Solution

Question 1

We define a random variable T which represents the time that will take the prisoner to obtain freedom. We are interested in computing $E[T]$. To that end, we define another random variable D which denotes the door that the prisoner choose. Now, using the law of total expectation we can write

$$E[T] = E[E[T|D]] = \sum_{i=1}^{3} E[T|D=i] \cdot P(D=i) = \frac{1}{3}\sum_{i=1}^{3} E[T|D=i],$$

where the last equality is due to the fact that $P(D=1) = P(D=2) = P(D=3) = \frac{1}{3}$.
Notice that if the prisoner chooses the first door he obtains freedom within 3 hours, otherwise, he must go through a tunnel and then repeat the process. Therefore

$$E[T|D=1] = 3,\ E[T|D=2] = 5 + E[T],\ E[T|D=3] = 7 + E[T],$$

which implies that

$$E[T] = \frac{1}{3}\big(3 + 5 + E[T] + 7 + E[T]\big) \Rightarrow E[T] = 15.$$

Question 2

(a) The MLE is given by

$$\begin{aligned}
\hat{y}_{\text{MLE}} &= \arg\max_{y\in\mathbb{R}} \quad p(D|y) \\
&= \arg\max_{y\in\mathbb{R}} \quad \log p(D|y) \\
&= \arg\max_{y\in\mathbb{R}} \quad \log \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-y)^2}{2}} \\
&= \arg\max_{y\in\mathbb{R}} \quad \sum_{i=1}^{n} \left[\log \frac{1}{\sqrt{2\pi}} - \frac{(x_i-y)^2}{2}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} x_i.
\end{aligned}$$

(b) It is given that $y \sim \mathcal{N}(z, 1/s)$, i.e.,

$$p(y) = \frac{1}{\sqrt{2\pi/s}} e^{-\frac{(y-z)^2}{2/s}}.$$

Hence, the MAP estimator $\hat{y}_1$ is given by

$$\begin{aligned}
\hat{y}_1 &= \arg\max_{y\in\mathbb{R}} \quad p(D|y)p(y) \\
&= \arg\max_{y\in\mathbb{R}} \quad \log p(D|y)p(y) \\
&= \arg\max_{y\in\mathbb{R}} \quad \log p(D|y) + \log p(y) \\
&= \arg\max_{y\in\mathbb{R}} \quad \sum_{i=1}^{n} \left[\log \frac{1}{\sqrt{2\pi}} - \frac{(x_i-y)^2}{2}\right] + \log \frac{1}{\sqrt{2\pi/s}} - \frac{(y-z)^2}{2/s}.
\end{aligned}$$

By taking the derivative with respect to y we get

$$\begin{aligned}
&\sum_{i=1}^{n}(x_i - y) - s(y-z) = 0. \\
\Rightarrow \quad \hat{y}_1 &= \frac{1}{n+s}\sum_{i=1}^{n} x_i + \frac{sz}{n+s} \\
&= \frac{n}{n+s}\hat{y}_{\text{MLE}} + \frac{sz}{n+s}.
\end{aligned}$$

(c) Now $y \sim U[-1,1]$, i.e.,

$$p(y) = \begin{cases} 1/2, & -1 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Hence, the MAP estimator $\hat{y}_2$ is given by

$$\begin{aligned}
\hat{y}_2 &= \arg\max_{y \in \mathbb{R}} \quad p(D|y)p(y) \\
\hat{y}_2 &= \arg\max_{y \in [-1,1]} \quad p(D|y)/2 \\
&= \arg\max_{y \in [-1,1]} \quad \log p(D|y) \\
&= \arg\max_{y \in [-1,1]} \quad \sum_{i=1}^{n} \Big[\log \frac{1}{\sqrt{2\pi}} - \frac{(x_i - y)^2}{2} \Big].
\end{aligned}$$

Computing the derivative with respect to y we have

$$n \Big(\frac{1}{n} \sum_{i=1}^{n} x_i - y \Big) = 0.,$$

which implies that $\hat{y}_2 = \frac{1}{n} \sum_{i=1}^{n} x_i = \hat{y}_{\text{MLE}}$. However, this is true only if $-1 \le \hat{y}_{\text{MLE}} \le 1$. Otherwise, if $\hat{y}_{\text{MLE}} > 1$ then the derivative is positive, the function is monotonically increasing and the maximum is obtained at $\hat{y}_2 = 1$. Similarly, if $\hat{y}_{\text{MLE}} > 1$ then the maximum is achieved at $\hat{y}_2 = -1$. Overall

$$\hat{y}_2 = \begin{cases} -1, & \hat{y}_{\text{MLE}} < -1, \\ \hat{y}_{\text{MLE}}, & -1 \le \hat{y}_{\text{MLE}} \le 1, \\ 1, & \hat{y}_{\text{MLE}} > 1. \end{cases}$$

(d)
- For $n \to \infty$ we have

$$\begin{aligned}
& \hat{y}_{\text{MLE}} \to y, \\
& \hat{y}_1 \to \hat{y}_{\text{MLE}}, \\
\Rightarrow \quad & \hat{y}_1 \to y.
\end{aligned}$$

  However, $\hat{y}_2 \to y$ only if $y \in [-1,1]$, hence, we will prefer to use $\hat{y}_1$.
- For $n = 1$ we get

$$\begin{aligned}
\hat{y}_{\text{MLE}} &= x_1, \\
\hat{y}_1 &= \frac{1}{11} x_1 + \frac{100}{11}.
\end{aligned}$$

  To compare the two estimators, we can compute their conditional expectation given y

$$\begin{aligned}
E[\hat{y}_1|y] &= \frac{1}{11} E[x_1|y] + \frac{100}{11} = \frac{1}{11} y + \frac{100}{11} \approx 0.1y + 9. \\
E[\hat{y}_2|y] &= -1 \cdot P(y < -1) + y \cdot P(-1 \le y \le 1) + 1 \cdot P(y > 1) \approx 0.4y.
\end{aligned}$$

  Therefore, $\hat{y}_1$ is strongly biased and we will prefer to use $\hat{y}_2$.

Question 3

(a) The MLE is

$$\begin{aligned}
\hat{\lambda}_{\text{MLE}} = \arg\max_{\lambda} \ \log p(D|\lambda) &= \arg\max_{\lambda} \ \sum_{i=1}^{n} \log \left(\frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right) \\
&= \arg\max_{\lambda} \ \sum_{i=1}^{n} (-\log x_i! + x_i \log \lambda - \lambda)
\end{aligned}$$

Setting the derivative w.r.t. λ to 0, we get

$$\begin{aligned}
& \sum_{i=1}^{n} \Big(\frac{x_i}{\lambda} - 1 \Big) = 0 \\
\Rightarrow \ & \hat{\lambda}_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^{n} x_i.
\end{aligned}$$

The expectation of $\hat{\lambda}_{\text{MLE}}$ is

$$E[\hat{\lambda}_{\text{MLE}}] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{1}{n}\sum_{i=1}^{n} \lambda = \lambda.$$

Hence, $\hat{\lambda}_{\text{MLE}}$ is an unbiased estimator.

(b) First, given a single sample x, we have that $\hat{\lambda}_{\text{MLE}} = x$. As we saw in class, since y is a function of λ we get

$$\hat{y}_{\text{MLE}} = e^{-2\hat{\lambda}_{\text{MLE}}} = e^{-2x}.$$

The expectation of $\hat{y}_{\text{MLE}}$ is

$$\begin{aligned} E[\hat{y}_{\text{MLE}}] = E[e^{-2x}] &= \sum_{k=0}^{\infty} e^{-2k}\frac{\lambda^k}{k!}e^{-\lambda} \\ &= \sum_{k=0}^{\infty} \frac{(\lambda e^{-2})^k}{k!}e^{-\lambda} \\ &= e^{-\lambda}e^{\lambda e^{-2}} \sum_{k=0}^{\infty} \frac{(\lambda e^{-2})^k}{k!}e^{-\lambda e^{-2}} \\ &= e^{-\lambda}e^{\lambda e^{-2}} = e^{-\lambda(1-e^{-2})} \neq e^{-2\lambda}. \end{aligned}$$

Hence, the estimator is biased.

(c) An unbiased estimator $y_U = g(x)$ should satisfy

$$E[g(x)] = e^{-2\lambda}.$$

Using the Taylor expansion of $e^{-\lambda} = \sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!}$ we get

$$E[g(x)] = \sum_{k=0}^{\infty} g(k)\frac{\lambda^k}{k!}e^{-\lambda} = e^{-2\lambda} = e^{-\lambda}e^{-\lambda} = \sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!}e^{-\lambda}.$$

By equating the coefficients of both sides we get an unbiased estimator

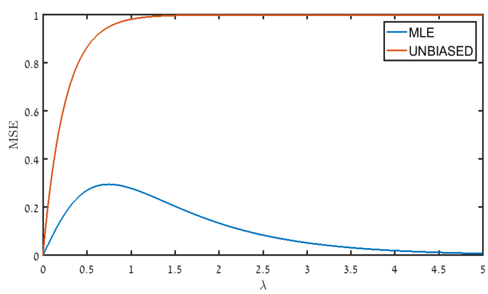
$$\begin{aligned} g(k)\lambda^k &= (-\lambda)^k = (-1)^k(\lambda)^k \\ \Rightarrow y_U &= g(x) = (-1)^x. \end{aligned}$$

(d) The MSE of the estimator are given by

$$\begin{aligned} E\left[(\hat{y}_{\text{MLE}} - e^{-2\lambda})^2\right] &= E\left[(e^{-2x} - e^{-2\lambda})^2\right] \\ &= E\left[e^{-4x} - 2e^{-2\lambda}e^{-2x} + e^{-4\lambda}\right] \\ &= e^{-\lambda(1-e^{-4})} - 2e^{-2\lambda}e^{-\lambda(1-e^{-2})} + e^{-4\lambda} \end{aligned}$$

$$\begin{aligned} E\left[(y_U - e^{-2\lambda})^2\right] &= E\left[((-1)^x - e^{-2\lambda})^2\right] \\ &= E\left[1 - 2e^{-2\lambda}(-1)^x + e^{-4\lambda}\right] \\ &= 1 - 2e^{-2\lambda}e^{-2\lambda} + e^{-4\lambda} = 1 - e^{-4\lambda} \end{aligned}$$

Therefore, $\hat{y}_{\text{MLE}}$ has a lower MSE, as illustrated below



Question 4

For brevity, we denote $\hat{\mu} = \hat{\mu}_{\mathrm{MLE}}$ and $\hat{\Sigma} = \hat{\Sigma}_{\mathrm{MLE}}$

- $\hat{\mu} = \frac{1}{n}\sum_{k=1}^{n} x_k$.
- $\hat{\Sigma} = \frac{1}{n}\sum_{k=1}^{n} (x_k - \hat{\mu})(x_k - \hat{\mu})^T$.

Now, we compute the expectation of $\hat{\Sigma}$

$$E\left[\hat{\Sigma}\right] = E\left[\frac{1}{n}\sum_{k=1}^{n}\left(\underline{x}_k - \underline{\hat{\mu}}\right)\left(\underline{x}_k - \underline{\hat{\mu}}\right)^T\right] = \frac{1}{n}\sum_{k=1}^{n} E\left[\left(\underline{x}_k - \underline{\hat{\mu}}\right)\left(\underline{x}_k - \underline{\hat{\mu}}\right)^T\right] =$$

$$= \frac{1}{n}\sum_{k=1}^{n} E\left[\underline{x}_k \cdot \underline{x}_k^T - \underline{x}_k \cdot \underline{\hat{\mu}}^T - \underline{\hat{\mu}} \cdot \underline{x}_k^T + \underline{\hat{\mu}} \cdot \underline{\hat{\mu}}^T\right] =$$

$$= \frac{1}{n}\sum_{k=1}^{n} E\left[\underline{x}_k \cdot \underline{x}_k^T - \underline{x}_k \cdot \frac{1}{n}\sum_{i=1}^{n}\underline{x}_i^T - \frac{1}{n}\sum_{i=1}^{n}\underline{x}_i \cdot \underline{x}_k^T + \frac{1}{n}\sum_{i=1}^{n}\underline{x}_i \cdot \frac{1}{n}\sum_{i=1}^{n}\underline{x}_i^T\right] =$$

$$= \frac{1}{n}\sum_{k=1}^{n}\left\{E\left[\underline{x}_k \cdot \underline{x}_k^T\right] - \frac{1}{n}\sum_{i=1}^{n}E\left[\underline{x}_k \cdot \underline{x}_i^T\right] - \frac{1}{n}\sum_{i=1}^{n}E\left[\underline{x}_i \cdot \underline{x}_k^T\right] + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left[\underline{x}_i \cdot \underline{x}_j^T\right]\right\} =$$

$$= \frac{1}{n}\sum_{k=1}^{n}\left\{E\left[\underline{x}_k \cdot \underline{x}_k^T\right] - \frac{2}{n}\sum_{\substack{i=1\\ i\neq k}}^{n}E\left[\underline{x}_k \cdot \underline{x}_i^T\right] - \frac{2}{n}E\left[\underline{x}_k \cdot \underline{x}_k^T\right] + \frac{1}{n^2}\sum_{i=j}E\left[\underline{x}_i \cdot \underline{x}_i^T\right] + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}E\left[\underline{x}_i \cdot \underline{x}_j^T\right]\right\} =$$

$$= \frac{1}{n}\sum_{k=1}^{n}\left\{\left(\frac{n-2}{n}\right)E\left[\underline{x}_k \cdot \underline{x}_k^T\right] - \frac{2}{n}\sum_{\substack{i=1\\ i\neq k}}^{n}\underline{\mu} \cdot \underline{\mu}^T + \frac{1}{n^2}\sum_{i=j}E\left[\underline{x}_i \cdot \underline{x}_i^T\right] + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\underline{\mu} \cdot \underline{\mu}^T\right\} =$$

$$= \frac{1}{n}\sum_{k=1}^{n}\left\{\left(\frac{n-2}{n}\right)E\left[\underline{x}_k \cdot \underline{x}_k^T\right] - \frac{2(n-1)}{n}\underline{\mu} \cdot \underline{\mu}^T + \frac{n}{n^2}E\left[\underline{x} \cdot \underline{x}^T\right] + \frac{n(n-1)}{n^2}\underline{\mu} \cdot \underline{\mu}^T\right\} =$$

$$= \left(\frac{n-1}{n}\right)\frac{1}{n}\sum_{k=1}^{n}\left\{E\left[\underline{x} \cdot \underline{x}^T\right] - 2\underline{\mu} \cdot \underline{\mu}^T + \underline{\mu} \cdot \underline{\mu}^T\right\} =$$

$$= \left(\frac{n-1}{n}\right)\frac{1}{n}\sum_{k=1}^{n}E\left[\underline{x} \cdot \underline{x}^T - \underline{x} \cdot \underline{\mu}^T - \underline{\mu} \cdot \underline{x}^T + \underline{\mu} \cdot \underline{\mu}^T\right]$$

$$= \left(\frac{n-1}{n}\right)\frac{1}{n}\sum_{k=1}^{n}E\left[\left(\underline{x} - \underline{\mu}\right) \cdot \left(\underline{x} - \underline{\mu}\right)^T\right] = \left(\frac{n-1}{n}\right)\frac{1}{n}\sum_{k=1}^{n}\Sigma = \left(\frac{n-1}{n}\right)\frac{n}{n}\Sigma$$

$$\Rightarrow \boxed{E\left[\hat{\Sigma}\right] = \frac{n-1}{n}\Sigma}$$

Therefore, $\hat{\Sigma}_{\mathrm{MLE}}$ is biased. An unbiased estimator is

$$\hat{\Sigma}_2 = \frac{n}{n-1}\hat{\Sigma}_{\mathrm{MLE}} = \frac{n}{n-1}\frac{1}{n}\sum_{k=1}^{n}(x_k - \hat{\mu})(x_k - \hat{\mu})^T = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \hat{\mu})(x_k - \hat{\mu})^T.$$

Question 5

(a) First, notice that N should satisfy

$$N \geq x_i,\ i = 1, ..., k \Rightarrow\ N \geq \max_{i=1,\ldots,k} x_i \triangleq x_{\max}.$$

The ML estimator is given by

$$\hat{N}_{\text{MLE}} = \underset{N \geq x_{\max}}{\arg\max}\, p(x_1, x_2, ..., x_k|N) = \underset{N \geq x_{\max}}{\arg\max}\, p(x_1|N)p(x_2|x_1, N) \cdots (x_k|x_1, x_2, ..., x_{k-1}, N)$$

When the first sample is drawn we have a total number of N number to choose x_1 from (i.e. $p(x_1|N) = \frac{1}{N}$, for the second we have $N - 1$ number to choose from, and so on. Hence. the likelihood becomes

$$\begin{aligned}
\hat{N}_{\text{MLE}} &= \underset{N \geq x_{\max}}{\arg\max}\, p(x_1, x_2, ..., x_k|N) \\
&= \underset{N \geq x_{\max}}{\arg\max}\, \frac{1}{N}\frac{1}{N-1} \cdots \frac{1}{N-(k-1)} \\
&= \underset{N \geq x_{\max}}{\arg\max}\, -\log\Big(N(N-1)\cdots(N-k+1)\Big) \\
&= \underset{N \geq x_{\max}}{\arg\max}\, -\log\left(\binom{N}{k}k!\right) \\
&= \underset{N \geq x_{\max}}{\arg\max}\, -\log\binom{N}{k} - \log(k!) \\
&= \underset{N \geq x_{\max}}{\arg\min}\, \log\binom{N}{k}
\end{aligned}$$

Hence, we need to minimize N which implies that $\hat{N}_{\text{MLE}} = x_{\max}$.

(b) The expectation of $\hat{N}_{\text{MLE}}$ is

$$\begin{aligned}
E[\hat{N}_{\text{MLE}}] &= \sum_{i=1}^{N} iP(\hat{N}_{\text{MLE}} = i) \\
&= \sum_{i=k}^{N} iP(\hat{N}_{\text{MLE}} = i) \\
&= \sum_{i=k}^{N} i\Big(P(\hat{N}_{\text{MLE}} \leq i) - P(\hat{N}_{\text{MLE}} \leq i-1)\Big) \\
&= k\Big(P(\hat{N}_{\text{MLE}} \leq k) - 0\Big) + \sum_{i=k+1}^{N-1} i\Big(P(\hat{N}_{\text{MLE}} \leq i) - P(\hat{N}_{\text{MLE}} \leq i-1)\Big) + N\Big(1 - P(\hat{N}_{\text{MLE}} \leq N-1)\Big) \\
&= k\frac{\binom{k}{k}}{\binom{N}{k}} + \sum_{i=k+1}^{N-1} i\left(\frac{\binom{i}{k}}{\binom{N}{k}} - \frac{\binom{i-1}{k}}{\binom{N}{k}}\right) + N\left(1 - \frac{\binom{N-1}{k}}{\binom{N}{k}}\right) \\
&= N - \sum_{i=k}^{N-1} \frac{\binom{i}{k}}{\binom{N}{k}} \\
&= N - \frac{\binom{N}{k+1}}{\binom{N}{k}} = N - \frac{N-k}{k+1} = \frac{k}{k+1}(N+1)
\end{aligned}$$

Hence, the estimator is biased.

(c) The unbiased estimator is simply $\hat{N}_U = \frac{k+1}{k}\hat{N}_{\text{MLE}} - 1$.

(d) Simulation code:

```
N = 300;
k = 5;
r = 1e4;

MLE = 0;
NU  = 0;

for ii = 1:r

    x = randperm(N,k);
    MLE = MLE + max(x);
    NU  = NU + (k+1)/k*max(x)-1;

end

MLE = MLE/r;
NU  = NU/r;
```

Listing 1: MATLAB example

The results are $\hat{N}_{\text{MLE}} = 250.9$ and $\hat{N}_U = 300.1$. Clearly, $\hat{N}_U$ is a better estimator of N.