MLE for the normal distribution. Suppose $\vec{X} = (X_1, \ldots, X_n)$ is a sample from a normal distribution with (unknown) parameters μ and σ^2. The maximum likelihood estimates for μ and σ^2 are worked out on page 321, by finding the critical points of the two-variable likelihood function. The results are:

$$MLE_\mu(x_1, \ldots, x_n) = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

$$MLE_{\sigma^2}(x_1, \ldots, x_n) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

It is noteworthy that MLE_{σ^2} is *not equal to* the sample variance:

$$S^2(x_1, \ldots, x_n) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

We showed previously that $E(S^2) = \sigma^2$. Thus, $E(MLE_{\sigma^2}) = \frac{n-1}{n}\sigma^2$. This shows that MLE_{σ^2} is *biased*—it systematically underestimates σ^2. Thus, we see a purely mathematical reason for caution in the use of MLEs. They should not be used without evaluating their behavior. We will study this problem in Section 7.3.

Homework. Chapter 7 (exercised on pages 355–367):
- hand in Friday, October 1: 1, 2(a), 6, 7, 9, 10;
- hand in Monday, October 4: 14, 19, 20, 21,22.

Hints.

1. MLE_θ is a function of x.

2(a). If X is gamma(α, β), then

$$f_X(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta},$$

so, if $\vec{x} = (x_1, \ldots, x_n)$, then

$$L(\beta|\vec{x}, \alpha) = \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x_i/\beta}.$$

Maximize this function of β.

6. Let $\vec{X} = (X_1, \ldots, X_n)$. Then

$$f_{\vec{X}}(\vec{x}|\theta) = \prod_{i=1}^{n} \theta x_i^2 = \theta^n t^2, \quad \text{with } t = \prod_{i=1}^{n} x_i.$$

7. Let $t = \prod_{i=1}^n x_i$, let

$$g(t,\theta) := \begin{cases} 1, & \text{if } \theta = 0; \\ 2^{-n}\sqrt{1/t} & \text{if } \theta = 1. \end{cases}$$

Also, let

$$h(\vec{x}) = \begin{cases} 1 & \text{if } 0 < x_i < 1 \text{ for } i = 1, 2, \ldots n; \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$L(\theta|\vec{x}) = g(t,\theta) \cdot h(\vec{x}).$$

9. See page 312 for a description of the method of moments. We describe it briefly here. Suppose we have a sample $X_1, \ldots, X_n$ that is iid $f_X(x|\theta_1, \ldots, \theta_k)$, where the distribution has k (unknown) scalar parameters. The moments of a random variable X with this distribution are:

$$m_j(\vec{\theta}) = E(X^j|\vec{\theta}),\ j = 1, 2, \ldots.$$

The method of moments tells us estimate $\theta_1, \ldots, \theta_k$ by solving the system of k equations:

$$\frac{1}{n}\sum_{i=1}^n X_i^j = m_j(\vec{\theta}),\ j = 1, 2, \ldots, k. \qquad (*)$$

If X is uniform on $(0,\theta)$, then the moments of X are:

$$m_i = E(X^i) = \frac{1}{\theta}\int_0^\theta x^i dx = \frac{\theta^i}{i+1}.$$

Now, in this example there is only one parameter to estimate. The system $(*)$ reduces to:

$$\overline{X} = m_1 = \theta/2,$$

so $\theta = 2\overline{X}$ is the method-of-moments estimate.

To find the MLE, we need to write the likelihood function. We have calculated this previously, in the examples we worked for sufficient statistics:

$$L(\theta|\vec{x}) = f_{\vec{X}}(\vec{x}|\theta) = \begin{cases} \theta^{-n} & \text{if } \max\{X_i \mid i = 1, 2, \ldots, n\} \le \theta; \\ 0 & \text{otherwise.} \end{cases}$$

Fixing the x_i, the maximum value of $L(\theta|\vec{x})$ clearly occurs when $\theta = \max_i\{x_i\}$, since θ^{-n} is a decreasing function of θ. (How would this problem change of the original distribution had been uniform on $(0,\theta)$ rather than $[0,\theta]$?)

To complete this problem, you must calculate the expected value and the variance of $2\overline{X}$ and of $\max\{X_i \mid i = 1, 2, \ldots, n\}$.

10. After the above, this should be within your capability.