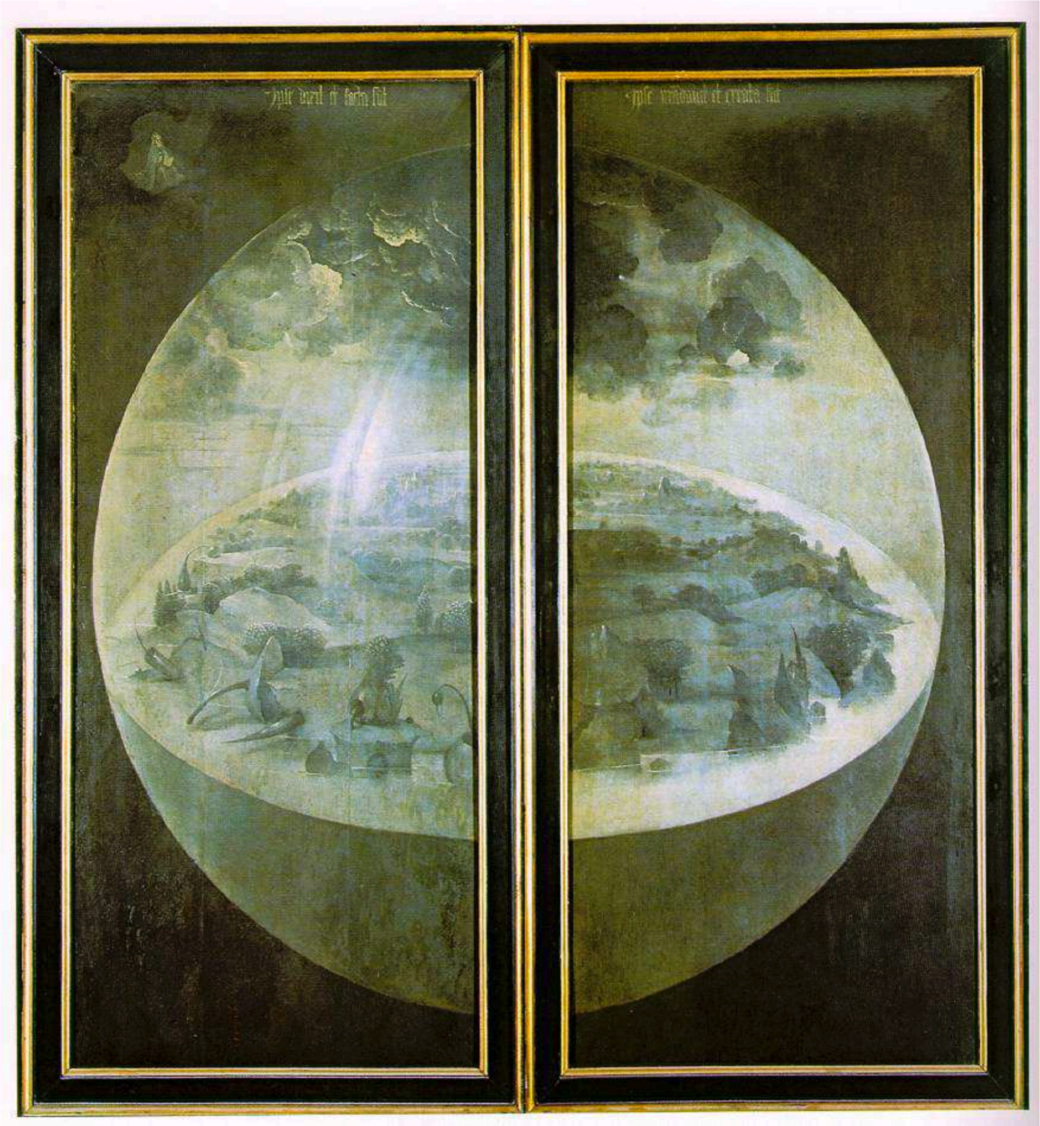
Comments on papers of Kuusela and Cranmer

Anthony Davison

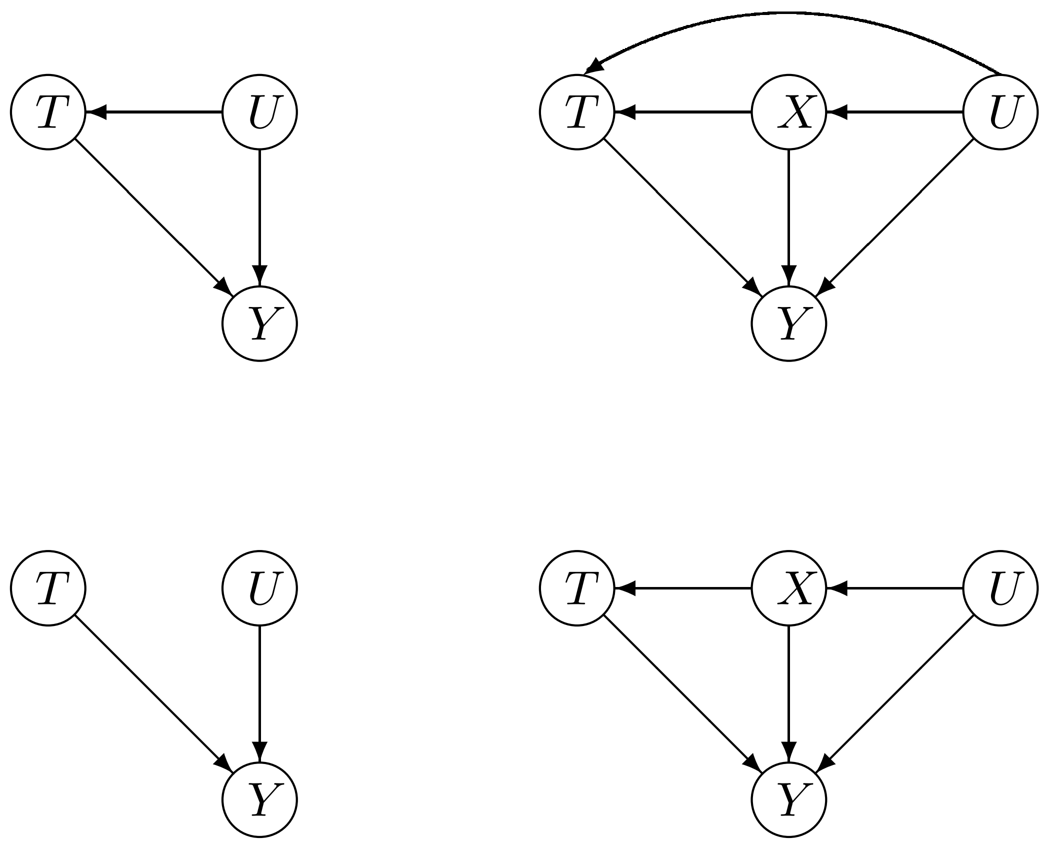
The Garden of Earthly Delights



- Bayesian approach:
  - practical (computational) and intellectual (coherence, single interpretation for all probabilities) advantages,
  - but using flat priors in high-dimensional settings can give poor inferences for interest parameters,
  - may sacrifice calibration for coherence — aided by matching priors?
- Frequentist approach:
  - calibrate confidence sets, p-values . . . relative to repeated sampling of data within a reference set of outcomes,
  - need to determine the reference set with care — does it take into account the forking path of analysis decisions taken to reach the current position in the statistical garden? What about selection effects?
  - can appear ad hoc in dealing with nuisance parameters — but r^* helps in likelihood settings!
  - Use of r^* is equivalent to certain bootstrap inferences (papers by Tom DiCiccio and Alastair Young)

Systematic errors and randomisation

- In agriculture, medicine, . . . , aim to reduce (or, ideally, eliminate) ‘biases’ due to unseen confounders U by randomized allocation of treatments T to units.
- What would the analogy in the present context be? What is a unit? What is a treatment?



Kuusela

- Four sources of systematics in unfolding:
  - regularisation bias
  - wide-bin bias
  - missing nuisance variables
  - uncertainty in response kernel
- Will mostly comment on first two. (next slide)
- Missing nuisance variables: what (if anything) can be learned from statistical design of experiments?
- Response kernel uncertainty: sounds like a ‘combination of estimators’ or ‘meta-analysis’ problem

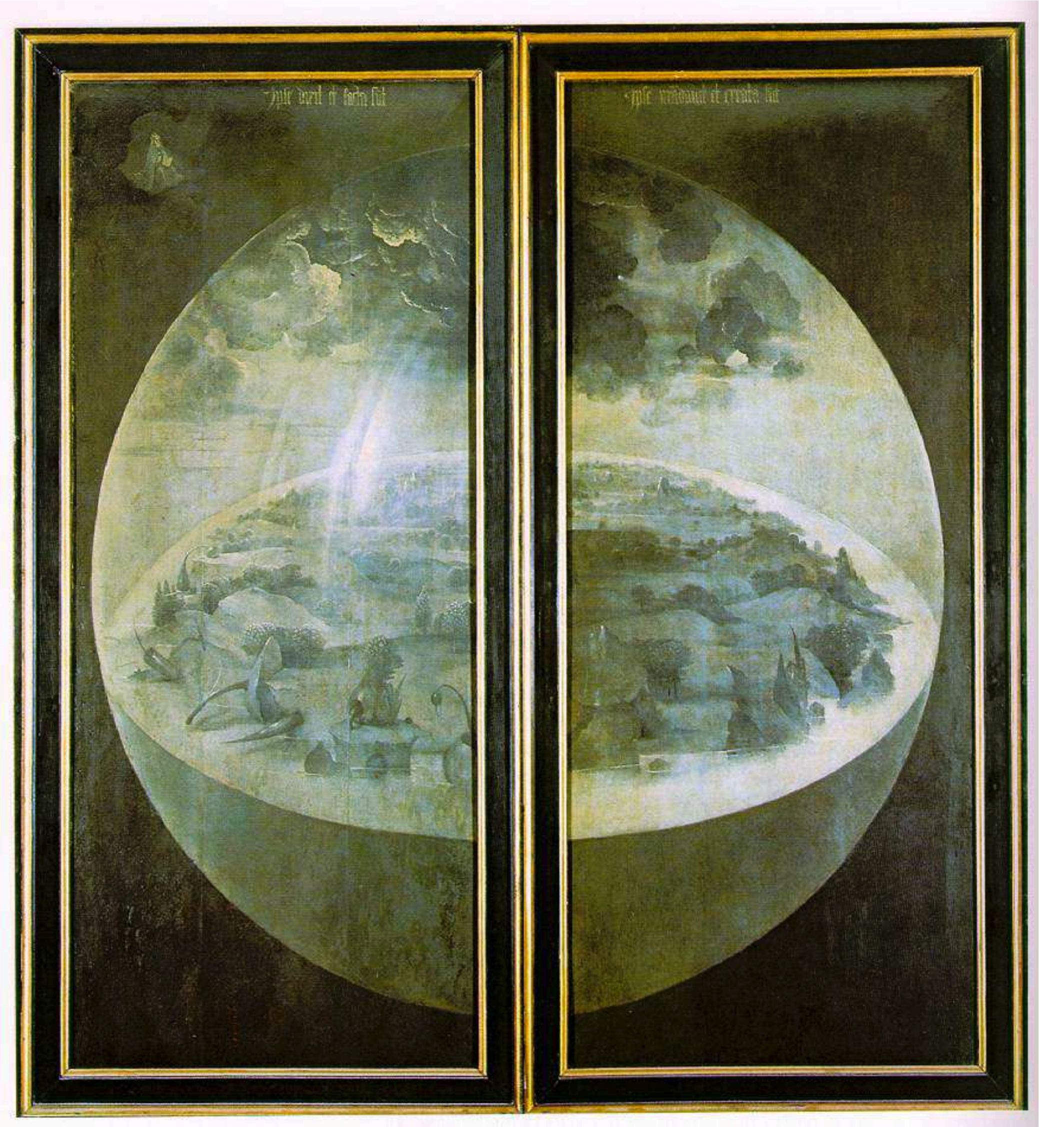
Regularisation and wide-bin biases

- '...splines are also sometimes used': relation to GAMs?
- Generalized additive models (GAMs) $\equiv$ stochastic partial differential equations fitted using INLA (form of Bayesian analysis avoiding MCMC) very widely used (Wood, 2017).

- GAMs use splines and estimate degree of smoothing automatically and efficiently using (generalized) cross-validation or marginal maximum likelihood
- It is claimed that good coverage for the underlying smooth curve ($\equiv$ spectrum) is achieved using a 'Bayesian' confidence interval approach
- Can unfolding be formulated in this way? Lot of advantages if so ...
- Is good coverage everywhere the goal, or do we care more about certain places? The numerical results in slides 9–10 look at the peak, but is this a realistic focus when looking for 'new physics'?
- Taking 'point estimate ± 2 SE' can fail badly, but point and interval estimation have different goals.
- For bootstrap confidence intervals for kernel regression estimators, theory around 1990 established that for a confidence set for a curve $\mu(x)$ estimated using bandwidth h should be based on residuals from a fit for bandwidth h' , added to an estimate with bandwith h'', where $h' \ll h \ll h''$. Is this (or can it be made) relevant?

Cranmer

- Nomenclature: ‘nuisance parameter’?
- In statistics, discrete nuisance parameter not often considered, as expect to be able to identify ‘true’ value with high probability in large samples. Sensitivity analysis in small samples?
- Baseline classifier: don’t forget (unexpected) success of naive Bayes classifier.
- Data augmentation: looks like Bayesian averaging, so surprising it’s not so good. Is it being done ‘right’?
- Adversarial training and Uncertainty-awareness: look like flip sides of a coin.
- Obvious that uncertainty awareness should be better, if information is available about the ‘nuisance parameter’ — relates to choice of ancillary statistic?



The Garden of ~~Earthly~~ Statistical Delights