Statistics 3858b Assignment 1

Handout January 22; Due date: February 2, 2018

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, λ. Find the method of moments estimator and the MLE. Are they the same estimator in this case? (ASIDE : Usually these two types of estimators are 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 calculate the 95% confidence interval, that is solving for λ 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 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 μ, σ^2. Are these unbiased estimators of μ and of σ^2 respectively? Aside : You can use your result in (b) to justify your answer for the bias part of the MLE estimator of σ^2.

(b) In this part you will show, despite that the sample variance is an unbiased estimator of σ^2, that the sample standard deviation is is a biased estimator of σ.

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

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 $\mathrm{E}(S^2)$.

Use Chebychev's inequality to show that S^2 converges in probability to σ^2 as $n \to \infty$. Aside : This is true more generally but you will only do this for the iid normal case.

Suppose $Y \sim \chi^2_{(k)}$. Show that

$$\mathrm{E}(\sqrt{Y}) < \sqrt{k} = \sqrt{\mathrm{E}(Y)} \ .$$

Show that $\mathrm{E}(\sqrt{S^2}) < \sigma$, and that the inequality is strict.

You will do this by using the following result, which is known as the Cauchy Schwarz inequality. Specifically you will apply this to $W = \sqrt{Y^2}$.

Suppose W is a random variable with finite mean and variance. Then

$$|E(W)| \leq \sqrt{\mathrm{E}(W^2)}$$

and there is equality iff $\mathrm{E}(W^2) = \mathrm{E}(W)^2$.

Aside (1) : The student should consider how the Cauchy Schwarz inequality is related to properties of the variance of W.

Aside (2) : In this problem you are showing in the iid normal setting that the sample standard deviation is a biased estimator of the population standard deviation σ. The result holds much more generally. In fact based on the Cauchy Schwarz inequality this will be the case provided that the sample variance r.v. S^2 does not itself have 0 variance.

Aside (3) : You are also showing that the sample variance r.v. is a consistent estimator of the population variance σ^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. 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 α 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.

2. 8.10.5 , 8.10.9, 8.10.10 , 8.10.17 (a) - (c) , 8.10.27, 8.10.31, 10.32