MALMQUIST BIAS

(also known as Eddington bias)

...... or Eddington's solution of Fredholm's integral equation of 1st kind

$$F(x) \;=\; \int U(x-z)\; K(z)\; dz$$

Expand the LH integral argument to give

$$U(x-z) \;=\; U(x) \;-\; U'(x)\; z \;+\; U''(x)\; \frac{z^2}{2!} \;-\; ......$$

Integrate term by term

$$F(x) \;=\; \sum_n \frac{\mu_n}{n!}\; U^{(n)}(x)(-1)^n$$

where μ_n are moments of integration kernel K. Now rewrite

$$U(x) \;=\; F(x) + \sum_n A_n F^{(n)}(x)$$

For a central Kernel (ie. $\mu_1 = 0$) and equating coefficients

$$U(x) \;=\; F(x) - \frac{\mu_2}{2!} F^{(2)}(x) + \frac{\mu_3}{3!} F^{(3)}(x) - \left[\frac{\mu_4}{4!} - \left(\frac{\mu_2}{2!}\right)^2\right] F^{(4)}(x) + ...$$

For a Gaussian kernel $\mu_{odd} = 0$, $\mu_2 = \sigma^2$, $\mu_4 = 3\sigma^4$, ... therefore

$$U(x) \;=\; F(x) - \frac{\sigma^2}{2} F^{(2)}(x) + \frac{\sigma^4}{8} F^{(4)}(x) + ...$$

and for say a luminosity function of the form eg. $N(m) = 10^{\alpha(m-m_o)}$

$$\frac{dN}{dm} \;=\; ln10\; \alpha\; N(m) \qquad\qquad \frac{d^2N}{dm^2} \;=\; (ln10)^2\; \alpha^2\; N(m)$$

$$N_{obs}(m) \;=\; N(m) + \frac{\sigma^2}{2}\; (ln10)^2\; \alpha^2\; N(m) \quad \text{equivalently} \quad \Delta m \;=\; -ln10\; \alpha\; \frac{\sigma^2}{2}$$

PRINCIPAL COMPONENTS ANALYSIS - PCA

Given a series of n-dimensional vectors $\mathbf{x}_k;\ k = 1, 2, ...m$ what is the optimal linear transformation to reduce the dimensionality of the data ? Define

$$\mathbf{x'}_k = \mathbf{x}_k - < \mathbf{x}_k >_k \qquad \mathbf{x'}_k \ = \ \sum_{j=1}^{p} a_{kj}\boldsymbol{\psi}_j \ + \ \boldsymbol{\epsilon}_j$$

where $< \boldsymbol{\epsilon}_k^{\tau}\ \boldsymbol{\epsilon}_k >_k$ is to be minimised subject to $|\boldsymbol{\psi}_j|^2 = 1$

$$\Rightarrow\ \boldsymbol{\psi}_j^{\tau}\boldsymbol{\psi}_k \ = \ \delta_{jk} \qquad \text{and} \quad \frac{1}{m}\sum_{jk} a_{jk}^2 \quad \text{maximised}$$

$$\Rightarrow\ a_{kj} \ = \ \boldsymbol{\psi}_j^{\tau}\ \mathbf{x'}_k \quad \text{and} \quad \frac{1}{m}\sum_{jk} a_{kj}^2 \ = \ \sum_{j} \boldsymbol{\psi}_j^{\tau}\ \mathbf{C}\ \boldsymbol{\psi}_j$$

where $\mathbf{C} = \frac{1}{m}\sum_{k=1}^{m}\ \mathbf{x'}_k\mathbf{x'}_k^{\tau}$ ie. $\mathbf{C}$ is symmetric and +ve definite.

$$\Rightarrow\ \mathbf{C}\ \boldsymbol{\psi}_j \ = \ \lambda_j\boldsymbol{\psi}_j \quad \text{and} \quad \frac{1}{m}\sum_{jk} a_{jk}^2 = \sum_{j=1}^{p} \lambda_j$$

Therefore sorting the eigenvectors of the data covariance matrix by eigenvalue defines the optimum compression/feature extraction scheme.

INDEPENDENT COMPONENT ANALYSIS - ICA

This is an alternative approach for identifying independent features (components) in the data but this time defined by the requirement that they are as statistically independent as possible. (ICA is closely related to *blind source separation* and *projection pursuit*.)

Start again from series of n-dimensional vectors $\mathbf{x}_k$, $k = 1, 2, ...m$ and define

$$\mathbf{x}'_k = \mathbf{x}_k - < \mathbf{x}_k >_k \qquad \mathbf{x}'_k = \sum_{j=1}^{m} a_{kj}\, \mathbf{s}_j$$

where $\mathbf{s}_j$ are the sought after independent components. Defining $\mathbf{X}$ and $\mathbf{S}$ as the matrices with column vectors $\mathbf{x_k}$, $\mathbf{s_k}$ respectively and $\mathbf{A}$ as the matrix with elements a_{kj} we have

$$\mathbf{X} = \mathbf{A}\ \mathbf{S} \qquad \mathbf{S} = \mathbf{A}^{-1}\ \mathbf{X} \qquad \mathbf{S} = \mathbf{W}\ \mathbf{X}$$

where the weight matrix $\mathbf{W}$ defines the independent components.

Independence already implies uncorrelated but ICA also aims to maximise the non-Gaussianity of the $\mathbf{s_k}$, which is equivalent to minimising the entropy of the distribution of the values of the components of $\mathbf{s_k}$, which in this case is also equivalent to minimising the mutual information of the vectors $\mathbf{s_k}$.

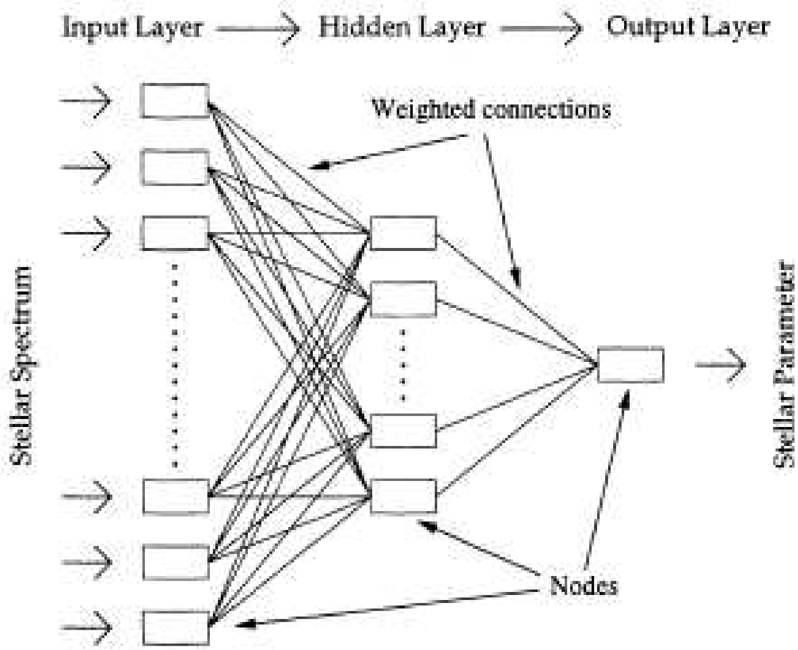
The simplest algorithm is FastICA which solves for $\mathbf{s_k}$ one at a time using a fixed point iteration scheme (Hyvärinen & Oja 1997).

$$\mathbf{s}_k = \sum_{j=1}^{m} w_{kj}\, \mathbf{x_j} = \mathbf{w}_k^{\tau}\, \mathbf{X}$$

which in practice is done by minimising $< G(\mathbf{w}_k^{\tau}\, \mathbf{X}) >$ subject to $\mathbf{w}_k^{\tau}\, \mathbf{w}_k = 1$ where $G(u) = tanh(u)$!!

ARTIFICIAL NEURAL NETWORKS - ANNs

Used for feature extraction, classification, data compression, prediction .....



Input layer: $\mathbf{x}\{1 \rightarrow m\}$; hidden layer(s): $\mathbf{X}\{1 \rightarrow h\}$; output layer: $\mathbf{y}\{1 \rightarrow p\}$

$$\mathbf{y}_k = f_k(\sum_{i=1}^{h} w_{ki}\ \mathbf{X}_i + w_{k0}) \quad eg. \quad f(z) = \frac{1}{1+e^{-z}} \quad \text{sigmoid}$$

$$\mathbf{X}_k = g_k(\sum_{i=1}^{m} w'_{ki}\ \mathbf{x}_i + w'_{k0}) \qquad f(z) = g(z) \quad \text{common}$$

$$\text{minimise} \quad < \sum_i [y_i(t) - d_i(t)]^2 >_t$$

where t denotes training set and $d_i(t)$ desired outcome.

Solution → back propagation of errors (Werbos 1974)

$$\text{output units} \ \ \epsilon_j = (d_j - y_j)\ y_j\ (1 - y_j) \quad \text{(sigmoid function)}$$

$$\text{hidden units} \quad \epsilon'_j = y_j\ (1 - y_j)\ \sum_k w_{jk}\ \epsilon_k$$

$$\text{adjust weights iteratively} \quad \Delta w(t)_{ij} = \eta\ \epsilon_j\ y_i + \alpha\ \Delta w(t-1)_{ij}$$

loop through entire training set → $n_{loop} >> 1$.

GENETIC ALGORITHMS

Generally used for NP hard problems ie. $\neq\ N^2,\ NlnN,\ N^3$ ....
but more of the variety, no. of solutions = $N!,\ N^M$....... ie. solution space is combinatorial or has complex topology.

Examples include: scheduling timetables, airline routes, travelling salesman-type problems, fiber configuration, χ^2 template minimisation ......

1. devise gene-like encoding scheme for parameters of interest (N_{gene})
2. randomly generate large nos. of trial solutions (eg. $N_{trial} = 1000+$)
3. devise a fitness score (0–1) to quantify them (eg. constraints, χ^2)
4. breed new offspring solutions $\propto$ fitness $P_{crossover} = 0.5 - 1.0, \quad P_{where}$
5. allow "genetic" mutations in offspring $P_{mutate} \approx \frac{1}{N_{trial}N_{gene}}$
6. test new generation and rescale fitness score to range 0–1
7. test convergence, end, or repeat from 4.

OUTLINE PROOF – MAXIMUM LIKELIHOOD METHOD

The likelihood is the probability of observing a particular dataset, therefore

$$\int L(x \mid \theta)\ dx = 1$$

differentiate with respect to θ

$$\int \frac{\partial L}{\partial \theta}\ dx \ = \ 0 \ = \ \int \frac{1}{L}\ \frac{\partial L}{\partial \theta}\ .L\ dx \ = \ \int \frac{\partial ln(L)}{\partial \theta}\ .L\ dx$$

differentiate RH term with respect to θ again

$$\int \frac{\partial^2 ln(L)}{\partial \theta^2}\ .L + \left(\frac{\partial ln(L)}{\partial \theta} \right)^2 .L\ dx = 0$$

therefore

$$\left\langle\ - \frac{\partial^2 ln(L)}{\partial \theta^2}\ \right\rangle \ = \ \left\langle\ \left(\frac{\partial ln(L)}{\partial \theta} \right)^2\ \right\rangle$$

Let t be an unbiased estimator of some function of θ, say $\tau(\theta)$, then

$$< \ t \ > = \int t\ L\ dx \ = \ \tau(\theta)$$

$$\tau'(\theta) \ = \ \frac{\partial \tau(\theta)}{\partial \theta} \ = \ \int t\ \frac{\partial ln(L)}{\partial \theta}\ L\ dx$$

therefore from above

$$\tau'(\theta) \ = \ \int (t - \tau(\theta))\ \frac{\partial ln(L)}{\partial \theta}\ L\ dx$$

Use Schwarz inequality on $\tau'^{\ 2}$ to generate

$$\tau'^{\ 2} \ \leq \int (t - \tau)^2\ L\ dx \ \times \ \int \left(\frac{\partial ln(L)}{\partial \theta} \right)^2 L\ dx$$

Therefore, for the case $\tau(\theta) = \theta$

$$var\{t\} \geq \frac{1}{\left\langle\ \left(\frac{\partial ln(L)}{\partial \theta} \right)^2\ \right\rangle} = \frac{1}{\left\langle\ - \frac{\partial^2 ln(L)}{\partial \theta^2}\ \right\rangle}$$

WORKED EXAMPLES

What are the correct ± 1-σ error bars to use for a Poisson distribution eg. number density of objects in various parameter ranges ?

Observe N objects in the interval $\Delta\Omega$ consider the MLE of the model density parameter ϕ.

$$Poisson \;=\; P(N|\phi) \;=\; \frac{(\phi\Delta\Omega)^N}{N!}\; e^{-\phi\Delta\Omega}$$

$$lnL(\phi) \;=\; -\phi\Delta\Omega + N\; ln(\phi\Delta\Omega) - lnN! \qquad \Rightarrow\; \hat{\phi} \;=\; \frac{N}{\Delta\Omega}$$

Error on estimate $\pm p\sigma$ when $lnL = lnL_{max} - \frac{1}{2} \times p^2$, substitute for $\hat{\phi}$

$$lnL(\phi) \;=\; -\frac{\phi}{\hat{\phi}}\; N + N\; ln(\frac{\phi}{\hat{\phi}}\; N) - lnN!$$

$$1 + \frac{p^2}{2N} \;=\; \frac{\phi}{\hat{\phi}} - ln(\frac{\phi}{\hat{\phi}})$$

For $N = 1$ the 1-σ range is $0.3 < \phi/\hat{\phi} < 2.4$

In the limit of large N let $\phi/\hat{\phi} = 1 + \epsilon$, then

$$1 + \frac{1}{2N} \;=\; 1 + \frac{\epsilon^2}{2} - \frac{\epsilon^3}{3} + ... \Rightarrow Lim_{\;N\rightarrow\infty}\;\; \epsilon = \pm\frac{1}{\sqrt{N}}$$

What is the optimum aperture to use for photometry of radially symmetric Gaussian, Exponential and Moffat profile images ?

$$Gaussian = I(r) = \frac{I_{tot}}{2\pi\sigma_G^2}\ e^{-r^2/2\sigma_G^2} \qquad FWHM = 2\sigma_G\sqrt{2ln(2)}$$

$$I(<R) = I_{tot}\ (1 - e^{-R^2/2\sigma_G^2}) \qquad MVB = \sqrt{4\pi\sigma_G^2}\ \sigma_{noise}$$

$$Efficiency\ \ = \ \ \sqrt{\frac{4\sigma_G^2}{R^2}}\ (1 - e^{-R^2/2\sigma_G^2})$$

$$Exponential = I(r) = \frac{I_{tot}}{2\pi a^2}\ e^{-r/a} \qquad FWHM = 2a\ ln(2)$$

$$I(<R) = I_{tot}\ (1 - e^{-R/a} - R/ae^{-R/a}) \qquad MVB = \sqrt{8\pi a^2}\ \sigma_{noise}$$

$$Efficiency\ \ = \ \ \sqrt{\frac{8a}{R^2}}\ (1 - e^{-R/a}[1 + r/a])$$

$$Moffat = I(r) = I_o\ [1 + (r/\alpha)^2]^{-\beta} \qquad FWHM = 2\alpha\ \sqrt{2^{1/\beta} - 1}$$

$$I(<R) = \frac{\pi\alpha^2}{\beta - 1}\ I_o\ \{1 - [1 + (R/\alpha)^2]^{-\beta+1}\} \qquad MVB = \sqrt{\frac{\pi\alpha^2(2\beta - 1)}{(\beta - 1)^2}}\ \sigma_{noise}$$

$$Efficiency\ \ = \ \ \sqrt{\frac{\alpha^2(2\beta - 1)}{R^2(\beta - 1)^2}}\ \{1 - [1 + (R/\alpha)^2]^{-\beta+1}\}$$