Chapter 3

Decision making under risk: Data problem

Key words: Decisions, sample information, prior distribution, likelihood, posterior distribution, expected value of perfect information, expected value of sample information, cost of sampling, expected net gain from sampling, optimal sample size.

Suggested readings:

1. Winkler R. L. and Hays W. L. (1975), **Statistics: Probability, Inference, and Decision** (2nd edition), Holt, Rinehart and Winston
2. Frame P. and Taylor R. (1776), **Statistical Analysis for Business Decisions**, McGraw-Hill
3. Berger J. O. (1980), **Statistical decision Theory: Foundations, Concepts and Methods**, Springer- Verlag.
4. Goon A.M., Gupta M.K. and Dasgupta B, (1987), **An Outline of Statistical Theory (volume II)**, The World Press Private Ltd.

3.1 Introduction

The problems which we have considered in chapter 2 were essentially non data problems in the sense that no sample information was associated with them as there was no information resulting from some statistical experiment or phenomenon under investigation. We had some prior information (or intuitive knowledge or belief) about the parameter of interest θ in from of prior distribution $\xi(\theta)$ and we utilized this information to take decisions.

However, the real life situation is that the prior information about θ is seldom precise. Hence it is reasonable to modify or reshape the beliefs in terms of the probabilities of various possible values of θ being true. This can be accomplished by utilizing the information generated from the data resulting from the statistical experiment or phenomenon under investigation.

3.2 Use of sample information in decision making

Consider the following situation:

Now-a-days, quality control is an essential process of any production process. With the advancement of technology, more and more companies are adopting for 3σ or even 6σ limits thereby eliminating practically any defectives being produced (or rather being marketed). One of the criteria followed in quality control is that of the proportion of defectives. A production manager (who is concerned about the defectives being produced) feels that the proportion of defectives θ can take four possible values 0.01, 0.05, 0.10 and 0.50 corresponding to the following states of nature:

(i) The process is under control. In this case $\theta = 0.01$.
(ii) There is a 'type a' error in the production. In this case $\theta = 0.05$.
(iii) There is a 'type b' error in the production. In this case $\theta = 0.10$.
(iv) There are both 'type a' and 'type b' errors in the production. In this case $\theta = 0.50$

For the probabilities of occurrences of these proportions, he has some beliefs in the form of the probability distribution given by

$$P(\theta = 0.01) = 0.50$$

$$P(\theta = 0.05) = 0.30$$

$$P(\theta = 0.10) = 0.18$$

$$P(\theta = 0.50) = 0.02$$

This defines the prior probability distribution of θ. Now it is up to the production manager that he decides to act upon the basis of this information only or he goes for getting some information. In the former case this will be a no data problem.

Now suppose that he assumes the quality control process to be a Bernoulli process (with two outcomes non-defective or defective with respective probabilities θ and 1-θ). The assumption seems to be reasonable as the probability of an item being defective is independent of the previous defects.

He collects a sample of size 8 and examines the sample for the number of defectives, say k. The likelihood that k =1 for various values of θ is given by

$$P(k=1 \mid n=8, \theta=0.01) = \binom{8}{1}(0.01)(0.99)^7 = 0.07456$$

$$P(k=1 \mid n=8, \theta=0.05) = \binom{8}{1}(0.05)(0.95)^7 = 0.2793$$

$$P(k=1 \mid n=8, \theta=0.10) = \binom{8}{1}(0.10)(0.90)^7 = 0.3826$$

$$P(k=1 \mid n=8, \theta=0.50) = \binom{8}{1}(0.50)(0.50)^7 = 0.03125$$

Now, we calculate the posterior probabilities with the help of the following formula:

$$\text{Posterior probability} = \frac{\text{(prior probability)(likelihood probability)}}{\sum \text{(prior probability)(likelihood probability)}}$$

Then the posterior probabilities are presented in the following table:

Table 3.1

θ (I)	Prior probability (II)	Likelihood (III)	(Prior probability) (Likelihood) (IV)	Posterior probability (V)
0.01	0.50	0.07456	0.03728	0.1956
0.05	0.30	0.2793	0.08379	0.4397
0.10	0.18	0.3826	0.068868	0.3614
0.50	0.02	0.03125	0.000625	0.0033
	1.00		**0.190563**	**1.00**

Column (II) of the above table defines the probability distribution based on the beliefs of the production manager whereas column (V) defines the probability distribution based on his beliefs as well as the sample information. In this sense it is a modified probability distribution of θ.

If we compare prior and posterior probabilities, the rationale of the whole exercise becomes clear. For θ = 0.01, the prior probability was 0.50 but the revised probability is 0.1956. The subjective belief of the production manager is being modified on the basis of the objective information. He was assuming his process to be almost fault free. However, this is not the situation; the probabilities have to be revised on the basis of the sample information.

One thing should be noted at this point. The concept of prior and posterior probabilities is relative. A set of probabilities which has been obtained as posterior probabilities may serve as a prior probability distribution at a later point. Consider, for example, the situation when the production manager wants to be more objective and in the process, he decides to pick another sample of size 7. The sample contains two defectives. Now the earlier set of posterior probabilities serves as the prior probability distribution at this stage. Let us obtain the posterior probability distribution of this prior probability distribution.

$$P(k=2 \mid n=7, \theta=0.1956) = \binom{7}{2}(0.1956)^2(0.8044)^5 = 0.270593$$

$$P(k=2 \mid n=7, \theta=0.4397) = \binom{7}{2}(0.4397)^2(0.5603)^5 = 0.2242$$

$$P(k=2 \mid n=7, \theta=0.3614) = \binom{7}{2}(0.3614)^2(0.6386)^5 = 0.2913$$

$$P(k=2 \mid n=7, \theta=0.0033) = \binom{7}{2}(0.0033)^2(0.9967)^5 = 0.000225$$

Then the posterior probabilities may be calculated as follows:

Table 3.2

θ (I)	Prior probability (II)	Likelihood (III)	(Prior probability) (Likelihood) (IV)	Posterior probability (V)
0.01	0.1956	0.270593	0.052928	0.206118
0.05	0.4397	0.2242	0.098581	0.383903
0.10	0.3614	0.2913	0.105276	0.409976
0.50	0.0033	0.000225	0.000000742	2.89E-06
	1.00		**0.256785**	**1.00**

Now compare the posterior probability distribution with the earlier two sets of probabilities. Earlier the production manager had thought that the probability of the process being under control was the highest. Does he still retain his belief? Let us draw the graphs of the three situations:

(i) **For the initial set of probabilities.**

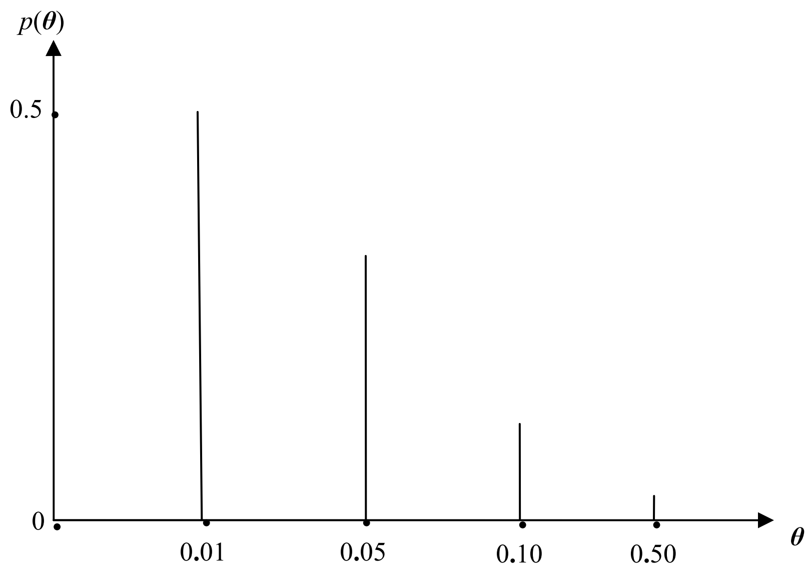

Fig. 3.1

(ii) **For the first set of posterior probabilities**.

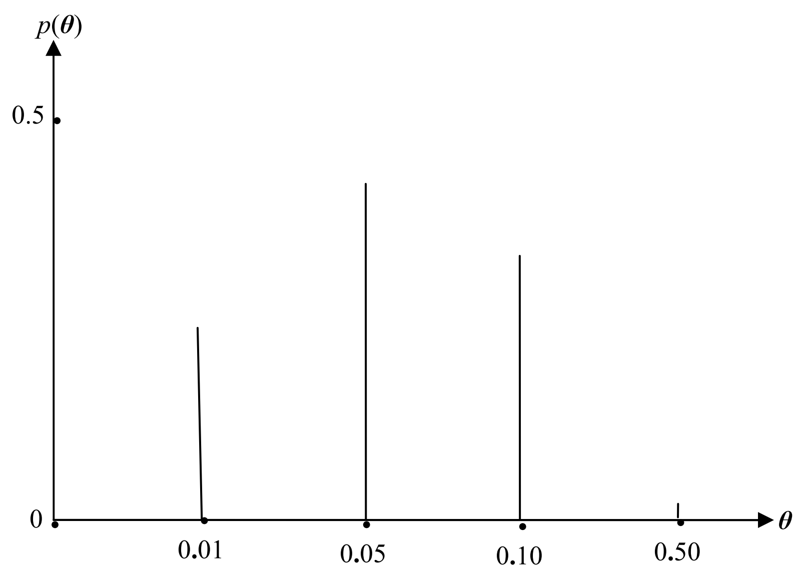

Fig. 3.2

(iii) **For the second set of posterior probabilities.**

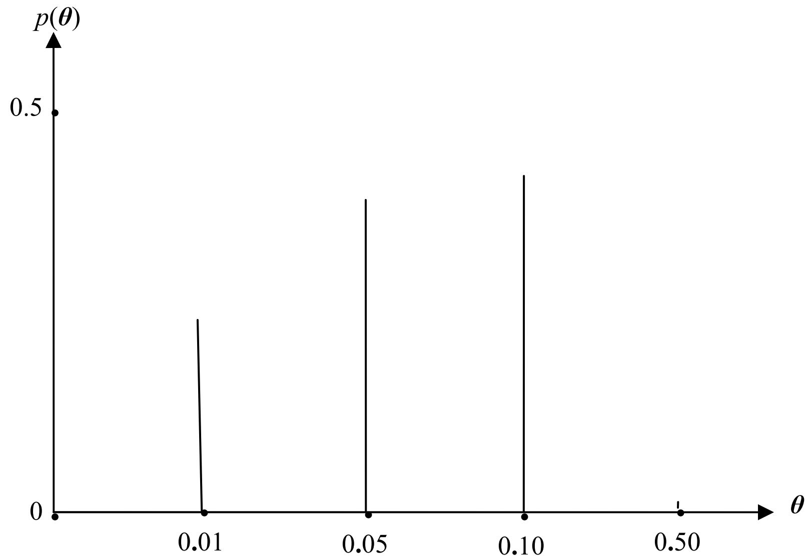

Fig. 3.3

Note the change in the shapes of the graphs.

The average proportion of defectives for the three sets of probabilities, i.e. $E(\theta)$ are given by

$$E(\theta) = (0.01)(0.50) + (0.05)(0.30) + (0.10)(0.18) + (0.50)(0.02) = 0.048$$

$$E(\theta) = (0.01)(0.1956) + (0.05)(0.4397) + (0.10)(0.3614) + (0.50)(0.0033) = 0.0617$$

$$E(\theta) = (0.01)(0.206118) + (0.05)(0.383903) + (0.10)(0.409976) + (0.50)(2.89\text{E-}06) = 0.0623$$

The actual proportion of defectives is much higher as compared to the belief of the production manager.

Now consider a situation when in place of taking two successive samples of total size 15 and finding respectively 1 and 2 defectives, the production manager takes a large sample of size 15 and finds 3 defectives. How would the results be affected? Compute the probabilities with this sample and you will find the same results!

How is this information related t the decision theory? The production manager is just being more objective as he is gathering more information.
In fact, the job of the production manager is to minimize the number of defectives. A defective item needs repair or replacement which results in increased production cost. Also, shipment of a defective item affects the goodwill of the firm and in some cases a defective item may be life threatening. Whereas the later two

consequences may be difficult to be compute numerically, the first consequence may be computed mathematically.

Suppose that the cost of correcting (repairing or replacing) a defective item to the firm is Rs. 75. The profit of a non-defective item is Rs. 15. One defective item 'eats up' the profit of 5 non-defective items**!!** Suppose that the firm has obtained an order of 5, 00,000 units of the product. The maximum profit that can be realized by the firm is, thus Rs. 75, 00,000. Let us compute the profit of the firm under different states of nature.

Table 3.3

Proportion of defective (θ) I	Number of defectives I (5,00,000) II	Cost of correcting defectives 75(II) III	Profit 75, 00,000-III IV
0.01	5,000	375000	7125000
0.05	25,000	1875000	5625000
0.10	75,000	5625000	1875000
0.50	2,50,000	18750000	-11250000

The expected profits for the three sets of probabilities are then given by

$$E(P_1) = (0.50)(7125000) + (0.30)(5625000) + (0.18)(1875000) + (0.02)(-11250000) \ = \ 5362500$$

$$E(P_2) = (0.1956)(7125000) + (0.4397)(5625000) + (0.3614)(1875000) + (0.0033)(-11250000) \ = \ 4507462.50$$

$$E(P_3) = (0.206118)(7125000) + (0.383903)(5625000) + (0.409976)(1875000) + (2.89\text{E-}06)\ (-11250000) = \ 4396717.60$$

Suppose that the production manager has the following possible course of actions from which he has to choose one:

(i) Leave the process as it is.

(ii) Adjust the process at a cost of Rs. 10,000 so that the maximum possible value of θ is 0.10.

(iii) Adjust the process at a cost of Rs. 50,000 so that the maximum possible value of θ is 0.05.

(iv) Adjust the process at a cost of Rs. 5, 00,000 so that the maximum possible value of θ is 0.01.

What should be the decision of the production manager? One thing is obvious. Any adjustment would be ineffective if the value of θ is already less than or equal to the adjusted value. In such situation(s), the adjustment would just increase the cost of production.

We prepare the following pay-off table:

Table 3.4: Conditional pay-off matrix

Strategies	States of nature (θ)			
	0.01	0.05	0.10	0.50
No adjustment	7125000	5625000	1875000	-11250000
First adjustment	7115000	5615000	1865000	1865000
Second adjustment	7075000	5575000	5575000	5575000
Third adjustment	6525000	6525000	6525000	6525000

We calculate the expected pay-offs of various decisions using final probabilities:

$$E(P \text{ for no adjustment}) = (0.206118)(7125000) + (0.383903)(5625000) + (0.409976)(1875000) + (2.89\text{E-}06)\ (-11250000) = 4396717.60$$

$$E(P \text{ for first adjustment}) = (0.206118)(7115000) + (0.383903)(5615000) + (0.409976)(1865000) + (2.89\text{E-}06)(1865000) = 4386755.54$$

$$E(P \text{ for second adjustment}) = (0.206118)(7075000) + (0.383903)(5575000) + (0.409976)(5575000) + (2.89\text{E-}06)(5575000) = 5884176.39$$

$$E(P \text{ for third adjustment}) = (0.206118)(6525000) + (0.383903)(6525000) + (0.409976)(6525000) + (2.89\text{E-}06)(6525000) = 6524999.28$$

If he did not have collected the data, then the expected pay-offs would have been

$$E(P \text{ for no adjustment}) = (0.5)(7125000) + (0.3)(5625000) + (0.18)(1875000) + (0.02)\ (-11250000) = 5362500.00$$

$$E(P \text{ for first adjustment}) = (0.5)(7115000) + (0.3)(5615000) + (0.18)(1865000) + (0.02)(1865000) = 5615000.00$$

$$E(P \text{ for second adjustment}) = (0.5)(7075000) + (0.3)(5575000) + (0.18)(5575000) + (0.02)(5575000) = 6325000.00$$

$$E(P \text{ for third adjustment}) = (0.5)(6525000) + (0.3)(6525000) + (0.18)(6525000) + (0.02)(6525000) = 6525000.00$$

Although his decision would have been the same, still the difference between the expected pay-offs is worth noting.

Now, consider situation of another firm where the production manager is facing somewhat similar situation but with the following information:

This production manager feels that the proportion of defectives θ can take four possible values 0.01, 0.05, 0.10 and 0.25 corresponding to the following states of nature:

(i) The process is under control. In this case $\theta = 0.01$.

(ii) There is a 'type a' error in the production. In this case $\theta = 0.05$.

(iii) There is a 'type b' error in the production. In this case $\theta = 0.10$.

(iv) There are both 'type a' and 'type b' errors in the production. In this case $\theta = 0.25$

For the probabilities of occurrences of these defectives, he has some beliefs in the form of the probability distribution given by

$$P(\theta = 0.01) = 0.80$$

$$P(\theta = 0.05) = 0.10$$

$$P(\theta = 0.10) = 0.08$$

$$P(\theta = 0.50) = 0.02$$

Definitely this man is more optimistic about his production process. Again suppose that he assumes the quality control process to be a Bernoulli process (with two outcomes non-defective or defective with respective probabilities θ and $1-\theta$.)

He collects a sample of size 5 and examines the sample for the number of defectives, say k. The likelihood that $k = 1$ for various values of θ is given by

$$P(k = 1 \mid n = 5, \theta = 0.01) = \binom{5}{1}(0.01)(0.99)^4 = 0.04803$$

$$P(k = 1 \mid n = 5, \theta = 0.05) = \binom{5}{1}(0.05)(0.95)^4 = 0.20363$$

$$P(k = 1 \mid n = 5, \theta = 0.10) = \binom{5}{1}(0.10)(0.90)^4 = 0.32805$$

$$P(k = 1 \mid n = 5, \theta = 0.25) = \binom{5}{1}(0.25)(0.75)^4 = 0.39551$$

The production manger obtains the following posterior probabilities:

Table 3.5

θ (I)	Prior probability (II)	Likelihood (III)	(Prior probability) (Likelihood) (IV)	Posterior probability (V)
0.01	0.80	0.04803	0.038424	0.413423
0.05	0.10	0.20363	0.020363	0.219093
0.10	0.08	0.32805	0.026244	0.282374
0.25	0.02	0.39551	0.00791	0.08511
	1.00		**0.092941**	**1.00**

Another sample of size 10 contains 2 defectives so he has the following table:

Table 3.6

θ (I)	Prior probability (II)	Likelihood (III)	(Prior probability) (Likelihood) (IV)	Posterior probability (V)
0.01	0.413423	0.015827	0.006543	0.234207
0.05	0.219093	0.025054	0.005489	0.196473
0.10	0.282374	0.002642	0.000746	0.026706
0.25	0.08511	0.178119	0.01516	0.542613
	1.00		**0.027938**	**1.00**

The average numbers of defectives, in the two cases are given by

$$E(\theta) = (0.01)(0.80) + (0.05)(0.10) + (0.10)(0.08) + (0.25)(0.02) \;=\; 0.026$$

$$E(\theta) = (0.01)(0.234207) + (0.05)(0.196473) + (0.10)(0.026706) + (0.25)(542613) \;=\; 0.1505$$

By now, we know that the second figure of average proportion of defectives is the sum of proportions of defectives in the two samples.

If the cost of correcting defectives is same as in the earlier case, then the production manager gets the following table:

Table 3.7

Proportion of defective (θ) I	Number of defectives I (5,00,000) II	Cost of correcting defectives 75(II) III	Profit 75, 00,000-III IV
0.01	5,000	375000	7125000
0.05	25,000	1875000	5625000
0.10	75,000	5625000	1875000
0.25	1,25,000	9375000	- 1875000

The expected profits are then given by

$$E(P) = (0.80)(7125000) + (0.10)(5625000) + (0.08)(1875000) + (0.02)(-1875000) = 6375000$$

$$E(P) = (0.206118)(7125000) + (0.196473)(5625000) + (0.026706)(1875000) + (0.542613)(-1875000) = 1806564.58$$

For the similar type of adjustments, his conditional pay-offs are given by

Table 3.8: Conditional pay-off matrix

Strategies	States of nature (θ)			
	0.01	0.05	0.10	0.25
No adjustment	7125000	5625000	1875000	-1875000
First adjustment	7115000	5615000	1865000	1865000
Second adjustment	7075000	5575000	5575000	5575000
Third adjustment	6525000	6525000	6525000	6525000

We calculate the expected pay-offs of various decisions using final probabilities:

$$E(P \text{ for no adjustment}) = (0.234207)(7125000) + (0.196473)(5625000) + (0.026706)(1875000) + (0.542613)\ (-1875000) = 1806560$$

$$E(P \text{ for first adjustment}) = (0.234207)(7115000) + (0.196473)(5615000) + (0.026706)(1865000) + (0.542613)(1865000) = 3831359$$

$$E(P \text{ for second adjustment}) = (0.234207)(7075000) + (0.196473)(5575000) + (0.026706)(5575000) + (0.542613)(5575000) = 5926305$$

$$E(P \text{ for third adjustment}) = (0.234207)(6525000) + (0.196473)(6525000) + (0.026706)(6525000) + (0.542613)(6525000) = 6524993$$

If he did not have collected the data, then the expected pay-offs would have been

$$E(P \text{ for no adjustment}) = (0.8)(7125000) + (0.1)(5625000) + (0.08)(1875000) + (0.02)\ (-1875000) = \ 6375000$$

$$E(P \text{ for first adjustment}) = (0.8)(7115000) + (0.1)(5615000) + (0.08)(1865000) + (0.02)(1865000) = 6440000$$

$$E(P \text{ for second adjustment}) = (0.8)(7075000) + (0.1)(5575000) + (0.08)(5575000) + (0.02)(5575000) = 6775000$$

$$E(P \text{ for third adjustment}) = (0.8)(6525000) + (0.1)(6525000) + (0.08)(6525000) + (0.02)(6525000) = 7112250$$

The decision of the production manager has changed!! Earlier he was satisfied with the second type of adjustment but after collecting the sample information, he will go for the type third adjustment.

3.3 Decisions- Terminal and Pre posterior

We have seen that the sampling affects the decision making process. Sample information can alter the decisions or substantiate the already made decisions. If a decision is made on the basis of the current information, it is called a **terminal decision**. The terminal decisions involve any latest information provided by the sample, (which in general is in terms of the posterior distribution). On the other hand, depending upon the circumstances, a decision maker can decide whether to collect any sample information or not. Such a decision may be necessary because additional sample information may be helpful to the decision maker but it involves cost. There has to be justification in incurring additional cost in the sense that the additional information should be useful enough. Such a decision (whether to collect additional information or not) is called a **pre posterior decision**. The name has been given because after this decision, a posterior distribution is obtained by collecting additional sample information. A pre posterior decision can be facilitated by expected value of perfect information or expected value of sample information.

Expected value of perfect information (EVPI)

What kind of information does a decision maker look for which is expected to be useful in making terminal decisions? As we have seen in chapter 1, the ideal situation is that the information is perfect. For example, if an investor knows with certainty about the prices of various stocks one year from now, he would decide with certainty about the stock(s) in which investment should be made. As we have seen that obtaining information involves cost. Then the next question is: how much is he willing to spend in obtaining this information? As we have seen in earlier chapters, the information is not worth more than the difference between the profits that he would make as a consequence of perfect information and the expected profit as a consequence of making decision under uncertainty.

But, as we have pointed out, this is an ideal situation. An investor does not know with certainty the next day prices of the stock(s); leave alone one year's time. This means, in place of computing values of perfect information, one should be interested in obtaining its estimates, i.e. expected value of the perfect information (**EVPI**). Recall the pay-off table of the second production manager:

Table 3.9: Conditional pay-off matrix

Strategies	States of nature (θ)			
	0.01	0.05	0.10	0.25
No adjustment	7125000*	5625000	1875000	-1875000
First adjustment	7115000	5615000	1865000	1865000
Second adjustment	7075000	5575000	5575000	5575000
Third adjustment	6525000	6525000*	6525000*	6525000*

*- The optimal action for the respective probabilities.

Using the prior probabilities, the expected rewards are

$$E(P \text{ for no adjustment}) = 6375000$$

$$E(P \text{ for first adjustment}) = 6440000$$

$$E(P \text{ for second adjustment}) = 6775000$$

$$E(P \text{ for third adjustment}) = 7112250$$

Here the optimal action is "to go for the second adjustment".
Then the values of the perfect information are given in the following table:

Table 3.10

θ	VPI(θ) (= $E(a_\theta)$-$E(a_\theta$*))
0.8	7125000-7075000 = 50000
0.1	6525000-5575000 = 950000
0.08	6525000-5575000= 950000
0.02	6525000-5575000= 950000
EVPI = $\Sigma_\theta (E(a_\theta)-E(a_\theta^*))\, P(\theta)$ = 2,30,000	

Thus the production manager would pay any amount up to Rs. 2, 30,000 to get the perfect information about the number of defectives.

For the first production manager, although further sample information just substantiates his decision, still he might be interested in incurring this additional cost in order to make his reasoning water-tight. How much would he spend on this account? The conditional pay-off table for this manager is given as

Table 3.11: Conditional pay-off matrix

Strategies	States of nature (θ)			
	0.01	0.05	0.10	0.50
No adjustment	7125000*	5625000	1875000	-11250000
First adjustment	7115000	5615000	1865000	1865000
Second adjustment	7075000	5575000	5575000	5575000
Third adjustment	6525000	6525000*	6525000*	6525000*

Using the prior probabilities, the expected rewards are

$$E(P \text{ for no adjustment}) = 5362500.00$$

$$E(P \text{ for first adjustment}) = 5615000.00$$

$$E(P \text{ for second adjustment}) = 6325000.00$$

$$E(P \text{ for third adjustment}) = 6525000.00$$

Here the optimal action is "to go for the third adjustment".

Then the values of the perfect information are given in the following table:

Table 3.12

θ	VPI(θ) (= $E(a_\theta)$-$E(a_\theta*)$)
0.5	7125000-6525000 = 600000
0.3	0
0.18	0
0.02	0
EVPI = Σ $_\theta$ $(E(a_\theta)$-$E(a_\theta*))$ $P(\theta)$ **= 3,00,000**	

This production manager also would pay any amount up to Rs. 3, 00,000 to get the perfect information about the number of defectives.

Expected value of sample information (EVSI)

As, we have seen, if a decision maker is able to obtain perfect information at a cost less than or equal to EVPI, the situation would be ideal for him. But this is not the situation always. Perfect information is seldom available, if any. Then how should the decision maker proceed? In such situations, the decision maker relies upon the sample information. This information would not be perfect but can be taken as an estimate of the perfect information. The expected cost that may be incurred in obtaining the sample information is called the expected value of sample information (**EVSI**).

In the quality control examples considered above, suppose the two production managers want to purchase some sample information. As simple cases, we will consider three samples of sizes 1, 2 and 3 which can be drawn from the population and are observed for the number of defectives.

(i) First production process

(a) Sample size is 1.

In this case, the number of defectives could be zero or one. Probability of a defective item is θ and that of a non defective item is 1- θ. We calculate the expected pay-offs associated with the sample information:

Table 3.13

Number of defectives	θ	$\pi(\theta)$	Likelihood	Likelihood ($\pi(\theta)$)	Posterior probability
1	0.01	0.5	0.01	0.005	0.104167
	0.05	0.3	0.05	0.015	0.3125
	0.1	0.18	0.1	0.018	0.375
	0.25	0.02	0.5	0.01	0.208333
				0.048	
0	0.01	0.5	0.99	0.495	0.519958
	0.05	0.3	0.95	0.285	0.29937
	0.1	0.18	0.9	0.162	0.170168
	0.25	0.02	0.5	0.01	0.010504
				0.952	

For a defective item, the expected pay-offs are given by

$$E(P \text{ for no adjustment} \mid \text{defective}) = 859381.125$$

$$E(P \text{ for first adjustment} \mid \text{defective}) = 3583751.75$$

$$E(P \text{ for second adjustment} \mid \text{defective}) = 5731250.5$$

$$E(P \text{ for third adjustment} \mid \text{defective}) = 6525000$$

The optimal decision is to make third adjustment.

For a non- defective item, the expected pay-offs are given by

$$E(P \text{ for no adjustment} \mid \text{non-defective}) = 5589552$$

$$E(P \text{ for first adjustment} \mid \text{non-defective}) = 5717417$$

$$E(P \text{ for second adjustment} \mid \text{non-defective}) = 6354937$$

$$E(P \text{ for third adjustment} \mid \text{non-defective}) = 6525000$$

The optimal decision is to make third adjustment.

Under prior distribution, the optimal decision was to go for the third adjustment. Then VSI for the two possibilities of the sample, i.e., defective or non-defective are both zero.
i.e.

$$\begin{aligned} EVSI\ (n = 1) &= VSI\ (\text{third adjustment} \mid \text{defective})\ p\ (\text{defective}) \\ &\quad + VSI\ (\text{third adjustment} \mid \text{non- defective})\ p\ (\text{non - defective}) \\ &= 0 \end{aligned}$$

(b) Sample size is 2.

In this case, the number of defectives could be zero, one or two. Probability of two defective items is θ^2 , that of one defective item is 2θ (1- θ) and that of a non defective item is (1- $\theta)^2$. We calculate the expected pay-offs associated with the sample information:

Table 3.14

Number of defectives	θ	$\pi(\theta)$	Likelihood	Likelihood ($\pi(\theta)$)	Posterior probability
2	0.01	0.5	0.0001	0.00005	0.006579
	0.05	0.3	0.0025	0.00075	0.098684
	0.1	0.18	0.01	0.0018	0.236842
	0.25	0.02	0.25	0.005	0.657895
1	0.01	0.5	0.0099	0.0099	0.122525
	0.05	0.3	0.0475	0.0285	0.352723
	0.1	0.18	0.09	0.0324	0.40099
	0.25	0.02	0.25	0.01	0.123762
0	0.01	0.5	0.9801	0.49005	0.537571
	0.05	0.3	0.9025	0.27075	0.297005
	0.1	0.18	0.81	0.1458	0.159939
	0.25	0.02	0.25	0.005	0.005485

For two defective items, the expected pay-offs are given by

$$E(P \text{ for no adjustment} \mid \text{two defectives}) = -6355267.125$$

$$E(P \text{ for first adjustment} \mid \text{two defectives}) = 2269604.75$$

$$E(P \text{ for second adjustment} \mid \text{two defectives}) = 5584868.5$$

$$E(P \text{ for third adjustment} \mid \text{two defectives}) = 6525000$$

The optimal decision is to make third adjustment.

For one defective item, the expected pay-offs are given by

$$E(P \text{ for no adjustment} \mid \text{one defective}) = 2216591.25$$

$$E(P \text{ for first adjustment} \mid \text{one defective}) = 3830967.5$$

$$E(P \text{ for second adjustment} \mid \text{one defective}) = 5758787.5$$

$$E(P \text{ for third adjustment} \mid \text{one defective}) = 6525000$$

The optimal decision is to make third adjustment.

For a non- defective item, the expected pay-offs are given by

$$E(P \text{ for no adjustment} \mid \text{non-defective}) = 5739025.875$$

$$E(P \text{ for first adjustment} \mid \text{non-defective}) = 5801016.5$$

$$E(P \text{ for second adjustment} \mid \text{non-defective}) = 6381356.5$$

$$E(P \text{ for third adjustment} \mid \text{non-defective}) = 6525000$$

The optimal decision is to make third adjustment.

Under prior distribution, the optimal decision was to go for the third adjustment. Then VSI for the three possibilities of the sample, i.e., two defectives, one defective or non-defective are all zero.

i.e.

$$\begin{aligned} EVSI\ (n = 2) &= VSI\ (\text{third adjustment} \mid \text{two defective})\ p\ (\text{two defective}) \\ &+ VSI\ (\text{third adjustment} \mid \text{one defective})\ p\ (\text{one defective}) \\ &+ VSI\ (\text{third adjustment} \mid \text{non- defective})\ p\ (\text{non - defective}) \\ &= 0 \end{aligned}$$

(c) Sample size is 3.

In this case, the number of defectives could be zero, one, two or three. Probability of three defective items is θ^3, that of two defective item is $3\theta^2$ $(1- \theta)$, that of one defective item is 3θ $(1- \theta)^2$ and that of a non defective item is $(1- \theta)^3$. We calculate the expected pay-offs associated with the sample information:

Table 3.15

Number of defectives	θ	$\pi(\theta)$	Likelihood	Likelihood ($\pi(\theta)$)	Posterior probability
3	0.01	0.5	0.000001	5E-07	0.000184
	0.05	0.3	0.000125	3.75E-05	0.013797
	0.1	0.18	0.001	0.00018	0.066225
	0.25	0.02	0.125	0.0025	0.919794
2	0.01	0.5	0.000297	0.000149	0.010139
	0.05	0.3	0.007125	0.002138	0.145944
	0.1	0.18	0.027	0.00486	0.331831
	0.25	0.02	0.375	0.0075	0.512085
1	0.01	0.5	0.029403	0.014702	0.137972
	0.05	0.3	0.135375	0.040613	0.381145
	0.1	0.18	0.243	0.04374	0.410496
	0.25	0.02	0.375	0.0075	0.070387
0	0.01	0.5	0.970299	0.48515	0.553772
	0.05	0.3	0.857375	0.257213	0.293594
	0.1	0.18	0.729	0.13122	0.14978
	0.25	0.02	0.125	0.0025	0.002854

For three defective items, the expected pay-offs are given by

$$E(P \text{ for no adjustment} \mid \text{three defectives}) = -10144591.5$$

$$E(P \text{ for first adjustment} \mid \text{three defectives}) = 1917704.75$$

$$E(P \text{ for second adjustment} \mid \text{three defectives}) = 5575276$$

$$E(P \text{ for third adjustment} \mid \text{three defectives}) = 6525000$$

The optimal decision is to make third adjustment

For two defective items, the expected pay-offs are given by

$$E(P \text{ for no adjustment} \mid \text{two defectives}) = -4245597.75$$

$$E(P \text{ for first adjustment} \mid \text{two defectives}) = 2465517.885$$

$$E(P \text{ for second adjustment} \mid \text{two defectives}) = 5590202.925$$

$$E(P \text{ for third adjustment} \mid \text{two defectives}) = 6525000$$

The optimal decision is to make third adjustment.

For one defective item, the expected pay-offs are given by

$$E(P \text{ for no adjustment} \mid \text{one defective}) = 3104817.375$$

$$E(P \text{ for first adjustment} \mid \text{one defective}) = 4018646.75$$

$$E(P \text{ for second adjustment} \mid \text{one defective}) = 5781958$$

$$E(P \text{ for third adjustment} \mid \text{one defective}) = 6525000$$

The optimal decision is to make third adjust

For a non- defective item, the expected pay-offs are given by

$$E(P \text{ for no adjustment} \mid \text{non-defective}) = 5845821.75$$

$$E(P \text{ for first adjustment} \mid \text{non-defective}) = 5873280.5$$

$$E(P \text{ for second adjustment} \mid \text{non-defective}) = 6405658$$

$$E(P \text{ for third adjustment} \mid \text{non-defective}) = 6525000$$

The optimal decision is to make third adjustment.

Under prior distribution, the optimal decision was to go for the third adjustment. Then VSI for the three possibilities of the sample, i.e., two defectives, one defective or non-defective are all zero.
i.e.

$$\begin{aligned} EVSI\ (n = 3) &= \sum_i VSI\ (\text{ third adjustment} \mid i \text{ defectives })\ p\ (i \text{ defectives}) \\ &= 0 \end{aligned}$$

(ii) Second production process

(a) Sample size is 1.

In this case, the number of defectives could be zero or one. Probability of a defective item is θ and that of a non defective item is 1- θ. We calculate the expected pay-offs associated with the sample information:

Table 3.16

Number of defectives	$\boldsymbol{\theta}$	$\boldsymbol{\pi(\theta)}$	**Likelihood**	**Likelihood ($\boldsymbol{\pi(\theta)}$)**	**Posterior probability**	**Pay-offs**
1	0.01	0.8	0.01	0.008	0.258065	2625002
	0.05	0.1	0.05	0.005	0.16129	3824681
	0.1	0.08	0.1	0.008	0.258065	5962103
	0.25	0.02	0.5	0.01	0.322581	6525007
0	0.01	0.8	0.99	0.792	0.813142	6452006
	0.05	0.1	0.95	0.095	0.097536	6499756
	0.1	0.08	0.9	0.072	0.073922	6794713
	0.25	0.02	0.75	0.015	0.0154	6525000

(b) Sample size is 2.

In this case, the number of defectives could be zero, one or two. Probability of two defective items is θ^2 , that of one defective item is 2θ (1- θ) and that of a non defective item is (1- $\theta)^2$. We calculate the expected pay-offs associated with the sample information:

Table 3.17

Number of defectives	$\boldsymbol{\theta}$	$\boldsymbol{\pi(\theta)}$	**Likelihood**	**Likelihood ($\boldsymbol{\pi(\theta)}$)**	**Posterior probability**	**Pay-offs**
2	0.01	0.8	0.0001	0.00008	0.033613	475836.4
	0.05	0.1	0.0025	0.00025	0.105042	2435374
	0.1	0.08	0.01	0.0008	0.336134	5625414
	0.25	0.02	0.0625	0.00125	0.52521	6524993
1	0.01	0.8	0.0099	0.01584	0.335309	3794136
	0.05	0.1	0.0475	0.0095	0.201101	4379501
	0.1	0.08	0.09	0.0144	0.304826	6077964
	0.25	0.02	0.1875	0.0075	0.158764	6525000
0	0.01	0.8	0.9801	0.78408	0.825017	6518056
	0.05	0.1	0.9025	0.09025	0.094962	6552445
	0.1	0.08	0.81	0.0648	0.068183	6812520
	0.25	0.02	0.5625	0.01125	0.011837	6525000

(c) Sample size is 3.

In this case, the number of defectives could be zero, one, two or three. Probability of three defective items is θ^3, that of two defective item is $3\theta^2$ (1- θ), that of one defective item is 3θ (1- $\theta)^2$ and that of a non defective item is (1- $\theta)^3$. We calculate the expected pay-offs associated with the sample information:

Table 3.18

Number of defectives	θ	$\pi(\theta)$	Likelihood	Likelihood ($\pi(\theta)$)	Posterior probability	Pay-off
3	0.01	0.8	0.000001	8E-07	0.001971	886957.9
	0.05	0.1	0.000125	1.25E-05	0.030803	1990857
	0.1	0.08	0.001	0.00008	0.197141	5577951
	0.25	0.02	0.015625	0.000313	0.770084	6524993
2	0.01	0.8	0.01	0.8	0.01	71250
	0.05	0.1	0.05	0.1	0.05	1004650
	0.1	0.08	0.1	0.08	0.1	2300750
	0.25	0.02	0.25	0.02	0.25	2675250
1	0.01	0.8	0.029403	0.023522	0.362232	4071232
	0.05	0.1	0.135375	0.013538	0.20847	4548481
	0.1	0.08	0.243	0.01944	0.299365	6118348
	0.25	0.02	0.421875	0.008438	0.129933	6525000
0	0.01	0.8	0.970299	0.776239	0.835803	6575086
	0.05	0.1	0.857375	0.085738	0.092317	6599155
	0.1	0.08	0.729	0.05832	0.062795	6828705
	0.25	0.02	0.421875	0.008438	0.009085	6525000

Find out EVSI for this manager for different sample sizes. Does this information help him to substantiate the decision or to change the decision? Which sample size do you recommend?

3.4 Linear pay-of function for continuous θ: Two action problem

In the examples considered above, we have assumed that there are only four possible states (θ = 0.01, 0.05. 0.10, 0.50 (0.25)) of the random variable θ where θ denotes the proportion of defectives in the production processes and probabilities have been assigned to the possible states of θ. However, as we know, the problem is too simplified for a real-life production process where θ would rarely be a discrete random variable taking only a finite number of values. The more realistic assumption would be to assume θ a continuous random variable taking values in the range 0 to 1.

As we have seen in chapter 2, in such cases, the pay-offs can be expressed as functions of the continuous random variable θ. To keep the matters simple, we assume the pay-offs to be linear functions of θ i.e.

$$F(a,\theta) = r + s\theta$$

can be assumed to be the linear pay-off function of the state θ corresponding to the action (decision) a where r and s are constants.

As we have seen in a two-action problem, there are only two actions. Let we denote these actions by a_1 and a_2. We have seen that the action a_1 will be optimal if

$$EF(a_1) > EF(a_2)$$

and action a_2 will be optimal if

$$EF(a_1) < EF(a_2)$$

The expected losses corresponding to these two actions are

$$EOL(a_1) = (s_2 - s_1)\int_{\theta_b}^{\infty} (\theta - \theta_b)\pi(\theta)d\theta$$

and

$$EOL(a_2) = (s_2 - s_1)\int_{\infty}^{\theta_b} (\theta_b - \theta)\pi(\theta)d\theta$$

where θ_b is the break-even value of θ and $\pi(\theta)$ is the prior distribution of θ.

Example 1: A producer wants to introduce a new variant of his product in the market. If he decides so, the fixed cost of this decision would be Rs. 1, 00,000 and every unit of the new variant will fetch a profit of Rs. 2.50 per unit. If he does not introduce the product, he will not earn or loose anything. The producer, however, is uncertain about the number of unit which he can sell in the market and associates his expectations with a normal distribution with mean 50,000 and standard deviation 25,000. His estimate is to sell 48,000 units. He has a network of 200 agents through whom he would sell his product. What should be the decision of the producer? How much do you think is he willing to spend to get exact information about his future sales?

In order to confirm his belief, the producer conducted a survey by selecting at random 25 of his 200 agents. On average every agent bought 280 units with a standard deviation of 100 per agent. Does this information make any effect on the decision taken by the producer?

Sol: Let θ be the random variable denoting the sale of the new variant per agent. Then

$$\theta \sim N\,(250,\,(125)^2) \qquad (50{,}000/200 \text{ and } 25{,}000/200)$$

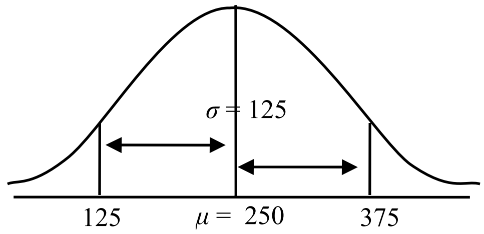

Fig. 3.4

We define the two actions of the producer, along with their pay-offs as follows:

d_1: Launch the new variant

$$P(d_1,\theta) = -1,00,000 + (2.50)200\theta$$

$$= -1,00,000 + 500\theta$$

d_2: Do not launch the new variant.

$$P(d_2,\theta) = 0$$

The break-even point is then given by

$$P(d_1,\theta) = P(d_2,\theta)$$

$$\Leftrightarrow \quad -1,00,000 + 500\theta = 0$$

$$\Leftrightarrow \quad \theta_b = 200 < 250 = \mu$$

As we had seen in chapter 2, the expected pay-offs (with prior mean 240 = 48000/200) are given as

$$EP(d_1) = -1,00,000 + (500)240 = Rs.\ 20,000$$

and

$$EP(d_1) = 0$$

The prior mean (250) is more than the break-even value (200). With this information, the producer should go for the new variant of his product. The expected profit of this decision will be Rs. 20,000.

The opportunity loss functions of the two decisions are given by

$$OL(d_1,\theta) \quad = \quad \begin{cases} (2.50)200(200-\theta) & \text{if } \theta < 200 \\ \\ 0 & \text{if } \theta \geq 200 \end{cases}$$

and

$$OL(d_2,\theta) \quad = \quad \begin{cases} (2.50)200(\theta-200) & \text{if } \theta > 200 \\ \\ 0 & \text{if } \theta \leq 200 \end{cases}$$

Then EVPI can be calculated as

$$\begin{aligned} EVPI &= \delta_2 \sigma L_N(D) \\ &= (500)(125)L_N\left(\frac{200-250}{125}\right) \\ &= (500)(125)L_N(0.40) \\ &= (500)(125)(0.2304) \\ &= 14400 \end{aligned}$$

This is the amount that the producer is willing to spend to get more exact information about his future sales in absence of any sample information.

Now, when the producer has collected sample information, the sample mean is $\overline{X} = 280$; $s.d. = 500$; and $n = 25$. Then the s.d. of the sample mean is

$$\sigma_{\overline{X}} \quad = \quad \frac{s.d.}{\sqrt{n}} \quad = \quad \frac{500}{5} \quad = \quad 100$$

Since sample mean is also above the prior mean, the producer should go ahead with the earlier decision.

For the posterior distribution the parameters are given as

$$\mu_1 \quad = \quad \frac{\dfrac{\mu}{\sigma^2}+\dfrac{\overline{X}}{\sigma_{\bar{x}}^2}}{\dfrac{1}{\sigma^2}+\dfrac{1}{\sigma_{\bar{x}}^2}} \quad = \quad \frac{\dfrac{250}{125^2}+\dfrac{280}{100^2}}{\dfrac{1}{125^2}+\dfrac{1}{100^2}} \quad = \quad \frac{250(10000)+280(15625)}{25625}$$

$$= \quad 268.29$$

$$\frac{1}{\sigma_1^2} \quad = \quad \frac{1}{\sigma^2}+\frac{1}{\sigma_{\bar{x}}^2} \quad = \quad \frac{1}{125^2}+\frac{1}{100^2} \quad = \quad \frac{25625}{156250000} \quad = \quad 0.000164$$

$$\Rightarrow \sigma_1^2 \quad = \quad 6097.56; \text{ and } \sigma_1 \quad = \quad 78.09$$

Now, for the posterior distribution, we calculate the expected profit and EVPI.

$$\begin{aligned} EP(d_1) &= -1,00,000 + (2.50)200(268.29) \\ &= -1,00,000 + 500(268.29) \\ &= Rs.\ 34145 \end{aligned}$$

and

$$\begin{aligned} EVPI &= \delta_2 \sigma L_N(D) \\ &= (500)(78.09) L_N\left(\frac{200-268.29}{78.09}\right) \\ &= (500)(125) L_N(0.88) \\ &= (500)(125) 0.0876 \\ &= 5475 \end{aligned}$$

In this case, the producer is willing to spend Rs. 5475 to get perfect information to substantiate his decision.

Now, we shall obtain the expected value of sample information (EVSI) and expected net gain from sampling (ENGS)

As we have seen the variance of the posterior distribution is given by

$$\frac{1}{\sigma_1^2} = \frac{1}{\sigma^2} + \frac{1}{\sigma_{\bar{X}}^2}$$

And the reduction in the variance as a consequence of adding more sample information is

$$\begin{aligned} \sigma^{*2} &= \sigma^2 - \sigma_1^2 \\ &= 125^2 - 6097.56 \\ &= 9527.44; \end{aligned}$$

$$\Rightarrow \sigma^* = \sqrt{9527.44} = 97.61$$

Then EVSI is calculated as

$$
\begin{aligned}
EVSI &= \delta_2 \sigma^* L_N(D) \\
&= (500)(97.61) L_N\left(\frac{200-268.29}{97.61}\right) \\
&= (500)(97.61) L_N(0.70) \\
&= (500)(97.61)(0.1398) \\
&= Rs.\ 6823
\end{aligned}
$$

Now, we calculate the expected net gain form sampling (ENGS).

If the cost of obtaining a sample of size n is C_n, then

$$ENGS = EVSI - C_n$$

Let the cost of sampling per unit is Rs. 20 and the fixed cost associated with the sampling process is Rs. 6000 then

$$
\begin{aligned}
C_n &= 6000 + 20n \\
&= 6000 + 20(25) \quad \text{for } n = 25 \\
&= 6500
\end{aligned}
$$

Then

$$
\begin{aligned}
ENGS &= 6823 - 6500 \\
&= Rs.\ 323
\end{aligned}
$$

Finally, we find out the optimal sample size which would ensure the maximum expected gain from sampling.

If we plot the cost function and the expected value of sample information on a graph then the point where the two curves intersect will be the point when the sample size will be optimal.

$$EVSI = \delta_2 \sigma^* L_N(D)$$

The following table represents EVSI for different values of n.

Table 3.19

n	$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$	$\sigma_1^2 = \frac{1}{\frac{1}{\sigma^2}+\frac{1}{\sigma_{\bar{X}}^2}}$	$\sigma^{*2} = \sigma^2 - \sigma_1^2$	$D = \frac{200-268.29}{\sigma^*}$	$EVSI = 500\sigma^* L_N(D)$
25	10000	6097.561	9527.439	0.699631	6828.0642
50	5000	3787.879	11837.12	0.627674	8746.0189
100	2500	2155.172	13469.83	0.588405	9990.5631
150	1666.667	1506.024	14118.98	0.574719	10464.263
200	1250	1157.407	14467.59	0.567752	10714.166
250	1000	939.8496	14685.15	0.563531	10868.597

Plotting EVSI curve and the cost line, we get the optimal sample range as follows:

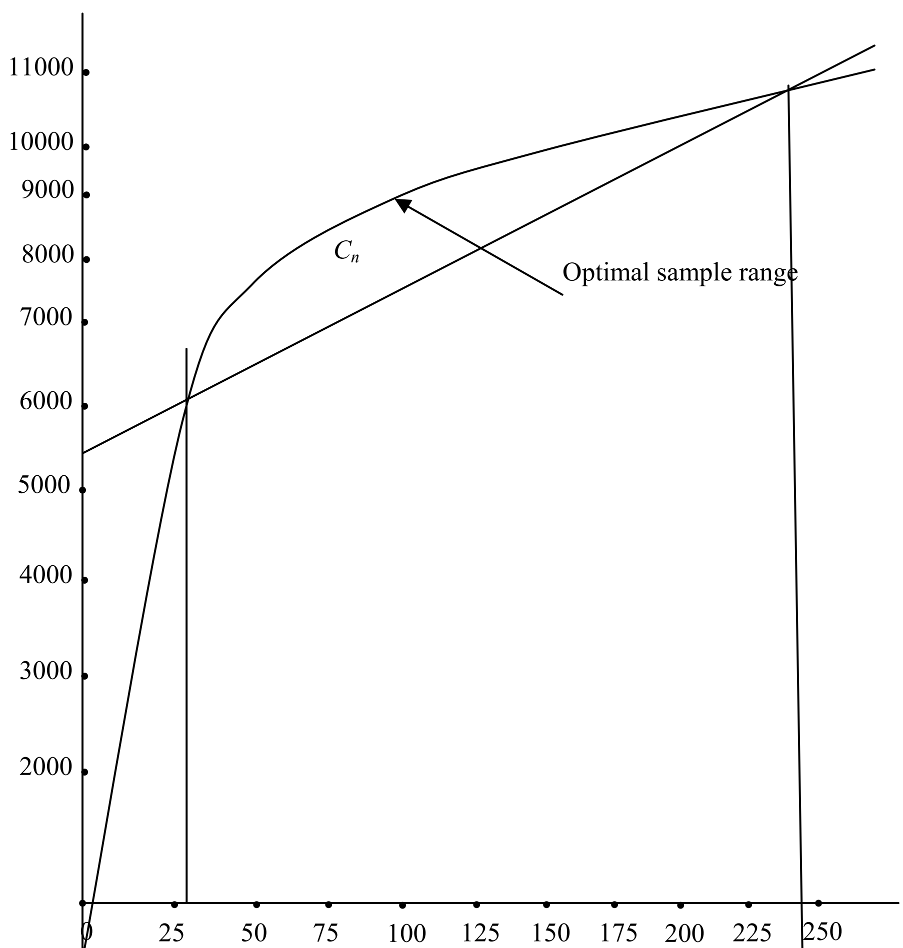

Fig. 3.5

The optimal sample range that can be used to get information is between $n = 30$ and $n = 245$. However, this range is very large. We can get an estimate of the optimal sample size is by considering ENGS.

Table 3.20

n	*EVSI*	C_n	ENGS
25	6828.0642	6500	328.064
50	8746.0189	7000	1746.019
100	9990.5631	8000	**1990.563**
150	10464.263	9000	1464.263
200	10714.166	10000	.714.166
250	10868.597	11000	-131.403

Plotting ENGS on the scale, we get the following graph:

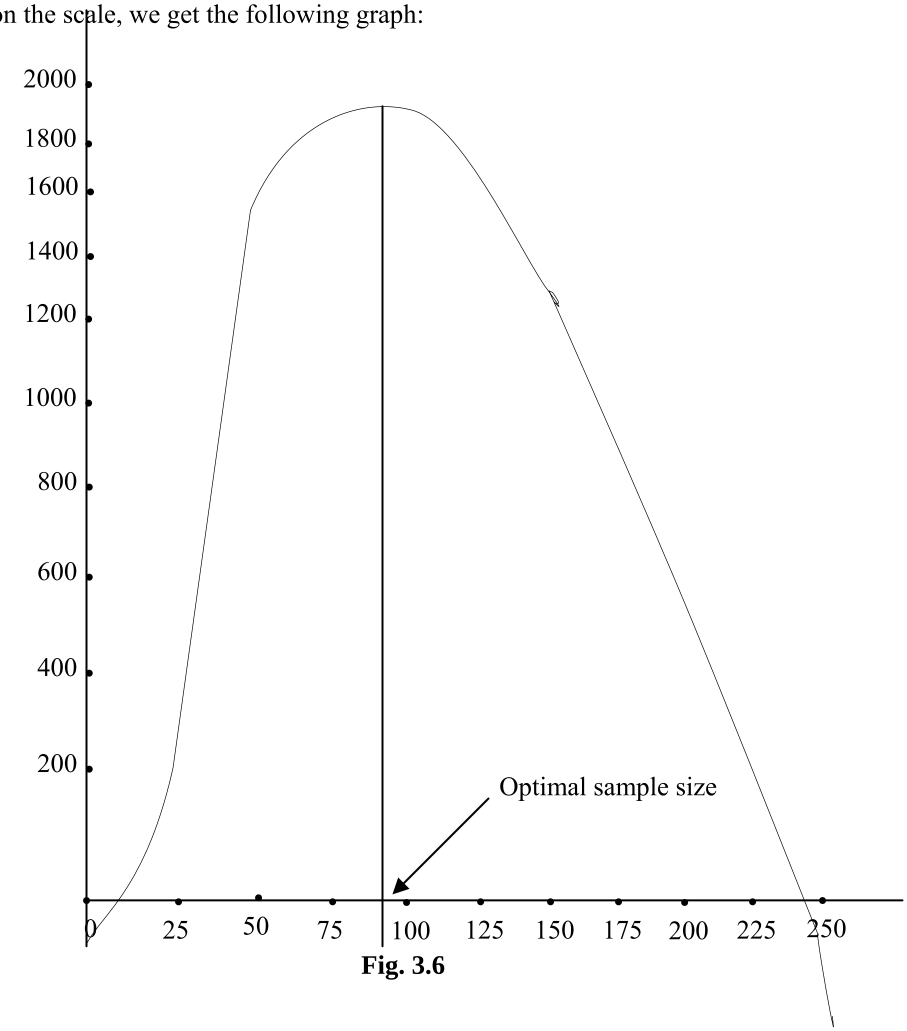

Fig. 3.6

A very small sample may not yield very reliable results and for very large samples, the cost of sample collection exceeds the expected value of the sample information. Thus a sample of size 90 to 100 will be an optimal sample.

Example 2: A production manager feels that the proportion p of defectives produced in the production process can be identifies with a beta distribution with parameters α = 2 and β =20. The production process is found to follow a Bernoulli distribution with two outcomes, a defective with probability p and a non-defective with probability 1- p. In order to confirm his belief about the nature of proportion of defective, he chooses a sample of size 20 and finds 1defective in the sample. Depending upon the outcomes of his findings, he can take any of the three possible actions:

(i) Leave the process as it is.

(ii) Adjust the process at a cost of Rs. 10,000 to reduce the proportion of defectives to one - fourth of what it is now.

(iii) Adjust the process at a cost of Rs. 100,000 to reduce the proportion of defectives to one - tenth of what it is now.

Further the cost of "correcting" a defective item is Rs. 50 whereas the profit of a non-defective unit is Rs. 20. At a time 5000 units may be produced by the system. What should be the decision of the manager?

(i) Before taking the sample;

(ii) After taking the sample.

Sol: For the prior distribution of p,

$$\alpha = 2 \text{ and } \beta = 20$$

$$\Rightarrow \text{Mean of the distribution} = \frac{\alpha}{\beta} = \frac{2}{20} = 0.1; \text{ and}$$

$$\text{Variance} = \frac{\alpha(\beta-\alpha)}{\beta^2(\beta+1)} = \frac{2(18)}{400(21)} = 0.0043$$

For the Bernoulli production process

$$P(\text{defective}) = p \text{ and}$$

$$P(\text{non-defective}) = 1 - p$$

For the posterior distribution of p which, again is a beta distribution

$$\alpha_1 = \alpha + k \quad \text{where } k \text{ is the number of defectives in the sample of size } n$$

$$= 2+1 = 3 \quad ; \text{and}$$

$$\beta_1 = \beta + n$$

$$= 20+20 = 40$$

$\Rightarrow$ Mean of the distribution $= \dfrac{\alpha_1}{\beta_1} = \dfrac{3}{40} = 0.075$; and

$$\text{Variance} = \frac{\alpha_1(\beta_1 - \alpha_1)}{\beta_1^2(\beta_1 + 1)} = \frac{3(37)}{1600(41)} = 0.00169$$

Now, expected number of defectives in a single production run is 5000 p.

Also, profit per unit is Rs. 20; and

The cost of "correcting" a defective is Rs. 50.

Define the three decisions as follows:

d_1: Leave the process as it is.

d_2: Adjust the process at a cost of Rs. 10,000

d_3: Adjust the process at a cost of Rs. 100,000

Then the pay-off function of the process may be written as

$$P(p,\ d) = 20(50000 - 50000p) - 50(50000p) - C(d)$$

$$= 1000000 - 3500000p - C(d); \quad \text{where}$$

$$C(d) = \begin{cases} 0; & \text{if } d = d_1 \\ 10000; & \text{if } d = d_2 \\ 50000; & \text{if } d = d_3 \end{cases}$$

Under prior distribution

$$EP(d_1) = 1000000 - 3500000(0.1)$$
$$= 650000$$
$$EP(d_2) = 1000000 - 3500000(0.025) - 10000$$
$$= 922500$$
$$EP(d_3) = 1000000 - 3500000(0.01) - 100000$$
$$= 965000$$

The optimal decision is to adjust the process at a cost of Rs. 100,000.

Under posterior distribution

$$EP(d_1) = 1000000 - 3500000(0.075)$$
$$= 737500$$
$$EP(d_2) = 1000000 - 3500000(0.025) - 10000$$
$$= 988125$$
$$EP(d_3) = 1000000 - 3500000(0.0075) - 100000$$
$$= 873750$$

The optimal decision is to adjust the process at a cost of Rs. 10,000.

Problems

1. Consider a prior normal distribution with parameters (80, $(15)^2$). A sample of size n is taken. What should be the minimum value of n so that the standard deviation of the posterior distribution does not exceed 4? If the mean and the standard deviation of the posterior distribution are 50 and 4, then what will be mean of the sample?

2. Refer to example 1 of the chapter. Find EVSI and ENSG for the following sample information:

   (i) $\overline{X} = 280;\quad s.d. = 500;\quad \text{and}\ \ n = 40.$

   (ii) $\overline{X} = 300;\quad s.d. = 500;\quad \text{and}\ \ n = 25.$

   (iii) $\overline{X} = 280;\quad s.d. = 750;\quad \text{and}\ \ n = 25.$

3. Refer to problem 1of chapter 2. The firm has decided to conduct an advanced trial of the drug when 2100 patients were treated with the drug. The drug was effective on 500 on them. Find the posterior mean and variance of the firm's estimates. If the cost of the trial was Rs. 2, 00,000, should the firm introduce the new drug. Find EVPI, EVSI and ENGS. What should be the optimal sample size?

4. A trainer is to train a group for high altitude mountaineering. In the past he has trained 20 such groups and has found the following success rates for the groups:

$$P(p) = \begin{cases} \dfrac{1}{5}\ ; & \text{for } p = 0.48 \\[2ex] \dfrac{1}{4}; & \text{for } p = 0.50 \\[2ex] \dfrac{1}{5}; & \text{for } p = 0.52 \\[2ex] \dfrac{3}{10}; & \text{for } p = 0.56 \\[2ex] \dfrac{1}{20}; & \text{if } p = 0.58 \end{cases}$$

where p is the proportion of those candidates who have completed the training successfully. However for this group of 100 candidates he is a bit unsure. So he selects 10 of them and trains. He finds that 4 of them have completed the training successfully. How should he modify his earlier belief about the success of the candidates? If the cost of training a candidate is Rs. 10,000 to him and he charges Rs. 14,000 per candidate, then what should be his decision?

5. A company is researching on a new anti- theft security system for which it estimates the success rates to be either 50%, or 60% or 80%. If the success rate is 50%, the security system will be termed as "poor", in case the success rate is 60%, the security system will be termed as "average" and for the success rate of 80%, the security system will be termed as "good". The probabilities of the three events are 0.25, 0.30 and 0.45. The trials have indicated that out of 20 times, the system was effective 14 times. Find the posterior probabilities that the system is poor, average or good.

6. A company researching in herbal drugs finds that not all the herbs it collects are useful for making drugs. In fact it has assigned a probability distribution to the proportion p of herbs which have been found to be useless:

$$P(p) = \begin{cases} 0.2\,; & \text{for } p = 0.02 \\ 0.7; & \text{for } p = 0.1 \\ 0.1; & \text{for } p = 0.15 \end{cases}$$

(a) In a sample of 20 herbs, 2 have been found to be useless. Obtain the posterior probabilities of the proportions of the useless herbs.

(b) Herbs are collected till the second useless herb is found and it turns out to be the 20th herb collected. Obtain the posterior probabilities of the proportions of the useless herbs.

7. Refer to problem 6(a) above. The company decides to take a second sample of size 15 and find 1 useless herb. How will you modify the posterior probabilities?

8. Refer to problem 4 above. If p is a continuous beta variable with parameters $\alpha = 2$ and $\beta = 20$ and the sample outcomes are according to a Bernoulli distribution, then what should be the decision of the company.

9. Consider the following information about a normal posterior distribution and about sample used in obtaining the posterior distribution:

$$\mu_1 = 50; \qquad \sigma_1 = 10$$

$$\overline{X_1} = 52; \qquad s.d. = 5; \qquad n = 5$$

Can the parameters of the prior distribution be determined?

10. In a play school, the authorities are interested in measuring the creative skills of the children at a scale of 1 to 20. Earlier experience shows that the distribution of the skill is normal with mean 12 and standard deviation 4. A test has been conducted to determine the skill but the test an inherent

error, which again is normally distributed with mean 0 and standard deviation 1. A kid chosen at random scores 14 points. What is the probability that the kid's actual skill is

(i) 10; (ii) 12; and (iii) 18 ?