140.840 Special Study: Decision Theory

Room 2008, Tuesday and Thursday, 3:30–4:50

Syllabus and links (in blue) to reference materials.

References and links in this file will be updated as the the course progresses.
Check for the latest links right after the class in question.

Th Aug 28 **Introduction**. Overview of the syllabus, goals of the class, ground rules and some fun examples.

Tu Sep 2 **Utility and Probability**. von-Neumann and Morgenstern's utility theory. Anscombe Aumann theory of subjective probability and utility.

(Jensen 1967)

(Anscombe and Aumann 1963)

Chapters 2 and 4.

Th Sep 4 **Utility theory in medical problems**. Meaning and elicitation of utilities in medical decision making. Multistage decision problems: the diagnostic test decision tree.

(Torrance et al. 1972) [Full Text]

(Parmigiani 2004) [Full Text]

Chapters 3 and 12.

Tu Sep 9 **Wald and Decision Functions**. We visit the birthplace of statistical decision theory, and discuss the first chapter of the first book on the subject. Statistical decision theory applies rational decision making to the choice of appropriate statistical strategies.

(Wald 1950)

(Savage 1951) [Full Text from Jstor]

Chapter 7.

Th Sep 11 No class.

Tu Sep 16 **Decision Functions for Testing and Estimation**. We revisit some of the traditional statistical problems, such as point estimation and testing, in the light of Wald's contribution. Where we reinterpret one of the milestones of statistical theory, the Neyman-Pearson lemma, as a decision theoretic result. Neyman-Pearson hypothesis testing played a key role in shaping Wald's theory, so this seems fair. We also go a little further and discuss the role of the likelihood ratio, Bayes factors, and posterior probabilities in testing.

(Neyman and Pearson 1933)

Chapters 7 and 8.

Th Sep 18 **Completeness and Sufficiency**. Rao-Blackwell theorem.

Chapter 8.

Tu Sep 23 **Admissibility**. Where we study admissibility, the most basic and influential rationality requirement of classical statistical decision theory. We discuss admissibility of Bayes rules and admissibility of $\bar{X}$ in estimating the mean of a normal.

Chapter 8.

(Blyth 1951) [Full Text from Jstor]

Th Sep 25 **Stein and Shrinkage Estimators**. Where we encounter the controversy stirred by the fact that $\bar{X}$ is inadmissible in estimating the mean of a multidimensional normal vector of observations. Stein was the first to realize this. We explore some of the important research direction that stemmed from Stein's paper, including shrinkage estimation, empirical Bayes estimation, and hierarchical modeling.

(Stein 1955)

Chapter 9.

Tu Sep 30 No Class.

Th Oct 2 No Class.

Tu Oct 7 **Scoring Rules.** Where we explore the implications of holding forecasters accountable for their predictions. We study the incentive systems that must be set in place for the forecasters to reveal their information/beliefs rather than using them to game the system. This leads to the study of proper scoring rules. We also discuss the relation between scoring rule and the axioms of probability, providing an alternative version of the "Dutch Book Theorem"

(Brier 1950)

(Winkler 1969) [Full Text from Jstor]

Chapter 10.

Th Oct 9 **The value of information**. Where we are reminded that the value of the information carried by a data set depends on what we intend to do with the data once we have collected it. And where we use decision trees to quantify this value. The life of a statistician is boiled down to a two-stage decision problem: what data to get, if any; and what use to make of the data in addressing the question at hand. Following Lindley, we explore a specific way of measuring the information in a data set, which tries to capture "generic learning" rather than specific usefulness in a given problem.

(Good 1967) [Full Text from Jstor]

(DeGroot 1984)

(Lindley 1956) [Full Text from Jstor]

Chapter 13.

Tu Oct 14 **Optimal Sample Size**. Where we finally come to terms with the single most common decision statisticians make in their daily activities: how big should a data set be? We try to understand how all the machinery we have been setting in place can help us and give some examples, both static and dynamic. When dynamic programming is applied to data collection, one's stopping rule may depend on past observations. Is this giving us a license to "sample to a foregone conclusion"? Good question.

(Raiffa and Schleifer 1961, Chapter 4)

(Wald 1947) (Selected chapters)

(Müller et al. 2005) [Preprint]

Chapter 14.

Th Oct 16 **Dynamic Programming**. Where we encounter a general approach for making decision dynamically, so that we can both learn from accruing knowledge and plan ahead to account for how present decisions will affect future decisions and future knowledge. This approach, called dynamic programming was developed in the 1950's by Bellman and has led to many great contributions, but we are still far from having general computational tools for the solution of complex dynamic decisions. We'll try to understand why the problem is so hard (the "curse of dimensionality").

Chapter 15.

Mo Oct 20 In the Genome Cafè. 1:30–3:00 and 3:30–5 Student presentations.

Tu Oct 21 Student presentations.

Lecture notes

The material in this class comes from many different sources and the lectures are organized as a collection of (related) individual seminars, many of which revisit classic papers. Lurder Inoue and I are organizing notes from previous incarnations of this class, working toward a book. I will create a mailing list and circulate the book by email immediately after each class. Please send me a note to be added to the mailing list.

We welcome all comments.

Good Books

Because of its very broad scope, our class will overlap with many existing textbooks, although no book I know covers more than about one half of the material. Here is a highly subjective list of references I used and loved over the years. You can use them to learn about topics in greater detail.

- Raiffa and Schleifer (1961) contributed enormously to defining practical Bayesian statistics and decision making in the earliest days of the field. The book was exploring new territory on almost every page and, even in describing the simplest practical ideas, is full of deep insight.
- DeGroot (1970) is another landmark book. It has extensive discussion of material in Part 3 and in depth treatment of foundations, which overlaps with Part 1, but gives a quite independent treatment of the material compared to the classical papers discussed in class.
- Berger (1985) is a more recent, comprehensive and complete reference for Bayesian statistical decision theory. It covers Part II in detail, and it includes material on Lectures 2,4,19 and 20, and minor additional overlaps.
- Ferguson (1967) is an excellent source for classical statistical decision theory.
- Robert (1994) covers Bayesian statistical decision theory at an advanced mathematical level and is the right place to learn stuff if you want to start proving theorems of your own.
- Bernardo and Smith (1994) focusses on Bayesian inference but approches are motivated with decision theoretic ideas throughout. It is a good source for scoring rules and decision theoretic model choice, among other topics.
- Schervish (1995) covers both statistical theory and its axiomatic foundation. It is very complete and accurate and has some of the most illuminating examples around.
- French (1988) and Smith (1987) cover some of the axiomatic foundations and then take off towards applications of decision theory to real world decision making.
- Lindley (1985) is a cristal clear introduction to basic decision making principles, decision trees et cetera.
- Kreps (1988) covers almost all of Part 1 in much greater technical detail. Nowhere else have I ever found as much insight into the meaning, goals and workings of the various axiomatizations. The style of the book is very informal —almost a transcription of the author's witty lectures, and is very enjoyable. Its digest of Savage's theory (Savage 1954) is unsurpassed.

References

Anscombe, F. J. and Aumann, R. J. (1963). A definition of subjective probability, *The Annals of Mathematical Statistics* **34**: 199–205.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, second edn, Springer-Verlag, New York.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*, Wiley, New York.

Blyth, C. R. (1951). On minimax statistical decision procedures and their admissibility, *The Annals of Mathematical Statistics* **22**: 22–42.

Brier, G. (1950). Verification of forecasts expressed in terms of probability, *Monthly Weather Review* **78**: 1–3.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*, Mc Graw-Hill.

DeGroot, M. H. (1984). Changes in utility as information, *Theory and Decision* **17**: 287–303.

Ferguson, T. S. (1967). *Mathematical statistics: a decision theoretic approach*, Academic Press.

French, S. (1988). *Decision theory: an introduction to the mathematics of rationality*, Ellis Horwood, Chichester.

Good, I. J. (1967). On the principle of total evidence, *British Journal for the Philosophy of Science* **17**: 319–321.

Jensen, N. E. (1967). An introduction to Bernoullian utility theory: I Utility functions., *Swedish Journal of Economics* **69**: 163–183.

Kreps, D. M. (1988). *Notes on the theory of choice*, Westview Press, Boulder.

Lindley, D. V. (1956). On a measure of the information provided by an experiment, *Annals of Mathematical Statistics* **27**: 986–1005.

Lindley, D. V. (1985). *Making Decisions (Second Edition)*, John Wiley & Sons.

Müller, P., Parmigiani, G., Robert, C. and Rousseau, J. (2005). Optimal sample size for multiple testing: the case of gene expression microarrays, *Journal of the American Statistical Association* **99**: 990–1001.

Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient test of statistical hypotheses, *Philosophical Transaction of the Royal Society (Series A)* **231**: 286–337.

Parmigiani, G. (2004). Uncertainty and the value of diagnostic information, with application to axillary lymph node dissection in breast cancer., *Stat Med* **23**(5): 843–855.
URL: *http://dx.doi.org/10.1002/sim.1623*

Raiffa, H. and Schleifer, R. (1961). *Applied Statistical Decision Theory*, Harvard University Press, Boston.

Robert, C. P. (1994). *The Bayesian Choice*, Springer-Verlag.

Savage, L. J. (1951). The theory of statistical decision, *Journal of the American Statistical Association* **46**: 55–67.

Savage, L. J. (1954). *The foundations of statistics*, Wiley, New York.

Schervish, M. J. (1995). *Theory of Statistics*, Springer-Verlag.

Smith, J. Q. (1987). *Decision Analysis: A Bayesian Approach*, Chapman & Hall.

Stein, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1**: 197–206.

Torrance, G., Thomas, W. and Sackett, D. (1972). A utility maximization model for evaluation of health care programs, *Health Services Research* **7**: 118–133.

Wald, A. (1947). *Sequential Analysis*, John Wiley & Sons, New York.

Wald, A. (1950). *Statistical Decision Functions*, John Wiley & Sons, New York.

Winkler, R. L. (1969). Scoring rules and the evaluation of probability assessors, *Journal of the American Statistical Association* **64**: 1073–1078.