VARIANCE COMPONENT MODELS FOR RELIABILITY AND CALIBRATION

Hal Stern
Department of Statistics
University of California, Irvine
sternh@uci.edu

BIRN AHM
San Diego, CA
October 17, 2005

Preliminaries

- Probability model (or statistical model)
  - mathematical model relating observable quantities (i.e., data) to underlying parameters
  - e.g., the traditional one way ANOVA model $y_{ij} = \mu_i + \epsilon_{ij}$ with $\epsilon_{ij} \sim N(0, \sigma^2)$
- Fixed vs random effects
  - fixed effects = factors in an experiment whose levels are set by investigator and of direct interest to the investigator
  - random effects = factors for which levels in the experiment are thought of as a random sample from an infinite population of possible levels
  - primarily a matter impacting interpretation of inferences
  - Bayesian approach to statistical inference does not make much use of this distinction

Variance components model

- Model used to decompose observed variation in a quantity into portions attributable to various factors
- Example would be to look at brain activation for a particular task and consider variation due to
  - subject
  - site
  - day
  - hemisphere
  - etc.....
- Often used in measurement context (as here)
  - to assess reliability/repeatability
  - to plan subsequent data collection

Variance components model

- Conceptual example (Snedecor & Cochran, 8th ed., Sect 13.8)
  - examine % calcium concentration in leaves
  - randomly sample 4 plants from field
  - randomly sample 3 leaves per plant
  - take two random samples of material per leave for measurement
  - y_{ijk} = calcium pct for sample k from leaf j of plant i
  - $y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijk}$ with
    $\alpha_i \sim N(0, \sigma_a^2),\ \beta_{j(i)} \sim N(0, \sigma_b^2),\ \epsilon_{ijk} \sim N(0, \sigma_e^2)$
  - variance components model with 3 sources of variation
  - once we know variance components we can decide how to best estimate in the future: more plants with few samples per plant, or fewer plants with lots of samples per plant

Variance components model

- There is not a unique definition for a variance components model
- Many different models can usually be created for a particular outcome of interest
- For example, in the above there may be other factors that could be incorporated
  - age of plant
  - soil characteristics near plant
- Alternative model might be $y_{ijk} = x_i\gamma + \alpha_i + \beta_{j(i)} + \epsilon_{ijk}$ which includes a regression term for plant factors in addition to variance components terms
- We will see this issue later in the fMRI context

Variance components model and ICCs

- Consider a simple repeatability study for fMRI
- I subjects on J days
- Let Y_{ij} = activation in ROI for subj i on day j
- Model: $Y_{ij} = \mu_i + \epsilon_{ij}$ with
  $\mu_i \sim N(\mu, \sigma^2_{indiv})$ and
  $\epsilon_{ij} \sim N(0, \sigma^2_{day-to-day})$ (a.k.a. σ^2_{error})
- Notice that Y_{ij} has mean μ and variance $\sigma^2_{indiv} + \sigma^2_{day-to-day}$
- Of course $Y_{ij'}$ (a diff't day) has the same mean and variance
- But Y_{ij} and $Y_{ij'}$ are correlated because they share a common subject effect
- $ICC = \sigma^2_{indiv}/(\sigma^2_{indiv} + \sigma^2_{day-to-day})$ measures this correlation
- This is absolute agreement ICC

Variance components model and ICCs

- Recall earlier comment about the possibilitiy of different models
- What if we expect there might be day (or judge) effects?
- Might like to look for ICC that does not require absolute agreement (only agreement in ranking)
- Can consider an alternative model:
  $Y_{ij} = \mu_i + d_j + \epsilon_{ij}$
  with d_j a fixed day effect and other terms as above
- This leads to an alternative version of ICC that accepts a day-to-day shift in measurements

Variance components model and ICCs

- Example 1: Human phantom study at a single **site**
  - have 5 subjects seen twice
  - response is mean activation (beta) in left occipital lobe to sensorimotor task (avg of four runs)
  - can fit simple variance components model and estimate ICC

Site	σ^2_{subj}	$\sigma^2_{day-to-day}$	ICC
1	.0090	.0013	.87
2	.0174	.0019	.90
3	.0227	.0124	.65
4	.0060	.0255	.19

  - caveat: very small sample (5 subjects seen twice) hence high variability

Variance components model and ICCs

- Example 2: Human phantom study for a single **subject**
  - two measurements at 10 sites
  - response is mean activation (beta) in left occipital lobe to sensorimotor task (avg of 4 runs)
  - can fit simple variance components model and estimate ICC

Subj	σ^2_{site}	$\sigma^2_{day-to-day}$	ICC
1	.0053	.0135	.28
2	.0304	.0006	.98
3	.0211	.0028	.88
4	.0071	.0035	.67
5	.0069	.0100	.41

  - note: this is a different ICC, measures day-to-day reliability across different sites for a single subject

Variance components model and ICCs

- Example 3: Human phantom study
  - 5 subjects at 10 sites
  - response is mean activation (beta) in left occipital lobe to sensorimotor task (avg of 8 runs over two days)
  - can fit simple variance components model and estimate ICC
    - variance for subjects = .0075
    - variance for sites = .0141
    - ICC = .35

Estimating variance components

- Staying with the simple (ICC) variance components model we can talk about how variance components analyses are done
- To fix ideas, consider a study at a single site (I subjects and J visits per subject)
- Different estimation strategies
  - ANOVA/Method of moments
  - Maximum likelihood (ML, REML, MIVQUE)
  - Bayesian inference

Estimating variance components - moments

- A traditional analysis of variance of the Y_{ij} (with i being subject and j being visit) is run
- $E(MSError) = \sigma^2_{day-to-day}$
- $E(MSSubj) = \sigma^2_{day-to-day} + 2\sigma^2_{subj}$
- method of moments (MOM) equates observed MS to expected MS and solves for estimates
- $\hat{\sigma}^2_{day-to-day} = MSError$ is MOM estimate of $\sigma^2_{day-to-day}$
- $\hat{\sigma}^2_{subj} = (MSSubj - MSError)/2$ is MOM estimate of σ^2_{subj}
- Comments:
  - very easy (especially for balanced data)
  - but can yield negative estimates (which we know are not right)

Estimating variance components - ML

- Not much detail today
- Maximum likelihood (ML) returns to the full normal likelihood for the Y_{ij}'s and chooses parameter estimates to maximize this likelihood (which typically involves the constraint that they be nonnegative)
- ML estimation is typically biased in this context if model includes fixed effects so people prefer REML (ML REstricted to the variance components)
- May also see a reference to MIVQUE (minimum variance quadratic unbaised estimation) which is related to ML

Estimating variance components - Bayes

- Don't actually need to talk about this today
- I have mentioned this in the past though and since this is a "teaching" session ....

The Bayesian approach to inference - variance components

- Full probability modeling
  - likelihood $p(y|\theta) = p(\text{data} \mid \text{parameters})$
  - prior distribution $p(\theta|\phi)$ (depending on other parameters)
  - hyperprior distribution $p(\phi)$
  - variance components:
    - likelihood is the normal likelihood for the $Y'_{ij}s$ (which depends on μ_i and $\sigma^2_{day-to-day}$)
    - prior distribution is the normal random effects distribution (which depends on μ and σ^2_{subj})
    - hyperprior distribution on μ and σ^2_{subj} (often noninformative)

The Bayesian approach to inference - variance components

- Posterior inference
  - Bayes' thm to derive posterior distribution

$$p(\theta, \phi|y) = \frac{p(y|\theta)p(\theta|\phi)p(\phi)}{p(y)}$$

  - probability statements about unknowns (θ, ϕ)
- Model checking/sensitivity analysis
  - does the model fit
  - are conclusions sensitive to choice of prior distn/likelihood

What to know about Bayes/traditional methods

- Common statistical appraoches are largely a collection of methods with good frequentist properties, developed over time for specific problems (e.g., t-test, REML)
- Bayesian approach can be thought of as a way of "automatically" generating statistical procedures
- But subjective Bayesian methods don't really provide for study of properties of procedures When viewed in this way, there is no
- A modern synthesis is for Bayesians to study the frequentist properties of their procedures

Variance components models – general

- So far discussion has really been a statistician's view of ICC and its relationship to variance components models
- Consider more sophisticated models
- First, a short digression – crossed and nested factors
  - Crossed factors - two factors in an experiment are crossed when each level of the first factor is seen in combination with each level of the second factor
  - Nested factors - one factor is nested within another if the levels of the nested factor don't mean the same thing within each level of the other factor

Variance components models – general

- Nested and crossed factors - examples
- Crossed: study with 3 drugs and 4 doses would have 12 combinations (and then see perhaps 5 patients at each combination)
- Nested: study with 3 schools and 4 teachers per school would consider teachers nested within schools because "teacher 1" doesn't mean the same thing at each school (more likely we have 4 randomly chosen teachers from each school)

Variance components models – general

- Consider the human phantom study in all its glory
- Let Y_{ijkl} = activation of left occipital lobe during sensorimotor task run l, visit k, site j, subject i
- A possible model

$$Y_{ijkl} = subj_i + site_j + subj.site_{ij} + visit_{k(ij)} + run_{l(ijk)}$$

  - considers 4 runs nested within visits (hence assumes no fatigue or regular run effect)
  - considers 2 visits nested within the subject/site interaction (hence assumes no consistent day 1 / day 2 pattern)
  - considers all items as random (a little odd for site)

Variance components models – general

- Example: human phantom study
  - 5 subj, 10 site, 2 visits, 4 sensorimotor runs
  - mean activation in left occipital lobe

Source	σ^2	proportion
subj	.0055	.19
site	.0087	.30
subj.site	.0055	.19
visit	.0051	.18
run	.0040	.14

Variance components models – general

- Again emphasize that alternative models are possible
- Alternative models I: include visit and run as crossed factors rather than nested factors

  $Y_{ijkl} = subj_i + site_j + visit_k + run_l \qquad + subj.site_{ij} + subj.visit_{ik} + ...$

  – results on next slide

Variance components model – fully crossed

Source	DF	Squares	Mean Square
subj	4	2.0476	0.5119
site	9	3.7343	0.4149
subj*site	36	2.4657	0.0685
visit	1	0.0309	0.0309
subj*visit	4	0.0399	0.0075
site*visit	9	0.1884	0.0209
subj*site*visit	36	0.9660	0.0268
run	3	0.0012	0.0004
subj*run	12	0.0950	0.0079
site*run	27	0.0916	0.0034
subj*site*run	108	0.3774	0.0035
visit*run	3	0.0005	0.0002
subj*visit*run	12	0.0244	0.0020
site*visit*run	27	0.1102	0.0041
subj*site*visit*run	108	0.4949	0.0046
Corrected Total	399	10.6580	

Variance components models – general

- Alternative models II
  - Run effects don't look significant
  - Can create hybrid that treats visit as crossed with other factors but leaves runs nested within visits
  - Results match those of original nested model
- Alternative models III
  - include hemisphere
  - Y_{ihjkl} is response for hemisphere h (left or right) of subject i at site j on run l of visit k
  - see UCSD variance components results

Application of variance components models

- Simple variance components model gave us ICCs
- What do we get from more complex variance component models?

  a) bored to death

  b) a major headache

  c) great flexibility

  d) all of the above

- Can ask a variety of questions such as
  - how similar would the activation be for two runs on the same subject, site, visit?
  - how similar would the activation be for two runs on same subject at the same site but on different days?

Application of variance components models

- How similar would the activation be for two runs on the same subject, site, visit?
  - correlation of two runs on the same subject, site, visit is

$$\frac{\sigma^2_{subj} + \sigma^2_{site} + \sigma^2_{subj.site} + \sigma^2_{visit}}{\sigma^2_{subj} + \sigma^2_{site} + \sigma^2_{subj.site} + \sigma^2_{visit} + \sigma^2_{run}}$$

- How similar would the activation be for two runs on same subject at the same site but on different days?
  - correlation of two runs on the same subject, same site but diff days is

$$\frac{\sigma^2_{subj} + \sigma^2_{site} + \sigma^2_{subj.site}}{\sigma^2_{subj} + \sigma^2_{site} + \sigma^2_{subj.site} + \sigma^2_{visit} + \sigma^2_{run}}$$

- Considerations like these lead us to generalizability theory (next talk)

Variance components and hierarchical models

- What do people (e.g., me) mean by hierarchical models?
  - model is specified in layers or stages
  - observable outcomes are modeled as depending on parameters, $p(y|\theta)$
  - parameters are given their own probability model involving other "hyper"-parameters, $p(\theta|\alpha)$ (can have more levels)
- Doesn't this also describe variance components models?
  - yes
  - variance components models are hierarchical models
- Lots of models, any that are specified hierarchically, can be thought of as hierarchical models

Variance components and calibration

- What role can variance components models play in calibration?
  - Good question – I'm not sure I know the answer yet
  - Need to know how consistent variance components are across different regions (both activated and inactive)
  - Could potentially estimate site effects on calibration tasks (SM, BH) in a study like the phantom study and use the estimated effects in subsequent analyses
- What role can hierarchical models play in calibration?
  - more general view of hierarchical models may allow for analysis of multisite data sets statistically (e.g., by controlling for activation on SM/BH tasks in a hierarchical model)