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.

```
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
```

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

Problem 3

3(a)

The possible outcomes summing two dice to 6 are

$$(1,5),(2,4),(3,3),(4,2),(5,1)$$

which are equally likely.

Only one of these outcomes has the first die equal to 2, which is $(2,4)$. Therefore,

$$P(\text{first die} = 2 \mid \text{sum} = 6) = \boxed{\frac{1}{5}}$$

3(b)

By Bayes' rule,

$$P(P \mid T) = \frac{P(T \mid P)P(P)}{P(T)}.$$

Substituting the given values,

$$P(P \mid T) = \frac{(0.30)(0.55)}{0.45} = \frac{11}{30} \approx \boxed{0.367}$$

Problem 4

4(a)

$D = (4,3,8,6,2,7,4,5)$.

We know that die A counts for $(4,3,8,6,2,7,4,5)$ are $(3,4,1,1,5,1,3,2)$, and we assume that each roll from D is independent of each other, so

$$P(D \mid A) = \frac{3 \cdot 4 \cdot 1 \cdot 1 \cdot 5 \cdot 1 \cdot 3 \cdot 2}{24^8} = \frac{360}{24^8}.$$

Die B counts are $(3,3,2,2,3,2,3,3)$, so

$$P(D \mid B) = \frac{3 \cdot 3 \cdot 2 \cdot 2 \cdot 3 \cdot 2 \cdot 3 \cdot 3}{24^8} = \frac{1944}{24^8}.$$

Die C is uniform with count 2 for every face, so

$$P(D \mid C) = \frac{2^8}{24^8} = \frac{256}{24^8}.$$

We have equal priors $P(A) = P(B) = P(C) = \frac{1}{3}$, and so the common factor $\frac{1}{3 \cdot 24^8}$ cancels, so

$$P(A \mid D) : P(B \mid D) : P(C \mid D) = 360 : 1944 : 256.$$

Sum is 2560. Therefore,

$$P(A \mid D) = \frac{360}{2560} = \boxed{\frac{9}{64} = 0.140625}, \quad P(B \mid D) = \frac{1944}{2560} = \boxed{\frac{243}{320} = 0.759375}, \quad P(C \mid D) = \frac{256}{2560} = \boxed{\frac{1}{10} = 0.1}$$

4(b)

We observe one more roll: $x_{\text{new}} = 12$.

Because rolls are conditionally independent given the die, we get:

$$P(D, 12 \mid d) = P(D \mid d)\, P(12 \mid d).$$

From part (a), we already had:

$$P(D \mid A) = \frac{360}{24^8}, \quad P(D \mid B) = \frac{1944}{24^8}, \quad P(D \mid C) = \frac{256}{24^8}.$$

Now multiply by the probability of rolling a 12 under each die. Since probabilities are counts divided by 24,

$$P(12 \mid A) = \frac{0}{24}, \quad P(12 \mid B) = \frac{1}{24}, \quad P(12 \mid C) = \frac{2}{24}.$$

So the extended likelihoods become

$$P(D, 12 \mid A) = \frac{360}{24^8} \cdot \frac{0}{24} = \frac{360 \cdot 0}{24^9} = 0,$$

$$P(D, 12 \mid B) = \frac{1944}{24^8} \cdot \frac{1}{24} = \frac{1944 \cdot 1}{24^9},$$

$$P(D, 12 \mid C) = \frac{256}{24^8} \cdot \frac{2}{24} = \frac{256 \cdot 2}{24^9} = \frac{512}{24^9}.$$

With equal priors $P(A) = P(B) = P(C) = \frac{1}{3}$, the unnormalized posteriors are

$$P(d \mid D, 12) \propto P(D, 12 \mid d)\, P(d).$$

But the factors $\frac{1}{24^9}$ and $\frac{1}{3}$ are common to all three models, so they cancel when we normalize. So:

$$w_A = 360 \cdot 0 = 0, \quad w_B = 1944 \cdot 1 = 1944, \quad w_C = 256 \cdot 2 = 512.$$

Normalize by the total weight:

$$w_A + w_B + w_C = 0 + 1944 + 512 = 2456.$$

Therefore,

$$\boxed{P(A \mid D, 12) = \frac{0}{2456} = 0},$$

$$\boxed{P(B \mid D, 12) = \frac{1944}{2456} \approx 0.7915},$$

$$\boxed{P(C \mid D, 12) = \frac{512}{2456} \approx 0.2085}.$$

4(c)

If a die assigns probability 0 to a face that actually appears, then the likelihood for that die becomes 0, so its posterior immediately becomes 0 and stays 0 forever under Bayes updating.

In log space this shows up as $\log(0) = -\infty$, so adding the log likelihoods gives $-\infty$ for that die. In practice, we avoid exact zeros by giving every face a small positive probability, so one unexpected roll does not immediately eliminate a die.

4(d)

```
import numpy as np

faces = np.arange(1, 13)
probs_b = np.array([3, 3, 3, 3, 3, 2, 2, 2, 1, 1, 0, 1]) / 24.0
probs_c = np.array([2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2]) / 24.0

def run_simulation(true_die_probs, target_odds=99):
    log_odds = 0
    rolls = 0
    # Log-odds form: log(P(B)/P(C)) = Sum(log(P(x|B)) - log(P(x|C)))
    threshold = np.log(target_odds)
```

```python
    while abs(log_odds) < threshold:
        roll = np.random.choice(faces, p=true_die_probs)
        idx = roll - 1

        term_b = probs_b[idx]
        term_c = probs_c[idx]

        if term_b == 0:
            log_odds = -np.inf # B is impossible
        elif term_c == 0:
            log_odds = np.inf # C is impossible
        else:
            log_odds += np.log(term_b) - np.log(term_c)

        rolls += 1

    return rolls

# Run Monte Carlo
n_sims = 10000
rolls_given_b = [run_simulation(probs_b) for _ in range(n_sims)]
rolls_given_c = [run_simulation(probs_c) for _ in range(n_sims)]

print(f"Avg rolls given we are rolling B: {np.mean(rolls_given_b):.2f}")
print(f"Avg rolls given we are rolling C: {np.mean(rolls_given_c):.2f}")
```

```
Avg rolls given we are rolling B: 29.06
Avg rolls given we are rolling C: 11.82
```

We get: True die B: average rolls ≈ 29.01. True die C: average rolls ≈ 11.75.

The results might vary slightly across each run.

Problem 5

5(a)

Each visitor is an independent Bernoulli(π) trial, so:

$$\boxed{P(k \mid \pi) = \binom{16}{k} \pi^k (1-\pi)^{16-k}}$$

(b)

We know that

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

The prior is uniform on $[0, 1]$, so $p(\pi) \propto 1$ for $0 \leq \pi \leq 1$. It is a constant with respect to π.

Thus,

$$p(\pi \mid k) \propto P(k \mid \pi)\, p(\pi) \propto \binom{16}{k} \pi^k (1-\pi)^{16-k}.$$

Dropping the constant $\binom{16}{k}$,

$$\boxed{p(\pi \mid k) \propto \pi^k (1-\pi)^{16-k}, \quad 0 \leq \pi \leq 1.}$$

(c)

Maximize the posterior expression,

$$f(\pi) = \pi^k (1-\pi)^{16-k}.$$

Take logs:

$$\ell(\pi) = k \ln \pi + (16-k) \ln(1-\pi).$$

Differentiate and set to zero:

$$\ell'(\pi) = \frac{k}{\pi} - \frac{16-k}{1-\pi} = 0 \Rightarrow k(1-\pi) = (16-k)\pi \Rightarrow k = 16\pi$$

Therefore,

$$\boxed{\hat{\pi}_{MAP} = \frac{k}{16}}$$

(d)

If $k = 11$, then

$$\widehat{\pi}_{MLE} = \boxed{\frac{11}{16} = 0.6875}.$$

Problem 6

6(a)

The prior is $\pi \sim \text{Beta}(\alpha, \beta)$. For this problem, $(\alpha, \beta) = (3, 2)$.

So the prior density is

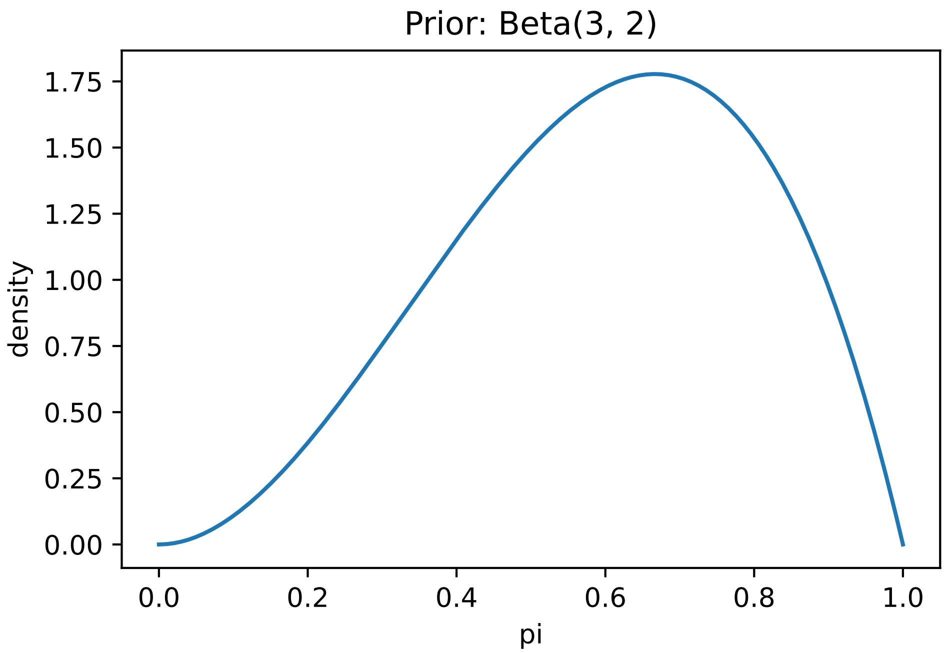
$$p(\pi) = \frac{\pi^{\alpha-1}(1-\pi)^{\beta-1}}{B(\alpha,\beta)} = \frac{\pi^{3-1}(1-\pi)^{2-1}}{B(3,2)} = \frac{\pi^2(1-\pi)^1}{B(3,2)}, \quad 0 \le \pi \le 1.$$

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import beta

alpha, beta_param = 3, 2

x = np.linspace(0, 1, 500)
y = beta.pdf(x, alpha, beta_param)

plt.figure()
plt.plot(x, y)
plt.title("Prior: Beta(3, 2)")
plt.xlabel("pi")
plt.ylabel("density")
plt.show()
```



6(b)

(b)

From Problem 5, the Binomial likelihood for observing k signups out of n visitors is

$$P(k \mid \pi) = \binom{n}{k} \pi^k (1-\pi)^{n-k}.$$

The Beta prior for π is

$$p(\pi) = \frac{1}{B(\alpha, \beta)} \pi^{\alpha-1} (1-\pi)^{\beta-1}.$$

The posterior distribution is proportional to the product of the likelihood and the prior:

$$p(\pi \mid k) \propto P(k \mid \pi)\, p(\pi).$$

Dropping constants that do not depend on π (namely $\binom{n}{k}$ and $1/B(\alpha, \beta)$),

$$p(\pi \mid k) \propto \pi^k (1-\pi)^{n-k}\, \pi^{\alpha-1} (1-\pi)^{\beta-1}.$$

Combine exponents:

$$\boxed{p(\pi \mid k) \propto \pi^{(\alpha+k)-1} (1-\pi)^{(\beta+n-k)-1}.}$$

This has the form of a Beta density. Therefore,

$$\boxed{\pi \mid k \sim \mathrm{Beta}(\alpha + k,\ \beta + n - k).}$$

6(c)

From part (b),

$$\pi \mid k \sim \mathrm{Beta}(\alpha + k,\ \beta + n - k).$$

Plug in $n = 16$, $k = 11$, $(\alpha, \beta) = (3, 2)$:

$$\alpha' = \alpha + k = 3 + 11 = 14, \qquad \beta' = \beta + n - k = 2 + (16 - 11) = 2 + 5 = 7.$$

So the posterior is $\boxed{Beta(14, 7)}$.

6(d)

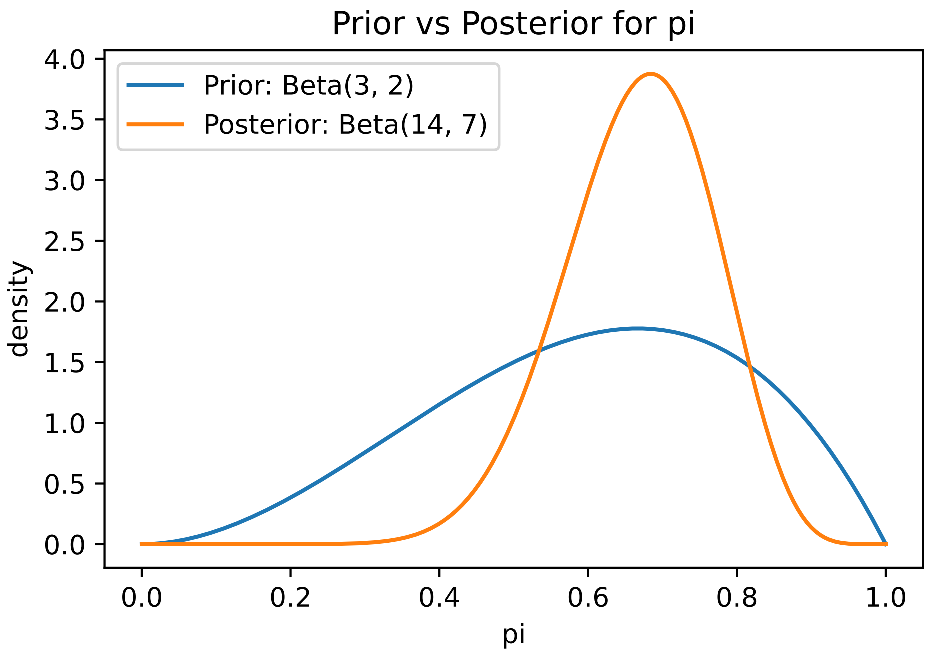
```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import beta

alpha, beta_param = 3, 2
n, k = 16, 11

alpha_post = alpha + k
beta_post = beta_param + (n - k)

x = np.linspace(0, 1, 500)
prior = beta.pdf(x, alpha, beta_param)
post = beta.pdf(x, alpha_post, beta_post)

plt.figure()
plt.plot(x, prior, label="Prior: Beta(3, 2)")
plt.plot(x, post, label=f"Posterior: Beta({alpha_post}, {beta_post})")
plt.title("Prior vs Posterior for pi")
plt.xlabel("pi")
plt.ylabel("density")
plt.legend()
plt.show()
```



The posterior curve will be narrower (lower variance) and shifted to the right compared to the prior. This shows the data ($\frac{11}{16}$ successes) pulling the estimate higher than the prior mean ($\frac{3}{5}$). The posterior is narrower because the data reduce uncertainty.

6(e)

We use the posterior from part (c): $\pi \mid k \sim \text{Beta}(\alpha', \beta') = \text{Beta}(14, 7)$.

MAP from maximizing the posterior density

From 6(b), we know that the posterior is proportional to the Beta density. A Beta(a, b) density has the form

$$p(\pi) = \frac{1}{B(a,b)}\pi^{a-1}(1-\pi)^{b-1}, \quad 0 < \pi < 1.$$

Dropping the constants:

$$p(\pi) \propto \pi^{a-1}(1-\pi)^{b-1}.$$

Take logs:

$$\ell(\pi) = (a-1)\log \pi + (b-1)\log(1-\pi).$$

Differentiate:

$$\ell'(\pi) = \frac{a-1}{\pi} - \frac{b-1}{1-\pi}.$$

Set $\ell'(\pi) = 0$:

$$\frac{a-1}{\pi} = \frac{b-1}{1-\pi}.$$

$$(a-1)(1-\pi) = (b-1)\pi \Rightarrow a-1 = (a+b-2)\pi$$

$$\boxed{\hat{\pi}_{MAP} = \frac{a-1}{a+b-2}.}$$

Here $a = 14$ and $b = 7$, so

$$\boxed{\hat{\pi}_{MAP} = \frac{14-1}{14+7-2} = \frac{13}{19} \approx 0.6842.}$$

Compare to MLE

From Problem 5, the MLE is

$$\hat{\pi}_{MLE} = \frac{k}{n} = \frac{11}{16} = 0.6875.$$

The MAP and MLE are very close, with the MAP slightly smaller due to the influence of the Beta$(3, 2)$ prior.