Bayes class 9

Sean van der Merwe

# Fixed effects

Fixed effects have a closed set of category levels.

It is assumed that the entire population of relevant category levels are represented in the data.

Future data will only present as one of the studied category levels.

For example, a business might mark customers in their data base as Gender = “Male/Female/Other” knowing that a future customer will be easily placed in one of these categories, as easily as existing and past customers. They might also classify customers as Risk = “LowRisk/HighRisk” knowing that all customers (past, present, and future) can be classified into these categories deterministically if desired.

Revision: what would the design matrix look like for a regression on the interaction of these factors? Give column headings and the values for a customer that is Female and High Risk.

# Random effects

Random effects occur when categorical effects are a random sample from a population of possible effects. Random effects are incorporated into a model when future data may include categories not previously sampled.

For example, a business might manufacture their products from a standard base, say cheap wood, that they source from various suppliers. They might choose to model the suppliers as random effects if they are open to using new suppliers in the near future. Even if they don’t, the quality of the supplied wood may vary randomly from month to month without warning or signs, so the business might model “Month” as a random effect in their input model.

How would the model differ if they could measure the quality of the wood prior to accepting delivery?

# Mixed effects

Mixed effects models occur when the design matrix includes both fixed and random effects. Usually you would have a treatment variable that is a fixed effect, a subject variable that is a random effect, and confounding variables (possibly continuous) that may interact with either or both sets of effects.

For example, in medical trials the medicine/treatment is a fixed effect because the statistician is only interested in a closed set of medicines being studied and future people will also know exactly which medicine they are taking. Confounding variables such as blood pressure, gender, or BMI might be included in the model as they could interact with the medicine and change its effect. However, even after accounting for the confounding variables, each person still reacts differently to the medicine due to random aspects of their bodies that are not measured. Critically, it is the effect of the medicine on random future people (subjects) that we care about, not the set already observed.

Consider a different example: Suppose you are studying child behaviour and how it is affected by a specific set of teaching approaches. What would be possible fixed effects and random effects in this case? What would be the treatment and what would be the subject?

Fixed effects:

Random effects:

# Uncertain/Debatable effects

Sometimes an effect can be either a fixed or a random effect based on problem context.

For example, a business might initially consider the town/city of origin of a customer as a fixed effect when they operate in a closed area of the country, but then consider origin a random effect when they want to expand to new parts of the country. They could possibly revert to setting it as a fixed effect again when they are done expanding.

The key issue is that random effects can help to account for random variation in groups due to factors that we cannot measure.

# Including a nuisance effect

Consider again the examples discussed so far. In each case it might be convenient to simply ignore the random effect variables. Including ‘subject’ in the model makes it more complicated, and this can seem unjustified when we don’t actually care about any of the specific ‘subjects’.

When is it important to account for such nuisance factors in an experiment and when can we just ignore them as part of the random variation captured in the residuals?

Hint: How many measurements are being taken per subject? Is it the same for each subject?