CSE 312: Foundations of Computing II
**Instructor**: Alex Tsun
**Date**: 8/03/20
**Lecture Topics**: 7.1 Maximum Likelihood Estimation, 7.2 MLE Examples

[**Tags:** MLE]

1. Suppose $x = (x_1, \dots, x_n)$ are iid samples from $\mathcal{N}(\theta_1, \theta_2)$ where $\theta_1$ is the mean and $\theta_2$ is the variance (both unknown). Let $\theta = (\theta_1, \theta_2)$ denote the parameter vector.
    a. What are the likelihood and log-likelihood of the data?
    b. What are the maximum likelihood estimates for $\theta_1, \theta_2$?

**Solution**:

a. The likelihood is

$$L(x|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta_2}} \exp\left(\frac{-(x_i - \theta_1)^2}{2\theta_2}\right) = \prod_{i=1}^{n} (2\pi\theta_2)^{-1/2} \exp\left(\frac{-(x_i - \theta_1)^2}{2\theta_2}\right)$$

Recall log properties: $\log(ab) = \log(a) + \log(b)$ and $\log(a^b) = b \log a$. The log-likelihood is

$$\ln L(x|\theta) = \ln\left(\prod_{i=1}^{n} (2\pi\theta_2)^{-\frac{1}{2}} \exp\left(\frac{-(x_i - \theta_1)^2}{2\theta_2}\right)\right) = \sum_{i=1}^{n}\left(-\frac{1}{2}\ln(2\pi\theta_2) - \frac{(x_i - \theta_1)^2}{2\theta_2}\right)$$

b. We'll take the partial derivatives with respect to (wrt) $\theta_1$ and $\theta_2$ (don't forget the chain rule)

$$\frac{\partial}{\partial\theta_1} \ln L(x|\theta) = \sum_{i=1}^{n}\left(\frac{2(x_i - \theta_1)}{2\theta_2}\right) = \frac{1}{\theta_2}\left(\sum_{i=1}^{n} x_i - n\theta_1\right)$$

Setting this to 0, we get

$$\frac{1}{\hat{\theta}_2}\left(\sum_{i=1}^{n} x_i - n\hat{\theta}_1\right) = 0 \rightarrow \sum_{i=1}^{n} x_i - n\hat{\theta}_1 = 0 \rightarrow \hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Now, with respect to $\theta_2$,

$$\frac{\partial}{\partial\theta_2} \ln L(x|\theta) = \sum_{i=1}^{n}\left(-\frac{1}{2}\frac{1}{2\pi\theta_2}2\pi + \frac{(x_i - \theta_1)^2}{2\theta_2^2}\right) = \sum_{i=1}^{n}\left(-\frac{1}{2\theta_2} + \frac{(x_i - \theta_1)^2}{2\theta_2^2}\right)$$

$$= -\frac{n}{2\theta_2} + \sum_{i=1}^{n}\left(\frac{(x_i - \theta_1)^2}{2\theta_2^2}\right)$$

Setting this to 0, we get

$$-\frac{n}{2\hat{\theta}_2} + \sum_{i=1}^{n}\left(\frac{\left(x_i - \hat{\theta}_1\right)^2}{2\hat{\theta}_2^2}\right) = 0 \rightarrow n\hat{\theta}_2 = \sum_{i=1}^{n}\left(x_i - \hat{\theta}_1\right)^2$$

$$\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \hat{\theta}_1\right)^2$$

We also need to check that this is in fact a maximum, since the first derivative will only give us a critical point. This is a bit out of scope for this class and it's prerequisites because we are using a multivariate function, so don't feel like you need to understand it, but we are including it in case you are curious and as a reminder that we need to always check that a point is in fact a maximum!

We will use $l(\theta_1, \theta_2)$ as shorthand for our log-likelihood function. We need to take the second derivatives which gives us the following:

For the second derivative in respect to $\theta_1$twice:

$$\frac{\partial^2}{\partial \theta_1^2} l(\theta_1, \theta_2) = -\frac{n}{\theta_2}$$

$$\frac{\partial^2}{\partial \theta_1^2} l(\hat{\theta}_1, \hat{\theta}_2) = -\frac{n^2}{\sum_{i=1}^{n} \left(x_i - \hat{\theta}_1\right)^2} < 0$$

For the second derivative in respect to $\theta_2$ twice:

$$\frac{\partial^2}{\partial \theta_2^2} l(\theta_1, \theta_2) = \frac{n}{2\theta_2^2} - \sum_{i=1}^{n} \left(\frac{(x_i - \theta_1)^2}{\theta_2^3}\right)$$

$$\frac{\partial^2}{\partial \theta_2^2} l(\hat{\theta}_1, \hat{\theta}_2) = -\left(\frac{n^3}{\left(\sum_{i=1}^{n} \left(x_i - \hat{\theta}_1\right)^2\right)^2}\right) < 0$$

And for the second derivative in respect to the first and second:

$$\frac{\partial^2}{\partial \theta_1 \partial \theta_2} l(\theta_1, \theta_2) = -\frac{1}{\theta_2^2}\left(\sum_{i=1}^{n} x_i - n\theta_1\right)$$

$$\frac{\partial^2}{\partial \theta_1 \partial \theta_2} l(\hat{\theta}_1, \hat{\theta}_2) = 0$$

So,

$$|H| = (-)(-) - 0^2 = + > 0$$

and since the second derivative in respect to $\theta_1$ twice and $\theta_2$ twice are negative, we have a local maximum.