Engineering Pt IIa, Module 3F3 - Estimation and Inference
Examples paper

Questions 1 and 2 are revising and extending some useful basic properties from Linear Algebra

1. Find an expression for the vector gradient $\partial J/\partial \mathbf{x}$ for a column vector $\mathbf{x}$ in each of the following cases:

   (a)
   $$J = \mathbf{b}^T\mathbf{x} = \mathbf{b}^T\mathbf{x}$$
   where $\mathbf{b}$ is a column vector

   (b)
   $$J = \mathbf{x}^T\mathbf{B}\mathbf{x}$$
   where $\mathbf{B}$ is a non-symmetric matrix

   (c)
   $$J = \mathbf{x}^T\mathbf{B}\mathbf{x}$$
   where $\mathbf{B}$ is a symmetric matrix

2. (a) Show that any matrix of the form $\mathbf{B} = \mathbf{G}^T\mathbf{G}$ is positive semi-definite (non-negative definite), i.e. that
   $$\mathbf{x}^T\mathbf{B}\mathbf{x} \geq 0$$
   for *any* vector $\mathbf{x}$.

   (b) Under what condition on $\mathbf{G}$ is the matrix $\mathbf{B}$ also positive definite, i.e.
   $$\mathbf{x}^T\mathbf{B}\mathbf{x} > 0$$
   for *any* non-zero vector $\mathbf{x}$?

   (c) Take the following expression and find the global minimum with respect to $\mathbf{x}$:
   $$J = \mathbf{x}^T\mathbf{B}\mathbf{x} - 2\mathbf{b}^T\mathbf{x}$$
   where $\mathbf{B}$ is a symmetric positive definite matrix.

(d) Complete the square for the following expression:

$$\mathbf{x}^T\mathbf{B}\mathbf{x} + 2\mathbf{b}^T\mathbf{x}$$

where $\mathbf{B}$ is a symmetric, invertible matrix, i.e. express it in the form:

$$(\mathbf{x} - \mathbf{m})^T\mathbf{M}(\mathbf{x} - \mathbf{m}) + C$$

where $\mathbf{m}$, $\mathbf{M}$ and C do not depend upon $\mathbf{x}$.

(e) Hence show that in the General Linear Model with Gaussian independent errors $e_n \sim \mathcal{N}(0, \sigma_e^2)$ and Gaussian prior distribution $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{m}_{\boldsymbol{\theta}}, \mathbf{C}_{\boldsymbol{\theta}})$, the posterior distribution is also Gaussian, with mean vector and covariance matrix as given in the lectures. You may assume that the design matrix $\mathbf{G}$ is full rank.

3. * In a ML estimation scenario, the data are generated independently as follows:

$$x_n = \mu + e_n,\ n = 0, 1, ..., N - 1$$

where e_n is zero mean, independent, Gaussian noise, with variance σ_n^2. Note that the variance is known, but is *different* for different data points n.

Determine the ML estimator for the unknown parameter μ. Determine also OLS estimator for the same model and comment on the relation between the two.

Determine the bias and variance for this ML estimator and compare it to that of the OLS estimator.

How do you think such a model could be used to deal with occasional large erroneous measurements ('outliers') in the measurement system?

4. In the random sinusoid model, a frequency component with frequency Ω can be expressed as:

$$X_n = A\sin(n\Omega) + B\cos(n\Omega)$$

If A and B are independent and distributed as $\mathcal{N}(0, \sigma^2)$, show that an equivalent model can be written in the form

$$X_n = R\cos(n\Omega + \Phi)$$

where Φ is uniformly distributed between 0 and 2π, and $R \geq 0$ is independently drawn as a random variable whose pdf you should determine.

Sketch the pdf of R.

[Hint: you should consider the Jacobian between the random variables involved, see also Ex. paper 1 Q.8].

5. The random time interval T between trades in an electronic trading system is modelled as an exponential distribution:

$$p_T(t) = \theta^{-1}\exp(-t/\theta),\ \ t > 0$$

where θ is an unknown parameter.

Show that the expected time between trades is θ and determine the variance of T.

Observations are made of the times t_i between the first $N+1$ trades in a given day, leading to a vector of measurements $\mathbf{t} = [t_1, t_2, ..., t_N]^T$.

If the times between successive trades are assumed to be independent of one another, show that the likelihood function for θ can be expressed as

$$p(\mathbf{t}|\theta) = \theta^{-N} \exp(-\sum_{i=1}^{N} t_i/\theta)$$

Hence determine the ML estimator for θ.

Determine the bias and variance of this estimator and comment on its behaviour for small and large N.

Now derive the Bayesian MAP estimator if the prior distribution can be assumed to be:

$$p(\theta) = \theta^{-2} \exp(-1/\theta)$$

By considering the form of the MAP estimator compared with the ML estimator, give an interpretation of the prior density as introducing one or more additonal observations t_i compared with the original likelihood function. What is the effect of the prior as the number of data points becomes very large?

6. * In an AR modelling problem it is required to fit a first order ($P = 1$) model to some scientific data $\mathbf{x} = [x_0, x_1, ..., x_{N-1}]^T$. It is known that all data prior to time $n = 0$ were zero. It is also known that the prediction error process $\{e_n\}$ is Gaussian with zero mean and variance 1.

   Show that the ML estimator of the single parameter is given by

$$a^{ML} = \frac{\sum_{n=1}^{N-1} x_{n-1}x_n}{\sum_{n=0}^{N-2} x_n^2} = S_1(\mathbf{x})/S_2(\mathbf{x})$$

   It is thought that the AR parameter lies around 0.9 for the type of data being measured, with an uncertainty of roughly 0.1 . Thus a Gaussian prior with mean 0.9 and variance 0.01 is chosen for the estimation task.

   Determine the Bayesian posterior distribution for a when $S_1(\mathbf{x}) = 95$ and $S_2(\mathbf{x}) = 97$ and find the posterior probability that the model corresponds to an unstable filter (poles outside the unit circle).

   Sketch the prior probability distribution, the likelihood function and the posterior density, all as a function of a. Comment on the relationship between them in terms of their mean values and their variances.

   It is now required to constrain the prior such that unstable filter models have probability zero, but otherwise the prior has the same Gaussian shape as before within the stable region.

   Determine the new posterior probability for a and determine the Bayesian MAP and MMSE estimator for this new prior. Hence comment on whether the assumption of a stable model has a significant impact on the inference task.

Answers:

1. (a) $\mathbf{b}$
   (b) $(\mathbf{B}+\mathbf{B}^T)\mathbf{x}$
   (c) $2\mathbf{B}\mathbf{x}$

2. (b) $\mathbf{G}$ must be full rank (i.e. it has a full set of independent columns)
   (c) $\mathbf{x} = \mathbf{B}^{-1}\mathbf{b}$

3. $$\mu^{ML} = \sum_{n=0}^{N-1} x_n w_n$$

   where
   $$w_n = \frac{1/\sigma_n^2}{\sum_{m=0}^{N-1} 1/\sigma_m^2}$$

   Compared with $\mu^{OLS} = 1/N \sum_{n=0}^{N-1} x_n$ (from lectures) we see that this is a weighted version of OLS in which terms with larger observation variance get optimally 'down-weighted' and vice versa.

   Bias is zero, as for OLS, variance is $1/(\sum_{n=0}^{N-1} 1/\sigma_n^2)$, so the variance depends very much on the values of σ_n^2 as $N \to \infty$.

4. $$p(r,\phi) = \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}}$$

   i.e. Uniform in ϕ, Rayleigh in r.

5. $$var[T] = \theta^2$$

   $\theta^{ML} = \frac{\sum_{i=1}^{N} t_i}{N}$

   Bias is zero, variance is θ^2/N

   $\theta^{MAP} = \frac{1+\sum_{i=1}^{N} t_i}{N+2}$

6. $$p(a|\mathbf{x}) = \mathcal{N}(a|0.94, 0.005)$$

   Stable model:
   $$p(a|\mathbf{x}) = \begin{cases} \mathcal{N}(a|0.94, 0.005)/0.8, & -1 < a < 1 \\ 0 & \text{Otherwise} \end{cases}$$

   MAP estimator: 0.94

   MMSE estimator: 0.92

Suitable past tripos questions:
3F3: past Q4s are sometimes on syllabus but the style of question has changed, e.g.
2012 Q4 (part (c) can be done but is not really on-syllabus)
2008 Q4
2006 Q4