1. **Estimating Parameter of Random Graph Given Average Degree**

Consider an Erdős–Rényi random graph on $n$ vertices, in which each edge appears independently with probability $p$. Let $D$ be the average degree of a vertex in the graph. Compute the maximum likelihood estimator of $p$ given $D$. You may approximate $\text{Binomial}(n, p) \approx \text{Poisson}(np)$.

**Solution:** Let $m$ be the number of edges in the graph, so that $D = \frac{2m}{n}$ by the handshake lemma. Write $M = \binom{n}{2}$. Since $m \sim \text{Binomial}(M, p) \approx \text{Poisson}(Mp)$,

$$
P(D = d; p) \approx \frac{M^{nd/2} p^{nd/2}}{(nd/2)!} e^{-Mp}.
$$

To obtain the log-likelihood, we take the logarithm and drop all terms which have no dependence on $p$, which gives the function

$$
\ell(d; p) \approx -\binom{n}{2} p + \frac{nd}{2} \ln p.
$$

Differentiating w.r.t. $p$, we see that the MLE for $p$ is $p = \frac{D}{n - 1}$, which agrees with intuition: the average degree of a node is binomial with $n - 1$ potential neighbors and probability $p$ for each edge, so the expected value of $D$ is $(n - 1)p$. 


2. Estimating Rate of Exponential Distribution

Given $X \in \{0, 1\}$, the random variable $Y$ is Exponentially distributed with rate $3X + 1$.

a. Suppose that $P(X = 1) = p \in (0, 1)$. Find the MAP estimate of $X$ given $Y$.

b. Find the MLE of $X$ given $Y$.

Solution:

a. The MAP maximizes the posterior distribution $f_{X \mid Y}(x \mid y)$ over $x$ for the given observation $y$, which is equivalent to maximizing the joint distribution $f_{X,Y}(x,y)$. We are given that

$$f_{X,Y}(0, y) = f_{Y \mid X}(y \mid 0) \cdot p_X(0) = (1 - p)e^{-y}$$
$$f_{X,Y}(1, y) = f_{Y \mid X}(y \mid 1) \cdot p_X(1) = 4pe^{-4y},$$

and $\text{MAP}(X \mid Y) = 1$ whenever $4pe^{-4Y} > (1 - p)e^{-Y}$. Thus

$$\text{MAP}(X \mid Y) = 1 \left\{ Y < \frac{1}{3} \ln \frac{4p}{1-p} \right\}.$$

b. The MLE is equal to the MAP with uniform prior, i.e. $p = \frac{1}{2}$:

$$\text{MLE}(X \mid Y) = 1 \left\{ Y < \frac{1}{3} \ln 4 \right\} \approx 1\{Y < 0.462\}.$$
3. **Gaussians and the MSE**

Suppose you draw \( n \) i.i.d. data points \((x_1, y_1), \ldots, (x_n, y_n)\), where the true relationship is given by \( Y = WX + \varepsilon \) for \( \varepsilon \sim \mathcal{N}(0, \sigma^2) \). In other words, \( Y \) has a linear dependence on \( X \) with additive Gaussian noise. Show that finding the MLE of \( W \) given the data points \( \{(x_i, y_i)\}_{i=1}^n \) is equivalent to minimizing the cost function

\[
J(w) = \sum_{i=1}^{n} (y_i - wx_i)^2.
\]

**Solution:** The likelihood of the data is

\[
L((x_1, y_1), \ldots, (x_n, y_n) \mid W = w) = \prod_{i=1}^{n} L((x_i, y_i) \mid W = w)
\]
as the data points are conditionally independent given \( W \);

\[
= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} e^{-(y_i - wx_i)^2/(2\sigma^2)}
\]
as the likelihood of \((x_i, y_i)\) given \( W = w \) is the density of \( \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \) evaluated at \( y_i - wx_i \);

\[
\propto \prod_{i=1}^{n} e^{-(y_i - wx_i)^2/(2\sigma^2)},
\]
discarding constant factors that do not depend on the data points or \( w \). We wish to maximize this expression w.r.t. \( w \), but we find it more convenient to work with the log-likelihood

\[
\ell((x_1, y_1), \ldots, (x_n, y_n) \mid W = w) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - wx_i)^2.
\]

We wish to maximize the log-likelihood, which is equivalent to minimizing the cost function

\[
J(w) = \sum_{i=1}^{n} (y_i - wx_i)^2.
\]
4. **Hypothesis Testing for Bernoulli Random Variables**

Suppose that

- If $X = 0$, then $Y \sim \text{Bernoulli}(\frac{1}{4})$.
- If $X = 1$, then $Y \sim \text{Bernoulli}(\frac{3}{4})$.

Using the Neyman–Pearson formulation of hypothesis testing, find the optimal randomized decision rule $r : \{0, 1\} \rightarrow \{0, 1\}$ with respect to the criterion

$$
\min_{r : \{0, 1\} \rightarrow \{0, 1\}} \mathbb{P}(r(Y) = 0 \mid X = 1)
$$

s.t. $\mathbb{P}(r(Y) = 1 \mid X = 0) \leq \beta$,

where $\beta \in [0, 1]$ is a given upper bound on the probability of false alarm (PFA).

**Solution:** The likelihood ratio is the discrete function

$$
L(y) = \frac{f_{Y \mid X}(y \mid 1)}{f_{Y \mid X}(y \mid 0)} = \begin{cases} 
3 & \text{if } y = 1 \\
\frac{1}{3} & \text{if } y = 0.
\end{cases}
$$

By Neyman–Pearson, the optimal decision rule with randomization $r$ is given by

- If $\mathbb{P}(Y = 1 \mid X = 0) = \frac{1}{4} \geq \beta$, then $r(0) = 0$ and $r(1) = 1$ with probability $\gamma = \beta / \frac{1}{4}$;
- Otherwise, $r(1) = 1$ and $r(0) = 1$ with probability $\gamma = \frac{4}{3} \beta - \frac{1}{3}$, which is chosen to make

$$
\text{PFA} = \mathbb{P}(Y = 1 \mid X = 0) + \gamma \cdot \mathbb{P}(Y = 0 \mid X = 0) = \frac{1}{4} + \frac{3}{4} \gamma = \beta.
$$