Estimation of farm specific parameters in a longitudinal model for litter size with variance components and random dropout

Nils Toft*  Erik Jørgensen†  Claus Dethlefsen‡
nt@dina.kvl.dk  Erik.Jorgensen@agrisci.dk  dethlef@math.auc.dk

Abstract

We seek to characterize the distribution of parameters in a litter size model among a population of herds, using a non-linear mean curve and variance components. Data consists of registrations from 26 herds, and is subject to biased selection due to culling. By assuming random dropout i.e., selection based on observed values only, the mean and covariance can be estimated independently of the dropout structure using maximum likelihood estimation. Preliminary results suggest that the farms should be partitioned into groups reflecting the differences in their parameter structure.

1 Introduction

Animal replacement problems is an area which has been developed through the last 30-40 years, see e.g. Kristensen [6] for a survey. However, most attention has been given to formulation of the problem in terms of structural considerations and methods for solving large scale problems. An example is the sow replacement model of Huirne [4], where the model is well founded in terms of structure in the state space, but use "guessed" parameters obtained from literature to simulate a typical Dutch farm. As demonstrated in Jørgensen [5] the variation among farms is too large to ignore. Using parameters common to all herds will result in erroneous results for some farms. Hence, it is necessary to use farm specific data to estimate the relevant parameters in a decision support system. Furthermore the traditional approach is to model the system as a Markov decision process (MDP), for an introduction to MDP see e.g. [10]. Within this framework the state space must be completely known to the decision maker, i.e., no latent or hidden variables can be used to combine traits. Recent developments within multistage decision networks gives hope for relaxing this assumption. Within the framework of, e.g. Limited Memory Influence Diagrams [7], the structure of the underlying system can be assessed independently of the decision problem regarding the observed traits.

In this paper we describe a model and method for estimation of the mean and covariance structure of the litter size at different parities for individual sows. It is important to realize that a joint distribution is essential, since it is the correlation structure between

*Dept. of Animal Science & Animal Health, The Royal Veterinary- & Agricultural University, Denmark
†Biometry Research Unit, Danish Institute of Agricultural Sciences, Research Center Foulum, Denmark
‡Department of Mathematical Sciences, Aalborg University, Denmark
litters that gives rise to the idea of culling based on litter size. If subsequent litters were independent, culling could be based solely on age. Simply because only the shape of the mean curve (common to all sows) would have impact on the expected litter size at a given parity. Our aim of this study is to gain initial insight in how the parameters distribute among the farms, and identify elements for future studies.

2 Data

Data for this study has been extracted from the data base of the Danish Applied Pig Research Scheme (DAPR) by the National Committee of Pig Breeding, Health and Production [8]. Herds are numbered to ensure anonymity of the participating pig producers. Data are registrations of different sow events and include traits like: day of farrowing, sow number, parity number, number of piglets born alive, herd number etc. The farms in DAPR usually participate for a couple of years, hence the data consists of many, short time series. High culling rates also reduce the length of the time series. Two herds have extremely few observations per sow (1 and 1.03). It seems reasonable to exclude these herds from the remainder of this study, since very little can be deduced about the covariance structure of series of length 1. In Table 1 some traits in the registrations are given. Further descriptions have been dropped for abbreviation. We will ignore the difference in time between successive litters and treat parities as equidistant. In general it is assumed that the possible variation in time between farrowings does not contribute much to variation in litter size.

Missing values can occur for at least two reasons: censoring and dropout. Censoring occurs as either left or right censoring when sows where present prior to the beginning of the study, or the study terminates before the sow is culled. In fact most observations are assumed to have either or both kinds of censoring. Dropout occur because the farmer is inclined to replace sows that perform at an unsatisfactory level. Hence the sows in a herd must be subject to a biased selection. The sample mean curve must be higher than the expected mean if no sows were culled. The mean curve in absence of culling may be regarded as the true mean. One objective of this study is to describe the evolution of the true mean as the parity increases.

3 The model

To model the litter size of sows we need three elements: a part describing the common mean of a herd, a part describing the common covariance structure and a part describing the dropout process (culling) in a herd.

Assume the herd has $m$ sows, with up to $n$ litters per sow. In order to model the dropout process we introduce $y_{i,j}^*$ as the hypothetical litter size for sow $i$ at litter $j$, which would have been observed in case of no missing values. Then $Y_i^*$ is an $n$-dimensional
vector of measurements for the $ith$ sow and we assume that
\[ Y_i^* \sim N_n(\mu, \Sigma), \quad (1) \]
i.e., a multivariate Normal distribution. Let $Y_i$ define the corresponding actual observations with missing values because of culling (drop out) encoded as 0. Let $d_i$ be the index of the first missing value due to culling in $Y_i$. This implies that $2 \leq d_i \leq n$ indicates the dropout time, whereas $d_i = n + 1$ identifies no dropout. If we ignore censoring, the relationship between $Y_{ij}$ and $Y_{ij}^*$ can be expressed as $Y_{ij} = Y_{ij}^*$ for $j < d_i$ and $Y_{ij} = 0$ elsewhere.

3.1 The mean curve

The curve describing the mean should be non-linear in order to allow a structure where the expected litter sizes increase for the first 5-6 litters and then start to decrease. An obvious approach could be to fit a polynomial, however, we are especially interested in the prediction of litter sizes outside the region of observations. Thus we adopt the curve suggested originally by Jörgensen [5] as a combination of a straight line and a Gaussian curve i.e.,
\[ \mu_j = -\theta_1 \exp(-((j^2 - 1)\theta_2) + \theta_3 - \theta_1 \times j \quad (2) \]
For certain values of $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$ the straight line dominates for large parities, while the Gaussian term causes the curve to bend downward for smaller parities. It is possible to reparametrize the model in terms of more easily interpretable parameters as suggested in Jörgensen [5]. The mean curve is a reasonable trade-off between complexity and available data, i.e., a complex model is used in the first parities, whereas a simpler linear model is applied in the section where observations are sparse.

3.2 The covariance structure

An unstructured multivariate approach to repeated measures leaves the variance matrix $\Sigma$ entirely unspecified, i.e., a total of $n(n + 1)/2$ parameters are required to specify the variance matrix ($n$ is the max number of litters of a sow, hence $\Sigma$ is an $n \times n$-dimensional matrix). It is a more or less straight forward approach to estimate these parameters as the sample covariance given we have enough data.

Here we adopt a different approach in order to obtain a parsimonious parametrization of $\Sigma$. In Diggle [2] it is suggested that this can be achieved by incorporating time series structure into the variance matrix of the multivariate Normal distribution of (1).

Let $y_{ij}$ be the number of piglets of the $ith$ sow’s $j$th litter. Assume that $y_{ij}$ can be decomposed as $y_{ij} = \mu_j + \epsilon_{ij}$, where $\mu_j$ is the deterministic mean described in (2) and $\epsilon_{ij}$ is a random, zero-mean component. Adopting the notation in Diggle [2] $\epsilon_{ij}$ can be decomposed into three components,
\[ \epsilon_{ij} = A_i + Z_{ij} + M_i(j). \quad (3) \]
In (3) the $A_i$ are mutually independent $N(0, \sigma^2)$ random variables, which represent the variation between sows. The $Z_{ij}$ are mutually independent $N(0, \tau^2)$ random variables representing measurement error, or short term random influences. Finally the $\{M_i(j)\}$ are independent stationary random processes with common autocovariance function $\gamma(u) = \sigma^2 \rho(u)$, where the correlation function $\rho(u)$ is such that $\rho(0) = 1$ and $\rho(u) \to 0$ as $u \to \infty$. Here we assume an exponential correlation function,
\[ \rho(u) = \exp(-\alpha u), \quad (4) \]
however we might elaborate on this structure later. The autocovariance function of \( \epsilon_{ij} \) can be derived as,

\[
\gamma(u) = \begin{cases} 
\nu^2 + \tau^2 + \sigma^2 & , u = 0 \\
\nu^2 + \sigma^2 \exp(-\alpha u) & , u > 0.
\end{cases}
\] (5)

Hence, we have reduced the required set of parameters from \( n(n + 1)/2 \) to just 4. Define \( \phi \) as the parameter vector describing the variance, hence \( \phi = (\nu^2, \tau^2, \sigma^2, \alpha) \).

### 3.3 The dropout structure

Let \( H_k = (y_{i1}, \ldots, y_{ik}) \) denote an observed sequence of litter sizes up to parity \( k - 1 \), and define \( p_d(\cdot) \) as the probability of dropping out between parity \( d-1 \) and \( d \). Diggle and Kenward [3] distinguish between three different types of dropout, i.e., completely random dropout where \( p_d(\cdot) \) depends neither on \( H_d \) nor \( Y^*_d \), random dropout where \( p_d(\cdot) \) depends on \( H_d \) but not on \( Y^*_d \) and informative dropout where \( p_d(\cdot) \) depends on \( Y^*_d \) (and possibly \( H_d \)). Informative dropout might be present in e.g. medical studies where the patient drops out of the study because she is too ill to attend the meeting (or dies). However, our dropout can be classified as random dropout, i.e., we expect culling to be based on previously observed litter sizes. This implies that the probability of dropout between parity \( d - 1 \) and \( d \) can be modeled as:

\[
P(d_i = d | \text{history}) = p_d(H_d, \beta) \quad (6)
\]

where \( \beta \) is a vector of parameters. The modeling of dropout must admit the distinction between involuntary culling and culling based on production parameters. In our case we could consider any culling not related to litter size as involuntary culling, i.e., reproductive failure etc.

### 3.4 The joint distribution

We can now describe the joint distribution for \( Y_i \) using the sequence of conditional distributions of \( Y_i \). We can ignore the subscript \( i \) since all sows are assumed to share the same distribution. Let \( f_j^*(y_j) \) denote the univariate Gaussian density of \( Y^*_j \) and \( f^*_j(y^*_k) \) the multivariate Gaussian density of the \( k \) to \( j \) elements of \( Y^* \). Define \( f^*_j(y_j|H_j) \) as the conditional univariate Gaussian distribution of \( y_j \) given \( H_j \). Similarly define \( f_j(y_j), f_j(y^*_k), f_j(y_j|H_j) \) as the distributions for the observed data.

We see that the relationship between \( Y^* \) and \( Y \) defined earlier and the dropout process (6) implies that

\[
P(y_j = 0 | H_j, y_{j-1} = 0) = 1 \quad (7)
\]

\[
P(y_j = 0 \mid H_j, y_{j-1} \neq 0) = p_j(H_j, \beta) \quad (8)
\]

\[
f_j(y_j|H_j) = \{1 - p_j(H_j, \beta)\} f^*_j(y_j|H_j, \beta, \phi), \text{for } y_j \neq 0 \quad (9)
\]

Using equations (8)-(9) we can determine the joint distribution of \( Y \). Suppressing dependence on the parameters \( \theta, \phi \) and \( \beta \) and assuming no data missing because of censoring the PDF for a complete sequence is

\[
f_n(y_n^*) = f_1^*(y_1) \prod_{k=2}^n f_k(y_k|H_k) = f^*_n(y_n^*) \prod_{k=2}^n \{1 - p_k(H_k)\} \quad (10)
\]
whereas for an incomplete sequence with dropout at time \(d\), i.e., \(Y = (Y_1, \ldots, Y_{d-1}, 0, \ldots, 0)\), the PDF for the joint distribution is

\[
f_n(y^n_1) = f_1(y_1) \left\{ \prod_{k=2}^{d-1} f_k(y_k | H_k) \right\} p_d(H_d)
\]

\[
= f_{d-1}(y_1^{d-1}) \left\{ \prod_{k=2}^{d-1} \{1 - p_k(H_k)\} \right\} p_d(H_d) \quad (11)
\]

3.4.1 Censoring

Because of censoring the data seldomly consist of complete series, hence the expressions in (10) and (11) should be modified to allow for right or left censoring. This is simply done by only considering the part of the series which contain data, i.e., start at time \(k\) instead of the first litter and likewise end at time \(s\) rather than the \(n\)th litter.

4 Inference

Inference from the model can be discussed at two levels: estimation of the parameters in \(\theta\), \(\phi\), and \(\beta\) and inference from a model with specified parameters about the expected future production from a sow having an observed history of litters. While the former is the subject of this paper the latter may be regarded as the true purpose of this exercise. In Figure 1 the two models are presented. Figure 1(a) describes the problem of learning the hyperparameters, i.e., learning about the parameters describing the distribution of the parameters. The model can be interpreted as follows: the model within the frame can be considered unique to each sow, i.e., for each sow a value for variance components and mean parameters are given, according to an underlying distribution shared by all parameters of this type. Our objective is to determine these common parameters. The decision network given in Figure 1(b) is given here for reference and will be discussed later. The model for learning about the hyperparameters can in theory be analyzed using several different techniques, however, the large proportion of missing data as well as the non-linear mean curve cause problems for most tools. A Markov Chain Monte Carlo approach could relax the assumption of the underlying posterior distributions, which would be an excellent reason for adopting this technique. Unfortunately, an implementation in WinBUGS [11] failed to achieve convergence, due to high correlation between parameters and within the variance structure. Fortuitously, it is a simple task to derive the analytical expressions for the log-likelihood function and obtain traditional ML-estimates.

4.1 Likelihood function

The joint distribution derived in Section 3.4 for the different scenarios (Equations (10)-(11)) and the similar derivations in the case of censoring all share the same property. The joint distribution can be written as the product of terms involving \(\phi\) and \(\theta\) describing the evolution of the true mean and \(\beta\) modeling the dropout structure. The consequence is that the log-likelihood for \(\theta\) and \(\phi\), and \(\beta\) can be written as the sum of independent terms:

\[
L(\theta, \phi, \beta) = L_1(\theta, \phi) + L_2(\beta) \quad (12)
\]

This is a special case of the general expression in Diggle and Kenward [3] and implies that we can derive ML estimates for the mean curve independently of the dropout structure since \(L_1()\) and \(L_2()\) can be optimized independently. For \(m\) sows, let \((y_{ij})_{k_i}^{d_i} = \{y_{ij} : j =
\( k_i, \ldots, d_i \) denote the observed litters of the \( i \)th sow. Recall that \( f^{*}_{d}(y^{d}_{k}) \) denotes the joint PDF of observed litters for a given sow. Let \( (f_{i})^{*}_{d}(y^{d}_{k}) \) denote the joint PDF for the \( i \)th sow, then

\[
\log(f_{i})^{*}_{d}(y^{d}_{k}) = -\{(d_{i} - k_{i} - 1)/2\} \log(2\pi) - \frac{1}{2} \log |V^{d}_{k}(\phi)| - \frac{1}{2} \left[(y_{i})^{d}_{k} - \mu^{(i)}\right]^T \left[V^{d}_{k}(\phi)\right]^{-1} \left[(y_{i})^{d}_{k} - \mu^{(i)}\right]
\]

where \( \mu^{(i)} = (\mu_{k_{i}}(\theta_{1}), \ldots, \mu_{d_{i}}(\theta)) \) and \( V^{d}_{k}(\phi) \) is the relevant part of the covariance matrix. \( L_{1}(\cdot) \) is just the sum of (13) for each sow, i.e.,

\[
L_{1}(\theta, \phi) = \sum_{i=1}^{m} \log(f_{i})^{*}_{d}(y^{d}_{k})
\]

This expression can now be optimized using a non-linear optimization algorithm, like the Powell algorithm from Press et al. [9].

5 Results

In Figure 2 the distribution of the standardized ML estimates for 23 herds are shown to give an impression of the underlying common distribution. From the initial 26 herds, 2 has been eliminated because only 1 observation per sow was available and 1 herd have been eliminated because the data from this herd caused the optimization algorithm to crash. Other than that, no selection or weeding has been performed, hence some spurious results may be contributed to few extreme values. The mean vector and covariance matrix for \( \theta = (\theta_{1}, \theta_{2}, \theta_{3}, \theta_{4}) \) and \( \phi = (\nu^{2}, \tau^{2}, \sigma^{2}, \alpha) \) has been calculated from the ML estimates.
Figure 2: The distribution of the parameters on a standardized scale with added max and min for each parameter. Individual estimates are marked by points.

for the individual herds under the assumption that $\theta$ and $\phi$ are independent as:

$$
\mu_\theta = \begin{pmatrix} 2.93 \\ 0.20 \\ 13.30 \\ 0.29 \end{pmatrix} \quad \Sigma_\theta = \begin{pmatrix} 0.945 & -0.073 & 1.072 & 0.144 \\ -0.579 & 0.017 & -0.086 & -0.013 \\ 0.959 & -0.572 & 1.322 & 0.167 \\ 0.931 & -0.615 & 0.912 & 0.025 \end{pmatrix}
$$

$$
\mu_\phi = \begin{pmatrix} 0.34 \\ 5.25 \\ 2.34 \\ 13.80 \end{pmatrix} \quad \Sigma_\phi = \begin{pmatrix} 0.465 & -0.876 & 0.393 & 20.557 \\ -0.716 & 3.225 & -1.971 & -0.823 \\ 0.457 & -0.870 & 1.590 & -14.973 \\ 0.610 & -0.009 & -0.240 & 2446.610 \end{pmatrix}
$$

where correlations are given below the diagonal. It appears that the covariance structure is a mixture of different distributions, see the distribution plots in Figure 2. Thence an example of a population litter size structure seems uninformative, but below we give an example the structure of the mean vector and covariance matrix for herd 1 and herd 24 for the first 5 parities.

$$
\mu_1 = \begin{pmatrix} 9.49 \\ 10.61 \\ 11.28 \\ 11.34 \\ 11.12 \end{pmatrix} \quad \Sigma_1 = \begin{pmatrix} 7.58 & 1.54 & 1.38 & 1.23 & 1.10 \\ 0.203 & 7.58 & 1.54 & 1.38 & 1.23 \\ 0.182 & 0.203 & 7.58 & 1.54 & 1.38 \\ 0.163 & 0.182 & 0.203 & 7.58 & 1.54 \\ 0.146 & 0.163 & 0.182 & 0.203 & 7.58 \end{pmatrix}
$$

$$
\mu_{24} = \begin{pmatrix} 10.31 \\ 11.48 \\ 11.64 \\ 11.64 \end{pmatrix} \quad \Sigma_{24} = \begin{pmatrix} 7.25 & 1.62 & 1.42 & 1.41 & 1.41 \\ 0.223 & 7.25 & 1.62 & 1.42 & 1.41 \\ 0.195 & 0.223 & 7.25 & 1.62 & 1.42 \\ 0.194 & 0.195 & 0.223 & 7.25 & 1.62 \end{pmatrix}
$$

where correlations again are given below the diagonal.

5.1 Variance components

To illustrate how well the theoretical autocovariance function with estimated parameters corresponds to the sample autocovariance calculated as the mean of covariances from
the sample covariance matrix we have plotted the estimates and the theoretical function with estimated parameters for herd 1 and 24 in Figure 3. Incidentally these herds also represent the two classes of herds apparently present in the data: Herds with and without an additive effect ($\nu^2$). The problem can be understood from the sample autocovariance for herd 24. The sample correlation in herd 24 increase as the lag increase. This is most likely the result of increased uncertainty due to fewer underlying observations towards the end. However, our covariance structure does not allow this behavior, we explicitly demand the correlation to decrease, thus create a problem for the ML estimation.

The population mean and covariance in (16) have been somewhat blurred by this behavior, since 5 herds actually produce such results. The distribution of points in Figure 2 for the variance components tends to suggest that the systematic effect might be a mixture of two groups, whereas the noise ($\tau^2$) seems common to all herds. The 2 outliers towards the left, is a result of problems separating noise and serial effect when the correlation between litters tends to zero everywhere in the model.

### 5.2 Mean curve

The estimated mean curve and the sample means for each parity is plotted for three herds in Figure 4. These herds have been chosen because sample means can be calculated for many parities, hence we can demonstrate how the curve lies below the samples for high parties. The presence of high parity litters in a herd, however, is the result of a moderate culling strategy, hence the effect from culling can be hard to see on the high parity litters of such herds. The population mean and covariance structure in (15) clearly show the high correlation expected by the current parametrization. Between $\theta_1, \theta_3$ and $\theta_4$ there is a correlation of more than 0.9. This clearly suggest an over-parametrization of the mean. Another interesting result is that of herd 24, where $\theta_4 = 0$. This is the result of
observations on the first 4 parties only. The first part of the mean curve is determined solely by the exponential term and the constant "intercept" term, hence no information of the slope is available. Observe how the mean structure for herd 1 in (17) lies below that of herd 24 in (18). Furthermore note that the expected increase followed by a decrease is present in herd 1, but due to insufficient data in herd 24 the decline in expected litter size is unobserved and the expected litter size remains constant at 11.64.

6 Discussion & Conclusion

Our primary reason for initiating this study was to estimate the parameters for the decision problem outlined in Figure 1(b). This model is different from previous sow replacement problems, in the sense that the model describing the evolution of litter sizes is based on unobserved variables. The decision whether to keep or cull the sow is still based on the observed history (or part of) of previous litters. Thus we have tried to model the underlying structure of the decision problem, using a model which naturally encompass the structure anticipated in data, as oppose to the necessary reduction in traditional Markov decision processes (MDP), where the transitions must be modeled in terms of observable traits. The consequence of this approach is that we need to specify the model in another framework than the MDP. A recent development within multistage decision networks is the Limited Memory Influence Diagrams (LimId) [7], which seems to adequately handle the trade-off between complexity and tractability. Considering our model in a LimId framework we have presented a scenario where culling decisions are based on the observed litter in previous parity only. Explicitly adding arrows from nodes $Y_1$ to $Keep_2$ and henceforth, will force more information to be considered when making inference about the optimal policy.

Using this approach our analysis can be conducted on two levels. We can explore the impact of the underlying model on the optimal culling policy, as well as the value of including more information in terms of previous litters.

As regards the underlying model, this study has shown that a reparametrization of the mean curve is imminent to reduce correlation between parameters. In our deployment of the model, the decrease of expected litter size for high parities ($\theta_4$) is crucial, yet this parameter is probably the one data contains the least information about. This problem is evidently bigger the less farm specific information available, i.e., a newly started farm will have to wait several years to gain enough insight of the mean curve at high parities. The high correlation between $\theta_1$, $\theta_3$ and $\theta_4$ could suggest a curve fixed in terms of "stiffness", i.e., the three parameters are reduced to two: one pertaining the summit of the curve, and one the curvature downwards. Perhaps a model like the one frequently used in studies of lactation curves would fit our needs.

The problem of newcomers and the uncertainty in variance structure for large lags between litters calls for an exploration of the similarities between herds to characterize the distributions of parameters, either common to all herds or common to families of herds distinguished by certain traits, e.g., age at weaning, housing system, dropout structure etc. Given distributions for $\theta$ and $\phi$ for certain types of farms, the natural approach for a new farm would be to adapt the parameters relevant for this type of farm as prior information and then adapt these parameters to his/her settings by learning as either batch learning or adaptive learning, see Cowell et al. [1] for an introduction to learning in graphical models. The technique used is essentially Bayes formula:

\[
\text{Posterior} \approx \text{Likelihood} \times \text{Prior},
\]
i.e., update your belief in the current state of the system in light of the newly received
information. In our case distributions like (15) and (16) would be our prior, the likelihood
is based on observed data and the posterior is farm specific estimates of $\theta$ and $\phi$.

Our conclusion is that there is a need for further studies to characterize the population
of farms with respect to the parameters required by a litter size model. The objective
should be to define relevant traits to identify groups of herds so that the prior can be as
informative and specific as possible. However, just as prominent is the need for further
analysis of the model structure in order to obtain a less correlated parametrization of the
mean and clearer distinction between variance components.

Acknowledgments

The authors wish to thank *The Federation of Danish Pig Producers and Slaughterhouses*
for providing data for this study.

References


