Computation of the Value of Security

Final Report

Dr. Mario Rios
Dr. Keith Bell
Dr. Daniel Kirschen
Prof. Ron Allan

October 1, 1999

EPSRC / ERCOS grant reference no. GR/K 80310

Manchester Centre for Electrical Energy,
Department of Electrical Engineering and Electronics,
UMIST,
PO Box 88,
Manchester M60 1QD

Computation of the Value of Security
Final Project Report
Volume I

Dr. Mario Rios
Dr. Keith Bell
Dr. Daniel Kirschen
Prof. Ron Allan

October 1, 1999

Contents

Chapter 1

Introduction

A power system is always operated with a significant security margin to ensure that the transmission network is capable of withstanding unpredictable events such as line and generator outages. Since the introduction of competition in the electricity supply industry in many countries, the cost of this security margin has become considerably more apparent. This transparency has quite naturally led some parties to question whether customers are paying for more security than they need or want.

The work in this project follows the premise that, instead of following fixed security standards for the operation of the power system, a cost/benefit analysis should be performed. While the cost of security is reflected directly and deterministically in the payments made to the generators, the benefit is related to the consequences of stochastic events and is therefore considerably more complex to evaluate. When exposed to the same disturbances under the same conditions, a more secure dispatch will lead to smaller voluntary or involuntary load disconnections and require fewer emergency actions than a less secure dispatch. The avoided societal costs of the load disconnections and the avoided cost of the emergency actions constitute the benefit of the more secure dispatch.

A method has been developed for systematically and efficiently estimating *a priori* these outage costs in the context of power system operation. Such estimates can be balanced against the deterministic production cost of various dispatches to determine the optimal level of security in the operational time frame.

Since unscheduled outages in power systems are stochastic phenomena, computing their cost requires a probabilistic approach. While probabilistic methods have been used extensively in power system planning, they have so far not been widely applied in the operational timeframe domain. This distinction is not justified by a qualitative difference in the nature of the issues but by different attitudes towards risks and responsibilities. Long-term planners accept the probability of outages because they would occur in a system that will not be in operation for a few years. On the other hand, operators are much more likely to get blamed if the lights go out. They are therefore less comfortable with uncertainty and prefer deterministic criteria. However, as the "Review of Security Standards" [1] performed in 1994 by the National Grid Company for the UK electricity regulator OFFER shows, competition may force utilities to accept more risks. In

order to manage this increased exposure to risk, it is likely that they will want to make wider use of probabilistic techniques.

Power system operation differs from power system planning in terms of the accuracy requirements and of the level of uncertainty. Since the analysis takes place much closer in time to the actual events in the operational time frame, the base case is known with much greater certainty. On the other hand, since the accuracy requirements are much higher, the models must be more detailed. Since the computational efficiency of state enumeration falls off rapidly as the system being studied becomes large and results are required for use by operators, implying a need for speed, Monte Carlo estimation forms the basis of the calculation of the value of security.

Monte Carlo simulation is defined in [2] and [3] as "any technique for the solution of a model using random numbers or pseudorandom numbers." (Others, such as Billinton and Allan [4], prefer to describe Monte Carlo simulation as applying only to a process that is completely random in all respects. They prefer to describe the application of "stochastic simulation" to any process related to time). While Monte Carlo methods may naturally be used to simulate systems which have some stochastic content, any system for which a probability density function (p.d.f.) can be formed can be simulated in the same way.

A Monte Carlo estimation (or simulation) has 5 main aspects (see figure 1.1):

- *probability density functions* - the system being modelled must be described in these terms
- a *random number generator* - a source of random numbers uniformly distributed in $[0, 1]$. No computer algorithm exists to generate genuinely random numbers, but many different algorithms for generating pseudo-random numbers are available. Of these, many can be found in pre-written libraries such as the NAG library [5, 6].
- a *sampling rule* for prescribing how samples from the given p.d.f.s are to be taken.
- *scoring* (or *tallying*) to accumulate the outcomes of the trial simulations.
- *error estimation* - an estimate of the variance is obtained as a function, principally, of the number of trials.
- *variance reduction* - a method for reducing the variance in the estimated solution in order to reduce the computation needed to obtain a given accuracy of result.

The Monte Carlo estimation developed here allows the operator to compare operational plans on the basis of the sum of the production cost of the scheduled plant configuration and the expected cost of unplanned outages. Unlike conventional security analysis, no assumption is made with regard to what outages or combinations of outages are possible, or what classes of outages must under all circumstances be secured against.

For each plan to be evaluated, the estimation starts from the state of the system which would result from the implementation of the plan. In this "planned" state:

- the load follows the forecast;

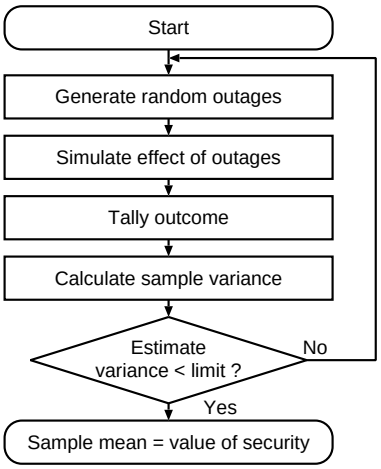

Figure 1.1: The Monte Carlo simulation

- all generators produce active power according to the specified dispatch;
- all equipment is in service except when scheduled for maintenance;
- all off-line reserves are in the appropriate state of readiness.

Individual "trials" are then generated by creating random contingency conditions based on this state. This is done by simulating the following types of random events:

- faults on lines, cables, transformers and busbars leading to an outage of the plant in question which is "permanent" in the short-term;
- unscheduled shutdowns and derations of generating units;
- large, abrupt and unforeseen changes in the load;
- cascade, sympathetic and inter-trips trippings.

The probabilities used to generate these random events should reflect not only the intrinsic characteristics of the equipment involved but also external factors such as weather conditions, which can be taken into account in the operational planning timeframe but which are often "averaged out" in planning studies. These external factors can have a very significant effect on the probability of faults and hence on the expected outage cost for a particular dispatch. To improve the accuracy of the outage cost estimate, the effect of weather conditions on fault rates are modelled on a regional basis.

Each generated random contingency condition must then simulated to determine what it would cost were it to occur. This is done using an AC load-flow computation to determine the power system state. Then, the response of operators to that state should be represented. If there are no violations of operating limits, this response will be to do nothing. Otherwise, generation may need to be rescheduled or load shed. The costs of these actions constitute the outage costs of the contingency, with the cost of shedding load determined by some suitable function or parameter, e.g. the "value of lost load".

Generation of new trials (random contingencies) then continues until the mean calculated from the random sample is judged to have reached some pre-determined precision.

The environment in which the computation described above can be achieved may be broken down into the following components:

- Input/output manager.
- Monte Carlo block comprising:
  - Monte Carlo simulation controller.
  - implementation of variance reduction.
  - probability of trip monitor.
  - random number generator.
- Load flow block comprising:
  - load flow engine.
  - automatic action monitor.
- Operator actions block comprising:
  - security monitor.
  - power system sensitivity analysis.
  - corrective action scheduler.
  - load restoration process.
- Load interruption cost calculation.
- Results analyser.

These are illustrated in figure 1.2.

This final report is divided in two volumes. The first one presents the modelling characteristics used in the development of the program Value of Security Assessor. The second volume shows testing this software in both a small portion of England-Wales southweast system and in the full large NGC power system.

The first volume includes:

Chapter 2 describes main characteristics of the modelling of the power system for the computation of the value of security; such as the load-flow and operator action blocks. Appendix A provides details of the power system analysis.

Issues surrounding how load interruptions might be costed are discussed in chapter 3. Chapter 4.2 shows how the weather effect on failure rates is considered in the *Value of Security Assessor* .

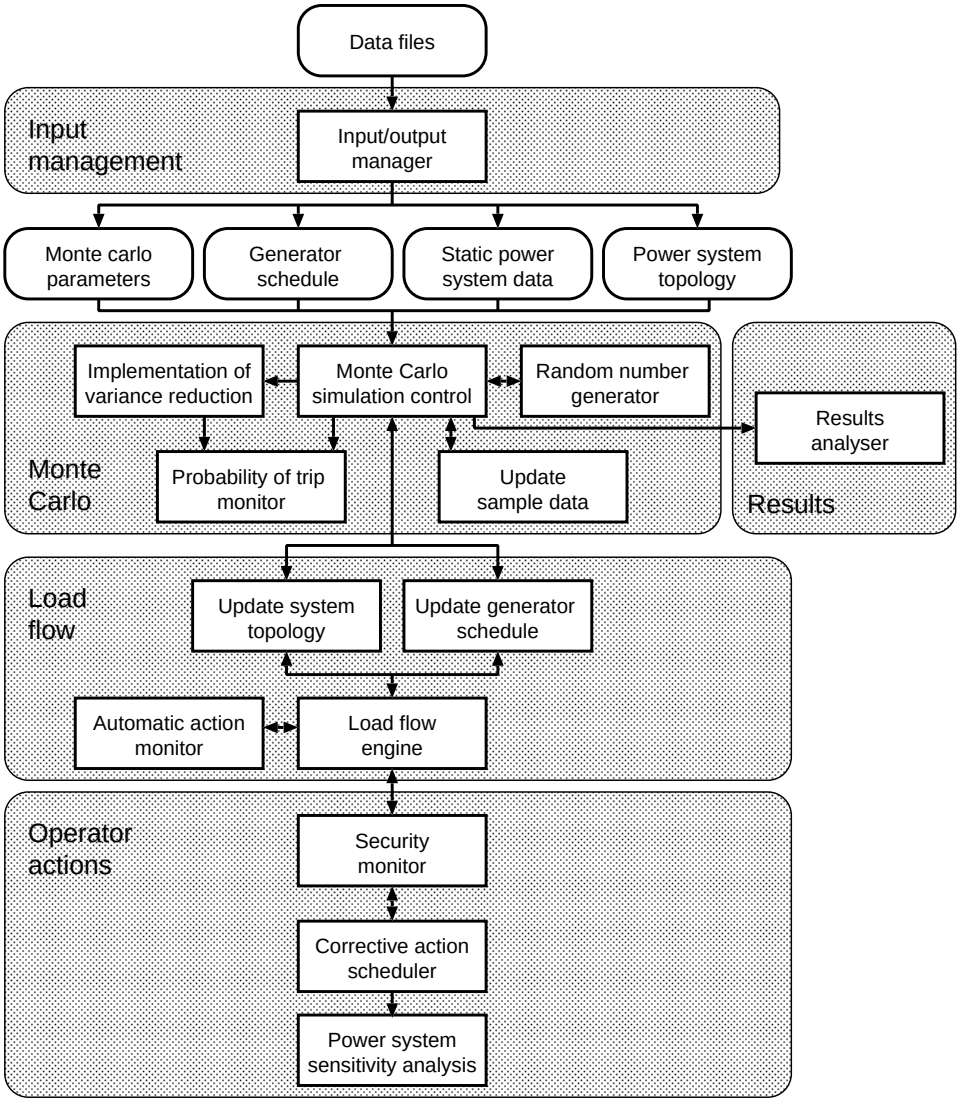

Figure 1.2: Value of Security Assessor structure

Chapters 5, 6 and 7 describe the basic statistical theory used in Monte Carlo estimation, the Monte Carlo method itself and the background of variance reduction respectively.

The application of variance reduction in this project is described in chapter 10 while underlying in terms of sampling methods and measures of component reliabilty in power systems are given in chapter 8. Some background in terms of application of variance reduction in power system reliability analysis is described in chapter 9.

Chapter 2

System Modelling

Computation of the value of security involves the analysis of the system's behaviour over a period of time (for example 1 hour or 1 day). This computation must include the cost associated with shed load and generation rescheduling, demand variations on the period and the corresponding schedule of generation. Therefore, the value of security must model a planned generation schedule over a period of time with unplanned events represented at the times at which they occur in the simulation. This implies a sequential simulation of temporal snapshots.

As chapter 1 states, a Monte Carlo sample simulation computes the variable of interest (the cost in this case) for a number of trials. This chapter shows the modelling of the power system behaviour that is used in the analysis of one trial.

The trial simulation is divided in sequentially snapshots. For each snapshot, a sequence of events occurs, such as random outages (section 2.2) and time-dependent phenomena. The term "time-dependent phenomena" is used to cover a range of influences and effects. These phenomena are:

- The evolution of an outage in terms of the response of protection. In particular, this concerns the possibility of "sympathetic" and cascading (section 2.3) outages.
- The impact of transient instability (section 2.3.5).
- The impact of voltage collapse (when it happens slowly enough for it to be seen and arrested by operators) (section 2.5).
- The staggering of the random outages over the period of study (section 2.4).

The two main functions of the load-flow and operator action blocks of the value of security analysis software are to derive the system state that would be seen by an operator, and to derive the corrective actions that an operator may take in response to any violations of operating limits found. These are described in sections 2.1 and 2.6 respectively.

The assessment of the duration of interruptions is also an important factor in the value of security calculation. Section 2.7 looks into the uncertain duration of plant outages and of the corrective

actions taken to soften their effects and their impact on subsequent time intervals of a daily generation schedule. Since a rigorous approach would require impractical amounts of computing time, suitable approximations will be needed. Hence, a heuristic approach, which takes into account the unserved load during the interruption, will be used. This approach will model the interruption duration as tables, rules or curves.

2.1 Derivation of the System State

Figure 2.1 shows the simulation of disturbance events for each trial in the Monte Carlo simulation used in the *Value of Security Assessor* . The sequence of events is divided in three parts:

1. Generation of a new system state.
2. Computation of an equilibrium point.
3. Corrective actions and cost evaluation.

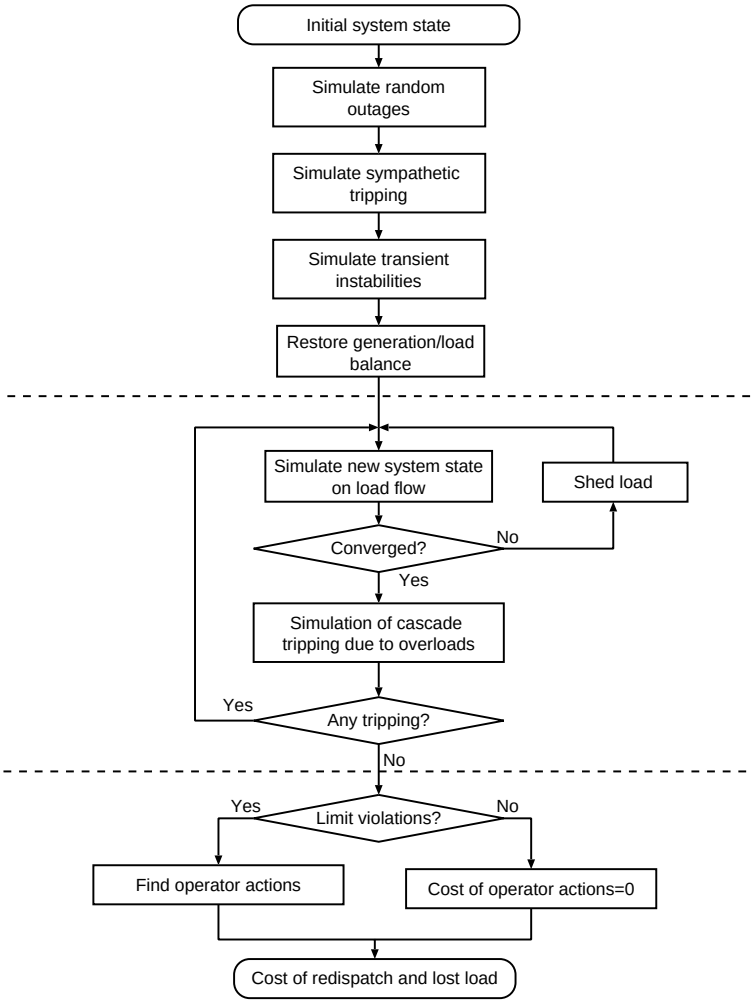

Figure 2.1: Simulation of one snapshot in a trial

The initial system state given by network topology, load demand and generation schedule is modified by random disturbances. These disturbances could also produce other disturbances such as sympathetic and transient instability trips.

The equilibrium point of the new system state must be calculated using a power flow computation (see appendix A for the details of the static modelling). Two outcomes are possible for this computation:

- The power flow converges.
- The power flow diverges. This indicates that the occurrence of this contingency state would result in voltage stability problems. A heuristic technique has been developed to determine how much load must be dropped to restore the feasibility of the power flow. In the *Value of Security Assessor* , if the load flow diverges, it is assumed that the system or a part of the system (an island) would have suffered a voltage collapse were the operator not have taken action. It is further assumed that the operator's response to an impeding collapse would have been to shed load in 5% blocks in the area of the biggest mismatch until convergence is achieved. If convergence has still not been achieved after all load has been shed, the island or the system is deemed to have "collapsed".

  Generation is reduced by matching amounts. The island is then checked for any violations of operating limits and further operator actions simulated if necessary (see section 2.6).

When the system has reached an equilibrium point (EP1, convergence of the load flow), a series of cascade tripping events may occur. In this case, a new load flow computation is required. A divergence of this new load flow indicates a severe problem (voltage collapse) has been caused by events ocurred after EP1. As in the computation of EP1, a load shedding is realised until a new equilibrium point (EP2) is reached.

As figure 2.1 shows, a sequence of load flow calculations and disturbance events may be established in an iterative way. This succession of calculation of EP(i)-disturbances-EP(i+1) can be interpreted as a succession of slow events that provokes a voltage collapse in the system.

Finally, the system reaches a last equilibrium point (EP). The system has a converged load flow with two possible outcomes:

- The resulting state of the system does not exhibit any major violation of normal operating limits. This state does not require any corrective action and has a cost of zero.
- The resulting state has some violations of normal operating constraints. Corrective actions must be taken to bring the system back within acceptable limits. The cost of these actions is computed and tallied.

2.2 Random Outages

The first step in the simulation of one trial in a complete Monte Carlo simulation process, is the generation of random outages. In the *Value of Security Assessor* program random outages are generated for each temporal snapshot. In this way, the electric power system could be subjected to coincident events (i.e. outages that take place at the same temporal snapshot) or random sequential events (i.e. random outages that take place in different snapshots).

Random outages are simulated for a specified time interval from the last interval system state condition. The set of possible independent failure events includes:

- generator failure;
- line outage;
- double circuit outage;
- busbar outage;
- SVC and shunt compensation failure.

Note that transient events are not considered. It may also be noted that double cicuit outages and busbar outages entail taking out of service more than one line.

If the failure rate of each type of failure is known and assumed to be constant, the probability $p_i(t)$ of the ith plant item going out of service unexpectedly in a time interval of duration t may be found from [4]

$$p_i(t) = 1 - e^{\lambda_i t} \approx \lambda_i t \tag{2.1}$$

where λ_i is the failure rate of the ith plant item.

If the system contains m items of plant and bulk supply points, the status of the system may be determined for the end of some time period t by taking a vector of random numbers, one for each of the m statuses. If a random number U_i is less than $p_i(t)$, the ith plant status is deemed to be "out of service"; otherwise, it remains in service.

2.2.1 Topology checking

If the contingency being simulated includes any outage of transmission plant, the topology of the system state being studied is checked before execution of the load-flow module. Each bus in the system is assigned to just one island and each island is solved in turn.

2.2.2 Modelling of frequency control

Since the aim of the model is to determine the state of the system that would be seen by an operator, and then the operator's response to that state, a number of functions have been included to represent generator primary frequency response and operator actions to maintain frequency.

Three levels of reserve are generally provided in large inter-connected power systems [7]:

- primary—fast response provided by governor action;
- secondary—generation operating at less than maximum output which may be called on to change output by the operator;
- standing—unsynchronized, fast start-up plant such as pumped storage and gas turbines.

Since the value of security assessment is designed for use in short operational timescales, "replacement" reserve—hot-standby plant or deferred-start plant—has not been modelled.

In order to model the different kinds of reserve available to the operator, all generators in the model are designated as having one of the following types:

1. frequency responsive;
2. available for secondary reserve;
3. inflexible;
4. gas turbine;
5. pumped storage/hydro.

The total spinning generation in each island is compared with the total demand. If an unplanned outage causes an unbalance between generation and load of $\Delta P = \Delta P_l - \Delta P_g$ where ΔP_l is the change in total load and ΔP_g is the change in total generation, a change in frequency will result. If the change in load due to a change in frequency is neglected (i.e. the load remaining after the unplanned outage is assumed to be constant), the change in frequency $\Delta f'$ when the system regains steady-state is [8]

$$\Delta f' = \frac{-\Delta P}{\sum_{i=1}^{m_g} (1/R_i)} \tag{2.2}$$

where R_i is the "regulation" or "droop" of the ith generator and m_g is the number of generators. The change in generation at the ith generator is then

$$\Delta P_{g_i} = \frac{\Delta f'}{R_i}. \tag{2.3}$$

If the change in frequency, which is assumed to take place quickly and before an operator has a chance to respond, is such that the system frequency goes outside under-frequency relay limits, the load-shedding action of these relays is modelled and the load-generation balance is rechecked.

If the sum of the P generation in an island is still too low with respect to the load, the generators not denoted as inflexible are moved in turn to their maximum generation (cheapest generator first), then, if necessary and depending on the level of water in the relevant reservoir, pumped storage generation is increased or pumping reduced. Open cycle gas turbines are then switched on if required. As a last resort, load is shed in 5% blocks and the load/generation balance again checked. If there is too much generation in an island, the most expensive generation is reduced in turn to its minimum stable generation or switched off, pumped storage generation is reduced or pumping started and finally, if necessary, expensive generation is shut-down altogether.

System active power losses calculated in the course of the load-flow solution are distributed among the generators providing primary response according to their respective droops and maxiumum limits.

2.3 Sympathetic and cascade tripping

A serious disturbance is characterized by the total or partial collapse of the electric power system. A region, or the whole territory served by the utility, is deprived of electric energy. This collapse is the result of a degradation process in the operation of the system, during which a least one of the following phenomena occurs:

1. Casacade tripping of transmission elements (lines, transformers)
2. Loss of static stability of generation units (resulting in the tripping of these units)
3. Loss of transient stability of generation units
4. Voltage instability in the power system

A number of causes can trigger this process of degradation [9, 10, 11]: malfuntion of protective systems; misoperation of the monitoring, operation and control system (EMS/SCADA); uncorrect decision taken by the operators, etc. A critical event will be defined [12] as an event which, in a given operating state of the power system, is potentially responsible for a serious disturbance. A serious disturbance is an event which results in the collapse of the power system.

Obviously, a critical state of the power system is a necessary but not sufficient condition for the collapse of the system to occur. This is because an element can always be operated outside of its allowable limits for a short time. It is only when no control action has enabled the element to regain an allowable operating state that the element will be taken out of service. If this were the case, the tripping might place the system in a critical state and, by cumulative effect, further line trippings could occur until the cascade tripping results in a partial or total collapse of the power system.

In order to model appropriately these phenomena in the *Value of Security Assessor* , they are divided in two types: "sympathetic" tripping and cascade tripping. They are defined as follow:

- **Cascade tripping**. Tripping of overloaded parallel lines after a fault on one of these lines.
- **"Sympathetic" tripping**. Unnecessary tripping of one or more elements caused by a fault in their neighbourhood.

2.3.1 Cascade tripping of parallel elements

When a set of parallel or quasi-parallel lines is heavily loaded, a fault and subsequent tripping of one of these lines may cause overloads on the other lines. Unless corrective action is taken promptly, overload protection relays on these lines (or faults due to sagging) may then cause the cascade tripping of the remaining lines.

In some situations the operator is not able to eliminate the overloads within the maximum period allowed. The element (line) is then correctly tripped by its protection system [12]. Alternatively, the overload elimination by the protection system is realised before the operator can take any action. An example of this is:

- New York Blackout, 9th november 1965. After the tripping of the line Q29BD, the other four lines arriving at Beck's station became loaded beyond their critical level and the protection system tripped out "in cascade" the four lines 2.7 seconds after the main disturbance [13].

In other cases, the operator fails to take action or takes an inappropriate action:

- WSCC System - Blackout August 10, 1996 [11]. "[...] Big Eddy-Ostrander, Jhon Day-Marion and Marion-Lane 500 kV lines were forced out of service. While none of these lines were individually judged to be crucial by BPA dispatchers, the cumulative impact resulted in a susbstantially weakened system. BPA did not communicate these outages to other WSCC members nor did they reduce loadings on lines or adjust local generation as precautionary measures ...".

2.3.2 "Sympathetic" tripping of transmission elements

In a large proportion of blackouts, it has been observed that protective system failures are a contributing factor in the degradation process. A failed or improperly set protection can make a bad situation worse. A study of significant disturbances reported by NERC in the period from 1984 through 1988 indicates that protective relays were involved in one way or another in 75 percent of major disturbances [14, 15].

A common scenario is that the relay has an undetected, hidden, defect that is activated under the conditions created by other disturbances. For example, nearby faults, overloads, or reverse power flows expose the defective relay and cause a false trip [14, 15]. The subsequent event occurs either before the system operator has had time to take any corrective action in the system. Some examples:

- New York City Blackout - November 9th 1965 [13]. "[...] the back-up relay protecting line Q29BD operated normally and caused the circuit breaker at Beck (Station) to trip this unfaulted line."
- New York City Blackout - July 1977 [14]. The blackout was caused initially by two lightning strikes and a series of relay and control operations and misoperations, including a malfunctioning directional control contact.
- WSCC System - Blackout July 2, 1996 [16]. The main disturbance was a flashover to a tree on the Jim Bridger-Kinport 345 kV-line. Misoperation of a ground unit of an analog electronic relay tripped the parallel Jim Bridger-Goshen 345 kV line.
- WSCC System - Blackout August 10, 1996 [11]. Approximately five minutes after the main disturbance (failure of the Keeler-Allston 500 kV line) the St. Johns-Merwin 115 kV line tripped due to zone 1 KD relay malfunction.
- WSCC System - Blackout August 10, 1996 [11]. "[...] units a McNary began tripping due to excitation equipment problems [...]", "The unit is supposed to stay connected to the line [...] but erroneous protective relay actions tripped the units off line",

2.3.3 Modelling protection system misoperation in the *Value of Security Assessor*

References [14, 15] present the concepts of hidden failure and vulnerability region. It defines a hidden failure in a protection system as a permanent defect that will cause a relay or a relay system to incorrectly and inappropriately remove circuit elements as a direct consequence of another swithching event (the initial disturbance). So, hidden failures play an important role in extending the disturbance.

Each hidden failure has a region of vulnerability associated with it. If an abnormal event occurs inside the region of vulnerability, the hidden failure will cause the relay to incorrectly remove the circuit element, thereby creating an additional abnormal state in which additional hidden failures may be exposed [14, 15].

Based on these concepts, it is proposed to use a probabilistic model in the *Value of Security Assessor* . In this model the following characteristics are associated with each transmission element:

1. A probability (p_1) associated with the misoperation of the element's protective system when a failure occurs in its vulnerability region (conditional probability).

2. A vulnerability region that defines the portion of the system where a fault may provoke the tripping of the element.

References [14, 15] define four types of vulnerability regions (RV) based upon the reach settings of the relays:

1. Type 1- Reverse local bus RV. The RV is behind the local bus (A) and extends to *half* of the line as determined by the setting of the relays at the remote bus (B) (See figure 2.2)
2. Type 2- Remote bus RV. The RV is beyond the remote bus (B) and extends to half of the line protected.
3. Type 3- Zone 2 RV. The RV is beyond bus B and has a reach of 0.2 of the line protected by the relay.
4. Type 4- Zone 3 RV. The RV is beyond bus B and has a reach of 1.2 times the longest line connected behind the remote bus.

References [14, 15] use these four regions to calculate the relative importance of each one in the tripping of elements. This calculation is not relevant for the value of security calculation in the *Value of Security Assessor* . A simpler definition of the vulnerability region based on the definition of the biggest region that includes the four types defined above will be used. Figure 2.2 shows the proposed RV for the *Value of Security Assessor* .

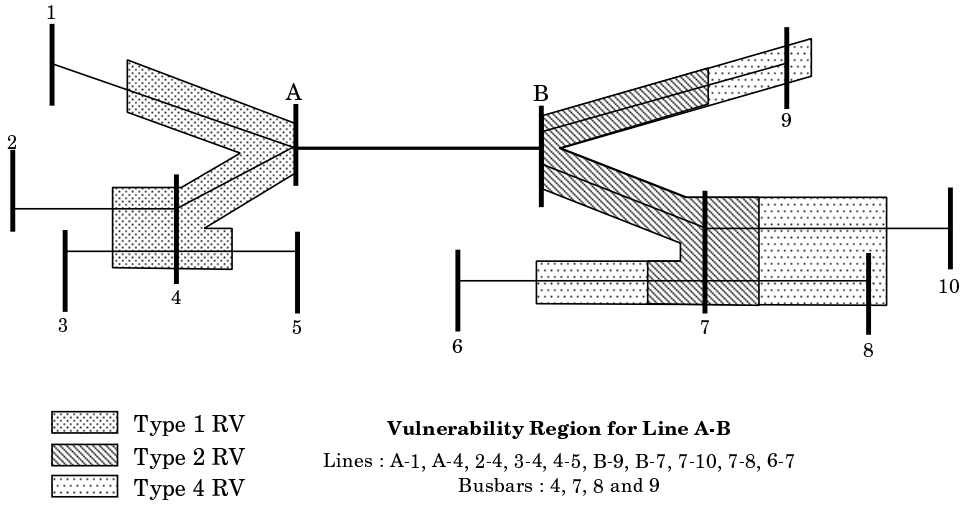

Figure 2.2: Vulnerability region of line A-B

When the *Value of Security Assessor* simulates a fault on an element (line, transformer, busbar), its effect on protection misoperation will be evaluated for each element which has the original faulted element in its vulnerability region. For example, suppose that a fault occurs on line 7-10 of figure 2.2. Since this line is in the vulnerability region of line A-B a Monte Carlo trial will indicate whether line A-B is tripped by sympathy. The probability used in this Monte Carlo trial is the value p_1.

2.3.4 Modelling of inability to correct overloads and human error in *Value of Security Assessor*

Lines in a transmission system can be overloaded for some time before relays disconnect them or before they sag and cause a fault. During this time the operator must eliminate the overload in the line. This task is not always completed properly or in time and line trips due to overloads do occur.

For modelling this characteristic in *Value of Security Assessor* , it is assumed that in $r\%$ of the situations encountered the overloaded lines are tripped [12]. "r" is the probability that the operator is unable to eliminate the overloads in time. The value of r is a function not only of the operator's ability to control the system but also of the control tools at the operators disposal. In practice, it may be assumed that this value increases with the number of overloads that the operators must correct.

2.3.5 Loss of stability of generation units

One of the major causes of degradation of a power system is the tripping or disconnection of generation units. Normally, this disconection is provoked by the loss of synchronism of the generators. The tripping of these units can be a consequence of either a line fault, a fault on the generator or a cascade of events in the transmission system.

In a high proportion of major blackouts, generation tripping is an important factor, as shown by the following examples:

- New York City Blackout 9th November 1965. "The instantaneous result (of the tripping of the conection lines between Toronto and Beck's Station) was the acceleration of generators a Beck and Pasny-Niagara, with a sharp drop in their electrical outputs ... The instantaneous drop in generation at Beck and Pasny ... resulted in putting this generation out-of-phase with most of the other generation attached to the interconnected transmission system ..." [13].
- WSCC blackout, 10th August 1996. Power and voltage oscillations produced by the initial cascade of events provoked the separation of the system in four islands. The generation lost was 25455 MW, involving 178 generation units [11].
- Sweden Blackout of 1983. During the cascade tripping of elements, the frequency of the system decreased very fast. This produced the tripping of nuclear plants in the South of the system [10].

This section deals with the modelling of generation tripping by loss of synchronism due to a fault or an event that does not directly affect the generator itself. Faults that affect the generator directly are simulated by the main Monte Carlo simulation process.

Two types of instabilities affect power systems: angular instability and voltage instability. Angular stability is further divided into: transient stability and steady-state stability.

Transient stability in the *Value of Security Assessor*

Much work has been done to include transient stability issues in reliability evaluation [17, 18, 19, 20]. Several factors make this assessment complex:

1. Type of fault: is the fault a three-phase fault, a double line-to-ground fault, a double line fault or a single line-to-ground fault?
2. Location of fault: is the fault at one end of the line or at an intermediate point?
3. Fault-clearing phenomena: the clearing time of faults is a stochastic variable.

The first and second factors are independent events while the critical clearing time (CCT) depends on the type and the location of the fault. Therefore, a conditional probability approach can be used to assess system transient stability.

A discrete probability density function (*pdf*) which assigns a probability for each type of fault i is used. So, for the set of types of fault I (3ϕ, 2ϕ, 1ϕ and double line faults), the following equation is satisfied:

$$\sum_{i=1}^{4} P_i = 1.0 \tag{2.4}$$

where P_i is the probability of a fault of type i. It is assumed that there are only 4 types of fault in the set I.

The location of fault is stochastically modeled by a discrete *pdf* which assigns a probability of fault for each portion of the line. If N_j is the number of possible locations (portions of the line) where the fault can occur, the following relation must be satisfied:

$$\sum_{j=1}^{N_j} P_j = 1.0 \tag{2.5}$$

where P_j is the probability of a fault at location j.

A possible *pdf* for the fault clearing time is a normal distribution [18, 19]. If the critical clearing time is the maximum time in which the fault must be cleared, then from its *pdf*, the probability of stability ($P(S/ij)$) for a fault type i at a location j is obtained.

Using the conditional probability approach, the probability of stability due to a fault on line k (P_{ST_k}) is given by the following equation [18, 19]:

$$P_{ST_k} = \sum_{j=1}^{N_j} \sum_{i=1}^{4} P_k(S/ij) \times P_i \times P_j \tag{2.6}$$

where $P_k(S/ij)$ is the probability of stability for a given fault of type i at location j of line k.

For the operational timescale evaluation, the transient stability effect on the value of security calculation can be modeled using the PST_k values. Naturally, off-line work is necesary to obtain the probabilities $P_k(S/ij)$. In the value of security calculation, a Monte Carlo simulation will determine if the fault provokes an instability in the system based on the set of probabilities PST_k.

On the other hand, if a stability problem is detected, it is necessary to determine what are the affected generators. Off-line stability studies could be used to determine the vulnerability regions associated with the stability of each generator. Thus, if a fault on line k provokes an instability and this line is in the vulnerability region of generator l, then the generator is disconnected.

In brief, including the effect of transient stability on the value of security calculation requires the following data:

1. Set of probabilities PST
2. Vulnerability regions of stability for each generator

The first and most difficult step in the probabilistic approach is to collect statistical data on system faults.

2.4 Staggering of random outages - Sequential Simulation

Computation of the value of security over a period of time (for example 1 day) must include the cost calculation associated with shed load and generation rescheduling, demand variations on the period and the corresponding schedule of generation. Therefore, the value of security must model a planned generation schedule over a period of time with unplanned events represented at the times at which they occur in the simulation. This implies a sequential simulation. For this purpose, the *Value of Security Assessor* uses a definition of sequential intervals. Each interval is characterized by a load, considered constant over the interval, and its generation schedule.

Sequential simulation allows the simulation of new random outages on each interval and the corresponding analysis described in figure 2.1. However, the effect produced by two or more random outages will be different if they occur at the same moment (beginning of the interval) or if they occur at different moments over the simulated interval. In order to properly represent the time at which stochastic events take place, this interval may be further broken down into ”subintervals”. This simulation subinterval is taken as the time it would take for an operator to realise that some event has taken place and to respond to it. As such, it may typically be regarded as, perhaps, 10 minutes or half-an-hour. Thus, events that take place within one ”subinterval” may be regarded as having taken place concurrently.

Figure 2.3 shows graphically the definition of terms used by *Value of Security Assessor* for the sequential simulation.

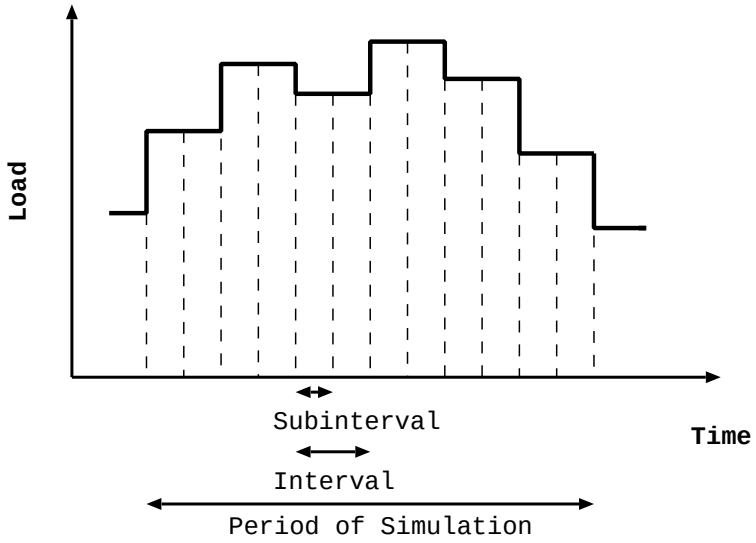

Figure 2.3: Division of time scale - Sequential Simulation -

In Figure 2.1, random outages are suppossed ”permanent” over the operational timescale. However, as it has been mentionned, a series of line trippings can be provoked by these random outages. These trippings cannot be considered to be ”permanent” faults. The *Value of Security Assessor* models a reconnection of tripped lines in the next simulation ”subinterval”.

Figure 2.4 shows graphically the *Value of Security Assessor* algorithm for sequential simulation. The blocks enclosed in the dashed line represents the algorithm of figure 2.1.

2.5 Voltage collapse considerations

Voltage stability has been defined by the System Dynamic Performance Comittee of the IEEE as being the ability of a system to mantain voltage so that when load admitance is increased, load power will increase, and both power and voltage are controllable. Voltage stability depends on the location and type of reactive power sources available in the system. A voltage collapse is a major problem in the power system caused by the degradation of the system following a serious disturbance.

2.5.1 Techniques of analysis of voltage collapse

Although the voltage collapse is a dynamic phenomenon, static analytical approaches have been developed for its study. The dynamic study using time-domain simulations is employed mainly for the coordination of control and protection systems and for the investigation of interactions among different power system components such as generators, ULTCs, SVCs, generator field current limiters, etc. Static approaches examine the viability of the equilibrium point represented by a specified operating condition of the power system. Static approaches also allow the identification of voltage-weak areas in the system and the mesure of a distance to instability. These approaches are [21]:

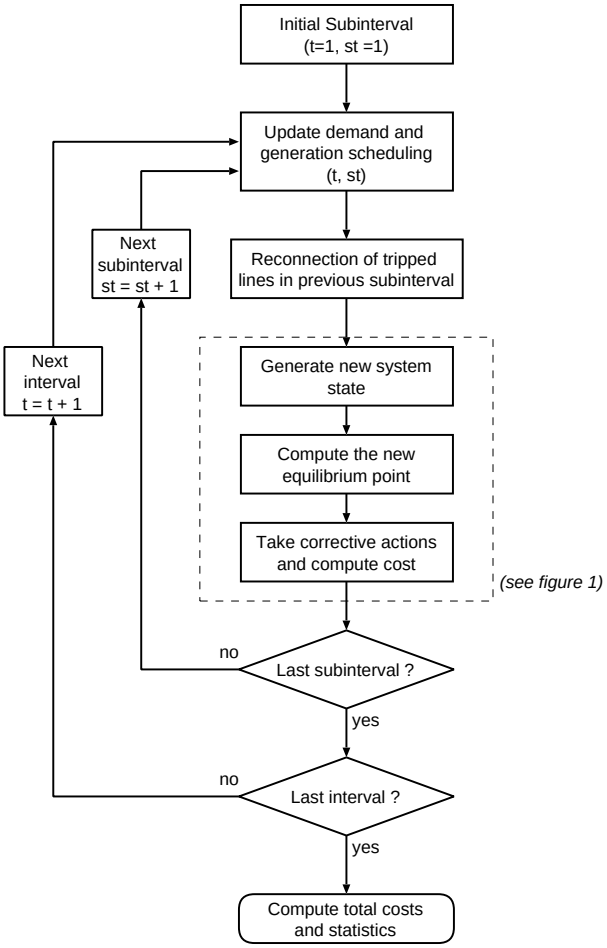

Figure 2.4: Sequential Simulation

1. PV and QV curves
2. VQ sensitivity analysis.
3. Modal analysis. This method develops an eigenvalue analysis of the reduced Jacobian matrix in order to determine the instable or critical modes of the system. This method determines the main busbars, branches and generators associated with each critical mode [21, 22] providing information about the mechanism of loss of stability.
4. Continuation load flow.

Formulation of indicators of the distance to the voltage collapse have been developped from these techniques, such as:

1. Risk of voltage instability indicator, L [23, 24]. The index L is calculated for each load node of the system. The value of L is between 0 (no load in the node) to 1 (voltage collapse).
2. Modal proximity to instability [22]. Each eigenvalue of the reduced Jacobian matrix corresponds to a mode of voltage/reactive power variation. If this value is positive the mode is stable. The magnitude of the eigenvalue gives an indication of the degree of stability.

3. Power (MW, MVAr and MVA) margin to voltage instability.

As reference [22] states, the calculation of L indices for all the load buses requires much less CPU time than modal analysis while the CPU time required to compute the sensitivities for all load buses is more than the time required in the modal analysis. On the other hand, modal analysis provides information about mechanism of instability and about the components associated with the main instability. However, in the operation timescale an indicator, such as the index L, which provides information on the voltage-weak area (busbars) for an operating condition and allows the calculation of the emergency load shedding to avoid risk of voltage instability is very useful [24].

2.5.2 Impact of voltage collapse in reliability evaluation

Reference [25] uses the indice L, proposed in [23] to incorporate voltage stability considerations in the adequacy assessment of composite power systems and proposes a bisection algorithm for the load-shedding calculation. Bus indices L are used as voltage stability constraints. If these constraints are violated, the load bus which has the highest voltage stability indicator is selected to carry out load shedding in order to reduce the indicator. The bisection algorithm is used to determine the minimum load to be shed (ΔP) to reduce the value of the indicator of the selected bus to the minimum acceptable voltage stability indicator (threshold value). After load shedding in the selected bus, indicators are recalculated for all the system load buses and the bisection process is repeated in the case of the another voltage violation until all buses do not violate the voltage stability constraints.

2.5.3 Indice L and modified indice B

Reference [23] defines the indicator L_j to assess the voltage stability of load bus j for a given condition of the system. A hybrid representation of a transmission system is used to define the indicator. This representation is given by:

$$\begin{bmatrix} V^L \\ I^G \end{bmatrix} = \begin{bmatrix} Z^{LL} & F^{LG} \\ K^{GL} & Y^{GG} \end{bmatrix} \quad \begin{bmatrix} I^L \\ V^G \end{bmatrix} \tag{2.7}$$

where V^L and I^L are vectors of voltages and currents at load nodes; V^G and I^G are vectors of voltages and currents at generation nodes; Z^{LL}, F^{LG}, K^{GL} and Y^{GG} are sub matrices of the hybrid matrix. The hybrid matrix is generated from the admitance matrix (Y_{bus}) by partial inversion.

The indicator L_j is given by [23]:

$$L_j = \left|1 + \frac{\sum_{i \in \alpha_G} \overline{F}_{ji} \overline{V}_i}{\overline{V}_j}\right| \tag{2.8}$$

where α_G is the set of generation nodes; $\overline{F}_{ji}$ is the complex component ji of the sub-matrix F^{LG}; $\overline{V}_i$ and $\overline{V}_j$ are the complex voltages at nodes i and j respectively.

From the indicator L, an approximation was developed in [24]. This approximation defines the risk indicator as:

$$B_j = \left|1 + \frac{\sum_{i \in \alpha_G} C_{ji} \overline{V}_i}{\overline{V}_j}\right| \tag{2.9}$$

where C_{ji} is the element ji of matrix C determined by:

$$[C] = -[B'']^{-1}[B_{LG}] \tag{2.10}$$

where $[B'']$ is the imaginary part of the matrix $[Y_{LL}]$ and $[B_{LG}]$ is the imaginary part of the matrix $[Y_{LG}]$. $[Y_{LL}]$ and $[Y_{LG}]$ are submatrices of the Y-bus matrix. As reference [24] shows, the approximated indice B_j has an error of less of 1% respect to indice L_j.

The indicator B_j can be separated into real and imaginary parts (B_j^R, B_j^I):

$$B_j^R = 1 - \frac{\sum_{i \in \alpha_G} C_{ji}|V_i|cos(\delta_i - \delta_j)}{|V_j|} \tag{2.11}$$

and

$$B_j^I = -\frac{\sum_{i \in \alpha_G} C_{ji}|V_i|sin(\delta_i - \delta_j)}{|V_j|} \tag{2.12}$$

where δ_i and δ_j are the voltage angles at buses i and j; $|V_i|$ and $|V_j|$ are the voltages magnitudes at buses i and j. Changes in the value of the indicator due to variations of the load (shedded) can be calculated as:

$$\begin{bmatrix} \Delta B^I \\ \Delta B^R \end{bmatrix} = [T] \begin{bmatrix} \Delta\delta \\ \Delta|V| \end{bmatrix} = [T][J]^{-1} \begin{bmatrix} \Delta P \\ \Delta Q \end{bmatrix} \tag{2.13}$$

where [J] is the Jacobian matrix and [T] is the sensitivity matrix between indicator changes and voltage angle and magnitude changes.

The variation of the indicator at node j will be given by:

$$\Delta B_j = \sqrt{(\Delta B_j^R)^2 + (\Delta B_j^I)^2} \tag{2.14}$$

2.5.4 Suggestion of an additional constraint to avoid voltage collapse in the *Value of Security Assessor*

As it was mentioned, the load flow divergence is an indication of voltage stability problems. In the *Value of Security Assessor* , if the load flow fails to converge, it is assumed that the system or a part of the system (an island) would have suffered a voltage collapse were the operator not have taken action. It is further assumed that the operator's response to impeding collapse would have been to shed load in 5% blocks in the area of the biggest mismatch until convergence is

achieved. If convergence has still not been achieved after all load has been shed, the island or the system is deemed to have "collapsed".

On the other hand, if the load flow converges for any contingency, system operators take corrective actions in response of violation of system operating limits. After that, the power flow of the system converges but the system could be located in a critical condition in the proximity of the voltage collapse. If this is the case, the voltage stability could occur if system operators do not take additional corrective actions.

The value of security calculation would determine if the system is on a viable operating condition or not and, if it is the case, to determine the emergency load shedding to avoid risk of voltage instability as an additional corrective action. That is an additional constraint that could be studied and considered in the *Value of Security Assessor* .

The *Value of Security Assessor* may follow some of the ideas proposed in [25] taking advantage of the calculation of indicators B_j, proposed in [24]. In order to avoid the iterative process of the bisection algorithm for the calculation of the load shedding, the *Value of Security Assessor* could use the relationship of equation (2.13) between indicator changes and load powers to be shed [24].

The following algorithm calculates the impact of risk of voltage collapse:

1. Calculation of indicator B_j for all load node j using equations (2.11), (2.12) and:

$$B_j = \sqrt{(B_j^R)^2 + (B_j^I)^2} \tag{2.15}$$

2. Evaluation of:

$$B_j \leq B_{threshold} \qquad\qquad j \in \alpha_L \tag{2.16}$$

   where α_L is the set of load buses. If the relation is satisfied for all nodes j then stop else continue.

3. Determination of the load bus j having the highest indicator, to carry out load shedding.

4. Calculation of load to be shed in order to reduce the indicator value at bus j to the target value using equation (2.13). As [24] suggests, a lower target value than threshold value must be used.

5. Calculation of the new state (load flow) and return to first step.

2.6 Modelling of Corrective Actions

In response to violations of system operating limits, operators can reschedule generation, change voltage set-points and tap ratios, and, as a last resort, shed load. Since operators reach decisions

about what actions to take based on advice given by planners, information gleaned from "what-if" load flow studies, and experience, their actions are represented in the value of security assessor by a fuzzy expert system with embedded load flow and linear sensitivity analysis ([26, 27] and appendix A of this report). Three types of corrective actions are modelled:

- active dispatch—dispatches settings of active power generation, shedding of load (active and reactive components in proportion) and changes to phase shifter settings in order to relieve overloads of transmission lines and cables;
- reactive dispatch—dispatches settings of reactive control devices in order to correct violations of voltage limits;
- dispatch of active controls for correction of voltage problems—this is activated to change the active generation and, if necessary, shed load in order to remove any outstanding violations of voltage limits [28].

Fig. 2.5 shows the procedure used to decide on corrective actions.

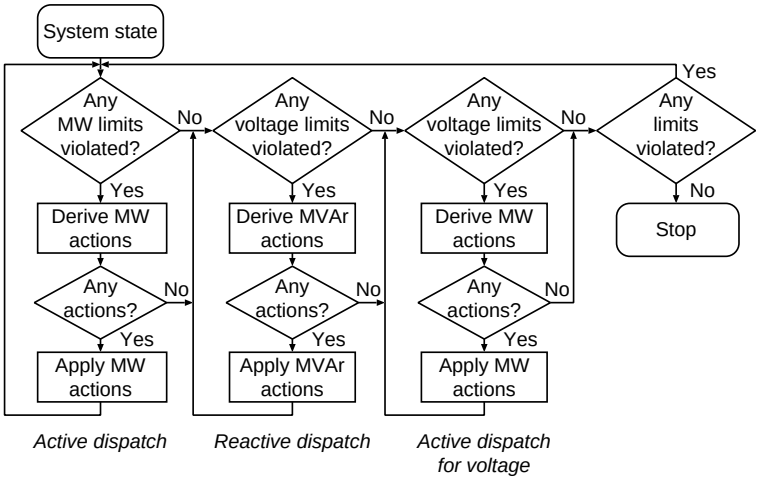

Figure 2.5: Modelling of corrective actions

The active dispatch, reactive dispatch and active dispatch for voltage are implemented as shown in fig. 2.6. In each, the operators' judgments are modelled by means of "qualitative reasoning" whereby the decisions are reached by balancing the criteria of control effectiveness, control margin, cost, simplicity and the possibility of unwanted secondary effects. More details of the operator action modelling can be found in [27, 29]

2.7 Duration of load interruption

This section discusses the duration of load interruptions in power systems. A heuristic method for the modelling of these interruption duration in the *Value of Security Assessor* is formulated. Determination of interruption duration is needed, not only for the calculation of the energy not supplied, but also for the calculation of the cost associated with the interrupted load.

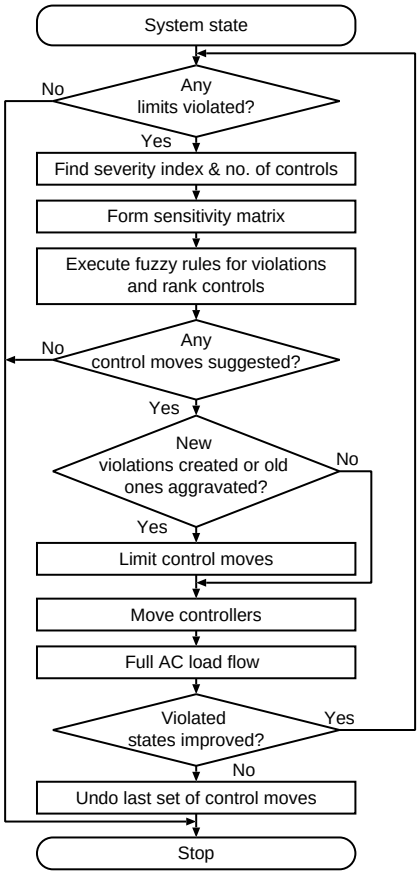

Figure 2.6: Operation of dispatch routine

Some factors that affect the duration of the interruption are:

- State of the system.
- Generation units start-up
- Load variations (load demand curve)
- Size (number of MW) and extent (number of affected busses) of the interrupted load

The duration of each load interruption can be divided in two phases:

1. The control phase. In this phase, the operators take corrective actions. Also in this phase, a degradation of the system involving cascade tripping can occur.
2. The load restoration phase. This phase involves the process of reconnection of the load that was lost or voluntarily disconnected during the control phase to maintain the system in operation.

2.7.1 Load restoration process

The overall goals of the restoration process is to first rebuild a stable electric system and then to restore all remaining unserved load [30, 31, 32]. In order to attain these goals, the following practices should be observed [30]:

1. Restore safely. Precautions must be taken to guard against human injury and equipment damage, avoiding both thermal overloads and overvoltages.
2. Restore smoothly and deliberately. To assure a stable rebuilt electric system, the balance between load and generations must be maintained. Repeated interruptions should be avoided.
3. Minimize overall restoration time. Both minimum and maximum restart times for generators should be examined when deciding on an unit startup sequence. Restoration actions should be taken immediately but a hasty restoration should be avoided.
4. Minimize adverse impact to public.
5. Maintain flexibility to respond to problems.

Before a restoration sequence starts, an assessment of the system should be performed to determine the extent and the characteristics of the problem and the overall status of the system. This assessment should determine to which of the following cases the problem belongs:

1. A *standard power failure* involving loss of load and regional transmission only. Load can be restored using straight-forward control actions.
2. An *unstable system* with significant degradation in voltage, frequency or power flow equilibrium. Emergency actions, such as load shedding, may help limit further degradation.
3. A *major power failure island formation* in which a portion of the power system becomes separated from the main synchronized power grid.
4. A *major power failure blackout* in which critical generation or bulk transmission facilities are lost resulting in a system collapse or blackout.

Different approaches to the restoration of the system are used depending on the size of de-energized area, the posibility to receive assistance from interconnected systems, the amount of blackstart capability in the system and the type of production in the system. Two major strategies [10] for restoring a power system following a blackout are known:

- *The build-up strategy.* It is often used when the system has suffered a total blackout and when it is impossible to receive assistance from the neighbouring systems. The system is divided into subsystems that include a least one station with blackstart capability, that can regulate both frequency and voltage of the subsystem itself and that have a good balance between production and consumption. After starting the station, emergency power

is supplied to the stations without blackstart equipment in order to start, if possible, the units in these stations and to synchronize them to the system. Loads are connected and more units are taken into operation. Each subsystem is then synchronized with the others.

- *The build-down strategy.* This strategy is mostly used in small systems without long high voltage lines or in hydro systems with good reactive absorbing capacity. The interrupted system has to include a least one station with blackstart capability. When this station becomes operative, it is connected to other stations in order to supply them with the emergency power necessary for their start up. The generators are synchronized with the system and more lines are connected. During these initial steps some load is connected but most of the load is connected when most of the generation and transmission systems are restored.

Based on the goals, objectives and power resources, the utilities make their restoration process planning. Figure 2.7 [33] shows the outcome of the restoration process following major blackouts.

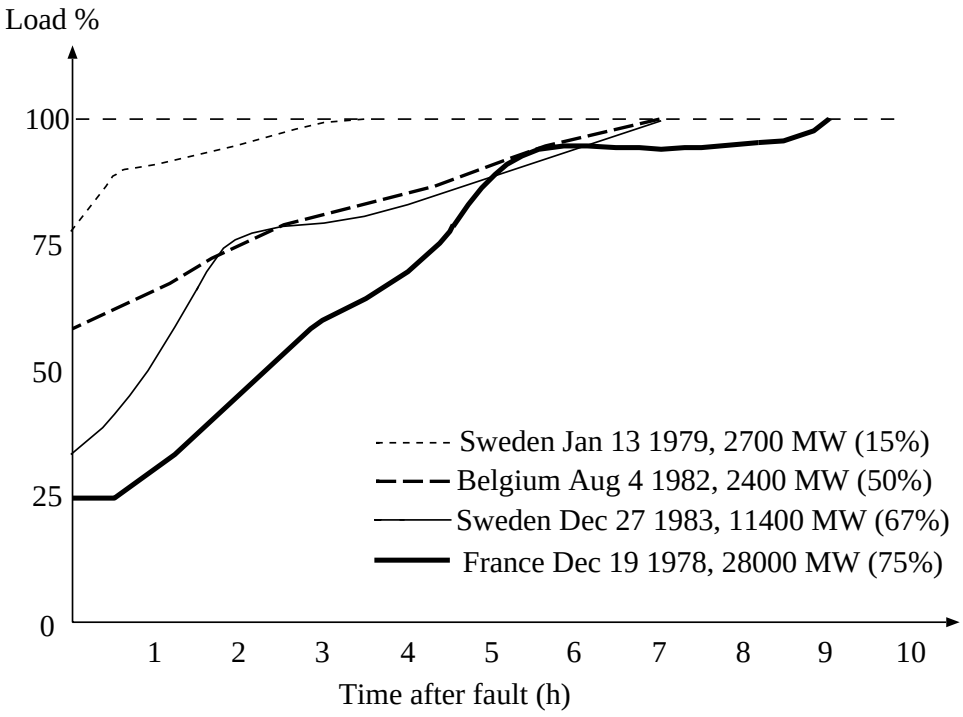

Figure 2.7: Load Restoration following Major Disturbances

2.7.2 A heuristic approach for the *Value of Security Assessor*

The modelling of the interruption duration in *Value of Security Assessor* takes into account the two phases of the interruption. The proposed modelling is used for the restoration of both partial and total blackouts.

It is clear that in the evaluation of the value of security, an exhaustive modelling of the restoration process is not necessary. This modelling must only capture the relevant characteristic for the evaluation of the value of security, that is the duration of interruption. As shown in figure 2.7, this duration is directly related to the unserved load. In fact, as the unserved load is related to the size (number of busbar) and the severity of the problem (cause), a relationship between unserved load and time of restoration could model the interruption duration, in a heuristic way.

A heuristic model is used in the *Value of Security Assessor* to model the interruption duration. This model gives the restored load as a function of time. This model divides the restoration process in two phases. If we measure the duration of the restoration from the moment where the first load is shed or disconnected, the first phase includes:

- the period during which the operator takes corrective actions to try to stabilise or to solve the rest of the system,
- the period during which the operator assurres the situation and decides an a restoration strategy,
- period during which the operator takes restoration actions that do not result in immediate load reconnections.

No load is thus reconnected during this first phase. It seems reasonable to assume that its duration is proportional to the severity of the outage. It can be modelled by:

$$t_1 = \frac{P_{disconnected}}{P_{total}} \times T_{total\,blackout} \tag{2.17}$$

where t_1 is the duration of the first phase. $T_{total\,blackout}$ is the expected duration of this first phase for a total blackout. Figure 2.7 (French blackout) suggests that 30 minutes is a reasonable value for $T_{total\,blackout}$. $P_{disconnected}$ and P_{total} are the unserved and the total load of the system, respectively.

The second phase represents the actual reconnection of the load. It appears reasonable to assume that the restoration rate increasses with time. Table 2.1 and Figure 2.8 show the restoration rates that will be used in this project.

Time Period (Min)	Restoration Rate [MW/min]
0 - 30	10.0
30 - 60	33.3
60 - 90	66.6
90 and more	83.3

Table 2.1: Restoration Rates - An example

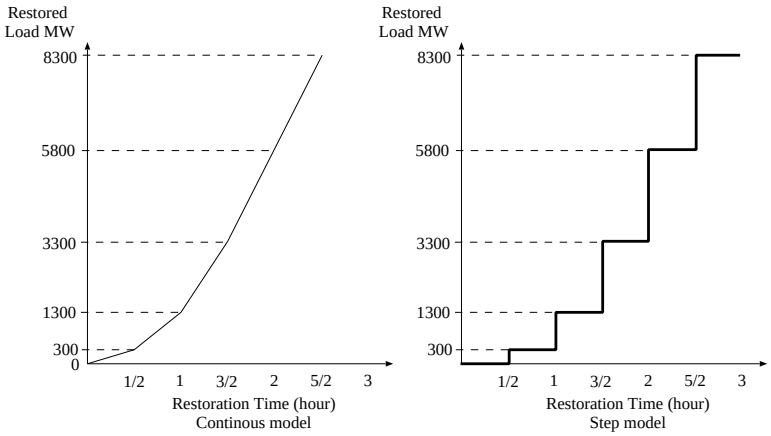

Figure 2.8: Aggregated Restored Load - An example

Chapter 3

Assignment of a Cost for a Load Interruption

The key idea of the *Value of Security Assessor* is the weighing up of the cost of security against the benefit [34]. This is achieved by assessing the costs of

- the base level of security i.e. the system and operating schedule with only planned maintenance outages.
- any corrective actions needed to remove violations of operating limits caused by unplanned outages.

and comparing these with the value of any load that may be lost due to unplanned outages. Obviously, since the aim is to establish the optimum level of security given the probabilities of unplanned events, the decision on which initial operating scenario from a number of possible schedules depends critically on using an appropriate measure of the cost to consumers, utilities and society of an interruption to electricity supply.

In the regime by which the Electricity Supply Industry (ESI) in England and Wales is currently managed, a standard value is attached to the cost of unsupplied electricity. This is known as the "value of lost load" (VOLL). In spite of considerable importance being attached to VOLL, and increasing attention being paid worldwide to the worth of supply continuity or the cost of interruption in order to perform cost-benefit analysis of system reinforcement, there is relatively little discussion in published literature of how suitable values may be found. As this discussion will reveal, the issues involved in deciding on a suitable measure of the cost of an interruption to electricity supply are far from simple and the data available is rather limited.

3.1 Terms used in the literature

Most of the literature concerning the value of lost load or the worth of supply falls into one of two categories:

- *a posteriori* studies of significant outage events to establish their overall costs [9]
- assessments of reliability used in planning and reliability studies [35, 36].

The most useful general result that the studies of specific events have yielded is that interruption costs are substantial. Often, indirect costs such as the costs of civil disorder and the costs of emergency services far out-weighed the direct costs [37]. These studies being specific to given customer mixes, economies and times of day and year, they do not provide information on general methodology.

In the context of system reliability, the question is often asked: "what is a consumer prepared to pay to avoid interruption?" or "how much reimbursement would the consumer accept in the event of being interrupted?". The answer is then used to find some expected cost of interruptions in the course of a year given conventional reliability analysis involving outage rates and rates of repair. As such, the reliability analysis often makes use of only one value for the cost of interruption for the whole year, or perhaps a value that is a function of only one variable, usually interruption duration [36]. As will be seen presently, this approach is limiting when put into the context of power system operation.

Among the terms used in the literature are:

Customer Interruption Cost (CIC): the cost resulting from an interruption as perceived by an individual customer [35].

Customer Damage Function (CDF): function relating some variable, often interruption duration, to a customer's costs caused by supply interruption [37].

Sector Customer Damage Function (SCDF): The normalised costs of supply interruption of a sector of customers, expressed as a function of interruption duration [37].

Composite Customer Damage Function (CCDF): a single function relating the costs of interruption of a particular mix of customers to some variable [37].

Customer Outage Cost (COC): The overall cost of interruptions in a given period e.g. one year [35].

Δ**COC**: "customers' marginal benefit" or "reliability worth" [35].

Value of Lost Load (VOLL): "the value an average customer puts on an unsupplied kWh" [1, 38]. In practice, the figure used may turn out to be, effectively, a function of the CCDF for an interruption of one hour [35].

3.2 Factors influencing interruption costs

The value of lost load figure as currently used in the UK implies that the net value a customer attaches to a supply interruption depends on the amount of *energy* not supplied. This has been disputed in submissions to the *Review of Transmission Security Standards* conducted by the National Grid Company plc in 1994 [1] and by various authors writing in the engineering journals. The principal objection is that "worth of supply" and "energy not supplied" are not directly related [35, 38], and that stress should be given to the cost per disconnection rather than per kWh [39]. Allan illustrates this in [38] by pointing out that the value of being unable to watch a live transmission of an international football match may be considerably higher than that of being unable to use a washing machine. Though the energy consumed by the latter is much higher, watching the live game is time critical. The washing could be done another time, or even by hand.

Extensive consideration of the cost of an interruption reveals that the costs perceived vary substantially from customer to customer. They further vary in the sorts of direct (resulting immediately from the cessation of supply) or indirect (arising from the response to an interruption) impacts the interruptions impose. For example, in [37], direct monetary impacts are said to include

- lost production.
- idle but paid-for resources e.g. labour, capital.
- equipment damage.
- process restart costs.
- spoilage of resources.
- costs of damage to health and/or safety.
- the utility's restoration costs.

Direct social impacts may include inconvenience due to

- lack of transportation.
- loss of leisure time.
- uncomfortable building temperatures.
- personal injury
- fear.

Indirect consequences may be either economic or social and may include

- crime
- necessitation of residential evacuation due to failure of safety equipment at a nearby industrial plant.
- political “fallout”.

It can be seen immediately that these vary from customer to customer. In addition, these costs will vary according to

- time of day.
- the duration of the interruption.
- the availability of advance warning.
- prevailing weather and hence time of year.
- availability of alternative energy sources.

A further, not inconsiderable, factor relating to a customer’s perception of the cost of an interruption will be the reputation of the utility and the recent history of interruptions [1].

3.3 Alternative Models of Value of Lost Load

3.3.1 VOLL currently in use in the ESI

VOLL represents the value an average consumer puts on an unsupplied kWh. Its initial value of £2/kWh adopted in 1989 by the Electricity Supply Industry was apparently derived from the results of a Finnish survey conducted in 1977/78 and has since been pegged to the retail price index (RPI). The NGC has used a value of around £2/kWh (1992/93), as the VOLL of the system. So, this model uses one value as VOLL for all the system.

3.3.2 VOLL derived from customer interruption costs

From the discussion of section 3.2, it is clear that ascertainment of the cost of an interruption is far from simple. The most popular means of gathering any sort of data has, to date, been by customer survey. A drawback of the surveys reported in the literature [1, 40] is that they are normally geared towards reliability studies which make little distinction between times of day and times of year i.e. they seek values which represent sectors of customers for “average” circumstances for different interruption durations. Seeing as the *Value of Security Assessor* will be concerned with operating conditions at a given time and possible unplanned events occurring

within a short time window, to obtain reliable costs for the analysis, the interruption costs for *those conditions* should be estimated.

One survey was conducted as part of a study at UMIST in 1994 [35] and the results are reproduced in table 3.1. In this study, some analysis of the sensitivity of the customer interruption costs to variations in time of day, time of week and season was performed on the UMIST data [41].

Sector	**No. of responses**	**Response rate (%)**	**CIC (£) for interruption duration of:**						
			moment	1 min	20 mins	1 hr	4 hrs	8 hrs	24 hrs
Residential	4014	19.1	-	-	0.19	0.70	4.78	-	-
Commercial	203	4.0	11.47	11.74	49.12	106	345	719	1.0k
Industrial	119	5.7	1.2k	1.5k	2.9k	4.3k	7.6k	12.0k	16.3k
Large user	19	29.2	216k	216k	219k	233k	329k	413k	581k

Table 3.1: UK Customer Interruption Costs (CIC)

A further difficulty is that survey results are very sensitive to the methods used for the calculation of the customer interruption cost. Ideally, the "Willingness to Pay (WTP)[1]" and the "Willingness to Accept (WTA)[2]" approaches are complementary and theory suggests that values derived from these methods should be nearly equal. However, actual valuations consistently yield WTP values significntly less than WTA values [37].

These methods are the so-called "direct methods" of customer survey. In the "indirect" approach of which the "preparatory action method (PAM)[3]" used at UMIST is an example [42], the customer is invited to select an option from one or more lists regarding responses to interruptions.

Clearly, the implication of the existence of so many approaches is that no survey can be regarded as definitive. In addition, limitations in the sample size especially for commercial and industrial customers, as in the UMIST study, constitute a drawback. However, sufficient data to overcome these drawbacks is unavailable and the survey results are reasonable for relative comparisions or studies.

An alternative to the survey approach to putting a value on interruptions suggested during the consultation conducted in the preparation of [1] is simply to leave RECs to "bid" the value of lost load to be used in their areas. However, in this case, the residential customers will always lose out.

A procedure is therefore presented below for using the UMIST survey results to assemble the interruption cost of a given outage. The method is based on that presented in [35].

[1]The WTP approach assesses customer willingness to pay premiums for an assured supply considering their current level of reliability. These premiums are derived from direct monetary estimates by customers or deduced from premiums expressed as percentage of their electricity bill [41].

[2]The WTA approach measures customers' willingness to accept compensation for reduction in reliability from the current level. This is measured as a percentage redunction in electricity bill [41].

[3]The PAM approach ask to customers to choose from a list, the likely mitigating actions they would take to alleviate the impacts of interruptions. The rationale here is that the worth of one good is equivalent to the amount the user is willing to pay for some other goods (preparatory actions) which provide the same benefits as the primary good [41].

1. Find individual customer interruption costs (CICs) as functions of interruption duration t. These are denoted $C_{I,x}(t)$ for each customer x.

2. Normalise the CIC functions to find the customer damage functions (CDFs) either by

   (a) the customer's annual energy consumption so that

   $$C_{E,x}(t) = \frac{C_{I,x}(t)}{E_x} \tag{3.1}$$

   where E_x is the annual energy consumption of customer x in MWh and $C_{E,x}(t)$ is the normalised customer damage function in £/MWh.

   (b) or the customer's peak demand so that

   $$C_{L,x}(t) = \frac{C_{I,x}(t)}{L_x} \tag{3.2}$$

   where L_x is the peak demand (load) of customer x in kW and $C_{L,x}(t)$ is the corresponding normalised customer damage function in £/kW.

   The normalisation is conducted in order to reduce the impact of a few extreme values on the overall value. It should be noted that the customer damage function normalised by peak demand is *not* the cost of energy not served.

3. Find the sector customer damage function (SCDF).

   (a) either

   $$SCDF_{E,y}(t) = \frac{\sum C_{E,x}(t)}{n_{x,y}} \tag{3.3}$$

   where $SCDF_{E,y}(t)$ is the sector customer damage function in £/MWh with costs normalised by annual energy and $n_{x,y}$ is the number of customers in sector y.

   (b) or

   $$SCDF_{L,y}(t) = \frac{\sum C_{L,x}(t)}{n_{x,y}} \tag{3.4}$$

   where $SCDF_{L,y}(t)$ is the sector customer damage function in £/kW with costs normalised by peak load and $n_{x,y}$ is the number of customers in sector y.

   As [35] shown, $SCDF_L(t)$ and $SCDF_E(t)$ are related by

   $$SCDF_{E,y}(t) = \frac{SCDF_{L,y}(t)}{8.76 \times LF_y} \tag{3.5}$$

   where LF_y is the load factor of sector y.

4. Find the VOLL associated to a particular interruption of duration t, according to the knowledge of load factors (LF). So,

(a) if the individual sector load factors are known (LF_y), then the VOLL for each sector y is calculated from the sectorial SCDFs as

$$VOLL_y(t) = \frac{SCDF_{L,y}(t)}{t \times LF_y} \tag{3.6}$$

and then, these sectorial VOLLs are weighted by energy consumption. Such as,

$$VOLL(t) = \frac{\sum VOLL_y(t) E_y}{\sum E_y} \tag{3.7}$$

(b) if the individual sector load factor information is unavailable, then an equivalent SCDF is calculated from sectorial SCDFs appropriately weighted by energy consumption

$$SCDF_{eq}(t) = \sum (SCDF_{L,y}(t) \frac{E_y}{\sum E_y}) \tag{3.8}$$

and the VOLL is calculated from $SCDF_{eq}(t)$ as

$$VOLL(t) = \frac{SCDF_{eq}(t)}{t \times LF} \tag{3.9}$$

The VOLL is in £/kWh

In this way, interruption duration is identified as the most signficant variable upon which the cost of an interruption depends. For use of the method, the approximate duration of the outage must be determined. This is likely to depend on the system conditions which gave rise to the interruption. For example:

1. Supply interrupted to arrest a decline in system frequency may only be restored when sufficient extra generation has become synchronised with the system or when system demand has fallen sufficiently.
2. Supply interrupted to prevent voltage instability may only be restored when demand in the area judged to be vulnerable to voltage collapse has fallen sufficiently, when sufficient extra local generation has become available or when sufficient other extra local MVAr reserves have become available.

It should be recalled that the "value of security" studies will be conducted at a transmission level and that load interruptions, should they arise, will generally be experienced evenly distributed within individual geographical areas. It may be, instead, that different cost functions, i.e. the VOLLs, should be assembled for each geographical area (coinciding with Regional Electricity Company areas) or for each node of the system.

An example

From a survey conducted as part of a study at UMIST in 1994 [35], the CIC costs (table 3.1) and SCDF functions (see table 3.2) for interruption duration for United Kingdom were calculated.

Sector	SCDF (£/kW) for interruption duration of:						
	moment	1 min	20 mins	1 hr	4 hrs	8 hrs	24 hrs
Residential	-	-	0.15	0.54	3.72	7.96	24.92
Commercial	0.99	1.02	3.89	10.65	39.04	78.65	99.98
Industrial	6.15	6.47	14.27	25.26	72.22	120.11	150.38
Large user	6.74	6.74	6.86	7.18	8.86	9.71	13.35

Table 3.2: UK Sector Customer Damage Functions (SCDF) in £/kW

These SCDFs values can be used for the calculation of VOLL of the system, one region or one node. The method is general for any case. For example, the VOLL associated to the bus x will be calculated. The following assumptions are made:

1. The energy consumption by sector at bus x is 35% for the residential sector, 25 % for the commercial sector, 30 % for industrial sector and 10% for the large users sector.
2. The load factors are: 0.333 for the residential sector, 0.40 for the commercial sector, 0.50 for the industrial sector and 0.666 for the large users sector.

Thus, using equations 3.6 and 3.7, the VOLL function for each sector are calculated as in table 3.3 and then the VOLL function for bus x is obtained (see table 3.4).

Sector	VOLL (£/kWh) for interruption duration of:						
	moment	1 min	20 mins	1 hr	4 hrs	8 hrs	24 hrs
Residential	-	-	1.35	1.62	2.79	2.99	3.12
Commercial		153.0	29.17	26.63	24.40	24.58	10.41
Industrial		776.4	85.62	50.52	36.11	30.03	12.53
Large user		606.6	30.87	10.77	3.32	1.82	0.83

Table 3.3: VOLL functions by customer sector

Interruption duration:	1 min	20 mins	1 hr	4 hrs	8 hrs	24 hrs
$VOLL_x$ (£/kWh)	341.4	36.5	23.4	18.2	16.4	7.5

Table 3.4: VOLL bus x - first method -

On the other hand, if the sector's LF are unknown but the bus LF is known then the VOLL is calculated using equations 3.8 and 3.9. For this example, the simultaneous maximum demand corresponds to 80% of residential peak, 40% of commercial peak, 65% of industrial peak and 85% of large user peak. Then, the LF for bus x is 0.42. We suppose that only the last value is known, then the VOLL results are as in table 3.5.

Variations of the method

In equation 3.7, the sectorial VOLLs were weighted by the energy consumption. Alternatively, the weighting factors may be the number of consumers in sector y ot their peak demand. As

interruption duration:	1 min	20 mins	1 hr	4 hrs	8 hrs	24 hrs
$VOLL_x$ (£/kWh)	417.5	42.8	26.5	20.0	17.7	8.0

Table 3.5: VOLL bus x - second method -

stated in [35], the VOLL curves derived have similar characteristics than the VOLL function presented in table 3.4, but their absolute values are very different. So, the weighting factors used for each node may affect their VOLL curves.

Some commentaries

Two methods of normalising CICs are considered since there appears to be no consensus in the literature as to which is more suitable. It has been argued that for short interruptions, to normalise by peak demand is more appropriate since it is the access to instantaneous power that has been lost, while normalisation by energy should be used for longer outages [43].

In addition, it will be noted that calculation of VOLL can be made for differents times of day, day of week or week of year in terms of sectorial weightings, i.e. these will change the mixture factors of SCDFs. These are likely to significantly influence the indirect costs of interruption. The sensitivity of the results to these factors will be investigated.

3.3.3 Customer Outage Costs -COC-

An essential weakness of the VOLL approach is that it calculates the cost of the outage based only on the energu not supplied [39]. A study of the factors affecting the perceived customer interruption costs (CIC) [35, 41] has shown that, in the event of supply interruptions, customers are much more concerned with the inability to use their equipment and the likely damage to this equipment than to the energy they have been unable to consume.

The evaluation of Customer Outage Cost (COC) is an alternative way of valuation of the reliability worth. The COC (in £) due to supply interruptions at the busbar j is given by:

$$COC_j = E_j \times C_j(r_j) \times \lambda_j \tag{3.10}$$

where E_j is the energy consumption at busbar j in the period of evaluation, r_j is the average outage duration, λ_j is the number of interruptions during the period of evaluation and $C_j(r_j)$ is the composite customer damage function (CCDF) calculated from the weighting of SCDF sectorial functions by the sectorial energy consumption at busbar j and expressed in £/kWh. When the load factors by sectors are known, then the $C_j(r_j)$ is given by:

$$C_j(r_j) = \sum_y \left(\frac{SCDF(r_j)}{LF_y \times T} \times \frac{E_{yj}}{\sum E_y}\right) \tag{3.11}$$

where T is the duration of the period of evaluation.

When only the load factor at busbar j is known, $C_j(r_j)$ is given by:

$$C_j(r_j) = \frac{1}{LF_j \times T} \sum_y (SCDF(r_j) \times \frac{E_{yj}}{\sum E_y}) \tag{3.12}$$

In equation (3.12), $\sum E_y = E_j$ at busbar j and $E_j = P_j \times LF_j \times T$. Equation (3.10) then becomes:

$$COC_j = P_j \times SCDF_{eq}(r_j) \times \lambda_j \tag{3.13}$$

where P_j is the peak load at busbar j during the evaluation period T.

3.4 VOLL modelling in the *Value of Security Assessor*

It must recalled that the principal objectives of the research grant do not include the arrival at a decision on what is the best means of costing interruptions to supply. That decision should rest with whoever uses the value of security analysis procedure. All that can realistically be provided is the flexibility of doing the costing by whatever means is chosen and to enable some degree of sensitivity analysis to be performed with regard to different costing approaches. This will be accomplished simply by keeping track of the amount and location of load lost in the course of the Monte Carlo simulation and then costing it all in post-processing by the chosen means after sufficient samples have been generated.

The current software implementation of the *Value of Security Assessor* provides this flexibility of modelling.

The discussion has introduced the need for careful consideration of how interruptions to supply are to be costed in the *Value of Security Assessor* under development at UMIST. Some of the factors contributing towards the cost of interruption have been outlined and in so-doing, some of the drawbacks of the "value of lost load" currently used in the UK electricity supply industry have been pointed out.

A proposal has then been presented for what might constitute a useful first approximation of a realistic value of lost load in the value of security analysis. Results of the application of *Value of Security Assessor* are also presented.

3.5 Effect od Duration of Interruption on the VOLL computation

In order to illustrate the effect of restoration load schemes in the evaluation of security value, let us use the following example. Consider an unserved load P_R of 300 MW that is restored at a rate of m_1 (600 MW/hour). The restoration process starts t_1 (10 min) after the failure occurs. Then, the first step is the calculation of the total time of restoration. This is given by:

$$T_{restoration} = \frac{P_R}{m_1} = 30min \tag{3.14}$$

Thus, the total load is restored between t_1 and $t_1 + T_{restoration} = t_2$. It is supposed, that the load restoration can be realised by one of the restoration schemes of figure 3.1:

1. The total load (P_R) is restores a t_2
2. The load is restored in n (6) steps of $\frac{P_R}{n}$ (50 MW)
3. The load is restored continously at the rate of m_1

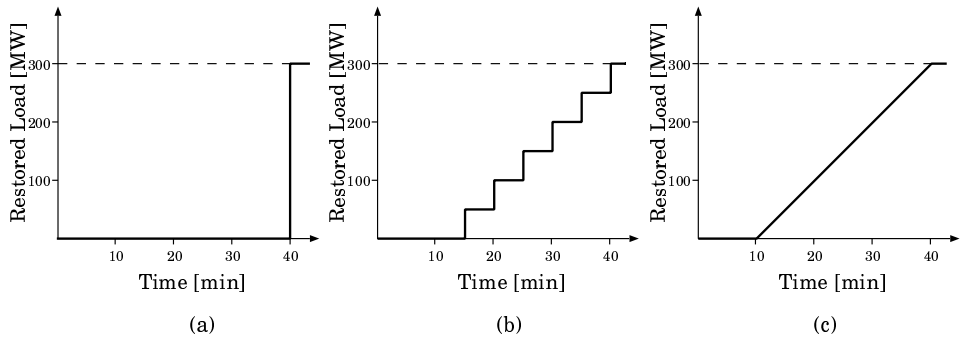

Figure 3.1: Schemes of Restoration Load

As, it was showed in [44], VOLL depends on the duration of the load interruption.

The total cost, for each scheme, is given as follow:

- Scheme 1. As the total load is restored at t_2, then:

$$TC = voll(t_2) \times t_2 \times P_R \tag{3.15}$$

  where TC is the total cost (in £) associated with the load interruption.

- Scheme 2. Each step (k) of the restored load has an associated time of restoration (t_k) and, hence an associated VOLL ($voll(t_k)$). Then, TC is given by:

$$TC = \sum_{k=1}^{n} (voll(t_k) \times t_k \times \frac{P_R}{n}) \tag{3.16}$$

  where t_k is given by

$$t_k = t_1 + \frac{T_{restoration}}{n} k \tag{3.17}$$

  Then,

$$TC = \sum_{k=1}^{n} (voll(t_k) \times (t_1 + \frac{T_{restoration}}{n} k) \times \frac{P_R}{n}) \tag{3.18}$$

- Scheme 3. The TC for this case is obtained from the last equation by making $n \to \infty$. Thus, TC is given by

$$TC = \int_{t_1}^{t_2} (voll(t) \times t \times m_1) dt \tag{3.19}$$

In the evaluation of the integral, it is important to take into account that the VOLL function is defined differently for different periods of time, i.e. $\frac{dvoll}{dt}$ is not continuous.
If a continuous first derivative voll function between t_1 and t_2 given by:

$$voll(t) = n \times (t - t_1) + voll(t_1) \qquad t_1 < t < t_2 \tag{3.20}$$

is assumed, TC will be given by:

$$TC = \int_{t_1}^{t_2} ((n \times (t - t_1) + voll(t_1)) \times t \times m_1) dt \tag{3.21}$$

or

$$TC = n \times m_1 \times (t_2^3 - t_1^3) + \frac{m_1(voll(t_1) - n \times t_1)}{2}(t_2^2 - t_1^2) \tag{3.22}$$

Chapter 4

Weather Modelling

4.1 Introduction

It is clear that in an operational timescale, the initial weather is quite predictable, with some random variation. In addition, studies of succeeding conditions will simply be further 'operation snapshots' which will be examined nearer to the time, but for which the weather can also be fairly predicted. In consequence, the weather needs not be regarded as a random variable, and the effects of predicted weather patterns can be reflected in the probabilities of different random events on the power systems.

Section 4.2 presents the basical concepts used in the weather modelling and the proposed modelling based on the definition of discrete weather states. Section 4.3 describes the implementation of this feature in the *Assessor.*

4.2 Weather concepts

As it is stated in [45] and [46], failure rates of some components such as overhead transmission lines depend on the weather conditions to which they are exposed. The weather aspects that affect the performance of the transmission lines could vary from one region to another and could be different from one season to another. Generally, wind speed, lighning, precipitation and others are the main contributing factors affecting the reliability of the system. In addition, the severity of the weather scenario is determined by the intensity of the constituent aspects; therefore the actual number of severity levels needed to represent weather aspects depends not only on the available data but also on the appreciation of the intensity of the individual constituent factors and their impact on the transmission network.

In previous work done at UMIST ([45], [46]), a weather model was developped to evaluate its effect in the composite system reliability evaluation. This model is general in nature and recognises the following characteristics of weather:

1. Constituent factors or variables of weather
2. Severity levels of individual variables
3. Interdependency between variables
4. Adversity or stress level of each weather state
5. Regional variation
6. Nonstationary attributes of the weather scenario
7. Seasonal variation of weather

It should be noted that the number of attributes, variables and severities can be easily reduced to include only those considered necessary and viable in any particular application with the available data. On the other hand, this model considers the weather as a random variable. For application in the operation timescale, the weather is considered to be predictable.

4.2.1 Definition of weather states

Different weather states can be defined for each region and each season. The number of variables needed to describe the weather could be function of the region, but in a general way the weather states can be defined by the following variables:

1. Wind speed (S)
2. Lightning (L)
3. Precipitation (P)

Therefore, a weather condition can be represented by the following relationship:

$$W = f(S, L, P) \tag{4.1}$$

To determine the effect of weather on the transmission network, the behaviour of each variable should be recognised. The number and kind of severity levels needed to represent the impact of each variable are determined by the relevant meteorological data.

As it is stated in [45], the weather and therefore the stress level that a transmission network can encounter could be associated with a continuous distribution. However, it is much easier to develop a discretised distribution from a data collection point of view. For this discretisation, severity levels are defined for each variable; for example, the number of severity levels for the S, L and P variables are s, l and p respectively. Therefore the total number of states required to represent the weather will be $s.l.p$, and equation (4.1) becomes:

$$W = f(S_i, L_j, P_k) \qquad i \in s, j \in l, k \in p \tag{4.2}$$

Table 4.1 shows the weather states for one region in a specified season. It is assummed that the weather conditions for this region are specified using three variables: wind speed, ligthning and precipitation. For each variable, three levels of severity are defined: low, medium and high. Hence, the number of weather states will be 27. These states are obtained by combination of the severity levels of each variable, as shown in the table.

Weather State	Speed Wind Level	Lightning Level	Precipitation Level
1	Low	Low	Low
2	Low	Low	Medium
3	Low	Low	High
27	High	High	High

Table 4.1: Weather states

However, some of these general weather categories are little frequent. In such case, it is better to identify special weather states that are more likely to lead to increased outage rates. The special weather states can be defined, for example:

- Normal weather
- Thunderstorm
- Freezing rain/wet snow
- High winds
- Dry spell followed by fog

Alternatively, a more simple definition can be employed using two possible states as function of data availability:

- Adverse weather
- Normal (or non-adverse) weather

4.2.2 Failure stress

The weather state 1 from table 4.1 (i.e. low severity of all varieties) represents the normal state for which the failure rates of the transmission components are specified. After determining the weather state in function of variables' severity levels, it is necessary to calculate the impact of this state on the failure rate of the transmission components. This impact is usually measured quantitatively in terms of increasing failure rate of individual transmission components. For example, if the weather is in the jth state, then the failure stress (FS) on the nth component can be specified as:

$$FS_j^n = \frac{\lambda_j^n}{\lambda_1^n} \tag{4.3}$$

where:

FS_j^n = failure stress on nth component in the jth weather state
λ_j^n = failure rate on nth component in the jth weather state
λ_1^n = failure rate on nth component in the normal weather state (1).

The calculation of the failure stress uses the concept of the proportion of failures that is explained in the following section.

4.2.3 Proportion of failures concept

The proportion of failures ocurring in each weather state [4] is defined as:

$$F_i^n = \frac{\frac{T_i}{T}\lambda_i^n}{\overline{\lambda}^n} \tag{4.4}$$

where:

F_i^n is the proportion of failures for the nth component in the ith weather state. It is such that $\sum_{i=1}^{N} F_i^n = 1.0$
$\overline{\lambda}^n$ is the average failure rate of the nth component
λ_i^n is the failure rate of the nth component in the ith weather state
T_i is the duration of the ith weather state
$T = \sum_{k=1}^{N} T_k$ is the duration of all weather states (N)

On the other hand, the failure rate for the normal weather state is known. In addition, from the data collection, the proportion of failure and the duration of the normal state and the total time of the sample (history) can be calculated. Hence, the average failure rate can be calculated using equation 4.4. Thus, the failure rate for the nth component is given by:

$$\overline{\lambda}^n = \frac{T_1}{T}\frac{\lambda_1^n}{F_1^n} \tag{4.5}$$

The failure rate of the nth component for the ith weather state is obtained replacing 4.5 in 4.4 and reordering. Thus:

$$\lambda_i^n = \frac{T_1}{T_i}\frac{F_i^n}{F_1^n}\lambda_1^n \tag{4.6}$$

It is convenient to use a normalised time T_i such that:

$$\sum_{i=1}^{N} Tn_i = 1.0 \tag{4.7}$$

Equation 4.6 then becomes:

$$\lambda_i^n = \frac{Tn_1}{Tn_i}\frac{F_i^n}{F_1^n}\lambda_1^n \tag{4.8}$$

4.3 Modelling in the *Assessor*

4.3.1 Prevailing weather scenario

In the *Assessor,* the value of security evaluation is realised either for one-hour dispatch or for one-day schedule. For the first condition, a prevailing weather scenario can be assumed and its effect on the failure rate is incorporated using equations 4.5 to 4.8.

A point of dificulty for implementation in the program is to define if the component's failure rate data (λ) is related either to the failure rate for normal weather condition (λ_1) or to the average failure rate ($\overline{\lambda}$).

When the known value is $\overline{\lambda}$, the adjustment equation to apply is:

$$\lambda_i = \frac{F_i}{Tn_i}\overline{\lambda} \tag{4.9}$$

(Since this adjustment is applicable to every component in the network, the superscript has been removed to clarify the equation)

When the known value is λ_1, the adjustment equation to apply is:

$$\lambda_i = \frac{F_i}{Tn_i}\frac{Tn_1}{F_1}\lambda_1 \tag{4.10}$$

In order to give flexibility to the *Assessor,* the adjustment equation uses three factors for each area:

1. The normalised duration of the weather state i (Tn_i)
2. The failure proportion for the weather state i (F_i)
3. An adjustment factor equal to:
   - 1.0 if the known value is $\overline{\lambda}$
   - $\frac{Tn_1}{F_1}$ if the known value is λ_1

These three data will be given by area and they will be different for each season of the year. Additional data processing must be performed to determine the duration and failure proportion factors for each weather state for each area and each season, as explained in section 4.4.

4.3.2 Traversing weather

The evaluation of value security in the one-day case requires that the traversing characteristics of the weather be included in the modelling. As it is stated, the weather can be considered as

predictable in the operation timescale. An alternative way for the required modelling is to use a succession of 'operation snapshots'. Then, each snapshot will be modelled as a prevailing weather condition, as presented in the last section.

4.3.3 Failure rate adjustment of multiregional lines

The adjustment of λ has been presented for the case of only one region traversed by a line. When a line traverses several regions, each one will have a different adjustment. So, the failure rate of the line will be adjusted by a factor that ponders the adjustment factors of the traversed regions based in the length of the line on each region. Thus, equation 4.9 becomes:

$$\lambda_i = \overline{\lambda} \sum_{k=1}^{nr} (\frac{F_i}{Tn_i})_k \frac{l_k}{L} \tag{4.11}$$

where nr is the total number of traversed regions by the line and $(\frac{F_i}{Tn_i})_k$ is the adjustment factor associated to the region k. l_k is the length of the line on region k and $L = \sum_{k=1}^{nr} l_k$ is the total length of the line.

4.4 Data requirements

Two type of data are necessary for weather modelling: historic data and forecasted weather data. The historic data allows the definition of weather states for each area of the system and their associated factors (duration and proportion of failures). The predicted data allows the determination of the forecasted weather state and so the determination of factors to use in the *Assessor.*

The historic data will be collected by:

- Area or region
- Season or period (summer, fall, winter and spring)

If the system is divided in 5 regions and 4 periods of weather are considered, then 20 data groups must be collected.

From each data group, i.e. for each region-area entity; the following processes must be developed:

1. Selection of variables that determine different weather states
2. Selection of severity levels for each variable
3. Definition of weather states

4. Calculation of duration and failure proportion of each weather state. This allow the construction of a table that relates weather state and the needed factors for value of security evaluation (this table will be called the Weather Table).

The predicted data will be used for determining the weather state of the region according to the previously defined weather states. The procedure is:

1. Identify the severity level of each forecasted variable
2. Determine the predicted weather state
3. Obtain the duration and failure proportion factors for the value of security evaluation from the Weather Table

Chapter 5

Key Ideas from Statistics

This section summarises key concepts and terms from statistics that are necessary to understanding and using Monte Carlo methods. Fuller descriptions may be in standard statistics texts such as [47] and [48].

A process that is subject to some degree of randomness is known as a *stochastic process.* Every time the process (or *experiment* as it sometimes known) takes place or is carried out, there will be a number of possible *outcomes.* After observing the process many times, it may be possible to recognise that some outcomes happen more often than others and that it is possible to assign probabilities to them.

The *sample space* of the stochastic process or experiment contains all the possible outcomes and may be denoted by S (corresponding to the "universal set"). Each realization of the experiment, known as a *trial*, will yield one outcome in the space S. The outcome may then be thought of as being or leading to an *event* which is either a categorization or consequence of the outcome. Since events can include sets of outcomes or more general consequences, many texts talk about events rather than outcomes.

In order to make sense of the different outcomes, real numbers are very often assigned to them. These are known as *random variables* since they can vary over the real axis and their values depend on processes subject to some randomness. Here, they will be denoted by X. A value of a random variable (i.e. a number generated by some random process) will be denoted by X'.

5.1 Expectation

If there are a finite number of possible outcomes and a finite *population* of trials, it will be possible to find the *population mean* value of some random variable. If there are N trials in the population, the population mean will be

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X'_i. \tag{5.1}$$

The mean is also known as the *expected value* or *expectation.*

If information on m possible outcomes were to be gathered, the mean could be found by

$$\mu = \sum_{i=1}^{m} P_i X_i' \tag{5.2}$$

where P_i is the probability of X_i' occurring as an outcome.

Very often, however, the whole population of trials cannot be observed and a sample of n trials of the population will be explored. This allows the calculation of the *sample mean* denoted by $\overline{X}$:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i'. \tag{5.3}$$

5.2 Distributions

Since the process being observed is stochastic, it would not be expected that all the random variable being recorded would have the mean value at each trial. There will be some scatter about the mean. The frequency by which each value occurs over a number of trials can be represented on a frequency function of the random variable. This is more commonly known as a *probability density function* or p.d.f. as it represents the probability that particular values of the random variable will be observed. When the random variable is continuous, this will roughly equal zero for any exact value of the random variable. It is therefore more usual to talk about the probability of the random variable lying between two bounds. This will be the area under the p.d.f. between those bounds i.e.

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx \tag{5.4}$$

where $f_X(x)$ is the p.d.f. of X.

If X is continuous, then an integral form of equation (5.2) could be written such that

$$\mu = \int_{-\infty}^{\infty} f_X(x) x dx. \tag{5.5}$$

One common form of p.d.f., a "normal distribution", is shown in figure 5.1 (that shown is the distribution of some measurement z^{meas} subject to Gaussian noise).

Another common distribution is that representing a fixed number of trials where each trial has only two possible outcomes - "success" or "failure", "hit" or "miss". If there are n trials and h of these are successful, there will have been $n-h$ failures. If p is the probability of getting a success in one trial, then $1-p$ is the probability of getting a failure. The probability of getting h successes *and* $n - h$ failures for a given order of successes and failures is $p^h(1-p)^{n-h}$. The number of

possible orders of h successes from n trials (i.e. combinations of successes and failures) is C_h^n or $\left(\begin{array}{c} n \\ h \end{array} \right)$ which is

$$\left(\begin{array}{c} n \\ h \end{array} \right) = \frac{n!}{h!(n-h)!}. \tag{5.6}$$

Thus the probability of getting h successes in n independent trials is

$$f(h) = \frac{n!}{h!(n-h)!} p^h (1-p)^{n-h} \quad \text{for } h = 0, 1, 2, \ldots, \text{or } n. \tag{5.7}$$

This is known as the *binomial distribution.* Its mean is simply

$$\mu = np. \tag{5.8}$$

Another representation of frequency information is a *cumulative distribution function* or c.d.f.. This describes the probability that X has a value equal to or less than some other value, say y and is

$$F_X(y) = P(X \leq y) = \int_{-\infty}^{y} f_X(x) dx \tag{5.9}$$

5.3 Variance

How values of the random variable from different trials vary can be described in terms of the difference between X_i from trial i and the mean. The overall quality of the random process (i.e. how random it is) can be seen from the differences between each X_i and the mean. Simply adding them together would not be very useful as they would sum to zero. There are also problems associated with adding the absolute values of the differences so adding the squares of the differences has been adopted as a standard. This is known as the *variance*

$$\sigma^2 = \sum_{i=1}^{m} P_i (X_i' - \mu)^2 \tag{5.10}$$

$$= \frac{1}{N} \sum_{i=1}^{N} (X_i' - \mu)^2. \tag{5.11}$$

Another way of calculating this is

$$\sigma^2 = \overline{X^2} - \overline{X}^2. \tag{5.12}$$

The variance of a continuous random variable is

$$\sigma^2 = \int_{-\infty}^{\infty} f_X(x)(x-\mu)^2 dx \tag{5.13}$$

In practice, a divisor of $n-1$ is used for the sample variance s^2 i.e.

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i' - \overline{X})^2. \tag{5.14}$$

This is because s^2 is an *estimator* of σ^2 (σ^2 is the variance of the population and s^2 is the variance of a sample of it which may be taken as an estimate of the overall variance). Using $n-1$ makes it unbiased i.e. the average of a series of such estimators will equal the quantity they are estimating. The idea of an estimator is quite important in Monte Carlo simulation. For example, $\overline{X}$ is an estimator of μ, an idea which can be expressed by

$$E(\overline{X}) = \mu. \tag{5.15}$$

The computing formula for the sample variance is then

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1}. \tag{5.16}$$

The variance of a p.d.f. $f_X(x)$ can be worked out by equation (5.10) recognising that $P_i = f_X(X_i')$ so that

$$\sigma^2 = \sum_{i=1}^{m}(X_i' - \mu)^2 f_X(X_i'). \tag{5.17}$$

The variance of a binomial distribution is

$$\sigma^2 = np(1-p) \tag{5.18}$$

where p is the probability of a "success" in one trial and n is the number of trials.

It may be usefully noted that

$$\text{var}(u_X(x)v_X(x)) = \text{var}(u_X(x))\text{var}(v_X(x)) \tag{5.19}$$

$$\text{var}\left(\frac{u_X(x)}{v_X(x)}\right) = \frac{\text{var}(u_X(x))}{\text{var}(v_X(x))}. \tag{5.20}$$

Related to the variance is the positive square root of the variance or *standard deviation*, σ for the population and s for a sample. Another useful term is the *coefficient of variation* which, for a population, is

$$\nu = \frac{\sigma}{\mu} \tag{5.21}$$

or for a sample is

$$CV = \frac{s}{\overline{X}}. \tag{5.22}$$

This allows the scatter of random variables to be compared regardless of the units in which they are measured.

5.4 Covariance

If two separate random variables X and Y are defined on the same probability space, the relationship may be described in terms of the *covariance* or the *correlation coefficient*. The covariance is defined by

$$\text{cov}[X, Y] = \sigma_{X,Y} = E[(X - \mu_X)(Y - \mu_Y)] \tag{5.23}$$

while the correlation coefficient $\rho[X, Y]$ or $\rho_{X,Y}$ is

$$\rho_{X,Y} = \frac{\text{cov}[X, Y]}{\sigma_X \sigma_Y} \tag{5.24}$$

provided that cov$[X, Y]$, σ_X and σ_Y all exist and $\sigma_X > 0$ and $\sigma_Y > 0$.

As a measure of the linear relationship between X and Y, cov$[X, Y]$ will be positive when $X - \mu_X$ and $Y - \mu_Y$ tend to have the same sign with high probability and negative when they tend to have opposite signs with high probability. If there is little relationship, the covariance will be roughly zero.

The magnitude of the covariance depends on the variability of each of X and Y. The correlation coefficient removes this dependence by dividing the covariance by the product of the individual standard deviations so that $-1 \leq \rho_{X,Y} \leq 1$.

Similar to the expression given in equation (5.12) is the following expression for computing the coefficient of correlation of a sample r:

$$r = \frac{S_{XY}^2}{S_{XX} \cdot S_{YY}} \tag{5.25}$$

where S_{XX} is the sum of squares of the deviation of X from the mean $\overline{X}$ or

$$S_{XX} = \sum_{i=1}^{n} X_i'^2 - \frac{1}{n}\left[\sum_{i=1}^{n} X_i'\right]^2, \tag{5.26}$$

S_{YY} is the sum of squares of the deviation of Y from $\overline{Y}$ or

$$S_{YY} = \sum_{i=1}^{n} Y_i'^2 - \frac{1}{n}\left[\sum_{i=1}^{n} Y_i'\right]^2 \tag{5.27}$$

and S_{XY} is n times the "first product moment" [48] i.e.

$$S_{XY} = \sum_{i=1}^{n} X_i' Y_i' - \frac{1}{n}\left[\sum_{i=1}^{n} X_i'\right]\left[\sum_{i=1}^{n} Y_i'\right] \tag{5.28}$$

$$= \sum_{i=1}^{n} (X_i' - \overline{X})(Y_i' - \overline{Y}). \tag{5.29}$$

S_{XY} is more commonly seen when divided by $n - 1$ and known as the *sample covariance*. The sample covariance is then most easily computed using equation (5.28) and dividing S_{XY} by $n - 1$.

Just as the mean, variance and standard deviation can be defined for random variables, they can also be defined for functions of random variables.

5.5 Estimate error

One of the most important distributions in statistics is the *normal* or *Gaussian distribution* (figure 5.1). This is described by two parameters, μ and σ^2 such that

$$g_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}. \tag{5.30}$$

It is a key component of the *central limit theorem* which says that the *sampling distribution* of the mean has an approximately normal distribution when n is large i.e.

$$f_{\overline{X}}(x) \approx g_{\mu,\sigma^2/n}(x) \quad \text{for } n \text{ large} \tag{5.31}$$

where $f_{\overline{X}}(x)$ is the p.d.f of $\overline{X}$.

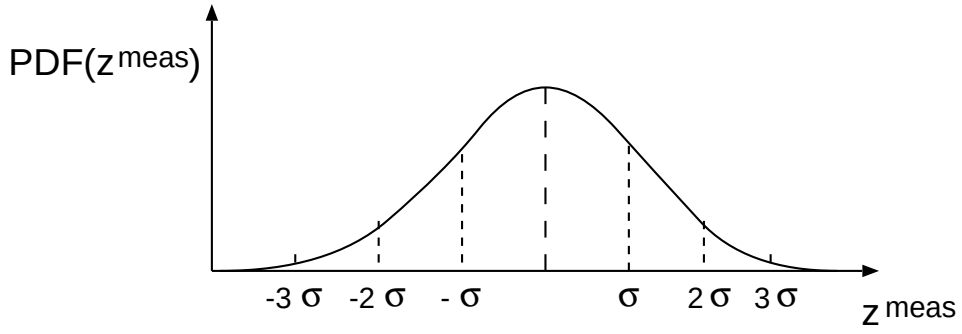

Figure 5.1: An example of a normal distribution

It can be shown that this represents the variation of the expectation (or mean) in a Monte Carlo simulation if the population is infinite. This is to say that the probability of finding a particular value for the mean of a random sample of n trials from a population is represented by a normal distribution. Moreover, the variance of this normal distribution describing the variation of the mean (or expected value) of a sample of size n is related to the variance of the population from which the n variables are taken, and n. This relationship is

$$\sigma^2_{\overline{X}} = \frac{\sigma^2_X}{n}. \tag{5.32}$$

where σ^2_X, or simply σ^2, is the variance of the population of values of the random variable X and $\sigma^2_{\overline{X}}$ is the variance of the mean (or expected value) $\overline{X}$ of X. The *standard error of the mean* is then defined to be

$$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}. \tag{5.33}$$

The mean of the sampling distribution of $\overline{X}$ is

$$\mu_{\overline{X}} = \mu_X \tag{5.34}$$

where μ_X, or simply μ, is the population mean of X.

If the population is finite, the standard error is [48]

$$\sigma\overline{X} = \frac{\sigma_X}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \tag{5.35}$$

where N is the population size. When the sample represents less than about 5% of the population, the term $\sqrt{\frac{N-n}{N-1}}$, known as the finite population correction factor, can be omitted as it will be very close to 1.

5.6 Standard scores and confidence

Since the purpose of the simulation is to find the expectation, it is not known in advance, so use can be made of the normal distribution to assign a probability that the estimated expectation is within a given range of the true one. This in turn enables an estimate to be made of the number of samples needed to find the expectation to a given specification.

Another quantity referred to in some texts is the *standard unit, standard score* or *z-score*. Yet another name that may be seen is "normal score". It is used to re-express the original scale of X to enable the forming of a *standard normal distribution* which has $\mu = 0$ and $\sigma = 1$. The standard unit is referred to as z and is

$$z = \frac{X - \mu}{\sigma}. \tag{5.36}$$

A particular value of z is z_α which is the value of z above which the area of the standard normal distribution is equal to α i.e. if the standard normal distribution is given by $g(z)$,

$$\int_{z_\alpha}^{\infty} g(z)dz = \alpha = P(z \geq z_\alpha). \tag{5.37}$$

Likewise, the areas to the right of $z_{\alpha/2}$ and the left of $-z_{\alpha/2}$ under a standard normal curve are both $\alpha/2$. The area between is $1 - \alpha$ (see figure 5.2).

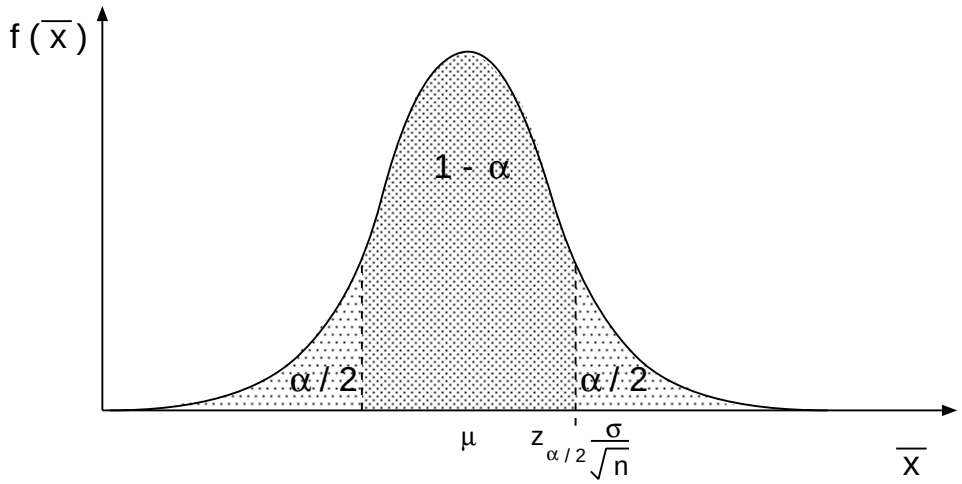

Figure 5.2: Sampling distribution of the mean

Making use of the central limit theorem, for large random samples from infinite populations, the sampling distribution[1] of $\overline{X}$ is roughly a normal distribution with $\mu_{\overline{X}} = \mu$ and $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$. This

[1] A sampling distribution is the distribution of a random variable within a sample taken from the the population

can be described in terms of the standard normal distribution $g(z)$ where

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{5.38}$$

is a value of a random variable having approximately the standard normal distribution.

Most importantly, the probability is $1 - \alpha$ that $\overline{X}$ found from a large sample will differ at most from the mean of the population μ by $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ i.e.

$$P\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha. \tag{5.39}$$

Thus, a *confidence interval* may be defined such that the probability of the true mean lying within that interval is the confidence level γ such that

$$\gamma = P(\overline{X} - L \leq \mu \leq \overline{X} + L) = \int_{\overline{X}-L}^{\overline{X}+L} g(x)dx \tag{5.40}$$

where

$$L = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{5.41}$$

$$\gamma = 1 - \alpha \tag{5.42}$$

and $g(x)$ is the Gaussian distribution. In natural language, this says that we can be $(1-\alpha)\times 100\%$ confident that the estimate of the mean $\overline{X}$ lies within $\pm L$ of the true population mean μ. (Note that $1 - \alpha = \gamma$ is also sometimes known as a confidence coefficient or a *degree of confidence*).

The above relationships are most useful in allowing a decision to be made about the size of sample needed to obtain an estimate of a given accuracy with a given degree of confidence. If the error that will be tolerated is L and $\gamma = 1 - \alpha$ is the degree of confidence wanted, $z_{\alpha/2}$ can be found from a standard set of tables (a part of which is reproduced in table 5.1) and the expression

$$n = \frac{\sigma^2 z_{\alpha/2}^2}{L^2} \tag{5.43}$$

used to find a suitable number samples.

Clearly, from equation (5.43) and table 5.1 it can be seen that the required sample size increases with the degree of confidence but is inversely proportional *to the square* of the confidence interval.

γ	0.90	0.95	0.99	0.999
α	0.10	0.05	0.01	0.001
$z_{\alpha/2}$	1.645	1.960	2.576	3.291

Table 5.1: Relationship of degree of confidence and $z_{\alpha/2}$ for some useful values of γ

This is an illustration of the *law of large numbers.* One of its formal statements says that for a random variable x with an expectation $\overline{X}$ and a sufficiently small positive number ε,

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i' - \overline{X}\right| < \varepsilon\right) = 1.0 \tag{5.44}$$

which means to say that when n is very large, the sample mean of X approaches its true mean with a very large probability.

An inevitable problem with equation (5.43) is the presence of the population variance σ^2 which will not be known. Even the sample variance s^2 which may make a good approximation for σ^2 will not be known until a number of trials have been carried which is unhelpful for knowing how many trials to do. A commonly used compromise is to do some small pilot studies to obtain a reasonable value for s^2 and consequently a required number of trials n.

If the sample variance s^2 is used as an approximation of σ^2, equation (5.43) may be used as a test for "convergence" of a sample to pre-defined confidence interval and confidence limit i.e. once n is greater than, say, 20, stop taking any more samples once

$$s^2 \leq z_{\alpha/2}^2 \frac{L^2}{n}. \tag{5.45}$$

The question may arise: can confidence limits be defined for *small* samples with, say, less than 30 observations? When the sample is small and the population variance is unknown, the sample variance must be used. Under these circumstances, the sampling distribution cannot be approximated by a normal distribution. However, if it can be assumed that the populations being sampled have roughly the shape of a normal distribution, the confidence limits can be based on the statistic

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}}. \tag{5.46}$$

This is the value of a random variable having *Student's t distribution.* While it is similar to the normal distribution—bell-shaped and with a zero mean but with longer tails—it's exact shape depends on the number of *degrees of freedom.* Using a number of degrees of freedom equal to $n-1$ where n is the sample size, the t distribution can be adopted to find the value $t_{\alpha/2}$ which replaces $z_{\alpha/2}$ in the calculation of a confidence interval.

It may be further noted that confidence intervals can be expressed for variance [49]. Using the definition of the variance of a sample of n trials from a normally distributed population as

$$s^2 = \frac{\sum_{i=1}^{n}(X_i' - \overline{X})^2}{n-1}, \tag{5.47}$$

and recognising that s^2 is an unbiased estimate of σ^2, it can be shown that the sampling distribution of $(n-1)s^2/\sigma^2$ is $\chi^2(n-1)$. Thus,

$$P\left[\chi^2(1-\alpha/2; n-1) \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi^2(\alpha/2; n-1)\right] = 1-\alpha \tag{5.48}$$

where $\chi^2(1-\alpha/2; n-1)$ is the upper $100(\alpha/2)$ percentage point of the $\chi^2(n-1)$ distribution.

Rearranging equation (5.48) enables the expression of a confidence interval for σ^2 such that

$$P\left[\frac{n-1)s^2}{\chi^2(\alpha/2;n-1)} \leq \sigma^2 \leq \frac{n-1)s^2}{\chi^2(1-\alpha/2;n-1)}\right] = 1-\alpha. \tag{5.49}$$

Confidence intervals for σ^2 (and hence for σ) are usually quite wide unless the sample size is very as an estimate of the a variance will always be less precise than that of a mean as an average of squared observations is more variable than an average of observations.

Chapter 6

Monte Carlo Algorithms

Perhaps the best way of describing how Monte Carlo schemes work is to show examples. A simple example is that of calculating the integral of a function. When the function is not regular, a Monte Carlo based estimation of the integral becomes more attractive than using a conventional numerical technique. Two such Monte Carlo estimation algorithms are offered - one which can be understood as returning 1 or 0 "hits" or "misses" in the required area, the other which returns an expected value from a p.d.f..

A third example of use of Monte Carlo simulation is offered in section 8 in terms of its application to reliability analysis of a power system.

6.1 The "Hit or Miss" method

The "Hit or Miss" method is based on a graphical interpretation of an integral as an area. It is possibly the easiest the understand and was originally the one usually propounded in discussion of Monte Carlo techniques though it is the least efficient [2] (a proof of this is offered in [50] as well as [2]).

6.1.1 "Hit or Miss" applied to a one-dimensional integral

The bounded integral of a one-dimensional function $h(x)$ is to be found [50] (figure 6.1). It may be supposed that

$$0 \leq \quad h(x) \quad \leq c \tag{6.1}$$

$$a \leq \quad x \quad \leq b. \tag{6.2}$$

The area of the rectangle bounded by $y = 0$, $y = c$, $x = a$ and $x = b$ is denoted A and is

$$A = \{(x, y) : a \leq x \leq b, 0 \leq y \leq c\}. \tag{6.3}$$

A random vector (X, Y) is supposed to be uniformly distributed over A with a p.d.f. $f_{XY}(x, y)$ such that

$$f_{XY}(x,y) = \begin{cases} \frac{1}{c(b-a)} & \text{if } (x,y) \in A \\ 0 & \text{otherwise} \end{cases} \tag{6.4}$$

i.e. $f_{XY}(x, y) = 1/\text{area}(A)$ when (x, y) falls within A, and 0 outside A. This in turn means to say the the probability of hitting any spot inside A is the same and is inversely proportional to the area of A, while there is no chance of (x, y) falling outside A.

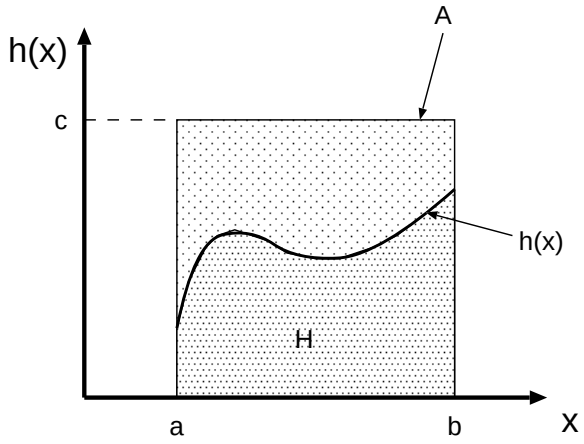

Figure 6.1: Illustration of "Hit or Miss" method

The task now is to find the probability p of a random vector (X, Y) falling in the area under the curve $h(x)$. If H is defined such that $H = \{(x, y) : y \leq h(x)\}$, then

$$\text{area under } h(x) = \text{area}(H) = \int_a^b h(x)dx. \tag{6.5}$$

Now it can be seen that

$$p = \frac{\text{area}(H)}{\text{area}(A)} = \frac{\int_a^b h(x)dx}{c(b-a)} = \frac{I}{c(b-a)}. \tag{6.6}$$

Let it be assumed that n independent random vectors $(X_1', Y_1'), (X_2', Y_2'), \ldots, (X_n', Y_n')$ are generated. p can be estimated by

$$\hat{p} = \frac{n_H}{n} \tag{6.7}$$

where n_H is the number of times that $Y_i' \leq h(X_i')$ for $i = 1, 2, \ldots, n$ i.e. the number of "hits" in area H, while $n - n_H$ is the number of "misses" where $Y_i' > h(X_i')$ for $i = 1, 2, \ldots, n$.

It follows, then, that equation (6.6) can be used to estimate the integral I i.e.

$$I \approx \theta = c(b-a)\frac{n_H}{n}. \tag{6.8}$$

A sample of n trials is taken from $f_{XY}(x, y)$, the number of hits n_H is counted and equation (6.8) is applied to find θ.

It remains for the required precision of the estimator θ and the required number of trials to be defined.

6.1.2 Precision of the "Hit or Miss" method

Since θ is an estimator of I and each of the trials is a *Bernoulli trial*[1], the following equalities are true as c, b, a and n are known and not estimated:

$$E(\theta) = c(b-a)E\left(\frac{n_h}{n}\right) = c(b-a)\frac{E(n_H)}{n} = pc(b-a) = I. \tag{6.9}$$

(This implies that θ is an unbiased estimator of I).

The variance of the estimated quantity $\hat{p}$ is

$$\sigma^2_{\hat{p}} = \text{var}\left(\frac{n_H}{n}\right) = \frac{1}{n^2}\text{var}(n_h) \tag{6.10}$$

where var(n_H) is the variance of n_H (which has a binomial distribution) and is

$$\text{var}(n_h) = np(1-p). \tag{6.11}$$

Hence, the variance of $\hat{p}$ is

$$\sigma^2_{\hat{p}} = \frac{1}{n}p(1-p). \tag{6.12}$$

Substituting p from (6.6),

$$\sigma^2_{\hat{p}} = \frac{1}{n}\frac{I}{[c(b-a)]^2}[c(b-a)-I]. \tag{6.13}$$

Now, finding the variance of the right hand side in equation (6.8),

$$\begin{aligned}
\sigma^2_{\theta} &= \text{var}\left(c(b-a)\frac{n_H}{n}\right) = [c(b-a)]^2\sigma^2_{\hat{p}} && (6.14)\\
&= [c(b-a)]^2\frac{1}{n}p(1-p) && (6.15)\\
&= \frac{I}{n}[c(b-a)-I]. && (6.16)
\end{aligned}$$

Now, the requirement of a certain degree of confidence α that the estimate θ of I lies within a certain bound can be stated

$$P\left(|\theta - I| < \varepsilon\right) \geq \alpha. \tag{6.17}$$

This can be expressed in terms of *Chebyshev's inequality* which says that the probability of the absolute difference between the estimated value and the true value being less than ε is at least

[1] A Bernoulli trial is a trial of an experiment with only two possible outcomes.

α with α expressed in terms of the variance of the estimate and ε [2]

$$P\left(|\theta - I| < \varepsilon\right) \geq 1 - \frac{\sigma_\theta^2}{\varepsilon^2} \tag{6.18}$$

so that, from equations (6.17) and (6.18)

$$\alpha \leq 1 - \frac{\sigma_\theta^2}{\varepsilon^2}. \tag{6.19}$$

Substituting equation (6.15) in to (6.19),

$$\alpha \leq 1 - \frac{p(1-p)[c(b-a)]^2}{n\varepsilon^2} \tag{6.20}$$

is obtained which can be solved for n:

$$n \geq \frac{p(1-p)[c(b-a)]^2}{(1-\alpha)\varepsilon^2}. \tag{6.21}$$

This gives the required number of trials.

When n is large enough, the central limit theorem can be applied which says the random variable $\hat{\theta}$ where

$$\hat{\theta} = \frac{\theta - I}{\sigma_\theta} \tag{6.22}$$

is distributed approximately according to the standard normal distribution i.e.

$$P(\theta \leq z) \approx g(z) \tag{6.23}$$

where $g(z)$ is the standard normal distribution.

It can be shown that, for a confidence level of $1 - \alpha$, the confidence interval for I is

$$\theta \pm z_{\alpha/2} \frac{c(b-a)\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \tag{6.24}$$

where

$$z_{\alpha/2} = g^{-1}(\alpha). \tag{6.25}$$

[2] Another way of stating it is to say that the probability of the absolute difference between the estimate and the true value being greater than ε is less than $1 - \alpha$ or

$$P\left(|x - \overline{X}| \geq \varepsilon\right) \leq \frac{\sigma_x^2}{\varepsilon^2}$$

where

$$\frac{\sigma_x^2}{\varepsilon^2} = 1 - \alpha$$

which is analogous to equation (5.37)

6.1.3 Summary of the "Hit or Miss" algorithm

1. Generate a sequence of $2n$ random numbers.
2. Arrange the random numbers into n pairs $(U_1', V_1'), (U_2', V_2'), \ldots, (U_n', V_n')$ in any manner such that each random number from the original sequence is used exactly once.
3. Calculate
$$\left.\begin{array}{l} X_i' = a + U_i'(b-a) \\ g(X_i') \end{array}\right\} i = 1, 2, \ldots, n \tag{6.26}$$
4. Count the number of cases n_H for which
$$g(X_i') > cV_i'. \tag{6.27}$$
5. Estimate the integral I by
$$\theta = c(b-a)\frac{n_H}{n}. \tag{6.28}$$

6.2 The "Sample-Mean" method

6.2.1 Computing an integral by the "Sample-Mean" method

An integral I such that
$$I = \int_a^b h(x)dx \tag{6.29}$$
may be represented as the expected value of some random variable. If the integral is re-written as
$$I = \int_a^b \frac{h(x)}{f_X(x)} f_X(x)dx \tag{6.30}$$
where $f_X(x)$ is any p.d.f. such that $f_X(x) > 0$ when $h(x) \neq 0$, then making use of equation (5.5),
$$I = E\left[\frac{h(X)}{f_X(X)}\right] \tag{6.31}$$
where the random variable X is distributed according to $f_X(x)$.

In a similar manner to that adopted in section 6.1, $f_X(x)$ may be defined as
$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases} \tag{6.32}$$

Thus

$$E[h(X)] = IE[f_X(X)] \tag{6.33}$$
$$= \frac{I}{b-a} \tag{6.34}$$

and

$$I = (b-a)E[h(X)]. \tag{6.35}$$

An unbiased estimator of I is its sample mean

$$\theta = (b-a)\frac{1}{n}\sum_{i=1}^{n} h(X_i'). \tag{6.36}$$

6.2.2 Precision of the "Sample-Mean" method

From equation (5.12), the variance of θ is $E(\theta^2) - [E(\theta)]^2$ so that

$$\sigma_\theta^2 = \text{var}\left[\frac{1}{n}(b-a)\sum_{i=1}^{n} h(X_i')\right] \tag{6.37}$$
$$= \frac{1}{n}\left[(b-a)^2\int_a^b \frac{h^2(x)}{b-a}dx - I^2\right] \tag{6.38}$$
$$= \frac{1}{n}\left[(b-a)\int_a^b h^2(x)dx - I^2\right]. \tag{6.39}$$

6.2.3 Summary of the "Sample-Mean" algorithm

1. Generate a sequence of n random numbers $\{U_i'\}_{i=1}^n$.
2. Compute $X_i' = a + U_i'(b-a), i = 1, 2, \ldots, n$.
3. Compute $h(X_i'), i = 1, 2, \ldots, n$.
4. Compute the sample mean θ according to equation (6.36) which estimates I.

Chapter 7

Variance Reduction

It can be seen from equation (5.43) that to obtain a given accuracy in the result of a sample estimate for a given degree of confidence may require many trials. In Monte Carlo simulation each trial may require significant computational effort. The greatest attention in use of Monte Carlo simulation has therefore been spent in reducing the number of samples needed for a given confidence interval and degree of confidence by means of *variance reduction*.

Hammersley and Handscomb [2] say: "If, at any point of a Monte Carlo calculation, we can replace an estimate by an exact value, we shall reduce the sampling error in the final result." This can be understood by considering the variance of an estimated parameter. If the variance is zero, the parameter is known perfectly and there would be no need for Monte Carlo simulation.

Variance reduction is a way of making better use of existing knowledge about a problem [50]. Exactly what form that existing knowledge takes determines what kind of variance reduction technique can be used. The more existing knowledge there is, the more effective variance reduction will be.

There are 5 main variance reduction methods:

- antithetic variates.
- correlated sampling.
- control variates.
- importance sampling.
- stratified sampling.

Of these, all except correlated sampling have found at least some application in power systems, examples of which are given in section 9.

Other variance reduction techniques include dagger sampling and selective sampling. These and those listed above are next described in the next sections. Other means of improving the efficiency of a Monte Carlo simulation are described in [50] and [3].

It may also be noted that the variance of an estimate can be reduced if some analytical integration can be performed with respect to *some* of the variables.

7.1 Antithetic variates

This technique is based on finding two unbiased estimators for the unknown parameter (e.g. the integral I) which have strong negative correlation. If the two estimators are ϕ_1 and ϕ_2, the estimator of I will be $\frac{1}{2}(\phi_1 + \phi_2)$ with variance

$$\text{var}\left[\frac{1}{2}(\phi_1 + \phi_2)\right] = \frac{1}{4}\text{var}(\phi_1) + \frac{1}{4}\text{var}(\phi_2) + \frac{1}{2}\text{cov}(\phi_1, \phi_2). \tag{7.1}$$

Clearly, from equation (7.1), if $\text{cov}(\phi_1, \phi_2)$ is strongly negative, the overall variance can be reduced.

By way of example, consider the integral

$$I = \int_0^1 h(x)dx \tag{7.2}$$

which is equal to

$$I = \frac{1}{2}\int_0^1 \left[h(x) + h(1-x)\right] dx. \tag{7.3}$$

The estimator of I is then

$$\phi = \frac{1}{2}(\phi_1 + \phi_2) = \frac{1}{2}\left[h(U) + h(1-U)\right] \tag{7.4}$$

where U is a uniformly distributed random number sequence between $[0, 1]$. ϕ is an unbiased estimator of I because both $\phi_1 = h(U)$ and $\phi_2 = h(1-U)$ are unbiased estimators of I.

To estimate I, a sample size of n is taken from the uniform distribution and

$$\theta = \frac{1}{2n}\sum_{i=1}^{n}\left[h(U_i') + h(1-U_i')\right] \tag{7.5}$$

is found.

When sampling, the effect is to have pairs of trials which are negatively correlated, the first one determined by a random number or vector of random numbers U, the second by 1 minus the same random number(s), $1 - U$, so that if one trial returns a large outcome, the other will be likely to return a small one. It would be hoped that the average of the two will be near to the population average. Thus, if many such pairs of trials are performed, the spread (or variance) of the averages of the pairs will be low, i.e. the variance of the sample will have been reduced.

When carrying out a Monte Carlo estimation using antithetic variates and tracking the sample variance to detect convergence of the estimate to some pre-defined limits, it should be noted that if the covariance of ϕ_1 and ϕ_2 is not to be computed (using equation (5.28)) in order to

evaluate the right hand side of (7.1), the average of pairs of trials (one of ϕ_1 and the other of ϕ_2) should be tallied in order to evaluate the left hand side of (7.1). From this, it should be clear that the computation time required by equation (7.5) is twice that required by the standard sample-mean method (since two trials are required to provide one tallied 'data point') so the estimator of (7.5) is more efficient only when its variance is less than half that of the straightforward sample-mean approach. A proof is included in [50] which shows that this can be guaranteed if $h(x)$ is a monotonic function.

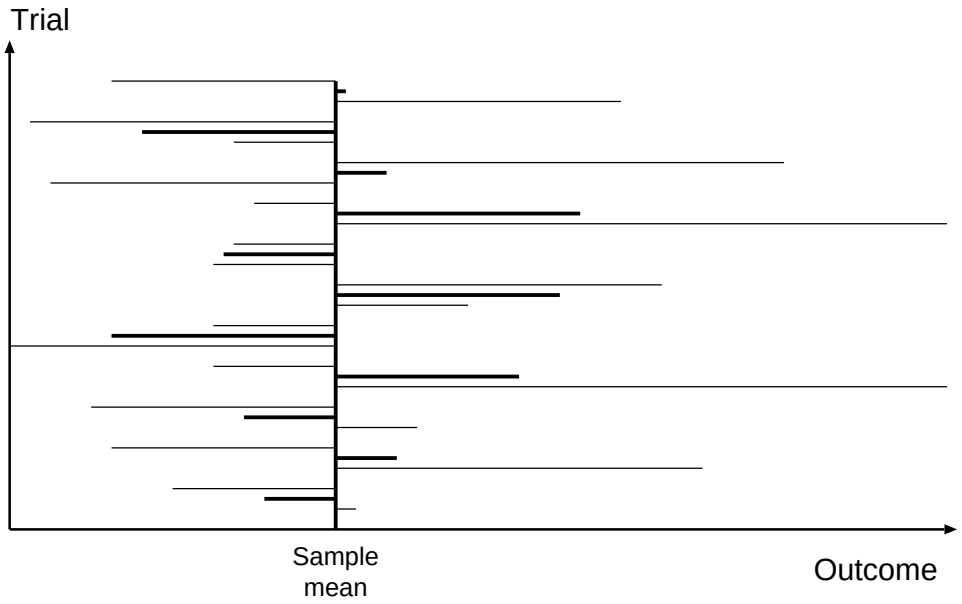

Figure 7.1: Illustration of use of antithetic variates

The benefit of antithetic variates is illustrated in figure 7.1. Individual trials' costs' differences from the mean are shown by light horizontal lines. The variance of simple sampling will be the sum of the square of these differences divided by the number of trials. In antithetic variates, however, the individual trials' outcomes are not logged as data points, but the averages of pairs of trials are. If there is significant negative correlation between pairs of trials, each pair will generally a comprise a lower outcome and a higher outcome the average of which will be closer to the sample mean. The differences between the pair averages and the sample average can be seen to be smaller meaning there is less spread of outcomes, i.e. there is smaller variance.

A further point worth noting is that when antithetic variates are applied to simulations of systems comprising two-state stochastic variables (such as power systems—see sections 7.6 and 8.3 for brief descriptions), the correlation of ϕ_1 and ϕ_2 and hence the efficiency of the method is likely to fall off as the probability of the 'out of service' state gets smaller. This is because if U_i' is such that the 'basic event' is 'in service', U_i' needs to be very large for the state to change when it is a function of $1 - U_i'$. Since this happens rarely, the negative correlation is reduced. Take a failure probability of 0.15 as an example. If $U_i' > 0.15$, the state is "in service" but U_i' has to be ≥ 0.85 for $1 - U_i'$ to give the "out of service" condition for the 'antithetic' state.

7.2 Correlated sampling

The aim of correlated sampling is to produce a high positive correlation between two similar processes so that the variance of the difference is considerably smaller than it would be if the two processes were statistically independent [50].

This may be used, for example, in assessing the effect of a small change in a system. A Monte Carlo run could be carried out for the system with the change and then for the system without it. The results of the two runs could then be subtracted. However, that difference may often be small in comparison with each individual result or with the variances of those results. The variance of the difference, meanwhile, would be the sum of the variances of each run.

A way, then, of reducing the variance of the difference would be to use the same random numbers for each run so that the individual results are highly positively correlated.

Take, by way of example, the difference ΔI of the integrals I_1 and I_2 of two functions, $\phi_1(x)$ and $\phi_2(x)$ i.e.

$$\begin{aligned} \Delta I &= I_1 - I_2 && (7.6)\\ &= \int \phi_1(x)dx - \int \phi_2(x)dx. && (7.7) \end{aligned}$$

If $h_1(x)$ and $h_2(x)$ are defined such that

$$\begin{aligned} h_1(x) &= \frac{\phi_1(x)}{f_1(x)} && (7.8)\\ h_2(x) &= \frac{\phi_2(x)}{f_2(x)} && (7.9) \end{aligned}$$

where $f_1(x)$ and $f_2(x)$ are p.d.f.s, then

$$\begin{aligned} I_1 &= \int h_1(x)f_1(x)dx && (7.10)\\ I_2 &= \int h_2(x)f_2(x)dx. && (7.11) \end{aligned}$$

If $X'_1, \ldots, X'_n$ are sampled from $f_1(x)$ and $Y'_1, \ldots, Y'_n$ from $f_2(x)$, ΔI can be estimated using

$$\begin{aligned} \Delta\theta &= \frac{1}{n}\sum_{i=1}^{n} h_1(X'_i) - \frac{1}{n}\sum_{i=1}^{n} h_2(Y'_i) && (7.12)\\ &= \frac{1}{n}\sum_{i=1}^{n} d_i && (7.13) \end{aligned}$$

where

$$d_i = h_1(X'_i) - h_2(Y'_i). \qquad (7.14)$$

The variance of $\Delta\theta$ is

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\text{cov}(\hat{\theta_1}, \hat{\theta_2}) \qquad (7.15)$$

where

$$\hat{\theta_1} = \frac{1}{n}\sum_{i=1}^{n} h_1(X_i') \tag{7.16}$$

$$\hat{\theta_2} = \frac{1}{n}\sum_{i=1}^{n} h_2(Y_i') \tag{7.17}$$

$$\sigma_1^2 = E(\hat{\theta_1} - I_1)^2 \tag{7.18}$$

$$\sigma_2^2 = E(\hat{\theta_2} - I_2)^2 \tag{7.19}$$

and

$$\text{cov}(\hat{\theta_1}, \hat{\theta_2}) = E\left[(\hat{\theta_1} - I_1)(\hat{\theta_2} - I_2)\right]. \tag{7.20}$$

If $\hat{\theta_1}$ and $\hat{\theta_2}$ are statistically independent, then

$$\text{cov}(\hat{\theta_1}, \hat{\theta_2}) = 0 \tag{7.21}$$

and

$$\sigma^2 = \sigma_1^2 + \sigma_2^2. \tag{7.22}$$

However, if the random variables X and Y are positively correlated and $h_1(x)$ has a similar shape to $h_2(x)$ then the variance of $\Delta\theta$ will be greatly reduced. Thus, the key to exploiting this form of variance reduction is in ensuring positive correlation between the estimates of $\hat{I_1}$ and $\hat{I_2}$. If $f_1(x)$ and $f_2(x)$ are similar, this can be achieved by taking a sequence of random numbers $U_1', \ldots, U_n'$ and finding $X_1', \ldots, X_n'$ and $Y_1', \ldots, Y_n'$ by $X_i' = F_1^{-1}(U_i')$ and $Y_i' = F_2^{-1}(U_i')$ where F^{-1} is the inverse of the cumulative distribution function of f used in the generation of random variates [50, 51].

According to [50], there is no general procedure that can be implemented in correlated sampling. However, it can be usefully employed when

- the effect of a small change in a system is to be calculated.
- the difference in a parameter in two or more similar cases is of more interest than its absolute value.

This form of variance reduction is also described in [3] where it is known as *common random numbers.*

7.3 Control variates

Also known as *regression sampling* [3], this technique replaces the direct estimation of a parameter with an estimate of the difference between the problem of interest and some analytical

model [50]. This is possible, for example, when the problem contains a part which can be solved by means of an analytical model so that the Monte Carlo simulation is used to calculate the difference between the solution of the problem and this part [2].

A random variate C is a control variate for Y if it is correlated with Y and if its expectation μ_C is known. C can then be used to estimate μ with a smaller variance than the estimator Y.

For any value of a,

$$Y(a) = Y - a(C - \mu_C) \tag{7.23}$$

is an unbiased estimator of μ [50].

The variance of $Y(a)$ is given by

$$\text{var}[Y(a)] = \text{var}[Y] - 2a\text{cov}[Y, C] + a^2\text{var}[C]. \tag{7.24}$$

Variance reduction is achieved if

$$2a\text{cov}[Y, C] > a^2\text{var}[C]. \tag{7.25}$$

The value of a that minimizes var$[Y(a)]$ is

$$a^* = \frac{\text{cov}[Y, C]}{\text{var}[C]} \tag{7.26}$$

and the minimum variance is equal to

$$\text{var}[Y(a^*)] = (1 - \rho_{YC}^2)\text{var}[Y] \tag{7.27}$$

where ρ_{YC} is the correlation coefficient between Y and C.

7.4 Importance sampling

To illustrate the concept, the problem of estimating I is considered where

$$I = \int h(x)dx, \quad x \in D \tag{7.28}$$

and it is supposed that $\int h^2(x)dx$ and I both exist [50, 52].

The basic idea of importance sampling is to concentrate the distribution of sample points in the parts of region D where $h(x)$ is most significant instead of spreading them evenly.

As was done in equations (6.30) and (6.31), the integral can be represented by

$$I = \int \frac{h(x)}{f_X(x)} f_X(x)dx = E\left[\frac{h(X)}{f_X(X)}\right] \tag{7.29}$$

where X is any random variable with p.d.f. $f_X(x)$ such that $f_X(x) > 0$ for each $x \in D$. The function $f_X(x)$ is called *the importance sampling distribution.* From equation (7.29), it is possible to say that

$$\phi = h(x)/f_x(x) \tag{7.30}$$

is an unbiased estimator of I. The variance of ϕ is

$$\sigma_\phi^2 = E(\phi^2) - [E(\phi)]^2 . \tag{7.31}$$

Since $E(\phi) = I$, this is

$$\sigma_\phi^2 = \int \phi^2 f_X(x)dx - I^2 \tag{7.32}$$

$$= \int \left(\frac{h(x)}{f_X(x)}\right)^2 f_X(x)dx - I^2. \tag{7.33}$$

In order to estimate the integral, a sample $\{X_1', X_2', \ldots, X_n'\}$ is taken from $f_X(x)$ and the sample-mean formula used whereby

$$\theta = \frac{1}{n}\sum_{i=1}^{n} \frac{h(X_i')}{f_X(X_i')}. \tag{7.34}$$

The task is to choose $f_X(x)$ to minimize the variance of ϕ which is the same as minimizing the variance of θ.

In [53], it is proved that σ_ϕ^2 is a minimum when

$$f_X(x) = \frac{|h(x)|}{\int |h(x)|dx} \tag{7.35}$$

and

$$\sigma_\phi^2 = \left(\int |h(x)|dx\right)^2 - I^2. \tag{7.36}$$

(A proof is also offered in [50]). It is further proved that if $h(x) > 0$, the optimal p.d.f. $f_X(x)$ is

$$f_X(x) = \frac{h(x)}{I} \tag{7.37}$$

and $\sigma_\phi^2 = 0$.

One serious problem with this should immediately be apparent: to choose an optimal $f_X(x)$ to reduce the variance of the estimate of I, I or $\int |h(x)|dx$ (which is practically equivalent to computing I) should be known. But if they are known, there is no need for the Monte Carlo estimate!

Although this may seem to be a terminal case, it has been shown that the variance can still be significantly reduced if $f_X(x)$ is chosen to have a shape similar to $\int |h(x)|dx$. When choosing

$f_X(x)$, care should be taken to ensure that it is relatively easy to sample from, especially if $|h(x)|$ is not well-behaved. Another difficulty may be that an inappropriate choice of the importance sampling distribution may theoretically *increase* the variance of the estimate of I.

Since the shape of $f_X(x)$ should follow $|h(x)|$, some computational time may be saved in finding the integral if random numbers are sampled from the subregion $D' = \{x : h(x) \neq 0\}$ of D i.e.

$$f_X(x) > 0, \quad \text{if } h(x) \neq 0 \tag{7.38}$$

$$f_X(x) = 0, \quad \text{if } h(x) = 0. \tag{7.39}$$

7.5 Stratified sampling

The main idea of stratified sampling is similar to that of importance sampling. The region of interest D is subdivided into m disjointed regions (i.e. they are adjacent but do not overlap) $D = \cup_{i=1}^{m} D_i, D_k \cap D_j = 0, k \neq j$. The integral (if the integral is to be found) of each region can then be found separately with more trials conducted in the regions of more interest.

To demonstrate the method, let the integral to be found be

$$I = \int_D h(x) f_X(x) dx. \tag{7.40}$$

Now sub-divide the region D into m sub-regions $D_1, D_2, \ldots, D_m$ so that

$$I_i = \int_{D_i} h(x) f_X(x) dx. \tag{7.41}$$

If

$$P_i = \int_{D_i} f_X(x) dx, \tag{7.42}$$

it should be clear that

$$\sum_{i=1}^{m} P_i = 1 \tag{7.43}$$

and

$$I = \int_D h(x) f_X(x) dx = \sum_{i=1}^{m} \int_{D_i} h(x) f_X(x) dx \tag{7.44}$$

$$= \sum_{i=1}^{m} I_i. \tag{7.45}$$

Now $h_i(x)$ is defined such that

$$h_i(x) = \begin{cases} h(x), & \text{if } x \in D_i \\ 0, & \text{otherwise.} \end{cases} \tag{7.46}$$

The integral I_i can be re-written as

$$I_i = \int_{D_i} P_i h(x) \frac{f_X(x)}{P_i} dx = P_i \int_{D_i} h_i(x) \frac{f_X(x)}{P_i} dx \tag{7.47}$$

$$= P_i E\left[h_i(X)\right] \tag{7.48}$$

where

$$\int_{D_i} f_X(x) dx = P_i. \tag{7.49}$$

A sample-mean estimator for I_i can be written

$$Y_i = P_i h(X_i) \tag{7.50}$$

where the random variable X_i is distributed according to $f_X(x)/P_i$ on D_i.

The integral I_i can be estimated by

$$\phi_i = \frac{P_i}{n_i} \sum_{k_i=1} n_i h(X_{k_i}), \quad k_i = 1, \ldots, n_i, i = 1, \ldots, m \tag{7.51}$$

and the final integral I by

$$\theta = \sum_{i=1}^{m} \phi_i \tag{7.52}$$

$$= \sum_{i=1}^{m} \frac{P_i}{n_i} \sum_{k_i=1}^{n_i} h(X_{k_i}). \tag{7.53}$$

The variance of θ is

$$\sigma_\theta^2 = \sum_{i=1}^{m} \frac{P_i^2}{n_i} \text{var}\left(h(X_i)\right) \tag{7.54}$$

$$= \sum_{i=1}^{m} \frac{P_i^2 \sigma_i^2}{n_i} \tag{7.55}$$

where

$$\sigma_i^2 = \text{var}\left(h(X_i)\right) = \frac{1}{P_i} \int_{D_i} h^2(x) f_X(x) dx - \frac{I_i^2}{P_i^2}. \tag{7.56}$$

Note the use of the population variance for each stratum. In many studies, this is not available and is approximated by the sample variance s_i^2.

If the stratification is carried out well, the variance of the estimate should be less than the variance of an estimate found using the basic sample-mean method.

7.5.1 Allocation of trials

It has been proved [50] that if the sample size n_i in each subregion D_i is proportional to P_i (effectively the size of the subregion) i.e. $n_i = nP_i$, the variance of the stratified sampling method is guaranteed to be less than or equal to the variance of the sample-mean method. (This rule for apportioning n_i is known as *proportionally stratified sampling* [3]).

The question remains: how many samples should be assigned to each subregion once the subregions have been chosen? If n_i is the number of samples assigned to subregion D_i such that

$$\sum_{i=1}^{m} n_i = n \tag{7.57}$$

where n is the total number of samples, the minimum of the variance of θ, i.e.

$$\min\left(\sum_{i=1}^{m} \frac{P_i^2}{n_i}\sigma_i^2\right) \tag{7.58}$$

occurs when

$$n_i = n\frac{P_i\sigma_i}{\sum_{j=1}^{m} P_j\sigma_j} \tag{7.59}$$

and is equal to

$$\frac{1}{n}\left[\sum_{i=1}^{m} P_i\sigma_i\right]^2. \tag{7.60}$$

Thus minimum variance of θ occurs when n_i are proportional to $P_i\sigma_i$. Of course, this is not of much use since the σ_i are not usually known in advance. However, it may be possible to conduct some small "pilot" run to get rough estimates for σ_i.

The particular case when $P_i = 1/m$ and $n_i = N/m$ is the so-called *systematic sampling* method and is described in detail in [54] and outlined in [50].

7.5.2 Stratification after sampling

Kleijnen describes a procedure for *stratification after sampling* [3]. In this, the number of samples in each stratum is not fixed beforehand but depends on the sample outcome. Let the estimate obtained by this means be ψ such that

$$\psi = \sum_{i=1}^{m} P_i z_i \tag{7.61}$$

where

$$z_i = \sum_{k=1}^{n_i} \frac{h(X_{k_i})}{n_i}. \tag{7.62}$$

If the estimate obtained by this method is the same as that by the simple sample-mean approach, the estimates can be equated:

$$\theta = \frac{1}{n}\sum_{i=1}^{n} h(X_i) = \frac{1}{n}\sum_{i=1}^{m}\sum_{k=1}^{n_i} h(X_{k_i}) \tag{7.63}$$

$$= \sum_{i=1}^{m}\frac{n_i}{n}\sum_{k=1}^{n_i}\frac{h(X_{k_i})}{n_i} \tag{7.64}$$

$$= \sum_{i=1}^{m}\frac{n_i}{n} z_i \tag{7.65}$$

$$= \frac{1}{n}\sum_{i=1}^{m}\sum_{k=1}^{n_i} h(X_{k_i}). \tag{7.66}$$

It can be seen that the "empirical" weights n_i/n have replaced the "theoretical" weights P_i. Since the number of samples per stratum depends on the sample results, it is known as a *stochastic variable.* n_i can then be further understood as being an estimator of $P_i n$ i.e.

$$E(n_i) = P_i n. \tag{7.67}$$

The variance of the estimate can be estimated by

$$s_\psi^2 = \sum_{i=1}^{m} P_i^2 s_{z_i}^2 \tag{7.68}$$

$$= \sum_{i=1}^{m}\left(\frac{n_i}{n}\right)^2 s_{z_i}^2 \tag{7.69}$$

where

$$s_{z_i}^2 = \frac{\sum_{k=1}^{n_i}\left[h(X_{k_i}) - z_i\right]^2}{(n_i - 1)}\cdot\frac{1}{n_i} \tag{7.70}$$

$$= \frac{1}{n_i} E\left[\text{var}(h(X_i))\right] \tag{7.71}$$

Hence

$$s_\psi^2 = \frac{1}{n^2}\sum_{i=1}^{m}\frac{n_i}{n_i - 1}\sum_{k=1}^{n_i}\left[h(X_{k_i}) - z_i\right]^2 \tag{7.72}$$

or, using the computing formula for $s_{z_i}^2$,

$$s_\psi^2 = \frac{1}{n^2}\sum_{i=1}^{m}\frac{n_i}{n_i - 1}\left\{\sum_{k=1}^{n_i}\left[h(X_{k_i})\right]^2 - \frac{\left[\sum_{k=1}^{n_i} h(X_{k_i})\right]^2}{n_i}\right\}. \tag{7.73}$$

It is demonstrated in [55] that stratification after sampling is nearly as precise as proportionally stratified sampling so long as the number of samples per subregion is reasonably large, say > 20. The method then becomes attractive if it is difficult or impossible to fix the number of trials per subregion in advance.

7.5.3 Choice of stratum boundaries

It may be noted that the preceeding discussion has assumed that the stratum boundaries have already been allocated. In practice, choice of stratum boundary depends, firstly, on the chosen *stratification variable*. In order that trials can be concentrated in the region of most interest, the stratification variable should be strongly correlated with the variable being estimated [55], ideally the variable itself though this is not always possible.

Cochran [55] discusses a rule concerning the optimum allocation of stratum boundaries which minimizes the variance of the estimate. The optimum allocation problem is shown in [55] to reduce to the minimization of $\sum P_i\sigma_i$. As would be expected from comparison of stratified sampling with importance sampling, the solution depends on knowledge of the probability density function of the stratification variable. In the following illustration, the stratification variable is assumed to be the variable being estimated.

Let

$$Z(x) = \int_{X_0}^{X} \sqrt{f_X(x)}dx. \tag{7.74}$$

If the strata are numerous and narrow, $f_X(x)$ should be approximately constant, i.e. rectangular in shape, within each stratum. Hence,

$$\begin{aligned} P_i &= \int_{X_{i-1}}^{X_i} f_X(x)dx \approx f_i(X_i - X_{i-1}) && (7.75)\\ \sigma_i &\approx \frac{1}{\sqrt{12}}(X_i - X_{i-1}) && (7.76)\\ Z_i - Z_{i-1} &= \int_{X_{i-1}}^{X_i} \sqrt{f_X(x)}dx && (7.77)\\ &\approx \sqrt{f_{X_i}}(X_i - X_{i-1}) && (7.78) \end{aligned}$$

where f_{X_i} is the "constant" value of f_X in stratum i. Substituting these approximations, it is found that

$$\begin{aligned} \sqrt{12}\sum_{i=1}^{m} P_i\sigma_i &\approx \sum_{i=1}^{m} f_{X_i}(X_i - X_{i-1})^2 && (7.79)\\ &\approx \sum_{i=1}^{m}(Z_i - Z_{i-1})^2. && (7.80) \end{aligned}$$

Since $(Z_m - Z_0)$ is fixed, it can be verified that the right hand side of equation (7.80) can be minimized by making $(Z_i - Z_{i-1})$ constant.

Thus, given $f_X(x)$, the rule is to form the cumulative of $\sqrt{f_X(x)}$ and choose X_i so that equal intervals are created on the cum $\sqrt{f_X(x)}$ scale. This is illustrated for 5 strata in table 7.2 for the data in table 7.1 where $f_X(x)$ is defined in terms of the frequency of outcomes in 10 classes (bands) of x. In table 7.2, it can be seen that, due to the need to fix stratum boundaries on class boundaries, the interval for each stratum on the cum $\sqrt{f_X(x)}$ scale is not constant. The stratification is done to make this as constant as possible.

x	$f_X(x)$	cum $\sqrt{f_X x}$
0-5	3464	58.9
5-10	2516	109.1
10-15	2157	155.5
15-20	1581	195.3
20-25	1142	229.1
25-30	746	256.4
30-35	512	279.0
35-40	376	298.4
40-45	265	314.7
45-50	207	329.1

Table 7.1: Frequency distribution data for stratum allocation

Stratum	1	2	3	4	5
Boundaries	0-5	5-10	10-20	20-30	30-50
Interval on cum $\sqrt{f_X(x)}$	58.9	50.2	86.2	61.1	72.7

Table 7.2: Stratum allocation for data in table 7.1

Evidence is quoted in [55] to suggest that little reduction in estimate variance can be expected beyond $m = 6$ strata.

7.5.4 Difficulties in stratified sampling

For many studies, fixing of the number of trials in each stratum in advance is impossible. Stratification after sampling then becomes attractive, but this, too has a number of difficulties. They are:

Errors in stratum weights. If the theoretical weights cannot be known, they must be estimated. In stratification after sampling, an estimate of $P_i = n_i/n$ is readily available, though Cochran believes this may lead to bias in the estimate of the mean [55] in stratified sampling in general, though the discussion of stratification after sampling in section 7.5.1 above shows that this is not the case here.

The stochastic nature of n_i in post-stratification. This has the effect of making the confidence limits more approximate than for cases where the theoretical weights are known and the n_i are fixed [3].

The approximate nature of confidence limits due to use of sample variance instead of population variance. This may be partially resolved by modelling the sampling distribution for each stratum by a Student t distribution with an underestimate of the degrees of freedom. The distribution of the variance of an estimate obtained by stratified sampling is too complex to allow an analytical modelling, but Cochran [55] quotes an approximate formula due to Satterthwaite for a corrected number of degrees of freedom. It requires

knowledge of the population size of each stratum, but its lower bound is equal to the smallest of the $n_i - 1$.

In addition, it may be intuitively understood that the confidence limits are more approximate with more strata since there are more variables subject to error involved. Also, with more strata, there will be fewer trials in each stratum implying greater error in the stratum sample variance.

These difficulties may be resolved by

- "robust estimation" or "resampling" to deal with trials which aren't normally distributed.
- tracking the variance of stratum variances. Force more trials until each sample stratum variance comes within some (arbitrary) confidence limits.

Two robust resampling techniques, the "jackknife" and the "bootstrap", are described in section 7.8.

7.6 Dagger sampling

This method is due to Kumamoto *et al* [56] and is described in [52] as being particularly well-suited to two-state variables and small probability events. As such, the method can be understood as an extension of antithetic variates overcoming the degradation of the negative correlation of the two estimators under such circumstances. Two-state variables and rare events will be considered in describing the procedure.

A uniformly distributed random number U' between $[0, 1]$ can be generated to test whether a system component with failure probability p is in the failure state or normal state. If $U' \leq p$, the component is in the failure state. If $U' > p$, it is in the normal state. Each random number therefore corresponds to one trial of the component state. This approach is known as *direct Monte Carlo sampling*.

Dagger sampling proceeds in the following way. If z is the largest integer not larger than $1/p$, the interval $[0, 1]$ may be divided into z subintervals each of length p. There may also be a remainder subinterval, for example in the case where $p = 0.15$, $1/p = 6.66$ and $z = 6$. There are then 6 subintervals of length 0.15 plus a remainder part of length 0.1.

In dagger sampling, each subinterval formed in the above way corresponds to a trial of a component. If a generated random number falls into the ith subinterval, the component failure is assumed to occur in trial i and not in the other trials. If the random number falls in the remainder part, the component is assumed not to fail in z trials.

In this procedure, only one random number determines the outcome of z trials of the component state (it is as if one random number pierces in z subintervals, hence the name "dagger"). If the generated random number falls in the ith sub-interval, then the "dagger number", i.e. the

number of the trial in which a failure takes place, will be i. Since z trials are determined by one random number after which another random number must be generated to determine the component's status in the next z trials, z may be known as the "dagger cycle length".

Combination of all component states yields the state vector $\mathbf{\Psi}$. Let $\mathbf{\Psi}_1, \mathbf{\Psi}_2, \ldots, \mathbf{\Psi}_n$ be n system vectors obtained by dagger sampling. The system unavailability can be estimated by

$$\theta = \frac{1}{n} \sum_{i=1}^{n} I(\mathbf{\Psi}_i) \tag{7.81}$$

where $I(\mathbf{\Psi})$ is an indicator function which equals 0 if the state $\mathbf{\Psi}$ is a non-failure state and 1 if it is a failure state.

The variance of θ, s_θ^2, can be calculated by [1]

$$s_\theta^2 = \frac{1}{n^2} \left\{ \text{var}[I(\mathbf{\Psi})] + \sum_{i,j,i \neq j}^{n} \text{cov}[I(\mathbf{\Psi}_i), I(\mathbf{\Psi}_j)] \right\}. \tag{7.82}$$

With direct Monte Carlo sampling, there is no correlation between $I(\mathbf{\Psi}_i)$ and $I(\mathbf{\Psi}_j), i \neq j$, while in dagger sampling, different random vectors within each group of z trials are correlated: if a failure occurs on one component in a trial, it definitely does not occur in another $z-1$ trials. If Ψ_{ik} and Ψ_{jk} are single elements of the state vectors $\mathbf{\Psi}_i$ and $\mathbf{\Psi}_j$ associated with component k, and the trials i and j are generated by a common random number for component k, Ψ_{ik} and Ψ_{jk} have negative covariance as one of Ψ_{ik} and Ψ_{jk} will always be zero, i.e.

$$\text{cov}(\Psi_{ik}, \Psi_{jk}) = E(\Psi_{ik}, \Psi_{jk}) - E(\Psi_{ik})E(\Psi_{jk}) = -p_k^2 < 0. \tag{7.83}$$

Thus the correlation between two system state random vectors is negative if they have some elements corresponding to a common random number. Since $I(\mathbf{\Psi})$ is an indicator variable, the negative correlation between $\mathbf{\Psi}_i$ and $\mathbf{\Psi}_j$ also applies to $I(\mathbf{\Psi}_i)$ and $I(\mathbf{\Psi}_j)$. Dagger sampling therefore has a smaller variance than direct sampling.

A concern may be that some bias will be introduced due to presence of a remainder subinterval in $[0, 1]$ implying that individual components' number of failures will not match the expected number. It will now be shown by means of an illustration that this is not the case.

Suppose that a component has a probability of failure of $p = 0.3$. $1/p = 3.333$ so the greatest integer not greater than $1/p$ is $z = 3$. Thus, there will be 3 intervals each of length 0.3 in the $[0, 1]$ interval with a remainder interval of length 0.1. This is illustrated in figure 7.2.

One random number can now be used to determine z trials which, in this case, is 3. If a random number U' falls in the first subinterval, a failure of the component will take place in the first of the next $z = 3$ trials. If U' falls in the second, the failure will be in the second trial, and if in the third subinterval the failure will be in the third trial. However, if U' falls in the remainder subinterval, no failure will take place in the set of z trials.

[1] It should be noted that the author believes the version of equation (7.82) that appears in [56] is erroneous. The expression that appears here is therefore different.

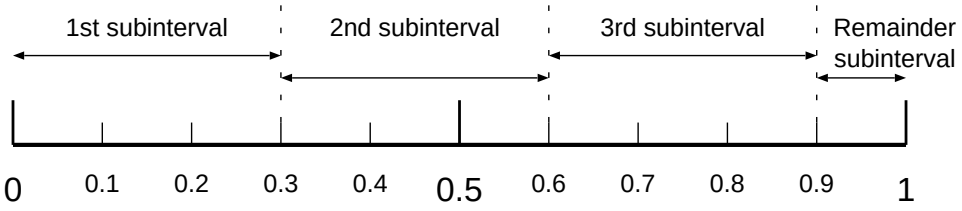

Figure 7.2: Partitioning of the $[0, 1]$ interval for $p = 0.3$ in dagger sampling

It can be seen from figure 7.2 that the probability of having one failure in three trials is $1 - P(U'$ falls in the remainder subinterval$)$ i.e. $1 - 0.1 = 0.9$. It is readily apparent that if the probability of having one failure in 3 trials is 0.9, the probability of having one failure in 1 trial is $0.9/3 = 0.3$ which is the original probability of failure of the component. Thus, the behaviour of the component has been correctly represented.

If variances are being monitored in the course of a Monte Carlo simulation to decide when to stop sampling, a potential disadvantage of dagger sampling is the need to calculate the covariances in equation (7.82). The number of covariances to be computed will be

$$\begin{pmatrix} z \\ r \end{pmatrix} = \frac{z!}{2!(z-2)!} \tag{7.84}$$

$$= \frac{z(z-1)}{2}. \tag{7.85}$$

Alternatively, it may be noted that the correlation between trials of *different* sets of z trials will be zero (each set of z trials is generated by a different vector of independent random numbers). Thus, if each tallied trial constitutes the average of all the trials in a set of z trials, the need for computation of the covariances is negated. In addition, this illustrates the mechanism by which the variance of the estimate is reduced as each average of each set of z trials, since it contains very nearly the expected number of failures of each component, will be expected to be near to the population average thus reducing the spread of the outcomes of the independent trials. Inevitably, however, this dictates that the number of trials required will be factor of z greater than the number required by direct Monte Carlo sampling meaning that the variance of the dagger sampling estimate needs to be z times less than that of direct sampling for any benefit to be seen.

This procedure also serves to illustrate the way in which dagger sampling can be regarded as an extension of antithetic variates. While with rare events modelled using antithetic variates, the 'antithetic event' cannot be guaranteed to happen (i.e. if the component state caused the random number U' is 'in service' on one trial, the use of $1 - U'$ will not guarantee that the component is 'out of service' on the next trial), with dagger sampling the 'out of service' state is virtually guaranteed to happen once in z trials.

The antithetic random number of $1 - U'$ can be thought of as being a reflection in the interval $[0, 1]$ about 0.5 of the initial number U'. In this way, the antithetic variates method causes the $[0, 1]$ interval to be divided into two with the out of service state comprising only a small part of one of them. Dagger sampling, on the other hand, divides $[0, 1]$ into sufficient intervals to virtually guarantee that one of them entirely includes the 'out of service' state.

A further advantage of dagger sampling is that it better guarantees that the number of occurences of a particular state for a particular component matches the expected number.

The question may now be raised as to what happens when not all the components comprising the modelled system have the same probability of failure. The necessary number of partitions of the $[0, 1]$ interval becomes less than obvious and reference [56] fails to address this possibility. It may be understood that the primary need, dictated by the need to avoid the calculation of whole sets of covariances, is to find the number of trials in which there is some correlation between each trial due to the use of similar vectors of random numbers. Tallies may then be kept of the outcomes of independent *sets* of trials. There are a number of possible ways of achieving this which will now be described in turn.

Suppose that three components have probabilities of failure of p_1, p_2 and p_3 such that

$$p_2 < p_3 < p_1. \tag{7.86}$$

The number of trials in which the states of each are determined by one random number for each component are then $z_1 = 1/p_1, z_2 = 1/p_2$ and $z_3 = 1/p_3$ where

$$z_2 > z_3 > z_1. \tag{7.87}$$

Every state of component 1 will be correlated to every other state within each cycle of z_1 trials. Every state of component 2 will be correlated to every other state within each cycle of z_2 trials, and likewise for component 3.

If cycles z_1, z_2 and z_3 are not all exactly the same, it can be seen that some correlation of the system state will exist *across* cycles, even if it is small, possibly due to the influence of only one component. Thus, two trials are guaranteed to be independent only if they exist in separate cycles, or sets, of z' trials where z' is the smallest integer of which all of z_1, z_2 and z_3 are divisors. Hence, if the estimate obtained by direct sampling is θ and that from dagger sampling is θ_z with variances of estimates of s^2_θ and $s^2_{\theta_z}$ respectively, for any benefit to be observed from use of dagger sampling over direct sampling,

$$s^2_{\theta_z} < \frac{s^2_\theta}{z'} \tag{7.88}$$

where

$$z' \geq \max(z_1, z_2, z_3) \tag{7.89}$$

$$\geq \frac{1}{\min(p_1, p_2, p_3)}. \tag{7.90}$$

Where z' is greater than all of z_1, z_2 and z_3, it can be seen that the benefit of dagger sampling is reduced compared to if z' had been equal to the longest cycle. As an alternative, no loss of accuracy will be found if

$$z' = \max(z_1, z_2, z_3). \tag{7.91}$$

In this case, if z_2 is the longest cycle, the dagger variables will be reset for the z_1 and z_3 cycles at the end of every $z' = z_2$ trials whether or not the z_1 and z_3 cycles are complete [57]. If the

dagger numbers for either fell into the remainder of the cycles not completed, then no outage will have taken place in the truncated cycle. Instead, an outage may already have taken in the truncated cycle, and may happen early in the next, too. Thus, over many such cycles, the outage rate of the component will converge to its expected value so that no bias should be found.

Using this approach, the effectiveness of dagger sampling will be determined by the rarest basic event in the modelled system.

There are two further alternatives to this scenario suggested by Kelen [57]:

1. reset all cycles at the end of each turn through the *shortest* cycle. If the shortest cycle is for component 1 and has z_1 trials, then if a random number U_2' for component 2 falls in the ith sub-interval such that $i > z_1$, the component will not fail in that cycle. If $i \leq z_1$, it will fail.

2. reset all cycles at the end of each cycle corresponding to the most significant component or component type.

In all of these methods, there is no correlation across cycles.

One final remark: the similarity between *rotation* or *reflective* sampling, described in [58] as being extensions of antithetic variates, and dagger sampling has been noted by Kelen [59] who has formulated rotation sampling in such a way that its behaviour is identical to that of dagger sampling.

7.7 Selective sampling

In this, the number of events of a particular type within a sample is fixed to equal the expected number of events of that type [60]. The work of the originator of the technique is summarised in [3] in the following way.

It is assumed that only one value is to be estimated, say $\overline{Y}$. It is further assumed that $\overline{Y}$ is discrete and has M possible values each of which has the same probability of occurring, q_i.

If there are m trials in a sample, the number of times Y has the value Y_i' is V_i. The expected value of V_i will be

$$E(V_i) = q_i m. \tag{7.92}$$

In selective sampling, the number of times that a particular value of Y is sampled is fixed to the expected number of times i.e.

$$V_i = q_i m \tag{7.93}$$

where

$$\sum_{i=1}^{M} V_i = m \sum_{i=1}^{M} q_i = m. \tag{7.94}$$

The number of times that a value occurs in a sample is, of course, an integer i.e. V_i is an integer. However, $q_i m$, in general, is not an integer. $q_i m$ is then made to be an integer in such a way that the sum of the squared deviations between the "theoretical" values $q_i m$ and V_i is minimized i.e. z is minimized where

$$z = \sum_{i=1}^{M} (V_i - q_i m)^2. \tag{7.95}$$

It remains then for each V_i to be chosen subject to the need to minimize z in equation (7.95). Brenner gives an algorithm for this purpose which sets V_i to the integer closest to $q_i m$. If it turns out that $\sum V_i > m$ then the terms which have the smallest fractional parts of $q_i m$ which are greater than 0.5 have the values V_i reduced by 1 until $\sum V_i = m$. Likewise, if $\sum V_i < m$, the terms which have the largest fractional parts of $q_i m$ which are less than 0.5 have the values V_i increased by 1 until $\sum V_i = m$.

The value of V_i which minimizes z may be denoted V_i^*. The number of times that Y_i' occurs is now fixed. In order to meet this number in the sample, *sampling without replacement* is performed whereby the probability of sampling the value Y_i' in the tth trial (where $t = 1, \ldots, m$) is

$$P(Y_i') = \frac{V_i^* - g_i}{m - t + 1} \tag{7.96}$$

where g_i is the number of times that Y_i' has already been sampled.

It is noted in [3] that the method may be biased. It is also noted that it may theoretically be applied to the estimation of continuous random variables whose values are divided into a finite number of classes or strata.

7.8 Resampling Techniques

The techniques described here are not, strictly speaking, variance reduction techniques. However, they are useful in estimating the variance of estimates of statistics derived from distributions that are far from normal or from small samples such that the sampling distribution cannot be assumed to be normal. They can also help to reduce any bias inherent in an estimate. The two techniques which will now be described are the "jackknife" and the "bootstrap".

7.8.1 The jackknife

The jackknife was the first method devised to estimate biases and standard errors. It was originally proposed by Quenouille in the 1940s [61] and was refined by Tukey [62]. It involves

recalculating the statistic of interest for the sample with one observation removed. This is done with the removal of each observation in turn, and the resulting statistics are recombined to give an estimate of the bias in the original sample. It can thus also be used to compensate for the bias to give a better estimate and to give a better estimate of the population variance than can be obtained by conventional techniques.

Suppose that a true mean μ is to be estimated by the mean $\overline{X}$ of a sample of size n. The estimate will be denoted by the statistic T_n. The bias in the estimate will be given by

$$\text{bias}(T_n) = E(T_n) - \mu \tag{7.97}$$

where

$$T_n = \overline{X}. \tag{7.98}$$

Now let $T_{n-1,i}$ be the statistic (i.e., in this case, the mean) based on $n-1$ observations $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$, i.e. with the ith observation from the original sample removed. Then, the average of the means of the samples with each observation removed in turn, i.e. the average of the $T_{n-1,i}$ for $i = 1, \ldots, n$, is denoted $\overline{T}_n$ such that

$$\overline{T}_n = \frac{1}{n}\sum_{i=1}^{n} T_{n-1,i}. \tag{7.99}$$

Quenouille then derives a *jackknife bias estimator* b_{jack} which is

$$b_{jack} = (n-1)(\overline{T}_n - T_n). \tag{7.100}$$

This leads to a bias-reduced *jackknife estimator* of μ,

$$\begin{aligned} T_{jack} &= T_n - b_{jack} && (7.101)\\ &= nT_n - (n-1)\overline{T}_n. && (7.102) \end{aligned}$$

(See [61, 63] for proofs of these relationships).

Tukey's principal contribution to use of the jackknife was in showing that it could be used to construct variance estimators [62]. With T_{jack} re-written as

$$T_{jack} = \frac{1}{n}\sum_{i=1}^{n}\left[nT_n - (n-1)T_{n-1,i}\right], \tag{7.103}$$

Tukey defined

$$\tilde{T}_{n,i} = nT_n - (n-1)T_{n-1,i} \tag{7.104}$$

as "jackknife pseudovalues" and conjectured that, for $i = 1, \ldots, n$, they may be treated as independent and identically distributed (i.i.d.), and as having approximately the same variance as $\sqrt{n}T_n$. In this way, he defined the *delete-1 jackknife variance estimator* for T_n given by

$$v_{jack} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\tilde{T}_{n,i} - \frac{1}{n}\sum_{j=1}^{n}\tilde{T}_{n,j}\right)^2 \tag{7.105}$$

$$= \frac{n-1}{n}\sum_{i=1}^{n}\left(T_{n-1,i} - \frac{1}{n}\sum_{j=1}^{n} T_{n-1,j}\right)^2 \tag{7.106}$$

$$= \frac{n-1}{n}\sum_{i=1}^{n}\left(T_{n-1,i} - \overline{T}_n\right)^2 \tag{7.107}$$

$$= (n-1)\left[\frac{1}{n}\left(\sum_{i=1}^{n} T_{n-1,i}^2\right) - \overline{T_n}^2\right] \tag{7.108}$$

$$= \frac{n-1}{n}\left[\left(\sum_{i=1}^{n} T_{n-1,i}^2\right) - \frac{1}{n}\left(\sum_{i=1}^{n} T_{n-1,i}\right)^2\right]. \tag{7.109}$$

The advantage of the jackknife is that it is less dependent on model assumptions (e.g. the normality of the sampling distribution) than traditional methods. Its disadvantage is, clearly, the extra computation needed to calculate the required statistic n times.

7.8.2 The bootstrap

The bootstrap method, generally attributed to Efron [64], is a "robust" statistical method in that it is claimed to deal adequately with non-smooth (i.e. discontinuous) and biased statistics, and with non-normal sampling distributions. It involves taking an existing sample and resampling it, drawing observations from that sample with replacement to obtain a series of *bootstrap samples*. These can then be compared with each other and with the original sample to gain more complete information about the population from which the original sample was drawn. In particular, it allows the formation of an *empirical distribution function* which can be used to derive confidence limits for underlying distributions which, for example, do not permit the application of the central limit theorem where the distribution of the mean is assumed to be normal. The disadvantage of the method is that the empirical distribution is obtained at the expense of considerably more computation.

Bootstrap-t confidence limits

According to standard statistics, the "standard score" of an observation $\overline{X}$ on a sampling distribution is denoted by z and is given by

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{7.110}$$

where μ is the mean of the underlying distribution, σ the standard deviation of the underlying distribution and n is the sample size.

The quantity $\sigma/\sqrt{n}$ is known as the standard error.

By the central limit theorem, i.e. when n is large enough, the standard scores of the sampling distribution are assumed themselves to follow a normal distribution. That is to say that, if a

number of samples of similar (large) size are taken from the underlying distribution, the means of each of those samples are assumed to follow a normal distribution. This is then used to derive confidence limits.

A difficulty of statistics is that samples are often too small for the characteristics of the underlying distribution for the sampling distribution to be normal. The simplest way of overcoming this using the bootstrap is to replace $\overline{X}$ and the standard error in equation (7.110) by the bootstrap equivalents.

Bootstrap quantities are generally identified by a $*$ superfix. Thus, the bootstrap standard score of the sampling distribution for the bth bootstrap sample is [65]

$$z_b^* = \frac{\overline{X}_b^* - \overline{X}}{se_b^*} \tag{7.111}$$

where $\overline{X}_b^*$ is the mean of the bth bootstrap sample and se_b^* is the standard error of the bth bootstrap sample. Then, the $\alpha/2$th percentile of z_b^* where $100(1-\alpha)$ is the percentage degree of confidence, is estimated by the value $t^{(\alpha/2)}$ such that number of values of z_b^* less than or equal to $t^{\alpha/2)}$ equals $B(\alpha/2)$ where B is the total number of bootstrap samples. Likewise, $t^{(1-\alpha/2)}$ is given as the $B(1-\alpha/2)$th value of z_b^*. This implies that, for example, if $B = 1000$ and $100(1-\alpha) = 90.0$, the estimate of the 5% point, i.e. $t^{(1-\alpha/2)}$, will be the 50th largest value of z_b^* while the 95% point, $t^{(\alpha/2)}$, will be the 950th largest.

Finally, the so-called "bootstrap-t" confidence interval is

$$(\overline{X} - t^{(1-\alpha/2)} \cdot se, \overline{X} - t^{(\alpha/2)} \cdot se) \tag{7.112}$$

where se is the standard error of the original sample.

While this would appear to be quite straightforward to calculate, three things should be noted:

- the number of bootstrap samples B must be large, $B \geq 1000$.
- the interval derived is valid only for the given original sample.
- the interval can be heavily influenced by a few outlying data points and can thus give somewhat erratic results.

These difficulties lead to consideration of the "bootstrap BC_a", or *bias-corrected and accelerated* interval. This is described below.

Bootstrap BC_a confidence limits

The bootstrap BC_a confidence intervals are based on "percentile methods" whereby results from a histogram of the bootstrap samples are used directly rather than using bootstrap standard scores to find "pivot points" as in the bootstrap-t method.

If $\overline{X}^{*(\alpha/2)}$ is the $100(\alpha/2)$th percentile of B bootstrap replications $\overline{X}^*_1, \overline{X}^*_2, \ldots, \overline{X}^*_B$, the interval $(\overline{X}_{lo}, \overline{X}_{hi})$ of the intended coverage $1-\alpha$ is obtained from

$$(\overline{X}_{lo}, \overline{X}_{hi}) = (\overline{X}^{*(\alpha/2)}, \overline{X}^{*(1-\alpha/2)}). \tag{7.113}$$

Thus, the percentile interval for $B = 2000$ and $\alpha = 0.1$ is the interval extending from the 100th to the 1900th ordered values of the 2000 numbers $\overline{X}^*_b$.

The BC_a interval end-points are given in a similar way but depend on the *acceleration* a and the *bias correction* z_0 [65]. These are used to obtain α_1 and α_2 and the interval is, in turn,

$$(\overline{X}_{lo}, \overline{X}_{hi}) = (\overline{X}^{*(\alpha_1)}, \overline{X}^{*(\alpha_2)}). \tag{7.114}$$

The values α_1 and α_2 are given by

$$\alpha_1 = \Phi\left(z_0 + \frac{z_0 + z^{(\alpha/2)}}{1 - a\left(z_0 + z^{(\alpha/2)}\right)}\right) \tag{7.115}$$

$$\alpha_2 = \Phi\left(z_0 + \frac{z_0 + z^{(1-\alpha/2)}}{1 - a\left(z_0 + z^{(1-\alpha/2)}\right)}\right) \tag{7.116}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function and $z^{(\alpha/2)}$ is the $100(\alpha/2)$th percentile point of the standard normal distribution.

It may be noted that if a and z_0 are zero, equations (7.115) and (7.116) reduce to the simple percentile method.

The bias correction z_0 is calculated by

$$z_0 = \Phi^{-1}\left(\frac{\text{number of } \overline{X}^*_b < \overline{X}}{B}\right) \tag{7.117}$$

where $\Phi^{-1}(\cdot)$ indicates the inverse function of a standard normal cumulative distribution function.

There are various ways to calculate the acceleration a. One uses a jackknife method [65] in which

$$a = \frac{\sum_{i=1}^{n}\left(\overline{X} - \overline{X}_{n-1,i}\right)^3}{6\left[\sum_{i=1}^{n}\left(\overline{X} - \overline{X}_{n-1,i}\right)^2\right]^{1.5}} \tag{7.118}$$

where $\overline{X}_{n-1,i}$ is the mean of the original sample with the ith value removed. n is the size of that original sample.

While the effect of outlying trials is reduced with respect to the bootstrap-t, the required number of bootstrap samples is still large.

7.8.3 Use of the bootstrap

As was noted above, the disadvantage of the bootstrap is the large computational overhead. This is especially large when the method is being used to derive confidence limits as at least 1000 bootstrap samples are likely to be needed [66, 65].

One possible approach may be to perform the bootstrap itself by a Monte Carlo estimation, apply classical confidence limits for the upper and and lower bounds being calculated by the bootstrap to ascertain when to stop the simulation.

The large computational overhead effectively precludes the bootstrap from use as a test for convergence of a Monte Carlo simulation to a given precision and degree of confidence. However, it *can* be used to quantify the confidence interval of a fixed, given sample where the statistic has a non-normal distribution.

Chapter 8

Application of Monte Carlo Simulation in Composite Power System Reliability

The operation and planning of a power system have long been recognised as having a considerable number of uncertainties, not least due to unexpected failures of plant, whether they be the sudden unavailability or "derating" (reducing of capacity) of generating plant or the tripping of lines or transformers due to faults, and attempts have been made to represent these uncertainties by probabilistic analysis since the 1940s. Some engineers have argued that, given all these uncertainties, a power system should be described in terms of its reliability i.e. its overall ability to perform its function.

In power systems, reliability is conventionally broken into two aspects: *adequacy* and *security* [52]. Adequacy relates to "the existence of sufficient facilities within the system to satisfy the consumer load demand or system operational constraints." Security, on the other hand, relates to "the ability of the system to respond to dynamic or transient disturbances arising within the system." Computational tools to assess the latter category are not yet well-developed, so "power system reliability" tends to imply adequacy.

8.1 Measures of component reliability

Individual components' reliabilities may be expressed in a number of ways. In essence, the reliability expresses the proportion of time the component is "in service" or "available" for. Data may also be used, however, to express the probability of the component being in service at a particular time.

The *availability* A of a component is expressed as [4]

$$A = \frac{\sum (\text{up time})}{\sum (\text{up time}) + \sum (\text{down time})} \tag{8.1}$$

where the "up time" is the total time the component is in service and the "down time" the total time when it is not. A may be re-expressed in terms of the "mean time to failure" F and the

"mean time to repair" R so that

$$A = \frac{F}{F+R}. \tag{8.2}$$

These mean times may further be re-expressed in terms of expected "failure rate" λ and the expected "repair rate" μ such that

$$\lambda = \frac{1}{F} = \text{number of failures per unit time} \tag{8.3}$$

$$\mu = \frac{1}{R} = \text{number of reapirs per unit time} \tag{8.4}$$

$$A = \frac{\mu}{\mu+\lambda}. \tag{8.5}$$

The sum of the mean time to failure and mean time to repair is known as the "mean time between failures".

The converse of the availability is the unavailability U and it can be defined in similar terms[1]:

$$U = \frac{\sum \text{(down time)}}{\sum \text{(down time)} + \sum \text{(up time)}} \tag{8.6}$$

$$= \frac{\lambda}{\lambda+\mu} \tag{8.7}$$

$$= \frac{R}{R+F} \tag{8.8}$$

In general, the failure and repair rates need not be constant but may be functions of time, though assumptions that they are constant normally suffice.

The probability $P(t)$ that the component is available at time t is sometimes known as the "cumulative failure distribution function" and is the integral of the "failure density function" f(t), i.e.

$$P(t) = \int_0^t f(t)dt. \tag{8.9}$$

Likewise, the "survivor function" S(t) expresses the probability that the component will still be in service at time t:

$$S(t) = 1 - P(t) = 1 - \int_0^t f(t)dt. \tag{8.10}$$

Clearly, the failure density function, $f(t)$ is

$$f(t) = \frac{dP(t)}{dt} \tag{8.11}$$

$$= -\frac{S(t)}{dt}. \tag{8.12}$$

[1]The unavailability is sometimes known as the "forced outage rate".

The failure density function may be fitted to any one of a number of standard functions such as the normal, Weibull or exponential. The "negative exponential" is the most well-known so that, when λ is assumed constant, the repair rate is infinite and the component is available at time $t = 0$, the probability of finding the component on outage at time $t = T$ is

$$P(t) = 1 - e^{-\lambda t}|_{t=T}. \tag{8.13}$$

If $\lambda t \ll 1$, this is

$$P(t) \approx \lambda t. \tag{8.14}$$

It may be noted that when $f(t)$ is taken to be negative exponential, the *a posteriori* probability $P_c(t)$, the probability of the component failing during t *given* that the component has survived up to T_0 is

$$\begin{aligned} P_c(t) &= \frac{\text{Prob(surviving up to } T_0 \text{ and failing during the period } T_0 \text{ to } T_0 + t)}{\text{Prob(surviving up to } T_0)} && (8.15)\\ &= \frac{\int_{T_0}^{T_0+t} f(t)dt}{\int_{T_0}^{\infty} f(t)dt} && (8.16)\\ &= \frac{e^{-\lambda T_0} - e^{\lambda(T_0+t)}}{e^{\lambda T_0}} && (8.17)\\ &= 1 - e^{-\lambda t} && (8.18)\\ &= P(t) && (8.19) \end{aligned}$$

i.e. the probability of failure is independent of how long the component has already been operating for and depends only on the length of the interval t. In other, $P(t)$ has no memory of the past.

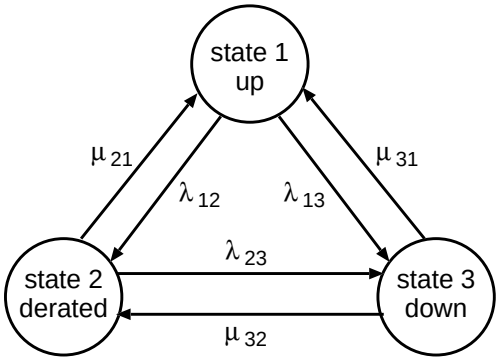

Figure 8.1: Three-state model of a generator

The notion of "failure" and "repair" rates may be extended to models where the components have more than two states. Figure 8.1 shows a three-state model for a generator where the three states are "up", "derated" (where the maximum output has been reduced) and "down". Transitions rates between the three states are shown. These may all be denoted by λ with the subscripts written in the appropriate order to suggest the order of the transition, but here transitions towards full generating capacity are represented by μ to maintain consistency with the idea of a repair rate. If all the transition density functions are assumed to be exponential or negative exponential and the time interval t is assumed to be short such that the probability of a transition in an interval of length t approximately equals the product of the transition rate and t, a "discrete Markov chain" can be depicted as in figure 8.2.

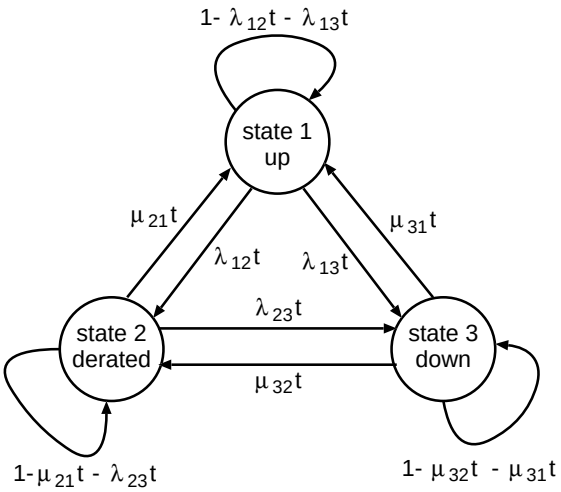

Figure 8.2: Discrete Markov chain for three-state model

If the period t is so short that it may be assumed that only one state transition occurs in that time, a state diagram of a system comprising a number of components may be drawn showing appropriate system state transition probabilities and the probability of the system being in any one state may be derived.

8.2 Power system reliability indices

Various indices have been used in the engineering literature to measure adequacy. Most of them are expected values of random variables (though some give a p.d.f.). Among these are [52]

Loss of load expectation ($LOLE$) in days/year or hours/year:

$$LOLE = \sum_{i \in S} p_i T \tag{8.20}$$

where p_i is the probability of system state i, T is the time unit of the index (i.e. either one day or one hour) and S is the set of all possible system states associated with loss of load. The index gives the expected (or mean) number of days or hours in a given period (usually one year) in which *the daily peak load or hourly load is exceeds the available generating capacity*.

Note that neither the severity of the generation deficiency nor the frequency or duration of the loss of load are indicated.

Loss of energy expectation ($LOEE$) in MWh/year:

$$LOEE = \sum_{i \in S} 8760 C_i p_i \tag{8.21}$$

where p_i and S are as above and C_i is the number of MWs of load lost for system state i. It gives *the expected energy not supplied by the generating system due to the load exceeding the available generating capacity*. When the energy actually supplied is divided by the

total energy demanded, a normalized index known as the *energy index of reliability* (EIR) is found.

Loss of load frequency ($LOLF$) in number of occurences per year:

$$LOLF = \sum_{i \in S} (F_i - f_i) \tag{8.22}$$

where F_i is the "frequency of departing system state i" and f_i is the "portion of F_i which does not change from being part of the 'loss-of-load state set' to being part of 'no-loss-of-load state set' ".

Loss of load duration ($LOLD$) in hours per occurrence:

$$LOLD = \frac{LOLE}{LOLF}. \tag{8.23}$$

This represents the duration of each state in which load had been lost.

The term *load curtailment* is sometimes used in the context of system states in which some load has been lost. Some authors seem to use it distinguish "composite system adequacy" (in which transmission outages are considered along with generation outages or deratings) from "generation adequacy" (where transmission system effects are not considered).

8.3 Simulation approaches in reliability evaluation

There are three main simulation approaches in reliability evaluation and another introduced in [52].

8.3.1 State sampling

The system state depends on the combination of all component states while each component state can be determined by sampling the probability that the component appears in that state. The p.d.f. of each component is usually modelled by a uniform distribution between $[0, 1]$ with two possible states - "success" and "failure" - and a probability of failure p_i.

If the state of a system $\mathbf{\Psi}$ is described by a vector comprising the states of the m components $\Psi_1, \Psi_2, \ldots, \Psi_m$ such that

$$\mathbf{\Psi} = [\Psi_1, \Psi_2, \ldots, \Psi_m], \tag{8.24}$$

then the expected value of some index $L(\mathbf{\Psi})$ will be

$$E(L) = \sum_{\mathbf{\Psi} \in S} L(\mathbf{\Psi}) P(\mathbf{\Psi}) \tag{8.25}$$

where S is the set of all possible system states. In a simulation, this can be estimated by

$$E(L) = \frac{1}{n} \sum_{\mathbf{\Psi} \in S} L(\mathbf{\Psi}) n_H(\mathbf{\Psi}) \tag{8.26}$$

where $n_H(\mathbf{\Psi})$ is the number of occurrences of state $\mathbf{\Psi}$ and n is the total number of trials.

Advantages of the "state sampling" approach include

- simplicity - only necessary to generate random numbers on a uniform distribution.
- only the component state probabilities are required.

The disadvantage is that dependency of a component's state or a system state on previous states as behaviour develops through time cannot be represented.

8.3.2 State duration sampling

This is based on sampling the p.d.f. of the duration of a component state. Each component has an initial state and the duration of each remaining in that state is sampled, for example from an exponential distribution. If the state of a component changes within the time span of the simulation, how long it remains in the next state is sampled repeatedly until the time span is reached. Thus descriptions of the system state for the whole time span are obtained and the desired index is calculated.

The advantages of this approach are

- it can easily be used to calculate a frequency index.
- any state duration p.d.f. can be considered.
- the p.d.f. of a desired reliability index can be calculated.
- the "history" and trend of a state can be taken into account.

Disadvantages are that more computation and storage are required (not least in the computation of each random variate[2]) and more parameters are needed to define the variates.

8.3.3 Sequential sampling

In common with state duration sampling, sequential simulation [67] allows the history of a state to be taken into account. Unlike state duration sampling where the history of each component's

[2] A *random variate* is a random variable which follows a given p.d.f.

state is assembled independently and different system states must then be identified and evaluated, in sequential sampling, the period of interest is divided into time intervals and each one is simulated in sequence with the state of each component at the start of each time interval generated randomly depending on the component's state at the start of the previous interval and the duration of the interval.

Conventional sequential Monte Carlo simulation is generally what may be called a "synchronous" sequential simulation [68] in that successive time slots are simulated with constant time intervals. The interval length may be chosen to be so short that a maximum one component state will have changed, or to track significant changes in particular variable (say the load in a power system reliability analysis).

Advantages of the method are that the evolution of both individual component states and the system state can be tracked and that any form of state duration distribution can be modelled. Disadvantages are the computational demand, the need to choose a suitable length of time interval, and the need to generate many random numbers at the start of each interval.

8.3.4 State transition sampling

This follows the transition of system state rather than of individual component states [69, 4] and is similar to sequential sampling.

It is assumed that the system starts with state $\mathbf{\Psi}^{(1)}$ and moves successively to states $\mathbf{\Psi}^{(2)}, \ldots, \mathbf{\Psi}^{(n)}$. The transition of the system state depends randomly on the state duration of the component which departs earliest from its present state. The duration of each state $\mathbf{\Psi}^{(k)}$ can be described in terms of the random variable $T^{(k)}$ which is the minimum of the state durations of each component T_i for $i = 1, \ldots, m$.

Since the state duration of each component follows an exponential distribution, it can be proved that each $T^{(k)}$ follows an exponential distribution. Further, each depends on a conditional probability. For example, if the transition from $\mathbf{\Psi}^{(k)}$ to $\mathbf{\Psi}^{(k+1)}$ takes place at time t_k, the probability that this transition is caused by the jth component is

$$P_j^{(k)} = P\left(T_j^{(k)} = t_k | T^{(k)} = t_k\right). \tag{8.27}$$

The state transition appraoch may be described as an "asynchronous" sequential simulation method [68] in that simulations of successive states are carried out but the time of the event which changes the system state varies.

The advantages of this approach are

- exact frequency index information can be calculated without the need to sample the distribution function and store chronological data.
- only one random number is required to produce the next system state unlike in the state sampling approach which requires one random number for each component.

- the “history” and trend of a state can be taken into account.

The disadvantages are that it is computationally expensive and it applies only to exponentially distributed component state durations.

8.3.5 Hybrids

It has already been noted that the three methods outlined above, state duration sampling, sequential simulation and state transition sampling, which deal with the evolution of system states, necessary when a system state depends on what happened before, have the significant disadvantage of being highly computationally intensive. This has led researchers to adopt simplifying assumptions and implement hybrids of the various sampling approaches.

A hybrid of state sampling and sequential simulation is reported in [70]. The basic premise of the work is that the generation of states is not computationally demanding, but the analysis of their effects is. The duration of load interruptions is, however, required, something which is not available from simple state sampling but only from some sampling approach that represents sequences of events. A large set of sequences of states is sampled but not simulated. Let this set of sequences be called S. Since only states where loss-of-load are of interest, simple sampling of sequences from within S is carried out. If there are n_S sequences in the set, this involves partitioning the [0,1] interval in n_S sub-intervals and generating a random number from a uniform distribution. If the random number falls in the ith sub-interval, the ith sequence is taken (with replacement). One time interval, say the jth, from that sequence is then sampled and simulated. If the jth snap-shot contains a loss of load, then the time intervals after it are simulated in sequence until the loss of load condition has gone, and simulated before it until the start of the loss of load condition. If, on the other hand, the jth interval of the ith sequence does not contain a loss of load, another sequence from S is sampled. Sampling from S then continues until sufficient precision in the estimate has been obtained.

A hybrid of state transition and the pseudo-sequential sampling described in [70] is reported in [71]. The method is essntially the same as in [70] but the paper describes the additional use of a variance reduction technique known as “conditional Monte Carlo” [72, 73].

8.4 Example of the “state sampling” method

The following is a description of a simple example given in [52].

The objective is to find the expected value of some reliability index from equation (8.26). As outlined above, the reliability index is invariably concerned with loss of load or “load curtailment”. The index need only be worked out for those states in which curtailment occurs and experience shows that they are relatively few in number. The key, then, is to identify them.

The complete set of states may be sub-divided in the manner shown in figure 8.3.

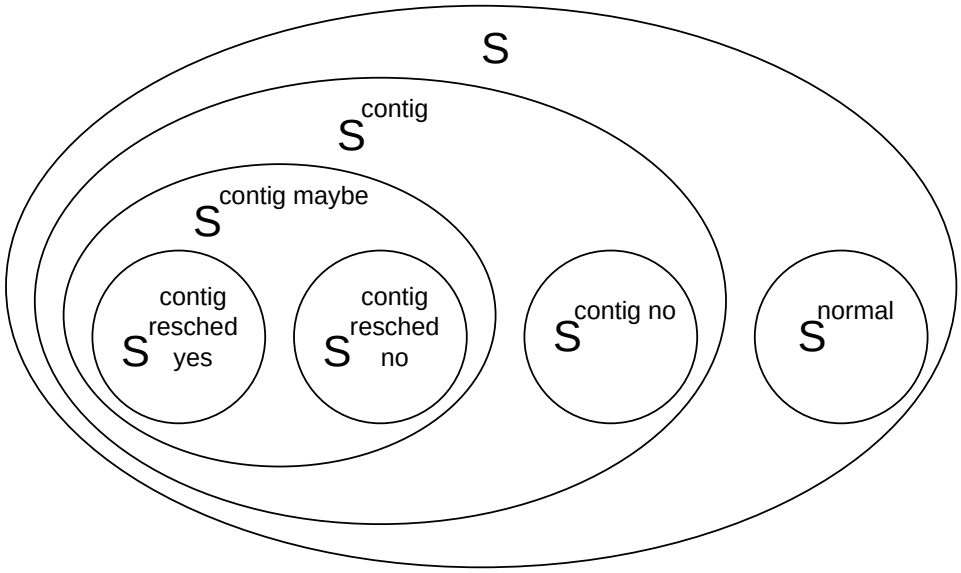

Figure 8.3: Classifications of system state in reliability evaluation

A system state is obtained by sampling the state of each individual component, transmission and generation. Where any component is in the "down" state, the resulting system state is denoted a "contingency state" and is a member of S^{contig}.

If a generating unit has been lost or derated, the capacity of the remaining generating units at the bus where it was lost is first worked out. This is done for each generator derating or loss. If there are n_i generating units at bus i, the initial (normal) active power outputs are $P_{ik}, k \in n_i$ and the maximum outputs in contingency system state are $P_{ik}^{max}, k \in n_i$, the truth of the following inequality is found:

$$\sum_{k \in n_i} (P_{ik}^{max} - P_{ik}) \geq 0 \tag{8.28}$$

If (8.28) is satisfied for all buses where generator loss or derating has taken place, whether any load curtailment will take place will depend on the effects of any transmission outages. If, however, (8.28) fails to be satisfied for at least one bus, the system state will belong to set $S^{contig\ maybe}$. Further analysis will be necessary to determine if the state is a member of $S^{contig\ resched\ no}$ or $S^{contig\ resched\ yes}$.

If a transmission component has been lost, it must be determined if any overloads have been created. If there are no overloads and equation (8.28) has been satisfied for all generator buses, the system state belongs to $S^{contig\ no}$. Otherwise, the state belongs to $S^{contig\ maybe}$ and rescheduling of generation must take place.

For any system state in set $S^{contig\ maybe}$, generation must rescheduled to meet demand and remove overloads while minimising load curtailment. This is often achieved by means of a linear program. Then, if any load has been curtailed, the system state belongs to $S^{contig\ resched\ yes}$ and the appropriate index should be worked out. Otherwise, the state belongs to $S^{contig\ resched\ no}$.

8.5 Solution by "state enumeration"

This method is the traditional one of carrying out reliability analysis and is the main alternative to using Monte Carlo simulation. A brief description is offered here so that the reasons for using Monte Carlo can be better understood.

State enumeration involves explicitly evaluating *every* case in a pre-defined sequence until a certain level of accuracy is judged to have been obtained. Its effect is to determine according to some rule which subset of the whole set of possible states is worthy of study.

The rule used often depends on use of so-called *capacity outage probability tables* [4]. A probability cut-off is defined such that any contingencies deemed to be less likely than the cut-off are also deemed to be not worthy of study having insignificant impact on the reliability index being used.

If a state is termed a "level 1 state" when one fault has occurred causing one component to go out of service, a "level 2 state" when 2 faults have occurred and so on, where the probability of each fault is the same, the probability of a trial having a particular level will be given by the binomial distribution $f_B(x)$ where

$$f_B(x) = \left(\begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{n-x} \tag{8.29}$$

and n is the number of components, x is the level of a particular trial and p is the probability of a component being out of service.

When the number of possible levels (i.e. number of components which may be faulted) is large, the binomial distribution may be approximated by the normal distribution. A graph may be drawn of the form shown in figure 8.4 of the cumulative probability $F_X(y)$ of being in particular state level y against the state level i.e. of the probability of being at that level or below.

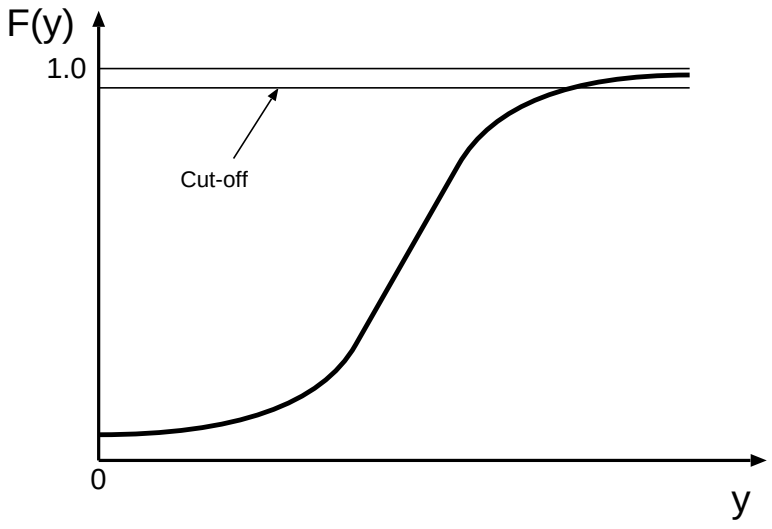

Figure 8.4: Cumulative distribution function of level probability

It can be seen that a cumulative cut-off probability may be fixed on the figure so that state levels

where the cumulative probability is greater than the cumulative cut-off are neglected. Implicit in this approach, however, is the assumption that the contributions of each state level to the overall reliability index follow the shape of $1 - F_X(y)$ i.e. the characteristic shown in figure 8.5 since the number of possible contingency scenarios at each state level follows that shape.

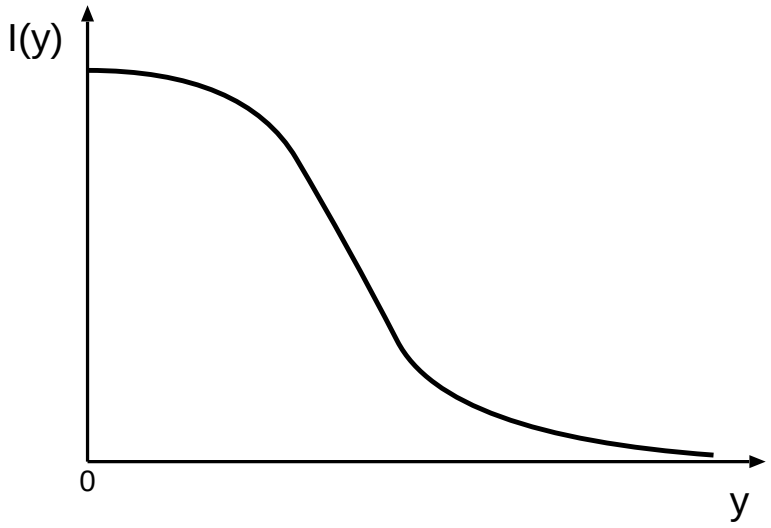

Figure 8.5: Contribution of state levels to reliability index

It will be recalled that these characteristics, derived from a binomial distribution, are based on the assumption that all faults have the same probability. While this may be approximately true for a transmission system, it is not true for a real power system comprising both transmission and generation where generator failure rates are generally significantly higher than branch failure rates. Add in the phenomenon of cascade tripping and the possibility of protection failures meaning that one fault can lead to the failures of many branches and it can be seen that the probabilities of each possible scenario must be evaluated explicitly. Further, consideration of the characteristics of a large power system reveals that the concurrent outage of, say, 5 lines due to the (not very probable) concurrence of 5 separate faults may lead to no unreliability if all the lines are remote from each other. Use of the enumeration approach when the system has 500 items of plant would make the binomial distribution of state level greatest at level 5 (and the c.d.f. steepest), implying that this would make most contribution to the reliability index. On the other hand, the concurrent outage of two lines in close proximity may lead to, for example, a great deal of load being shed.

Thus it can be seen that for a reasonably realistic model of a large power system, the state enumeration approach with a probability cut-off has serious limitations. A Monte Carlo estimation of system reliability, even with its great computational demands, becomes extremely attractive.

Chapter 9

Applications of Variance Reduction in Power Systems

9.1 Application of importance sampling to power system planning

The National Grid Company has developed a number of tools for assessment of the value of investment in transmission system plant [74]. The main criterion for the investment is that system reliability is improved or the operating cost reduced. When assessing the reliability, work has been done in exploiting importance sampling [75]. The following offers two illustrations of that work showing the influence of system outage probabilities on the number of trials needed in a Monte Carlo evaluation of reliability [1].

9.1.1 Outline of method

In common with section 8.2, a power system is considered with n_l line components and n_g generator components giving a total of m components subject to random failures. The set of all possible system states is S and an individual state is $\mathbf{\Psi} \in S$ comprising the states of each component $\{\Psi_1, \Psi_2, \ldots, \Psi_m\}$.

A system failure index $I(\mathbf{\Psi})$ is defined such that $I(\mathbf{\Psi}) = 1$ if any consumer disconnections occur in state $\mathbf{\Psi}$ and 0 otherwise. The system failure probability is then

$$f = E[I(\mathbf{\Psi})]. \tag{9.1}$$

If $p_{\mathbf{\Psi}}$ is the probability that state $\mathbf{\Psi}$ will occur,

$$f = \sum_{\mathbf{\Psi} \in S} p_{\mathbf{\Psi}} I(\mathbf{\Psi}). \tag{9.2}$$

[1] Importance sampling is also mentioned in [76] in combination with control variates and in [77] where the issue of "cascade tripping" in power system disturbances is considered.

f may be estimated in a Monte Carlo simulation by $\hat{f}$ such that

$$\hat{f} = \frac{1}{n}\sum_{i=1}^{n} I(\mathbf{\Psi}_i) \tag{9.3}$$

where n is the number of trials and $\mathbf{\Psi}_i$ is the system state in the ith trial.

The variance of f is given by[2]

$$\sigma_I^2 = \sum_{\mathbf{\Psi}\in S} I^2(\mathbf{\Psi}) p_{\mathbf{\Psi}} - \left[\sum_{\mathbf{\Psi}\in S} I(\mathbf{\Psi}) p_{\mathbf{\Psi}}\right]^2. \tag{9.4}$$

Since $I^2(\mathbf{\Psi}_i) = I(\mathbf{\Psi}_i)$ as I is always either 0 or 1, this is

$$\sigma_I^2 = \sum_{\mathbf{\Psi}\in S} I(\mathbf{\Psi}) p_{\mathbf{\Psi}} - \left[\sum_{\mathbf{\Psi}\in S} I(\mathbf{\Psi}) p_{\mathbf{\Psi}}\right]^2. \tag{9.5}$$

As $\sum_{\mathbf{\Psi}\in S} I(\mathbf{\Psi}_i) p_{\mathbf{\Psi}} = f$ and $\left[\sum_{\mathbf{\Psi}\in S} I(\mathbf{\Psi}_i) p_{\mathbf{\Psi}}\right]^2 = f^2$, this gives

$$\sigma_I^2 = f - f^2. \tag{9.6}$$

In a simulation, f is approximated by $\hat{f}$ and the variance of the population may be approximated by

$$\begin{aligned} s_I^2 &= \frac{1}{n}\sum_{i=1}^{n} I^2(\mathbf{\Psi}_i) - \left[\frac{1}{n}\sum_{i=1}^{n} I(\mathbf{\Psi}_i)\right]^2 & (9.7)\\ &= \hat{f} - \hat{f}^2. & (9.8) \end{aligned}$$

Now let state $\mathbf{\Psi}$ be sampled with probability $p^*_{\mathbf{\Psi}}$. f now becomes

$$f^* = \sum_{\mathbf{\Psi}\in S} I(\mathbf{\Psi}) w_{\mathbf{\Psi}} p^*_{\mathbf{\Psi}} \tag{9.9}$$

where the weighting $w_{\mathbf{\Psi}}$ is

$$w_{\mathbf{\Psi}} = \frac{p_{\mathbf{\Psi}}}{p^*_{\mathbf{\Psi}}}. \tag{9.10}$$

The variance of f^* is then

$$\begin{aligned} \sigma_f^{*^2} &= \sum_{\mathbf{\Psi}\in S} I^2(\mathbf{\Psi}) w_{\mathbf{\Psi}}^2 p^*_{\mathbf{\Psi}} - f^2 & (9.11)\\ &= \sum_{\mathbf{\Psi}\in S} I^2(\mathbf{\Psi}) w_{\mathbf{\Psi}} p_{\mathbf{\Psi}} - f^2. & (9.12) \end{aligned}$$

[2] A factor of $p_{\mathbf{\Psi}}$ in a summation of all the possible states is equivalent to a factor of $1/N$ where N is the population size in a summation over all the outcomes in the population.

f^* can be approximated in the simulation by

$$\hat{f}^* = \frac{1}{n}\sum_{i=1}^{n} I(\mathbf{\Psi}_i) w_{\mathbf{\Psi}_i} \tag{9.13}$$

of which the variance $s_{f*}^{*^2}$ is

$$s_I^{*^2} = \frac{1}{n}\sum_{i=1}^{n} I^2(\mathbf{\Psi}_i) w_{\mathbf{\Psi}_i}^2 - \hat{f}^{*^2}. \tag{9.14}$$

The aim of the importance sampling strategy is then to choose $w_{\mathbf{\Psi}}$ to minimise $s_I^{*^2}$.

Suppose that bias factors w_l and w_g are applied to individual line and generator outage probabilities p_l and p_g. Applied to a complete system in which all lines have uniform outage probability p_l and all generators p_g, the probability of sampling a state $\mathbf{\Psi}$ with $t_l(\mathbf{\Psi})$ lines and $t_g(\mathbf{\Psi})$ generators tripped is

$$p_{\mathbf{\Psi}}^* = (p_l w_l)^{t_l(\mathbf{\Psi})}(1-p_l w_l)^{n_l - t_l(\mathbf{\Psi})} \cdot (p_g w_g)^{t_g(\mathbf{\Psi})}(1-p_g w_g)^{n_g - t_g(\mathbf{\Psi})} \tag{9.15}$$

compared with an intrinsic probability of

$$p_{\mathbf{\Psi}} = p_l^{t_l(\mathbf{\Psi})}(1-p_l)^{n_l - t_l(\mathbf{\Psi})} \cdot p_g^{t_g(\mathbf{\Psi})}(1-p_g)^{n_g - t_g(\mathbf{\Psi})}. \tag{9.16}$$

The sampling weight is then

$$w_{\mathbf{\Psi}} = \frac{(1-p_l)^{n_l - t_l(\mathbf{\Psi})}(1-p_g)^{n_g - t_g(\mathbf{\Psi})}}{w_l^{t_l(\mathbf{\Psi})}(1-p_l w_l)^{n_l - t_l(\mathbf{\Psi})} \cdot w_g^{t_g(\mathbf{\Psi})}(1-p_g w_g)^{n_g - t_g(\mathbf{\Psi})}} \tag{9.17}$$

$$= \frac{p_l^{t_l(\mathbf{\Psi})}(1-p_l)^{n_l - t_l(\mathbf{\Psi})} \cdot p_g^{t_g(\mathbf{\Psi})}(1-p_g)^{n_g - t_g(\mathbf{\Psi})}}{p_l^{*^{t_l(\mathbf{\Psi})}}(1-p_l^*)^{n_l - t_l(\mathbf{\Psi})} \cdot p_g^{*^{t_g(\mathbf{\Psi})}}(1-p_g^*)^{n_g - t_g(\mathbf{\Psi})}} \tag{9.18}$$

where p_l^* is the biased line failure probability and p_g^* is the biased generator failure probability.

States where any given line fails are then over-represented in the simulation by a factor w_l while states in which it does not fail are over-represented by a factor $(1-p_l w_l)/(1-p_l)$ or under-represented by a factor $(1-p_l)/(1-p_l w_l)$.

It can be seen that application of the method to systems with non-uniform outage probabilities will be straightforward since separate factors from each component are simply multiplied together.

9.1.2 First illustration of method

Consider the system in figure 9.1 [75]. Let there be uniform outage probabilities such that $p_g = 0$ and $p_l = \alpha$. A suitable value of w_l is to be found.

It can be seen that a system failure state $I(\mathbf{\Psi}) = 1$ will result if

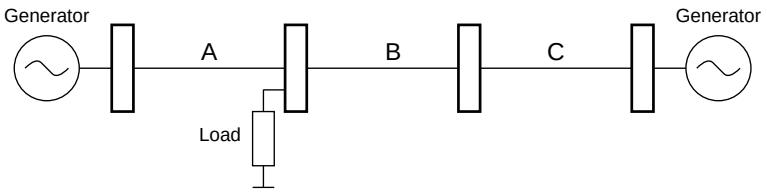

Figure 9.1: Four bus, three line example system

1. lines A and B fail but not line C. $p_{\mathbf{\Psi}} = \alpha^2(1-\alpha)$.
2. lines A and C fail but not line B. $p_{\mathbf{\Psi}} = \alpha^2(1-\alpha)$.
3. lines A, B and C all fail. $p_{\mathbf{\Psi}} = \alpha^3$.

Averaging over all system states gives

$$f = \sum_{\mathbf{\Psi} \in S} I(\mathbf{\Psi}) p_{\mathbf{\Psi}} \tag{9.19}$$

$$= 2\alpha^2 - \alpha^3. \tag{9.20}$$

When α is small,

$$f \approx 2\alpha^2. \tag{9.21}$$

The variance of a Monte Carlo estimate[3] may be found by making use of equation (5.32) with σ_I^2 approximated by s_I^2. With f approximated by $\hat{f}$, this gives a coefficient of variation of the estimate of

$$CV_{\hat{f}} = \frac{s_{\hat{f}}}{\hat{f}} = \frac{s_I}{\hat{f}\sqrt{n}} \tag{9.22}$$

$$= \sqrt{\frac{1-\hat{f}}{n\hat{f}}} \tag{9.23}$$

$$\approx \frac{1}{\sqrt{n\hat{f}}} \approx \frac{1}{\alpha\sqrt{2n}}. \tag{9.24}$$

From this, it can be seen that to obtain an estimate coefficient of variation of 10% when $\alpha = 0.01$ and $\hat{f} = 2 \times 10^{-4}$ would require $n = 5 \times 10^5$ trials.

Performing the same simulation with importance sampling, it may be found that the variance is

$$s_I^{*^2} = \frac{1}{n^*}\left[\frac{\alpha^3}{w_l^3} + \frac{2\alpha^2(1-\alpha)^2}{w_l^2(1-w_l\alpha)} - (2\alpha^2 - \alpha^3)^2\right]. \tag{9.25}$$

$s_I^{*^2}$ is smallest when

$$w_l = \frac{2}{3\alpha}. \tag{9.26}$$

[3]Note that this is the variance of the estimate not the variance of the sample.

For this value of w_l,

$$s_I^{*^2} = \frac{19\alpha^4}{2n^*} \tag{9.27}$$

and the estimate coefficient of variation is

$$CV_{\hat{f}} = \sqrt{\frac{19}{8n^*}}. \tag{9.28}$$

It can now be seen that a coefficient of variation of 10% can be achieved with only $n^* = 240$ simulations.

9.1.3 Second illustration of method

This illustration is based on that described in [78]. Consider the system shown in figure 9.2. With $\mathbf{\Psi}$ and $I(\mathbf{\Psi})$ defined as before, it can be seen that failure states result when

1. only line C fails. $p_{\mathbf{\Psi}} = \alpha(1-\alpha)^2$.
2. lines A and B fail but not line C. $p_{\mathbf{\Psi}} = \alpha^2(1-\alpha)$.
3. lines A and C fail but not line B. $p_{\mathbf{\Psi}} = \alpha^2(1-\alpha)$.
4. lines B and C fail but not line A. $p_{\mathbf{\Psi}} = \alpha^2(1-\alpha)$.
5. lines A, B and C all fail. $p_{\mathbf{\Psi}} = \alpha^3$.

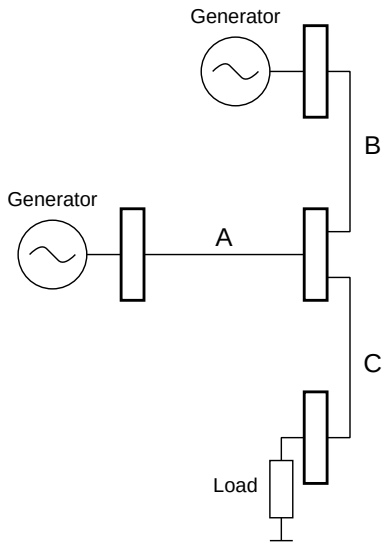

Figure 9.2: Alternative four bus, three line example system

If $p_l = \alpha$ and $p_g = 0$,

$$f = \alpha(1-\alpha)^2 + 3\alpha^2(1-\alpha) + \alpha^3 \tag{9.29}$$

$$= \alpha(1+\alpha-\alpha^2). \tag{9.30}$$

This may be approximated to α.

As in the first example in the preceding section, $\sigma_I^2 = f - f^2$ and $CV_{\hat{f}} = \sqrt{(1-f)/nf} \approx 1/\sqrt{nf}$. Let the number of trials n be found for which the estimate coefficient of variation is 10% when $\alpha = 0.01$. With $\hat{f} = f = \alpha$, $n \approx 1.0 \times 10^4$.

Now using importance sampling, it can be seen from equations (9.12) and (9.18) with $p_l = \alpha$ that

$$\sigma_I^{*^2} = \sum_{\mathbf{\Psi} \in S} I(\mathbf{\Psi}) \frac{\alpha^{t_l(\mathbf{\Psi})}(1-\alpha)^{n_l - t_l(\mathbf{\Psi})}}{p_l^{*^{t_L(\mathbf{\Psi})}}(1-p_l^*)^{n_l - t_l(\mathbf{\Psi})}} p_{\mathbf{\Psi}} - f^2 \tag{9.31}$$

$$= \sum_{\mathbf{\Psi} \in S} I(\mathbf{\Psi}) \frac{\alpha^{t_l(\mathbf{\Psi})}(1-\alpha)^{n_l - t_l(\mathbf{\Psi})}}{p_l^{*^{t_L(\mathbf{\Psi})}}(1-p_l^*)^{n_l - t_l(\mathbf{\Psi})}} \alpha^{t_l(\mathbf{\Psi})}(1-\alpha)^{n_l - t_l(\mathbf{\Psi})} - f^2 \tag{9.32}$$

$$= \sum_{\mathbf{\Psi} \in S} I(\mathbf{\Psi}) \frac{\alpha^{2t_l(\mathbf{\Psi})}(1-\alpha)^{2(n_l - t_l(\mathbf{\Psi}))}}{p_l^{*^{t_l(\mathbf{\Psi})}}(1-p_l^*)^{n_l - t_l(\mathbf{\Psi})}} - f^2 \tag{9.33}$$

where $n_l = 3$ for the three-line network under discussion [4].

Since α is small, the lowest power of α dominates so

$$\sigma_I^{*^2} \approx \sum_{\mathbf{\Psi} \in S} I(\mathbf{\Psi}) \frac{\alpha^{2t_l(\mathbf{\Psi})}}{p_l^{*^{t_l(\mathbf{\Psi})}}(1-p_l^*)^{n_l - t_l(\mathbf{\Psi})}} - f^2. \tag{9.34}$$

The only one of the three states where $t_l(\mathbf{\Psi}) = 1$ for which $I(\mathbf{\Psi}) \neq 0$ is the state corresponding to the tripping of line C with lines A and B still in service. Thus, since $f = \alpha$,

$$\sigma_I^{*^2} = \alpha^2 \left(\frac{1}{p_l^*(1-p_l^*)^2} - 1 \right). \tag{9.35}$$

From here, it can be found that $\sigma_I^{*^2}$ is minimized when $p_l^* = 1/3$.

Recalling that $CV_{\hat{f}} = \sigma_{\hat{f}}/\hat{f}$ and $\sigma_{\hat{f}}^2 = \sigma_I^{2^*}/n^*$,

$$CV_{\hat{f}} = \frac{1}{f}\sqrt{\frac{\sigma_I^{*^2}}{n^*}}. \tag{9.36}$$

With $p_l^* = 1/3$ and $f = \alpha$, this gives

$$CV_{\hat{f}} = \frac{1}{\alpha}\sqrt{\frac{5.75\alpha^2}{n^*}}. \tag{9.37}$$

Thus, the number of trials using importance sampling n^* can be found for an estimate coefficient of variation of 10%. It is found to be independent of α and equal to 575.

[4] The term $\begin{pmatrix} n_l \\ t_l(\mathbf{\Psi}) \end{pmatrix}$ does not appear as each state with the same $t_l(\mathbf{\Psi})$ is counted separately in the summation as $I(\mathbf{\Psi})$ may be different for each.

9.1.4 Remarks

A comparison can be made between the relative improvement in the number of trials needed for a given coefficient of variation between an estimate made with and without importance sampling.

From the first example in section 9.1.2, the ratio of n to n^* is

$$\frac{n}{n^*} = \frac{5 \times 10^5}{240} \approx 2080 \tag{9.38}$$

while for the second example in section 9.1.3 the relative improvement in the number of trials is

$$\frac{n}{n^*} = \frac{1.0 \times 10^4}{575} \approx 17. \tag{9.39}$$

The fact that the first example gives a better improvement is consistent with the idea that the importance sampling works better the rarer the failure states are. Consideration of equations (9.12) and (9.18) shows that for minimum variance, $w_{\mathbf{\Psi}}$ should be minimized and, in turn, the denominator of (9.18) maximized. This occurs for $0 < p_l^* < 1$ when

$$p_l^* = \frac{t_l(\mathbf{\Psi})}{n_l}. \tag{9.40}$$

For $p_l^* > p_l$ (which is necessary to achieve a reduction in variance),

$$p_l < \frac{t_l(\mathbf{\Psi})}{n_l}. \tag{9.41}$$

This can be interpreted as a restriction on the number of lines tripped [78]:

$$t_l(\mathbf{\Psi}) > p_l n_l. \tag{9.42}$$

For $p_l = 0.01$, this provides no difficulty on a small network, but for a system with around 500 lines, $t_l(\mathbf{\Psi})$ should be more than 5. In other words, importance sampling should be used to bias the Monte Carlo simulation towards the extreme conditions which lead to loss of supply [5]. However, even having the number of line trips for a particular trial greater than or equal to 5 does not guarantee the occurrence of a failure state: the tripped lines may be geographically distributed such that all load can continue to be supplied.

Another important point is that the derivation of the necessary weights for importance sampling in the above examples depended on analytical knowledge of the estimated quantity. In practice, of course, this quantity to be estimated is not known.

It is clear, then, that some form of additional knowledge is necessary to find weightings which give significant improvement to the efficiency of the Monte Carlo algorithm. The information relevant to the power system reliability problem is: which terms from the series by which the

[5] For an index that was continuous rather than simply 0 or 1, this may be interpreted as biassing the simulation to have more trials around the expected estimate to better "confirm" the validity of that estimate.

expected value is calculated contribute most to the mean probability of failure to supply. In the two illustrations given above, it has been assumed that certain terms dominate the series making approximations possible. However, this is only true for small networks. In larger networks, while individual terms corresponding to outages of single components may be large, there are many terms corresponding to different combinations of multiple outages.

It may be necessary, as suggested in [75], to run a trial simulation to get a preliminary estimate of the system failure rate in order to calculate the importance sampling weightings and the number of trials needed, and then run the full study. More qualitative knowledge gained from operators may be useful in this context to weight different system components separately according to empirical knowledge of which ones tend to cause most problems, though it may be difficult to ensure that this knowledge is "generic" i.e. can be applied similarly to all large power systems, rather than system specific. Alternatively, some representation of that empirical knowlegde may be accessible from sensitivity analysis of the intact system - changes in generation or in injections caused by losses of lines which have greater impact on other system components operating near to their limits may be weighted in proportion to those sensitivities.

9.2 Application of control variates to power system reliability

An application of control variates (or "regression sampling") is outlined in [79], [80] and [76].

In [79, 80, 76], contributions to "loss of load probability" (LOLP) and "expected power not supplied" (EPNS) indices are broken down into those caused by

- *generation* outages severe enough to lead to load curtailment even if there were no limitations in transmission capacity.
- *transmission* outages severe enough to lead to load curtailment even if all generation were available.
- *composite* outages where the effects of generation and transmission alone would not be enough to cause load curtailment but whose combined effect leads to system problems.

The expected energy not supplied caused solely by generation outages is denoted in [79] as UDG. The expected energy not supplied caused by outages of either generation or transmission or both is denoted $UDGT$. The control variate C from equation (7.23) where

$$Y(a) = Y - a(C - \mu_C) \tag{9.43}$$

is chosen to be UDG. $Y(a)$ is a new function which is an unbiased estimator of μ_C. μ_C is calculated from enumeration of UDG (in general, this enumeration will provide an estimate of μ_C). Taking $a = 1$ (an approximation which is often used in power system composite reliability analysis) and the expected values of equation (9.43) it is found that

$$E[Y(a)] = E(Y) - E(C) + \mu_C \tag{9.44}$$

$$= E(Y) \tag{9.45}$$

meaning that an estimate of a "residual", here $E(Y) - E(C) + \mu_C$, would give the desired result when $E(C) \approx \mu_C$ with, as was shown in section 7.3, a reduced estimate variance.

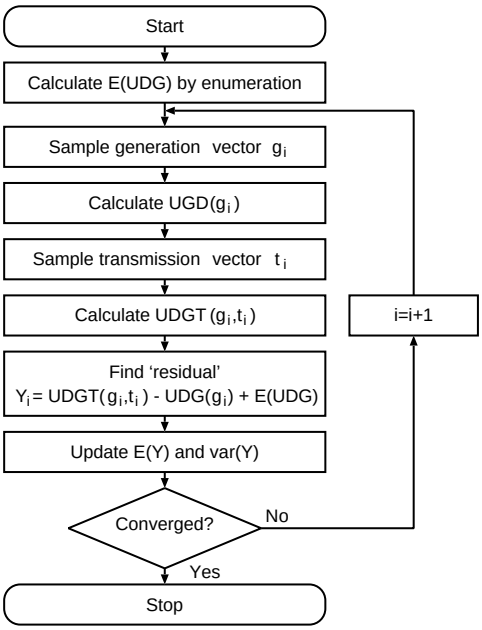

Figure 9.3: Illustration of the use of a control variate

In the context of [79], then, the Monte Carlo simluation proceeds in the manner shown in figure 9.3. First μ_C, here denoted by the expected value of UDG, is found. Then a random generation vector g_i is sampled and the EPNS caused by this vector of generation states (with the transmission system intact), $UDG(g_i)$, is found. This represents the random variate C from equation (9.43). A transmission vector t_i is then sampled and the EPNS calculated for the composite system described by g_i and t_i. Denoted $UDGT(g_i, t_i)$ here, this represents the random variate Y from (9.43). Finally, the residual is found for the trial in question. If this is the ith trial, this may be denoted Y_i and represents the random variate $Y(a)$ from (9.43).

Two test systems were used in [80] to find the relative contributions. The first had 465 buses, 679 branches and 104 generating units. This gave the relative contributions to each index being approximately distributed with between 7 and 12 % caused by generation outages, 25-30% by transmission and 60% by composite effects. The second system with 519 buses, 706 branches and 51 generating units gave contributions to LOLP as follows: generation 24%, transmission 62% and composite effects 14%. For the EPNS index, however, this system gave generation 78%, transmission 16% and composite 6%.

The results show that the speed-up gained by use of a particular control variate is likely to depend significantly on the system being analysed. For the first test system above, for example, it may be useful to adopt two control variates (or "regression variables"): generation and single circuit transmission outages. The final choice of regression variable should be dictated by the expected correlation between the composite system index and the candidate regression variable or by prior numerical tests. It should also be recalled that calculation of the regression variable itself has a computational overhead.

A speed-up of 12 times for the EPNS index and 30 times for the LOLP index is claimed in [80]. Further speed-up is claimed in [76] due to combination of control variates and importance sam-

pling schemes. It is also noted in [80] that the effectiveness of load curtailment cuased simply by generation outages as a control variate will depend on the system being studied and the loading level. For example, if there is plenty of spare generating capacity relative to the load, the impact of the loss of any one generator will be less.In this circumstance, load curtailment due solely to transmission outages may be much more correlated to the composite system index of interest and would make a better choice as control variate.

In [81], it is noted that a in equation (9.43) is very often taken simply as 1 in power system reliability studies. It is proposed instead that greater variance reduction can be achieved by using estimates of a^* based on estimates of cov$[Y, C]$ and var$[C]$ obtained as the Monte Carlo simulation proceeds. It is noted that the estimate for μ need not then be unbiased but that the bias may be expected to decrease as the number of trials increases.

9.3 Application of stratified sampling to power system reliability

It is noted in [52] that in reliability evaluation requiring analysis over a period of time, high load level points in the daily or annual load curves make greater contributions to unreliability indices than low load level points.

The work reported in [82] proposes running a full Monte Carlo study for production costing in one simulation period to estimate the population parameters necessary to define the number of strata, their boundaries and their weights. Other simulation periods are therefore estimated using the appropriately applied stratification. A disadvantage of this method, however, as well the overhead of performing the initial crude Monte Carlo simulation and the possibility the stratification found may less than ideal for other simulation periods, is the need for expert knowledge to interpret the results of the crude simulation. Various papers by Huang and associates, for example [83], therefore propose a more automated procedure along the same lines. The method is, however, only applied to production simulation.

9.3.1 Stratification after sampling

One of the difficulties associated with stratified sampling is the need to be able to choose in advance of each trial which stratum it should fall into. Having chosen the stratum, the trial is then randomly generated within it. This permits control over the number of trials in each stratum which can enable considerable reduction in variance. However, it is not always possible to fix the stratum in advance. Fortunately, as was shown in section 7.5, significant reduction in variance can still be achieved by stratification after sampling when the number of trials in each stratum has been determined by the sampling process itself.

This is what was done in [84]. However, the theoretical probabilities of a trial occurring in the kth stratum must be known in advance, which is possible given the stratification carried out in [84]: the strata are

1. those trials with no outages.
2. those trials with only line outages.
3. those trials with only generator outages.
4. those trials with both line and generator outages.

It may also be envisaged that stratification could be carried out on a basis of the "contingency levels" outlined in section 8.5 i.e.

1. those trials with no outages.
2. those trials with only one outage ("level 1").
3. those trials with two concurrent outages ("level 2").
4. those trials with three concurrent outages ("level 3").
5. those trials with four or more concurrent outages ("level 4 or above").

The above two stratification schemes could be combined to provide subdivisions of line outage, generator outage and combined line and generator outage strata.

It can be seen from [84] that the choice of strata is critical as for "extreme emergencies", an improvement in efficiency of stratified sampling over a crude Monte Carlo estimation of only 1.056 is reported.

9.4 Application of antithetic variates to power system reliability

Use of antithetic variates has been mentioned in [84]. Little description is offered in [84] of its use, and certainly no analytical justification. Suffice to say that the method involves generating one set of random numbers $U'_1, \ldots, U'_n$ in the interval $[0, 1]$ and carrying out one simulation using $U'_1, \ldots, U'_n$ and then another using $1 - U'_1, \ldots, 1 - U'_n$. The results reported show a significantly improved computational efficiency using antithetic variates over the crude Monte Carlo method.

Use of antithetic variates in [85] is claimed to reduce the number of trials needed by more than 3 times. Antithetic variates were also used in [86]

A simple example is also offered in [87].

9.5 Application of correlated sampling to power system reliability

An example is reported in [88] of use of correlated sampling. The variance reduction technique is used to improve the variance of the difference in cost between two operating policies for a fixed number of trials.

Chapter 10

Application of Variance Reduction in Value of Security Assessment

Each variance reduction method discussed in chapter 7 has its own distinctive characteristics. Quite how they fit with a particular problem largely cannot be known in advance so, in this project, a number of variance reduction techniques have been implemented for the purposes of detailed comparison. How they behave also depends on the frequency of underlying events, so how the random events are generated and used in obtaining an estimate is outlined in the first section of this chapter. Detailed discussion of each variance reduction method in turn then follows.

10.1 General estimation model

The Monte Carlo estimation process is used so that, when the sample is large enough, the sample mean cost can be taken as a good estimate of the expected cost.

Individual trials are created by generating random outages for a specified time interval assuming that the system state is known at the start of this interval. The set of possible independent failure events includes

- generator failure;
- line outage;
- double circuit outage;
- busbar outage;
- SVC failure

Note that transient events are not considered. It may also be noted that double cicuit outages and busbar outages entail taking out of service more than one line.

If the failure rate of each type of failure is known and assumed to be constant, the probability $p_i(t)$ of the ith plant item going out of service unexpectedly in a time interval of duration t may be found from [4]

$$p_i(t) = 1 - e^{\lambda_i t} \approx \lambda_i t \tag{10.1}$$

where λ_i is the failure rate of the ith plant item (see chapter 8 for more on failure rates and outage probabilities).

If the system contains m individual basic events, i.e. line faults, busbar faults, double circuit faults, generator faults and SVC faults (deration of generation and sudden changes of load have been neglected in this project to date), the status of the system may be determined for the end of some time period t by taking a vector of random numbers $\mathbf{U} = [U_1, \ldots, U_m]$. Each $U_i, i = 1, \ldots, m$ is taken from a uniform distribution in the interval $[0, 1]$[1]. If a random number U_i is less than $p_i(t)$, the ith plant status Ψ_i (where the plant is regarded as a line, busbar, double circuit, generator or SVC) is deemed to be "out of service"; otherwise, it remains in service. A vector of plant statuses $\mathbf{\Psi}(\mathbf{U}) = [\Psi_1(U_1), \ldots, \Psi_m(U_m)]$ may thus be obtained and may be called the contingency state.

Deration of generation and sudden reduction of demand at a given bulk supply point may be modelled in a similar way with sets of state transition rates describing the number of changes from one state to another per unit time where a state may be considered to be the generator output or the demand. The probability of changing from one state to another can then calculated in the same way as in equation (10.1).

The cost of a given plan at a given time will be a function of the contingency state $\mathbf{\Psi}$ and will comprise the sum of load shedding costs $s(\mathbf{\Psi})$ and generation costs $g(\mathbf{\Psi})$ so that the total cost $f(\mathbf{\Psi})$ is $s(\mathbf{\Psi}) + g(\mathbf{\Psi})$. If $f(\mathbf{\Psi}(\mathbf{U}))$ is represented by $h(\mathbf{U})$, the expected total cost for a given time and plan may then be estimated by θ such that

$$\theta = \frac{1}{n} \sum_{i=1}^{n} h(\mathbf{U}_i) \tag{10.2}$$

where n is the number of trials.

10.2 Antithetic variates

This technique is based on finding two unbiased estimators for the unknown parameter $f(\mathbf{\Psi})$ which have strong negative correlation [50]. If the two estimators are ϕ_1 and ϕ_2, the estimator of $f(\mathbf{\Psi})$ will be $\frac{1}{2}(\phi_1 + \phi_2)$ with

$$\begin{aligned} &\text{var}\left[\frac{1}{2}(\phi_1 + \phi_2)\right] = \\ &\quad \frac{1}{4}\text{var}(\phi_1) + \frac{1}{4}\text{var}(\phi_2) + \frac{1}{2}\text{cov}(\phi_1, \phi_2). \end{aligned} \tag{10.3}$$

[1]Random numbers may be generated by any suitable random number generator such as that included in the NAG library [6] (random number generation algorithms are discussed in [51]).

Clearly, from equation (10.3), if cov(ϕ_1, ϕ_2) is strongly negative, the overall variance can be reduced.

To estimate the expected value of $f(\mathbf{\Psi})$, $E[f(\mathbf{\Psi})]$, a sample of n trials is taken with $\mathbf{\Psi}$ determined by a vector of random numbers $\mathbf{U}$ taken from the uniform distribution. The two estimators may be taken as $\phi_1 = f[\mathbf{\Psi}(\mathbf{U})] = h(\mathbf{U})$ and $\phi_2 = f[\mathbf{\Psi}(1-\mathbf{U})] = h(1-\mathbf{U})$. The antithetic variates estimator θ_a is then

$$\theta_a = \frac{1}{2n} \sum_{i=1}^{n} \left[h(\mathbf{U}_i) + h(1 - \mathbf{U}_i) \right]. \tag{10.4}$$

When sampling, the effect is to have pairs of trials where the first trial has its contingency condition $\mathbf{\Psi}_j$ determined by $\mathbf{U}_j$ and the second one has $\mathbf{\Psi}_{j+1}$ given by $\mathbf{U}_{j+1} = 1 - \mathbf{U}_j$. The two trials are negatively correlated so that if one trial returns a large outcome, the other will be likely to return a small one. It would be hoped that the average cost of the two, $\frac{1}{2}[h(\mathbf{U}_j) + h(\mathbf{U}_{j+1})]$, will be near to the population mean. Thus, if many such pairs of trials are performed, the spread (or variance) of the averages of the pairs will be low.

Since the average of pairs of trials is tallied, it should be clear that the computation time required by equation (10.4) is twice that required by the simple Monte carlo sampling (since two trials are required to provide one independent tallied 'data point') so the estimator of (10.4) is more efficient only when its variance is less than half that of the simple approach.

10.3 Dagger sampling

In the context of power system value of security assessment, the application of dagger sampling requires an appropriate choice of system dagger cycle length. This means that any item of plant that, according to its own outage probability, would have a dagger cycle longer than the system one would have its cycle truncated. Items of plant with shorter cycles would have repeated cycles contained within system cycles. Any cycle that is incomplete when the system cycle is finished will also be truncated. An illustration of this will now be given.

Suppose component a has an outage probability of 0.1 so that the dagger cycle length z_a is 10; component b has an outage probability of 0.33 so that $z_b = 3$; and component c has an outage probability of 0.15 so that $z_c = 6$. Choosing the system dagger cycle length as $z = 6$ would mean that two cycles could be completed for component b for each one for component c. This means that for every single random number which must be generated to determine the status of c in each of a cycle of $z_c = 6$ trials, two random numbers must be generated for b. For component a, however, one random number would determine its status in each of $z_a = 10$ trials. If the system cycle length is 6, a new random number will be generated for component a after 6 trials regardless of the dagger number for a determined before. This will fix the next 10 trials, only the first 6 of which will be modelled according to that (new) random number.

As was seen in section 7.6, where different components in a system have different outage probabilities, a choice must be made as to what the system dagger cycle should be.

If a minimum number of tallied data points needed for a reasonable estimate of the sample variance is taken arbitrarily as 10, the minimum number of trials needed to be simulated is $10z$. With an absolute maximum number of trials set as n_{max} in order to prevent excessive solution times, it can be seen that a limit will be imposed on z such that

$$z \leq \frac{n_{max}}{10}. \tag{10.5}$$

In general, a longer dagger cycle will give a greater reduction in variance although, as z trials are needed for each data point, this may imply a large number of trials. Which dagger method is best then depends on the system being evaluated, the outage probabilities and the relative gain in variance reduction for a particular z.

In order to remove this system dependence, then, two refinements of dagger sampling have been developed. These have been called "automatic dagger sampling" and "alternative dagger sampling".

In the first of these, the system dagger cycle length is initially set equal to the shortest cycle in the system. After some minimum number of tallied points have been collected (say, 10), the sample variance is calculated. Based on the assumption that the calculated sample variance will not change significantly as sampling continues, a judgment is then made as to how many more data points need to be collected for variance of the sample mean to reach the required level. This in turn, since the system dagger cycle length is known, implies a prediction of the required number of simulated trials. If this number is greater than the minimum number of trials that would need to be simulated were the system dagger cycle length to be equal to the greatest component cycle lenght (this is based on an arbitrary choice of a minimum number of tallied data points, such as 10 above), it would appear that a result could be reached more quickly using the longer dagger cycle. The system dagger cycle length is then switched to equal the longest component cycle length in the system and the costs of the trials already simulated are included in the calculation of the next tallied outcome.

Since the outage probability of the rarest type of component failure may so long as to make the switching described above very inadvantageous, another method, "alternative cycling", has been offered. This follows the same principle as the "automatic cycling" but the user of the software chooses which plant type the initial system dagger cycle should be fixed to, and which plant type to switch to where possible. Clearly, the outage probability of the first plant type should be higher than the second[2].

10.4 Stratification by change in MVA

Stratification is carried out after sampling (or, more precisely, as sampling proceeds) on a basis of the change in MVA capacity. This enables line and generator outages to be treated on the same basis (unlike the work in [84]) and also takes into account MVArs as well as MWs. As well as

[2]Further variations of the "automatic" and "alternative" cycles could clearly be thought of where the system dager cycle length steps up in order of component cycle lenghts wherever it is estimated to be useful to do so.

giving some account of voltage problems, this means that SVC outages can also be included. It has the further advantage of being simple to implement as the range of possible changes in MVA is known in advance allowing the stratum boundaries to be determined before sampling commences.

While extensive analysis could be carried out to find the correct stratum weights, this is likely to be almost as computationally demanding as the estimation of the expected cost itself. Thus, it was felt that each stratum weight (which is the probability of an individual trial falling into the stratum) would be better approximated by the ratio of the number of trials in the stratum to the total number of trials. As was noted by Kleijnen [3], this makes each stratum weight stochastic since the number of trials in each stratum is stochastic, and implies that the confidence interval is more approximate. Cochran [55] goes further to suggest that errors in stratum weights can introduce bias to the estimate. This is not felt by the author of this report to be the case in stratification after sampling as the estimate is not a function of the number of trials in each stratum (with other stratification methods, it is).

10.5 "Adaptive" stratification by cost

For this method, it was noted that the ideal stratification variable would be the trial outcome itself. However, this cannot be known *a priori*. Further, it was noted that certain rules exist for optimum determination of stratum boundaries. Of these rules, the best established is the so-called "cum $\sqrt{f}$" rule [55].

The method exploits the fact that as sampling proceeds, more knowledge of the problem is continually gained in the form of the frequency distribution of the outcomes (costs). Given that when the number of trials per stratum is "large", say greater than 20, stratification after sampling is nearly as precise as proportionally stratified sampling, this knowledge is used "adaptively" to determine stratum boundaries whenever the sample is so large that at least 20 trials would be in each stratum.

At the start of sampling, there is just one stratum. By monitoring the cost frequency distribution, the "cum $\sqrt{f}$" rule is used to split the stratum in such a way that the cum $\sqrt{f(\text{cost})}$ interval is the same in each half and each half has at least 20 trials. As each new trial is evaluated, its outcome is added to the appropriate stratum. The frequency distribution of each stratum is continually monitored until it, too, can be split, creating a new stratum until a maximum number of strata has been obtained. This maximum number of strata has been set at 8 since it lends itself well to such "binary" splitting and is close to the figure of 6 suggested by Cochran [55].

Such adaptive stratification allows the most appropriate stratification to be carried out at the earliest possible moment so that a minimum number of trials needed for a given estimate variance can be obtained as nearly as is practicable. However, a stratification carried out early on in the sampling, ideal according to the the knowledge at the time, i.e. the frequency distribution of the sample to date, may be seen to be less than ideal as new knowledge is gained. Hence, the entire sample is restratified periodically, say every 1000 or n_r trials.

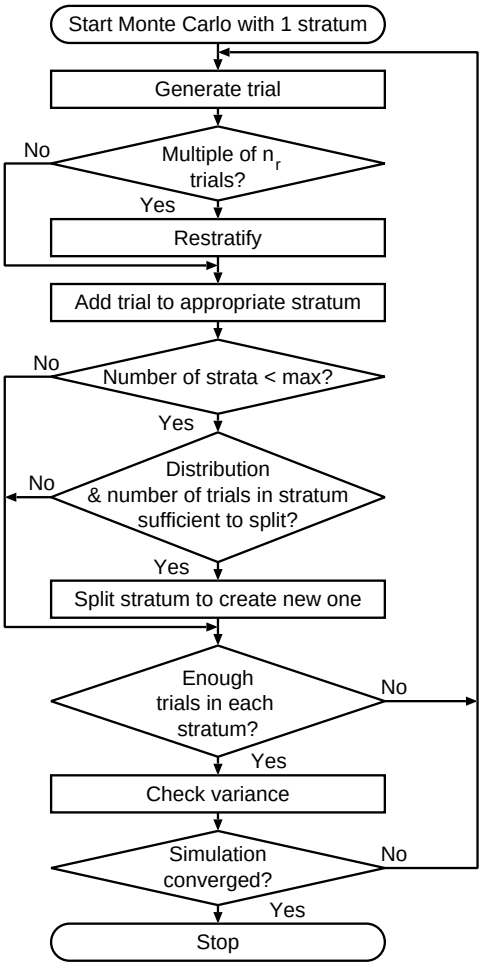

Figure 10.1: Flow chart of "adaptive" stratification

The method is illustrated in figure 10.1.

The main disadvantage of the method is that while the theory regarding the confidence interval requires the use the population variance of each stratum, only the sample variance of each stratum is available. In addition, the theoretical weights of each stratum are not known and are approximated by the ratios of the number of trials in each stratum to the total number of trials. These circumstances make the confidence interval approximate meaning that sampling may stop earlier than necessary leading to imprecision in the estimate. (No bias is introduced as can be seen from the results shown chapter **??**. The mean of the sample means is equal to population mean.)

Experiments have shown that fewer strata than 8 lead to less imprecision (according to [55], 6 is an ideal number of strata) but, inevitably, as the benefit of stratification is reduced by reducing the number of strata, more trials are needed. This was felt to indicate the dependence of the stopping criterion on the estimates of population variances. With fewer strata, the error accruing from the error in each stratum variance is less, and there are more trials in each stratum making each estimate more reliable. Further, with more trials, the central limit theorem which states that when a sample is large, the sampling distribution may be approximated by a normal distribution regardless of the underlying population distribution, holds with greater certainty than with smaller samples.

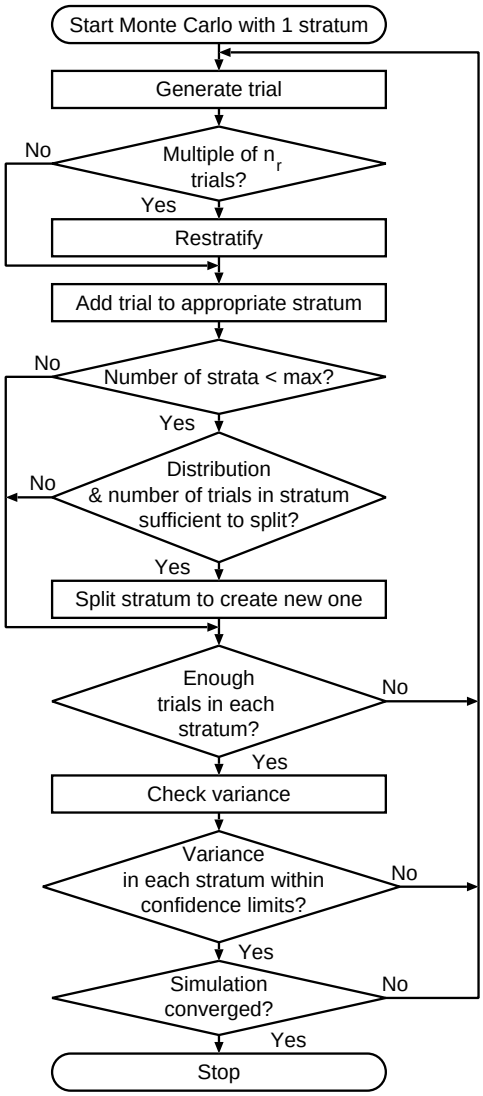

Figure 10.2: Flow chart of "adaptive" stratification with convergence checking for estimates of stratum variances

Theory has shown that confidence intervals can be set not only for means, but also for variances. Since the maximum number of strata must be set in advance of sampling, this has been exploited to ensure more reliable estimates of the stratum variances by forcing sampling to continue while pre-set confidence limits are not met by any one stratum sample variance, so giving less aproximate confidence limits for the mean cost. This approach is illustrated in figure 10.2. What tends to happen, in this case, is that the stratum with largest costs will have the greatest within-stratum variance and hence the least reliable estimate of the stratum population variance. Where allocation of trials in advance is possible, this would be the stratum which would have proportionally more trials. With stratification after sampling, sampling must simply continue until sufficient precision in the estimate of the stratum variance is obtained.

Even in the case where sampling is forced to continue until certain confidence limits on stratum variances have been reached, sampling can stop inappropriately due to error in the variance of the estimate. It is already been noted that the underlying distribution in each stratum is not normal and that this leads to imprecision in the confidence limits. One means tried to voercome this has been to split a stratum only at the minimum turning point in the stratum's frequency

distribution that is nearest to the break-piint suggested by the cum$\sqrt{f}$ rule, though this still has limited effectiveness in the presence of extreme outliers.

It has been noted in the literature that resampling methods such as the jackknife or bootstrap (see section 7.8) can derive better estimates of the variance of the estimate under such conditions. Some investigation of application of both the jackknife and the bootstrap has therefore been carried out. This is described in the next two sections.

10.5.1 Use of the jackknife in stratified sampling

The weights, $P_i, i = 1, \ldots, m$ where m is the number of strata, are kept equal to the empirical weights found in the original sample, i.e. $P_i = n_i/n$ where n_i is the number of trials in the ith stratum and n is the number of trials in the whole sample. (If the weights were allowed to vary with the deletion of each observation, the statistic would reduce to the simple average and the advantage of the stratification would be lost.)

In the conventional approach, the estimates of the variances of the weighted stratum means are summed to give the variance of the final estimate. With the jackknife, the variance of the final estimate is found directly since the statistic evaluated by the jackknife method is the sum of the weighted stratum means where the weights P_i are given by n_i/n.

The estimated cost before the jackknife is

$$\theta = T_n = \sum_{i=1}^{m} P_i \overline{y}_i \tag{10.6}$$

$$= \sum_{i=1}^{m} \frac{n_i}{n} \overline{y}_i \tag{10.7}$$

where

$$\overline{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{i_j}. \tag{10.8}$$

If the deleted observation k falls in the hth stratum, the jackknife statistic, $T_{n-1,k}$, is then

$$T_{n-1,k} = \left(\sum_{i=1, i\neq h}^{m} \frac{n_i}{n} \overline{y}_i \right) + \frac{n_h}{n} \cdot \frac{1}{n_h - 1} \cdot \sum_{j=1, j\neq j_k}^{n_h} y_{h_j} \tag{10.9}$$

$$= \frac{1}{n} \left[\left(\sum_{i=1}^{m} \sum_{j=1}^{n_i} y_{i_j} \right) - \sum_{j=1}^{n_h} y_{h_j} \right] + \frac{n_h}{n} \cdot \frac{1}{n_h - 1} \left[\left(\sum_{j=1}^{n_h} y_{h_j} \right) - y_k \right] \tag{10.10}$$

$$= \frac{1}{n} \left[\left(\sum_{i=1}^{m} \sum_{j=1}^{n_i} y_{i_j} \right) + \frac{1}{n_h - 1} \left(\sum_{j=1}^{n_h} y_{h_j} \right) - \frac{n_h}{n_h - 1} y_k \right] \tag{10.11}$$

where the j_kth observation of the hth stratum corresponds to the kth observation of the sample.

All the $T_{n-1,k}, k = 1, \ldots, n$ are then used to obtain v_{jack}, the jacknife estimate of the variance of the final estimated mean which is then used in conjunction with the classical normal model of the sampling distribution to ascertain confidence limits.

The jackknife procedure also permits the final mean itself to be calculated with some compensation for any error incurred by the ratios n_i/n not being, in general, equal to P_i.

Since the model of convergence based on a jackknife estimate of the mean will generally imply a need to generate more trials than the conventional model, the computational burden of the jackknife procedure is reduced by not carrying out the jackknife to calculate more accurate variances until the confidence limits determined by the conventional model have already been met. This is illustrated in figure 10.3.

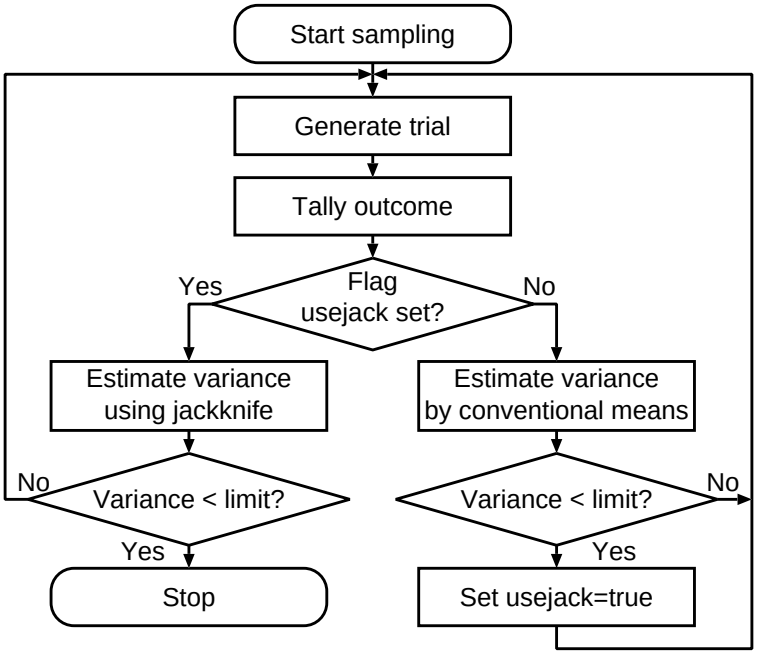

Figure 10.3: Use of the jackknife

10.5.2 Use of the bootstrap in stratified sampling

The bootstrap may be exploited to model the confidence interval better, being based, as it is, on an "empirical distribution".

The decision must be made as to whether to apply the bootstrap to the whole sample, or to treat each stratum as a separate i.i.d. sample. The effect of this difference is that in the latter case the number of trials in each stratum is the same as in the original sample, while in the former it need not be. In common with the application of the jackknife, the weights are taken as the n_i/n from the original sample.

As was noted in section 7.8, a major difficulty with the bootstrap is the large amount of extra computation needed to generate the bootstrap samples. It is also not clear from the literature how many bootstrap samples are needed, though it is asserted in [66, 65] that at least 1000 are needed.

One possible approach may be to perform the bootstrap itself by a Monte Carlo estimation, apply classical confidence limits for the upper and and lower bounds being calculated by the bootstrap to ascertain when to stop the simulation.

The large computational overhead effectively precludes the bootstrap from use as a test for convergence of a Monte Carlo simulation to a given precision and degree of confidence. However, it *can* be used to quantify the confidence interval of a fixed, given sample where the statistic has a non-normal distribution, such the sample mean worked out by stratification. As such, it confirms that considerable reduction in variance may be achieved by the "adaptive stratified sampling" method and indicates that relatively few additional trials would be needed in addition to those suggested by classical confidence intervals for the degree of confidence specified to be met.

10.6 Stratification by outage type

This is similar to the method used in [84], but achieves greater reduction in variance by grouping together more similar outages in the two-way stratification shown in figure 10.4 where the numbers in italics are the stratum numbers. It has the advantage over the other stratification methods described above of having the theoretical stratum weights available. Its disadvantage is that the stratification variables may not be particularly straongly correlated with the outcome of a trial.

Number of line outages \ Number of generator outages	0	1	2 or more
0	*1*	*2*	*3*
1	*4*	*5*	*6*
2 or more	*7*	*8*	*9*

Figure 10.4: Two-way stratification by outage type

10.7 Control variates

Two variations on the method have been implemented thus far in this project and are based on the work in [79, 80]. One uses the expected cost of contingencies comprising only generation outages as the regression (or control) variable (figure 10.5) while the other uses the expected cost of transmission outages (figure 10.6).

The expected cost of the regression variable is first evaluated by enumeration up to a given "out-

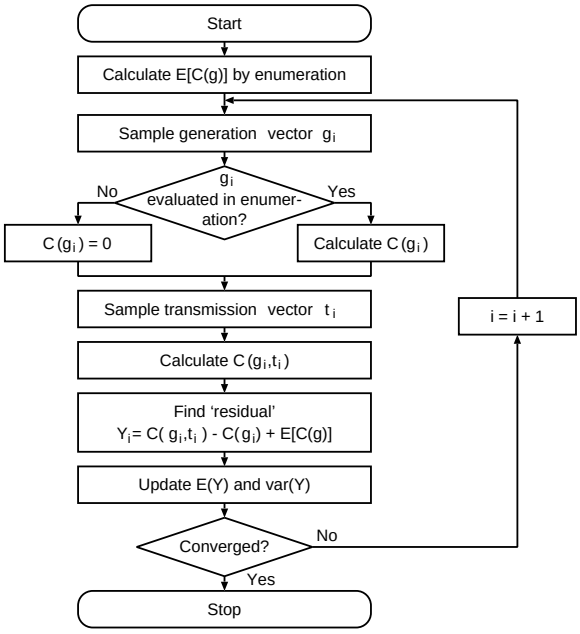

Figure 10.5: The use of the expected cost of generation outages as a control variable

age level"[3]. Monte Carlo simulation then proceeds. If a randomly generated vector of regression variable plant statuses was evaluated in the enumeration, its cost is either calculated or looked up from previously stored values, i.e. for enumeration up to level 3, any vector comprising 1, 2 or 3 outages will have its cost found. Otherwise, since the scenario was not covered in the enumeration, it must be taken as having cost 0. Random statuses are then taken for all the other plant and cost of all the plant outages is found. The "residual" is then found and tallied.

The average residual gives an unbiased estimate of the expected outage cost for all plant outages. The variance reduction can be seen through the tallied values being the average of the regression variable's expected value plus the costs incurred by the other variables. Thus, since the element of variability due to the regression variable (in the two cases studied, either generation outages or transmission outages) has been removed, the overall spread of tallied costs should be less.

10.8 Correlated sampling

This is a simple method in which the variance of the difference in a statistic calculated from two similar distributions can be reduced by performing Monte Carlo sampling of each using the same set of random numbers [50]. Although it has few theoretical difficulties, its implementation in this project has required considerable re-writing of the software to enable the necessary comparison between different initial scenarios.

[3]The outage level is the number of concurrent independent outages. Evaluation of generation outages up to level of 3, for example, would entail calculating the cost times the probability of every possible single generator outage and all combinations of pairs of generator outages and of 3 concurrent outages. The expected value is then the sum of all these products. The accuracy of the enumeration depends on the assumption that the costs of scenarios above the given level are insignificant.

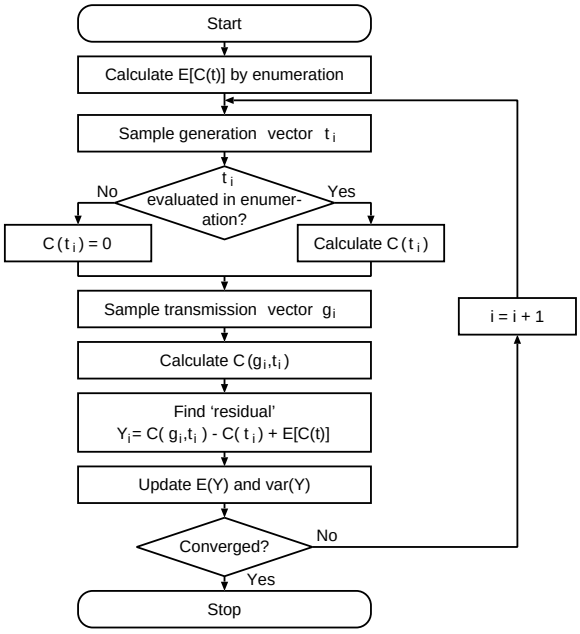

Figure 10.6: The use of the expected cost of transmission outages as a control variable

The action of correlated sampling in the context of the assessment of the value of security is illustrated in figure 10.7. In the figure, two pairs of power system scenarios are shown. The scenarios on the left concern a part of a system where the only generator is generating 1000 MW. In those on the right, the same generator is generating 1270 MW. Let it be supposed that the cheapest scenario is to be found - that with the illustrated generator generating 1270 MWs, or that where it is generating 1000 MWs and some other generator is generating an extra 270 MWs. All other plant statuses may be assumed to be the same in each case.

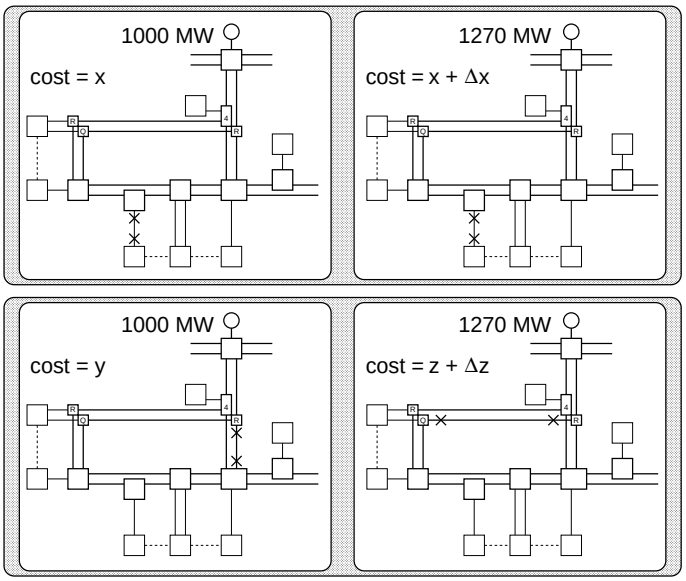

Figure 10.7: Illustration of correlated sampling in comparing scenarios

In estimating the value of the security afforded by the different levels of generation at the generator in question, random outages may applied to the system and the resulting costs and cost differences tallied. In the pair of situations at the top of the figure, the same outage has been

applied to each case. In the bottom pair, there are different outages. Another way of expressing the aim of the simulation is to discover the effect of the change in generation on the expected outage cost. It should be clear that the cost difference Δx between the top pair of situations is due solely to the difference in generation patterns, while the cost difference between the bottom pair is $(z + \Delta z) - y$ and is due both to the different outage in each case and the difference in generation.

The effect of correlated sampling is to tally the differences Δx for similar outage scenarios. Variations due to factors other than the difference in initial condition are removed and the variance of the tallied cost differences is reduced so that a decision on which of two scenarios is cheaper can be made quickly.

The software has been written so that correlated sampling can be exploited to make comparisons between scenarios in one of two ways:

1. by forcing sampling to continue until a degree of confidence specified by the user has been obtained for the sign of the difference in cost between two scenarios.

2. by forcing sampling to continue until a given degree of confidence in the sign of the difference has been obtained, *and* an estimate of each individual scenario's cost has been obtained within a pre-specified confidence interval.

The necessary degree of confidence in the sign of the cost difference is detected quite simply. With knowledge of the sample variance (of the cost difference), the sample size and an assumption that the sampling distribution is normal, the standard expressions relating degree of confidence and confidence interval can be used:

$$L = t_{\alpha/2}\frac{s}{\sqrt{n}} \tag{10.12}$$

$$\gamma = 1 - \alpha \tag{10.13}$$

where L is the confidence interval, s is the sample variance, n is the sample size and γ is the desired degree of confidence. $t_{\alpha/2}$ is found from the t-distribution with $n-1$ degrees of freedom. (Note that L is always positive.)

If $X_1, X_2, \ldots, X_n$ are the tallied cost differences, the expected cost difference can be estimated by

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{10.14}$$

Since the meaning of the confidence interval is that the true population mean lies within $\pm L$ of the estimated mean with $100\gamma\%$ confidence, only when $|\overline{X}| \geq L$ can there be $100\gamma\%$ confidence in the sign of the actual cost difference. This is because, if $|\overline{X}| < L$ and $\overline{X}$ is, say, negative, the confidence interval describing the possible range of the true mean (with $100\gamma\%$ confidence) is $\overline{X} - L$, which is clearly negative, to $\overline{X} + L$, which is positive when $|\overline{X}| < L$, meaning that the true population mean *could* be positive rather than negative implying a different answer to the question "which scenario is cheaper?" This is illustrated in figure 10.8. When the sample is

still quite small, the sample mean $\overline{X}_1$ is found and the confidence interval L_1 calculated. It can be seen that $L_1 > |\overline{X}_1|$. Sampling is then continued and at some later point the sample mean and confidence interval are recalculated. These are denoted $\overline{X}_2$ and L_2. Now, $L_2 < |\overline{X}_2|$ so sufficient precision has been obtained and sampling can stop.

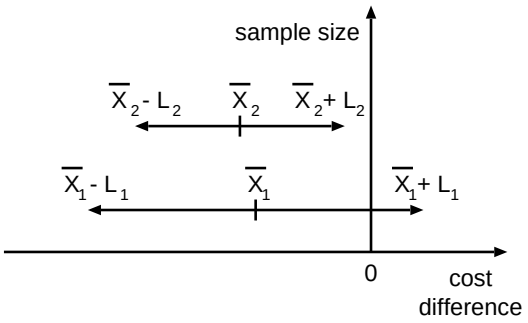

Figure 10.8: Illustration of confidence limits in correlated sampling

In general, there will be a need to compare more than 2 scenarios. The method for doing this adopted here resembles a "knock-out" competition. The first two scenarios are compared, generating one trial set of random outages which is applied to both scenarios, simulating the outcome of those outages on each and tallying the individual costs and the difference in cost before generating another set of outages until sufficient precision has been achieved. The outage set associated with each trial is recorded. The cheapest scenario is then regarded as the "winner" and is compared to the next scenario. The same outages from the first "contest" are applied to the new scenario and the costs of those outages on that scenario tallied. The winner from the previous contest does not need to be simulated again since all the costs have already been tallied, which enables the cost differences between the new scenario and the previous "winner" to be tallied. The recorded outages are applied until sufficient precision has been obtained. If, however, the precision is not yet sufficient when the record of outage sets is exhausted, new sets of random outages must be generated and simulated on both the new scenario and the previous "winner". A winner will then be found for this "contest" (which may be the same as the previous winner), and another new scenario simulated for the same common outage sets. In this way, any number of scenarios may be compared in order to find the single cheapest one. The process is illustrated in figure 10.9.

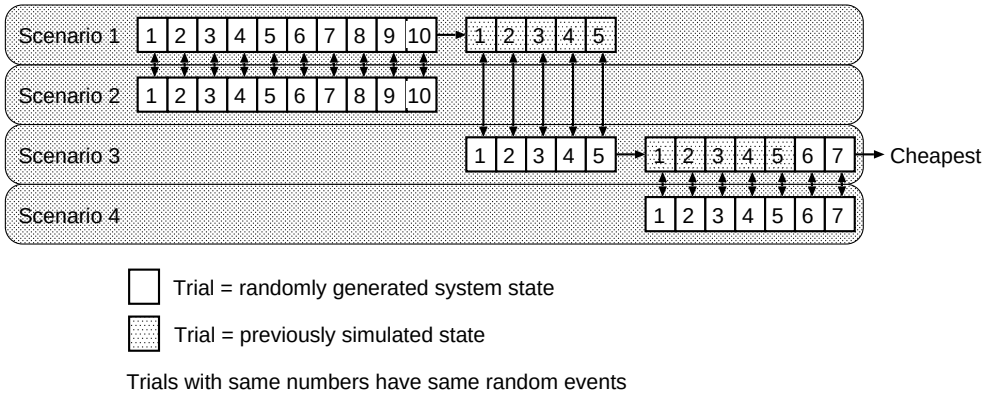

Figure 10.9: Comparison of scenarios in correlated sampling

10.9 Importance sampling

Work on importance sampling has focused on biasing not only individual plant outages but also the expected number of outages per trial of each plant type.

It may further be found after a simluation that the average number of outages per trial for different plant types for trials with outcomes close to the mean is something else. Thus, in order that there be more trials with costs around the expected cost, in further simulations of similar scenarios, the outage probabilties may be biased to achieve the expected number of outages for outcomes close to the mean.

A further investigation has been based on the observation that over-biasing may cause there to be too many trials in a simulation with very high cost. In particular, it was previously felt that individual plant items which are associated with high cost should be weighted more highly than others. This has the effect of moving the cost frequency distribution too far to the right and may result in a *higher* rather than lower variance. Instead, individual plant items whose outages are associated with costs that are found afterwards to be close to the mean (or expected) cost are weighted more heavily in subsequent simulations of similar scenarios.

Early findings indicate that these approaches to importance sampling do not significantly reduce the number of trials needed for a given precision of estimate. While importance sampling may still, in theory, yield large reductions in variance, on the basis of the results found so far, it is felt by the author that for significant rewards, too much time would need to be spent in the course of this project to warrant further investigation. It should be noted, however, that an opportunity may present itself in the course of some other project to return to the subject, and that such an opportunity may well be worth taking.

Appendix A

Static Analysis in the Value of Security Assessor

Two basic tools included in the "Value of Security" assessor package are described here for static analysis of power systems (i.e. concerned with the algebraic rather than differential equations). The first is the load flow tool which solves the non-linear equations concerned with complex nodal voltages, power injections and power flows. The second is sensitivity analysis in which the algebraic load flow equations are linearised to show changes in dependent variables for changes in control variables. The equations are introduced in the first section.

A.1 Basics

A power system can be described by a set of differential equations and a set of algebraic equations to form a differential-algebraic system which can be written

$$\dot{x} = F(x, u, p) \tag{A.1}$$
$$0 = G(x, u, p) \tag{A.2}$$

where x is the vector of power system state (dependent) variables, u is the vector of independent variables and p is a vector of system parameters. Alternatively, the system can be described in terms of differential equations, algebraic equations and difference equations describing discrete time events such as changes to on-load tap changing transformer tap ratios and changes in switchable shunt susceptances. In this case, the system can be written

$$\dot{x} = F(x, u, z, p) \tag{A.3}$$
$$0 = G(x, u, z, p) \tag{A.4}$$
$$z(k+1) = H(x, u, z(k), p) \tag{A.5}$$

where z is the vector of variables which can be changed in discrete steps at time k.

In this study, only static controls and their algebraic constraints are being considered as even when constraints are originally defined according to dynamic criteria, standard practice in op-

eration of large power systems determines that the constraints are expressed in terms of static quantities. Hence, only equation A.2 is of interest.

The function G is the sum of the set of power injection equations. These equations are functions of the complex nodal voltages and complex branch currents. The total complex power S should be zero for each node. For a node i it is

$$S_i = V_i I_i^* \tag{A.6}$$

where V_i is the complex voltage at node i and I_i is the current at node i. For every branch from node i to a node j, $j = 1 \ldots N, j \neq i$,

$$S_i = P + \mathrm{j}Q = \sum_{j=1}^{N, j \neq i} V_i I_{ij}^* \tag{A.7}$$

where P is the active power and Q is the reactive power. If the complex admittance of the branch connecting nodes i and j is y_{ij},

$$S_i = \sum_{j=1}^{N, j \neq i} V_i \left(y_{ij}(V_i - V_j)\right)^* \tag{A.8}$$

A.2 Load Flow

Load flow routines have been used to solve the algebraic power flow equations in analysis of static power system conditions for many years and have taken many different forms [89]. In the late 1970s and early 1980s, fast decoupled routines became popular [90], although trends now again favour fully coupled implementations for analysis of systems run under a wider variety of conditions.

The load flow routine implemented as part of this project is based on a standard formulation such as that described in [91]. It has been further developed to provide an interactive shell through which the user can easily change parameters such as transformer tap ratios, generator P and V set points and Q generation limits, and line switched in/out status. Network sensitivity matrices can be calculated from each new system condition (see section A.3).

The operating state of an interconnected power system can be described in terms of four sets of quantities which relate to each node of the system. These are the nodal voltage magnitude V, voltage angle θ, active power injection P and reactive power injection Q. The load flow allows two of these quantities to be found for each node once the others have been defined.

Three different bus conditions are defined depending on which two of the four parameters are pre-defined. These are

PV **or voltage-controlled (regulated) bus.** Such a bus has the facility to maintain a fixed voltage magnitude V which is specified and as such will be a variable source of reactive power. The active power injected is also specified. In practice, this represents a generator or compensator bus, and maximum or minimum limits on the reactive power injected Q_{max} and Q_{min} may be set.

PQ **or unregulated bus.** The total injection P+jQ is specified corresponding to the load at a load bus.

Slack or swing bus. This bus defines the voltage angle reference and also has the voltage magnitude defined. P and Q are allowed to vary since all powers cannot be defined in advance as system losses are unknown. The slack bus can be regarded as analogous to a generator responsible for maintenance of system frequency.

The set of simultaneous power equations which defines the system's state is non-linear. It is therefore solved via a set of successive linear approximations based on first order Taylor expansions of the power equations. The most common method is that of Newton which is what has been used in this study.

The solution routine requires the assignment of some initial estimate to all the busbar voltage magnitudes and angles (the slack bus angle is typically assigned to 0.0 degrees and the calculation of the initial real and reactive power mismatches. Should any of these be above the set tolerance, the Jacobian is formed and solved for updates of voltages and angles upon which new estimates of the power mismatches are obtained. While convergence is not obtained, the Jacobian is again formed and new updates of V and θ are found.

The power equation for bus k is

$$S_k = P_k + jQ_k = V_k I_k^* \tag{A.9}$$

$$= V_k \sum_{m \epsilon k} y_{km}^* V_m^* \tag{A.10}$$

where V_k is the voltage at bus k and

$$E_k = V_k^{re} + jV_k^{im} = \mathrm{V_k}\angle\theta_k, \tag{A.11}$$

y_{km} is the admittance between buses k and m and I_k is the current injected at bus k. In polar co-ordinates, P_k and Q_k are

$$P_k = \sum_{m \in k} \mathrm{V_k V_m} \left\{G_{km} \cos(\theta_k - \theta_m) + B_{km} \sin(\theta_k - \theta_m)\right\} \tag{A.12}$$

$$Q_k = \sum_{m \in k} \mathrm{V_k V_m} \left\{G_{km} \sin(\theta_k - \theta_m) - B_{km} \cos(\theta_k - \theta_m)\right\} \tag{A.13}$$

where

$$y_{km} = G_{km} + jB_{km}. \tag{A.14}$$

Linear relationships for small changes in V and θ are found so that for a PQ bus

$$\Delta P_k = \sum_{m \in k} \frac{\partial P_k}{\partial \theta_m} \Delta\theta_m + \sum_{m \in k} \frac{\partial P_k}{\partial \mathrm{V_m}} \Delta \mathrm{V_m} \tag{A.15}$$

$$\Delta Q_k = \sum_{m \in k} \frac{\partial Q_k}{\partial \theta_m} \Delta\theta_m + \sum_{m \in k} \frac{\partial Q_k}{\partial \mathrm{V_m}} \Delta \mathrm{V_m} \tag{A.16}$$

For a PV bus, only equation A.15 is used since Q_k is not specified, and there are no equations for the slack bus.

With the voltages in rectangular form, the partial derivatives are

$$H_{km} = \frac{\partial P_k}{\partial \theta_m} = G_{km}(V_k^{im}V_m^{re} - V_k^{re}V_m^{im}) - B_{km}(V_k^{re}V_m^{im} + V_k^{im}V_m^{re}) \quad \text{(A.17)}$$

$$N_{km} = \mathrm{V_m}\frac{\partial P_k}{\partial \mathrm{V_m}} = G_{km}(V_k^{re}V_m^{im} - V_k^{im}V_m^{re}) - B_{km}(V_k^{im}V_m^{re} + V_k^{re}V_m^{im}) \quad \text{(A.18)}$$

$$J_{km} = \frac{\partial Q_k}{\partial \theta_m} = -N_{km} \quad \text{(A.19)}$$

$$L_{km} = \mathrm{V_m}\frac{\partial Q_k}{\partial \mathrm{V_m}} = H_{km} \quad \text{(A.20)}$$

$$H_{kk} = \frac{\partial P_k}{\partial \theta_k} = -Q_k - B_{kk}\mathrm{V_k}^2 \quad \text{(A.21)}$$

$$N_{kk} = \mathrm{V_k}\frac{\partial P_k}{\partial \mathrm{V_k}} = P_k + G_{kk}\mathrm{V_k}^2 \quad \text{(A.22)}$$

$$J_{kk} = \frac{\partial Q_k}{\partial \theta_k} = P_k - G_{kk}\mathrm{V_k}^2 \quad \text{(A.23)}$$

$$L_{kk} = \mathrm{V_k}\frac{\partial Q_k}{\partial V_k} = Q_k - B_{kk}\mathrm{V_k}^2 \quad \text{(A.24)}$$

These are assembled in a matrix equation of the form

$$\begin{bmatrix} \Delta P^{p-1} \\ \Delta Q^{p-1} \end{bmatrix} = \begin{bmatrix} H^{p-1} & N^{p-1} \\ J^{p-1} & L^{p-1} \end{bmatrix} \begin{bmatrix} \Delta\theta^p \\ \frac{\Delta \mathrm{V}^p}{\mathrm{V}^{p-1}} \end{bmatrix} \quad \text{(A.25)}$$

which is solved for $\Delta\theta^p$ and $\Delta\mathrm{V}^p$ at the pth iteration where ΔP^{p-1} are the P mismatches for all PQ and PV busbars, ΔQ^{p-1} are the Q mismatches for the PQ busbars, $\Delta\theta^p$ are the θ corrections for all PQ and PV busbars and $\Delta\mathrm{V}^p$ are the V corrections for all PQ buses.

A.2.1 PV bus reactive limits

Maximum and minimum limits on reactive generation are often defined for PV buses. Once the maximum mismatch is less than some pre-defined quantity (meaning that the solution is being approached), the Q injections at all the PV buses are compared with their limits. When a limit is exceeded, Q at that bus is fixed to the limiting value and the bus type is switched to PQ with V to be found. The effect is then for V to 'float' to meet the required system conditions.

If, at subsequent iterations, the bus voltage magnitude V is higher than its target value and Q is at its maximum value (meaning that too many MVArs are being generated when there is freedom to generate fewer) or V is below its target and Q is at its minimum value (meaning that too many MVArs are being absorbed when there is freedom to absorb fewer), the type is switched back to PV with V set to its target value.

Only enforcing the Q limits once the biggest mismatch becomes less than a certain value prevents early divergence of the iterative process due to excessively large changes.

A.2.2 Distributed slack bus

The slack bus is conventionally used to provide a reference voltage angle and to take up the system losses which cannot be predicted prior to running the study. It could also be thought of compensating any difference between load and generation and in this way is a like a generator with governor control. It is often chosen as the "largest" PV node on the system.

In practice, there will not be only one generator operating with governor control. The effect of modelling more can be represented by a "distributed slack" where one bus is retained as providing a reference, but adjustments to active power generation due to differences between the total load (including losses) and total generation are divided among a number of PV buses in proportion to the maximum generation at the bus since this can be taken to be proportional to the inertia of the machine(s).

Once the maxmimum mismatch becomes less than some pre-defined threshold, the difference between the reference bus active generation and its scheduled generation is distributed among all the "free governor" buses changing the scheduled P setting at the start of the next iteration. Any subsequent changes in reference bus active generation from one iteration to the next are similarly distributed. While this might be expected to increase the number of iterations required, it is found that only one extra iteration is needed. ("Free governor buses" are indicated in the input study file and the bus type is set accordingly. See section A.2.6 for details of bus types.)

A.2.3 Modelling of SVCs

An SVC (static VAr compensator—so-called as it provides inductive or capacitve VArs but, unlike a synchronous compensator, does not rotate) can be modelled like a PV bus with a given voltage and limits set for reactive power. In such a model, the $V - I$ characteristic would be like that in figure A.1. However, in reality, some droop is present in the SVC characteristic, essentially for reasons of stability. This is to ensure that the switching characteristic of the SVC controller is such that a stable voltage is found. This can be understood by reference to figure A.2. Curve A represents the voltage at the SVC bus as the reactive load changes and can be thought of as the characteristic of the system. If the SVC characteristic were to be as flat as that in A.1, it can be seen that the cross-over between the "system curve" and the SVC curve is less well-defined. The result is that for any given V_{ref} to the SVC, some error between V_{ref} and the final V_{SVC} will be present.

The operating area of the SVC will be given by the design of its components and is shown in A.2 by the shaded area. By varying a gain in the controller, the slope of the characteristic can be varied. V_{ref} can also be varied and with a slope given, one line is defined for operation. This line can be defined in terms of points p_1 and p_2 with pairs of values $(Q_{min}, \mathrm{V}_{min})$ and $(Q_{max}, \mathrm{V}_{max})$.

In the load flow implementation, the SVC characteristic is defined by $(Q_{min}, \mathrm{V}_{min})$ and $(Q_{max}, \mathrm{V}_{max})$ in the input study file since it is assumed that V_{ref} will not be changed by the operator. The SVC node is then modelled as a PQ bus as follows.

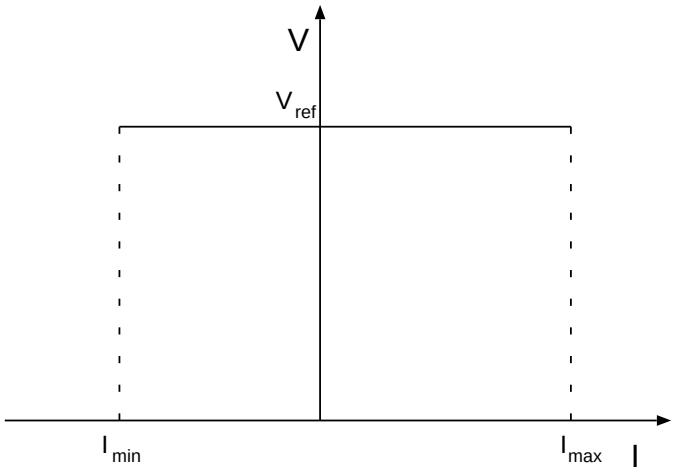

Figure A.1: Terminal voltage V against current magnitude I for PV bus

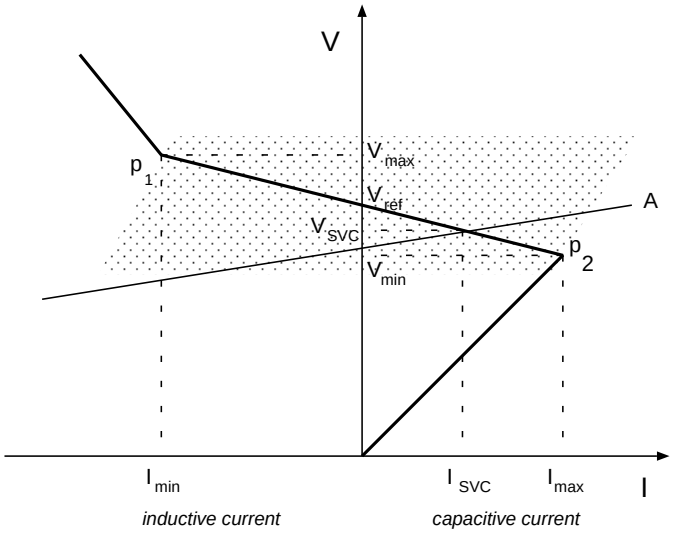

Figure A.2: Terminal voltage V against current magnitude I for SVC

For $\mathrm{V}_{SVC} < \mathrm{V}_{min}$,

$$Q_{SVC} = \frac{\mathrm{V}_{SVC}^2}{\mathrm{V}_{min}^2} Q_{max}. \tag{A.26}$$

For $\mathrm{V}_{min} < \mathrm{V}_{SVC} < \mathrm{V}_{max}$,

$$Q_{SVC} = \frac{\mathrm{V}_{SVC}}{\mathrm{V}_{min} - \mathrm{V}_{max}} \left[\frac{Q_{max}}{\mathrm{V}_{min}} (\mathrm{V}_{SVC} - \mathrm{V}_{max}) + \frac{Q_{min}}{\mathrm{V}_{max}} (\mathrm{V}_{min} - \mathrm{V}_{SVC}) \right]. \tag{A.27}$$

For $\mathrm{V}_{SVC} > \mathrm{V}_{max}$,

$$Q_{SVC} = \frac{\mathrm{V}_{SVC}^2}{\mathrm{V}_{max}^2} Q_{min} \tag{A.28}$$

where

Q_{SVC} = reactive power generation of the SVC;
Q_{max} = maximum reactive power generation of the SVC;
Q_{min} = minimum reactive power generation of the SVC;
V_{SVC} = terminal voltage of the SVC;
V_{max} = maximum control voltage of the SVC;
V_{min} = minimum control voltage of the SVC.

The following L_{kk} terms in the Jacobian corresponding to the SVC bus are modified with extra derivative terms added for the changes due to the SVC.

For $\mathrm{V}_{SVC} < \mathrm{V}_{min}$,

$$\mathrm{V}_{SVC}\frac{\partial Q_{SVC}}{\partial \mathrm{V}_{SVC}} = \frac{2\mathrm{V}_{SVC}^2}{\mathrm{V}_{min}^2}Q_{max}. \tag{A.29}$$

For $\mathrm{V}_{min} < \mathrm{V}_{SVC} < \mathrm{V}_{max}$,

$$\mathrm{V}_{SVC}\frac{\partial Q_{SVC}}{\partial \mathrm{V}_{SVC}} = \frac{\mathrm{V}_{SVC}}{\mathrm{V}_{min}-\mathrm{V}_{max}}\left[\frac{Q_{max}}{\mathrm{V}_{min}}(2\mathrm{V}_{SVC}-\mathrm{V}_{max})+\frac{Q_{min}}{\mathrm{V}_{max}}(\mathrm{V}_{min}-2\mathrm{V}_{SVC})\right] \tag{A.30}$$

For $\mathrm{V}_{SVC} > \mathrm{V}_{max}$

$$\mathrm{V}_{SVC}\frac{\partial Q_{SVC}}{\partial \mathrm{V}_{SVC}} = \frac{2\mathrm{V}_{SVC}^2}{\mathrm{V}_{max}^2}Q_{min}. \tag{A.31}$$

Further discussion of the SVC model can be found in [92].

A.2.4 Treatment of transformers

Figure A.3 shows the basic equivalent circuit of a transformer with the currents I_{km} and I_{mk}, voltages V_k, V_m and V', admittance y_{km} and turns ratio a shown.

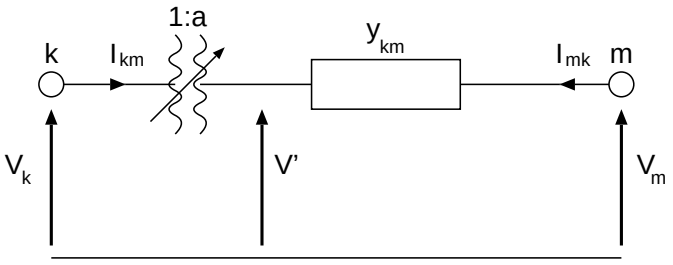

Figure A.3: Transformer equivalent circuit

With $a = 1$,

$$I_{km} = y_{km}(V_k - V_m) \tag{A.32}$$
$$I_{mk} = y_{km}(V_m - V_k) \tag{A.33}$$

i.e. $I_{km} = -I_{mk}$.

With $a \neq 1$,

$$V' = aV_k \tag{A.34}$$
$$I_{mk} = y_{km}(V_m - V') \tag{A.35}$$
$$I_{km} = -aI_{mk} \tag{A.36}$$

Eliminating V' from (A.35) and replacing I_{mk} in (A.36),

$$I_{km} = a^2 y_{km} V_k - a y_{km} V_m \tag{A.37}$$
$$I_{mk} = -a y_{km} V_k + y_{km} V_m \tag{A.38}$$

or in admittance matrix form,

$$\begin{bmatrix} I_{km} \\ I_{mk} \end{bmatrix} = \begin{bmatrix} a^2 y_{km} & -a y_{km} \\ -a y_{km} & y_{km} \end{bmatrix} \begin{bmatrix} V_k \\ V_m \end{bmatrix} \tag{A.39}$$

This corresponds to equivalent circuit shown in figure A.4.

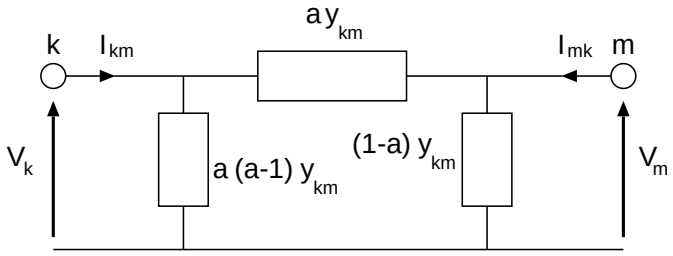

Figure A.4: Transformer equivalent circuit with winding removed

On-load tap changing transformers

On-load tap changing transformers (OLTCs) automatically change their tap ratios in discrete steps while under load. The changes take place in order to try to attain a target voltage on either the HV or LV side. If the voltage error on the side of the transformer that is scheduled is $\mathrm{V}^0 - \mathrm{V}$ at time k, a function $f(\Delta \mathrm{V})$ [93] can be used to determine the new tap ratio a_{k+1} where

$$a_{k+1} = a_k - f(\Delta \mathrm{V})\Delta a \tag{A.40}$$
$$f(\Delta \mathrm{V}) = \begin{cases} 1 & \text{if } \mathrm{V}^0 - \mathrm{V} > \varepsilon \\ 0 & \text{if } |\mathrm{V}^0 - \mathrm{V}| \leq \varepsilon \\ -1 & \text{if } \mathrm{V}^0 - \mathrm{V} < -\varepsilon \end{cases} \tag{A.41}$$

and Δa is the maximum tap step size, V^0 is the scheduled voltage magnitude, and ε is a dead-band.

Such a function $f(\Delta \mathrm{V})$ can be implemented in a load-flow routine at the end of each iteration so that a new tap ratio can be found and the solution adjusted. However, the solution can tend to zig-zag around due to the discrete nature of the change. Instead, either adjustment of tap ratios can be delayed until beyond the first iteration of the load flow, or a continuous model should be implemented in which V is kept constant and is replaced as the state variable by a, and the

partial derivatives corresponding to the tap-controlled bus are taken with respect to a. When the bus being controlled is bus k i.e. a replaces V_k and the transformer is as in figure A.3, the relevant partial derivatives are then [94]

$$\frac{\partial P_k}{\partial a}a = \quad N_{kk} = \quad -2G_{km}{\mathrm{V_k}}^2a^2 + (e_m V_k^{re} + f_m V_k^{im})a \tag{A.42}$$

$$\frac{\partial Q_k}{\partial a}a = \quad L_{kk} = \quad 2B_{km}{\mathrm{V_k}}^2a^2 + (e_m V_k^{im} - f_m V_k^{re})a \tag{A.43}$$

$$\frac{\partial P_m}{\partial a}a = \quad N_{mk} = \quad (e_k V_m^{re} + f_k V_m^{im})a \tag{A.44}$$

$$\frac{\partial P_m}{\partial a}a = \quad L_{mk} = \quad (e_k V_m^{im} - f_k V_m^{re})a \tag{A.45}$$

where

$$y_{km} = G_{km} + jB_{km} \tag{A.46}$$

$$V_k = V_k^{re} + jV_k^{im} \tag{A.47}$$

$$V_m = V_m^{re} + jV_m^{im} \tag{A.48}$$

$$\mathrm{V}_k = |V_k| \tag{A.49}$$

$$e_k + jf_k = V_k y_{km}. \tag{A.50}$$

In order to avoid excessive interaction with other variables, taps are regarded as fixed until the maximum mismatch falls below some given level. The tap-controlled voltage is then set to the target voltage and the state to solved for becomes the tap ratio. In addition, oscillation of the load flow solution is prevented by updating the tap ratio only when the change exceeds the minimum tap step.

If the tap ratio a reaches an upper or lower limit, a is replaced by V_k as the state variable and updates of V_k are obtained from subsequent iterations of the load flow with a fixed.

A.2.5 Treatment of quadrature boosters

The basic equivalent circuit for a quadrature booster or phase-shifting transformer is similar to that shown in figure A.3, but the turns ratio a is replaced by the complex turns ratio α, as in figure A.5, where

$$\alpha = a + jb \tag{A.51}$$

and

$$V' = \alpha V_k \tag{A.52}$$

Since power loss in the ideal transformer is negligible,

$$V_k I_k^* = -V' I'^* \tag{A.53}$$

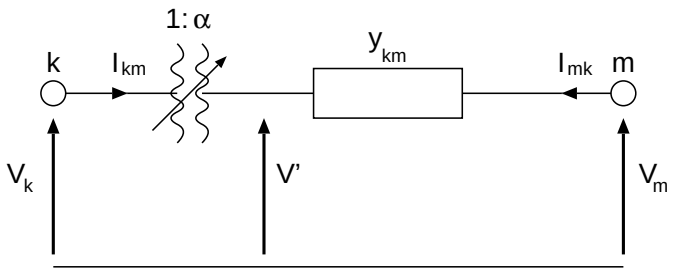

Figure A.5: Quadrature booster equivalent circuit

From (A.52) and (A.53),

$$I_k^* = -(a + jb)I'^* \tag{A.54}$$

or

$$I_k = -I'(a - jb) = -I'\alpha^* \tag{A.55}$$

Thus it can be seen that there are two different turns ratios, one for voltage, $\alpha_v = a + jb$, and one for current, $\alpha_i = a - jb$.

Solving for terminal currents,

$$\begin{aligned} I_k = \quad \alpha_i I' = \quad & \alpha_i(V' - V_m)y_{km} = \alpha_i(V' - V_m)y_{km} && \text{(A.56)} \\ = \quad & \alpha_i(\alpha_v V_k - V_m)y_{km} = \alpha_i\alpha_v y_{km}V_k - \alpha_i y_{km}V_m \\ -I_m = \quad I' = \quad & \alpha_v y_{km}V_k - y_{km}V_m && \text{(A.57)} \end{aligned}$$

or in admittance matrix form,

$$\begin{bmatrix} I_k \\ I_m \end{bmatrix} = \begin{bmatrix} \alpha_v\alpha_i y_{km} & -\alpha_i y_{km} \\ -\alpha_v y_{km} & y_{km} \end{bmatrix} \begin{bmatrix} V_k \\ V_m \end{bmatrix} \tag{A.58}$$

It can be seen that, unlike for a standard transformer whose admittance matrix is described by (A.39), the admittance matrix for a quadrature booster is non-symmetric and the equivalent circuit is not readily available. The expression for the active power entering the booster, however, can be easily found. Considering the power at node k, $S_k = V_k I_k^*$ or

$$S_k = V_k(V_k Y_{kk} + V_m Y_{km})^* \tag{A.59}$$

If the complex tap ratio α is written as $\alpha = a + jb = t\angle\psi$ and the voltages V_k and V_m as $V_k = \mathrm{V_k}\angle\theta_k, V_m = \mathrm{V_m}\angle\theta_m$, then from equations A.58 and A.59

$$\begin{aligned} S_k \quad &= \quad V_k\left(t^2 y_{km}V_k - t y_{km}V_m\angle - \psi\right)^* && \text{(A.60)} \\ &= \quad y_{km}^*\left(t^2 \mathrm{V_k}^2 - t\mathrm{V_k}\mathrm{V_m}\angle(\theta_k - \theta_m + \psi)\right) && \text{(A.61)} \end{aligned}$$

With $y_{km} = g + jb$, the active power P_k is the real part of S_k and is

$$P_k = g\left(t^2 \mathrm{V_k}^2 - t\mathrm{V_k}\mathrm{V_m}\cos\phi\right) - tb\mathrm{V_k}\mathrm{V_m}\sin\phi \tag{A.62}$$

where $\phi = \theta_k - \theta_m + \psi$.

A.2.6 Summary of sub-categories of busbar types

In the code, different busbar types are denoted by an integer in which bits are set denoting properties or combinations of properties. All the busbar types are one of

- slack (reference) bus;
- PV bus;
- PQ bus.

Some buses have on-load tap-changing transformers (OLTCs). While most of these will be PQ buses, some are PV buses. For these, the tap is not moved until the reactive generation reaches a limit and the bus type switches to PQ. When an OLTC is "active" i.e. the tap-changing is moving to meet a target voltage, the voltage magnitude at the bus is assumed constant and is replaced as the variable for which an update is found in the load flow solution by the tap ratio. If the tap ratio hits an upper or lower limit, the tap ratio becomes fixed (the "tap limited" bit in the type integer is set) and the voltage magnitude once again becomes the state variable so that it is updated at each iteration.

The bus types are summarised in table A.1.

Description	Bit set								Type
	slack	Free gov.	PV	PQ	SVC	OLTC	Q limited	tap limited	
Slack	√								slack
Free governor		√	√						PV
PV			√						PV
PV + OLTC			√			√			PV
Free gov., Q = limit		√	√				√		PQ
PV, Q = limit			√				√		PQ
PV + OLTC, Q = limit			√			√	√	?	PQ
PQ				√					PQ
PQ + OLTC				√		√		?	PQ
SVC				√	√				PQ

Table A.1: Summary of bus types. "?" signifies bit is sometimes set

A.2.7 Summary of categories of line/branch types

The types of lines or branches that are modelled are

- A transmission line.
- A transformer with a fixed tap ratio.

- A transformer with a tap ratio that is fixed in each load flow study but which can nevertheless be varied between solutions.
- An on-load tap-changing transformer (OLTC) for which a target voltage may be set for one terminal of the transformer implying that the load flow state variable is the tap ratio while it remains within the upper and lower bounds.
- A quadrature booster (phase-shifting transformer) where the phase shift is fixed in each load flow study but may be changed between studies.
- A shunt compensator, usually a capacitor with fixed susceptance.

A.3 Power System Sensitivity Analysis

Consider a power system with N interconnected nodes. If the total power injected into the system is described by the vector G of nodal power injections and is a function of the vector of N dependent complex variables x, M control variables u and the vector of power system parameters such as line impendances and shunt susceptances p, then for balanced operation [26],

$$G(x, u, p) = 0 \tag{A.63}$$

If a small change to the control vector Δu is applied then a small change in the vector of dependent variables Δx will result. For balanced operation to continue

$$G(x + \Delta x, u + \Delta u, p) = 0 \tag{A.64}$$

Using a Taylor series expansion and neglecting higher order terms, A.64 can be re-expressed as

$$G(x_0, u_0, p) + G_x(x_0, u_0, p)\Delta x + G_u(x_0, u_0, p)\Delta u = 0 \tag{A.65}$$

where x_0 and u_0 are the original x and u vectors before the change and G_x and G_u are the Jacobians of G with respect to x and u respectively. If the system was balanced before the change, the first term vanishes leaving

$$G_x(x_0, u_0, p)\Delta x + G_u(x_0, u_0, p)\Delta u = 0 \tag{A.66}$$

From this, the change Δx resulting from Δu can be found

$$\Delta x = -G_x(x_0, u_0, p)^{-1} G_u(x_0, u_0, p)\Delta u \tag{A.67}$$

so that the sensitivity matrix S relating Δx to Δu is

$$S = -G_x(x_0, u_0, p)^{-1} G_u(x_0, u_0, p) \tag{A.68}$$

Once S is known, the change Δx_i resulting from the change in control Δu_j can be estimated simply as $S_{ij}\Delta u_j$.

A.3.1 Relation of reactive power controls to voltage magnitudes

If the well-known principle of decoupling the active and reactive power equations of a power system is utilised, equation A.66 can be re-expressed for the reactive subsystem as

$$G_{Qx_Q}(x_{Q_0}, u_{Q_0}, p)\Delta x_Q + G_{Qu_Q}(x_{Q_0}, u_{Q_0}, p)\Delta u_Q = 0 \tag{A.69}$$

where the Jacobians G_{Qx} and G_{Qu} are

$$G_{Qx_Q} = \left.\frac{\partial G_Q}{\partial x_Q}\right|_{x=x_{Q_0}} \tag{A.70}$$

$$G_{Qu_Q} = \left.\frac{\partial G_Q}{\partial u_Q}\right|_{u=u_{Q_0}} \tag{A.71}$$

and G_Q comprises only the reactive power equations. x_Q contains dependent voltage magnitudes V_l and reactive powers Q_g. u_Q contains controllable voltage magnitudes at generators, synchronous compensators or SVCs V_g, adjustable transformer tap ratios t and adjustable shunt susceptances B.

The $N \times M$ reactive sensitivity matrix S_Q is found by

$$S_Q = -G_{Qx_Q}(x_{Q_0}, u_{Q_0}, p)^{-1} G_{Qu_Q}(x_{Q_0}, u_{Q_0}, p) \tag{A.72}$$

In practice, S_Q is found by re-expressing equation A.72 as

$$G_{Qx_Q}(x_{Q_0}, u_{Q_0}, p)S_Q = -G_{Qu_Q}(x_{Q_0}, u_{Q_0}, p) \tag{A.73}$$

and factorising G_{Qx_Q} into lower and upper triangular factors so that

$$LUS_Q = -G_{Qu_Q}(x_{Q_0}, u_{Q_0}, p) \tag{A.74}$$

If the matrices S_Q and G_{Qu_Q} are described in terms of column vectors $s_1 \ldots s_M$ and $g_1 \ldots g_M$ where

$$S_Q = [s_1 \ldots s_M] \tag{A.75}$$

$$G_{Qu_Q} = [g_1 \ldots g_M] \tag{A.76}$$

then each equation

$$LUs_i = -g_i \tag{A.77}$$

for $i = 1 \ldots M$ can be solved for s_i by forward and backward substitution.

Finally, S_Q can be represented as

$$\begin{bmatrix} \Delta V_l \\ \Delta Q_g \end{bmatrix} = \begin{bmatrix} S_{Vt} & S_{VB} & S_{VV} \\ S_{Qt} & S_{QB} & S_{QV} \end{bmatrix} \begin{bmatrix} \Delta t \\ \Delta B \\ \Delta V_g \end{bmatrix} \tag{A.78}$$

where V_l is the vector of N_l load bus voltage bus magnitudes, Q_g is the vector of N_g reactive injections at generator or compensation buses, V_g is the vector of N_g generator or static voltage compensator (SVC) voltage set-points, B is the vector of N_b shunt susceptances and t is the vector of N_t transformer tap ratios.

A.3.2 Relation of active power controls to active power flows

Most references describing the formulation of a matrix of sensitivities of active power flows to active controls have required some pseudo-inverse matrix to be found, for example [95]. The approach adopted in this study and described in this section, however, avoids that and allows fast sparse matrix techniques to be used. This is achieved by finding the (sparse) Jacobian of derivatives of active power flow with respect to voltage angles and pre-multiplying it with an inverted square matrix.

If u_P is defined as the $1 \times M$ vector of controllable active power injections and quadrature booster angles, bus 1 is the slack bus, buses 2 to L are load buses and buses $L+1$ to N are generation buses, a sensitivity matrix A_P can be found relating small changes in u_P to small changes in the active state vector x_P comprising the active power generation at the slack bus and the nodal voltage angles θ at all the others. A_P is such that

$$\Delta x_P = A_P \Delta u_P \tag{A.79}$$

and

$$A_P = -G_{Px_P}^{-1} G_{Pu_P} \tag{A.80}$$

where

$$u_P = \{P_{L+1}, \ldots, P_N, \phi_1, \ldots, \phi_{M-(N-L)}\}^T \tag{A.81}$$

$$x_P = \{P_1, \theta_2, \ldots, \theta_N\}^T \tag{A.82}$$

$$G_{Px_P} = \left.\frac{\partial G_P}{\partial x_P}\right|_{x_P = x_{P_0}} \tag{A.83}$$

$$G_{Pu_P} = \left.\frac{\partial G_P}{\partial u_P}\right|_{u_P = u_{P_0}} \tag{A.84}$$

G_P comprises only the active power equations.

The changes in slack bus power and nodal voltage angles can be related to changes in active power transmitted along each transmision line or through each transformer P_{km} from general node k to node m by

$$\Delta P_{km} = \frac{\partial P_{km}}{\partial x_P} \Delta x_P \tag{A.85}$$

Hence, a matrix directly relating change in active power generation and quadrature booster angles (the vector of quantities which can be set) and change in active power flow (which is to be controlled) is found from

$$\Delta P_{km} = \frac{\partial P_{km}}{\partial x_P} A_P \Delta u_P \tag{A.86}$$

It is denoted S_P such that

$$S_P = \frac{\partial P_{km}}{\partial x_P} A_P \tag{A.87}$$

If equation A.80 is re-expressed as

$$\frac{\partial G_P}{\partial x_P} A_P = -\frac{\partial G_P}{\partial u_P} \tag{A.88}$$

then A_P can be found and pre-multiplied by $\partial P_{km}/\partial x_P$ to obtain S_P. Since $\Delta P_{km} = -\Delta P_{mk}$, S_P can then be used to determine the effects of changes in active power generation and quadrature booster angles on active power flows measured at each end of items of transmission plant. Note that change in slack bus power is not available as a control variable.

A.3.3 Relation of active power controls to voltage magnitudes

Whenever it has been found that movement of conventional reactive controls such as generator terminal voltages, transformer tap ratios, SVC set points and switchable shunt susceptances has been insuficient to remove violations of load bus voltage magnitude limits, it becomes apparent that the contribution of the transmission system itself to reactive demand must be altered. This can only be achieved by moving MW controls (active generation and quadrature booster angles) to alter flows. However, that such measures are necessary suggests that the system is highly stressed and use of load shedding to amend MW and MVAr demand as well as to change power flows on the system should be considered too.

In order to make use of these controls effectively, some sensitivity of the quantities to be controlled to movements in MW generation, quadrature booster angles and shedding of load should be found. The load shedding will be achieved by shedding both active and reactive load in equal proportions since it may in general be assumed that load power factors will remain constant.

As in sections A.3.1 and A.3.2, it is convenient to describe the power balance equation in terms of control vector u and state vector x partitioned into active and reactive components. In this case, however, u_P will comprise active generation, quadrature booster angles and active load while u_Q will comprise only reactive loads. x_P consists of busbar voltage angles and slack bus active power, and x_Q includes busbar voltage magnitudes for PQ buses and reactive generation for PV buses.

The power balance equation is then

$$G(x_Q, x_P, u_Q, u_P, p) = 0 \tag{A.89}$$

If u_P is changed by Δu_P and u_Q by Δu_Q, for balance to be maintained, there will be small changes in the states Δx_Q and Δx_P such that

$$G(x_Q + \Delta x_Q, x_P + \Delta x_P, u_Q + \Delta u_Q, u_P + \Delta u_P, p) = 0 \tag{A.90}$$

Taking the Taylor series expansion of this expression and neglecting higher order terms,

$$\begin{aligned} G(x_{Q_0}, x_{P_0}, u_{Q_0}, u_{P_0}, p) + G_{x_Q}(x_{Q_0}, x_{P_0}, u_{Q_0}, u_{P_0}, p)\Delta x_Q + G_{x_P}(x_{Q_0}, x_{P_0}, u_{Q_0}, u_{P_0}, p)\Delta x_P \\ + G_{u_Q}(x_{Q_0}, x_{P_0}, u_{Q_0}, u_{P_0}, p)\Delta u_Q + G_{u_P}(x_{Q_0}, x_{P_0}, u_{Q_0}, u_{P_0}, p)\Delta u_P = 0 \end{aligned} \tag{A.91}$$

where

$$G_{x_Q} = \frac{\partial G}{\partial x_Q} \tag{A.92}$$

$$G_{x_P} = \frac{\partial G}{\partial x_P} \tag{A.93}$$

$$G_{u_Q} = \frac{\partial G}{\partial u_Q} \tag{A.94}$$

$$G_{u_P} = \frac{\partial G}{\partial u_P} \tag{A.95}$$

Since the system was initially in balance, the first term, $G(x_{Q_0}, x_{P_0}, u_{Q_0}, u_{P_0}, p)$, is zero leaving

$$\frac{\partial G}{\partial x_Q}\Delta x_Q = -\frac{\partial G}{\partial x_P}\Delta x_P - \frac{\partial G}{\partial u_Q}\Delta u_Q - \frac{\partial G}{\partial u_P}\Delta u_P \tag{A.96}$$

If, now, G is also partitioned into active and reactive equations it can be seen that

$$\begin{bmatrix} \frac{\partial P}{\partial x_Q} \\ \frac{\partial Q}{\partial x_Q} \end{bmatrix} \Delta x_Q = -\begin{bmatrix} \frac{\partial P}{\partial x_P} \\ \frac{\partial Q}{\partial x_P} \end{bmatrix} \Delta x_P - \begin{bmatrix} \frac{\partial P}{\partial u_Q} \\ \frac{\partial Q}{\partial u_Q} \end{bmatrix} \Delta u_Q - \begin{bmatrix} \frac{\partial P}{\partial u_P} \\ \frac{\partial Q}{\partial u_P} \end{bmatrix} \Delta u_P \tag{A.97}$$

This represents two equations:

$$\frac{\partial P}{\partial x_Q}\Delta x_Q = -\frac{\partial P}{\partial x_P}\Delta x_P - \frac{\partial P}{\partial u_Q}\Delta u_Q - \frac{\partial P}{\partial u_P}\Delta u_P \tag{A.98}$$

$$\frac{\partial Q}{\partial x_Q}\Delta x_Q = -\frac{\partial Q}{\partial x_P}\Delta x_P - \frac{\partial Q}{\partial u_Q}\Delta u_Q - \frac{\partial Q}{\partial u_P}\Delta u_P \tag{A.99}$$

Equation (A.98) can be re-arranged to give

$$\frac{\partial P}{\partial x_P}\Delta x_P = -\frac{\partial P}{\partial x_Q}\Delta x_Q - \frac{\partial P}{\partial u_Q}\Delta u_Q - \frac{\partial P}{\partial u_P}\Delta u_P \tag{A.100}$$

and then

$$\Delta x_P = -\left[\frac{\partial P}{\partial x_P}\right]^{-1}\left\{\frac{\partial P}{\partial x_Q}\Delta x_Q + \frac{\partial P}{\partial u_Q}\Delta u_Q + \frac{\partial P}{\partial u_P}\Delta u_P\right\} \tag{A.101}$$

Substituting (A.101) into (A.99),

$$\frac{\partial Q}{\partial x_Q}\Delta x_Q = -\frac{\partial Q}{\partial u_Q}\Delta u_Q + \frac{\partial Q}{\partial x_P}\left[\frac{\partial P}{\partial x_P}\right]^{-1}\left\{\frac{\partial P}{\partial x_Q}\Delta x_Q + \frac{\partial P}{\partial u_Q}\Delta u_Q + \frac{\partial P}{\partial u_P}\Delta u_P\right\} - \frac{\partial Q}{\partial u_P}\Delta u_P \tag{A.102}$$

Grouping terms involving Δx_Q, Δu_Q and Δu_P together,

$$\left\{\frac{\partial Q}{\partial x_Q} - \frac{\partial Q}{\partial x_P}\left[\frac{\partial P}{\partial x_P}\right]^{-1}\frac{\partial P}{\partial x_Q}\right\}\Delta x_Q = \left\{\frac{\partial Q}{\partial x_P}\left[\frac{\partial P}{\partial x_P}\right]^{-1}\frac{\partial P}{\partial u_Q} - \frac{\partial Q}{\partial u_Q}\right\}\Delta u_Q + \left\{\frac{\partial Q}{\partial x_P}\left[\frac{\partial P}{\partial x_P}\right]^{-1}\frac{\partial P}{\partial u_P} - \frac{\partial Q}{\partial u_P}\right\}\Delta u_P \tag{A.103}$$

Since Δu_Q consists only of changes to reactive load and these are in proportion to the changes in active load, Δu_Q can be represented by

$$\Delta u_Q = \alpha^T \Delta u_P \tag{A.104}$$

where α is a column vector where the terms corresponding to changes in active generation and quadrature booster angles in u_P are zero and the others, as load power factors are assumed to be constant, are $\tan \phi_i$ where ϕ_i is the power factor angle of the load at the ith bus. Further, $\frac{\partial P}{\partial u_Q}$ under this circumstance is zero. Hence, equation (A.103) reduces to

$$\left\{ \frac{\partial Q}{\partial x_Q} - \frac{\partial Q}{\partial x_P} \left[\frac{\partial P}{\partial x_P} \right]^{-1} \frac{\partial P}{\partial x_Q} \right\} \Delta x_Q = \left\{ \frac{\partial Q}{\partial x_P} \left[\frac{\partial P}{\partial x_P} \right]^{-1} \frac{\partial P}{\partial u_P} - \frac{\partial Q}{\partial u_P} - \frac{\partial Q}{\partial u_Q} \alpha^T \right\} \Delta u_P \tag{A.105}$$

From equation (A.105), it can be seen that a sensitivity matrix relating changes in MW controls and load shedding to load bus voltage magnitudes and PV bus reactive generation can be found such that

$$\Delta x_Q = S_{QP} \Delta u_P \tag{A.106}$$

where

$$S_{QP} = \left\{ \frac{\partial Q}{\partial x_Q} - \frac{\partial Q}{\partial x_P} \left[\frac{\partial P}{\partial x_P} \right]^{-1} \frac{\partial P}{\partial x_Q} \right\}^{-1} \left\{ \frac{\partial Q}{\partial x_P} \left[\frac{\partial P}{\partial x_P} \right]^{-1} \frac{\partial P}{\partial u_P} - \frac{\partial Q}{\partial u_P} - \frac{\partial Q}{\partial u_Q} \alpha^T \right\} \tag{A.107}$$

In practice, S_{QP} is found by solving

$$A S_{QP} = B \tag{A.108}$$

where

$$A = \left\{ \frac{\partial Q}{\partial x_Q} - \frac{\partial Q}{\partial x_P} \left[\frac{\partial P}{\partial x_P} \right]^{-1} \frac{\partial P}{\partial x_Q} \right\} \tag{A.109}$$

$$B = \left\{ \frac{\partial Q}{\partial x_P} \left[\frac{\partial P}{\partial x_P} \right]^{-1} \frac{\partial P}{\partial u_P} - \frac{\partial Q}{\partial u_P} - \frac{\partial Q}{\partial u_Q} \alpha^T \right\} \tag{A.110}$$

by triangular factorisation of A and repeated forward and backward substitution for each column of S_{QP}.

Under certain circumstances, likely to be when the transmission system is not very heavily loaded, an approximation *may* be possible by considering that the term $\frac{\partial Q}{\partial x_P} \left[\frac{\partial P}{\partial x_P} \right]^{-1} \frac{\partial P}{\partial x_Q}$ has two cross-coupling terms between the active and reactive subsystems and is therefore likely to be insignificant [28]. A will then be simply $\frac{\partial Q}{\partial x_Q}$. This speeds the finding of the sensitivity matrix considerably since the triangular factors of $\frac{\partial Q}{\partial x_Q}$ will have already been found in the use of reactive controls for voltage control which preceeds the invocation of any MW controls for voltage. Care must be taken, though, to ensure that the approximation is valid under the system conditions being investigated.

A.3.4 Example of Jacobians for voltage control

Consider the small test power system shown in figure A.6 with $N = 6$ nodes, $n_{0\ldots5}$ where there is generation at n_4 and n_5, static compensation at n_3 and a tap-changing transformer between n_0 and n_3 so that $M = 4$. Using standard loadflow analysis, n_4 and n_5 are designated as 'PV' nodes, i.e. nodes with controllable active power injection P and voltage magnitude V. Load nodes $n_{0\ldots2}$ are designated 'PQ' nodes with fixed P and Q injections and have voltage magnitudes which are to be controlled. Node n_3 can be regarded in one of two ways, either as a load 'PQ' node with a shunt capacitor, or a 'PV' node with variable set voltage magnitude V [91]. These will be considered in turn.

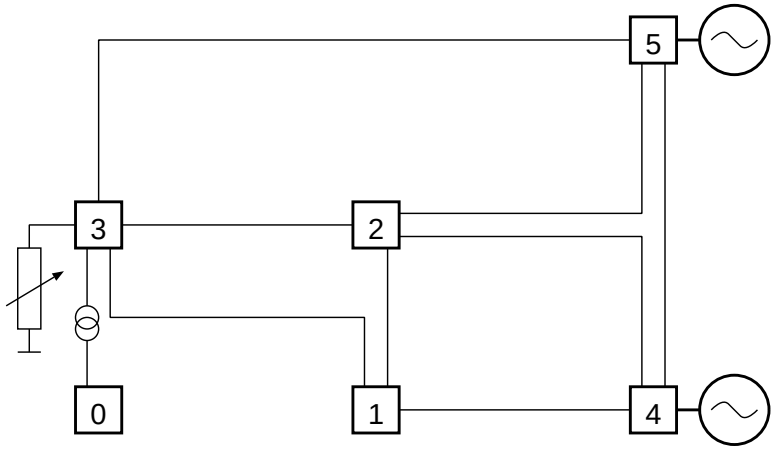

Figure A.6: Example 6 bus system

Static compensator node treated as a 'PV' node

For the test system described and node n_3 treated as a 'PV' node, x_Q is a column vector of order N and u_Q is a column vector of order M such that,

$$x_Q = \{V_0, V_1, V_2, Q_3, Q_4, Q_5\}^T \tag{A.111}$$

$$u_Q = \{t_{03}, V_3, V_4, V_5\}^T \tag{A.112}$$

where t_{03} is the tap ratio of the transformer between nodes n_0 and n_3.

$$G_{Qx} = \begin{bmatrix} \partial Q_0/\partial V_0 & & & & & \\ & \partial Q_1/\partial V_1 & \partial Q_1/\partial V_2 & & & \\ & \partial Q_2/\partial V_1 & \partial Q_2/\partial V_2 & & & \\ \partial Q_3/\partial V_0 & \partial Q_3/\partial V_1 & \partial Q_3/\partial V_2 & -1 & & \\ & \partial Q_4/\partial V_1 & \partial Q_4/\partial V_2 & & -1 & \\ & & \partial Q_5/\partial V_2 & & & -1 \end{bmatrix} \tag{A.113}$$

$$G_{Qu} = \begin{bmatrix} \partial Q_0/\partial t_{03} & \partial Q_0/\partial V_3 & & \\ & \partial Q_1/\partial V_3 & \partial Q_1/\partial V_4 & \\ & \partial Q_2/\partial V_3 & \partial Q_2/\partial V_4 & \partial Q_2/\partial V_5 \\ \partial Q_3/\partial t_{03} & \partial Q_3/\partial V_3 & & \partial Q_3/\partial V_5 \\ & & \partial Q_4/\partial V_4 & \partial Q_4/\partial V_5 \\ & \partial Q_5/\partial V_3 & \partial Q_5/\partial V_4 & \partial Q_5/\partial V_5 \end{bmatrix} \quad \text{(A.114)}$$

Static compensator node treated as a 'PQ' node

With node n_3 treated as a 'PQ' node, there are two possibilities for the control at bus 3: an MSC with a variable shunt susceptance B_{s3} and an SVC modelled as having some "droop" and a variable V_{ref}. The latter case is not considered here since the SVCs on the studied systems are regarded as having V_{ref} fixed. In the former case i.e. with the control variable regarded as B_{s3},

$$x_Q = \{V_0, V_1, V_2, V_3, Q_4, Q_5\}^T \quad \text{(A.115)}$$

$$u_Q = \{t_{03}, B_{s3}, V_4, V_5\}^T \quad \text{(A.116)}$$

and

$$G_{Qx} = \begin{bmatrix} \partial Q_0/\partial V_0 & & & \partial Q_0/\partial V_3 & & \\ & \partial Q_1/\partial V_1 & \partial Q_1/\partial V_2 & \partial Q_1/\partial V_3 & & \\ & \partial Q_2/\partial V_1 & \partial Q_2/\partial V_2 & \partial Q_2/\partial V_3 & & \\ \partial Q_3/\partial V_0 & \partial Q_3/\partial V_1 & \partial Q_3/\partial V_2 & \partial Q_3/\partial V_3 & & \\ & \partial Q_4/\partial V_1 & \partial Q_4/\partial V_2 & & -1 & \\ & & \partial Q_5/\partial V_2 & \partial Q_5/\partial V_3 & & -1 \end{bmatrix} \quad \text{(A.117)}$$

$$G_{Qu} = \begin{bmatrix} \partial Q_0/\partial t_{03} & & & \\ & & \partial Q_1/\partial V_4 & \\ & & \partial Q_2/\partial V_4 & \partial Q_2/\partial V_5 \\ \partial Q_3/\partial t_{03} & \partial Q_3/\partial B_3 & & \partial Q_3/\partial V_5 \\ & & \partial Q_4/\partial V_4 & \partial Q_4/\partial V_5 \\ & & \partial Q_5/\partial V_4 & \partial Q_5/\partial V_5 \end{bmatrix} \quad \text{(A.118)}$$

With on-load tap-changing transformer

With the transformer between nodes 0 and 3 regarded as an OLTC with the tap changing to control the voltage at bus 0 (a 'PQ' bus) and bus 3 modelled as a 'PV' bus, the state vector x_Q and the control vector u_Q are

$$x_Q = \{t_{03}, V_1, V_2, Q_3, Q_4, Q_5\}^T \quad \text{(A.119)}$$

$$u_Q = \{V_0, V_3, V_4, V_5\}^T \quad \text{(A.120)}$$

and the Jacobians are

$$G_{Qx} = \begin{bmatrix} \partial Q_0/\partial t_{03} & & & & & \\ & \partial Q_1/\partial V_1 & \partial Q_1/\partial V_2 & & & \\ & \partial Q_2/\partial V_1 & \partial Q_2/\partial V_2 & & & \\ \partial Q_3/\partial t_{03} & \partial Q_3/\partial V_1 & \partial Q_3/\partial V_2 & -1 & & \\ & \partial Q_4/\partial V_1 & \partial Q_4/\partial V_2 & & -1 & \\ & & \partial Q_5/\partial V_2 & & & -1 \end{bmatrix} \tag{A.121}$$

$$G_{Qu} = \begin{bmatrix} \partial Q_0/\partial V_0 & \partial Q_0/\partial V_3 & & \\ & \partial Q_1/\partial V_3 & \partial Q_1/\partial V_4 & \\ & \partial Q_2/\partial V_3 & \partial Q_2/\partial V_4 & \partial Q_2/\partial V_5 \\ \partial Q_3/\partial V_0 & \partial Q_3/\partial V_3 & & \partial Q_3/\partial V_5 \\ & & \partial Q_4/\partial V_4 & \partial Q_4/\partial V_5 \\ & \partial Q_5/\partial V_3 & \partial Q_5/\partial V_4 & \partial Q_5/\partial V_5 \end{bmatrix} \tag{A.122}$$

Bibliography

[1] The National Grid Company plc, *A review of transmission security standards*, National Grid Company plc, U.K., 1994.

[2] J.M. Hammersley and D.C. Handscomb, *Monte Carlo Methods*, Methuen, 1964.

[3] J.P.C. Kleijnen, *Statistical techniques in simulation—part 1*, Marcel Dekker, 1974.

[4] R. Billinton and R.N. Allan, *Reliability Evaluation of Power Systems*, Plenum Press, second edition, 1996.

[5] T. Hopkins, *Numerical methods in practice: using the NAG Library*, Addison-Wesley, 1988.

[6] Numerical Algorithms Group Ltd., *NAG Fortran library manual. Mark 16*, Oxford: NAG, 1993.

[7] B.M. Weedy and B.J. Cory, *Electric Power Systems*, John Wiley and Sons, fourth edition, 1998.

[8] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation and Control*, Wiley Interscience, second edition, 1996.

[9] J. Corwin and W. Miles, "Impact assessment of the 1977 New York blackout", Tech. Rep., U.S. Department of Energy, Washington DC, USA, July 1977.

[10] E. Agneholm, "The restoration process following a major breakdown in a power system", Tech. Rep., Chelmer University of Technology - School of Electrical and COmputer Engineering, Technical Report No. 230L, May 1996.

[11] WSCC Investigative Task Force, "Wscc preliminary system disturbance report august 10, 1996", Tech. Rep., WSCC for the DOE, WSCC, August 1996.

[12] A. Merlin J.C. Dodu, "New probabilistic approach taking into account reliability and operation security in ehv power system planning at edf", *IEEE Trans on Power Systems*, vol. 1, no. 3, pp. 175–181, 1986.

[13] G.D. Friedlander, "The great blackout of '65", *IEEE Spectrum*, , no. 10, pp. 83–86, 1976.

[14] A.G. Phadke and L.S. Thorp, "Expose hidden failures to prevent cacading outages", *Computer Applications on Power Systems*, vol. CAP-11, no. 3, pp. 20–23, 1996.

[15] C.Tamronglak, S.H. Horowitz, A.G. Phadke, and J.S. Thorp, "Anatomy of power system blackouts: preventive relaying strategies", *Trans. on Pow. Sys.*, vol. 11, no. 2, pp. 708–715, 1995.

[16] C.W. Taylor and D.C. Erickson, "Recording and analyzing the july 2 cascading outage", *Computer Applications on Power Sistems*, vol. 13, no. 1, pp. 26–30, 1997.

[17] S. Aboreshaid and R. Billinton, "Probabilistic transient stability studies using the method of bisection", *IEEE Trans on Power Systems*, vol. 11, no. 4, pp. 1990–1995, 1996.

[18] R. Billinton and S. Aboreshaid, "Stochastic modelling of high-speed reclosing in probabilistic transient stability studies", *IEE Proc Gen, Trans and Dist*, vol. 142, no. 4, pp. 350–354, 1995.

[19] R. Billinton and S. Aboreshaid, "Security evaluation of composite power systems", *IEE Proc Gen, Trans and Dist*, vol. 142, no. 5, pp. 511–516, 1995.

[20] J.D.McCalley, A.A.Fouad, V.Vittal, A.A.Irizarry-Rivera, B.L.Agrawal, and R.G.Farmer, "A risk-based security index for determining operating limits in stability-limited electric power systems", *IEEE Transactions on Power Systems*, vol. 12, no. 3, pp. 1210–1219, 1997.

[21] P. Kundur, *Power System Stability and Control*, McGraw-Hill, 1994.

[22] B.Gao, G.K. Morison, and P. Kundur, "Voltage stability evaluation using modal analysis", *IEEE Trans on Power Systems*, vol. 7, no. 4, pp. 1529–1542, 1992.

[23] P. Kessel and H. Glavitsch, "Estimating the voltage stability of a power system", *IEEE Trans. on Power Delivery*, vol. 1, no. 3, pp. 346–354, 1986.

[24] T.Quoc Tran, J.Fandino, N.Hadjsaid, J.C. Sabonnadiere, and H.Vu, "Emergency load shedding to avoid risks of voltage instability indicators", *IEEE Trans. on Power Systems*, vol. 9, no. 1, pp. 341–351, 1994.

[25] R. Billinton and S. Aboreshaid, "Voltage stability considerations in composite power system reliability evaluation", *IEEE Trans. on Power Systems*, vol. 13, no. 2, pp. 655–660, 1998.

[26] I. Hano, Y. Tamura, S. Narita, and K. Matsumoto, "Real time control of system voltage and reactive power", *IEEE Trans. on Power Apparatus and Systems*, vol. PAS-88, no. 10, pp. 1544–1559, 1969.

[27] K.R.W. Bell, A.R. Daniels, and R.W. Dunn, "Modelling of operator heauristics in dispatch for security enhancement", *to appear in IEEE Trans. on Power Systems*, 1998.

[28] D.S. Kirschen and H.P. Van Meeteren, "MW/voltage control in a linear programming based optimal power flow", *IEEE Trans. on Power Systems*, vol. 3, no. 2, pp. 481–489, 1988.

[29] K.R.W. Bell, A.R. Daniels, and R.W. Dunn, "A fuzzy expert system for overload alleviation", *Proc, PSCC 12, Dresden, Germany*, pp. 1177–1183, August, 1996.

[30] J.J. Ancona, "A framework for power system restoration following a major power failure", *IEEE Trans on Power Systems*, vol. 10, pp. 1480–1485, 1995.

[31] IEEE Comittee Report, "Power system restoration - a task force report", *IEEE Trans on Power Systems*, vol. 2, no. 2, pp. 271–277, 1987.

[32] IEEE Comittee Report, "Power system restoration - the second task force report", *IEEE Trans on Power Systems*, vol. 2, no. 4, pp. 927–933, 1987.

[33] R. Kearsley, "Restoration in sweden and experience gained from the blackout of 1883", *IEEE Trans. on Power Systems*, vol. 2, no. 2, pp. 422–428, 1987.

[34] R.N. Allan and D.S. Kirschen, "Assessment of the value of security—case for support", Tech. Rep., Manchester Centre for Electrical Energy, UMIST, November 1994.

[35] K.K. Kariuki and R.N. Allan, "Evaluation of reliability worth and value of lost load", *IEE Proc. on Gen., Tran. and Dist.*, vol. 143, no. 2, pp. 171–180, 1996.

[36] K.K. Kariuki and R.N. Allan, "Applications of customer outage costs in system planning, design and operation", *IEE Proc. on Gen., Trans. and Dist.*, vol. 143, no. 2, pp. 305–312, 1996.

[37] G. Wacker and R. Billinton, "Customer cost of electric service interruptions", *Proc. IEEE*, vol. 77, no. 6, pp. 919–930, 1989.

[38] R.N. Allan, "VOLL—fact or fiction?", *IEE Power Engineering Journal*, vol. 9, no. 1, pp. 2, 1995.

[39] R.N. Allan and R. Billinton, "Power system reliability and its assessment", *IEE Power Engineering Journal*, August 1993.

[40] G. Tollefson, R. Billinton, G. Wacker, E. Chan, and J. Aweya, "A Canadian customer survey to assess power system reliability worth", *IEEE Trans. Pow. Sys.*, vol. 9, no. 1, pp. 443–450, 1994.

[41] K.K. Kariuki, *Assessment of customer outage costs due to electric service interruptions*, PhD thesis, Department of Electrical Engineering and Electronics, UMIST, Manchester, October 1995.

[42] K.K. Kariuki and R.N. Allan, "Assessment of customer outage costs due to electric service interruptions—residential sector", *IEE Proc. on Gen., Tran. and Dist.*, vol. 143, no. 2, pp. 163–170, 1996.

[43] D.C. Goushleff, "Use of interruption costs in regional supply planning", Tech. Rep., Ontario Hydro, 80-301-K.

[44] M.A. Rios, "Assignment of value of lost of load in value of security assessor", Tech. Rep., Manchester Centre for Electrical Energy, UMIST, Manchester, January 1999.

[45] M.R. Bhiuyun and R.N. Allan, "Inclusion of weather in composite system reliability evaluation using sequential simulation", *IEE Proc. Gen., Trans. and Dist.*, vol. 141, no. 6, pp. 575–584, 1994.

[46] M.R. Bhiuyun, *Reliability assessment in power system planning*, PhD thesis, Department of Electrical Engineering and Electronics, UMIST, Manchester, October 1993.

[47] A.M. Mood, F.A. Graybill, and D.C. Boes, *Introduction to the Theory of Statistics*, McGraw-Hill, 1974.

[48] J. E. Freund and G.A. Simon, *Modern Elementary Statistics*, Prentice-Hall, 1992.

[49] R.V. Hogg and J. Ledolter, *Applied statistics for engineers and physical scientists*, Macmillan, 1992.

[50] R.Y. Rubinstein, *Simulation and the Monte Carlo Method*, Wiley, 1981.

[51] D.E. Knuth, *The Art of Computer Programming—volume 2: Seminumerical Algorithms*, Addison-Wesley, 1981.

[52] R. Billinton and W. Li, *Reliability Assessment of Electric Power Systems Using Monte Carlo Methods*, Plenum Press, 1994.

[53] M. Kahn and A.W. Marshall, "Methods of reducing sample size in Monte Carlo computations", *Operations Research*, vol. 1, pp. 263–278, 1953.

[54] E.I. McGrath, *Fundamental of Operations Research*, West Coast University, 1970.

[55] W.G. Cochran, *Sampling Techniques*, Wiley, 1966.

[56] H. Kumamoto, K. Tanaka, K. Inoue, and E.J. Henley, "Dagger-sampling Monte Carlo for system unavailability evaluation", *IEEE Trans. on Reliability*, vol. R-29, no. 2, 1980.

[57] Peter Kelen, "Private correspondance", Consulting Engineer, 24 Woodside Avenue, Bucks HP9 1JJ, U.K., 12th December 1997.

[58] G.S. Fishman and B.D. Huang, "Antithetic variates revisited", *Communication of the ACM*, vol. 26, no. 11, pp. 964–971, 1983.

[59] Peter Kelen, "Private correspondance", Consulting Engineer, 24 Woodside Avenue, Bucks HP9 1JJ, U.K., 7th November 1997.

[60] M.E. Brenner, "Selective sampling—a technique for reducing sample size in simulation of decision making problems", *J. Ind. Eng.*, vol. 14, pp. 291–296, 1963.

[61] M. Quenouille, "Approximation tests of correlation in time series", *J. Royal Statistical Society B*, vol. 11, pp. 18–84, 1949.

[62] J.W. Tukey, "Bias and confidence in not-quite large samples", *Annals of Mathematical Statistics*, vol. 29, 1958.

[63] J. Shao and D. Tu, *The Jackknife and Bootstrap*, Springer, 1995.

[64] B. Efron, "Bootstrap methods: another look at the jackknife", *Ann. Statistics*, vol. 7, pp. 1–26, 1979.

[65] B. Efron and R.J. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall, 1993.

[66] B. Efron and R. Tibshirani, "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy", *Statistical Science*, vol. 1, no. 1, pp. 54–77, 1986.

[67] J.R. Ubeda and R.N. Allan, "Sequential simulation applied to composite system reliability evaluation", *IEE Proceedings Generation, Transmission and Distribution*, vol. 139, no. 2, pp. 81–86, 1992.

[68] C. Singh and R. Billinton, *System Reliability Modelling and Evaluation*, Hutchinson, London, 1977.

[69] R. Billinton and W. Li, "System state transition sampling method for composite system reliability evaluation", *IEEE Trans on Power Systems*, vol. 8, no. 3, pp. 761–770, 1993.

[70] J.C.O. Mello, M.V.F. Pereira, and A.M. Leite da Silva, "Evaluation of reliability worth in composite systems based on pseudo-sequential Monte Carlo simulation", *IEEE Trans on Power Systems*, vol. 9, no. 3, pp. 1318–1326, 1994.

[71] J.C.O. Mello, A.M. Leite da Silva, and M.V.F. Pereira, "Efficient loss-of-load cost evaluation by combined pseudo-sequential and state transition simulation", *IEE Proc. Gen. Trans. and Dist.*, vol. 144, no. 2, pp. 147–154, 1997.

[72] P. Bratley, B.L. Fox, and L.E. Schrage, *A Guide to Simulation*, Springer-Verlag, 1987.

[73] H. Kumamoto, K. Tanaka, K. Inoue, and E.J. Henley, "State transition Monte Carlo for evaluating large, repairable systems", *IEEE Trans. on Reliability*, vol. R-29, no. 5, 1980.

[74] U. Bryan and D.T.Y. Cheng, "New economic assessment tools for maximising open trading opportunities", *CIGRE, Paris*, 1997.

[75] J.H. Pickles, I.H. Russell, and J.F. Macqueen, "Importance sampling for power system security assessment", *Third International Conference on Probabilistic Methods Applied to Electric Power Systems*, pp. 47–52, 3-5 July 1991.

[76] M.V.F. Pereira, M.E.P. Maceira, G.C. Oliveira, and L.M.V.G. Pinto, "Combining analytical models and monte carlo techniques in probabilistic power system analysis", *IEEE Trans on Power Systems*, vol. 7, no. 1, pp. 265–272, 1992.

[77] J.S. Thorp, A.G. Phadke, S.H. Horowitz, and C. Tamronglak, "Anatomy of power system disturbances: importance sampling", *Proc 12th Power Systems Computation Conference, Dresden*, pp. 350–357, 1996.

[78] D.H. Brereton, "A note on importance sampling with special reference to large networks", Tech. Rep., National Grid Research and Development Centre, Kelvin Avenue, Leatherhead, Surrey KT22 7ST, March 1991.

[79] G.C. Oliveira, M.V.F. Pereira, and S.H.F. Cunha, "A technique for reducing the computational effort in monte-carlo based composite reliability evaluation", *IEEE Trans on Power Systems*, vol. 4, no. 4, pp. 1309–1315, 1989.

[80] M.V.F. Pereira and L.M.V.G. Pinto, "A new computational tool for composite reliability evaluation", *IEEE Trans on Power Systems*, vol. 7, no. 1, pp. 258–264, 1992.

[81] R. Billinton and A. Jonnavithula, "Composite system adequacy assessment using sequential Monte Carlo simulation with variance reduction techniques", *IEE Proc Gen, Trans and Dist*, vol. 144, no. 1, pp. 1–6, 1997.

[82] A. Breipohl, F.N. Lee, J. Huang, and Q. Feng, "Sample size reduction in stochastic production simulation", *IEEE Trans. on Power Systems*, vol. 5, no. 3, pp. 984–992, 1990.

[83] S.R. Huang and S.L. Chen, "Evaluation and improvement of variance reduction in monte carlo production simulation", *IEEE Transactions on Energy Conversion*, vol. 8, no. 4, pp. 610–617, 1993.

[84] M.E. Khan and R. Billinton, "A hybrid model for quantifying different operating states of composite power systems", *IEEE Trans. on Power Systems*, vol. 7, no. 1, pp. 187–193, 1992.

[85] A. Sankarakrishnan and R. Billinton, "Sequential Monte Carlo simulation for composite power system reliability analysis with time varying loads", *IEEE Trans on Power Systems*, vol. 10, no. 3, pp. 1540–1545, 1995.

[86] C. Marnay and T. Strauss, "Effectiveness of antithetic sampling and stratified sampling in Monte Carlo chronological production cost modeling.", *IEEE Transactions on Power Systems*, vol. 6, no. 2, pp. 669–675, 1991.

[87] G.J. Anders, *Probability Concepts in Electric Power Systems*, Wiley, 1990.

[88] R.N. Allan and J. Roman, "Reliability assessment of hydrothermal generation systems containing pumped storage plant", *IEE Proceedings, Part C: Generation, Transmission and Distribution*, vol. 138, no. 6, pp. 471–478, 1991.

[89] B. Stott, "Review of load flow calculation methods", *Proc. IEEE*, vol. 62, pp. 916–929, 1974.

[90] B. Stott and O. Alsac, "Fast decoupled load flow", *IEEE Trans. on Power Apparatus and Systems*, vol. PAS-93, no. 3, pp. 859–869, 1974.

[91] J.Arrillaga and C.P.Arnold, *Computer Analysis of Power Systems*, Wiley, 1990.

[92] Keith Bell, "SVC modelling in the Value of Security Assessor", Tech. Rep., Manchester Centre for Electrical Energy, UMIST, PO Box 88, Manchester M60 1QD, UK, 1997.

[93] S-K. Chang and V. Brandwajn, "Adjusted solutions in fast decoupled load flow", *IEEE Trans. on Power Systems*, vol. 3, no. 2, pp. 726–733, 1988.

[94] N.M. Peterson and W.S. Meyer, "Automatic adjustment of transformer and phase-shifter taps in newton power flow", *IEEE Trans Pow App Sys*, vol. PAS-90, no. 1, pp. 103–108, 1971.

[95] A.O. Ekwue and M.J. Short, "Interactive security evaluation in power system operation", *Proc. IEE Part C*, vol. 130, no. 2, pp. 61–70, 1983.