Tina Memo No. 2013-006
Published in Annals of the BMVA 2014 Vol 1

Linear Poisson Models: A Pattern Recognition Solution to the Histogram Composition Problem

P. Tar, N.A. Thacker.

Last updated
15 / 2 / 2014

ISBE, Medical School,
University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT, UK.

Linear Poisson Models: A Pattern Recognition Solution to the Histogram Composition Problem

Foreword

I believe the following is a quite familiar scenario which many scientists will recognise;

urged on by the enthusiasm of other researchers for this new apporach to data analysis, you have decided that Pattern Recogntion (PR, training a classifier using sample data) is the way to solve a difficult problem and (using open software) have quickly extracted your first ROC curve. You now wish to use this system to analyse new data and estimate how many of a limited number of possible classes are present. Ideally, as you are doing science, you would like to compare these measurements with a theoretical prediction and make some statement regarding conformity. However, as someone more used to conventional statistical methods you suspect there is a problem.

You realise that you don't understand how the PR system solved the problem (one motivation for using the appoach is that you would not need to), but as a consequence you cannot extrapolate the previous ROC result to new data. You now require either a detailed simulation or very large quantities of ground truth data in order to perform an error assessment, neither of which are available. Alternatively, what is needed is a way of keeping track of the effects of errors present during training and also those inherrent to any sample later being measured, so that the reliability of your quantity estimates can be estimated. But no such method currently exists, and if you ask those who encouraged you to use PR you will probably be told that this can't be done. You may even be told that, given the assumptions underpinning PR, it should not even be attempted; the assumed data density distributions can never be expected to match new data.

This paper gives a method, designed for linear models with Poisson samples, which allows these sources of error to be propegated through to uncertainties on quantity estimates. Statistical tests are also provided which allow estimated models to be checked against data in order to help ensure that this process is valid. The overall result is that error bars can be constructed which allow the results of quantites estimated using PR to be used in scientific testing of theories. The approach contradicts the common assertion that PR is inevitably based on poor (and untestable) distribution assumptions and a version of probability theory which must remain non-quantitative.

The idea is relatively straight forward but ultimately quite involved. As a consequence this paper is the culmination of 4 years of sustained effort and forms the theoretical cornerstone of Paul Tar's PhD thesis. Previous work can be found in Tina Memos's 2010-008, 2011-003, 2011-004, 2012-003, but this is our first journal publication. We both hope that this demonstration of quantitative validity will help to encourage a more principled approach to machine learning which accords better with more conventional statistics. Enjoy.

Neil Thacker

Abstract

The use of histogram data is ubiquitous within the sciences and histograms composed of linear combinations of sub-processes are common. The statistical properties of recorded frequencies are well understood, i.e. Poisson statistics can model the probabilities of observing discrete events, from which an in-depth theoretical analysis of histogram components and their errors can be derived. Despite this, linear decomposition techniques such as Independent Component Analysis, Principal Component Analysis, and their variants, have largely focused on continuous data, with little or no account of error characteristics on estimated parameters. We argue that the properties of histograms, i.e. Poisson, non-negative discrete frequencies, requires a linear decomposition technique for making quantitative measurements. We also argue that a solution appropriate for scientific analysis tasks must involve a good understanding of noise so that uncertainties in measurements can be reported to end users. We present an approach to the analysis of histograms capable of summarising data in terms of Probability Mass Functions, weighting quantities and associated covariance matrices which we believe is suitable for scientific applications. The end result is a quantitative pattern recognition system capable of learning distributions from training data, then estimating the quantity of similar distributions in new incoming data.

1 Introduction

The need to process and understand multivariate data is a problem common to many fields. Disciplines including signal processing, neural networks, and the physical and biological sciences regularly require reduction techniques for representing, summarising and visualising complex data. A popular approach is to decompose data using linear transformations as found in methods including Principal Component Analysis (PCA) [15][16], Factor Analysis (FA) [12][16], Gaussian mixture modelling (GMM) [8] and Independent Component Analysis (ICA) [6]. Increasingly sophisticated methods have also been investigated such as Gaussian Process Latent Variable Models and its variants [14]. These techniques are primarily designed for continuous data and often make the assumption of a Gaussian residual distribution (noise) which is either uniform, isotropic or negligible. Some formulations of these problems, especially ICA, explicitly exclude consideration of data noise for simplicity and tractability [13].

Although the statistical definitions of histogram data are well known there are fewer linear decomposition techniques designed for them specifically. Histogram data has the properties of discrete, non-negative frequencies and non-uniform, non-isotropic distributed residuals. Whilst it is possible computationally to apply continuous techniques to histograms, it would be difficult to interpret results in a physically meaningful way. Results may include non-physical negative components and unknown destabilising effects due to propagated non-uniform noise for which no error theory exists. We believe such outputs are inappropriate for making scientific measurements where a good understanding of noise is required to provide estimates of uncertainties (error bars) allowing researchers to interpret data with confidence. Some nonparametric and semi-parametric regression methods [9][5] may be more applicable here, but these tend not to attribute physical quantitative meanings to parameters, whereas we aim to use model components as the basis of multi-class measurements, accompanied by error predictions, in a supervised learning scenario.

In this paper we show theoretically that quantitative measurements may be made of histogram components, such as tissue types in MRI or chemicals in mass spectra, by decomposing histograms in terms of probability mass functions (PMFs) of independent components, weighting quantities

(the measurements, i.e. how much of each component is in the data) and an error covariance matrix. Whilst the linear models we apply are relatively commonplace, our comprehensive estimation of errors is a novel contribution to this type of methodology. All aspects of the theory are corroborated in Monte-Carlo simulation using simple artificial distributions and also realistic distributions drawn from Martian terrain texture descriptor histograms. It must be emphasised that the focus of this paper is on making valid measurements from histograms within theoretically predictable accuracies. This is in contrast to other methods which focus on summarising data without any comprehensive error theories. As such, it is difficult to make direct performance comparisons between our approach and common alternatives, e.g. it is not possible to compare our error predictions with methods which do not produce their own error predictions. Instead we develop a predictive theory and design experiments to corroborate those predictions, consistent with the scientific method, as opposed to conducting empirical comparisons of alternative methods.

1.1 Formal Specification

The linear histogram composition problem can be compactly summaried as follows:

$$\mathbf{H} = \mathbf{P}\mathbf{Q} + \mathbf{e}_H \tag{1}$$

where $\mathbf{H} = \{H_{X=1}, H_{X=2}, \ldots, H_{X=m}\}^\intercal$ is a histogram as a vector with bin frequencies which will be denoted $\mathbf{H}_X$; $\mathbf{P}$ is an m by n matrix describing the PMFs of n components with elements $\mathbf{P}_{ij} = P(X = i|k = j)$, i.e. the probability of an entry in bin X given component k; $\mathbf{Q}$ is a column vector of n quantities corresponding to the amount of each component present within the histogram vector; and $\mathbf{e}_H$ is a column vector of noise caused by independent Poisson perturbations around the expected value of each $\mathbf{H}_X$. It is instructive to also consider the inverted problem:

$$\mathbf{Q} = \mathbf{P}^{-1}\mathbf{H} + \mathbf{e}_Q \tag{2}$$

where $\mathbf{P}^{-1}$ is an n by m matrix with elements consistent with Bayes Theorem, $\mathbf{P}^{-1}_{ij} = P(k = i|X = j)$, i.e. the probability that component k was the source of an entry in bin X; and $\mathbf{e}_Q$ is noise on the quantities which can be summarised with a covariance matrix $\mathbf{C}$.

It may be sufficient to attribute each linear histogram component uniquely to an underlying effect, e.g. one PMF per MRI tissue type. However, a class of physical process can be more complex requiring itself to be approximated by a set of linear sub-processes (as is commonly done in imaging techniques such as appearance modelling [7])). A histogram based upon image texture descriptors for example may model a single surface terrain type using several linear components approximating a more complex PMF, e.g. using multiple grey level co-occurrence histograms [10] to describe a single texture. Quantities associated with such processes can be measured by adding an additional layer to the histogram models which sums subsets of the n histogram components into z physical class sets:

$$\mathbb{Q} = \mathbb{K}\mathbf{Q} + \mathbf{e}_\mathbb{Q} \tag{3}$$

where $\mathbb{Q}$ is a column vector containing z quantities of the z physical generation processes for which measurements are sought; $\mathbb{K}$ is an o by n mapping matrix with mutually exclusive binary elements; and $\mathbf{e}_\mathbb{Q}$ is the noise on the total class quantities which can be summarised with the covariance matrix $\mathbf{C}_\mathbb{Q}$. The elements of $\mathbb{K}$ are then:

$$\mathbb{K}_{ij} = \delta(k_j \in K_i) \tag{4}$$

where k_j is the jth model component and K_i is the ith physical process for which quantity measurements are sought; and the function δ is 1 when component k is within the set K and 0 otherwise.

In the simplified case, when both $\mathbf{H}$ and $\mathbf{P}$ are known, a solution to the linear histogram composition problem involves estimating quantities, $\mathbf{Q}$, and an associated covariance, $\mathbf{C}$. In the more general case when a set of N histograms for each physical class is the only available information a solution involves estimating a set of quantities, covariances and a common definition of $\mathbf{P}$.

We show in the simplified case where $\mathbf{H}$ and $\mathbf{P}$ are known that Expectation Maximisation (EM) can be applied to solve an Extended Maximum Likelihood (EML) problem providing estimates of $\mathbf{Q}$ and $\mathbf{P}^{-1}$ (section 2.1). We also show in this case that the Cramer Rao Bound (CRB) can be applied to the EML function providing an estimate of covariance, $\mathbf{C}$, describing the noise, $\mathbf{e}_Q$, on estimated quantities (section 2.2). In the general case we provide an extension to the EM algorithm which can estimate the necessary sub-matrices of $\mathbf{P}$ from exemplar histograms containing examples of each physical class of process and show how a Chi-squared per degree of freedom test can be used in model selection to determine how many components are required to give good approximations to the exemplar histograms (section 2.3). In this latter case systematic errors caused by imprecise component determination (i.e. when $\mathbf{P}$ is estimated once with sampling errors then used multiple times) are accounted for in the covariance, $\mathbf{C}$, by applying error propagation, which also provides an alternative method for computing the statistical error equivalent to applying the CRB (section 2.4). Errors on total class quantities, $\mathbb{Q}$, are computed giving $\mathbf{C}_{\mathbb{Q}}$ using error propagation (section 2.5). Finally, an algorithm is outlined in pseudo-code to illustrate how the method should be used (section 2.6).

2 Methodology

Histograms are typically accumulated by counting discrete Poisson events which occur with some expected mean frequency per bin. The statistical modelling of this scenario is usually attributed to Fermi in the form of Extended Maximum Likelihood [4], which in terms of PMFs and quantities becomes:

$$\ln \mathcal{L} = \sum_X \ln \left[\sum_k P(X|k)\mathbf{Q}_k \right] \mathbf{H}_X - \sum_k \mathbf{Q}_k \tag{5}$$

where $P(X|k)$ (i.e. $\mathbf{P}$) and $\mathbf{Q}$ are parameters to be estimated.

2.1 Simple Case: Quantity estimation when component PMFs are known

In cases where $\mathbf{P}$ is known the EM algorithm can be used to optimise (5) with respect to free parameters, $\mathbf{Q}$. EM iteratively updates the elements of $\mathbf{Q}$ by weighting them with the current estimate of the posteriori probability $P(k|X)$, (i.e. $\mathbf{P}^{-1}$), starting from some initial (perhaps randomised) estimate at iteration $t = 0$:

$$\mathbf{Q}_{(t)} = \mathbf{P}^{-1}_{(t-1)}\mathbf{H} \tag{6}$$

where $\mathbf{Q}_{(t)}$ is the quantity vector estimate at time t; and $\mathbf{P}^{-1}_{(t-1)}$ is the last estimate of the posteriori probabilities. For a single quantity, k, this update is implemented using Bayes Theorem [18] [3]:

$$\mathbf{Q}_{k(t)} = \sum_X \frac{P(X|k)\mathbf{Q}_{k(t-1)}}{\sum_l P(X|l)\mathbf{Q}_{l(t-1)}}\mathbf{H}_X \tag{7}$$

where $P(X|k) = \mathbf{P}_{i=X,j=k}$; and the Bayesian 'prior' is substituted for the quantity being estimated. Upon convergence the quantities, $\hat{\mathbf{Q}} = \mathbf{Q}_{(t=\infty)}$, provide the maximum likelihood solutions assuming the PMFs of the model components were correct for the incoming histogram data.

2.2 Simple case: Covariance estimation when component PMFs are known

The accuracy of estimated quantities can be determined using the CRB on equation (5):

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \mathbf{Q}_i \partial \mathbf{Q}_j} \geq \mathbf{C}_{ij}^{-1} \tag{8}$$

giving (see appendix A for derivation):

$$\mathbf{C}_{ij}^{-1} \approx \frac{\sum_X P(i|X)P(j|X)\mathbf{H}_X}{\mathbf{Q}_i\mathbf{Q}_j} \tag{9}$$

where $\mathbf{C}_{ij}^{-1}$ is the inverse covariance between quantities $\mathbf{Q}_i$ and $\mathbf{Q}_j$. This assumes all noise is statistical, originating from the histograms under analysis and that the PMFs are correct, being free of any errors. From this the covariance matrix can be computed using a standard matrix inversion algorithm. Error predictions made using this method are corroborated in Monte-Carlo in section 3.

2.3 General Case: Estimating PMFs from multiple histograms

Given a source of data composed of n unknown linearly combined sub-histograms (with unknown PMFs), and given a set of N ($n \ll N$) independent exemplars, the EM algorithm can be extended to estimate the common PMFs of the constituent sub-histograms, i.e. perform a histogram ICA. The exemplar linear histograms can be viewed as being generated by:

$$\mathbf{H}_{X(r)} = \sum_{k=1}^{n} R_r(X|k) \tag{10}$$

where $\mathbf{H}_{X(r)}$ is the X bin of the rth independent example; and $R_r(X|k)$ is the sub-histogram generation process (distributed as $P(X|k)$, which will be estimated) contributing to the X bin in example r from unknown component k.

Initial estimates of component PMFs (perhaps randomised) are generated at iteration 0 giving $\mathbf{P}_{(t=0)}$. Weighting quantities are estimated (via estimation process of section 2.1) for each of the N examples giving modelled approximations:

$$\mathbf{H}_{(r)} \approx \mathbf{M}_{(r)(t)} = \mathbf{P}_{(t)}\mathbf{Q}_{(r)(t)} \tag{11}$$

where all histogram models (i.e. each r) share a common definition of $\mathbf{P}$; and each histogram has its own estimate of weighting quantities, $\mathbf{Q}_{(r)}$. In line with the EM algorithm the estimated posteriori probabilities, $\mathbf{P}^{-1}$, are used to provide a new estimate of the contribution to each histogram, r, from each component, k, i.e. an estimate of the underlying sub-histogram generators:

$$\hat{R}_{r(t)}(X|k) = P_{r(t-1)}(k|X)\mathbf{H}_{X(r)} \tag{12}$$

These independent estimates are combined and normalised to create a new common estimate of $\mathbf{P}_{(t)}$:

$$\mathbf{P}_{i=X,j=k,(t)} = P_{(t)}(X|k) = \frac{\sum_r \hat{R}_{r(t)}(X|k)}{\sum_r \mathbf{Q}_{k(r)(t)}} \tag{13}$$

This new common estimate of $\mathbf{P}$ is used to reestimate quantities and posteriori probabilities for exemplar histograms and the process continues until convergence giving $\hat{\mathbf{P}} = \mathbf{P}_{(t=\infty)}$, maximising (5) consistent with the convergence theorem of EM [18] [3]. To avoid the risk of converging into a local minimum the algorithm can be restarted multiple times from different random PMF initialisations.

Knowledge of the Poisson noise in incoming data can assist in assessing the quality of the extracted components. If the residuals between model and data are observed to be greater than predicted by the Poisson noise then a better model must be sought. If the residuals are too small then this is a sign of over-training. Training can continue using different random seeds and numbers of components until a goodness-of-fit criterion is met. Starting with Chi-squared statistics, assuming normally distributed residuals with widths σ_X, the fit between a model and observed data will have a Chi-squared per degree of freedom, $\chi^2_D = \frac{1}{D}\sum_X \frac{(\mathbf{M}_X - \mathbf{H}_X)^2}{\sigma^2_X}$ of unity [4]. However, the models sought have Poisson residuals with discrepancies between Poisson and Gaussian distributions especially evident at low sample statistics. A square-root transform [1] can be performed to both model and data to transform the residuals into something better approximating a Gaussian with uniform width of $\sigma^2 = \frac{1}{4}$, thereby widening the scope of the test to such cases. A Chi-squared per D degrees of freedom function, χ^2_D, can then be defined as:

$$\chi^2_{(N(m-n))} = \frac{1}{N}\sum_r \frac{4}{m-n}\sum_X \left(\sqrt{\mathbf{M}_{X(r)}} - \sqrt{\mathbf{H}_{X(r)}}\right)^2 \tag{14}$$

where m is the number of bins in the examplar histograms; n is the number of components in the model; and $\mathbf{M}_{X(r)}$ is the modelled frequency in the rth example histogram's X bin accounted for by the selected components. The number of model degrees of freedom are the total number of X bins per histogram times the number of histograms, minus the total number of component quantity parameters per histogram. A goodness-of-fit of unity confirms data is being modelled within the limits of the Poisson histograms and therefore allows the Poisson noise assumption to be used to derive further error information in section 2.4.1.

2.4 General Case: Covariance estimation using EM ICA components

The CRB covariance is inappropriate when component PMFs have been estimated from exemplar histograms, as the CRB only accounts for statistical perturbations in incoming data, $\mathbf{H}$, and not additional error from training data introduced into $\mathbf{P}$. PMFs containing noise become a systematic source of error. Error propagation can be used in this case to estimate a new covariance taking into account both statistical and systematic effects. This propagation of uncertainty must deal with the iterative nature of the EM algorithm. The new covariance will estimate uncertainties in the measured quantity vector using the following three steps:

1. Estimation of Error Sources: variance in $\mathbf{H}$ and $\mathbf{P}$;

2. Single EM Step Error: a single EM step estimate of how noise affects one instance of the update function;

3. Error Amplification: an amplification stage accounting for the accumulative effects of the iterative feedback of errors in subsequent EM steps.

2.4.1 Estimation of Error Sources

The statistical source of error in final quantity estimates stems from Poisson variability in incoming histogram bin frequencies. This error is assumed to be independent from bin to bin. The variance from a single X bin is therefore:

$$\sigma^2_{\mathbf{H}_X} =< \mathbf{H}_X > \approx \mathbf{H}_X \tag{15}$$

where $< \mathbf{H}_X >$ indicates the statistical expectation of the bin frequency; and for implementation and notational convenience can be approximated by the observed frequency.

The noise embedded within extracted histogram components can be considered as a type of systematic error. This is because whichever pattern of noise is found within the training data, that instance of the noise remains fixed during any subsequent usage, i.e. the system is only trained once and then the same noise is observed in the components each time they are fitted to new incoming data. The noise from training data can be seen to give overall Poisson behaviour in the noise in components. The resulting variances can be determined via a two step argument:

1) For any given approximation of $P(k|X)$, if the total content of an example histogram bin, $\mathbf{H}_{X(r)}$, were fixed then the variance on $\hat{R}_r(X|k)$ would follow a Binomial distribution with each bin entry having a probability of $P(k|X)$ of being included in the sample and a probability of $1 - P(k|X)$ of being excluded. This variance is approximated by:

$$\mathbf{H}_{X(r)}P(k|X)[1 - P(k|X)] = \mathbf{H}_{X(r)}P(k|X) - \mathbf{H}_{X(r)}P(k|X)^2 \tag{16}$$

2) However, as $\mathbf{H}_{X(r)}$ is not fixed there is also an independent Poisson contribution from the originating histogram sample:

$$\left(\frac{\partial \hat{R}_r(X|k)}{\partial \mathbf{H}_{X(r)}}\right)^2 \sigma^2_{\mathbf{H}_{X(r)}} = P(k|X)^2\mathbf{H}_{X(r)} \tag{17}$$

which adds to the Binomial variance to give a total variance of:

$$\sigma^2_{R_r(X|k)} = P(k|X)\mathbf{H}_{X(r)} = R_r(X|k) \tag{18}$$

giving an overall Poisson behaviour.

The sum of the independent example histogram errors therefore also combine like Poisson variances, giving a systematic error equivalent to using a raw histogram sample, $\mathbf{H}_{X|k} = \sum_r R_r(X|k)$, containing all the information for component k from all exemplars:

$$\sigma^2_{\mathbf{H}_{X|k}} =< \mathbf{H}_{X|k} > \approx \mathbf{H}_{X|k} \tag{19}$$

These approximations to variances, using the assumptions made, are corroborated in Monte-Carlo with results contributing to section 3.

2.4.2 Single EM Step Error

For the purposes of error propagation it is convenient to consider the EM update function in terms of a single histogram bin, X, and every other bin which is not X, which will be denoted $\bar{X}$, e.g. $\mathbf{H}_{\bar{X}} = \sum_{Y \neq X} \mathbf{H}_Y$. It is also convenient to consider components in a similar notation using k and all other components, $\bar{k}$, e.g. $\mathbf{Q}_{\bar{k}} = \sum_{l \neq k} \mathbf{Q}_l$. The update function can then be stated in terms of incoming histogram, $\mathbf{H}_X$, (data) and exemplar sub-histogram, $\mathbf{H}_{X|k}$, (model) bins giving:

$$\mathbf{Q}'_k = P(k|X)\mathbf{H}_X + P(k|\bar{X})\mathbf{H}_{\bar{X}} \tag{20}$$

$$= \frac{\left(\frac{\mathbf{H}_{X|k}\mathbf{Q}_k}{\mathbf{H}_{X|k}+\mathbf{H}_{\bar{X}|k}}\right)\mathbf{H}_X}{\left(\frac{\mathbf{H}_{X|k}\mathbf{Q}_k}{\mathbf{H}_{X|k}+\mathbf{H}_{\bar{X}|k}} + \frac{\mathbf{H}_{X|\bar{k}}\mathbf{Q}_{\bar{k}}}{\mathbf{H}_{X|\bar{k}}+\mathbf{H}_{\bar{X}|\bar{k}}}\right)} + \frac{\left(\frac{\mathbf{H}_{\bar{X}|k}\mathbf{Q}_k}{\mathbf{H}_{X|k}+\mathbf{H}_{\bar{X}|k}}\right)\mathbf{H}_{\bar{X}}}{\left(\frac{\mathbf{H}_{\bar{X}|k}\mathbf{Q}_k}{\mathbf{H}_{X|k}+\mathbf{H}_{\bar{X}|k}} + \frac{\mathbf{H}_{\bar{X}|\bar{k}}\mathbf{Q}_{\bar{k}}}{\mathbf{H}_{X|\bar{k}}+\mathbf{H}_{\bar{X}|\bar{k}}}\right)} \tag{21}$$

where $\mathbf{Q}'$ is the updated quantity; and $\mathbf{Q}$ is the previous quantity. Uncertainty from the two sources of error can be propagated through this update function by considering how small changes in the inputs affect the estimated vector of quantities. As the two sources are independent their contributions to the covariance can be derived separately then summed:

$$\mathbf{C}_{EMStep} = \mathbf{C}_{data} + \mathbf{C}_{model} \tag{22}$$

$$\mathbf{C}_{ij(data)} = \sum_X \left[\left(\frac{\partial \mathbf{Q}'_i}{\partial \mathbf{H}_X}\right)\left(\frac{\partial \mathbf{Q}'_j}{\partial \mathbf{H}_X}\right)\sigma^2_{\mathbf{H}_X}\right] \tag{23}$$

$$\mathbf{C}_{ij(model)} = \sum_X \left[\sum_k \left(\frac{\partial \mathbf{Q}'_i}{\partial \mathbf{H}_{X|k}}\right)\left(\frac{\partial \mathbf{Q}'_j}{\partial \mathbf{H}_{X|k}}\right)\sigma^2_{\mathbf{H}_{X|k}}\right] \tag{24}$$

where $\mathbf{C}_{data}$ is the statistical contribution from the incoming histogram data; and $\mathbf{C}_{model}$ is the systematic contribution from the training exemplar histograms used to construct the component models.

The statistical contribution is straightforward giving:

$$\mathbf{C}_{ij(data)} = \sum_X P(i|X)P(j|X)\mathbf{H}_X \tag{25}$$

In contrast, the systematic contribution involves relatively complex derivatives (found in appendix B). Defining the total quantity of training data for component k as $\mathbf{T}_k = \mathbf{H}_{X|k} + \mathbf{H}_{\bar{X}|k}$, in the cases where $i = k$ or $j = k$ the derivatives are:

$$\frac{\partial \mathbf{Q}'_{i=k}}{\partial \mathbf{H}_{X|k}} = \frac{\partial \mathbf{Q}'_{j=k}}{\partial \mathbf{H}_{X|k}} = \frac{P(k|X)P(\bar{k}|X)P(\bar{X}|k)\mathbf{H}_X}{\mathbf{T}_k P(X|k)} - \frac{P(\bar{k}|\bar{X})P(k|\bar{X})\mathbf{H}_{\bar{X}}}{\mathbf{T}_k} \tag{26}$$

In the cases where $i \neq k$ (or similarly when $j \neq k$, substituting j for i in the following) the same terms become:

$$\frac{\partial \mathbf{Q}'_{i\neq k}}{\partial \mathbf{H}_{X|k}} = \frac{P(i|\bar{X})P(k|\bar{X})\mathbf{H}_{\bar{X}}}{\mathbf{Q}_k} - \frac{P(i|X)P(k|X)P(\bar{X}|k)\mathbf{H}_X}{\mathbf{T}_k P(X|k)} \tag{27}$$

Substituting these results back into the covariance calculation (24) completes the estimate of the error contribution on quantities after performing a single EM step. These approximations are corroborated in Monte-Carlo with results contributing to section 3.

2.4.3 Error Amplification

The initial error from an EM step will be amplified in the final converged quantity estimates. This amplification can be approximated by a convergent geometric series (derived in appendix C) and applied to the single step covariance using an amplification matrix:

$$\mathbf{C} = \mathbf{A}^\intercal \mathbf{C}_{EMStep} \mathbf{A} \tag{28}$$

$$\mathbf{A} = [\mathbf{I} - \nabla \mathbf{Q}]^{-1} \tag{29}$$

where $\mathbf{I}$ is the identity matrix; and $\nabla \mathbf{Q}$ is the Jacobian:

$$\nabla \mathbf{Q}_{ij} = \frac{\partial \mathbf{Q}_i}{\partial \mathbf{\Delta}_j} \tag{30}$$

where $\mathbf{\Delta}$ is an initial small error in quantity. The diagonal terms are given by:

$$\nabla \mathbf{Q}_{ii} = \sum_X P(X|i) - P(X|i)P(i|X) \tag{31}$$

and off-diagonal terms given by:

$$\nabla \mathbf{Q}_{ij} = \sum_X -P(X|j)P(i|X) \tag{32}$$

Whilst this formulation of $\mathbf{C}$ looks considerably different from the original CRB covariance estimate, the two are consistent. It will be shown numerically in section 3 that when model error terms are all zero the error propagation version described above produces equivalent results to the CRB version. Importantly, this confirms that the approach taken provides an appropriate approximation of the EM amplification effect.

2.5 Total Measurement Error: Class covariance estimation

Physical classes being measured (i.e. the linear combination of a set of related histogram subcomponents) have quantity estimates given by (3) which is computable after the EML estimation process has determined a histogram's composition in terms of its individual components. This simple linear addition of component quantities into superordinate class quantities imply a simple linear combination of errors. Error propagation can be applied to give final measurement accuracies:

$$\mathbf{C}_{\mathbb{Q}} = \mathbb{K} \mathbf{C} \mathbb{K}^\intercal \tag{33}$$

thereby completing the process of quantitative measurement, providing both estimates of the physical processes present in data and the accuracy to which they can be measured.

2.6 Pseudo-code

The following algorithm is for the general case where PMFs must be learned from training examples before they can be applied to make measurements, $\mathbb{Q}$, in new data. An implementation in C will be available in the Tina-Vision (www.tina-vision.net) open source vision software libraries.

2.6.1 Training algorithm

For each class of physical process $\{K = 0, K = 1, \ldots, K = z\}$:

1. Gather N exemplar histograms of class K $\{\mathbf{H}_{(r=1)}, \mathbf{H}_{(r=2)}, \ldots, \mathbf{H}_{(r=N)}\}$
2. Randomly initialise common definition of $\mathbf{P}$
3. Using an increasing number of components, whilst model fit, χ^2_D equ. (14), is greater than unity and $\mathbf{P}$ has not converged:
   (a) For each histogram $\mathbf{H}_{(r)}$:
      i. Iterate equ. (6) and equ. (7) until $\mathbf{P}^{-1}_{(r)}$ and $\mathbf{Q}_{(r)}$ converge
      ii. Use $\mathbf{P}^{-1}_{(r)}$ to estimate individual component contributions, $\hat{R}_r$, using equ. (12)
   (b) Sum and normalise each component to give new common estimate of $\mathbf{P}$ using equ (13)
4. Update component-class mapping matrix, $\mathbb{K}$, to indicate which components belong to class K

2.6.2 Measurement algorithm

For incoming data histogram $\mathbf{H}$:

1. Iterate equ. (6) and equ. (7) until $\mathbf{P}^{-1}$ and $\mathbf{Q}$ converge
2. Compute total class quantities, $\mathbb{Q}$, using equ. (3)
3. Compute statistical and systematic EM covariance terms using equ. (23) and equ. (24)
4. Compute amplification matrix using equ. (29)
5. Compute component covariances, $\mathbf{C}$, using equ. (28)
6. Compute total class covariances, $\mathbf{C}_{\mathbb{Q}}$, using equ. (33)
7. Square-root of diagonal terms give +/- 1 standard deviation error on measurements, i.e. $\mathbb{Q}_K \pm \sqrt{\mathbf{C}_{\mathbb{Q}(K)}}$

3 Experiments

Simulated histogram data was used to test the following aspects of the method:

- The ability of the ICA training algorithm to extract subcomponent distributions, given known reference generating distributions;
- The ability of the EM fitting algorithm and CRB to estimate the composition of histograms, within predicted statistical errors;

- The ability of the EM fitting algorithm and error propagation to estimate the composition of histograms, within predicted systematic and statistical errors;

A Monte-Carlo simulation was created to generate histogram data containing linearly combined reference distribution, perturbed by independent Poisson noise. 64 bin reference distributions were coded by hand and can be seen in figure 1. These distributions were used as templates for creating 10 different components by shifting the distributions along the x-axis. For example, figure 2 shows the approximate distributions extracted during 64 bin tests at the point where 3 components per model were generated. 4096 and 16384 bin histogram component reference distributions were generated using a texture sampling scheme. To summarise, the distributions were sampled from Martian image data using 12 bit and 14 bit binary descriptors, involving a random selection of pixels making the organisation of the histogram bins essentially random, giving a complex range of synthetic data.

The scaling quantities were all real valued uniform random numbers between 1,000 and 1,000,000. For each histogram bin, a random Poisson number generator was used to draw independent samples from each component before summing them to give total histogram frequencies. Model selection was tested for histograms containing 64, 4096 and 16384 bins. Data was synthesised for each sized histogram containing known linear combinations of predefined reference distributions. This was performed for increasingly complex models containing between 1 to 10 linear components. Figure 3 plots the goodness-of-fit for different numbers of extracted components for the 4096 bin histograms. These results are indicative of the other histogram sizes, reaching unity at the correct number of components.

The process of applying the CRB then combining related errors together via error propagation was tested as a predictor of errors on class quantities. Histograms were grouped into 3 classes each containing 3 components. Known random quantities of each component were repeatedly combined to generate a series of new histograms. The EM model fitting algorithm was used to fit the extracted components to this new incoming data and estimated quantity parameters, $\mathbf{Q}$, were inspected. These quantities were summed into class quantities, $\mathbb{Q}$, with the differences between true class values and estimated values being measured over multiple independent trials.

The incoming testing histograms were generated using a range of relative quantities from 0.01 to 100 times the training amounts. Each ratio of training to testing data was repeatedly tested over 1,000 trials. At each trial the difference between the true class quantities and estimated class quantities were divided by the predicted standard deviation (via CRB and error prop.) and recorded in a Pull distribution. If errors were predicted correctly then the standard deviation of the Pull distribution around its mean should equal unity for the CRB, i.e. the mean ratio of observed to predicted errors should be one. If the spread of the systematic and statistical errors, via error propagation, were predicted correctly then the standard deviation of the Pull distribution about zero should be equal to unity, i.e. the deviation from zero (systematic error) should also be accounted for.

Figure 4 plots the standard deviations of the Pull distributions about their means for each histogram size and relative quantity of testing data using the CRB. The ratio of observed to predicted total errors can be seen in figure 5 for the errror propagation version. The change in relative contributions from statistical and systematic effects to the total error as the ratio of training to testing data increases can be seen in figure 6. Finally, the relationship between statistical errors computed using the CRB verses the same errors computed using error propagation can be seen in figure 7, showing that they are self-consistent.

We believe the set of simulated problems cover a sufficient dynamic range of quantities and complexities to demonstrate our method's general utility.

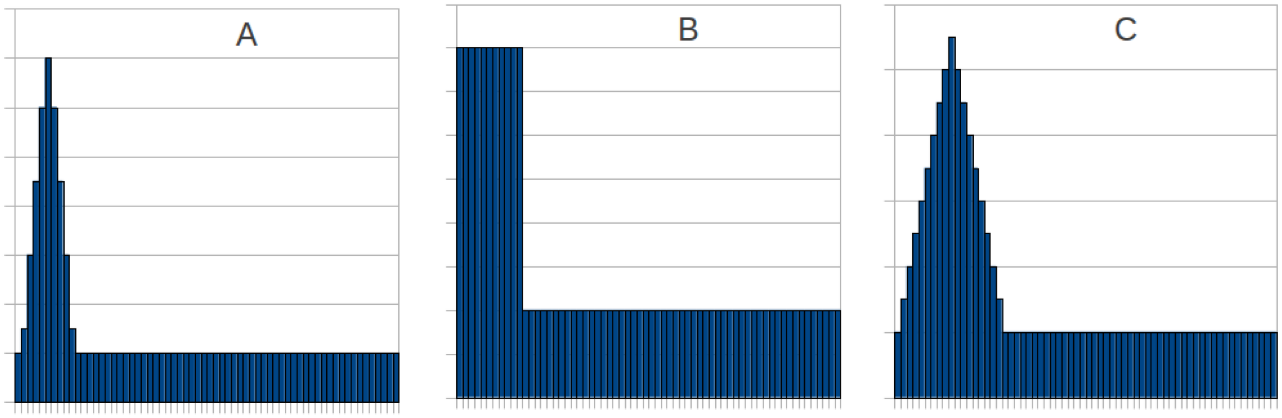

Figure 1: These are example distributions used as generators within the Monte-Carlo simulation. Shifted and scaled versions of these distributions were used to generate the 64 bin models.

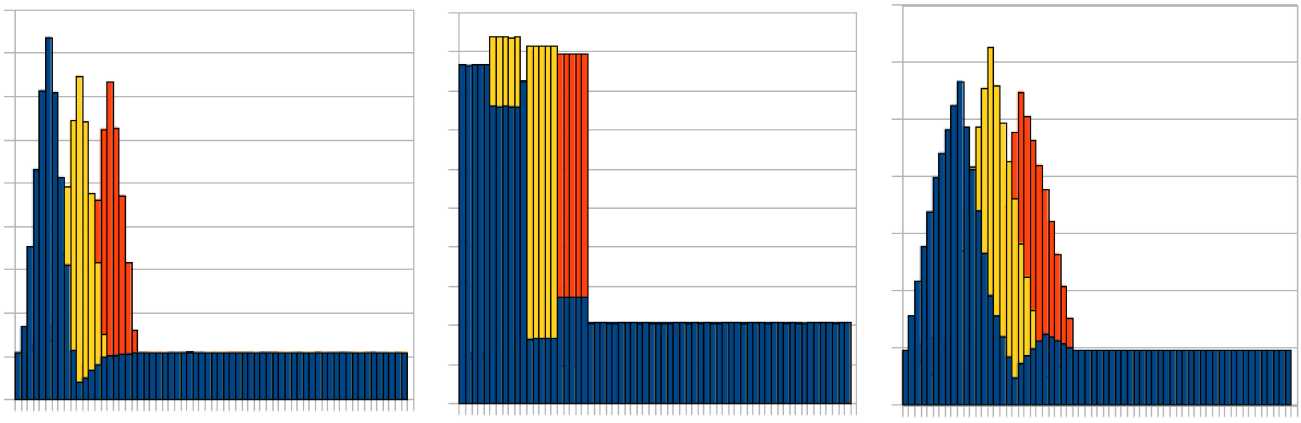

Figure 2: These are examples of components extracted from 64 bin training histograms created during the Monte-Carlo simulations. Note that the dependences between reference distributions and extracted components at the location of overlaps need not be considered a problem, assuming the linear sub-space defined by these histograms spans the full range of training and anticipated future data.

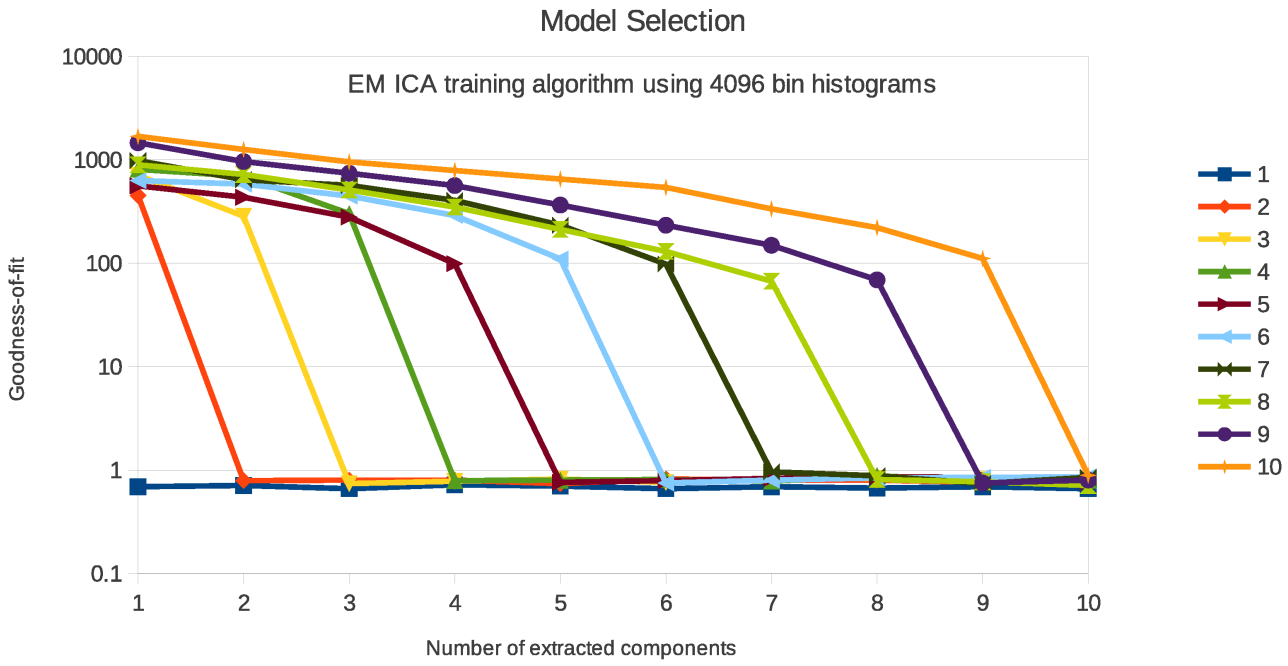

Figure 3: This plot shows that by using the χ^2_D goodness-of-fit function the EM ICA algorithm can extract an appropriate number of linear components to describe a set of example training histograms. Each curve represents a Monte-Carlo data source generated with between 1 and 10 components. Each curve crosses unity at the most appropriate point, i.e. when the number of extracted components equals the number of generating components.

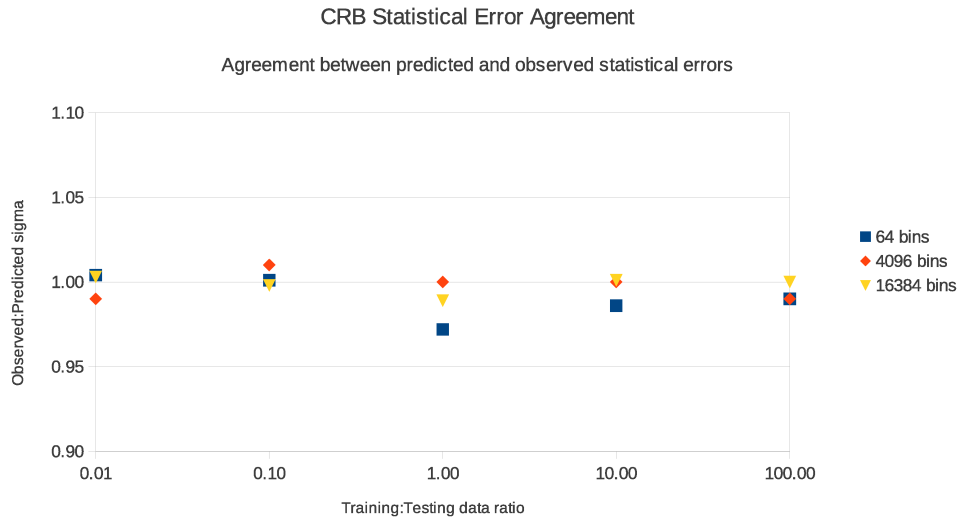

Figure 4: This plot compares CRB error estimates to the observed spread of estimated quantities for varying ratios of training to testing data. The y-axis shows the ratio of observed to predicted errors.

Figure 5: This plot compares total error estimates, incorporating both systematic and statistical sources of uncertainty, with observed errors around true values for varying ratios of training to testing data. The y-axis shows the ratio of observed to predicted errors.

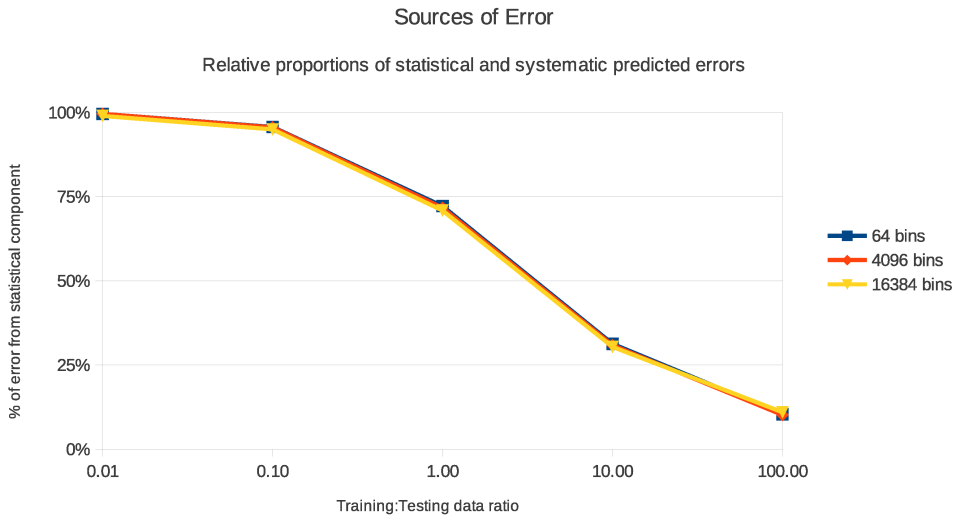

Figure 6: This plot shows how contributions from statistical and systematic sources to the total error change in proportion as the relative quantity of testing data increases. The contribution from the statistical component is below the curve.

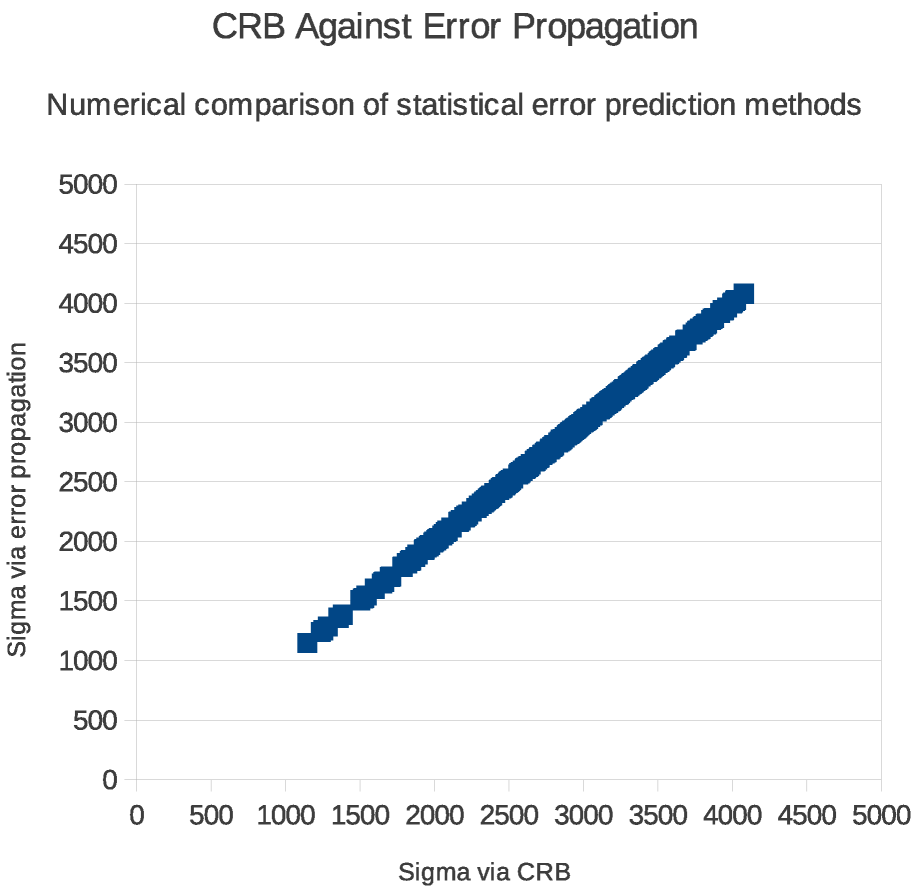

Figure 7: The plot shows the equivalence between the CRB predicted statistical error and the error propagation alternative over a range of quantities.

4 Discussion

The quantitative criteria of success whilst assessing our technique's performance are that:

- empirically observed quantity errors match predicted quantity errors, i.e. the covariance, $\mathbf{C}_{\mathbb{Q}}$, describes the actual spread of estimated $\mathbb{Q}$ over repeated trials;
- and model fit values approach unity when the correct components are fitted to data, i.e. if the incoming histograms can be described by the selected PMFs then the chi-squared per degree of freedom should be 1.

These criteria can only be met if both quantity and error estimation processes work correctly. Error predictions necessitate the estimation of quantities and the model fit necessitates correct extraction of PMFs with predictable bin errors. As such, testing errors and model fits is sufficient for drawing conclusions regarding the overall validity of our methods. As noted in the introduction we do not provide comparative results from alternative ICA style methods because they do not provide error assessments which can be meaningfully compared to our own. Whilst it would be possible to build alternative models using other noted methods, those methods do not provide predictions of parameter error covariances and therefore lack the necessary outputs for comparison to our quantitative technique.

Figures 4 and 5 show that for the range of problems selected the simple and general cases of quantity and error estimation work correctly. In all experiments the predicted errors were in agreement with the observed deviations from known ground-truth within the statistical limits of our ability to measure agreement over 1,000 trials. Fig 6, with a common x-axis to Fig 5, illustrates that both statistical and systematic effects are being estimated correctly and that the statistical effect is only dominant at relatively small quantities of incoming data, making the systematic errors a limiting factor in the ability to make large quantity measurements. Fig 7 shows that the two contrasting methods of statistical error prediction (CRB vs error propagation) are self-consistent, and confirms that the error amplification theory is an appropriate method of modelling iterative error feedback in EM.

Inspection of the covariance calculations show that the statistical error component grows proportionally to the square-root of the quantity of data being analysed. This is in contrast to the systematic component which grows linearly. In the best case scenario where model components are unambiguous, i.e. have non-overlapping PMFs, the statistical component behaves like simple Poisson noise and the systematic effect disappears. In the more usual case of ambiguity between components, the greater the overlap between PMFs the larger both statistical and systematic errors become. This is analogous to the conventional observation that ICA cannot be applied to multivariate Gaussian distributions, as the degeneracy is total and errors become excessively large. We extend this conventional view by adding that any combination of highly ambiguous components cannot be measured accurately, however, with the benefit of a covariance estimate the extent of correlated uncertainties can be quantified, thus avoiding the risk of over-interpreting results.

It is known that linear decompositions of data can have multiple equally valid solutions spanning a common subspace. This is also true of our method where extracted components do not necessarily appear identical to the underlying distributions generated in Monte-Carlo, as can be seen by the triplets of extracted components in Fig 2. What is important is that the components extracted are on the same manifold as the original sources and are capable of describing the subspace covered by the exemplar histograms encountered during training. It is also important that the extracted components have the property of predictable noise. These properties are what allow our model selection method to work correctly, as seen in Fig 3 where the chi-squared per degree of freedom

model fit function approaches unity at an appropriate number of components. These predictable noise properties are also required for our covariance estimates to be valid.

Many similar techniques have avoid detailed considerations of noise or attempts have been made to remove noise in a preprocessing stage [13]. For quantitative use, any noise in incoming data will affect the operation and choice of algorithm. For instance, it may appear that the inverse probability matrix in (2) could be solved using a standard least-squared pseudo-inverse approximation. However, such a method assumes uniform Gaussian isotropic noise. Probabilities derived from Poisson histogram samples do not have such noise characteristics, necessitating the use of EM or an equivalent alternative. Also, noise does not need to be removed or reduced for an algorithm to be successful. What is important is that users are given enough information to understand how the noise in inputs affects the stability of outputs for the purposes of setting confidence intervals or plotting error bars. We have shown that the CRB and error propagation can be applied to provide this key information. Our emphasis on understanding the effects of noise also facilitates the fully quantitative testing of algorithms whereby the performance of an algorithm can be predicted through theoretical considerations, then predicted performances can be corroborated empirically [11].

Our Monte-Carlo testing comprehensively shows that our techniques are appropriate for a wide range of linear models composed from histograms with independent Poissson bins. We therefore expect our techniques to be applicable to real scientific analysis tasks involving data that exhibits, or closely approximates, such properties. Potential applications are wide-ranging and varied, including problems such as image segmentation from texture histograms [10], mass spectrum analysis [17] and grain size frequency analysis for sedimentology [2]. Raw data in the form of histograms is an extremely common format found in almost all of the sciences, many of which would be expected to have the properties described. However, it will be important to confirm the method's appropriateness on an application specific basis, as nonlinearities or unexpected correlations within a histogram, for example, could violate the assumptions made within the theory.

Algorithmically, training and model fitting run-time grows linearly with the number of histogram bins and with the number of histogram components. Covariance calculation run-time grows linearly with the number of histogram bins and grows with the square of the number of components. In absolute terms, histograms containing a few thousand bins and a dozen components can be used to build models and make measurements within a small number of minutes.

5 Conclusion

We have taken techniques and ideas from linear modelling of complex data which are usually applied to reduce and summarise results and created a method of quantitative measurement. Our method provides the means to conduct ICA on histograms, to group components into meaningful physical classes given appropriate training data, then to measure the quantities of classes present in new data. A significant part of the theory provides the necessary error estimates for scientifically useful interpretation of measurement in the form of covariances. All aspects have been corroborated in Monte-Carlo simulation demonstrating validity over a wide range of scenarios. Our methods are now sufficiently well developed to turn attention to their practical use in real scientific applications.

Appendix A

The accuracy of estimated quantities can be determined using the CRB on the EML function:

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \mathbf{Q}_i \partial \mathbf{Q}_j} \geq \mathbf{C}_{ij}^{-1} \tag{34}$$

$$\mathbf{C}_{ij}^{-1} \approx \frac{\sum_X P(i|X)P(j|X)\mathbf{H}_X}{\mathbf{Q}_i \mathbf{Q}_j} \tag{35}$$

where $\mathbf{C}_{ij}^{-1}$ is the inverse covariance between quantities $\mathbf{Q}_i$ and $\mathbf{Q}_j$. This assumes all noise is statistical, originating from the histograms under analysis and that the PMFs are correct, being free of any errors.

Computing the first and second order derivatives gives:

$$\frac{\partial \ln \mathcal{L}}{\partial \mathbf{Q}_j} = \sum_X \frac{P(X|j)\mathbf{H}_X}{\sum_k P(X|k)\mathbf{Q}_k} - 1 \tag{36}$$

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \mathbf{Q}_i \partial \mathbf{Q}_j} = \sum_X \frac{P(X|i)P(X|j)\mathbf{H}_X}{[\sum_k P(X|l)\mathbf{Q}_k]^2} \tag{37}$$

then the similarity with Bayes Theorem can be used to give:

$$\mathbf{C}_{ij}^{-1} \approx \frac{\sum_X P(i|X)P(j|X)\mathbf{H}_X}{\mathbf{Q}_i \mathbf{Q}_j} \tag{38}$$

From this the covariance matrix can be computed using a standard matrix inversion algorithm.

Appendix B

For the purposes of error propagation it is convenient to consider the EM update function in terms of a single histogram bin, X, and every other bin which is not X, which will be denoted $\bar{X}$, e.g. $\mathbf{H}_{\bar{X}} = \sum_{Y \neq X} \mathbf{H}_Y$. It is also convenient to consider components in a similar notation using k and all other components, $\bar{k}$, e.g. $\mathbf{Q}_{\bar{k}} = \sum_{l \neq k} \mathbf{Q}_l$. The update function can then be stated in terms of incoming histogram, $\mathbf{H}_X$, (data) and exemplar sub-histogram, $\mathbf{H}_{X|k}$, (model) bins giving:

$$\mathbf{Q}'_k = P(k|X)\mathbf{H}_X + P(k|\bar{X})\mathbf{H}_{\bar{X}} \tag{39}$$

$$= \frac{\left(\frac{\mathbf{H}_{X|k}\mathbf{Q}_k}{\mathbf{H}_{X|k}+\mathbf{H}_{\bar{X}|k}}\right)\mathbf{H}_X}{\left(\frac{\mathbf{H}_{X|k}\mathbf{Q}_k}{\mathbf{H}_{X|k}+\mathbf{H}_{\bar{X}|k}} + \frac{\mathbf{H}_{X|\bar{k}}\mathbf{Q}_{\bar{k}}}{\mathbf{H}_{X|\bar{k}}+\mathbf{H}_{\bar{X}|\bar{k}}}\right)} + \frac{\left(\frac{\mathbf{H}_{\bar{X}|k}\mathbf{Q}_k}{\mathbf{H}_{X|k}+\mathbf{H}_{\bar{X}|k}}\right)\mathbf{H}_{\bar{X}}}{\left(\frac{\mathbf{H}_{\bar{X}|k}\mathbf{Q}_k}{\mathbf{H}_{X|k}+\mathbf{H}_{\bar{X}|k}} + \frac{\mathbf{H}_{\bar{X}|\bar{k}}\mathbf{Q}_{\bar{k}}}{\mathbf{H}_{X|\bar{k}}+\mathbf{H}_{\bar{X}|\bar{k}}}\right)} \tag{40}$$

where $\mathbf{Q}'$ is the updated quantity; and $\mathbf{Q}$ is the previous quantity. Uncertainty from the two sources of error can be propagated through this update function by considering how small changes in the inputs affect the estimated vector of quantities. As the two sources are independent their contributions to the covariance can be derived separately then summed:

$$\mathbf{C}_{EMStep} = \mathbf{C}_{data} + \mathbf{C}_{model} \tag{41}$$

$$\mathbf{C}_{ij(data)} = \sum_X \left[\left(\frac{\partial \mathbf{Q}'_i}{\partial \mathbf{H}_X}\right)\left(\frac{\partial \mathbf{Q}'_j}{\partial \mathbf{H}_X}\right)\sigma^2_{\mathbf{H}_X}\right] \tag{42}$$

$$\mathbf{C}_{ij(model)} = \sum_X \left[\sum_k \left(\frac{\partial \mathbf{Q}'_i}{\partial \mathbf{H}_{X|k}} \right) \left(\frac{\partial \mathbf{Q}'_j}{\partial \mathbf{H}_{X|k}} \right) \sigma^2_{\mathbf{H}_{X|k}} \right] \tag{43}$$

where $\mathbf{C}_{data}$ is the statistical contribution from the incoming histogram data; and $\mathbf{C}_{model}$ is the systematic contribution from the training exemplar histograms used to construct the component models.

The statistical contribution is straightforward giving:

$$\mathbf{C}_{ij(data)} = \sum_X P(i|X)P(j|X)\mathbf{H}_X \tag{44}$$

In contrast, the systematic contribution involves relatively complex derivatives (found in appendix B) and is best approached in parts. The derivative of a quantity with respect to an exemplar histogram bin can be divided into two independent terms:

$$\frac{\partial \mathbf{Q}'_j}{\partial \mathbf{H}_{X|k}} = \frac{\partial P(j|X)\mathbf{H}_X}{\partial \mathbf{H}_{X|k}} + \frac{\partial P(j|\bar{X})\mathbf{H}_{\bar{X}}}{\partial \mathbf{H}_{X|k}} \tag{45}$$

Defining the total quantity of training data for component k as $\mathbf{T}_k = \mathbf{H}_{X|k} + \mathbf{H}_{\bar{X}|k}$, in the cases where $j = k$ the two terms are given by

$$\begin{aligned}
\frac{\partial P(k|X)\mathbf{H}_X}{\partial \mathbf{H}_{X|k}} &= \frac{\left(\frac{\mathbf{H}_{X|k}\mathbf{Q}_k}{\mathbf{T}_k} + \frac{\mathbf{H}_{X|\bar{k}}\mathbf{Q}_{\bar{k}}}{\mathbf{T}_{\bar{k}}}\right)\frac{P(\bar{X}|k)\mathbf{Q}_k}{\mathbf{T}_k} - \left(\frac{\mathbf{H}_{X|k}\mathbf{Q}_k}{\mathbf{T}_k}\right)\frac{P(\bar{X}|k)\mathbf{Q}_k}{\mathbf{T}_k}}{\left(\frac{\mathbf{H}_{X|k}\mathbf{Q}_k}{\mathbf{T}_k} + \frac{\mathbf{H}_{X|\bar{k}}\mathbf{Q}_{\bar{k}}}{\mathbf{T}_{\bar{k}}}\right)^2}\mathbf{H}_X \\
&= \frac{P(X|\bar{k})\mathbf{Q}_{\bar{k}}P(\bar{X}|k)\mathbf{Q}_k\mathbf{H}_X}{\mathbf{T}_k[P(X|k)\mathbf{Q}_k + P(X|\bar{k})\mathbf{Q}_{\bar{k}}]^2} \times \frac{P(X|k)}{P(X|k)} \\
&= \frac{P(k|X)P(\bar{k}|X)P(\bar{X}|k)\mathbf{H}_X}{\mathbf{T}_k P(X|k)}
\end{aligned} \tag{46}$$

and

$$\begin{aligned}
\frac{\partial P(k|\bar{X})\mathbf{H}_{\bar{X}}}{\partial \mathbf{H}_{X|k}} &= \frac{-\left(\frac{\mathbf{H}_{\bar{X}|k}\mathbf{Q}_k}{\mathbf{T}_k} + \frac{\mathbf{H}_{\bar{X}|\bar{k}}\mathbf{Q}_{\bar{k}}}{\mathbf{T}_{\bar{k}}}\right)\frac{P(\bar{X}|k)\mathbf{Q}_k}{\mathbf{T}_k} + \left(\frac{\mathbf{H}_{\bar{X}|k}\mathbf{Q}_k}{\mathbf{T}_k}\right)\frac{P(\bar{X}|k)\mathbf{Q}_k}{\mathbf{T}_k}}{\left(\frac{\mathbf{H}_{\bar{X}|k}\mathbf{Q}_k}{\mathbf{T}_k} + \frac{\mathbf{H}_{\bar{X}|\bar{k}}\mathbf{Q}_{\bar{k}}}{\mathbf{T}_{\bar{k}}}\right)^2}\mathbf{H}_{\bar{X}} \\
&= -\frac{\mathbf{H}_{\bar{X}|\bar{k}}\mathbf{Q}_{\bar{k}}P(\bar{X}|k)\mathbf{Q}_k\mathbf{H}_{\bar{X}}}{\mathbf{T}_{\bar{k}}\mathbf{T}_k[P(\bar{X}|k)\mathbf{Q}_k + P(\bar{X}|\bar{k})\mathbf{Q}_{\bar{k}}]^2} \\
&= -\frac{P(\bar{k}|\bar{X})P(k|\bar{X})\mathbf{H}_{\bar{X}}}{\mathbf{T}_k}
\end{aligned} \tag{47}$$

giving

$$\frac{\partial \mathbf{Q}'_k}{\partial \mathbf{H}_{X|k}} = \frac{P(k|X)P(\bar{k}|X)P(\bar{X}|k)\mathbf{H}_X}{\mathbf{T}_k P(X|k)} - \frac{P(\bar{k}|\bar{X})P(k|\bar{X})\mathbf{H}_{\bar{X}}}{\mathbf{T}_k} \tag{48}$$

In the case where $j \neq k$ the same terms become

$$\frac{\partial P(j|X)\mathbf{H}_X}{\partial \mathbf{H}_{X|k}} = \frac{-\left(\frac{\mathbf{H}_{X|j}\mathbf{Q}_j}{\mathbf{T}_j}\right)\frac{P(\bar{X}|k)\mathbf{Q}_k}{\mathbf{T}_k}}{\left(\frac{\mathbf{H}_{X|k}\mathbf{Q}_k}{\mathbf{T}_k} + \frac{\mathbf{H}_{X|\bar{k}}\mathbf{Q}_{\bar{k}}}{\mathbf{T}_{\bar{k}}}\right)^2}\mathbf{H}_X$$

$$= -\frac{P(X|j)\mathbf{Q}_j P(\bar{X}|k)\mathbf{Q}_k \mathbf{H}_X}{\mathbf{T}_k[P(X|k)\mathbf{Q}_k + P(X|\bar{k})\mathbf{Q}_{\bar{k}}]^2} \times \frac{P(X|k)}{P(X|k)}$$

$$= -\frac{P(j|X)P(k|X)P(\bar{X}|k)\mathbf{H}_X}{\mathbf{T}_k P(X|k)} \tag{49}$$

and

$$\frac{\partial P(j|\bar{X})\mathbf{H}_{\bar{X}}}{\partial \mathbf{H}_{X|k}} = \frac{P(\bar{X}|j)\mathbf{Q}_j P(\bar{X}|k)\mathbf{Q}_k \mathbf{H}_{\bar{X}}}{\mathbf{Q}_k[P(\bar{X}|k)\mathbf{Q}_k + P(\bar{X}|\bar{k})\mathbf{Q}_{\bar{k}}]^2}$$

$$= \frac{P(j|\bar{X})P(k|\bar{X})\mathbf{H}_{\bar{X}}}{\mathbf{Q}_k} \tag{50}$$

giving

$$\frac{\partial \mathbf{Q}'_j}{\partial \mathbf{H}_{X|k}} = \frac{P(j|\bar{X})P(k|\bar{X})\mathbf{H}_{\bar{X}}}{\mathbf{Q}_k} - \frac{P(j|X)P(k|X)P(\bar{X}|k)\mathbf{H}_X}{\mathbf{T}_k P(X|k)} \tag{51}$$

Substituting these results back into the covariance calculation provides an estimate of the error on quantities after performing a single EM step.

Appendix C

If $\mathbf{\Delta}$ is a vector of quantity errors evaluated at a particular time, t, then the accumulation of error from one step to the next is given by

$$\mathbf{\Delta}|_t \approx (\mathbf{\Delta}|_{t-1})^T \nabla \mathbf{Q}|_{t-1} \tag{52}$$

where $\nabla \mathbf{Q}$ is the Jacobian

$$\nabla \mathbf{Q}_{ij}|_{t-1} = \frac{\partial \mathbf{Q}_i|_{t-1}}{\partial \mathbf{\Delta}_j} \tag{53}$$

The diagonal terms of the Jacobian are given by

$$\nabla \mathbf{Q}_{ii} = \frac{\partial \mathbf{Q}_i}{\partial \mathbf{\Delta}_i} \tag{54}$$

$$= \sum_X \frac{P(X|i)[P(X|i)\mathbf{Q}_i + P(X|\bar{i})\mathbf{Q}_{\bar{i}}] - P(X|i)^2\mathbf{Q}_i}{[P(X|i)\mathbf{Q}_i + P(X|\bar{i})\mathbf{Q}_{\bar{i}}]^2}\mathbf{H}_X \tag{55}$$

where $P(X|i)\mathbf{Q}_i + P(X|\bar{i})\mathbf{Q}_{\bar{i}} \approx \mathbf{H}_X$ giving

$$= \sum_X \frac{P(X|i)\mathbf{H}_X - P(X|i)^2\mathbf{Q}_i}{\mathbf{H}_X^2}\mathbf{H}_X \tag{56}$$

$$= \sum_X P(X|i) - \frac{P(X|i)^2\mathbf{Q}_i}{\mathbf{H}_X} \tag{57}$$

which via Bayes Theorem becomes

$$\nabla \mathbf{Q}_{ii} = \sum_X P(X|i) - P(X|i)P(i|X) \tag{58}$$

Similar treatment gives the off-diagonal terms

$$\nabla \mathbf{Q}_{ij} = \frac{\partial \mathbf{Q}_i}{\partial \mathbf{\Delta}_j} \tag{59}$$

$$= \sum_X \frac{-P(X|i)\mathbf{Q}_i P(X|j)}{[P(X|j)\mathbf{Q}_j + P(X|\bar{j})\mathbf{Q}_{\bar{j}}]^2}\mathbf{H}_X \tag{60}$$

$$= \sum_X \frac{-P(X|i)\mathbf{Q}_i P(X|j)}{\mathbf{H}_X^2}\mathbf{H}_X \tag{61}$$

$$= \sum_X -P(X|j)\frac{P(X|i)\mathbf{Q}_i}{\mathbf{H}_X} \tag{62}$$

i.e.

$$\nabla \mathbf{Q}_{ij} = \sum_X -P(X|j)P(i|X) \tag{63}$$

Assuming the derivative computed at any time is approximately equal, i.e. $\frac{\partial \mathbf{Q}_i|_{t-1}}{\partial \mathbf{\Delta}_j} \approx \frac{\partial \mathbf{Q}_i|_t}{\partial \mathbf{\Delta}_j}$, such that for all t then $\nabla \mathbf{Q}|_t \approx \nabla \mathbf{Q}$, then the error accumulation from the first step onwards becomes

$$\mathbf{\Delta}|_1 \approx \mathbf{\Delta}|_0 \nabla \mathbf{Q} \tag{64}$$

$$\mathbf{\Delta}|_2 \approx \mathbf{\Delta}|_1 \nabla \mathbf{Q} \approx \mathbf{\Delta}|_0 \nabla \mathbf{Q}^2 \tag{65}$$

$$\mathbf{\Delta}|_t \approx \mathbf{\Delta}|_0 \nabla \mathbf{Q}^t \tag{66}$$

The total vector change in the quantity $\mathbf{Q}$ is then given by

$$\mathbf{\Delta} \approx \sum_{t=0}^{\infty} \mathbf{\Delta}|_t = \mathbf{\Delta}|_0 + \sum_{t=1}^{\infty} \mathbf{\Delta}|_0 \nabla \mathbf{Q}^t \tag{67}$$

$$\mathbf{\Delta} \approx \mathbf{\Delta}^T|_0 \left[\mathbf{I} + \sum_{t=1}^{\infty} \nabla \mathbf{Q}^t\right] \;=\; \mathbf{\Delta}^T|_0 \left[\mathbf{I} - \nabla \mathbf{Q}\right]^{-1} \tag{68}$$

so that the total error amplification can be approximated by a single linear process

$$\mathbf{\Delta} \approx \mathbf{\Delta}^T|_0 \mathbf{A} \tag{69}$$

where the amplification matrix is

$$\mathbf{A} = \left[\mathbf{I} - \nabla \mathbf{Q}\right]^{-1} \tag{70}$$

Finally, the covariance matrix for the quantity vector can be given by scaling the one step covariance by the amplification matrix

$$\mathbf{C} = \mathbf{A}^{\intercal}\mathbf{C}_{EMStep}\mathbf{A} \tag{71}$$

Whilst this formulation looks considerably different from the original CRB covariance estimate the two are consistent.

Acknowledgements

We thank J. Gilmour and M. Jones, of the School of Earth, Atmospheric and Environmental Sciences, University of Manchester for their input and support. We also thank the Science and Technology Facilities Council for providing project funding.

References

[1] F.J. Anscombe, The transformation of Poisson, binomial and negative-binomial data, Biometrika 35, 3-4, 1948

[2] P.J. Ashworth, J.L. Best, M.A. Jones, The relationship between channel avulsion, flow occupancy and aggradation in braided rivers: insights from an experimental model, Sedimentology 54, 497-513, 2007

[3] C.M. Bishop, Neural Networks for Pattern Recognition, Clarendon Press, Oxford, 66 ff, 1995

[4] R.J. Barlow, Statistics: A Guide to the use of Statistical Methods in the Physical Sciences. John Wiley and Sons, U.K., 1989

[5] Y.H. Chen, Maximum likelihood analysis of semicompeting risks data with semiparametric regression models, Lifetime Data Analysis, 18(1):36 57, 2012

[6] P. Comon, Independent Component Analysis - a new concept? Signal Processing 36, 287-314, 1994

[7] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, Training models of shape from sets of examples, Proceedings of BMVC, 266-275, 1992

[8] A.P. Dempster, N. M. Laird, and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society, 39, 1-38, 1977

[9] D. Duvenaud, J.R. Lloyd, R. Grosse, J.B. Tenenbaum, Z. Ghahramani, Structure Discovery in Nonparametric Regression through Compositional Kernel Search, 30th International Conference on Machine Learning, W&CP volume 28, 2013

[10] R.M. Haralick, Land use Classification Using Texture Information in ERTS-A MSS Imagery, Remote Sensing Laboratory, NASA Technical Reports, E73-10461, 1973

[11] R.M. Haralick, On the Use of Error Propagation for Statistical Validation of Computer Vision Software, (with Xufei Liu and Tapas Kanungo), IEEE Pattern Analysis and Machine Intelligence 27, No. 10, pp. 1603-1614, 2005

[12] H.H. Harman, Modern Factor Analysis 2nd edition, University of Chicago Press, 1967

[13] A. Hyvarinen, Survey on Independent Component Analysis, Neural Computing Surveys 2, 94-128, 1999

[14] X. Jiang, J. Gao, T. Wang, L. Zheng, Supervised Latent Linear Gaussian Process Latent Variable Model for Dimensionality Reduction Systems, Man and Cybernetics Part B: Cybernetics, IEEE Transactions volume 42 issue: 6, 2012

[15] I.T. Jolliffe, Principal Component Analysis, Springer-Verlag, 1986.

[16] M. Kendall, Multivariate Analysis, Charles Griffin&Co., 1975

[17] F.W. McLafferty, F.T. Tureek, Interpretation of Mass Spectra, University Science Books, CA, 1993

[18] B.D. Ripley, Appendix A in Pattern Recognition and Neural Networks, Cambridge University Press, 1996