Random Effects Models

This is a very interesting class of models but even a partial understanding is fairly difficult to achieve. These notes will try to convey the general notion of random effects models and the idea of "variance components." We begin by considering a population where the conditional survival parameters are *known* over a 25 year period (i.e., assume a species, where the annual survival probabilities of, say, adult males, are known exactly). These population parameters are denoted as S_1, S_2, S_3, ..., S_{25}. These parameters have some distribution (log-normal?, perhaps reasonably normal?). Clearly the variation in the survival parameters across years is measured by the usual population variance,

$$\sigma^2 = \frac{\sum(S_i - \mu_s)^2}{24},$$

where μ_s is the mean of the 25 survival parameters. Surprisingly, we can estimate σ^2 from the MLEs and the sampling variance-covariance matrix from an appropriate model! This seems counter-intuitive as one might first think that the parameters S_1, S_2, S_3, ..., S_{25} would be needed to estimate σ^2.

The name "random effects" comes from the notion that the model is based on

$$S_i = \mu_s + \epsilon_i,$$

where the ϵ_i is an independent random variable with mean 0 and variance σ^2. The population parameters fluctuate "randomly" around their mean μ_s. Deviations from this mean are "random." Other names found in the general literature include variance components, errors-in-variables regression, empirical Bayes, and "shrinkage" methods. The notes provided here are taken from an original ms. by Kenneth Burnham who has developed the theory for the methods implemented in program *MARK*.

To understand the concept here, we consider a set of band recovery data, or open population capture-recapture data, or known fate data, or other similar data set. We assume the data set represents a long time frame (e.g., 25 occasions) and the data are reasonably "good." We also assume (for the tutorial here, only) that a suitable structure of the sampling probabilities (the r_j, p_j *or* f_j) is known or given and, therefore, not address this issue directly.

Consider model $S(t)$ with the 25 year-survival parameters S_1, S_2, S_3, ..., S_{25}. We can get MLEs for each of these 25 parameters and their estimated conditional sampling variance-covariance matrix. From these estimated values, we can estimate both μ_s and σ^2 under the so-called *random effects model*.

There are several reasons for wanting valid estimates of σ^2, including the need for this quantity in population viability models. In the past, people have often mistakenly used the estimated sampling variance in these models. This is in error. Note, the sampling variance can be driven toward 0 as sample size increases, while the variation in the parameter values of

a population is hardly a function of sample size. σ^2 is a characteristic of the population. We will consider three primary issues in these notes.

1. Estimation of the Population Variance

We now consider the MLEs $\hat{S}_j$ (j = 1, 2, ..., 25) from model $S(t)$ and their conditional sampling variances, $\widehat{\text{var}}(\hat{S}_j \mid \textit{Model } t)$. It turns out that if one takes the MLEs and computes the quantity,

$$\frac{\sum(\hat{S}_i - \bar{S})^2}{24},$$

(where $\bar{S}$ is the mean of the 25 estimates), one has an estimate of the *total* variance [σ^2 + $\sum\widehat{\text{var}}(\hat{S}_j \mid \textit{Model } t)/25$]. By total variance we mean the variance among the 25 parameters (i.e., σ^2) plus the sampling variances, given the model (i.e., $\sum\widehat{\text{var}}(\hat{S}_j \mid \textit{Model } t)/25$).

The notion behind getting an estimate of σ^2 is to realize that estimates of the conditional sampling variances have been obtained through the usual likelihood methods. Thus, we proceed in the obvious way to estimate σ^2,

$$\frac{\sum(\hat{S}_i - \bar{S})^2}{24} = \sigma^2 + \widehat{\text{var}}(\hat{S}_j \mid \textit{Model } t).$$

Then,

$$\sigma^2 = \frac{\sum(\hat{S}_i - \bar{S})^2}{24} - \widehat{\text{var}}(\hat{S}_j \mid \textit{Model } t) \ .$$

Thus, if one has estimates of the total variance and the sampling variance, they can estimate σ^2 by subtraction. The subject is given in its simplest forms in Burnham et al. (1987:260-278). Of course, the method rests on the notion that one has “good" estimates of the sampling variances of the estimates of survival probability for a good model. Thus, variance inflation and other issues must be carefully considered for fear of bias in estimating σ^2.

A profile likelihood interval can be set on σ^2 as a way to assess precision. Under this approach, the interval endpoints will always lie $\geq$ 0.

2. Estimation of Functions of Survival

In ANOVA, CANOVA and regression one can “partition" variation and sums of squares into “components" (e.g., partitions by treatments, blocks, residual error, etc.). In the same way, the $\sum(\hat{S}_i - \bar{S})$ can be partitioned for group effects, time trends, group covariates, etc. Thus, one might consider a logit(S) model as a linear function of these effects or factors. Such relationships will often decrease the “error" by *explaining* some of the $\sum(\hat{S}_i - \bar{S})$.

This analysis feature will be demonstrated using program *MARK*. Here, the idea is the consider the population parameters fluctuating around means, conditional on the value of one or more covariates; i.e., $\mu_s \mid X_k$ (where X is a covariate, or matrix of covariates). Thus, just as in regression, the mean response is a function of another variable.

3. Shrinkage Estimation

This is a difficult subject and we will try only to give some insights. If one believes the random effects model is a valid approximation, based on

$$S_i = \mu_s + \epsilon_i,$$

then alternative estimates can be computed (denoted as $\tilde{S}_i$). Such estimators are termed “shrinkage" estimators. These estimates are not MLE but have some excellent theoretical properties of their own (e.g., smaller expected mean squared error). The shrinkage estimators have the property

$$|\tilde{S}_i - \hat{\mu}_s| < |\hat{S}_i - \hat{\mu}_s|, \text{ where } |x| \text{ is absolute value.}$$

The estimator $\tilde{S}$ is “shrunk" toward the estimated mean ($\hat{\mu}_s$). The degree of this shrinkage (shriveling) depends on the variance components proportion,

$$\frac{\sigma^2}{[\sigma^2 + \text{var}(\hat{S}_i | \textit{Model } t)]} .$$

An individual estimate $\tilde{S}_i$ may not happen to improve upon the corresponding MLE $\hat{S}_i$ in the sense of being nearer to the parameter S_i in a given case, but overall the shrinkage estimators are a set to be preferred as being closer to the true S_i if the random effects model apples with $\sigma^2 > 0$.

If σ^2 is a large proportion of the total variance, then the shrunk estimate will be much like the MLE. However, if σ^2 is a small proportion of the total variance, then $\tilde{S}$ will shrink much more toward μ_s. Biologists have had almost no experience with the concept of random effects models, but the potential is quite exciting (see recent PhD dissertation by Alan Franklin).

Summary

Often, with such a long (25 occasions) data set model $S(.)$ will be selected. Clearly, this is a *model* as we know that conditional survival probability cannot remain exactly the same over a quarter century. While model $S(.)$ might be “best" in the sense of a bias–variance trade-off, it might leave the investigator wondering about the variation in the parameters. Thus, the estimation of σ^2 has relevance. At the other extreme, assume that model $S(t)$ is selected; here the investigator has 25 estimates of the survival parameters, each perhaps with substantial sampling variation. This makes it difficult to see patterns (e.g., time trends or associations) or understand the variation in the parameters. Burnham's “random effects" models are interesting in such cases.