DATA 221 Homework 1

Chris Low

Problem 1

1(a)

For each 30 minute interval,

$$P(X_i = x_i \mid \lambda) \ = \ e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}, \qquad x_i \geq 0.$$

Assuming independence across the 16 intervals, the likelihood is the joint probability:

$$L(\lambda \mid x_1, \dots, x_{16}) \ = \ \prod_{i=1}^{16} P(X_i = x_i \mid \lambda) \ = \ \prod_{i=1}^{16} \left(e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} \right).$$

Collect terms:

$$\prod_{i=1}^{16} e^{-\lambda} = e^{-16\lambda}, \qquad \prod_{i=1}^{16} \lambda^{x_i} = \lambda^{\sum_{i=1}^{16} x_i}, \qquad \prod_{i=1}^{16} \frac{1}{x_i!} = \frac{1}{\prod_{i=1}^{16} x_i!}.$$

Therefore,

$$\boxed{L(\lambda \mid x_1, \dots, x_{16}) = e^{-16\lambda}\ \frac{\lambda^{\sum_{i=1}^{16} x_i}}{\prod_{i=1}^{16} x_i!}}.$$

1(b)

From 1(a),

$$L(\lambda \mid x_1, \dots, x_{16}) = e^{-16\lambda} \frac{\lambda^{\sum_{i=1}^{16} x_i}}{\prod_{i=1}^{16} x_i!}.$$

Take logs:

$$\ell(\lambda) = \log L(\lambda \mid x_1, \dots, x_{16}) = \log\left(e^{-16\lambda}\right) + \log\left(\lambda^{\sum_{i=1}^{16} x_i}\right) - \log\left(\prod_{i=1}^{16} x_i!\right).$$

Simplify:

$$\log\left(e^{-16\lambda}\right) = -16\lambda, \qquad \log\left(\lambda^{\sum x_i}\right) = \left(\sum_{i=1}^{16} x_i\right)\log\lambda, \qquad \log\left(\prod x_i!\right) = \sum_{i=1}^{16}\log(x_i!).$$

Therefore,

$$\boxed{\ell(\lambda) = -16\lambda + \left(\sum_{i=1}^{16} x_i\right)\log\lambda - \sum_{i=1}^{16}\log(x_i!)}.$$

1(c)

Differentiate:

$$\frac{d}{d\lambda}\ell(\lambda) = -16 + \left(\sum_{i=1}^{16} x_i\right)\frac{1}{\lambda}.$$

Set to zero and solve:

$$-16 + \frac{\sum_{i=1}^{16} x_i}{\lambda} = 0 \;\Rightarrow\; \hat{\lambda} = \frac{1}{16}\sum_{i=1}^{16} x_i.$$

Second derivative:

$$\ell''(\lambda) = -\frac{\sum_{i=1}^{16} x_i}{\lambda^2} < 0 \quad \text{for } \lambda > 0,$$

So we know that $\ell'(\lambda)$ is concave, and thus the critical point is a maximizer of the log-likelihood.

Problem 2

2(a)

From Problem 1,

$$\hat{\lambda} = \frac{1}{16}\sum_{i=1}^{16} x_i.$$

Here,

$$\sum_{i=1}^{16} x_i = 442 \quad \Rightarrow \quad \hat{\lambda} = \frac{442}{16} = \boxed{27.625}.$$

2(b)

Define the negative log-likelihood for Poisson data (from 1(b)):

$$\mathrm{NLL}(\lambda) = -\sum_{i=1}^{16} \log\left(e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}\right) = \sum_{i=1}^{16} \left(\lambda - x_i \log\lambda + \log(x_i!)\right).$$

Minimizing this function over $\lambda > 0$ gives us the MLE.

```python
import numpy as np
from scipy.optimize import minimize_scalar
from scipy.special import gammaln

x = np.array([28, 33, 21, 27, 24, 35, 26, 30, 18, 29, 31, 22, 34, 25, 27, 32], dtype=float)

def neg_log_likelihood(lam):
    if lam <= 0:
        return float("inf")
    # log-likelihood: sum(-lam + x_i log(lam) - log(x_i!))
    log_like = np.sum(-lam + x * np.log(lam) - gammaln(x + 1))
    return -log_like

result = minimize_scalar(neg_log_likelihood, bounds=(1e-8, 200), method="bounded")
lam_optimizer = result.x
lam_formula = np.mean(x)

print(lam_formula)
print(lam_optimizer)
```

```
27.625
27.62500006114126
```

The optimizer returns $\hat{\lambda} \approx 27.625$, which matches our closed-form estimate from part (a).

2(c)

We have a Poisson rate λ per 30 minutes. One hour is two consecutive 30 minute calls. We let X_1 be calls in the first half hour and let X_2 be calls in the second half hour. The model assumes that X_1 and X_2 are independent and each is Poisson($\hat{\lambda}$).

$$Y = X_1 + X_2 \sim \text{Poisson}(2\hat{\lambda}).$$

$\hat{\lambda} = 27.625$, so we have $2\hat{\lambda} = 55.25$. T

$$P(Y \geq 60) = 1 - P(Y \leq 59) = 1 - \sum_{k=0}^{59} e^{-55.25} \frac{55.25^k}{k!}.$$

```python
from scipy.stats import poisson

mu = 2 * result.x
print(poisson.sf(59, mu))
```

```
0.278650610261531
```

$$P(Y \geq 60) \approx 0.279.$$