Bayesian Updating: Probabilistic Prediction
Class 12, 18.05, Spring 2014
Jeremy Orloff and Jonathan Bloom

1 Learning Goals

1. Be able to use the law of total probability to compute prior and posterior predictive probabilities.

2 Introduction

In the previous class we looked at updating the probability of hypotheses based on data. We can also use the data to update the probability of each possible outcome of a future experiment. In this class we will look at how this is done.

2.1 Probabilistic prediciton; words of estimative probability (WEP)

There are many ways to word predictions:

- Prediction: "It will rain tomorrow."

- Prediction using words of estimative probability (WEP): "It is likely to rain tomorrow."

- Probabilistic prediction: "Tomorrow it will rain with probability 60% (and not rain with probability 40%)."

Each type of wording is appropriate at different times.

In this class we are going to focus on probabilistic prediction and precise quantitative statements. You can see http://en.wikipedia.org/wiki/Words_of_Estimative_Probability for an interesting discussion about the appropriate use of words of estimative probability. The article also contains a list of *weasel words* such as 'might', 'cannot rule out', 'it's conceivable' that should be avoided as almost certain to cause confusion.

There are many places where we want to make a probabilistic prediction. Examples are

- Medical treatment outcomes

- Weather forecasting

- Climate change

- Sports betting

- Elections

- . . .

These are all situations where there is uncertainty about the outcome and we would like as precise a description of what could happen as possible.

3 Predictive Probabilities

Probabilistic prediction simply means assigning a probability to each possible outcomes of an experiment.

Recall the coin example from the previous class notes: there are three types of coins which are indistinguishable apart from their probability of landing heads when tossed.

- Type A coins are fair, with probability .5 of heads

- Type B coins have probability .6 of heads

- Type C coins have probability .9 of heads

You have a drawer containing 4 coins: 2 of type A, 1 of type B, and 1 of type C. You reach into the drawer and pick a coin at random. We let A stand for the event 'the chosen coin is of type A'. Likewise for B and C.

3.1 Prior predictive probabilities

Before taking data we can compute the probability that our chosen coin will land heads (or tails) if flipped. Let D_H be the event it lands heads and let D_T the event it lands tails. We can use the *law of total probability* to determine the probabilities of these events. Either by drawing a tree or directly proceeding to the algebra, we get:
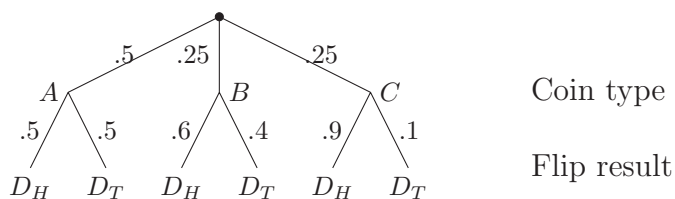
$$P(D_H) = P(D_H|A)P(A) + P(D_H|B)P(B) + P(D_H|C)P(C)$$
$$= .5 \cdot .5 + .6 \cdot .25 + .9 \cdot .25 = .625$$
$$P(D_T) = P(D_T|A)P(A) + P(D_T|B)P(B) + P(D_T|C)P(C)$$
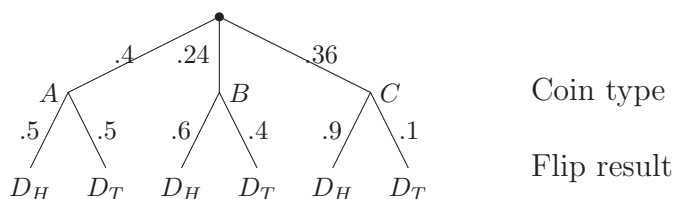$$= .5 \cdot .5 + .4 \cdot .25 + .1 \cdot .25 = .375$$

Definition: These probabilities give a (probabilistic) prediction of what will happen if the coin is tossed. Because they are computed before we collect any data they are called *prior predictive probabilities*.

3.2 Posterior predictive probabilities

Suppose we flip the coin once and it lands heads. We now have data D, which we can use to update the prior probabilities of our hypotheses to posterior probabilities. Last class we learned to use a Bayes table to facilitate this computation:

| | | | unnormalized | |
| hypothesis | prior | likelihood | posterior | posterior |
H	$P(H)$	$P(D\|H)$	$P(D\|H)P(H)$	$P(H\|D)$
A	.5	.5	.25	.4
B	.25	.6	.15	.24
C	.25	.9	.225	.36
total	1		.625	1

Having flipped the coin once and gotten heads, we can compute the probability that our chosen coin will land heads (or tails) if flipped a second time. We proceed just as before, but using the posterior probabilities $P(A|D)$, $P(B|D)$, $P(C|D)$ in place of the prior probabilities $P(A)$, $P(B)$, $P(C)$.

$$P(D_H|D) = P(D_H|A)P(A|D) + P(D_H|B)P(B|D) + P(D_H|C)P(C|D)$$
$$= .5 \cdot .4 + .6 \cdot .24 + .9 \cdot .36 = .668$$
$$P(D_T|D) = P(D_T|A)P(A|D) + P(D_T|B)P(B|D) + P(D_T|C)P(C|D)$$
$$= .5 \cdot .4 + .4 \cdot .24 + .1 \cdot .36 = .332$$

Definition: These probabilities give a (probabilistic) prediction of what will happen if the coin is tossed again. Because they are computed after collecting data and updating the prior to the posterior, they are called *posterior predictive probabilities*.

Note that heads on the first toss increases the probability of heads on the second toss.

3.3   Review

Here's a succinct description of the preceding sections that may be helpful:

Each hypothesis gives a different probability of heads, so the total probability of heads is a weighted average. For the prior predictive probability of heads, the weights are given by the prior probabilities of the hypotheses. For the posterior predictive probability of heads, the weights are given by the posterior probabilities of the hypotheses.

Remember: Prior and posterior probabilities are for hypotheses. Prior predictive and posterior predictive probabilities are for data. To keep this straight, remember that the latter **predict** future data.

18.05 Introduction to Probability and Statistics
Spring 2014