

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	vi
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 Summary	19
1.8 Complements	22
1.9 Problems	22
2 Multivariate Distributions	31
2.1 Joint, Marginal and Conditional Distributions	31
2.2 Expectation, Covariance and Independence	35
2.3 Conditional Expectation and Variance	45
2.4 Location–Scale Families	48
2.5 Transformations	49
2.6 Sums of Random Variables	58
2.7 Random Vectors	64
2.8 The Multinomial Distribution	67
2.9 The Multivariate Normal Distribution	69
2.10 Elliptically Contoured Distributions	72
2.11 Summary	76
2.12 Complements	78
2.13 Problems	79

3	Exponential Families	92
3.1	Regular Exponential Families	92
3.2	Properties of $(t_1(Y), \dots, t_k(Y))$	102
3.3	Summary	107
3.4	Complements	108
3.5	Problems	110
4	Sufficient Statistics	114
4.1	Statistics and Sampling Distributions	114
4.2	Minimal Sufficient Statistics	121
4.3	Summary	135
4.4	Complements	138
4.5	Problems	139
5	Point Estimation	147
5.1	Maximum Likelihood Estimators	147
5.2	Method of Moments Estimators	161
5.3	Summary	163
5.4	Complements	164
5.4.1	Two “Proofs” of the Invariance Principle	166
5.5	Problems	169
6	UMVUEs and the FCRLB	178
6.1	MSE and Bias	178
6.2	Exponential Families, UMVUEs and the FCRLB.	181
6.3	Summary	192
6.4	Complements	193
6.5	Problems	193
7	Testing Statistical Hypotheses	203
7.1	Exponential Families, the Neyman Pearson Lemma, and UMP Tests	204
7.2	Likelihood Ratio Tests	212
7.3	Summary	215
7.4	Complements	219
7.5	Problems	221

8	Large Sample Theory	228
8.1	The CLT, Delta Method and an Exponential Family Limit Theorem	228
8.2	Asymptotically Efficient Estimators	236
8.3	Modes of Convergence and Consistency	240
8.4	Slutsky's Theorem and Related Results	245
8.5	Order Relations and Convergence Rates	249
8.6	Multivariate Limit Theorems	252
8.7	More Multivariate Results	257
8.8	Summary	259
8.9	Complements	260
8.10	Problems	262
9	Confidence Intervals	274
9.1	Introduction	274
9.2	Some Examples	281
9.3	Bootstrap and Randomization Tests	300
9.4	Complements	307
9.5	Problems	307
10	Some Useful Distributions	314
10.1	The Beta Distribution	316
10.2	The Beta–Binomial Distribution	317
10.3	The Binomial and Bernoulli Distributions	318
10.4	The Birnbaum Saunders Distribution	321
10.5	The Burr Type III Distribution	322
10.6	The Burr Type X Distribution	323
10.7	The Burr Type XII Distribution	324
10.8	The Cauchy Distribution	325
10.9	The Chi Distribution	326
10.10	The Chi–square Distribution	328
10.11	The Discrete Uniform Distribution	329
10.12	The Double Exponential Distribution	331
10.13	The Exponential Distribution	333
10.14	The Two Parameter Exponential Distribution	335
10.15	The F Distribution	337
10.16	The Gamma Distribution	337
10.17	The Generalized Gamma Distribution	340

10.18	The Generalized Negative Binomial Distribution	341
10.19	The Geometric Distribution	342
10.20	The Gompertz Distribution	343
10.21	The Half Cauchy Distribution	344
10.22	The Half Logistic Distribution	344
10.23	The Half Normal Distribution	345
10.24	The Hypergeometric Distribution	347
10.25	The Inverse Exponential Distribution	347
10.26	The Inverse Gaussian Distribution	348
10.27	The Inverse Weibull Distribution	351
10.28	The Inverted Gamma Distribution	352
10.29	The Largest Extreme Value Distribution	353
10.30	The Logarithmic Distribution	354
10.31	The Logistic Distribution	355
10.32	The Log-Cauchy Distribution	356
10.33	The Log-Gamma Distribution	356
10.34	The Log-Logistic Distribution	356
10.35	The Lognormal Distribution	357
10.36	The Maxwell-Boltzmann Distribution	360
10.37	The Modified DeMoivre's Law Distribution	361
10.38	The Negative Binomial Distribution	362
10.39	The Normal Distribution	364
10.40	The One Sided Stable Distribution	368
10.41	The Pareto Distribution	369
10.42	The Poisson Distribution	372
10.43	The Power Distribution	373
10.44	The Two Parameter Power Distribution	375
10.45	The Rayleigh Distribution	376
10.46	The Smallest Extreme Value Distribution	378
10.47	The Student's t Distribution	379
10.48	The Topp-Leone Distribution	379
10.49	The Truncated Extreme Value Distribution	381
10.50	The Uniform Distribution	382
10.51	The Weibull Distribution	383
10.52	The Zero Truncated Poisson Distribution	385
10.53	The Zeta Distribution	385
10.54	The Zipf Distribution	386
10.55	Complements	387

11 Stuff for Students	388
11.1 R/Splus Statistical Software	388
11.2 Hints and Solutions to Selected Problems	390
11.3 Tables	426

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 \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 (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, \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 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.