

STA 2101/442 Assignment Five¹

Please bring your R printouts to the quiz. **Do not write anything on your printouts in advance except possibly your name and student number.** Bring a calculator to the quiz.

- For each of the following distributions and associated data sets, obtain the maximum likelihood estimate numerically with R. *Bring your printout for each problem to the quiz;* you may be asked to hand it in. There are links to the data from the course web page in case the ones from this document do not work.

(a) $f(x) = \frac{1}{\pi[1+(x-\theta)^2]}$ for x real, where $-\infty < \theta < \infty$. Data:

-3.77 -3.57 4.10 4.87 -4.18 -4.59 -5.27 -8.33 5.55 -4.35 -0.55
5.57 -34.78 5.05 2.18 4.12 -3.24 3.78 -3.57 4.86

You might want to read the data from [cauchy.data](#). For this one, try at least two different starting values and *plot the minus log likelihood function!*

(b) $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$ for $0 < x < 1$, where $\alpha > 0$ and $\beta > 0$. Data:

0.45 0.42 0.38 0.26 0.43 0.24 0.32 0.50 0.44 0.29 0.45 0.29 0.29 0.32 0.30
0.32 0.30 0.38 0.43 0.35 0.32 0.33 0.29 0.20 0.46 0.31 0.35 0.27 0.29 0.46
0.43 0.37 0.32 0.28 0.20 0.26 0.39 0.35 0.35 0.24 0.36 0.28 0.32 0.23 0.25
0.43 0.30 0.43 0.33 0.37

You might want to read the data from [beta.data](#). If you are getting a lot of warnings, maybe it's because the numerical search is leaving the parameter space. If so, try `help(nlminb)`.

(c) $f(x) = \frac{\theta e^{\theta(x-\mu)}}{(1+e^{\theta(x-\mu)})^2}$ for x real, where $-\infty < \mu < \infty$ and $\theta > 0$. Data:

4.82 3.66 4.39 1.66 3.80 4.69 1.73 4.50 9.29 4.05 4.50 -0.64 1.40
4.18 2.70 5.65 5.47 0.55 4.64 1.19 2.28 7.16 4.80 3.19 2.33 2.57
2.31 0.35 2.81 2.35 2.52 3.44 2.71 -1.43 7.61 0.93 2.52 6.86 6.14
4.37 3.79 5.04 4.50 1.92 3.25 -0.06 2.81 3.09 2.95 3.69

You might want to read the data from [mystery.data](#).

(d) $f(x) = \frac{1}{m!}e^{-x}x^m$ for $x > 0$, where the unknown parameter m is a positive integer. *This means your estimate will be an integer.* Data:

8.34 7.65 6.72 3.84 7.12 1.88 5.07 2.69 4.50 5.78 4.88 5.23 6.17
11.76 7.84 5.87 5.23 6.55 8.34 5.35 4.98 13.81 8.62 7.88 6.34 5.16
6.64 4.35 6.77 5.83 5.85 2.46 8.33 3.74 5.10 3.95 7.84 4.70 6.09
5.23 1.44 6.11 4.88 7.24 7.89 8.98 1.78 5.46 5.34 4.25

¹Copyright information is at the end of the last page.

You might want to read the data from [gamma.data](#).

For each distribution, be able to state (briefly) why differentiating the log likelihood and setting the derivative to zero does not work. For the computer part, bring to the quiz one sheet of printed output for each distribution. The sheets should be separate, because you may hand only one of them in. Each printed page should show the following, *in this order*.

- Definition of the function that computes the likelihood, or log likelihood, or minus log likelihood or whatever.
 - How you got the data into R – probably a `scan` statement.
 - Listing of the data for the problem.
 - The `nlm` or `nlminb` statement and resulting output.
 - For the Cauchy example, a plot of the minus log likelihood.
2. For the data of Problem [1b](#), conduct a large-sample likelihood ratio test of $H_0 : \alpha = \beta$, using R. Your printout should display the value of G^2 , the degrees of freedom and the p -value. Do you reject H_0 at the 0.05 significance level? If yes, which parameter seems to be higher based on $\hat{\alpha}$ and $\hat{\beta}$? *Bring your printout to the quiz.* This can be on the same paper as the printout of Problem [1b](#).
 3. Based on a random sample of size n from a p -dimensional multivariate normal distribution, derive a formula for the large-sample likelihood ratio test statistic G^2 for the null hypothesis that Σ is diagonal (all covariances between variables are zero). You may use the likelihood function on the formula sheet. You may also use without proof the fact that the unrestricted MLE is $\hat{\theta} = (\bar{\mathbf{y}}, \hat{\Sigma})$, where

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^\top.$$

Hint: Because zero covariance implies independence for the multivariate normal, the joint density is a product of marginals under H_0 . To be direct, I am suggesting that you *not* use the likelihood function on the formula sheet to calculate the numerator of the likelihood ratio. You'll eventually get the right answer if you insist on doing it that way, but it's a lot more work.

4. This question uses the SAT data again: [sat.data](#). There is a link on the course web page in case the one in this document does not work.

Using R, carry out a large-sample likelihood ratio test to determine whether there are any non-zero covariances among the three variables; guided by the usual $\alpha = 0.05$ significance level, what do you conclude? Are the three variables all independent of one another? Answer Yes or No. *Bring your printout to the quiz.*

5. 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 λ , the expected number of dead pixels. Naturally, λ 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

- (a) 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 five lines of code: One line to enter the means, one line to enter the sample sizes, one line to compute G^2 , one line to compute the p -value, and one other line. Here are some little questions to get you started.

- i. Is this a between-cases design or a within-cases design?
- ii. Denote the parameter vector by $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)^\top$. What is p ?
- iii. What is the null hypothesis?
- iv. What is the distribution of a sum of independent Poisson random variables?
- v. What is the distribution of $n_j \bar{Y}_j$?
- vi. What is the likelihood function? Write it down and simplify.
- vii. What is the unrestricted MLE $\hat{\boldsymbol{\lambda}}$? It's a vector. Work it out if you need to.
- viii. What is the restricted MLE $\hat{\boldsymbol{\lambda}}_0$? It's a vector. Work it out if you need to.
- ix. Now you are ready to write the test statistic. There are a lot of cancellations. Keep simplifying!
- x. Now use R to compute the test statistic and p -value. For comparison, my p -value is 0.01169133. *Bring your printout to the quiz.*

- (b) Clearly we need to follow up this result to see where it came from, but we'll do Wald tests because they are a little easier. As preparation (and to get some exercise), carry out a Wald test of the same null hypothesis you just tested above. You'll need an estimated asymptotic covariance matrix of $\hat{\lambda}$, and in this case expressing it in closed form is easier than obtaining and inverting the Hessian. So please do it the easy way. The questions below serve as a guide.
- The asymptotic variance of \bar{Y}_j is just its variance. What is $Var(\bar{Y}_j)$?
 - The *estimated* asymptotic variance of \bar{Y}_j is the most natural thing you can imagine. What is it?
 - So what's the estimated asymptotic variance-covariance matrix of the random vector $\hat{\lambda}$?

Now you can carry out the Wald test. Do it with R, obtaining the test statistic, the degrees of freedom and the p -value. *Bring your printout to the quiz.*

- (c) Finally carry out all pairwise comparisons with a Bonferroni correction, protecting the entire family of tests against Type I error at the *joint* $\alpha = 0.05$ significance level. In plain language, what do you conclude? Remember, with pairwise comparisons you can always draw directional conclusions. *Bring your printout to the quiz.*

This assignment was prepared by [Jerry Brunner](#), Department of Statistics, 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^AT_EX source code is available from the course website: <http://www.utstat.toronto.edu/~brunner/oldclass/appliedf14>