AMSI VACATION RESEARCH SCHOLARSHIPS 2019–20

EXPLORE THE MATHEMATICAL SCIENCES THIS SUMMER

The L_2 Method for Robust Estimation of Mixtures

Mr. George L. Malone

Supervised by Dr. Brenton Clarke and Dr. Alethea Rea

Murdoch University

Vacation Research Scholarships are funded jointly by the Department of Education and the Australian Mathematical Sciences Institute.

Abstract

A condition in horses known as pituitary pars intermedia dysfunction (PPID) can be diagnosed by examining adrenocorticotropic hormone (ACTH) levels in blood plasma. Due to the gradual nature of PPID, it is possible for a number of horses with the condition to be diagnosed as healthy, as their ACTH levels appear normal. This results in a single set of natural-logged ACTH levels, encompassing all horses examined, with an underlying distribution consisting of two mixed normal distributions. Further confusion possibly caused by influence from outlying values has made the clinical threshold of diagnosis difficult to define to date. This report seeks to justify the usage of a robust statistical method in estimating the parameters of ACTH level distribution in mixed samples of horses with and without PPID, for the purpose of providing greater clarity in defining said clinical thresholds. Monte Carlo simulation is used to assess the accuracy and precision of the robust parameter estimation method against a maximum likelihood estimation method. The robust parameter estimation method is found to demonstrate greater accuracy and precision than the maximum likelihood method, as well as lower generalised and total variances implying greater reliability, and lower calculation time despite greater mathematical complexity. This exploration supports work currently under review for publication whereby clinical thresholds for ACTH concentrations for PPID diagnosis are defined across an annual period with a weekly temporal resolution.

1 Introduction

1.1 Background – Biology

Adrenocorticotropic hormone (ACTH), in the equine system, stimulates cortisol hormone synthesis in the subject's adrenal glands (Bowen, 2019; Abercrombie et al., 1970, p. 9). If cortisol synthesis is abnormally high, this can be associated with a number of symptoms, such as lack of shedding winter coat, lethargy, excess sweating, increased water intake and urination, loss of muscle mass, infertility, and infections due to immunosuppression, including laminitis (Graves, 2015; McGowan et al., 2013; Schott, 2002; Okada et al., 1997). Abnormal ACTH regulation in horses can be caused by pituitary pars intermedia dysfunction (PPID), a condition whereby the pars intermedia in the pituitary enlarges, resulting in excess production of ACTH (AAEP, 2019; Okada et al., 1997). PPID can be diagnosed by examining the ACTH levels in the equine subject, typically through blood plasma testing.

ACTH levels tend to remain consistent over the course of the year, but depart notably around autumn, with ACTH starting to rise around the summer solstice, peaking mid-autumn, and decreasing more rapidly towards the onset of winter. Subjects with normal pituitary function see a rise in median ACTH concentration of approximately 20 pg/ml. However, subjects diagnosed with PPID see ACTH levels depart drastically more so than healthy horses, with median levels rising by approximately 59 pg/ml (Copas and Durham, 2012). This is used as part of the diagnostic process for PPID – if the initial blood sample results are unclear, take another sample in autumn around the peak period and check the ACTH concentration. This can result in a wait of up

to 10 months to collect a sample, during which time the possible PPID will progress.

The separation between what defines a normal or abnormal level of plasma ACTH is not entirely clear. This is partially due to the nature of PPID – because it is gradual, a horse that has PPID may have minor or major effects. A more accurate estimation of the parameters of ACTH distributions could contribute to clarifying a diagnostic threshold for assessing occurrence of PPID. Samples of the logarithm of ACTH concentration (in picograms per millilitre) are assumed to be a mixture of two normal distributions, with different parameterisations per component, where the sample is a single mixed dataset due to the unclear nature of PPID diagnosis. That is, a horse with relatively normal blood plasma ACTH concentration may actually have PPID, but will be classified as healthy because the ACTH levels are not extreme (Durham et al., n.d.), but horses with more advanced PPID will demonstrate a different set of parameters defining logged ACTH concentration. Hence, the data are a mixture of two distinct populations – PPID and non-PPID horses.

1.2 Background – Robustness

There are two main parameter estimation methods used in this exploration – the EM algorithm, for maximum likelihood estimates, and the L_2 minimum distance estimation method, a robust parameter estimation technique. The EM algorithm is susceptible to influence from outlying values, warping the "trueness" of the estimated parameters. The L_2 method is asserted to be a better alternative, as it is not susceptible to such errors from influence by outlying values.

The EM algorithm estimates are susceptible to outlying values because maximum likelihood estimates have unbounded influence. This means that a single outlying point has potentially unlimited affect on contaminating the estimated values. The L_2 method has a bounded influence function, meaning that the influence of outlying points is greatly restricted or certainly bounded, resulting in limited influence affecting the estimated values (Clarke et al., 2017). This makes the L_2 parameter estimates robust to outlying values, possibly resulting in parameters that are more accurate to true parameter values and more precise by demonstrating lower variability in estimation.

Figure 1 demonstrates the effect of a single perturbed point on the estimated parameters of a two-component normal mixture. Parameters are estimated by the EM algorithm, for maximum likelihood estimates, and the L_2 method, for robust parameter estimates. It is clear that the maximum likelihood estimates can be significantly influenced by a single datum, whereas the L_2 estimates are not influenced once this point is no longer near the centre of mass of the data.

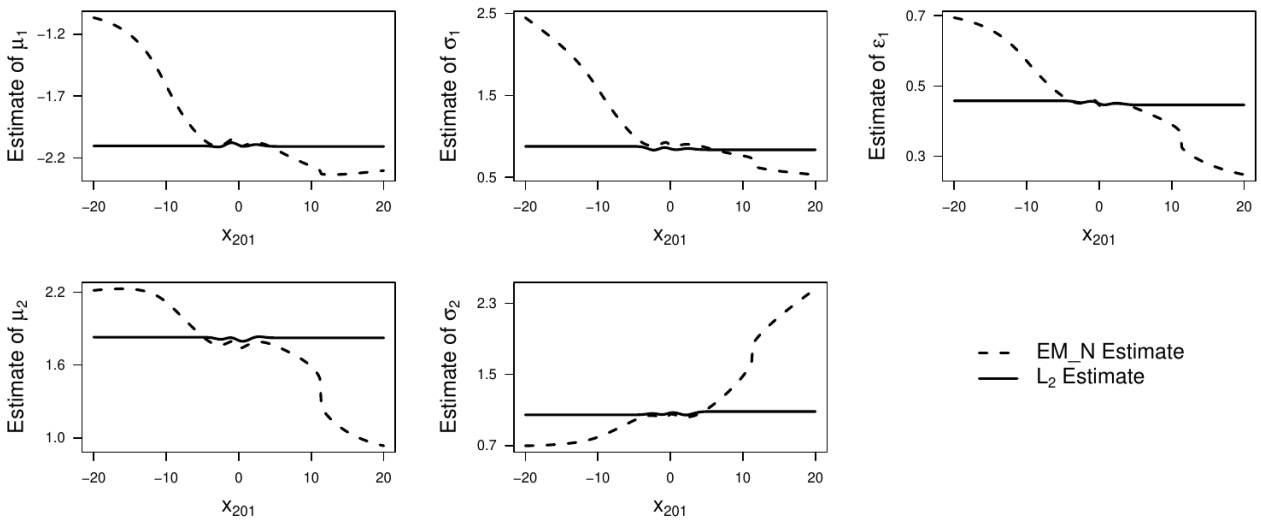

Figure 1: Adapted from Clarke et al. (2017). Effect of single-point influence on estimated values via EM algorithm (maximum likelihood) and L_2 (robust) methods. Data are a set of 200 randomly sampled observations from 2-component normal mixture with parameters: $\epsilon_1 = 0.5$, $\mu_1 = -2$, $\mu_2 = 2$, $\sigma_1 = 1$, $\sigma_2 = 1$.
The x-axis shows the value of x_{201}, the 201^{st} point, varied between -20 and 20 to observe the effect on the maximum likelihood parameter estimates (dotted line) and robust parameter estimates (solid line).

1.3 Background – Mixtures

Mixtures of probability densities are essentially a weighted sum of component densities, as shown in Equation 2. In a k-component mixture, all ϵ components must sum to 1, therefore in a two-component mixture, as in Equation 3, $\epsilon_2 = 1 - \epsilon_1$. The ϵ_2 component is not estimated, hence not displayed in the results, due to this property.

Given that the density function of the normal distribution is defined by

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{1}$$

then the k-component normal mixture is defined by:

$$\sum_{i=1}^{k} \frac{\epsilon_i}{\sigma_i} \cdot \phi\left(\frac{x - \mu_i}{\sigma_i}\right) \tag{2}$$

hence, two-component normal mixture:

$$\frac{\epsilon_1}{\sigma_1} \cdot \phi\left(\frac{x - \mu_1}{\sigma_1}\right) + \frac{\epsilon_2}{\sigma_2} \cdot \phi\left(\frac{x - \mu_2}{\sigma_2}\right) \tag{3}$$

The ϵ_i element can therefore be considered as the proportion of the overall probability density attributed to component i.

1.4 Background – Data

The original data, provided by Dr. Brenton Clarke, from Professor Andy Durham of the Liphook Equine Hospital, were assumed to consist of a two-component mixture of normal densities. The data provided for this exploration were the set of L_2 parameter estimates of the two-component normal mixture ln (ACTH concentration) (pg/mL) of $75,892$ individual subjects across the year, consisting of a mixture of horses and ponies of varying breed and age, and of varying sample size per week (Durham et al., n.d.). The parameter estimates are representative of the means, standard deviations, and ϵ components of both PPID and non-PPID components.

Figure 2 displays the set of weekly mean ACTH levels across one year, whereby the data are from the United Kingdom, hence week 1 is in winter. It can be seen that the ACTH levels for horses diagnosed as non-PPID are not particularly variable, whereas horses diagnosed with PPID see much more variable mean ACTH levels. This is likely due to the nature of PPID, whereby the possible effect of PPID on ACTH levels can be minor or major – that is, a horse diagnosed with PPID may be in very early stages of the condition or may have a significantly advanced condition. It is also clear that the separation between ACTH levels in the two groups is significantly greater within the autumn period (around weeks 30 to 50). This is also due to the nature of PPID – horses with even minor PPID demonstrate much higher ACTH levels in the autumn period.

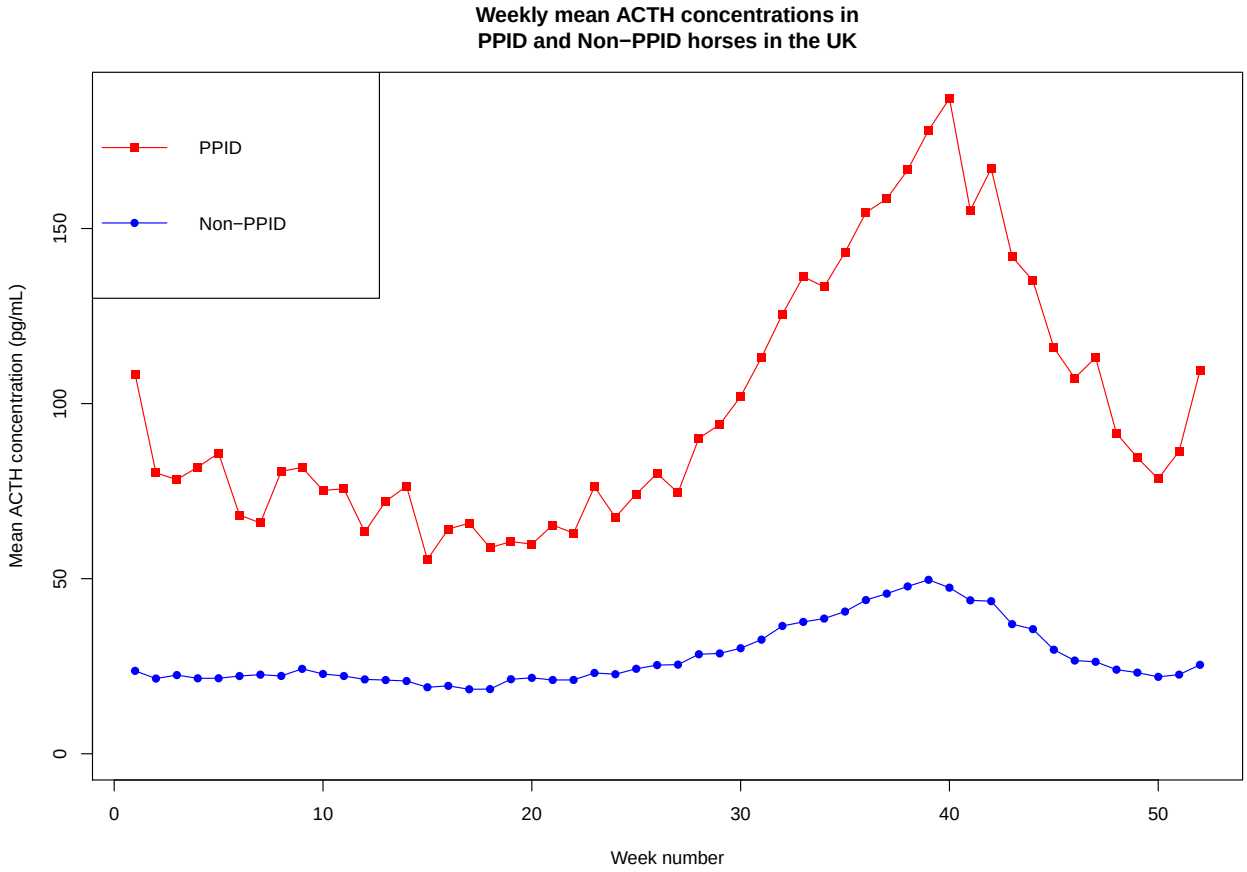

Figure 2: Weekly mean ACTH levels, picograms/millilitre. Data provided by Professor Andy Durham of the Liphook Equine Hospital, Hampshire, United Kingdom.

2 Statement of Authorship

The paper by Clarke, Davidson and Hammarstrand Clarke et al. (2017) provides the statistical concept and backing for the comparison between the EM algorithm for maximum likelihood estimation and the L_2 method for robust parameter estimation of mixtures. This work is a supplementary exploration to support the paper by Durham, Clarke, Potier and Hammarstrand (n.d.), which is currently undergoing review for publication. The original data were provided to Clarke by Durham, from which the weekly L_2 parameter estimates for the two-component mixtures of normal densities were provided by Clarke. The *R* scripts for simulating data and parameter estimation were written by Davidson, Hammarstrand and Clarke. These scripts were restored to working condition in consultation with Hammarstrand and Clarke, mainly due to deprecation of one package dependency. Missing components regarding number of algorithm failures and time taken to calculate were restored and the scripts converted to usage as a function by the author. Scripts were written by the author to produce raw LaTeXtable entries with bold highlighting from parameter estimation results.

3 Method

The method used to demonstrate the preference for the L_2 method over the EM algorithm is Monte Carlo simulation. Monte Carlo simulation is repeated random sampling of a stochastic event such that the range of possible outcomes of the event may be observed. In this case, this is used to assess the accuracy and precision of the parameter estimation techniques. One hundred simulations of size n, whereby n is chosen to imitate the original data, are performed given each weekly set of L_2 parameter estimates, provided by Dr. Brenton Clarke. The estimated parameter is the mean value of the parameter across one hundred simulations. Accuracy is measured by the proximity of the estimated parameter to the input, or "true", value. Single-parameter precision is measured by the mean squared error of the parameter estimate across one hundred simulations. Overall precision of the methods, considered a snapshot of the overall variability of the method, is given by the determinant and trace of the covariance matrix of estimated parameters, giving the generalised variance and total variance respectively.

Scripts were provided to perform Monte Carlo simulation of mixtures of normally-distributed data. Additional modifications were made to restore an additional two output components for the L_2 method scripts, in particular the number of algorithm failures and calculation time. Once these scripts were restored, they were converted into functions for ease of use and for producing a more compact script for output production and analysis. The result of these efforts is that, for the simulation aspect of the computation, two functions are prepared, for simulation and estimation using the EM algorithm and the L_2 method. Seed values and sampling method for randomised data sampling are identical between both simulation methods.

A subset of L_2 parameter estimates of weekly ACTH levels was selected by pseudo-randomisation, and was found to be suitably representative of the trends within the data. Five weeks within the autumn period were randomly selected and five weeks outside the autumn period were randomly selected, and the set of week indices combined and sorted in ascending order. The data used as input values in the simulations are the L_2 two-component normal mixture parameter estimates of weekly $\ln(\text{ACTH concentration})$. The subset of parameters is provided in Table 1. Distributions representative of the logged data are simulated using the modified scripts, and the simulated parameters estimated by the EM algorithm and L_2 method assuming the data are a mixture of two normal distributions. The output of the simulations is a list of values containing the mean estimated parameter value, mean squared error per parameter value, and determinant and trace of covariance matrix of parameter estimates representing the generalised variance and total variance respectively.

Graphs of mixed two-component normal densities are provided in Figure 3, with individual components highlighted, to provide an overview of the typical behaviour of the data. These sets of graphs clearly demonstrate the shifting location of the two components over the annual cycle, and show the high variability of the PPID component in context of the relatively low variability of the non-PPID component. Also notable is the changing category, visible as changing area of component densities, of a portion of subjects previously considered non-PPID outside the autumn period but now classified as potential PPID subjects within the autumn period (approx. weeks 30-45).

Week	ϵ_1	μ_1	μ_2	σ_1	σ_2
1	0.7548	3.1644	4.6863	0.4530	1.0400
13	0.7571	3.0471	4.2779	0.4320	0.9780
24	0.7091	3.1231	4.2126	0.4300	0.9200
29	0.6069	3.3554	4.5435	0.4410	1.0160
33	0.5429	3.6284	4.9144	0.5260	1.0130
34	0.4989	3.6535	4.8929	0.4510	1.0050
37	0.3967	3.8228	5.0658	0.4370	0.9910
40	0.4731	3.8592	5.2318	0.5220	1.0130
43	0.5059	3.6113	4.9556	0.4210	1.0410
45	0.5762	3.3915	4.7534	0.4150	1.0850

Table 1: L_2 parameter estimates of $\ln$ (ACTH concentration) per corresponding week.

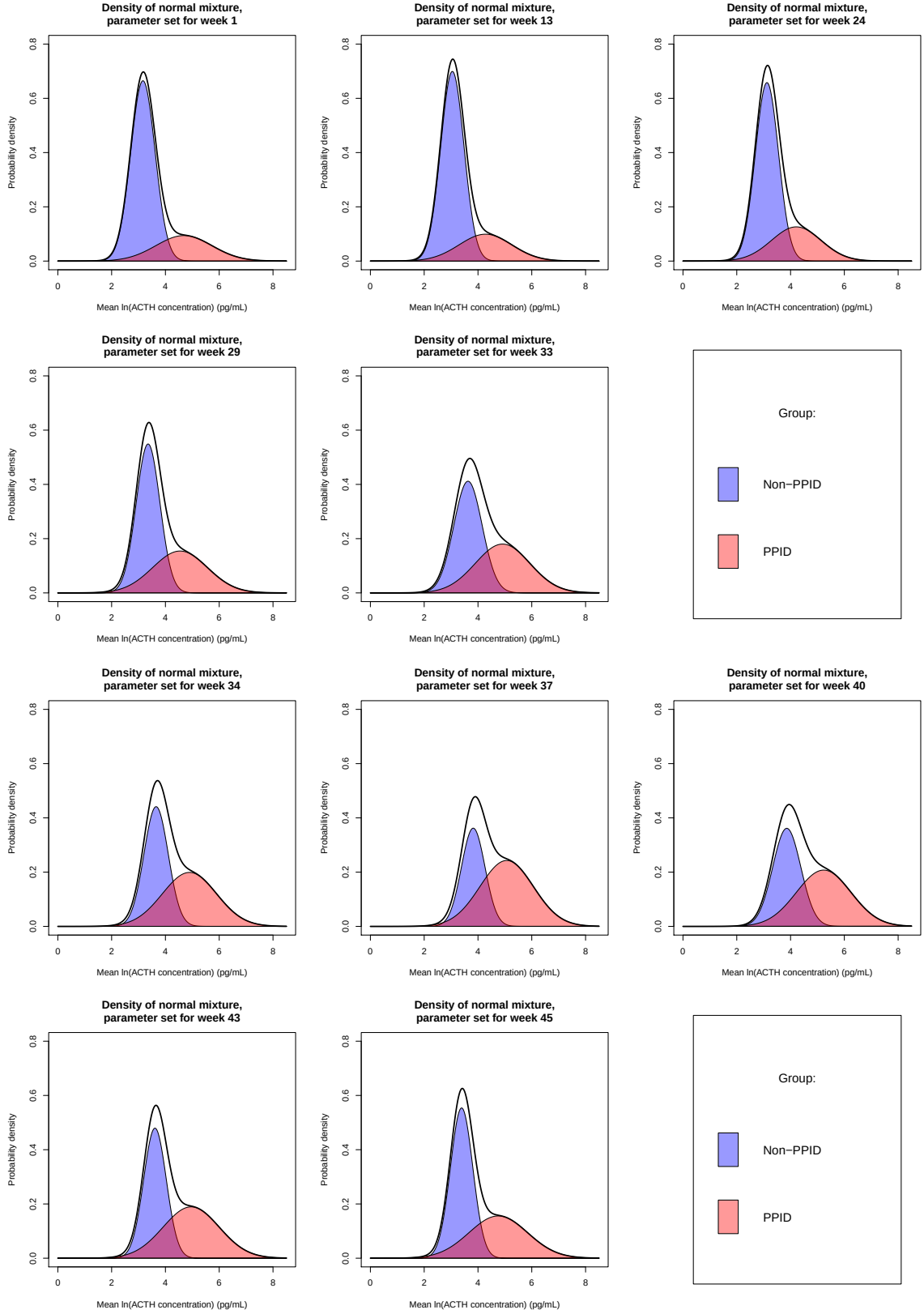

Figure 3: Representative probability densities of ln (ACTH concentration), across weeks randomly selected for simulation. PPID component highlighted in red, non-PPID component highlighted in blue.

4 Results

The results for EM algorithm maximum likelihood estimates are compared to the L_2 method minimum distance estimates. In measurements of accuracy, the L_2 method was found to be closer to the true or input values for most parameters. In measurements of precision, the L_2 method demonstrated lower mean squared error for most parameters, and showed minimal generalised and total variances in all parameter sets estimated. Tables 2, 3, and 4 provide the results for accuracy, precision by mean squared error, and precision by generalised and total variances, respectively. Overall, in direct comparison to the EM algorithm for maximum likelihood estimation, the L_2 parameter estimation method demonstrates greater accuracy by trueness in most individual parameters, greater precision by minimal mean squared error in most individual parameters, and greater precision by minimal generalised and total variances in all parameter sets. Of further interest may be the time taken to calculate – the mean time per set for the L_2 parameter sets was 1.88 seconds, against a mean time per set of 3.07 seconds for the EM algoritm.

Week	n	Set	ϵ_1	μ_1	μ_2	σ_1	σ_2
1	781	True	0.7548	3.1644	4.6863	0.4530	1.0400
		EM	0.7575	3.1730	**4.7040**	0.4557	1.0270
		L_2	**0.7527**	**3.1690**	4.6660	**0.4529**	**1.0400**
13	842	True	0.7571	3.0471	4.2779	0.4320	0.9780
		EM	0.7722	3.0510	4.3840	0.4360	0.9261
		L_2	**0.7542**	**3.0450**	**4.2790**	**0.4295**	**0.9762**
24	2018	True	0.7091	3.1231	4.2126	0.4300	0.9200
		EM	0.7215	3.1270	4.2730	**0.4310**	0.8913
		L_2	**0.7062**	**3.1220**	**4.2080**	0.4265	**0.9180**
29	1499	True	0.6069	3.3554	4.5435	0.4410	1.0160
		EM	**0.6122**	3.3570	4.5710	**0.4397**	1.0100
		L_2	0.6011	**3.3540**	**4.5340**	0.4338	**1.0210**
33	1381	True	0.5429	3.6284	4.9144	0.5260	1.0130
		EM	0.5595	3.6370	4.9860	**0.5285**	0.9821
		L_2	**0.5328**	**3.6230**	**4.9080**	0.5176	**1.0100**
34	1445	True	0.4989	3.6535	4.8929	0.4510	1.0050
		EM	**0.5054**	3.6570	4.9280	**0.4540**	0.9891
		L_2	0.4893	**3.6500**	**4.8880**	0.4453	**1.0040**
37	2053	True	0.3967	3.8228	5.0658	0.4370	0.9910
		EM	0.4014	3.8300	5.0750	**0.4350**	0.9879
		L_2	**0.3972**	**3.8260**	**5.0660**	0.4325	**0.9903**
40	2612	True	0.4731	3.8592	5.2318	0.5220	1.0130
		EM	0.4824	3.8650	5.2590	**0.5224**	1.0020
		L_2	**0.4695**	**3.8580**	**5.2270**	0.5163	**1.0130**
43	2611	True	0.5059	3.6113	4.9556	0.4210	1.0410
		EM	0.5103	3.6140	4.9650	0.4230	1.0330
		L_2	**0.5093**	**3.6130**	**4.9620**	**0.4222**	**1.0340**
45	1745	True	0.5762	3.3915	4.7534	0.4150	1.0850
		EM	**0.5764**	**3.3920**	**4.7580**	**0.4120**	**1.0840**
		L_2	0.5728	3.3900	4.7450	0.4100	1.0860

Table 2: Results for accuracy, whereby the results marked in bold are those which are closer to the input or "true" values, and are hence considered to be the more accurate result.

Week	n	Set	ϵ_1	μ_1	μ_2	σ_1	σ_2
1	781	EM	0.0015	0.0007	0.0466	0.0005	0.0089
		L_2	**0.0010**	**0.0006**	**0.0122**	**0.0004**	**0.0053**
13	842	EM	0.0023	0.0005	0.0616	0.0005	0.0134
		L_2	**0.0012**	**0.0005**	**0.0112**	**0.0004**	**0.0044**
24	2018	EM	0.0019	0.0003	0.0294	0.0002	0.0054
		L_2	**0.0006**	**0.0002**	**0.0047**	**0.0001**	**0.0015**
29	1499	EM	0.0014	**0.0005**	0.0143	**0.0004**	0.0025
		L_2	**0.0010**	0.0006	**0.0038**	0.0005	**0.0017**
33	1381	EM	0.0047	0.0020	0.0365	0.0012	0.0065
		L_2	**0.0019**	**0.0012**	**0.0069**	**0.0009**	**0.0018**
34	1445	EM	0.0024	0.0008	0.0149	0.0008	0.0038
		L_2	**0.0012**	**0.0007**	**0.0046**	**0.0007**	**0.0022**
37	2053	EM	0.0016	0.0008	0.0063	0.0009	0.0013
		L_2	**0.0013**	**0.0008**	**0.0040**	**0.0009**	**0.0011**
40	2612	EM	0.0023	0.0012	0.0115	0.0008	0.0023
		L_2	**0.0012**	**0.0009**	**0.0056**	**0.0007**	**0.0013**
43	2611	EM	0.0010	0.0004	0.0066	0.0004	**0.0014**
		L_2	**0.0008**	**0.0003**	**0.0052**	**0.0003**	0.0015
45	1745	EM	0.0007	0.0004	0.0083	0.0003	0.0017
		L_2	**0.0005**	**0.0004**	**0.0040**	**0.0003**	**0.0017**

Table 3: Results for precision by mean squared error (MSE), whereby the results marked in bold are those with minimal MSE, and are hence considered to be the more precise result.

Week	n	Set	Determinant	Trace
1	781	EM	1.159×10^{-14}	0.0583
		L_2	$\mathbf{5.336 \times 10^{-15}}$	**0.0192**
13	842	EM	8.151×10^{-15}	0.0647
		L_2	$\mathbf{2.228 \times 10^{-15}}$	**0.0178**
24	2018	EM	1.036×10^{-16}	0.0328
		L_2	$\mathbf{3.072 \times 10^{-17}}$	**0.0073**
29	1499	EM	3.118×10^{-16}	0.0184
		L_2	$\mathbf{1.259 \times 10^{-16}}$	**0.0074**
33	1381	EM	5.748×10^{-15}	0.0448
		L_2	$\mathbf{1.198 \times 10^{-15}}$	**0.0126**
34	1445	EM	1.175×10^{-15}	0.0214
		L_2	$\mathbf{4.525 \times 10^{-16}}$	**0.0093**
37	2053	EM	2.654×10^{-16}	0.0108
		L_2	$\mathbf{2.149 \times 10^{-16}}$	**0.0082**
40	2612	EM	3.097×10^{-16}	0.0173
		L_2	$\mathbf{1.553 \times 10^{-16}}$	**0.0096**
43	2611	EM	5.442×10^{-17}	0.0097
		L_2	$\mathbf{4.773 \times 10^{-17}}$	**0.0080**
45	1745	EM	1.571×10^{-16}	0.0115
		L_2	$\mathbf{9.154 \times 10^{-17}}$	**0.0069**

Table 4: Results for precision by generalised and total variances (determinant and trace respectively), whereby the results marked in bold are those with minimal variances, and are hence considered to be the more precise results.

5 Discussion

In the context of estimating the parameters of the mixed two-component normal density of ACTH concentration in blood plasma, the L_2 method is shown to be more applicable than the EM algorithm. The robust parameter estimates produced by the L_2 method demonstrate greater accuracy and precision per parameter, as well as greater overall precision, than the EM algorithm maximum likelihood estimates. Further, although the L_2 estimation method is mathematically more complex than maximum likelihood estimation, and the maximum likelihood method is theoretically more efficient, the L_2 estimates were also quicker to calculate, implying greater practical efficiency. The efficiency of the L_2 method is shown to be dependent on the parameters estimated, whereby it is more efficient than the EM algorithm in calculation when the component densities are not well separated (Clarke et al., 2017). The component densities in these data are not well separated, hence the L_2 method was found to be more efficient in this case.

This work comprises a supplementary analysis to support the paper by Durham et al. (n.d.), which is currently undergoing the review process for publication. Durham et. al. intends to provide clinical reference values for PPID diagnosis by ACTH concentration in blood plasma with a temporally-specific resolution, in particular weekly resolution but broadly categorised as within and outside of the autumn period. The supplementary evidence within this report is in support of the usage of the L_2 robust parameter estimation method for defining the clinical thresholds in Durham et. al.

For the parameter values covered in this study, the L_2 method exhibited greater accuracy and precision. This robust estimation method is shown to be significantly better in defining the clinical thresholds in Durham et al. (n.d.), in particular comparison to the more traditional EM algorithm for maximum likelihood estimates. The demonstrated results of greater accuracy and precision in defining the said thresholds could allow for greater confidence in positive diagnosis of pituitary pars intermedia dysfunction, potentially reducing the expense of materials and time for veterinarians, and allowing for earlier treatment of PPID, reducing the possible pain and distress to the horse.

Bibliography

Abercrombie, M., Hickman, C. J. and Johnson, M. L. (1970), *A Dictionary of Biology*, fifth ed. edn, Penguin Books Ltd., Harmondsworth, Middlesex, England.

American Association of Equine Practitioners (2019), 'Equine Cushing's Disease'.
URL: *https://aaep.org/horsehealth/equine-cushings-disease-equine-pituitary-pars-intermedia-dysfunction*

Bowen, R. (2019), 'Glucocorticoids'.
URL: *http://www.vivo.colostate.edu/hbooks/pathphys/endocrine/adrenal/gluco.html*

Clarke, B. R., Davidson, T. and Hammarstrand, R. (2017), 'A comparison of the L2 minimum distance estimator and the EM-algorithm when fitting k-component univariate normal mixtures', *Statistical Papers* **58**(4), 1247–1266.
URL: *https://doi.org/10.1007/s00362-016-0747-x*

Copas, V. E. N. and Durham, A. E. (2012), 'Circannual variation in plasma adrenocorticotropic hormone concentrations in the UK in normal horses and ponies, and those with pituitary pars intermedia dysfunction', *Equine Veterinary Journal* **44**(4), 440–443.
URL: *https://beva.onlinelibrary.wiley.com/doi/abs/10.1111/j.2042-3306.2011.00444.x*

Durham, A., Clarke, B. R., Potier, J. and Hammarstrand, R. (n.d.), 'Temporally specific diagnostic thresholds for plasma ACTH in the horse', *Equine Veterinary Journal* . Manuscript submitted for publication.

Graves, E. A. (2015), 'Equine Endocrine Diseases'.
URL: *https://aaep.org/horsehealth/equine-endocrine-diseases-basics*

McGowan, T. W., Pinchbeck, G. P. and McGowan, C. M. (2013), 'Prevalence, risk factors and clinical signs predictive for equine pituitary pars intermedia dysfunction in aged horses', *Equine Veterinary Journal* **45**(1), 74–79.
URL: *https://beva.onlinelibrary.wiley.com/doi/abs/10.1111/j.2042-3306.2012.00578.x*

Okada, T., Yuguchi, K., Kiso, Y., Morikawa, Y., Nambo, Y., Oikawa, M. and Sasaki, F. (1997), 'A case of a pony with Cushing's Disease', *Journal of Veterinary Medical Science* **59**(8), 707–710.

Schott, H. C. (2002), 'Pituitary pars intermedia dysfunction', *Veterinary Clinics of North America: Equine Practice* **18**(2), 237–270.
URL: *http://www.sciencedirect.com/science/article/pii/S0749073902000184*