

# Foundations and Frontiers of Statistical Science

## From Probability to Learning, Complex Systems, and Philosophy

Shen Han

December 2, 2025



# Preface



# Contents

<b>I</b>	<b>Probability and Classical Statistical Inference</b>	<b>1</b>
1	What Is Statistical Science?	3
2	Events, Probability Spaces, and Randomness	5
3	Random Variables and Expectations	7
4	Sequences of Random Variables and Limit Theorems	9
5	Classical Probability Distributions	11
6	Joint, Marginal, and Conditional Distributions	13
7	From Populations to Samples: Data as Random Objects	15
8	Principles of Point Estimation	17
9	Constructing Estimators: Method of Moments and Maximum Likelihood	19
10	Confidence Intervals and Uncertainty Quantification	21
11	Hypothesis Tests, $p$ -Values, and Error Probabilities	23
12	The Neyman–Pearson Paradigm and Likelihood Ratio Tests	25
13	Classical Parametric Tests	27
14	Sufficiency, Exponential Families, and Classical Optimality	29
15	An Introduction to Decision-Theoretic Statistics	31
16	Linear Models as an Application of Classical Inference	33
<b>II</b>	<b>Bayesian, Computational, and High-Dimensional Statistics</b>	<b>35</b>
17	Bayesian Statistical Inference: Priors, Likelihoods, and Posteriors	37
18	Conjugate Priors and Exponential Families	39

19 Bayesian Decision Theory and Loss Functions	41
20 Hierarchical and Multilevel Models	43
21 Empirical Bayes and Shrinkage Estimation	45
22 Bayesian Model Checking and Model Comparison	47
23 Nonparametric Bayesian Methods	49
24 Monte Carlo Integration and Importance Sampling	51
25 The Bootstrap and Other Resampling Methods	53
26 Markov Chain Monte Carlo: Basic Algorithms	55
27 Advanced MCMC: Hamiltonian and Langevin Methods	57
28 Sequential Monte Carlo and Particle Methods	59
29 Variational Inference and Optimization-Based Approximation	61
30 Probabilistic Programming and Automatic Inference	63
31 High-Dimensional Data and the Curse of Dimensionality	65
32 Penalized Estimation: Ridge, Lasso, and Elastic Net	67
33 Sparse Models and Feature Selection	69
34 Dimension Reduction: PCA, Factor Models, and ICA	71
35 Covariance Estimation and Graphical Models in High Dimensions	73
36 Multiple Testing and False Discovery Rate Control	75
37 Supervised Learning: Risk, Empirical Risk, and Generalization	77
38 Kernel Methods and Support Vector Machines	79
39 Ensemble Methods: Bagging, Random Forests, and Boosting	81
40 Neural Networks and Deep Learning: A Statistical Perspective	83
41 Model Assessment, Cross-Validation, and Hyperparameter Tuning	85
42 Uncertainty Quantification in Predictive Modeling	87
 III Graphical Models, Causal Inference, and Complex Systems	 89
43 Probabilistic Graphical Models: Directed and Undirected	91

44 Latent Variable Models and the EM Algorithm	93
45 Foundations of Causal Inference	95
46 Causal Graphs and Structural Equation Models	97
47 Experiments, Observational Studies, and Identification Strategies	99
48 Causal Discovery and Causal Machine Learning	101
49 Sparse Model Discovery and Equation Reconstruction	103
50 Simulation-Based and Likelihood-Free Inference	105
51 Probabilistic Numerics: A Bayesian View on Computation	107
52 Topological Data Analysis and the Shape of Data	109
53 Information Geometry and Statistical Manifolds	111
54 Statistical Modeling of Complex Networks and Structures	113
55 Hierarchical Worlds and Structural Causal Models	115
<b>IV Modern AI and Statistical Frontiers</b>	<b>117</b>
56 Conformal Prediction and Distribution-Free Uncertainty Quantification	119
57 Distribution Shift, Robustness, and Trustworthy AI	121
58 Causal Representation Learning and Causal AI	123
59 Calibration, Selective Prediction, and Human–Model Collaboration	125
60 Generative Models: Diffusions, Flows, and Score-Based Methods	127
61 Federated Learning and Differential Privacy	129
62 Simulation-Based Inference in Scientific Applications	131
<b>V Philosophy and Foundations of Statistical Worlds</b>	<b>133</b>
63 What Is Probability? Interpretations and Debates	135
64 What Does Statistical Inference Claim? Frequency, Bayesian, and Likelihoodist Views	137
65 Induction, Learning, and Information	139
66 What Is a Statistical Model? Idealization, Approximation, and Tools	141

67 Model Complexity, Occam's Razor, and Information Criteria	143
68 Explanation vs Prediction in Statistical Science	145
69 The Concept of Causation: From Regularities to Structural Models	147
70 Hierarchical Worlds, Coarse-Graining, and Emergence	149
71 Reconstructing Structure: From Data to Equations, Topology, and Geometry	151
72 Reproducibility, Misuse of Significance, and Scientific Practice	153
73 Statistics in the Age of AI: Protocols, Norms, and Human Judgment	155
74 A Personal Worldview: Banach Worlds, Infinity-Categories, and Statistical Science	157



**Part I**

**Probability and Classical Statistical  
Inference**



# Chapter 1

## What Is Statistical Science?



## Chapter 2

# Events, Probability Spaces, and Randomness



## Chapter 3

# Random Variables and Expectations





## Chapter 4

# Sequences of Random Variables and Limit Theorems



## Chapter 5

# Classical Probability Distributions



## Chapter 6

# Joint, Marginal, and Conditional Distributions



## Chapter 7

# From Populations to Samples: Data as Random Objects





## Chapter 8

# Principles of Point Estimation



## Chapter 9

# Constructing Estimators: Method of Moments and Maximum Likelihood



## Chapter 10

# Confidence Intervals and Uncertainty Quantification



## Chapter 11

# Hypothesis Tests, $p$ -Values, and Error Probabilities





## Chapter 12

# The Neyman–Pearson Paradigm and Likelihood Ratio Tests



# Chapter 13

## Classical Parametric Tests



## Chapter 14

# Sufficiency, Exponential Families, and Classical Optimality



## Chapter 15

# An Introduction to Decision-Theoretic Statistics





## Chapter 16

# Linear Models as an Application of Classical Inference



## Part II

# Bayesian, Computational, and High-Dimensional Statistics



## Chapter 17

# Bayesian Statistical Inference: Priors, Likelihoods, and Posteriors



## Chapter 18

# Conjugate Priors and Exponential Families





## Chapter 19

# Bayesian Decision Theory and Loss Functions



## **Chapter 20**

# **Hierarchical and Multilevel Models**



## Chapter 21

# Empirical Bayes and Shrinkage Estimation



## Chapter 22

# Bayesian Model Checking and Model Comparison





## Chapter 23

# Nonparametric Bayesian Methods



## Chapter 24

# Monte Carlo Integration and Importance Sampling



## Chapter 25

# The Bootstrap and Other Resampling Methods



## Chapter 26

# Markov Chain Monte Carlo: Basic Algorithms





## Chapter 27

# Advanced MCMC: Hamiltonian and Langevin Methods



## Chapter 28

# Sequential Monte Carlo and Particle Methods



## Chapter 29

# Variational Inference and Optimization-Based Approximation



## Chapter 30

# Probabilistic Programming and Automatic Inference





## Chapter 31

# High-Dimensional Data and the Curse of Dimensionality



## Chapter 32

### Penalized Estimation: Ridge, Lasso, and Elastic Net



## Chapter 33

# Sparse Models and Feature Selection



## Chapter 34

# Dimension Reduction: PCA, Factor Models, and ICA





## Chapter 35

# Covariance Estimation and Graphical Models in High Dimensions



## Chapter 36

# Multiple Testing and False Discovery Rate Control



## Chapter 37

# Supervised Learning: Risk, Empirical Risk, and Generalization



## Chapter 38

# Kernel Methods and Support Vector Machines





## Chapter 39

# Ensemble Methods: Bagging, Random Forests, and Boosting



## Chapter 40

# Neural Networks and Deep Learning: A Statistical Perspective



## Chapter 41

# Model Assessment, Cross-Validation, and Hyperparameter Tuning



## Chapter 42

# Uncertainty Quantification in Predictive Modeling





## Part III

# Graphical Models, Causal Inference, and Complex Systems



## Chapter 43

# Probabilistic Graphical Models: Directed and Undirected



## Chapter 44

# Latent Variable Models and the EM Algorithm



## Chapter 45

# Foundations of Causal Inference





## Chapter 46

# Causal Graphs and Structural Equation Models



## Chapter 47

# Experiments, Observational Studies, and Identification Strategies



## Chapter 48

# Causal Discovery and Causal Machine Learning



## Chapter 49

# Sparse Model Discovery and Equation Reconstruction





## Chapter 50

### Simulation-Based and Likelihood-Free Inference



## Chapter 51

# Probabilistic Numerics: A Bayesian View on Computation



## Chapter 52

# Topological Data Analysis and the Shape of Data



## Chapter 53

# Information Geometry and Statistical Manifolds





## Chapter 54

# Statistical Modeling of Complex Networks and Structures



## Chapter 55

# Hierarchical Worlds and Structural Causal Models



## **Part IV**

# **Modern AI and Statistical Frontiers**



## Chapter 56

# Conformal Prediction and Distribution-Free Uncertainty Quantification





## Chapter 57

# Distribution Shift, Robustness, and Trustworthy AI



## Chapter 58

# Causal Representation Learning and Causal AI



## Chapter 59

# Calibration, Selective Prediction, and Human–Model Collaboration



## Chapter 60

# Generative Models: Diffusions, Flows, and Score-Based Methods





## Chapter 61

# Federated Learning and Differential Privacy



## Chapter 62

# Simulation-Based Inference in Scientific Applications



## Part V

# Philosophy and Foundations of Statistical Worlds



## Chapter 63

### What Is Probability? Interpretations and Debates





## Chapter 64

### What Does Statistical Inference Claim? Frequency, Bayesian, and Likelihoodist Views



## Chapter 65

# Induction, Learning, and Information



## Chapter 66

### What Is a Statistical Model? Idealization, Approximation, and Tools



## Chapter 67

# Model Complexity, Occam's Razor, and Information Criteria





## Chapter 68

# Explanation vs Prediction in Statistical Science



## Chapter 69

# The Concept of Causation: From Regularities to Structural Models



## Chapter 70

# Hierarchical Worlds, Coarse-Graining, and Emergence



## Chapter 71

# Reconstructing Structure: From Data to Equations, Topology, and Geometry





## Chapter 72

# Reproducibility, Misuse of Significance, and Scientific Practice



## Chapter 73

# Statistics in the Age of AI: Protocols, Norms, and Human Judgment



## Chapter 74

### A Personal Worldview: Banach Worlds, Infinity-Categories, and Statistical Science



## Notes and Further Reading