**Information Collection - Key Strategy**

- **Motivation**
  - To reduce uncertainty which makes us choose “second best” solutions as insurance

- **Concept**
  - Insert an information-gathering stage (e.g., a test) before decision problems, as a possibility

```
      Decision Problem
       /           |
      /     
Decision Problem Test
```

- **Operation of Test**

```
New Information
\downarrow
Revision of Prior Probabilities in Decision Problem
\downarrow
New Expected Values in Decision Problem
```

- EV (after test) > EV (without test) Why?
  - Because we can avoid bad choices and take advantage of good ones, in light of test results

- **Questions:**
  - Since test generally has a cost, is the test worthwhile?
  - What is the value of information?
  - Does it exceed the cost of the test?
Essential Concept

- Value of information is an expected value

- Expected Value of information
  \[= \text{EV (after test)} - \text{EV (without test)}\]

Concept application is complex!

- Expected Value of information
  \[= \sum p_k [\text{EV(D}_k^*)] - \text{EV (D*)}\]

Where:
- D* is optimal decision without test
- D_k^* are optimal decisions after test, based on k test results, 
  TR_k, each of which revise probabilities from p_j to p_{jk}
- p_k = probability, after test, of kth observation

- Example:
  Test \[\rightarrow\] Good
  Medium
  Poor

- Revise probabilities after each test result
... as an example

- Expected Value of information
  \[ \sum p_k [EV(D_k^*)] - EV(D^*) \]

- Current best design, \( D^* \)
- ... has an expected value of \( EV(D^*) \)
- We run lab tests with possible outcomes, \( k \)
  (e.g., success, failure, ...), each with prior, \( p_k \)
- Each test leads different best designs, \( D_k^* \)
- Each with a different expected value, \( EV(D_k^*) \)
- For a total value, post test, \( \sum p_k [EV(D_k^*)] \)

Expected Value of Perfect Information  \( EVPI \)

- Perfect information is hypothetical – but simplifies!
- Use: Establishes upper bound on value of any test
- Concept: Imagine a “perfect” test which indicated exactly which Event, \( E_j \), will occur (Cassandra)
  - By definition, this is “best” possible information
  - So, the “best” possible decisions can be made
  - and, the EV gain over the “no test” EV must be the maximum possible; upper limit on the value of any test!
EVPI Example (1)

- Question: Should I wear a raincoat?
  - RC – Raincoat
  - RC - No Raincoat

- Two possible Uncertain Outcomes
  - Rain: \( p = 0.4 \)
  - No Rain: \( p = 0.6 \)

Remember that better choice is to take raincoat, \( EV = 0.8 \)

EVPI Example (2)

- Perfect test
  - Says Rain \( p = 0.4 \) Take R/C 5
  - Says No Rain \( p = 0.6 \) No R/C 4

Why these probabilities? Because these are best estimates of results. Every time it rains, perfect test will say “rain”

- EVPI

\[
EV (\text{after test}) = 0.4(5) + 0.6(4) = 4.4 \\
EVPI = 4.4 - 0.8 = 3.6
\]
Application of EVPI

- A major advantage: EVPI is simple to calculate

- Notice:
  - Prior probability (occurrence of uncertain event) MUST EQUAL probability (associated perfect test result)
  - For “perfect test”, the posterior probabilities are either 1 or 0 (no doubt remains)
  - Optimal choice generally obvious, once we “know” what will happen

- Therefore, EVPI can generally be written directly
- No need to use Bayes’ Theorem

Expected Value of Sample Information  EVSI

- Sample information are results taken from an actual test

- Real Tests can improve estimate, some doubt generally remains

- Value of actual test not as good as hypothetical perfect test: 0 ≤ EVSI ≤ EVPI

- Complex Calculations needed to account for persisting doubts…
EVSI Calculations required

- Obtain probabilities of each result $TR_k$, $p_k$
- For each test result $TR_k$
  - Revise prior probabilities $p_j \Rightarrow p_{jk}$
  - Calculate best decision $D_{k^*}$ (Note: this is a k-fold repetition of the original decision problem!!)
- Calculate $EV$ (after test) = $\Sigma_k p_k(D_{k^*})$
- Calculate EVSI as the difference between $EV$ (after test) - $EV$ (without test)

A BIG JOB

EVSI Example

- Test consists of listening to forecasts
- Two possible test results
  - Rain predicted = RP
  - Rain not predicted = NRP
- Assume probability of a correct forecast = 0.7
  - $p(RP/R) = p(NRP/NR) = 0.7$
    says “rain” or “no rain” and that is what occurs
  - $p(NRP/R) = p(RP/NR) = 0.3$
    the prediction of “rain” or “no rain” is wrong
EVSI Example (1)

- First calculation: probabilities of test results
  \[
  P(RP) = p(RP/R) \cdot p(R) + p(RP/NR) \cdot p(NR) \\
  = (0.7) (0.4) + (0.3) (0.6) = 0.46
  \]
  
  The prediction of “rain” occurs both from correct forecasts when it is to rain AND wrong forecasts when it does not rain.

- \[
  P(NRP) = 1.00 - 0.46 = 0.54
  \]
  Since in this case it either rains or does not.

EVSI Example (2)

- Next: Posterior Probabilities
  \[
  P(R/RP) = p(R) \cdot (p(RP/R)/p(RP)) = 0.4(0.7/0.46) = 0.61
  \]
  \[
  P(NR/NRP) = 0.6(0.7/0.54) = 0.78
  \]

  Therefore,
  - \[
  p(NR/RP) = 0.39 \text{ (false positive – says it will happen and it does not)}
  \]
  - \[
  p(R/RNP) = 0.22 \text{ (false negative – says it will not happen, yet it does)}
  \]
False Positive Example

- Prior Probability of Disease = 0.0001
- Accuracy of Test = \( P(\text{Disease if test predicts}) = P(D|DP) \). Assume = 0.001 and = \( P(\neg D|DNP) \)
- What is \( P(\text{Disease/after test}) \)?

- What is probability that test will report disease?

- Almost all the times you have it: (~1) (0.0001)
- In Error (0.001) (~1) ~ 0.001
- In this case, false positive ~ 10x true positive

---

EVSI Example (3)

- Best decisions conditional upon test results
- First, if rain predicted:

```
  RP ---- D
     /     \
   /       \
 RC 0.61  C
     |    
     |    
     R 5
     NR -2

  RC 0.39  C
     |    
     |    
     R -10
     NR 4
```

\[
EV (RC) = (0.61) (5) + (0.39) (-2) = 2.27
\]
\[
EV (RC) = (0.61) (-10) + (0.39) (4) = -4.54
\]

\[\text{EV} = 2.27 \]
\[\text{EV} = -4.54 \]
EVSI Example (4)

- Best decisions if No Rain Predicted

\[
\begin{align*}
EV &= -0.48 \\
EV &= 0.92
\end{align*}
\]

\[
\begin{align*}
EV (RC) &= (0.22) (5) + (0.78) (-2) = -0.48 \\
EV (RC) &= (0.22) (-10) + (0.78) (4) = 0.92
\end{align*}
\]

EVSI Example (5)

- EV (after test)
  \[
  = p(\text{rain predicted}) \cdot (EV(\text{strategy/RP})) + P(\text{no rain predicted}) \cdot (EV(\text{strategy/NRP}))
  \]
  \[
  = 0.46 (2.27) + 0.54 (0.92) = 1.54
  \]

- EVSI = 1.54 - 0.8 = 0.74
- EVSI < EVPI = 3.6 as indicated earlier
Practical Example: Is a Test Worthwhile? (1)

- If value is Linear (i.e., probabilistic expectations correctly represent value of uncertain outcomes)
  - Calculate EVPI
  - If EVPI < cost of test → Reject test
  - Pragmatic rule of thumb
    - If cost > 50% EVPI → Reject test
    - (Real tests are not close to perfect)
  - Calculate EVSI
  - EVSI < cost of test → Reject test
  - Otherwise, accept test

Is Test Worthwhile? (2)

- If Value Non-Linear (i.e., EV of outcomes does NOT reflect attitudes about uncertainty)
- Theoretically, cost of test should be deducted from EACH outcome that follows a test
  - If cost of test is known
    - A) Deduct costs
    - B) Calculate EVPI and EVSI (cost deducted)
    - C) Proceed as for linear EXCEPT
      - Question is if EVPI(cd) or EVSI(cd) > 0?
  - If cost of test is not known
    - A) Use iterative, approximate pragmatic approach
    - B) Focus first on EVPI
    - C) Use this to estimate maximum cost of a test