PART IIB 2011 4F6 2011

PROF W J FITZGERALD

Signal Detection + Estimation

1) For a probability distribution $p(x)$ a 'good' definition of information that satisfies certain desirable properties is

$$I = -\int \ln p(x)\,dx$$

and the expectation of this is defined as entropy (H)

$$H = -\int p(x) \ln p(x)\,dx.$$

Nature seems to like having distribution which have maximum entropy and to assign probabilities that have maximum entropy subject to certain constraints

1 cont) Using the two constraints gives

$$p(x) = \beta e^{-\beta x} \; ; \; \beta = e^{1-\alpha} .$$

Using the 2nd constraint gives

$$\mu = \int x \, p(x) dx = \int_0^\infty x \beta e^{-\beta x} dx .$$

$$\therefore \mu = \frac{1}{\beta} .$$

Therefore the distribution having max. ent is

$$p(x) = \beta e^{-\beta x}$$

where $\beta = \dfrac{1}{\text{mean}}$.

2) If the likelihood function is $p(x|\theta)$, the Fisher Information associated with the parameter θ is defined as;

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

Where the expectation is w.r.t the probability measure $p(x|\theta)$.

For an unbiased estimator $\exists$ θ we have by definition

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

Differentiating w.r.t θ we have

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

$$\int dx \left[\frac{\partial}{\partial\theta} p(x|\theta)\right]\left(\hat{\theta}(x) - \theta\right) - \int dx\; p(x|\theta) = 0.$$

2 cont.)

Using $\frac{\partial}{\partial\theta} p(x|\theta) = \frac{\partial}{\partial\theta} \ln p(x|\theta) \cdot p(x|\theta)$

and $\int p(x|\theta)\, dx = 1$ (normalization)

we have

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

Using the Schwarz inequality

$$\left| \int f(x)^* g(x)\, dx \right|^2 \leq \int f(x)^* f(x)\, dx \int g^*(x) g(x)\, dx$$

with equality iff $g(x) = A f(x)$ (A = constant,

we obtain

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

2 cont.

Using the definition of the Fisher information

and the mean square error $\varepsilon^2 \triangleq \int (\hat{\theta}(x) - \theta)^2 p(x|\theta) dx$ ← power 2!

we have

$$\varepsilon^2 \geq I_\theta^{-1}$$

The condition for an unbiased efficient estimator is found when the above inequality is an equality (ie $g(x) = A f(x)$)

In the above,

$$g(x) \triangleq \frac{\partial}{\partial \theta} \ln p(x|\theta) \cdot \sqrt{p(x|\theta)}.$$

$$f(x) \triangleq (\hat{\theta}(x) - \theta) \sqrt{p(x|\theta)}.$$

$$\therefore \boxed{\frac{\partial}{\partial \theta} \ln p(x|\theta) = A(\theta) (\hat{\theta}(x) - \theta).} \quad -①$$

To find $A(\theta)$, differentiate ①

$$\frac{\partial^2}{\partial \theta^2} \ln p(x|\theta) = - A(\theta) + (\hat{\theta}(x) - \theta) \frac{\partial A(\theta)}{\partial \theta}.$$

2 cont)

Taking expectations, we have

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

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

for an efficient unbiased estimator.

Integrating gives

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

where $\ln h(x)$ is an arbitrary function of x.

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

which is the Neyman Fisher factorization theorem and $T(x)$ is a sufficient statistic for θ.

3

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

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

$$c_1 \leftarrow p(c | m, D)$$

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 First part is book work. 40%

—//—

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 $\underline{d} = \underline{s} + \underline{w}$

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

B4 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

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

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

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

$$\therefore \frac{1}{\sigma^2} d^T \underline{H} \underline{\theta} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda'$$

$$\therefore \boxed{\frac{1}{\sigma^2} \sum d(n)(A + Bn) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda'}$$

60%