2008 IIB 4F6 2008 PROJ W J FITZGERALD

4F6 SIGNAL DETECTION AND ESTIMATION

1) Neyman Fisher Factorization theorem.

For an unbiased efficient estimator,

$$\frac{\partial}{\partial\theta} \ln p(x|\theta) = k(\theta)\left(\hat{\theta}(x) - \theta\right)$$

$$\therefore \ln p(x|\theta) = \int k(\theta')\left(\hat{\theta}(x) - \theta'\right) d\theta' + \ln h(x).$$

$$\therefore p(x|\theta) = g\left(T(x), \theta\right)\, h(x).$$

Where $T(x)$ is a sufficient statistic.

This is useful for finding sufficient statistics for unbiased estimators

20%.

1 ~~cont~~ cont.

For the scalar exponential family, the likelihood for the data $\underline{x}$ is given by

$$p(\underline{x} \mid \theta) = \prod_{n=0}^{N-1} \exp\left(A(\theta) B(x) + C(x) + D(\theta)\right)$$

$$= \exp\left(A(\theta) \sum_n B(x_n) + \sum_n C(x_n) + N D(\theta)\right)$$

$$= \exp\left(A(\theta) \sum_n B(x_n) + N D(\theta)\right) \times \exp\left(\sum_n C(x_n)\right).$$

$$\equiv g\left(T(\underline{x}), \theta\right) \times h(\underline{x})$$

Hence using the Neyman-Fisher factorization theorem, the sufficient statistic is given by

$$T(x) = \sum_n B(x_n).$$

40%.

1 (cont.)

For the Gaussian case, we can write

$$p(x|\mu) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-\mu)^2\right)$$

(assuming unit variance for simplicity – makes no difference

$$\therefore p(x|\mu) = \exp\Big(\underbrace{x\mu}_{A(\mu)B(x)} \underbrace{-\frac{1}{2}x^2}_{C(x)} + \underbrace{\left(-\frac{1}{2}\mu^2 + \ln 1/\sqrt{2\pi}\right)}_{D(\mu)}\Big)$$

$$\therefore \quad T(x) = \sum_{n=0}^{N-1} x_n$$

60%

~~For the exponential case,~~

$$\cancel{p(x|\lambda) = \lambda e^{-\lambda x}}$$

$$\cancel{= \exp -\underbrace{\lambda x}_{A(\lambda)B(x)} + \underbrace{0}_{C(x)} + \underbrace{\ln \lambda}_{D(\lambda)}}$$

$$\cancel{\therefore \quad T(x) = \sum_{n=0}^{N-1} x_n}$$

40%

Q2

Outline ~~proof~~ of CRCB

for unbiased estimator

$$\theta = \int \hat{\theta}(x)\, p(x|\theta)\, dx$$

$$\therefore \int (\hat{\theta}(x) - \theta)\, p(x|\theta)\, dx = 0$$

Diff w.r.t θ.

$$\int \frac{\partial}{\partial \theta} (\hat{\theta}(x) - \theta)\, p(x|\theta)\, dx = 0.$$

$$\int dx \left(\frac{\partial}{\partial \theta} p(x|\theta)\right)(\hat{\theta} - \theta) - \underbrace{\int dx\, p(x|\theta)}_{=1} = 0$$

use $\frac{\partial}{\partial \theta} p(x|\theta) = p(x|\theta) \frac{\partial}{\partial \theta} \ln p(x|\theta)$. and

Schwarz inequality $\left(\left|\int f g\, dx\right|^2 \leq \int f^2 dx \int g^2 dx. \right)$

$$\therefore \int dx\, p(x|\theta) \left(\frac{\partial}{\partial \theta} \ln p(x|\theta)\right)^2 \times \int (\hat{\theta} - \theta)\, p(x|\theta)\, dx \geq 1$$

$$\therefore I_\theta \times \varepsilon^2 \geq 1.$$

$$I_\theta = E\left[\left(\frac{\partial}{\partial \theta} \ln p(x|\theta)\right)^2\right].$$

$$\frac{\partial^2 \ln p(x|\theta)}{\partial \theta^2} = \frac{\partial}{\partial \theta}\left(\frac{1}{p(x|\theta)} \frac{\partial}{\partial \theta} p(x|\theta)\right)$$

$$= -\frac{1}{p(x|\theta)^2}\left(\frac{\partial p(x|\theta)}{\partial \theta}\right)^2 + \frac{1}{p} \frac{\partial^2 p}{\partial \theta^2}.$$

$$\therefore E\left(\frac{\partial \ln p(x|\theta)}{\partial \theta}\right)^2 = -E\left(\frac{\partial^2}{\partial \theta^2} \ln p(x|\theta)\right)$$

40%

2/ first part is book work. $\frac{38}{0}$%

—//—

parameterizing A, and B as $\Theta = \begin{bmatrix} A \\ B \end{bmatrix}$,

the 2×2 Fisher information is

$$I(\theta) = \begin{bmatrix} -E\frac{\partial^2 \ln p(d|\theta)}{\partial A^2} & -E\frac{\partial^2 \ln p(d|\theta)}{\partial A \partial B} \\ -E\frac{\partial^2 \ln p(d|\theta)}{\partial B \partial A} & -E\frac{\partial^2 \ln p(d|\theta)}{\partial B^2} \end{bmatrix}.$$

and $p(d|\theta) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(d(n) - A - Bn)^2\right).$

$$\frac{\partial \ln p(d|\theta)}{\partial A} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(d(n) - A - Bn) \; ; \; \frac{\partial^2 \ln p(d|\theta)}{\partial A^2} = -\frac{N}{\sigma^2}$$

$$\frac{\partial \ln p(d|\theta)}{\partial B} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(d(n) - A - Bn)\,n \; ; \; \frac{\partial^2 \ln p(d|\theta)}{\partial B^2} = -\frac{1}{\sigma^2}\sum_{n=0}^{N-1} n^2$$

$$\frac{\partial \ln p(d|\theta)}{\partial A \partial B} = -\frac{1}{\sigma^2}\sum_{n=0}^{N-1} n$$

6/

-464-

2 cont

$$I(\theta) = \frac{1}{\sigma^2}\begin{bmatrix} N & \sum n \\ \sum n & \sum n^2 \end{bmatrix}.$$

$$= \frac{1}{\sigma^2}\begin{bmatrix} N & \frac{N(N-1)}{2} \\ \frac{N(N-1)}{2} & \frac{N(N-1)(2N-1)}{6} \end{bmatrix}.$$

80%

$$\therefore I^{-1} = \sigma^2\begin{bmatrix} \frac{2(2N-1)}{N(N+1)} & -\frac{6}{N(N+1)} \\ -\frac{6}{N(N+1)} & \frac{12}{N(N^2-1)} \end{bmatrix}.$$

$$\therefore \operatorname{var}(\hat{A}) \geq \frac{2(2N-1)\sigma^2}{N(N+1)}$$

$$\operatorname{var}(\hat{B}) \geq \frac{12\sigma^2}{N(N^2-1)}$$

$$\therefore \frac{\operatorname{var}(\hat{A})}{\operatorname{var}(\hat{B})} = \frac{(2N-1)(N-1)}{6} > 1 \quad \text{for } N \geq 3.$$

20%

$\therefore$ easier to estimate B.

Q3

i) Bayesian inference requires multidimensional integration for

i) Expectations
ii) Marginalization
iii) Evidence calculations
(denominator in Bayes theorem for model selection)

These integrals can only be carried out in simple situations – linear models, Gaussian noise and low dimensionality

Therefore need numerical techniques:

i) Accept – Reject
ii) Importance sampling
iii) Monte carlo
iv) Markov Chain Monte Carlo.

Special case of Metropolis – Hastings is the Gibbs sampler.

20%

Parameter vector $(a_1, a_2, \ldots a_k)$

initial guess is $(a_1^0, a_2^0, \ldots a_k^0)$

Gibbs sampler algorithm is:

$$a_1^1 \leftarrow p(a_1 | a_2^0, a_3^0 \ldots a_k^0)$$

$$a_2^1 \leftarrow p(a_2 | a_1^1, a_3^0, a_4^0, \ldots a_k^0)$$

$$\vdots$$

etc.

Requires conditional distributions

For the case of linear regression, with known (noise) variance

$$m_1 \leftarrow p(m \mid c_0, D)$$

$$c_1 \leftarrow p(c \mid m_1 D)$$

Q3 cont

For the model $d_i = c + m x_i + n_i$

where the additive noise is Gaussian, the likelihood is

$$p(d|m,c) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_i (d_i - m x_i - c)^2\right)$$

Assuming uniform priors, the posterior can be written

$$p(m,c|d) \propto \sigma^{-(N+1)} \exp(\text{ as above. })$$

Conditioning on m and c respectively we find the conditionals

$$p(m|c,d) \propto \exp -\frac{1}{2}\left(\frac{\sqrt{X_2}}{\sigma}\right)^2\left(m - \frac{X_D - cX_1}{X_2}\right)^2$$

$$p(c|m,d) \propto \exp -\frac{1}{2}\left(\frac{\sqrt{N}}{\sigma}\right)^2\left(c - \frac{D_1 - mX_1}{N}\right)^2$$

Both are Gaussians and the Gibbs sampler can be obtained by drawing Gaussian variates.

40%

Q4

MAP

Conditional probabilities are $p(H_0/y)$ & $p(H_1/y)$

Choose H_0 if $p(H_0/y) > p(H_1/y)$

ie $\frac{p(H_0/y)}{p(H_1/y)} > 1$

Use Bayes theorem $p(H_0/y) = \frac{p(y/H_0)\,p(H_0)}{p(y)}$ etc.

$\therefore$ MAP criteria is $\frac{p(y/H_0)\,p(H_0)}{p(y)} > \frac{p(y/H_1)\,p(H_1)}{p(y)}$

$$\therefore \frac{p(y/H_0)}{p(y/H_1)} \geq \frac{p(H_1)}{p(H_0)}.$$

For Bayes criteria, introduce costs $C_{10} - C_{00} > 0$
$C_{01} - C_{11} > 0$.

Neyman Pearson based on maximising the prob of detection for some specified prob of false alarm.
– gets away from specifying arbitrary thresholds.

25

Threshold for N-P detector found from the slope of the ROC curve at the specified prob of false alarm.

15

A3 — Bookwork. (in the course notes)

Q4 First part is book work. ~~20%~~ 15%

—— // ——

threshold from slope of ROC curve

The general linear model may be written

$$\underline{d} = \underline{G}\,\underline{\theta} + \underline{w}$$

where the model parameters are contained in the vector $\underline{\theta}$. $\underline{d}$ is the observed data vector, $\underline{w}$ is the noise vector and $\underline{G}$ is a matrix.

For the signal

$$s(n) = A + Bn$$

the observed data is $\underline{d} = \underline{s} + \underline{w}$

$$\therefore \underline{d} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & N-1 \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} + \underline{w} = \underline{G}\,\underline{\theta} + \underline{w}.$$

30

8/

3 cont

The likelihoods for the observed noisy data $d(n)$ are

$$p(d|H_0) = \frac{1}{(2\pi)^{N/2}|C|^{1/2}} e^{-\frac{1}{2}d^T C^{-1} d}.$$

$$p(d|H_1) = \frac{1}{(2\pi)^{N/2}|C|^{1/2}} e^{-\frac{1}{2}(d-s)^T C^{-1}(d-s)}.$$

and the NP detector is

$$L(d) = \frac{p(d|H_1)}{p(d|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda.$$

$$\therefore L(d) = e^{-\frac{1}{2}[-2d^T C^{-1} s + s^T C^{-1} s]}$$

$$\therefore d^T C^{-1} s \underset{H_0}{\overset{H_1}{\gtrless}} \frac{1}{2} s^T C^{-1} s + \log(\lambda).$$

$$\therefore \frac{1}{\sigma^2} d^T \underline{H} \underline{\theta} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda' \qquad \therefore \boxed{\frac{1}{\sigma^2} \sum d(n)(A+Bn) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda'}$$

30%