SOLUTION: DO NOT DISTRIBUTE

Data Fundamentals (H)\\COMPSCI 4073

Total marks: 50

1. (a) **Solution:**

Unseen problem, but similar to seen lab problem

1/2 mark for code correct; 1/2 mark for comment correct

```
# [+1] for converting prior to log (or log-lik to lik)
# [+1] for using arange or similar to compute prior
# compute log-prior array
log_prior_theta = np.log(prior(np.arange(112)))
# [+1] for calling llik
# compute log-likelihood
log_lik = llik(scan_data)
# [+1] for Bayes Rule (top half)
posterior_un = np.exp(log_prior_theta + log_lik)
# [+1] for integrating over evidence
# [+1] for correct use of axis
posterior = posterior_un / np.sum(posterior_un, axis=1)
# [+1] for using argmax
# [+1] for correct use of axis
# compute maximum
max_indices = np.argmax(posterior, axis=0)
```

NOTE: normalisation is not needed! give 2 marks if Bayes' Rule used correctly but no integration over evidence!

any formulation, including writing functions, writing it all in one line, or similar is acceptable. **No marks** for elements that use explicit iteration (`for` loops).

(b) **Solution:**

For each sub-point, up to two marks for any of the listed points.

Why is not Bayesian? Any of (for [+1] for any one)

- does not quantify uncertainty/single point estimate
- does not incorporate a prior
- does not treat parameters as random variables or any similar wording

What would be required?

- A prior distribution over the parameters
- Some form of inference approach (e.g. Monte Carlo sampling)

What advantages and disadvantages?

- Would give full uncertainty on parameters
- May be harder or slower to fit
- Specifying a good prior might require additional research or expert input

(c) Solution:
Unseen problem

- **[2]** for a sensible choice
- **[2]** for a reasonable justification

I would expect an answer *not* to use raw probabilities or log probabilities.

- Odds can be argued to be interpretable and work well for both very likely and very unlikely values. However, very large odds ratios may not be informative.
- Logits arguably have better scaling for extreme events, but could be hard for humans to interpret.

(d) (i) **Solution:**
Unseen problem
These are logarithmic [+1] polar [+1] coordinates. Polar coordinates correctly represent the wraparound at phi=0 to phi=111 [+1] (or any similar wording). The logarithmic scale allows values of ranging over many magnitudes to be displayed. [+1]

- Log-probabilities on a linear coord system would look the same as this representation; using a log coord would not be useful **[1]**
- Logits would require using *signed* values **[1]** which are hard to represent on a polar plot without an obvious "zero x axis"
- The same applies to odds, which would probably require two "sides" broken at 1:1 and would scale **[1]**
- The plot shows raw probabilities with a log scale, giving the advantages of log-probability representation without a change of units; a linear scale would make it very hard to distinguish smaller probabilities **[1]**

2. **(a)** **Solution:**

Unseen problem

- This can be solved with power iteration **[2]**
- Illuminate a random point $\vec{x_0}$ on the projector **[1]** for random init. **[1]** for illuminating a point
- Compute the inverse of B, B^{-1} **[1]**

  This is possible because we know B is non-singular [1]

1. Measure $\vec{z_0}$ on the camera **[1]**
2. Compute the corresponding point $\vec{y_0}$ via $\vec{y_0} = B^{-1}\vec{z_0}$ **[2]**
3. Normalise $\vec{y_0}$ via $\vec{y'_0} = \frac{\vec{y_0}}{\|\vec{y_0}\|_\infty}$ **[1]** (any norm can be used, not just infinity).
4. illuminate a point $x_1 = y'_0$ (or $\vec{x_1} = \lambda \vec{y'_0}$ for some arbitrary fixed λ) **[1]**

- Repeat steps 1-4 until $\vec{x_n}$ close to $\vec{x_{n-1}}$ **[1]** using a vector norm to compute similarity **[1]**
- Take $\vec{x_n}$ and compute $\vec{y_n}$ via $B^{-1}\vec{z_n}$ **[1]**
- Compute $\lambda = \vec{y_n}/\vec{x_n}$ to get the eigenvalue (assumes the value is real, but we can ignore this) **[1]**

(b) (i) **Solution:**

- A linear model can be fitted very quickly using standard algorithms which will converge
- Powerful analysis tools like eigendecomposition can be applied to the model
- Linear transforms of coordinates cannot represent distortions like perspective or curved changes
- Fitting a nonlinear model may require expensive optimisation which may or may not converge

(ii) **Solution:**

Unseen problem

This problem *is* a sum of simpler problems [+1] and stochastic gradient descent could be applied [+1]. This would allow for better resource management (e.g. smaller memory requirements) [+1] as well as potentially providing some form of stochastic relaxation [+1]. This would require some way of estimating derivatives efficiently [+1], for example automatic differentiation [+1]

3. (a) Solution:

Unseen problem

At 600Hz, frequencies up to the Nyquist rate [+1] of 300Hz could be captured without aliasing [+1]. Microphone A is suitable [+1] and microphone B is not cost-effective in this application, similarly with Microphone C, which would also require very fast sampling hardware and a lot of storage [+1]. If too low of a sampling rate were chosen (like Microphone D), aliasing [+1] would occur; high frequencies would appear as phantom low-frequency components in the measurement [+1]

(b) Solution:

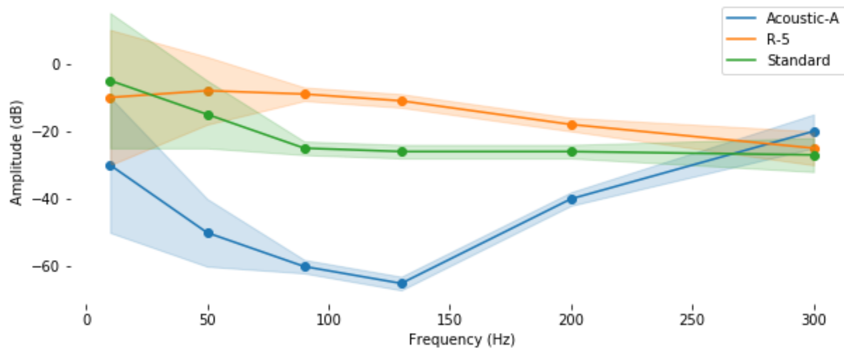
unseen problem

The Dirac delta function is the identity with respect to convolution [+1]. An approximate delta impulse could be produced by tapping/striking/etc. near the glass [+1] and the corresponding signal recorded [+1]. This would be the convolution kernel desired. This might be difficult both because of background noise [+1] and because generating a perfect delta function is not feasible by physical means [+1]. Also, the glass response might be nonlinear [+1] and thus not representable as a convolution kernel [+1].

(c) (i) **Solution:**

For each point: [+1] for correct term; [+0.5] for problem; [+0.5] for fix. Up to 6 separate points may be identified for marking.

- The **guides** (ticks) are both ugly and poorly spaced; they should be spaced evenly and span the data range
- The plot is **faceted** but would better be **layered**, or at least faceted with consistent **coords**
- The **mapping** is the wrong way around: **freq** should be on the x-axis and the amplitude on the y-axis
- The **coords** are very badly scaled and inconsistent. A single set of coords covering the data should be chosen.
- There are no **guides** to label the axes. Labels *with units* should be present. (only 0.5 if units not mentioned!)
- There are no **geoms** to represent uncertainty, which is present in the table. Error bars or ribbon plots could show this.
- There are no **geoms** to represent measurement points. Point geoms could provide this.
- There is no **guide** to identify the lines/facets. A legend or titles could provide this. A good plot would look like:



CONTINUED OVERLEAF

Mark table

Question	Points	Score
1	25	
2	25	
3	25	
Total:	75	