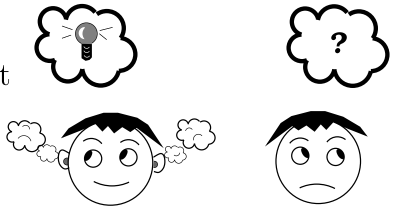
STATS 210 Extra Examples 3

Maximum likelihood in action

We know that maximum likelihood estimation gives sensible answers for simple problems like estimating the Binomial parameter p. What happens in a more complicated case?

In this tutorial we see an example of maximum likelihood estimation where the final answer is not obvious. We want to estimate a *mixture* probability when we have a mixture of two distributions. This type of model is called a *mixture model*.

Guessers and Swots



Professor Partridge has a class of N students. Every student in the class is either a **Swot** or a **Guesser**. Swots study hard, while Guessers only guess. The proportion of Swots in the class, s, is unknown. Partridge wants to use some test results to estimate the proportion of Swots, s.

- There are N students in the class.
- Each student is a Swot with probability s, or a Guesser with probability $1 - s$.
- Define $\Omega = \{\text{All students}\}$, and events $S = \{\text{Swot}\}$; $G = S^c = \{\text{Guesser}\}$.
- **Mixture of two distributions.** On a *test of 10 questions,* let X be the *number of correct answers* a student gets. The distribution of X depends on whether the student is a Guesser or a Swot. Look carefully at the following notation. We write:

  $$X \,|\, G \sim \text{Binomial}(10, 0.5)\,; \qquad X \,|\, S \sim \text{Binomial}(10, 0.75)$$

- Using the Partition Theorem, you will show in Q1(a) below that:

  $$\mathbb{P}(X = x) = (1 - s)\binom{10}{x}(0.5)^x(0.5)^{10-x} + s\binom{10}{x}(0.75)^x(0.25)^{10-x}.$$

  The overall distribution of X is a ***mixture*** of a Bin$(10, 0.5)$ distribution for Guessers, and a Bin$(10, 0.75)$ distribution for Swots, where s is the ***mixture probability***.

- Professor Partridge wants to estimate the proportion of Swots in the class, s.

1.(a) We said above that $\mathbb{P}(X = x) = (1 - s)\binom{10}{x}(0.5)^x(0.5)^{10-x} + s\binom{10}{x}(0.75)^x(0.25)^{10-x}$. Show this using proper probability notation. Your answer should be two lines long. The first line should be of the form $\mathbb{P}(X = x) = \mathbb{P}(X = x \,|\, \ldots)\mathbb{P}(\ldots) + \mathbb{P}(X = x \,|\, \ldots)\mathbb{P}(\ldots)$, and the second line should fill in all quantities to produce the desired result.

(b) Use the formula to show that $\mathbb{P}(X = 4) = 0.21 - 0.19s$. Round all decimals to 2 d.p.

(c) The same procedure as in (b) using $X = 8$ gives: $\mathbb{P}(X = 8) = 0.04 + 0.24s$.

If two students take the test ($N = 2$), and they score marks 4 and 8 respectively, the likelihood function is $L(s\,; 4, 8) = \mathbb{P}(X = 4)\mathbb{P}(X = 8)$. Use the information in (b) and (c) to write down the likelihood explicitly in terms of s. Remember to give the range of values of s over which the likelihood is defined.

(d) Using (c), solve $\frac{dL}{ds} = 0$ to find the maximum likelihood estimate of s when there are two students who score 4 and 8 marks respectively. For convenience, use the rounded values 0.21, 0.19, 0.04, and 0.24 from (b) and (c), not their unrounded values.

Instead of maximizing the likelihood function, statisticians usually maximize the **log-likelihood function**. In this example, the log-likelihood is:

$$\log\Big(\ L(s\,;x_1,x_2)\ \Big) = \log\Big(\ (0.21 - 0.19s)(0.04 + 0.24s)\Big)\,,$$

where log refers to the **natural logarithm,** $\log_e$. Maximizing the log-likelihood should always give the same answer as maximizing the likelihood, because $\log(L)$ is a strictly increasing function of L. Using the log-likelihood has the convenient effect of transforming awkward products, like $(0.21 - 0.19s)(0.04 + 0.24s)$, into sums:

$$\log\Big(\ L(s\,;x_1,x_2)\ \Big) = \log(0.21 - 0.19s) + \log(0.04 + 0.24s).$$

(e) Differentiate $\log\big(\ L(s\,;x_1,x_2)\ \big)$ from above, and solve the equation $\frac{d\log(L)}{ds} = 0$ for s. The value of s that you get maximizes the **log-likelihood**. Is this the same value of s that maximized the **likelihood** from part (d)?
[Hint: remember that if $y = \log(x)$, then $\frac{dy}{dx} = \frac{1}{x}$. In this question you need to consider something like $\frac{dy}{dx}$ when $y = \log(a + bx)$.]

Does it work?

We can't solve this problem by common sense like the previous Binomial problems. However, we can use simulations to discover whether the MLE method is giving us sensible answers for s.

In this question we work through some computer simulations where we *know* the true value of s. We will see whether or not the MLE method gives us an estimate of s close to the true value.

- Set the true value of s to be $s = 0.6$. This means that the true proportion of Swots in the class is 0.6.
- Set $N = 100$: 100 students in the class.

2.(a) How many Swots are there in the class if $N = 100$ and $s = 0.6$? How many Guessers are there?
[Hint: these answers are not random, because we are told that s is exactly 0.6.]

(b) For our computer simulation, we need to generate a random score out of 10 for each of 60 Swots and 40 Guessers. Give two *R* commands, one for the Swots and one for the Guessers, that will generate the required 100 scores. For the Swots, we need 60 random numbers from the Binomial(10, 0.75) distribution, while for the Guessers we need 40 random numbers from the Binomial(10, 0.5) distribution. [Hint: see Section 5.2 of lecture notes.]

The simulated scores for a class of 100 students are shown below. Note that the Swots are higher than the Guessers, on the whole, but there is also a lot of overlap.

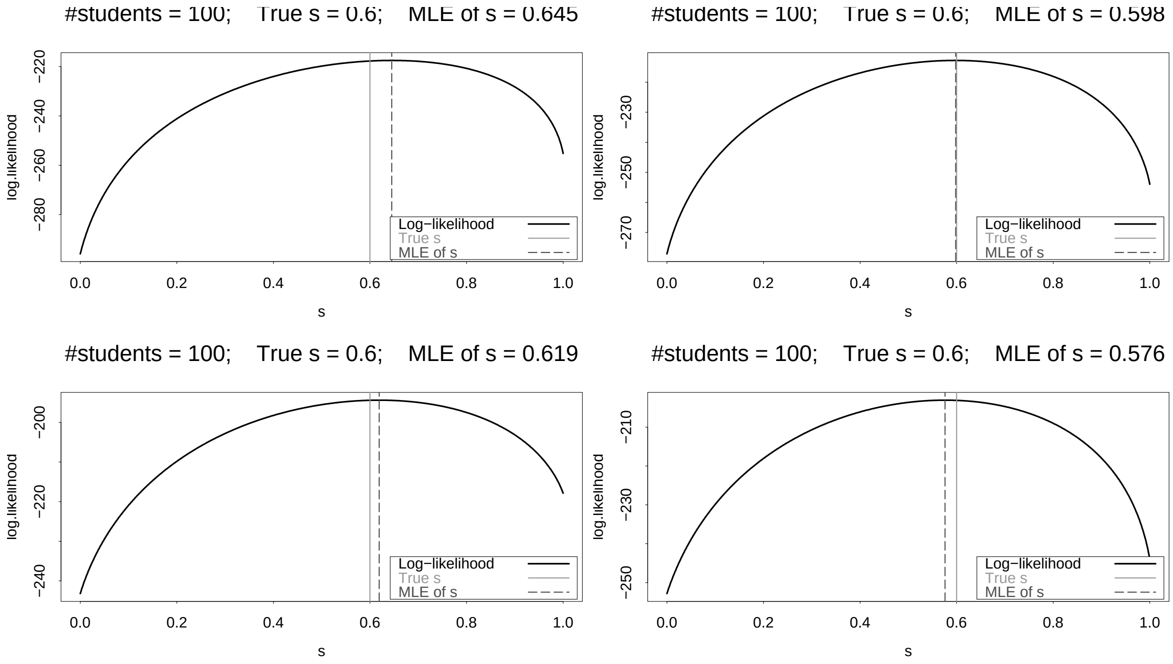
```
Swots:
  4  6  6  7  8  7  9  9  8  8  5  8  7  8  6  5  7  6  9  9  8  6 10  7  9
  9 10  6  9  7  4  7  6  6  7  8 10  9  9  8  8  8  9  9  8  7  7  7  9  4
  6  8  9  8  6  7 10 10  7  6
Guessers:
 5  5  5  3  5  7  7  7  5  5  1  4  6  5  6  4  4  9  8  6  7  2  6  5  2
 4  6  6  5  7  4  7  6  6  4  7  4  7  6  4
```

Let X_i be the score for student i. For each of the 100 observations x_i, we can find $\mathbb{P}(X_i = x_i)$ as we did in question 1(b) when $x_i = 4$. We can then calculate the log-likelihood as a function of s as we did in question 1(e), but with 100 terms instead of just 2 terms. We can plot the log-likelihood and find its maximum using a computer.

Although we know the true value of s is 0.6, we will get a different MLE of s every time we do this. If the method is sensible, the MLE of s will be close to the true value of 0.6.

Here are four log-likelihood curves using $N = 100$ students and the true value $s = 0.6$.

Log-likelihood curves from 4 classes of 100 students when $s = 0.6$



#students = 100; True s = 0.6; MLE of s = 0.645

#students = 100; True s = 0.6; MLE of s = 0.598

#students = 100; True s = 0.6; MLE of s = 0.619

#students = 100; True s = 0.6; MLE of s = 0.576

The curves show $\log\big(L(s\,; x_1, \ldots, x_{100})\big)$ for each of the four simulations. The solid vertical line shows the position of the true value of s, $s = 0.6$, which would be unknown in real life. The dashed vertical line shows the position of the maximum log-likelihood, which in real life would be our estimate of s.

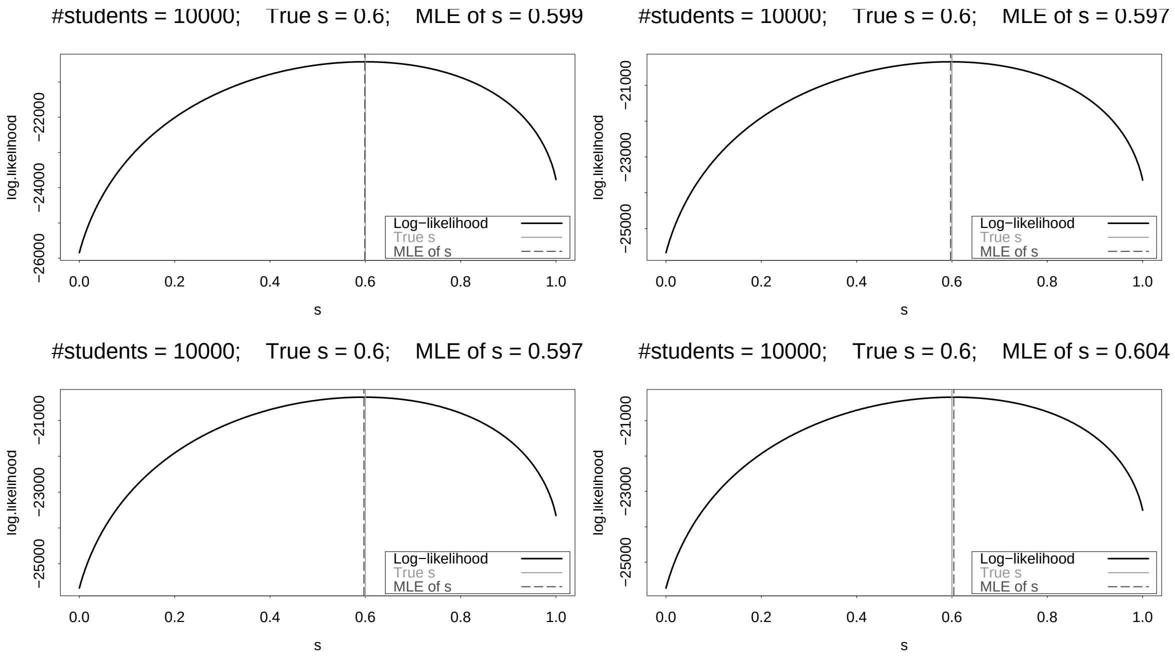
3.(a) By referring to the graphs, write down the four maximum likelihood estimates of s from the four simulations above. Do you think that the results from the maximum likelihood method are sensible? Why?

(b) Which of the dashed line and the solid line should correspond to the peak of the curve? State in words what this line represents.

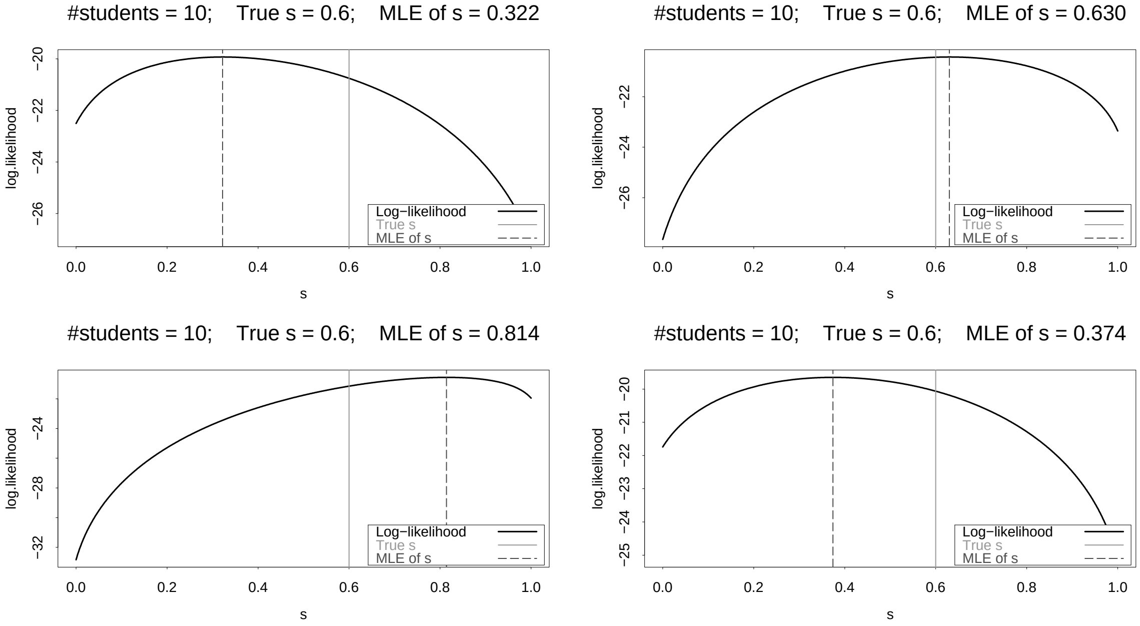
Overleaf are two more trials. The first shows the MLEs of s from four classes of $N = 10,000$ students. The second shows the MLEs of s from four classes of $N = 10$ students.

(c) Write down the four maximum likelihood estimates of s when $N = 10,000$, and write down the four maximum likelihood estimates of s when $N = 10$. What do you notice about the MLEs for large and small sample sizes?

Log-likelihood curves from 4 classes of 10,000 students when $s = 0.6$



Log-likelihood curves from 4 classes of 10 students when $s = 0.6$


#students = 10;   True s = 0.6;   MLE of s = 0.322
#students = 10;   True s = 0.6;   MLE of s = 0.630
#students = 10;   True s = 0.6;   MLE of s = 0.814
#students = 10;   True s = 0.6;   MLE of s = 0.374
log.likelihood
s
Log−likelihood
True s
MLE of s