

A Course in Statistical Theory

David J. Olive

Southern Illinois University
Department of Mathematics

Mailcode 4408

Carbondale, IL 62901-4408

dolive@siu.edu

January 2008, notes

September 29, 2013

Contents

Preface	v
1 Probability and Expectations	1
1.1 Probability	1
1.2 Counting	3
1.3 Conditional Probability and Independence	4
1.4 The Expected Value and Variance	7
1.5 The Kernel Method	13
1.6 Mixture Distributions	16
1.7 Complements	19
1.8 Problems	22
2 Multivariate Distributions	29
2.1 Joint, Marginal and Conditional Distributions	29
2.2 Expectation, Covariance and Independence	33
2.3 Conditional Expectation and Variance	43
2.4 Location–Scale Families	46
2.5 Transformations	47
2.6 Sums of Random Variables	56
2.7 Random Vectors	62
2.8 The Multinomial Distribution	65
2.9 The Multivariate Normal Distribution	67
2.10 Elliptically Contoured Distributions	70
2.11 Complements	74
2.12 Problems	75
3 Exponential Families	88
3.1 Regular Exponential Families	88

3.2	Properties of $(t_1(Y), \dots, t_k(Y))$	96
3.3	Complements	101
3.4	Problems	102
4	Sufficient Statistics	106
4.1	Statistics and Sampling Distributions	106
4.2	Minimal Sufficient Statistics	112
4.3	Summary	124
4.4	Complements	127
4.5	Problems	127
5	Point Estimation	133
5.1	Maximum Likelihood Estimators	133
5.2	Method of Moments Estimators	145
5.3	Summary	147
5.4	Complements	148
5.4.1	Two “Proofs” of the Invariance Principle	149
5.5	Problems	152
6	UMVUEs and the FCRLB	160
6.1	MSE and Bias	160
6.2	Exponential Families, UMVUEs and the FCRLB.	163
6.3	Summary	172
6.4	Complements	173
6.5	Problems	173
7	Testing Statistical Hypotheses	182
7.1	Exponential Families, the Neyman Pearson Lemma, and UMP Tests	183
7.2	Likelihood Ratio Tests	190
7.3	Summary	194
7.4	Complements	197
7.5	Problems	198
8	Large Sample Theory	203
8.1	The CLT, Delta Method and an Exponential Family Limit Theorem	203
8.2	Asymptotically Efficient Estimators	211

8.3	Modes of Convergence and Consistency	214
8.4	Slutsky's Theorem and Related Results	220
8.5	Order Relations and Convergence Rates	223
8.6	Multivariate Limit Theorems	226
8.7	More Multivariate Results	231
8.8	Summary	233
8.9	Complements	234
8.10	Problems	236
9	Confidence Intervals	247
9.1	Introduction	247
9.2	Some Examples	253
9.3	Complements	269
9.4	Problems	269
10	Some Useful Distributions	274
10.1	The Beta Distribution	276
10.2	The Beta–Binomial Distribution	277
10.3	The Bernoulli and Binomial Distributions	278
10.4	The Burr Distribution	281
10.5	The Cauchy Distribution	282
10.6	The Chi Distribution	283
10.7	The Chi–square Distribution	285
10.8	The Discrete Uniform Distribution	286
10.9	The Double Exponential Distribution	287
10.10	The Exponential Distribution	290
10.11	The Two Parameter Exponential Distribution	291
10.12	The F Distribution	293
10.13	The Gamma Distribution	294
10.14	The Generalized Gamma Distribution	297
10.15	The Generalized Negative Binomial Distribution	297
10.16	The Geometric Distribution	298
10.17	The Gompertz Distribution	299
10.18	The Half Cauchy Distribution	300
10.19	The Half Logistic Distribution	300
10.20	The Half Normal Distribution	301
10.21	The Hypergeometric Distribution	303
10.22	The Inverse Gaussian Distribution	303

10.23	The Inverted Gamma Distribution	306
10.24	The Largest Extreme Value Distribution	306
10.25	The Logarithmic Distribution	308
10.26	The Logistic Distribution	308
10.27	The Log-Cauchy Distribution	309
10.28	The Log-Logistic Distribution	309
10.29	The Lognormal Distribution	310
10.30	The Maxwell-Boltzmann Distribution	313
10.31	The Negative Binomial Distribution	314
10.32	The Normal Distribution	316
10.33	The One Sided Stable Distribution	319
10.34	The Pareto Distribution	321
10.35	The Poisson Distribution	323
10.36	The Power Distribution	325
10.37	The Rayleigh Distribution	326
10.38	The Smallest Extreme Value Distribution	328
10.39	The Student's t Distribution	329
10.40	The Topp-Leone Distribution	330
10.41	The Truncated Extreme Value Distribution	331
10.42	The Uniform Distribution	332
10.43	The Weibull Distribution	333
10.44	The Zeta Distribution	335
10.45	Complements	335
11	Stuff for Students	336
11.1	R/Splus Statistical Software	336
11.2	Hints and Solutions to Selected Problems	339

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), 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 \mathbf{T} from the exponential family parameterization. Similarly, if Y_1, \dots, Y_n are iid from a one parameter regular exponential family with complete sufficient statistic $T(\mathbf{Y})$, 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 (2001), Wackerly, Mendenhall and Scheaffer (2002) or Walpole, Myers, Myers and Ye (2002). 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 (2001), Hogg, Craig and McKean (2004), 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).

A course in Real Analysis at the level of Bartle (1964), Gaughan (1993), Rosenlicht (1985), Ross (1980) or Rudin (1964) would be useful for the large sample theory chapters.

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 Hinkley (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 (1976) are at a higher level as are Poor (1994) and Zacks (1971). The texts Bierens (2004), Cramér (1946), Lehmann and Romano (2005), Rao (1965), 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 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 R.A. Fisher. See, for example, David (1995), Fisher (1922), Savage (1976) and Stigler (2008). The 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 44 univariate distributions, many of which are exponential families. The MLEs and UMVUEs for the parameters are derived for several of these distributions. 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, \dots, 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, \dots, Y_n are independent.
- Theorems 2.17 and 2.18 give the distribution of $\sum Y_i$ when Y_1, \dots, 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 7.3 provides a simple method for finding uniformly most powerful tests for a large class on 1-parameter exponential families.

- 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 on many more “brand name” distributions than is typical.

Much of the course 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.

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 was very useful. Some of the Chapter 8 material came from a reading course in Large Sample Theory taught to 2 SIU students. Some of the SIU QUAL problems were written by Bhaskar Bhattacharya, Sakthivel Jeyaratnam, and Abdel Mugdadi, who also contributed several solutions.