EGT2
ENGINEERING TRIPOS PART IIA

Module 3F3 - Worked solutions

STATISTICAL SIGNAL PROCESSING

Answer not more than ***three*** *questions.*

All questions carry the same number of marks.

The ***approximate*** *percentage of marks allocated to each part of a question is indicated in the right margin.*

Write your candidate number ***not*** *your name on the cover sheet.*

STATIONERY REQUIREMENTS
Single-sided script paper

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
CUED approved calculator allowed
Engineering Data Book

10 minutes reading time is allowed for this paper at the start of the exam.

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

You may not remove any stationery from the Examination Room.

1 Consider the following Markov sequence $X_0, X_1, \ldots$ which describes the population count of prey, where $0 \leq X_k \leq M$,

$$P\left(X_{k+1} = j | X_k = i\right) = \begin{cases} \theta_1 i & \text{if } j = i + 1 \text{ and } i < M \\ \theta_2 i & \text{if } j = i - 1 \text{ and } i > 0 \\ ? & \text{if } j = i \\ 0, & \text{otherwise} \end{cases}$$

and M denotes the maximum population count. (The value in the $j = i$ row is intentionally left as '?'.)

(a) Find the transition probability matrix for this Markov chain. [20%]

(b) Find the range of values for θ_1 and θ_2 that ensures the transition probability matrix is valid. [10%]

(c) For $X_0 = M$, find the probability mass functions (pmfs) $P(X_1 | X_0 = M)$ and $P(X_2 | X_0 = M)$. [20%]

(d) For $X_0 = M$, find the values of θ_2 and θ_1 that maximise the probability of observing the following population count: $X_1 = M - 1, X_2 = M - 2, \ldots, X_{M-1} = 1$. [20%]

(e) Find $P(X_k > X_{k-1})$ in terms of the marginals $P(X_k = i)$. Hence, in a data set $X_1, X_2, \ldots, X_T$ of length T, find the expected number of transitions where the population increases. [30%]

Part a

The transition probability matrix is

$$P = \begin{bmatrix} 1 & 0 & \cdots & & & & 0 \\ \theta_2 & 1 - \theta_1 - \theta_2 & \theta_1 & 0 & & \cdots & 0 \\ 0 & 2\theta_2 & 1 - 2\theta_1 - 2\theta_2 & 2\theta_1 & 0 & \cdots & 0 \\ & & & & & & \\ 0 & \cdots & & 0 & (M-1)\theta_2 & 1 - (M-1)(\theta_1 + \theta_2) & (M-1)\theta_1 \\ 0 & \cdots & & & 0 & M\theta_2 & 1 - M\theta_2 \end{bmatrix}$$

Part b

Last row: $M\theta_2 \leq 1$

Penultimate row: $(M-1)(\theta_1+\theta_2) \leq 1$

Second row: $\theta_2 \geq 0, \theta_1 \geq 0$

Part c

$P(X_1 = M) = P(X_1 = M|X_0 = M)P(X_0 = M) = 1 - M\theta_2$

$P(X_1 = M-1) = M\theta_2$

$P(X_2 = M) = (1 - M\theta_2)^2 + M\theta_2(M-1)\theta_1$

$P(X_2 = M-1) = (1 - M\theta_2)\, M\theta_2 + M\theta_2\,(1 - (M-1)(\theta_1+\theta_2))$

$P(X_2 = M-2) = M\theta_2(M-1)\theta_2$

Part d

The probability of the data set is

$$M\theta_2 \times (M-1)\theta_2 \times \cdots \times 2\theta_2 = \theta_2^{M-1} \times M!$$

The log of the probability is $(M-1)\log\theta_2 + \log(M!)$. This is maximised by the largest possible θ_2 value which is $\theta_2 = 1/M$. Also, no extra restriction on θ_1 beyond part (b).

Part e

$$P(X_k > X_{k-1}) = \sum_{i=1}^{M-1} P(X_{k-1} = i) \times \theta_1 \times i$$

The expected number of transitions where the population increases is

$$\mathbf{E}\left\{\mathbb{I}[X_1 > X_0] + \cdots + \mathbb{I}[X_T > X_{T-1}]\right\} = \sum_{k=1}^{T} P(X_k > X_{k-1}).$$

2 Let $\{X_n\}$ be a wide sense stationary random process generated by the AR(2) model

$$X_n = 0.1X_{n-1} + 0.7X_{n-2} + W_n$$

where $\{W_n\}$ is an independent sequence with zero mean and variance 1. Let $R_X(k) = \mathbf{E}(X_n X_{n+k})$, which is the autocorrelation function of $\{X_n\}$.

(a) Find $R_X(k)$ in terms of $R_X(k-1)$ and $R_X(k-2)$ for $k \geq 1$. [30%]

(b) Find $R_X(k)/R_X(0)$ for $k = 1, 2, 3$ and hence sketch the autocorrelation function. [30%]

(c) A modeller suggests that the model

$$Z_n = \alpha Z_{n-2} + W_n$$

for $|\alpha| < 1$ can approximate $R_X(k)$ well.

(i) Find $R_Z(k)/R_Z(0)$ and comment on the appropriateness of this model for data $X_1, X_2, ...$ [30%]

(ii) For an appropriately chosen α, justify why the model $Z_n = \alpha Z_{n-1} + W_n$ may be better suited to explain this data. [10%]

Part a

Multiply by X_{n-1}:

$$\begin{aligned} X_n X_{n-1} &= 0.1X_{n-1}X_{n-1} + 0.7X_{n-2}X_{n-1} + W_n X_{n-1} \\ R_X(1) &= 0.1R_X(0) + 0.7R_X(1) \\ R_X(1) &= \frac{1}{3}R_X(0) \end{aligned}$$

Multiply by X_{n-2} and find the expectation:

$$R_X(2) = 0.1R_X(1) + 0.7R_X(0)$$

Multiply by X_{n-k} and find the expectation:

$$R_X(k) = 0.1R_X(k-1) + 0.7R_X(k-2)$$

Part b

$$
\begin{aligned}
R_X(1)/R_X(0) &= 1/3 \\
R_X(2)/R_X(0) &= 0.1 \times \frac{1}{3} + 0.7 = 22/30 = 0.73 > 1/3 \\
R_X(3)/R_X(0) &= 0.1R_X(2)/R_X(0) + 0.7R_X(1)/R_X(0) \\
&= 0.3
\end{aligned}
$$

Plot these or just declare.

Part c.i

Multiply by Z_{n-2} and find the expectation:

$$R_Z(2) = \alpha R_Z(0)$$

Multiply by Z_{n-4} and find the expectation:

$$R_Z(4) = \alpha R_Z(2) = \alpha^2 R_Z(0)$$

This reveals even lag trend.

Multiply by Z_{n-1} and find the expectation:

$$R_Z(1) = \alpha R_Z(1) = 0$$

for $|\alpha| < 1$.

Multiply by Z_{n-3} and find the expectation:

$$R_Z(3) = \alpha R_Z(1) = 0$$

This reveals odd lag trend, which is 0. Not suitable approximation $R_X(k)$ since odd lags are not zero.

Part c.ii

Odd lags will not be zero. This implies Z_n will be correlated with the random variables $Z_{n-1}, Z_{n-3}, \ldots$ as well, just like the data X_n is correlated with its entire past.

3 Examiner's comments:

A less popular question. Part a) many candidates did not know the definition of joint pdf. In particular many stated that is was the probability p(A=a,B=b) which would generally be zero of course. Part b) well answered in general, Jacobians are well known and applied. Part c) should then have been straightforward, but there was quite a lot of confusion and poor graph sketching. Part d) well answered by many.

It is desired to fit a single sinusoid to noisy oscillating data measured from a distant galaxy. The model is posed as

$$x_n = A\sin(\omega n) + B\cos(\omega n) + \epsilon_n$$

where A and B are unknown coefficients, ω is a *known* frequency and ϵ_n is a random zero-mean noise process. n is an integer time index.

(a) Define the term *joint probability density function* $f_{A,B}(a,b)$ for the random variables (A,B) evaluated at $A=a$ and $B=b$, in terms of an event $(a_l < A \le a_u,\ b_l < B \le b_u)$. [10%]

Solution:

$$Pr(Event) = \int_{a_l}^{a_u}\int_{b_l}^{b_u} f_{A,B}(a,b)\,da\,db$$

Equivalent answers in terms of partial derivatives of joint CDF also acceptable

(b) Suppose the joint probability density function distribution of (A,B) is *circularly symmetric*, which means that it may be expressed as

$$f_{A,B}(a,b) = f(a^2+b^2)$$

where $f()$ is a scalar function of a^2+b^2. Show that the cumulative distribution function (CDF) for $R=\sqrt{A^2+B^2}$ is given by

$$F_R(r) = 2\pi\int_0^r \rho f(\rho^2)d\rho$$

and hence show that the probability density function for R is

$$f_R(r) = 2\pi r f(r^2)$$

What is the distribution of the phase, $tan^{-1}(B/A)$? [30%]

Solution:

Integrate pdf over the interior of circle radius r:

$$F_R(r) = Pr(A^2+B^2 \le r^2) = \int_{a^2+b^2\le r} f(a^2+b^2)\,da\,db$$
$$= \int_0^{2\pi}\int_0^r f(\rho^2)\rho\, d\rho\, d\theta = 2\pi\int_0^r f(\rho^2)\rho\, d\rho$$

To get pdf, take derivative wrt r:

$$d/dr(2\pi \int_0^r f(\rho^2) r d\rho) = 2\pi r f(\rho^2)$$

Phase is uniformly distributed on any 2π interval and independent of r. Can get this by inspection of the integrand after change of variables (A, B) to (ρ, θ) with Jacobian r.

(c) Suppose that A and B are independent, zero mean Gaussian random variables with variance equal to 1. Determine and sketch the probability density functions for both $R = \sqrt{A^2 + B^2}$ and $R^2 = A^2 + B^2$. [30%]

Solution:

Independent Gaussians:

$$f_{A,B}(a, b) = \frac{1}{2\pi} \exp(-a^2/2) \exp(-b^2/2) = \frac{1}{2\pi} \exp(-(a^2 + -b^2)/2)$$

So, $f(r^2) = \frac{1}{2\pi} \exp(-(r^2/2)$ and we have a circularly symmetric density.

Then pdf of $R = \sqrt{A^2 + B^2}$ is, from (b):

$$f_R(r) = 2\pi r \frac{1}{2\pi} \exp(-(r^2/2) = 2\pi r \frac{1}{2\pi} \exp(-(r^2/2) = r \exp(-r^2/2)$$

Now, take $Y = R^2$, so $f_Y(y) = f_R(r)|dy/dr|^{-1} = r \exp(-(r^2/2)/(2r) = \exp(-y/2)/2$.

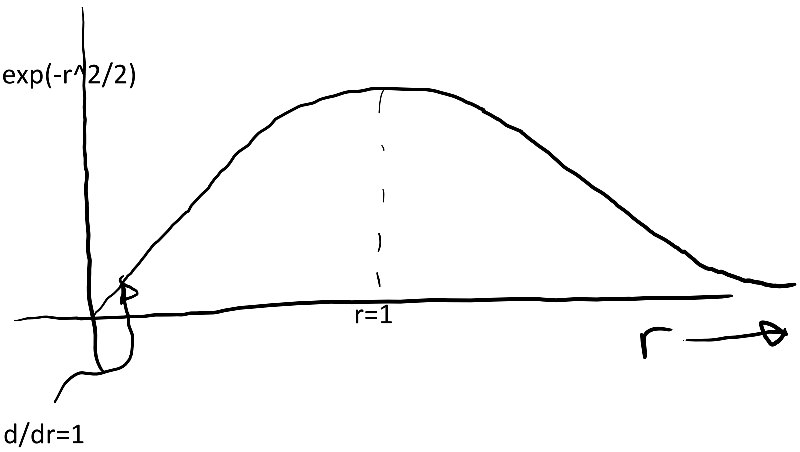
Sketches...

$r \exp(-r^2/2)$:

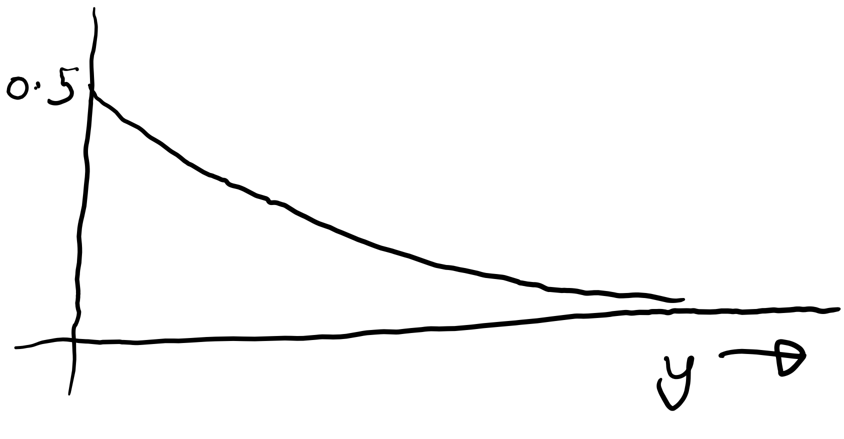
$d/dr = \exp(-r^2/2) - r^2 \exp(-r^2/2)$

$r=1$ for turning point.

Initial gradient $=1$



$\exp(-y/2)/2$:



(d) Suppose there are N data measurements, $n = 0, 1, ..., N-1$. An unbiased estimator $\hat{\theta}(\mathbf{x}) = [\hat{A}, \hat{B}]^T$ is constructed for $\theta = [A, B]^T$. An estimate for R^2 is now constructed as $\hat{R^2} = |\hat{\theta}(\mathbf{x})|^2 = \hat{A}^2 + \hat{B}^2$. Show that this estimator for R^2 is biased since

$$E[\hat{A}^2] = A^2 + \text{var}\hat{A}$$

Explain why the OLS estimator would give the smallest bias in R^2 of any linear unbiased estimator under this scheme. What do you think would happen to this bias as $N \to \infty$? [30%]

Solution:

To get the bias, calculate $E[\hat{R}^2] = E[|\hat{\theta}|^2]$:

$$E[\hat{R}^2] = E[|\hat{\theta}|^2] = E[\hat{A}^2] + E[\hat{B}^2]$$

Now, $E[X^2] = var[X] + E[X]^2$, so

$$E[\hat{A}^2] = var(\hat{A}) + E[\hat{A}]^2 = var(\hat{A}) + A^2$$

since $E[\hat{A}] = A$ (unbiased). But, $var(\hat{A}) \geq 0$ and so $E[\hat{A}^2]] \geq A^2$, which means that the estimator is always bised on the positive side (too large on average), unless it is unbiased with zero variance (trivial case).

Since the model at start of the question is a general linear model, the unbiased solution with lowest $var(\hat{A})$ is the OLS solution (see lecture notes). Hence this is the best you can do with an unbiased estimator $\hat{\theta}$. We would expect the bias to get smaller and smaller as N increases, since $\hat{\theta}$ would get closer and closer to the true values with smaller variance.

---

4 Examiner's comments:

a) Impulse response was obtained by most but few could sketch the response of the filter to s_n. b) was pleasingly well handled by many. c) caused some issues, but a good number of candidates came up with a good numerical estimate (and no one found a closed form answer – neither did I!) d) well known by most.

A smoothing fillter is to be used in enhancement of a signal pulse in additive noise. The fiter has the following recursive form:

$$x_n = \alpha x_{n-1} + u_n$$

where $0 < \alpha < 1$ is a parameter of the filter, u_n is its input signal and x_n its output. Suppose that the input signal is a noisy rectangular pulse of the form:

$$u_n = s_n + v_n$$

where $\{v_n\}$ is zero-mean white noise having variance σ_v^2 and, for $n = -\infty, \ldots, -1, 0, 1, \ldots\infty$,

$$s_n = \begin{cases} 1, & n = 0, \ldots, N-1 \\ 0, & \text{otherwise} \end{cases}$$

(a) Show that the impulse (unit pulse) response of the filter is

$$h_p = \alpha^p, \ p \geq 0$$

and sketch the output of the filter in response to the pulse s_n alone, i.e. setting $v_n = 0$. [20%]

Solution:

Can do this either using z-transform, or by directly substituting in $u_0 = \delta_0 = 1$:

$$h_0 = x_0 = \alpha \times 0 + 1 = 1, \ h_n = x_n = \alpha h_{n-1} + 0 = \alpha h_{n-1}$$

So by induction, $h_n = \alpha h_{n-1} = \alpha\alpha^{n-1} = \alpha^n$, and this proves the result.

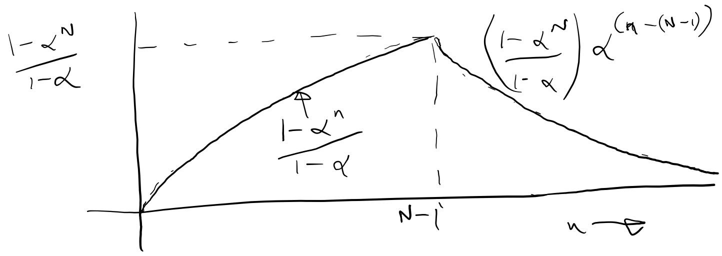
Output in response to just s_n is

$$\sum_{m=0}^{n} h_m = \frac{1-\alpha^n}{1-\alpha}, \ \ n \leq N-1$$

and

$$\alpha^{n-(N-1)} x_{N-1}, \ \ n > N-1$$

Sketch...



(b) Show that the power Signal-to-noise ratio (SNR) at the output of the filter at time index $n = N - 1$ can be written as

$$\frac{(1-\alpha^N)^2(1+\alpha)}{(1-\alpha)\sigma_v^2}$$

Determine value of α that gives best output SNR when $N = 2$ and comment on the result compared with the SNR of the unfiltered signal u_n at $n = N - 1$. [30%]

Signal component at output at time $N - 1$:

$$\sum_{p=0}^{\infty} h_p s_{N-1-p} = \sum_{p=0}^{N-1} \alpha^p = (1-\alpha^N)/(1-\alpha)$$

[summing a GP with N terms].

and hence signal output power is $(1-\alpha^N)^2/(1-\alpha)^2$.

Noise output power is:

$$E[(\sum_{p=0}^{\infty} h_p v_{N-p})^2] = \sum_p \alpha^{2p} E[v_{N-p}^2] = \sigma_v^2/(1-\alpha^2)$$

since cross terms are $h_p h_q E[v_{N-p}v_{N-q}] = 0$ for $p \neq q$, as v_n is zero mean and white (=uncorrelated).

Therefore signal-to-noise power ratio at output is:

$$\frac{(1-\alpha^N)^2}{\sigma_v^2(1-\alpha)^2}(1-\alpha^2) = \frac{(1-\alpha^N)^2}{\sigma_v^2(1-\alpha)}(1+\alpha)$$

as required.

Optimal enhanced pulse has highest SNR. With $N = 2$, this gives:

$$SNR = (1+\alpha)^3(1-\alpha)$$

and

$$dSNR/\alpha = 2(1+\alpha)^2(1-\alpha) - (1+\alpha)^3$$

and solving for maximum find $\alpha = 0.5$.

SNR at this optimal α is $3 \times 0.75^2/\sigma_v^2 = 1.7$.

For the unfiltered signal, $SNR = 1/\sigma_v^2$, so we are getting an improvement already.

(c) Now, set $N = 10$. Determine, using approximate numerical optimisation if necessary, the value of α which gives best output SNR at $n = N - 1$. [20%]

A few trial values of α leads to optmal value $\alpha \approx 0.9$ and SNR around 8 (only approx. answers req'd.) Thus we are getting further improvement for a larger N compared to (b).

(d) Design the optimal matched filter for this problem and comment on its performance compared with the smoothing filter above for $N = 2$ and $N = 10$, using the optimal values of α computed in (b) and (c) above. [30%]

Solution:

The optimal filter just uses $h_p = s_{N-1-p}$ (time-reversed pulse s_n) and has SNR N/σ_v^2 (from lecture notes).

Thus for both $N = 2$ and $N = 10$ the ad hoc smoothing filter gets quite close to the SNR performance of the optimal matched filter.

END OF PAPER