

David J. Olive

# Large Sample Theory

January 14, 2022





# Preface

Many statistics departments offer a one semester graduate course in large sample theory. A nice review of large sample theory is Chernoff (1956). There are several PhD level texts on large sample theory including, in roughly increasing order of difficulty, Lehmann (1999), Ferguson (1996), Sen and Singer (1993), and Serfling (1980). Cramér (1946) is also an important reference, and White (1984) considers asymptotic theory for econometric applications. The online text Hunter (2014) is useful. Also see DasGupta (2008), Davidson (1994), Jiang (2010), Polansky (2011), Sen, Singer, and Pedrosa De Lima (2010), and van der Vaart (1998).

For some roughly Master's level large sample theory, see Bickel and Doksum (1977, section 4.4), Casella and Berger (2002, section 5.5), Hoel, Port, and Stone (1971, sections 8.2-8.4), Lehmann (1983, ch. 5), Olive (2014, ch. 8), Rohatgi (1976, ch. 6), Rohatgi (1984, ch. 9), and Woodroffe (1975, ch. 9).

The prerequisite for this text is a Master's level course in Statistics (USA) such as Casella and Berger (2002) or Olive (2014).

Some highlights of this text follow.

- The large sample theory for the elastic net, lasso, and ridge regression is greatly simplified.

**Downloading the book's R functions** *lsamppack.txt* and data files *lsampdata.txt* into *R*: The commands

```
source("http://parker.ad.siu.edu/Olive/lsamppack.txt")
source("http://parker.ad.siu.edu/Olive/lsampdata.txt")
```

## Acknowledgements

Teaching large sample theory as Math 582 in 2022 and as a reading course at Southern Illinois University was useful.



# Contents

<b>1</b>	<b>Introduction</b> .....	1
1.1	Probability, Expected Value, CDF .....	1
1.2	Multivariate Distributions .....	6
1.3	Characteristic Function, MGF, CGF .....	14
1.4	Sums of Random Variables .....	18
1.5	The Multivariate Normal Distribution .....	23
1.6	Exponential Families .....	26
1.6.1	Properties of $(t_1(Y), \dots, t_k(Y))$ .....	32
1.7	MSE, Information Number, MLE, UMVUE .....	34
1.8	Mixture Distributions .....	39
1.9	Elliptically Contoured Distributions .....	41
1.10	Some Useful Distributions .....	41
1.11	Summary .....	47
1.12	Complements .....	47
1.13	Problems .....	47
<b>2</b>	<b>Univariate Limit Theorems</b> .....	49
2.1	The CLT, Delta Method and an Exponential Family Limit Theorem .....	49
2.2	Asymptotically Efficient Estimators .....	57
2.3	Modes of Convergence and Consistency .....	61
2.4	Slutsky's Theorem and Related Results .....	65
2.5	Order Relations and Convergence Rates .....	69
2.6	Summary .....	72
2.7	Complements .....	74
2.8	Problems .....	75
<b>3</b>	<b>Multivariate Limit Theorems</b> .....	85
3.1	Multivariate Limit Theorems .....	85
3.2	More Multivariate Results .....	90
3.3	Summary .....	94

3.4	Complements	94
3.5	Problems	94
4	Prediction Intervals and Prediction Regions	97
4.1	Prediction Intervals	97
4.2	Prediction Regions	102
4.3	Prediction Regions If $n/p$ Is Small	109
4.4	Summary	109
4.5	Complements	109
4.6	Problems	111
5	Confidence Regions and the Bootstrap	113
5.1	Confidence Intervals	113
5.2	Large Sample CIs and Tests	116
5.3	Some CI Examples	119
5.4	Bootstrap Confidence Regions and Hypothesis Tests	128
5.4.1	The Bootstrap	131
5.4.2	Bootstrap Confidence Regions for Hypothesis Testing	134
5.4.3	Theory for Bootstrap Confidence Regions	138
5.5	Summary	143
5.6	Complements	143
5.7	Problems	144
6	Regression: GLMs, GAMs, Statistical Learning	149
6.1	Multiple Linear Regression	150
6.1.1	Ordinary Least Squares	152
6.1.2	$L_1$	161
6.2	Bootstrapping OLS MLR	162
6.2.1	The Parametric Bootstrap	162
6.2.2	The Residual Bootstrap	163
6.2.3	The Nonparametric Bootstrap	165
6.3	Statistical Learning Methods for MLR	166
6.3.1	Ridge Regression	167
6.3.2	Lasso	170
6.3.3	The Elastic Net	172
6.3.4	Ridge Type Regression Estimators	175
6.4	Weighted Least Squares	176
6.5	GLMs and Related Regression Models	176
6.6	Survival Regression	178
6.7	Bootstrapping Some Regression Models	179
6.7.1	Parametric Bootstrap	179
6.7.2	Nonparametric Bootstrap	180
6.8	Variable Selection	180

6.8.1	Large Sample Theory for Variable Selection Estimators .....	182
6.8.2	Bootstrapping Variable Selection Estimators ...	186
6.9	Data Splitting .....	189
6.10	Prediction Intervals .....	189
6.11	Multivariate Linear Regression .....	189
6.11.1	Testing Hypotheses .....	192
6.11.2	Asymptotically Optimal Prediction Regions ....	203
6.12	Summary .....	208
6.13	Complements .....	208
6.14	Problems .....	209
<b>7</b>	<b>Experimental Design and One Way MANOVA .....</b>	<b>211</b>
7.1	Summary .....	211
7.2	Complements .....	211
7.3	Problems .....	211
<b>8</b>	<b>Robust Statistics .....</b>	<b>213</b>
8.1	The Location Model .....	213
8.1.1	Robust Confidence Intervals .....	217
8.1.2	Some Two Stage Trimmed Means .....	219
8.1.3	Asymptotics for Two Stage Trimmed Means ...	223
8.1.4	Asymptotic Theory for the MAD .....	226
8.1.5	Truncated Distributions .....	229
8.1.6	Asymptotic Variances for Trimmed Means ....	234
8.2	The Multivariate Location and Dispersion Model ....	238
8.2.1	Affine Equivariance .....	240
8.2.2	Breakdown .....	241
8.2.3	The Concentration Algorithm .....	244
8.2.4	Theory for Practical Estimators .....	248
8.2.5	The RMVN and RFCH Sets .....	256
8.2.6	MLD Outlier Detection if $p > n$ .....	258
8.3	Resistant Multiple Linear Regression .....	259
8.4	Robust Regression .....	265
8.4.1	MLR Breakdown and Equivariance .....	265
8.4.2	A Practical High Breakdown Consistent Estimator .....	273
8.5	The Robust <code>rmreg2</code> Estimator .....	277
8.6	Summary .....	278
8.7	Complements .....	280
8.8	Problems .....	282

<b>9</b>	<b>Time Series</b> .....	285
9.1	<b>Prediction Intervals</b> .....	285
9.2	<b>Inference after Model Selection</b> .....	285
9.3	<b>Summary</b> .....	285
9.4	<b>Complements</b> .....	285
9.5	<b>Problems</b> .....	285
<b>10</b>	<b>Graphical Diagnostics</b> .....	287
10.1	<b>1D Regression</b> .....	287
10.2	<b>Plots for MLR</b> .....	288
10.2.1	<b>Plots for Variable Selection</b> .....	290
10.2.2	<b>Plots for Response Transformations</b> .....	295
10.3	<b>Plots for GLMs and GAMs</b> .....	295
10.4	<b>Outlier Detection for the MLD Model</b> .....	295
10.5	<b>Summary</b> .....	300
10.6	<b>Complements</b> .....	300
10.7	<b>Problems</b> .....	300
<b>11</b>	<b>More Results</b> .....	301
11.1	<b>Hints and Solutions to Selected Problems</b> .....	301
11.2	<b>Tables</b> .....	306
11.3	<b>Summary</b> .....	308
11.4	<b>Complements</b> .....	308
11.5	<b>Problems</b> .....	308
	<b>Index</b> .....	327