WORKED EXAMPLES

For a series of quasar spectra, model the number of LLS as no. per unit redshift $f(z) = N_o(1+z)^\gamma$, find the MLE for N_o and γ.

For each quasar there is a maximum redshift z_{max} for detection of a LLS, which we can take to be the redshift of the quasar z_{em}, and a minimum redshift z_{min} for which LLS are observable. This latter redshift either corresponds to the blue cutoff of the quasar spectrum z_{minobs} or to a higher redshift LLS cutting out the flux ie. $z_{min} = max\{z_{minobs}, z_{lls}\}$.

$$ln(L) = -\sum_{j=1}^{n} \{\int_{z_{minj}}^{z_{maxj}} f(z)\ dz\} \ + \ \sum_{i=1}^{m} ln\ [f(z_i)]$$

$$ln(L) = \sum_{j=1}^{n} \frac{N_o}{1-\gamma} \left[(1+z)^{-\gamma+1}\right]_{z_{minj}}^{z_{maxj}} \ + \ \sum_{i=1}^{m} ln\ [N_o\ (1+z_i)^{-\gamma}]$$

Note the general form for likelihood problems such as this usually involves a normalising constant η (ie. no. per unit "volume" of parameter space). Explicitly write the ML model for the predicted distribution as $\eta\ \Phi(\underline{\theta})$ then

$$ln(L) \ = \ \sum_{j=1}^{n} ln(L_j) \ = \ -\sum_{j=1}^{n} \int \eta\ \Phi(\underline{\theta})_j\ d\underline{\theta} + \sum_{i=1}^{m} ln[\eta\ \Phi(\underline{\theta})_i]$$

$$\frac{\partial ln(L)}{\partial \eta} = \ -\sum_{j=1}^{n} \int\ \Phi(\underline{\theta})_j\ d\underline{\theta} \ + \ \sum_{i=1}^{m} \frac{1}{\eta} \ = \ 0$$

$$\Rightarrow \quad \eta \ = \ m \ / \ \sum_{j=1}^{n} \int\ \Phi(\underline{\theta})_j\ d\underline{\theta}$$

You have two data frames taken on different nights at almost the same orientation and position and with almost the same pixel scale, and you want to match the coordinate systems, (x', y') (x, y), prior to combining them.

$$
\begin{aligned}
x' = \; a\; x \; + \; b\; y \; + \; c \qquad &\equiv \qquad a\; cos\theta\; x \; + \; a\; sin\theta\; y \; + \; c \\
y' = \; d\; x \; + \; e\; y \; + \; f \qquad &\equiv \quad -a\; sin\theta\; x \; + \; a\; cos\theta\; y \; + \; f
\end{aligned}
$$

Pair up the objects from the frames $\{x'_i, y'_i\} \; \leftrightarrow \; \{x_i, y_i\}$ solve using linear least-squares.

The Information matrix for $[a, b, c]$ in this case is given by

$$
I \;\; = \;\; < \left[\begin{array}{ccc}
\Sigma_i \;\; x_i^2/\sigma_i^2 & \Sigma_i \;\; x_i y_i/\sigma_i^2 & \Sigma_i \;\; x_i/\sigma_i^2 \\
\Sigma_i \;\; x_i y_i/\sigma_i^2 & \Sigma_i \;\; y_i^2/\sigma_i^2 & \Sigma_i \;\; y_i/\sigma_i^2 \\
\Sigma_i \;\; x_i/\sigma_i^2 & \Sigma_i \;\; y_i/\sigma_i^2 & \Sigma_i \;\; 1/\sigma_i^2
\end{array} \right] >
$$

For a random distribution of N points $0 < x_i < L_x$; $0 < y_i < L_y$; $\sigma_i = \sigma$

$$
V \;\; = \;\; I^{-1} \;\; = \;\; \frac{\sigma^2}{N} \left[\begin{array}{ccc}
1/3 \; L_x^2 & 1/4 \; L_x L_y & 1/2 \; L_x \\
1/4 \; L_x L_y & 1/3 \; L_y^2 & 1/2 \; L_y \\
1/2 \; L_x & 1/2 \; L_y & 1
\end{array} \right]^{-1}
$$

LIKELIHOOD & CONFIDENCE INTERVALS

Interval estimation $\rightarrow$ find range $\theta_a \leq \theta \leq \theta_b$ which contains true value θ_o with probability β, or alternatively has probability $\alpha = 1 - \beta$ of lying outside the range.

Confidence interval $\equiv$ error of the parameter θ

For example, 1-D Gaussian parameter error distribution
$\pm 1\sigma$ $\beta = 68.3\%$, for $\pm 2\sigma$ β =95.5%

In particular Maximum Likelihood estimators lead to this type of parameter error distribution for large N and/or exponential family PDFs – eg. in 1-D

$$lnL(\bar{x} \mid \mu) = const - \sum_i (x_i - \mu)/2\sigma_i^2$$

$$\sigma_i = \sigma \qquad lnL = const - (\mu - \bar{x})^2 / \frac{2\sigma^2}{N}$$

Now consider a 2-D Gaussian parameter error distribution what is the equivalent confidence interval for $\pm p\sigma$ errors ?

First normalise the distribution to N(0,1) form on both axes

$$\beta \;=\; \frac{1}{2\pi} \int_{s_x} \int_{s_y} e^{-(x^2+y^2)/2}\, dxdy \;=\; \int_0^p \; re^{-r^2/2}\, dr$$

Hence $\beta = 1 - e^{-p^2/2}$ and $1\sigma \rightarrow \beta = 39.3\%$, for $2\sigma \rightarrow \beta$ =86.5%

To cover 68.3% and 95.5% interval requires 1.52σ and 2.49σ ??

To generalise to many parameters, note CLT $\Rightarrow$ distribution multivariate Gaussian for likelihood estimator for large N

Defined by mean vector θ and covariance matrix C

$$P(\hat{\theta}) = \frac{1}{(2\pi)^{m/2} \mid C \mid^{1/2}} \; exp - \frac{1}{2}(\hat{\theta} - \theta)^{\tau} C^{-1} (\hat{\theta} - \theta)$$

The corresponding confidence interval for normalised σ is

$$\beta = \frac{1}{2^{m/2-1}(m/2-1)!} \int_0^p X^{m-1} e^{-X^2/2} dX$$

INVARIANCE OF ML ESTIMATE

- if $\hat{\theta}$ is ML estimate then $\tau(\hat{\theta})$ is also
- only one function $\tau(\theta)$ will be unbiased for finite N
- $\Rightarrow$ transform [in principle] any-shaped lnL to parabolic shape
- transform method exact only to $O(\frac{1}{N})$ since experimental likelihood is made normal not theoretical distribution
- interval obtained is central in transformed variable not original
- multiple maxima cause problems

HYPOTHESIS TESTING

Often set up two possible hypotheses, H_0 – the null hypothesis, usually formulated to be rejected; and H_1 the alternative, or reseach hypothesis – simple or composite

Reject H_0 if test yields value of statistic whose probability of occurrence under H_0 is $\leq \alpha =$ the level of significance.

When H_0 is true:

$$Type\ I\ error = P(rejection\ of\ H_0) = \alpha$$

When H_0 is false:

$$Type\ II\ error = P(failure\ to\ reject\ H_0) = \beta$$

$$Power = P(rejection\ H_0) = 1 - \beta$$

Problem = form of the sampling distribution of the test–statistic

Parametric tests eg. Students' t, Fisher F, χ^2 assume: normally distributed populations; observations independent; measured on a viable scale

Non-parametric tests make fewer assumptions concerning data; underlying distribution can be unknown; small sample size; cope with measures such as ranks; can treat multiple population types – lack of power in special cases.

NEYMAN-PEARSON LEMMA

In hypothesis testing minimise a type II error (failure to reject hypothesis) by choosing the boundary of the acceptance region such that

$$likelihood\ ratio \ = \ \frac{L(X|H_1)}{L(X|H_0)} \ = \ C_\alpha$$

If w_α represents some enclosed "volume" of the data space

$$Power \ = \ 1-\beta \ = \ \int_{w_\alpha} \ L(X|\theta_1)\ dX$$

to find the region w_α that maximises $1-\beta$ consider

$$1-\beta \ = \ \int_{w_\alpha} \ \frac{L(X|\theta_1)}{L(X|\theta_0)}\ L(X|\theta_0)\ dX \ = < \frac{L(X|\theta_1)}{L(X|\theta_0)} >$$

Therefore the best critical region w_α consists of points having the largest values of $L(X|\theta_1)/L(X|\theta_0)$ ie. $\geq C_\alpha$

- simple hypotheses have fixed parameters postulated prior to the experiment and lead to exact distributions for the test statistic
- composite hypotheses use a continous family of parameters but since the pdf of the likelihood is the same form can compare eg. s parameters -v- r parameters, but involve correction factors $\approx 1 + a/N + O(1/N^2)$ since estimate the parameters from the data
- for composite hypotheses belonging to separate families the above applies but also have to construct a comprehensive parametric family eg. $\lambda f(X|\theta) + (1-\lambda) g(X|\psi)$ and test this -v- one or other of the individual hypotheses

SUMMARY OF HYPOTHESIS TESTS

Parametric tests – see for example Kendall & Stuart "Advanced Theory Of Statistics"

Non-parametric – best bet is Siegel & Castellan "Non-parametric Statistics for the Behavioral Sciences"

- Student's t – observed mean consistent with theoretical value ?
- Fisher's F-test – compares variances of two distributions
- Correlation coefficient – are two or more variables correlated ?
- Pearson χ^2 – used for comparing general distributions/models
- Likelihood ratio – general use in hypothesis testing H_0 -v- H_1
- Mann-Whitney U – general comparison of distributions
- Cramer-Smirnov-von Mises – general comparison of distributions
- Kolmogorov-Smirnov (KS) – general comparison of distributions
- Spearman (+Kendall) rank correlation coefficent – correlation ?
- Monte Carlo methods – lots of CPU test for almost anything

SPEARMAN RANK CORRELATION COEFFICIENT

Are two sets of variables correlated ?

Rank observations $\{x_i\}$ and $\{y_i\}$ separately noting paired order

The usual Pearson correlation coefficient for mean–corrected x, y is

$$r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$$

For ranks

$$\sum_i x_i = N(N+1)/2 \qquad \sum_i x_i^2 = N(N+1)(2N+1)/6$$

$$\Rightarrow \sum_i (x_i - \bar{x}_i)^2 = (N^3 - N)/12$$

Construct the following using $d = x - y$ (mean–corrected)

$$r_s = \sum_i x_i^2 + \sum_i y_i^2 - \sum_i d_i^2 / 2\sqrt{\sum x_i^2 \sum y_i^2}$$

$$r_s = 1 - 6\sum_i d_i^2/(N^3 - N) \qquad -1 \le r_s \le 1$$

For $N \ge 10$

$$t = r_s \sqrt{\frac{N-2}{1-r_s^2}}$$

is distributed as Students' t with $N-2$ degrees of freedom, which in turn is distributed as N(0,1) as $N >> 1$.

PEARSON'S CHI-SQUARE TEST

Test on binned data and as such is distribution–free from original

H_0: the proportion of "objects" in each bin is that "expected" from either a model or the presumed population.

The chi–square statistic is

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

with $n - 1$ degrees of freedom.
The modified chi–square $E_i \to O_i$ is not $P(\chi^2)$.

Choose bin size by redefining ordinate such that probability content of bins $\approx$ same.

Advantages: easy to understand, model testing is for free since can vary the model parameters to turn testing to model fitting.
Disadvantages: binning loses power, unstable for ≤ 5 counts per bin.

Note that for model fitting m parameters on naturally sampled (binned) data n samples then

$$\chi^2_{n-m} = \sum_i (d_i - m_i)^2/\sigma_i^2$$

$$< \chi^2 >= n - m \qquad var\{\chi^2\} = 2(n - m)$$

then known as the Goodness-of-FIT (GoF) test.

NEYMAN-PEARSON TEST

.....otherwise known as the likelihood ratio test.

Can either be two competing hypotheses H_0 -v- H_1 or $goodness-of-fit$ testing H_0 -v- *everything* (eg. parameter estimation, number of model parameters....)

$$\lambda = \frac{L(data \mid H_0)}{L(data \mid H_1)}$$

Suppose H_0 has r free parameters fit from data and H_1 has s free parameters where $s > r$, then in the asymptotic limit of large N $-2ln(\lambda)$ is distributed as χ^2_{s-r} to $O(1/N)$

$$< -2ln(\lambda) >= (s-r)(1 + \frac{a}{N} + ...)$$

Note that for model fitting with Gaussian errors the above is exact for all N since

$$-2ln(\lambda) = \chi^2_s - \chi^2_r$$

This gives a simple recipe for deciding if extra parameters are "necessary"

$$\chi^2_{n-m+1} - \chi^2_{n-m} \qquad \chi^2_1 \;\; distribution$$

$$t^2 = \frac{\chi^2_{n-m+1} - \chi^2_{n-m}}{\chi^2_{n-m}}(N-m-1) \qquad Students't$$