Let’s look at a new situation in which conditional probability can be used to help make a business decision. This time, the decision is whether or not a bank should foreclose on a business loan.

1 Foreclosures and workouts

Business loans can differ from personal loans in several ways. In a personal loan (such as a mortgage on a house), the borrower typically pays off a portion of the principal, along with the accumulated interest, with each payment. In business loans, by contrast, the payments often consist only of interest; at the end of the loan period, the borrower pays off the whole amount of the principal in a lump sum.

The value of the collateral for a personal loan typically exceeds the outstanding principal on the loan. For example, unless property values plummet, the bank can usually recover all of its principal by foreclosing on a home mortgage if the borrowers default on their payments. This is not always the case with business loans, since they are spent not only on tangible assets (buildings, equipment, etc.), but on operating expenses.

This complicates the decision whether or not to foreclose on a business loan. If the bank does foreclose, they probably won’t recover all of their money: only a lesser amount, the foreclosure value. If the business is fundamentally sound, it might be better to allow the borrower a workout. For example, they might allow him to make smaller payments over a longer period of time; or they might allow him to defer his payments for some period, then resume them later. If the workout succeeds, then the bank recovers the full value of the loan. On the other hand, there’s risk in allowing a workout: if the business fails after all, the bank only recovers a default value, which would probably be smaller than the foreclosure value.

We’ll use v with subscripts to indicate these three values:

\[ v_{\text{FC}} = \text{foreclosure value} \quad v_{\text{full}} = \text{full value} \quad v_{\text{def}} = \text{default value} \]

\[ v_{\text{def}} < v_{\text{FC}} < v_{\text{full}} \]

You can see the difficulty of the bank’s position: should they foreclose and be sure of getting the foreclosure value; or should they allow a workout, which will give them more money if it succeeds, but less if it fails?

It would help the bank if they knew the probability that a workout would succeed or fail. Let’s call \( S \) the event that the workout succeeds; and \( F \) the event that it fails. The tree diagram indicates the different possible outcomes of the decision:
If the bank forecloses, it’s sure to get \( v_{FC} \). If it allows a workout, there’s a probability of \( \mathcal{P}(S) \) that the workout will succeed and the bank will get \( v_{full} \); and a probability of \( \mathcal{P}(F) \) that the workout will fail, and it will get \( v_{def} \). The expected return from a workout is

\[
\mathcal{P}(S) v_{full} + \mathcal{P}(F) v_{def}
\]

The bank should allow a workout if this expected return is greater than the foreclosure value:

\[
\mathcal{P}(S) v_{full} + \mathcal{P}(F) v_{def} > v_{FC}
\]

If it’s significantly less, they should foreclose.

Note that every workout either succeeds or fails, and no workout does both. That means that \( S \) and \( F \) are a partition of the sample space; so

\[
\mathcal{P}(F) = 1 - \mathcal{P}(S)
\]

The bank should allow a workout if

\[
\mathcal{P}(S) v_{full} + [1 - \mathcal{P}(S)] v_{def} > v_{FC}
\]

**Example 1.** A bank must decide whether to foreclose or allow a workout on a business loan of $500,000. If it forecloses, it will recover a foreclosure value of $300,000. If it allows a workout and the workout fails, it will recover a default value of $100,000. From the records of all its workouts, the probability of success is \( \mathcal{P}(S) = 0.35 \). What should it do in this case?

**Solution:** The expected return to the bank from a workout is

\[
\mathcal{P}(S) v_{full} + [1 - \mathcal{P}(S)] v_{def} = (0.35)(500,000) + (0.65)(100,000) = 240,000
\]

This is significantly lower than the foreclosure value of $300,000. The bank should probably foreclose on this loan.

In making this calculation, the bank could look at the record of all workouts and estimate \( \mathcal{P}(S) \) based on that. However, there are factors that might affect the success or failure of a business loan workout. Three of these are

- The type of business: restaurant, retail clothing, convenience store, etc.
The location of the business.

- The owner’s business experience.

Using this information about the business seeking the workout might help the bank arrive at a better estimate of $P(S)$. Let’s define three events:

- $T$: The business is of a particular type
- $L$: The business is in a particular location
- $B$: The owner has a particular amount of business experience

Instead of using $P(S)$ in its calculations, the bank might get more accurate results by using $P(S | T \cap L \cap B)$: the probability that a workout will succeed for a business of a particular type, in a particular location, with a particular amount of business experience for the owner.

**Example 2.** Consider the loan in Example 1. Suppose that the business is a home-and-garden store, located in a rapidly-developing suburb, whose owner has over ten years of business experience. We’ll suppose that for such a situation, $P(S | T \cap L \cap B) = 0.7$. Should the bank foreclose or allow a workout?

**Solution:** The expected return from a workout is

$$P(S | T \cap L \cap B)v_{\text{full}} + [1 - P(S | T \cap L \cap B)]v_{\text{def}}$$

$$= (0.7)(500,000) + (0.3)(100,000)$$

$$= 380,000$$

This is much larger than the foreclosure value of $300,000; so the bank should probably allow a workout.

**Example 3.** Now, let’s assume instead that the business is an independent bookstore, located in a depressed part of town, owned by someone with less than a year of business experience. In this case, we’ll assume that $P(S | T \cap L \cap B) = 0.15$. Now what should the bank do?

**Solution:** The expected return from a workout is

$$P(S | T \cap L \cap B)v_{\text{full}} + [1 - P(S | T \cap L \cap B)]v_{\text{def}}$$

$$= (0.15)(500,000) + (0.85)(100,000)$$

$$= 160,000$$

This is a lot less than the foreclosure value of $300,000. The bank should almost certainly foreclose in this case.

How could the bank estimate $P(S | T \cap L \cap B)$? It would be nice if they could obtain records of past loan workouts, each record including the business type, location, and owner’s experience level. However, such complete records might not be available.
2 Our situation

We need to advise a bank whether to foreclose on a business loan or allow a workout. We know the foreclosure value, full value, and default value. We also know the business type, the location, and the owner’s experience level. We also have records of past loan workouts from three other banks. Unfortunately, those records are incomplete. Bank 1 kept records of business types and workout successes or failures. Bank 2 recorded business locations and the outcomes of workouts. Bank 3 kept track of owners’ experience levels and whether the workouts succeeded or failed. We have no records tying workout success to all three factors at once. In the language of events, Bank 1’s records tell us about \( S \) and \( T \). Bank 2 can tell us about \( S \) and \( L \). Bank 3 has information about \( S \) and \( B \). We need to calculate

\[
P(S \mid T \cap L \cap B)
\]

from these three sets of records.

If we had a set of records with complete information on each loan workout, this would be easy. We’d take that set of records as the solution space \( \Omega \); and we’d calculate

\[
P(S \mid T \cap L \cap B) = \frac{\#(S \cap T \cap L \cap B)}{\#(T \cap L \cap B)}
\]

Unluckily for us, the records are not complete. We can’t look at a single workout in any of our three databases and say whether it’s a member of the set \( T \cap L \cap B \).

There are ways to get around this problem, but they’ll require that we make certain assumptions.

First, we’ll assume that Banks 1, 2, and 3 made loans to similar types of borrowers. Probabilities calculated from one bank’s data apply to the other two banks’ borrowers.

Second, we’ll assume that each bank’s database is a good sample. Probabilities calculated from a bank’s data are close to real-world probabilities.

Third, we’ll assume that the events \( T \), \( L \), and \( B \) are independent, and remain so when conditioned on \( S \). This is probably the most questionable assumption: for example, if we say that \( T \) and \( L \) are independent, we’re saying that knowing the type of a business tells you nothing about where in town it’s located, and vice versa. In reality, of course, certain types of businesses tend to occur in certain neighborhoods: you would go to some parts of town to look for pawnshops or adult bookstores, and to others to find Scandinavian furniture outlets or nouvelle-cuisine restaurants. However, since we can’t do our calculation without this assumption, we’ll make it; and we’ll trust that for the business type and location categories we’re using, it holds true for the most part.
3 The calculations

According to the formula for conditional probability,

\[
P(S \mid T \cap L \cap B) = \frac{P(S \cap T \cap L \cap B)}{P(T \cap L \cap B)}
\]

We’ve assumed that the events \( T, L, \text{ and } B \) are independent; so we can immediately rewrite the denominator

\[
P(S \mid T \cap L \cap B) = \frac{P(S \cap T \cap L \cap B)}{P(T)P(L)P(B)} \quad (3.1)
\]

We can get the factors of the denominator from our three bank databases: Bank 1 gives us \( P(T) \), Bank 2 gives us \( P(L) \), and Bank 3 gives us \( P(B) \). To get the numerator, we have to be clever. By the formula for conditional probability,

\[
P(T \cap L \cap B \mid S) = \frac{P(T \cap L \cap B \cap S)}{P(S)} \quad (3.2)
\]

Remember that the intersection of sets is commutative: so

\[
P(T \cap L \cap B \cap S) = P(S \cap T \cap L \cap B)
\]

Thus if we multiply both sides of Equation (3.2) by \( P(S) \), we get

\[
P(S \cap T \cap L \cap B) = P(S)P(T \cap L \cap B \mid S) \quad (3.3)
\]

We’ve assumed that \( T, L, \text{ and } B \) remain independent when conditioned on \( S \); so

\[
P(T \cap L \cap B \mid S) = P(T \mid S)P(L \mid S)P(B \mid S) \quad (3.4)
\]

We can get the three factors on the right side of this equation from our three banks’ databases: \( P(T \mid S) \) from Bank 1, \( P(L \mid S) \) from Bank 2, and \( P(B \mid S) \) from Bank 3.

Substituting Equation (3.4) into Equation (3.3), we get

\[
P(S \cap T \cap L \cap B) = P(S)P(T \mid S)P(L \mid S)P(B \mid S)
\]

We can substitute this into Equation (3.1) to get our final equation:

\[
P(S \mid T \cap L \cap B) = \frac{P(S)P(T \mid S)P(L \mid S)P(B \mid S)}{P(T)P(L)P(B)} \quad (3.5)
\]

We can combine the three bank databases to estimate \( P(S) \); so we can get all the factors we need for this calculation.
4 Examples

Let’s look at some examples. For all of them, we’ll assume that the full value of the loan is $500,000; the foreclosure value is $300,000; and the default value is $100,000.

Example 4. For our first example, we’ll look at an auto-repair shop in an older blue-collar part of town, whose owner has five to seven years of business experience. Here are the probabilities estimated from the various databases:

Bank 1:  \( P(T) = 0.11 \)  \( P(T|S) = 0.16 \)
Bank 2:  \( P(L) = 0.15 \)  \( P(L|S) = 0.13 \)
Bank 3:  \( P(B) = 0.10 \)  \( P(B|S) = 0.12 \)
All banks:  \( P(S) = 0.40 \)

From Equation (3.5), the probability of a successful workout is

\[ P(S|T \cap L \cap B) = \frac{(0.40)(0.16)(0.13)(0.12)}{(0.11)(0.15)(0.10)} = 0.6051 \]

The expected return from a workout is

\[ P(S|T \cap L \cap B) v_{\text{full}} + [1 - P(S|T \cap L \cap B)] v_{\text{def}} \]
\[ = (0.6051)(500,000) + (0.3949)(100,000) \]
\[ = 342,036.36 \]

(We’ve recorded the probability of success to four decimal places here. However, we calculated that and the expected return in an Excel sheet; from that, we give the expected return to the nearest cent.)

This is considerably larger than the foreclosure value of $300,000; so the bank would probably be better off if they arranged a workout.

Example 5. In our next example, we’ll look at an independent restaurant in a 10-year-old suburb, whose owner has one to three years of business experience. We’ll assume that the databases give these probabilities:

Bank 1:  \( P(T) = 0.21 \)  \( P(T|S) = 0.13 \)
Bank 2:  \( P(L) = 0.08 \)  \( P(L|S) = 0.11 \)
Bank 3:  \( P(B) = 0.19 \)  \( P(B|S) = 0.12 \)
All banks:  \( P(S) = 0.40 \)

Using Equation (3.5) to find the probability of a successful workout, we get

\[ P(S|T \cap L \cap B) = \frac{(0.40)(0.13)(0.11)(0.12)}{(0.21)(0.08)(0.19)} = 0.2150 \]

For this case, the expected return from a workout would be

\[ P(S|T \cap L \cap B) v_{\text{full}} + [1 - P(S|T \cap L \cap B)] v_{\text{def}} \]
\[ = (0.2150)(500,000) + (0.7850)(100,000) \]
\[ = 186,015.04 \]
This is much less than the foreclosure value of $300,000. The bank should definitely foreclose in this case.

**Example 6.** Our third example is a retail furniture store in a midtown neighborhood with a mix of property values, whose owner has one to three years of business experience. We assume that the database gives these probabilities:

- **Bank 1:** \( P(T) = 0.085 \) \( P(T|S) = 0.125 \)
- **Bank 2:** \( P(L) = 0.08 \) \( P(L|S) = 0.105 \)
- **Bank 3:** \( P(B) = 0.19 \) \( P(B|S) = 0.12 \)
- **All banks:** \( P(S) = 0.40 \)

The probability of a successful workout is

\[
P(S | T \cap L \cap B) = \frac{(0.40)(0.125)(0.105)(0.12)}{(0.085)(0.08)(0.19)} = 0.4876
\]

For this case, the expected return from a workout would be

\[
P(S | T \cap L \cap B) v_{\text{full}} + [1 - P(S | T \cap L \cap B)] v_{\text{def}}
\]

\[= (0.4876)(500,000) + (0.5124)(100,000) = $295,046.44
\]

This is less than the foreclosure value of $300,000, but the difference is small. While the strict numbers indicate foreclosure, it might be in the bank’s interest to look more closely into the business’s situation. If their financial difficulties appear to result from an extraordinary and temporary situation (for example, road construction that has made access to their store difficult for prospective customers), the bank might opt for a workout anyhow. On the other hand, if their financial troubles arise from a long-term situation (for example, a major competitor that’s opened nearby and is drawing away business), then foreclosure might be called for.

**Example 7.** Our final example is a plumbing business located downtown, operated by someone with 0-1 years of business experience. For this example, instead of giving you the probabilities, we’ll estimate them from the tables of past loan workouts on page 9. The first column of each table contains either the business type, the location, or the owner’s experience; in the second column of each table, “Yes” indicates a successful workout; “No” indicates that the workout failed. (These databases should only be used for this example; they won’t give the probabilities that we used in earlier ones.)

The “Bank 1” table contains information on \( T \). To calculate \( P(T) \) and \( P(T|S) \), we’ll treat the 30 workouts in that table as our sample space \( \Omega_1 \). For 9 of those workouts, the business type is “Plumbing”. Hence

\[
P(T) = \frac{\#(T)}{\#(\Omega_1)} = \frac{\# \text{ plumbing workouts for Bank 1}}{\# \text{ all workouts for Bank 1}} = \frac{9}{30} = 0.3
\]
There are 15 successful workouts in “Bank 1”. We’ll treat these as the set $S_1$. For 5 of these successful workouts, the business type is “Plumbing”. Thus:

$$P(T|S) = \frac{\#(T \cap S_1)}{\#(S_1)} = \frac{\text{# successful plumbing workouts for Bank 1}}{\text{# all successful workouts for Bank 1}} = \frac{5}{15} = 0.3333$$

We’ll use the same approach to calculate $P(L)$ and $P(L|S)$ from the “Bank 2” table. There are 28 workouts in that table; we’ll treat them as the sample space $\Omega_2$. Then

$$P(L) = \frac{\#(L)}{\#(\Omega_2)} = ?$$

The set $S_2$ consists of the successful workouts in the “Bank 2” table. We’ll estimate

$$P(L|S) = \frac{\#(L \cap S_2)}{\#(S_2)} = ?$$

In the same way, you should use the “Bank 3” table to determine $P(B)$ and $P(B|S)$. The final number we need is $P(S)$. All three tables contain information on workout success; so we’ll combine them into the sample space

$$\Omega_{123} = \Omega_1 \cup \Omega_2 \cup \Omega_3$$

The collection of successful workouts from all three tables is

$$S_{123} = S_1 \cup S_2 \cup S_3$$

and we’ll calculate

$$P(S) = \frac{\#(S_{123})}{\#(\Omega_{123})} = \frac{\text{# successful workouts in all tables}}{\text{# workouts in all tables}} = \frac{15 + 7 + 8}{30 + 28 + 32} = \frac{30}{90} = 0.3333$$

Now, calculate $P(S \mid T \cap L \cap B)$ from Equation (3.5). Use that to calculate the expected return from a workout, and decide whether or not to foreclose on the loan. (You should get an expected return of about $136,000.)
<table>
<thead>
<tr>
<th>Bank 1</th>
<th>Bank 2</th>
<th>Bank 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Location</td>
<td>Experience</td>
</tr>
<tr>
<td>Plumbing</td>
<td>No</td>
<td>West Central</td>
</tr>
<tr>
<td>Plumbing</td>
<td>Yes</td>
<td>E Suburbs</td>
</tr>
<tr>
<td>Office Supplies</td>
<td>No</td>
<td>Downtown</td>
</tr>
<tr>
<td>Pharmacy</td>
<td>Yes</td>
<td>NW Suburbs</td>
</tr>
<tr>
<td>Clothing</td>
<td>Yes</td>
<td>West Central</td>
</tr>
<tr>
<td>Hardware</td>
<td>Yes</td>
<td>Downtown</td>
</tr>
<tr>
<td>Housewares</td>
<td>No</td>
<td>Downtown</td>
</tr>
<tr>
<td>Plumbing</td>
<td>No</td>
<td>Near South</td>
</tr>
<tr>
<td>Plumbing</td>
<td>Yes</td>
<td>Downtown</td>
</tr>
<tr>
<td>Auto Repair</td>
<td>No</td>
<td>Downtown</td>
</tr>
<tr>
<td>Plumbing</td>
<td>Yes</td>
<td>Downtown</td>
</tr>
<tr>
<td>Pharmacy</td>
<td>No</td>
<td>Downtown</td>
</tr>
<tr>
<td>Bookstore</td>
<td>No</td>
<td>Downtown</td>
</tr>
<tr>
<td>Restaurant</td>
<td>Yes</td>
<td>NW Suburbs</td>
</tr>
<tr>
<td>Plumbing</td>
<td>No</td>
<td>Northeast</td>
</tr>
<tr>
<td>Bookstore</td>
<td>No</td>
<td>Downtown</td>
</tr>
<tr>
<td>Housewares</td>
<td>No</td>
<td>Downtown</td>
</tr>
<tr>
<td>Clothing</td>
<td>Yes</td>
<td>Downtown</td>
</tr>
<tr>
<td>Plumbing</td>
<td>Yes</td>
<td>Downtown</td>
</tr>
<tr>
<td>Auto Repair</td>
<td>No</td>
<td>Northside</td>
</tr>
<tr>
<td>Restaurant</td>
<td>Yes</td>
<td>Far West</td>
</tr>
<tr>
<td>Clothing</td>
<td>Yes</td>
<td>South Central</td>
</tr>
<tr>
<td>Furniture</td>
<td>No</td>
<td>Near South</td>
</tr>
<tr>
<td>Hardware</td>
<td>No</td>
<td>Downtown</td>
</tr>
<tr>
<td>Electronics</td>
<td>No</td>
<td>Downtown</td>
</tr>
<tr>
<td>Restaurant</td>
<td>Yes</td>
<td>Northside</td>
</tr>
<tr>
<td>Pharmacy</td>
<td>Yes</td>
<td>Downtown</td>
</tr>
<tr>
<td>Plumbing</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Plumbing</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Hardware</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>