Abstract: Design of Computer Experiments for Packet Communication Networks

Ben M Parker

February 28, 2009

1 Overview

We investigate uses of computer experiments in the measurement of packet communication networks. The general idea behind packet communication networks is to transmit information from a sender to a receiver by gathering the information into small fragments called packets; these packets are sent across a transmission medium where they are reassembled by the receiver according to a pre-determined protocol. The exact mechanisms used for this information transfer are beyond the scope of this work, but suffice it to say that packet communication networks account for the vast majority of information transferred today through the Internet, and other communication channels.

In an industrial context, manufacturers of networking equipment and also internet service providers expend a great deal of resources in trying to measure networks, both in order to assess any faults with a service, and to provide accurate service level information to clients for use in billing, or assessing success against contractual agreements for service provision, for example.

Perhaps strangely, there is little understanding of how to measure these communications networks on a fundamental level. In our work, we show that by regarding measurement procedures on packet communication networks as experiments that can be optimised in the traditional sense, we can improve current industrial best practice in measurement.

Consider, for example, the case of a broadband user who wants to confirm that his Internet service provider is providing the agreed bandwidth: having no privileged access to the system, he may only send exploratory or "probe" packets at whichever rate he chooses, and measure the time that these packets take to arrive at their destination. What rate best allows him to determine the bandwidth? The problem of users seemingly not getting contractually agreed bandwidth has been recently discussed in the popular media, e.g. [1].

Essentially here, we have a classic bias-variance trade-off. By increasing the rate of the probe packets, our broadband user will gather more data which will allow him accurately to determine the bandwidth more precisely (lower variance). Conversely, sending more probe packets means that there will be more packets in the system; the system may become overloaded, and probe packets may become delayed, which decreases the accuracy of the probing regime (higher bias).

In previous work[2], we have studied a system (a buffer in a communications network) which we can measure as a empty/full binary state. The system evolves, according to a transition matrix P which we parameterise via $P = \begin{pmatrix} 1-p & p \\ r & 1-r \end{pmatrix}$. Our question was then, given we can observe the system at discrete time points 0,1,2,..., and we have a fixed number of observations T, at what time points should we measure to allow us to make the most inference about our parameters under study (here p and r)?

In the work we present here, we seek to examine these probing experiments in the context of computer experiments. We model a (perhaps large) network as a computer simulation. There may be many unknown parameters, representing different usage patterns of users on the network. We wish to measure these unknown parameters, or more usually a subset of them. We have some design factors which we may vary, such as the times and the patterns of any monitoring or probing we do on the network. We wish to choose our design factors such that we get the most information on the unknown parameters.

This can be thought of as a typical computer experiment problem, in the context of a regression problem. Given vectors of input parameters, some of which may be design factors, $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}$, we get a set of outputs $\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_n}$. Typically in computer experiments, the simulations are deterministic in the sense that the same input generates the same outputs, i.e. if $\mathbf{x_i} = \mathbf{x_j}$ then $\mathbf{y_i} = \mathbf{y_j}$. The aim of these computer experiments is to be able to describe $\mathbf{y}$ well across the whole (normally multi-dimensional) range of possible inputs, given that we have observed a small number of them. Due to the deterministic nature of the process, we know with certainty the value of $\mathbf{y}$ at the design points, and various methods are used to interpolate between them: Gaussian process regression, otherwise known as Kriging, is a popular technique. Space filling designs, and sequential designs, are both the subject of current research.

The difference in the problems for networks that we consider here, is that networks are stochastic (i.e. non-deterministic) processes and we feel that deterministic simulation is not appropriate. In many cases, we are actually interested in modelling the variation in the output.

There seems to be little research on designs for stochastic simulation. We therefore focus on two questions in this presentation. Firstly, computer experiments on packet communications networks are an interesting new application of statistics to a real engineering problem, and we discuss how we can apply the optimal design framework in this context. Secondly, we discuss some of the problems and opportunities we find in applying computer experiments to stochastic systems where the variability itself is interesting to measure.

References

[1] BBC. Broadband speeds under scrutiny. `http://news.bbc.co.uk/1/hi/technology/7003113.stm`, 2007.

[2] B. Parker, S. Gilmour, and J. Schormans. Measurement of packet loss probability by optimal design of packet probing experiments. *IET Communications*, Special Issue, To appear.

[3] W. van Beers and J. Kleijnen. Customized sequential designs for random simulation experiments: Kriging metamodeling and bootstrapping. *European Journal of Operational Research*, 186(3):1099–1113, 2008.