NCJRS
National Criminal Justice Reference Service

Office of Justice Programs
BJA BJS NIJ OJJDP OVC SMART

The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title:	**Statistical Methods for Combining Multivariate and Categorical Data in Postmortem Interval Estimation**
Author(s):	**Lynn R. LaMotte**
Document Number:	**250467**
Date Received:	**December 2016**
Award Number:	**2013-DN-BX-K042**

This resource has not been published by the U.S. Department of Justice. This resource is being made publically available through the Office of Justice Programs' National Criminal Justice Reference Service.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Final Summary Overview of Award 2013-DN-BX-K042

Statistical Methods for Combining Multivariate and Categorical Data in Postmortem Interval Estimation [1]

Principal Investigator: Lynn R. LaMotte

1 Purpose of the Project

Inferring time since death is a standard goal of a murder investigation. Many physical or biological variables that change as a function of postmortem interval (PMI) have been proposed as proxy clocks for PMI estimation. The questions inherent in such inferences are statistical; probabilistic properties of the statistical procedures must be known in order to represent uncertainty correctly. During this project we focused on estimating PMI using insect evidence because this field provided the largest data sets for model exploration. However our statistical methods are broadly applicable, and we hope to see them incorporated into all scientific disciplines concerned with PMI estimation.

The purpose of this project is to address the statistical questions that arise in the following examples.

1. A larva of *L. sericata* weighs 3.5mg and its length is 11.5mm ($y_* = (3.5, 11.5)$), how old is it (x_*)?

2. Insect species A and B are found on a decaying body, and species C and D are not (this response might be coded $y_* = (1, 1, 0, 0)$): how long (x_*) has the body been decaying?

3. A larva of *C. vicina* is 9mm long, and it is in the Larva II stage: how old is it?

[1]This project was supported by Award No. 2013-DN-BX-K042, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, conclusions or recommendations expressed in this publication/program/exhibition are those of the author(s) and do not necessarily reflect those of the Department of Justice.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice

Because sizes and succession combinations differ even when observed under identical conditions, there is no defensible, unique answer to each of these questions. These are questions that can only be addressed statistically. Only estimates of x_* are possible. Estimates stated as ranges communicate the uncertainty inherent in the estimate. These are called confidence intervals or confidence sets. Associated with each is a level of confidence, most commonly chosen to be 95%, that states the nominal probability that such intervals cover the true value. In applications, confidence sets entail several layers of approximations. Thus intervals that nominally have the 95% confidence level may in fact cover the true value only 85% of the time, or 99% of the time, and the true coverage rate is not known.

Inverse prediction (IP) is a statistical methodology that can be used to obtain confidence sets on x_*. In outline, here is how IP works. In a controlled experimental setting, subjects are sampled at each x over a range of values, and the response y is measured on each subject. The resulting data set is called the training data, abbreviated here TD. A model relating y to x is fit to the TD. This generally entails specifying a class of models for first and second moments, that is, means and variance-covariances of y, in terms of x and any other covariates that might be involved (such as temperature, for example). Given a mystery specimen (MS) with response y_*, its response is compared statistically to the model at each of a grid of potential values of the unknown x_* from which it originated. This results in a probability statement, a p-value, for each potential x_*. Those values for which the p-value is not small ($\geqslant 5\%$ is the conventional guide) constitute a confidence set on x_*.

IP was developed and most widely used for settings in which the response is quantitative and univariate, like weight, and where the assumed relation to x is linear. Further, it assumes that the variance of the response – the source of uncertainty – is constant under all conditions. Both growth data and succession data violate this restriction. Treatments of this setting that had appeared in the literature were not comprehensive, and they required complex computational algorithms that were not available in statistical computing packages. They entailed approximations, and their statistical properties (like coverage probabilities of

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice

confidence sets) were unknown.

When we began this project, there was only one basic result for IP for categorical responses. LaMotte and Wells (2000) derived exact p-values for comparing a given categorical outcome to training data.

The objective of this project is to extend and adapt methods of IP for use in forensically important settings. We proposed three main parts to this research. 1. To develop methods of IP for multivariate quantitative responses with heterogeneous variance-covariance structure, including flexible, adaptive modeling of both means and variances. 2. To develop methods of IP for categorical responses, with the same modeling flexibility. 3. To integrate these results to create IP methods for hybrid quantitative-categorical responses. A common objective in each of these is to assess and quantify the inherent uncertainty accurately and defensibly.

The over-arching objective is to develop a set of statistical methods and versatile, accessible tools for IP in settings important in assessing time since death and other forensically important applications. We have accomplished that objective by making it possible for IP methodology to be implemented within the broad context of mixed linear models, so that that its computations leading to p-values and confidence sets can be more easily performed because an investigator can now do the analysis using a variety of widely-available statistical computing packages. Within that framework, we have demonstrated that polynomial splines provide flexible, adaptive models for both the mean vectors and variance-covariance matrices.

It is equally important to assess and report the statistical properties of IP methods. An essential part of that assessment is to use real data in cross-validation and re-sampling studies. Dr. Wells has assembled a wide variety of growth data and succession data, both from his students' experiments and contributed by other scientists. His efforts will continue through the conclusion of this project. We have gathered those data centrally, and we have been putting them into a common format. In the end, we will have assembled a voluminous repository of data, in a form that can be accessed readily by us and by other researchers.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice

We have also submitted for publication scientific manuscripts to illustrate the immediate practical value of our inverse prediction methods for guiding PMI research design. In particular we have defined objective standards, never before proposed, for sampling methods, minimum required replication, and method validation.

2 Project Design, Methods, and Data Analysis

As described in our proposal, the project design was planned in three stages, for quantitative responses, for categorical responses, and for hybrid quantitative-categorical responses. All models relating y to x were to be linear-interpolation models, also called linear splines; and these were to be used to model both means and variances.

Following Wells and LaMotte (1995, 2010), these models interpolated between TD sampling points. Ad hoc programs were required for the complex computations of approximate denominator degrees of freedom required for the test statistic that provided the p-values.

With both real data and simulated data, the resulting model fits were often jerky from sample point to sample point, particularly for variances. Smoother results could be had by using fewer knots (join points) for the linear interpolation and locating them strategically, or by using higher-order interpolation, like cubic splines.

These considerations led to a major development in the modeling strategy, one that simultaneously took advantage of existing algorithms and programs. Formulating the models for the means and the variance-covariance matrices separately in terms of polynomial splines, the model combining both can be expressed mathematically as a mixed (fixed- and random-effects) multivariate linear model. In that way, all the computations required for IP can be accomplished using PROC MIXED in SAS, a program written by professional statisticians, numerical analysts, and programmers that has been widely used and vetted for three decades.

The program PROC MIXED doesn't include an option that one can simply choose for these models. Instead, it is set up to analyze data for models of the form $Y = X\beta + Z\gamma + \epsilon$, where Y is the observation vector of multivariate responses, X and Z are fixed matrices,

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice

and γ and ϵ are independent vector-valued random variables with zero mean vectors and variance-covariance matrices G and R, respectively. The program has syntax, a language, by which the user can affect the structures of X and Z and the parametric forms of G and R. The vocabulary is extensive; combinations of a few statements can produce millions of different model specifications. The challenge now is not to write the program, but to communicate with the program to specify precisely the desired model.

The choices that must be made require mastery of the statistical methodology. The choices are very much setting-dependent, and so it is neither feasible nor desirable to provide a general, one-size-fits-all macro package. We can describe the methods clearly in precise statistical and mathematical terms that can be implemented by a statistician who is fully conversant with mixed models theory and methods.

The objectives and some of the progress to date of the project are illustrated in the first appendix. Figure 1 shows linear-interpolation models for means and variances fit to simulated training data. It shows IP p-values computed for each possible value of the condition variable t and the approximate 95% confidence interval on the true t_*. This follows the method first described by Wells and LaMotte (1995).

That method required special programming, and it relied on there being multiple specimens observed at each age. By placing these methods and models in the context of mixed models, as noted above, all the computations were done using PROC MIXED in SAS. Other than data handling and preparation for the mixed models computations, no special programming was required.

Figure 2 illustrates an additional milestone in this project. We proposed originally to use linear interpolation models. More flexibility can be had using polynomial splines, of which linear splines are a special case. We have shown how polynomial spline models in both means (already known and used) and variance-covariance matrices (not used before) fit as special cases into the general framework of mixed linear models. Figure 2 also illustrates that the modeling methods work even when observations are all at different values of t (denoted x

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice

above), thus not requiring, as formerly, that variances be estimable at each sampled time in the training data.

Figure 4 illustrates the assessment of coverage probabilities of nominally 95% IP confidence sets on t_*. The essential feature for $t_* = 5$, for example, is that the coverage rate at $t_* = 5$ is quite close to 95%. For specimens from $t_* = 2$, the coverage rate is a little greater than 95%, and for true $t_* = 8$, it is almost exactly 95%. The curves also indicate how discriminatory the confidence sets are: the confidence sets for $t_* = 2$ do not distinguish the values less than 2 from 2, but the coverage rate falls off steeply for values greater than 2.

Table 1 and Figure 5 illustrate the IP methods that we have developed in this project for categorical responses. For p-values, the basic result derived in LaMotte and Wells (2000) is used. See the captions for explanations.

Appendix 2 is the manuscript accepted for publication by *Communications in Statistics – Simulation and Computation* and alread published online. Appendix 3 comprises the presentation at the International Workshop on Matrices and Statistics in Haikou, China, cited among presentations below. It details the steps required to specify the models and to obtain p-values for IP using PROC MIXED in SAS. See particularly the graphical animations showing IP for a bivariate quantitative response, using models in linear and cubic splines. Appendix 4 lists publications cited here.

3 Scholarly Products

Published articles

Wells JD, Lecheta MC, Moura MO, LaMotte LR. 2015. An evaluation of sampling methods used to produce insect growth models for postmortem interval estimation. *International Journal of Legal Medicine* **129**(2): 405-410, Mar. 2015.

LaMotte LR, Wells JD. 2016. Inverse Prediction for Heteroscedastic Response Using Mixed Models Software. Accepted for publication by *Communications in Statistics – Simu-*

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice

lation and Computation.

`http://www.tandfonline.com/eprint/kEWiz9SMejyJW4pZQTaW/full`

Articles under review

LaMotte LR, Wells JD. 201?. Inverse Prediction for Multivariate Mixed Models with Standard Software. Under review by *Statistical Papers.*

LaMotte LR, Roe A, Wells JD, Higley LG. 201?. Inverse prediction of Lucilia sericata age based on developmental stage (Diptera: Calliphoridae). Under review by *Journal of Forensic Sciences.*

Presentations

Jeffrey Wells, Melise Lecheta, Mauricio Moura, Lynn LaMotte. "An evaluation of sampling methods used to produce forensic entomology growth models." Oral presentation at the Entomological Society of America annual meeting, Portland, OR, November 19, 2014.

LaMotte LR. 2014. Conducted a four-hour workshop on inverse prediction presented at the 12th annual meeting of the North American Forensic Entomology Association, July 15-16, 2014.

LaMotte LR, Wells JD. 2015. Mixed Interpolation Models for Inverse Prediction Using Standard Software. Invited plenary presentation at the 24th International Workshop on Matrices and Statistics, Haikou, China, May 25-28, 2015.

LaMotte LR, Wells JD. 2015. Mixed Interpolation Models for Inverse Prediction Using Standard Software. Plenary lecture at the 7th International Conference on Probability and Statistics, Smolenice Castle, Slovakia, June 29 - July 3, 2015.

LaMotte LR, Wells JD. 2015. Multivariate Inverse Prediction with Mixed Models. Invited lecture at the 2015 Lloyd Roeling Conference, University of Louisiana at Lafayette, November 2021, 2015.

LaMotte LR, Wells JD. 2016. Multivariate Inverse Prediction (Calibration) Using Standard Software for Mixed Models. Accepted for presentation at the 2016 American Statistical

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice

Association Conference on Statistical Practice, Feb. 18-20, 2016, San Diego, California.

Articles in preparation

Wells JD and LaMotte LR. 201?. Assessing the Efficacy of ADH in Capturing Effects of Age and Temperature on Sizes of Developing Insect Larvae.

LaMotte LR and Wells JD. 201?. Inverse Prediction Based Simultaneously on Quantitative and Categorical Traits.

Watters CG, LaMotte LR. 201?. On Accuracy of P-Values Comparing Heterogeneous Categorical Responses.

It is clear that both time and temperature affect growth and decay. It is widespread practice to reduce these two factors to one dimension, in the form of accumulated degree days or hours (ADD or ADH). We have seen, in multiple real data sets, both quantitative and categorical, that observations at the same ADH but different time-temperature combinations differ significantly. It remains to be assessed whether ADH can be used without materially distorting inferences. We expect to produce a paper on this important issue.

Another question that has appeared regards basic p-values for categorical responses as shown in LaMotte and Wells (2000). A basic assumption of that computation is that the training data result from independent multi-category Bernoulli trials, all with the same probability distribution. Often rearing experiments that yield development data are conducted on batches of larvae, which are processed together. The batch is the basic experimental unit. With multiple batches at a given time-temperature combination, as in the data depicted in the second section of the first appendix, the question arises, how combining counts from several batches into one frequency distribution (of stages, for example) affects the accuracy of the p-value. The graduate research assistant on this project, C. G. Watters, is pursuing that question, as indicated in the list above.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice

4 Implications for Policy and Practice

To our knowledge there is no conventional, generally accepted statistical approach to estimate PMI from postmortem decay data and to quantify the attendant uncertainty. We believe that the value to the criminal justice system of developing robust statistical methods for calculating a confidence interval about a PMI estimate is obvious. While this project is an example of basic research, we have developed tools that are of immediate practical value to scientists involved in PMI estimation research and casework.

Research: Because almost all PMI research has been conducted without an explicit model in mind for how the data would be used to predict PMI, researchers have never considered how to design experiments so as to optimize predictive model performance. Consequently a great deal of published death investigation research was not as valuable as it could have been, sometimes with little or no extra effort by the researchers. Using our models we have demonstrated that most (but not all) forensic entomology development models are likely to produce inaccurate estimates of specimen age (= minimum PMI) because of the sampling scheme used (Wells et al. 2015), i.e. we have proposed an objective mathematically-supported standard for the design of such experiments. Our models also provide standards for the amount of replication needed in a training experiment in order for the resulting model to have enough statistical power to be used in casework. We had earlier done this for insect succession (LaMotte and Wells 2000), but as part of our collaboration with PIs from another recent NIJ-funded study (LaMotte, Roe, Wells, Higley, under review) we have defined minimum sample sizes needed for predicting PMI using other categorical responses such as insect life stage or corpse tissue color.

Our methods also open the door to a properly designed PMI prediction model validation study. Obviously a predictive model should not be expected to correctly predict the exact time since death, but how close is close enough? The 95% confidence interval defines an objective standard for whether or not the prediction can be considered to have included the true age, and for the proportion of validation observations that must be correct in order to

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice

support the model.

Casework: From our discussion with other forensic scientists, it appears that a main obstacle to the adoption of inverse prediction for PMI estimation was that such models are not implemented in standard statistical software packages. We have remedied that in two ways. We have shown how IP for one or more continuous quantitative variables can be implemented for a wide variety of postmortem settings in the context of the widely used and understood PROC MIXED procedure in SAS (LaMotte and Wells, 2016). For a categorical response, we have shown how a 95% inverse prediction interval about a PMI estimate can be calculated by simply consulting a table (LaMotte, Roe, Wells, Higley, under review). These are methods that a death investigator can use at the present time to evaluate 1) whether an existing reference data set provides enough statistical power for PMI estimation, 2) whether it has been sufficiently validated, and then 3) apply these methods to an actual death investigation.

Lynn LaMotte 12/09/2015

(Signature) (Date)

Lynn R. LaMotte, PhD

Principal Investigator

Joseph Moerschbaecher 12/14/15

(Signature) (Date)

Joseph M. Moerschbaecher, III PhD

Official Authorized to Sign for Institution

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice