Random Effects Models in Fishing Gear Selectivity Analysis

René Holst
DIFRES/Uni. of Southern Denmark

August 18, 2005

Abstract

Selectivity of fishing nets are modelled using GLMM. It is demonstrated how the model simplifies to a log-linear mixed effects model for a particular choice of selection curve and population curve.

Presented at Dina research summer school on "Likelihood-based inference for hierarchical/mixed statistical models"

0.1 Introduction

An important tool in the management of fisheries is the selective properties of fishing gear. Only a proportion of the fish getting in contact with a fishing gear is caught by the gear. This proportion depends on properties of the fish (e.g. species, morphology, length) as well as properties of the the gear (e.g. mesh size).

For a given species the size of the fish (relative to the mesh size of the gear) is the most important factor determining the selectivity. Usually the size is measured by the length of the fish due to practical convenience.

Small fish entering a trawl gear will easily escape, if it's body size allow to pass through a mesh. For larger fish only a proportion will escape and the probability of being retained by the trawl will increase with length. In practice. 100% of the fish above a certain length class will be retained. In experimental use a cover can be mounted over the trawl. The cover is made of small meshes that do permit escapement. Fish escaping from the trawl will thus end up in the cover. Following the above description of the catch process it is appropriate to assume the size-selectivity described by some sigmoid shaped curve. A logistic curve is a common choice

Nets operate by a different mechanism. Small fish will penetrate the meshes and are unlikely to be caught. Larger fish can perhaps penetrate a mesh with their head, but cannot go through the meshes with their body. They will typically be caught by their gills (hence the name gill nets). Fish above a certain length class cannot even get their heads into the mesh and are consequently less likely to be caught. A size selection curve is therefor typically assumed to attain some sort of bell shaped form. This can be uni-modal or multi-modal, reflecting different ways of capture. In practice a mixture of two Gaussian shaped curves appears to be appropriate for most cases. Unlike trawl gears it is not possible to determine the number of fish that encounter a net, because escapees cannot be collected. Experiments with nets therefor consist of a range of nets of varying mesh sizes and inference is based on the relationship between numbers of fish at different length classes caught in the nets of different mesh sizes. By way of example, assuming a uni-modal selection curve, the probability for a length ℓ fish to be caught in a net of mesh size m:

$$r(\ell, m; k_1, k_2) = \exp\left(-\frac{(\ell - k_1 \cdot m)^2}{2(k_2 \cdot m)^2}\right)$$

The unknown parameters k_1 and k_2 describe the location and spread respectively.

An experiment with a particular gear typically consists of a number of sets and considerable random variation is commonly observed between replications. A proper model for an experiment must therefor accommodate both the within-set (Poisson) variation as well as the random between-set variation.

Data from experiments with fishing gear are often collected in more complicated structures. In essence they are multi-level data and often with more than the two levels described so far. Apart from the random effect attributed to individual sets, experiments sometimes consists of multiple cruises and/or make use of multiple vessels. From a modeling point of view it is natural to assign random effects to these factors too. Fryer's model of between haul variation (Fryer, 1991) accommodates only two levels and analysis of more complex structures calls for new methodological developments. The present state of the art can be considered a fortunate combination of two independently developed models which are merged into a rigorous two-step approach. The splitting of the analysis into two steps ensures that data can be virtually considered within the framework of Linear Mixed Models. On the other hand it is also clear that this two-step approach is a proxy for more cohesive models and in particular the need for modelling and analysis of more complex structures naturally points in the direction of generalized mixed effects models.

0.2 Model

The number of length ℓ fish caught in a net (of a given mesh size) is the product of the abundance of length ℓ fish at the fishing location and the selective properties of the gear. Both of these are obviously unknown. To put the model more formally we assume that the number of length ℓ fish contacting the gear is a Poisson:

$$\Lambda_\ell \tilde{\ } Po(\lambda_\ell)$$

If $r_j(\ell;\theta)$ denotes the selectivity, i.e. the probability that a length ℓ fish will be retained by the j'th mesh size, the catch can be modelled as (Feller, 1968)

$$C_{\ell,j}\tilde{}\,Po\left(\lambda_\ell \cdot r_j(\ell;\theta)\right) \quad \ell=\ell_1,...,\ell_L \text{ and } j=1,..,J$$

The objective of the analysis is to estimate θ, which characterizes the gear. The λ_ℓ paramters are generally of no interest and considered nuissance parameters. A more parsimoneous approach impose a parametric shape $g(\ell;\nu)$ on the abundance rather than considering it by levels of length classes.

$$C_{\ell,j}\tilde{}\,Po\left(g(\ell;\nu)\cdot r_j(\ell;\theta)\right)$$

The parameters are estimated by maximising the likelihood function. By conditioning on the total catch across all mesh sizes in each length class the model can be given a multinomial form. This approach has the advantage of eliminating the λ_ℓ parameters, but will not be considered here.

For a range of choice of uni modal selectivity curves the model reduces to a log-linear model. These include Gaussian, gamma, log-normal and inverse Gaussian. Assume r is as given above and that the abundance is described by a scaled gamma density function

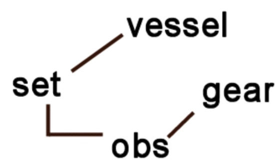
$$g(\ell;\nu)=N\cdot\left(\frac{\ell}{\tau(\alpha-1)}\right)\exp\left(\alpha-1-\frac{\ell}{\tau}\right)$$

Note that the use of Gaussian and gamma for the selectivity and abundance respectively, refers only to the functional shape of the curves. Using the log link function we get

$$\begin{aligned}\log\left(\mu_{\ell,j}\right) &= \log\left((\ell;\nu)\cdot r_j(\ell;\theta)\right)\\ &= \log(N)+(\alpha-1)\cdot[1-\log(\tau(\alpha-1))]- \\ &\quad \frac{k_1}{2k_2^2}-\frac{1}{\tau}\cdot\ell+(\alpha-1)\cdot\log(\ell)+\frac{k_1}{k_2^2}\cdot\left(\frac{\ell}{m}\right)-\frac{1}{k_2^2}\cdot\left(\frac{\ell}{m}\right)^2\\ &= \beta_0+\beta_1\cdot\ell+\beta_2\cdot\log(\ell)+\beta_3\cdot x+\beta_4\cdot x^2\end{aligned}$$

This form allows for estimation by most general statistical software packages. The main parameters of interest $k_1$and k_2 are subsequently derived from β_3 and β_4. Information on their standard errors can be assessed by use of the delta theorem (Lehmann, 1983) or by simulation, using the asymptotic bivariate normal distribution of $\left(\widehat{\beta}_3,\widehat{\beta}_4\right)$.

The above model applies to individual sets within a cluster and involves no random effects (beyound the Poisson errors). It can however easily be ammended with random effects and fits well in the framework of GLMM. A number of different directions can be motivated. There will likely be considerable variation in the local abundance of fish between different sets. Similarly there is known to be variation in selectivity between sets. A couple of these directions will be explored below.



0.3 Data

The data were collected in an experiment in the North Sea during 2001. Three vessels (similar but of slightly different sizes) operated in parallel and in close proximity during 11 days. Each vessel deployed two types of nets: gillnets and trammel-nets and each of these in five different mesh sizes. Two species, Plaice and Sole, were measured at length classes of 1 cm intervals.

length	Length of the fish fish in the record
mesh	Mesh size used for catching the fish in the record.
set	Set number. A factor describing the level-1 ”cluster” to which the record belongs to. Correspon
vessel	A factor giving the vessel-cluster for the set.
species	A factor indicating the species: Plaice or Sole.
gear	A factor indicating the net type: Gillnet or Trammel net.
freq	The number fish measured in this record.

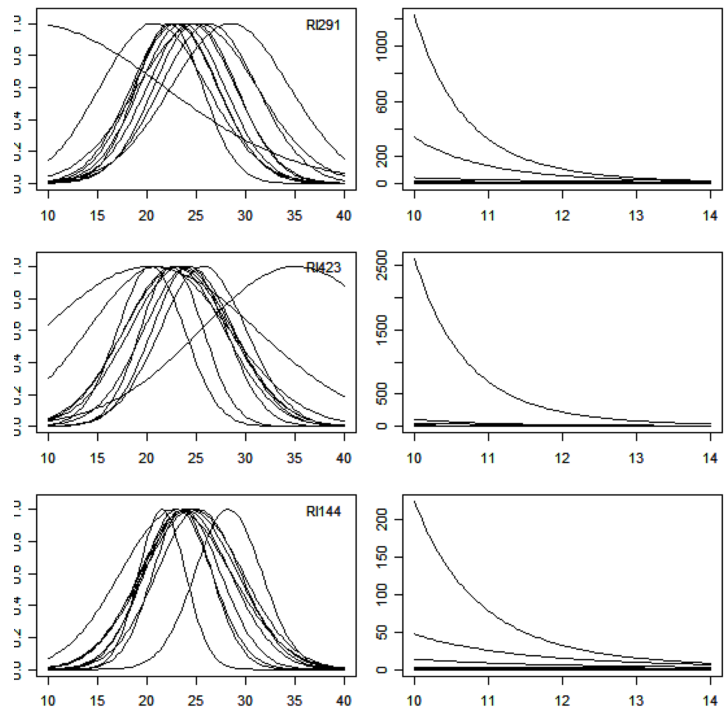
Only parts of the data will be examined here and in particular the analysis will only consider "plaice".

Estimation of selectivity aims at making inference about the mean selectivity for the fleet population. Estimates for individual vessels are therefor of less interest. Ideally the "vessel" factor should therefor enter as a top-level in the hierachical structure. With only three vessel it does however make limited sense to treat it as a random effect.

The structure considered here can therefor be depicted by the following figure:

0.4 Results

The extension to random effects models were conducted in steps from the simple basic fixed effects model. The following primarily aims at demonstrating the process of building a more comprehensive model. It is not intended to provide a full analysis of the data here.



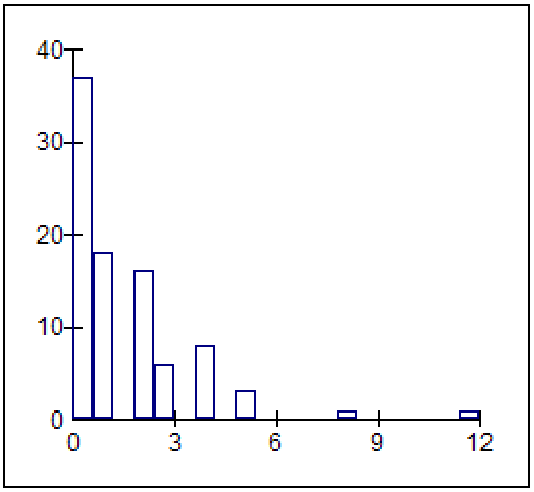
0.4.1 Exploration of data

The basic log-linear model was estimated for each set and for each of three vessels

The following figure plots the estimated selectivity and abundance curves

There is seen to be considerable variation between sets for both types of curves. Furthermore the plot indicates differences in variation for the three vessels. For the selectivity curves the random variation seems most predominant for the modes, whereas the variation in the spread of the curves appears less distinct.

The table below gives the back transformed paramter estimates from the RI291 sets.



Set	τ	α	k_1	k_2
1	4.279	6.399	2.526	0.487
2	-263.84	-0.048	2.916	0.542
3	2.251	9.827	2.717	0.498
4	-5.719	-5.472	3.16	0.666
5	-3.131	-7.48	2.615	0.478
6	-0.944	-24.053	2.467	0.376
7	17.715	3.232	2.293	0.602
8	32.063	3.445	0.902	1.501
9	-1.589	-15.624	2.786	0.67
10	1.852	15.525	2.545	0.476
11	1.598	13.969	2.762	0.466

Estimation of the dispersion parameter showed extra-Poisson variation in all sets around with estimated dispersion parameters varying between 3 and 9. Previous work has demonstrated that bi-modal selection curves provide better fits. These are however not applicable in the context of log-linear models. It is therefor unlikely that the overdispesion can be remidied. A histogram from a single set demonstrates the problem by a few extreme observations:

The following two figures list output from MLwiN for the same set fitted with and without extra-poisson variation respectively

Accomodation of over-dispersion is not unexpectedly seen to have a considerable effect on the estimated standard errors.

$$\text{freq}_i \sim \text{Poisson}(\pi_i)$$

$$\log(\pi_i) = -89.090(16.359)\text{const} + -1.311(0.281)\text{length}_i + 35.127(7.287)\text{logl}_i + \ldots$$

$$-1.496(0.324)\text{xsqr}_i$$

$$\text{var}(\text{freq}_i|\pi_i) = \pi_i$$

$$\text{freq}_i \sim \text{Poisson}(\pi_i)$$

$$\log(\pi_i) = -89.055(49.277)\text{const} + -1.310(0.846)\text{length}_i + 35.109(21.947)\text{logl}_i + \ldots$$

$$-1.497(0.975)\text{xsqr}_i$$

$$\text{var}(\text{freq}_i|\pi_i) = 9.013(1.351)\pi_i$$

$$\text{freq}_{ij} \sim \text{Poisson}(\pi_{ij})$$
$$\log(\pi_{ij}) = -22.023(10.605)\text{const} + -0.219(0.202)\text{length}_{ij} + 4.856(4.940)\text{logl}_{ij} + \beta$$
$$\beta_{3j} = 9.301(2.078) + u_{3j}$$
$$\beta_{4j} = -1.745(0.391) + u_{4j}$$

$$\begin{bmatrix} u_{3j} \\ u_{4j} \end{bmatrix} \sim \text{N}(0, \ \Omega_u) \ : \ \Omega_u = \begin{bmatrix} 0.000(0.000) & \\ 0.000(0.000) & 0.000(0.000) \end{bmatrix}$$

$$\text{var}(\text{freq}_{ij}|\pi_{ij}) = 8.264(0.397)\pi_{ij}$$

0.5 RANDOM EFFECTS MODEL

The sets taken by a single vessel, "RI291" formed the cluster for the first extension of the model

Random selectivity Model - Fixed Pop - Gamma This model showed

no random effects on the selectivity parameters. Subsequently a model with only random effect on the selectivity location parameter β_3 was estimated.

Removal of the random effect resulted in a considerable increase of the dispersion parameter.

Even though most parameters appear insignificant the model cannot be reduced due to the structural roles they play in the model.

Fixed selectivity Model - Random Pop - Gamma

$$\text{freq}_{ij} \sim \text{Poisson}(\pi_{ij})$$

$$\log(\pi_{ij}) = -21.965(15.090)\text{const} + -0.218(0.288)\text{length}_{ij} + 4.829(7.035)\text{logl}_{ij} + \beta$$

$$-1.759(0.557)\text{xsqr}_{ij}$$

$$\beta_{3j} = 9.283(2.954) + u_{3j}$$

$$\begin{bmatrix} u_{3j} \end{bmatrix} \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 0.030(0.031) \end{bmatrix}$$

$$\text{var}(\text{freq}_{ij}|\pi_{ij}) = 15.651(0.758)\pi_{ij}$$

$$\text{freq}_{ij} \sim \text{Poisson}(\pi_{ij})$$

$$\log(\pi_{ij}) = -22.034(10.611)\text{const} + \beta_{1j}\text{length}_{ij} + \beta_{2j}\text{logl}_{ij} + 9.307(2.080)\text{x}_{ij} + -1.7$$

$$\beta_{1j} = -0.219(0.202) + u_{1j}$$

$$\beta_{2j} = 4.857(4.942) + u_{2j}$$

$$\begin{bmatrix} u_{1j} \\ u_{2j} \end{bmatrix} \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 0.000(0.000) & \\ 0.000(0.000) & 0.000(0.000) \end{bmatrix}$$

$$\text{var}(\text{freq}_{ij}|\pi_{ij}) = 8.278(0.398)\pi_{ij}$$

$$
\begin{aligned}
&\text{freq}_{ij} \sim \text{Poisson}(\pi_{ij}) \\
&\log(\pi_{ij}) = -27.747(3.954)\text{const} + -0.368(0.075)\text{length}_{ij} + 8.242(1.853)\text{logl}_{ij} + \beta \\
&\qquad -0.043(0.339)\text{tram.x}_{ij} + 0.050(0.093)\text{tram.xsqr}_{ij} \\
&\beta_{3j} = 7.883(0.742) + u_{3j} \\
&\beta_{4j} = -1.460(0.146) + u_{4j}
\end{aligned}
$$

$$
\begin{bmatrix} u_{3j} \\ u_{4j} \end{bmatrix} \sim \text{N}(0,\ \Omega_u) \ : \ \Omega_u = \begin{bmatrix} 0.417(0.183) & \\ -0.095(0.048) & 0.022(0.013) \end{bmatrix}
$$

$$
\text{var}(\text{freq}_{ij}|\pi_{ij}) = 2.472(0.082)\pi_{ij}
$$

Inclusion of a fixed effect: There appears to be no significant difference in

selectivity parameters between the two gear variants. A new model was fitted in which

0.6 Conclusion:

The data analysed in this document have sampled in a hierarchical structure suitable for a mixed effects model. The top level consisted however only of three units. Consequently it was treated as a fixed effect. By imposing certain assumptions on the underlying population of fish and the selectivity it was demonstrated how the model simplified to a log-linear model, which could be ammended with random effects. Random effects were introduced in different ways, primarily referring to either the selectivity or to the population. Considerable over dispersion was observed in each model. It was not possible to remedy

$$\text{freq}_{ij} \sim \text{Poisson}(\pi_{ij})$$

$$\log(\pi_{ij}) = -27.824(3.836)\text{const} + -0.370(0.073)\text{length}_{ij} + 8.288(1.798)\text{logl}_{ij} + \beta$$
$$0.040(0.021)\text{tram.xsqr}_{ij}$$

$$\beta_{3j} = 7.884(0.690) + u_{3j}$$

$$\beta_{4j} = -1.461(0.130) + u_{4j}$$

$$\begin{bmatrix} u_{3j} \\ u_{4j} \end{bmatrix} \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 0.367(0.161) & \\ -0.079(0.041) & 0.017(0.011) \end{bmatrix}$$

$$\text{var}(\text{freq}_{ij}|\pi_{ij}) = 2.478(0.082)\pi_{ij}$$

$$\text{freq}_{ij} \sim \text{Poisson}(\pi_{ij})$$

$$\log(\pi_{ij}) = -27.696(3.782)\text{const} + -0.367(0.072)\text{length}_{ij} + 8.220(1.771)\text{logl}_{ij} + \beta$$

$$\beta_{3j} = 7.874(0.679) + u_{3j}$$

$$\beta_{4j} = -1.441(0.127) + u_{4j}$$

$$\begin{bmatrix} u_{3j} \\ u_{4j} \end{bmatrix} \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 0.377(0.168) & \\ -0.085(0.045) & 0.020(0.012) \end{bmatrix}$$

$$\text{var}(\text{freq}_{ij}|\pi_{ij}) = 2.283(0.076)\pi_{ij}$$

this by the introduction of random effects. On the contrary, the random effects appeared to be insignificant. The over-dispersion was consequently interpretated as a consequence of an inappropriate choice of model. In particular it is suspected that the choice of a uni-modal seection curve is in-adequate