TESTS FREE OF BINNING

Test hypothesis H_0 that a series of observations $\{x_i\}$ have been drawn from a population with a given cumulative distribution $F(x)$. Pdf of the sample is $F^*(x) = \nu/n$ where ν is number of sample values $\leq x$.

Smirnov-Cramer-von Mises test considers

$$w^2 = \int_{-\infty}^{\infty} [F^*(x) - F(x)]^2 \, dF(x)$$

Frequency ratio $F^*(x)$ satifies $< F^* >= F$ and $< (F^* - F)^2 >= F(1 - F)/N$

$$< w^2 >= \frac{1}{6N} \qquad var\{w^2\} = \frac{4N-3}{180N^3}$$

ie. the distribution of w^2 is independent of F.

Kolmogorov- Smirnov test considers maximum deviation of $F^*(x)$ from $F(x)$

$$D_N = max \mid F^*(x) - F(x) \mid \ _x$$

or maximum ± deviation

$$lim_{N\rightarrow\infty} \ \ P(\sqrt{N} D_N > z) = 2 \sum_{r=1}^{\infty} (-1)^{r-1} exp(-2r^2z^2)$$

For small N known tabulated distribution independent of F.

For comparing two distributions M, N observations

$$D_{MN} = max \mid F_1^*(x) - F_2^*(x) \mid \ _x \qquad \sqrt{\frac{MN}{M+N}} D_{MN}$$

MANN–WHITNEY U TEST (WILCOXON)

Powerful non-parametric test to decide if pairs of observations

$\{x_i\}, i = 1, 2, ...m \quad \{y_i\}, i = 1, 2, ...n$

are drawn from same population.

H_0: samples X, Y are drawn from same distribution; H_1: X is stochastically larger than Y, ie. directional hypothesis.

Order the observations $\quad x_1, x_2, y_1, x_3, y_2 .... x_m, y_n$
assign running rank $\quad$ 1 2 3 4 j m+n

Construct the test statistic, U

$$U = mn + \frac{m(m+1)}{2} - \sum_{j \,\epsilon\, m} j$$

$$or = mn + \frac{n(n+1)}{2} - \sum_{j \,\epsilon\, n} j$$

For $m, n \lesssim 20$ use Tables to look up exact distribution, otherwise for large m, n

$$P(U) = N(\mu, \sigma^2) \qquad z = \frac{U - \mu}{\sigma}$$

$$\mu = \frac{mn}{2}$$

$$\sigma^2 = \frac{mn(m+n+1)}{12}$$

For example: U test is 95% efficiency of T test for comparing means of two Gaussian distributions.

BAYES THEOREM & OCCAM'S RAZOR

Bayesian viewpoint of parameter estimation for a particular hypothesis/model

$$P(\theta\ |data, M) \ = \ \frac{P(data\ |\theta, M)\ P(\theta|M)}{\int P(data\ |\theta, M)\ P(\theta|M)\ d\theta}$$

where the term in the denominator is known as the Bayesian evidence.

Now consider a Bayesian view of testing different models M_1, M_2

$$P(M_1\ |data) \ = \ \frac{P(data\ |M_1)\ P(M_1)}{P(data)}$$

$$P(M_2\ |data) \ = \ \frac{P(data\ |M_2)\ P(M_2)}{P(data)}$$

$$\frac{P(M_1\ |data)}{P(M_2\ |data)} \geq \ C_\alpha \quad \equiv \frac{P(data\ |M_1)\ P(M_1)}{P(data\ |M_2)\ P(M_2)} \quad \equiv \frac{P(data, M_1)}{P(data, M_2)}$$

The key feature here is the combination of the ratio of the global likelihoods of the models (Bayes factor) and the ratio of the model prior odds (the last part of the RHS is the ratio of joint PDFs of data and model).

The global likelihood of the model is given by

$$P(data\ |M) \ = \ \int P(data\ |\theta, M)\ P(\theta|M) d\theta$$

which is none other than the Bayesian Evidence *i.e.* the average of the likelihood weighted by the model parameters prior PDF.

If the model priors favours neither M_1 nor M_2 then Bayesian hypothesis/model reduces to examining the ratio of the global likelihoods.

Note that this likelihood ratio favours simpler hypotheses since in general they will make a more precise prediction as $P(data\ |\theta, M)\ P(\theta|M)$ is not so spread out over parameter space $\Rightarrow$ Occam's Razor.

DIGITAL FILTERING

Aim: reduce "noise" and keep signal ≈ same

Linear filters $\quad \hat{L}(A+B) = \hat{L}(A) + \hat{L}(B)$

Non-recursive filters take the form $y_l = \sum_k a_k x_{l-k}$

Recursive filters feedback the previous output $y_l = \sum_j b_j y_{l-j}$ and in 1-D are causal if $j > 0$.

$$y_l = \sum_k a_k \; x_{l-k} + \sum_j b_j \; y_{l-j} \qquad\qquad Y(\omega) = \frac{A(\omega)\; X(\omega)}{1 - B(\omega)}$$

For non-recursive filters (common) constraint $\sum_k a_k = 1$

$\Rightarrow$ constant signal level and a random noise reduction of $\sum_k a_k^2$

In 1-D timeseries analysis <u>z-transforms</u> are used to design and implement digital filters – z is a complex variable

$$S(z) \;=\; \sum_k s_k z^k \qquad s_k \;=\; \oint_c S(z) z^{-k-1} dz$$

ARMA (AutoRegressive Moving Average) process is most popular

1-D model eg. modelling speech, linear prediction, properties →

$$Y(z) \;=\; \frac{A(z)}{1 - B(z)}\, X(z) \;=\; \frac{\Pi_k (z - A_k)}{\Pi_j (z - B_j)}\, X(z)$$

WIENER FILTERING

Observed data $d(x) = s(x) + n(x)$ ie. signal plus random noise component.

In Fourier domain $D(w) = S(w) + N(w)$.

Is there a "best" linear filter such that $Y(w) = H(w)\ D(w)$ is optimal ?

In data domain $y(x) = h(x) \otimes d(x)$ and usually $h(x)$ symmetric about $x = 0$

$\Rightarrow$ no shift and $H(w)$ real.

$$minimise < [s(x) - h(x) \bigotimes d(x)]^2 > \qquad < [S(w) - H(w)D(w)]^2 >$$

$$\hat{H}(w) \ = \ \frac{< |S(w)|^2 >}{< |D(w)|^2 >} \ = \ \frac{< |S(w)|^2 >}{< |S(w)|^2 + |N(w)|^2 >}$$

Generalise to a deconvolution problem and ask the same question.

$$d(x) = s(x) \bigotimes b(x) + n(x) \qquad D(w) = S(w)B(w) + N(w)$$

where $b(x)$ is the blurring function. The solution is

$$\hat{H}(w) \ = \ \frac{B^*(w) < |S(w)|^2 >}{|B(w)|^2 < |S(w)|^2 + \lambda |N(w)|^2 >}$$

A parametric Wiener filter, λ, derived from a constraint on the variance of the output. Note that for zero-noise $\hat{H}(w) = B^{-1}(w)$ the Inverse filter.

ENTROPY AS A MEASURE OF INFORMATION

Shannon & Weaver (The Mathmatical Theory of Communication – 1949) introduced the concept of entropy as a measure of information.

Given a signal source with N possible outputs probability P_i define the information gain $\Delta I_i = h(P_i)$ as a monotonically decreasing function of P_i

$$\Delta I_{i,j} = h(P_i) + h(P_j) = h(P_i P_j) \quad \Rightarrow \; \Delta I_i = -lnP_i$$

Hence the average or expected information is

$$< \Delta I_i >= \; H \; = \; -\sum_{i=1}^{N} \; P_i \; lnP_i$$

The entropy $H \equiv$ average <u>a priori uncertainty</u> regarding the source and is a measure of the channel capacity.

Jaynes (1957) → entropy as a measure of randomness of distribution cf. statistical thermodynamics.

$$H = -\sum_{i=1}^{N} P_i \; lnP_i \qquad H = -\int \; P(x) \; lnP(x) \; dx$$

Maximum Entropy Principle: choose PDF with maximum entropy subject to whatever constraints apply.

Example I – Dice problem

$$H \ = \ -\sum_{i=1}^{6} P_i \ lnP_i \quad constraints \ \ \sum_i \ P_i \ = \ 1 \ ; \ \sum_i \ i \ P_i \ = \ 2$$

$$max \ H \ \Rightarrow \ P_j \ = \ \frac{e^{-j\lambda}}{z} \ ; \ z \ = \ \sum_j e^{-j\lambda}$$

NB. constraint recovered by $-\partial lnz/\partial\lambda = 2 = \sum_j \ j \ P_j$

P_1	P_2	P_3	P_4	P_5	P_6
0.51	0.25	0.13	0.06	0.03	0.02

Example II – Gaussian Distribution

Series of constraints of the form $< f_k(x) >$ for $k = 1, 2, ...m$

Maximum Entropy solution defined by

$$P(x) = \frac{e^{-\sum_k f_k(x)\lambda_k}}{z} \qquad\qquad z = \int_{-\infty}^{\infty} e^{-\sum_k f_k(x)\lambda_k} \ dx$$

$$< f_k(x) > = \ -\frac{\partial}{\partial\lambda_k} lnz(\lambda_1, \lambda_2, ...\lambda_m)$$

Eg. constraints $< x >= \mu$ and $< x^2 >= \mu^2 + \sigma^2$

$$P(x) = \frac{e^{-\lambda_1 x - \lambda_2 x^2}}{z} \qquad \rightarrow Gaussian$$

$$H_{max} = \frac{1}{2} + ln\sqrt{2\pi\sigma^2}$$

IMAGE RESTORATION

Problem:- estimate underlying image pixel intensities (ie. parameters θ_i) from observations $\{d_i, i = 1, 2, ...n\}$ in the presence of various types of image degradation (eg. blurring $\otimes b$) and additive "measurement" noise ϵ_i

$$d_i = f(\theta) + \epsilon_i \qquad eg.\ \ d_i = \theta_i \bigotimes b + \epsilon_i$$

Considering the distribution of pixel intensity parameters θ_i as equivalent to state space variables in statisical thermodynamics suggests

$$maximise\ entropy = -\sum_i \theta_i\ ln\ \theta_i \quad \Leftarrow \quad \sum_i \theta_i = total\ flux; \ \ \sum_i \chi^2 = N$$

Surprisingly $\partial/\partial\theta_j$ yields a simple iterative MaxEnt scheme

$$\theta_j \ = \ exp\ \{\lambda\ b \bigotimes [d_j - \theta_j \bigotimes b]\}$$

From a Bayesian viewpoint image restoration translates to maximising

$$P(\theta, M|D) \ = \ P(D|\theta, M)\ P(\theta|M)\ P(M)\ /\ P(D)$$

with respect to model M with parameters θ.

$P(D|\theta, M)$ is usually taken as the GoF measure χ^2 and the prior

$$P(\theta|M) \ = \ \frac{N!}{m^N \prod_{i=1}^m \ \theta_i!} \ = \ e^{\alpha S}$$

where m is the no. of parameters "pixels" and N total counts for image.

Note that for θ_i large (usually the case) using Stirling's approximation gives

$$ln\ P(\theta|M) \ = \ ln\ N! - N\ ln\ m - \sum_i ln\ \theta_i! \ = \ const - \sum_i \ \theta_i\ ln\ \theta_i \Rightarrow S = entropy$$

and ignoring constants the MAP estimator is equivalent to maximising

$$MAP = e^{-\chi^2/2}\ e^{\alpha s}$$