Version SJG/3

EGT2

ENGINEERING TRIPOS PART IIA

April 2018

Module 3F3

STATISTICAL SIGNAL PROCESSING

WORKED SOLUTIONS

Assessors' comments:

Q1: Markov chains

Part a answered well except the proof of the Markov property. Part b surprisingly not. Some candidates could not compute the characteristic function of Yn, which is the sum of independent n random variables variables. Part c(i). Surprisingly some candidates failed to use the Gaussian approximation from part b to calculate the probability of wealth being positive after n bets. Part c(ii). Most understood that the approximation from c(i) overestimates the probability since it allows the gambler to go into debt.

Q2: AR models/ power spectrum

Parts a and b. The power spectral density is most straightforwardly calculated via the z transform of the impulse response. Taking the DTFT of the autocorelation function directly is more lengthy, especially for the AR(2) process. Parts c and d very well done. Part e was a challenge for many. Part f was recognized as straightforward and the power spectrum of X could be calculated using the parts a+e or by calculating the power spectrum of Y and then using parts c+d.

Q3: Matched filtering

This question was quite well answered, although many candidates were unable to provide full detail in their solutions. The precise definition of white noise was not well known, but reasonable attempts were not penalised heavily. The sketch of filter output in response to just signal was poorly done by many âĂŞ showing that not many students have a good insight into how a FIR filter works. The matched filter was well known, but many made simple errors in calculating the SNRs âĂŞ e.g. not remembering to square the maximum output value.

Q4: Maximum likelihood/ Bayes

This question was answered pleasingly well. Many spotted the non-standard form of the likelihood estimates in the first part and were able to comment successfully on the likely bias of the solutions. ML and Bayesian estimates were very well answered in part (b), although surprisingly few people got the normalising constant for the posterior, even given the appropriate gamma integral in the hint. Good sketching of the densities in the final part.

STATIONERY REQUIREMENTS

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM

10 minutes reading time is allowed for this paper.

You may not start to read the questions printed on the subsequent pages of this question paper until instructed to do so.

1 (a) Let $f(i)$ be a probability mass function (pmf), i.e. $f(i) \geq 0$ and $\sum_{-\infty}^{\infty} f(i) = 1$. Let

$$X_k = X_{k-1} + W_k, \quad \text{for } k = 1, 2, \ldots$$

where $X_0 = i_0$ and $W_1, W_2, \ldots$ are independent and identically distributed random variables and each W_k has pmf f.

(i) Find $p\left(X_{k+1} = j \mid X_k = i\right)$; [10%]

(ii) Find $p(X_1 = i_1, \ldots, X_n = i_n)$; [10%]

(iii) Hence show that $X_1, X_2, \ldots$ is a Markov chain. [20%]

(b) Let $f(-1) = 1/2$, $f(1) = 1/2$ (thus $f(i) = 0$ for all other values of i) and let

$$Y_n = \left(\sum_{j=1}^{n} W_j \right) / \sqrt{n}$$

(i) By computing the characteristic function of Y_n, which is $E\left\{\exp\left(iY_n t\right)\right\}$, show that Y_n tends to a Gaussian random variable as $n \to \infty$. (Hint: you may use the fact that $\cos(t/\sqrt{n})^n \to \exp\left(-t^2/2\right)$ as n tends to infinity.) [25%]

(c) A gambler, with initial wealth R, wagers one pound for each bet and the probability of winning the bet is 0.5. Their wealth increases by 1 if the bet is won; otherwise it decreases by 1.

(i) The gambler is allowed to make n successive bets, potentially going into debt. Find an approximation for $\Pr(X_n > 0)$. [25%]

(ii) In a change of the rules, the gambler is allowed to make n successive bets but must stop as soon as their wealth is zero. Give the Markov chain that describes the change in wealth of the gambler and comment on how well the answer in (c)(i) approximates the probability that the gambler's wealth is positive. [10%]

SOLUTION:

$$\begin{aligned} p(X_{k+1} = j \mid X_k = i) &= p(X_k + W_{k+1} = j \mid X_k = i) \\ &= p(W_{k+1} = j - i) \\ &= f(j - i). \end{aligned}$$

$$
\begin{aligned}
&p(X_1 = i_1, \ldots, X_n = i_n) \\
&= p(W_1 = i_1 - i_0, \ldots, W_n = i_n - i_{n-1}) \\
&= p(W_1 = i_1 - i_0) \cdots p(W_n = i_n - i_{n-1}) \\
&= f(i_1 - i_0) \cdots f(i_n - i_{n-1})
\end{aligned}
$$

Proof of Markov property by showing

$$p(X_n = i_n \mid X_1 = i_1, \ldots, X_{n-1} = i_{n-1}) = p(X_n = i_n \mid X_{n-1} = i_{n-1}).$$

To do so, use previous two derived results as follows:

$$
\begin{aligned}
&p(X_n = i_n \mid X_1 = i_1, \ldots, X_{n-1} = i_{n-1}) \\
&= \frac{p(X_1 = i_1, \ldots, X_n = i_n)}{p(X_1 = i_1, \ldots, X_{n-1} = i_{n-1})} \\
&= f(i_n - i_{n-1}) \\
&= p(X_n = i_n \mid X_{n-1} = i_{n-1})
\end{aligned}
$$

When $f(-1) = 0.5$ and $f(1) = 0.5$ then given $X_k = i_k$,

$$X_{k+1} = \begin{cases} i_k - 1 & \text{w.p } 0.5 \\ i_k + 1 & \text{w.p } 0.5 \end{cases}$$

where w.p. abbreviates with probability.

$$
\begin{aligned}
E\left(\exp(iY_n t)\right) &= E\left(\exp(iW_1 t/\sqrt{n}) \cdots \exp(iW_n t/\sqrt{n})\right) \\
&= E\left(\exp(iW_1 t/\sqrt{n})\right) \cdots E\left(\exp(iW_n t/\sqrt{n})\right)
\end{aligned}
$$

where the second line follows since W_k are independent.

$$
\begin{aligned}
E\left(\exp(iW_1 t/\sqrt{n})\right) &= \frac{1}{2}\exp(it/\sqrt{n}) + \frac{1}{2}\exp(-it/\sqrt{n}) \\
&= \cos(t/\sqrt{n}).
\end{aligned}
$$

Thus

$$E\left(\exp(iY_n t)\right) = \cos(t/\sqrt{n})^n \to \exp(-\frac{t^2}{2})$$

(TURN OVER

which is the characteristic function of a zero mean unit variance Gaussian random variable.

Note that $X_n = R + \sqrt{n} Y_n$.Thus

$$\begin{aligned}\Pr\left(R + \sqrt{n} Y_n > 0\right) &= \Pr\left(Y_n > -R/\sqrt{n}\right) \\ &\approx \int_{-R/\sqrt{n}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-0.5x^2) dx\end{aligned}$$

The new Markov chain $X_0', X_1', \ldots, X_n$' that describes the change in wealth when gambler must stop betting once their wealth is zero is: given $X_k' = i_k$,

$$X_{k+1}' = \begin{cases} i_k & \text{if } i_k = 0. \\ i_k + 1 & \text{w.p. } 0.5 \\ i_k - 1 & \text{w.p. } 0.5 \end{cases}$$

The event $\left\{X_n' > 0\right\}$ is a strict subset of the event $\{X_n > 0\}$ since X_n permits the wealth to dip below zero. So the solution to the previous part over estimates the probability.

2 Consider the following autoregressive process

$$X_n + a_1 X_{n-1} + a_2 X_{n-2} = \sigma W_n$$

where $\{W_n\}$ is a zero-mean white noise process with variance 1 and σ a positive constant.

(a) Find the power spectrum of $\{X_n\}$. [10%]

(b) Let $\{V_n\}$ be the moving average process

$$V_n = b_0 E_n + b_1 E_{n-1}$$

where $\{E_n\}$ is a zero-mean white noise process with variance 1. Find the power spectrum of $\{V_n\}$. [10%]

(c) Let Y_n be the noisy measurement of X_n given by

$$Y_n = X_n + V_n.$$

Assume the noise sequences $\{W_n\}$ and $\{E_n\}$ are independent. Find the power spectrum of $\{Y_n\}$. [15%]

(d) Based on measurements of $\{V_n\}$, the power spectrum of $\{V_n\}$ is estimated to be

$$\widehat{S}_V(\omega) = 2 + 2\cos\omega$$

Show that valid estimates of b_0 and b_1 are $b_0 = b_1 = 1$ and $b_0 = b_1 = -1$. [25%]

(e) Show that

$$\begin{bmatrix} R_{XX}[0] & R_{XX}[1] \\ R_{XX}[1] & R_{XX}[0] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = -\begin{bmatrix} R_{XX}[1] \\ R_{XX}[2] \end{bmatrix},$$

$$R_{XX}[0] + a_1 R_{XX}[1] + a_2 R_{XX}[2] = \sigma^2.$$

[30%]

(f) Based on measurements of Y_n as in (c), the following estimates are made for its autocorrelation function:

$$\widehat{R}_{YY}[0] = 4.74, \qquad \widehat{R}_{YY}[1] = 0.54, \qquad \widehat{R}_{YY}[2] = 1.41$$

Use these values to estimate the power spectrum of $\{X_n\}$. [10%]

SOLUTION:

The power spectrum of X_n:

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2}, \qquad S_X(e^{j\omega}) = \sigma^2 |A(e^{j\omega})|^{-2}$$

The power spectrum of V_n:

$$B(z) = b_0 + b_1 z^{-1}, \qquad S_V(e^{j\omega}) = |B(e^{j\omega})|^2$$

The power spectrum of Y_n:

$$\begin{aligned} E\{y_n y_{n+k}\} &= E\{(x_n + v_n)(x_{n+k} + v_{n+k})\} \\ &= E\{x_n x_{n+k}\} + E\{v_n v_{n+k}\} + \text{crossterms} \end{aligned}$$

Note that the cross terms have zero expectation. So

$$R_{YY}[k] = R_{XX}[k] + R_{VV}[k]$$

and

$$S_Y(e^{j\omega}) = S_X(e^{j\omega}) + S_V(e^{j\omega})$$

$$\begin{aligned} S_V(e^{j\omega}) &= |b_0 + b_1(\cos\omega - j\sin\omega)|^2 \\ &= b_0^2 + b_1^2\cos^2\omega + 2b_1 b_0\cos\omega + b_1^2\sin^2\omega \\ &= b_0^2 + b_1^2 + 2b_1 b_0\cos\omega \end{aligned}$$

Just verify the stated values of b_0 and b_1 solve this equation.

Multiply X_n with $X_n + a_1 X_{n-1} + a_2 X_{n-2} = \sigma W_n$ and take the expectation to get

$$R_{XX}[0] + a_1 R_{XX}[1] + a_2 R_{XX}[2] = \sigma^2.$$

Multiply X_{n-1} with $X_n + a_1 X_{n-1} + a_2 X_{n-2} = \sigma W_n$ and take the expectation to get

$$R_{XX}[1] + a_1 R_{XX}[0] + a_2 R_{XX}[1] = 0.$$

Multiply X_{n-2} with $X_n + a_1 X_{n-1} + a_2 X_{n-2} = \sigma W_n$ and take the expectation to get

$$R_{XX}[2] + a_1 R_{XX}[1] + a_2 R_{XX}[0] = 0.$$

Use $b_0 = b_1 = 1$, which implies $v_n = e_n + e_{n-1}$. So

$$R_{VV}[0] = 2,\ R_{VV}[1] = 1,\ R_{VV}[2] = 0, \ldots$$

We can estimate $\widehat{R}_{XX}$ using the given $\widehat{R}_{YY}$ and calculated $\widehat{R}_{VV}$

$$\widehat{R}_{XX}[0] = 2.74,\ \widehat{R}_{XX}[1] = -0.46,\ \widehat{R}_{XX}[2] = 1.41,$$

Now use the derived (Yule-Walker) equations

$$\begin{aligned}
\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} &= - \begin{bmatrix} 2.74 & -0.46 \\ -0.46 & 2.74 \end{bmatrix}^{-1} \begin{bmatrix} -0.46 \\ 1.41 \end{bmatrix} \approx \begin{bmatrix} 1/12 \\ -1/2 \end{bmatrix}, \\
\sigma^2 &= 2.74 - \frac{0.46}{12} - \frac{1.41}{2} \approx 2
\end{aligned}$$

3 (a) Define the term white noise. What is the form of the autocorrelation function for white noise, and what is its power spectrum? [15%]

SOLUTION:

White noise has autocovarince function as follows:

$$E(W_n - \mu)(W_{n+m} - \mu) = \sigma_W^2 \delta[m]$$

where $\mu = EW_n$ (wide-sense ststionary).

Autocorrelation function has the form:

$$r_{XX}[m] = \sigma_W^2 \delta[m] + \mu^2$$

[note the possible constant term when non-zero mean].

Power spectrum:

$$\sigma_W^2 + \mu^2 \delta[0]$$

(b) A pulse waveform $s_n = n/N$, $n = 0, ...N$ is buried in noise at a sample time n_0, i.e. the noisy signal is:

$$x_n = \begin{cases} s_{n-n_0} + v_n, & n - n_0 = 0, 1, ..., N \\ v_n, & \text{otherwise,} \end{cases}$$

where v_n is white, zero-mean noise with variance σ_v^2.

A FIR smoothing filter with N coefficients [1, 1, ..., 1] is applied to the noisy waveform.

(i) Show that the output y_n of the filter has variance $N\sigma_v^2$ when only noise v_n is input to the filter. [15%]

SOLUTION:

$$E[y_n^2] = E[(\sum_{i=0}^{N-1} (v_{n-i}.1))^2] = N\sigma_v^2$$

(ii) Determine and sketch the output of the filter when $\sigma_v = 0$, i.e. the noise is not present and just the pulse s_{n-n_0} is filtered.

SOLUTION: Output is

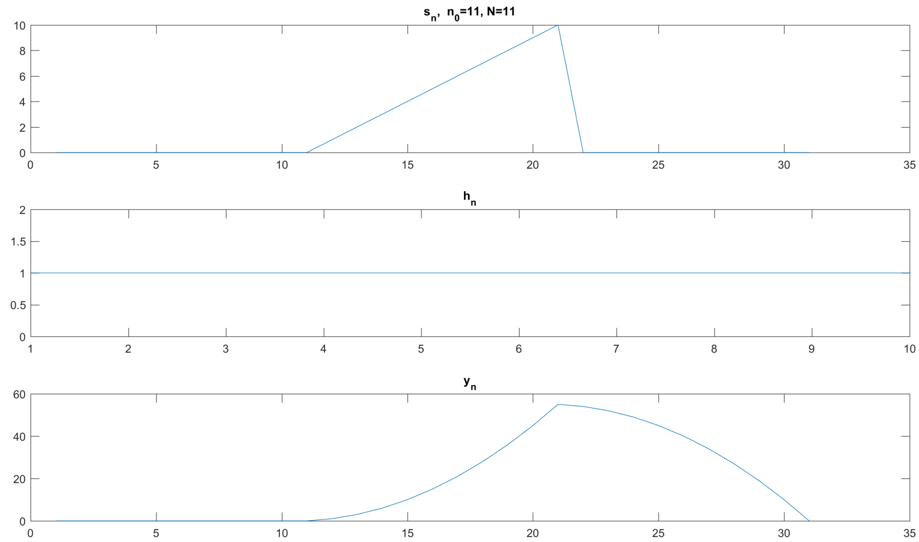
$$y_n = \sum_{i=1}^{N} h_i s_{n-i}$$

So output is 0 up to $n = n_0$.

Then, for $n = n_0, ..., n_0 + N$:

$$y_n = \sum_{i=0}^{n-n_0} h_i s_{n-n_0-i} = 1/N \sum_{i=0}^{n-n_0} i = 1/(2N)(n - n_0 + 1)(n - n_0)$$

and similarly for $n = n_0 + N + 1 : n_0 + 2N$, see plot:



[20%]

(iii) What is the maximum expected signal-to-noise ratio at the output of the filter, when applied to noisy pulse data x_n? (i.e. now $\sigma_v > 0$.), and at what value of n does this occur? [20%]

SOLUTION: The maximum signal output from the last part is at $n = n_0 + N$, at which point we have signal value $(N+1)/2$. Hence, the maximum SNR is:

$$(N+1)^2/(4(N\sigma_v^2))$$

(iv) Now design the optimal filter for detection of the location n_0 (no derivation is required) and compare its performance (in terms of SNR) with that of the FIR smoothing filter as N becomes large. [30%]

SOLUTION: The optimal choice is the Matched filter, so we choose the time-reversed signal pulse for the FIR coefficients:

$$h_p = (N-p)/N,\ \ p = 0, \dots, N-1$$

The maximum SNR for the matched filter is just (from the lectures):

$$1/\sigma_v^2 \sum_{p=0}^{N} s_p^2 = 1/\sigma_v^2 \sum_{p=1}^{N} p^2/N^2 = (N+1)(2N+1)/(6N\sigma_v^2)$$

using the given summation formula.

We can thus see that for large N the SNR tends to $N/(3\sigma_v^2)$, which is a modest improvement compared to $N/(4\sigma_v^2)$ from the FIR smoothing filter.

4 (a) A symmetric uniform probability distribution is defined as

$$f_Y(y|a) = \begin{cases} 1/(2a), & -a \le y \le a \\ 0, & \text{Otherwise} \end{cases}$$

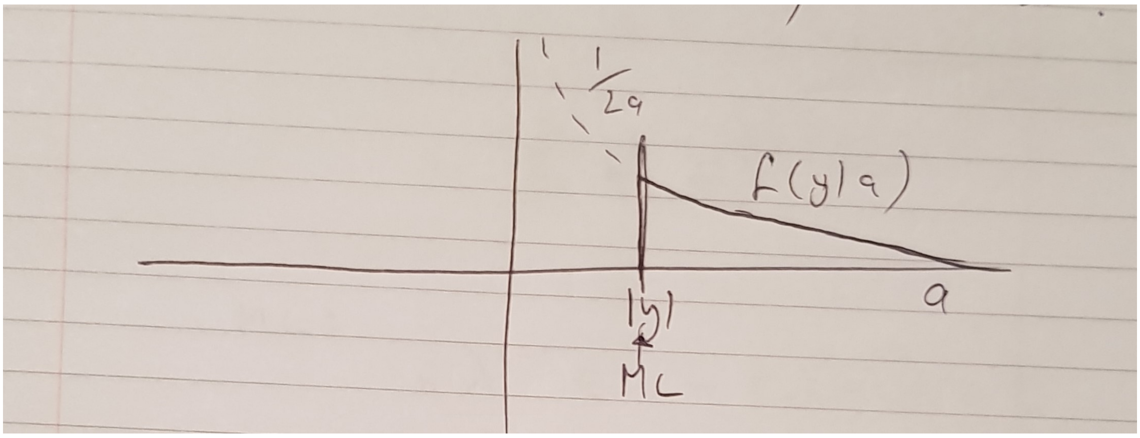
A sequence of discrete-time measurements $y_1, y_2, y_3, \ldots, y_n$ is made at the output of a, i.i.d. symmetric uniform noise source, but the scaling of the noise, a, is unknown.

(i) Determine the likelihood function for a when $n = 1$ and $n = 2$ and sketch it as a function of a, marking on the maximum likelihood estimator in each case. [15%]

SOLUTION:

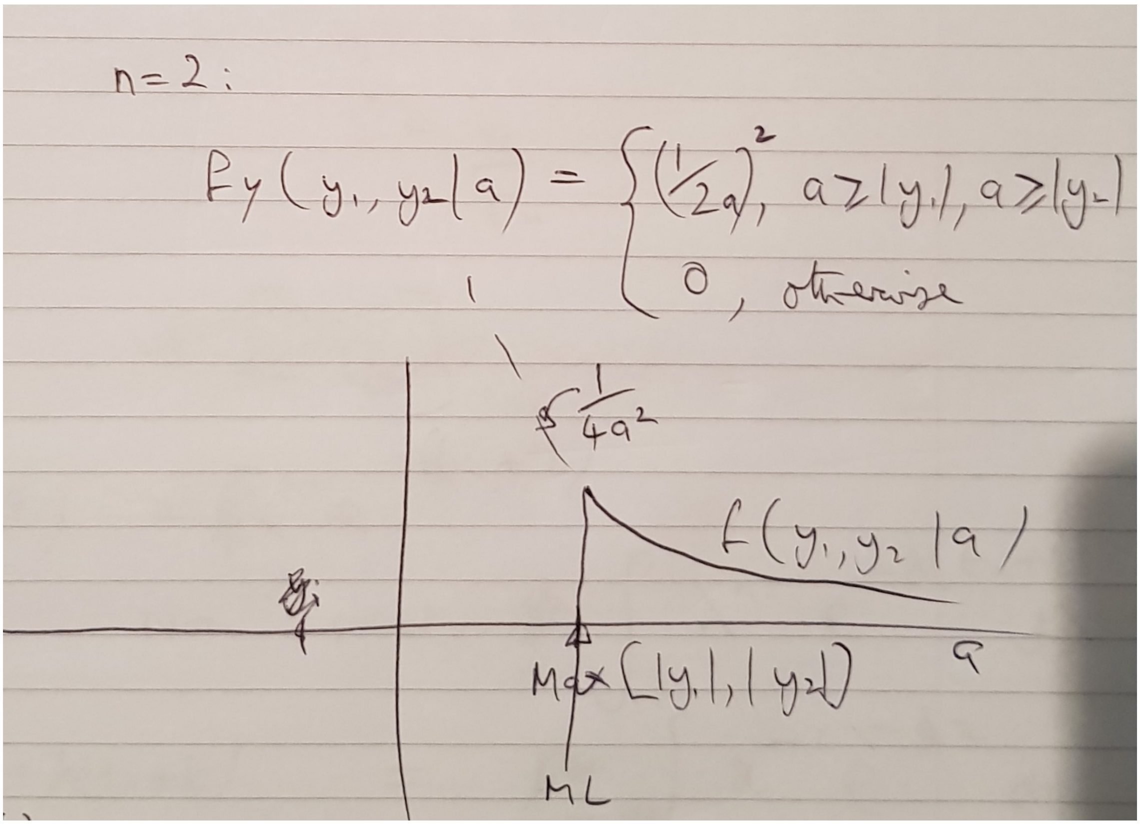
$n = 1$, just rearrange the limits as follows:

$$f_Y(y|a) = \begin{cases} 1/(2a), & -a \le a \ge |y| \\ 0, & \text{Otherwise} \end{cases}$$



$n = 2$,

$$f_Y(y_1, y_2|a) = \begin{cases} 1/(2a)^2, & -a \le a \ge |y_1|, a \ge |y_2| \\ 0, & \text{Otherwise} \end{cases}$$



(ii) Determine the maximum likelihood estimator for an arbitrary number of measurements n. Is this estimate likely to be unbiased for finite n? What do you think would happen as $n \to \infty$? [20%]

SOLUTION:

$$a^{ML} = \max_i\{|y_i|\}$$

This will not be unbiased since a^{ML} is always less than a. As $n \to \infty$ though we can expect to see the largest y_i approaching a, hence we mig1ht expect a^{ML} to be a consistent estimator.

(b) A communications network is monitored. It is desired to find the average rate of symbols, λ. Prior information about the network traffic states that λ is distributed in the following way:

$$f(\lambda) = \begin{cases} \lambda b^2 \exp(-\lambda b), & \lambda > 0 \\ 0, & \text{otherwise.} \end{cases}$$

The time τ between each symbol is independently and randomly distributed as an exponential random variable with mean $1/\lambda$:

$$f(\tau|\lambda) = \lambda \exp(-\lambda \tau)$$

(i) The times of arrival of n successive symbols are now measured as $t_1, t_2, t_3, ..., t_n$, where t_0, the first symbol's arrival time, is zero. Show that the likelihood function for λ is:

$$\lambda^n \exp(-\lambda t_n)$$

and find the ML estimate of λ. [20%]

SOLUTION:

$$p(t_1, ..., t_n|\lambda) = \prod_{i=1}^{n} \lambda \exp(-\lambda(t_i - t_{i-1})) = \lambda^n \exp(-\lambda t_n)$$

ML estimate:
Differentiate the likelihood and equate to 0:

$$n\lambda^{n-1} \exp(...) - t_n \lambda^n \exp(...) = 0$$

So,

$$n = t_n \lambda,\ \lambda = n/t_n$$

(ii) Determine the Bayesian posterior density for λ (including its normalising constant). [25%]

SOLUTION:
Multiply the prior by the likelihood:

$$\lambda^n \exp(-\lambda t_n)\lambda b^2 \exp(-\lambda b) = \lambda^{n+1} b^2 \exp(-\lambda(t_n + b))$$

Now compute normalising constant:

$$\begin{aligned}\int \lambda^{n+1} b^2 \exp(-\lambda(t_n + b)) d\lambda &= \frac{b^2}{(t_n + b)^{n+2}} \int x^{n+1} \exp(-x) dx \\ &= \frac{b^2}{(t_n + b)^{n+2}} (n+1)!\end{aligned}$$

So, combining expression with normalising constant:

$$p(\lambda|t_1, ..., t_n) = \frac{(t_n + b)^{n+2}}{(n+1)!} \lambda^{n+1} \exp(-\lambda(t_n + b))$$

(iii) Sketch the prior density, likelihood function and posterior density, marking the MAP and ML estimators clearly on the sketch and commenting on their relationship to the prior. Use the following values $b = 1$, $n = 4$ and $t_n = 5$. [20%]

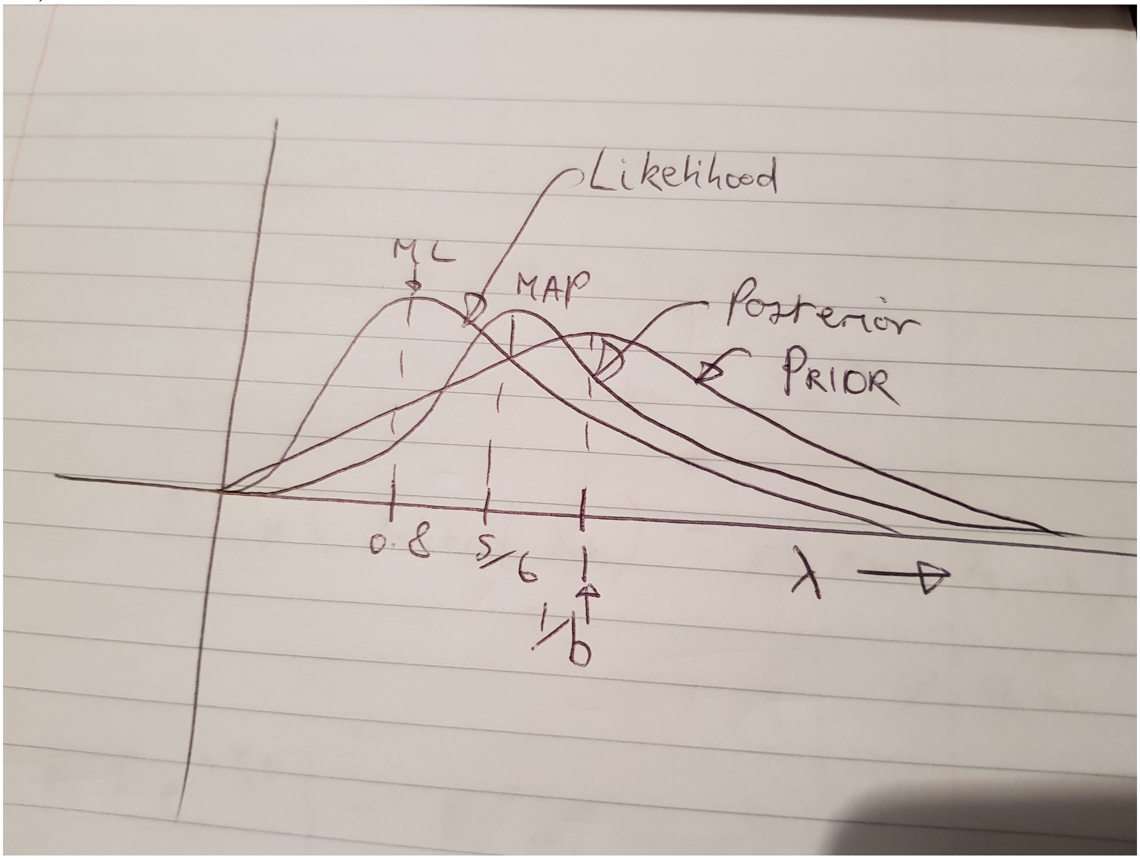
SOLUTION:
ML estimator:

$$\lambda = n/t_n = 0.8$$

MAP estimator:
Differentiate posterior and set to zero as before:

$$\lambda = \frac{n+1}{t_n + b} = 5/6$$

Similarly, the prior has a maximum at $\lambda = 1/b = 1$.
So, sketch is as follows:



We observe that the MAP estimator (5/6) 'pulls' the ML estimator (4/5) towards the maximum of the prior (1). This is the classic Bayesian regularisation effect...

[Note that, for any integer n:

$$\int_0^\infty x^n \exp(-x) dx = n!$$

]

END OF PAPER

THIS PAGE IS BLANK