Version SJG/4

Worked Solutions
Worked Solutions

EGT3
ENGINEERING TRIPOS PART IIB

Wednesday 30 April 2025   9.30 to 11.10

Module 4F7

STATISTICAL SIGNAL AND NETWORK MODELS

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 **Examiner's comment:** A popular question, well answered by most. Part a)i) The state space model well known but many missed the initial conditions and stated in terms of Gaussians rather than expectations. a)ii) Well known but many missed the initial conditions a)iii) Good a)iv) good although many only got that it was BLUE b)i) Good. b)ii) Good b)iii) OK but some got the variance term wrong.

The Kalman filtering recursions for a linear state-space model, based on data $y_1, y_2, \ldots y_t$, are expressed as

$$
\begin{aligned}
\mu_{t|t-1} &= A\mu_{t-1|t-1} \\
P_{t|t-1} &= \Sigma_v + AP_{t-1|t-1}A^T \\
\mu_{t|t} &= \mu_{t|t-1} + K_t(y_t - B\mu_{t|t-1}) \\
P_{t|t} &= (I - K_t B)P_{t|t-1} \\
K_t &= P_{t|t-1}B^T(P_{t|t-1}B^T + \Sigma_w)^{-1}
\end{aligned}
$$

where the terms $\mu_{t|t-1}$ and $\mu_{t|t}$ are estimators of a hidden state variable x_t.

(a) (i) Write down an appropriate linear state-space model that is solved by these recursions, including the initialisation at $t = 0$. What do the terms Σ_v and Σ_w signify? [10%]

Solution:

$$x_t = Ax_{t-1} + v_t,\ \ E[v_t] = 0,\ \ cov[v_t] = \Sigma_v$$

$$y_t = Bx_t + w_t,\ \ E[w_t] = 0,\ \ cov[w_t] = \Sigma_w$$

w_t and v_t serially uncorrelated. $Ex_0 = \mu_{0|0}$, $cov[x_0] = P_{0|0}$.

(ii) Under what conditions on the state-space model do $\mu_{t|t-1}$ and $\mu_{t|t}$ give optimal Bayesian estimators of the state? [10%]

Solution:

Conditions are as in (i) (linear), plus Gaussian model,

$$v_t \overset{iid}{\sim} \mathcal{N}(0, \Sigma_v)$$

$$w_t \overset{iid}{\sim} \mathcal{N}(0, \Sigma_w)$$

$$x_0 \sim \mathcal{N}(\mu_{0|0}, P_{0|0})$$

(iii) Under these conditions, what are the conditional distributions $p(x_t|y_{1:t})$ and $p(x_t|y_{1:t-1})$? [15%]

Solution:
$p(x_t|y_{1:t}) = \mathcal{N}(\mu_{t|t}, P_{t|t})$ and $p(x_t|y_{1:t-1}) = \mathcal{N}(\mu_{t|t-1}, P_{t|t-1})$

(iv) When these conditions may not be assumed, what form of optimality is achieved by the Kalman filter recursions? [10%]

Solution:
The Kalman filter in this case is the best linear estimator for x_t in a MMSE sense.

(b) A stochastic differential equation is defined as

$$\dot{x}(t) = ax(t) + W(t)$$

where $\{W(t)\}$ is continuous-time zero-mean Gaussian white noise with autocorrelation function $R_{WW}(\tau) = \delta(\tau)$ and $a \leq 0$.

(i) Using Laplace transforms, or otherwise, show that

$$x(\delta t) = \exp(a\delta t)\left(x(0) + \int_0^{\delta t} \exp(-a\tau)W(\tau)d\tau\right)$$

[20%]

Solution:
Take Laplace transforms of both sides:

$$s\bar{X}(s) - x(0) = a\bar{X}(s) + \bar{W}(s)$$

Rearranging:

$$\bar{X}(s) = \frac{1}{(s-a)}x(0) + \frac{1}{(s-a)}\bar{W}(s)$$

and taking ILT:

$$x(t) = \exp(at)x(0) + \int_0^t \exp(a(t-\tau))W(\tau)d\tau$$

Now, shifting time interval to start at t and end at $t + \delta t$, since the SDE is time-homogeneous:

$$x(t+\delta t) = \exp(a\delta t)x(t) + \int_0^{\delta t} \exp(a(\delta t - \tau))W(\tau)d\tau = \exp(a\delta t)\left(x(t) + \int_0^{\delta t} \exp(-a\tau)W(\tau)d\tau\right)$$

(ii) Determine the mean and variance of the integral term $\int_0^{\delta t} \exp(-a\tau)W(\tau)d\tau$. [25%]

Solution:
$E\int_0^{\delta_t} \exp(-a\tau)W(\tau)d\tau = \int_0^{\delta_t} \exp(-a\tau)EW(\tau)d\tau = 0.$

$$\begin{aligned} var\int_0^{\delta_t} \exp(-a\tau)W(\tau)d\tau &= E\int_0^{\delta_t} \exp(-a\tau)W(\tau)d\tau \int_0^{\delta_t} \exp(-a\tau')W(\tau')d\tau' \\ &= \int_0^{\delta_t}\int_0^{\delta_t} \exp(-a\tau)E[W(\tau)W(\tau')]\exp(-a\tau')d\tau' d\tau \\ &= \int_0^{\delta_t}\int_0^{\delta_t} \exp(-a\tau)\delta(\tau-\tau')\exp(-a\tau')d\tau' d\tau \\ &= \int_0^{\delta_t} \exp(-a\tau)\exp(-a\tau)d\tau \\ &= \frac{1}{-2a}[\exp(-2a\tau)]_0^{\delta_t} \\ &= \frac{1}{-2a}(\exp(-2a\delta t)-1) = \frac{1}{2a}(1-\exp(-2a\delta t)) \end{aligned}$$

(iii) Hence write down a linear Gaussian dynamical model in discrete time for samples $x_n = x(n\delta t)$. [10%]

Solution:
As in a)(i) with:
$A = \exp(a\delta t)$ and
$\Sigma_v = \exp(2a\delta_t)\frac{1}{2a}(1-\exp(-2a\delta t)) = \frac{1}{2a}(\exp(2a\delta t)-1)$
The extra term $\exp(2a\delta_t)$ arises because the integral is multiplied by $\exp(a\delta_t)$ and $cov(\alpha X) = \alpha^2 covX$.

2 **Examiner's comment:** A less popular question, but quite well answered. Parts a) – c) well handled in general – good skill in linearising the state space model. Part d) generally well done, a few very vague summaries of the EKF update. e) again well answered although many (most) neglected to give the actual formula for the incremental weight.

A nonlinear dynamical system obeys the following dynamical model for $t = 1, 2, ...,$:

$$x_t = \pi/2\sin(x_{t-1}) + v_t$$

and the state is observed in noise:

$$y_t = x_t^2 + w_t$$

where $v_t \sim \mathcal{N}(0, \sigma_v^2)$ and $w_t \sim \mathcal{N}(0, \sigma_w^2)$ are mutually uncorrelated white noise processes.

(a) Derive a linear and Gaussian approximation of the dynamical system expanded around a fixed point $x_{t-1} = x^*$. [15%]

Answer:

$$\pi/2 \frac{d \sin(x)}{dx} = \pi/2 \cos(x)$$
$$x_t \approx \pi/2 \sin(x^*) + (x_{t-1} - x^*)\pi/2 \cos(x^*) + v_t$$

(b) Derive a linear and Gaussian approximation of the observation model, expanded around a fixed point $x_t = x^+$. [10%]

Answer:

$$dx^2/dx = 2x$$
$$y_t \approx x^{+2} + 2x^+(x_t - x^+) + w_t$$

(c) Suppose that $p(x_{t-1}|y_{1:t-1}) = \mathcal{N}(\mu_{t-1}, p_{t-1}^2)$ where p_{t-1} is a standard deviation. Show that the prediction step under the approximate dynamical system is:

$$p(x_t|y_{1:t-1}) \approx \mathcal{N}(\pi/2 \sin(x^*) + (\mu_{t-1} - x^*)\pi/2 \cos(x^*),\ \pi^2/4 \cos^2(x^*) p_{t-1}^2 + \sigma_v^2)$$

[20%]

Answer:

In the approximation we have:

$$x_t \approx \pi/2 \sin(x^*) + (x_{t-1} - x^*)\pi/2 \cos(x^*) + v_t$$

But $x_{t-1} \sim \mathcal{N}(\mu_{t-1}, p_{t-1}^2)$, and we know that if $X \sim \mathcal{N}(\mu, \sigma^2)$ then $Y = aX + b + V \sim \mathcal{N}(a\mu + b, a^2\sigma^2 + \sigma_v^2)$. Combining the two:

$$x_t \sim \mathcal{N}(\pi/2 \sin(x^*) + (\mu_{t-1} - x^*)\pi/2 \cos(x^*), \pi^2/4 \cos^2(x^*) p_{t-1}^2 + \sigma_v^2)$$

as required.

(d) Explain how sequential estimation of x_t in this approximate model can be implemented as an Extended Kalman Filter, stating carefully how x^* and x^+ are chosen and detailing how the update step of the Kalman Filter is modified compared with the standard KF. [35%]

Answer:

First, we expand the prediction step around $x^* = \mu_{t-1}$:

$$x_t \sim \mathcal{N}(\pi/2\sin(\mu_{t-1})+(\mu_{t-1}-\mu_{t-1})\pi/2\cos(x^*), \pi^2/4\cos^2(x^*)p_{t-1}^2+\sigma_v^2) = \mathcal{N}(\pi/2\sin(\mu_{t-1}), \pi^2/4\cos^2(x$$

Thus the EKF prediction step is: $\mu_{t|t-1} = \pi/2\sin(\mu_{t-1})$ and $p_{t|t-1}^2 = \pi^2/4\cos^2(x^*)p_{t-1}^2 + \sigma_v^2$

Now, express the approximate observation equation around $x^+ = \mu_{t|t-1}$:

$$y_t = \mu_{t|t-1}{}^2 + 2\mu_{t|t-1}(x_t - \mu_{t|t-1}) + w_t$$

This can be arranged as a pseudo-observation equation:

$$\tilde{y}_t = y_t + \mu_{t|t-1}{}^2 = 2\mu_{t|t-1}x_t + w_t$$

We now carry out the standard Kalman update with observation matrix $B = 2\mu_{t|t-1}$, observation $\tilde{y}_t = y_t + \mu_{t|t-1}{}^2$ and observation variance σ_w^2.

(e) If this model were also implemented using a particle filter, using the 'prior' proposal distribution $p(x_t|x_{t-1})$, compare and contrast the likely performance of the two filters. Determine the incremental weight of the particle filter with this proposal distribution. [20%]

Answer:

The EKF is sub-optimal and we cannot adjust its performance. But it is cheap to implement. The PF is adjustable and tends towards optimal as number of particles gets very large. Can then be very expensive to implement - trade-off! Another important point is that the observation density is multimodal because of the x^2 term. Hence the EKF is likely to track the wrong mode and the particle filter at least has a chance of covering both modes.

Incremental weight in this case ('bootstrap' filter):

$$w_t^{(i)} \propto p(y_t|x_t^{(i)})\mathcal{N}(x_t^{(i)2}, \sigma_w^2)$$

Weights to be normalised to sum to 1.

3 **Examiner's comment:** This question was generally answered to a high standard. Part (a) was straightforward. Some answers got confused in (b) and (c) when taking derivatives of a sum, while some answers to part (d) appeared to fudge the θ_i term. Part (e) was more variable and distinguished the good from the very good answers. Some of the good answers were still not fully incisive. For example, some argued that when using MLE parameters the model could not scale appropriately – this essentially gets at the issue but doesn't quite show that no set of parameters can have the desired scaling.

Consider a model for *directed* graphs on n nodes, where the edge from i to j occurs with probability p_{ij}. The adjacency matrix is defined such that

$$A_{ij} \sim \text{Bernoulli}(p_{ij})$$

and we allow self-edges to occur.

(a) Show that the log-likelihood for p_{ij} given adjacency matrix A is

$$\sum_{i,j} \left(A_{ij} \log p_{ij} + (1 - A_{ij}) \log \left(1 - p_{ij}\right)\right)$$

[20%]

Solution: The probability of each edge is $P(A_{ij}) = p_{ij}^{A_{ij}} \left(1 - p_{ij}\right)^{1-A_{ij}}$. Each edge is independent, so the probability of the full matrix A is the product over all i and j. Taking the logarithm gives

$$\log P(A|p) = \sum_{i,j} \left(A_{ij} \log p_{ij} + (1 - A_{ij}) \log(1 - p_{ij})\right)$$

(b) Find the maximum likelihood estimate (MLE) for the parameters p_{ij}. [20%]

Solution: The MLE are the values of p_{ij} that maximize $\log P(A|p)$, so we need

$$\frac{d \log P(A|p)}{dp_{ij}} = \frac{A_{ij}}{p_{ij}} - \frac{1 - A_{ij}}{1 - p_{ij}} = 0$$

This is solved with $\hat{p}_{ij} = A_{ij}$.

(c) We now constrain the parameter matrix p_{ij} to be of the form

$$p_{ij} = \theta_i$$

Show that the MLE is

$$\hat{\theta}_i = \frac{1}{n} \sum_j A_{ij} = \frac{k_i}{n}$$

where $k_i = \sum_j A_{ij}$ is the out-degree of node i. [20%]

Solution: The log-likelihood is now

$$\sum_{i,j} \big(A_{ij} \log \theta_i + (1 - A_{ij}) \log(1 - \theta_i) \big)$$

and taking derivative wrt θ_i we get

$$\sum_j \left(\frac{A_{ij}}{\theta_i} - \frac{1 - A_{ij}}{1 - \theta_i} \right) = \frac{k_i}{\theta_i} - \frac{n - k_i}{1 - \theta_i} = 0$$

from which simple algebra establishes $\hat{\theta}_i = k_i/n$.

(d) A *reciprocal* edge is an edge that exists in both directions, i.e., $A_{ij} = A_{ji} = 1$ (note, we will not count self-edges as reciprocal). Show that the expected number of reciprocal edges at node i is

$$\theta_i \left(\sum_j \theta_j - \theta_i \right)$$

and the total number is thus

$$\left(\sum_i \theta_i \right)^2 - \sum_i \theta_i^2$$

[20%]

Solution: A reciprocal edge exists if $A_{ij} A_{ji} = 1$, so the expected number is $\sum_{j \neq i} E[A_{ij} A_{ji}]$. By independence we have

$$\sum_{j \neq i} E[A_{ij} A_{ji}] = \sum_{j \neq i} E[A_{ij}] E[A_{ji}] = \sum_{j \neq i} \theta_i \theta_j = \theta_i \left(\sum_j \theta_j - \theta_i \right)$$

The total number is this quantity summed over all i, which is precisely the formula given.

(e) For online social networks, both the number of edges and the number of reciprocal edges typically scale linearly with n. Show that in this model, one cannot have both. [20%]

Solution: The expected total number of edges is $\sum_{i,j} \theta_i = n \sum_i \theta_i$. For this to be $O(n)$ we must ask that $\sum_i \theta_i$ is $O(1)$.

However, the number of reciprocal edges is less than $(\sum_i \theta_i)^2$, which is an $O(1)$ quantity squared, and hence $O(1)$.

Hence, in this model if the number of edges is $O(n)$ then the number of reciprocal edges is $O(1)$.

---

4 **Examiner's comment:** The weakest question on average but also a high variance. This was the only question with a handful of very incomplete answers. Removing these attempts raises the mean in line with other questions. Parts (a) and (b) had no major issues for most candidates. Using the variable u instead of w in part (c) caused some confusion but generally did not cause substantive issues – several students reverted to using w but gave otherwise correct answers. Part (d) was well answered by students who had correctly answered (c) but others failed to use the solution given in (c) to answer (d). For part (e), some answers correctly computed u = 1/2 but then incorrectly provided that as the final answer.

Consider a configuration model network that contains only nodes of degree 1 and degree 3. A fraction α of the nodes have degree 1, while a fraction $1-\alpha$ have degree 3. Hence, the degree distribution is

$$p_k = \alpha\, \delta_{k,1} + (1-\alpha)\, \delta_{k,3}$$

(a) What is the mean degree? [10%]

The mean degree is

$$\sum_k k p_k = \alpha + 3(1-\alpha) = 3 - 2\alpha$$

(b) Provide an expression for the excess degree distribution, q_k. [15%]

The excess degree distribution is $q_k = (k+1)p_{k+1}/E[k]$, hence

$$q_k = \frac{(k+1)\alpha\delta_{k+1,1} + (k+1)(1-\alpha)\delta_{k+1,3}}{3-2\alpha} = \frac{\alpha\delta_{k,0} + 3(1-\alpha)\delta_{k,2}}{3-2\alpha}$$

(c) Let u be the probability that a randomly chosen edge does not lead to the giant component. Show that u obeys the self-consistent equation

$$\left(u-1\right)\left(u - \frac{\alpha}{3-3\alpha}\right) = 0$$

[35%]

The equation for u is $u = \sum_k u^k q_k$. Inserting the formula for q_k we get

$$u = \frac{\alpha}{3-2\alpha} + \frac{3-3\alpha}{3-2\alpha} u^2$$

Re-arranging to be in a standard quadratic form we have

$$u^2 - \left(\frac{3-2\alpha}{3-3\alpha}\right)u + \frac{\alpha}{3-3\alpha} = 0$$

which factorizes to the stated quadratic.

(d) Hence, argue that a giant component exists if and only if $\alpha < 3/4$. [15%]

The quantity u is the probability that a randomly chosen edge does not lead to the giant component. The non-trivial root is at $u = \frac{\alpha}{3-3\alpha}$. For a giant component this must be less than 1, hence $\frac{\alpha}{3-3\alpha} < 1$. This occurs when $\alpha < 3/4$.

(e) Derive an expression for the fractional size of the giant component when $\alpha = 3/5$. [25%]

When $\alpha = 3/5$ then $u = \frac{\alpha}{3-3\alpha} = \frac{1}{2}$. Let s be the probability a randomly chosen node is not in the giant component. We have $s = \sum_k p_k u^k$, so

$$s = \frac{\alpha}{2} + \frac{1-\alpha}{2^3} = \frac{3}{10} + \frac{2}{40} = \frac{7}{20}$$

or 35%.

END OF PAPER

THIS PAGE IS BLANK