Chapter 2

Bayesian Updating and Beliefs

Section 2.1 introduces the canonical Bayesian framework and the definition of a signal. Section 2.2 reviews Bayes' rule and key properties of Bayesian posteriors. Section 2.3 provides closed-form expressions for posterior beliefs in the special case of Bayesian updating to normal signals, with applications.

2.1 Preliminaries

There is a set of *parameters* Θ endowed with a σ-algebra Σ. An agent has a *prior* $p \in \Delta(\Theta)$, where $\Delta(\Theta)$ denotes the set of Σ-measurable probability measures on Θ. The prior describes the agent's belief at an "ex-ante" stage in the absence of any information, where what is ex-ante is understood in the context of a specific model.

The focus of this chapter is the object that we will call an *information structure*, *experiment*, or a *signal*, which can be formalized in either of several ways:

(a) We can define the signal to be a mapping $\sigma : \Theta \to \Delta(S)$ from the set of parameters to distributions over a set of signal realizations S. See for example de Oliveira (2019).

(b) We can define a signal to be an $(S, \mathcal{S})$-valued random variable X on an underlying probability space (Ω, Σ, P), where $\Omega = \Theta \times E$ for some set E. For example, we might define the signal to be $X = \theta + \varepsilon$ for an E-valued noise term ε that is independent of θ, as we do in Section 2.3.

(c) We can define a signal S to be a finite partition of $\Omega = \Theta \times [0, 1]$, whose elements are non-empty and measurable with respect to the Lebesgue σ-algebra on Ω. Conditional on parameter θ, the probability of observing $s \in S$ is the Lebesgue measure of $\{x \in [0, 1] \mid (\theta, x) \in s\}$. See for example Frankel and Kamenica (2019).

Remark 2.1. It is straightforward to see that the first two formalisms nest one another when all the relevant sets are finite. Suppose we are given a prior $p \in \Delta(\Theta)$ and a signal $\sigma : \Theta \to \Delta(S)$. Define the expanded state space to be $\Omega = \Theta \times S$ and let $P(\theta, s) = p(\theta)\sigma(s \mid \theta)$. Then the random variable

$X : \Omega \to S$ satisfying $X(\theta, s) = s$ is equivalent to σ in the sense that posterior beliefs about θ are the same whether we condition on the realization of X or the realization of $\sigma(\theta)$. In the other direction, if we start with a random variable $X : \Theta \times E \to S$ and a distribution $P \in \Delta(\Theta \times E)$, then we can define $\sigma : \Theta \to \Delta(S)$ to satisfy $\sigma(s \mid \theta) = P(X^{-1}(s) \mid \theta)$. The formalism in (c) is a special case of (b), where $E = [0, 1]$, the random variable $X : \Omega \to S$ maps each ω into the partition element of S to which it belongs, and the probability distribution P is the Lebesgue measure.

Example families of signals include:

Example 2.1 (Aumann (1976)'s Partitional Information Structures). For each agent i, let Π_i be a finite partition of Θ into measurable elements of strictly positive measure. Index these partition elements to $S = \{1, \ldots, n\}$ where n is the size of Π_i. Then let σ map each θ with probability 1 to the index of the partition element to which θ belongs.

Example 2.2 (Finite Information Structures). Suppose $|\Theta|, |S| < \infty$. Then we can express σ as a $|\Theta| \times |S|$ matrix where (1) all entries are nonnegative, and (2) all rows sum to 1. For example, suppose a drug is either good (g) or bad (b). The drug is administered to a patient who is either cured (C) or not (N). The patient is cured with probability 3/4 if the drug is good and with probability 1/4 if the drug is bad. Then $\Theta = \{g, b\}$ and $S = \{C, N\}$ and the information structure is

$$\begin{array}{ccc} & C & N \\ g & 3/4 & 1/4 \\ b & 1/4 & 3/4 \end{array}$$

with each row depicting the probability over the signal realizations in the associated state.

Example 2.3 (Gaussian Information). The signal is $X = \theta + \varepsilon$, where $\theta \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$, $\varepsilon \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$, and $\theta \perp\!\!\!\perp \varepsilon$.

2.2 Posterior Beliefs

2.2.1 Bayes' Rule

The agent updates his prior to the realization of the signal using Bayes' rule.

Definition 2.1 (Bayes' Rule, Finite Case). *Suppose $|\Theta| < \infty$. Fix any distribution $P \in \Delta(\Theta)$ and any events $A, B \subseteq \Theta$ where $P(A), P(B) > 0$. Then*

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

Remark 2.2. Rather than memorizing this formula, it is easier to remember that by the law of total probability, we can rewrite $P(A \cap B)$ as $P(A \mid B)P(B)$ or as $P(B \mid A)P(A)$, so

$$P(A \mid B)P(B) = P(B \mid A)P(A).$$

Dividing through by $P(B)$ yields (2.1).

Remark 2.3. Applying (2.1) twice for the pairs of events (A,E) and (B,E), we have

$$\frac{P(A \mid E)}{P(B \mid E)} = \frac{P(E \mid A)}{P(E \mid B)} \cdot \frac{P(A)}{P(B)}$$

so the relative conditional probabilities of events A and B is determined by their relative probabilities under the prior, $\frac{P(A)}{P(B)}$, and the *likelihood ratio* of E under A and B, $\frac{P(E|A)}{P(E|B)}$. *Base-rate neglect* is the tendency to falsely equate $\frac{P(A|E)}{P(B|E)}$ with $\frac{P(E|A)}{P(E|B)}$, neglecting the prior distribution. This can lead to compelling but inaccurate statistical conclusions.

For example, suppose $\Omega = \{p,n\}$, where $\omega = p$ indicates that an individual is positive for a medical condition while $\omega = n$ indicates that the individual is negative, with $P(\omega = p) = 0.01$. Let $X \in \{+,-\}$ be the outcome of a test where $P(X = + \mid \omega = p) = 0.95$ and $P(X = + \mid \omega = n) = 0.05$. Since the likelihood of observing $X = +$ is much higher when the individual has the condition than when he does not—indeed, the likelihood ratio is $\frac{P(X=+|\omega=p)}{P(X=+|\omega=n)} = 19$—it is tempting to conclude from a positive test result that the individual has the condition. But correctly applying Bayes' rule yields that $\frac{P(\omega=p|X=+)}{P(\omega=n|X=+)} < 1$; that is, even with a positive test it is more likely that the individual is negative for the condition.

A useful rewriting of Bayes' rule is

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

where the conditional distribution $P(\cdot \mid X = x)$ is precisely the agent's posterior belief upon observing $X = x$.

Example 2.4. A drug is either effective ($\theta = A$) or not ($\theta = B$), where the prior probability that the drug is effective is $p \in (0,1)$. The signal is

$$\begin{array}{ccc} & a & b \\ A & q & 1-q \\ B & 1-q & q \end{array}$$

for some $q \in (0,1)$. Then upon observing a, the agent assigns to $\theta = A$ a posterior probability of

$$\frac{pq}{pq + (1-p)(1-q)} = \frac{1}{1 + \frac{1-p}{p}\left(\frac{1-q}{q}\right)}$$

which exceeds the prior belief of p if and only if $q \geq \frac{1}{2}$.

More generally, when θ and X are (not necessarily finite-valued) random variables with densities f_θ and f_X and conditional densities $f_{\theta|X=x}$ and $f_{X|\theta=t}$, then the posterior belief given $X = x$ is

$$f_{\theta|X=x}(t) = \frac{f_{X|\theta=t}(x)f_\theta(t)}{\int_{\theta'\in\Theta} f_{X|\theta=t'}(x)f_\theta(t')dt'} \quad \forall t \in \Theta. \tag{2.3}$$

Somewhat more generally, we may suppose that the joint distribution of (θ, X) is such that for every realization x of X, there is a (measurable) function q_x satisfying

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

Then this q_x is the posterior belief.

2.2.2 Bayes' Plausibility

Outside of special cases (such as the one we will cover in Section 2.3), posterior beliefs often cannot be expressed in closed-form. Nevertheless, there are certain properties they must satisfy. One important property is that beliefs are a martingale, i.e., the expected posterior is equal to the prior. Intuitively, if you expect to change your mind given more information, then why haven't you done so already?

Fact 2.1 (Beliefs are a martingale.). *Let $p \in \Delta(\Theta)$ denote the agent's prior belief, and choose any event A. Then the posterior probability assigned to this event conditional on the realization of random variable X is $\mathbb{E}(\mathbb{1}_A \mid X)$. By the law of iterated expectations,*

$$\mathbb{E}(\mathbb{E}(\mathbb{1}_A \mid X)) = \mathbb{E}(\mathbb{1}_A)$$

so the expected posterior probability of A is equal to the prior probability of A. Since the event A was arbitrarily chosen, we can conclude that the expected posterior belief is equal to the prior belief. (In the case of a finite state space Θ, choosing $A = \{\theta\}$ yields $\mathbb{E}(p(\theta \mid X)) = p(\theta)$ for every θ.)

Since any signal X induces a distribution $\tau \in \Delta(\Delta(\Theta))$ over posterior beliefs, Fact 2.1 implies that this distribution must average to the prior.

Definition 2.2. *Fixing a prior $p \in \Delta(\Theta)$, say that a distribution of posteriors τ is* Bayes plausible *if*

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

i.e. the expected posterior is equal to the prior. We'll use

$$\mathcal{T}(p) \equiv \left\{\tau \in \Delta(\Delta(\Theta)) \;\middle|\; \int q d\tau(q) = p\right\}$$

to denote the set of Bayes plausible posterior distributions given prior p.

EXERCISE 2.1 (U). *The state space is* $\Theta = \{\theta_1, \theta_2\}$ *and the prior is* $(\mu, 1-\mu)$ *for some* $\mu \in [0,1]$. *The signal structure is*

	s_1	s_2
θ_1	p	$1-p$
θ_2	q	$1-q$

where $p, q \in [0,1]$. *What is the distribution over posterior beliefs induced by this signal structure? Verify that the expected posterior belief is equal to the prior belief.*

Not only are we guaranteed that any signal induces a Bayes-plausible distribution over posterior beliefs, but also any Bayes-plausible distribution over posterior beliefs can be induced by some signal.

DEFINITION 2.3. *For any signal* $X \sim P_X$, *let* $\tau_X \in \Delta(\Delta(\Theta))$ *satisfy* $\tau_X(q) = P_X(\{x : q_x = q\})$. *Say that* $\tau \in \Delta(\Delta(\Theta))$ *is* induced by X *if* $\tau = \tau_X$.

Proposition 3. *Suppose the prior* p *belongs to the interior of the set* $\Delta(\Theta)$. *Then every Bayes-plausible distribution* $\tau \in \mathcal{T}(p)$ *is induced by some signal* X.

The proof (demonstrated in Kamenica and Gentzkow (2011) and Shmaya and Yariv (2016) among others) proceeds by construction. For any distribution τ, index the distinct posterior beliefs in the support of τ to be $\{q_x\}_{x \in \mathcal{X}}$, where $\mathcal{X}$ may not be finite. Then define $\sigma : \Theta \to \Delta(\mathcal{X})$ to satisfy

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

We have constructed a signal σ whose realizations x are identified with posterior beliefs q_x, where the conditional distribution over signal realizations mimics Bayes' rule $p(x \mid \theta) = \frac{p(\theta|x)p(x)}{p(\theta)}$, setting $q_x(\theta) = p(\theta \mid x)$ and $\tau(q_x) = p(x)$. This is a valid signal structure since

$$\int_{\mathcal{X}} \sigma(x \mid \theta) dx = \int_{\mathcal{X}} \frac{q_x(\theta)\tau(q_x)}{p(\theta)} dx = 1$$

by (2.4) and the definition of Bayes-plausibility. Moreover,

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

so $q_x(\cdot)$ is precisely the posterior belief when updating to the signal σ.

Thus the probability that the posterior belief is q_x is exactly the probability that the realization of the constructed signal σ is x, so τ is induced by σ as desired.

EXERCISE 2.2 (U). *Suppose the prior is 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.*

Together, Fact 2.1 and Proposition 3 imply:

Corollary 2.1. *Fix any prior belief $p \in Int(\Delta(\Theta))$. Then a distribution over posteriors $\tau \in \Delta(\Delta(\Theta))$ is induced by some signal if and only if it is Bayes-plausible, i.e., $\tau \in \mathcal{T}(p)$.*

2.2.3 Application of Bayes' Rule: Incompatibility of Fairness Definitions

Here we take a detour to demonstrate the power of Bayes' rule. Individuals in a population are each described by a covariate vector $C \in \mathcal{C}$, a group membership $G \in \{g_1, g_2\}$, and a type $\theta \in \{0, 1\}$. For example, we might interpret θ as the individual's creditworthiness (whether the individual would pay back a loan if approved), G as a demographic group, and C as the individual's credit history. Across individuals, the random vector (C, G, θ) is distributed according to P, and we use $p_g = P(\theta = 1 \mid G = g)$ for the base rate of $\theta = 1$ in each group g. A *scoring rule* is any mapping $S : \mathcal{C} \to \{0, 1\}$ that predicts the type given the covariate vector.

Definition 2.4 (Equality of False Positives). *A scoring rule S has equal false positive rates if*

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

In words, the probability of being incorrectly assessed to pay back the loan is independent of group membership. Equivalently: $S \perp\!\!\!\perp G \mid \theta = 0$, i.e., the score is conditionally independent of group membership given type $\theta = 0$.

Definition 2.5 (Equality of False Negatives). *A scoring rule S has equal false negative rates if*

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

In words, the probability of being incorrectly assessed to not pay back the loan is independent of group membership. Equivalently: $S \perp\!\!\!\perp G \mid \theta = 1$, i.e., the score is conditionally independent of group membership given type $\theta = 1$.

Definition 2.6 (Calibrated). *A score S is calibrated if for each $s \in \{0, 1\}$,*

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

In words, among those assessed to pay back the loan (or, to not pay back the loan), the probability of paying back the loan is independent of group membership. Equivalently: $\theta \perp\!\!\!\perp G \mid S$, i.e., type is independent of group membership conditional on the score.

The following impossibility result demonstrates that (outside of edge cases) these fairness criteria cannot be simultaneously satisfied.

Proposition 4 (Kleinberg, Mullainathan and Raghavan (2017),Chouldechova (2017))**.** *Suppose $p_{g_1} \neq p_{g_2}$. Then no scoring rule S can simultaneously satisfy calibration, equal false positive rates, and equal false negative rates.*

Proof. Choose either group g and define $FP_g = P(S = 1 \mid \theta = 0, G = g)$, $FN_g = P(S = 0 \mid \theta = 1, G = g)$, and $PPV_g = P(\theta = 1 \mid S = 1, G = g)$. We'll show that these quantities are related by the following identity:

$$FP_g = \frac{p_g}{1 - p_g} \times \frac{1 - PPV_g}{PPV_g} \times (1 - FN_g). \tag{2.5}$$

To simplify notation, let Q denote the joint distribution over (C, G, S) after conditioning on $G = g$. Then, expanding (2.5), we have

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

Multiplying both sides by $Q(\theta = 0)$ and applying Bayes' rule,

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

Thus, (2.5) is equivalent to

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

Again using Bayes' rule, the RHS can be rewritten

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

so (2.6) is equivalent to

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

and is therefore trivially true.

The identity (2.5) holds for both groups $g \in \{g_1, g_2\}$. So if $FP_{g_1} = FP_{g_2}$ (as required by equality of false positive rates), $FN_{g_1} = FN_{g_2}$ (as required by equality of false negative rates), and also $PPV_{g_1} = PPV_{g_2}$ (as required by calibration), it must also hold that $p_{g_1} = p_{g_2}$. ■

2.3 Gaussian Information

Gaussian information environments are unusually tractable, since the posterior belief can be expressed in closed-form. We'll cover the main formulae for Bayesian updating in these environments, and show how these can be used to derive results in three applications.

2.3.1 Formulae

We'll start with the simplest case. The state is $\theta \sim \mathcal{N}(\mu, \sigma_\theta^2)$ and the signal is $X = \theta + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, $\theta \perp\!\!\!\perp \varepsilon$, and $\sigma_\theta^2, \sigma_\varepsilon^2 > 0$. Then:

FACT 2.2. *The agent's posterior belief about θ conditional on signal realization $X = x$ is normally distributed with mean*

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

and variance

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

A key property worth remembering is that the posterior mean is a convex combination of the prior mean μ and the signal realization x, where the weights are proportional to prior precision and signal precision. Additionally, while the posterior mean depends on the signal realization, the posterior variance is a constant.

Fact 2.2 is also sometimes written as:

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

where $\tau_\theta = 1/\sigma_\theta^2$ is the precision of the prior belief and $\tau_\varepsilon = 1/\sigma_\varepsilon^2$ is the precision of the signal. This restatement makes it apparent that the posterior precision is the sum of the prior precision and signal precision.

We can use Fact 2.2 to derive the distribution of the posterior mean.

EXERCISE 2.3 (U). *Let $\theta \sim \mathcal{N}(\mu, 1)$ and define two signals*

$$\begin{aligned} Y_1 &= \theta + \varepsilon_1 \\ Y_2 &= \theta + \varepsilon_2 \end{aligned}$$

where θ, ε_1, and ε_2 are all independent of one another, and $\varepsilon_1 \sim \mathcal{N}(0, 1)$ while $\varepsilon_2 \sim \mathcal{N}(0, 2)$.

(a) Solve for the conditional distributions $\theta \mid Y_1 = y$ and $\theta \mid Y_2 = y$.

(b) What are the values of (μ, y) for which it the case that $\mathbb{E}(\theta \mid Y_1 = y) > \mathbb{E}(\theta \mid Y_2 = y)$? Provide intuition for the condition you derive.

EXERCISE 2.4 (U). *Suppose we write the posterior belief as $\mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$, where $\hat{\mu}$ is a random variable that depends on the realization of the signal X. Prove that*

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

i.e. the expected posterior mean is the prior mean, and the variance of the posterior mean is equal to the prior variance (σ_θ^2), reduced by the posterior variance, $\left(\frac{\sigma_\theta^2 \sigma_\varepsilon^2}{\sigma_\theta^2 + \sigma_\varepsilon^2}\right)$.

This characterization implies that the more informative the signal is, the more variable the posterior mean is.

Remark 2.4. More generally (i.e., for θ and X that are not necessarily normally-distributed), the law of total variance says that

$$\mathrm{Var}(\mathbb{E}[\theta \mid X]) = \mathrm{Var}(\theta) - \mathbb{E}[\mathrm{Var}(\theta \mid X)]$$

so the variance of the posterior mean is equal to the difference of the prior variance and the expectation of the posterior variance.

Similar closed-forms exist for multivariate Gaussian states and signals. Suppose Z is a $1 \times K$ vector distributed according to $\mathcal{N}(\mu, \Sigma)$, where Σ has full rank. Partition the vector as follows:

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

Fact 2.3. *The conditional distribution of Z_1 given $Z_2 = z_2$ is $\mathcal{N}(\hat{\mu}, \widehat{\Sigma})$ where*

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

Again, the posterior mean depends on the signal realization, but the posterior covariance matrix does not.

Example 2.5. Let $\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right)$. Then $(Z_1 \mid Z_2 = z_2) \sim \mathcal{N}(\hat{\mu}, \widehat{\Sigma})$ where

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

Exercise 2.5 (U). *Let $Z_1 = \theta$ and $Z_2 = X$ where θ and X are as defined at the beginning of this section. Show that Fact 2.3 implies Fact 2.2.*

Sections 2.3.2-2.3.4 demonstrate three applications of these Bayesian updating formulae.

2.3.2 Application 1: Career Concerns

Our first application is solving the two-period version of Holmstrom (1999) model of career concerns.

There is a single agent and a manager. The agent has a type $\theta \sim \mathcal{N}(\mu, \sigma_\theta^2)$ that is unknown to both the agent and the manager. In period 1, the agent chooses an effort level $a \in \mathbb{R}_+$ at cost $c(a) = \frac{1}{2}a^2$. This effort is not observed by the manager. The agent's type and effort jointly determine the realization of a performance signal

$$X = \theta + a + \varepsilon \tag{2.7}$$

where $\theta \perp\!\!\!\perp \varepsilon$ and $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. In period 2, the manager observes the realization of X and forms an expectation about the agent's type. Since the manager does not observe a, this expectation is taken with respect to the manager's possibly misspecified perception about the distribution of X (more soon). The agent receives the manager's expectation of his type.

For arbitrary $a \in \mathbb{R}_+$, write $\mathbb{E}^a(\theta \mid X)$ for the conditional expectation of θ with respect to $X = \theta + a + \varepsilon$. If the manager expects the agent to choose effort a^* while the agent in fact chooses effort a, then the agent's total expected payoff is

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

where the inner expectation $\mathbb{E}^{a^*}(\theta \mid X)$ is the manager's expectation of the agent's type, and $\mathbb{E}^a[\mathbb{E}^{a^*}(\theta \mid X)]$ is the agent's expectation of the manager's expectation.

Claim 1. *There is a unique equilibrium in which the agent chooses effort $a^* = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\varepsilon^2}$.*

Corollary 2.2. *Equilibrium effort a^* is decreasing in σ_ε^2 (i.e., it is less valuable to manipulate a noisier signal) and is increasing in σ_θ^2 (i.e., it is more valuable to manipulate information about a more uncertain unknown).*

We'll now prove Claim 1. Equilibrium effort a^* must satisfy the first-order condition

$$\left.\frac{\partial \mathbb{E}^a[\mathbb{E}^{a^*}(\theta \mid X)]}{\partial a}\right|_{a=a^*} = a^* \tag{2.8}$$

equating the marginal value of increasing effort (over a^*) to the marginal cost of increasing effort (over a^*). Applying Fact 2.2, the manager's expectation of θ with respect to the de-biased signal $X - a^* = \theta + \varepsilon$ is

$$\mathbb{E}^{a^*}(\theta \mid X) = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\varepsilon^2}(X - a^*) + \frac{\sigma_\varepsilon^2}{\sigma_\theta^2 + \sigma_\varepsilon^2}\mu$$

The agent's expectation of this expectation (with respect to $X = \theta + a + \varepsilon$) is

$$\mathbb{E}^a\left[\mathbb{E}^{a^*}(\theta \mid X)\right] = \mu + \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\varepsilon^2}(a - a^*)$$

So (2.8) implies that equilibrium effort is $a^* = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\varepsilon^2}$. (Uniqueness follows from strict concavity of the agent's payoff function.)

Exercise 2.6 (U). *Consider the model described in this section, and set $\sigma_\theta^2 = \sigma_\varepsilon^2 = 1$.*

(a) *Suppose that in addition to the worker's performance signal, the firm (through collection of additional data about the worker's type) is able to separately observe a signal*

$$S = \theta + \delta$$

where θ and δ are jointly normal and independent, and $\delta \sim \mathcal{N}(0,1)$. The firm's expectation about the worker's type is based both on S as well as on the worker's performance signal X (as defined in (2.7)). Solve for the worker's equilibrium action and compare it with the previous solution $a^* = \frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\varepsilon^2}$. Does the worker exert more or less effort in equilibrium? Provide intuition for your result.

(b) *Suppose that the firm instead acquires data that allows it to more accurately monitor the performance shocks that the worker experiences (e.g., whether the worker had a rough day, or had help at work). Formally, the firm observes*

$$S = \varepsilon + \delta$$

where ε and δ are jointly normal and independent, and $\delta \sim \mathcal{N}(0,1)$. The firm's expectation about the worker's type is based both on S as well as on the worker's performance signal X (as defined in (2.7)). Solve for the worker's equilibrium action and compare it with the previous solution $a^* = \frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\varepsilon^2}$. Does the worker exert more or less effort in equilibrium? Provide intuition for your result.

EXERCISE 2.7 (G). *Consider a variation on Holmstrom (1999)'s career concerns model, in which the type θ and noise term ε are correlated. Specifically, the type is decomposed as $\theta = \theta_1 + \theta_2$, the signal is $X = \theta + \varepsilon + a$, and we suppose that*

$$\theta_2 = \alpha\theta_1 + z$$
$$\varepsilon = \beta\theta_1 + w$$

where $\alpha, \beta \in \mathbb{R}$ are known constants, and $\theta_1 \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$, $z \sim \mathcal{N}(0, \sigma_z^2)$, and $w \sim \mathcal{N}(0, \sigma_w^2)$ are mutually independent and unknown to both the agent and the manager.

(a) *Solve for equilibrium effort. How does this compare to Claim 1 in the special case $\alpha = \beta = 0$?*

(b) *Suppose $\alpha, \beta > 0$. How does equilibrium effort change in the parameters α and β? Provide intuition.*

2.3.3 Application 2: Linear-Quadatic Coordination Games

Our second application is solving for equilibrium in a two-agent linear-quadratic coordination game (Morris and Shin, 2002).

Let $\theta \sim \mathcal{N}(\mu, \sigma_\theta^2)$ be an unknown state. Each agent $i = 1, 2$ receives a private signal about the state

$$X_i = \theta + \varepsilon_i$$

where $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ is independent of the state and across agents. Each agent chooses an action $a_i \in \mathbb{R}$ given their signal realization x_i. Agent i's payoff is

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

where $\beta \in (0,1)$ controls how much the agent cares about matching the state versus matching the other agent's action.

We'll solve for a symmetric linear Bayesian Nash equilibrium (a_1^*, a_2^*) in which each agent's strategy satisfies

$$a_i^*(x_i) = cx_i + \kappa \tag{2.9}$$

for some constants $c, \kappa \in \mathbb{R}$. Let's first conjecture that such an equilibrium exists. Given agent j's strategy $a_j(x_j) = cx_j + \kappa$, agent i's expected payoff (conditional on $X_i = x_i$) is

$$\mathbb{E}[-(1-\beta)(a_i - \theta)^2 - \beta(a_i - (cX_j + \kappa))^2 \mid X_i = x_i]$$

Taking a derivative with respect to a_i, agent i's best reply is

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

Plugging in the expression for $\mathbb{E}(\theta \mid X_i = x_i)$ from Fact 2.2, and matching coefficients with (2.9), we have $c = \frac{\sigma_\theta^2(1-\beta)}{\sigma_\varepsilon^2 + \sigma_\theta^2(1-\beta)}$ and $\kappa = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_\theta^2(1-\beta)}\mu$. Thus a symmetric linear equilibrium exists in which each agent i chooses

$$a_i^*(x_i) = \frac{\sigma_\theta^2(1-\beta)}{\sigma_\varepsilon^2 + \sigma_\theta^2(1-\beta)} x_i + \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_\theta^2(1-\beta)}\mu \tag{2.10}$$

Morris and Shin (2002) further show that this is the unique pure-strategy equilibrium.

Suppose we interpret the common prior $\mathcal{N}(\mu, \sigma_\theta^2)$ as informed by a public signal, where a more informative signal implies a smaller σ_θ^2. Then we see from (2.10) that the more informative the public signal is, the less weight agents place on their private signal.

2.3.4 Application 3: Data Sharing

Our final application is an example from Acemoglu et al. (2022) regarding why online platforms don't compensate users for the data that they give up.

There is a single platform and two agents $i = 1, 2$ with types distributed

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

Each agent i privately observes the realization of a signal $X_i = \theta_i + \varepsilon_i$, where $\varepsilon_i \sim \mathcal{N}(0, 1)$ is independent across agents and independent of both types.

The platform chooses a payment p_i to offer to each agent i for sharing their data. After receiving these offers, each agent i chooses whether to share $(a_i = 1)$ or withhold $(a_i = 0)$ their signal realization. Write $X_{\mathbf{a}}$ for the signals shared under action profile $\mathbf{a} = (a_1, a_2)$. For example, if $\mathbf{a} = (1, 0)$, then $X_{\mathbf{a}} = X_1$, while if $\mathbf{a} = (1, 1)$, then $X_{\mathbf{a}} = (X_1, X_2)$.

Each agent i's payoff is determined by the platform's posterior uncertainty about his type, a privacy parameter $v \in \mathbb{R}_+$, and his payment via

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

The platform's payoff is $u_P(\mathbf{a}, \mathbf{p}) = -u_1(\mathbf{a}, \mathbf{p}) - u_2(\mathbf{a}, \mathbf{p})$. So the agents prefer for the platform to be more uncertain about their types, while the platform prefers to be less uncertain.

We'll show that when agent types are sufficiently correlated, i.e., ρ is large, then the platform can induce both agents to share their data at a lower total payment than what is required to induce exactly one agent to share.

Let's first solve for payment vectors (p_1, p_2) given which it is an equilibrium for both agents to share their signals. Suppose agent j chooses to share. Then if agent i does not share, the platform's belief about θ_i is updated only to X_j. Since

$$\begin{pmatrix} \theta_i \\ X_j \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 2 \end{pmatrix} \right)$$

the platform's posterior variance of θ_i is $1 - \rho^2/2$ (by Fact 2.3). So agent i's payoff is $v \cdot \left(1 - \rho^2/2\right)$. If agent i does share, then beliefs about θ_i are further updated to the signal X_i, and (by Fact 2.2) the platform's posterior variance of θ_i reduces to $\frac{2-\rho^2}{4-\rho^2}$. So agent i's payoff is $v \cdot \left(\frac{2-\rho^2}{4-\rho^2}\right) + p_i$. Thus, agent i's best reply to $a_j = 1$ is to share if and only if

$$p_i \geq v \cdot \left(\frac{(2-\rho^2)^2}{2(4-\rho^2)} \right)$$

and the action profile $(a_1, a_2) = (1, 1)$ is an equilibrium if the above display holds for both agents i. The minimum total payment is twice the right-hand-side, i.e., $v \cdot \left(\frac{(2-\rho^2)^2}{(4-\rho^2)}\right)$.

Let's now solve for payment vectors (p_1, p_2) given which it is an equilibrium for exactly one agent to share his data. Without loss, fix $a_2 = 0$. If agent 1 chooses $a_1 = 0$, then the platform's uncertainty about θ_1 is its prior uncertainty, 1, so agent 1's payoff is v. If agent 1 chooses $a_1 = 1$, then the platform's belief about θ_1 updates to the signal X_1. Applying Fact 2.2, the platform's posterior variance about θ_1 is $1/2$ and so agent 1's payoff is $v \cdot (1/2) + p_1$. Thus, $a_1 = 1$ is a best reply to $a_2 = 0$ if and only if $p_1 \geq v/2$. So the platform can induce (exactly) one agent to share if it offers one agent a payment of at least $v/2$ (which is accepted) and another a payment of no more than $v/2$ (which is rejected), at a total payment of $v/2$.

When $\rho^2 \geq \frac{7-\sqrt{17}}{4} \approx 0.71$, then $v \cdot \left(\frac{(2-\rho^2)^2}{(4-\rho^2)}\right) < v/2$, so the platform pays less to induce two users share than one. Intuitively, each agent's choice to share their data exerts a negative externality on other agent: When both users share, each of their signals is less valuable in view of the signal revealed by the other. Agents paid their marginal value thus receive lower compensation, and in a limiting version of this model with a growing number of agents, the amount of compensation needed to induce all agents to share vanishes to zero.

2.4 Additional Exercises

Exercise 2.8 (U). *Suppose $\theta \sim \mathcal{N}(0, \sigma_\theta^2)$ and*

$$\begin{aligned} Y_1 &= \theta + b + \varepsilon_1 \\ Y_2 &= b + \varepsilon_2 \end{aligned}$$

where θ, b, ε_1, and ε_2 are all independent of one another, $b \sim \mathcal{N}(0, \sigma_b^2)$, $\varepsilon_1 \sim \mathcal{N}(0, \sigma_1^2)$, and $\varepsilon_2 \sim \mathcal{N}(0, \sigma_2^2)$. We can interpret Y_1 as a biased signal about θ and Y_2 as a signal about the size of the bias.

Your friend says: "The only value of Y_1 and Y_2 for learning about θ is to provide information about the size of b. Since $Y_1 - Y_2$ is an unbiased signal about b, it is equally valuable to learn the outcome of $Y_1 - Y_2$ as it is to learn the pair of signals (Y_1, Y_2)."

Show that your friend is wrong: The distribution of $\theta \mid Y_1, Y_2$ is different from the distribution of $\theta \mid Y_1 - Y_2$. Also provide an intuition explaining to your friend the error in their reasoning.

Exercise 2.9. *Consider the two-player game described in Section 2.3.4, but suppose that the types are distributed*

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}\right)$$

As in Section 2.3.4, let a_1 and a_2 denote the actions of players 1 and 2, where an action of '0' means that the player does not share their data, while '1' means that they do.

(a) Suppose player 2 chooses $a_2 = 0$. What is player 1's payoff from choosing $a_1 = 0$ and what is player 1's payoff from choosing $a_1 = 1$? Provide a condition that characterizes when it is the case that player 1's best reply is to share data (i.e., $a_1 = 1$).

(b) Suppose player 2 chooses $a_2 = 1$. What is player 1's payoff from choosing $a_1 = 0$ and what is player 1's payoff from choosing $a_1 = 1$? Provide a condition that characterizes when it is the case that player 1's best reply is to share data (i.e., $a_1 = 1$).

(c) Suppose player 1 chooses $a_1 = 0$. What is player 2's payoff from choosing $a_2 = 0$ and what is player 2's payoff from choosing $a_2 = 1$? Provide a condition that

characterizes when it is the case that player 2's best reply is to share data (i.e., $a_2 = 1$).

(d) *Suppose player 1 chooses $a_1 = 1$. What is player 1's payoff from choosing $a_2 = 0$ and what is player 2's payoff from choosing $a_2 = 1$? Provide a condition that characterizes when it is the case that player 2's best reply is to share data (i.e., $a_2 = 1$).*

(e) *Suppose $v = 1$. For what set of values of (p_1, p_2) is it the case that:*

  (i) $(a_1, a_2) = (1, 1)$ *is a Nash equilibrium?*
  (ii) $(a_1, a_2) = (1, 0)$ *is a Nash equilibrium?*
  (iii) $(a_1, a_2) = (0, 1)$ *is a Nash equilibrium?*
  (iv) $(a_1, a_2) = (0, 0)$ *is a Nash equilibrium?*

(f) *Again let $v = 1$. What is the smallest total payment the firm must make to induce an equilibrium where both players share their data? (That is, what is the smallest sum $p_1 + p_2$ such that $(a_1, a_2) = (1, 1)$ is an equilibrium given the payment profile (p_1, p_2)?) Comment on whether it is the case that one player receives the higher payment, and why this answer makes sense.*

(g) *Again let $v = 1$. Suppose there is a firm 1 and firm 2, where firm 1 only interacts with player 1, and firm 2 only interacts with player 2. What is the smallest amount p_1 that firm 1 must pay to induce player 1 to choose $a_1 = 1$? What is the smallest amount p_2 that firm 2 must pay to induce player 2 to choose $a_2 = 2$? Compare the sum of these values $p_1 + p_2$ to your answer from part (f).*

EXERCISE 2.10 (G). *Suppose θ is normally distributed. For each $i = 1, \ldots, n$, let $X_i = \theta + \varepsilon_i$ where ε_i is independent of θ, the vector $(\varepsilon_1, \ldots, \varepsilon_n)$ is jointly normal, and the signals $X_1, \ldots, X_n$ are exchangeable. Define $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$. Prove that $\theta \mid X_1, \ldots, X_n$ is identical in distribution to $\theta \mid \overline{X}$.*

HINT. Recall that $\mathbb{E}(\theta \mid X)$ minimizes $\mathbb{E}[(\hat{\theta} - \theta)^2]$ among all $\sigma(X)$-measurable random variables $\hat{\theta}$.

EXERCISE 2.11 (G). *Consider two processes of social learning about an unknown state $\theta \sim \mathcal{N}(0, 1)$.*

Scenario 1: *At $t = 0$, a single agent privately observes the signal*

$$Y = \theta + \delta, \quad \delta \sim \mathcal{N}(0, 1/\tau)$$

where θ and δ are independent of one another, and the precision $\tau \in \mathbb{R}_+$ is a known constant. The agent chooses an action y and receives the payoff $-\mathbb{E}[(y - \theta)^2]$. At $t = 1$, each of n agents, indexed by i, privately observes a signal

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

as well as the action y of the first agent. The error terms ε_i are independent across agents, and independent of θ and δ. Each agent i from this generation then takes an *action a_i to maximize the payoff $-\mathbb{E}[(a_i - \theta)^2]$. At $t = 2$, you arrive, observe the actions $(a_1, \dots, a_n)$ of the preceding generation (but* not *the action of the first agent), and choose an action a^* with payoff $-\mathbb{E}[(a^* - \theta)^2]$.*

Scenario 2: *At $t = 1$, each of m agents observes a private signal*

$$Z_i = \theta + \eta_i, \quad \eta_i \sim \mathcal{N}(0, 1)$$

where the error terms η_i are independent across agents and of θ. Each agent i takes an action b_i with payoff $-\mathbb{E}[(b_i - \theta)^2]$. At $t = 1$, you arrive, observe the actions $(b_1, \dots, b_m)$ of the preceding generation, and choose an action a^* with payoff $-\mathbb{E}[(a^* - \theta)^2]$.

Characterize the function $h(n)$ such that your expected payoff is higher in scenario 1 if and only if $m < h(n)$. As clearly as you can, write out an intuition for this result.

Hint. Use the fact given in Exercise 2.10.