# Statistics 3858b Assignment 1

Handout January 21; Due date: February 4, 2019

Assignments are due at the beginning of class.

These problems are all from the course text unless otherwise stated.

Most students will do their programming in R. If a student wishes to program in some other language such as MatLab please see me about this. The R or other code should be well documented. This will help the marker in reading your code.

1. Use the data from 8.10.3. The model for a given group or concentration is Poisson.

   (a) The data $X_i, i = 1, n$ are iid Poisson, $\lambda$. Find the method of moments estimator and the MLE. Are they the same estimator in this case? (ASIDE : Sometimes method of moments and maximum likelihood produce the same estimator, and sometimes not the same.)

   (b) Obtain the approximate sampling distrib.ution for this estimator, using the CLT. Justify your answer.

   (c) For each group (concentration) find the method of moments estimate. Aside : You will treat each group as an independent experiment, so there is a different Poisson parameter for each group.

   (d) In your CLT approximation in (b), replace the standard deviation term in the denominator by its estimate. The standard deviation term is $\sqrt{\lambda}$, so this will be replaced by $\sqrt{\hat{\lambda}_n}$. The normal approximation continues to hold. In order to simplify or clarify notation for your calculation below we use $\hat{\lambda}_{\text{obs}}$ for the observed value of the estimator $\hat{\lambda}_n$.

   Using this approximation calculate the 95% confidence interval, that is solving for $\lambda$ from

   $$-1.96 = z_{.025} \leq \frac{\sqrt{n}(\hat{\lambda}_{\text{obs}} - \lambda)}{\sqrt{\hat{\lambda}_{\text{obs}}}} \leq z_{.975} = 1.96 \ .$$

   The term $z_q$ is the $q$-th quantile of the standard normal distribution. This interval is called the asymptotic or approximate 95% confidence interval.

   Give these intervals for each of the 4 concentration levels.

(e) Discuss if there is evidence that the population value of the parameter for the different Poisson parameters are the same or different for the different concentrations. (Aside : This is a method based on pairwise comparisons and may not be the most efficient way of examining or testing equality of the Poisson parameters.)

*Aside : this notion is discussed more later in the course when we discuss confidence intervals in more details.*

2. 8.10.4 (a), (c), (d)

3. 8.10.7 (a) - (c).

Questions continue on the next page

4. $X_i, i = 1, ..., n$ are iid $N(\mu, \sigma^2)$.

(a) Find the MLE of $\mu, \sigma^2$. Are these unbiased estimators of $\mu$ and of $\sigma^2$ respectively? Aside : You can use your result in (b) to justify your answer for the bias part of the MLE estimator of $\sigma^2$.

(b) In this part you will show, despite that the sample variance is an unbiased estimator of $\sigma^2$, that the sample standard deviation is is a biased estimator of $\sigma$. This fills in some parts of a comment in the lectures.

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \ .$$

State an appropriate result from Chapter 6 of the text which gives the distribution of $(n-1)S^2/\sigma^2$.

Use this to find $\mathrm{E}(S^2)$.

(c) Suppose that $Y$ has a Gamma( $\frac{d}{2}, \frac{1}{2}$) distribution in the parametrization in Rice. See aside (1) below. Find the formula for $\mathrm{E}(\sqrt{Y})$

Use this to obtain the formula $\mathrm{E}(\sqrt{S^2})$. The answer will involve a ratio Gamma functions. See for example problem Rice 2.5.49 for some relevant properties.

Specifically if $n = 5$ give $\mathrm{E}(\sqrt{S^2})$ and simplify the ratio of the Gamma functions. You should use some of the properties of the Gamma function.

Aside (1) : The chi square distribution is a special case of a gamma distribution. A chisquare degrees of freedom $d$ has a Gamma distribution with parameters $\frac{d}{2}, \frac{1}{2}$ is the parametrization in the Rice text Appendix A page A2, or with parameter $\frac{d}{2}, 2$ in the parametrization most of you used in the fall semester. The two parametrizations are of course equivalent.

Aside (2) : You are also showing that the sample variance r.v. is a consistent estimator of the population variance $\sigma^2$. This result holds more generally, although it might be more complicated to find the variance of $S^2$ than in this normal sampling setting.

## Suggested Problems

1. Show that parameters are identifable for the the Gamma family of distributions. *Hint* For this family you need to use properties for the ratio of the pdfs that are equal (or a variation of this idea) and note that limits as the argument tends to 0 or infinity must be equal.

   Do this also for the Poisson family, for the bivariabe normal. For these you may use the ideas suggested in class.

2. 8.10.13 + the additional part : Consider the pdf from Section 8.4 Example D
$$f(x; \alpha) = \frac{1}{2}(1 + \alpha x) I_{[-1,1]}(x) \ .$$
   Find the cdf for this pdf.

   For a given value of $\alpha$ find the cdf $F$. Notice the cdf is a quadratic in $x$ for $x \in [-1, 1]$ for $\alpha \neq 0$ and is linear for $\alpha = 0$. Find the inverse function for this cdf; that is for a given $u \in (0, 1)$ solve $F(x) = u$, where the solution also has to satisfy $x \in [-1, 1]$. You will need to be careful which solution for the quadratic you use (in the cases $\alpha \neq 0$). It is easier to consider $-1 \leq \alpha < 0$, $\alpha = 0$ and $0 < \alpha \leq 1$ as three separate cases for your algebra.

   Suppose that the actual or true value of $\alpha = .3$. Write an R program to study the sampling distribution of $\hat{\alpha}$ in the case $n = 50$. Include comments so that the marker can determine what you intend the code to do.

   *Remark: How can you simulate a random variable with the distribution f in example D? Recall the probability integral transform method of Proposition 2.3D page 63.*

   Compare this with the normal approximation. You may do this by making a qqnorm plot of the Monte Carlo simulation replicates of $\hat{\alpha}$, say $\hat{\alpha}_{n,m}$, for $m = 1, 2, \ldots, M$. Do this with $M = 1000$ Monte Carlo replicates. Discuss in one or two sentences how the true sampling distribution of $\hat{\alpha}$ and the normal approximation to the sampling distribution of $\hat{\alpha}$ compare.

   Finally compare the cases $n = 50, 100, 500$, commenting if the relative frequency histograms of $\hat{\alpha}$ seem to more or less close to the true value .3.

3. 8.10.5 , 8.10.9, 8.10.10 , 8.10.17 (a) - (c) , 8.10.27, 8.10.31, 10.32