Module – 6.3

MAX LIKELIHOOD METHOD

S. Lakshmivarahan

School of Computer Science

University of Oklahoma

Norman, Ok – 73069, USA

varahan@ou.edu

- $z = Hx + v$ or $z = h(x) + v$
  - Assume x and v are not correlated
- $P(z|x)$ is the conditional distribution of z
- Likelihood function $L(x|z) = p(z|x)$

  function of x — dual interpretation

- Fisher's (1920) principle: given z, what is the value of x that minimize $p(z|x)$
  - (ie) Maximize the prob. of observing the sample z

THE ML METHOD

- z = h(x) + v
- E(v) = 0, E(x^Tv) = 0, E(vv^T) = 0
- Let v ~ N(0, Σ) => p(z|x) ~ N(h(x), Σ)
- By def. $L(\hat{\mathbf{x}}_{ML}|\mathbf{z}) \geq L(\hat{\mathbf{x}}|\mathbf{z})$ ← for any other estimate
- => $ln\ L(\hat{\mathbf{x}}_{ML}|\mathbf{z}) \geq ln\ L(\hat{\mathbf{x}}|\mathbf{z})$
- A necessary condition is:

$$\nabla_{\mathbf{x}}[ln\ L(\mathbf{x}|\mathbf{z})] = \frac{1}{L(\mathbf{x}|\mathbf{z})} \nabla_{\mathbf{x}}\ L(\mathbf{x}|\mathbf{z}) = 0$$

EXAMPLE 15.1.1

- $z = H\mu + v$, $\mu \in R$
- $H = [1, 1, \ldots\ldots, 1]^T$
- $z_i = \mu + v_i$, for $i=1, 2, \ldots\ldots, m$
- $E[vv^T] = N(0, \sigma^2)$, $R = \sigma^2 \mathrm{I}$
- $z = N(H\mu, \sigma^2 \mathrm{I})$
- $L(x|z) = p(z|x) = N(H\mu, \sigma^2 \mathrm{I})$

$$= (2\pi)^{-\frac{m}{2}} (\sigma^2)^{-\frac{m}{2}} exp\ [-\frac{1}{2\sigma^2} (\mathbf{z} - \mathbf{H}\mu)^T (\mathbf{z} - \mathbf{H}\mu)]$$

EXAMPLE 15.1.1 (CONT'D)

- => $0 = \nabla_{\mathbf{x}}\ ln\ L(\mathbf{x}|\mathbf{z}) = \begin{bmatrix} \frac{\partial ln\ L(\mathbf{x}|\mathbf{z})}{\partial \mu} \\ \\ \frac{\partial ln\ L(\mathbf{x}|\mathbf{z})}{\partial \sigma^2} \end{bmatrix}$

$$= \begin{bmatrix} \frac{1}{\sigma^2}\sum_{i=1}^{m}(z_i - \mu) \\ \\ -\frac{m}{2}(\frac{1}{\sigma^2}) + \frac{1}{2\sigma^4}\sum_{i=1}^{m}(z_i - \mu)^2 \end{bmatrix}$$

- $\therefore$ $\hat{\mu}_{ML} = \frac{1}{m}\sum_{i=1}^{m} z_i = \bar{z}$ and $\hat{\sigma}^2_{ML} = \frac{1}{m}\sum_{i=1}^{m}(z_i - \bar{z})^2$
  - Both are unbiased

CRAMER-RAO BOUND: (SECTION 13.2) (LLD (2006))

- $\ln L(x|z)$ – Log of likelihood function
- $\nabla_x^2[\ \ln L(x|z)\ ]$ – Hessian of the log of likelihood
- $I(x) = -E[\nabla_x^2 \ln L(x|z)]$ – Called Information matrix

  $= E[\nabla_x \ln L(x|z)]\ [\nabla_x \ln L(x|z)]^T$
- $\hat{x}$ be an estimate of x, then $\mathrm{cov}(\hat{x}|x) \geq I^{-1}(x)$
- Non-linear case:

  $l(x|z) = -(m/2)\log(2\pi) - (m/2)\ln(\sigma^2) - (1/2\sigma^2)(z - h(x))^T(z - h(x))$
- Solved iteratively

EXERCISES

- Verify that $L(\hat{x}_{ML}|z)$ is a maximum by computing second derivative.

REFERENCES

- J. L. Melsa and D. L. Cohn (1978) Decision and Estimation Theory *McGraw Hill*
- Also refer to chapter 15 in LLD (2006)