Things to consider when seeking additional information:

- Information gathering requires an investment in time and money.
  - The situation may have changed by the time the information is acquired.
  - The information may not be worth its cost of acquisition.
- In decision making, additional information only has value when it is likely to alter a decision.

So, the questions are:

1. How do you decide if additional information gathering is worthwhile?
2. What would you pay for additional information?

This can be summarized in one question:
  
  What is the value of additional information?
**Definitions:**

The **Expected Value of Perfect Information** (EVPI) 
= the upper limit of what you would pay for *any* supplemental information

It is calculated before you actually acquire the new information ("preposterior" analysis).

The **Expected Value of Sample Information** (EVSI) 
= what you would be willing to pay for sample information

Any sample information will be less than perfect. 
EVSI does not include the cost of obtaining information (sampling costs)

The **Expected Net Gain of Sampling** (ENGS) 
= EVSI – Sampling Cost

**Example:**

Drilling is costly, and it is not known whether oil is present or not.
So, it looks like you should go ahead with drilling.

What would it be worth to eliminate uncertainty about finding oil?

If drilling, expected payoff = $4M + (0.5)($0k) + (0.3)($10M) + (0.2)($30M) = $5M

If not drilling, sure payoff = $0

Expected Value of Perfect Information (EVPI):

This is the amount that you gain by using perfect information.

- Variable node first, followed by decision
- When does it change your decision? How much is that change worth?
- Sets upper limit on what you should pay for any additional information.
Drill — $4M
Don’t drill — $0M

P = 0.5 Dry
$0k
Don’t drill — $0M
Don’t drill — $0M

P = 0.3 Wet
$10M
Drill — $4M
Don’t drill — $0M

P = 0.2 Gushing
$30M
Drill — $4M
Don’t drill — $0M

Expected payoff = (0.5)($0k) + (0.3)($10M − $4M) + (0.2)($30M − $4M) = $7M

No drilling dry holes saves $4M half of the time!
EVPI = $7M (with knowledge) − $5M (without knowledge) = $2M

But, how can you know ahead of time? Data must exist for you to know.

So, suppose now that you have the opportunity to buy seismic data for $100k and to base your decision on this predictor of oil.

Should you purchase the data?
Seismic data do not detect the presence of oil. They detect rock formations (pockets) that may be favorable to holding oil.

Suppose the seismic data, which you may purchase for $100k, report that the rock formation is one of three types:

“No Structure” – No rock formation favorable to holding oil

“Open Structure” – Incomplete rock formation, somewhat hopeful

“Closed Structure” – Enclosing rock formation, most favorable to holding oil.

None of these findings provide a certainty that there is oil or not, only a likelihood that oil is there.

Keep in mind:
The geological state is first (with dry, wet or gushing condition) before seismic data are obtained from it, not the other way round.

Thus, the probabilities are of the type:
What is the chance that the presence or absence of oil is accompanied by a certain structure?

<table>
<thead>
<tr>
<th>Geological State</th>
<th>Seismic Result</th>
<th>No Structure</th>
<th>Open Structure</th>
<th>Closed Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dry</td>
<td></td>
<td>0.6</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>Wet</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>Gushing</td>
<td>0.1</td>
<td>0.2</td>
<td></td>
<td>0.7</td>
</tr>
</tbody>
</table>

We denote these probabilities as $P(\text{Seismic Result} | \text{Geological State})$ and called them conditional probabilities.

Note: Their sum from left to right = 1, but not necessarily so from top to bottom.
At this stage, you are only contemplating the possible purchase of the seismic data. So, you have not yet seen them and do not know what they will reveal.

You need to ascribe a probability to each of the possible outcomes. For this, we use the probabilities of the geological state $P(\text{Geological State})$.

**Joint probability table**
Certainly, the probability that the well will be found dry is not the same once you know whether the seismic data reveal a lack of structure (highest), an open structure (intermediate), or a closed structure (lowest).

To get the reverse conditional probabilities, \( P(\text{Geological State} | \text{Seismic Result}) \),

\[
\begin{array}{c|c|c|c}
\text{Geological State} & \text{No Structure} & \text{Open Structure} & \text{Closed Structure} \\
\hline
\text{Dry} & 0.30 / 0.41 & 0.15 / 0.31 & 0.05 / 0.28 \\
\text{Wet} & 0.09 / 0.41 & 0.12 / 0.31 & 0.09 / 0.28 \\
\text{Gushing} & 0.02 / 0.41 & 0.04 / 0.31 & 0.14 / 0.28 \\
\hline
\text{TOTAL} & 0.41 / 0.41 & 0.31 / 0.31 & 0.28 / 0.28
\end{array}
\]

\[
\sum = 1
\]
Drop the eliminated options and consolidate the remaining options

Purchase data?

No (no cost)

Yes

- $0.1M

“No Structure”

Drill?

Yes

+$5M

- $4M

$0

0.41

“Open Structure”

Drill?

Yes

+$3.74M

- $4M

$0

0.31

“Closed Structure”

Drill?

Yes

+$14.21M

- $4M

$0

0.28

No

$0

0.41

0.31

0.28
Continue the cleanup

Purchase data?

No

$5M

Yes

−$0.1M

$5.140M

Expected Value of Sample Information:
EVSI = $5.140M − $5M = $0.140M = $140,000

Expected Net Gain of sampling:
ENGS = EVSI − cost of data = $140,000 − $100,000 = $40,000

Since EVSI > 0, it is advantageous to purchase the seismic data.

Since EVSI is modest, the data are only slightly informative.

A hint that this might be the case was that the EVPI was slim ($2M).

Question: If you could negotiate the purchase price, how high would you go?

In closing, the important points to remember about the Value of Information (VoI) Analysis are:

• Increases in value come from changes in decision.

• EVPI can only be calculated if a decision has no effect on the variable being assessed. The variable must be known with certainty, or its various states known with given probabilities.

• If the EVPI is low, it is not likely that gathering more information will be worth it, or the cost of obtaining the data has to remain modest.

• VoI calculations depend on probability estimates. Estimates that are not representative may lead to false impression of VoI. A sensitivity analysis is recommended (= shaking the numbers a bit to see what happens).

• Caution: Watch out for rounding errors, as they may propagate.
Summary of Bayesian Statistics

We start with two different but not independent populations, say $G_i$ (for $i = 1$ to $M$), such as a collection of possible geological states, $S_j$ (for $j = 1$ to $N$), such as a collection of possible seismic results.

We know
- the probabilities $P(G_i)$ from accumulated past evidence or data;
- the conditional probabilities $P(S_j | G_i)$ from science, deduction, or past sampling.

Requirements are: $\sum_{i=1}^{M} P(G_i) = 1$ $\sum_{j=1}^{N} P(S_j | G_i) = 1$ for every $i$

We calculate the joint probabilities: $P(S_j \& G_i) = P(S_j | G_i) \times P(G_i)$

Next, we calculate the $S_j$ probabilities: $\sum_{i=1}^{M} P(S_j \& G_i) = P(S_j)$ with $\sum_{j=1}^{N} P(S_j) = 1$

Finally, we obtain the reversed conditional probabilities:

$$P(G_i | S_j) = \frac{P(S_j \& G_i)}{P(S_j)} = \frac{P(S_j | G_i) \times P(G_i)}{P(S_j)}$$

with $\sum_{i=1}^{M} P(G_i | S_j) = 1$ for every $j$

Bayes’ Theorem