Data Collection in Steady State Conditions

We briefly discuss in this section some of the common methods used to gather steady state observation data during a simulation run. The problem specifically considered is the way one can get *n* independent simulation runs over which the output parameters may be observed and estimated so that the results from the runs may be used to get a suitable confidence estimate. The three techniques commonly proposed for this are *(i) The Subinterval Method*, *(ii) The Regenerative Method* and *(iii) The Replication Method*. In our earlier discussions, we had in a way implied the use of the replication method as that is indeed the one that is most commonly used.

7.6.1 The Subinterval Method or the Method of Batch Mean

This method really involves only one very long simulation run which is suitably subdivided into an initial transient period and *n* batches. Each of these batches is then treated as an independent run of the simulation experiment while no observations are made during the transient period which is treated as the warm-up interval. An example of this has been shown in Figure 7.10. Note that one would normally choose batch intervals of equal size but this is not mandatory. One can choose batch intervals of different lengths if there is a reasonable logic on which this may be based.

Choosing a large batch interval size would effectively lead to independent batches and hence, independent runs of the simulation as required for confidence estimation purposes. One advantage of this method is that only one transient (warm-up) interval needs to be accounted for and discounted and removed during the process of recording observations. Since the simulation would tend to run for a long simulation time in this approach, this would also tend to reduce the ill effects, if any, of not properly removing the transient period during the warm-up interval.

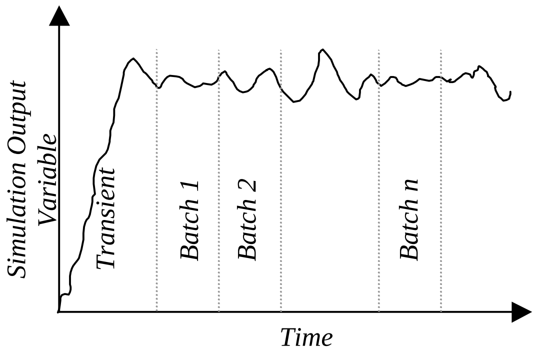

Figure 7.10. The Subinterval Method or The Method of Batch Means

One disadvantage of this approach is that the batches, as illustrated in Figure 7.10, may not really be independent. Statistically, one can make the observation that high values of an observed parameter will tend to follow high values and low values will tend to follow low values. This would be especially true if the batch sizes are not large enough and will lead to significant correlation between successive batches. Since the confidence estimation measures are based on the independence of the individual simulation runs, this lack of independence between successive batches, if sufficiently serious, may have a serious impact on the accuracy of the confidence estimation procedures.

7.6.2 The Regenerative Method

This method is basically intended to reduce the problem of correlation between the batches that one may encounter in the Subinterval Method of Section 7.6.1. We still use one long run as before but select an appropriately identified state of the system as the *regenerative state* and the time instants when this occurs as the *regenerative points*. The batches start and end at these regenerative points once steady state has been reached. This method is illustrated in Figure 7.11.

Choosing the regenerative state appropriately may not be as difficult as it may first appear. Consider the variable N representing the number in a queue, which is being simulated. This will typically build up from a *zero* value but, if the system is stable, then it will also eventually return to a *zero* value. This will happen repeatedly with the system state fluctuating between positive, non-zero integer values in between the zero-value points. The zero-value points are interesting because once the system reaches this state, its future evolution does not depend on how it actually reached this state. This behaviour of the system implies that the way the system evolves in one *idle*-

to-idle cycle of this type will not depend on previous cycles of this type and will not affect future cycles of this type either. Since this is the case, we can indeed choose these *idle-to-idle cycles* as the independent batches suggested in the earlier description and illustrated in Figure 7.11.

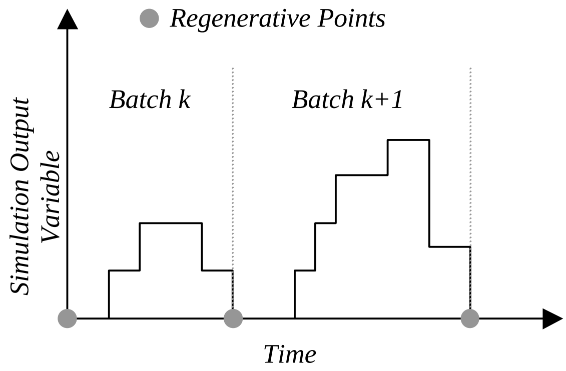

Figure 7.11. The Regenerative Method of choosing Independent Batches

This method is qualitatively better at ensuring independence between successive batches than the earlier *Subinterval Method*. Apart from the fact that one must correctly identify the regenerative states, this method does have a few problems that should be considered. Firstly, the time interval between regenerative points may be long - if this happens then one will need to run the simulation for a long time (in terms of the simulation time) in order to get a sufficient number of independent batches. The adverse effect of this, if it happens, will also be that only a few batches may be obtainable even in a long simulation run. If this is the case then it will adversely affect the computation of the confidence intervals for confidence estimation purposes. One other problem is that the batch sizes and their duration will themselves be random in nature. It may be necessary to account for this in our computations.

7.6.3 The Replication Method

This method is the one most popularly used and had been implied in a way in our earlier descriptions. This was basically the suggested technique for getting *n* independent runs of the simulation experiment by running the simulator *n* times with different initial random seeds for the simulator's random number generator. We have illustrated an example of this in Figure 7.12 where the independent runs for the two batches, batch *k* and batch *k+1* are shown with their own transient (warm up) intervals and the time intervals during which the system reaches steady state in these two runs.

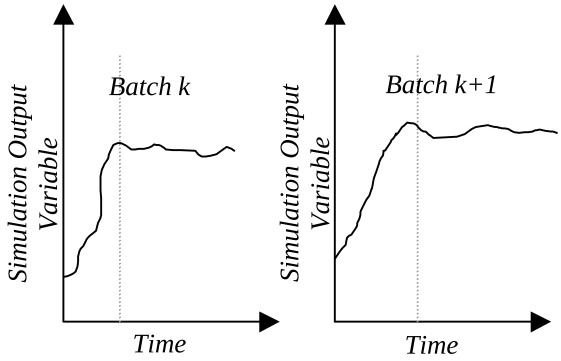

Figure 7.12. The Replication Method

In this case, *n* independent replications of the simulation run are done. In each case, an appropriately selected transient (warm up) period is removed from each run before the data collection is started. For the observed intervals after the warm up period (i.e. the intervals during which the system reaches steady state), data is collected and processed for the point estimates of the variables being observed and for their subsequent confidence estimation.

The *n* simulation runs must be truly independent. In practice, this may generally be satisfactorily achieved by selecting different seeds for starting the random number generator used by the simulator. In order to achieve independent replications in this fashion, one should ensure the use of a good random number generator and sufficiently statistically different starting seeds for each run. This method does require separate warm up periods to be estimated and removed from each run to eliminate the effects of transients. Apart from this, the replication method is simple and easy to use and is the method typically used to generate multiple simulation runs of the simulation experiment for confidence estimation.