. I found that value numerically, and tested  $H_0 : \alpha = \alpha_0$ , obtaining  $Z = -22.71$  and  $G^2 = 381.89$ . Note that the  $Z$ -tests are 2-sided, so the difference in sign is not a problem.

**Bring your printout for Question 2 to the quiz.** The printout should show your complete R input and output. Do not write anything on your printout in advance except possibly your name and student number.

**Bring a calculator** with natural log and exponential functions.

Again, Question 3e of this assignment *will not be on Quiz 3*