Practice questions

1. A random variable X is Normal$(1, 1)$ with probability p and Normal$(-1, 1)$ with probability $1 - p$, where the parameter p is unknown.

   (a) What is the maximum likelihood estimate of p given that $X = x$?

   (b) (**Optional**) What is the maximum likelihood estimate of p given independent samples $X_1 = x_1$ and $X_2 = x_2$?

2. A batch of light bulbs is either all normal or all defective, each with expected lifetime of 5 years and 2 years respectively. Lifetimes of two light bulbs from the same batch are tested to determine whether the batch is defective. Propose a test with false negative probability 5%.

3. A food processing company packages honey in glass jars. The volume of honey in a random jar is a Normal$(\mu, 5)$ millilitre random variable for an unknown value of μ. The government wants to verify that μ is at least 100 millilitres.

   (a) The government proposes the following test: Choose a random jar and verify that the jar has at least t millilitres of honey. Which value of t should be chosen so that a complying company passes the test with probability at least 95%?

   (b) The ACME company jars contain Normal$(95, 5)$ millilitres of honey. What is the probability that ACME passes the test?

4. You want to estimate the parameter θ of a Uniform$(0, \theta)$ random variable.

   (a) What is the maximum likelihood estimate $\hat{\Theta}_n$ given independent samples $X_1, \ldots, X_n$?

   (b) Calculate $\mathrm{E}[\hat{\Theta}_1]$. Is $\hat{\Theta}_1$ unbiased?

   (c) **(Optional)** Calculate $\mathrm{E}[\hat{\Theta}_n]$.

   (d) Is $\hat{\Theta}_n$ consistent? (**Hint:** Calculate the probability that $|\hat{\Theta}_n - \theta| \leq \varepsilon$.)

5. On April 23 the Guardian published this text about a Stanford study which estimated that 4.16% of Santa Clara county's population is infected with Covid-19:

   > The biggest criticism was that it estimated cases for the whole county's population based on detecting only 50 positives out of 3,300 people sampled. And since the tests had a false positive rate in one assessment of 2 out of 371, critics argued all the Covid-19 cases detected by the tests in Santa Clara could conceivably have been false positives.

   Is the critics' argument valid? You can model the number of positives as Binomial$(3300, p)$ random variable with unknown p. Assume the false negative rate is zero.