A Course in Statistical Theory

David J. Olive
Southern Illinois University
Department of Mathematics
Mailcode 4408
Carbondale, IL 62901-4408
dolive@math.siu.edu

May 13, 2012

Contents

Preface

Many math and some statistics departments offer a one semester graduate course in statistical inference using texts such as Casella and Berger (2002), Bickel and Doksum (2007) or Mukhopadhyay (2000, 2006). The course typically covers minimal and complete sufficient statistics, maximum likelihood estimators (MLEs), method of moments, bias and mean square error, uniform minimum variance estimators (UMVUEs) and the Fréchet-Cramér-Rao lower bound (FCRLB), an introduction to large sample theory, likelihood ratio tests, and uniformly most powerful (UMP) tests and the Neyman Pearson Lemma. A major goal of this text is to make these topics much more accessible to students by using the theory of exponential families.

This material is essential for Masters and PhD students in biostatistics and statistics, and the material is often very useful for graduate students in economics, psychology and electrical engineering (especially communications and control).

The material is also useful for actuaries. According to (www.casact.org), topics for the CAS Exam 3 (Statistics and Actuarial Methods) include the MLE, method of moments, consistency, unbiasedness, mean square error, testing hypotheses using the Neyman Pearson Lemma and likelihood ratio tests, and the distribution of the max. These topics make up about 20% of the exam.

One of the most important uses of exponential families is that the theory often provides two methods for doing inference. For example, minimal sufficient statistics can be found with either the Lehmann-Scheffé theorem or by finding $\boldsymbol{T}$ from the exponential family parameterization. Similarly, if $Y_1, ..., Y_n$ are iid from a one parameter regular exponential family with complete sufficient statistic $T(\boldsymbol{Y})$, then one sided UMP tests can be found by using the Neyman Pearson lemma or by using exponential family theory.

The prerequisite for this text is a calculus based course in statistics at the level of Hogg and Tanis (2005), Larsen and Marx (2011), Wackerly, Mendenhall and Scheaffer (2008) or Walpole, Myers, Myers and Ye (2006). Also see Arnold (1990), Gathwaite, Joliffe and Jones (2002), Spanos (1999), Wasserman (2004) and Welsh (1996).

The following intermediate texts are especially recommended: DeGroot and Schervish (2011), Hogg, McKean and Craig (2012), Rice (2006) and Rohatgi (1984).

A less satisfactory alternative prerequisite is a calculus based course in probability at the level of Hoel, Port and Stone (1971), Parzen (1960) or Ross (1984, 2009).

A course in Real Analysis at the level of Bartle (1964), Gaughan (2009), Rosenlicht (1985), Ross (1980) or Rudin (1964) would be useful for the large sample theory chapter.

The following texts are at a similar to higher level than this text: Azzalini (1996), Bain and Engelhardt (1992), Berry and Lindgren (1995), Cox and Hinckley (1974), Ferguson (1967), Knight (2000), Lindgren (1993), Lindsey (1996), Mood, Graybill and Boes (1974), Roussas (1997) and Silvey (1970).

The texts Bickel and Doksum (2007), Lehmann and Casella (2003) and Rohatgi and Ehsanes Saleh (2001) are at a higher level as are Poor (1994) and Zacks (1971). The texts Bierens (2004), Cramér (1946), Keener (2010), Lehmann and Romano (2005), Rao (1973), Schervish (1995) and Shao (2003) are at a much higher level. Cox (2006) would be hard to use as a text, but is a useful monograph.

Some other useful references include a good low level probability text Ash (1993) and a good introduction to probability and statistics Dekking, Kraaikamp, Lopuhaä and Meester (2005). Also see Ash (2011), Spiegel (1975), Romano and Siegel (1986) and see online lecture notes by Ash at (www.math.uiuc.edu/∼r-ash/).

Many of the most important ideas in statistics are due to Fisher, Neyman, E.S. Pearson and K. Pearson. For example, David (2006-7) says that the following terms were due to Fisher: consistency, covariance, degrees of freedom, efficiency, information, information matrix, level of significance, likelihood, location, maximum likelihood, multinomial distribution, null hypothesis, pivotal quantity, probability integral transformation, sampling distribution, scale, statistic, Student's t, studentization, sufficiency, sufficient statistic, test of significance, uniformly most powerful test and variance.

David (2006-7) says that terms due to Neyman and E.S. Pearson include alternative hypothesis, composite hypothesis, likelihood ratio, power, power function, simple hypothesis, size of critical region, test criterion, test of hypotheses, type I and type II errors. Neyman also coined the term confidence interval.

David (2006-7) says that terms due to K. Pearson include binomial distribution, bivariate normal, method of moments, moment, random sampling, skewness, and standard deviation.

Also see, for example, David (1995), Fisher (1922), Savage (1976) and Stigler (2007). The influence of Gosset (Student) on Fisher is described in Zabell (2008) and Hanley, Julien and Moodie (2008). The influence of Karl Pearson on Fisher is described in Stigler (2008).

This book covers some of these ideas and begins by reviewing probability, counting, conditional probability, independence of events, the expected value and the variance. Chapter 1 also covers mixture distributions and shows how to use the kernel method to find $E(g(Y))$. Chapter 2 reviews joint, marginal, and conditional distributions; expectation; independence of random variables and covariance; conditional expectation and variance; location–scale families; univariate and multivariate transformations; sums of random variables; random vectors; the multinomial, multivariate normal and elliptically contoured distributions. Chapter 3 introduces exponential families while Chapter 4 covers sufficient statistics. Chapter 5 covers maximum likelihood estimators and method of moments estimators. Chapter 6 examines the mean square error and bias as well as uniformly minimum variance unbiased estimators, Fisher information and the Fréchet-Cramér-Rao lower bound. Chapter 7 covers uniformly most powerful and likelihood ratio tests. Chapter 8 gives an introduction to large sample theory while Chapter 9 covers confidence intervals. Chapter 10 gives some of the properties of 54 univariate distributions, many of which are exponential families. Chapter 10 also gives over 30 exponential family parameterizations, over 28 MLE examples and over 27 UMVUE examples. Chapter 11 gives some hints for the problems.

Some highlights of this text follow.

- Exponential families, indicator functions and the support of the distribution are used throughout the text to simplify the theory.

- Section 1.5 describes the kernel method, a technique for computing $E(g(Y))$, in detail rarely given in texts.

- Theorem 2.2 shows the essential relationship between the independence of random variables $Y_1, ..., Y_n$ and the support in that the random variables are dependent if the support is not a cross product. If the support is a cross product and if the joint pdf or pmf factors on the support, then $Y_1, ..., Y_n$ are independent.

- Theorems 2.17 and 2.18 give the distribution of $\sum Y_i$ when $Y_1, ..., Y_n$ are iid for a wide variety of distributions.
- Chapter 3 presents exponential families. The theory of these distributions greatly simplifies many of the most important results in mathematical statistics.
- Corollary 4.6 presents a simple method for finding sufficient, minimal sufficient and complete statistics for k-parameter exponential families.
- Section 5.4.1 compares the "proofs" of the MLE invariance principle due to Zehna (1966) and Berk (1967). Although Zehna (1966) is cited by most texts, Berk (1967) gives a correct elementary proof.
- Theorem 6.5 compares the UMVUE and the estimator that minimizes the MSE for a large class of one parameter exponential families.
- Theorem 7.3 provides a simple method for finding uniformly most powerful tests for a large class of 1–parameter exponential families. Power is examined in Example 7.10.
- Theorem 8.4 gives a simple proof of the asymptotic efficiency of the complete sufficient statistic as an estimator of its expected value for 1–parameter regular exponential families.
- Theorem 8.21 provides a simple limit theorem for the complete sufficient statistic of a k–parameter regular exponential family.
- Chapter 10 gives information for 54 "brand name" distributions.

Much of the text material is on parametric frequentist methods, but the most used methods in statistics tend to be semiparametric. Many of the most used methods originally based on the univariate or multivariate normal distribution are also semiparametric methods. For example the t-interval works for a large class of distributions if σ^2 is finite and n is large. Similarly, least squares regression is a semiparametric method. Multivariate analysis procedures originally based on the multivariate normal distribution tend to also work well for a large class of elliptically contoured distributions.

In a semester, I cover Sections 1.1–1.6, 2.1–2.9, 3.1, 3.2, 4.1, 4.2, 5.1, 5.2, 6.1, 6.2, 7.1, 7.2 and 8.1-8.4.

Warning: For parametric methods that are not based on the normal distribution, often the methods work well if the parametric distribution is a good approximation to the data, but perform very poorly otherwise.

Acknowledgements

Teaching the material to Math 580 students at Southern Illinois University in 2001, 2004, 2008 (when the text was first used) and 2012 was very useful. Some of the Chapter 8 material came from a reading course in Large Sample Theory taught to SIU students on two occasions. Some of the SIU QUAL problems were written by Bhaskar Bhattacharya, Sakthivel Jeyaratnam, and Abdel Mugdadi, who also contributed several solutions. A small part of the research in Chapters 9 and 10 was supported by NSF grant DMS 0600933.