Contents

The previous chapter modeled once-and-for-all decisions. The decider (or *player*, as we shall now say) chooses a feasible action, then Nature reveals the true state and the player experiences the final consequences.

Life is more dynamic than that. A person makes a decision, events unfold influenced in part by that decision, then new decisions loom with new consequences, followed by more events and more decisions.

In this chapter we offer simple models of these dynamic situations. The first part considers how to combine new information with old; the technique is called Bayesian updating. The rest of the chapter draws on Bayesian updating to model optimal decision making when different choices have not just different immediate consequences but also different impacts on future information and choice opportunities.

1 Bayes' Theorem

How should we combine different pieces of information? It turns out that probability theory has a clear answer, one that may or may not accord with naive intuition. To explain that answer, we begin with notation.

Notation	*Classic name*	*Bayesian name*	*Meaning*
s	random variable	state of Nature	possible true state
z	random variable	message	possible signal or message
$p(s)$	marginal prob	prior prob of state s	prob of s before message received
$p(z)$	marginal prob	message prob	overall probability of signal z
$p(s, z)$	joint probability	...of given signal and state	Sometimes written $p(z \cap s)$.
$p(z\|s)$	conditional prob	likelihood	measures accuracy of message
$p(s\|z)$	conditional prob	posterior probability	prob of s after z received.

The idea is that the player knows the possible states $s \in S$, and given all previous relevant information, she assesses the probabilities as $p(s)$. Then she receives a new message z and updates the probabilities to $p(s|z)$.

For example, you think that the probability of rain tomorrow is $p(r) = 0.4$ and the probability of sun (or, more precisely, of no measurable rainfall) is $p(s) = 0.6$. But then you hear a forecast $z = \rho$ that it will rain. You update your belief that the true state will be rain to $p(r|\rho)$.

But what is the correct updated probability $p(r|\rho)$ of rain? Obviously it depends on the reliability of the forecast, e.g., on the likelihood $p(\rho|r)$.... but what else does it depend on, and how do we calculate the value?

The answer is obtained via classical probability algebra that goes back to the 1700s (worked out originally by Reverend Thomas Bayes and later by Pierre-Simon Laplace). Twentieth Century philosophers and statisticians developed the ideas (and terminology) to show how to leverage that answer.

Here's the algebra, along with commentary.

- Begin with the joint probabilities $p(s, z)$ (sometimes written $p(z, s)$ or $p(z \cap s)$) that the true state is s and the message received is z.

- For example, given true states $r, s \in S$ and possible messages $\rho, \sigma \in Z$, there are four possible joint events (s, z), whose probabilities $p(s, z)$ sum to 1.0.

  - Say $p(s, \sigma) = 0.42, p(s, \rho) = 0.18, p(r, \sigma) = 0.36, p(s, \rho) = 0.04$.

- Summing over messages, we get the marginal[1] state probabilities

$$p(s) = \sum_{z \in Z} p(s, z). \tag{1}$$

- You can check that in the example $p(r) = .36 + .04 = 0.4$ and $p(s) = .42 + .18 = 0.6$.

[1]This classic terminology is, of course, inconsistent with standard econ jargon, where marginal refers to first derivatives, not integrals or sums.

- The Bayesian jargon emphasizes that these are the probabilities you have *prior* to the arrival of the message (sometimes called *the news* or *a signal.*)

- Summing over states, we get the marginal probabilities

$$p(z) = \sum_{s \in S} p(s, z). \tag{2}$$

- In the example $p(\rho) = .18 + .04 = 0.22$ and $p(\sigma) = .42 + .36 = 0.78$.

- The Bayesian jargon emphasizes that these are the probabilities of the possible messages that you might receive later.

- By definition[2] our two sorts of conditional probabilities are:

$$p(s|z) = \frac{p(s, z)}{p(z)} \tag{3}$$

$$p(z|s) = \frac{p(s, z)}{p(s)} \tag{4}$$

  Cross-multiplying equation (4) gives us

$$p(s, z) = p(z|s)p(s) \tag{5}$$

  Substituting equation (5) into (3) gives us a first version of Bayes' Theorem (also called Bayes' rule):

$$p(s|z) = \frac{p(z|s)p(s)}{p(z)} \tag{6}$$

- This equation gives us what we sought.

  - The LHS is the Bayesian posterior probability of state s occurring after observing signal z.

  - The numerator of the RHS is the likelihood that the signal corresponds to the true state, times the prior probability.

[2]These formulas assume that none of the marginal probabilities are zero. If any of them is, that state or message will never occur, and can be dropped from the list.

- – The denominator normalizes so that the posterior probabilities sum to 1.0. Of course, equation (3) tells us that the denominator is the message probability.
- – In turn, plugging equation (5) into the definition (2) of the message probability gives us

$$p(z) = \sum_{t \in S} p(t, z) = \sum_{t \in S} p(z|t)p(t) \tag{7}$$

- Note the Bayesian posterior is increasing in the accuracy of the signal (likelihood) and the prior of the state, and is decreasing in the the message probability.

1.1 Four versions of the theorem

Theorem 1 (Bayes) *Using notation introduced above, the following formulas are all valid.*

$$p(s|z) = \frac{p(z|s)p(s)}{p(z)} \tag{8}$$

$$p(s|z) = \frac{p(z|s)p(s)}{\sum_{t \in S} p(z|t)p(t)} \tag{9}$$

$$\frac{p(s|z)}{p(t|z)} = \frac{p(z|s)p(s)}{p(z|t)p(t)} \tag{10}$$

$$\ln \frac{p(s|z)}{p(t|z)} = \ln \frac{p(z|s)}{p(z|t)} + \ln \frac{p(s)}{p(t)}. \tag{11}$$

The proof is simple.

- Equation (8) is a repeat of (6), derived via classic probability algebra and interpreted in terms of updating prior probabilities in the light of new information.
- Equation (9) is just (8) with (7) substituted into the denominator. It is handy when the message probabilities $p(z)$ are not given, or needed explicitly.

- Equation (10) is just the quotient of equation (8) for one particular state s and the same equation for another particular state r.[3] The denominators in equation (8) cancel when you take the quotient.

- Equation (10) says that the posterior odds $\frac{p(s|z)}{p(t|z)}$ of any two possible states are the product of the likelihood ratio and the two states' prior odds.

- Equation (11) is the easiest one for us to remember. It says that, expressed in logit form, Bayes' theorem is just a simple sum:
  posterior log odds = log likelihood + prior log odds.

Bayes theorem is written above for finite state spaces S and finite message spaces Z. There are two important extensions that should be mentioned immediately, so that you can do your homework problems.

1. Multiple messages. For example, you might hear several different forecasts about tomorrow's weather, or get several different medical diagnostic test results. With two distinct messages, for example, you might receive messages $z_1 \in Z_1,\ z_2 \in Z_2$.

- One way to extend Bayes theorem is to apply it to conjunctions of messages, and work out the joint probabilities and likelihoods for each possible conjunction, e.g., $z = (z_1, z_2)$ for two messages.

- Another way is to apply Bayes theorem sequentially. Treat the posterior probabilities $p(s|z_1)$ after receiving the first message as the prior probabilities when processing the second message, etc.

- Either way, the algebra is straightforward if the likelihoods of subsequent messages are not affected by those of earlier messages. The key property is called *conditional*

[3]If you prefer, you can take quotients using (9) instead of (8), once you remember that t is just a dummy variable for summation in (9) rather than a particular state.

independence:

$$p(z_1, z_2|s) = p(z_1|s)p(z_2|s) \quad \forall\ z_1 \in Z_1,\ z_2 \in Z_2,\ s \in S. \tag{12}$$

- With messages $z = (z_1, ..., z_n)$ conditionally independent, equation (10) becomes a simple product. Likewise, equation (11) is then a simple sum

$$\ln \frac{p(s|z)}{p(t|z)} = \ln \frac{p(s)}{p(t)} + \sum_{i=1}^{n} \ln \frac{p(z_i|s)}{p(z_i|t)} \tag{13}$$

2. Infinite State and Message Spaces. For example, the possible messages might be the number of geiger counter clicks per minute which, practically speaking, has no finite upper bound. Fortunately, the Bayes formulas given earlier require no modification when there are countably infinite states or messages.

Slight modifications are needed when we apply Bayes theorem to continuous state or message spaces.

- Suppose that we have joint probability density $f(s, z)$ over some rectangle $S \subset \Re$ of states and $Z \subset \Re$ of messages, and the rectangle $S \times R$ contains the support of the joint distribution.

- Replace sums by integrals in the definitions (1) and (2) of marginal probabilities, e.g., the prior density for states $s \in S$ is $f(s) = \int_{z \in Z} f(s, z) dz$.

- Provided that the joint density is continuous and positive at the realized signal and state, all of the formulas in Bayes Theorem remain valid when the symbol p (for probability mass) is replaced everywhere by f (for probability density function).

- Readers who enjoy formal mathematics will be able to prove this by chopping up the support of the joint distribution using a fine grid (mesh $\delta > 0$, say), applying the discrete Bayes formulas, taking the limit as $\delta \to 0$, and applying the definitions of density and continuity at (s, z).

1.2 Using Bayes theorem

One way to visualize the elements of Bayes theorem is to draw a tree in which Nature first determines the true state and then determines which message to send, as in Fig 1.

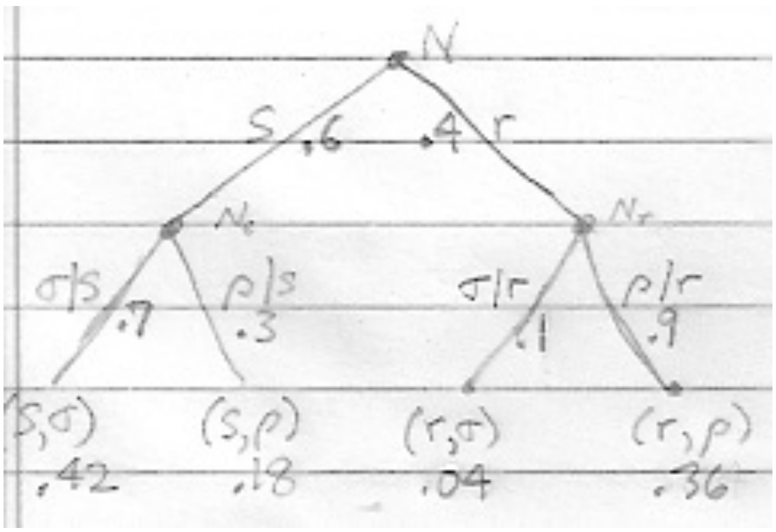

Figure 1: Nature determines the state (with prior probability $p(s) = .6$) and then determines the message (with likelihoods $p(\sigma|s) = .7$ and $p(\rho|r) = .9$). Those given probabilities imply the joint probabilities and posterior probabilities.

To return to our earlier example,

- the given prior probabilities of the states are $p(r) = 0.4$, $p(s) = 0.6$. The message accuracy is given in the form of the likelihoods $p(\rho|r) = 0.90$ and $p(\sigma|s) = 0.70$.
- Of course, the complementary likelihoods are $p(\rho|s) = 1 - .70 = .30$ and $p(\sigma|r) = 1 - .90 = .10$.
- You can find the joint probabilities by following tree branches to the end, which is the interpretation of equation (5).
  - For example, follow the s branch and the $\sigma|s$ branch to obtain the joint probability $p(s, \sigma) = p(s)p(\sigma|s)$.

– To spell it all out:

$$
\begin{aligned}
p(r,\rho) &= p(\rho|r)p(r) = (0.9)(0.4) = 0.36 && (14)\\
p(r,\sigma) &= p(\sigma|r)p(r) = (0.1)(0.4) = 0.04\\
p(s,\rho) &= p(\rho|s)p(s) = (0.3)(0.6) = 0.18\\
p(s,\sigma) &= p(\sigma|s)p(s) = (0.7)(0.6) = 0.42\\
sum\ of\ joints &= 1
\end{aligned}
$$

- Read the first line as: "the joint probability of a rainy forecast and a rainy day is equal to the likelihood of a rainy forecast conditional on a rainy day, times the probability of a rainy day."
- Note that the sum of the joint probabilities equals one since they cover all possible combinations.

Next, you can obtain the unconditional probability of a given message (ex. a rainy forecast ρ or a sunny forecast σ).

- Recall that it is found by summing the joint probabilities containing that signal across the possible states.

$$p(z) = \sum_{s\in S} p(s,z) \tag{15}$$

- Thus, the message probabilities in the example are:

$$
\begin{aligned}
p(\rho) &= \sum_{s\in S} p(\rho,s) = p(\rho,r) + p(\rho,s) = 0.36 + 0.18 = 0.54 && (16)\\
p(\sigma) &= \sum_{s\in S} p(\sigma,s) = p(\sigma,r) + p(\sigma,s) = 0.04 + 0.42 = 0.46
\end{aligned}
$$

- Note that even though there is a higher prior for sunny days ($p(s) = 0.60$), there is a lower prior for a sunny forecast ($p(\sigma) = 0.46$). The reason is that a rainy forecast is more accurate: $p(\rho|r) = 0.90$. There is only a 10% chance of observing a sunny forecast on a rainy day, but a 30% of observing a rainy forecast on a sunny day. Sunny days are more likely (prior of 60%), and a rainy forecast occurs 30% of the time on a sunny day. This makes a sunny forecast more believable, since it only occurs 10% of the time on rainy days.

Finally, you get to the main goal — to find the Bayesian posterior probabilities of each possible state having seen a particular message.

- Just use version 1 of Bayes Rule:

$$p(s|z) = \frac{p(s,z)}{p(z)} = \frac{p(z|s)p(s)}{p(z)}.$$

- In the current example,

$$\begin{aligned} p(r|\rho) &= \frac{p(r,\rho)}{p(\rho)} = \frac{0.36}{0.54} = 0.67 \\ p(s|\rho) &= \frac{p(s,\rho)}{p(\rho)} = \frac{0.18}{0.54} = 0.33 \\ p(r|\sigma) &= \frac{p(r,\sigma)}{p(\sigma)} = \frac{0.04}{0.46} = 0.087 \\ p(s|\sigma) &= \frac{p(s,\sigma)}{p(\sigma)} = \frac{0.42}{0.46} = 0.913 \end{aligned} \tag{17}$$

- Note that the sum of the posteriors for any given message is 1.

To do these sorts of calculations routinely, download from the class website a spreadsheet that spells it all out.

1. Begin by filling in your given information regarding prior probabilities and message likelihoods.
2. Use those givens to calculate complementary prior probabilities and likelihoods.
3. Note that likelihoods over all messages for a given state sum to 1.0, but that the likelihoods over all states for a given message can sum to more or less than 1.0.

4. Then use the spreadsheet to compute the joint probabilities.

5. Alternatively, if joint probabilities are given, then enter them and proceed with subsequent steps.

6. The spreadsheet formulas then automatically calculate the message probabilities.

7. Then the spreadsheet formulas automatically calculate all the posterior probabilities.

2 Decision trees

Decision trees are a great device for posing and solving dynamic decisions. These trees involve of one player, plus possibly Nature.

Decision trees consist of:

1. an initial node, sometimes called a root

2. subsequent nodes connected by branches

3. terminal nodes called leaves, which give the utility of the outcomes they represent.

4. each non-terminal node is owned either by Nature or the player, and is so labelled.

5. each terminal node gives the payoff if reached: the player's utility for that outcome.

Example: rain and umbrellas, Figure 2 .

- Action set $A = \{\text{carry umbrella } (c), \text{ don't carry umbrella } (d)\}$.

- Set of states $S = \{\text{rain } (r), \text{ sun}(s)\}$.

- Nature (N) determines which state is realized according to $p(r) = 0.4$, $p(s) = 0.6$.

- The fact that Nature moves after the Player in this tree indicates that the Player does not know Nature's move when choosing his action.

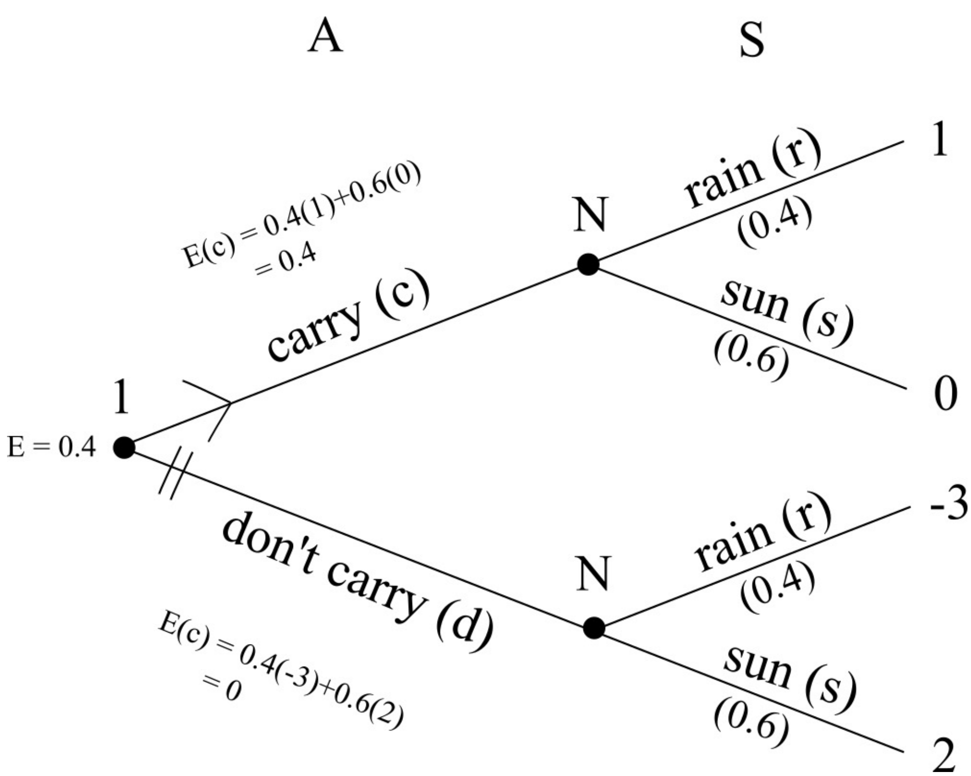

Figure 2: Umbrella decision tree, solved.