Week 2 — Levels of Analysis + Bayes Continued

Friday, April 24, 2026

Prof. Joseph Austerweil

ChibaTech
School of
Design & Science

Agenda

Welcome + meet Chibany	0:00
Marr’s three levels	0:08
Notation lock-in	0:15
Joint, marginal, conditional, independence	0:18
Expected value + discrete	0:30
Continuous prob + Gaussian	0:45
Break	1:05
Gaussian-Gaussian update	1:15

Welcome back + meet Chibany

Two changes since Week 1

Paper presentations are in.

Each of you presents one reading once. More at 1:40.

Reflections: 6 of 12, not 8 of 13.

Class is now 2 h × 12 weeks. More at 1:50.

Meet Chibany

Chibany is the Chiba Tech mascot. They hang out at the cafeteria.

Students bring them two bentos a day — lunch and dinner — as offerings.

Chibany loves tonkatsu.

Chibany is the protagonist of our textbook,

“A Narrative Introduction to Probability” — Tutorial 3 Ch 1.

Meet Chibany (2/4)

Chibany is the Chiba Tech mascot. They hang out at the cafeteria.

Students bring them two bentos a day — lunch and dinner — as offerings.

Chibany loves tonkatsu.

Chibany is the protagonist of our textbook,

“A Narrative Introduction to Probability” — Tutorial 3 Ch 1.

The catch this week: the bento boxes are OPAQUE.

Chibany can’t peek — and it would be rude to open one while the student watches.

Meet Chibany (3/4)

Chibany is the Chiba Tech mascot. Students offer them two bentos a day.

Chibany loves tonkatsu.

Textbook pointer: T3 Ch 1 — “A Narrative Introduction to Probability.”

The catch this week: the bento boxes are OPAQUE.

Chibany can’t peek — and it would be rude to open one while the student watches.

What Chibany knows (from last week’s transparent bentos):

- Tonkatsu bentos weigh about 500 g
- Hamburger bentos weigh about 350 g
- About 70% of offerings are tonkatsu
  (overheard two students chatting)

Meet Chibany (4/4)

Chibany is the Chiba Tech mascot. Students offer them two bentos a day.
Chibany loves tonkatsu.
Textbook pointer: T3 Ch 1 — “A Narrative Introduction to Probability.”
The catch this week: the bento boxes are OPAQUE.
Chibany can’t peek — and it would be rude to open one while the student watches.
What Chibany knows (from last week’s transparent bentos):

- Tonkatsu bentos weigh about 500 g
- Hamburger bentos weigh about 350 g
- About 70% of offerings are tonkatsu
  (overheard two students chatting)

Chibany’s plan: WEIGH THE BENTO.
“The weight won’t tell me for sure if it’s tonkatsu,
but it will UPDATE my belief about today’s meal.”
That update is Bayes’ rule. Same rule as last week — new anchor.

Agenda

Welcome + meet Chibany	0:00
Marr’s three levels	0:08
Notation lock-in	0:15
Joint, marginal, conditional, independence	0:18
Expected value + discrete	0:30
Continuous prob + Gaussian	0:45
Break	1:05
Gaussian-Gaussian update	1:15

Marr’s three levels — frame for the whole course

ChibaTech School of Design & Science

Marr L1 — what problem is being solved?

For Chibany: Compute P(meal | weight).

L1 specifies the CORRECT computation — not how to do it,
not how Chibany's brain actually does it.

Different goal ⇒ different L1 ⇒ different normative answer.

Marr L2 — what algorithm?

Same L1 — multiple algorithms:

- Enumerate (today: 2 hypotheses, sum-and-normalize)
- Sample (Week 7 Monte Carlo)
- “Always guess tonkatsu” — an algorithm. A BAD one.

Bayes is L1. Algorithms are L2.

Marr L3 — what implementation?

Chibany: neurons, experience, gut feeling.

Our simulation: JAX arrays, floating-point math, Colab.

Three levels — mostly independent.

Right L1 + wrong L2 → systematic human biases.

(Full L3 treatment: Week 11, Deep NNs.)

Check-in — the cab problem

Last week: witness says “blue”; 15% blue cabs;
witness accurate 80%. Most of you guessed ~80%.

At which level was the mistake?

Target: L1 = P(blue | report). Actual L1 answer ≈ 0.41.
80% is a wrong-L2 claim (ignore the base rate).

Agenda

Welcome + meet Chibany	0:00
Marr’s three levels	0:08
Notation lock-in	**0:15**
Joint, marginal, conditional, independence	0:18
Expected value + discrete	0:30
Continuous prob + Gaussian	0:45
Break	1:05
Gaussian-Gaussian update	1:15

Notation lock-in

Notation — one rule, starting today

From today:

H = hypothesis (hidden — what we want to know)

D = data (observed — what we see)

For Chibany: H = meal, D = today’s observation.

Notation — one rule, starting today (2/3)

From today:

H = hypothesis (hidden — what we want to know)

D = data (observed — what we see)

For Chibany: H = meal, D = today’s observation.

$$P(H = h \mid D = d) \;=\; \frac{P(D = d \mid H = h)\, P(H = h)}{P(D = d)}$$

Notation — one rule, starting today (3/3)

From today:

H = hypothesis (hidden — what we want to know)

D = data (observed — what we see)

For Chibany: H = meal, D = today's observation.

$$P(H = h \mid D = d) \;=\; \frac{P(D = d \mid H = h)\, P(H = h)}{P(D = d)}$$

posterior = likelihood · prior / evidence

ChibaTech School of Design & Science

Agenda

Welcome + meet Chibany	0:00
Marr’s three levels	0:08
Notation lock-in	0:15
Joint, marginal, conditional, independence	**0:18**
Expected value + discrete	0:30
Continuous prob + Gaussian	0:45
Break	1:05
Gaussian-Gaussian update	1:15

Joint, marginalization, conditional, independence

Last week's sick-friend problem, in H/D notation

Last week: smoker friend, cough heard, three hypotheses.

In the new notation: H ∈ { cold, stomach virus, lung cancer }, D = cough.

	cold	stomach virus	lung cancer
Prior P(H)	?	?	?
Likelihood P(D ∣ H)	?	?	?
Numerator P(H) · P(D ∣ H)	?	?	?
Posterior P(H ∣ D)	?	?	?

Last week's sick-friend problem, in H/D notation (2/3)

Last week: smoker friend, cough heard, three hypotheses.

In the new notation: $H \in \{$ cold, stomach virus, lung cancer $\}$, D = cough.

	cold	stomach virus	lung cancer
Prior $P(H)$	0.45	0.10	0.10
Likelihood $P(D \mid H)$	0.90	0.45	0.90
Numerator $P(H) \cdot P(D \mid H)$	0.405	0.045	0.090
Posterior $P(H \mid D)$	?	?	?

Evidence $P(D) = 0.405 + 0.045 + 0.090 = 0.540$

Last week's sick-friend problem, in H/D notation (3/3)

Last week: smoker friend, cough heard, three hypotheses.

In the new notation: $H \in \{ \text{cold, stomach virus, lung cancer} \}$, $D = \text{cough}$.

	cold	stomach virus	lung cancer
Prior $P(H)$	0.45	0.10	0.10
Likelihood $P(D \mid H)$	0.90	0.45	0.90
Numerator $P(H) \cdot P(D \mid H)$	0.405	0.045	0.090
Posterior $P(H \mid D)$	**0.75**	**0.083**	**0.167**

Evidence $P(D) = 0.540$ → posteriors = numerators / evidence.

Same answer as last week. Same rule. New layout.

Everything we do today uses this SAME structure.

Setup — two meals, two variables

Today Chibany gets two bentos: one for lunch, one for dinner.

A = lunch meal, $A \in$ {tonkatsu, hamburger}

B = dinner meal, $B \in$ {tonkatsu, hamburger}

Same outcome space as Week 1’s two-coin-flip grid: Ω = {HH, HT, TH, TT}.

Now the “coin” has a biased and asymmetric joint.

The joint P(A, B) — build the 2×2 table

Tanaka-san tells Chibany that students coordinate: they try to bring at least one tonkatsu, but not two (to keep it special).

	B = H	B = T
A = H	?	?
A = T	?	?

The joint P(A, B) — build the 2×2 table (2/3)

Tanaka-san tells Chibany that students coordinate: they try to bring at least one tonkatsu, but not two (to keep it special).

	B = H	B = T
A = H	0.04	0.43
A = T	0.43	0.10

HH is rare (both offerings hamburger).
HT and TH are common (exactly one tonkatsu).
TT is moderate (students avoid giving two tonkatsu in a row).

The joint P(A, B) — build the 2×2 table (3/3)

Tanaka-san tells Chibany that students coordinate: they try to bring at least one tonkatsu, but not two (to keep it special).

	B = H	B = T
A = H	0.04	0.43
A = T	0.43	0.10

Sanity check: 0.04 + 0.43 + 0.43 + 0.10 = 1.00 ✓

Recognize the shape? Week 1’s 2-coin-flip grid had 4 cells too. Different numbers, same structure.

Marginalization — sum over what you don’t care about

$P(A = T) = ?$

Sum the joint over B:

$$P(A = T) = P(A = T, B = H) + P(A = T, B = T)$$

$$= 0.43 + 0.10 = 0.53$$

Marginalization — sum over what you don't care about (2/3)

P(A = T) = ?

Sum the joint over B:

$P(A = T) = P(A = T, B = H) + P(A = T, B = T)$

$\mathbf{= 0.43 + 0.10 = 0.53}$

$\mathbf{P(A = H) = 0.04 + 0.43 = 0.47}$

$0.53 + 0.47 = 1$ ✓

Marginalization — sum over what you don't care about (3/3)

$P(A = T) = ?$

Sum the joint over B:

$$P(A = T) = P(A = T, B = H) + P(A = T, B = T)$$

$$\mathbf{= 0.43 + 0.10 = 0.53}$$

$P(A = H) = 0.04 + 0.43 = 0.47$

$0.53 + 0.47 = 1$ ✓

The named move: MARGINALIZATION.

Given the joint, sum over the variable you're not asking about.

Week 1 called this the sum rule.

Conditional — restrict and renormalize

Chibany learns: today's DINNER was tonkatsu ($B = T$).

What's $P(A \mid B = T)$? What should Chibany now believe about lunch?

Same move as Week 1's 2×2 coin-flip grid:

"Restrict to the slice where the observation is true; renormalize."

Conditional — restrict and renormalize (2/3)

Chibany learns: today's DINNER was tonkatsu ($B = T$).

What's $P(A \mid B = T)$? What should Chibany now believe about lunch?

Same move as Week 1's 2×2 coin-flip grid:

"Restrict to the slice where the observation is true; renormalize."

Step 1: keep only the "B = T" column of the joint.

$P(A = H, B = T) = 0.43$ $P(A = T, B = T) = 0.10$

Step 2: renormalize by $P(B = T) = 0.53$.

Conditional — restrict and renormalize (3/3)

Chibany learns: today's DINNER was tonkatsu ($B = T$).

What's $P(A \mid B = T)$? What should Chibany now believe about lunch?

Same move as Week 1's 2×2 coin-flip grid:

"Restrict to the slice where the observation is true; renormalize."

Step 1: keep only the "B = T" column of the joint.

$P(A = H, B = T) = 0.43$ $P(A = T, B = T) = 0.10$

Step 2: renormalize by $P(B = T) = 0.53$.

$\mathbf{P(A = H \mid B = T) = 0.43 / 0.53 \approx 0.811}$

$\mathbf{P(A = T \mid B = T) = 0.10 / 0.53 \approx 0.189}$

Learning dinner was tonkatsu made Chibany MORE confident lunch was hamburger.

Makes sense — students avoid two tonkatsu in a row.

Independence — definition

A and B are **independent** (written $A \perp B$) iff:

$P(A \mid B) = P(A)$ for every value of A and every value of B with $P(B) > 0$

“Learning B doesn’t change what I believe about A.”

Equivalently: **$P(A, B) = P(A) \cdot P(B)$**

The joint factors into the product of marginals.

Independence — check today’s joint

Are A (lunch) and B (dinner) independent?

Marginal: **$P(A = T) = 0.53$**

Conditional: **$P(A = T \mid B = T) \approx 0.189$**

$0.189 \neq 0.53$. A and B are DEPENDENT: not ($A \perp B$).

If students acted independently — flipping a coin each meal for T with some p —
the joint would FACTOR: $P(A, B) = P(A) \cdot P(B)$. It doesn’t.
The coordination IS the dependence.

Summary — four operations on the joint

Joint = full picture, $P(A, B)$. All cells sum to 1.

Marginal = sum the joint over the thing you don't care about.

Conditional = slice on your observation, renormalize.

Independence = check if $P(A, B) = P(A) \cdot P(B)$ (equivalently, $P(A|B) = P(A)$).

Same Bayes as last week. Now you can SEE it as a table operation.

Agenda

Welcome + meet Chibany	0:00
Marr’s three levels	0:08
Notation lock-in	0:15
Joint, marginal, conditional, independence	0:18
Expected value + discrete	**0:30**
Continuous prob + Gaussian	0:45
Break	1:05
Gaussian-Gaussian update	1:15

ChibaTech School of Design & Science

Expected value + discrete distributions

Expected value E[X]

For a discrete RV X with values $\{x_i\}$ and probabilities $\{p_i\}$:

$$\mathbb{E}[X] = \sum_i x_i \, p_i$$

“Weight each value by its probability, then add.”

Expected value E[X] (2/4)

For a discrete RV X with values $\{x_i\}$ and probabilities $\{p_i\}$:

$$\mathbb{E}[X] = \sum_i x_i \, p_i$$

"Weight each value by its probability, then add."

Chibany: encode meal numerically. tonkatsu = 1, hamburger = 0.

$$\mathbb{E}[\text{meal}] = 1 \cdot 0.7 + 0 \cdot 0.3 = 0.7$$

The expected value of an indicator IS the probability.

Expected value E[X] (3/4)

For a discrete RV X with values $\{x_i\}$ and probabilities $\{p_i\}$:

$$\mathbb{E}[X] = \sum_i x_i \, p_i$$

“Weight each value by its probability, then add.”

Chibany: encode meal numerically. tonkatsu = 1, hamburger = 0.

$$\mathbb{E}[\mathrm{meal}] = 1 \cdot 0.7 + 0 \cdot 0.3 = 0.7$$

The expected value of an indicator IS the probability.

Variance: $\mathrm{Var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

For $\mathrm{Bernoulli}(p)$: $\mathrm{Var} = p(1-p)$

Bento meal: Var = 0.21, σ ≈ 0.46

Expected value E[X] (4/4)

For a discrete RV X with values $\{x_i\}$ and probabilities $\{p_i\}$:

$$\mathbb{E}[X] = \sum_i x_i \, p_i$$

“Weight each value by its probability, then add.”

Chibany: encode meal numerically. tonkatsu = 1, hamburger = 0.

$$\mathbb{E}[\text{meal}] = 1 \cdot 0.7 + 0 \cdot 0.3 = 0.7$$

The expected value of an indicator IS the probability.

Variance: $\mathrm{Var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

For $\mathrm{Bernoulli}(p)$: $\mathrm{Var} = p(1-p)$

Bento meal: Var = 0.21, σ ≈ 0.46

$\mathbb{E}[X]$ **= center of mass.** $\mathrm{Var}[X]$ **= spread.**

Bernoulli distribution

$X \sim \mathrm{Bernoulli}(p)$ means:

$$P(X = 1) = p, \quad P(X = 0) = 1 - p$$

One flip of a weighted coin. One parameter: p.

Bernoulli distribution (2/3)

$X \sim \mathrm{Bernoulli}(p)$ means:

$$P(X = 1) = p, \quad P(X = 0) = 1 - p$$

One flip of a weighted coin. One parameter: p.

Chibany: $\mathrm{meal}_{\mathrm{today}} \sim \mathrm{Bernoulli}(0.7)$**.**

P(meal_today = tonkatsu) = 0.7

P(meal_today = hamburger) = 0.3

Bernoulli distribution (3/3)

$X \sim \mathrm{Bernoulli}(p)$ means:

$$P(X = 1) = p, \quad P(X = 0) = 1 - p$$

One flip of a weighted coin. One parameter: p.

Chibany: $\mathrm{meal}_{\mathrm{today}} \sim \mathrm{Bernoulli}(0.7)$.

P(meal_today = tonkatsu) = 0.7

P(meal_today = hamburger) = 0.3

Today we treat p = 0.7 as GIVEN.

Next week: we ask where p comes from, and infer it from data.

The conjugate prior on p is the Beta distribution — Week 3.

Binomial distribution

Chibany counts tonkatsu across 5 days.

Each day: $\mathrm{Bernoulli}(0.7)$, independent.

Let Y = total tonkatsu in 5 days. $Y \sim \mathrm{Binomial}(5, 0.7)$.

Binomial distribution (2/3)

Chibany counts tonkatsu across 5 days.

Each day: $\mathrm{Bernoulli}(0.7)$, independent.

Let Y = total tonkatsu in 5 days. $Y \sim \mathrm{Binomial}(5, 0.7)$.

$$P(Y = k) = \binom{n}{k} p^k \,(1 - p)^{n-k}$$

- n = 5 (trials)
- k (count of successes)
- p = 0.7 (per-trial success prob)
- C(n, k) counts orderings of k successes among n trials

Binomial distribution (3/3)

Chibany counts tonkatsu across 5 days.

Each day: $\mathrm{Bernoulli}(0.7)$, independent.

Let Y = total tonkatsu in 5 days. $Y \sim \mathrm{Binomial}(5, 0.7)$.

$$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

n = 5 (trials)
k (count of successes)
p = 0.7 (per-trial success prob)
C(n, k) counts orderings of k successes among n trials

$P(Y = 5) = 1 \cdot 0.7^5 \cdot 1 \approx 0.168$

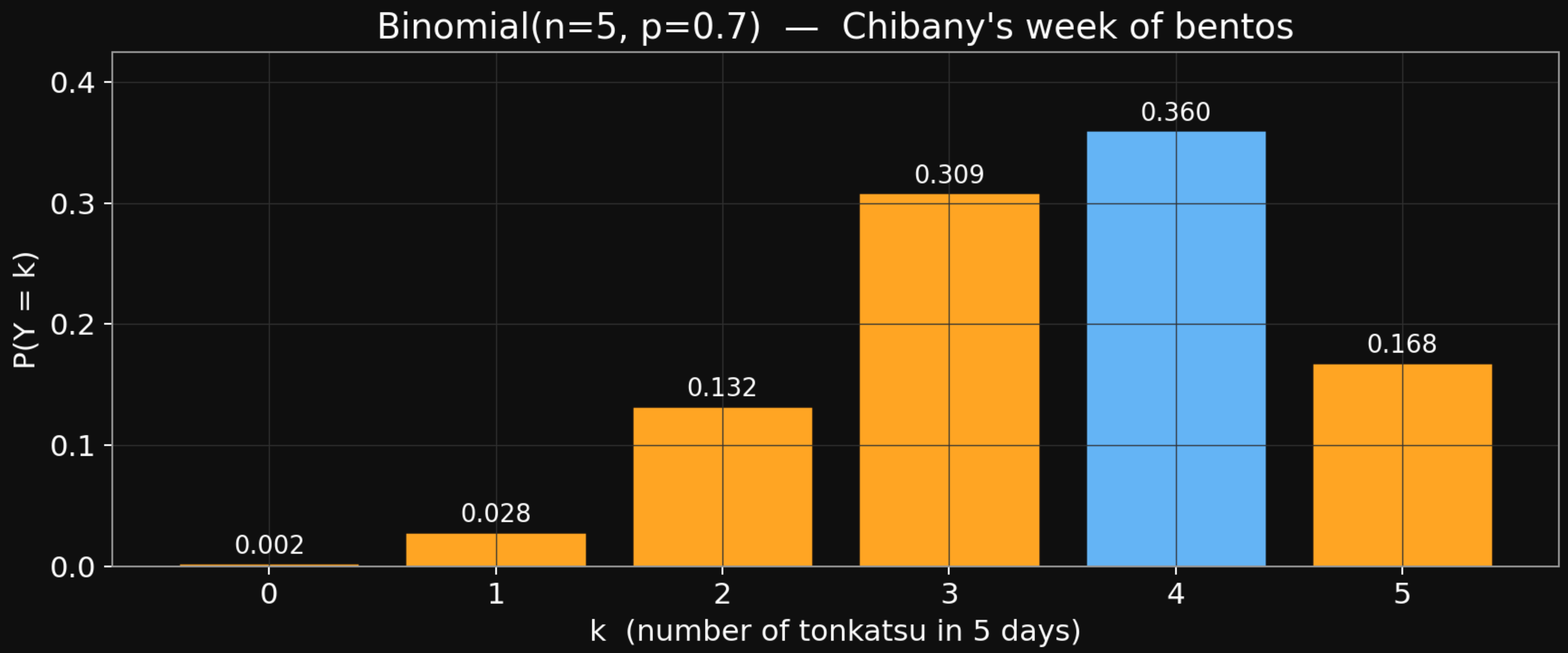
$P(Y = 4) = 5 \cdot 0.7^4 \cdot 0.3 \approx 0.360$ ← the MODE

$P(Y = 3) = 10 \cdot 0.7^3 \cdot 0.3^2 \approx 0.309$

Why is k=4 more likely than k=5, even though p = 0.7?

There are 5 ways to get 4 tonkatsu in 5 days, but only 1 way to get all 5.
The extra orderings outweigh the smaller (0.7·0.3) factor.

Binomial(5, 0.7) — full PMF



Mode at k=4 (0.360). In Week 3 this PMF becomes the LIKELIHOOD

for inferring p from observed counts. Beta × Binomial = Beta.

ChibaTech School of Design & Science

Agenda

Welcome + meet Chibany	0:00
Marr’s three levels	0:08
Notation lock-in	0:15
Joint, marginal, conditional, independence	0:18
Expected value + discrete	0:30
Continuous prob + Gaussian	**0:45**
Break	1:05
Gaussian-Gaussian update	1:15

Continuous probability + the Gaussian

From PMF to PDF — going continuous

Chibany’s bentos come in all weights — not a finite menu of values, a real number.

We need a different object — a probability DENSITY.

From PMF to PDF — going continuous (2/4)

Chibany’s bentos come in all weights — not a finite menu of values, a real number.

We need a different object — a probability DENSITY.

Thought experiment — shrink the bins:

30g bins: coarse histogram, PMF over bins

10g bins: finer, still PMF

2g bins: finer still

bin → 0: PMF → PDF

From PMF to PDF — going continuous (3/4)

Chibany's bentos come in all weights — not a finite menu of values, a real number.

We need a different object — a probability DENSITY.

Thought experiment — shrink the bins:

30g bins: coarse histogram, PMF over bins

10g bins: finer, still PMF

2g bins: finer still

bin → 0: PMF → PDF

The density f(x) is NOT a probability — it's probability PER UNIT.

$$P(X \in [a, b]) = \int_a^b f(x)\,dx \qquad \text{“area under the density”}$$

PDF values can exceed 1. Probabilities cannot.

From PMF to PDF — going continuous (4/4)

Chibany's bentos come in all weights — not a finite menu of values, a real number.

We need a different object — a probability DENSITY.

Thought experiment — shrink the bins:

- 30g bins: coarse histogram, PMF over bins
- 10g bins: finer, still PMF
- 2g bins: finer still
- **bin → 0: PMF → PDF**

The density f(x) is NOT a probability — it's probability PER UNIT.

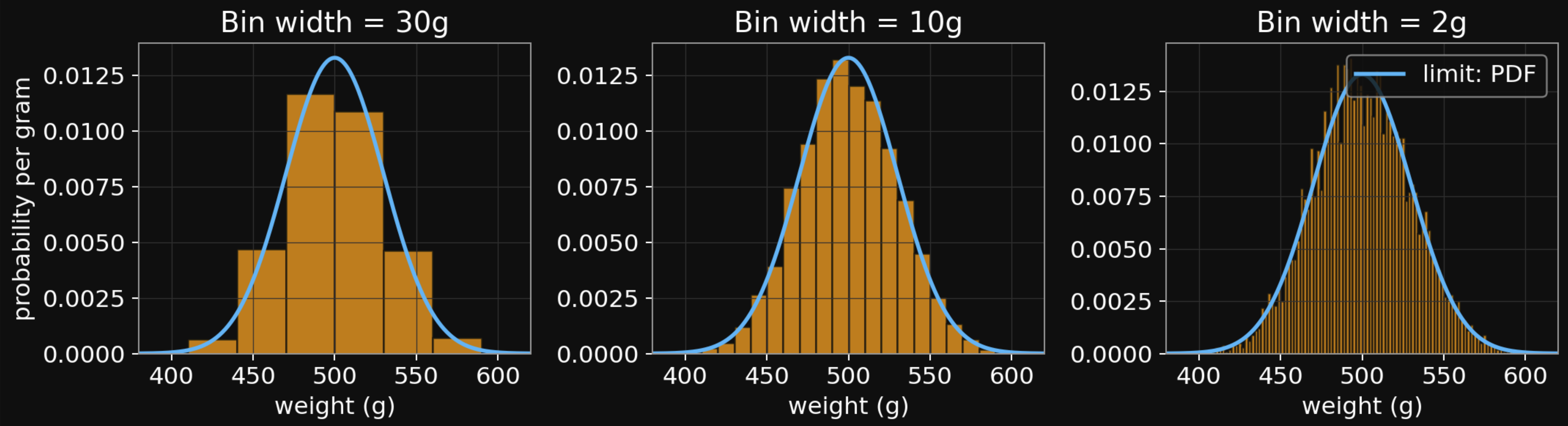
$$P(X \in [a, b]) = \int_a^b f(x)\,dx \qquad \text{“area under the density”}$$

PDF values can exceed 1. Probabilities cannot.

Why P(X = exactly 450g) = 0 for continuous X:

$\int_{450}^{450} f(x)\,dx = 0$ (integrate over a single point)

Shrinking bins → PDF as the limit



Same underlying data, three bin widths. Blue curve: limit (PDF).

The Gaussian (normal) distribution

Chibany's bento weights vary — even within one meal type.

Tonkatsu weight is roughly $N(500, 30^2)$ — centered at 500, spread ≈ 30g.

The Gaussian (normal) distribution (2/5)

Chibany’s bento weights vary — even within one meal type.

Tonkatsu weight is roughly $N(500, 30^2)$ — centered at 500, spread ≈ 30g.

$$N(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

The Gaussian (normal) distribution (3/5)

Chibany’s bento weights vary — even within one meal type.

Tonkatsu weight is roughly $N(500, 30^2)$ — centered at 500, spread ≈ 30g.

$$N(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Name each piece:

- μ mean — where the curve peaks
- σ^2 variance — how wide the curve is (σ = std dev)
- $\exp(-(x-\mu)^2/(2\sigma^2))$ peaks at μ, falls off fast outside $\pm\sigma$
- $1/\sqrt{2\pi\,\sigma^2}$ the normalizer — makes total area = 1

The Gaussian (normal) distribution (4/5)

Chibany's bento weights vary — even within one meal type.

Tonkatsu weight is roughly $N(500, 30^2)$ — centered at 500, spread ≈ 30g.

$$N(x \mid \mu, \sigma^2) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}}\;\exp\!\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Name each piece:

- μ mean — where the curve peaks
- σ^2 variance — how wide the curve is (σ = std dev)
- $\exp(-(x-\mu)^2/(2\sigma^2))$ peaks at μ, falls off fast outside $\pm\sigma$
- $1/\sqrt{(2\pi\,\sigma^2)}$ the normalizer — makes total area = 1

Shape facts:

- symmetric around μ
- peak height = $1/\sqrt{(2\pi\,\sigma^2)}$ — sharper curves are TALLER
- ~68% of mass within $\mu \pm \sigma$; ~95% within $\mu \pm 2\sigma$

The Gaussian (normal) distribution (5/5)

Chibany's bento weights vary — even within one meal type.

Tonkatsu weight is roughly $N(500, 30^2)$ — centered at 500, spread ≈ 30g.

$$N(x \mid \mu, \sigma^2) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}}\;\exp\!\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Name each piece:

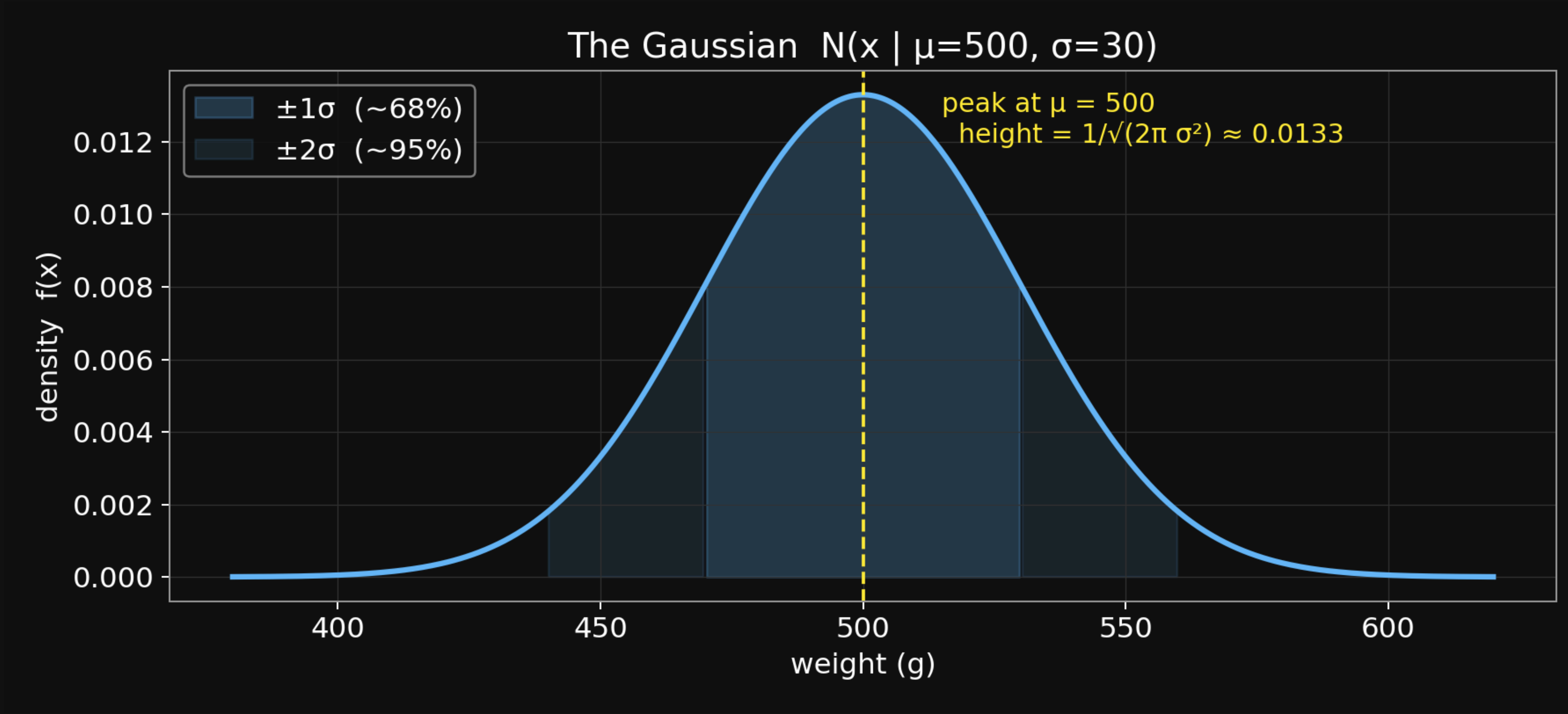
- μ mean — where the curve peaks
- σ^2 variance — how wide the curve is (σ = std dev)
- $\exp(-(x-\mu)^2/(2\sigma^2))$ peaks at μ, falls off fast outside $\pm\sigma$
- $1/\sqrt{(2\pi\,\sigma^2)}$ the normalizer — makes total area = 1

Shape facts:

- symmetric around μ
- peak height = $1/\sqrt{(2\pi\,\sigma^2)}$ — sharper curves are TALLER
- ~68% of mass within $\mu \pm \sigma$; ~95% within $\mu \pm 2\sigma$

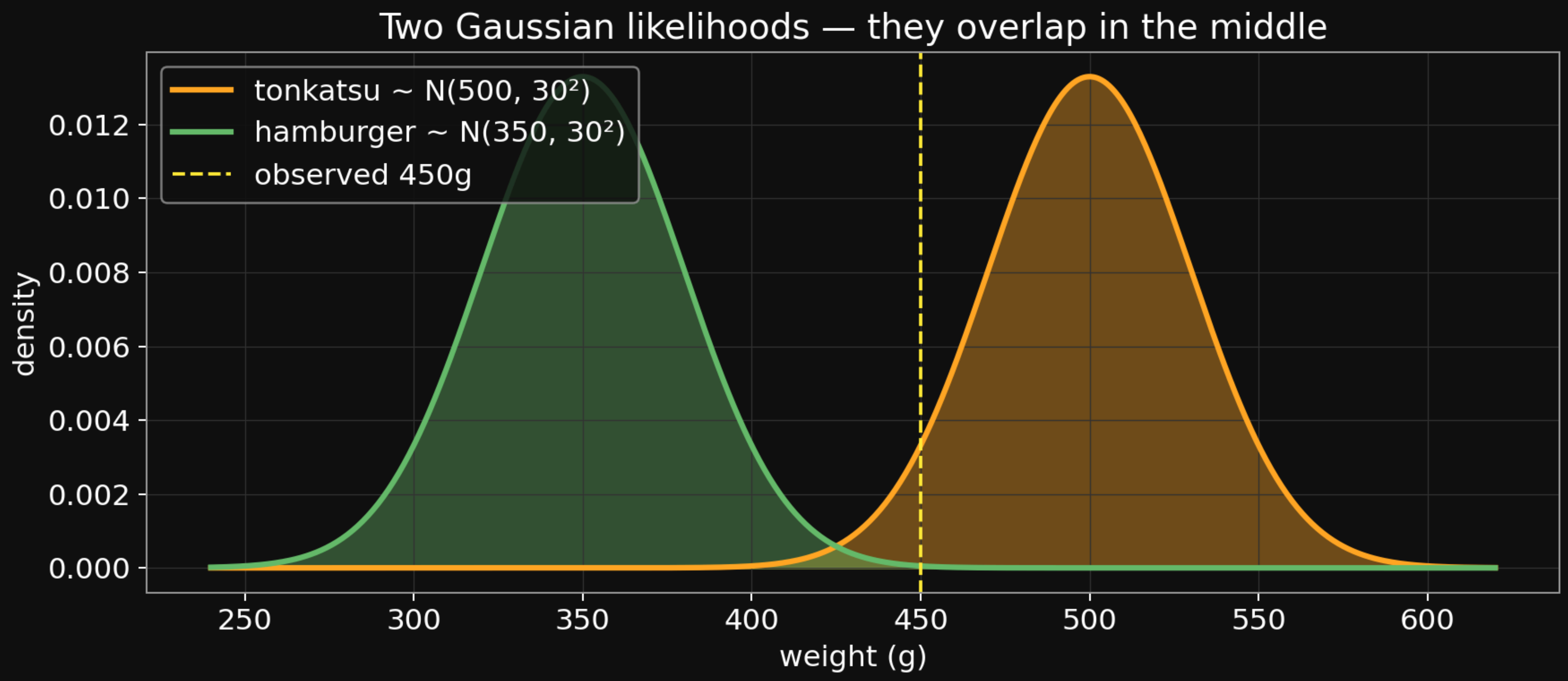
ChibaTech School of Design & Science

The Gaussian — shape and the 68-95 rule



Dark fill: ±1σ (~68%). Light fill: ±2σ (~95%).

Chibany’s two bento-weight likelihoods



Yellow line: observed 450g. Closer to tonkatsu, but within hamburger tail.

ChibaTech School of Design & Science

Bayes with a continuous likelihood — worked

Today's observation: D = weight = 450g.

What's P(meal | weight = 450g) ?

Bayes with a continuous likelihood — worked (2/6)

Today's observation: D = weight = 450g.

What's P(meal | weight = 450g) ?

Prior stays discrete: P(tonkatsu) = 0.7, P(hamburger) = 0.3

Bayes with a continuous likelihood — worked (3/6)

Today's observation: D = weight = 450g.

What's P(meal | weight = 450g) ?

Prior stays discrete: P(tonkatsu) = 0.7, P(hamburger) = 0.3

Likelihood is a Gaussian density:

$$f_T = N(450 \mid 500, 30^2)$$

$$= \frac{1}{\sqrt{2\pi \cdot 900}} \cdot \exp\left(-\frac{(-50)^2}{1800}\right)$$

$$= 0.01330 \cdot \exp(-1.389) \approx 0.00332$$

$$f_H = N(450 \mid 350, 30^2) \approx 0.00044$$

Density values, not probabilities. (Integral → probability.)

ChibaTech School of Design & Science

Bayes with a continuous likelihood — worked (4/6)

Today's observation: D = weight = 450g.
What's P(meal | weight = 450g) ?
Prior stays discrete: P(tonkatsu) = 0.7, P(hamburger) = 0.3
Likelihood is a Gaussian density:

$$
\begin{aligned}
f_T &= N(450 \mid 500, 30^2) \\
&= \frac{1}{\sqrt{2\pi \cdot 900}} \cdot \exp\left(-\frac{(-50)^2}{1800}\right) \\
&= 0.01330 \cdot \exp(-1.389) \approx 0.00332
\end{aligned}
$$

$$f_H = N(450 \mid 350, 30^2) \approx 0.00044$$

Density values, not probabilities. (Integral → probability.)

Numerator = prior · likelihood:

tonkatsu: $0.7 \cdot 0.00332 = 0.00232$

hamburger: $0.3 \cdot 0.00044 = 0.000132$

Evidence = sum = 0.00246

Bayes with a continuous likelihood — worked (5/6)

Today's observation: D = weight = 450g.

What's P(meal | weight = 450g) ?

Prior stays discrete: P(tonkatsu) = 0.7, P(hamburger) = 0.3

Likelihood is a Gaussian density:

$$f_T = N(450 \mid 500, 30^2)$$
$$= \frac{1}{\sqrt{2\pi \cdot 900}} \cdot \exp\left(-\frac{(-50)^2}{1800}\right)$$
$$= 0.01330 \cdot \exp(-1.389) \approx 0.00332$$

$$f_H = N(450 \mid 350, 30^2) \approx 0.00044$$

Density values, not probabilities. (Integral → probability.)

Numerator = prior · likelihood:

tonkatsu: $0.7 \cdot 0.00332 = 0.00232$

hamburger: $0.3 \cdot 0.00044 = 0.000132$

Evidence = sum = 0.00246

Posterior:

$$P(\text{tonkatsu} \mid 450\text{g}) = 0.00232/0.00246 \approx 0.946$$
$$P(\text{hamburger} \mid 450\text{g}) = 0.000132/0.00246 \approx 0.054$$

Bayes with a continuous likelihood — worked (6/6)

Today's observation: D = weight = 450g.
What's P(meal | weight = 450g) ?
Prior stays discrete: P(tonkatsu) = 0.7, P(hamburger) = 0.3
Likelihood is a Gaussian density:

$$f_T = N(450 \mid 500, 30^2)$$
$$= \frac{1}{\sqrt{2\pi \cdot 900}} \cdot \exp\left(-\frac{(-50)^2}{1800}\right)$$
$$= 0.01330 \cdot \exp(-1.389) \approx 0.00332$$
$$f_H = N(450 \mid 350, 30^2) \approx 0.00044$$

Density values, not probabilities. (Integral → probability.)

Numerator = prior · likelihood:

tonkatsu: $0.7 \cdot 0.00332 = 0.00232$

hamburger: $0.3 \cdot 0.00044 = 0.000132$

Evidence = sum = 0.00246

Posterior:

$$P(\text{tonkatsu} \mid 450\text{g}) = 0.00232/0.00246 \approx 0.946$$
$$P(\text{hamburger} \mid 450\text{g}) = 0.000132/0.00246 \approx 0.054$$

Prior → Posterior: 0.70 → 0.946 (tonkatsu)
0.30 → 0.054 (hamburger)

A weight of 450g is closer to the hamburger mean (350) than the tonkatsu mean (500),

Break — 10 minutes

Agenda

Welcome + meet Chibany	0:00
Marr’s three levels	0:08
Notation lock-in	0:15
Joint, marginal, conditional, independence	0:18
Expected value + discrete	0:30
Continuous prob + Gaussian	0:45
Break	1:05
Gaussian-Gaussian update	**1:15**

ChibaTech School of Design & Science

Gaussian-Gaussian update — launching Week 3

Why would Chibany want to infer μ ?

So far Chibany asks: “what’s in TODAY’s bento?”

Today’s weight, 450g → posterior over today’s meal.

Why would Chibany want to infer μ ? (2/3)

So far Chibany asks: “what’s in TODAY’s bento?”

Today’s weight, 450g → posterior over today’s meal.

But to DO that well, Chibany needs to know:

How much DO tonkatsu bentos typically weigh?

(We wrote “500” as if it were known. But is it?)

Why would Chibany want to infer μ ? (3/3)

So far Chibany asks: “what’s in TODAY’s bento?”

Today’s weight, 450g → posterior over today’s meal.

But to DO that well, Chibany needs to know:

How much DO tonkatsu bentos typically weigh?

(We wrote “500” as if it were known. But is it?)

Shift in perspective:

- Yesterday’s Bayes: H = today’s MEAL (discrete hypothesis)
- Today’s Bayes: H = tonkatsu’s MEAN WEIGHT μ (continuous parameter)

Why Bayesian inference on μ instead of just averaging?

Because Chibany has PRIOR knowledge (last week), plus LIMITED data.

Averaging throws the prior away. Bayes combines both.

Step 1 — the prior $\mu \sim N(\mu_0, \sigma_0^2)$

Chibany's belief about μ BEFORE observing any new weights:

$$\mu \ \sim \ N(\mu_0, \sigma_0^2)$$

"I think μ is around μ_0, give or take σ_0."

Step 1 — the prior $\mu \sim N(\mu_0, \sigma_0^2)$ (2/3)

Chibany's belief about μ BEFORE observing any new weights:

$$\mu \ \sim \ N(\mu_0, \sigma_0^2)$$

"I think μ is around μ_0, give or take σ_0."

From last week's memory:

$\mu_0 = 500$ **(best guess)** $\sigma_0 = 20$ **(uncertainty)**

Prior: $\mu \sim N(500, 20^2)$

Step 1 — the prior $\mu \sim N(\mu_0, \sigma_0^2)$ (3/3)

Chibany's belief about μ BEFORE observing any new weights:

$$\mu \sim N(\mu_0, \sigma_0^2)$$

"I think μ is around μ_0, give or take σ_0."

From last week's memory:

$\mu_0 = 500$ **(best guess)** $\sigma_0 = 20$ **(uncertainty)**

Prior: $\mu \sim N(500, 20^2)$

The conceptual leap: priors over PARAMETERS.

The Gaussian is NOT over weights now. It's over μ.

A prior is a distribution over anything hidden —

including parameters of OTHER distributions.

Step 2 — the likelihood

Chibany weighs one bento: $D_1 = 510$g.

(Assume $\sigma = 30$ is known — we infer μ only.)

Step 2 — the likelihood (2/3)

Chibany weighs one bento: $D_1 = 510$g.

(Assume $\sigma = 30$ is known — we infer μ only.)

Likelihood:

$$P(D_1 = 510 \mid \mu) = N(510 \mid \mu, 30^2)$$

Step 2 — the likelihood (3/3)

Chibany weighs one bento: $D_1 = 510$g.

(Assume $\sigma = 30$ is known — we infer μ only.)

Likelihood:

$P(D_1 = 510 \mid \mu) = N(510 \mid \mu, 30^2)$

As a function of μ, this is ALSO a Gaussian —

peaked at $\mu = 510$, width $\sigma = 30$.

So we have two Gaussian functions of μ:

prior (peaked at 500, tight)

likelihood (peaked at 510, broader)

Why Gaussian × Gaussian = Gaussian (derivation sketch)

Multiply the two densities (drop the normalizers — collect at end):

$$P(\mu \mid D) \;\propto\; \exp\left(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right)\cdot\exp\left(-\frac{(D-\mu)^2}{2\sigma^2}\right)$$

Why Gaussian × Gaussian = Gaussian (derivation sketch) (2/4)

Multiply the two densities (drop the normalizers — collect at end):

$$P(\mu \mid D) \;\propto\; \exp\!\left(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right)\cdot\exp\!\left(-\frac{(D-\mu)^2}{2\sigma^2}\right)$$

Combine exponents:

$$\propto\; \exp\!\left(-\left[\frac{(\mu-\mu_0)^2}{2\sigma_0^2} \;+\; \frac{(D-\mu)^2}{2\sigma^2}\right]\right)$$

The thing in brackets is a quadratic in μ — $A\mu^2 - 2B\mu + \mathrm{const}$.

Why Gaussian × Gaussian = Gaussian (derivation sketch) (3/4)

Multiply the two densities (drop the normalizers — collect at end):

$$P(\mu \mid D) \;\propto\; \exp\!\left(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right)\cdot\exp\!\left(-\frac{(D-\mu)^2}{2\sigma^2}\right)$$

Combine exponents:

$$\propto\; \exp\!\left(-\left[\frac{(\mu-\mu_0)^2}{2\sigma_0^2} \;+\; \frac{(D-\mu)^2}{2\sigma^2}\right]\right)$$

The thing in brackets is a quadratic in μ — $A\mu^2 - 2B\mu + \text{const}$.

$$\text{Complete the square} \;\longrightarrow\; \exp\!\left(-\frac{(\mu-\mu_{\text{post}})^2}{2\sigma_{\text{post}}^2}\right) \;+\; \text{const}.$$

That’s the kernel of a Gaussian in μ. Posterior is Gaussian.

Why Gaussian × Gaussian = Gaussian (derivation sketch) (4/4)

Multiply the two densities (drop the normalizers — collect at end):

$$P(\mu \mid D) \;\propto\; \exp\!\left(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right)\cdot \exp\!\left(-\frac{(D-\mu)^2}{2\sigma^2}\right)$$

Combine exponents:

$$\propto\; \exp\!\left(-\left[\frac{(\mu-\mu_0)^2}{2\sigma_0^2} \;+\; \frac{(D-\mu)^2}{2\sigma^2}\right]\right)$$

The thing in brackets is a quadratic in μ — $A\mu^2 - 2B\mu + \text{const}$.

$$\text{Complete the square} \;\longrightarrow\; \exp\!\left(-\frac{(\mu-\mu_{\text{post}})^2}{2\sigma_{\text{post}}^2}\right) \;+\; \text{const.}$$

That’s the kernel of a Gaussian in μ. Posterior is Gaussian.

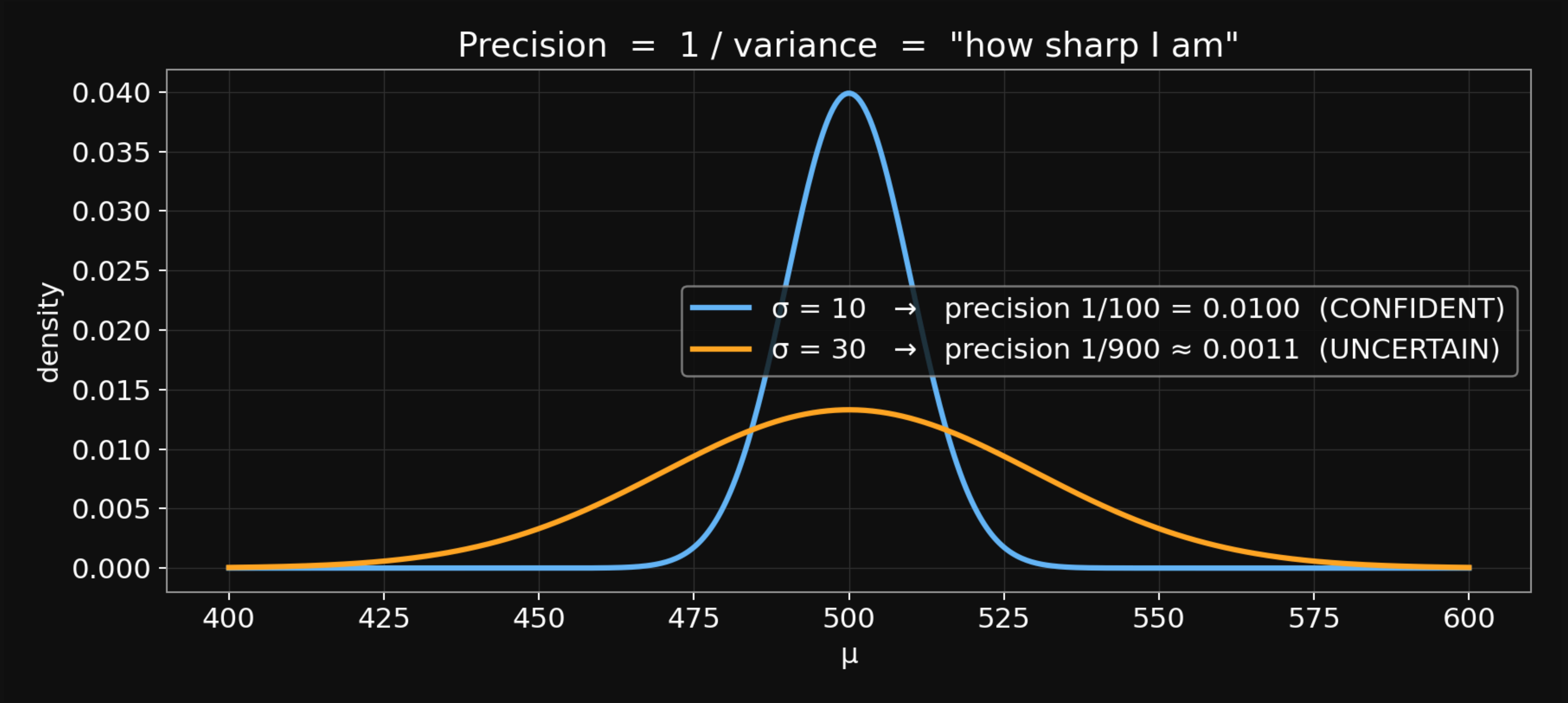
The new mean and variance come from matching coefficients:

$$\frac{1}{\sigma_{\text{post}}^2} \;=\; \frac{1}{\sigma_0^2}+\frac{1}{\sigma^2} \qquad (\text{coefficient of } \mu^2)$$

$$\mu_{\text{post}} \;=\; \sigma_{\text{post}}^2\left(\frac{\mu_0}{\sigma_0^2}+\frac{D}{\sigma^2}\right) \qquad (\text{coefficient of } \mu)$$

ChibaTech School of Design & Science

Precision = 1 / variance = “how sharp I am”



Sharper curve = higher precision = more confident (more weight in the update).

Plug in the numbers — posterior = prior × likelihood, rescaled

The update formulas (from the derivation):

$$\frac{1}{\sigma_{\text{post}}^2} = \frac{1}{\sigma_0^2} + \frac{1}{\sigma^2} \qquad \mu_{\text{post}} = \sigma_{\text{post}}^2 \left(\frac{\mu_0}{\sigma_0^2} + \frac{D}{\sigma^2} \right)$$

Intuition: precisions add. μ_{post} *= precision-weighted average.*

Plug in the numbers — posterior = prior × likelihood, rescaled (2/4)

The update formulas (from the derivation):

$$\frac{1}{\sigma^2_{\text{post}}} = \frac{1}{\sigma_0^2} + \frac{1}{\sigma^2} \qquad \mu_{\text{post}} = \sigma^2_{\text{post}}\left(\frac{\mu_0}{\sigma_0^2} + \frac{D}{\sigma^2}\right)$$

Intuition: precisions add. μ_{post} *= precision-weighted average.*

Numbers:

$\sigma_0^2 = 400$ (prior), $\sigma^2 = 900$ (data), $\mu_0 = 500$, $D = 510$

Plug in the numbers — posterior = prior × likelihood, rescaled (3/4)

The update formulas (from the derivation):

$$\frac{1}{\sigma^2_{\text{post}}} = \frac{1}{\sigma_0^2} + \frac{1}{\sigma^2} \qquad \mu_{\text{post}} = \sigma^2_{\text{post}}\left(\frac{\mu_0}{\sigma_0^2} + \frac{D}{\sigma^2}\right)$$

Intuition: precisions add. μ_{post} = precision-weighted average.

Numbers:

$\sigma_0^2 = 400$ (prior), $\sigma^2 = 900$ (data), $\mu_0 = 500$, $D = 510$

Posterior variance:

$$\frac{1}{\sigma^2_{\text{post}}} = \frac{1}{400} + \frac{1}{900} = \frac{9}{3600} + \frac{4}{3600} = \frac{13}{3600}$$

$\sigma^2_{\text{post}} \approx 277$, $\sigma_{\text{post}} \approx 16.6$

Tighter than both inputs (20 and 30). Adding precisions REDUCES uncertainty.

ChibaTech School of Design & Science

Plug in the numbers — posterior = prior × likelihood, rescaled (4/4)

The update formulas (from the derivation):

$$\frac{1}{\sigma^2_{\text{post}}} = \frac{1}{\sigma_0^2} + \frac{1}{\sigma^2} \qquad \mu_{\text{post}} = \sigma^2_{\text{post}}\left(\frac{\mu_0}{\sigma_0^2} + \frac{D}{\sigma^2}\right)$$

Intuition: precisions add. μ_{post} = precision-weighted average.

Numbers:

$\sigma_0^2 = 400$ (prior), $\sigma^2 = 900$ (data), $\mu_0 = 500$, $D = 510$

Posterior variance:

$$\frac{1}{\sigma^2_{\text{post}}} = \frac{1}{400} + \frac{1}{900} = \frac{9}{3600} + \frac{4}{3600} = \frac{13}{3600}$$

$\sigma^2_{\text{post}} \approx 277$, $\sigma_{\text{post}} \approx 16.6$

Tighter than both inputs (20 and 30). Adding precisions REDUCES uncertainty.

Posterior mean:

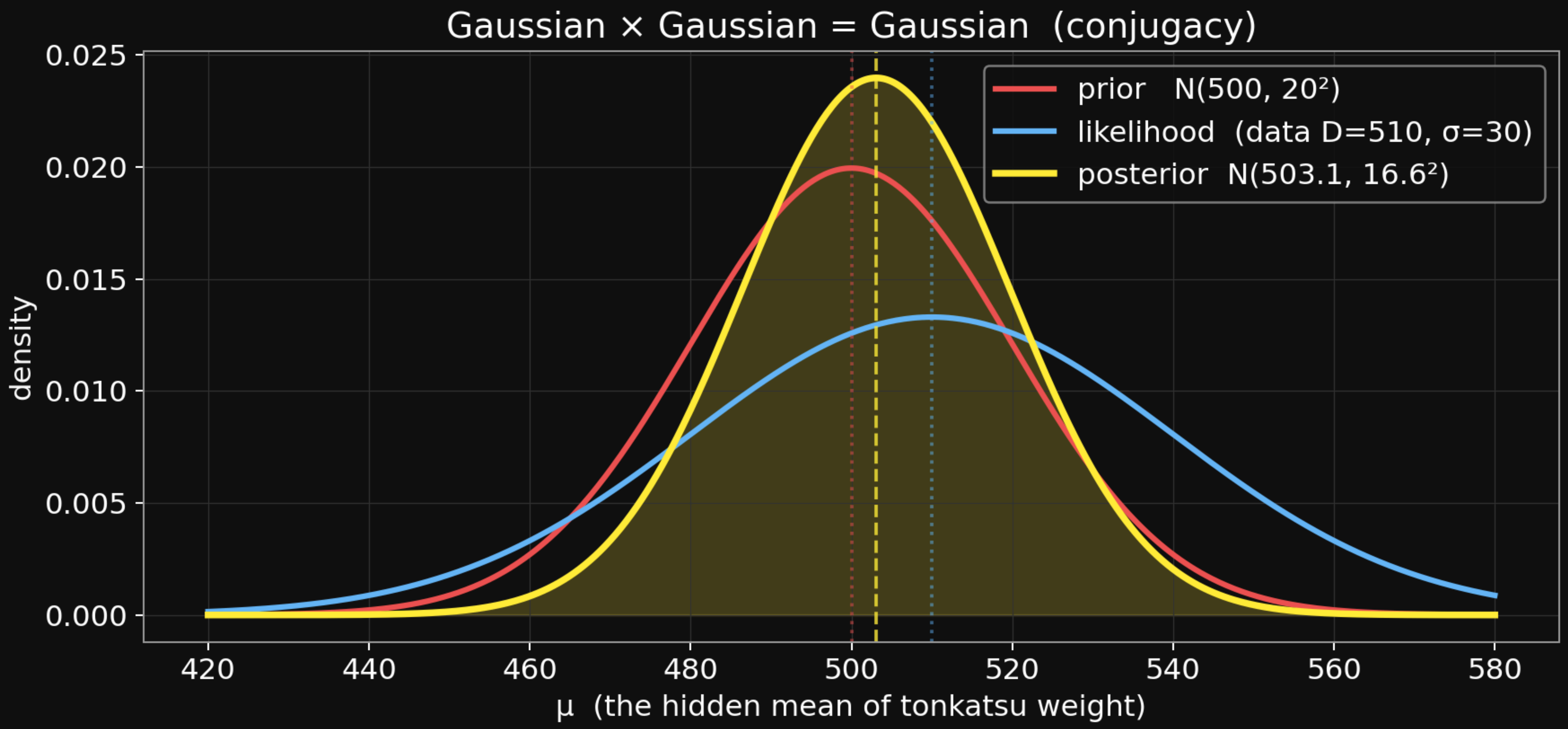
$$\mu_{\text{post}} = 277 \cdot \left(\frac{500}{400} + \frac{510}{900}\right) = 277 \cdot (1.250 + 0.567)$$

$$= 277 \cdot 1.817 \approx 503.1$$

Moved from prior (500) toward data (510), but stayed closer to prior

ChibaTech School of Design & Science

Prior × Likelihood = Posterior



Red = prior. Blue = likelihood. Yellow = posterior (tighter; between both).

N observations — precisions keep adding

One observation: precision added $= 1/\sigma^2$.

Two observations: precision added $= 2/\sigma^2$.

N observations: precision added $= N/\sigma^2$.

N observations — precisions keep adding (2/3)

One observation: precision added $= 1/\sigma^2$.

Two observations: precision added $= 2/\sigma^2$.

N observations: precision added $= N/\sigma^2$.

$$\frac{1}{\sigma_{\text{post}}^2} = \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2} \qquad \mu_{\text{post}} = \sigma_{\text{post}}^2 \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_i D_i}{\sigma^2} \right)$$

N observations — precisions keep adding (3/3)

One observation: precision added $= 1/\sigma^2$.

Two observations: precision added $= 2/\sigma^2$.

N observations: precision added $= N/\sigma^2$.

$$\frac{1}{\sigma_{\text{post}}^2} = \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2} \qquad \mu_{\text{post}} = \sigma_{\text{post}}^2 \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_i D_i}{\sigma^2} \right)$$

As N grows:

- $1/\sigma_{\text{post}}^2 \to \infty$ (posterior gets sharp)
- $\mu_{\text{post}} \to$ mean of observed weights (data dominates)

More data → sharper posterior, pulled toward sample mean.

Closing the arc — same Bayes, new objects

Block 4: two discrete variables. Joint is a 4-cell table; operations: marginalize, condition, check independence.

Block 6: H discrete, D continuous. Likelihood is a density; prior stays discrete.

Block 7: H continuous, D continuous. Prior, likelihood, posterior all densities. Updates via precision arithmetic.

Same Bayes rule throughout. The OBJECTS generalized.

Closing the arc — same Bayes, new objects (2/2)

Block 4: two discrete variables. Joint is a 4-cell table; operations: marginalize, condition, check independence.

Block 6: H discrete, D continuous. Likelihood is a density; prior stays discrete.

Block 7: H continuous, D continuous. Prior, likelihood, posterior all densities. Updates via precision arithmetic.

Same Bayes rule throughout. The OBJECTS generalized.

The named move: CONJUGACY.

Gaussian × Gaussian = Gaussian (today)

Beta × Binomial = Beta (Week 3)

Dirichlet × Multinomial = Dirichlet (Week 3)

Conjugacy = algebra works out; the posterior stays in the prior's family.

Agenda

Welcome + meet Chibany	0:00
Marr’s three levels	0:08
Notation lock-in	0:15
Joint, marginal, conditional, independence	0:18
Expected value + discrete	0:30
Continuous prob + Gaussian	0:45
Break	1:05
Gaussian-Gaussian update	1:15

Paper presentations — rubric + signup

Presentations — why and what

Enrollment is 6. Each of you presents ONCE.

20 min presentation + 10-15 min discussion you facilitate.

7.5% of grade. Reflections dropped 15% → 12.5% to make room.

Meet with me in office hours the week before your slot.

Rubric — 5 points

Understanding of the paper 1.5

Covering key aspects 1.5

Presentation clarity 1.0

Discussion questions (at least 3) 0.5

Appropriate use of time 0.5

Focus: how the math connects to cognitive science.

Full guidelines: resources/classPresentationGuidelines.pdf (emailed today).

Signup — Weeks 4 through 12

9 slots, 6 students. Three stay open for contemporary-ML content.

- Week 4 (May 22) — Generalization + hierarchical Bayes
- Week 5 (May 29) — Bayes nets + causal Bayes nets
- Week 6 (Jun 5) — Markov chains + networks
- Week 7 (Jun 12) — Monte Carlo
- Week 8 (Jun 19) — SDT / MDP / RL
- Week 9 (Jun 26) — Inverse RL
- Week 10 (Jul 3) — Bayesian nonparametrics
- Week 11 (Jul 10) — Deep NNs
- Week 12 (Jul 17) — Ethics + adversarial ML + project presentations

Agenda

Welcome + meet Chibany	0:00
Marr’s three levels	0:08
Notation lock-in	0:15
Joint, marginal, conditional, independence	0:18
Expected value + discrete	0:30
Continuous prob + Gaussian	0:45
Break	1:05
Gaussian-Gaussian update	1:15

Admin + Week 3 homework

Syllabus — three things changed

① 12 sessions, not 13.

No class May 1 AND May 8. Week 3 is Friday May 15.

② Reflections: 6 of 12, not 8 of 13.

~200 words, pre-class, pass/fail.

③ Paper presentations exist → 7.5%.

Reflections dropped 15% → 12.5% to make room.

Week 3 — readings + homework

Next class: Friday May 15 — Week 3: Conjugate Bayes + Topics.

Required reading:

- Textbook T3 Ch 1 (Chibany’s mystery bentos — the full story)
- Textbook T3 Ch 4 (Bayesian learning with Gaussians)

  — formalizes what we did in Block 7.

Ungraded self-check: “Intro Probability Theory 1” quiz.

Preview — the arc from here

Week 3: Conjugacy proper. Beta-Binomial + Gaussian-Gaussian.

Week 4: Hierarchical Bayes. Chibany’s distribution of tonkatsu rates.

Week 5: Bayes nets + causal.

Weeks 6-7: Markov chains, Monte Carlo (sampling-based L2).

Weeks 8-10: Decision, RL, IRL, nonparametrics.

Weeks 11-12: Deep NNs, ethics — contemporary ML.

Thanks — see you May 15.