

# STAT 510: Homework 06

David Dalpiaz

Due: Monday, March 28, 11:59 PM

## Exercise 1 (Method of Moments)

Let  $X_1, X_2, \dots, X_n \sim \text{Gamma}(\alpha, \beta)$ . Find the method of moments estimator of  $\alpha$  and  $\beta$ .

## Exercise 2 (“Numeric” Maximum Likelihood)

Let  $X_1, X_2, \dots, X_n \sim \text{Exponential}(\lambda)$ . That is

$$f(x) = \lambda e^{-\lambda x}.$$

Consider each of the following potential values of  $\lambda$ .

```
lambda = seq(0.001, 1, by = 0.001)
```

We create some data and store it in a vector named `some_data`.

```
set.seed(42)
some_data = rexp(n = 100, rate = 0.2)
```

For each value  $\lambda$ , calculate the log-likelihood given the data above. Plot the results and report the “MLE” based on this procedure.

## Exercise 3 (Estimating Allele Frequency)

In genetics, [single nucleotide polymorphisms](#) (SNPs) are locations in the (human) genome that exhibit variation across the population. SNPs cause the differences we see in traits such as hair color. Each SNP typically has two possible alleles – say  $A$  and  $a$  – and each person’s genotype at the SNP is either  $AA$ ,  $Aa$ , or  $aa$ , where one allele comes from the person’s mother and one from the father. Let  $X$  be the number of  $A$  alleles at a particular SNP, and suppose we collect a random sample of people from some population. Under some assumptions (such as “random mating” and “no selection”) we may assume that

$$X_1, X_2, \dots, X_n \sim \text{Binom}(2, p),$$

where  $p$  is called the *allele frequency* of allele  $A$ . What is the maximum likelihood **estimator** of  $p$ ? What is the maximum likelihood **estimate** of the allele frequency of allele  $A$  if our sample consists of five people with genotypes

$AA, aa, Aa, aa, Aa$

at this particular SNP?

### Exercise 4 (Corn!)

Consider two corn varieties, A and B, both grown in the [Morrow Plots](#). Illinois is very serious about our corn. Rumor has it, if a student is found trespassing in the Morrow Plots, they will be expelled...

Suppose that  $X_1, X_2, \dots, X_n$ , representing yields per acre for corn variety A, constitute a random sample from a normal distribution with mean  $\mu_1$  and variance  $\theta$ . (In more usual notation,  $\theta = \sigma^2$ , but we are using  $\theta$  here to make the notation easier in this problem.) Also,  $Y_1, Y_2, \dots, Y_m$ , representing yields for corn variety B, constitute a random sample from a normal distribution with mean  $\mu_2$  and variance  $\theta$ . If the  $X_i$  and  $Y_j$  are all mutually independent, find the maximum likelihood **estimator** for the common variance  $\theta$ . Assume that  $\mu_1$  and  $\mu_2$  are **known**.

### Exercise 5 (A “Fun and Easy” MLE)

Let  $X_1, X_2$  be independent random variables from Poisson distributions with parameters  $\lambda_1$  and  $\lambda_2$  respectively. That is

$$f(x_i) = \frac{\lambda_i^{x_i} e^{-\lambda_i}}{x_i!}, \quad x_i = 0, 1, 2, \dots$$

- When  $\theta = -1$ , we have  $\lambda_1 = 2.3$  and  $\lambda_2 = 5.6$ .
- When  $\theta = 1$ , we have  $\lambda_1 = 4.4$  and  $\lambda_2 = 3.2$ .

Suppose we observe  $x_1 = 3$  and  $x_2 = 4$ . Based on this data, what is the maximum likelihood estimate of  $\theta$ ? Justify your answer!

### Exercise 6 (Method of Moments with Uniform)

Let  $X_1, X_2, \dots, X_n \sim \text{Uniform}(a, b)$  where  $a < b$ . Find the method of moments estimators for  $a$  and  $b$ .

### Exercise 7 (Maximum Likelihood with Uniform)

Let  $X_1, X_2, \dots, X_n \sim \text{Uniform}(a, b)$  where  $a < b$ . Find the maximum likelihood estimators for  $a$  and  $b$ . Also find the MLE of

$$\tau = \int x dF(x)$$

### Exercise 8 (Maximum Likelihood versus Empirical Distribution)

Let  $X_1, X_2, \dots, X_n \sim \text{Poisson}(\lambda)$  and consider the following data generated according to this model:

```
set.seed(7)
some_data = rpois(n = 20, lambda = 2)
some_data
```

```
## [1] 6 1 0 0 1 3 1 5 1 2 1 1 3 0 2 0 2 0 6 1
```

Consider two probabilities:

- $P(X > 3)$
- $P(X > 7)$

For both:

- Provide an estimate using maximum likelihood.
- Provide an estimate using the empirical distribution.
- Provide the true value.

## Exercise 9 (Parametric Bootstrap)

Let  $X_1, X_2, \dots, X_n \sim \text{Normal}(\mu, \sigma)$ . Given the data below, find the MLE of  $P(X > 5)$ . Use the parametric bootstrap to estimate the standard error of this MLE.

```
set.seed(42)
some_data = rnorm(n = 100, mean = 4, sd = 2)
```

## Exercise 10 (Numeric MLE)

Let  $X_1, X_2, \dots, X_n \sim \text{Exponential}(\lambda)$ . That is

$$f(x) = \lambda e^{-\lambda x}.$$

We create some data and store it in a vector named `some_data`.

```
set.seed(42)
some_data = rexp(n = 100, rate = 0.2)
```

Use Newton–Raphson to find the MLE numerically. (Note that numerical optimization is not actually necessary in this example, so you can easily check your work analytically.) Consider three different initial values for  $\lambda$ :

- $1e-10$
- $0.3$
- $0.5$

Use any reasonable stopping criteria. Comment on the differences based on initial values.

## Exercise 11 (EM for Mixture of Normals)

**This is a challenge question.** You will likely need to do some “Googling” to complete this question.

The following code generates data according to a mixture model. In particular, we have a mixture of three univariate normals.

```
mu = c(0, 5, 10)
sd = sqrt(c(2, 1, 0.5))
mix = c(0.7, 0.1, 0.2)

set.seed(42)
components = sample(1:3, prob = mix, size = 1000, replace = TRUE)
some_data = rnorm(n = 1000, mean = mu[components], sd = sd[components])
```

Use the EM algorithm assuming a three component mixture of normals to estimate the mixing parameters, means, and standard deviations. State how you initialized the parameters, and how you decided to stop iterating. Plot a histogram of the data. Overlay both the true and estimated densities. Do not use any built in functions or packages for fitting mixture models except to check your work.