Appendix 2: Core PhD Curriculum and key textbooks

Sequence	Description	Background	Core Topics	References
651-4	The primary objective of this series is to introduce statistical methods that are commonly adopted in biomedical research. Both exploratory and confirmatory approaches will be emphasized. Students are expected to master the material by learning through homework and discussion how to analyze biomedical data, interpret and summarize results.		• Philosophy: exploratory versus confirmatory • Some basic concepts in probability • Exploratory data analysis: summarizing and describing (graphically) data • Issues in design ⇒ Sample size and statistical power ⇒ Role of randomization • Formal inferences ⇒ Approaches * P-value and confidence interval * Frequentist, Bayesian, and likelihood paradigms * Multiple hypothesis testing ⇒ Methods * Large sample likelihood inference * Delta method * Sample-reuse method • Inferences for one sample: normal/binomial/Poisson responses • Inferences for two-sample problem: normal • Statistical inferences in epidemiology ⇒ Designs ⇒ Odds ratio and relative risk ⇒ Two by two tables ⇒ Exact versus large sample inference ⇒ Matching • Analysis of count data • Linear regression models ⇒ Least squares versus maximum likelihood ⇒ Parameter interpretation ⇒ Confounding and interaction ⇒ Model checking and selection ⇒ Prediction • Generalized linear models for discrete responses ⇒ Logistic regression models ⇒ Conditional logistic regression for matched designs ⇒ Poisson regression models ⇒ Overdispersion ⇒ Introduction to survival analysis	• Rice JA: *Mathematical Statistics*. Boston: Duxbury Press, 1988. • Rosner B: *Fundamentals of Biostatistics*. Boston: Duxbury Press, 1990. • Dobson A: *An Introduction to Generalized Linear Models*. 2nd ed. Boca Raton, Florida: Chapman and Hall, 2002. • Weisberg S: *Applied Linear Regression*. New York: John Wiley, 1980.

PhD Curriculum
Johns Hopkins Department of Biostatistics

Sequence	Description	Background	Core Topics	References
671-2	This two- term sequence introduces basic probability concepts that are essential to statistical theory to be covered in 673-4. Students are expected to master the basic concepts with elementary proofs of the most important results.	• Set theory (basic concepts)	• Probability space ⇒ σ-field, probability measure, fundamental probability properties • Random variables/vectors and distributions ⇒ General concepts; discrete and continuous random variables/vectors; cumulative distribution function; probability density function; hazard function • Simple Poisson process ⇒ Poisson process and Poisson distribution; relationship between Poisson process and other probability distributions • Moments, probability inequalities, moment generating functions, some covariance formulas • Transformation of random variables/vectors; Jacobian formulas • Markov inequality; Chebyshev inequality; Jensen inequality; associated applications • Independence and conditional distributions • WLLN, SLLN, central limit theorem • Convergence modes: convergence in probability/in L^p/in distribution/with probability one • O_p, o_p; O/o with probability one	• Grimmett G, Stirzaker D: *Probability and Random Processes*. New York: Oxford University Press, 1982. • Roussas GG: *A Course in Mathematical Statistics*. 2nd ed. San Diego: Academic Press, 1997.

PhD Curriculum
Johns Hopkins Department of Biostatistics

Sequence	Description	Background	Core Topics	References
673-4	This two-term sequence builds upon probability concepts introduced in 671-2 to lay out a foundation and to discuss classical theory/methods for drawing statistical inference. Emphasis is on statistical reasoning and theoretical justification behind two main streams in inference: hypothesis testing and estimation (point and interval).	• Probability theory (671-2)	• Statistics for science ⇒ Statistics in the path towards decisions or actions ⇒ Loss functions, likelihood function, and decision strategies ⇒ Relationship between statistics and criteria for decisions • Principles of data reduction ⇒ Sufficient statistics; minimal sufficient statistics ⇒ Ancillary and complete statistics ⇒ Likelihood principle ⇒ Exponential family • Methods of estimation ⇒ Methods of moments ⇒ Methods of least squares ⇒ Maximum likelihood estimation: MLE and conditional MLE ⇒ Likelihood score functions; Fisher information • Evaluation of estimators ⇒ Mean square error ⇒ Unbiased estimators: UMVUE; information inequality ⇒ Unbiased estimating functions: finite sample optimality • Interval estimation ⇒ Finding interval estimators ⇒ Methods of evaluating interval estimators • Asymptotic theory ⇒ Review of convergence modes ⇒ Asymptotic theory for MLE • Hypothesis testing ⇒ Neyman-Pearson lemma ⇒ Uniformly most powerful tests ⇒ Likelihood ratio tests; Wald test	• Casella G, Berger RL: *Statistical Inference*. Pacific Grove, California: Wadsworth, 1990. • Bickel PJ, Doksum KA: *Mathematical Statistics*. Englewood Cliffs, New Jersey: Prentice-Hall, 2001. • Cox DR, Hinkley DV: *Theoretical Statistics.* London: Chapman and Hall, 1974

PhD Curriculum
Johns Hopkins Department of Biostatistics

Sequence	Description	Background	Core Topics	References
751-4	The primary objective of this series is to discuss the theoretical underpinning and applications of regression techniques that are commonly used in medical/public health research.		• Matrix algebra relevant to statistics • General linear models for continuous responses • Hierarchical models/variance components • REML • Introduction to multivariate analysis • Maximum likelihood and least squares • Generalized linear models for discrete responses • Over-dispersion • Conditional logistic regression • Quasi-likelihood and estimating functions • Generalized additive models • Smoothing techniques • Model selection • Classification, clustering and algorithmic statistics • Introduction to time series	• Nelder JA, Wedderburn RWM: Generalized linear models. *Journal of the Royal Statistical Society A* 135(Part 3): 370-384, 1972 • McCullagh P, Nelder JA: *Generalized Linear Models*. 2nd ed. London: Chapman & Hall, 1989. • Ripley BD, Venables WN: *Modern Applied Statistics with S*. New York: Springer Verlag, 2002. • Rao CR: *Linear Statistical Inference and Its Applications.* 2nd ed. New York: John Wiley, 1973. • Graybill F: *Theory and Application of the Linear Model.* North Scituate, Massachusetts: Duxbury Press, 1976.

PhD Curriculum
Johns Hopkins Department of Biostatistics

Sequence	Description	Background	Core Topics	References
771-2	This two-term sequence provides in-depth treatments on both finite and large sample theory for statistical inference. While not exclusive, frequentist interpretations will be emphasized on the development of the theory considered.		• Finite sample theory ⇒ Unbiased estimation, unbiased estimating functions ⇒ Fisher information and its generalization (using Hilbert spaces) ⇒ Cramer-Rao inequality and its generalization (using Hilbert spaces) ⇒ Sufficiency, ancillarity ⇒ Conditional, marginal, and partial likelihood • Large sample theory ⇒ Convergence in distribution and probability for random vectors ⇒ Multivariate delta method, continuous mapping theorem, Cramer-Wold device ⇒ Comparing estimators ⇒ Maximum likelihood – consistency, asymptotic normality and efficiency ⇒ Generalized method of moments/M-estimators – consistency ,asymptotic normality and efficiency ⇒ Regularity and contiguity ⇒ Influence functions ⇒ Hypothesis testing ⇒ Inference in semi-parametric models ⇒ One and two sample U-statistics ⇒ Order statistics, quantiles, and extreme values ⇒ L- and R-statistics	• Gourieroux C, Monfort A: *Statistics and Econometric Models* (vol 1 and 2). Cambridge, England: Cambridge University Press, 1995. • Lehmann EL, Casella G: *Theory of Point Estimation*. 2nd ed. New York: Springer, 1998. • Lehmann EL: *Elements of Large-Sample Theory*. New York: Springer, 1998. • Serfling RJ: *Approximation Theorems of Mathematical Statistics.* New York: John Wiley, 1980. • Van der Vaart AW: *Asymptotic Statistics*. Cambridge, England: Cambridge University Press, 1998.

PhD Curriculum
Johns Hopkins Department of Biostatistics

Sequence	Description	Background	Core Topics	References
773-4	This two-term sequence discusses foundations of statistical inference. Fundamental concepts such as sufficiency, ancillarity and law of likelihood will be closely examined. Approaches for inference including frequentist, Bayesian and likelihood paradigm will be critiqued and contrasted.		• Background: The inference problem • Decision theoretic approach including loss functions, risk functions, admissibility, etc. • Sufficiency, conditioning and ancillarity • The likelihood function and the Law of Likelihood • Criticism of Neyman-Pearson and the repeated sampling approach to inference • Comments on Bayesian approaches to inference • The likelihood approach in more complex problems involving nuisance parameters, random effects, prediction, etc.	• Royall R: *Statistical Evidence: A Likelihood Paradigm*. London: Chapman and Hall, 1997. • Berger JO, Wolpert RL: *The Likelihood Principle*. 2^{nd} ed. Hayward, California: Institute of Mathematical Statistics, 1988. • Birnbaum A: On the foundations of statistical inference. *JASA* 57(298): 269-306, 1962. • Fisher RA: On the mathematical foundation of theoretical statistics. *Phil. Trans. Roy. Soc*. Series A 222: 309-368, 1922.

PhD Curriculum
Johns Hopkins Department of Biostatistics

Sequence	Description	Background	Core Topics	References
Advanced Probability: Mathematical Analysis		• Calculus • Advanced calculus • Fundamental set theory • Linear algebra	• σ-field, Borel σ-field • Concepts on measures; measurable functions • Lebesgue-Stieltjes measures, Lebesgue integral • Fatou's lemma, monotone convergence theorem, dominated convergence theorem • Absolute continuity, convex function, differentiation of Lebesgue integral • Radon-Nikodym theorem • L^p-spaces, Holder and Minkowski inequalities • Metric spaces, uniform continuity, continuous function in metric spaces • Models of convergence: almost everywhere convergence; almost uniform convergence; convergence in measure; convergence in L^p • Basic concepts and properties of Banach and Hilbert spaces, Frechet differentiation	• Rudin W: *Principles of Mathematical Analysis*. New York: McGraw Hill, 1976. • Royden HL: *Real Analysis*. Upper Saddle River, New Jersey: Macmillan, 1988.

PhD Curriculum
Johns Hopkins Department of Biostatistics

Sequence	Description	Background	Core Topics	References
Advanced Probability: Probability Theory	This full-year course presents the theory of probability from a measure theoretic perspective. In contrast to 671-2, students are expected to master the concepts through engaging in formal proofs.	• Calculus • Advanced calculus • Fundamental set theory • Linear algebra	• Probability space ⇒ σ-field ⇒ Probability measure ⇒ Fundamental probability properties • Random variables/vectors and distributions ⇒ General concepts ⇒ Discrete and continuous random variables/vectors ⇒ Cumulative distribution function ⇒ Probability density function ⇒ Hazard function • Poisson process ⇒ Poisson process and Poisson distributions ⇒ Relationship between Poisson process and other probability distributions • Moments, probability inequalities, moment generating functions, some covariance formulas • Characteristic functions • Transformation of random variables/vectors; Jacobian formulas • Markov inequality; Chebyshev inequality; Jensen inequality; associated applications • Independence, conditional distributions, conditional expectation • WLLN, SLLN, central limit theorem, law of iterated logarithm • Convergence modes: convergence in probability/in L^p/in distribution/with probability one, continuous mapping theorem, delta method • O_p, o_p; O/o with probability one • Stochastic processes and associated convergence results, functional delta method, brownian motion • Martingale theory, martingale limit theorems	• Billingsley P: *Probability and Measure*. New York: John Wiley, 1995. • Chung K: *A Course in Probability Theory*. 3rd ed. San Diego: Academic Press, 2001.