Decision Theory and Bayesian Inference I

PURPOSE
To equip the students with skills to build statistical models for non-trivial problems when data is sparse and expert opinion needs to be incorporated and to use the key features of a Bayesian problem and algorithms for Bayesian Analysis.

OBJECTIVES
By the end of this course the student should be able to;
(i) Explain and apply Bayes’ rule and other decision rules.
(ii) Explain the likelihood principle and derive a posterior distribution from a prior distribution.
(iii) Perform classification, hypothesis testing and estimation.
(iv) Explain the subjectivism point of view.
(v) Apply Bayesian inference and analysis for the normal and binomial distributions.
(vi) Apply the basic concepts in decision analysis.

DESCRIPTION

PRE-REQUISITES: STA 2300 Theory of Estimation, STA 2304 Decision theory

COURSE TEXT BOOKS

COURSE JOURNALS
i) Bayesian Analysis
ii) Journal of the Royal Statistical Society (Series B)

FURTHER REFERENCE TEXT BOOKS AND JOURNALS:
(viii) American Statistician
(ix) Journal of the American Society
(x) Biometrics

TEACHING METHODS
a) Lecture: oral presentation generally incorporating additional activities, e.g. writing on a chalk-board, exercises, class questions and discussions, or student presentations.
b) Tutorial: to give the students more attention.
Decision Theory

Introduction
A decision may be defined as the process of choosing an action (solution) to a problem from a set of feasible alternatives. In choosing the optimal solution, it means we have a set of possible other solutions. In decision theory, the focus is on the process of finding the action yielding the best results. Statistical decision theory is concerned with which decision from a set of feasible alternatives is optimal for a particular set of conditions.

In many situations, selection of a course of action is vital especially where the situation being considered has a significant monetary or social impact and is accompanied by some degree of uncertainty. Under this situation, decision theory becomes an important analytical tool that can be used to provide a favorable approach to the selection of a particular course of action.

Decision analysis utilizes the concept of gain and loss (profit & cost) associated with every possible action the decision-maker can select. Nearly all facets/fields of life require decision-making e.g. marketing, production, finance, etc.

Example (marketing Decision Problem)
A motor company must decide whether to purchase assembled door locks for the new vehicle model or manufacture & assemble the parts at their plant. If the sales of the new model continue to increase, it will be profitable to manufacture & assemble the parts. However, if the sales level continue to decrease, it will be more profitable to purchase already assembled parts. Which decision should the management make?

Elements of a Decision
There are 3 elements or components for any decision-making situation. They include:-

1. The alternatives/actions/decisions-They are the choices available to the decision-maker and they are assured to be mutually exclusive and exhaustive events.
   - The decision-maker has control over the alternatives.

2. States of nature:-These are uncontrollable future events that affect the outcome of a decision.
   - They are outside the control of the decision-maker and they are also assured to be mutually exclusive & exhaustive. Eg. In the previous example, the company does not know whether the demand will be high or low.

3. Pay offs:- Each combination of a course of action and the state of nature is associated with a pay-off which measures the net benefit for the decision-maker.
   - Pay-off can be defined as a quantitative measure of the value of the outcome of an action.

Note If an action does not involve risk, the pay-off will be the same regardless of which state of nature occurs.

Decision Making Environment
Selection of a terminal (final) decision is mainly influenced by the state of nature surrounding the problem. Usually, actions & their associated pay-offs are known in advance to their decision-maker. However, the decision maker does not know which alternative will be the best in each case. He/she knows with certainty the state of nature that will be in effect.
We have 3 different types of decision-making environments namely:

(i) Decision-making under the environment of certainty
(ii) Decision-making under conditions of risk
(iii) Decision-making under conditions of uncertainty

1. Decision-Making under Conditions of Certainty

This is a situation where the states of nature are known, and if they are not known, they have no influence on the outcome of the decision. In this case, it is easy for the decision-maker to maximize monetary gains or any other type of utility. Decision-making under conditions of certainty is straightforward unless the number of alternatives is very large.

The availability of partial/imperfect information about a problem results in the other two decision-making environments.

2. Decision-Making under Conditions of Risk

Here the decision-making process is conducted in an environment of risk and the degree of knowledge associated with each state of nature is unknown. However, an appropriate estimate of the likelihood can be determined based on a subjective judgment or use of some mathematical relations.

3. Decision-Making under Conditions of Uncertainty

Here, more than one outcome can be rendered from only a single decision and probabilities can be associated with values. This complicates the decision making because the decision-maker is constrained due to lack of knowledge about probabilities or the occurrences of the outcomes of possible course of action only pay-offs are known and nothing known about the likelihood (probability) of each state of nature.

Note Data availability should point certainty & uncertainty about a problem. Risk lies between certainty and uncertainty.

Pay-off Tables

Let $A_1, A_2, \ldots, A_n$ be the possible actions and $S_1, S_2, \ldots, S_m$ be the states of nature. Also, let the return associated with action $A_i$ and state of nature $S_j$ be $V_{ij}$ or $V(A_i S_j)$ for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$.

Note for each combination of action $A_i$ and state of nature $S_j$ the decision maker knows what the resulting pay-off will be. The following is the possible pay-off:

<table>
<thead>
<tr>
<th>Alternatives</th>
<th>State of Nature</th>
<th>$S_1$</th>
<th>$S_2$</th>
<th>$\ldots$</th>
<th>$S_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>$V_{11}$</td>
<td>$V_{12}$</td>
<td>$\ldots$</td>
<td>$V_{1m}$</td>
<td></td>
</tr>
<tr>
<td>$A_2$</td>
<td>$V_{21}$</td>
<td>$V_{22}$</td>
<td>$\ldots$</td>
<td>$V_{2m}$</td>
<td></td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\ldots$</td>
<td>$\vdots$</td>
<td></td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\ldots$</td>
<td>$\vdots$</td>
<td></td>
</tr>
<tr>
<td>$A_n$</td>
<td>$V_{n1}$</td>
<td>$V_{n2}$</td>
<td>$\ldots$</td>
<td>$V_{nm}$</td>
<td></td>
</tr>
<tr>
<td>Prob</td>
<td>$P(S_1)$</td>
<td>$P(S_2)$</td>
<td>$\ldots$</td>
<td>$P(S_m)$</td>
<td></td>
</tr>
</tbody>
</table>
To use the above pay-off table to reach an optimal action, the state of nature is taken to be a random variable, an additional element i.e the probability distribution for each state of nature is needed in decision-making frame-work.

**Example**  Company A owns a track of land that contains oil. A consulting geologist  has reported to the company management that she believes that there is one chance in four of land having oil. Due to this prospect another company B has offered to purchase the land for $90,000. However company A can hold the land and drill for oil itself. If oil is found the company’s expected profit is $700,000. A loss of $100,000 will be incurred if the land is dry.

The management of company A has 2 possible actions.

- A₁: sell the land to company B at $ 90,000
- A₂: drill for oil at a cost of $ 100,000, and make a profit of $700,000 if oil is found.

The possible states of nature are

- S₁: the land contains oil with corresponding probability of 0.25
- S₂: the land is dry with corresponding probability of 0.75

The Pay-off table is as shown alongside:

<table>
<thead>
<tr>
<th>State of Nature</th>
<th>Actions</th>
<th>S₁</th>
<th>S₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁</td>
<td>$90</td>
<td>$90</td>
<td></td>
</tr>
<tr>
<td>A₂</td>
<td>$700</td>
<td>$-100</td>
<td></td>
</tr>
</tbody>
</table>

The Prior Probabilities are 0.25 for S₁ and 0.75 for S₂.

**Exercise 1**

1. Getz Products Company is investigating the possibility of producing and marketing backyard storage sheds. Undertaking this project would require the construction of either a large or a small manufacturing plant. The market for the product produced—storage sheds—could be either favorable or unfavorable. Getz, of course, has the option of not developing the new product line at all. With a favorable market, a large facility will give Getz Products a net profit of $200,000. If the market is unfavorable, a $180,000 net loss will occur. A small plant will result in a net profit of $100,000 in a favorable market, but a net loss of $20,000 will be encountered if the market is unfavorable. Construct a payoff table, indicating the events and alternative courses of action.

2. You are a marketing manager for a food products company, considering the introduction of a new brand of organic salad dressings. You need to develop a marketing plan for the salad dressings in which you must decide whether you will have a gradual introduction of the salad dressings (with only a few different salad dressings introduced to the market) or a concentrated introduction of the salad dressings (in which a full line of salad dressings will be introduced to the market). You estimate that if there is a low demand for the salad dressings, your first year’s profit will be $1 million for a gradual introduction and -$5 million (a loss of $5 million) for a concentrated introduction. If there is high demand, you estimate that your first year’s profit will be $4 million for a gradual introduction and $10 million for a concentrated introduction. Suppose that the probability is 0.60 that there will be low demand, construct a payoff table for these two alternative courses of action.

3. The DellaVecchia Garden Center purchases and sells Christmas trees during the holiday season. It purchases the trees for $10 each and sells them for $20 each. Any trees not sold by Christmas day are sold for $2 each to a company that makes wood chips. The garden center estimates that four levels of demand are possible: 100, 200, 500, and 1,000 trees. Compute the payoffs for purchasing 100, 200, 500, or 1,000 trees for each of the four levels of demand. Construct a payoff table, indicating the events and alternative courses of action. Construct a decision tree.
4. Zed and Adrian and run a small bicycle shop called "Z to A Bicycles". They must order bicycles for the coming season. Orders for the bicycles must be placed in quantities of twenty (20). The cost per bicycle is $70 if they order 20, $67 if they order 40, $65 if they order 60, and $64 if they order 80. The bicycles will be sold for $100 each. Any bicycles left over at the end of the season can be sold (for certain) at $45 each. If Zed and Adrian run out of bicycles during the season, then they will suffer a loss of "goodwill" among their customers. They estimate this goodwill loss to be $5 per customer who was unable to buy a bicycle. Zed and Adrian estimate that the demand for bicycles this season will be 10, 30, 50, or 70 bicycles with probabilities of 0.2, 0.4, 0.3, and 0.1 respectively.

5. Finicky's Jewelers sells watches for $50 each. During the next month, they estimate that they will sell 15, 25, 35, or 45 watches with respective probabilities of 0.35, 0.25, 0.20, and ... (figure it out). They can only buy watches in lots of ten from their dealer. 10, 20, 30, 40, and 50 watches cost $40, 39, 37, 36, and 34 per watch respectively. Every month, Finicky's has a clearance sale and will get rid of any unsold watches for $24 (watches are only in style for a month and so they have to buy the latest model each month). Any customer that comes in during the month to buy a watch, but is unable to, costs Finicky's $6 in lost goodwill. Find the best action under each of the four decision criteria.

6. An author is trying to choose between two publishing companies that are competing for the marketing rights to her new novel. Company A has offered the author $10,000 plus $2 per book sold. Company B has offered the author $2,000 plus $4 per book sold. The author believes that five levels of demand for the book are possible: 1,000, 2,000, 5,000, 10,000, and 50,000 books sold.
   a) Compute the payoffs for each level of demand for company A and company B.
   b) Construct a payoff table, indicating the events and alternative courses of action.
   c) Construct a decision tree.
   d) Construct an opportunity loss table.

7. A supermarket chain purchases large quantities of white bread for sale during a week. The stores purchase the bread for $0.75 per loaf and sell it for $1.10 per loaf. Any loaves not sold by the end of the week can be sold to a local thrift shop for $0.40. Based on past demand, the probability of various levels of demand is as follows:

<table>
<thead>
<tr>
<th>Demand (Loaves)</th>
<th>6,000</th>
<th>8,000</th>
<th>10,000</th>
<th>12,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.10</td>
<td>0.50</td>
<td>0.30</td>
<td>0.10</td>
</tr>
</tbody>
</table>

   a) Construct a payoff table, indicating the events and alternative courses of action.
   b) Construct a decision tree.
   c) Compute the expected monetary value for purchasing 6,000, 8,000, 10,000, and 12,000 loaves.
   d) Compute the expected opportunity loss for purchasing 6,000, 8,000, 10,000, and 12,000 loaves
   e) Compute the coefficient of variation for each purchase level.
   f) Compute the return-to-risk ratio for each purchase level. What action would you choose? Why?
   g) Suppose that new information changes the probabilities associated with the demand level. Use the following probabilities to repeat (c) through (j):

<table>
<thead>
<tr>
<th>Demand (Loaves)</th>
<th>6,000</th>
<th>8,000</th>
<th>10,000</th>
<th>12,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.30</td>
<td>0.40</td>
<td>0.20</td>
<td>0.10</td>
</tr>
</tbody>
</table>

8. The owner of a company that supplies home heating oil would like to determine whether to offer a solar heating installation service to its customers. The owner of the company has determined that a start-up cost of $150,000 would be necessary, but a profit of $2,000 can be made on each solar heating system installed. The owner estimates the probability of various demand levels as follows: How would your answers to (a) through (h) be affected if the startup cost were $200,000?

<table>
<thead>
<tr>
<th>No of Units Installed</th>
<th>50</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.40</td>
<td>0.30</td>
<td>0.30</td>
</tr>
</tbody>
</table>

   a) Construct a payoff table, indicating the events and alternative courses of action.
b) Construct a decision tree.

c) Construct an opportunity loss table.

d) Compute the expected monetary value for offering this solar heating system installation service.

e) Compute the expected opportunity loss for offering this solar heating system installation service.

f) Explain the meaning of the expected value of perfect information (EVPI) in this problem.

g) Compute the return-to-risk ratio (RTRR) for offering this solar heating system installation service.

h) Based on the results of (d), (e), and (g), should the company offer this solar heating system installation service? Why?

9. H. Weiss, Inc., is considering building a sensitive new airport scanning device. His managers believe that there is a probability of 0.4 that the ATR Co. will come out with a competitive product. If Weiss adds an assembly line for the product and ATR Co. does not follow with a competitive product, Weiss’s expected profit is $40,000; if Weiss adds an assembly line and ATR follows suit, Weiss still expects $10,000 profit. If Weiss adds a new plant addition and ATR does not produce a competitive product, Weiss expects a profit of $600,000; if ATR does compete for this market, Weiss expects a loss of $100,000.

10. Ronald Lau, chief engineer at South Dakota Electronics, has to decide whether to build a new state-of-the-art processing facility. If the new facility works, the company could realize a profit of $200,000. If it fails, South Dakota Electronics could lose $180,000. At this time, Lau estimates a 60% chance that the new process will fail. The other option is to build a pilot plant and then decide whether to build a complete facility. The pilot plant would cost $10,000 to build. Lau estimates a 50-50 chance that the pilot plant will work. If the pilot plant works, there is a 90% probability that the complete plant, if it is built, will also work. If the pilot plant does not work, there is only a 20% chance that the complete project (if it is constructed) will work. Lau faces a dilemma. Should he build the plant? Should he build the pilot project and then make a decision?

11. Joseph Biggs owns his own sno-cone business and lives 30 miles from a California beach resort. The sale of sno-cones is highly dependent on his location and on the weather. At the resort, his profit will be $120 per day in fair weather, $10 per day in bad weather. At home, his profit will be $70 in fair weather and $55 in bad weather. Assume that on any particular day, the weather service suggests a 40% chance of foul weather. Construct Joseph’s pay off table.

12. Kenneth Boyer is considering opening a bicycle shop in North Chicago. Boyer enjoys biking, but this is to be a business endeavor from which he expects to make a living. He can open a small shop, a large shop, or no shop at all. Because there will be a 5-year lease on the building that Boyer is thinking about using, he wants to make sure he makes the correct decision. Boyer is also thinking about hiring his old marketing professor to conduct a marketing research study to see if there is a market for his services. The results of such a study could be either favorable or unfavorable. Develop a decision tree for Boyer.

13. Kenneth Boyer (of Problem A.18) has done some analysis of his bicycle shop decision. If he builds a large shop, he will earn $60,000 if the market is favorable; he will lose $40,000 if the market is unfavorable. A small shop will return a $30,000 profit with a favorable market and a $10,000 loss if the market is unfavorable. At the present time, he believes that there is a 50-50 chance of a favorable market. His former marketing professor, Y. L. Yang, will charge him $5,000 for the market research. He has estimated that there is a .6 probability that the market survey will be favorable. Furthermore, there is a .9 probability that the market will be favorable given a favorable outcome of the study. However, Yang has warned Boyer that there is a probability of only .12 of a favorable market if the marketing research results are not favorable. Expand the decision tree of Problem A.18 to help Boyer decide what to do.

14. Dick Holliday is not sure what he should do. He can build either a large video rental section or a small one in his drugstore. He can also gather additional information or simply do nothing. If he gathers additional information, the results could suggest either a favorable or an unfavorable market, but it
would cost him $3,000 to gather the information. Holliday believes that there is a 50-50 chance that the information will be favorable. If the rental market is favorable, Holliday will earn $15,000 with a large section or $5,000 with a small. With an unfavorable video-rental market, however, Holliday could lose $20,000 with a large section or $10,000 with a small section. Without gathering additional information, Holliday estimates that the probability of a favorable rental market is .7. A favorable report from the study would increase the probability of a favorable rental market to .9. Furthermore, an unfavorable report from the additional information would decrease the probability of a favorable rental market to .4. Of course, Holliday could ignore these numbers and do nothing. What is your advice to Holliday?

15. Louisiana is busy designing new lottery “scratch-off” games. In the latest game, Bayou Boondoggle, the player is instructed to scratch off one spot: A, B, or C. A can reveal “Loser,” “Win $1,” or “Win $50.” B can reveal “Loser” or “Take a Second Chance.” C can reveal “Loser” or “Win $500.” On the second chance, the player is instructed to scratch off D or E. D can reveal “Loser” or “Win $10.” The probabilities at A are .9, .09, and .01. The probabilities at B are .8 and .2. The probabilities at C are .999 and .001. The probabilities at D are .5 and .5. Finally, the probabilities at E are .95 and .05. Draw the decision tree that represents this scenario. Use proper symbols and label all branches clearly. Calculate the expected value of this game.

Criteria for Decision-Making under Uncertainty
Here we consider decision-making under the assumptions that no probability distributions are available. The methods to be presented here include,

a) Maximax /Optimistic criterion
b) Maximin (Minimax)/Conservative/Pessimistic Approach.
c) Laplace/Equally likely Approach
d) Minimax Regret/Savage criterion
e) Hurwicz/Realism/Approach

a) Maximax/Optimistic Approach
The approach would be used by an optimistic decision-maker. Based on this approach we choose the decision/action with the largest profit if the pay-offs are gains, otherwise, if the pay-offs are costs/losses, choose the action with the lowest loss/costs. we choose A_i which satisfies.

$$\max_{A_i} \left( \max_{S_j} V_{ij} \right) \quad \text{if } V_{ij} \text{ are profits}$$
$$\min_{A_i} \left( \min_{S_j} V_{ij} \right) \quad \text{if } V_{ij} \text{ are costs}$$

Example 1 Consider the following problem with two alternatives and two states of nature. The payoff table below consists of profits.

<table>
<thead>
<tr>
<th>State of Nature</th>
<th>S_1</th>
<th>S_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A_1</td>
<td>$20</td>
<td>$6</td>
</tr>
<tr>
<td>A_2</td>
<td>$25</td>
<td>$3</td>
</tr>
</tbody>
</table>

Using the optimistic approach, obtain the optimal decision.

<table>
<thead>
<tr>
<th>Actions</th>
<th>S_j</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_1</td>
<td>$20</td>
<td>- maximax</td>
</tr>
<tr>
<td>A_2</td>
<td>$25</td>
<td></td>
</tr>
</tbody>
</table>

Optimal course of action is A_2 NB if V_{ij} were costs the best course of action would be A_1
b) Maximin/Conservative Approach

Here, we list the minimum pay off for each action (across all the states of nature) and then select the action corresponding to the maximum of these minimums. This is called ‘maximizing the lowest possible profit’. However, if the payoffs are in terms of cost, list the maximum costs for each alternative and then select the action with the lowest maximum (i.e., maximum possible cost is minimized). This criterion is based on making the best out of the worst possible condition, and we choose $A_i$ which satisfies.

$$\max_{A_i} \left( \min_{S_j} V_{ij} \right) \quad \text{if } V_{ij} \text{ are profits and } \min_{A_i} \left( \max_{S_j} V_{ij} \right) \quad \text{if } V_{ij} \text{ are costs}$$

**Example 2** Re-do example 1 using the conservative approach. If $v_{ij}$ are costs

**Solution:**

<table>
<thead>
<tr>
<th>Actions</th>
<th>Row Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>$6$</td>
</tr>
<tr>
<td>$A_2$</td>
<td>$3$</td>
</tr>
</tbody>
</table>

Therefore choose action $A_1$.

c) Laplace Criterion

This criteria is based on the principal of equally likely events. Basically, we average the pay-offs for $A_i$ over $S_j$ and then pick the action with the highest average. However, if the pay-offs are costs, we pick the action with the lowest average. i.e, we select $A_i$ which satisfies.

$$\max_{A_i} \left( \frac{1}{m} \sum_{j=1}^{m} V_{ij} \right) \quad \text{if } V_{ij} \text{ are profits and } \min_{A_i} \left( \frac{1}{m} \sum_{j=1}^{m} V_{ij} \right) \quad \text{if } V_{ij} \text{ are costs}$$

**Example 3** Re-do example 1 using the laplace approach

**Solution:**

<table>
<thead>
<tr>
<th>Actions</th>
<th>Row Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>$13$</td>
</tr>
<tr>
<td>$A_2$</td>
<td>$14$</td>
</tr>
</tbody>
</table>

Choose action $A_2$ as the optimal.

d) Savage Minimax Regret criterion

This approach requires the calculation regret/opportunity loss table as following

Let $R_{ij}$ be the regret/loss incurred by taking action $A_i$ under state $S_j$, i.e.

$$R_{ij} = \begin{cases} \max_{S_j} V_{ij} - V_{ij} & \text{if are profits or gain} \\ V_{ij} - \min_{S_j} V_{ij} & \text{if are costs or losses} \end{cases}$$

The table generated by $R_{ij}$ is called regret matrix. Specifically, for each state or nature obtain the difference between each pay-off and the largest pay-off if pay-offs are gains. Using the regret table, we list the maximum regret for each possible action and then select the action with the lowest regret. Effectively, Select $A_i$ such that:

$$\min_{A_i} \left( \max_{S_j} R_{ij} \right) \quad \text{no matter whether } V_{ij} \text{ are profit or costs}$$

**Example 4** Re-do example 1 using the minimax regret criterion.
Time is precious, but we do not know yet how precious it really is. We will only know when we are no longer able to take advantage of it...

**Proverbs 21:5** The plans of the diligent lead to profit as surely as haste leads to poverty

---

**Solution**

Regret matrix::- Since we have profits  \( R_{ij} = \max_{S_j} V_{ij} - V_{ij} \)

Regret Table

<table>
<thead>
<tr>
<th>Actions</th>
<th>state of Nature</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( S_1 )</td>
<td>( S_2 )</td>
</tr>
<tr>
<td>( A_1 )</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

\( A_2 \) has the minimax regret therefore it’s the best course of action.

**e) Hurwicz /Realism Approach**

Also called the weighted average/realism approach; and it’s a compromise between the most optimistic and the most pessimistic decision. It strikes a balance between extreme optimism and extreme pessimism by weighing the two conditions using respective weights \( \alpha \) and \( 1 - \alpha \) where \( 0 < \alpha < 1 \). The parameter \( \alpha \) is called the index of optimism. Select action \( \alpha \) that yields:

\[
\max_{A_i} \{ \alpha \max_{S_j} V_{ij} + (1 - \alpha) \min_{S_j} V_{ij} \} \quad \text{if } V_{ij} \text{ are profits or gains}
\]

\[
\min_{A_i} \{ \alpha \min_{S_j} V_{ij} + (1 - \alpha) \max_{S_j} V_{ij} \} \quad \text{if } V_{ij} \text{ are costs or losses}
\]

**Note** When \( \alpha = 1 \), the decision-maker is optimistic and if \( \alpha = 0 \), he/she is pessimistic about the future.

**Example 5** Re-do example 1 using hurwicz criterion with \( \alpha = 0.55 \)

**Solution:**

\[
\begin{array}{l|ll}
\text{Action} & 0.55\max + 0.45\min & \text{Row Average} \\
\hline
A_1 & 0.55(20)+0.45(6)=13.7 & 18.8 \\
A_2 & 0.55(25)+0.45(3)=15.1 & 15.4 \\
\end{array}
\]

The best course of action is action \( A_2 \) since it has the maximum value

**Example 6** A recreation facility must decide on the level of supply it must stock to meet the needs of its customers during one of the holidays. The exact number of customers is unknown but it’s expected to be in one of the categories 200, 250, 300, or 350. Four levels of supply are suggested and the payoff table below gives the pay-offs for each combination of the level of stock \( A_i \) and the category of customers \( S_j \) use all the 5 criterions above (for Hurwicz, use \( \alpha = 0.4 \) ) to find the optimal course of action.

The values in pay-off table are cost.

**Solution**

<table>
<thead>
<tr>
<th>Action</th>
<th>Row Min</th>
<th>Row Max</th>
<th>Row Average</th>
<th>0.4min +0.6max</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1 )</td>
<td>5</td>
<td>28</td>
<td>15.25</td>
<td>18.8</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>7</td>
<td>23</td>
<td>11.5</td>
<td>16.6</td>
</tr>
<tr>
<td>( A_3 )</td>
<td>12</td>
<td>21</td>
<td>18</td>
<td>17.4</td>
</tr>
<tr>
<td>( A_4 )</td>
<td>15</td>
<td>30</td>
<td>21</td>
<td>24</td>
</tr>
</tbody>
</table>

---

**Category of customers**

<table>
<thead>
<tr>
<th>( S_1 )</th>
<th>( S_2 )</th>
<th>( S_3 )</th>
<th>( S_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1 )</td>
<td>5</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>8</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>( A_3 )</td>
<td>21</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>( A_4 )</td>
<td>30</td>
<td>22</td>
<td>17</td>
</tr>
</tbody>
</table>

**Proverbs 21:5** The plans of the diligent lead to profit as surely as haste leads to poverty
a). Optimistic approach (minimum)
\[
\min_{A_i} \left( \min_{S_j} V_{ij} \right) = 5 \text{ corresponding to } A_1 \Rightarrow A_1 \text{ is the best course of action}
\]
\[
\min_{A_i} \left( \max_{S_j} V_{ij} \right) = 21 \text{ corresponding to } A_3 \Rightarrow A_3 \text{ is the best course of action}
\]
c). Laplace Criterion
\[
\min_{A_i} \left\{ \text{Average} \right\} = 11.5 \text{ corresponding to } A_2 \Rightarrow A_2 \text{ is the optimal course of action}
\]
d). Savage Regret approach.
\[
\begin{array}{cccccc}
\text{Category of customers} & S_1 & S_2 & S_3 & S_4 & \text{Max} \\
\hline
A_1 & 0 & 3 & 10 & 13 & 13 \\
A_2 & 3 & 0 & 0 & 8 & 8 \\
A_3 & 16 & 11 & 4 & 6 & 16 \\
A_4 & 25 & 15 & 9 & 0 & 25 \\
\end{array}
\]
\[
\min_{A_i} \left( \max_{S_j} R_{ij} \right) = 8 \text{ corresponding to } A_2 \Rightarrow A_2 \text{ is the best course of action}
\]
e). Hurwicz Criterion See table 1
\[
\min_{A_i} \left\{ 0.4 \min_{S_j} V_{ij} + 0.6 \max_{S_j} V_{ij} \right\} = 16.6 \text{ corresponding to } A_2 \Rightarrow A_2 \text{ is the optimal action}
\]

Exercise 2

1. For the following payoff table, determine the optimal action based on the maximax, the maximin and the Laplace criterions

\[
\begin{array}{cccc}
\text{Actions} & S_1 & S_2 & \text{Max} \\
A_1 & 50 & 200 & \min_{A_i} \left( \max_{S_j} V_{ij} \right) = 5 \text{ corresponding to } A_1 \Rightarrow A_1 \\
A_2 & 100 & 125 & \min_{A_i} \left( \min_{S_j} V_{ij} \right) = 10 \text{ corresponding to } A_2 \Rightarrow A_2 \\
\end{array}
\]

2. Obtain the optimal decision for the following pay off tables ($V_{ij}$ are profits) using the 5 criterions above

\[
\begin{array}{cccc}
\text{Actions} & S_1 & S_2 & \text{Max} \\
A_1 & 100 & 0 & \min_{A_i} \left( \max_{S_j} V_{ij} \right) = 35 \text{ corresponding to } A_1 \Rightarrow A_1 \\
A_2 & 50 & 20 & \min_{A_i} \left( \min_{S_j} V_{ij} \right) = 35 \text{ corresponding to } A_2 \Rightarrow A_2 \\
\end{array}
\]

\[
\begin{array}{cccc}
\text{Actions} & S_1 & S_2 & S_3 & \text{Max} \\
A_1 & 16 & 9 & 3 & 6 \min_{A_i} \left( \max_{S_j} V_{ij} \right) = 12 \text{ corresponding to } A_1 \Rightarrow A_1 \\
A_2 & 3 & 4 & 2 & 0 \\
A_3 & -6 & -2 & 1 & 0 \\
\end{array}
\]

Proverbs 21:5 The plans of the diligent lead to profit as surely as haste leads to poverty
3. Redo to question 2, treating $V_{ij}$ as costs obtain the optimal course of action.

4. A group of friends are planning a recreational outing and have constructed the following payoff table to help them decide which activity to engage in. Assume that the payoffs represent their level of enjoyment for each activity under the various weather conditions.

<table>
<thead>
<tr>
<th>Alternatives</th>
<th>Cold</th>
<th>Warm</th>
<th>Rainy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bike: $A_1$</td>
<td>10</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Hike: $A_2$</td>
<td>14</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>Fish: $A_3$</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>

a) If the group is (i) optimistic (ii) conservative, what decision should they make?
b) If the group chooses to minimize their maximum regret, what activity will they choose?

5. An investor is considering 4 different opportunities, A, B, C, or D. The payoff for each opportunity will depend on the economic conditions, represented in the payoff table below.

<table>
<thead>
<tr>
<th>Economic Condition</th>
<th>Poor $S_1$</th>
<th>Average $S_2$</th>
<th>Good $S_3$</th>
<th>Excellent $S_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>50</td>
<td>75</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>B</td>
<td>80</td>
<td>15</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>C</td>
<td>-100</td>
<td>300</td>
<td>-50</td>
<td>10</td>
</tr>
<tr>
<td>D</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

What decision would be made under each of the five criterions above?

6. A milk vender must decide the no of casto stock using each of the five criterions and the following profits pay-off table obtain the best course of action.

<table>
<thead>
<tr>
<th>Stock</th>
<th>Demand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk cases</td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>39</td>
</tr>
<tr>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>18</td>
<td>6</td>
</tr>
</tbody>
</table>

7. A business owner is trying to decide whether to buy, rent, or lease office space and has constructed the following payoff table based on whether business is brisk or slow. Using each of the five criterions, find the optimal strategy for the business owner.

<table>
<thead>
<tr>
<th>state of Business</th>
<th>Alternatives</th>
<th>Brisk</th>
<th>Slow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buy</td>
<td>90</td>
<td>-10</td>
<td></td>
</tr>
<tr>
<td>Rent</td>
<td>70</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>Lease</td>
<td>60</td>
<td>55</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of Workers</th>
<th>Compliance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>150</td>
</tr>
</tbody>
</table>

8. The local operations manager for the IRS must decide whether to hire 1, 2, or 3 temporary workers. He estimates that net revenues will vary with how well taxpayers comply with the new tax code. How many new workers will he hire using each of the five criterions?

**Decision-Making under Risk Environment**

The problem of decision making under risk is to choose an action (or decision) among many different available actions which gives (possibly) maximum expected profit or maximum expected revenue or minimum expected losses or minimum expected costs as the case may be, under uncertain situations.
Under this environment, the decision maker has some knowledge regarding the state of nature and he/she is able to assign subject probability estimates for each state of nature. Various criterions are involved under this environment.

a) **Maximum likelihood criterion**
   Its also called the most probable state of nature criterion. Here, the decision-maker identifies the state of nature with the highest prior probability and then selects the action corresponding to the maximum pay off in this particular state of nature. The disadvantage of this criterion is that it ignores some useful information as it only concentrates on the most probable state of nature.

b) **Expected Pay-off (Expected Monetary Value) Criterion.**
   --simply called EMV Criterion
   Choose the action that yields maximum expected value. For each action compute
   \[
   \text{EMV}(A_i) = \sum_{j=1}^{m} p_j V_{ij}
   \]
   Criterion: select the alternative \(A_i\) that yields
   \[
   \max_{A_i}\{\text{EMV}(A_i)\} \text{ if } V_{ij} \text{ are profits} \quad \text{or} \quad \min_{A_i}\{\text{EMV}(A_i)\} \text{ if } V_{ij} \text{ are costs}
   \]

**Example 7** The Robotic Micro computer company manufacturers micro-computers. The company is contemplating the expansion of its manufacturing facilities. The following table shows this pay-off in million dollars for the company.

<table>
<thead>
<tr>
<th>Alternatives</th>
<th>Demand</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
</tr>
<tr>
<td>Expand current facility</td>
<td>10</td>
</tr>
<tr>
<td>Build new facility</td>
<td>25</td>
</tr>
<tr>
<td>License another manufacturer</td>
<td>15</td>
</tr>
<tr>
<td>Prior Probability</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Obtain the best course of action using the maximum likelihood and EMV criterion.

**Solution**

a) Maximum likelihood Approach.
   Most probable state of nature is either high or moderate demand. The best course of action is building a new facility.

b) EMV Criterion
   \[
   \text{EMV (A_1) } = 10(0.4) + 5(0.4)-1(0.2)= 5.8
   \]
   \[
   \text{EMV(A_2) =25(0.4)+15(0.4)-5(0.2)=15} \quad \text{EMV(A_3)=15(0.4)+8(0.4)-0.5(0.2)=9.1}
   \]
   Best EMV (-one with maximum-) is EMV(A_2). Thus choose A_2 i.e build new facility

**Question** The manufacturer of summering pools must determine in advance of the summer season the number of pools that must be produced for the coming year. Each summering pool costs $ 500 and sells at $1000. Any summering pool that is unsold at the end of the season may be disposed off at $300 each. In
order to decide the number of pools to produce, the manufacturer must obtain information pertaining to the demand for the summering pool. Based on past experience, the following levels of demand are postulated (assumed).

(i) Low demand = 1 000 pods demanded
(ii) Moderate demand = 5 000 pods demanded
(iii) High demand = 10 000 pods demanded

If the probabilities of low, moderate & high demand are 0.2, 0.5 and 0.3 respectively. Prepare the gains pay-off table hence determine the best course of action using both maximum likelihood and the ENV criterions.

**Example 8** A Company is considering purchasing a new machine that is expected to save a considerable amount of operational cost associated with the current machine. The new machine costs $10 000 and is expected to save half a dollar per hour over the current. There is a considerable amount of uncertainty concerning the expected number of hours that the company will actually use the new machine. The management of the company has expected the uncertainty in the basis of the following probability distribution. Should the company buy the machine?

<table>
<thead>
<tr>
<th>No of hours</th>
<th>10,000</th>
<th>20,000</th>
<th>30,000</th>
<th>40,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.1</td>
<td>0.3</td>
<td>0.5</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**Solution:**

Let $X$ be the no. of hrs of use of the new machine serves and Let $S$ be the savings due to the new machine.

Thus, $S = 0.5x - 10 000 \Rightarrow E(S) = 0.5 E(X) - 10 000$

But $E(X) = 0.1(10 000) + 0.3(20 000) + 0.5(30 000) + 0.1(40 000) = 26 000$

$\Rightarrow E(S) = 0.5(26 000) - 10 000 = $ 3 000

Conclusion: since $E(S)$ is positive, the company should buy the machine.

**Question:** At a distance of 100 yards kibue hit target at an average of 60% of the time. In a local rifle contest, each contestant will shoot on the target 3 times and receive the following rewards depending on the number targets hit.

<table>
<thead>
<tr>
<th>No of targets hit</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reward</td>
<td>$100</td>
<td>$50</td>
<td>$10</td>
<td>-$50</td>
</tr>
</tbody>
</table>

Should Kibue enter the contest?

c) **Expected Opportunity loss (EOL) Criterion**

Payoffs and opportunity losses can be viewed as two sides of the same coin, depending on whether you wish to view the problem in terms of maximizing expected monetary value or minimizing expected opportunity loss..

Here we compute the expected regret for each course of action as follows;

$$EOL(A_i) = \sum_{j=1}^{m} p_j R_{ij}$$

Where $R_{ij}$ is the regret or loss incurred by choosing action $A_i$ and the state of nature $S_j$ occurs or prevails over the other states of nature. The decision criterion is to select the course of action with the smallest $EOL$. 

---

Proverbs 21:5 The plans of the diligent lead to profit as surely as haste leads to poverty
Selecting the course of action with the smallest $EOL$ is equivalent to selecting the course of action with the largest $EMV$

ie Choose action $A_i$ with

$$\min_{A_i} \{EOL(A_i)\}$$

**Example 9**  Consider the following pay-off table and obtain the best course of action using the Expected Opportunity loss ($EOL$) criterion

<table>
<thead>
<tr>
<th>Actions</th>
<th>$S_1$</th>
<th>$S_2$</th>
<th>$S_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>60</td>
<td>100</td>
<td>120</td>
</tr>
<tr>
<td>B</td>
<td>90</td>
<td>60</td>
<td>110</td>
</tr>
<tr>
<td>C</td>
<td>80</td>
<td>110</td>
<td>90</td>
</tr>
<tr>
<td><strong>Prob</strong></td>
<td><strong>0.5</strong></td>
<td><strong>0.3</strong></td>
<td><strong>0.2</strong></td>
</tr>
</tbody>
</table>

**Regret Table**

<table>
<thead>
<tr>
<th>Actions</th>
<th>$S_1$</th>
<th>$S_2$</th>
<th>$S_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>30</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>C</td>
<td>10</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td><strong>Prob</strong></td>
<td><strong>0.5</strong></td>
<td><strong>0.3</strong></td>
<td><strong>0.2</strong></td>
</tr>
</tbody>
</table>

**Solution**

Regret matrix:::- Since we have profit

$EOL$ (A) $=30(0.5) + 10(0.3)+0(0.2)= 18$

$EOL$ (B) $=0(0.5) + 50(0.3)+10(0.2)= 17$

$EOL$ (C) $=10(0.5) + 0(0.3)+30(0.2)= 11$  ---This is the best EOL thus C is the optimal course of action.

**Question**

Consider the following pay-off table and obtain the best course of action using $EOI$ criterion

<table>
<thead>
<tr>
<th>Alternatives</th>
<th>Demand</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Moderate</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>80</td>
<td>110</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>120</td>
<td>70</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>80</td>
<td>120</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td><strong>Probability</strong></td>
<td><strong>0.2</strong></td>
<td><strong>0.5</strong></td>
<td><strong>0.3</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Actions</th>
<th>$S_1$</th>
<th>$S_2$</th>
<th>$S_3$</th>
<th>$S_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>70</td>
<td>100</td>
<td>100</td>
<td>130</td>
</tr>
<tr>
<td>B</td>
<td>220</td>
<td>50</td>
<td>90</td>
<td>120</td>
</tr>
<tr>
<td>C</td>
<td>90</td>
<td>80</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td><strong>Prob</strong></td>
<td><strong>0.32</strong></td>
<td><strong>0.25</strong></td>
<td><strong>0.28</strong></td>
<td><strong>0.15</strong></td>
</tr>
</tbody>
</table>

**Expected Value of Perfect Information (EVPI)**

In most cases the decision maker does not know the state of Nature with certainty. EVPI also called the cost of uncertainty is the expected opportunity loss associated with the optimal decision. It is actually the maximum amount that the decision maker is willing to pay in order to know precisely the states of nature.
that will be in effect. It also represents the difference between expected profit under certainty (EVUC) and the EMV of the best alternative.

**Example 10** Consider the pay-off table in example 7 and obtain the Expected Value of Perfect Information (EVPI)

**Solution**

Best EMV (-one with maximum-) is EMV(A₂) = 15. 

EVUC = 25(0.4) + 15(0.4) - 0.5(0.2) = 15.9 

Therefore EVPi = EVUC - Best EMV = $0.9

d) **Return-to-Risk Ratio**

Unfortunately, neither the expected monetary value nor the expected opportunity loss criterion takes into account the variability of the payoffs for the alternative courses of action under different states. Example 11 illustrates the return-to-risk ratio (RTRR) criterion.

**Example 11** The manager of Reliable Fund has to decide between two stocks to purchase for a short-term investment of one year. An economist at the company has predicted returns for the two stocks under four economic conditions: recession, stability, moderate growth, and boom. The table below presents the predicted one-year return of a $1,000 investment in each stock under each economic condition.

<table>
<thead>
<tr>
<th>Economic Condition</th>
<th>A</th>
<th>B</th>
<th>EMV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recession</td>
<td>30</td>
<td>-50</td>
<td></td>
</tr>
<tr>
<td>Stability</td>
<td>70</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Moderate Growth</td>
<td>100</td>
<td>250</td>
<td></td>
</tr>
<tr>
<td>Boom</td>
<td>150</td>
<td>400</td>
<td>162</td>
</tr>
</tbody>
</table>

From the above table, you see that the return for stock A varies from $30 in a recession to $150 in an economic boom, whereas the return for stock B (the one chosen according to the expected monetary value and expected opportunity loss criteria) varies from a loss of $50 in a recession to a profit of $400 in an economic boom.

To take into account the variability of the states of nature (in this case, the different economic conditions), we can compute the variance and standard deviation of each stock. Using the information presented in above table for stock A, $EMV(A) = \mu_A = 91$ and the variance is

$$\sigma_A^2 = \sum_{j=1}^{m} (V_{ij} - \mu_A)^2 \times P(S_j) = 0.1(-61)^2 + 0.4(-21)^2 + 0.3(9)^2 + 0.2(59)^2 = 1,269 \Rightarrow \sigma_A = 35.62.$$ 

For stock A, $EMV(B) = \mu_B = 162$ and the variance is

$$\sigma_B^2 = \sum_{j=1}^{m} (V_{ij} - \mu_B)^2 \times P(S_j) = 0.1(-212)^2 + 0.4(-132)^2 + 0.3(88)^2 + 0.2(238)^2 = 25,116 \Rightarrow \sigma_B = 158.48.$$ 

**Remark:** Notice that $EMV(A) = \mu_A = \sum_{j=1}^{m} V_{ij} \times P(S_j)$ and the variance is

$$\sigma^2 = E(V_{ij} - \mu_A)^2 = \sum_{j=1}^{m} (V_{ij} - \mu_A)^2 \times P(S_j) \quad \text{or} \quad \sigma^2 = E(V_{ij} - \mu_A) = \sum_{j=1}^{m} V_{ij}^2 \times P(S_j) - \mu^2$$

Because you are comparing two stocks with different means, you should evaluate the relative risk associated with each stock. Once you compute the standard deviation of the return from each stock, you
compute the coefficient of variation discussed in STA 2100. Substituting $\sigma$ for $S$ and EMV for $\bar{x}$ in the formula for the coefficient of variation we get

$$CV(A_i) = \left( \frac{\sigma_{A_i}}{EMV(A_i)} \right) \times 100\%$$

$$CV(A) = \left( \frac{35.62}{91} \right) \times 100\% \approx 39.1\% \quad \text{and} \quad CV(B) = \left( \frac{158.48}{162} \right) \times 100\% \approx 97.8\%$$

Thus, there is much more variation in the return for stock $B$ than for stock $A$.

When there are large differences in the amount of variability in the different events, a criterion other than $EMV$ or $EOL$ is needed to express the relationship between the return (as expressed by the $EMV$) and the risk (as expressed by the standard deviation). The following equation defines the **return-to-risk ratio ($RTRR$)** as the expected monetary value of action $A_i$ divided by the standard deviation of action $A_i$.

$$RTRR(A_i) = \frac{EMV(A_i)}{\sigma_{A_i}}$$

**Criterion**: Select the course of action with the largest $RTRR$.

For each of the two stocks discussed previously, you compute the return-to-risk ratio as follows.

For stock $A$, the return-to-risk ratio is equal to

$$RTRR(A) = \frac{91}{35.62} \approx 2.55$$

For stock $B$, the return-to-risk ratio is given by

$$RTRR(B) = \frac{162}{158.48} \approx 1.02$$

Thus, relative to the risk as expressed by the standard deviation, the expected return is much higher for stock $A$ than for stock $B$. Stock $A$ has a smaller expected monetary value than stock $B$ but also has a much smaller risk than stock $B$. The return-to-risk ratio shows $A$ to be preferable to $B$.

**Exercise 3**

1. For the following payoff table, the probability of event $S_1$ is 0.8, the probability of event $S_2$ is 0.1, and the probability of event $S_3$ is 0.1:

<table>
<thead>
<tr>
<th>Actions</th>
<th>$S_1$</th>
<th>$S_2$</th>
<th>$S_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>50</td>
<td>300</td>
<td>500</td>
</tr>
<tr>
<td>B</td>
<td>10</td>
<td>100</td>
<td>200</td>
</tr>
</tbody>
</table>

a) Determine the optimal action using the $EMV$, the $(EOL)$ and the return to risk ratio criterions
b) Explain the meaning of EVPI in this problem.

2. Consider the following costs payoff table. Using the $EMV$, the $(EOL)$ and the return to risk ratio criterion find the optimal action.

<table>
<thead>
<tr>
<th>Actions</th>
<th>$S_1$</th>
<th>$S_2$</th>
<th>$S_3$</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>60</td>
<td>110</td>
<td>120</td>
<td>0.25</td>
</tr>
<tr>
<td>B</td>
<td>110</td>
<td>60</td>
<td>110</td>
<td>0.45</td>
</tr>
<tr>
<td>C</td>
<td>80</td>
<td>90</td>
<td>100</td>
<td>0.30</td>
</tr>
<tr>
<td>D</td>
<td>80</td>
<td>100</td>
<td>110</td>
<td>0.30</td>
</tr>
</tbody>
</table>

3. Refer to question 4 in exercise 2. If the probabilities of cold weather ($S_1$), warm weather ($S_2$), and rainy weather ($S_3$) are 0.2, 0.4, and 0.4, respectively, then what decision should be made using the $EMV$, the $EOL$ and the return to risk ratio criterion? What is the EVPI for this situation?

4. Refer to question 5 in exercise 2. If the probabilities of each economic condition are 0.5, 0.1, 0.35, and 0.05 respectively, what investment would be made using the $EMV$, the $(EOL)$ and the return to risk ratio criterion? What is the EVPI for this situation?
5. Refer to question 7 in exercise 2. If the probability of brisk business is .40 and for slow business is .60, find the expected value of perfect information.

6. If the local operations manager in question 8 exercises 2 thinks the chances of low, medium, and high compliance are 20%, 30%, and 50% respectively, what are the expected net revenues for the number of workers he will decide to hire?

7. An entrepreneur must decide how many tones of plastic shipping pallets to produce each week for sale. His manufacturing manager has determine with an accounts assistance that the total cost per tone to produce pallets is $25. There is also an additional cost of $2 per tone of pallets sold to shippers. The pallets are sold for $50 per tone> pallets not sold at the end of the week are disposed off to a local farmer who will pay $5 per tonne. It’s assumed that the weekly demand for of plastic shipping pallets will range from 0 to 5 tonnes. Construct the manufacturers profit pay-off table. If after consultation with an experienced market analyst entrepreneur assesses the accompanying probability distribution to weekly demand to shippers of plastic pallets as;

<table>
<thead>
<tr>
<th>Demand</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.05</td>
<td>0.1</td>
<td>0.25</td>
<td>0.4</td>
<td>0.15</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Determine the number of tones to produce if the entrepreneur wishes to maximize expected profit. What is the entrepreneur Expected Value of Perfect Information (EVPI)? In simple language explain the meaning of EVPI to the entrepreneur.

8. Inflation tends to boast interest rates of loans of all types. As a partial hedge against inflation, some lending institutions offer variable rates loans to borrowers with high credit ratings. The borrower can then accept either a fixed rate loan with certain interest obligations or a variable rate loan that may result in either interest savings or additional interest costs to the borrower depending on the future state if the economy. Suppose an individual wishes to borrow $ 10,000 for 1 year when the prevailing interest rate is 9%. After consulting an economist, the individual attributes the probability distribution for the prevailing lending rates over the next 1 year to be;

<table>
<thead>
<tr>
<th>Interest Rate</th>
<th>8%</th>
<th>8.50%</th>
<th>8%</th>
<th>9.50%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.1</td>
<td>0.25</td>
<td>0.5</td>
<td>0.1</td>
<td>0.05</td>
</tr>
</tbody>
</table>

The borrower must decide whether to accept the prevailing interest rate of 9% or undertake a variable rate, a) Construct an opportunity loss table for the borrower  

b) If the borrower wishes to minimize the expected opportunity loss, which type of loan should he undertake?  
c) How much would the borrower be willing to pay to know exactly the interest rate that would be in effect at the end of the year?

9. For a potential investment of $1,000, if a stock has an EMV of $100 and a standard deviation of $25, what is the return-to-risk ratio?

10. A stock has the following predicted returns under the following economic conditions:

<table>
<thead>
<tr>
<th>Economic Condition</th>
<th>Probability</th>
<th>Return ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recession</td>
<td>0.3</td>
<td>50</td>
</tr>
<tr>
<td>Stable Economy</td>
<td>0.3</td>
<td>100</td>
</tr>
<tr>
<td>Moderate Growth</td>
<td>0.3</td>
<td>120</td>
</tr>
<tr>
<td>Boom</td>
<td>0.1</td>
<td>200</td>
</tr>
</tbody>
</table>

Compute the  
a) expected monetary value.  
b) standard deviation.  
c) coefficient of variation.  
d) return-to-risk ratio
11. The following are the returns ($) for two stocks: Which stock would you choose and why?

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMV</td>
<td>90</td>
<td>60</td>
</tr>
<tr>
<td>SD</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

12. A vendor at a local baseball stadium must determine whether to sell ice cream or soft drinks at today’s game. She believes that the profit made will depend on the weather. The payoff table (in $) is as shown.

<table>
<thead>
<tr>
<th>Weather</th>
<th>Cool</th>
<th>Warm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sell Soft Drinks</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>Sell Ice Cream</td>
<td>30</td>
<td>90</td>
</tr>
</tbody>
</table>

a) Based on her past experience at this time of year, the vendor estimates the probability of warm weather as 0.60.

b) Determine the optimal action based the EMV, EOL and the return to risk ratio criterions.

c) Explain the meaning of the expected value of perfect information (EVPI) in this problem.

13. The Islander Fishing Company purchases clams for $1.50 per pound from fishermen and sells them to various restaurants for $2.50 per pound. Any clams not sold to the restaurants by the end of the week can be sold to a local soup company for $0.50 per pound. The company can purchase 500, 1,000, or 2,000 pounds. The probabilities of various levels of demand are as follows:

<table>
<thead>
<tr>
<th>Demand (Pounds)</th>
<th>500</th>
<th>1,000</th>
<th>2,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.2</td>
<td>0.4</td>
<td>0.4</td>
</tr>
</tbody>
</table>

a) For each possible purchase level (500, 1,000, or 2,000 pounds), compute the profit (or loss) for each level of demand.

b) Determine the optimal action based the EMV, EOL and the return to risk ratio criterions.

c) Suppose that clams can be sold to restaurants for $3 per pound. Repeat (a) with this selling price for clams and compare the results with those in a

d) What would be the effect on the results in (a) if the probability of the demand for 500, 1,000, and 2,000 clams were 0.4, 0.4, and 0.2, respectively?

e) What would be the effect on the results in (a) if the probability of the demand for 500, 1,000, and 2,000 clams were 0.4, 0.4, and 0.2, respectively?

14. An investor has a certain amount of money available to invest now. Three alternative investments are available. The estimated profits ($) of each investment under each economic condition are indicated in the following payoff table. Based on his own past experience, the investor assigns the following probabilities to each economic condition:

<table>
<thead>
<tr>
<th>State of the Economy</th>
<th>Actions</th>
<th>Economy declines</th>
<th>No Change</th>
<th>Economy expands</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>$1,500</td>
<td>$1,000</td>
<td>$2,000</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>-$2,000</td>
<td>$2,000</td>
<td>$5,000</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>-$7,000</td>
<td>-$1,000</td>
<td>$20,000</td>
</tr>
<tr>
<td>Prior prob</td>
<td>0.3</td>
<td>0.5</td>
<td>0.2</td>
<td></td>
</tr>
</tbody>
</table>

a) Based on her past experience at this time of year, the vendor estimates the probability of warm weather as 0.60.

b) Determine the optimal action based the EMV, EOL and the return to risk ratio criterions.

c) Explain the meaning of the expected value of perfect information (EVPI) in this problem.
15. Consider the payoff table you developed on building a small factory or a large factory for manufacturing designer jeans. Given the results of that problem, suppose that the probabilities of the demand are as follows:

<table>
<thead>
<tr>
<th>Demand</th>
<th>10,000</th>
<th>20,000</th>
<th>50,000</th>
<th>100,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.1</td>
<td>0.4</td>
<td>0.2</td>
<td>0.3</td>
</tr>
</tbody>
</table>

a) Compute the expected monetary value (EMV) and the expected opportunity loss (EOL) for building a small factory and building a large factory.

b) Explain the meaning of the expected value of perfect information (EVPI) in this problem.

c) Based on the results of a) would you choose to build a small factory or a large factory? Why?

d) Compute the coefficient of variation for building a small factory and a large factory.

e) Compute the return-to-risk ratio (RTRR) for building a small factory and building a large factory.

f) Based on (d) and (e), would you choose to build a small factory or a large factory? Why?

g) Compare the results of (f) and (c) and explain any differences.

h) Suppose that the probabilities of demand are 0.4, 0.2, 0.2, and 0.2, respectively. Repeat (a) through (g) with these probabilities and compare the results with those in (a)–(g).

16. Consider the payoff table you developed to assist an author in choosing between signing with company A or with company B. Given the results computed in that problem, suppose that the probabilities of the levels of demand for the novel are as follows:

<table>
<thead>
<tr>
<th>Demand</th>
<th>1,000</th>
<th>2,000</th>
<th>5,000</th>
<th>10,000</th>
<th>50,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.45</td>
<td>0.20</td>
<td>0.15</td>
<td>0.10</td>
<td>0.10</td>
</tr>
</tbody>
</table>

a) Compute the expected monetary value (EMV) and the expected opportunity loss (EOL) for signing with company A and with company B.

b) Explain the meaning of the expected value of perfect information (EVPI) in this problem.

c) Based on the results of a) would you choose to sign with company A or company B? Why?

d) Compute the coefficient of variation for signing with company A and with company B.

e) Compute the return-to-risk ratio (RTRR) for signing with company A and with company B.

f) Based on (d) & (e), would you choose to sign with company A or with company B? Why?

g) Compare the results of (f) and (c) and explain any differences.

h) Suppose that the probabilities of demand are 0.3, 0.2, 0.2, 0.1, and 0.2 respectively. Repeat (a)–(g) with these probabilities and compare the results with those in (a)–(g).

17. Consider the payoff table you developed to determine whether to purchase 100, 200, 500, or 1,000 Christmas trees. Given the results of that problem, suppose that the probabilities of the demand for the different number of trees are as follows:

<table>
<thead>
<tr>
<th>Demand</th>
<th>100</th>
<th>200</th>
<th>500</th>
<th>1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.2</td>
<td>0.5</td>
<td>0.2</td>
<td>0.1</td>
</tr>
</tbody>
</table>

a) Compute the expected monetary value (EMV) and the expected opportunity loss (EOL) for purchasing 100, 200, 500, and 1,000 trees

b) Explain the meaning of the expected value of perfect information (EVPI) in this problem.

c) Based on the results of part a) would you choose to purchase 100, 200, 500, and 1,000 trees? Why?

d) Compute the coefficient of variation for purchasing 100, 200, 500, and 1,000 trees

e) Compute the return-to-risk ratio (RTRR) for purchasing 100, 200, 500, and 1,000 trees

f) Based on (d) and (e), would you choose to purchase 100, 200, 500, and 1,000 trees? Why?

g) Compare the results of (f) and (c) and explain any differences.

h) Suppose that the probabilities of demand are 0.4, 0.2, 0.2, and 0.2 respectively. Repeat (a) through (g) with these probabilities and compare the results with those in (a)–(g).
18. An oil company has some land that is reported to possibly contain oil. The company classifies such land into four categories by the total number of barrels that are expected to be obtained from the well, i.e. a 500,000 – barrel well, 200,000 – barrel well, 50,000 – barrel well, and a dry well. The company is faced with deciding whether to drill for oil, to unconditionally lease the land or to conditionally lease the land at a rate depending upon oil strike. The cost of drilling the well is $100,000; if it is a producing well and the cost of drilling is $75,000 if it is a dry well. For producing well, the profit per barrel of oil is $1.50, (after deduction of processing and all other costs except drilling costs). Under the unconditional lease agreement, the company receives $45,000 for the land whereas for the conditional lease agreement the company receives 50 cents for each barrel of oil extracted if it is a 500,000 or 200,000 barrel oil strike and nothing if otherwise. The probability for striking a 500,000 – barrel well is 0.1, probability for striking a 200,000 – barrel well is 0.15, probability for striking a 50,000 – barrel well is 0.25, and probability for a dry well is 0.5.

a) Make a profit payoff table for the oil company
b) Find the optimal act according to Expected opportunity loss criteria
c) Find the EVPI

Sensitivity Analysis
Sensitivity analysis is an essential element of decision analysis. The principle of sensitivity analysis is also directly applied to areas such as meta-analysis and cost effectiveness analysis most readily but is not confined to such approaches.

Sensitivity analysis evaluates the stability of the conclusions of an analysis to assumptions made in the analysis. When a conclusion is shown to be invariant to the assumptions, confidence in the validity of the conclusions of the analysis is enhanced. Such analysis also helps identify the most critical assumptions of the analysis (Petitti, 2000).

When the assumed value of a variable affects the conclusion of the analysis, the analysis is said to be “sensitive” to that variable. When the conclusion does not change, when the sensitivity analysis includes the values of the variables that are within a reasonable range, the analysis is said to be “insensitive” to that variable.

An area of emphasis for sensitivity analysis is the sensitivity with respect to the prior, when applying the EMV criterion. The prior probabilities in this model are most questionable and will be the focus of the sensitivity analysis to be conducted

Example 12 (The oil company on pg 3)
Basic probability theory suggests that the sum of the two prior probabilities must equal 1, so increasing one of these probabilities automatically decreases the other one by the same amount, and vice versa. The oil company’s management team feels that the true “chances” of having oil on the tract of land are more likely to lie between the range from 0.15 to 0.35. Thus, the corresponding prior probability of the land being dry would range from 0.85 to 0.65, respectively.

Conducting a sensitivity analysis in this situation requires the application of the EMV rule twice – once when the prior probability of the oil is at the lower end (0.15) of this range and next when it is at the upper end (0.35). When the prior probability is conjectured to be 0.15, we find

$$EMV(A_1) = 0.15(700) + 0.85(-100) = 20$$
$$EMV(A_2) = 0.15(90) + 0.85(90) = 90$$

and when the prior probability is thought to be 0.35, the Bayes’ decision rule finds that
EMV(A₁) = 0.35(700) + 0.65(-100) = 180
EMV(A₂) = 0.35(90) + 0.65(90) = 90

Thus, the decision is very sensitive to the prior probability of oil as the expected payoff \((\text{to drill})\) shifts from 20 to 100 to 180 when the prior probabilities are 0.15, 0.25, and 0.35, respectively. Thus, if the prior probability of oil is closer to 0.15, the optimal action would be to sell the land rather than to drill for oil as suggested by the other prior probabilities. This suggests that it would be more plausible to determine the true value of the probability of oil.

Let \(p\) = prior probability of oil. Then the expected payoff from drilling for any \(p\) is

\[
EMV(A₁) = 700p - 100(1 - p) = 800p - 100
\]

A graphical display of how the expected payoff for each alternative changes when the prior probability of oil changes for the oil company’s problem of whether to drill or sell is illustrated in Figure 1. The point in this figure where the two lines intersect is the threshold point where the decision shifts from one alternative (selling the land) to the other (drill for oil) as the prior probability increases. Algebraically, this is simple to determine as we set

\[
EMV(A₁) = EMV(A₂) \Rightarrow 800p - 100 = 90 \Rightarrow p = \frac{190}{800} = 0.2375
\]

Thus, the conclusion should be to sell the land if \(p < 0.2375\) and should drill for oil if \(p > 0.2375\). Because, the decision for the oil company decides heavily on the true probability of oil, serious consideration should be given to conducting a seismic survey to estimate the probability more accurately. This is considered in the next subsection.

Figure 1

### Decision Making with Experimentation/Sample Information

**Introduction**

In the section of decision making under risk, we assigned the probabilities of the various events using the past experience and/or the subjective judgment of the decision maker. In that section no current information
describing the change or outcome of the states of nature is assumed to be available. However the decision maker has the option of collecting sample information prior to making the final decision. The additional information about the states of nature is very important especially in dynamic environment where technology is changing rapidly resulting in constant change of the states of nature. The sample information reduces the uncertainty about the states of nature and is available to the decision maker either for purchase or at the cost of experimentation. Eg a retailer whose business depends on weather may consult meteorologist before making a decision or an investor may hire a market analysis before investing. This section introduces decision making when sample information is available to estimate probabilities

The **Bayesian** view of probability is related to degree of belief. It is a measure of the plausibility of an event given incomplete knowledge. *Prior* information can be based on the results of previous experiments, or expert opinion, and can be expressed as probabilities. If it is desirable to improve on this state of knowledge, an experiment can be conducted. Bayes' Theorem is the mechanism used to update the state of knowledge with the results of the experiment to provide a *posterior* distribution

**Definitions**

Knowledge of sample or survey information can be used to revise the probability estimates for the states of nature. Prior to obtaining this information, the probability estimates for the states of nature are called *prior probabilities*. With knowledge of *conditional probabilities* for the outcomes or indicators of the sample or survey information, these prior probabilities can be revised by employing Bayes' Theorem. The outcomes of this analysis are called *posterior probabilities*.

**Do we have to Test?**

Often, companies have the option to perform market tests/surveys, usually at a price, to get additional information prior to making decisions. However, some interesting questions need to be answered before this decision is made:

– How will the tests/surveys results be combined with prior information?
– How much should you be willing to pay to test/sample?

If the sample information is available, Bayes’ Theorem can be used to combine the prior and current information as follows. Suppose before choosing an action, an outcome from an experiment is observed $Z=z$. Therefore for each $i=1,2,...,m$, the corresponding posterior probability is

$$p(s_i / z_i) = \frac{p(z_i / s_j) p(s_j)}{p(z_i)} = \frac{p(z_i / s_j) p(s_j)}{\sum_{j=1}^{n} p(z_i / s_j) p(s_j)}$$

Thus, the question being addressed is: Given what is $p(s_j)$ and $p(z_i / s_j)$ for $i=1,2,...,n$ ? What is $p(s_j / z_i)$? The answer is solved by following standard formulas of probability theory which states that the conditional probability

$$p(s_j / z_i) = \frac{p(z_i, s_j)}{p(z_i)} = \frac{p(z_i / s_j) p(s_j)}{\sum_{j=1}^{n} p(z_i / s_j) p(s_j)}$$

The pair of propositions “if $z_i$ then $a_j$” is called the Bayes Strategy. Such a strategy constitutes a contingency plan since the decision maker knows the proper reaction to each survey result before sampling. Example 12 illustrates decision making with sample information.

**Proverbs 21:5** The plans of the diligent lead to profit as surely as haste leads to poverty
Example 13  For the oil company problem, an available option before making a decision is to conduct a detailed seismic survey of the land to obtain a better estimate of the probability of oil. The cost is $30,000. A seismic survey obtains seismic soundings that indicate whether the geological structure is favorable to the presence of oil. Dividing the possible findings of the survey into the following two categories:

- USS: unfavorable seismic soundings; oil is unlikely, and
- FSS: favorable seismic soundings; oil is likely to be found.

Based on past experience, if there is oil, then the probability of unfavorable seismic soundings (USS) is $P(\text{USS} \mid \text{State} = \text{Oil}) = 0.4$, so $P(\text{FSS} \mid \text{State} = \text{Oil}) = 1 - 0.4 = 0.6$.

Similarly, if there is no oil (state of nature = Dry), then the probability of unfavorable seismic soundings is estimated to be $P(\text{USS} \mid \text{State} = \text{Dry}) = 0.8$, so $P(\text{FSS} \mid \text{State} = \text{Dry}) = 1 - 0.8 = 0.2$.

This data is used to find the posterior probabilities of the respective states of nature given the seismic readings.

If the finding of the seismic survey is unfavorable, then the posterior probabilities are

$$p((\text{State Oil} \mid \text{Finding USS})) = p(s_1 / z_1) = \frac{0.4(0.25)}{0.4(0.25) + 0.8(0.75)} = \frac{1}{7}$$

$$\Rightarrow p(\text{State Dry} \mid \text{Finding USS}) = p(s_2 / z_1) = 1 - \frac{1}{7} = \frac{6}{7}$$

Note that USS is $z_1$ and FSS is $z_2$. Similarly, if the seismic survey is favorable, then

$$p((\text{State Oil} \mid \text{Finding FSS})) = p(s_1 / z_2) = \frac{0.6(0.25)}{0.6(0.25) + 0.2(0.75)} = \frac{1}{2}$$

$$\Rightarrow p(\text{State Dry} \mid \text{Finding FSS}) = p(s_2 / z_2) = 1 - \frac{1}{2} = \frac{1}{2}$$

The expected payoffs if finding is unfavorable seismic soundings

$$E(A_1 \mid z_1) = \frac{1}{7}(700) + \frac{6}{7}(-100) - 30 \approx -15.7 \quad \Rightarrow \quad E(A_1 \mid z_1) = \frac{1}{7}(90) + \frac{6}{7}(90) - 30 = 60$$

And the expected payoffs if the finding is favorable seismic surroundings

$$E(A_1 \mid z_1) = \frac{1}{2}(700) + \frac{1}{2}(-100) - 30 = 270 \quad \Rightarrow \quad E(A_1 \mid z_1) = \frac{1}{2}(90) + \frac{1}{2}(90) - 30 = 60$$

<table>
<thead>
<tr>
<th>Finding from Seismic Survey</th>
<th>Optimal Action</th>
<th>Expected Payoff Excluding Cost of Survey</th>
<th>Expected Payoff Including Cost of Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>USS</td>
<td>Sell</td>
<td>90</td>
<td>600</td>
</tr>
<tr>
<td>FSS</td>
<td>Drill</td>
<td>300</td>
<td>270</td>
</tr>
</tbody>
</table>

Since the objective is to maximize the expected payoff, these results yield the optimal policy shown in Table 14-3. However, this analysis has not yet addressed the issue of whether the expense of conducting a seismic survey is truly valuable or whether one should choose the optimal solution without experimentation. This issue is addressed next.

Expected Value of Experimentation

Now we want to determine expected increase directly which is referred to as the expected value of experimentation. Calculating this quantity requires first computing the expected payoff with experimentation (excluding the cost of experimentation). The logic behind the economic theory of information is straightforward. Since $Z$ are random variables, with probabilities determined by $p(z_i)$, the
decision maker can attach a value to a Bayes strategy by calculating the expected value of the strategy. The value is usually referred to as the Gross Value of the information system or the expected value with sample information.

Obtaining this latter quantity requires the previously determined posterior probabilities, the optimal policy with experimentation, and the corresponding expected payoff (excluding the cost of experimentation) for each possible finding from the experiment. Then each of these expected payoff needs to be weighted by the probability of the corresponding finding, that is,

\[
\text{Expected payoff with experimentation} = \sum_{j=1}^{m} p(z_j) \times \text{EMV}(\text{Best} \mid z_j)
\]

Then the net value of experimentation is the expected value with sample information less the expected value without sample information and we can compare this latter value to the cost of a one-trial experiment to determine if the experiment is cost effective.

The Expected Value of Sample Information (EVSI), is the additional expected profit possible through knowledge of the sample or survey information.

**EVSI Calculation**

- Determine the optimal decision and its expected return for the possible outcomes of the sample using the posterior probabilities for the states of nature.
- Compute the expected value of these optimal returns.
- Subtract the EV of the optimal decision obtained without using the sample information from the amount determined in step (2).

\[
\text{EVSI} = \left( \text{expected value of best decision with sample information assuming no cost to collect it} \right) - \left( \text{expected value of best decision without sample information} \right)
\]

This calculation ignores the cost of the test. Once you compute the EVSI, you can compare it to the cost of the test to determine the desirability to test. The ratio of the EVSI to the EVPI times 100 will also give you a measure of “efficiency” of the proposed test, expressed as a percentage. As the EVPI provides an upper bound for the EVSI, efficiency is always a number between 0 and 1.

**Example 14** For the example above much of the work has been done for the right side of the equation. The values of \( p(z_j) \) for each of the two possible findings from the seismic survey – unfavorable or favorable – \( p(z_j) \) were calculated to be \( p(z_1) = 0.7 \) and \( p(z_2) = 0.3 \). For the optimal policy with experimentation, the corresponding expected payoff for each finding can be obtained from Table 14.3 as

\[
\text{Best EMV} \mid z_1 = 90 \text{ and Best EMV} \mid z_2 = 270
\]

Then the expected payoff with experimentation is determined to be

\[
0.7(90) + 0.3(300) = 153.
\]

Similar to before the expected value of experimentation (EVE or EVSI) can be calculated as

\[
\text{EVE or EVSI} = 153-100 = 53
\]

and identifies the potential value of the experimentation. Since this value exceeds 30, the cost of conducting a detailed seismic survey shows valuable.

**Question** Consider a very simple decision problem where there are two states and two acts, and where the payoffs are given in the following table A: Suppose we have prior probabilities as \( P(s_1) = 0.5 \) and \( P(s_2) = 0.5 \), and an information system with two signals where the matrix of reliability probabilities is shown in table...
Time is precious, but we do not know yet how precious it really is. We will only know when we are no longer able to take advantage of it.

Prverbs 21:5 The plans of the diligent lead to profit as surely as haste leads to poverty

<table>
<thead>
<tr>
<th>States of Nature</th>
<th>predicted States of Nature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual states</td>
<td>$Z_1$</td>
</tr>
<tr>
<td>$S_1$</td>
<td>0.8</td>
</tr>
<tr>
<td>$S_2$</td>
<td>0.4</td>
</tr>
</tbody>
</table>

In other words, the information system is 80% reliable in detecting that the first state will occur and 60% reliable in detecting that the second state will occur. The prior and reliability probabilities are used to generate posterior probabilities:

- a) Calculate the posterior probabilities
- b) Obtain the optimal act for each signal
- c) the expected value of perfect information
- d) the expected value of experimentation

**Question** A company makes stuff and they hope for an increase in demand. In the first year, there are two possible actions: A1 (new machinery) and A2 (overtime). In the first year, sales can either be good (g1) or bad (b1) and experts judge that P(g1) = 0.6, P(b1) = 0.4. In the second year, the company has options which depend upon the choices made in year one. If they chose new machinery in year one, then they may choose either more machinery or more overtime. If overtime is chosen in year one, then overtime continues in year two. The sales in year two will either be high (h2), medium (m2) or low (l2). If sales are good in year one, then the probability of high sales in year two is 0.5, of medium sales is 0.4 and of low sales is 0.1. If sales are bad in year one, then the probability of high sales in year two is 0.4, of medium sales is 0.4 and of low sales is 0.2. Thus, the probabilities for year two are conditional upon what happens in year one. More formally, we have: The payoffs for the problem are shown in the decision tree in Figure 4. The decision tree is solved by rollback. Notice that the probabilities are conditional upon the states to the left of them are the tree (that have thus occurred).

**Example 15** Thompson Lumber Company is trying to decide whether to expand its product line by manufacturing and marketing a new product which is “backyard storage sheds.” The courses of action that may be chosen include:

1. Large plant to manufacture storage sheds,
2. Small plant to manufacture storage sheds, or
3. Build no plant at all.

<table>
<thead>
<tr>
<th>Alternatives</th>
<th>Favourable market $S_1$</th>
<th>Unfavourable market $S_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>construct large plant</td>
<td>200,000</td>
<td>-180,000</td>
</tr>
<tr>
<td>construct small plant</td>
<td>100,000</td>
<td>-20,000</td>
</tr>
<tr>
<td>Do nothing</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Probability of favorable market is same as probability of unfavorable market. Each state of nature has a 0.50 probability. A Reliable Market Survey Predicting Actual States of Nature is as show below

| Results of Survey                  | $P(z_i|s_j)$ | States of Nature |
|------------------------------------|--------------|------------------|
| Positive (Predict favorable market) | Pr(Survey positive/ $S_1$)=0.7 | Favorable market $S_1$ |
| Negative (Predict unfavorable market of) | Pr(Survey negative/ $S_1$)=0.3 | Unfavorable market $S_2$ |
| Positive (Predict unfavorable market) | Pr(Survey positive/ $S_2$)=0.2 | Favorable market $S_1$ |
| Negative (Predict favorable market of) | Pr(Survey negative/ $S_2$)=0.8 | Unfavorable market $S_2$ |
Using the above information, obtain;

a) The Expected Value of perfect Information (EVPI),
b) the posterior probabilities of the states of nature

c) The conditional Expected Value of Sample Information (EVSI),

**Solution**

a) $EMV(\text{Large plant}) = 0.5(200) + 0.5(-180) = 10$  
   $EMV(\text{Do nothing}) = 0$  
   $EMV(\text{Small plant}) = 0.5(100) + 0.5(-20) = 40$  

Best action is build a small plant

Therefore, expected value under certainty  

$$EVUC = (200,000)(0.50) + (0)(0.50) = 100,000$$

$\Rightarrow$ If one had perfect information, an *average* payoff of $100,000 could be achieved in the long run.

However, the maximum EMV (EV<sub>BEST</sub>) or expected value without perfect information is $40,000.

Therefore,  

$$EVPI = 100,000 - 40,000 = 60,000.$$  

b) Probability Revisions (joint probability)

<table>
<thead>
<tr>
<th>Results of Survey</th>
<th>Favorable market S&lt;sub&gt;1&lt;/sub&gt;</th>
<th>Unfavorable market S&lt;sub&gt;2&lt;/sub&gt;</th>
<th>p(Survey Results)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favorable Market (FM)</td>
<td>0.7(0.5)=0.35</td>
<td>0.2(0.5)=0.1</td>
<td>0.45</td>
</tr>
<tr>
<td>Unfavorable Market (UM)</td>
<td>0.3(0.5)=0.15</td>
<td>0.8(0.5)=0.4</td>
<td>0.55</td>
</tr>
</tbody>
</table>

The posterior probabilities are got by dividing row 1 of the above table by 0.45 and row 2 by 0.55. The results are as tabulated below.

<table>
<thead>
<tr>
<th>Results of Survey</th>
<th>Favorable market S&lt;sub&gt;1&lt;/sub&gt;</th>
<th>Unfavorable market S&lt;sub&gt;2&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favorable Market (FM)</td>
<td>0.77778</td>
<td>0.22222</td>
</tr>
<tr>
<td>Unfavorable Market (UM)</td>
<td>0.27273</td>
<td>0.72727</td>
</tr>
</tbody>
</table>

c) If the survey predicts Favorable Market then

$EMV(\text{Large plant}) = 0.77778(200) + 0.22222(-180) = 115.56$  
$EMV(\text{Do nothing}) = 0$  
$EMV(\text{Small plant}) = 0.77778(100) + 0.22222(-20) = 73.33$  

Best action is a large plant

If the survey predicts Unfavorable Market then

$EMV(\text{Large plant}) = 0.27273(200) + 0.72727(-180) = -76.36$  
$EMV(\text{Do nothing}) = 0$  
$EMV(\text{Small plant}) = 0.27273(100) + 0.72727(-20) = 12.73$  

Best action is a small plant

Therefore the Expected Value with Sample Information For Thompson Lumber Company is;

$$EVWSI = 115.56(0.45) + 12.73(0.55) = 59.0035$$  
$$EVSI = EVWSI - EVPI = 59,003.5 - 60,000 = -0.9965$$

**Exercise 4**

1) Consider the payoff table below.

<table>
<thead>
<tr>
<th>Actions</th>
<th>S&lt;sub&gt;1&lt;/sub&gt;</th>
<th>S&lt;sub&gt;2&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>50</td>
<td>200</td>
</tr>
<tr>
<td>B</td>
<td>100</td>
<td>125</td>
</tr>
</tbody>
</table>

$P(S_1) = 0.5 = P(S_2)$ , $P(F/S_1) = 0.6$ and $P(F/S_2) = 0.4$ suppose that you are informed that event $F$ occurs.

a) Revise the probabilities $P(S_1)$ and $P(S_2)$ now that you know that event $F$ has occurred.
b) Based on these revised probabilities, compute the expected monetary value of action A and action B.

c) Based on these revised probabilities, compute the expected opportunity loss of action A and action B.

2) Consider the payoff table on question 1 of exercise 3, if in this problem 

\[ P(S_1) = 0.8, \ P(S_2) = 0.1, \ P(S_3) = 0.1 \]

\[ P(F/S_1) = 0.2, \ P(F/S_2) = 0.4 \text{ and } P(F/S_3) = 0.4 \]

Suppose that you are informed that event F occurs.

a) Revise the probabilities \( P(S_1), P(S_2) \) and \( P(S_3) \) now that you know that event F has occurred.

b) Based on these revised probabilities, compute the expected monetary value of action A and action B.

c) Based on these revised probabilities, compute the expected opportunity loss of action A and action B.

3) Consider the following problem with two decision alternatives \( d_1 \) & \( d_2 \) and two states of nature \( S_1 \) (Market Receptive) and \( S_2 \) (Market Unfavorable) with the following payoff table representing profits (in $1000):

<table>
<thead>
<tr>
<th>Decisions</th>
<th>States of Nature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( S_1 )</td>
</tr>
<tr>
<td>( D_1 )</td>
<td>20</td>
</tr>
<tr>
<td>( D_2 )</td>
<td>25</td>
</tr>
</tbody>
</table>

Assume the probability of the market being receptive is known to be 0.75. Use the expected monetary value criterion to determine the optimal decision. It is known from past experience that of all the cases when the market was receptive, a research company predicted it in 90 percent of the cases. (In the other 10 percent, they predicted an unfavorable market). Also, of all the cases when the market proved to be unfavorable, the research company predicted it correctly in 85 percent of the cases. (In the other 15 percent of the cases, they predicted it incorrectly.) Answer the following questions based on the above information.

a) Draw a complete probability table.

b) Find the posterior probabilities of all states of nature.

c) Using the posterior probabilities, which plan would you recommend?

d) How much should one be willing to pay (maximum) for the research survey? That is, compute the expected value of sample information (EVSI).

4) In question 12 of exercise 3, a vendor at a baseball stadium is deciding whether to sell ice cream or soft drinks at today’s game. Prior to making her decision, she decides to listen to the local weather forecast. In the past, when it has been cool, the weather reporter has forecast cool weather 80\% of the time. When it has been warm, the weather reporter has forecast warm weather 70\% of the time. The local weather forecast is for cool weather.

a) Revise the prior probabilities for cool and warm weather on the bases of the weather forecast.

b) Use these revised probabilities to repeat question 12 and compare the results in (b) to those in question 12.

c) What is the efficiency of the saple information?

5) In question 14 of exercise 3, an investor is trying to determine the optimal investment decision among three investment opportunities. Prior to making his investment decision, the investor decides to consult with his financial adviser. In the past, when the economy has declined, the financial adviser has given a rosy forecast 20\% of the time (with a gloomy forecast 80\% of the time). When there has been no change in the economy, the financial adviser has given a rosy forecast 40\% of the time. When there has been an expanding economy, the financial adviser has given a rosy forecast 70\% of the time.

a) Revise the probabilities of the investor based on this economic forecast by the financial adviser.

b) Use these revised probabilities to redo question 14.

c) Compare the results in (b) to those in question 14.
d) What is the efficiency of the sample information?

6) In question 16 of exercise 3, an author is deciding which of two competing publishing companies to select to publish her new novel. Prior to making a final decision, the author decides to have an experienced reviewer examine her novel. This reviewer has an outstanding reputation for predicting the success of a novel. In the past, for novels that sold 1,000 copies, only 1% received favorable reviews. Of novels that sold 5,000 copies, 25% received favorable reviews. Of novels that sold 10,000 copies, 60% received favorable reviews. Of novels that sold 50,000 copies, 99% received favorable reviews. After examining the author’s novel, the reviewer gives it an unfavorable review.
   a) Revise the probabilities of the number of books sold in light of the review.
   b) Use these revised probabilities to repeat question 16
   c) Compare the results in (b) to those in Problem 16
   d) What is the efficiency of the sample information?

7) Jenny Lind may hire a market research firm to conduct a survey at a cost of $100,000. The result of the survey would be either a favorable (F) or unfavorable (U) public response to the movie. The firm’s ability to assess the market is:
   \[ P(F/S) = 0.3 \quad P(U/S) = 0.7 \]
   \[ P(F/M) = 0.6 \quad P(U/M) = 0.4 \]
   \[ P(F/L) = 0.8 \quad P(U/L) = 0.2 \]
   i) Should Jenny conduct the survey?
   ii) What is the most she should be willing to pay for the survey?
   iii) What is the efficiency of this survey?

8) A manufacturer produces a product in lots of fixed sizes. Because of occasional malfunctions in the production process, bad lots with an unacceptable number of defectives may be produced. Past experience indicates that the probability of producing a bad lot is 5%. The manufacturer realizes that a penalty may result from shipping out a bad lot. Suppose that the percentage of defectives in a good lot is 4% while a bad lot has 15% defectives items. Using the binomial distribution obtain the conditional probabilities of the test given that the lot is either good or bad if a test sample of 2 items is taken from the lot ie \( p(Z_i/S_j) \). Obtain the posterior probabilities \( p(S_j/Z_i) \) and hence the conditional Expected Value of Sample Information (EVSI).
   (Hint \( S_1 \) = the lot is good, \( S_2 \) = the lot is bad, \( Z_1 \) = the 2 sampled items are good, \( Z_2 \) = 1 good and 1 defective items are sampled and \( Z_3 \) = the 2 sampled items are defective)

9) Ahmed has inherited $1,000 He has to decide how to invest the money for one year. A broker has suggested five potential investments. \( A_1 \) – Gold \( A_2 \) – Company A \( A_3 \) – Company B \( A_4 \) – Company C \( A_5 \) – Company D The return on each investment depends on the (uncertain) market behavior during the year. Ahmed would build a payoff table to help make the investment decision.

<table>
<thead>
<tr>
<th>Alternatives ( A_i )</th>
<th>Large Rise</th>
<th>Small Rise</th>
<th>NoChange</th>
<th>Small Fall</th>
<th>Large Fall</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1 )</td>
<td>-100</td>
<td>100</td>
<td>200</td>
<td>300</td>
<td>0</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>250</td>
<td>200</td>
<td>150</td>
<td>-100</td>
<td>-150</td>
</tr>
<tr>
<td>( A_3 )</td>
<td>500</td>
<td>250</td>
<td>100</td>
<td>-200</td>
<td>-600</td>
</tr>
<tr>
<td>( A_4 )</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>( A_5 )</td>
<td>200</td>
<td>150</td>
<td>150</td>
<td>-200</td>
<td>-150</td>
</tr>
<tr>
<td>prior prob</td>
<td>0.2</td>
<td>0.3</td>
<td>0.3</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Ahmed can purchase econometric forecast results for $50. The forecast predicts “negative” or “positive” econometric growth. Statistics regarding the forecast are:

Proverbs 21:5 The plans of the diligent lead to profit as surely as haste leads to poverty
Time is precious, but we do not know yet how precious it really is. We will only know when we are no longer able to take advantage of it...

<table>
<thead>
<tr>
<th>The Forecast Predicted</th>
<th>When the Company showed a...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Large Rise</td>
</tr>
<tr>
<td>Positive econ. growth</td>
<td>80%</td>
</tr>
<tr>
<td>Negative econ. growth</td>
<td>20%</td>
</tr>
</tbody>
</table>

a) The Expected Value of perfect Information  
b) The posterior probabilities of the states of nature  
c) The revised conditional expected values for each decision  
d) Should Ahmed purchase the forecast?  
e) What is the value of the sample information?

10) A small entrepreneurial company is trying to decide between developing two different products that they believe they can sell to two potential companies, one large and one small. If they develop Product A, they have a 50% chance of selling it to the large company with annual purchases of about 20,000 units. If the large company won't purchase it, then they think they have an 80% chance of placing it with a smaller company, with sales of 15,000 units. On the other hand if they develop Product B, they feel they have a 40% chance of selling it to the large company, resulting in annual sales of about 17,000 units. If the large company doesn't buy it, they have a 50% chance of selling it to the small company with sales of 20,000 units.

a) What is the probability that Product A will be purchased by the smaller company?  
b) What is the probability that Product B will be purchased by the smaller company?  
c) How many units of Product A can they expect to sell?  
d) How many units of Product A can they expect to sell?  
e) How many units can they expect to sell for the optimum alternative?

### Decision Trees

Decisions that can be analyzed using pay off table can be display in a decision tree. We will therefore analyze some decisions using decision trees. Although a Payoff Table (Decision Table) is convenient in problems having one set of decisions and one set of states of nature, many problems include sequential decisions and states of nature. When there are two or more sequential decisions, and later decisions are based on the outcome of prior ones, the decision tree approach becomes appropriate. A decision tree is a graphic display of the decision process that indicates decision alternatives, states of nature and their respective probabilities, and payoffs for each combination of decision alternative and state of nature.

Expected monetary value (EMV) is the most commonly used criterion for decision tree analysis. One of the first steps in such analysis is to graph the decision tree and to specify the monetary consequences of all outcomes for a particular problem.

### How to draw

A decision tree is a chronological representation of the decision process/problem. Each decision tree has two types of nodes; round nodes correspond to the states of nature while square nodes correspond to the decision alternatives. The branches leaving each round node represent the different states of nature while the branches leaving each square node represent the different decision alternatives. At the end of each limb of a tree are the payoffs attained from the series of branches making up that limb.

Note: End nodes usually represent the final outcomes of earlier risky outcomes and decisions made in response.
Decision tree software is a relatively new advance that permits users to solve decision analysis problems with flexibility, power, and ease. Programs such as DPL, Tree Plan, and Supertree allow decision problems to be analyzed with less effort and in greater depth than ever before.

**Example 16** Getz Products Company is investigating the possibility of producing and marketing backyard storage sheds. Undertaking this project would require the construction of either a large or a small manufacturing plant. The market for the product produced—storage sheds—could be either favorable or unfavorable. Getz, of course, has the option of not developing the new product line at all. A decision tree for this situation is presented in Figure 1.

![Decision Tree Diagram](image)

**FIGURE A.1 Getz Products Decision Tree**

Analyzing problems with decision trees involves five steps:

a) Define the problem.
b) Structure or draw the decision tree.
c) Assign probabilities to the states of nature.
d) Estimate payoffs for each possible combination of decision alternatives and states of nature.
e) Solve the problem by computing expected monetary values (EMV) for each state-of-nature node. This is done by working backward—that is, by starting at the right of the tree and working back to decision nodes on the left. This is called rollback (or backward dynamic programming).

f) At each chance node reached, mark the node with EMV of the tree to the right of the node and remove tree to right.

ii) At each decision node reached, choose decision with highest EMV, replace all of tree to right by this EMV.

**Example 17** Thompson Lumber Company

A completed and solved decision tree for the Products is presented in Figure A.2. Note that the payoffs are placed at the right-hand side of each of the tree’s branches. The probabilities are placed in parentheses next to each state of nature. The expected monetary values for each state-of-nature node are then calculated and placed by their respective nodes. The EMV of the first node is $10,000. This represents the branch from the decision node to “construct a large plant.” The EMV for node 2, to “construct a small plant,” is $40,000. The option of “doing nothing” has, of course, a payoff of $0. The branch leaving the decision node leading to the state-of-nature node with the highest EMV will be chosen. In this case, a small plant should be built.
A More Complex Decision Tree

When a sequence of decisions must be made, decision trees are much more powerful tools than are decision tables. Let’s say that Getz Products has two decisions to make, with the second decision dependent on the outcome of the first. Before deciding about building a new plant, Getz has the option of conducting its own marketing research survey, at a cost of $10,000. The information from this survey could help it decide whether to build a large plant, to build a small plant, or not to build at all. Getz recognizes that although such a survey will not provide it with perfect information, it may be extremely helpful.

Getz’s new decision tree is represented in Figure A.3 of Example A7. Take a careful look at this more complex tree. Note that all possible outcomes and alternatives are included in their logical sequence. This procedure is one of the strengths of using decision trees. The manager is forced to examine all possible outcomes, including unfavorable ones. He or she is also forced to make decisions in a logical, sequential manner.

See figure 3 below

Examining the tree in Figure 3, we see that Getz’s first decision point is whether to conduct the $10,000 market survey. If it chooses not to do the study (the lower part of the tree), it can either build a large plant, a small plant, or no plant. This is Getz’s second decision point. If the decision is to build, the market will be either favorable (0.50 probability) or unfavorable (also 0.50 probability). The payoffs for each of the possible consequences are listed along the right-hand side. As a matter of fact, this lower portion of Getz’s tree is identical to the simpler decision tree shown in Figure 2.

The upper part of Figure A.3 reflects the decision to conduct the market survey. State-of-nature node number 1 has 2 branches coming out of it. Let us say there is a 45% chance that the survey results will indicate a favorable market for the storage sheds. We also note that the probability is 0.55 that the survey results will be negative.

The rest of the probabilities shown in parentheses in Figure 3 are all conditional probabilities. For example, 0.78 is the probability of a favorable market for the sheds given a favorable result from the market survey. Of course, you would expect to find a high probability of a favorable market given that the research indicated that the market was good. Don’t forget, though: There is a chance that Getz’s $10,000 market survey did not result in perfect or even reliable information. Any market research study is subject to error. In this case, there remains a 22% chance that the market for sheds will be unfavorable given positive survey results.

The short parallel lines mean “prune” that branch, as it is less favorable than another available option and may be dropped.

Likewise, we note that there is a 27% chance that the market for sheds will be favorable given negative survey results. The probability is much higher, 0.73, that the market will actually be unfavorable given a negative survey.

Finally, when we look to the payoff column in Figure A.3, we see that $10,000—the cost of the marketing study—has been subtracted from each of the top 10 tree branches. Thus, a large plant constructed in a favorable market would normally net a $200,000 profit. Yet because the market study was conducted, this figure is reduced by $10,000. In the unfavorable case, the loss of $180,000 would increase to $190,000.

Similarly, conducting the survey and building no plant now results in a _$10,000 payoff.
With all probabilities and payoffs specified, we can start calculating the expected monetary value of each branch. We begin at the end or right-hand side of the decision tree and work back toward the origin. When we finish, the best decision will be known.

1. Given favorable survey results,
EMV (node 2) = $190,000(0.78) - $190,000(0.22) = $106,400
EMV (node 3) = $90,000(0.78) - $30,000(0.22) = $63,600
The EMV of no plant in this case is -$10,000. Thus, if the survey results are favorable, a large plant should be built.

2. Given negative survey results,

   EMV (node 4) = $190,000(0.27) - $190,000(0.73) = -$87,400
   EMV (node 5) = $90,000(0.27) - $30,000(0.73) = $2,400

   The EMV of no plant is again -$10,000 for this branch. Thus, given a negative survey result, Getz should build a small plant with an expected value of $2,400.

3. Continuing on the upper part of the tree and moving backward, we compute the expected value of conducting the market survey.

   EMV(node 1) = (.45)($106,400) + (.55)($2,400) = $49,200

4. If the market survey is not conducted.

   EMV (node 6) = $200,000(0.5) - $180,000(0.5) = -$10,000
   EMV (node 7) = $100,000(0.5) - $20,000(0.5) = $40,000

   The EMV of no plant is $0. Thus, building a small plant is the best choice, given the marketing research is not performed.

5. Because the expected monetary value of conducting the survey is $49,200—versus an EMV of $40,000 for not conducting the study—the best choice is to seek marketing information. If the survey results are favorable, Getz should build the large plant; if they are unfavorable, it should build the small plant.

**Exercise**

1) Consider the following decision tree.

   ![Decision Tree Diagram]

   a) What is the expected value at node 4?
   b) What is the value associated with node 3?
   c) Which

2) decision, A or B, is best? What is the expected value of this decision?

3) If a student attends every management science class, the probability of passing the course is 0.80; but if the student only attends randomly, then the probability of passing the course is 0.50. If a student fails, they can
take a makeup exam where the probability of passing is 0.60 if the student has attended every class. This probability of passing the makeup exam drops to 0.10 if the student has attended at random. Passing the course is worth 5 credits. Full time attendance "costs" 3 credits in terms of energy and time whereas random attendance "costs" only 1 credit. Use a decision tree to decide which is the best attendance pattern to adopt. Assume that all failing students take the make up exam and that the payoff for failing is equal to 0.

4) A firm is considering a final “GO” decision on a new product. If the product is introduced and it is successful, the profit is $500,000 and if it is unsuccessful the loss is $300,000. There is no profit or loss if the product is not introduced. The management believes that there is a 0.20 probability (the odds are 2 to 8) that the product will be successful.

a) Based on the above information, should the firm introduce the product? Use a decision tree
b) A consulting firm specializing in new product introduction has offered its services to firm. Its betting average in similar situations is as follows. (i) When advice was given (either by client firm or by others in the market place) on products that later proved to be successful, then firm gave “GO” advice eight times out of ten. (ii) When advice was given (either by client firm or by others in the market place) on products that later proved to be unsuccessful, then firm gave “STOP” advice fifteen times out of twenty. The Firm charges $5,000 as consulting fee to give advice. Should the firm be hired in order to maximize the expected profit?

5) Mike Dirr, vice-president of marketing for Super-Cola, is considering which of two advertising plans to use for a new caffeine-free cola. Plan I will cost $500,000 while a more conservative approach, Plan II, will cost only $100,000. Table shows the projected gross (before advertising) profits on the new cola for each advertising plan under two possible states-of-nature - complete acceptance of the new cola and limited acceptance. Both the scenarios are equally likely.

<table>
<thead>
<tr>
<th>Advertising Plan</th>
<th>Limited Acceptance</th>
<th>Complete Acceptance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plan I</td>
<td>$400,000</td>
<td>$1,000,000</td>
</tr>
<tr>
<td>Plan II</td>
<td>$500,000</td>
<td>$500,000</td>
</tr>
</tbody>
</table>

Mike estimate equal chances of complete and limited acceptance of the new cola.

a) Set up a net profit payoff matrix.
b) Find the optimal act using a decision tree with EOL criterion
c) What is the value of perfect information in this situation?
d) A survey costing $50,000 can be run to test-market the product. In past uses, this survey has been shown to predict complete acceptance in 60% of the cases where there was acceptance and predicted limited acceptance 70% of the time when there was limited acceptance. Use this information to determine whether this survey should be used to help decide on an advertising plan. Use a decision tree to show your analysis.
e) What is the efficiency of the sample information?

6) An oil company has some land that is reported to possibly contain oil. The company classifies such land into four categories by the total number of barrels that are expected to be obtained from the well, i.e. a 500,000 – barrel well, 200,000 – barrel well, 50,000 – barrel well, and a dry well. The company is faced with deciding whether to drill for oil, to unconditionally lease the land or to conditionally lease the land at a rate depending upon oil strike. The cost of drilling the well is $100,000; if it is a producing well and the cost of drilling is $75,000 if it is a dry well. For producing well, the profit per barrel of oil is $1.50, (after deduction of processing and all other costs except drilling costs). Under the unconditional lease agreement,
the company receives $45,000 for the land whereas for the conditional lease agreement the company
receives 50 cents for each barrel of oil extracted if it is a 500,000 or 200,000 barrel oil strike and nothing if
otherwise. The probability for striking a 500,000 – barrel well is 0.1, probability for striking a 200,000 –
barrel well is 0.15, probability for striking a 50,000 – barrel well is 0.25, and probability for a dry well is
0.5.

a) Make a profit payoff table for the oil company
b) Find the optimal act using a decision tree with expected monetary value criterion
c) Find the EVPI

7) Safeway Pembina highway location must decide how many cases of Juice to stock each week to meet
demand. The probability distribution of demand during a week is shown as follows.

<table>
<thead>
<tr>
<th>Demand (Cases)</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.20</td>
<td>0.25</td>
<td>0.40</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Each case cost the grocer $10 and sells for $12. Unsold cases are sold to a local farmer for $2 per case. If
there is a shortage, the grocer considers the cost of customer ill will and lost profit to be $4 per case. The
grocer must decide how many cases of Juice to order each week.

a) construct the payoff table for this decision situation
b) Find the best decision according to EMV criteria via a decision tree.
c) Find EVPI

8) Stella Yan Hua is considering the possibility of opening a small dress shop on Fairbanks Avenue, a few
blocks from the university. She has located a good mall that attracts students. Her options are to open a
small shop, a medium-sized shop, or no shop at all. The market for a dress shop can be good, average, or
bad. The probabilities for these three possibilities are .2 for a good market, .5 for an average market, and .3
for a bad market. The net profit or loss for the medium-sized or small shops for the various market
conditions are given in the following table. Building no shop at all yields no loss and no gain. Daily
demand for cases of Tidy Bowl cleaner at Ravinder Nath’s Supermarket has always been 5, 6, or 7 cases.

Develop a decision tree that illustrates her decision alternatives as to whether to stock 5, 6, or 7
cases. Hence find the best decision according to EOL criterion

<table>
<thead>
<tr>
<th>Market State</th>
<th>Good</th>
<th>Average</th>
<th>Bad</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small shop</td>
<td>$750,000</td>
<td>$25,000</td>
<td>$40,000</td>
</tr>
<tr>
<td>Medium size shop</td>
<td>$100,000</td>
<td>$35,000</td>
<td>$60,000</td>
</tr>
<tr>
<td>No shop</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

| Probability | 0.2 | 0.5 | 0.3 |

9) Karen Villagomez, president of Wright Industries, is considering whether to build a manufacturing plant in
the Ozarks. Her decision is summarized in the following table:

<table>
<thead>
<tr>
<th>State-of-Market</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favorable</td>
</tr>
<tr>
<td>Unfavourable</td>
</tr>
<tr>
<td>Build Large plant</td>
</tr>
<tr>
<td>Build Large plant</td>
</tr>
<tr>
<td>Don’t Build</td>
</tr>
<tr>
<td>Probability</td>
</tr>
</tbody>
</table>
a) Construct a decision tree.
b) Determine the best strategy using expected monetary value (EMV).
c) What is the expected value of perfect information (EVPI)?

10) A manufacturer must decide whether to build a small or a large plant at a new location. Demand at the location can be either small or large, with probabilities estimated to be 0.4 and 0.6 respectively. If a small plant is built, and demand is large, the production manager may choose to maintain the current size or to expand. The net present value of profits is $223,000 if the firm chooses not to expand. However, if the firm chooses to expand, there is a 50% chance that the net present value of the returns will be $330,000 and 50% chance the estimated net present value of profits will be $210,000. If a small facility is built and demand is small, there is no reason to expand and the net present value of the profits is $200,000. However, if a large facility is built and the demand turns out to be small, the choice is to do nothing with a net present value of $40,000 or to stimulate demand through local advertising. The response to advertising can be either modest with a probability of .3 or favorable with a probability of .7. If the response to advertising is modest the net present value of the profits is $20,000. However, if the response to advertising is favorable, then the net present value of the profits is $220,000. Finally, when large plant is built and the demand happens to be high, the net present value of the profits is $800,000. Draw a decision tree and determine the optimal strategy.

**UTILITY**

**Introduction**

The methods we have discussed so far for decision making assume that each *incremental* amount of profit or loss has the same value as the previous amounts of profits attained or losses incurred. In fact, under many circumstances in the business world, this assumption of incremental changes is not valid. Most companies, as well as most individuals, make special efforts to avoid large losses. At the same time, many companies, as well as most individuals, place less value on extremely large profits than on initial profits. Such differential evaluation of incremental profits or losses is referred to as utility, a concept first discussed by Daniel Bernoulli in the eighteenth century. To illustrate this concept, suppose that you are faced with the following two choices:

**Choice 1:** A fair coin is to be tossed. If it lands on heads, you will receive $0.60; if it lands on Tails, you will pay $0.40.

**Choice 2:** Do not play the game.

What decision should you choose? The expected value of playing this game is $0.60 \times 0.5 - 0.40 \times 0.5 = +$1. and the expected value of not playing the game is 0.

Most people will decide to play the game because the expected value is positive, and only small amounts of money are involved. Suppose, however, that the game is formulated with a payoff of $600,000 when the coin lands on heads and a loss of $400,000 when the coin lands on tails. The expected value of playing the game is now +$100,000. With these payoffs, even though the expected value is positive, most individuals will not play the game because of the severe negative consequences of losing $400,000. Each additional dollar amount of either profit or loss does not have the same utility as the previous amount. Large negative amounts for most individuals have severely negative utility. Conversely, the extra value of each incremental dollar, of profit, decreases when high enough profit levels are reached. (In other words, the difference between 0 and $100,000 is much more than the difference between $1,000,000 and $1,100,000.)
Shortcomings of EMV

Usually the consequences of decisions are expressed in monetary terms. Additional factors such as reputation, time, etc. are also usually translated into money. However EMV has a number of shortcomings.

Under the EMV approach the following assumptions are made:

1) The utility of money received by the decision maker is a linear fiction of the amount of money received.

2) No constraints are imposed on the decision maker when evaluating all possible alternatives.

The assumptions may not be realistic because it is not always true that the utility of money is linear to the amount of money. This implies that it is not always that the optimal action is the one with the highest EMV. (Insurance policies cost more than the present value of the expected loss the insurance company pays to cover insured losses). Furthermore external constraints may prevail which can affect the decision to be made. Eg limited market size in relation to the optimal production level, non availability of outside capital to finance an optimal venture, etc.

The EMV criterion avoids recognising that a decision maker’s attitude to the prospect of gaining or losing different amounts of money is influenced by the risks involved. A person’s valuation of a risky venture is not the expected return of that venture but rather the expected utility from that venture.

Axioms of Utility

Before making an optimal decision based on utility theory, we need to construct the utility index. The key to constructing the utility function for money to fit the decision maker is based on the following axioms.

a) Completeness/Preference Regardless of the monetary gain or loss, the decision maker is either indifferent to 2 or more outcomes or he may prefer one outcome over the other. ie for any two prospects x and y, either $x \sim y \Rightarrow U(x) = U(y)$ or $x < * y \Rightarrow U(x) < U(y) \text{ or } y < * x \Rightarrow U(x) > U(y)$

b) Transitivity if $x < * y$ and $y < * z$ then $x < * z$ Also if $x \sim * y$ and $y \sim * z$ then $x \sim * z$.

c) Critical probability if $x < * y$ and $y < * z$ then there should be a critical probability that can be constructed so that $x \sim * z$ with uncertainty and that $x \sim * z$ with certainty. Eg Consider the lottery

- $L_1$: receive £0 for certain
- $L_2$: receive £100 with probability $p$ and -£50 with $(1-p)$
- $100p - 50(1 - p) = 0 \Rightarrow 150p - 50 \Rightarrow p = \frac{2}{3}$

d) Preference of higher probability if a decision maker has 2 or more jobs that pays individual monetary returns but have different probabilities of success, he would prefer to have the one with the highest probability of success.

e) Subjectivity since utility is strictly a subjective concept; utility index of event A for one individual may be different from the utility index derived from the same event by another individual. This is because each individual’s personal and business values differ in many ways including the attitude towards utility of money.

Shapes of Utility Index Curves

An important part of the decision-making process is to develop a utility curve for a decision maker that represents the utility of each specified dollar amount. We have many types of utility index curves whose shapes depends on individuals attitude towards risk. In general we say people have one of three attitudes toward risk. People can be risk avoiders, risk seekers (or risk lover), or indifferent toward risk (risk neutral). Utility values are assigned to monetary values and the general shape for each type of person is shown below.
Time is precious, but we do not know yet how precious it really is. We will only know when we are no longer able to take advantage of it…

Pr 21:5 The plans of the diligent lead to profit as surely as haste leads to poverty

Note that for equal increments in dollar value the utility either rises at a decreasing rate (avoider), constant rate (risk neutral) or increasing rate (risk seekers). A risk averter will avoid any large gamble in return for a relatively small return with certainty while a risk taker prefers a large gain/loss to an equal choice with a small return under uncertainty. Many people are risk averse but what can we do to account for the fact that many people are risk averse? We can use the concept of expected utility.

**Remark:** It is possible to exhibit a mixture of these kinds of behavior. For example, an individual might be essentially risk-neutral with small amounts of money, then become a risk seeker with moderate amounts, and then turn risk-averse with large amounts. Moreover, one’s attitude toward risk can change over time and circumstances. In this case the utility curve is partly concave and partly convex. The point where one changes from a risk taker at one extreme to a risk avoider at the other extreme is called an aspiration level or satisfying level of monetary reward.

The risk averter’s curve shows a rapid increase in utility for initial amounts of money followed by a gradual leveling off for increasing dollar amounts. This curve is appropriate for most individuals or businesses because the value of each additional dollar is not as great after large amounts of money have already been earned.

The risk seeker’s curve represents the utility of someone who enjoys taking risks. The utility is greater for large dollar amounts. This curve represents an individual who is interested only in “striking it rich” and is willing to take large risks for the opportunity of making large profits.

The risk-neutral curve represents the expected monetary value approach. Each additional dollar of profit has the same value as the previous dollar.

After a utility curve is developed in a specific situation, you convert the dollar amounts to utilities. Then you compute the utility of each alternative course of action and apply the decision criteria of expected utility value, expected opportunity loss, and return-to-risk ratio to make a decision.

**Utility Function**

A utility function is a twice-differentiable function of wealth \( U(w) \) defined for \( w > 0 \) which has the properties of non-satiation (the first derivative \( U'(w) > 0 \)) and risk aversion (the second derivative \( U''(w) < 0 \)). A utility function measures an investor’s relative preference for different levels of total wealth and the units of utility are called utiles.
Time is precious, but we do not know yet how precious it really is. We will only know when we are no longer able to take advantage of it.

*Proverbs 21:5*  
The plans of the diligent lead to profit as surely as haste leads to poverty

The non-satiation property states that utility increases with wealth, i.e., that more wealth is preferred to less wealth, and that the investor is never satiated - he never has so much wealth that getting more would not be at least a little bit desirable.

The risk aversion property states that the utility function is concave or, in other words, that the marginal utility of wealth decreases as wealth increases. To see why utility functions are concave, consider the extra (marginal) utility obtained by the acquisition of one additional dollar. For someone who only has one dollar to start with, obtaining one more dollar is quite important. For someone who already has a million dollars, obtaining one more dollar is nearly meaningless. In general, the increase in utility caused by the acquisition of an additional dollar decreases as wealth increases. It may not be obvious what this concavity of utility functions or decreasing marginal utility of wealth” has to do with the notion of risk aversion." This should become clear in the examples.

Different investors can and will have different utility functions, but we assume that any such utility function satisfies the two critical properties of non-satiation and risk aversion. A function is concave if a line joining any two points of the function lies entirely below the curve.

The principle of expected utility maximization states that a rational investor, when faced with a choice among a set of competing feasible investment alternatives, acts to select an investment which maximizes his expected utility of wealth. (note $\sum_w U(w_j) p(s_j)$)

Expressed more formally, for each investment $I$ in a set of competing feasible investment alternatives $F$, let $X(I)$ be the random variable giving the ending value of the investment for the time period in question. Then a rational investor with utility function $U$ faces the optimization problem of finding an investment $I_{opt} \in F$ for which:

$$E\{U(X(I_{opt}))\} = \max_{I \in F} E\{U(X(I))\}$$

**Example 1 (A Fair Game)** Consider an investor with a square root utility function: $U(x) = \sqrt{x} = x^{0.5}$

$\Rightarrow U'(x) = 0.5x^{-0.5} > 0$ and $U(x) = -0.25x^{-1.5} < 0$ Assume that the investor's current wealth is $5 and that there is only one investment available. In this investment a fair coin is flipped. If it comes up heads, the investor wins (receives) $4, increasing his wealth to $9. If it comes up tails, the investor loses (must pay) $4, decreasing his wealth to $1. Note that the expected gain is $0. This is called a fair game."

This game may seem more like a “bet” than an “investment." To see why it's an investment, consider an investment which costs $5 (all of our investor's current wealth) and which has two possible future values: $1 in the bad case and $9 in the good case. This investment is clearly exactly the same as the coin flipping game.

Note that we have chosen a very volatile investment for our example. In the bad case, the rate of return is -80%. In the good case, the rate of return is +80%. Note that the expected return is 0%, as is the case in all fair games/bets/investments.

We assume that our investor has only two choices (the set of feasible investment alternatives has only two elements). The investor can either play the game or not play the game (do nothing). Which alternative does the investor choose if he follows the principle of expected utility maximization? Figure 1 show our investor's current wealth and utility, the wealth and utility of the two possible outcomes in the fair game, and the expected outcome and the expected utility of the outcome in the fair game.
Time is precious, but we do not know yet how precious it really is. We will only know when we are no longer able to take advantage of it.

Proverbs 21:5 The plans of the diligent lead to profit as surely as haste leads to poverty
If an investor with utility function $U$ has current wealth less than $c$, he will consider the investment attractive (although some other investment may be even more attractive). If his current wealth is greater than $c$, he will consider the investment unattractive, because doing nothing has greater expected utility than the investment. If his current wealth is exactly $c$, he will be indifferent between undertaking the investment and doing nothing.

Note that because $U$ is an increasing function, maximizing expected utility is equivalent to maximizing the certainty equivalent. The certainty equivalent is always less than the expected value of the investment. In our example, the certainty equivalent is $4$, while the expected value (expected outcome) is $5$.

**Positive Affine Transformations**

Utility functions are used to compare investments to each other. For this reason, we can scale a utility function by multiplying it by any positive constant and/or translate it by adding any other constant (positive or negative). This kind of transformation is called a positive affine transformation.

For example, with the square root utility function we used above, we could have used any of the following functions instead:

\[
U(w) = 100\sqrt{w}, \quad U(w) = 50\sqrt{w} + 83, \quad U(w) = \sqrt{w} - 413 \quad \text{and} \quad U(w) = \frac{3\sqrt{w} + 10}{8}
\]

The specific numbers appearing as the utility function values on our graphs and in our calculations would be different, but the graphs would all look the same when scaled and translated appropriately, and all our results would be the same. It is easy to see why this is true in general. Suppose we have constants $a > 0$ and $b$ and a utility function $U$. Define another utility function $V$: $V(w) = aU(w) + b$

Note that:

\[
V'(w) = aU'(w) > 0 \quad \text{because } a > 0 \text{ and } U'(w) > 0
\]

\[
V''(w) = aU''(w) < 0 \quad \text{because } a > 0 \text{ and } U''(w) < 0
\]

so $V$ is a valid utility function.

Consider an investment $I$ with outcome given by a random variable $X$. We can easily see that the certainty equivalent for $I$ is the same under utility function $V$ as it is under utility function $U$. Let $c$ be the certainty equivalent under $U$. Then: $c = U^{-1}(E(U(w))) \Rightarrow U(c) = E(U(w))$

\[
V(c) = aU(c) + b = aE(U(w)) + b = E(V(w)) \Rightarrow c = V^{-1}[E(V(w))]
\]

When we talk about utility functions we often say that two functions are the same" when they differ only by a positive affine transformation. More generally utility functions are unique only up to a linear transformation. This is an important attribute of utility analysis.

**Qn** Suppose an entrepreneur owns a new factory worth $2,000,000. The probability that this factory will burn down this year is 0.001. Using the expected utility, should he buy the insurance policy at $2,250 per year against fire?

**Example 2 An All-or-Nothing Investment**

For our second example we assume that the investor's current wealth is $100 and we begin by using the following utility function:

\[
U(w) = \frac{-1,000,000}{w^3} = -1,000,000w^{-3} \Rightarrow U'(w) = 3,000,000w^{-4} > 0 \quad \text{and} \quad U''(w) = -12,000,000w^{-5} < 0
\]

Our utility function is the same as $\frac{3}{w}$. We use the scale factor $1,000,000$ to make the function values
easier to read in the wealth neighbourhood of $100 which we are investigating. Without the scale factor, the numbers are very small, messy to write, and not easy to interpret at a glance.

As in the first example, we assume that our investor has only one alternative to doing nothing. He may use his entire wealth of $100 to purchase an investment which returns -10% with probability 50% and +20% with probability 50%. We will continue to use this hypothetical investment as a running example through the rest of these notes.

Note that the expected return on this investment is +5% and the standard deviation of the returns is 15%. This is similar to many common real-life financial investments, except that in real life there are many more than only two possible outcomes. We continue to assume that there are only two possible outcomes to make the maths easier for the sake of exposition.

Figure 2 shows a graph similar to the one in our first example. The expected utility of the investment is -0.98, which is larger than the utility of doing nothing, which is -1.00. Thus, in this case the investor chooses to make the investment.

The decision looks like a close call in this example. The expected utility of the investment is only slightly larger than that of doing nothing. This leads us to wonder what might happen if we change the exponent in the utility function.

Figure 2: Example 2 - An All-or-Nothing Investment, \( U(w) = -1,000,000w^{-3} \)

<table>
<thead>
<tr>
<th>Wealth</th>
<th>Utility</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$90</td>
<td>-1.37</td>
<td>Bad outcome</td>
</tr>
<tr>
<td>$100</td>
<td>-1.00</td>
<td>Current wealth</td>
</tr>
<tr>
<td>$105</td>
<td>-0.98</td>
<td>Expected outcome</td>
</tr>
<tr>
<td>$120</td>
<td>-0.58</td>
<td>Good outcome</td>
</tr>
</tbody>
</table>
Figure 3 shows how the situation changes if we use an exponent of -5 instead of -3 in the utility function. \( U(w) = -10^{10} w^{-5} \Rightarrow U'(w) = 5 \times 10^{10} w^{-6} > 0 \) and \( U''(w) = -30 \times 10^{10} w^{-7} < 0 \)

Figure 3: Example 2 - An All-or-Nothing Investment, \( U(w) = -10^{10} w^{-5} \)

<table>
<thead>
<tr>
<th>Wealth</th>
<th>Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>$90</td>
<td>-1.69</td>
</tr>
<tr>
<td>$100</td>
<td>-1.00</td>
</tr>
<tr>
<td>$105</td>
<td>-1.05</td>
</tr>
<tr>
<td>$120</td>
<td>-0.40</td>
</tr>
</tbody>
</table>

In this case, the expected utility of the investment is -1.05, which is smaller than the utility of doing nothing, which is -1.00. Thus, in this case the investor chooses not to make the investment.

This example shows that an investor with a utility function of \(- w^{-5}\) is somewhat more risk-averse than is an investor with a utility function of \(- w^{-3}\).

**Example 3 - Optimizing a Portfolio**

In the first two examples we assumed that the investor had only two options; do nothing or invest all of his money in a risky asset. In this example we use the same kind of utility functions as in examples 1 and 2 and the same investment as in example 2. This time, however, we permit investing any desired portion of the investor's total wealth of $100 in the risky asset.

The investor may choose to do nothing, invest everything in the risky asset, or do nothing with part of his money and invest the rest. We also generalize the utility function in example 2 to permit exponents other than -3 and -5. For reasons which will become clear later we use the following parameterized form of these utility functions:

For any \( \lambda < 1, \lambda \neq 0 \): \( U_\lambda(w) = \frac{w^\lambda - 1}{\lambda} \Rightarrow U'_\lambda(w) = w^{\lambda - 1} > 0 \) and \( U''_\lambda(w) = (\lambda - 1)w^{\lambda - 2} < 0 \)
Note that we must have $\lambda < 1$ to guarantee $U_\lambda'(w) < 0$. If $\lambda = 1$, we get a risk neutral utility function. And if $\lambda > 1$, we get a risk-loving utility function. We assume risk aversion and do not investigate these alternatives here.

Consider the utility functions we used in examples 1 and 2: $\sqrt{w} = w^{0.5}$ and $-w^{-3}$. These functions are the same as $U_{0.5}, U_{-3}$ and $U_{-5}$. Our investor has a current wealth of $100$ and may choose to invest any part of it in the risky asset. Let $\alpha$ be the amount invested in the risky asset. Then $100 - \alpha$ is the amount with which the investor does nothing. The two possible outcomes are:

- Bad outcome: $w = 0.9\alpha + (100 - \alpha) = 100 - 0.1\alpha$
- Good outcome: $w = 1.2\alpha + (100 - \alpha) = 100 + 0.2\alpha$

The expected utility of the outcome is:

$$f(\alpha) = 0.5U(100 - 0.1\alpha) + 0.5U(100 + 0.2\alpha) = 0.5\left\{\left(\frac{(100 - 0.1\alpha)^4}{\lambda} - 1\right) + \left(\frac{(100 + 0.2\alpha)^4}{\lambda} - 1\right)\right\}$$

$$= \frac{1}{\lambda} \left[(100 - 0.1\alpha)^4 + (100 + 0.2\alpha)^4 - 2\right]$$

Figure 4 displays the certainty equivalent of this function $U_{-3}^{-1}[f(\alpha)]$ for $\lambda = -3$. Recall that maximizing expected utility is equivalent to maximizing the certainty equivalent. Using certainty equivalents in optimization problems like this one is often more natural and intuitive than working directly with the expected utility values.

Figure 4: Example 3 - Utility Hill for $\lambda = -3$

As we saw in example 2, given an all-or-nothing choice, our investor would choose to invest his $100 in the risky asset. Given the opportunity to invest an arbitrary portion of his total wealth, however, we see that neither extreme choice is optimal. In this case, with $\lambda = -3$, the optimal amount to invest in the risky asset is about $59. The investor does nothing with the remaining $41.

Figure 5 shows the corresponding graph for $\lambda = -5$.

Figure 5: Example 3 - Utility Hill for $\lambda = -5$
Time is precious, but we do not know yet how precious it really is. We will only know when we are no longer able to take advantage of it…

Pr 21:5 The plans of the diligent lead to profit as surely as haste leads to poverty

In this case, the optimal portfolio for our somewhat more risk-averse investor is to invest $39 in the risky asset and do nothing with the remaining $61.

The curves in these graphs are called “utility hills.” In general, when we mix together multiple possible risky assets, the curves becomes surfaces in two or more dimensions. The general asset allocation portfolio optimization problem is to climb the hill to find the particular portfolio at its peak. In the very simple kind of portfolio we’re examining in this example, we can easily use a bit of calculus and algebra to find the exact optimal portfolio.

The principle of expected utility maximization tells us that we need to find the value of $\alpha$ for which $f(\alpha)$ attains its maximum value. To do this we take the derivative of $f$, set it equal to 0, and solve for $\alpha$. i.e

$$f'(\alpha) = \frac{1}{2x} \left[ -0.1\lambda(100 - 0.1\alpha)^{x-1} + 0.2\lambda(100 + 0.2\alpha)^{x-1} \right] = 0.1(100 + 0.2\alpha)^{x-1} - 0.05(100 - 0.1\alpha)^{x-1} = 0$$

Rearranging and simplifying we get:

$$0.1(100 + 0.2\alpha)^{x-1} = 0.05(100 - 0.1\alpha)^{x-1} \Rightarrow 2 = \left( \frac{100 - 0.1\alpha}{100 + 0.2\alpha} \right)^{x-1} = \left( \frac{100 + 0.2\alpha}{100 - 0.1\alpha} \right)^{1-x}$$

$\lambda < 1$, so the exponent $1 - \lambda$ in this last equation is greater than 0. This number is called the coefficient of risk aversion and is often denoted by the variable A:

$$A = 1 - \lambda = \text{coeficient of risk aversion}$$

As we saw in example 2, as $\lambda$ decreases, investors become more risk-averse. Thus, as A increases, investors become more risk-averse. For the two examples we saw above, in the case $\lambda = -3$ we have $A = 4$, and for $\lambda = -5$ we have $A = 6$. The investor with $A = 6$ is more risk-averse than is the investor with $A = 4$.

We now rewrite the last equation above using our new coefficient of risk aversion and solve the result for $\alpha$:

$$2 = \left( \frac{100 + 0.2\alpha}{100 - 0.1\alpha} \right)^A \Rightarrow 100 + 0.2\alpha = (100 - 0.1\alpha) \times 2^A \Rightarrow (0.2 + 0.1 \times 2^A)\alpha = 100(2^\lambda - 1) \Rightarrow \alpha = \frac{100(2^\lambda - 1)}{0.2 + 0.1 \times 2^A}$$

We graph this function for the coefficient of risk aversion $A$ ranging from 2 to 20 in Figure 6.
As expected, as risk aversion increases, the portion of the optimal portfolio which is invested in the risky asset gets smaller. Something interesting happens at the left edge of this graph. For an investor with a low coefficient of risk aversion $A = 2$, the optimal amount to invest in the risky asset is about $121. This is $21 more than our investor's total wealth!

Figure 6: Example 3 - Optimizing a Portfolio

Suppose the investor is able to borrow an extra $21 from someone without having to pay any interest on the loan. If this is the case, the optimal portfolio for our investor is to borrow the $21, put it together with his $100, and invest the resulting $121 in the risky asset. For this investor with $A = 2$, however, it would not be optimal to borrow any more than $21. Borrowing money to help finance a risky investment is called “leverage.” If our investor is unable to borrow money, his optimal portfolio is to invest his entire current wealth in the risky asset.

The Logarithmic Utility Function
In example 3 we looked at the class of utility functions:

$$U_\lambda(w) = \frac{w^\lambda - 1}{\lambda}, \quad \lambda < 1, \text{ and } \lambda \neq 0$$

There is a conspicuous hole in this collection at $\lambda = 0$, corresponding to $A = 1$. Fortunately, this hole is quite easily and elegantly filled by taking the limit of $U_\lambda$, as $\lambda \to 0$ using L'Hopital's rule:

$$\lim_{\lambda \to 0} \left( \frac{w^\lambda - 1}{\lambda} \right) = \lim_{\lambda \to 0} \left( \frac{e^{\lambda \ln w} - 1}{\lambda} \right) = \lim_{\lambda \to 0} \left( \frac{\ln w e^{\lambda \ln w}}{1} \right) = \ln w$$

We are thus led to consider the natural logarithm function as a utility function:

$$U(w) = \ln w \Rightarrow U'(w) = \frac{1}{w} = w^{-1} > 0 \text{ and } U''(w) = -\frac{1}{w^2} = -w^{-2} < 0$$

We can rework example 3 using this new utility function:
\[ f(\alpha) = 0.5 \ln(100 - 0.1\alpha) + 0.5 \ln(100 + 0.2\alpha) \Rightarrow -0.05 \frac{1}{100 - 0.1\alpha} + \frac{0.1}{100 + 0.2\alpha} = 0 \]

\[ \Rightarrow \frac{0.05}{100 - 0.1\alpha} = \frac{0.1}{100 + 0.2\alpha} \Rightarrow 2(100 - 0.1\alpha) = 100 + 0.2\alpha \Rightarrow 100 = 0.4\alpha \Rightarrow \alpha = 250 \]

This result agrees exactly with the result we got for \( A = 1 \) in the equation we derived in example 3. Thus the natural logarithm function fills the hole quite nicely.

It’s interesting to note that, at least in our example, an investor with a logarithmic utility function has very low risk-aversion, since his optimal portfolio is highly leveraged.

**The Iso-Elastic Utility Functions**

All of the utility functions we’ve examined so far are members of a class called the iso-elastic utility functions:

\[ U_\lambda(w) = \begin{cases} \frac{w^\lambda - 1}{\lambda}, & \lambda < 1, \text{ and } \lambda \neq 0 \\ \ln w & \text{the limiting case for } \lambda = 0 \end{cases} \]

These functions have the property of iso-elasticity, which says that if we scale up wealth by some constant amount \( k \), we get the same utility function (modulo a positive affine transformation). Stated formally, for all \( k > 0 \):

\[ U(kw) = f(k)U(w) + g(k) \text{ for some function } f(k) > 0 \text{ and } g(k) \text{ which are independent of } w. \]

We can check this formal definition with our utility functions. First consider the case where \( \lambda \neq 0 \):

\[ U_\lambda(kw) = \frac{(kw)^\lambda - 1}{\lambda} = k^\lambda \left( \frac{w^\lambda - 1}{\lambda} \right) + k^{\lambda-1} - \frac{1}{\lambda} = k^\lambda U_\lambda(w) + k^{\lambda-1} - \frac{1}{\lambda} \]

Now consider the log function: \( U(kw) = \log(kw) = \log(k) + \log(w) = U(w) + \log(k) \)

This property of iso-elasticity has a very important consequence for portfolio optimization. It implies that if a given percentage asset allocation is optimal for some current level of wealth, that same percentage asset allocation is also optimal for all other levels of wealth. We can illustrate this fact by reworking example 3 with initial wealth \( w_0 \) a parameter and the amount \( \alpha \) invested in the risky asset expressed as a fraction of \( w_0 \). That is, we invest \( \alpha w_0 \) dollars in the risky asset and we do nothing with the remaining \((1 - \alpha)w_0\) dollars. In this case the two possible outcomes for ending wealth \( w \) are:

- Bad outcome: \( w = 0.9\alpha w_0 + (1 - \alpha)w_0 = (1 - 0.1\alpha)w_0 \)
  - Good outcome: \( w = 1.2\alpha w_0 + (1 - \alpha)w_0 = (1 + 0.2\alpha)w_0 \)

The expected utility is:

\[ f(\alpha) = 0.5 \left( \frac{(1 - 0.1\alpha)^2 w_0^\lambda - 1}{\lambda} + \frac{(1 + 0.2\alpha)^2 w_0^\lambda - 1}{\lambda} \right) = \frac{1}{2\lambda} \left( (1 - 0.1\alpha)^2 w_0^\lambda + (1 + 0.2\alpha)^2 w_0^\lambda - 2 \right) \]

To find the optimal portfolio we take the derivative with respect to \( \alpha \), set it equal to 0, and solve for \( \alpha \):

\[ f'(\alpha) = \frac{1}{2\lambda} \left[ -0.1\lambda(1 - 0.1\alpha)^{\lambda-1} w_0^\lambda + 0.2\lambda(1 + 0.2\alpha)^{\lambda-1} w_0^\lambda \right] = 0 \]

\[ \Rightarrow 0.2(1 + 0.2\alpha)^{\lambda-1} w_0^\lambda = 0.1(1 - 0.1\alpha)^{\lambda-1} w_0^\lambda \]
\[ 2 = \left( \frac{1 - 0.1\alpha}{1 + 0.2\alpha} \right)^{\lambda^{-1}} = \left( \frac{1 + 0.2\alpha}{1 - 0.1\alpha} \right)^{\lambda^{-1}} = \left( \frac{1 + 0.2\alpha}{1 - 0.1\alpha} \right)^A \Rightarrow 2^\lambda = \left( \frac{1 + 0.2\alpha}{1 - 0.1\alpha} \right) \]

\[ \Rightarrow 1 + 0.2\alpha = 2^\lambda \cdot 0.1\alpha \times 2^\lambda \Rightarrow (0.2 + 0.1 \times 2^\lambda)\alpha = 2^\lambda - 1 \Rightarrow \alpha = \frac{2^\lambda - 1}{0.2 + 0.1 \times 2^\lambda} \]

This equation for the optimal fraction \( \alpha \) to invest in the risky asset is independent of the initial wealth \( w_0 \). Thus the optimal portfolio is the same, regardless of wealth. For example, if the investor's current wealth is $1,000 and his optimal portfolio is 50% risky and 50% risk-free, then the investor will have exactly the same optimal percentage asset allocation with a current wealth of $1,000,000. Thus, investors with iso-elastic utility functions have a constant attitude towards risk expressed as a percentage of their current wealth. This property is called constant relative risk aversion.

**The Negative Exponential Utility Function**

Up to this point all of the utility functions we have looked at have been iso-elastic. We now examine a different kind of utility function:

\[ U(w) = -e^{-Aw} \quad \text{ (for any coefficient of risk aversion } A > 0) \]

\[ \Rightarrow U'(w) = Ae^{-Aw} > 0 \quad \text{ and } \quad U''(w) = -A^2e^{-Aw} < 0 \]

This class of utility functions has the interesting property that it is invariant under any translation of wealth. I.e. for any constant \( k \), \( U(k + w) = f(k)U(w) + g(k) \) for some function \( f(k) > 0 \) which is independent of \( w \) and some function \( g(k) \) which is also independent of \( w \).

We can easily verify this \( U(k + w) = -e^{-A(w+k)} = -e^{-Ak}e^{-Aw} = U(w)e^{-Ak} \)

To continue our running example, we compute the optimal portfolio in our now familiar investment universe using this new utility function:

\[ f(\alpha) = 0.5\left(-e^{-A(1-0.1\alpha)w_0}\right) + 0.5\left(-e^{-A(1+0.2\alpha)w_0}\right) = -0.5\left(e^{-A(1-0.1\alpha)w_0} + e^{-A(1+0.2\alpha)w_0}\right) \]

\[ \Rightarrow f'(\alpha) = -0.5\left(0.1Aw_0e^{-A(1-0.1\alpha)w_0} - 0.2Aw_0e^{-A(1+0.2\alpha)w_0}\right) = 0 \]

\[ \Rightarrow e^{-A(1-0.1\alpha)w_0} = 2e^{-A(1+0.2\alpha)w_0} \Rightarrow e^{A(1+0.2\alpha)w_0} - A(1-0.1\alpha)w_0 = 2 \]

\[ \Rightarrow e^{0.3Aw_0} = 2 \Rightarrow \alpha = \frac{\ln 2}{0.3Aw_0} = \alpha = \frac{2.31}{Aw_0} \]

Suppose for sake of example that the coefficient of risk aversion \( A = 0.00231 \). Our equation becomes simply \( \alpha = \frac{1000}{w_0} \). In this example, with a current wealth of $1,000, the investor's optimal strategy is to invest 100% = $1,000 in the risky asset and keep back nothing. With a current wealth of $2,000 his optimal strategy is to invest 50% = $1,000 in the risky asset and do nothing with the other 50% = $1,000.

In general, as our investor's wealth increases his portfolio rapidly becomes more conservative. He invests the same absolute amount of money ($1,000) in the risky asset no matter what his wealth is. For example, with a wealth of $1,000,000, he would invest the same $1,000 in the risky asset and do nothing with $999,000.

Thus, investors with negative exponential utility functions have a constant attitude towards risk expressed in absolute dollar terms. This property is called constant absolute risk aversion.
Risk Aversion Functions

We saw earlier that two utility functions are the “same” if they differ by a positive affine transformation: $V(w) = aU(w) + b$ for constants $a > 0$ and $b$

Differentiate both sides of this equation twice: ie $V'(w) = aU'(w)$ and $V''(w) = aU''(w)$

Now divide the second equation by the first equation: ie $\frac{V''(w)}{V'(w)} = \frac{aU''(w)}{aU'(w)}$

Conversely, suppose we have any pair of functions $U$ and $V$ for which this last equation holds (the ratio of their second to their first derivatives is the same). Let $f(w) = \frac{V'(w)}{U'(w)}$

Take the derivative: $f'(w) = \frac{V''(w)U'(w) - V'(w)U''(w)}{[U'(w)]^2} = 0$

$f'(w) = 0$, so we must have $f(w) = a$ for some constant $a$:

$$f(w) = \frac{V'(w)}{U'(w)} = a \Rightarrow V'(w) = aU'(w)$$

Integrate both sides of this equation:

$$\int V'(w)dw = \int aU'(w)dw \Rightarrow V(w) = aU(w) + b \text{ for some constant } b$$

We have shown that two utility functions are the “same” if and only if the ratios of their second derivatives to their first derivatives are the same.

For any utility function $U$ this leads us to consider the following function: $A(w) = -\frac{U''(w)}{U'(w)}$

Note that for utility functions we always have $U'(w) > 0$ and $U''(w) < 0$, so $A(w) > 0$.

This function is called the Pratt-Arrow absolute risk aversion function. It completely characterizes the utility function. It provides a measure of the absolute risk aversion of the investor as a function of the investor's wealth.

Let's evaluate this function for the utility functions we've looked at so far.

Iso-elastic, $A = \text{coefficient of risk aversion}$

$$U(w) = \frac{w^\lambda - 1}{\lambda} \quad A(w) = -\frac{(\lambda - 1)w^{\lambda - 2}}{w^\lambda} = 1 - \frac{\lambda}{w} = \frac{A}{w} \quad \text{and for } U(w) = \ln w \quad A(w) = -\frac{w^{-2}}{w^1} = \frac{1}{w} = \frac{A}{w}$$

Negative exponential, $A = \text{coefficient of risk aversion}$

$$U(w) = -e^{-Aw} \quad A(w) = -\frac{A^2e^{-Aw}}{Ae^{-Aw}} = A$$

For the negative exponential functions we have constant absolute risk aversion, as we saw before. For the iso-elastic functions we have decreasing absolute risk aversion.

A related function is the relative risk aversion function: $A(w) w = -\frac{U''(w)}{U'(w)}$

For this function we have: Iso-elastic: $A(w) w = A$ and Negative exponential: $A(w) w = Aw$

Thus for the iso-elastic functions we have constant relative risk aversion. For the negative exponential functions we have increasing relative risk aversion.

**Example 4 - Variable Risk Aversion**

In this example we show how to work backwards from a desired relative risk aversion function to the corresponding utility function. Our target relative risk aversion function is linear and increasing in wealth:
Time is precious, but we do not know yet how precious it really is. We will only know when we are no longer able to take advantage of it.

Pr overbs 21:5  

The plans of the diligent lead to profit as surely as haste leads to poverty

\[
A(w) w = - \frac{wU''(w)}{U'(w)} = a + bw \quad \text{for constants } a > 0 \text{ and } b > 0 \quad \Rightarrow \quad wU''(w) + (a + bw)U'(w) = 0
\]

Let \( V = U' \Rightarrow wV'(w) + (a + bw)V(w) = 0 \) This first-order ordinary differential equation has the solution: \( V(w) = w^{-a}e^{bw} \) and we can integrate to get \( U: \) \( U(w) = \int_0^w x^{-a}e^{bx}dx \)

Now \( U'(w) = w^{-a}e^{bw} > 0 \) and \( U''(w) = -aw^{-a-1}e^{bw} - bw^{-a}e^{bw} = -(a + bw)w^{-a-1}e^{bw} < 0 \)

\[
\Rightarrow A(w) = - \frac{U''(w)}{U'(w)} = \frac{a}{w} + b \quad \Rightarrow \quad wA(w) = a + bw
\]

Note that we haven’t attempted to simplify the integral in the definition of \( U. \) It turns out that this isn’t even necessary for the purpose of our example. This utility function has decreasing absolute risk aversion and increasing relative risk aversion. In this sense it is in between’ the iso-elastic family and the negative exponential family.

Once again we use our running example of determining the optimal portfolio with our usual investment opportunity, using the same notation as before. \( f(\alpha) = 0.5U[(1 - 0.1\alpha)w_0] + 0.5U[(1 + 0.2\alpha)w_0] \)

\[
\Rightarrow f'(\alpha) = -0.05w_0U'[(1 - 0.1\alpha)w_0] + 0.1w_0U'[(1 + 0.2\alpha)w_0] = 0
\]

\[
\Rightarrow 0 = U'[(1 - 0.1\alpha)w_0] = 2U'[(1 + 0.2\alpha)w_0] \Rightarrow w_0^{-a}(1 - 0.1\alpha)^{-a}e^{b(1 - 0.1\alpha)w_0} = w_0^{-a}(1 + 0.2\alpha)^{-a}e^{b(1 + 0.2\alpha)w_0}
\]

\[
\Rightarrow e^{-b(1 + 0.2\alpha)w_0} = 2\left(\frac{1 + 0.2\alpha}{1 - 0.1\alpha}\right)^{-a} \Rightarrow w_0 = \frac{1}{0.3b\alpha} \left\{ \ln 2 - a\ln \left(\frac{1 + 0.2\alpha}{1 - 0.1\alpha}\right) \right\}
\]

Figure 7: Example 4 - Variable Risk Aversion

As an example we use \( a = 2.408 \) and \( b = 0.000003. \) The graph in Figure 7 displays the function above, with \( \alpha \) ranging from 0% to 100% displayed on the y axis and initial wealth \( w_0 \) displayed using logarithmic scaling on the x axis.

With the values we chose for \( a \) and \( b \) this investor has low relative risk aversion of 2.438 at wealth $1,000, increasing linearly to 32.408 at wealth $10,000,000. Up to around $10,000 the investor is quite aggressive, placing nearly 100% of his portfolio in the risky asset. After his portfolio has grown in
value to $10,000 he gradually becomes more conservative, with 88% risky at $100,000, 43% risky at $1,000,000, and only 8% risky at $10,000,000.

This investor clearly becomes increasingly concerned about preserving his wealth as his wealth increases. This change is gradual, however, unlike the negative exponential utility functions. We do not mean to imply that all or even most investors have this characteristic. This is a totally contrived example. Indeed, the same technique can be used to generate any desired pattern of risk aversion as a function of wealth.

**Utility and Real-Life Investors**

What kind of utility function is the best kind to use? What kind of utility function is typical of most investors? What kind best describes the mythical “average” investor? These, of course, are the $64,000 questions. There do not appear to be definitive answers. Indeed, it seems that the questions are quite controversial.

We have taken a detailed look at several kinds of utility functions, including the iso-elastic family with constant relative risk aversion, the negative exponential family with constant absolute risk aversion, and an example in between with increasing relative and decreasing absolute risk aversion.

Without some kind of evidence there is no reason to believe that any of these particular utility functions which we have examined describes all investors or even any individual investor or the average investor. It is entirely reasonable for an investor's attitudes towards risk to vary with the amount of wealth the investor has accumulated, and it's reasonable for different investors to have different patterns of risk aversion as functions of wealth.

William Sharpe says the following about the notions of constant relative risk aversion and constant absolute risk aversion in reference. The assumption of constant relative risk aversion seems much closer to the preferences of most investors than does that of constant absolute risk aversion. Nonetheless, it is by no means guaranteed to reflect every Investor's attitude. Some may wish to take on more risk ... as their wealth increases. Others may wish to take on less.

Many analysts counsel a decrease in such risk as one ages. Some strategies are based on acceptance of more or less risk, based on economic conditions. And so on.

For these and other reasons it is important to at least consider strategies in which an Investor's risk tolerance... changes from time to time. However, such changes, if required at all, will likely be far more gradual than those associated with a constant risk tolerance expressed in terms of end-of-period value.

**APPENDICES**

In 1728, Nicholas Bernoulli clearly demonstrated a weakness of using expected value as the sole measure of preferences with his “St. Petersberg Paradox” and in 1738, Daniel Bernoulli, a cousin of Nicholas, provided an explanation for the St. Petersberg Paradox by stating that people cared about the expected “utility” of a gamble’s payoff, not the expected value of its payoff. As an individual’s wealth increases, the “utility” that one receives from the additional increase in wealth grows less than proportionally. In the St. Petersberg Paradox, prizes go up at the same rate that the probabilities decline. In order to obtain a finite valuation, the trick would be to allow the “value” or “utility” of prizes to increase slower than the rate probabilities decline.

---

Proverbs 21:5 The plans of the diligent lead to profit as surely as haste leads to poverty
The St. Petersburg Paradox
Consider the following game. A coin is tossed until a head appears. If the first head appears on the $j^{th}$ toss, then the payoff is £$2^j$. How much should you pay to play this game?

Let $S_j$ denote the event that the first head appears on the $j^{th}$ toss. Then $p(S_j) = 1/2^j$ and the payoff for $S_j$ is $V_j = 2^j$. \[ \text{EMV}(A) = \sum_{j=1}^{\infty} V_j p(S_j) = \sum_{j=1}^{\infty} 2^j \left( \frac{1}{2^j} \right) = \infty \]

The expected monetary payoff is infinite. However much you pay to play the game, you may expect to win more. Would you risk everything that you possess to play this game? One would suppose that real-world people would not be willing to risk an infinite amount to play this game.

BAYESIAN ANALYSIS

Review on Approaches to Probability
There are 3 approaches to probability, namely classical, empirical and subjective probability. The underlying idea behind the Classical or theoretical probability is symmetry. Empirical (or frequuntist) probability is based on observed data. The empirical probability of an event $A$ is the relative frequency of event $A$, that is
\[ p(A) = \frac{\text{Frequency of event } A}{\text{Total number of observations}} \]

Subjective probabilities result from intuition, educated guesses, and estimates. For example: given a patient's health and extent of injuries a doctor may feel that the patient has a 90% chance of a full recovery.

Subjectivity means two people can assign different probabilities to the same event.

Review on Conditional Probability and Independence
Humans often have to act based on incomplete information. If your boss has looked at you gloomily, you might conclude that something's wrong with your job performance. However, if you know that she just suffered some losses in the stock market, this extra information may change your assessment of the situation. Conditional probability is a tool for dealing with additional information like this. Conditional probability is the probability of an event occurring given the knowledge that another event has occurred. The conditional probability of event $A$ occurring, given that event $B$ has occurred is denoted by $P(A/B)$ and is read “probability of $A$ given $B$” and is given by
\[ P(A/B) = \frac{P(A \cap B)}{P(B)} \quad \text{provided } P(B) > 0 \]
Similarly $P(B/A) = \frac{P(A \cap B)}{P(A)}$ provided $P(A) > 0$.

\[ \Rightarrow P(A \cap B) = P(B \cap A) = P(A) \times P(B) = P(B) \times P(A) \]

Remark: Another way to express independence is to say that the knowledge of $B$ occurring does not change our assessment of $P(A)$. This means that if $A$ and $B$ are independent then $P(A \cap B) = P(A)$ and $P(B \cap A) = P(B)$.

Bayes’ Rule
Events $B_1, B_2, \ldots, B_k$ are said to be a partition of the sample space $S$ if the following two conditions are satisfied. i) $B_i \cap B_j = \emptyset$ for each pair $i, j$ and ii) $B_1 \cup B_2 \cup \ldots \cup B_k = S$

This situation often arises when the statistics are available in subgroups of a population. For example, an insurance company might know accident rates for each age group $B_j$. This will give the company conditional probabilities $P(A/B_j)$ (if we denote $A = \{\text{event of accident}\}$).
Question: If we know the conditional probabilities \( P(A/B_i) \), how do we find the unconditional \( P(A) \)?
Consider a case when \( k = 2 \):
The event \( A \) can be written as the union of mutually exclusive events \( AB_1 \) and \( AB_2 \), that is
\[
A = AB_1 \cup AB_2 \text{, it follows that } P(A) = P(AB_1) + P(AB_2)
\]
If the conditional probabilities \( P(A/B_1) \) and \( P(A/B_2) \) are known, that is
\[
P(AB_1) = \frac{P(AB_1)}{P(B_1)} \quad \text{and} \quad P(AB_2) = \frac{P(AB_2)}{P(B_2)} \text{ then } P(A) = P(B_1) \times P(AB_1) + P(B_2) \times P(AB_2)
\]
Suppose we want to find probability of the form \( P(B/A) \), which can be written as
\[
P(B/A) = \frac{P(AB)}{P(A)} = \frac{P(AB_1) \times P(A/B_1)}{P(A)} = \frac{P(B_1) \times P(AB_1)}{P(B_1) \times P(A/B_1) + P(B_2) \times P(A/B_2)}
\]
This calculation generalizes to \( k > 2 \) events as follows.

**Theorem** If \( B_1, B_2, \ldots, B_k \) form a partition of the sample space \( S \) such that \( P(B_i) \neq 0 \) for \( i = 1, 2, \ldots k \); then for any event \( A \subseteq S \),
\[
P(A) = \sum_{i=1}^{k} P(A|B_i) \sum_{i=1}^{k} P(B_i) \times P(A|B_i)
\]
Subsequently, \( P(B_1/A) \) equals
\[
P(B_1/A) = \frac{P(AB_1)}{P(A)} = \frac{P(B_1) \times P(AB_1)}{P(B_1) \times P(A/B_1) + P(B_2) \times P(A/B_2)}
\]
This last equation is often called Law of Total Probability.

**Example 1** A rare genetic disease (occurring in 1 out of 1000 people) is diagnosed using a DNA screening test. The test has false positive rate of 0.5%, meaning that \( P(\text{test positive/ no disease}) = 0.005 \). Given that a person has tested positive, what is the probability that this person actually has the disease? First, guess the answer, then read on.

**Solution**
Let’s reason in terms of actual numbers of people, for a change. Imagine 1000 people, 1 of them having the disease. How many out of 1000 will test positive? One that actually has the disease, and about 5 disease-free people who would test false positive. Thus,
\[
P(\text{disease/test positive}) \approx \frac{1}{6}
\]
It is left as an exercise for the reader to write down the formal probability calculation.

**Example 2** At a certain assembly plant, three machines make 30%, 45%, and 25%, respectively, of the products. It is known from the past experience that 2%, 3% and 2% of the products made by each machine, respectively, are defective. Now, suppose that a finished product is randomly selected.

a) What is the probability that it is defective?
b) If a product were chosen randomly and found to be defective, what is the probability that it was made by machine 3?

**Solution** Consider the following events:
A: the product is defective and \( B_i \); the product is made by machine \( i=1, 2, 3 \),
Applying additive and multiplicative rules, we can write
(a) \( P(A) = P(B_1) \times P(A/B_1) + P(B_2) \times P(A/B_2) + P(B_3) \times P(A/B_3) \)
\[
= (0.3)(0.02) + (0.45)(0.03) + (0.25)(0.02) = 0.006 + 0.0135 + 0.005 = 0.0245
\]
(b) Using Bayes’ rule \( P(B_3/A) = \frac{P(B_3) \times P(A/B_3)}{P(A)} = \frac{0.005}{0.0245} = 0.2041
\]
Bayesian Estimation

Introduction and Notations

The subject of statistics concerns itself with using data to make inferences and predictions about the world. The goal of statistical inference is to learn \( \theta \) using the data. Three major types of inference: point estimator (educated guess’), confidence interval and hypothesis testing. Given data \( X \), we denote the estimator of \( \theta \) by \( \hat{\theta}(X) \). It is a function of data.

Statistical decision theory is concerned with the problem of making decisions, in the presence of statistical knowledge which sheds light on the uncertainties involved in the problem. The uncertainties are presented by \( \theta \) (scalar, vector, or matrix)

The majority of statistics you have learned so are called classical or Frequentist. The probability for an event is defined as the proportion of successes in an infinite number of repeatable trials. By contrast, in Subjective Bayesian inference, probability is a measure of the strength of belief. We treat parameters as random variables. Before collecting any data we assume there is uncertainty about the value of a parameter. This uncertainty can be formalised by specifying a pdf (or pmf) for the parameter. We then conduct an experiment to collect some data that will give us information about the parameter. We then use Bayes Theorem to combine our prior beliefs with the data to derive an updated estimate of our uncertainty about the parameter.

The history of Bayesian Statistics

Bayesian methods originated with Bayes and Laplace (late 1700s to mid 1800s). In the early 1920’s, Fisher put forward an opposing viewpoint, that statistical inference must be based entirely on probabilities with direct experimental interpretation i.e. the repeated sampling principle.

In 1939 Jeffrey’s book ’The theory of probability’ started a resurgence of interest in Bayesian inference. This continued throughout the 1950-60s, especially as problems with the Frequentist approach started to emerge. The development of simulation based inference has transformed Bayesian statistics in the last 20-30 years and it now plays a prominent part in modern statistics.

Problems with Frequentist Inference

1) Frequentist Inference generally does not condition on the observed data
   A confidence interval is a set-valued function \( C(X) \subseteq \Theta \) of the data \( X \) which covers the parameter \( \theta \in C(X) \) a fraction \( 1-\alpha \) of repeated draws of \( X \) taken under the null \( H_0 \).
   This is not the same as the statement that, given data \( X = x \), the interval \( C(x) \) covers \( \theta \) with probability \( 1-\alpha \). But this is the type of statement we might wish to make.
   **Example**
   Suppose \( X_1, X_2 \sim U(\theta - \frac{1}{2}, \theta + \frac{1}{2}) \) so that \( X_{(1)} \) and \( X_{(2)} \) are the order statistics. Then \( C(X) = [X_{(1)}, X_{(2)}] \) is \( \alpha = \frac{1}{8} \) level CI for \( \theta \). Suppose in your data \( X = x \), \( X_{(2)} - X_{(1)} > \frac{1}{2} \), (this happens in an eighth of data sets), then \( \theta \in [X_{(1)}, X_{(2)}] \) with probability one.

2) Frequentist Inference depends on data that were never observed
   **The likelihood principle** Suppose that two experiments relating to \( \theta \), \( E_1, E_2 \) give rise to data \( y_1, y_2 \) such that the corresponding likelihoods are proportional, that is, for all \( \theta \)
   \[ L(\theta, y_1, E_1) = cL(\theta, y_2, E_2) \]
then the two experiments lead to identical conclusions about $\theta$. Key point, MLE’s respect the likelihood principle. i.e. the MLEs for $\theta$ are identical in both experiments. But significance tests do not respect the likelihood principle.

Consider a pure test for significance where we specify just H0. We must choose a test statistic $T(x)$, and define the p-value for data $T(x) = t$ as

$$p\text{-value} = P(T(X) \text{ at least as extreme as } t/H0):$$

The choice of $T(X)$ amounts to a statement about the direction of likely departures from the null, which requires some consideration of alternative models.

**Note**

(i) The calculation of the p-value involves a sum (or integral) over data that was not observed, and this can depend upon the form of the experiment.

(ii) A p-value is not $P(H_0/T(X) = t)$.

**Example:** (1975 British national referendum on whether the UK should remain part of the European Economic Community). Suppose 52% of voters supported the Labour Party and 48% the Conservative Party. Suppose 55% of Labour voters wanted the UK to remain part of the EEC and 85% of Conservative voters wanted this. What is the probability that a person voting “Yes” to remaining in EEC is a Labour voter?

$$p(L/Y) = \frac{p(Y/L)p(L)}{p(Y)} = \frac{p(Y/L)p(L)}{p(Y/L)p(L) + p(Y/L')p(L')} = \frac{(0.55)(0.52)}{(0.55)(0.52) + (0.85)(0.48)} = 0.41$$

Since $p(Y) = p(Y/L)p(L) + p(Y/L')p(L')$

**Bayes’ Rule**

Let D represent some observed data and let A, B, and C be mutually exclusive (and exhaustive) events conditional on D.

Note that $p(D) = p(DA) + p(DB) + p(DC) = p(D/A)p(A) + p(D/B)p(B) + p(D/C)p(C)$

By Bayes’ Law,

$$p(A/D) = \frac{p(D/A)p(A)}{p(D)} = \frac{p(D/A)p(A)}{p(D/A)p(A) + p(D/B)p(B) + p(D/C)p(C)}$$

Denoting A, B, C by $\theta_1$, $\theta_2$, $\theta_3$ we can write this more generally as

$$p(\theta_i/D) = \frac{p(D/\theta_i)p(\theta_i)}{\sum_{i=1}^{k} p(D/\theta_i)p(\theta_i)}$$

If there are k distinct discrete outcomes $\theta_1$, $\theta_2$, $\ldots$, $\theta_k$, we have, for any $\square = \{1,\ldots, k\}$:

$$p(\theta_i/D) = \frac{p(D/\theta_i)p(\theta_i)}{\sum_{i=1}^{k} p(D/\theta_i)p(\theta_i)}$$

The denominator equals P(D), the marginal distribution of the data.

**Note** if the values of $\theta$ are portions of the continuous real line, the sum may be replaced by an integral.

**Relationship between the Likelihood, Prior and the Posterior**

Bayesian statistical inference uses Bayes’ Law (Bayes’Theorem) to combine prior information and sample data to make conclusions about a parameter of interest. Generically denote an unobserved parameter of interest as $\theta$ and the observed data as x. Now given data, our goal then becomes to determine which probability distribution generated the data.
Let \( x = \{x_1, x_2, \ldots, x_n\} \) denote the full set of data. Thus \( x \) is a random variable, whose probability density function would in probability theory typically be denoted as \( f_x(\{x_1, x_2, \ldots, x_n\}) \). We will use a standard (in Bayesian analysis) shorthand notation for probability density functions, and denote the probability density function of the random variable \( x \) as simply \( p(x) \).

We will assume that the data were generated from a probability distribution that is described by some parameters \( \theta \) (not necessarily scalar). We treat \( \theta \) as a random variable. We will use the shorthand notation \( f(x|\theta) \) to represent the family of conditional density functions over \( x \), parameterized by the random variable \( \theta \). We call this family \( f(x|\theta) \) a likelihood function for the data \( x \), as it tells us how likely the data \( x \) are given the model specified by any value of \( \theta \).

We specify a prior distribution over \( \theta \), denoted \( \pi(\theta) \). This distribution represents any knowledge we have about how the data are generated prior to observing them.

Our end goal is the conditional density function over \( \theta \), given the observed data, which we denote as \( \pi(\theta|x) \). We call this the posterior distribution, and it informs us which parameters are likely given the observed data.

We, the modeler, specify the likelihood function (as a function of \( x \) and \( \theta \)) and the prior (we completely specify this) using our knowledge of the system at hand. We then use these quantities, together with the data, to compute the posterior. The likelihood, prior, and posterior are all related via Bayes’ rule:

\[
\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{p(x)} = \frac{\int f(x|\theta)\pi(\theta)d\theta}{p(x)} \Rightarrow \pi(\theta|x) \propto f(x|\theta)\pi(\theta)
\]

where the second step uses the law of total probability.

We call \( \int \theta f(x|\theta)\pi(\theta)d\theta \) the marginal likelihood or the normalizing constant and it’s often intractable. This is what makes Bayesian analysis difficult.

Note \( p(x) \) does not depend on \( \theta \) and thus carries no information about \( \theta \). It is simply a normalizing constant which makes \( \pi(\theta|x) \) sum (or integrate) to 1.

\( \pi(\theta|x) \) represents a compromise between the prior information about \( \theta \) in \( \pi(\theta) \) and the information from the sample about \( \theta \) in \( f(x|\theta) \).

Some useful summaries of the posterior are the posterior Mean and the posterior variance ie

\[
E(\theta|x) = \int \theta \pi(\theta|x)d\theta \quad \text{and} \quad \text{var}(\theta|x) = \left[ E[\theta^2] - E[(\theta|x)]^2 \right] = \int \theta^2 p(\theta|x)d\theta - E[(\theta|x)]^2
\]

If the values of \( \theta \) are discrete, sums would replace the integrals.
Likelihood Function
The likelihood function $L(\theta|x)$ is a function of $\theta$ that shows how “likely” are various parameter values $\theta$ to have produced the data $X$ that were observed. In classical statistics, the specific value of $\theta$ that maximizes $L(\theta|x)$ is the maximum likelihood estimator (MLE) of $\theta$.

In many common probability models, when the sample size $n$ is large, $L(\theta|x)$ is unimodal in $\theta$.

**Definition:** Let $X_1, X_2, ..., X_n$ have a joint density function $f(X_1, X_2, ..., X_n / \theta)$. Given $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$ is observed, the function of $\theta$ defined by:

$$L(\theta|x) = L(\theta|x_1, x_2, ..., x_n) = f(x_1, x_2, ..., x_n / \theta)$$

Mathematically, if $X_1, X_2, ..., X_n$ are independently and identically distributed as $X_i \sim f(x_i / \theta)$, then

$$L(\theta|x) = \prod_{i=1}^{n} f(x_i / \theta) \quad (\text{where } x_1, x_2, ..., x_n \text{ are the } n \text{ data vectors}).$$

The Likelihood Principle of Birnbaum states that (given the data) all of the evidence about $\theta$ is contained in the likelihood function. Often it is more convenient to use the log likelihood

$$l(\theta|x) = \ln[L(\theta|x)].$$

This definition almost seems to be defining the likelihood function to be the same as the pdf or pmf. The only distinction between these two functions is which variable is considered fixed and which is varying. When we consider the pdf or pmf $f(x / \theta)$, we are considering $\theta$ as fixed and $x$ as the variable; when we consider the likelihood function $L(\theta|x)$, we are considering $x$ to be the observed sample point and $\theta$ to be varying over all possible parameter values.

**Remarks**

- The likelihood function is not a probability density function.
- It is an important component of both frequentist and Bayesian analyses
- It measures the support provided by the data for each possible value of the parameter. If we compare the likelihood function at two parameter points and find that $L(\theta_1|x) > L(\theta_2|x)$, then the sample we actually observed is more likely to have occurred if $\theta = \theta_1$ than if $\theta = \theta_2$ which can be interpreted as saying that $\theta_1$ is a more plausible value for the true value of $\theta$ than is $\theta_2$. We carefully use the word “plausible” rather than “probable” because we often think of $\theta$ as a fixed value.

**Example:** Normal distribution. Assume that $x_1, x_2, ..., x_n$ is a random sample from $N(\mu, \sigma^2)$, where both $\mu$ and $\sigma^2$ are unknown parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. With $\theta = \{\mu, \sigma^2\}$, the likelihood is

$$L(\theta|x) = \prod_{i=1}^{n} \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left\{\frac{x_i - \mu}{\sigma}\right\}^2\right\} = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\right\}$$

and the log-likelihood is

$$l(\theta|x) = -\frac{n}{2} \ln(2\pi) - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

**Example:** Poisson distribution. Assume that $x_1, \ldots, x_n$ is a random sample from Poisson($\theta$), with unknown $\theta > 0$; then the likelihood is

$$L(\theta|x) = \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!} = \frac{\theta^{\sum_{i=1}^{n} x_i} e^{-n\theta}}{\prod_{i=1}^{n} x_i!}$$
and the log-likelihood is \( l(\theta / x) = -n\theta + (\ln \theta) \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln(x_i!) \)

**Example:** M&M’s sold in the United States have 50% red candies compared to 30% in those sold in Canada. In an experimental study, a sample of 5 candies was drawn from an unlabelled bag and 2 red candies were observed. Is it more plausible that this bag was from the United States or from Canada?

**Solution**

The likelihood function is: \( L(\theta / x) = \theta^2(1-\theta)^3 \), \( \theta = 0.3 \) or 0.5

\( L(0.3 / x) = 0.03087 < 0.03125 = L(0.5 / x) \) suggesting that it is more plausible that the bag used in the experiment was from the United States.

**Likelihood Principle**

If \( x \) and \( y \) are two sample points such that \( L(\theta / x) \propto L(\theta / y) \ \forall \theta \) then the conclusions drawn from \( x \) and \( y \) should be identical. Thus the likelihood principle implies that likelihood function can be used to compare the plausibility of various parameter values. For example, if \( L(\theta_2 / x) = 2L(\theta_1 / x) \) and \( L(\theta / x) \propto L(\theta / y) \ \forall \theta \), then \( L(\theta_2 / y) = 2L(\theta_1 / y) \). Therefore, whether we observed \( x \) or \( y \) we would come to the conclusion that \( \theta_2 \) is twice as plausible as \( \theta_1 \).

**Example:** Consider the distribution Multinomial(\( n = 6, \theta, \theta, 1-2\theta \)). The following two samples drawn from this distribution have the same likelihood:

- \( X = (1, 3, 2) \Rightarrow \frac{6!}{1!3!2!} \theta^3(1-2\theta)^2 \) and
- \( X = (1, 3, 2) \Rightarrow \frac{6!}{2!2!2!} \theta^2(1-2\theta)^2 \)

This means both samples would lead us to the same conclusion regarding the relative plausibility of different values of \( \theta \).

**The Bayesian Framework**

Suppose we observe an iid sample of data \( x = (x_1, x_2, \ldots, x_n) \). Now \( x \) is considered fixed and known. We also must specify \( \pi(\theta) \), the prior distribution for \( \theta \), based on any knowledge we have about \( \theta \) before observing the data.

Our model for the distribution of the data will give us the likelihood

\[ L(\theta / x) = \prod_{i=1}^{n} p(x_i / \theta) \]

Then by Bayes’ Law, our posterior distribution

\[ \pi(\theta / x) = \frac{\pi(\theta)L(\theta / x)}{p(x)} = \frac{\pi(\theta)L(\theta / x)}{\int_0^1 \pi(\theta)L(\theta / x) d\theta} \]

Note that the marginal distribution of \( X \), \( p(x) \), is simply the joint density \( p(\theta, x) \) (i.e., the numerator) with \( \theta \) integrated out. With respect to \( \theta \), it is simply a normalizing constant. Therefore \( \pi(\theta / x) \propto \pi(\theta)L(\theta / x) \)

Often we can calculate the posterior distribution by multiplying the prior by the likelihood and then normalizing the posterior at the last step, by including the necessary constant.
Summarizing the Posterior Distribution
How do we extract a Bayes estimator for some unknown parameter $\theta$? If our mindset is to use some sort of point estimator (as is usually done in classical statistics), there are a number of candidates. We could follow maximum likelihood and use the mode of the distribution (its maximal value), with $\hat{\theta} = \text{Max } \{\pi(\theta / x)\}$

We could take the expected value of $\theta$ given the posterior, $\hat{\theta} = E_{\theta}[\theta / x] = \int_0^\infty \theta \pi(\theta / x) d\theta$

Another candidate is the medium of the posterior distribution, where the estimator satisfies $Pr(\theta < \hat{\theta} / x) = Pr(\theta > \hat{\theta} / x) = 0.5$ hence $\int_0^\hat{\theta} \pi(\theta / x) d\theta = \int_{\hat{\theta}}^\infty \pi(\theta / x) d\theta = 0.5$

However, using any of the above estimators, or even all three simultaneously, loses the full power of a Bayesian analysis, as the full estimator is the entire posterior density itself. If we cannot obtain the full form of the posterior distribution, it may still be possible to obtain one of the three above estimators. However, as we will see later, we can generally obtain the posterior by simulation using Gibbs sampling, and hence the Bayes estimate of a parameter is frequently presented as a frequency histogram from (Gibbs) samples of the posterior distribution.

Posterior Intervals
The ideal summary of $\theta$ is an interval (or region) with a certain probability of containing $\theta$. Note that a classical (frequentist) confidence interval does not exactly have this interpretation.

Bayesian Credible Intervals
A credible interval (or in general, a credible set) is the Bayesian analogue of a confidence interval. A $100(1-\alpha)$% credible set $C$ is a subset of $\Theta$ such that $\int_C \pi(\theta / x) d\theta = 1-\alpha$

If the parameter space $\Theta$ is discrete, a sum replaces the integral.

Quantile-Based Intervals
If $\theta_L$ is the $\alpha$ posterior quantile for $\theta$, and $\theta_U$ is the $1-\alpha$ posterior quantile for $\theta$, then $(\theta_L, \theta_U)$ is a $100(1-\alpha)$% credible interval for $\theta$.

Note $p(\theta \leq \theta_L / x) = \%$ and $p(\theta \geq \theta_U / x) = \%$

$\Rightarrow p[\theta \in (\theta_L, \theta_U) / x] = 1 - p[\theta \not\in (\theta_L, \theta_U) / x] = 1 - [p(\theta \leq \theta_L / x) + p(\theta \geq \theta_U / x)] = 1-\alpha$

Example 1: Quantile-Based Interval
Suppose $X_1, \ldots, X_n$ are the durations of cabinets for a sample of cabinets from Western European countries. We assume the $X_i$'s follow an exponential distribution

$p(x / \theta) = \theta e^{-\theta x}, \quad x_i \geq 0 \quad \Rightarrow \quad L(\theta / x) = \theta^n e^{-\theta \sum x_i}$

Suppose our prior distribution for $\theta$ is $p(\theta) \propto \frac{1}{\theta}, \quad \theta > 0$) Larger values of $\theta$ are less likely a priori

Then

$\pi(\theta / x) \propto p(\theta) L(\theta / x) \propto \frac{1}{\theta^n} \times \theta^n e^{-\theta \sum x_i} = \theta^{n-1} e^{-\theta \sum x_i}$

This is the kernel of a gamma distribution with “shape” parameter $n$ and “rate” parameter $\sum x_i$

So including the normalizing constant,
Now, given the observed data $x_1, \ldots, x_n$, we can calculate any quantiles of this gamma distribution. The 0.05 and 0.95 quantiles will give us a 90\% credible interval for $\theta$.

Suppose we feel $\pi(\theta) = \frac{1}{\theta}$, $\theta > 0$ is too subjective and favors small values of $\theta$ too much. Instead, let’s consider the non-informative prior $\pi(\theta) = 1$, $\theta > 0$ (favors all values of $\theta$ equally).

Then our posterior is $\pi(\theta / x) \propto p(\theta) L(\theta / x) = 1 \times \theta^n e^{-\theta \sum x_i} = \theta^{(n+1)-1} e^{-\theta \sum x_i}$. This posterior is a gamma with parameters $(n + 1)$ and $\sum_{i=1}^{n} x_i$

We can similarly find the equal-tail credible interval.

Example 2: Quantile-Based Interval
Consider 10 flips of a coin having $P\{\text{Heads}\} = \theta$. Suppose we observe 2 “heads”.

We model the count of heads as binomial

$$p(x/\theta) = \binom{10}{x} \theta^x (1-\theta)^{10-x}, \ x = 0, 1, 2, \ldots, 10$$

Let’s use a uniform prior for $\theta$ ie $p(\theta) = 1$, $0 \leq \theta \leq 1$.

Then the posterior is: $\pi(\theta / x) \propto p(\theta) L(\theta / x) = \binom{10}{x} \theta^x (1-\theta)^{10-x} \propto \theta^x (1-\theta)^{10-x}, \ 0 \leq \theta \leq 1$

This is a beta distribution for $\theta$ with parameters $x + 1$ and $10 - x + 1$. Since $x = 2$ here, $\pi(\theta / X = 2) = \text{beta}(3,9)$. The 0.025 and 0.975 quantiles of a beta(3,9) are (.0602, .5178), which is a 95\% credible interval for $\theta$.

HPD Intervals / Regions
The equal-tail credible interval approach is ideal when the posterior distribution is symmetric. 
Definition: A 100(1-\alpha)\% HPD region for $\theta$ is a subset $C \in \Theta$ defined By $C = \{\theta : \pi(\theta / x) \geq k\}$ where $k$ is the largest number such that $\int_C \pi(\theta / x) d\theta = 1 - \alpha$

The value $k$ can be thought of as a horizontal line placed over the posterior density whose intersection(s) with the posterior define regions with probability $1 - \alpha$

Remarks
- The HPD region will be an interval when the posterior is unimodal.
- If the posterior is multimodal, the HPD region might be a discontiguous set.

Bayesian Hypothesis Testing
Before we go into the details of Bayesian hypothesis testing, let us briefly review frequentist hypothesis testing. Recall that in the Neyman-Pearson paradigm characteristic of frequentist hypothesis testing, there is an asymmetric relationship between two hypotheses: the null hypothesis H0 and the alternative hypothesis H1. A decision procedure is devised by which, on the basis of a set of collected data, the null hypothesis will either be rejected in favour of H1, or accepted.
In Bayesian hypothesis testing, there can be more than two hypotheses under consideration, and they do not necessarily stand in an asymmetric relationship. Rather, Bayesian hypothesis testing works just like any other type of Bayesian inference. Let us consider the case where we are comparing only two hypotheses: Then the Bayesian hypothesis testing can be done as follows.

Suppose that \( \Box \in \Box \) and we want to test

\[
H_0: \theta = \theta_0 \ vs \ H_1: \theta \neq \theta_0.
\]

If we really believe that there is a positive prior probability that \( H_0 \) is true then we can use a prior of the form \( a \delta_{\theta_0} + (1-a)g(\theta) \) where \( 0 < a < 1 \) is the prior probability that \( H_0 \) is true and \( g \) is a smooth prior density over \( \theta \) which represents our prior beliefs about \( \theta \) when \( H_0 \) is false. It follows from Bayes' theorem that

\[
P(\theta = \theta_0/X_1, X_2, ..., X_n) = \frac{aP(X_1, X_2, ..., X_n/\theta_0)}{aP(X_1, X_2, ..., X_n/\theta_0) + (1-a)\int P(X_1, X_2, ..., X_n/\theta)g(\theta)d\theta} = \frac{aL(\theta_0)}{aL(\theta_0) + (1-a)m}
\]

where \( m = \int P(X_1, X_2, ..., X_n/\theta)g(\theta)d\theta \). It can be shown that \( P(\theta = \theta_0/X_1, X_2, ..., X_n) \) is very sensitive to the choice of \( g \).

Sometimes, people like to summarize the test by using the Bayes factor \( B \) which is defined to be the posterior odds divided by the prior odds: Now

\[
\text{posterior odds} = \frac{P(\theta = \theta_0/X_1, X_2, ..., X_n)}{1 - P(\theta = \theta_0/X_1, X_2, ..., X_n)} = \left( \frac{aL(\theta_0)}{aL(\theta_0) + (1-a)m} \right) \frac{(1-a)m}{aL(\theta_0) + (1-a)m} = \frac{aL(\theta_0)}{(1-a)m}
\]

and the Bayes factor is \( B = \frac{\text{posterior odds}}{\text{prior odds}} = \left( \frac{aL(\theta_0)}{(1-a)m} \right) \frac{a}{1-a} = \frac{L(\theta_0)}{m} \).

**Example**

Let \( X_i \)'s be iid RV's with \( X_i \sim N(\theta, 1) \). Let's test \( H_0: \theta = 0 \ vs \ H_1: \theta \neq 0 \). Suppose we take \( g(\theta) \) to be \( N(0; 1) \). Thus \( g(\theta) = \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2} \theta^2 \). Let us further take \( a = \frac{1}{2} \). Then, after some tedious integration to compute \( m(X_1, X_2, ..., X_n) \), we get

\[
P(\theta = \theta_0/X_1, X_2, ..., X_n) = \frac{L(\theta_0)}{L(\theta_0) + m} = \frac{\exp \left\{-\frac{1}{2}n\overline{x}^2\right\}}{\exp \left\{-\frac{1}{2}n\overline{x}^2\right\} + \sqrt{\frac{\pi}{2n+1}} \exp \left\{-\frac{2n+1}{2n+1}\overline{x}^2\right\}}
\]

On the other hand, the p-value for the usual test is \( p = 2\Phi(-\sqrt{n}\overline{x}) \). The figure below shows the posterior of \( H_0 \) and the p-value as a function of \( X \) when \( n = 100 \). Note that they are very different. Unlike in estimation, in testing there is little agreement between Bayes and frequentist methods.

**Conclusion**

Bayesian and frequentist inference are answering two different questions. Frequentist inference answers the question: How do I construct a procedure that has frequency guarantees? Bayesian inference answers the question: How do I update my subjective beliefs after I observe some data? In parametric models, if \( n \) is large and the dimension of the model is fixed, Bayes and frequentist procedures will be similar. Otherwise, they can be quite different.
Conjugate Priors
In the Bayesian setting it is important to compute posterior distributions. This is not always an easy task. The main difficulty is to compute the normalizing constant in the denominator of Bayes theorem. The appropriate likelihood function (Binomial, Gaussian, Poisson, Bernoulli,...) is typically clear from the data, but there is a great deal of flexibility when choosing the prior distribution. However, for certain parametric families there are convenient choices of prior distributions. Particularly convenient is when the posterior belongs to the same family of distributions as the prior. Such families are called conjugate families.

Definition (Conjugate Priors) a prior $\pi(\theta)$ for a sampling model is called a conjugate prior if the resulting posterior $\pi(\theta|x)$ is in the same distributional family as the prior. For example, in Beta Prior and binomial likelihood $\Rightarrow$ Posterior is beta (with different parameter values!)

Remarks
- The parameters of the prior distribution are called prior hyperparameters. We choose them to best represent our beliefs about the distribution of $\theta$. The parameters of the posterior distribution are called posterior hyperparameters.
- Any time a likelihood model is used together with its conjugate prior, we know the posterior is from the same family of the prior, and moreover we have an explicit formula for the posterior hyperparameters. A table summarizing some of the useful conjugate prior relationships follows. There are many more conjugate prior relationships that are not shown in the following table but that can be found in reference books on Bayesian statistics.
We will now discuss a few of these conjugate prior relationships to try to gain additional insight.

**Complete Derivation of Beta/Binomial Bayesian Model**

Suppose we observe $X_1, X_2, \ldots, X_n$ which are iid Bernoulli($\theta$) r.v.’s and put $Y = \sum_{i=1}^{n} X_i$. Then

$Y \sim \text{Bin}(n, \theta)$. Let the prior distribution be Beta($\alpha, \beta$) which is a conjugate prior to the binomial likelihood. Then the posterior of $\theta$ given $Y = y$ is Beta($y + \alpha, n - y + \beta$).

We first write the joint density of $Y$ and $\theta$

$$
\pi(\theta, y) = f(y \mid \theta) \pi(\theta) = \left( \frac{n!}{y!(n-y)!} \right) \theta^y (1-\theta)^{n-y} \left[ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \right]
$$

$$
= \frac{\Gamma(n+1)}{\Gamma(y+1) \Gamma(n-y+1)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \theta^{\alpha+y-1} (1-\theta)^{n-y-\beta+1}
$$

Although it is not really necessary, let’s derive the marginal density of $Y$:

$$
p(y) = \int_{0}^{1} \pi(\theta, y) d\theta = \frac{\Gamma(n+1)}{\Gamma(y+1) \Gamma(n-y+1)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \int_{0}^{1} \theta^{\alpha+y-1} (1-\theta)^{n-y-\beta+1} d\theta
$$

$$
= \frac{\Gamma(n+1) \Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(y+1) \Gamma(n-y+1)} \times \frac{\Gamma(\alpha + y) \Gamma(n-y + \beta)}{\Gamma(n+\alpha + \beta)}
$$

Then the posterior $\pi(\theta \mid x)$ is

$$
\frac{\pi(\theta \mid x)}{p(y)} = \frac{\Gamma(n+1) \Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(y+1) \Gamma(n-y+1)} \theta^{\alpha+y-1} (1-\theta)^{n-y-\beta+1}
$$

$$
= \frac{\Gamma(n+\alpha + \beta)}{\Gamma(\alpha + y) \Gamma(n-y + \beta)} \theta^{\alpha+y-1} (1-\theta)^{n-y-\beta+1}, \quad 0 \leq \theta \leq 1
$$

Clearly this posterior is a Beta($\alpha + y, n - y + \beta$) distribution.

**Inference with Beta/Binomial Model**

As an interval estimate for $\theta$, we could use a (quantile-based or HPD) credible interval based on this posterior.

*Proverbs 21:5* The plans of the diligent lead to profit as surely as haste leads to poverty
As a point estimator of $\theta$, we could use:

i) The posterior mean $E(\theta/x)$ (the usual Bayes estimator)

ii) The posterior median

iii) The posterior mode

The mean of the (posterior) beta distribution is $E(\theta/x) = \frac{\alpha + y}{(\alpha + y) + (n - y + \beta)} = \frac{\alpha + y}{n + \alpha + \beta}$

Note $E(\theta/x) = \frac{y}{n + \alpha + \beta} + \frac{\alpha}{n + \alpha + \beta} = \frac{y}{n} \left( \frac{n}{n + \alpha + \beta} \right) + \frac{\alpha}{\alpha + \beta} \left( \frac{\alpha + \beta}{n + \alpha + \beta} \right)$

So the Bayes estimator $E(\theta/x)$ is a weighted average of the usual frequentist estimator (sample mean) and the prior mean. As $n$ increases, the sample data are weighted more heavily and the prior information less heavily. In general, with Bayesian estimation, as the sample size increases, the likelihood dominates the prior.

The Gamma/Poisson Bayesian Model

If our data $X_1, X_2, \ldots, X_n$ are iid Poisson($\lambda$), then a gamma $\text{gamma}(\alpha, \beta)$ prior on $\lambda$ is a conjugate prior.

Likelihood: $L(\lambda/x) = \prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda} \lambda^{\sum x_i}}{\prod (x_i!)}$ and prior $p(\lambda) = \frac{\beta^\alpha \lambda^{a-1} e^{-\beta \lambda}}{\Gamma(\alpha)}$, $\lambda > 0$. Thus posterior $\pi(\lambda/x) \propto \lambda^{a-1} \sum e^{-(n+\beta)\lambda}$, $\lambda > 0$ which is a gamma$(\alpha + \sum x_i + n + \beta)$ (Conjugate!)

The posterior mean is $E(\lambda/x) = \frac{\alpha + \sum x_i}{n + \beta} = \frac{\sum x_i}{n + \beta} + \frac{\alpha}{n + \beta} = \frac{n}{n + \beta} \left( \frac{\sum x_i}{n} \right) + \frac{\beta}{n + \beta} \left( \frac{\alpha}{\beta} \right)$

Again, the data get weighted more heavily as $n \to \infty$.

Bayesian Learning

We can use the Bayesian approach to update our information about the parameter(s) of interest sequentially as new data become available. Suppose we formulate a prior for our parameter $\theta$, and observe a random sample $x_1$. Then the posterior is $\pi(\theta/x) \propto p(\theta)L(\theta/x)$. Then we observe a new (independent) sample $y$.

We can use our previous posterior as the new prior and derive a new posterior:

$\pi(\theta/x,y) \propto \pi(\theta/x)L(\theta/y) \propto p(\theta)L(\theta/x)L(\theta/y) = p(\theta)L(\theta/x,y)$ (since $x, y$ independent)

Note this is the same posterior we would have obtained had $x$ and $y$ arrived at the same time! This “sequential updating” process can continue indefinitely in the Bayesian setup.

Loss and Risk Functions and Minimax Theory

Suppose we want to estimate a parameter $\theta$ using data $X = (X_1, X_2, \ldots, X_n)$. What is the best possible estimator $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n) \text{ for } \theta$? Minimax theory provides a framework for answering this question.

Loss Function

Let $\hat{\theta} = \hat{\theta}(X)$ be an estimator for the parameter $\theta$. We start with a loss function $L(\theta, \hat{\theta})$ that measures how good the estimator is. Effectively $L(\theta, \hat{\theta})$ is used to quantify the consequence that would be incurred for each possible decision for various possible values of $\theta$.

Examples of Loss Functions include;
i) squared error loss function $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$

ii) absolute error: $L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$

iii) $L_q$ loss: $L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|^q$

iv) Zero-one loss function $L(\theta, \hat{\theta}) = \begin{cases} 0 & \text{if } \theta = \hat{\theta} \\ 1 & \text{if } \theta \neq \hat{\theta} \end{cases}$

In general, we use a non-negative loss $L(\theta, \hat{\theta}) > 0$

**Risk Function**

Intuitively, we prefer decision rules with small “expected loss" resulting from the use of $\hat{\theta}(x)$ repeatedly with varying $x$. This leads to the risk function of a decision rule.

The risk function of an estimator $\hat{\theta}$ is

$$R(\theta, \hat{\theta}) = E[L(\theta, \hat{\theta})] = \begin{cases} \sum_{x \in X} L(\theta, \hat{\theta}) f(x/\theta) & \text{x is discrete} \\ \int_{x} L(\theta, \hat{\theta}) f(x/\theta) dx & \text{x is continuous} \end{cases}$$

where $X$ is the sample space (the set of possible outcomes) of $x$)

When the loss function is squared error, the risk is just the MSE (mean squared error):

$$R(\theta, \hat{\theta}) = E[(\theta - \hat{\theta})^2] = \text{var}(\hat{\theta}) + \text{bias}^2(\hat{\theta})$$

**Note** If the loss function is unspecified, assume the squared error loss function.

**Bias-Variance Decomposition of MSE**

Consider the squared loss function. The risk is known as the mean squared error (MSE) $\text{MSE} = E_o(\theta - \hat{\theta})^2$

We show that MSE has the following decomposition

$$\text{MSE} = E_o(\hat{\theta}(x) - \theta)^2 = E_o\{\hat{\theta}(x) - E_o[\hat{\theta}(x)] + E_o[\hat{\theta}(x)] - \theta\}^2$$

$$= E_o\{\hat{\theta}(x) - E_o[\hat{\theta}(x)]\}^2 + \{E_o[\hat{\theta}(x)] - \theta\}^2 = \text{Var}[\hat{\theta}(x)] + \text{Bias}^2[\hat{\theta}(x)]$$

This is known as bias-variance trade-offs.

**Risk Comparison**

How do we compare two estimators?

Given $\hat{\theta}_1(x)$ and $\hat{\theta}_2(x)$, if $R(\theta, \hat{\theta}_1) < R(\theta, \hat{\theta}_2)$ $\forall \theta \in \Theta$ then $\hat{\theta}_1(x)$ is the preferred estimator.

Ideally, we would like to use the decision rule $\hat{\theta}(x)$ which minimizes the risk $R(\theta, \hat{\theta})$ for all values of $\theta$.

However, this problem has no solution, as it is possible to reduce the risk at a specific $\theta_0$ to zero by making $\hat{\theta}(x)$ equal to $\theta_0$ for all $X$.

**Minimax rules**

A rule $\hat{\theta}$ is a minimax rule if $\max_\theta R(\theta, \hat{\theta}) < \max_\theta R(\theta, \hat{\theta}')$ for any other rule $\hat{\theta}'$. It minimizes the maximum risk. Sometimes this doesn’t produce a sensible choice of decision rule.

Since minimax minimizes the maximum risk (ie, the loss averaged over all possible data $X$) the choice of rule is not influenced by the actual data $X = x$ (though given the rule $\hat{\theta}$, the action $\hat{\theta}(x)$ is data-dependent).

It makes sense when the maximum loss scenario must be avoided, but can can lead to poor performance on average.
The minimax risk is  
\[ R_n = \inf_{\theta} \sup_{\theta} \left( R(\theta, \hat{\theta}) \right) \]
where the infimum is over all estimators. An estimator \( \hat{\theta} \) is a minimax estimator if  
\[ \sup_{\theta} \left( R(\theta, \hat{\theta}) \right) = \inf_{\theta} \sup_{\theta} \left( R(\theta, \hat{\theta}) \right) \]

**Bayes Rule and the Posterior Risks**

**Definition** Suppose we have a prior probability \( \pi(\theta) \) for \( \theta \). Denote  
\[ r(\theta, \hat{\theta}) = \int_{\theta} R(\theta, \hat{\theta}) \pi(\theta) d\theta \]
the Bayes risk of rule \( \hat{\theta} \). A Bayes rule is a rule that minimizes the Bayes risk. A Bayes rule is sometimes called a Bayes procedure.

Let  
\[ \pi(\theta / x) = \frac{L(\theta / x) \pi(\theta)}{p(x)} \]
denote the posterior following from likelihood \( L(\theta / x) \) L and prior \( \pi(\theta) \). The expected posterior loss (posterior risk) is defined as  
\[ \int_{\theta} L_s(\theta, \hat{\theta}) \pi(\theta / x) d\theta \]

**Lemma** A Bayes rule minimizes the expected posterior loss.

**Proof**  
\[
\int_{\theta} R(\theta, \hat{\theta}) \pi(\theta) d\theta = \int_{\theta} \int_{\theta} L_s(\theta, \hat{\theta}) L(\theta / x) \pi(\theta) d\theta dx = \int_{\theta} \int_{\theta} L_s(\theta, \hat{\theta}) \pi(\theta / x) p(x) d\theta dx
\]
That is for each \( x \) we choose \( \hat{\theta}(x) \) to minimize the integral  
\[ \int_{\theta} L_s(\theta, \hat{\theta}) \pi(\theta / x) d\theta \]
The form of the Bayes rule depends upon the loss function in the following way:
- Zero-one loss (as \( b \to \infty \)) leads to the posterior mode.
- Absolute error loss leads to the posterior median.
- Quadratic loss leads to the posterior mean.

**Note** These are not the only loss functions one could use in a given situation, and other loss functions will lead to different Bayes rules.

**Example** \( X \sim \text{Bin}(n, \theta) \), and the prior \( \pi(\theta) \) is a Beta(\( \alpha, \beta \)) distribution. The distribution is unimodal if \( \alpha, \beta > 1 \) with mode \( \frac{\alpha-1}{\alpha+\beta-2} \) and \( E(\theta) = \frac{\alpha}{\alpha+\beta} \). The posterior distribution of \( \theta / x \) is
\[ \text{Beta}(\alpha+x, n-x+\beta) \]  

With zero-one loss and \( b \to \infty \) the Bayes estimator is  
\[ \hat{\theta} = \frac{\alpha+x-1}{\alpha+\beta+n-2} \]

For a quadratic loss function, the Bayes estimator is  
\[ \hat{\theta} = \frac{\alpha+x}{\alpha+\beta+n} \]

For an absolute error loss function is the median of the posterior.

**Example 1** Let \( X_1, X_2, \ldots, X_n \sim N(\theta, 1) \). We will see that \( \bar{x} \) is minimax with respect to many different loss functions. The risk is \( 1/n \).

**Example 2** Let \( X_1, X_2, \ldots, X_n \) be a sample from a density \( f \). Let \( F \) be the class of smooth densities (defined more precisely later). We will see (later in the course) that the minimax risk for estimating \( f \) is \( Cn^{-\frac{1}{2}} \).
The Choice of a Prior

Obviously, a critical feature of any Bayesian analysis is the choice of a prior. The key here is that when the data have sufficient signal, even a bad prior will still not greatly influence the posterior. In a sense, this is an asymptotic property of Bayesian analysis in that all but pathological priors will be overcome by sufficient amounts of data. As mentioned above, one can check the impact of the prior by seeing how stable to posterior distribution is to different choices of priors. If the posterior is highly dependent on the prior, then the data (the likelihood function) may not contain sufficient information. However, if the posterior is relatively stable over a choice of priors, then the data indeed contain significant information.

The location of a parameter (mean or mode) and its precision (the reciprocal of the variance) of the prior is usually more critical than its actual shape in terms of conveying prior information. The shape (family) of the prior distribution is often chosen to facilitate calculation of the prior, especially through the use of conjugate priors that, for a given likelihood function, return a posterior in the same distribution family as the prior (i.e., a gamma prior returning a gamma posterior when the likelihood is Poisson). We will return to conjugate priors and the end of these notes, but we first discuss other standard approaches for construction of priors.

Diffuse Priors

One of the most common priors is the flat, uninformative, or diffuse prior where the prior is simply a constant,
\[ p(\theta) = k = \frac{1}{b-a} \text{ for } a \leq \theta \leq b \]
This conveys that we have no a priori reason to favor any particular parameter value over another. With a flat prior, the posterior is just a constant times the likelihood, \( p(\theta/x) = c \times l(\theta/x) \) and we typically write that \( p(\theta/x) \propto l(\theta/x) \). In many cases, classical expressions from frequentist statistics are obtained by Bayesian analysis assuming a flat prior. If the variable of interest ranges over \((0, +\infty)\) or \((-\infty, +\infty)\), then strictly speaking a flat prior does not exist, as if the constant takes on any non-zero value, the integral does not exist. In such cases a flat prior (assuming \( p(\theta/x) \propto l(\theta/x) \)) is referred to as an improper prior.

The Jeffreys’ Prior

Suppose we cannot easily find the natural scale on which the likelihood is in data-translated format, or that such a decomposition does not exist. Jeffreys (1961) proposed a general prior in such cases, based on the Fisher information \( I \) of the likelihood. Recall that
\[ I(\theta/x) = -E \left[ \frac{\delta^2 \ln l(\theta/x)}{\delta \theta^2} \right] \]
Jeffreys’ rule (giving the Jeffreys’ Prior) is to take as the prior \( p(\theta) \propto \sqrt{I(\theta/x)} \)

Example Consider the likelihood for \( n \) independent draws from a binomial, \( L(\theta/x) = C \theta^n (1-\theta)^{n-x} \)
where the constant \( C \) does not involve \( \theta \). Taking logs gives
\[ l(\theta/x) = \ln[L(\theta/x)] = \ln C + x \ln \theta + (n-x) \ln(1-\theta) \]
Thus \[ \frac{\delta l(\theta/x)}{\delta \theta} = \frac{x}{\theta} - \frac{n-x}{1-\theta} \] and likewise \[ \frac{\delta^2 l(\theta/x)}{\delta \theta^2} = -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2} = \left( \frac{x}{\theta^2} + \frac{n-x}{(1-\theta)^2} \right) \]
Since \( E(x) = n \theta \) we have
\[ I(\theta/x) = -E \left[ \frac{\delta^2 \ln l(\theta/x)}{\delta \theta^2} \right] = \frac{n \theta}{\theta^2} + \frac{n(1-\theta)}{(1-\theta)^2} = n \theta^{-1}(1-\theta)^{-1} \]
Time is precious, but we do not know yet how precious it really is. We will only know when we are no longer able to take advantage of it…

**Pr**overbs 21:5

The plans of the diligent lead to profit as surely as haste leads to poverty

Hence, the Jeffreys’ Prior becomes \( p(\theta) \propto \sqrt{\theta^{-1}(1-\theta)^{-1}} \) which is a Beta Distribution (which we discuss later).

When there are multiple parameters, \( I \) is the Fisher Information matrix, the matrix of the expected second partials,

\[
I(\Theta / x)_{ij} = -E \left[ \frac{\partial^2 \ln l(\Theta / x)}{\partial \theta_i \partial \theta_j} \right]
\]

In this case, the Jeffreys’ Prior becomes \( p(\Theta) \propto \sqrt{|I(\Theta / x)|} \)

**Example** Suppose our data consists of \( n \) independent draws from a normal distribution with unknown mean and variance, \( \mu \) and \( \sigma^2 \). In earlier notes on maximum likelihood estimation, we showed that the information matrix in this case is

\[
I = n \begin{pmatrix}
\frac{1}{\sigma^2} & 0 \\
0 & \frac{1}{2\sigma^4}
\end{pmatrix}
\]

Since the determinant of a diagonal matrix is the product of the diagonal elements, we have \( \det(I) = \sigma^{-6} \) giving the Jeffreys’ Prior for \( \mu \) and \( \sigma^2 \) as \( p(\Theta) \propto \sqrt{\sigma^{-6}} = \sigma^{-3} \)

Since the prior does not involve \( \mu \) we assume a flat prior for \( \mu \) (i.e. \( p(\mu) = \text{constant} \)). Note that the prior distributions of \( \mu \) and \( \sigma^2 \) are independent, as \( p(\mu, \sigma^2) = \text{constant} \ast \sigma^{-3} = p(\mu) \ast p(\sigma^2) \)

**Normal Models**

Why is it so common to model data using a normal distribution?

- Approximately normally distributed quantities appear often in nature.
- CLT tells us any variable that is basically a sum of independent components should be approximately normal.
- Note \( \bar{x} \) and \( S^2 \) are independent when sampling from a normal population — so if beliefs about the mean are independent of beliefs about the variance, a normal model may be appropriate.
- The normal model is analytically convenient (exponential family, sufficient statistics \( \bar{x} \) and \( S^2 \))
- Inference about the population mean based on a normal model will be correct as \( n \to \infty \) even if the data are truly non-normal.
- When we assume a normal likelihood, we can get a wide class of posterior distributions by using different priors.

**A Conjugate analysis with Normal Data (variance known)**

Simple situation: Assume data \( X_1, \ldots, X_n \) are iid \( N(\mu, \sigma^2) \), with \( \mu \) unknown and \( \sigma^2 \) known. We will make inference about \( \mu \). The likelihood is;

\[
L(\mu / x) = \prod_{i=1}^{n} \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left( \frac{X_i - \mu}{\sigma} \right)^2 \right\} = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2 \right\}
\]
A conjugate prior for $\mu$ is $\mu \sim N(\delta, \tau^2)$ \quad \Rightarrow \quad p(\mu) = \frac{1}{\tau \sqrt{2\pi}} \exp \left\{-\frac{1}{2} \left( \frac{\mu - \delta}{\tau} \right)^2 \right\}$ So the posterior is:

$$\pi(\mu | x) \propto L(\mu | x) p(\mu) \propto \exp \left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\} \times \exp \left\{-\frac{1}{2} \left( \frac{\mu - \delta}{\tau} \right)^2 \right\}$$

$$= \exp \left\{-\frac{1}{2} \left[ \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 + \frac{1}{\tau^2} (\mu - \delta)^2 \right]\right\} = \exp \left\{-\frac{1}{2} \left[ \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i^2 - 2\mu x_i + \mu^2) + \frac{1}{\tau^2} (\mu^2 - 2\delta \mu + \delta^2) \right]\right\}$$

So the posterior is:

$$\pi(\mu | x) \propto \exp \left\{-\frac{1}{2\sigma^2\tau^2} \left[ \tau^2 \sum_{i=1}^{n} x_i^2 - 2\mu \tau^2 n \bar{x} + n \tau^2 \mu^2 \right] + \left( \mu^2 \sigma^2 - 2\delta \mu \sigma^2 + \delta^2 \sigma^2 \right) \right\}$$

$$= \exp \left\{-\frac{1}{2\sigma^2\tau^2} \left[ \mu^2 (n \tau^2 + \sigma^2) - 2\mu (\delta \sigma^2 + \tau^2 n \bar{x}) + (\tau^2 \sum_{i=1}^{n} x_i^2 + \delta^2 \sigma^2) \right] \right\}$$

$$= \exp \left\{-\frac{1}{2} \left[ \mu^2 \left( \frac{n}{\sigma^2} + \frac{1}{\tau^2} \right) - 2\mu \left( \frac{n \bar{x}}{\sigma^2} + \frac{\delta}{\tau^2} \right) + k \right] \right\}$$

(\text{where } k \text{ is some constant})

Hence

$$\pi(\mu | x) \propto \exp \left\{-\frac{1}{2} \left[ \frac{n}{\sigma^2} + \frac{1}{\tau^2} \right] \left( \frac{n \bar{x}}{\sigma^2} + \frac{\delta}{\tau^2} \right)^2 \right\} \times \exp \left\{-\frac{1}{2} \left[ \frac{n}{\sigma^2} + \frac{1}{\tau^2} \right] \left( \mu - \frac{n \bar{x}}{n \tau^2 + \sigma^2} \right)^2 \right\}$$

Hence the posterior for $\mu$ is a normal distribution with;

Mean \( \frac{n \bar{x} \tau^2 + \delta \sigma^2}{n \tau^2 + \sigma^2} \) and variance \( \left( \frac{n}{\sigma^2} + \frac{1}{\tau^2} \right)^{-1} = \frac{\sigma^2 \tau^2}{n \tau^2 + \sigma^2} \)

The precision is the reciprocal of the variance. Here, $\frac{1}{\tau^2}$ is the prior precision, $\frac{n}{\sigma^2}$ is the data precision

$$\frac{n}{\sigma^2} + \frac{1}{\tau^2}$$ is the posterior precision

**Note** the posterior mean $E(\mu | x)$ is simply $E(\mu | x) = \frac{n \bar{x} \tau^2 + \delta \sigma^2}{n \tau^2 + \sigma^2} = \frac{\sigma^2}{n \tau^2 + \sigma^2} \delta + \frac{n \tau^2}{n \tau^2 + \sigma^2} \bar{x}$ a combination of the prior mean and the sample mean. If the prior is highly precise, the weight is large on $\delta$ and if the data are highly precise (e.g., when $n$ is large), the weight is large on $\bar{x}$.

Clearly as $n \to \infty$, $E(\mu | x) = \bar{x}$, and $\text{Var}(\mu | x) = \frac{\sigma^2}{n}$ if we choose a large prior variance $\tau^2$.

$\Rightarrow$ for $\tau^2$ large and $n$ large, Bayesian and frequentist inference about $\mu$ will be nearly identical

**A Conjugate analysis with Normal Data (mean known)**

I Now suppose $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$ with $\mu$ known and $\sigma^2$ unknown. We will make inference about $\sigma^2$. Our likelihood is
L(\sigma^2 / x) \propto (\sigma^2)^{\nu / 2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\} = \left(\sigma^2\right)^{\nu / 2} \exp \left\{ -\frac{n}{2\sigma^2} \left[ \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \right] \right\}

Let W denote the sufficient statistic.

The conjugate prior for \( \sigma^2 \) is the inverse gamma distribution. If a r.v. \( Y \sim \text{gamma} \), then \( 1/Y \sim \text{inverse gamma} \) (IG).

The prior for \( \sigma^2 \) is \( \pi(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-(\alpha+1)/2} e^{-\beta/\sigma^2} \) for \( \sigma^2 > 0 \) where \( \alpha > 0 \), \( \beta > 0 \).

Note the prior mean and variance are:

\[ E(\sigma^2) = \frac{\beta}{\alpha - 1} \text{ provided that } \alpha > 1, \text{ and } \Var(\sigma^2) = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)} \text{ provided that } \alpha > 2 \]

So the posterior for \( \sigma^2 \) is:

\[ \pi(\sigma^2 / x) \propto L(\sigma^2 / x) \pi(\sigma^2) \propto \left(\sigma^2\right)^{\nu / 2} e^{-n\sigma^2/2} (\sigma^2)^{-(\alpha+1)/2} e^{-\beta/\sigma^2} = \left(\sigma^2\right)^{-(\alpha+\nu/2+1)/2} \exp \left\{ -(\beta + \frac{\nu}{2} w) \right\} \]

Hence the posterior is clearly an IG(\( \alpha + \frac{\nu}{2} \), \( \beta + \frac{n}{2} w \)) distribution, where \( w = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \).

How to choose the prior parameters \( \alpha \) and \( \beta \): Note

\[ \alpha = \frac{[E(\sigma^2)]^2}{\Var(\sigma^2)} + 2 \text{ and } \beta = E(\sigma^2) \left[ \frac{[E(\sigma^2)]^2}{\Var(\sigma^2)} + 1 \right] \]

so we could make guesses about \( E(\sigma^2) \) and \( \Var(\sigma^2) \) and use these to determine \( \alpha \) and \( \beta \).

A Model for Normal Data (mean and variance both unknown)

When \( X_1, \ldots, X_n \) are iid \( N(\mu, \sigma^2) \) with both \( \mu \) and \( \sigma^2 \) unknown, the conjugate prior for the mean explicitly depends on the variance:

\[ p(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)/2} e^{-\beta/\sigma^2} \text{ and } p(\mu/\sigma^2) = (\sigma^2)^{-\nu/2} \exp \left\{ -\frac{1}{2\sigma^2/s_0} (\mu - \delta)^2 \right\} \]

The prior parameter \( s_0 \) measures the analyst’s confidence in the prior specification. When \( s_0 \) is large, we strongly believe in our prior.

The joint posterior for \( (\mu, \sigma^2) \) is:

\[ \pi(\mu, \sigma^2 / x) \propto L(\mu, \sigma^2 / x) p(\sigma^2) p(\mu / \sigma^2) \]

\[ \propto \left(\sigma^2\right)^{-(\alpha+\nu/2+1)/2} \exp \left\{ -\frac{\beta}{\sigma^2} - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 - \frac{1}{2\sigma^2/s_0} (\mu - \delta)^2 \right\} \]

\[ = \left(\sigma^2\right)^{-(\alpha+\nu/2+1)/2} \exp \left\{ -\frac{\beta}{\sigma^2} - \frac{1}{2\sigma^2} \left[ \sum_{i=1}^{n} x_i^2 - n\mu x + \mu^2 \right] - \frac{1}{2\sigma^2/s_0} (\mu^2 - 2\delta \mu - \delta^2) \right\} \]

\[ \pi(\mu, \sigma^2 / x) \propto \left(\sigma^2\right)^{-(\alpha+\nu/2+1)/2} \exp \left\{ -\frac{\beta}{\sigma^2} - \frac{1}{2\sigma^2} \left[ \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \right] \right\} \times \left(\sigma^2\right)^{1} \exp \left\{ -\frac{1}{2\sigma^2} \left[ (n + s_0) \mu^2 - 2(n\bar{x} + \delta s_0) \mu + (n\bar{x}^2 s_0 \delta^2) \right] \right\} \]

Note the second part is simply a normal kernel for \( \mu \).

To get the posterior for \( \sigma^2 \), we integrate out \( \mu \):
\[
\pi(\sigma^2 / x) = \int_0^\infty \pi(\mu, \sigma^2, x) d\mu \propto \left(\sigma^2\right)^{-\frac{\alpha + \frac{\nu}{2} + \frac{1}{2}}{2}} \exp\left\{-\frac{\beta}{\sigma^2} - \frac{1}{2\sigma^2} \left(\sum_{i=1}^{n} x_i^2 - nx^2\right)\right\}
\]
since the second piece (which depends on \(\mu\)) just integrates to a normalizing constant.
Hence since \(-(\alpha + \frac{\nu}{2} + \frac{1}{2}) = -(\alpha + \frac{n}{2} - \frac{1}{2}) - 1\), we see the posterior for \(\sigma^2\) is inverse gamma:
\[
\sigma^2 / x \overset{\text{IG}}{\sim} \alpha + \frac{n}{2} - \frac{1}{2}, \quad \beta + \frac{1}{2} \sum_{i=1}^{n} (x_i - \bar{x})^2
\]
Note that \(\pi(\mu / \sigma^2, x) = \pi(\mu, \sigma^2 / x) / \pi(\sigma^2 / x)\) and after lots of cancellation,
\[
\pi(\mu / \sigma^2, x) \propto \sigma^{-2} \exp\left\{-\frac{1}{2\sigma^2} \left[(n + s_0)\mu^2 - 2(n\bar{x} + \bar{s}_0)\mu + (n\bar{x}^2 s_0 \delta^2)\right]\right\}
\]
\[
= \sigma^{-2} \exp\left\{-\frac{1}{2\sigma^2 / (n + s_0)} \left[\mu^2 - 2 \frac{n\bar{x} + \bar{s}_0}{n + s_0} \mu + \frac{n\bar{x}^2 s_0 \delta^2}{n + s_0}\right]\right\}
\]
Clearly \(\pi(\mu / \sigma^2, x)\) is normal:
\[
\mu / \sigma^2, x \sim N\left(\frac{n\bar{x} + \bar{s}_0}{n + s_0}, \frac{\sigma^2}{n + s_0}\right), \quad \text{Note as } s_0 \to \infty, \quad \mu / \sigma^2, x \sim N\left(\bar{x}, \frac{\sigma^2}{n}\right)
\]
Note also the conditional posterior mean is
\[
\frac{n}{n + s_0} \bar{x} + \frac{s_0}{n + s_0} \delta
\]
The relative sizes of \(n\) and \(s_0\) determine the weighting of the sample mean \(\bar{x}\) and the prior mean \(\delta\)
The marginal posterior for \(\mu\) is:
\[
\pi(\mu / x) = \int_0^\infty \pi(\mu, \sigma^2 / x) d\sigma^2 = \int_0^\infty \left(\sigma^2\right)^{-\frac{\alpha + \frac{\nu}{2} + \frac{1}{2}}{2}} \exp\left\{-\frac{2\beta + (n + s_0)(\mu - \delta)^2}{2\sigma^2}\right\} d\sigma^2
\]
Letting \(A = 2\beta + (n + s_0)(\mu - \delta)^2\), \(z = \frac{A}{2\sigma^2} \Rightarrow \sigma^2 = \frac{A}{2z} \quad , \quad d\sigma^2 = -\frac{A}{2z^2} dz\) Thus
\[
\pi(\mu / x) = \int_0^\infty \left(\frac{A}{2z}\right)^{-(\alpha + \frac{\nu}{2} + \frac{1}{2})} \frac{A}{2z^2} e^{-z} dz = \int_0^\infty \left(\frac{A}{2z}\right)^{-(\alpha + \frac{\nu}{2} + \frac{1}{2})} e^{-z} dz \propto \frac{A^{-\alpha - \frac{\nu}{2} - \frac{1}{2}}}{2^{\alpha + \frac{\nu}{2} - 1}}
\]
This integrand is the kernel of a gamma density and thus the integral is a constant. So
\[
\pi(\mu / x) \propto A^{-(\alpha + \frac{\nu}{2} + \frac{1}{2})} = \left[2\beta + (n + s_0)(\mu - \delta)^2\right]^{-\alpha - \frac{\nu}{2} - \frac{1}{2}} \propto \left[1 + \frac{(n + s_0)(\mu - \delta)}{2\beta}\right]^{-\alpha - \frac{\nu}{2} - \frac{1}{2}}
\]
which is a (scaled) noncentral t kernel having noncentrality parameter \(\delta\) and degrees of freedom \(n + 2\alpha\).

**Appendix**

Bayesian inference differs from classical inference (such as the MLE, least squares and the method of moments) in that whereas the Classical approach treats the parameter \(\theta\) as fixed quantity and draws on a repeated sampling principle, the Bayesian approach regards \(\theta\) as the realized value of a random variable \(\Theta\) and specifies a probability distribution for the parameter(s) of interest. This makes life easier because it is clear that if we observe data \(X=x\), then we need to compute the conditional density of \(\Theta\) given \(X=x\) (“the posterior”)

**Why use Bayesian methods?** Some reasons:
Time is precious, but we do not know yet how precious it really is. We will only know when we are no longer able to take advantage of it…

Pr 21:5 The plans of the diligent lead to profit as surely as haste leads to poverty

i) We wish to specifically incorporate previous knowledge we have about a parameter of interest.
ii) To logically update our knowledge about the parameter after observing sample data
iii) To make formal probability statements about the parameter of interest.
iv) To specify model assumptions and check model quality and sensitivity to these assumptions in a straightforward way.

Why do people use classical methods?
• If the parameter(s) of interest is/are truly fixed (without the possibility of changing), as is possible in a highly controlled experiment
• If there is no prior information available about the parameter(s)
• If they prefer “cookbook”-type formulas with little input from the scientist/researcher

Many reasons classical methods are more common than Bayesian methods are historical:
• Many methods were developed in the context of controlled experiments.
• Bayesian methods require a bit more mathematical formalism.
• Historically (but not now) realistic Bayesian analyses had been infeasible due to a lack of computing power.

Motivation for Bayesian Modeling
1) Bayesians treat unobserved data and unknown parameters in similar ways.
2) They describe each with a probability distribution.
3) As their model, Bayesians specify:
   (i) A joint density function, which describes the form of the distribution of the full sample of data (given the parameter values)
   (ii) A prior distribution, which describes the behavior of the parameter(s) unconditional on the data
4) The prior could reflect:
   (i) Uncertainty about a parameter that is actually fixed
   (ii) the variety of values that a truly stochastic parameter could take.

Interpretations of Prior Distributions
• As frequency distributions
• As normative and objective representations of what is rational to believe about a parameter, usually in a state of ignorance
• As a subjective measure of what a particular individual, “you,” actually believes

Prior Frequency Distributions
Sometimes the parameter value may be generated by a stable physical mechanism that may be known, or inferred from previous data. e.g. a parameter that is a measure of a properties of a batch of material in an industrial inspection problem. Data on previous batches allow the estimation of a prior distribution. It has a physical interpretation in terms of frequencies

Normative/Objective Interpretation
The central problem in Bayesian estimation is specifying a prior distribution for a parameter about which nothing is known. If θ can only have a finite set of values, it seems natural to assume all values are equally likely a priori. This can have odd consequences. For example specifying a uniform prior on regression models:
Time is precious, but we do not know yet how precious it really is. We will only know when we are no longer able to take advantage of it…

[92x749]{}[]

Pr

overbs

21:5

The plans of the diligent lead to profit as surely as haste leads to poverty

assigns prior probability 6/16 to 3-variable models and prior probability only 4/16 to 2-variable models

Subjective Degrees of Belief
Under this view probability represents a subjective degree of belief held by a particular person at a particular time. There are various techniques for eliciting subjective priors.
For example, Good’s device of imaginary results. • e.g. binomial experiment. beta prior with $a=b$.
“Imagine” the experiment yields 1 tail and $n-1$ heads. How large should $n$ be in order that we would just give odds of 2 to 1 in favor of a head occurring next? (eg $n=4$ implies $a=b=1$)

Problems with Subjectivity
• What if the prior and the likelihood disagree substantially?
• The subjective prior cannot be “wrong” but may be based on a misconception
• The model may be substantially wrong
• Often use hierarchical models in practice:

Exponential families and conjugate priors
There is an important connection between exponential families (not to be confused with the exponential distribution) and conjugate priors.

Definition. The family of distributions $F$ is an exponential family if every member of $F$ has the form:

Essentially all of the distributions that we typically work with (normal, exponential, Poisson, beta, gamma, binomial, Bernoulli,...) are exponential families. The next theorem tells us when we can expect to have a conjugate prior.

Theorem 1 (Exponential families and conjugate priors). If the likelihood model is an exponential family, then there exists a conjugate prior.

Proof. Consider the likelihood of our iid data