BEHAVIORAL ECONOMICS · LECTURE 7

Bayesian updating and beliefs

Niels Boissonnet – Bielefeld University

Master Economics – Winter 2026/27

By the end of this lecture you will be able to

1. Formalise a **signal** in three equivalent ways and translate between them.
2. Apply **Bayes' rule** correctly — and recognise base-rate neglect when you see it.
3. State and prove that posteriors form a **martingale**, and characterise which distributions over posteriors a signal can induce.
4. Compute **Gaussian posteriors** in closed form, in one and several dimensions.
5. Use these tools to solve three **economic models**: career concerns, coordination under private information, and data sharing.
6. Explain why three natural **fairness criteria** for a predictor cannot hold simultaneously.

Why Bayesian updating now?

Looking back

Lectures 5–6: the decision maker does not look at every option (search, consideration sets).

Looking ahead

Salience (Lecture 8) and rational inattention (Lecture 9) treat attention as **information processing**: what you learn, and how you update, depends on what you attend to. Part 3 then asks whether beliefs are **correct** (Assumption 2).

Today is the **benchmark**: how a perfectly rational agent revises beliefs after receiving information, and what every posterior must satisfy. Every behavioral model of beliefs is a departure from this.

Five parts, two sessions

The framework (§2.1)
Posterior beliefs (§2.2)
Application: incompatible fairness definitions (§2.2.3)
Gaussian information (§2.3.1)
Three applications (§2.3.2–2.3.4)
Closing48

Prerequisites

Conditional probability, expectations, and the normal distribution — each is recalled on a *Refresher* slide right before it is used.

Exercise labels

(U) advanced-undergraduate level, **(G)** graduate level, following the notes. Exercises 2.1–2.3 are solved in class.

Suggested split

Session 1: Parts 1–3. Session 2: Parts 4–5.

1

The framework (§2.1)

Parameters, priors, and three equivalent ways to say what a signal is.

The primitives: a parameter space and a prior

Setup

A set of **parameters** Θ (the unknown "state of the world"), endowed with a σ-algebra Σ. An agent holds a **prior**

$$p \in \Delta(\Theta),$$

where $\Delta(\Theta)$ is the set of Σ-measurable probability measures on Θ.

How to read the prior

- It is the agent's belief at an **ex-ante** stage — before any information arrives.
- "Ex ante" is relative to a model: the prior at the start of one model may be a posterior from an earlier one.
- Everything the chapter does is: *take p, add information, produce a new belief in $\Delta(\Theta)$.*

The object of study is the **information structure** — also called an **experiment** or a **signal**.

Θ, θ	$p \in \Delta(\Theta)$	S, s	X, x	$\sigma(s \mid \theta)$	q, τ
parameter space, parameter	prior	realisations	signal r.v., realisation	$P(s \mid \theta)$	posterior; law of posteriors

What is a signal? Three equivalent formalisms

(a) A kernel

$$\sigma : \Theta \to \Delta(S)$$

A map from each parameter θ to a **distribution over signal realisations** S. Think "experiment": nature draws s from $\sigma(\cdot \mid \theta)$.

de Oliveira (2018)

(b) A random variable

$$X : \Omega \to S, \quad \Omega = \Theta \times E$$

An S-valued random variable on an enlarged probability space (Ω, Σ, P). The extra coordinate E carries the **noise**, e.g. $X = \theta + \varepsilon$ with $\varepsilon \perp\!\!\!\perp \theta$.

Used in §2.3 (Gaussian signals)

(c) A partition

$$P(s \mid \theta) = \mathrm{Leb}\{x : (\theta, x) \in s\}$$

A finite partition S of $\Theta \times [0,1]$ into measurable cells. Given θ, the probability of cell s is the **Lebesgue measure** of the x's landing in s.

Frankel and Kamenica (2019)

All three describe the same thing: **a joint distribution of (parameter, signal)** whose θ-marginal is the prior. (c) is a special case of (b) with $E = [0,1]$ and P = prior $\times$ Lebesgue; (a) and (b) nest each other — next slide.

Why (a) and (b) are the same object (finite case)

From a kernel to a random variable

Given a prior p and a kernel $\sigma : \Theta \to \Delta(S)$, expand the state space:

$$\Omega = \Theta \times S, \qquad X(\theta, s) = s,$$
$$P(\theta, s) = p(\theta)\,\sigma(s \mid \theta).$$

Conditioning on the realisation of X gives the same posterior about θ as conditioning on the realisation of $\sigma(\theta)$.

From a random variable to a kernel

Given $X : \Theta \times E \to S$ and a distribution P on $\Theta \times E$, read off the conditional law of the signal:

$$\sigma(s \mid \theta) = P(X^{-1}(s) \mid \theta),$$

the P-probability, given θ, of the noise draws that produce s.

Practical rule. Whenever Θ and S are finite, a signal is a $|\Theta| \times |S|$ **stochastic matrix**: one row per parameter, each row a probability distribution over realisations.

Three families of signals you will meet again

Partitional (Aumann, 1976)

Agent i has a finite partition Π_i of Θ into cells of positive measure. The signal reveals, with probability 1, **which cell** contains θ. Deterministic: every row of the matrix is a unit vector. This is the model of Chapter 1 of the notes.

Finite information structure

A drug is good (g) or bad (b); the patient is cured (C) or not (N). Cure probability is $3/4$ if good, $1/4$ if bad:

	C	N
g	$3/4$	$1/4$
b	$1/4$	$3/4$

Rows: parameters; columns: realisations; rows sum to 1.

Gaussian information

The signal is the parameter plus independent normal noise:

$$X = \theta + \varepsilon,$$

$\theta \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$, $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, $\theta \perp\!\!\!\perp \varepsilon$. Infinite Θ and S, but — Part 4 — posteriors in **closed form**.

Check

In the drug example, what are Θ, S and σ? Which formalism is the table — (a), (b) or (c)?

2

Posterior beliefs (§2.2)

Bayes' rule, the base-rate trap, and two properties every posterior must satisfy: beliefs are a martingale, and any Bayes-plausible distribution over posteriors can be induced.

Three facts we will use constantly

1 · Conditional probability

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) > 0.$$

"The share of B-worlds that are also A-worlds." Defined only when $P(B) > 0$ — which is why the notes assume every signal realisation has positive probability.

2 · Multiplication rule & total probability

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

$$P(B) = \sum_i P(B \mid A_i)\,P(A_i) \quad \text{for a partition } \{A_i$$

Two ways of writing one joint probability; the second line sums over a partition.

3 · Law of iterated expectations

$$\mathbb{E}\big[\mathbb{E}(Y \mid X)\big] = \mathbb{E}(Y).$$

For an indicator $Y = \mathbf{1}(A)$: *the average of the conditional probabilities of A equals $P(A)$.*

Bayes' rule = facts 1 + 2.
The martingale property of beliefs = fact 3.
Part 2 needs nothing more.

Bayes' rule

Definition 2.1 — Bayes' rule, finite case

Suppose $|\Theta| < \infty$. Fix any $P \in \Delta(\Theta)$ and events $A, B \subseteq \Theta$ with $P(A), P(B) > 0$. Then

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}. \qquad (2.1)$$

Don't memorise it — derive it (Remark 2.2)

Write the joint probability of A and B in both orders,

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

and divide by $P(B)$. That is all Bayes' rule is: the two factorisations of a joint probability.

In our language: A is a statement about the parameter, B the signal realisation. Bayes' rule converts "how likely is the signal given the parameter" (which the model gives us) into "how likely is the parameter given the signal" (which the agent wants).

Bayes' rule in odds form: what a signal actually does

Apply (2.1) twice, to the pairs (A, E) and (B, E), and divide. The normalising constant $P(E)$ cancels:

$$\underbrace{\frac{P(A \mid E)}{P(B \mid E)}}_{\text{posterior odds}} = \underbrace{\frac{P(E \mid A)}{P(E \mid B)}}_{\text{likelihood ratio}} \cdot \underbrace{\frac{P(A)}{P(B)}}_{\text{prior odds}}$$

Prior odds

How much more likely A was than B *before* the evidence. Comes from the prior only.

Likelihood ratio

How much more likely the evidence E is under A than under B. Comes from the signal only. This is the **entire** contribution of the signal.

Posterior odds

Their product. A likelihood ratio of 19 does **not** mean A is 19 times more likely than B — only 19 times more likely *than it was*.

A positive test for a rare condition

A condition affects 1% of the population. A test is positive for 95% of those who have it and 5% of those who do not:

$$P(\omega = p) = 0.01,$$
$$P(X = + \mid \omega = p) = 0.95,$$
$$P(X = + \mid \omega = n) = 0.05.$$

The tempting reasoning

"A positive result is 19 times more likely if the person has the condition — so they almost surely have it."

$$\frac{P(X = + \mid \omega = p)}{P(X = + \mid \omega = n)} = \frac{0.95}{0.05} = 19.$$

The correct reasoning: multiply by the prior odds

$$\frac{P(\omega = p \mid X = +)}{P(\omega = n \mid X = +)} = 19 \times \frac{0.01}{0.99} \approx 0.19 < 1,$$

$$P(\omega = p \mid X = +) = \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.05 \cdot 0.99} \approx 0.$$

Even after a positive test, the person is **more likely not to have** the condition: the 1% base rate dominates the 19-to-1 likelihood ratio.

Base-rate neglect

$=$ equating the posterior ratio $P(A \mid E)/P(B \mid E)$ with the likelihood ratio $P(E \mid A)/P(E \mid B)$, i.e. silently setting the prior odds to 1.

The posterior belief: three levels of generality

Finite Θ — Bayes' rule with $A = \{\theta\}$, $B = \{X = x\}$; the denominator is total probability

$$P(\theta \mid X = x) = \frac{P(X = x \mid \theta)\, P(\theta)}{\sum_{\theta' \in \Theta} P(X = x \mid \theta')\, P(\theta')} \qquad \forall\, \theta \in \Theta. \tag{2.2}$$

Densities — sums become integrals (Part 4)

$$f_{\theta \mid X = x}(t) = \frac{f_{X \mid \theta = t}(x)\, f_\theta(t)}{\int_\Theta f_{X \mid \theta = t'}(x)\, f_\theta(t')\, \mathrm{d}t'}. \tag{2.3}$$

General — whatever the joint law of (θ, X), the posterior after x is the measurable map q_x with

$$q_x(A) = \mathbb{E}(\mathbf{1}_A \mid X = x) \qquad \text{for all events } A \subseteq \Theta.$$

In every case the conditional distribution $P(\cdot \mid X = x)$ **is** the posterior after observing $X = x$; we write it $q_x \in \Delta(\Theta)$.

Worked example: a binary signal about a binary state

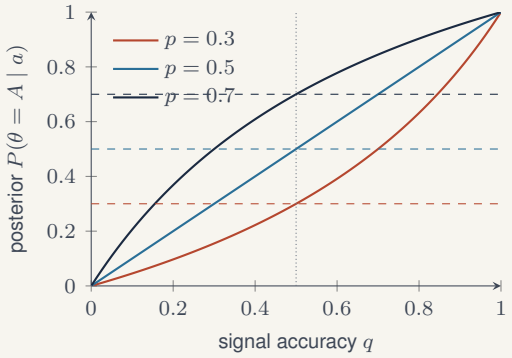
A drug is effective ($\theta = A$) with prior probability $p \in (0,1)$, otherwise not ($\theta = B$). A test returns a or b with accuracy $q \in (0,1)$:

	a	b
A	q	$1-q$
B	$1-q$	q

After observing a, Bayes' rule gives

$$P(\theta = A \mid a) = \frac{pq}{pq + (1-p)(1-q)} = \frac{1}{1 + \frac{1-p}{p}\,\frac{1-q}{q}},$$

$$P(\theta = A \mid a) \;\geq\; p \quad\Longleftrightarrow\quad q \;\geq\; \tfrac{1}{2}.$$



The signal a raises the belief in A exactly when $q > \frac{1}{2}$; at $q = \frac{1}{2}$ it is pure noise. Curves are steeper near $p = \frac{1}{2}$, where there is more to learn.

The distribution over posteriors induced by a signal

The state space is $\Theta = \{\theta_1, \theta_2\}$ and the prior is $(\mu, 1-\mu)$, $\mu \in [0,1]$. The signal structure is

	s_1	s_2
θ_1	p	$1-p$
θ_2	q	$1-q$

$p, q \in [0,1]$.

Your task

1. Compute the probability of each signal realisation.
2. Compute the posterior probability of θ_1 after each realisation.
3. Write down the induced **distribution over posterior beliefs**.
4. Verify that the **expected posterior equals the prior**.

Hint: a "distribution over posteriors" lists each possible posterior with the probability that it is reached.

Two posteriors, weighted by the signal probabilities

After s_1

$$P(s_1) = \mu p + (1-\mu)q, \qquad P(\theta_1 \mid s_1) = \frac{\mu p}{\mu p + (1-\mu)q}.$$

After s_2

$$P(s_2) = \mu(1-p) + (1-\mu)(1-q), \quad P(\theta_1 \mid s_2) = \frac{\mu(1-p)}{\mu(1-p) + (1-\mu)(1-q)}.$$

The induced distribution over posteriors

The posterior on θ_1 equals $P(\theta_1 \mid s_1)$ with probability $P(s_1)$ and $P(\theta_1 \mid s_2)$ with probability $P(s_2)$.

Expected posterior = prior

$$P(s_1)P(\theta_1 \mid s_1) + P(s_2)P(\theta_1 \mid s_2)$$
$$= \mu p + \mu(1-p) = \mu. \checkmark$$

The denominators cancel — a general fact, next slide.

Beliefs are a martingale: the expected posterior is the prior

Fact 2.1

Let $p \in \Delta(\Theta)$ be the prior and A any event. The posterior probability of A given X is $\mathbb{E}(\mathbf{1}_A \mid X)$. By the law of iterated expectations,

$$\mathbb{E}\big[\mathbb{E}(\mathbf{1}_A \mid X)\big] = \mathbb{E}(\mathbf{1}_A) = P(A).$$

Since A was arbitrary, **the expected posterior belief equals the prior**. For finite Θ, taking $A = \{\theta\}$:

$$\mathbb{E}\big[p(\theta \mid X)\big] = p(\theta) \quad \forall \theta \in \Theta.$$

The one-line intuition

"If you expect to change your mind after seeing more information, why haven't you changed it already?"

What it does and does not say

- It constrains the **average** posterior, not any single posterior.
- Updates are **unpredictable** on average: a Bayesian cannot expect to revise upward.
- Any signal X induces a distribution $\tau \in \Delta(\Delta(\Theta))$ over posteriors; Fact 2.1 says τ must **average to** p. Next: this is the *only* constraint.

Bayes plausibility: a condition on distributions over posteriors

Definition 2.2

Fix a prior $p \in \Delta(\Theta)$. A distribution of posteriors $\tau \in \Delta(\Delta(\Theta))$ is **Bayes plausible** if

$$\int_{\Delta(\Theta)} q \, \mathrm{d}\tau(q) = p,$$

i.e. the expected posterior equals the prior. The set of Bayes-plausible distributions given p is

$$\mathcal{T}(p) \equiv \left\{ \tau \in \Delta(\Delta(\Theta)) : \int q \, \mathrm{d}\tau(q) = p \right\}.$$

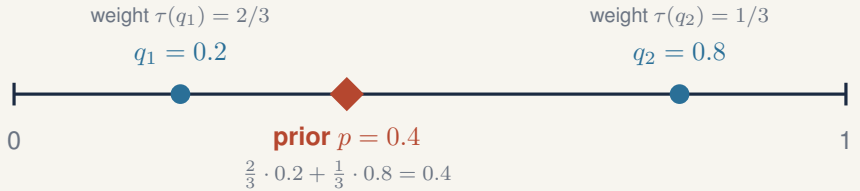
Reading the objects

- $q \in \Delta(\Theta)$ is a **posterior**.
- $\tau \in \Delta(\Delta(\Theta))$ is a **distribution over posteriors**: which posteriors arise, and how often.
- $\int q \, \mathrm{d}\tau(q) \in \Delta(\Theta)$ is the τ-average of the posteriors, coordinate by coordinate.

Fact 2.1 says: **every signal induces a Bayes-plausible** τ. The converse (next slides): every Bayes-plausible τ comes from some signal — so we may think in terms of posteriors instead of signals.

The prior is the centre of mass of the posteriors

With $\Theta = \{\theta_1, \theta_2\}$ a belief is a number $q = P(\theta_1) \in [0, 1]$. Take prior $p = 0.4$ and a signal leading to posterior $q_1 = 0.2$ with probability $2/3$ and to $q_2 = 0.8$ with probability $1/3$:



Bayes plausible

The weighted average of the posteriors is p — the prior "balances" on the posteriors: $\tau(q_1)q_1 + \tau(q_2)q_2 = p$.

Not plausible

Posteriors 0.2 and 0.8 with weights $\frac{1}{2}, \frac{1}{2}$ — their average is $0.5 \neq 0.4$. No signal can produce this from prior 0.4.

The converse: every Bayes-plausible τ is induced by a signal

Definition 2.3 — induced distribution

For a signal X with law P_X, let τ_X be the law of the posterior:

$$\tau_X(q) = P_X(\{x : q_x = q\}).$$

τ is **induced by** X if $\tau = \tau_X$.

Proposition 3

Suppose the prior p is in the interior of $\Delta(\Theta)$. Then every Bayes-plausible $\tau \in \mathcal{T}(p)$ is induced by some signal X.

Kamenica and Gentzkow (2011); Shmaya and Yariv (2016)

Proof by construction — let the signal *be* the posterior

Index the posteriors in the support of τ as $\{q_x\}_{x \in \mathcal{X}}$. Define a signal whose realisation x "announces" q_x, by reverse-engineering Bayes' rule, $\sigma(x \mid \theta) = P(\theta \mid x)P(x)/P(\theta)$:

$$\sigma(x \mid \theta) = \frac{q_x(\theta)\,\tau(q_x)}{p(\theta)}. \tag{2.4}$$

It is a valid signal — each row integrates to 1 by Bayes plausibility:

$$\int_{\mathcal{X}} \sigma(x \mid \theta)\,\mathrm{d}x = \frac{1}{p(\theta)} \int q_x(\theta)\,\mathrm{d}\tau(q_x) = 1.$$

It induces the right posterior — updating to x returns q_x:

$$\frac{\sigma(x \mid \theta)\,p(\theta)}{\int_\Theta \sigma(x \mid \theta')\,p(\theta')\,\mathrm{d}\theta'} = \frac{q_x(\theta)\,\tau(q_x)}{\tau(q_x)\int_\Theta q_x(\theta')\,\mathrm{d}\theta'} = q_x(\theta).$$

And $P(x) = \int \sigma(x \mid \theta)p(\theta)\,\mathrm{d}\theta = \tau(q_x)$: q_x arises with exactly the frequency τ prescribes. □

Signals and Bayes-plausible distributions over posteriors are the same thing

$$\tau \text{ is induced by some signal} \iff \tau \in \mathcal{T}(p)$$

for any prior p in the interior of $\Delta(\Theta)$. One direction is Fact 2.1 (the martingale property), the other is Proposition 3 (the construction).

Why economists care

Signals are messy objects (kernels, matrices, noise). Distributions over posteriors are clean: a single linear constraint. Problems about *which signal to use* become problems about *which distribution over posteriors to pick*.

Where it leads

Bayesian persuasion (Chapter 8 of the notes) and the comparison of information structures (Chapters 4–5) both work directly in $\mathcal{T}(p)$. The only constraint a designer faces is Bayes plausibility.

Build the signal that induces a given distribution over posteriors

The prior over $\Theta = \{\theta_1, \theta_2\}$ is $(1/3, 2/3)$. Provide a set S and a signal structure $\sigma : \Theta \to \Delta(S)$ that induces

the belief $(0, 1)$ with probability $1/3$, and
the belief $(1/2, 1/2)$ with probability $2/3$.

Suggested steps

1. Check that the target τ is Bayes plausible — otherwise no signal exists.
2. Let S have one realisation per posterior: $x_1 \mapsto (0, 1)$, $x_2 \mapsto (1/2, 1/2)$.
3. Use $\sigma(x \mid \theta) = q_x(\theta)\tau(q_x)/p(\theta)$ to fill the 2×2 matrix.
4. Verify by applying Bayes' rule to your matrix.

Reverse-engineering Bayes' rule

1 · Plausibility

$$\tfrac{1}{3}(0,1) + \tfrac{2}{3}\left(\tfrac{1}{2},\tfrac{1}{2}\right) = \left(\tfrac{1}{3},\tfrac{2}{3}\right) = p. \checkmark$$

2–3 · The construction, entry by entry

With $q_{x_1} = (0,1)$, $\tau(q_{x_1}) = \frac{1}{3}$, $q_{x_2} = (\frac{1}{2},\frac{1}{2})$, $\tau(q_{x_2}) = \frac{2}{3}$:

$$\sigma(x_1 \mid \theta_1) = \frac{0 \cdot \frac{1}{3}}{\frac{1}{3}} = 0, \quad \sigma(x_1 \mid \theta_2) = \frac{1 \cdot \frac{1}{3}}{\frac{2}{3}} =$$

and rows sum to one:

	x_1	x_2
θ_1	0	1
θ_2	1/2	1/2

4 · Verify with Bayes' rule

After x_1, θ_1 is impossible, so the posterior is $(0,1)$. After x_2:

$$P(\theta_1 \mid x_2) = \frac{\frac{1}{3} \cdot 1}{\frac{1}{3} \cdot 1 + \frac{2}{3} \cdot \frac{1}{2}} = \frac{1}{2},$$

$$P(x_1) = \tfrac{1}{3} \cdot 0 + \tfrac{2}{3} \cdot \tfrac{1}{2} = \tfrac{1}{3}. \checkmark$$

Interpretation: the signal **reveals θ_2 half of the time** it occurs, and otherwise stays silent (x_2), leaving the two states equally likely.

3

Application: incompatible fairness definitions (§2.2.3)

A detour to see the power of Bayes' rule: three natural fairness requirements on a predictor cannot all hold unless the groups have the same base rate.

Scoring individuals who belong to two groups

Each individual is described by three things:

Symbol	Meaning	Credit example
$C \in \mathcal{C}$	covariates (observable)	credit history
$G \in \{g_1, g_2\}$	group membership	a demographic group
$\theta \in \{0, 1\}$	type (unobserved)	would repay the loan

The triple (C, G, θ) is drawn from P. Each group has a **base rate** of good types, and a **scoring rule** predicts the type from the covariates:

$$p_g = P(\theta = 1 \mid G = g), \qquad S : \mathcal{C} \to \{0, 1\}.$$

The question

What should it mean for the score S to treat the two groups **fairly**? Three answers are natural, and each is used in practice.

Where this comes from

The debate over risk scores in criminal justice and lending. Kleinberg et al. (2017) and Chouldechova (2017) showed the three criteria are incompatible. The proof is Bayes' rule.

Three fairness criteria

Equal false positives (Definition 2.4) $S \perp\!\!\!\perp G \mid \theta = 0$

$$P(S = 1 \mid \theta = 0, G = g_1) = P(S = 1 \mid \theta = 0, G = g_2)$$

Among those who would **not** repay, the chance of being wrongly approved is the same in both groups.

Equal false negatives (Definition 2.5) $S \perp\!\!\!\perp G \mid \theta = 1$

$$P(S = 0 \mid \theta = 1, G = g_1) = P(S = 0 \mid \theta = 1, G = g_2)$$

Among those who **would** repay, the chance of being wrongly rejected is the same in both groups.

Calibration (Definition 2.6) $\theta \perp\!\!\!\perp G \mid S$

$$P(\theta = 1 \mid S = s, G = g_1) = P(\theta = 1 \mid S = s, G = g_2) \quad \forall s \in \{0, 1\}$$

Among those with a given score, the share who repay is the same in both groups: the score **means the same thing** for everyone.

The impossibility result and the identity behind it

Proposition 4 — Kleinberg et al. (2017); Chouldechova (2017)

Suppose the base rates differ, $p_{g_1} \neq p_{g_2}$. Then **no imperfect scoring rule** S can simultaneously satisfy calibration, equal false-positive rates and equal false-negative rates.

For each group g write the false-positive rate, the false-negative rate and the **positive predictive value**:

$$FP_g = P(S{=}1 \mid \theta{=}0, G{=}g), \quad FN_g = P(S{=}0 \mid \theta{=}1, G{=}g), \quad PPV_g = P(\theta{=}1 \mid S{=}1, G{=}g).$$

The key identity (2.5)

$$FP_g = \frac{p_g}{1-p_g} \times \frac{1-PPV_g}{PPV_g} \times (1-FN_g).$$

Equal FP, equal FN and calibration (equal PPV) pin down everything in (2.5) **except** the base rate — so the base rates must be equal too. The identity itself is just Bayes' rule, next slide.

Proof: the identity is Bayes' rule rearranged

Fix a group g and let Q be the joint distribution of (C, θ, S) conditional on $G = g$.

1. Expand each factor of (2.5):

$$Q(S{=}1 \mid \theta{=}0) = \frac{Q(\theta{=}1)}{Q(\theta{=}0)} \times \frac{Q(\theta{=}0 \mid S{=}1)}{Q(\theta{=}1 \mid S{=}1)} \times Q(S{=}1 \mid \theta{=}1).$$

2. Multiply both sides by $Q(\theta = 0)$ and use $Q(S{=}1 \mid \theta)Q(\theta) = Q(S{=}1, \theta)$. The claim becomes

$$\frac{Q(S{=}1, \theta{=}0)}{Q(S{=}1, \theta{=}1)} = \frac{Q(\theta{=}0 \mid S{=}1)}{Q(\theta{=}1 \mid S{=}1)}. \tag{2.6}$$

3. Bayes' rule again on the right-hand side — the $Q(S = 1)$ cancels:

$$\frac{Q(\theta{=}0 \mid S{=}1)}{Q(\theta{=}1 \mid S{=}1)} = \frac{Q(\theta{=}0, S{=}1)/Q(S{=}1)}{Q(\theta{=}1, S{=}1)/Q(S{=}1)} = \frac{Q(S{=}1, \theta{=}0)}{Q(S{=}1, \theta{=}1)}.$$

4. So (2.5) is an identity, in both groups. Impose the three criteria:

$$FP_{g_1} = FP_{g_2}, \quad FN_{g_1} = FN_{g_2}, \quad PPV_{g_1} = PPV_{g_2} \quad \Longrightarrow \quad p_{g_1} = p_{g_2}.$$

Contrapositive: if the base rates differ, at least one criterion fails. The only inputs were the multiplication rule and Bayes' rule. □

4

Gaussian information (§2.3.1)

When the prior and the noise are normal, the posterior is normal too — and its mean and variance have formulas you can carry in your head.

What we need about Gaussians

Density and precision

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

$\mathcal{N}(\mu, \sigma^2)$ is pinned down by mean and variance. Its **precision** is $\tau = 1/\sigma^2$: high precision means a tight distribution, i.e. a lot of information.

Linear combinations stay normal

For independent X, Y,

$$aX + bY \sim \mathcal{N}(m, s^2),$$

$$m = a\mu_X + b\mu_Y,$$

$$s^2 = a^2\sigma_X^2 + b^2\sigma_Y^2.$$

So $X = \theta + \varepsilon$ is normal with mean μ and variance $\sigma_\theta^2 + \sigma_\varepsilon^2$. Means add; **variances** add (not standard deviations).

Jointly normal vectors

(Z_1, Z_2) is **jointly normal** if every linear combination of its coordinates is normal; it is described by a mean vector and a covariance matrix Σ. Two facts: **uncorrelated** $\Rightarrow$ **independent**, and **conditioning on one block leaves the other block normal** (Fact 2.3).

Everything in Part 4 follows from one computation: write the posterior density as prior $\times$ likelihood, complete the square in θ, read off the mean and variance.

The Gaussian updating formulae

Setup — the simplest case

$$\theta \sim \mathcal{N}(\mu, \sigma_\theta^2), \qquad X = \theta + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2), \qquad \theta \perp\!\!\!\perp \varepsilon, \qquad \sigma_\theta^2, \sigma_\varepsilon^2 > 0.$$

Fact 2.2 — the posterior is normal, with

$$\mathbb{E}(\theta \mid X = x) = \Big(\frac{\sigma_\varepsilon^2}{\sigma_\theta^2 + \sigma_\varepsilon^2}\Big)\mu + \Big(\frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\varepsilon^2}\Big)x,$$

$$\mathrm{Var}(\theta \mid X = x) = \frac{\sigma_\theta^2\, \sigma_\varepsilon^2}{\sigma_\theta^2 + \sigma_\varepsilon^2}.$$

Sanity checks

- $\sigma_\varepsilon^2 \to 0$ (perfect signal): mean $\to x$, variance $\to 0$. You learn θ.
- $\sigma_\varepsilon^2 \to \infty$ (useless signal): mean $\to \mu$, variance $\to \sigma_\theta^2$. Nothing changes.
- The posterior variance is **smaller than both** σ_θ^2 and σ_ε^2.

Three properties worth memorising, and the precision form

1 · Convex combination

The posterior mean is a weighted average of the prior mean μ and the signal x, with weights **proportional to the precisions**: the more precise source gets more weight.

2 · Constant variance

The posterior variance depends on the two variances but **not on the realisation** x. How much you learn is known before the signal; only where you end up is random.

3 · Precisions add

The posterior precision is the prior precision plus the signal precision. Each independent normal signal adds its precision to what you already have.

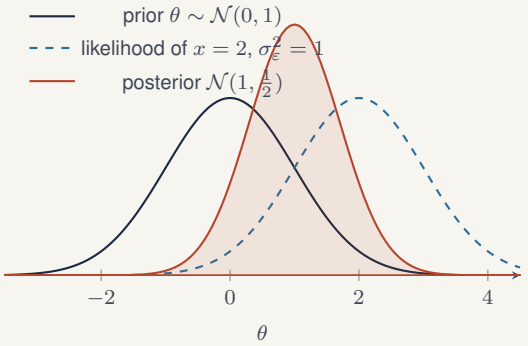
Fact 2.2 in precision form, $\tau_\theta = 1/\sigma_\theta^2$, $\tau_\varepsilon = 1/\sigma_\varepsilon^2$

$$(\theta \mid X = x) \sim \mathcal{N}\Big(\frac{\tau_\theta}{\tau_\theta + \tau_\varepsilon}\,\mu + \frac{\tau_\varepsilon}{\tau_\theta + \tau_\varepsilon}\,x,\ \frac{1}{\tau_\theta + \tau_\varepsilon}\Big).$$

Generalisation. With n independent signals $X_i = \theta + \varepsilon_i$ of precisions τ_i:

$$\theta \mid X_1, \ldots, X_n \sim \mathcal{N}\Big(\frac{\tau_\theta \mu + \sum_i \tau_i x_i}{\tau_\theta + \sum_i \tau_i},\ \frac{1}{\tau_\theta + \sum_i \tau_i}\Big).$$

Prior $\times$ likelihood $=$ posterior, in a numerical example



The numbers

Prior $\theta \sim \mathcal{N}(0,1)$; noise variance 1; observed $x = 2$. Both precisions equal 1, so the weights are $\frac{1}{2}, \frac{1}{2}$: mean $= \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 2 = \mathbf{1}$; variance $= \frac{1 \cdot 1}{1+1} = \mathbf{\frac{1}{2}}$; precision $= 1 + 1 = \mathbf{2}$.

What to see

- The posterior sits **between** prior mean and signal.
- It is **narrower** than either curve.
- Doubling the signal precision would pull the mean to $4/3$ and shrink the variance to $1/3$.

Before the signal arrives: the distribution of the posterior mean

Exercise 2.4 (U) — result

Write the posterior as $\mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$, where $\hat{\mu}$ depends on the realisation of X and is random *ex ante*. Then

$$\hat{\mu} \sim \mathcal{N}\Big(\mu,\ \sigma_\theta^2 - \frac{\sigma_\theta^2 \sigma_\varepsilon^2}{\sigma_\theta^2 + \sigma_\varepsilon^2}\Big).$$

The posterior mean is centred on the prior mean (the martingale property again) with variance equal to the **prior variance minus the posterior variance**.

Why — the law of total variance (Remark 2.4)

$$\mathrm{Var}\left(\mathbb{E}[\theta \mid X]\right) = \mathrm{Var}(\theta) - \mathbb{E}\big[\,\mathrm{Var}(\theta \mid X)\big].$$

Holds for **any** (θ, X). In the Gaussian case the posterior variance is a constant, so its expectation is itself.

Uncertainty about θ splits in two: the part the signal will **resolve** (variance of the posterior mean) and the part that will **remain** (posterior variance). A more informative signal moves the posterior mean around **more**.

Two signals of different precision

Question

Let $\theta \sim \mathcal{N}(\mu, 1)$ and define $Y_1 = \theta + \varepsilon_1$, $\varepsilon_1 \sim \mathcal{N}(0,1)$, and $Y_2 = \theta + \varepsilon_2$, $\varepsilon_2 \sim \mathcal{N}(0,2)$, all independent. (a) Find $\theta \mid Y_1 = y$ and $\theta \mid Y_2 = y$. (b) For which (μ, y) is $\mathbb{E}(\theta \mid Y_1 = y) > \mathbb{E}(\theta \mid Y_2 = y)$? Give the intuition.

(a) Apply Fact 2.2 twice: weights $\frac{1}{2}, \frac{1}{2}$ for Y_1; $\frac{2}{3}, \frac{1}{3}$ for Y_2

$$\theta \mid Y_1 = y \sim \mathcal{N}\Big(\frac{\mu + y}{2}, \frac{1}{2}\Big), \qquad \theta \mid Y_2 = y \sim \mathcal{N}\Big(\frac{2\mu + y}{3}, \frac{2}{3}\Big).$$

(b) Compare the means

$$\frac{\mu + y}{2} > \frac{2\mu + y}{3} \iff 3\mu + 3y > 4\mu + 2y \iff y > \mu$$

Intuition

Both posteriors move from μ toward y; the **more precise** signal Y_1 moves further. So the Y_1-posterior is higher exactly when the signal is above the prior mean.

The multivariate case: conditioning a jointly normal vector

Fact 2.3

Let Z be jointly normal with full-rank covariance, partitioned into two blocks:

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathcal{N}\Big(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \Big).$$

Then Z_1 given $Z_2 = z_2$ is normal with

$$\hat{\mu} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(z_2 - \mu_2),$$
$$\hat{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Again: the mean depends on the realisation, the covariance does not.

Example 2.5 — two scalars with correlation ρ

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathcal{N}\Big(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \Big)$$

gives the familiar regression formulas

$$\hat{\mu} = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(z_2 - \mu_2), \qquad \hat{\Sigma} = \sigma_1^2(1-\rho^2).$$

Exercise 2.5: set $Z_1 = \theta$, $Z_2 = X = \theta + \varepsilon$. Then $\Sigma_{12} = \sigma_\theta^2$, $\Sigma_{22} = \sigma_\theta^2 + \sigma_\varepsilon^2$, and Fact 2.3 reduces to Fact 2.2.

5

Three applications (§2.3.2–2.3.4)

Career concerns (Holmström 1999) · Linear-quadratic coordination games (Morris & Shin 2002) · Data sharing on platforms (Acemoglu et al. 2022)

Career concerns: working hard to look talented

Period 1

An agent of unknown talent $\theta \sim \mathcal{N}(\mu, \sigma_\theta^2)$ — unknown to **everyone**, including the agent — chooses effort $a \geq 0$ at cost $c(a) = \frac{1}{2}a^2$. Effort is **not observed** by the manager.

Signal

Talent and effort jointly produce a noisy performance measure observed by the manager:

$$X = \theta + a + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2). \tag{2.7}$$

Period 2

The manager forms an expectation of θ from X and **pays the agent that expectation** (a competitive wage equal to expected talent).

The agent's problem

The manager **conjectures** an effort a^* and de-biases the signal accordingly; the agent actually chooses a. Writing $\mathbb{E}^a$ for expectations under true effort a, the agent maximises

$$\mathbb{E}^a\left[\mathbb{E}^{a^*}(\theta \mid X)\right] - c(a) :$$

the expected wage (the agent's expectation of the manager's expectation) minus the cost of effort.

Why would anyone work?

Effort cannot change true talent. But the manager cannot separate θ from a in X — so **extra effort looks like extra talent**, and that raises the wage. In equilibrium the manager is not fooled. Yet the agent still works.

Three steps: de-bias, take expectations, first-order condition

1. **The manager updates.** Expecting a^*, she treats $X - a^* = \theta + \varepsilon$ as a Gaussian signal about θ and applies Fact 2.2:
$$\mathbb{E}^{a^*}(\theta \mid X) = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\varepsilon^2}(X - a^*) + \frac{\sigma_\varepsilon^2}{\sigma_\theta^2 + \sigma_\varepsilon^2}\,\mu.$$
2. **The agent anticipates.** Under true effort a, X has mean $\mu + a$, so the expected wage is
$$\mathbb{E}^{a}\left[\mathbb{E}^{a^*}(\theta \mid X)\right] = \mu + \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\varepsilon^2}(a - a^*).$$
3. **First-order condition.** Marginal expected wage equals marginal cost at $a = a^*$:
$$\frac{\partial\, \mathbb{E}^{a}[\mathbb{E}^{a^*}(\theta \mid X)]}{\partial a}\Big|_{a=a^*} = a^*. \tag{2.8}$$

Claim 1

The derivative of the expected wage with respect to a is the signal weight, so the unique equilibrium effort is $a^* = \dfrac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\varepsilon^2}$. Uniqueness: the agent's objective is strictly concave in a (linear benefit, quadratic cost).

Equilibrium effort is the weight the market puts on the signal

Equilibrium effort

$$a^* = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\varepsilon^2} \in (0, 1)$$

$=$ the coefficient on the signal in the manager's posterior mean.

Corollary 2.2 — comparative statics

- a^* is **decreasing in** σ_ε^2: a noisier performance measure gets less weight, so manipulating it is less valuable.
- a^* is **increasing in** σ_θ^2: when talent is very uncertain, the market relies more on the signal, so effort pays more.

The economics

Nobody is fooled in equilibrium, yet effort is positive: if the agent shirked, the market would *underestimate* talent. Effort is a **rat race** driven by Bayesian updating — it vanishes if talent is known ($\sigma_\theta^2 \to 0$) or if performance is pure noise.

A linear-quadratic coordination game with private signals

Information

Two agents share a common prior about a state and each receives a private, conditionally independent signal:

$$\theta \sim \mathcal{N}(\mu, \sigma_\theta^2),$$

$$X_i = \theta + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2) \text{ i.i.d.}$$

The prior is **public information**; the X_i are **private information**.

Payoffs

Each agent chooses $a_i \in \mathbb{R}$ after seeing x_i, and wants to be close both to the state and to the other's action:

$$U_i(a_1, a_2) = -(1-\beta)(a_i - \theta)^2 - \beta(a_i - a_j)^2,$$

$$\beta \in (0, 1).$$

β measures the **coordination motive**.

Solution concept

A symmetric **linear Bayesian Nash equilibrium**: each agent's strategy is affine in her own signal, with the same constants for both,

$$a_i^*(x_i) = c\, x_i + \kappa. \tag{2.9}$$

Best reply, then match coefficients

1. **Best reply.** If j plays $cx_j + \kappa$, differentiating i's expected payoff gives (using $\mathbb{E}(X_j \mid x_i) = \mathbb{E}(\theta \mid x_i)$)

$$a_i^*(x_i) = (1-\beta)\,\mathbb{E}(\theta \mid X_i = x_i) + \beta\big(c\,\mathbb{E}(\theta \mid X_i = x_i) + \kappa\big).$$

2. **Insert Fact 2.2** for $\mathbb{E}(\theta \mid x_i)$ and match the coefficients of x_i and the constant with (2.9):

$$c = \frac{\sigma_\theta^2(1-\beta)}{\sigma_\varepsilon^2 + \sigma_\theta^2(1-\beta)}, \qquad \kappa = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_\theta^2(1-\beta)}\,\mu.$$

Equilibrium (2.10)

$$a_i^*(x_i) = \frac{\sigma_\theta^2(1-\beta)}{\sigma_\varepsilon^2 + \sigma_\theta^2(1-\beta)}\,x_i + \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_\theta^2(1-\beta)}\,\mu.$$

Reading (2.10). Compared with the pure-guessing posterior mean ($\beta = 0$), the coordination motive **shrinks the weight on the private signal** and raises the weight on the public prior. If the prior comes from a public signal, a more informative public signal (smaller σ_θ^2) makes agents rely **less** on their private information.

Data sharing: why platforms pay so little for your data

Players and information

A platform and two users $i = 1, 2$ with **correlated** types and private signals:

$$\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix} \sim \mathcal{N}\Big(\begin{pmatrix}0\\ 0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\Big),$$

$$X_i = \theta_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, 1).$$

The platform offers each user a payment p_i; user i then **shares** ($a_i = 1$) or **withholds** ($a_i = 0$) her signal. $X_{\mathbf{a}}$ = the shared signals.

Payoffs

Users value **privacy**: utility rises with the platform's residual uncertainty about their type, scaled by $v > 0$, plus the payment if they share. The platform wants the opposite:

$$u_i(\mathbf{a}, \mathbf{p}) = v \cdot \mathrm{Var}(\theta_i \mid X_{\mathbf{a}}) + p_i \cdot \mathbf{1}(a_i = 1),$$

$$u_P = -u_1 - u_2.$$

Claim

When types are sufficiently correlated, the platform can induce **both** users to share at a lower total payment than it needs to induce **exactly one** of them. Each user's data reveals something about the other — a negative externality that depresses the price of data.

Cheapest payments: both share vs. exactly one shares

Both share — step 1. Suppose j shares. If i withholds, the platform learns about θ_i only through X_j. Fact 2.3 with $\mathrm{Cov}(\theta_i, X_j) = \rho$, $\mathrm{Var}(X_j) = 2$:

$$\begin{pmatrix} \theta_i \\ X_j \end{pmatrix} \sim \mathcal{N}\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 2 \end{pmatrix} \Big) \implies \mathrm{Var}(\theta_i \mid X_j) = 1 - \frac{\rho^2}{2}.$$

Step 2. If i also shares, the platform conditions on both signals (Fact 2.3 again):

$$\mathrm{Var}(\theta_i \mid X_i, X_j) = \frac{2-\rho^2}{4-\rho^2}.$$

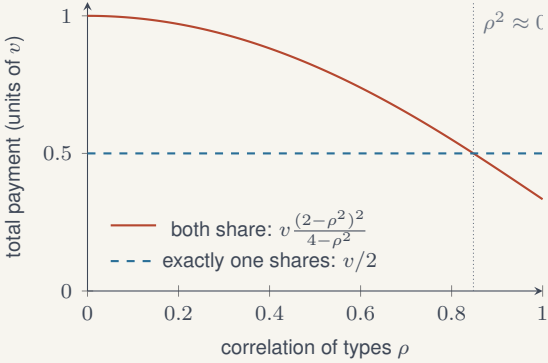
Step 3. Sharing is a best reply to $a_j = 1$ iff payoff from sharing $\geq$ payoff from withholding; the minimum total payment is twice the right-hand side:

$$v \cdot \frac{2-\rho^2}{4-\rho^2} + p_i \;\geq\; v\Big(1 - \frac{\rho^2}{2}\Big) \iff p_i \;\geq\; v \cdot \frac{(2-\rho^2)^2}{2(4-\rho^2)}, \qquad p_1 + p_2 = v \cdot \frac{(2-\rho^2)^2}{4-\rho^2}.$$

Exactly one shares

Fix $a_2 = 0$. Withholding leaves the platform at the prior variance 1; sharing X_1 alone reduces it to $\frac{1}{2}$ (Fact 2.2), so $p_1 \geq v/2$. User 2 is offered nothing ($p_2 = 0$), so she withholds. **Total payment:** $v/2$.

Correlated data is cheap data



Comparing the two totals

$$\rho^2 > \frac{7-\sqrt{17}}{4} \approx 0.72 \implies v \cdot \frac{(2-\rho^2)^2}{4-\rho^2} < \frac{v}{2}$$

Beyond this correlation, inducing **both** users to share costs the platform **less** than inducing one.

Why: a negative externality

When both share, each signal is worth less to its owner *in view of the signal revealed by the other*. Users paid their marginal value receive little — and with many correlated users the compensation needed to induce universal sharing **vanishes**.

Six things to take away from Chapter 2

1. A **signal** is a joint distribution of (parameter, realisation); kernels, random variables and partitions are three notations for it.
2. **Bayes' rule** is the two factorisations of a joint probability. In odds form: posterior odds = likelihood ratio $\times$ prior odds. Forgetting the prior odds is base-rate neglect.
3. Posteriors are a **martingale**: their expectation is the prior. Conversely, any distribution over posteriors that averages to the prior is induced by some signal.
4. Three fairness criteria for a score (equal FP, equal FN, calibration) are **jointly impossible** when base rates differ — a consequence of Bayes' rule alone.
5. **Gaussian updating**: posterior mean = precision-weighted average of prior mean and signal; posterior precision = sum of precisions; posterior variance does not depend on the realisation.
6. The **signal weight** in the posterior mean is the economic quantity in all three applications: equilibrium effort, the slope on private information, and the price of data.

Exercises for the problem set

Exercise	Level	Topic	What you are asked to do
2.4	U	Gaussian updating	Prove that the posterior mean is normal with mean μ and variance equal to the prior variance minus the posterior variance.
2.5	U	Multivariate	Show that Fact 2.3 implies Fact 2.2.
2.6	U	Career concerns	Add a second signal to the firm — about talent in (a), about the noise in (b) — and compare equilibrium effort with the a^* of Claim 1.
2.7	G	Career concerns	Correlate type and noise ($\theta = \theta_1 + \theta_2$, $\theta_2 = \alpha\theta_1 + z$, $\varepsilon = \beta\theta_1 + w$); solve for effort and its dependence on α, β.
2.8	U	Biased signals	$Y_1 = \theta + b + \varepsilon_1$, $Y_2 = b + \varepsilon_2$: show that $\theta \mid Y_1, Y_2$ differs from $\theta \mid Y_1 - Y_2$, and explain the friend's error.
2.9	—	Data sharing	Redo the two-user game with covariance matrix $\left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$: best replies, equilibria, cheapest payments, comparison with two separate firms.
2.10	G	Sufficiency	With exchangeable jointly normal signals, prove that $\theta \mid X_1, \ldots, X_n$ has the same distribution as $\theta \mid \bar{X}$.
2.11	G	Social learning	Compare learning from one well-informed predecessor versus m independent ones; characterise the threshold $h(n)$.

U = advanced undergraduate, G = graduate (as in the notes). 2.1–2.3 were solved in class. Suggested set: 2.4, 2.6, 2.8, 2.9; bonus 2.10–2.11.

Source and further reading

Source

Liang (2022), Chapter 2: Bayesian Updating and Beliefs. All numbered definitions, facts, propositions and exercises refer to this text.

- **Foundations:** Aumann (1976); de Oliveira (2018); Frankel and Kamenica (2019); Kamenica and Gentzkow (2011); Shmaya and Yariv (2016).
- **Applications:** Kleinberg et al. (2017); Chouldechova (2017); Holmström (1999); Morris and Shin (2002); Acemoglu et al. (2022).

Full references on the next slides.

References I

Acemoglu, D., Makhdoumi, A., Malekian, A., and Ozdaglar, A. (2022). Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics*.

Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics*, 4(6):1236–1239.

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5(2):153–163.

de Oliveira, H. (2018). Blackwell's informativeness theorem using diagrams. *Games and Economic Behavior*, 109:126–131.

Frankel, A. and Kamenica, E. (2019). Quantifying information and uncertainty. *American Economic Review*, 109(10):3650–3680.

Holmström, B. (1999). Managerial incentive problems: A dynamic perspective. *Review of Economic Studies*, 66(1):169–182.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Kleinberg, J., Mullainathan, S., and Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. In *8th Innovations in Theoretical Computer Science Conference (ITCS)*.

Liang, A. (2022). Information and learning in economic theory. Lecture notes (in progress), Chapter 2: Bayesian updating and beliefs. arXiv:2212.07521.

Morris, S. and Shin, H. S. (2002). Social value of public information. *American Economic Review*, 92(5):1521–1534.

Shmaya, E. and Yariv, L. (2016). Experiments on decisions under uncertainty: A theoretical framework. *American Economic Review*, 106(7):1775–1801.