STATISTICS 512: TECHNIQUES OF MATHEMATICS FOR STATISTICS

We view topics 1 - 3, *including the sub-headings*, as essential material which must be included each time this course is taught. For these topics, motivating examples are to be drawn from common problems in Statistics and Probability. It is anticipated that the sub-headings under topic 4 will vary from year to year; those given here are examples of applications which have been taught in previous years.

(1) Matrix theory (6 lecture hours)

- Vector spaces, column spaces, rank
- Orthogonality, projections, Gram-Schmidt Theorem
- Least squares estimation, spectral decomposition

(2) Calculus and Analysis (15 hours)

- Limits and continuity
- Probability spaces, random variables, distributions
- Convergence in probability, Jensen's inequality
- Differentiaion, Mean Value and Taylor's Theorems
- Transformations of random variables, variance stabilization
- Sequences and series
- Power series, moment and probability generating functions
- Riemann and Riemann-Stieltjes integration
- Chebyshev's inequality, Weak Law of Large Numbers, Central Limit Theorem

(3) Multidimensional calculus and optimization (3 hours)

- Taylor's and Inverse Function Theorems
- Implicit Function Theorem, extrema, Lagrange multipliers
- Integration in $\mathbb{R}^n$

(4) Specific applications in Statistics (12 hours)

- Transformations of random vectors, distribution of mean and variance in Normal samples
- Newton-Raphson and Gauss-Newton methods
- Regression design, simulated annealing
- Maximum likelihood estimation, Information inequality
- Minimax variance M-estimation

A suitable textbook covering (most of) this material is *Advanced Calculus with Applications in Statistics* by A.I. Khuri.

STATISTICS 532: SURVIVAL ANALYSIS

We view topics 1 - 4 as <u>essential</u> material which must be included each time this course is taught. The sub-headings may vary from year to year; those given here have been taught in previous years, and are recommended.

(1) General introduction (3 lecture hours)
- Survival function, hazard function
- Censoring, truncation with practical applications
- Other biases with discussion of practical situations

(2) Non-parametric approaches (10 hours)
- Kaplan-Meier estimator, life tables and their variants
- Comparison of survival functions
- Estimation of quantities such as median survival time

(3) Parametric approaches (10 hours)
- Models for survival functions and hazard functions and inference
- Regression models and interpretation of the parameters
- Exponential and Weibull distribution
- Regression diagnostics

(4) Semi-parametric approaches (10 hours)
- Cox proportional hazards model
- Partial likelihood

(5) Additional topics as time permits
- Stratified Cox proportional model
- Frailty models
- Discrete data survival models

Suitable textbooks covering (most of) this material are *Statistical Methods for Survival Data Analysis* by E.T. Lee and J.W. Wang and *Applied Survival Analysis* by Hosmer and Lemeshow.

STATISTICS 561: SAMPLE SURVEY METHODOLOGY

We view topics 1 - 6 as essential material which must be included each time this course is taught. The sub-headings may vary from year to year; those given here have been taught in previous years, and are recommended.

(1) Review of basic sampling schemes (4 lecture hours)
  - Simple random sampling
  - Systematic sampling
  - Stratified random sampling

(2) Unequal Probability Sampling Designs (8 hours)
  - With replacement design: Probability proportional to size sampling, Selection schemes estimation of population totals
  - Without replacement unequal probability sampling schemes: IPPS and Non-IPPS schemes, PPS systematic sampling, Rao-Hartley-Cochran procedure
  - Estimation of variance of linear estimators

(3) Cluster and Multistage Sampling (6 hours)
  - Estimation of finite population totals
  - Variance estimation

(4) Model based and model-assisted methods (8 lecture hours)
  - Ratio and Regression methods of estimation
  - Estimating equations
  - Distribution function and quantile estimation

(5) Variance estimation for functions of linear estimators (5 hours)
  - Linearization method
  - Jackknife method
  - Balanced repeated replication method

(6) Imputation methods (4 hours)

(7) Additional topics as time permits
  - Small area estimation
  - Network sampling
  - Response errors
  - Spatial sampling

Suitable textbooks covering (most of) this material are *Sampling: Design and Analysis* by S. Lohr together with *Theory of Sample Surveys* by M. E. Thompson.

STATISTICS 562: DISCRETE DATA ANALYSIS

We view topics 1 - 4 as essential material which must be included each time this course is taught. The sub-headings may vary from year to year; those given here have been taught in previous years, and are recommended.

(1) Review of basic discrete models (3 lecture hours)
- Binomial, Poisson and Multinomial Models
- Estimation and asymptotic theory

(2) Two-way contingency tables (4 hours)
- Measures of association
- Goodness-of-fit
- Testing independence
- Exact tests for small samples

(3) Log-linear models (8 hours)
- Two dimensional tables
- Three and higher dimensional tables
- Measures of association, goodness-of-fit and model selection

(4) Loglinear-logit models for ordinal variables (8 hours)
- Logit models
- Linear-by-linear association
- Row effects and column effects models
- Testing independence for ordinal classification

(5) Additional topics as time permits
- Mulitnomial response models
- Models for matched pairs
  - Tests for symmetry
  - Marginal homogeneity
  - Measures of agreement
- Analysis of repeated categorical response data

A suitable textbook covering (most of) this material is *Categorical Data Analysis* by A. Agresti.

STATISTICS 566: METHODS OF STATISTICAL INFERENCE

We view topics 1 - 4, *including the sub-headings*, as essential material which must be included each time this course is taught. The sub-headings may vary from year to year; those given here have been taught in previous years, and are recommended.

(1) Principles of data reduction (6 lecture hours)

- The sufficiency principle (sufficient statistics, minimal sufficient statistics, ancillarity, completeness)
- The likelihood principle (the likelihood function, the formal likelihood principle)

(2) Point estimation (12 hours)

- Basu's Theorem, Rao-Blackwell Theorem, Lehmann-Scheffé Theorem
- Information inequality
- Methods of finding estimators (method of moments, maximum likelihood estimators, Bayes estimators)
- Methods of evaluating estimators (mean squared error, best unbiased estimators, sufficiency and unbiasedness, loss function optimality, bootstrap and other Monte Carlo methods)

(3) Hypothesis testing (12 hours)

- Neymann-Pearson Lemma
- Methods of finding tests (likelihood ratio tests, Bayesian tests)
- Methods of evaluating tests (error probabilities and power function, most powerful tests, p-values, bootstrap and other Monte Carlo methods)

(4) Interval estimation (6 hours)

- Methods of finding interval estimators (inverting test statistics, pivotal quantities, pivoting the CDF, Bayesian intervals)
- Methods of evaluating interval estimators (size and coverage probability, bootstrap and other Monte Carlo methods)

A suitable textbook covering (most of) this material is *Statistical Inference* by G. Casella and R.L. Berger.

STATISTICS 568: DESIGN AND ANALYSIS OF EXPERIMENTS

We view topics 1 - 3, *including the sub-headings*, as essential material which must be included each time this course is taught. The sub-headings under topic 4 will vary from year to year; those given here have been taught in previous years, and are recommended.

(1) Basic principles, single factor experiments (6 lecture hours)

- Basic principles of replication, randomization, and blocking
- The general linear model
- Experiments with a single factor
- Multiple comparisons

(2) Experiments with several factors (6 hours)

- Randomized block designs
- Two-way layout
- Multi-way layout
- Latin square design
- Analysis of covariance

(3) Factorial designs (9 hours)

- Full factorial and fractional factorial experiments at two levels
- Confounding
- Full factorial and fractional factorial experiments at three levels

(4) Additional topics as time permits (15 hours)

- Response surface methodology
- Introduction to robust parameter design
- Random effects models and mixed models
- Nested designs
- Randomization restrictions

A suitable textbook covering (most of) this material is *Experiments: Planning, Analysis, and Parameter Design Optimization* by C. F. J. Wu and M. Hamada.

STATISTICS 571: APPLIED MEASURE THEORY FOR PROBABILITY

We view topics 1 - 3, *including the sub-headings*, as essential material which must be included each time this course is taught. For these topics, motivating examples are, where possible, to be drawn from common problems in Statistics.

(1) Measure and integration (15 lecture hours)

- Fundamentals of measure and integration
- Probability fields, random variables
- Product measure, independence of random variables, stochastic processes

(2) Convergence of random variables and probability measures (16 hours)

- Convergence in L_p, almost surely and in probability; relationships among these
- Borel-Cantelli Lemma
- Weak and Strong Laws of Large Numbers
- Weak convergence of distribution functions, Glivenko-Cantelli Theorem
- Weak and vague convergence of (sub-) probability measures
- Skorohod's Theorem
- Characteristic functions
- Central Limit Theory

(3) Conditional probability and expectation (5 hours)

Suitable textbooks covering (most of) this material are *Probability and Measure* by P. Billingsley, *Probability: Theory and Examples* by R. Durrett, *Probability for Statisticians* by G.R. Shorack and *A Course in Probability Theory* by K.L. Chung.

STATISTICS 575: MULTIVARIATE ANALYSIS

We view topics 1 - 3, *including the sub-headings*, as essential material which must be included each time this course is taught. It is anticipated that the sub-headings under topic 4 will vary from year to year; those given here are examples of applications which have been taught in previous years.

(1) Multivariate Normal theory (10 lecture hours)

- Matrix algebra and random vectors
- Sampling geometry and random sampling
- Multivariate normal distribution
- Inferences about a mean vector
- Comparison of several multivariate means
- Regression with multivariate response

(2) Classification and grouping (10 hours)

- Discriminant analysis and classification
- Clustering

(3) Analysis of covariance structure (10 hours)

- Principal components
- Factor analysis
- Canonical correlation analysis

(4) Additional topics as time permits

- Structural equations modelling
- Seemingly unrelated regression models
- Multivariate discrete models
- Correspondence analysis

A suitable textbook covering (most of) this material is *Applied Multivariate Statistical Analysis* by R.A. Johnson and D.W. Wichern.

STATISTICS 578: REGRESSION ANALYSIS

We view topics 1 and 2, *including the sub-headings*, as essential material which must be included each time this course is taught. As well, the remaining topic(s) are to be chosen from the sub-headings within topic 3.

(1) Linear regression (9 lecture hours)

- Matrix approach to least squares estimation
- Canonical reduction of linear model for estimation and testing theory
- Multicollinearity and ridge regression
- Regression diagnostics
- Model selection

(2) Nonlinear regression (9 hours)

- Nonlinear models
- Iterative estimation, numerical methods
- Inference in nonlinear regression: testing, confidence regions

(3) Further topics (18 hours) chosen from:

- Further topics in linear regression
- Robust regression
- Nonparametric regression
- Binary data regression
- Count data regression and other generalized linear models

Suitable textbooks covering (most of) topics 1 and 2 are *Introduction to Linear Regression Analysis* by Montgomery and Peck, *Applied Regression Analysis* by Draper and Smith, and *Nonlinear Regression Analysis and Its Applications* by Bates and Watts.

STATISTICS 580: STOCHASTIC PROCESSES

We view topics 1 - 5, *including the sub-headings*, as essential material which must be included each time this course is taught.

(1) Finite Markov chains (7 lecture hours)
- Basic definitions
- Calculating probabilities
- Invariant probabilities
- Classification of states
- Hitting times

(2) Countable Markov chains (7 hours)
- Recurrence and transience
- Difference equations and random walks on $\mathbb{Z}$
- General random walks
- Branching processes

(3) Martingales (5 hours)
- Conditional expectation
- Martingales
- Optional sampling theorem
- Martingale convergence theorem

(4) Continuous time Markov chains (5 hours)
- Poisson processes
- Birth and death processes
- Queues

(5) Brownian motion (7 hours)
- Basic properties
- Reflection principle
- Dirichlet problem
- Stochastic integration
- Integration with respect to Brownian motion
- Ito's formula

(6) Additional topics as time allows. A recommendation is
- Optimal stopping
  - Strategies for winning
  - Algorithms to find optimal strategies
  - Binomial pricing model

A suitable text covering (most of) this material is *Introduction to Stochastic Processes* by G.F. Lawler.

STATISTICS 664: ADVANCED STATISTICAL INFERENCE

Prerequisites: STAT 512, 566

This course looks at the topic of statistical inference from the non-likelihood perspective. In most statistical inference courses, we learn that the likelihood approach works very well in most situations. However, it is well known that likelihood inference fails in a number of practical situations, most famously in the presence of infinitely many nuisance parameters. Likelihood inference also suffers from lack of robustness to model specification. Several techniques have been developed to address and correct the problems with the likelihood approach in the presence of nuisance parameters.

Topics:

(1) Various versions of the likelihood and their applications: Conditional likelihood, Marginal likelihood, Integrated likelihood, Profile likelihood, Partial likelihood, Empirical Likelihood, etc.

(2) Theory of estimating functions and their unifying role in statistical inference.

(3) Semi-parametric models. Concept of information in semi-parametric models. Asymptotic theory for semi-parametric models.

(4) Likelihood principle and the Bayesian approach to statistical inference, with nuisance parameters situation as the exemplar.

(5) Foundational issues in statistical inference. Belief, Decision and Evidence.

Reference books:

1) Small and McLeish: *Hilbert space methods in probability and statistical inference*
2) Royall: *Statistical evidence*
3) Barnett: *Comparative Statistical Inference*
4) Vandervaart: *Asymptotic statistics*
5) Kass and Vos: *Geometric approach to statistical inference*

STATISTICS 665: ASYMPTOTIC METHODS IN STATISTICAL INFERENCE

We view topics 1 - 5, *including the sub-headings*, as <u>essential</u> material which must be included each time this course is taught.

(1) Probabilistic preliminaries (6 lecture hours)
- Convergence in law and in probability
- Central Limit theory: Edgeworth expansions, Berry-Esseen Theorem, Liapounov's Theorem
- Slutsky's Theorem, delta method

(2) Asymptotic testing and estimation (12 hours)
- Asymptotic level-α tests, power, sample size
- Level robustness
- Confidence intervals, point estimation
- Asymptotic relative efficiency, Pitman closeness

(3) Multivariate extensions (3 hours)
- Multivariate Central Limit Theorem, Cramér-Wold device
- Multivariate delta method

(4) Nonparametric methods (7.5 hours)
- U- and V-statistics
- Statistical functionals, influence functions, limit theory
- Bootstrapping

(5) Likelihood methods (7.5 hours)
- Likelihood estimation
- Consistency and asymptotic normality, efficiency
- Multiparameter ML estimation
- Wald's test, scores test, likelihood ratio test

A suitable textbook covering (most of) this material is *Elements of Large-Sample Theory* by E.L. Lehmann.