

## Information Gathering as a Strategy

Two Topics for today

### 1. Value of Information

As a concept

Methods of Calculation

- Definition of Upper Bound (EVPI)
- Calculation of “exact” value (EVS)

### 2. Information Collection as Basis for Flexible Strategy: It introduces flexible decision points

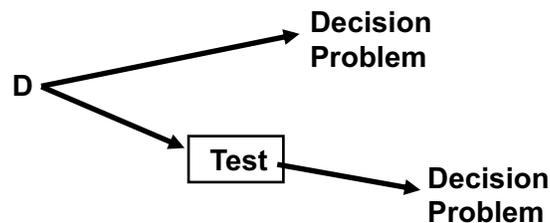
## Information Collection – Key Strategy before major commitment

### Motivation

To reduce uncertainty before we make a major initial commitment to a system design or roll-out

### Concept

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



## What does Test do?

**Starting point: When facing a decision problem, we have uncertainties = “prior probabilities”**

**For example: these may concern**

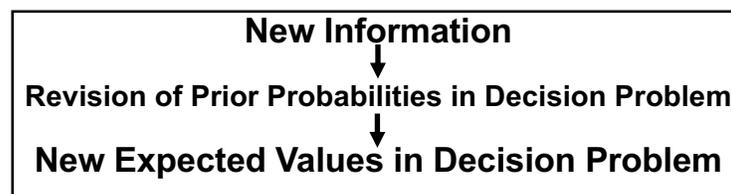
- **Cost of production**
- **Likelihood of sales**
- **Reaction of competitors ...**

**What does test do?**

**It gets information on these issues, for example:**

- **Run a test plant**
- **Carry out a market analysis**
- **Run a test market...**

## Operation of Test



**EV (after test)  $\geq$  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?**

## Value of I. -- Essential Concept

**Value of information is an expected value**

**Expected Value of information  
= EV (after test) - EV (without test)**

## Simple concept – Full Analysis has Complicated Calculations

**Calculations Complex because:**

- **Any test can have many different results. For example, prototype plant might work very well, well, so-so, poorly, fail...**
- **These results will not prove what will happen.... WHY?**
- **They are “Samples”, not “proofs”**
- **They “update” prior probabilities using Bayes’ Theorem – not a simple topic at all**
- **Each test result, implies a different value of project, each with a different probability**

## ... In equation form

---

**Expected Value of information**

$$= \sum p_k [EV(D_k^*)] - EV(D^*)$$

$$= EV(\text{after test}) - EV(\text{without test})$$

**Where in  $EV(D^*)$ ,  $D^*$  is**

- **optimal set of decisions**
- **that is, the optimal strategy**
- **calculated without test**
- **that is, without additional information**

➤ **Likewise,  $D_k^*$  is the optimal decision given test result “k”**

## Meaning of Left-Hand part of equation

$$\sum p_k [EV(D_k^*)] \quad \sim EV(D^*) \text{ but indexed to “k”}$$

**The “k” refer to the different observations or test results  $TR_k$ , that might be possible**

- **Each test result  $TR_k$  has a probability of occurring:  $p_k$**
- **... and also revises the prior probabilities,  $p_j$ ,**
- **... to “posterior” probabilities,  $p_{jk}$**
- **$D_k^*$  are then the optimal decisions, after test, calculated using the revised probabilities,  $p_{jk}$**

## Example

Suppose a company decides to run a prototype plant (or some small-scale system)

They can anticipate possible results, for example, that operation is “good”, “medium” or “poor”

Test ↙ Good  
Medium  
Poor

For each test result, they can calculate how it would affect prior estimates of value of system and their best choice consequentially

E.g.: “poor result” case implies “system failure” implying “optimal decision” = “do not invest”

## Example continued

How should one analyze overall value of test process, in terms of delivering information?

- Estimate probability of each possible Test Result (“good”, “medium”, or “poor”)
- Calculate the value of optimal decision,  $D_k^*$ , resulting from each test result
- Multiply these two respectively and sum these products to get the expected value after the test:

$$\sum p_k [EV(D_k^*)]$$

## ... put another way

**Expected Value of information**

$$= \sum p_k [EV(D_k^*)] - EV(D^*)$$

**Current best design,  $D^*$**

**... has an estimated 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 estimated value,  $EV(D_k^*)$**

**For a total value, post test,  $\sum p_k [EV(D_k^*)]$**

## Why have I shown you this?

**Your take-aways should be:**

- **Full analysis is a complicated process with many possible outcomes**
- **Involves many assumptions (what are the probabilities of outcome of tests)**
- **So full analysis cannot be accurate even if math is correct.**

**Be skeptical of such detailed analyses!**

**Need another way to get value of information!**

## **What do we do in practice?**

---

**Do you feel confident you could calculate value of information?**

**Intuition not obvious**

**Complicated at best -- likely to be tedious**

**A simple approach:**

**Calculate upper bound:**

**“Value of Perfect Information”**

## **Expected Value of Perfect Information EVPI**

---

**Perfect information is hypothetical – but simplifies!**

**Establishes upper bound on value of any test**

**Concept: Imagine a “perfect” test which indicated exactly which Event,  $E_j$ , will occur**

**This is a “Cassandra” machine**

**Who was Cassandra?**

## The Cassandra Machine

- **Concept:** Imagine a Cassandra machine
- This is a black box that predicts exactly
- It indicates exactly which Test Result will occur
- This is “best” possible information

Therefore:

- the “best” possible decisions can be made
- the EV gain over the “no test” EV must be the maximum possible
- Perfect test gives upper limit on value of test!

## Example: Should I wear raincoat?

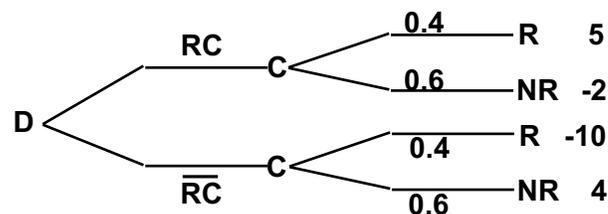
Two possible decisions     

RC – Raincoat

$\bar{RC}$  - No Raincoat

Two possible Uncertain Outcomes

Rain: ( $p = 0.4$ )    No Rain: ( $p = 0.6$ )



- Remember that better choice is to WHAT?

Take raincoat, EV = 0.8

## Organizing Cassandra Machine

- Cassandra Machine



What probabilities for test results?

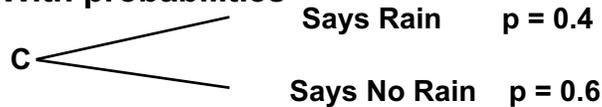
Reason thus: Every time it rains, perfect test will say "rain".

- Our prior estimate of "rain" = estimate that CM will say "rain"



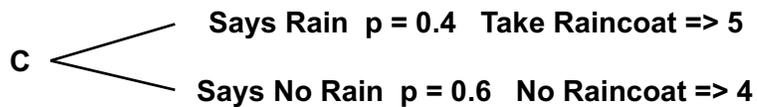
## Set up for EVPI

- With probabilities



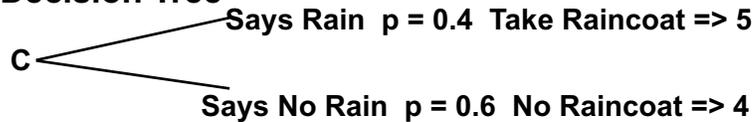
What are the outcomes?

In general, best decisions obvious, given perfect information. If I knew it would rain, best decision would be to wear raincoat.



## Calculation of EVPI

- **Decision Tree**



The Expected Value after the perfect test is easy to calculate. In this case = 4.4

- **EVPI = Difference between and value before. Thus:**

$$\begin{aligned} \text{EV (after test)} &= 0.4(5) + 0.6(4) = 4.4 \\ \text{EVPI} &= 4.4 - 0.8 = 3.6 \end{aligned}$$

## 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 (the mechanism for updating prior probabilities given new info)**

## Is test worthwhile?

Previous analyses have focused on the value of having information

This is not enough to know if it is worthwhile to run the test

Why is this?

Because we must compare the value of the test to its cost:

is the expected value of information from a test sufficiently greater than its cost?

## Practical Example: Is a Test Worthwhile? (1)

If value is Linear (i.e., probabilistic expectations correctly represent value of uncertain outcomes)

Calculate EVPI

If  $EVPI < \text{cost of test}$  → Reject test

Pragmatic rule of thumb

If  $\text{cost} > 50\% EVPI$  → Reject test

(Real test are not close to perfect)

Calculate EVSI

$EVSI < \text{cost of test}$  → Reject test

Otherwise, accept test

## **Information Collection As Basis for Flexible Strategy**

---

**Note that Decision to Collect Information  
... is a decision to insert flexibility into  
development strategy**

**Why is this?**

**Because rationale for “test” is that you might  
change your decision once you have results**

**This is a decision to insert a certain flexibility**

**Value of Information = Value of this flexibility**

## **Information Collection Only one form of flexibility**

---

**Information Collection changes the process of  
design development**

**It does not inherently change the physical  
design itself**

## Take-Aways

**Information Collection can increase value significantly**

**Thus we need to estimate the expected value of the information – in terms of improved value of out optimal strategies**

**Expected Value of Perfect Information is a useful way to establish upper bound on value**

**Information Collection is one way to insert flexibility in the development process**

## Sonny Reyes exercises

### 16.9 *Sonny's PV's*

Sonny Reyes (see Problem 15.6) must decide how to manufacture the PV panels. He has three choices:

- Develop a new method
- Alter existing methods
- Get an outside firm to produce them

Developing the new method would yield a profit of \$11 million if successful, and if it cannot be developed, the outside firm must be used for a profit of \$2M. There is a 70% chance that the new method will be successfully developed.

Altering the existing method successfully will yield a profit of \$7M, and there is a 90% chance of the alterations being successful. If not successful, the outside firm must be used for a profit of \$3M. If used immediately, the outside firm will definitely be able to produce the panels and this would lead to a profit of \$5M.

### 17.9. *Sonny and Ex Reyes*

Sonny's former wife, Ex, comes along and says she can do research that will investigate the probability of success for the new method of producing PV panels (see Problem 16.9). Ex wants \$1 million for her research.

- (a) Draw the decision after the test.
- (b) Should Sonny pay Ex Reyes's price? Explain.

---

This page is intentionally left blank.

## **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:  $0 \leq \text{EVSI} \leq \text{EVPI}$**

**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) =  $\sum_k p_k (D_k^*)$

Calculate EVSI as the difference between  
EV (after test) - EV (without test)

**A BIG JOB**

## EVSI Example

---

Test consists of getting forecasts from Internet

Two possible test results

- Rain predicted = RP
- Rain not predicted = NRP

Assume probability of a correct forecast = 0.7  
(as might be obtained from data analysis)

- $p(RP/R) = p(NRP/NR) = 0.7$   
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: Probabilities of test results

**Crucial Idea:** prediction “rain” occurs from  
correct forecasts when it is to rain AND  
wrong forecasts when it does not rain  
that is: correct answers and “mistakes”

$$\begin{aligned} P(RP) &= p(RP/R) p(R) + p(RP/NR) p(NR) \\ &= (0.7) (0.4) + (0.3) (0.6) = 0.46 \end{aligned}$$

$$P(NRP) = 1.00 - 0.46 = 0.54$$

Since in this simple case it either rains or does not

## EVSI Example: Posterior Probabilities

This is the revision of prior probabilities of  
rain or not, 0.4 and 0.6 respectively

Bayes theorem:  $P(E/O) = P(E) [P(O/E) / P(O)]$

$$P(R/RP) = p(R) (p(RP/R)/p(RP)) = 0.4(0.7/0.46) = 0.61$$

$$P(NR/NRP) = 0.6(0.7/0.54) = 0.78$$

Therefore,

$p(NR/RP) = 0.39$  (false positive – says it will happen  
and it does not)

$p(R/RNP) = 0.22$  (false negative – says it will not  
happen, yet it does)

## False Positive Example

**Prior Probability of Disease = 0.0001**

**Accuracy of Test = P(Disease if test predicts) =  
P (D/DP). Assume = 0.001 and = P(ND/DNP)**

**What is P(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: Value of EVSI

**Expected value after the realistic, imperfect,  
test = 1.54**

**Expected value without test = 0.8**

**Thus: EVSI = 1.54 - 0.8 = 0.74**

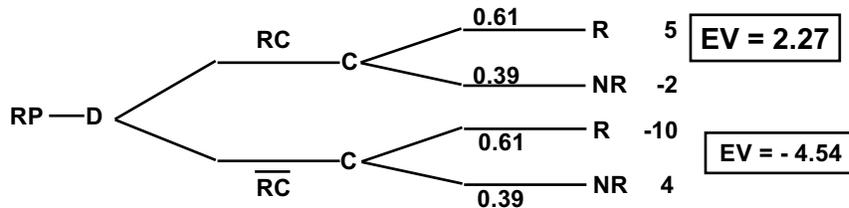
**Note that, as indicated earlier :**

**EVSI = 0.74 < EVPI = 3.6**

## EVSI Example: Best decision, test says “rain”

Best decision conditional upon test results

First, if rain predicted:



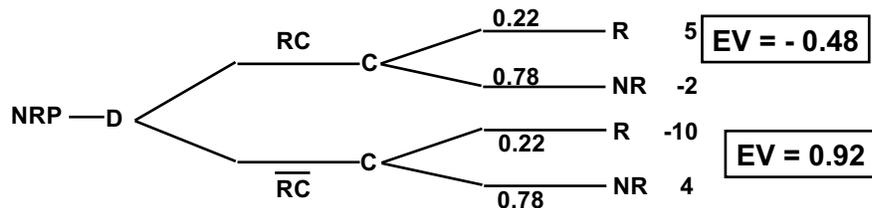
$$EV(RC) = (0.61)(5) + (0.39)(-2) = 2.27$$

$$EV(\overline{RC}) = (0.61)(-10) + (0.39)(4) = -4.54$$

**TAKE RAINCOAT**

## EVSI Example: Best decision, test says “no rain”

• Best decision if No Rain Predicted



$$EV(RC) = (0.22)(5) + (0.78)(-2) = -0.48$$

$$EV(\overline{RC}) = (0.22)(-10) + (0.78)(4) = 0.92$$

**DO NOT TAKE RAINCOAT**

## **EVSI Example: Expected Value after test**

**This is Expected Value over best decisions for each test result**

**EV (after test)**

$$\begin{aligned} &= P(\text{rain predicted}) (\text{EV}(\text{strategy}/\text{RP})) \\ &+ P(\text{no rain predicted}) (\text{EV}(\text{strategy}/\text{NRP})) \end{aligned}$$

$$= 0.46 (2.27) + 0.54 (0.92) = 1.54$$

**Note: EV(after test) > EV (without test)**

$$1.54 > 0.8$$

MIT OpenCourseWare  
<https://ocw.mit.edu/>

IDS.333 Risk and Decision Analysis  
Fall 2021

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.