Probability and Statistics

Guide to AI and Data Courses · Part V: Mathematical Foundations

Probability and Bayes | Distributions | Estimation | Hypothesis testing | Where in AI | Neighboring courses

Sercan KÜLCÜ · Giresun University · Computer Engineering

© Sercan Külcü. All rights reserved. · Interactive HTML version of these slides:
https://sercankulcu.github.io/files/ai_courses/slides/Chapter_32_Probability_and_Statistics.html

INTRODUCTION

What Question Does It Ask?

> **How is uncertainty measured, and how do we draw reliable conclusions from limited data?**

Probability is the language for measuring uncertainty; statistics uses it to draw general conclusions from **limited data**. Every model in AI and data courses works under uncertainty, and only this course can say how far a prediction can be trusted.

At a glance

APPROACH probability models and inference	**INPUT** sample data	**OUTPUT** estimate, confidence interval, test result
SUCCESS MEASURE unbiasedness, power, significance	**MATH LOAD** the theory itself	**DATA NEEDS** a sample

INTRODUCTION

A Short History

1654
Pascal and Fermat exchange letters on dividing the stakes of an unfinished game of chance.

1763
Bayes's essay is published after his death: probability updated by evidence.

1908
Gosset, writing as "Student", publishes the t distribution for small samples.

1925
Fisher's Statistical Methods for Research Workers popularizes significance testing.

1933
Kolmogorov builds probability on axioms; probability theory rests on measure theory.

1979
Efron proposes the bootstrap: measuring uncertainty by resampling.

THE COURSE

Core Concepts

Conditional probability
How a probability changes given information.

Bayes' rule
Turning a prior belief into a posterior with evidence.

Random variable
Expected value, variance and distribution.

Normal distribution
Central limit theorem: averages tend to normal.

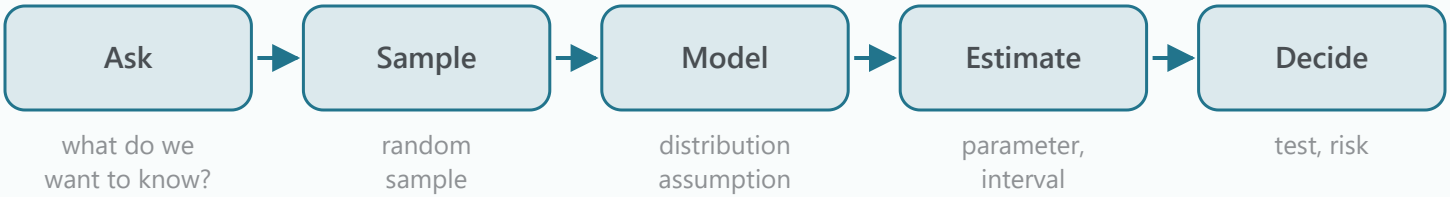
Maximum likelihood
Choosing the parameter that makes the data most likely.

Confidence interval
Reporting an estimate's uncertainty as a range.

Hypothesis testing
Is the observed difference chance or real?

Regression
Explaining one variable by others, and testing it.

A Typical Workflow



Example: An AI team asks whether a new recommendation model raises clicks. Half of the users see the new model at random (an A/B test), click rates are estimated, a confidence interval for the difference is computed, and if chance cannot explain the difference the model ships.

THE COURSE

What Does the Course Cover?

1. Counting, probability axioms, conditional probability
2. Bayes' rule and independence
3. Discrete distributions: Bernoulli, binomial, Poisson
4. Continuous distributions: uniform, exponential, normal
5. Expectation, variance and covariance
6. Joint and conditional distributions
7. Law of large numbers and central limit theorem
8. Point estimation: method of moments, maximum likelihood
9. Confidence intervals and the bootstrap
10. Hypothesis testing: t-test, chi-square, p-values
11. Linear regression and inference
12. Bayesian inference: priors and posteriors

Skills you will gain

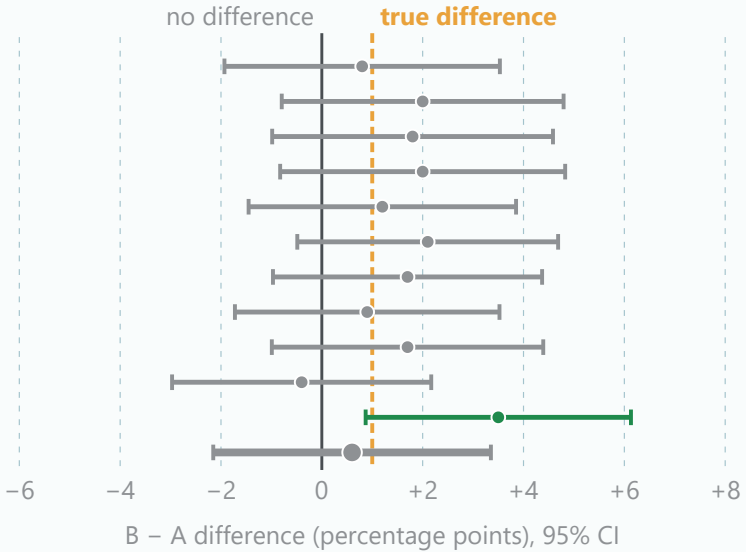
running tests with scipy.stats · regression in R or statsmodels · computing confidence intervals · checking probabilities by simulation · designing an A/B test

A typical semester; the order and coverage vary between universities.

TRY IT

Does the New Recommender Work?

Does the new recommendation model raise the purchase rate? Half of the visitors see the old model (A), half the new one (B). Change the sample size and the true difference, and repeat the experiment.



Visitors per group n = **1000**

True difference (B − A) = **+1.0** points

New experiment | Repeat 20 times

Last experiment: A 10.8%, B 11.4%
Difference +0.6 points, CI [−2.2, +3.4]
p = **0.669** → not significant

With this setting **1** of 12 experiments were significant (8%). Green: the interval excludes zero.

With a small sample the intervals are wide: a real improvement comes out “not significant” in most experiments. Computing the required n before starting the test keeps you from throwing away a good model.

CONTEXT

Prerequisites

Courses worth knowing before this one, and why they are needed.

Course	Why it is needed
Calculus	integrals and derivatives: continuous distributions and maximum likelihood
Discrete Mathematics	sets and counting (permutations, combinations)

CONTEXT

Are You Ready for This Course?

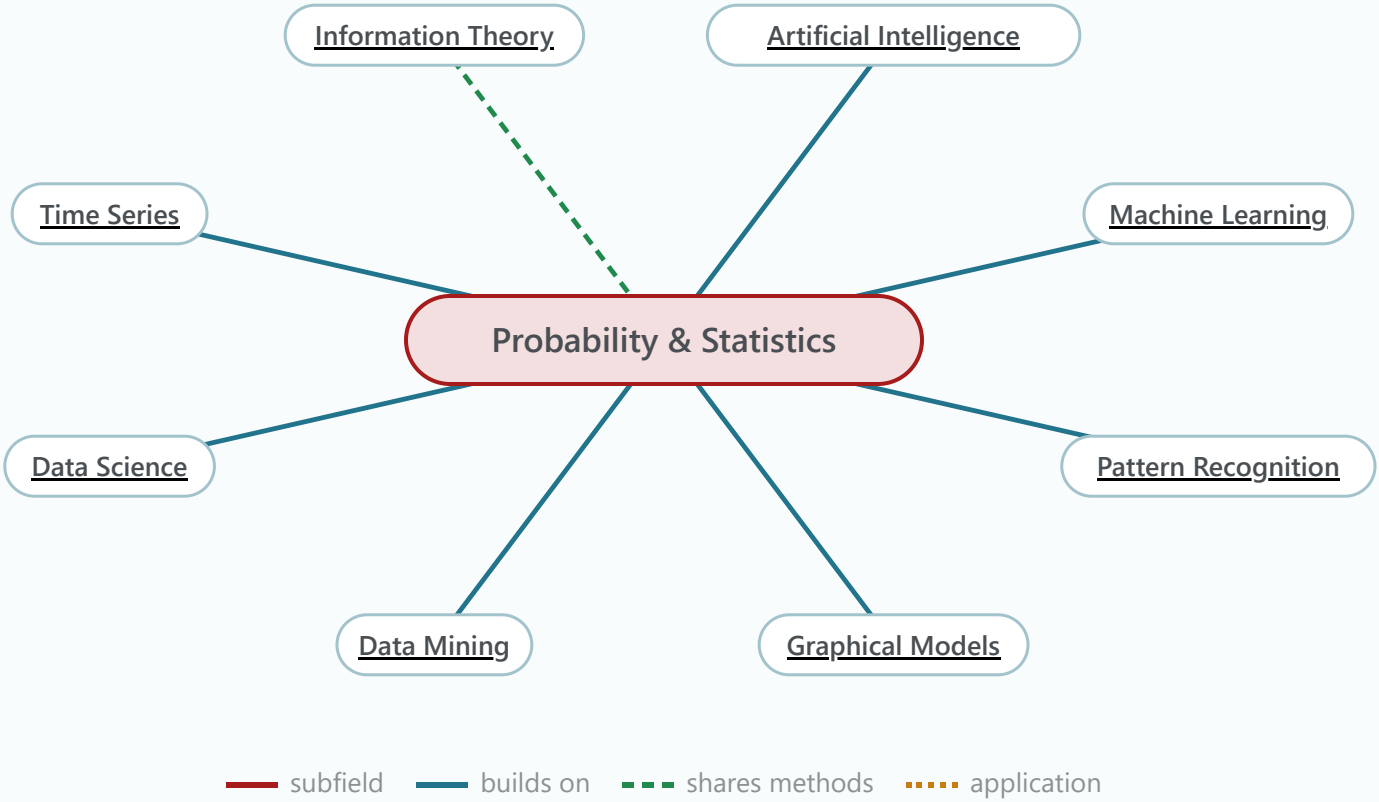
Ask yourself; for every "no", the right-hand column shows where to start.

Ask yourself	If not, start with
Can you integrate a function to find the area under a curve?	Definite integrals from calculus
Can you find the maximum of a function by taking its derivative?	Derivatives and extreme values from calculus
Can you count selections with permutations and combinations?	Counting from discrete mathematics
Can you compute the mean of a list in Python and plot its histogram?	An intro programming course, then a short NumPy tutorial

If you can say "yes" to most of these, you are ready; one or two gaps can be closed during the semester.

CONTEXT

Neighbouring Courses



Overlapping Courses

Relation	Course	Shared ground
builds on	Artificial Intelligence builds on this course	reasoning under uncertainty, Bayes rule
builds on	Machine Learning builds on this course	generalization error, bias-variance, probabilistic models
builds on	Pattern Recognition builds on this course	Bayesian decision theory, density estimation
builds on	Probabilistic Graphical Models builds on this course	conditional independence, joint distribution
builds on	Data Mining builds on this course	support, confidence, significance
builds on	Data Science builds on this course	hypothesis testing, regression, visualization
builds on	Time Series Analysis and Forecasting builds on this course	stationarity, autocorrelation, ARIMA
shares methods	Shares methods with Information Theory	entropy, mutual information

Where It Differs (1/2)

Criterion	Probability & Statistics	Machine Learning	Data Science
Core question	How is uncertainty measured, and how do we draw reliable conclusions from limited data?	How does a program generalize from examples?	How do we ask the right questions of data and turn the answers into decisions?
Approach	probability models and inference	statistical learning	question, data, analysis, story
Input	sample data	labeled or unlabeled examples	raw, messy data
Output	estimate, confidence interval, test result	a model that makes predictions	report, visualization, model
Success measure	unbiasedness, power, significance	error on unseen data	impact on decisions, accuracy
Math load	the theory itself	high: linear algebra, probability, optimization	medium: statistics
Data needs	a sample	moderate to large	moderate to large

- **Machine Learning:** ML puts predictive performance first; statistics asks about a model's meaning, assumptions and uncertainty.
- **Data Science:** Data science covers the whole process (data, analysis, communication); this course gives the theory of inference behind it.

Where It Differs (2/2)

Criterion	Probability & Statistics	Graphical Models	Information Theory
Core question	How is uncertainty measured, and how do we draw reliable conclusions from limited data?	How do we model and query many interdependent, uncertain variables?	How is information measured, and how is it compressed and transmitted with the least loss?
Approach	probability models and inference	probability and graphs combined	measuring information with probability
Input	sample data	variables, dependencies, data	source and channel model
Output	estimate, confidence interval, test result	conditional probability, best explanation	bounds, codes, compression rate
Success measure	unbiasedness, power, significance	likelihood, inference accuracy	bits, channel capacity
Math load	the theory itself	high: probability	high: probability
Data needs	a sample	moderate	none

- **Graphical Models:** Probabilistic graphical models structure multivariate distributions with graphs; this course starts with one or a few variables.
- **Information Theory:** Information theory measures uncertainty with entropy; this course is built on estimation and decisions with probability.

PRACTICE

Applications

A/B testing
Whether a new model is really better.

Naive Bayes
Text and spam classification.

Uncertainty
Confidence intervals and calibration of predictions.

Model comparison
Is the gap between two models significant?

Generative models
Learning a data distribution and sampling it.

Medical tests
Sensitivity, specificity and base rate.

PRACTICE

Risks and Ethics

p-hacking
Trying many analysis choices and reporting only the "significant" one makes chance look like a finding.

Multiple testing
At the 0.05 level, a hundred comparisons yield about five false positives; without a correction, spurious findings appear.

Simpson's paradox
A trend can reverse when groups are combined; the 1973 Berkeley graduate admissions data is the well-known example.

Selection bias
A non-random sample misleads however large it is, as the 1936 Literary Digest poll showed.

In depth: Explainable AI and AI Ethics.

PRACTICE

Tools and Datasets

Tools and libraries

R | SciPy (scipy.stats) | statsmodels | pandas | PyMC | Stan

Classic example problems

Monty Hall | Birthday paradox | Medical test and base rate | Galton's height data | A/B test | Gambler's fallacy

PRACTICE

Common Misconceptions

Common belief	In fact
If p = 0.03, the hypothesis is 97% likely to be true.	The p-value is the probability of data this extreme if the null hypothesis holds; it is not the probability of the hypothesis.
Correlation implies causation.	A common cause or chance also produces correlation; causation needs an experiment.
If a test is 99% accurate, a positive means 99% disease.	If the disease is rare, most positives may be false alarms; Bayes' rule brings in the base rate.
With enough data, statistics is unnecessary.	In big data even trivial differences come out "significant"; effect size and bias still matter.

NEXT STEPS

After This Course

Courses that build on or extend this one

Artificial Intelligence | Machine Learning | Pattern Recognition | Probabilistic Graphical Models | Data Mining | Data Science | Time Series Analysis and Forecasting

Related careers

Statistician | Data scientist | Machine learning researcher | Biostatistician | Actuary

After this course, machine learning, probabilistic graphical models, pattern recognition and information theory become easier to approach. Bayesian statistics and causal inference are two natural next steps that treat uncertainty and cause and effect in more depth.

Further Reading

- Wasserman, *All of Statistics*, Springer, 2004.
- DeGroot and Schervish, *Probability and Statistics*, 4th ed., Pearson, 2012.
- Blitzstein and Hwang, *Introduction to Probability*, 2nd ed., CRC Press, 2019.
- Gelman, Carlin, Stern, Dunson, Vehtari and Rubin, *Bayesian Data Analysis*, 3rd ed., CRC Press, 2013.

The End

Chapter 32: Probability and Statistics

← Previous chapter · Map of the field · Next chapter →