Human and Machine Learning

Week 1 — Introduction and Basic Bayes

Spring 2026 · Apr 17 · Prof. Joseph Austerweil

Chiba Tech
School of Design & Science

Outline

Welcome	Block 1 · 5 min
Why this course exists	Block 2 · 25 min
Basic Bayes	Block 3 · 25 min
Bayes in action	Block 4 · 35 min
GenJAX, admin, homework	Block 5 · 20 min

Welcome

Today

- Who you are, what you're hoping to get out of the course
- Why this course exists — inductive inference under uncertainty
- Basic probability and Bayes' rule
- Two worked examples: sick friend, Kahneman–Tversky cab problem
- What is GenJAX, and how will we use it?
- Logistics and homework

Why This Course Exists

Deduction vs. induction

Deduction

$$1 + 4 = ?$$

Preserves truth. One answer.

Induction

$$? + ? = 5$$

Underdetermined. Many plausible answers.

The problems people excel at — where we outperform machines — are inductive.

They feel easy. They are not.

Three inductive problems

(… that the mind solves constantly)

1. Perception

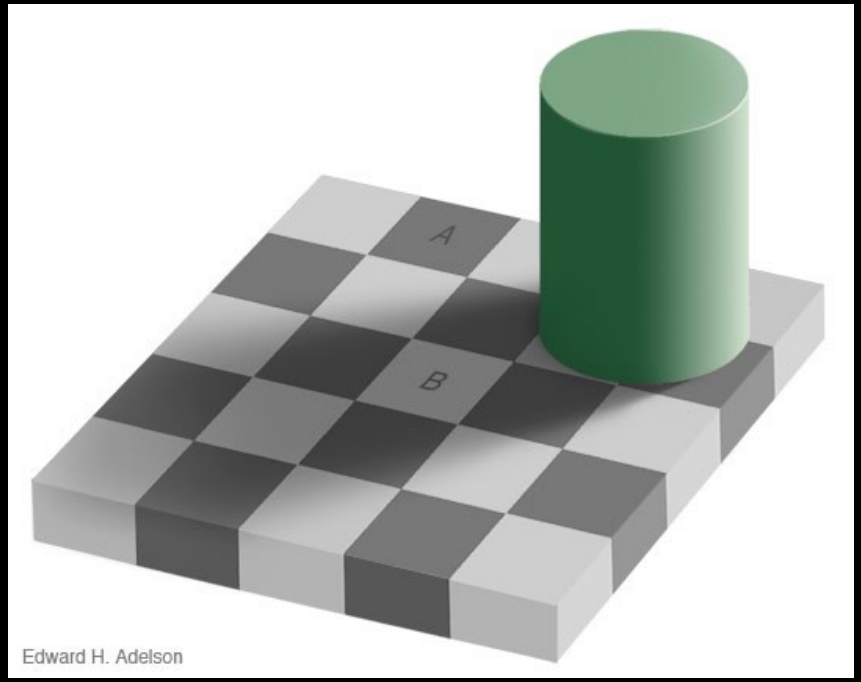
Which square is darker — A or B?

The visual system solves

```
square color + shadow = intensity
```

for square color, using priors about how retinal images are generated.

You are literally looking at an inductive inference right now.



2. Mental states

Data: other people's behavior (motion, speech, gaze).

Hypotheses: their goals and beliefs.

Heider & Simmel (1944) — animated geometric shapes.

Watch ~90 seconds, then describe what happened:

https://www.youtube.com/watch?v=VTNmLt7QX8E

2. Mental states — debrief

Everyone spontaneously narrates the shapes as agents with intentions, grudges, fears.

Those intentions were never in the video. Shapes moved on a plane.

- Data: 2D motion trajectories.
- Hidden: goals, beliefs, relationships.

Your mind solved the inverse problem — behavior → mental states — without you noticing.

3. Word learning

You're in the Australian outback. Someone points at a hopping animal and says "jumbuck."

What does jumbuck mean?

- the animal itself?
- undetached kangaroo-parts?
- kangaroo temporal-stages?
- any mammal?
- dinner?

All consistent with the data. Children pick one — fast.

They bring strong prior expectations about what words mean.

Check-in

What's the common structure across these three problems?

- Hidden variable we care about: h
- Noisy / sparse data we observe: d
- Prior knowledge we bring to bear

$$P(h \mid d) \propto P(d \mid h) \cdot P(h)$$

(We'll unpack every symbol in the next hour.)

Basic Bayes

Chibany is hungry

Chibany, mascot of Chiba Tech, gets two bento offerings per day:

one for lunch, one for dinner. Each is either Hamburger (H) or Tonkatsu (T).

Outcome space: $\Omega = \{HH, HT, TH, TT\}$ $|\Omega| = 4$

- Event: any subset of Ω

  e.g., "at least one tonkatsu" = {HT, TH, TT}

- Event space: the set of all events (powerset of Ω). $2^4 = 16$ events.

(This example is from the course textbook, Ch 2.)

Probability is counting

Chibany wants tonkatsu. What's the probability of at least one T?

$$P(A) = |A| / |\Omega|$$

$A = \{HT, TH, TT\}, \quad |A| = 3, \quad |\Omega| = 4$

$$P(\text{at least one T}) = 3/4 = 0.75$$

Count the outcomes in the event. Divide by total outcomes. That's it.

Random variables

Chibany wants to know: how much tonkatsu?

A random variable is a function from outcomes to numbers.

$$f : \Omega \to \{0, 1, 2\} \quad \text{"count the tonkatsus"}$$

$$f(HH) = 0, \quad f(HT) = 1, \quad f(TH) = 1, \quad f(TT) = 2$$

$$P(f = 1) = |\{HT, TH\}| \,/\, |\Omega| = 2/4 = 1/2$$

The word random is about the input, not the mapping.

Joint probability

What's the chance of hamburger for lunch and tonkatsu for dinner?

$$\mathrm{P}(\mathrm{lunch} = \mathrm{H},\ \mathrm{dinner} = \mathrm{T}) \ = \ \mathrm{P}(\{\mathrm{HT}\}) \ = \ 1/4$$

Joint = "both of these at once."

Conditional probability — Tanaka-san visits

Tanaka-san tells Chibany: "Tomorrow you'll get at least one tonkatsu!"

Chibany asks: what's the probability dinner is also tonkatsu?

Tanaka-san says 1/2. Chibany disagrees.

Conditioning = restricting the outcome space.

Cross out HH. New universe: {HT, TH, TT} (3 outcomes).

Of those, 2 have dinner = T.

$$P(\text{dinner} = T \mid \geq 1\ T) = 2/3$$

Dependence vs. independence

Tanaka-san protests: "But what if I just tell you the first meal is T?"

$$P(\text{dinner} = T \mid \text{lunch} = T) = |\{TT\}| \,/\, |\{TH, TT\}| = 1/2$$

That's the same as the unconditional P(dinner = T) = 1/2. No change!

Dependent: $P(A \mid B) \neq P(A)$ — learning B changes your belief about A.

Independent: $P(A \mid B) = P(A)$ — learning B tells you nothing new.

"At least one T" and "lunch = T" give different information about dinner!

Marginalization

Chibany's dinner student is sick. Only one meal today: $\Omega_1 = \{H, T\}$.

P(lunch = T) = 1/2.

Next day, both meals are back. What's P(lunch = T) now?

$$P(\text{lunch} = T) = P(TH) + P(TT) = 1/4 + 1/4 = 1/2 \quad \text{Same!}$$

$$P(X) = \Sigma_c \; P(X, C = c)$$

Sum the joint over the values of whatever you want to get rid of.

The ratio definition → Bayes' rule

$$P(A \mid B) = P(A \cap B) \;/\; P(B)$$

Chibany checks: $P(\text{dinner}=T \mid \geq 1T) = (2/4) \;/\; (3/4) = 2/3$ ✓

From the product rule, both directions:

$$P(A \mid B) \cdot P(B) = P(A, B) = P(B \mid A) \cdot P(A)$$

Divide:

$$P(A \mid B) = P(B \mid A) \cdot P(A) \;/\; P(B)$$

That's it. The rest of the course is understanding what this means.

Check-in

For the problem I'm about to give you — the cab problem —

- What's d (the data, the observation)?
- What's h (the hidden variable)?

Bayes in Action

Sick friend

Your 50-year-old friend has been a chain smoker since 18.

There's a nasty cold going around.

You go over to their house. You hear them cough.

Three possibilities:

Cold · **Stomach virus** · **Lung cancer**

How likely is each?

First pass — qualitatively

	Prior	P(cough | h)	Prior × Likelihood
Cold	**high**	**high**	**high**
Stomach virus	medium	low	low
Lung cancer	medium	high	medium

Before any math: cold is most likely.

With numbers

	P(h)	P(cough | h)	Numerator	Posterior
Cold	**0.45**	**0.9**	**0.405**	**0.750**
Stomach virus	0.10	0.45	0.045	0.083
Lung cancer	0.10	0.9	0.090	0.167

Sum of numerators: 0.405 + 0.045 + 0.090 = 0.540.

Divide each numerator by the sum to normalize.

The pedagogical punchline

$$P(\text{cold} \mid \text{cough}) = (0.9 \cdot 0.45) / (0.9 \cdot 0.45 + 0.45 \cdot 0.10 + 0.9 \cdot 0.10)$$
$$= 0.405 / 0.540 = 0.75$$

Even with a chain-smoking friend, P(lung cancer | cough) ≈ 17% — not dominant.

The cold wins because:

1. Priors matter. Cold has a ~4.5× higher prior than cancer.
2. Likelihoods matter. Cough is more likely from a cold (0.9) than a stomach virus (0.45).
3. **Neither alone is enough. You need the whole rule.**

The Kahneman–Tversky cab problem

Setup

- A city has two cab companies: Blue and Green.
- 85% of cabs are Green. 15% are Blue.
- At night, a hit-and-run happens. A witness says the cab was Blue.
- The witness correctly identifies cab colors at night 80% of the time.

What is the probability the cab was actually Blue?

(Commit to a number before we compute.)

Pass 2 — the area diagram

Imagine 100 cabs.

- Blue cabs: 15. Witness correctly says "blue" on 80% of them → 12.
- Green cabs: 85. Witness incorrectly says "blue" on 20% of them → 17.

Total "blue" reports: 12 + 17 = 29.

$$P(\text{Blue} \mid \text{witness says Blue}) = 12/29 \approx 0.41$$

Below 50%. The witness is more likely wrong than right, despite being 80% accurate.

Pass 3 — formal Bayes

Hypotheses h ∈ {B, G}. Observation o = "witness says Blue."

$$P(B \mid o) = P(o \mid B) \cdot P(B) \;/\; [\, P(o \mid B) \cdot P(B) + P(o \mid G) \cdot P(G) \,]$$

$$= (0.80 \cdot 0.15) \;/\; (0.80 \cdot 0.15 + 0.20 \cdot 0.85)$$

$$= 0.12 \;/\; (0.12 + 0.17) = 0.12 \;/\; 0.29 \approx 0.41$$

Same number as the area diagram. The equation and the counting are the same calculation.

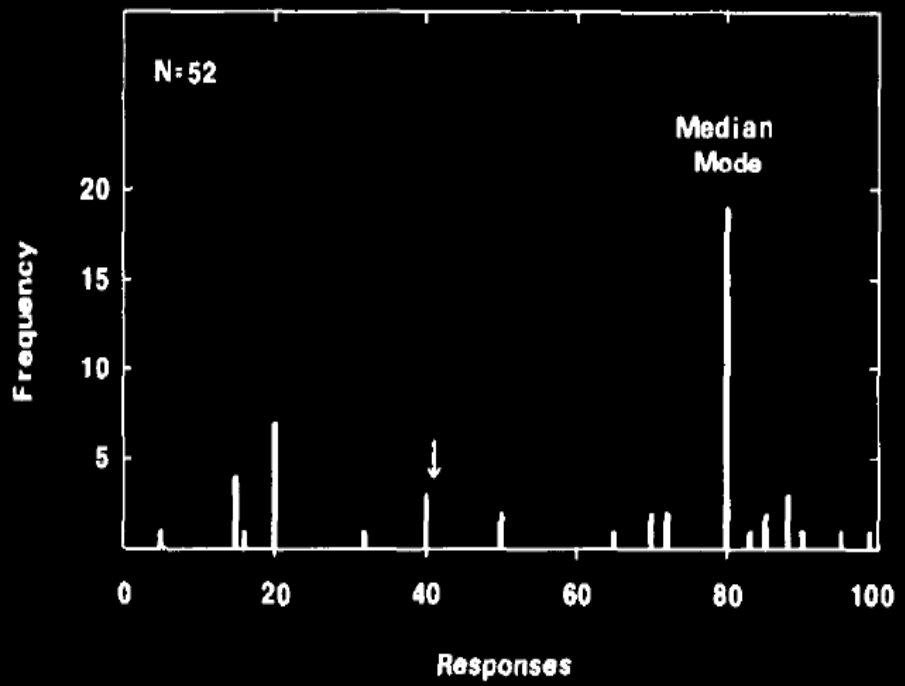
Base-rate neglect

Most people report something close to 80%.

They anchor on the likelihood and ignore the prior.

Kahneman & Tversky (1972); Bar-Hillel (1980).

Medical doctors fail this too, even for diagnoses (Casscells, Schoenberger, Grayboys, 1978).



Heuristics and biases

Heuristic. A rule-of-thumb. Usually works. Cheap to apply.

Bias. Systematic error introduced when a heuristic is misapplied.

Kahneman & Tversky argued that the mind is full of heuristics — good enough for most everyday inference but wrong in predictable ways when base rates are extreme, evidence is unfamiliar, or causal structure is unusual.

This is a running theme for the semester.

Synthesis — the 5-step recipe

This is the course's method.

1. Formalize the problem people face.
2. Formalize the knowledge they bring to it.
3. Apply Bayes' rule — compute what an ideal learner would infer.
4. Characterize the ideal learner's behavior.
5. Identify what knowledge and constraints must be assumed for model and human behavior to match.

Every week we pick a different cognitive domain and run this loop.

GenJAX, Admin, Homework

What is GenJAX?

Bayes' rule is easy to write. Enumerating hypothesis spaces by hand is not.

For continuous distributions you can't enumerate at all.

GenJAX is a probabilistic programming language.

You write the generative process as code. The machine does the enumeration and counting.

GenJAX by example

```
import genjax
from genjax import gen, flip
@gen
def chibany_day():
    lunch  = flip(0.5) @ "lunch"
    dinner = flip(0.5) @ "dinner"
    return (lunch, dinner)
```

- @gen — this is a generative model, not a regular function.
- flip(0.5) — a Bernoulli random variable.
- @ "lunch" — name the random choice so we can condition on it later.

This function IS the outcome space Ω from Block 3. Every call samples a point.

Preemptive callouts

- Use `flip(p)`, not `bernoulli(p)`. `bernoulli` takes a logit, not a probability.
- The `@` operator here is not matrix multiplication. GenJAX overloads it.
- Output displays as `Array(0, dtype=int32)`. Treat these as 0s and 1s.
- Everything runs in Google Colab. No local installation.

Admin — grading

Final project — 50%
proposal 5% · talk 7.5% · paper 37.5%

Programming assignments (4) — 30%
Clusters 7.5% · Gen 7.5% · MC 10.5% · RL 4.5%

Weekly written reflections — 15%
~200 words · 8 of 13 · pass / fail

Participation — 5%

Quizzes — 0%
self-check only

All four assignments are completed in GenJAX.

Admin — the rest

- Textbook: A Narrative Introduction to Probability
  https://josephausterweil.github.io/probintro/
- Tooling: Google Colab, GenJAX. No local setup.
- Office hours: TBD — I'll poll.
- Email: best-effort 36 h response (48 h on weekends).
- AI tools: welcome as a technical resource; must cite; not for quiz/assignment answers.
- Full syllabus on the course website (coming up).

Semester readings — Weeks 3–7

Each week includes 1-2 papers alongside the textbook. These may change.

Wk 3 Conjugate Bayes & Topic Models

Griffiths & Tenenbaum 2001 (Randomness) · Steyvers & Griffiths 2000 (Topic Models)

Wk 4 Generalization & Hierarchical Bayes

Tenenbaum 1999 (Concept Learning) · Xu & Tenenbaum 2000

Wk 5 Hierarchical Bayes & Bayes Nets

DeZoete 2019

Wk 6 Causal Bayes Nets

Lagnado 2002 · Griffiths et al. 2004 (Hidden Causes) · Gerstenberg 2015

Wk 7 Markov Chains & Networks

Abbott et al. 2012 (Random Walks) · Mukherjee 2017

Semester readings — Weeks 8–13

Wk 8 Monte Carlo Methods

Vul et al. 2009 (One and Done) · Griffiths & Sanborn 2007

Wk 9 SDT, MDPs & Reinforcement Learning

Schultz 1997 (Dopamine & Reward) · Daw et al. 2005 · Ho et al. 2015

Wk 10 Inverse Reinforcement Learning

Baker et al. 2007 (Theory of Mind) · Ho et al. 2016, 2017

Wk 11 Bayesian Nonparametrics

Austerweil et al. 2015 (BNP Book Chapter)

Wk 12 Deep Neural Networks

LeCun et al. 2015 (Deep Learning) · Lake et al. 2015 · Pereira et al. 2018

Wk 13 Ethics & Adversarial ML

Caliskan et al. 2017 (Bias) · Buolamwini 2018 · Zhou & Firestone 2019

Homework for Week 2

1. **Textbook T1 Ch 1-3 — reinforces everything from Block 3.**
   `intro/01_goals.md` · `02_hungry.md` · `03_prob_count.md`
2. **Textbook T2 Ch 0-1 — gets GenJAX running in Colab.**
   `genjax/00_getting_started.md` · `01_python_basics.md`
3. **Start thinking about a final-project topic. We'll talk Week 2.**

Week 2 has a short optional self-check quiz (Intro Probability Theory 1).

Questions?

Thanks — see you next Friday.