Statistics for Data Science
Wintersemester 2024/25
Solutions to the 10^{th} exercise

1–2. Please see the scripts `w10e1.py` and `w10e2.py` on the course webpage.

3. Let $y \sim \text{Binomial}(n, p)$ be an observation, where $n \in \mathbb{Z}_+$ is assumed to be known and $p \in (0, 1)$ is assumed to be unknown. The likelihood function is

$$\mathcal{L}(p) = \binom{n}{y}(1-p)^{n-y}p^y \propto (1-p)^{n-y}p^y.$$

The maximum likelihood (ML) estimator is

$$\begin{aligned}\hat{p}_{\text{ML}} &= \arg\max_{p\in(0,1)} (1-p)^{n-y}p^y \qquad \text{(take the logarithm)}\\ &= \arg\max_{p\in(0,1)} \underbrace{\big((n-y)\log(1-p) + y\log p\big)}_{=:f(p)}.\end{aligned}$$

The maximum is obtained precisely at[†]

$$0 = f'(p) = \frac{y-n}{1-p} + \frac{y}{p} = \frac{yp - np + y - yp}{p(1-p)} = \frac{y-np}{p(1-p)} \quad \Leftrightarrow \quad p = \frac{y}{n}.$$

Therefore

$$\hat{p}_{\text{ML}} = \frac{y}{n}.$$

In the present problem, $n = 70$ and $y = 58$, so

$$\hat{p}_{\text{ML}} = \frac{58}{70} = \frac{29}{35}.$$

4. (a) Since we assumed additive Gaussian noise $\eta \sim \nu(\cdot) = \mathcal{N}(0, \gamma^2 I)$, there holds

$$f(y_j|x) = \nu(y_j - F(x)) \propto \exp\left(-\frac{1}{2\gamma^2}\|y_j - F(x)\|^2\right), \quad j = 1, \ldots, n.$$

By independence, we conclude that

$$f(y_1, \ldots, y_n|x) = \prod_{j=1}^n f(y_j|x) \propto \exp\left(-\frac{1}{2\gamma^2}\sum_{j=1}^n \|y_j - F(x)\|^2\right).$$

[†]The function f is concave since $f''(p) = \frac{y-n}{(1-p)^2} - \frac{y}{p^2}$ is negative for all $p \in (0, 1)$ as long as $n > 0$.

(b) Denoting $\overline{y} := \frac{1}{n}\sum_{j=1}^{n} y_j$, we obtain by direct computation that

$$\begin{aligned}
&\sum_{j=1}^{n} \|y_j - F(x)\|^2 \\
&= \sum_{j=1}^{n} y_j^{\mathrm{T}} y_j - 2F(x)^{\mathrm{T}} \underbrace{\sum_{j=1}^{n} y_j}_{=n\overline{y}} + n\|F(x)\|^2 \\
&= n(\|F(x)\|^2 - 2F(x)^{\mathrm{T}}\overline{y} + \overline{y}^{\mathrm{T}}\overline{y}) + n\left(\frac{1}{n}\sum_{j=1}^{n} y_j^{\mathrm{T}} y_j - \overline{y}^{\mathrm{T}}\overline{y}\right) \\
&= n\|F(x) - \overline{y}\|^2 + n\left(\frac{1}{n}\sum_{j=1}^{n} y_j^{\mathrm{T}} y_j - \overline{y}^{\mathrm{T}}\overline{y}\right) \\
&= n\|F(x) - \overline{y}\|^2 + C,
\end{aligned}$$

where $C := n\big(\frac{1}{n}\sum_{j=1}^{n} y_j^{\mathrm{T}} y_j - \overline{y}^{\mathrm{T}}\overline{y}\big)$ is a constant depending on n and $y_1, \ldots, y_n$ (but *not* on x).

(c) We can rewrite the likelihood density as

$$\begin{aligned}
f(y_1, \ldots, y_n|x) &\propto \exp\left(-\frac{1}{2\gamma^2}\sum_{j=1}^{n} \|y_j - F(x)\|^2\right) \\
&= \exp\left(-\frac{n}{2\gamma^2}\|\overline{y} - F(x)\|^2 - \frac{C}{2\gamma^2}\right) \\
&\propto \exp\left(-\frac{1}{2(\gamma^2/n)}\|\overline{y} - F(x)\|^2\right).
\end{aligned}$$

Note that this is precisely the likelihood density $f(\overline{y}|x)$ corresponding to the measurement model

$$\overline{y} = F(x) + \eta, \quad \eta \sim \mathcal{N}\left(0, \frac{\gamma^2}{n} I\right).$$

Averaging n *independent* measurements of a *static* target reduces the uncertainty.