## Practice questions

1. $X$ is a Geometric($\Theta$) random variable, where $\Theta$ itself is a random variable with PDF $f_\Theta(\theta) = 2\theta$ where $0 \le \theta \le 1$. What are the MAP (Maximum a Posteriori) estimator and ML (Maximum Likelihood) estimates for $\Theta$?

   **Solution:**

   (a) The conditional PDF of $X$ is $f_{X|\Theta}(x|\theta) = \theta(1-\theta)^{x-1}$. Using Bayes' rule, the posterior PDF is
   $$f_{\Theta|X}(\theta|x) \propto f_\Theta(\theta) f_{X|\Theta}(x|\theta) \propto \theta \cdot \theta(1-\theta)^{x-1}$$
   when $0 \le \theta \le 1$. The MAP rule selects $\hat\theta$ that maximizes the posterior (or equivalently its numerator, since the denominator is a positive constant w.r.t. $\theta$). The derivative $(d/d\theta)\theta^2(1-\theta)^{x-1}$ equals zero when $\theta = 0$, 1, or $2/(1+x)$. The maximum is attained at $\hat\theta = 2/(1+x)$ for all $x$. This is the MAP estimate.

   (b) The ML rule chooses the value of $\theta$ that maximizes $f_{X|\Theta}(x|\theta) = \theta(1-\theta)^{x-1}$. The derivative of this function is zero when $\theta = 1$ or $1 - \theta - \theta x + \theta = 0$, which has solution $\theta = 1/x$. The maximum is in fact attained at $\hat\theta = 1/x$ for all $x$, so this is the ML estimate.

2. Jason has two 4-sided dice in a bag. Die A has sides 1, 2, 3, 4 and die B has sides 2, 2, 3, 3. Jason picks one of the dice randomly, rolls it twice, and reports the sum $S$ of the rolls. Your task is to guess which die Jason rolled based on the value of $S$.

   (a) For which values of $S$ would you guess that Jason rolled die A?

   (b) If you guess like in part (a), what is the probability that your guess is wrong?

   **Solution:**

   (a) Let $\Theta$ be the die and $S$ be the sum. By the MAP rule for uniform priors, you should guess that value of $\Theta$ for which $\mathrm{P}(S = s \mid \Theta = \theta)$ is larger. The PMFs are

| $\theta \setminus s$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|
| A | 1/16 | 1/8 | 3/16 | 1/4 | 3/16 | 1/8 | 1/16 |
| B | 0 | 0 | 1/4 | 1/2 | 1/4 | 0 | 0 |

   so the MAP guess is A when $S$ is 2, 3, 7, or 8, and B when $S$ is 4, 5, or 6.

   (b) The event of a wrong guess is $\Theta = \mathrm{A}$ and $S \in \{4, 5, 6\}$. The probability of this event is
   $$\mathrm{P}(S \in \{4,5,6\} \mid \Theta = \mathrm{A})\,\mathrm{P}(\Theta = \mathrm{A}) = \frac{5}{8} \cdot \frac{1}{2} = \frac{5}{16}.$$

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 $\mu$. The government wants to verify that $\mu$ 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) Let $N$ be standard normal and $V$ be the volume of a jar of honey. So, $V = 5N + \mu$. A complying company passes the test if $\mathrm{P}(V \geq t) \geq 0.95$. That is $\mathrm{P}(N \geq (t-\mu)/5) \geq 0.95$. Obviously if a company has honey with mean volume 100 passes the test with probability 95% then all complying companies will pass the test with higher probability. Then we have $t = 5\Phi^{-1}(0.05) + 100 \approx 91.775$, where $\Phi$ is the CDF of $N$.

(b) $\Phi((91.775 - 95)/5) = 0.2595$ 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 $\approx 1 - 0.2595 = 0.7405$.

4. 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 PDF of $X$?

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

   (c) Let $X_1$ and $X_2$ be independent samples of $X$. What is the maximum likelihood estimate of $p$ given that $X_1 = x_1$ and $X_2 = x_2$?

**Solution:**

(a) Let $Y$ be the random variable indicating whether $X$ has mean 1. Then the PDF of $X$ is

$$\begin{aligned} f_X(x;p) &= f_{X|Y}(x|1)f_Y(1) + f_{X|Y}(x|0)f_Y(0) \\ &= \frac{p}{\sqrt{2\pi}}e^{-\frac{(x-1)^2}{2}} + \frac{1-p}{\sqrt{2\pi}}e^{-\frac{(x+1)^2}{2}} \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2+1}{2}}(p(e^x - e^{-x}) + e^{-x}) \end{aligned}$$

(b) The PDF of $X$ is linear in $p$. It has non-negative slope if and only if $e^x \geq e^{-x}$, that is when $x \geq 0$. So the ML estimate of $p$ is

$$\hat{p} = \arg\max_p f_X(x;p) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

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

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

where $C$ is a constant independent of $p$ that doesn't affect the maximizer. The function is differentiable for all values of $x_1, x_2$ because $p \leq 1$. So we could set its derivative to 0 to find its maximizer:

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

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 $\log f_{X_1,X_2}$ 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 $\log f_{X_1,X_2}$ must be negative at the above value of $p$, thus it gives the maxima. The ranges of $x_1$ and $x_2$ such that $0 \leq p \leq 1$ is a bit complicated, but one can verify that

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

5. Coin A has probability of heads 40%. Coin B has probability of tails 40%. One of these coins is tossed is $n$ times. How large does $n$ need to be so that you can identify the coin with probability about 99%? (**Hint:** Use a normal approximation, or write a computer program.)

   **Solution:** Let $A$ be the event that coin A was tossed and $H$ be the number of heads in $n$ tosses. We have $\mathrm{E}[H \mid A] = 0.4n$ and $\mathrm{E}[H \mid \bar{A}] = 0.6n$. The standard deviations are $\sqrt{0.24n}$. Assume $H$ is normal, then $H = N\sqrt{0.24n} + 0.4n$ where $N$ is standard normal. Suppose we identify the coin as A if there are less than $t$ heads in $n$ tosses and as B otherwise. As in both cases have the same standard deviation and the CDFs of $H$ are symmetric along $0.5n$, if $t = 0.5n$ then $\mathrm{P}(H > 0.5n \mid A) = \mathrm{P}(H < 0.5n \mid \bar{A})$. Therefore,

$$\begin{aligned}\mathrm{P}(N\sqrt{0.24n} + 0.4n \geq 0.5n) &= 0.01\\ \Phi(\tfrac{0.1}{\sqrt{0.24}}\sqrt{n}) &= 0.01\\ \tfrac{0.1}{\sqrt{0.24}}\sqrt{n} &= 2.327\\ n &\approx 130\end{aligned}$$

   where $\Phi$ is the CDF of $N$.