NOTE

Random effect multiplicative model

The following random effect multiplicative heteroscedastic linear model :

$$Y_{ij} = \alpha_i \mu_j + \varepsilon_{ij}\sigma_j, \quad i = 1,2,\cdots,I, j = 1,2,\cdots,J, \qquad (1)$$

I : number of wells, *J* number of sampled instants along the time range and Y_{ij} is the absorbance for well *i* at time *j*. The basic assumptions for this model are the following. The terms of the well component, α_i , account for a random fluctuation factor, with mean 1. The actual value of α_i $_i$accounts for an overfitting or an underfitting of the mean absorbance curve along time. Its variance is σ_α^2 and its distribution is assumed to be normal. The time mean effects, μ_j, are unknown values that model the overall well mean absorbance along time. The positive constants σ_j^2 are the absorbance variances for the *j*-th time instant and the errors,ε_{ij}, are standard normal random variables that account for experimental error. As a consequence, the Y_{ij} are normally distributed with mean μ_j and variance $\mu_j^2\sigma_\alpha^2 + \sigma_j^2$. Model (1) is an extension of a principal mixed effect model.

Parameter estimation

A very simple approach has been adopted for parameter estimation. The method of moments has been used to obtain estimators for the mean (μ_j) and the variance (σ_j^2) time effect, as well as for the variance of the well random effect (σ_α^2):

$$\hat{\mu}_j = \hat{Y}_j = \frac{1}{I}\sum_{i=1}^{I} Y_{ij}, for \ \ j = 1,2,\cdots,J,$$

$$\bar{\sigma}_\alpha^2 = \frac{1}{I}\sum_{i=1}^{I}(\hat{\alpha}_i - \bar{\hat{\alpha}})^2,$$

$$\bar{\sigma}_j^2 = \frac{1}{I}\sum_{i=1}^{I}(\hat{e}_{ij} - \bar{\hat{e}}_j)^2, for \ \ j = 1,2,\cdots J,$$

where • denotes average along the pertaining index,

$$\bar{\hat{\alpha}} = \frac{1}{I}\sum_{i=1}^{I}\hat{\alpha}_i, with \ \ \hat{\alpha}_i = \frac{1}{J}\sum_{j=1}^{J}\frac{Y_{ij}}{\hat{\mu}_j}, for \ \ i = 1,2,\cdots,I,$$

$$\overline{\hat{e}_{\cdot j}} = \frac{1}{I}\sum_{i=1}^{I}\hat{e}_{ij} \qquad for \ \ j = 1,2,\cdots J \quad and$$

$$\hat{e}_{ij} = Y_{ij} - \hat{\alpha}_i\hat{\mu}_j, \qquad for \ \ i = 1,2,\cdots I, j = 1,2,\cdots,J,$$

In practice, outliers may seriously affect the estimators $\bar{\sigma}_\alpha^2$ and $\bar{\sigma}_j^2$ above. For this reason, robust versions of these have been used:$\bar{\sigma}_\alpha^2 = (MAD(\hat{\alpha}_\cdot)/\Phi^{-1}(0.75))^2$ and $\bar{\sigma}_j^2 = (MAD(\hat{e}_{\cdot j})/\Phi^{-1}(0.75))^2$, where $MAD(x\bullet) = Median\ (|x\,i - Me|)$ and $Me = Median\ (x\bullet)$ and Φ is the standard normal cumulative distribution function. These

robust versions are based on the fact that, for a normal distribution with mean μ and standard deviation σ, the following relationship, between its dispersion and its *MAD*, holds: $\sigma = MAD \cdot \Phi - 1\,(0.75)$.

Boostrap resampling plan

In order to construct simultaneous prediction intervals a bootstrap resampling method has been considered to mimic the joint probability distribution of the random vector $(Yi\,1,\ Yi\,2,\ldots,\ Y\,iJ\,)$. To that aim, the following procedure has been designed:

1. Given the original absorbance sample, $Y\,ij\ (i = 1,2,\ldots,\ I,\ j = 1,2,\ldots,\ J),$ compute the estimations $\hat{\mu}_j$, $\bar{\sigma}_j^2 (j = 1,2,\ldots,\ J)$ and $\bar{\sigma}_\alpha^2$ detailed in the previous subsection.

2. Fix the number of bootstrap resamples, *B*, typically a large number (*B* = 1000 or 5000, for example).

3. For every $b = 1,2,\ldots,\ B,$ draw bootstrap random well effect replications, α_b^* , from a normal distribution with mean 1 and variance $\hat{\sigma}_\alpha^2$, and the bootstrap version of the experimental error, $\overline{\overline{\varepsilon_{bJ}}}$, (*j* = 1, 2,..., *J*) from a standard normal distribution.

4. Using the bootstrap analogue of the well effect (α_b^*), the bootstrap random errors ($\overline{\overline{\varepsilon_{bJ}}}$) and the estimators from the original sample ($\hat{\mu}_j$,$\hat{\sigma}_j^2$), the bootstrap version of the absorbance is easily defined via (1):

The sample of simulated vectors $(Y_{b1}^*, Y_{b2}^*, \cdots, Y_{bJ}^*,)$ $(b = 1,2,\ldots,\ B)$ can be used to approximate the joint distribution of the random vector $(Yi\,1,\ Yi\,2,\ldots,\ Y\,iJ\,)$, which is needed to construct the prediction band.

Bootstrap prediction band

Since the number of sampled time instants is usually moderate or high, correction for multiple prediction intervals is an important issue. Given an initial prediction level,$1 - \alpha$, for a small $\alpha\,(\alpha = 0.01\ or\ 0.05, typically)$, marginal $(1 - \alpha) - prediction$ intervals, $(\ell\, j\ ,\ u\, j\ )$, for every time instant $j = 1,2,\ldots,\ J$ can be easily constructed. Their endpoints, $\ell\, j$ and $u\, j$, are the $\left\lceil \frac{\alpha}{2} \right\rceil B$ th and $\left\lceil (1 - \frac{\alpha}{2}) \right\rceil B$ th ordered statistics of the resample $\left\{ \overline{\overline{Y_{bJ}}} \ / b = 1,2,\cdots, B \right\}$, where $\lceil x \rceil$ denotes the integer part of *x*. In other terms, $\ell\, j$ and $u\, j$, are the values that are in positions $\left\lceil \frac{\alpha}{2} \right\rceil B$ and $\left\lceil (1 - \frac{\alpha}{2}) \right\rceil B$, when sorting the bootstrap resample in an increasing order.

Individual prediction intervals have approximately the nominal coverage probability $(1 - \alpha)$ when they are considered separately (for a particular sampled instant). However, the probability that the whole growth curve is included in the band depicted by the whole set of intervals is much smaller. This is known as the multiple range testing problem or the false discovery rate in high dimensional statistical problems.

A classical way to correct for multiple testing is the popular Bonferroni approach. In a hypothesis testing context, the idea behind this approach is to consider a new significance level,$\alpha_{Bonf} = \frac{\alpha}{3}$, and compute individual tests using this new level. The resulting multiple test has a multiple level which is much closer to the desired α. However, it is well known that the Bonferroni approach is a conservative procedure. In our context, this means that the joint coverage probability of the prediction band would be larger than the desired $1 - \alpha$.

Starting from the conservative Bonferroni approach and the anticonservative individual testing approach, the following algorithm finds an approximate $(1 - \alpha) - prediction$ interval, with a given approximation error δ (typically δ is small in comparison with the nominal α, for instance $\delta = \frac{\alpha}{10}$):

1. Fix $\alpha_{low}^{(0)} = \alpha_{Bonf} = \frac{\alpha}{3}$ and $\alpha_{high}^{(0)} = \alpha$. Fix the iteration number, $k = 0$.

2. Compute $\alpha_{mean}^{(k)} = \frac{\alpha_{low}^{(k)} + \alpha_{high}^{(k)}}{2}$

3. Use the bootstrap resamples to compute individual predictions intervals with $1 - \alpha_{low}^{(k)}$, $1 - \alpha_{mean}^{(k)}$ and $1 - \alpha_{high}^{(k)}$ prediction levels.

4. Compute with the same bootstrap resamples, the proportion of simulated growth curves that are included in each of these confidence bands. These proportions satisfy $p_{low}^{(k)} \geq p_{mean}^{(k)} \geq p_{high}^{(k)}$, $p_{low}^{(k)} \geq 1 - \alpha \geq p_{high}^{(k)}$and $p_{low}^{(k)} > p_{high}^{(k)}$

5. If $p_{mean}^{(k)} \geq 1 - \alpha$, then define$\alpha_{low}^{(k+1)} = \alpha_{mean}^{(k)}$ and $\alpha_{high}^{(k+1)} = \alpha_{high}^{(k)}$. Otherwise define $\alpha_{low}^{(k+1)} = \alpha_{low}^{(k)}$ and $\alpha_{high}^{(k+1)} = \alpha_{mean}^{(k)}$.

6. Stop at step *k* if $\left| p_{mean}^{(k)} - (1 - \alpha) \right| < \delta$. Otherwise increase *k* in one unit and repeat Steps 2-5

The final approximate $(1 - \alpha)$ simultaneous prediction intervals are those obtained for level $1 - \alpha_{mean}^{(k)}$ in the last iteration.