

Charles Valentine
Homework 5
10/09/2016

Note:

All code is in script hw05.R --- I have made use of cat and print statements to display information easily!

Problem 1

- a) The analytic Maximum Likelihood Estimation (MLE) function for the exponential distribution $\lambda e^{-x\lambda}$ is derived by first finding the likelihood estimator for λ . The likelihood estimator for λ is simply the product of all realized densities $f_X(x_j; \lambda)$ for $j = 1, \dots, n$. Setting the first order derivative of the log of the likelihood estimator to zero will allow us to find a value for λ that is a maxima and therefore most probable.

$$L(\lambda) = f(x_1, \dots, x_n; \lambda) = \prod_{j=1}^n f_X(x_j; \lambda) \quad \text{and} \quad \text{dexp}(x, \lambda) = \lambda e^{-x\lambda}$$

We can then express these two equations together as:

$$L(\lambda) = \lambda^n \exp(-\lambda \sum_{j=1}^n x_j)$$

The log likelihood function is thus:

$$\log(L(\lambda)) = l(\lambda) = n \ln(\lambda) - \lambda \sum_{j=1}^n x_j$$

The first order derivative, set to zero, of the log likelihood is:

$$l' = \frac{d}{d\lambda} l(x_1, \dots, x_n; \lambda) = \frac{d}{d\lambda} \left(n \ln(\lambda) - \lambda \sum_{j=1}^n x_j \right) = 0$$

Taking the derivative is:

$$\frac{n}{\lambda} - \sum_{j=1}^n x_j = 0 \quad \text{equivalent to} \quad \hat{\lambda} = \frac{n}{\sum_{j=1}^n x_j}$$

The Method of Moments (MoM) method uses the sample moments to estimate the population moments. Simply put, the mean of all sample realizations' means is an estimator for the population mean. The population mean of the exp. distribution is $\frac{1}{\lambda}$, with a sample mean \bar{X} .

$$\hat{\lambda}_{\text{MLE}} = \frac{n}{\sum_{j=1}^n x_j} \quad \text{is equivalent to} \quad \bar{\lambda}_{\text{MoM}} = \frac{n}{\bar{X}} = \frac{n}{\sum_{j=1}^n x_j}$$

- b) The maximum likelihood estimator for λ was calculated in R using the `optimize()` maximization solver. The analytic result was calculated by using the formula $\lambda = \frac{n}{\bar{X}}$ where $n = 1$. See code for implementation.

OUTPUT:

```
Numerical Optimization result: lambda=0.9999
Analytical Calculation result: lambda=0.8996
```

Problem 2

- a. The point estimator of m for the chi-square distribution is equal to the sample mean \bar{X} using the Method of Moments. Therefore $m = \bar{X} = 100.8$.
- b. A one-sided 90% lower CI interval of m can be calculated using the t-interval since the population standard deviation σ is unknown. The sample size is 53. See code for implementation using `qt()`. The interval is then calculated as follows:

$$1 - \alpha = 0.90 \quad \text{therefore} \quad \alpha = 0.10$$

$$\text{Lower 90\% CI} = (\bar{X} + t_{\alpha, n-1} \cdot \text{se}(\bar{X}), \infty) \quad \text{where} \quad n = 53 \quad \text{and} \quad \text{se}(\bar{X}) = \frac{s}{\sqrt{n}}$$

$$\text{Lower 90\% CI} = (100.8 + t_{0.1, 52} \cdot 1.7033, \infty)$$

$$\text{Lower 90\% CI} = (98.5891, \infty)$$

Problem 3

When looking at the sampling of the Zyxin gene in the golub dataset when restricted to either 'ALL' or 'AML' patients it is easily discernable that there is a difference in expression. The 'ALL' patients have Zyxin gene expression that is far less than the 'AML' patients. The confidence intervals of the mean, median, and variance qualities all confirm that these two distributed samples are quite different. A figure on the following page confirms our conclusion by showing both densities with minimal overlap. See code for implementation.

OUTPUT

```
CLASSIFICATION ALL
#####
```

```
The sample mean is: -0.2948
The sample median is: -0.2167
The sample variance is: 0.5225
```

```
The bootstrapped two-sided 95% CI of the sample mean is:
( -0.5928 , -0.0346 )
The parametric two-sided 95% CI of the sample mean is:
( -0.5807 , -0.0088 )
```

```
The bootstrapped two-sided 95% CI of the sample median is:
( -0.7351 , 0.3143 )
The parametric two-sided 95% CI of the sample median is:
( -0.4639 , 0.3967 )
```

```
The bootstrapped two-sided 95% CI of the sample variance is:
( 0.3381 , 0.6609 )
The parametric two-sided 95% CI of the sample variance is:
( 0.324 , 0.9813 )
```

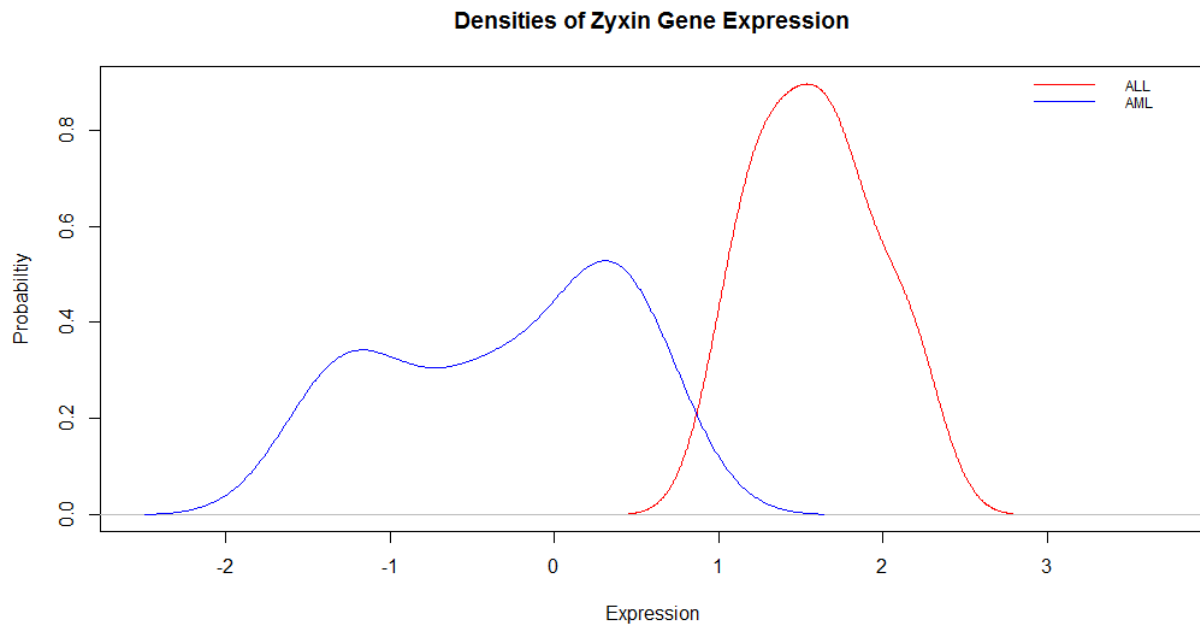
```
CLASSIFICATION AML
#####
```

```
The sample mean is: 1.5867
The sample median is: 1.5763
The sample variance is: 0.1351
```

```
The bootstrapped two-sided 95% CI of the sample mean is:
( 1.3836 , 1.8053 )
The parametric two-sided 95% CI of the sample mean is:
( 1.3397 , 1.8336 )
```

```
The bootstrapped two-sided 95% CI of the sample median is:
( 1.2281 , 1.8283 )
The parametric two-sided 95% CI of the sample median is:
( 1.476 , 4.948 )
```

```
The bootstrapped two-sided 95% CI of the sample variance is:
( 0.0508 , 0.2011 )
The parametric two-sided 95% CI of the sample variance is:
( 0.066 , 0.4162 )
```



Problem 4

- c. The Confidence Interval formula for λ that should be used to is the sample mean formula as it more accurately represents the intended CI of 90%. We have verified this using a Monte Carlo simulation of 1000 iterations based on a Poisson distribution with a sample size of 50. For rates $\lambda = [0.1, 1, 10]$ the CI sample mean formula performs similarly to the nominal 90% level at 90.9%, 92.2%, and 92.9%. The CI sample variance formula performs poorly at a low rate and converges to 90% as the rate increases with coverage probability 44.7%, 81.8%, and 87%. See code for implementation.