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 $X_2 = x_2$?

Solution:

(a) Let Θ be the indicator that X is Normal$(1,1)$. Then by total probability theorem,

$$\begin{aligned} f_X(x) &= f_{X|\Theta}(x|1)f_\Theta(1) + f_{X|\Theta}(x|0)f_\Theta(0) \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-1)^2}{2}} \cdot p + \frac{1}{\sqrt{2\pi}}e^{-\frac{(x+1)^2}{2}} \cdot 1 - p \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2+1}{2}}\left(p(e^x - e^{-x}) + e^{-x}\right) \end{aligned}$$

The maximum likelihood estimator finds $\hat{p}$ between 0 and 1 for which $f_X(x)$ is maximized. Since $f_X(x)$ is a linear function in p with positive slope iff $e^x - e^{-x} > 0$, i.e. when $x > 0$, the ML estimate is

$$\hat{p} = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

(b) The random variable (X_1, X_2) has PDF $f_{X_1,X_2}(x_1,x_2)$. By independence it is equal to $f_{X_1}(x_1)f_{X_2}(x_2)$. We wish to maximize this quantity w.r.t. p so it is sensible to maximize the log of it instead (ln is an increasing function so the maximizing p remains unchanged). The log of the PDF equals

$$\begin{aligned} \ln f_{X_1,X_2}(x_1,x_2) &= \ln f_{X_1}(x_1) + \ln f_{X_2}(x_2) \\ &= \ln\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{x_1^2+1}{2}}(p(e^{x_1} - e^{-x_1}) + e^{-x_1})\right) \\ &\quad + \ln\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{x_2^2+1}{2}}(p(e^{x_2} - e^{-x_2}) + e^{-x_2})\right) \\ &= \ln\left(p(e^{x_1} - e^{-x_1}) + e^{-x_1}\right) + \ln\left(p(e^{x_2} - e^{-x_2}) + e^{-x_2}\right) + C, \end{aligned}$$

where C is a constant in p. The function is differentiable for all values of x_1, x_2 because $p \le 1$. So we could set its derivative to 0 to find its maximizer:

$$\begin{aligned} &\frac{d}{dp}(\ln(p(e^{x_1} - e^{-x_1}) + e^{-x_1}) + \ln(p(e^{x_2} - e^{-x_2}) + e^{-x_2})) \\ &= \frac{e^{x_1} - e^{-x_1}}{p(e^{x_1} - e^{-x_1}) + e^{-x_1}} + \frac{e^{x_2} - e^{-x_2}}{p(e^{x_2} - e^{-x_2}) + e^{-x_2}} \\ &= \frac{e^{2x_1} - 1}{p(e^{2x_1} - 1) + 1} + \frac{e^{2x_2} - 1}{p(e^{2x_2} - 1) + 1} \\ &= \left(p(e^{2x_2} - 1) + 1\right)(e^{2x_1} - 1) + \left(p(e^{2x_1} - 1) + 1\right)(e^{2x_2} - 1) \end{aligned}$$

from where, assuming $x_1 \neq 0$ and $x_2 \neq 0$, zero derivative is attained at

$$p^* = \frac{2 - e^{2x_1} - e^{2x_2}}{2(e^{2x_1} - 1)(e^{2x_2} - 1)}. \tag{1}$$

If $x_1 = 0$ then $\ln f_{X_1,X_2}$ is a linear function of p with slope $e^{x_2} - e^{-x_2}$. In this case $\hat{p} = 1$ if $x_2 > 0$ and $\hat{p} = 0$ if $x_2 < 0$. The same is true for x_2. If $x_1 = x_2 = 0$ then f_{X_1,X_2} is constant and any p maximizes it. The derivative of fraction of the form $\frac{a}{ax+b}$ is $-(\frac{a}{ax+b})^2$, so the second derivative of $\ln f_{X_1,X_2}$ must be negative at the above value of p, thus it gives the maxima. It remains to find values of x_1, x_2 for which p^* is greater than 1 (resp. 0), in that case $\hat{p}$ is 1 (resp. 0). Solving the inequalities, we finally have

$$\hat{p} = \begin{cases} 0, & \text{if } x_2 \leq \frac{1}{2}\ln(2 - e^{2x_1})) \\ 1, & \text{if } x_2 \geq x_1 - \frac{1}{2}\ln(2e^{2x_1} - 1)) \\ p^*, & \text{otherwise.} \end{cases}$$

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%.

   Solution: Let X_1, X_2 be the lifetimes of the two light bulbs. Let Θ indicate whether the batch is defective. The Neyman-Pearson lemma tells us to choose a hypothesis such that the likelihood ratio is below a threshold. The likelihood ratio is

$$\frac{f_{X_1,X_2|\Theta}(x_1,x_2|1)}{f_{X_1,X_2|\Theta}(x_1,x_2|0)} = \frac{f_{X_1|\Theta}(x_1|1)\cdot f_{X_2|\Theta}(x_2|1)}{f_{X_1|\Theta}(x_1|0)\cdot f_{X_2|\Theta}(x_2|0)} = \frac{\frac{1}{2}e^{-x_1/2}\cdot\frac{1}{2}e^{-x_2/2}}{\frac{1}{5}e^{-x_1/5}\cdot\frac{1}{5}e^{-x_2/5}} = 6.25e^{-3(x_1+x_2)/10},$$

   which is an decreasing function in $x_1 + x_2$. So we should estimate $\hat{\Theta} = 1$ (the test is positive) when $X_1 + X_2 \leq t$ for some t. Since the false negative probability requirement is 5%, we set

$$0.05 = P(\hat{\Theta} = 0|\Theta = 1) = \mathrm{P}(X_1 + X_2 > t|\Theta = 1) = \mathrm{P}(Y > t|\Theta = 1), \tag{2}$$

   where $Y = X_1 + X_2$. By convolution, the PDF of $Y = X_1 + X_2$ conditioned on $\Theta = 1$ is

$$f_{Y|\Theta}(y|1) = \int_{-\infty}^{\infty} f_{X_1|\Theta}(x_1|1) f_{X_2|\Theta}(y - x_1|1)dx_1 = \int_0^y \frac{1}{2}e^{-x_1/2}\cdot\frac{1}{2}e^{(x_1-y)/2}dx_1 = \frac{1}{4}ye^{-y/2}$$

   Therefore, its CDF is

$$\mathrm{P}(Y \leq y|\Theta = 1) = \int_0^y \frac{1}{4}ze^{-z/2}dz = 1 - \frac{1}{2}e^{-y/2}(y+2)$$

   Therefore $\mathrm{P}(Y > t|\Theta = 1) = 1 - \frac{1}{2}(t+2)e^{-t/2}$, and plugging into (2) this happens when $(t+2)e^{-t/2} = 0.1$. This equation does not have a closed form solution for t, but a numerical calculator reveals that $t \approx 9.48773$. Therefore, the test guesses the batch is defective if the sum of the lifetimes of the tested light bulbs is less than 9.48773.

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?

Solution:

(a) The probability that a company has honey with mean volume more than 100 passes the test is greater than those having honey with mean volume exactly 100, so a company with $\mu = 100$ passes the test with probability exactly 95%. A Normal$(0, 1)$ random variable takes value greater than -1.645 with probability 95%. So $t \approx 100 - 1.645 \cdot 5 = 91.775$

(b) Let Φ be the Normal(0,1) CDF, then $\Phi((91.775 - 95)/5) = 0.2595$ is the probability that the sampled jar has less than 91.775 millilitres of honey, i.e. the ACME company fails the test. Therefore, the probability that the company passes the test is approximately $1 - 0.2595 = 0.7405$.

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 $|\hat{\Theta}_n - \theta| \le \varepsilon$.)

Solution:

(a) The conditional PDF is $f_{X_1,\ldots,X_n|\Theta}(x_1, \ldots, x_n|\theta) = \theta^{-n}$ if $0 \le x_1, \ldots, x_n \le \theta$. Maximum likelihood is achieved when θ^{-n} is as small as possible, so θ itself should be as small as possible. This happens when θ is the largest one among the $x_1, \ldots, x_n$. Therefore the maximum likelihood estimate is $\hat{\Theta}_n = \max\{X_1, \ldots, X_n\}$.

(b) $\hat{\Theta}_1$ is simply X_1. So, $\mathrm{E}[\hat{\Theta}_1] = \mathrm{E}[X_1] = \theta/2 \neq \theta$. Therefore $\hat{\Theta}_1$ is not unbiased.

(c) The CDF of $\hat{\Theta}_n$ at x is the probability that the maximum of $X_1, \ldots, X_n$ takes value at most x. This happens exactly when all of $X_1, \ldots, X_n$ are at most x. By independence,

$$F_{\hat{\Theta}_n}(x) = \mathrm{P}(X_1, \ldots, X_n \le x) = \mathrm{P}(X_1 \le x) \cdots \mathrm{P}(X_n \le x) = (x/\theta)^n \quad \text{if } 0 < x < \theta$$

We can calculate its PDF by differentiation:

$$f_{\hat{\Theta}_n}(x) = \frac{d}{dx}\left(\frac{x}{\theta}\right)^n = \frac{x^{n-1}}{\theta^n}n \quad \text{if } 0 < x < \theta$$

The expectation is by definition,

$$\mathrm{E}[\hat{\Theta}_n] = \int_{-\infty}^{\infty} x f_{\hat{\Theta}_n}(x)dx = \int_0^\theta \frac{x^n}{\theta^n} n dx = \frac{n}{n+1}\theta \neq \theta$$

We can see that $\hat{\Theta}_n$ is also not unbiased. As n increases, however, the bias becomes smaller.

(d) Since $\hat{\Theta}_n$ takes one of the values $X_1, \ldots, X_n$, $\hat{\Theta}_n$ can never be larger than θ. So the event $|\hat{\Theta}_n - \theta| > \varepsilon$ occurs when $\hat{\Theta}_n$ is less than $\theta - \varepsilon$, that is when all of the samples $X_1, \ldots, X_n$ are less than $\theta - \varepsilon$ (assuming $\varepsilon < \theta$). These samples are independent so

$$\mathrm{P}(|\hat{\Theta}_n - \theta| > \varepsilon) = \mathrm{P}(X_1, \ldots, X_n < \theta - \varepsilon) = \mathrm{P}(X_1 < \theta - \varepsilon) \cdots \mathrm{P}(X_n < \theta - \varepsilon) = \left(\frac{\theta - \varepsilon}{\theta}\right)^n.$$

For any $\varepsilon > 0$, $(\theta - \varepsilon)/\theta$ is strictly between 0 and 1, and so $((\theta - \varepsilon)/\theta)^n$ becomes arbitrarily small as the number of samples n increases. So $\hat{\Theta}_n$ is consistent.

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 $\text{Binomial}(3300, p)$ random variable with unknown p. Assume the false negative rate is zero.

Solution: Let q be the true fraction of the population infected with Covid-19 and $f = 2/371$ be the false positive rate. Assuming a false negative rate of zero, by the total probability theorem $p = 1 \cdot q + f \cdot (1 - q)$.

The Stanford study estimated that q equals $q_S = 4.16\%$ from where p equals $p_S \approx 4.68\%$. The critics' hypothesis is that q equals zero from where p equals $p_C = f \approx 0.54\%$. The likelihood ratio for 50 positives is

$$\frac{\mathrm{P}(\text{Binomial}(3300, p_S) = 50)}{\mathrm{P}(\text{Binomial}(3300, p_C) = 50)} = \frac{\binom{3300}{50} p_S^{50} (1 - p_S)^{3250}}{\binom{3300}{50} p_C^{50} (1 - p_C)^{3250}} = \left(\frac{p_S}{p_C}\right)^{50} \left(\frac{1 - p_S}{1 - p_C}\right)^{3250} \approx e^{-30}.$$

Therefore (under these assumptions) the critics' conclusion is much more likely to be true than the Stanford study conclusion.

As shown in Slide 7 of Lecture 11, the maximum likelihood estimate for p is in fact $p_{ML} = 50/3300$, from where the maximum likelihood value of q is $q_{ML} = (p_{ML} - f)/(1 - f) \approx 0.98\%$.

(There may be some important assumption missing from this analysis because even in the absence of false positives, i.e. assuming $f = 0$, it is unclear how to arrive at the 4.16% estimate based on 50 positives out of 3300.)