CHAPTER 2

BAYESIAN STATISTICS

CONTENTS

1 INTRODUCTION

1.1 Data that is used in Reliability and Maintainability (R&M) analysis is usually observed data, and does not take into account how random variations can affect data collection, or indeed that information exists on what the outcome is most likely to be. The classical approach to R&M estimates was to generate probabilities through a measure of what is true, based purely on observed data. The data would then be used through various probability theories to predict the likelihood of an event happening or not (success or failure, for example). However this is rarely true to life, since any R & D tests prior to the trials will provide evidence as to what the outcome should be. Bayesian Statistics attempts to make use of this knowledge and incorporate it into the posterior estimates and confidence intervals after the trials have taken place.

2 GENERAL CONCEPTS

2.1 The real value and the observed value will never match exactly (in the real world), usually due to uncertainty; it could even be argued that certain probabilities are never exact and are simply a number generated from experiments. Naturally, this observed data could be incorrect, as the random nature of data collection has not been accounted for. This approach to generating assumptions from statistical data is governed primarily by fixed parameters; in other words, it is assumed that the probability of occurrence is not distributed randomly and is assumed to always be constant. However, the Bayesian approach to statistics treats these parameters as unobserved **random** variables, as opposed to values that are purely observed. These can be *discrete* (i.e. integer) or *continuous* (any positive value), and can vary in value for every single trial that is performed. Bayes Theorem is used to provide a **subjective** probability distribution for the probability that an event will occur in future trials, taking into account this random variation.

2.2 Assume that mutually exclusive events (trials) $X_1, X_2, \ldots, X_n$ occur, such that $X_1 \cup X_2 \cup \ldots \cup X_n = S$. This is shown graphically in Figure 1.

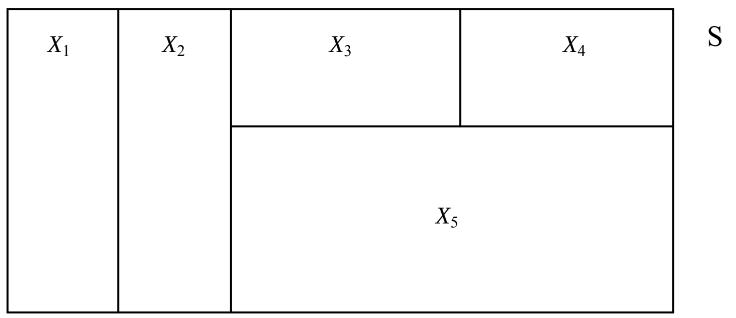


Figure 1: $X_1 \cup X_2 \cup \ldots \cup X_n = S$

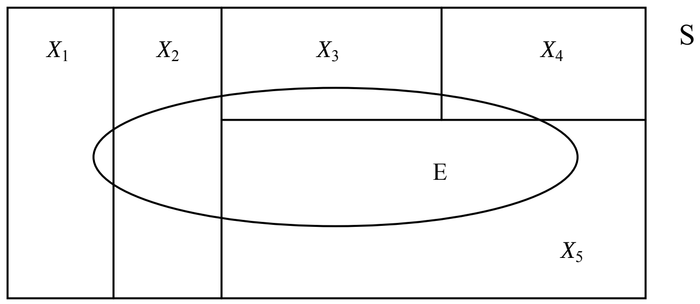


Figure 2: $E \subset S$

If X_1, X_2, …, X_n represents a partition of the sample space S, and $E \subset S$ ("E is within the set of S") represents an arbitrary event, as shown in Figure 2, then the theorem of total probability states that:

$$\mathbf{P}(E) = \mathbf{P}(X_1)\mathbf{P}(E \mid X_1) + \mathbf{P}(X_2)\mathbf{P}(E \mid X_2) + \ldots + \mathbf{P}(X_n)\mathbf{P}(E \mid X_n) \quad \text{Equation 1}$$

Hence, if the data X_1, X_2, …, X_n are known then the probability of event E can be found. Whilst the probability of event E cannot be directly observed, Bayes Theorem can use the partitioning events to compute the conditional probabilities. Bayes Theorem is based upon the same conditions of partitioning and events as the theorem of total probability, and is useful in computing the reverse probability of the type $P(X_i \mid E)$, for i=1,2,…,n. The reverse probability can be computed by dividing $P(X_i)P(E \mid X_i)$ by Equation 1 above. This constitutes the basic principle of Bayes Theorem.

2.3 Let us assume that a parameter exists, γ, that has an *unknown* value for its probability of occurrence (Rigdon Pg.146). For this example we shall say that γ is the tossing of a coin. Let us then assume that the random variables X_1, $X_2 \ldots X_n$ are observed that are determined by the unknown parameter γ; this could be the success rates of individual trials, or in this case 'heads' and 'tails' results from tossing the coin, γ. The joint probability density function is then:

$$\boldsymbol{f\,(x_1, x_2, \ldots, x_n \mid \gamma)} \quad \text{Equation 2}$$

2.3.1 Before observing the data x_1, x_2, … , x_n , the probability distribution for the situation γ has to be pre-determined that accurately reflects what our beliefs are on γ actually occurring. This information can be obtained from past data and records, and can be continuous or discrete. This is known as a **prior distribution**. Denoted as $p(\gamma)$, the prior distribution is based predominantly upon the objective or subjective views of the user, and depending upon the accuracy of these beliefs, will greatly determine what outcome is obtained by using Bayes Theorem. Hence the results that are obtained from Bayes Theorem must be respected in terms of the priors that are given.

2.3.2 Once data x_1, x_2, … , x_n is collected (i.e. trials), then the **posterior distribution** can be ascertained. This is the distribution of the unknown parameter γ

given the observed data $x_1, x_2, \dots, x_n$. The posterior distribution is calculated using Bayes Theorem and represents how the probability distribution has altered given the trials, taking into account the prior distributions. It could be said that these trials are attempting to 'push' the level of success higher (or indeed lower) in the posterior distribution, hence strengthening the belief of a success and thus a hypothesis. The example towards the end of this subsection will clarify the above.

2.4 Engineers often make decisions regarding probabilities using limited information. This takes the form of *objectivity* (i.e. observed data from experimental 'trials') or *subjectivity* (i.e. personal experience or opinion). The majority of decisions encapsulating prior probabilities constitute both objective and subjective information, even though the posterior probability in practical terms cannot be determined. Understanding this phenomenon is critical in appreciating Bayes Theorem. A posterior can be calculated using objective and subjective prior probabilities on the partitioning events and conditional probabilities, even though the posterior cannot be practically determined. The application of Bayes Theorem thus serves to determine this posterior probability from the priors, and in this way hypotheses can be created. However, it is critical to realise that Bayes Theorem is totally dependent upon the prior probability distribution that is supplied; naturally, as this is not an *exact* value (as stated earlier), the outcome through Bayes for future likelihoods is totally dependent upon these priors, and so the results must be viewed with caution. An over-confident prior probability distribution could give a future "success" probability that is in excess of what it actually *should* be, and vice versa. The belief that the posterior probability distribution is correct should always be treated with caution. Good examples of such are lawyers in court, who, through supplying inaccurate prior information, are able to 'deduce' the outcome that *they* want, and one that is not necessarily the *actual* answer.

2.5 In Figure 2, the sample space S is partitioned into the mutually exclusive events $X_1, X_2, \dots, X_n$ such that $X_1 \cup X_2 \cup \dots \cup X_n = S$. If this is the case then:

$$\boldsymbol{E = (E \cap X_1) \cup (E \cap X_2) \cup (E \cap X_3) \cup \dots\dots \cup (E \cap X_n)} \qquad \text{Equation 3}$$

Using the addition rule it is seen that:

$$\mathbf{P}(\boldsymbol{E}) = (\boldsymbol{E \cap X_1}) + (\boldsymbol{E \cap X_2}) + (\boldsymbol{E \cap X_3}) + \dots\dots + (\boldsymbol{E \cap X_n}) \qquad \text{Equation 4}$$

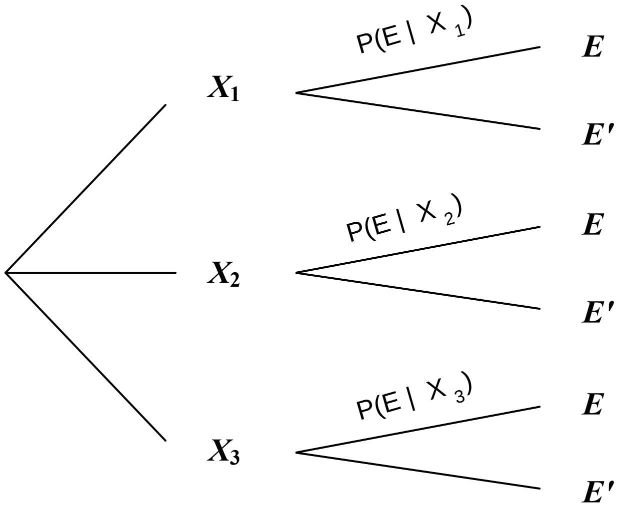
And using the multiplication rule:

$$\mathbf{P}(\boldsymbol{E \cap X_i}) = \mathbf{P}(\boldsymbol{X_i})\mathbf{P}(\boldsymbol{E} \mid \boldsymbol{X_i}) \qquad \textbf{for i=1,2,3,\dots..,n} \qquad \text{Equation 5}$$

Substituting Equations 5 and 4 reveals the 'total probability':

$$\mathbf{P}(\boldsymbol{E}) = \sum_{i=1}^{n} \mathbf{P}(\boldsymbol{X_i})\mathbf{P}(\boldsymbol{E} \mid \boldsymbol{X_i}) \qquad \text{Equation 6}$$

Equation 6 is the total probability rule for finding P(E) in terms of the probabilities of mutually exclusive events X_i to X_n and the corresponding conditional probabilities. This can be shown as a tree diagram, Figure 3. It is assumed that X_n in Figure 3 is represented by the three mutually exclusive events X_1, X_2, X_3, and that E' stands for "E not occurring".



$P(E \cap X_1) = P(X_1)P(E \mid X_1)$

$P(E' \cap X_1) = P(X_1)P(E' \mid X_1)$

$P(E \cap X_2) = P(X_2)P(E \mid X_2)$

$P(E' \cap X_2) = P(X_2)P(E' \mid X_2)$

$P(E \cap X_3) = P(X_3)P(E \mid X_3)$

$P(E' \cap X_3) = P(X_3)P(E' \mid X_3)$

Figure 3: Tree diagram representing Equation 6 for X_1, X_2, X_3

The expression "$P(E \mid X_1)$" is read as "the probability that E will happen, *given* that X_1 has already happened).

2.6 It has been shown how to compute $P(E \mid X_i)$, but not $P(X_i \mid E)$. In other words, how can we find out what the probability is of X_i happening given that E happened (or indeed E')? As stated earlier, to work out the reverse probabilities requires the use of Bayes Theorem, which is now derived as follows:

$$\mathbf{P(E \cap X_i) = P(X_i)P(E \mid X_i) = P(E)P(X_i \mid E)} \qquad \text{Equation 7}$$

And so if we imagine X_r is a specific event of one the events X_i, then rearrangement of Equations 6 and 7 reveals:

$$\mathbf{P(X_r \mid E) = \frac{P(X_r)P(E \mid X_r)}{\sum_{i=1}^{n} P(X_i)P(E \mid X_i)}} \qquad \textbf{for r=1,2,3,…..,n}$$

$$\mathbf{= \frac{P(X_r)P(E \mid X_r)}{P(E)}} \qquad \textbf{"Bayes Theorem"} \qquad \text{Equation 8}$$

Hence, by calculating the probability of $P(E \mid X_r)$ from the tree diagram and using the processes mentioned earlier, we have been able to calculate the reverse probability using Equation 8. This is Bayes Theorem. $P(E)$ is simply the total probability that can be calculated from the tree diagram.

2.7 It is also worth noting that Equation 8 gives rise to further formulae:

$$\mathbf{P}(X_r) = \mathbf{P}(X_r \mid E).\ \mathbf{P}(E) + \mathbf{P}(X_r \mid E').\ \mathbf{P}(E') \qquad \text{Equation 9}$$

$$\mathbf{P}(E) = \mathbf{P}(E \mid X_r).\ \mathbf{P}(X_r) + \mathbf{P}(E \mid X_r').\ \mathbf{P}(X_r') \qquad \text{Equation 10}$$

3 APPLICATION TO R & M

3.1 In classical statistics, parameter estimation and confidence interval calculations (see Part D Chapter 7) are conducted on the basis that, prior to the sample measurements, nothing whatever is known about the outcome. In practice this is rarely true. For example, in demonstration trials, if a comprehensive Reliability programme has been conducted during the R & D phase, one probably has some idea of the likely outcome of the trials.

3.1.1 Bayesian statistics has great intuitive appeal, especially to the engineer. It seems on the face of it absurd to ignore the fact that one has prior knowledge from R & D testing when conducting demonstration trials. One might, for example, feel 95% sure that a new missile will enter its demonstration or evaluation trials with a reliability (probability of individual firing success) greater than 60%. Therefore, why not use this knowledge to improve the estimate of Reliability from the trials and narrow the confidence intervals that would be obtained by classical methods?

3.1.2 Using prior information can enable demonstration tests to a specified level of confidence to be shorter than would otherwise be necessary (assuming that the item being assessed actually has a reliability which exceeds the target), thus saving time and money.

3.1.3 Bayesian statistics is equally applicable to Maintainability estimates; the principle of using prior information to obtain a posterior estimate can be used to calculate preventive maintenance schedules, for example. The theory in this chapter is such that Bayesian statistics can be used in a whole array of R&M functions that involve prior information. Obtaining posterior probabilities on equipments' reliability is necessary for effective R&M in every way. Since R&M govern when an item will be ready for service, Bayesian statistics are also effective in Availability estimates. Example 1 below concerns diagnostics, which itself is an important R&M function.

3.2 **Example 1**: *Equipment is in service and has integral Built In Test Equipment (BITE) diagnostics. The BITE is 90% effective at diagnosing a system fault correctly, displaying the message "FAULT" when this is so. However, the BITE contains a false error indication rate of 5%; that is, there is a 0.05 probability that equipment is diagnosed as faulty, given that it is actually healthy. At any point during a mission, past data and records have shown that there is an 8% chance that the equipment will actually be faulty.*

a) *What is the probability that equipment selected at random will read "FAULT" in the BITE system?*

b) If the equipment is diagnosed with the "FAULT" message, what is the probability that the equipment will actually be faulty?

3.2.1 General: Whilst a) and b) look very similar situations, there is a significant difference in their wording that greatly alters how the calculations should proceed. A tree diagram provides a diagrammatic version of the problem:

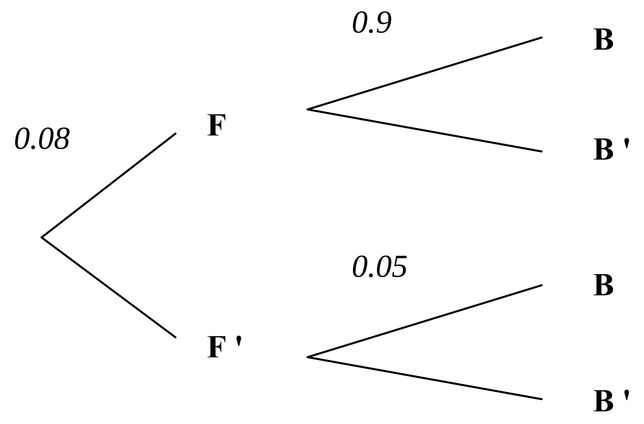


Figure 4: Tree diagram for Example 1

F = Faulty equipment and B = BITE reading "FAULT". Hence F' = Non-faulty equipment and B' = BITE not reading "FAULT".

Thus, from the given information: $P(F) = 0.08$ $P(B|F) = 0.9$ $P(B|F') = 0.05$ and this is shown in Figure 4.

3.2.2 Solution a) If we consider the above tree diagram, situation a) is not dependent upon other information; it is simply a case of adding up the values on the arms on the tree diagram for "FAULT" by multiplication. In this way, it is worth placing the value of P(F) on that arm, and thus P(F'), (1-0.08 = 0.92).

The same can then be applied for the arms of B and B'.

From here, the same process as used to formulate Equation 6 can be used; that is, to calculate the total probability for $P(B)$ it is required to multiply the branches of the tree for B. Thus the working is:

$$P(F)P(B|F) = 0.08 \times 0.9 = 0.072$$

$$P(F')P(B|F') = 0.92 \times 0.05 = 0.046$$

And thus $P(B) = 0.072 + 0.046 = 0.118$

Therefore, there is an 11.8% chance that a system selected at random will read "FAULT" in the BITE, regardless of whether or not the equipment is faulty.

3.2.3 Solution b) It is required to find the posterior probability (actually a fault), **given that** something has happened prior to it ("FAULT" message), denoted as $P(F|B)$.

It was calculated that the probability of a "FAULT" message displayed, $P(B)$, is 0.118. Using this information, Bayes Theory can be used to calculate $P(F|B)$, seeing as the value of $P(B|F)$ is already known. Likewise, this equation can always be rearranged to ascertain the required information. Notice how the process used in part a) cannot be used in part b), since "F" comes *before* "B" in the tree diagram. It is not possible to read $P(F|B)$ directly off the tree as it requires the *reverse* probability of $P(B|F)$, which demands Bayes Theory.

Substituting the calculated values into Equation 8, and using equations 9 and 10 to say that P(B) = :

$$\mathbf{P}(\boldsymbol{X}_\mathrm{r} \mid \boldsymbol{E}) = \frac{\mathbf{P}(\boldsymbol{X}_r)\mathbf{P}(\boldsymbol{E} \mid \boldsymbol{X}_r)}{\mathbf{P}(\boldsymbol{E})}$$

$$\longrightarrow \qquad \mathbf{P}(\boldsymbol{F} \mid \boldsymbol{B}) = \frac{\mathbf{P}(\boldsymbol{F})\mathbf{P}(\boldsymbol{B} \mid \boldsymbol{F})}{\mathbf{P}(\boldsymbol{B})}$$

$$= \frac{\mathbf{0.072}}{\mathbf{0.118}}$$

$$= \underline{\mathbf{0.61}}$$

Thus, there is a 61% chance that if the BITE displays "FAULT", the system will actually be faulty. This answer is, of course, totally dependent upon what prior information is used.

3.2.4 Example 1 was a simple walk-through of how a single prior reliability can be used to work out the posterior estimate. Yet what if there is more than one probability for the prior? Let us consider the following example, which covers aspects of R&M.

3.3 **Example 2:** *Past experience has shown that turbine blades, over a given period of service, are supplied in batches that are either of high reliability (0.99) or low reliability (0.9). A batch is received, a sample of 10 blades is tested under in-service conditions, and one fails. What is the probability that this is a high reliability batch, if it is known that these account for 25% of the total?*

3.3.1 Solution: In this case, the prior information is that 25% of batches have 99% reliability and 75% have 90% reliability.

Let event '*A*' be the occurrence of a high reliability batch.

Let event '*B*' be the test result 'one failure in ten'.

Then the answer to the problem is P(*A*|*B*).

3.3.2 To compute this using Bayes Theorem (Equation 8), P(*A*), P(*B*) and P(*B*|*A*) must be calculated.

$$P(A) = 0.25$$

P(*B*|*A*) = probability of getting 1 defective item in a sample of 10 items taken from a batch with reliability 0.99

This probability may be obtained from the Binomial Distribution (Part D Chapter 3) as:

$$_{10}C_1 0.99^9 \text{ x } 0.01 = 0.0914$$

and thus:

$$P(B|A) = 0.0914$$

3.3.3 The evaluation of P(B) is less straightforward, and use must be made of Equation 10. Replacing X_r and E with A and B respectively from Equation 10:

$$P(A) = 0.25$$

$$P(A') = 0.75$$

$$P(B|A) = 0.0914 \quad \text{(from above)}$$

P(*B*|*A'*)= probability of getting one failure in a sample of ten taken from a low reliability batch (since "A" is "not a high reliability batch"; that is, A is "a low reliability batch").

Using the Binomial Distribution again yields:

$$P(B|A') = {}_{10}C_1 0.09^9 \text{ x } 0.1 = 0.3874$$

$$\therefore \quad P(B) = 0.0914 \text{ x } 0.25 + 0.3874 \text{ x } 0.75 \text{ (from Equation 10)}$$

$$= 0.3134$$

3.3.4 It is now possible to evaluate Equation 8 since all the probabilities are known.

$$\therefore \quad P(A|B) = \frac{0.0914 \times 0.25}{0.3134} = 0.073$$

3.3.5 Thus the probability that the batch was a high reliability one is **<u>0.073</u>**.

4 CONTINUOUS PRIOR DISTRIBUTIONS

4.1 In Example 2 the 'prior' consisted of two batch reliabilities, with known probabilities of occurrence. One could easily conceive of priors with 3 possibilities for the same example, e.g. batches with reliabilities of .9, .95 and .99, with known probabilities of occurrence 0.1, 0.4 and 0.5 say. In general, there could be *n* possibilities. If *n* becomes very large, one can conceive of a prior which is, to all intents and purposes, a continuous function. Similarly, the two state posterior probability function (given in the last sentence of Example 2) can be generalised to a continuous posterior function when the prior is continuous.

4.1.1 These continuous functions are in fact Probability Density Functions (PDF), which can be expressed as f(p). In the notation of the example, there is now a PDF in which a batch may have any reliability between 0 and 1, and where the probability that the batch reliability lies between p and p δ p is f(p) δ p. The process embodied in Bayes Theorem (Equation 8) therefore is:

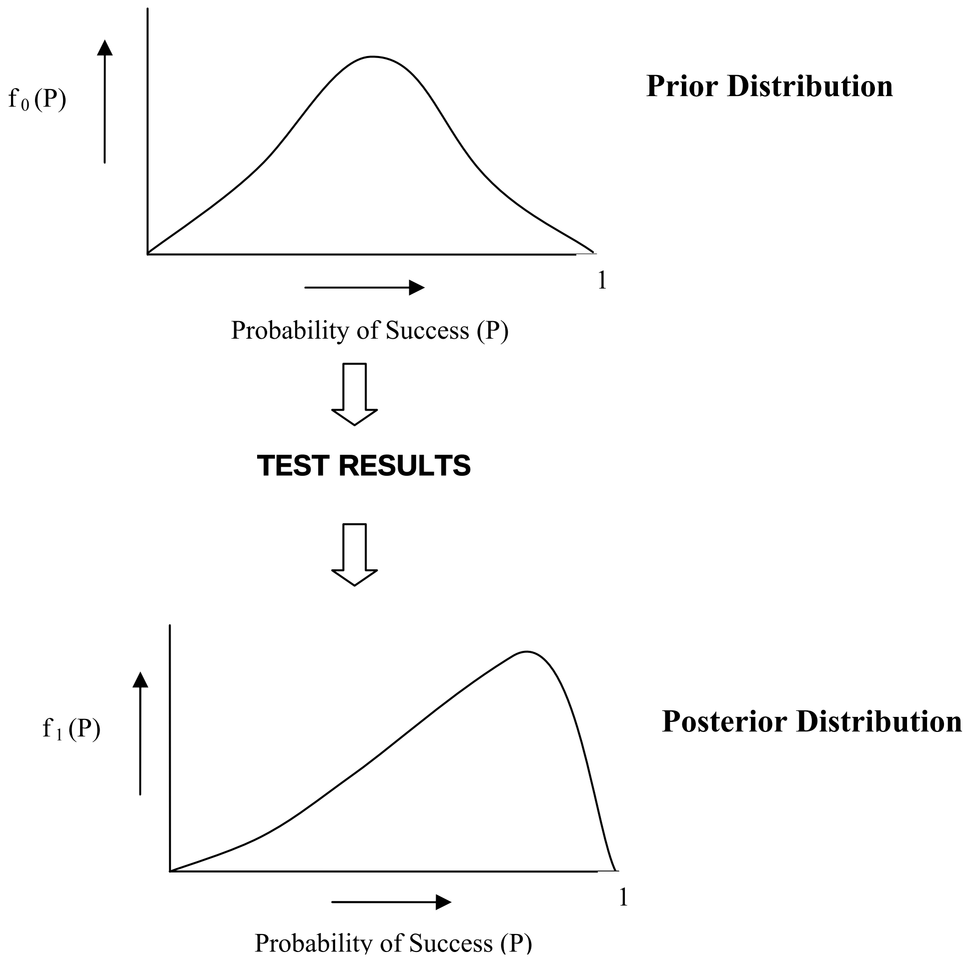


Figure 5: Bayesian Method

4.2 So far this section has outlined the principle of Bayesian statistics and its application in R&M. No attempt has been made here to introduce the subject in a rigorous mathematical fashion. The basic concepts have been described as simply as possible.

4.2.1 Particular techniques are applied to Reliability and failure rate assessment in the following sections, with illustrative examples embedded.

5 APPLICATION TO SUCCESS/FAILURE TYPE TRIALS

5.1 The method of analysis will first be illustrated by means of a simple example, and then generalised by induction.

5.2 Analysis Based on a Single Test Result

5.2.1 Suppose the prior distribution for probability of success is $f(p)_0$. What is the posterior distribution, $f_1(p)$, for probability of success given that 1 trial was conducted and the item operated successfully?

5.2.2 Let event A be ‘the true value of probability of success lies between p and p+δ p’, (δ p very small). Let event B be ‘the single trial is successful’.

Then

$$P(A|B) = f_1(p)\,\delta p \text{ (this is the definition of } f_1(p))$$
$$P(B|A) = p$$
$$P(A) = f_0(p)\,\delta p$$
$$P(B) = \text{Expected value of p from the prior distribution, i.e. } \int_0^1 pf_0(p)dp \text{ (see Part D Chapter 3)}$$

5.2.3 Thus from Equation 8:

$$f_1(p) = \frac{pf_0(p)}{\int_0^1 pf_0(p)dp} \qquad \text{Equation 11}$$

5.2.4 To proceed further it is necessary to assume a distribution for $f_0(p)$. For reasons that will become clear in what follows, the **Beta** distribution is chosen. The Beta distribution is defined by two parameters, V and W, and its PDF (f(p)) is:

$$f(p) = \frac{\Gamma(V+W)}{\Gamma(V)\Gamma(W)} p^{V-1}(1-p)^{W-1} \text{ where } \Gamma(X) = \text{Gamma function}$$

In this application V and W are usually integers, and since $\Gamma(X) = (X-1)!$ for X integer:

$$f(p) = \frac{(V+W-1)!}{(V-1)!(W-1)!} p^{V-1}(1-p)^{W-1} \qquad \text{Equation 12}$$

Also, the mean of a Beta distribution, which is the denominator of Equation 11, is V/(V+W).

5.2.5 Substituting these expressions in Equation 11 yields:

$$f_1(p) \quad = \quad p\frac{(V+W-1)!}{(V-1)!(W-1)!}p^{V-1}(1-p)^{W-1}\frac{(V+W)}{V}$$

$$\therefore \qquad f_1(p) \quad = \quad \frac{\{(V+1)+W-1\}!}{\{(V+1)-1\}!(W-1)!}p^{(V+1)-1}(1-p)^{W-1}$$

That is, if $f_0(p)$ is a Beta distribution with parameters V and W, $f_1(p)$ is also a Beta distribution, in which the V value of the prior has increased by 1.

5.2.6 Had the test result in paragraph 5.2.1 been a failure, a similar argument would have resulted in $f_1(p)$ being a Beta distribution in which W was increased by 1.

5.3 Generalisation to S Successes in S+F Trials

5.3.1 The foregoing can be extended to the general case by induction. If one thinks of the S+F trials occurring sequentially, then the posterior Beta distribution of the i^{th} trial can be thought of as the prior distribution for the $i+1^{th}$ trial. Since at each trial either V or W is increased by 1 depending on whether the result was a success or failure respectively, at the end of S+F trials we will have a posterior distribution where:

$$V \Rightarrow V+S$$
$$W \Rightarrow W+F$$

where V and W were the parameters of the prior.

5.3.2 It is the fact that this very simple modification to the prior, by the test results, yields the posterior distribution, which commends the Beta distribution in this application.

5.3.3 The expected or mean value of the posterior distribution is the reliability estimate after the trials, i.e.:

$$\frac{V+S}{V+W+S+F} \qquad \text{Equation 13}$$

The posterior *un*reliability estimate is therefore:

$$\frac{W+F}{V+W+S+F}$$

5.4 Quantifying the Prior

5.4.1 It is a common misapprehension that, since one adds successes to V and failures to W, V and W could be thought of as the successes and failures of a 'notional' prior trial. In fact, this is not the case, although it is approximately true for large V and W. The reason is that the Beta distribution is of interest, not the Binomial distribution. For low values of V and W the distributions differ quite considerably, as can be seen in Figure 6.

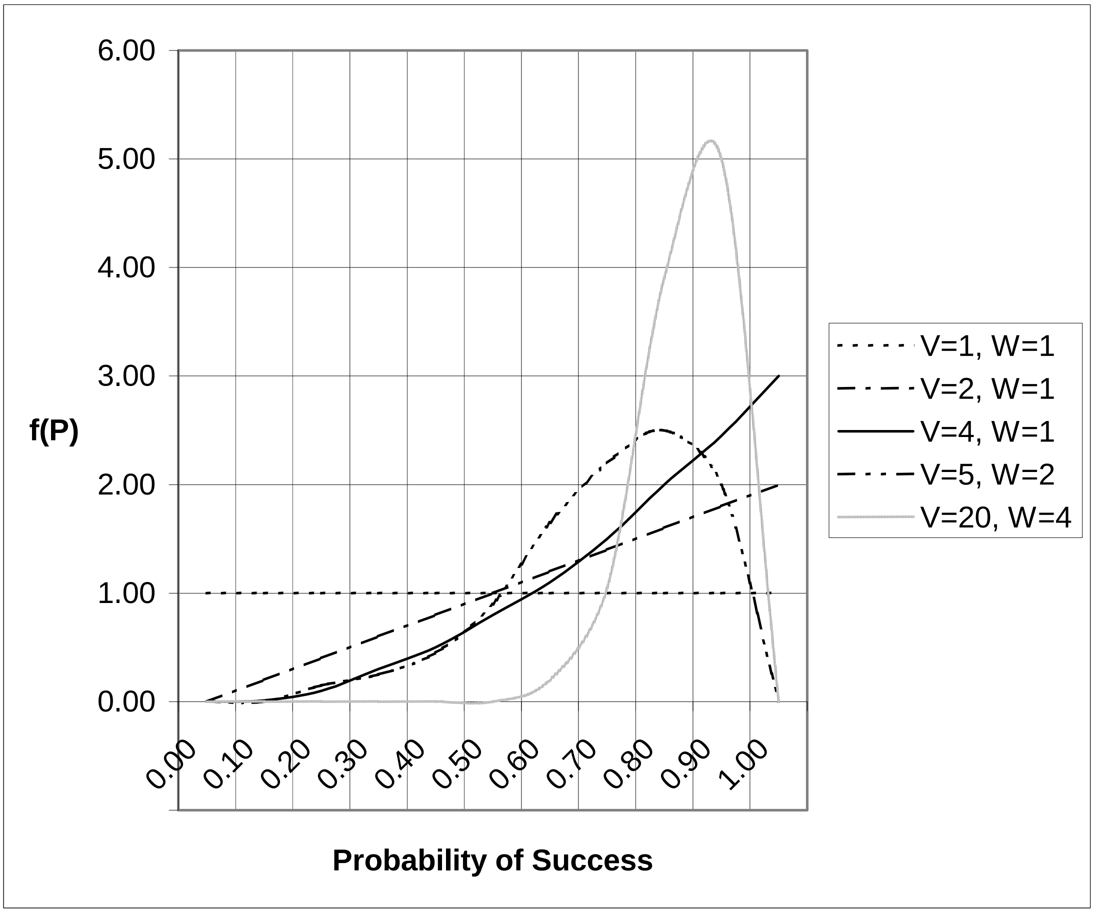


Figure 6: PDF of Beta Distribution for various values of V and W

5.4.2 The easiest way to recognise that V and W should not strictly be thought of as the results of a 'notional' prior trial is to observe that V = 1, W = 1 give the uniform distribution; that is, all values of p between 0 and 1 are equally likely. This would clearly not be the 'classical' or intuitive conclusion drawn from two trials, one of which failed. In fact, the prior V = 1, W = 1 is of special interest because it is the uniform prior. It can be considered to represent 'prior ignorance'. The posterior estimate of reliability (R) with this prior is:

$$R = \frac{S+1}{S+F+2} \qquad \text{(from Equation 13)}$$

This compares with the 'classical' estimate of R:

$$R = \frac{S}{S+F}$$

'Bayesians' tend to use the former expression for R when they have no strong prior information. Apart from philosophical considerations, it is claimed to be intuitively more reasonable, since estimates of R which are 0 or 1 are avoided.

5.4.3 The choice of prior in other situations involves looking at tables of the Beta distribution to obtain values for V and W which roughly meet the requirement. It should be noted that a one-sided prior, such as 'the probability that the reliability exceeds 60% is 0.90', is not strictly sufficient, since this condition could be met by either a weak or strong prior, as in Figure 7. (Each area to the left of the black line = 10% of PDF area.) When a one-sided prior is unavoidable, choose the smallest value of V to meet the condition as this will result in the weakest prior.

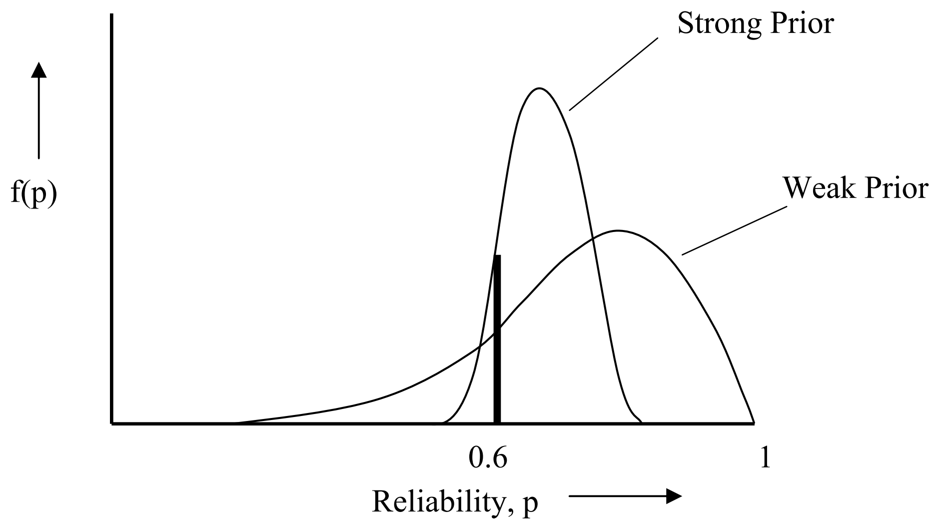


Figure 7: Possible Prior Distribution

In order to assist in quantifying the prior, tables of the Beta distribution are included in Part D Chapter 3*. Their use is illustrated by means of the following, Example 3.

5.5 **Example 3:** *On the basis of past experience there is estimated to be a probability of 0.9 that the reliability (R) of a hose seal lies between 0.6 and 0.9. (This may be interpreted as a 5% chance that R<0.6 and a 5% chance that R>0.9.) In a*

* To be included in a future amendment.

demonstration trial 15 seals are tested and 7 fail. What is the estimated reliability of the seal and the 80% CI?

5.5.1 Solution: The prior is expressed in terms of a 90% confidence statement, therefore to obtain the appropriate prior Beta Distribution, examine the Beta table (Part D Chapter 3, Table…)* for the lower 5% points (each 'tail' = 5%).

Firstly, look in the table to identify the values of V and W which will give a lower limit of ~0.6. These run diagonally down the table, taking in such values as V = 6, W = 1; V = 9, W = 2; etc.

Now identify values of V and W which will give a 5% upper tail above 0.9. To do this make use of the fact that the Beta distribution reflects about the line p = 0.5 when V and W are reversed. Thus, to identify such values of V and W, look in the table for values of 1-0.9 (=0.1), but reverse the V and W notation in the table. Again, this provides a 'line' across the table taking in such values (before V, W reversal) as V = 2, W = 3; V = 3, W = 6; etc. It then remains to match these two sets of pairs as closely as possible to obtain the V_0 and W_0 values for the prior.

In this case reasonable values are $V_0 = 15$, $W_0 = 5$, since this gives a 5% tails to the Beta distribution at p = 0.58 and p = 0.58 and p = 0.11 (=1-0.89).

In this example, S = 8, F = 7; therefore, from equation (9), the posterior reliability estimate is:

$$R = \frac{15+8}{5+15+7+8} = 0.657$$

To obtain the 80% CI look at the Table of lower 10% points for the Beta distribution (Part D Chapter 3, Table…)*, with $V_1 = V_0+S$, $W_1 = W_0+F$ (i.e. $V_1 + 23$, $W_1 = 12$) and also with $V_1 = 12$, $W_1 = 23$; this giving the lower, and 1 minus the upper, confidence limits respectively on reliability. Where necessary, interpolation must be used.

The 80% CI obtained from Table… is:

0.552 to 1-0.240 on the reliability;

that is: 0.552 to 0.760, about the reliability estimate of 0.657.

This compares with the classical 80% CI from Binomial Tables for 8 successes in 15 trials of 0.34 to 0.72, with a reliability estimate of 8/15 = 0.533.

5.5.2 The way in which demonstration trials can be shortened is observed by noting in the above example that using the Bayesian CI it can be said that there is a probability of 0.9 that the reliability exceeds 55.2%. With the classical approach it

* To be included in a future amendment.

can only be said that the lower one-sided 90% confidence limit is 34%. In effect, the prior has raised the lower confidence limit by about 20%.

6 APPLICATION TO CONSTANT FAILURE RATE ESTIMATION

6.1 General Theory

6.1.1 An approach similar to section 5 can be adopted here, except that instead of using the Beta distribution the **Gamma** distribution is used.

6.1.2 The Gamma distribution is defined as:

$$\text{PDF } (f(\lambda)) = \frac{\left[\frac{\lambda}{b}\right]^{c-1}\left[\exp(-\frac{\lambda}{b})\right]}{b\Gamma(c)} \qquad \text{Equation 14}$$

The parameters defining the distribution are b and c. For the Gamma distribution:

$$\text{Mean} = bc \qquad \text{Equation 15}$$

$$\text{Variance} = b^2c \qquad \text{Equation 16}$$

In this application it is simpler to constrain c to be an integer ≥ 1.

6.1.3 It is not proposed to derive here the relationship between the posterior and prior distributions. Suffice it to say that, if the prior Gamma distribution has parameters b_0 and c_0, and in a trial f failures are observed in time t, then application of Bayes theorem results in a posterior Gamma distribution with parameters b_1 and c_1, where:

$$\frac{1}{b_1} = \frac{1}{b_0} + t \qquad \text{Equation 17}$$

$$c_1 = c_0 + f \qquad \text{Equation 18}$$

6.1.4 Thus, the posterior estimate of failure rate is:

$$b_1c_1$$

and the CI about this estimate must be obtained using the properties of the Gamma distribution. Tables of the Gamma distribution are rarely provided, because it is closely related to the χ^2 distribution, and can be obtained from χ^2 tables. Denoting the Gamma distribution by Γ, the relationship is:

$$\Gamma : b, c \; = \; \frac{b}{2} \cdot \chi^2_{2c} \qquad \text{Equation 19}$$

The use of Equation 19, and the whole Bayesian approach to failure rate estimation, is illustrated in Example 4.

6.1.5 As in section 5, the quantification of the prior is the crucial factor in the process. One approach is to formulate it as: 'The mean of the prior is θ (prior failure rate estimate), with a standard error of σ'. The advantage of this is that equations 15 and 16 can then be used to obtain b_0 and c_0 as follows:

$$\theta = b_0 c_0$$

$$\sigma^2 = b_0^2 c_0$$

$$\therefore \qquad b_0 = \frac{\sigma^2}{\theta} \qquad \text{Equation 20}$$

$$c_0 = \frac{\theta^2}{\sigma^2} \qquad \text{Equation 21}$$

6.1.6 It is interesting to note that failure rate estimation, using Bayesian statistics, illustrates a fundamental difficulty in the subject, namely quantifying prior ignorance. It was coped with in section 5 by having a uniform prior, but this is not possible here, since failure rate has an infinite range (0 to ∞). In practice, an 'improper' uniform prior distribution is often used for λ, where $b_0 = \infty$ and $c_0 = 1$.

6.2 Example 4: *From R & D experience it is estimated that the failure rate of light bulbs is 0.025 failures/hour, with a standard error of 0.01. The bulbs undergo a demonstration trial in which 5 failures occur in 500 hours of testing. What is the posterior estimate of failure rate, and the 90% CI?*

6.2.1 Solution: In the notation of paragraph 6.1.5:

$\theta \quad = \quad 0.025$

$\sigma \quad = \quad 0.01$

From equations 20 and 21:

$b_0 \quad = \quad 0.004$

$c_0 \quad = \quad 6$ (to the nearest integer)

In the notation of equations 17 and 18:

$$t = 500$$

$$f = 5$$

$$\therefore \quad \frac{1}{b_1} = \frac{1}{0.004} + 500$$

$$\therefore \quad b_1 = \frac{1}{750} = 0.0013$$

$$c_1 = 6 + 5 = 11$$

Therefore the posterior failure rate estimate is:

$$b_1 c_1 = \frac{11}{750} = 0.0147$$

To obtain the 90% CI, use Equation 19. The CI is:

Lower limit: $\frac{b_1}{2}\chi^2_{95\%,2c_1}$

Upper limit: $\frac{b1}{2}\chi^2_{5\%,2c_1}$

since the two 'tails' of the PDF are each 5%.

This yields:

Lower limit: $\left[\frac{0.0013}{2}\right]33.9 = 0.0221$

Upper limit: $\left[\frac{0.0013}{2}\right]12.3 = 0.0080$

This compares with the 'classical' estimates of:

$$\lambda = 5/500 = 0.01$$

90% CI (see Part D Chapter 7):

Lower limit: 0.0227

Upper limit: 0.00392

7 CAUTIONARY NOTES

7.1 Bayesian Statistics has been and continues to be the subject of controversy. This chapter has indicated some of the reasons underlying this and is included, not necessarily to endorse its use, but to acknowledge the fact that it is currently used, and that the pressures inherent in development and demonstration programmes encourage its use. However, it must be recognised that, in the wrong hands, these statistical techniques can produce misleading results, and that some eminent statisticians have strongly deprecated their use for Defence Equipment. It is imperative that prior knowledge be assessed with complete honesty and placed formally on record before the trials take place. The technique should not be used without the advice of a competent statistician.

7.2 It is difficult to transform an intuitive subjective assessment of prior information into the parameters of a prior statistical distribution. In practice, the type of prior distribution is chosen on the grounds of its mathematical convenience, rather than because the prior knowledge is actually of that form.

7.3 There is a strong possibility that wishful thinking will colour the expression of the prior knowledge, particularly if the reliability development programme leaves something to be desired, and/or there are several as yet untested modifications in the items about to undergo trials. If a 'prior' is chosen which indicates a strong degree of belief, then the prior will dominate the test results for some time into the trial. If the prior happens to be optimistic (as is the tendency) the consequences could be serious. For example, a few early successes could be sufficient, with a strong prior, to provide a 'demonstration' to the required confidence level, even when the true reliability is below target.

7.4 Further problems, partly of an ethical nature, arise when one is faced with test results that clearly conflict with the chosen prior, because priors once chosen should be unalterable. Also, if the prior has not previously been placed on record, the temptation must exist for the unscrupulous, when writing up the test report, to post a prior after the trials, which produces the overall required demonstration.