Statistica Sinica Preprint No: SS-2016-0282	
Title	Discussion: Dissecting Multiple Imputation from a Multi-phase Inference Perspective: What Happens When God's Imputer's and Analyst's Models are Uncongenial?" by Xie and Men
Manuscript ID	SS-2016-0282
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0282
Complete List of Authors	Anthony Desmond
Corresponding Author	Anthony Desmond
E-mail	tdesmond@uoguelph.ca

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0282

Discussion of " Dissecting Multiple Imputation from a Multi-phase Inference Perspective: What Happens When God's, Imputer's and Analyst's Models are Uncongenial?"

Anthony F. Desmond

Department of Mathematics and Statistics, University of Guelph, Ontario, Canada

1 Introduction

When I was first invited to discuss this paper, I was stimulated by the challenge of looking at this novel area of multi-phase inference and was particularly interested in the role of estimating functions in missing data situations. At first glance, the paper looked quite congenial to me (congenial in the conventional sense of being stimulating and thought-provoking). At second reading, the brow became increasingly furrowed and I realized that this new area of multi-phase inference was going to be considerably more taxing or thorny than anticipated. I hasten to add that on subsequent readings the paper remains congenial to me, again in the conventional sense of being thought-provoking and stimulating. However, as a newcomer to this area, I needed to understand ideas relatively new to me such as uncongeniality, self-efficiency etc. Also, old ideas, such as validity, which are (arguably, see below) relatively straightforward in single-phase inference, take on a more difficult aspect in the multi-phase paradigm as discussed in Section 1.3. I found this section particularly challenging and initially at least somewhat mystifying (perhaps even mystical!); the great varieties of actors here, Gods, demi-Gods, imputers, and analysts I found rather daunting.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0282

2 Validity

Xie and Meng reference the classic paper by Neyman (1937) as justification for the notion of confidence validity, which implies conservative coverage properties for confidence intervals. This led me to look more closely at the early work of Neyman. Many texts such as Lehman's (1959) classic, and also texts such as Bickel and Doksum (1977) and Casella and Berger (2002), do indeed define confidence coefficients in conservative terms, i.e. a $1-\alpha$ confidence interval has *at least* coverage probability of $1-\alpha$. In teaching I always felt this was just a means to allow for the difficulties with discrete distributions in attaining exact coverage probabilities for a given α; Bickel and Doksum, for example, explicitly mention this difficulty. However, it is interesting to trace the evolving attitudes of Neyman through the years via the original classic paper Neyman (1934) and subsequent papers Neyman (1935), Neyman (1937), Neyman (1941), and finally Neyman (1977).

In the 1934 paper (read before the Royal Statistical Society) Neyman presents another classic, both for its pioneering contributions to survey sampling and for the first development in English of Neyman's approach to interval estimation. This appears to be the paper in which the terms confidence interval and confidence coefficient are introduced into the English language for the first time, although earlier work in Polish, Pitkowsky (1932), contains the equivalent terms, and Neyman had been using the equivalent terms in lectures in Poland for some time ; see Reid (1982). Bowley, in the discussion, sees nothing new and refers to a 'confidence trick'. However, Neyman here, p562, does indeed use the conservative definition and defines the confidence coefficent as the lower bound for the confidence coverage. He suggests that he is merely providing alternative derivations to Fisher's fiducial intervals describing "its main lines in a way somewhat different to that followed by Fisher" and says: "Thus the new solution of the problems of estimation consists mainly in a rigorous justification of what has been generally considered correct on more or less intuitive grounds."

On p 586 equation (43) he again uses the conservative condition although later:

"On the contrary, erroneous judgments of the form (43) must happen, but it is known how often they will happen in the long run: their probability is *equal to* ϵ (my italics)." This

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0282

is possibly just a *lapsus linguae*, to use a phrase that Neyman was fond of using in referring to Fisher's way of explaining the fiducial argument. The formal theory underlying the confidence intervals is delegated to Note I of appendix VI, pp 589-593. Thus, one of the most influential ideas (for good or bad!) in modern statistics is relegated to a note in an appendix to an admittedly pioneering paper! Neyman's illustrative example here is the binomial , so conservatism is inevitable. There is a quasi-Bayesian flavour as admitted by the later Neyman (1977), in that an unknown prior probability distribution $\phi(\theta)$ is assumed for the collective character (what we would now call a parameter) in the population. Fisher's response is generally positive in his comments "on those applications of inductive logic which constituted so illuminating and refreshing an aspect of the evening's paper." He compliments Neyman on his generalization of the fiducial argument for "its perfect clarity". He then takes issue with three issues : (1) lack of uniqueness, (2) the uses of inequalities for discrete distributions, and (3) the difficulties in the multi-parameter case. It is (2) that is of most interest here as it appears to be an inspiration for Neyman (1935). Of course, fiducial intervals for discrete distributions posed a major difficulty for Fisher.

In the 1935 paper, Neyman revisits the problem of confidence intervals. He begins by mentioning Fisher's criticism that his (Neyman's) extension of his (Fisher's) work concerning the fiducial argument to the case of *discontinuous* (my italics) distributions is obtained at great expense, namely the replacement of equalities by inequalities. He then shows that exact equalities are not possible in general for discontinuous distributions. In particular for the binomial he suggests that the well-known Clopper-Pearson intervals involving inequalities are best possible. This is now known to be false. Recent work by Agresti and Coull (1998) and Brown et al (2001) and references therein to earlier work, show that the 'gold-standard' exact Clopper-Pearson intervals are extremely conservative and much better alternatives are available. There is an interesting discussion in the Brown et al paper as to whether one should insist on conservatism or whether being close to the nominal level is a preferable criterion, suggesting that modern statisticians are not entirely in agreement with the textbook definition. It is interesting that Neyman in 1935 continues to invoke a prior distribution in his argument and that he appears to still regard his work as an extension of Fisher's approach to

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0282

fiducial intervals.

In his classic 1937 paper, the one to which the authors refer, Neyman gives a treatment of confidence intervals, which gives a solution to the problem of confidence intervals without recourse to any a priori distribution and answers the question posed in the last sentence of his 1935 paper at least for continuous distributions. Equalities of confidence coverage and confidence coefficient are maintained throughout and in the general treatment and the examples only continuous distributions are used. In his review of previous attempts at interval estimation the fiducial argument is studiously avoided, although a footnote indicates that this review is incomplete! In later work, his notes of 1952, for example, he maintains the equality of coverage and confidence coefficient , although the conservatism for discrete distributions is touched on briefly. The later Neyman (1977), in a delightfully contentious paper in Synthese, returns to equality of coverage probabilities and again considers only continuous distributions.

My reading of Neyman (1934) is that the conservative definition of the confidence coefficient is mainly to allow for the discrete nature of the binomial example used there. My main point is that I doubt the 1934 paper could be used as a general justification for the notion of confidence validity. That is not to say that the concept of confidence validity is not a useful one. The arguments in Rubin (1996), for example, seem sound to me.

3 The Multiphase Paradigm

As I understand it, the multi-phase inference paradigm is quite a general one and the multiple imputation special case is illustrative. As an academic statistician, I tried to think of situations in which I, or colleagues of mine, might have been part of a phase (or phases) to which the multi-phase paradigm might have (retrospectively) brought some useful perspective. It is often said that applications of statistics in science and technology are piece-meal ; but the multi-phase-paradigm may be an attractive way of adding a useful formalism to counter-act this. Like most academic statisticians, I have been involved in consulting projects with colleagues from the life sciences, engineering etc. The closest I may have been to a multi-phase

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0282

situation was a collaborative project with General Motors Canada. This did involve much data pre-processing, but Xie and Meng make me wonder whether we were as useful to the clients, in retrospect, as we might have been. We had a large amount of data on worker behaviour at a plant in Oshawa, of a very messy nature which needed to be cleaned etc. I was not personally involved in in the pre-processing phase. The data collection phase involved engineers and technicians at the plant with little statistical knowledge. The pre-processing of the data was done by colleagues in a group at the University of Guelph involving students, a colleague and research assistants. The ultimate motivation was to simulate the process and see how we might improve productivity and increase profits for the company. One of my tasks was to apply a plausible, but in some ways overly simple model, to a small part of this large data set., which resulted in Desmond and Chapman(1993). Other tasks resulted in several technical reports, which the team at GM seemed to find helpful. Although aspects of the data-preprocessing and the difficulties with the original very large data set were discussed with me , I was not involved in the earlier phase. In retrospect, I would now possibly consider some statistical learning techniques for Big Data. However, this was in the early nineties and the Big Data revolution had not yet begun! The warnings on the perils of preprocessing by Blocker and Meng (2013), cited here, is definitely on my reading list.

In the Multiple Imputation case discussed in this paper, we have large public-use data-bases with statistically sophisticated imputers relative to inexperienced users (analysts), who may indeed be non-statisticians unfamiliar with missing data procedures and likely to use off-the-shelf complete-data packages. The importance of some encompassing Multiphase paradigm makes a good deal of sense here, and the multiple imputation approach, despite its critics , seems a sensible pragmatic approach to this situation. Also, arguably, Bayesian imputation and frequentist analysis is reasonable at the present time, although with the increasing acceptance and use of Bayesian methods, and more importantly, the increasing availability of Bayesian-based software, that situation may change in the future. In the case of multi-party use of large public data bases, it is unlikely that the analyst would be as sophisticated as a Meng or a Xie say! In other applications, however, where the analyst is a scientist with considerable subject-matter knowledge is the Bayesian as Imputer versus Analyst as frequentist dichotomy really necessary? Why not consider Bayesian validity at the analysis phase along

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0282

with the effect of uncongeniality on inferences? For example in the Tu, Meng, Pagano (1993) example, cited in the paper, one could use the Bayesian approach to impute the delayed cases; but should this prevent a principled Bayesian from using a Bayesian version of the Cox or other analysis, which are available in survival analysis; one could even possibly use informative priors for relative risks, baseline hazard based on the first phase?

4 Use of Estimating Functions

The use of estimating functions in the context of multiple imputation and more generally multiphase inference is fascinating. The key decomposition result in Section 3 is very interesting. This seems to fit the multiple imputation particularly well. Extensions and applications in the more general multiphase scenario presents an area for future research. Some earlier work on the use of estimating for missing data is referenced in a recent encyclopedia article by Desmond (2016). Also, conditional expectations such as those of Xie and Meng, regarded as projections, in L_2 spaces are useful analytical tools in deriving optimality results in the search for good estimating functions. Small and McCleish (1994)is a good introduction to this approach. The EM algorithm, so ably summarized in Meng and van Dyck (1997), can be extended (generalized) to the estimating function situation. A particularly interesting example, when no likelihood is available, but second order assumptions are made, is Heyde and Morton (1996). They develop what they call the Projection-Solution approach, in which the E-step is replaced by a projection into a space of estimating functions determined by the second order assumptions; The M-step is then replaced by the solution step on solving the estimating equation obtained from the Projection stage. One wonders whether the Projection-Solution approach could be useful in multiphase inference? The projection idea is a very powerful one in statistics generally. Pythagoras for estimating functions, rather than estimators may be more fruitful?

5 Conservatism

Xie and Meng let $m = \infty$, for the number of imputations. My, admittedly limited, reading of the multiple imputation literature suggests m quite small, say 2 to 5, is adequate for validity

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0282

and this has been advocated as an advantage of the multiple imputation approach in terms of data storage e.g. Rubin (1996) and elsewhere. Under uncongeniality how does small m effect the double the variance rule? Intuitively a more conservative rule might seem appropriate? Yet the transition from (5.14) to (5.15) for the standard error rule for a scalar estimand suggests this is not the case? As Xie and Meng mention, the 'double the variance' rule is reminiscent of the discussed paper by Copas and Eguchi (2005). Some of the discussants of that paper expressed reservations about this rule. One issue is that practitioners might use such a rule automatically, and of course, there are dangers in that, which presumably will be even more challenging in the multi-phase case. Others, e.g. Little, suggested that a lower bound for uncertainty is not very useful. Copas and Eguchi are appropriately cautious (see their reply and page 484 of their paper). Those authors dealt with the twin issues of incomplete data and model misspecification but within a single-phase paradigm. It is interesting that this results in lower bounds for uncertainty, as opposed to the current paper, which gives conservative inferences. Of course, the former paper deals with model misspecification, whereas the current paper does not. In the multi-phase case, with incomplete data and uncongeniality, if we add in model misspecification, the statistician may have to throw his/her hands up and admit defeat! Sometimes no valid inference is possible, and a range of sensitivity analyses, possibly not very informative will be all that is feasible. In the simpler case, of single-phase, nonignorable missing data, such sensitivity analysis seems to be the only recourse. However, Xie and Meng have presented a substantial challenge to the statistics profession to deal with the unholy trinity of missingness, misspecification and uncongeniality in the multi-phase paradigm. They have made substantial advances in illustrating this paradigm in the multiple imputation case and presented many open problems for future research. They have given us much to think about. It has been a great pleasure to have had the opportunity to comment on this excellent paper, which I expect to re-read frequently for its thoughtful and challenging contributions to a new paradigm in statistical methodology.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0282

References

Agresti, A., and Coull, B. A. (1998). Approximate is better than 'exact' for interval estimation of binomial proportions, *The American Statistician* **52**, 119-126.

Bickel, P. and K.Doksum (1977). *Mathematical Statistics: Basic Ideas and Selected Topics.* Holden Day, San Francisco.

Blocker, A. and Meng, X.-L. (2013). The potential and perils of preprocessing: Building new foundations. *Bernoulli* **19**, 1176-1211.

Brown, L. D., Cai, T. T. and DasGupta, A. (2001). Interval estimation for a binomial proportion (with discussion). *Statist. Sci.* **16**, 101-133.

Casella, G. and Berger, R. L. (2002). *Statistical Inference.* Duxbury Press, New York, NY.

Copas, J. B. and Eguchi, S. (2005). Local model uncertainty and incomplete data bias (with discussion). *Journal of the Royal Statistical Society* B **67**, 459-513.

Desmond, A.F. (2016). *Estimating Functions.* Wiley StatsRef: Statistics Reference Online, 1-11, Wiley, New York.

Desmond, A.F. and Chapman G.R. (1993). Modelling task completion data with inverse Gaussian mixtures. *Journal of the Royal Statistical Society* C **42**, 603-613.

Heyde, C.C. and Morton, R. (1996) Quasi-likelihood and generalizing the EM algorithm. *Journal of the Royal Statistical Society* B **58**, 317-327.

Lehmann, E. L. (1959), *Testing Statistical Hypotheses.* John Wiley, New York.

Meng, X.L. and van Dyk, D.A. (1997). The EM Algorithm- An old folk song sung to a fast new tune (with discussion). *Journal of the Royal Statistical Society* B **59**, 511 - 567.

Neyman, J. (1934). On the two different aspects of the representative method. *Journal of the Royal Statistical Society* B **97**, 558-625.

Neyman, J. (1935). On the problem of confidence intervals, *Annals of Mathematical Statistics* **6**,111-116.

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philos. Trans. R. Soc. Lond.* A **236**, 333-380.

Neyman, J. (1941). Fiducial argument and the theory of confidence intervals. *Biometrika*,

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0282

32 128-150.

Neyman, J. (1977). Frequentist probability and frequentist statistics. *Synthese* **36**, 97-131.

Reid, C. (1982). *Neyman from Life*. Springer, New York.

Rubin, D. B. (1996) Multiple imputation after 18+ years. *Journal of American Statistical Association* **91**, 473-489.

Small, C. G. and McLeish, D. L. (1994). *Hilbert Space Methods in Probability and Statistical Inference*. Wiley, New York.

Tu, X.M., Meng, X.-L. and Pagano, M. (1993). The AIDS epidemic: Estimating survival after AIDS diagnosis from surveillance data. *Journal of American Statistical Association* **88**, 26-36.