Chapter 2
Recursive Estimation: A Simple Tutorial

Introduction

Figure 2.1 is a plot of the annual rainfall $y_i, i = 1, 2, \ldots, 132$, for the town of Walgett in New South Wales, Australia, over the period 1879 to 2010. It is a typical example of time-series data and the human eye (which is an extremely good and underrated filter of data) can discern various characteristics and patterns which could be described verbally: for example, the rainfall is extremely variable; it averages about 472 mm a year; and there seems to be indications of some changes in the mean and variance over the last 30 to 40 years.

But such a description is largely qualitative; if we are to be more quantitative and precise in our evaluation, and particularly if we wish to compare the rainfall at Walgett with that measured at other stations in N.S.W. or elsewhere, then the data must be compressed in some manner to yield a reduced and hopefully small number of ‘statistics’ that can be computed easily from the original data and which collectively provide a good description of these data. The most obvious statistics in the case of the Walgett data are the first two statistical moments, in the form of the sample mean or average rainfall $\bar{y}(N)$, the sample variance $\sigma^2(N)$ and the sample standard deviation $\sigma(N)$ about this mean value (the sample covariance is considered in Appendix B), where these are defined as follows:

\[
\bar{y}(N) = \frac{1}{N} \sum_{i=1}^{N} y(i)
\]

\[
\sigma^2(N) = \frac{1}{N} \sum_{i=1}^{N} [y(i) - \bar{y}(N)]^2
\]

\[
\sigma(N) = \sqrt{\sigma^2(N)}
\]

and $N$ is the total number of samples available, in this case 132. This yields values of $\bar{y}(N) = 471.9$ mm, $\sigma^2(N) = 26409$ mm$^2$ and $\sigma(N) = 162.5$ mm.

There is an implicit assumption in the formulae (2.1) that the true mean and variance are constant. If we trust the evidence of our eyes, we might suspect that this is not the case and that there are possibly changes in both the mean and variance over
Recursive Estimation: A Simple Tutorial Introduction

The computations involved in the calculation of $\bar{y}(N)$ and $\sigma^2(N)$ from (2.1) are extremely simple, so that the evaluation of these statistics for any subset of the $N$ samples is straightforward. If the statistics had been computed over some $k - 1$ samples within the data set, however, and then we wished to obtain them for $k$ samples, it seems reasonable to assume that there is a simple relationship between $\bar{y}(k - 1)$ and $\bar{y}(k)$. Similarly, if the estimates are available for $N$ samples and an additional sample is received, as might be the case in the present example when the total rainfall measurement for 2010 becomes available, then it might be suspected that some combination of the statistic already computed for $N$ samples and the new sample at $N + 1$ would yield the new value of the statistic at $N + 1$. 

Fig. 2.1 Rainfall at Walgett, N.S.W, 1879 to 2009, showing also the recursive estimate of the mean value generated by the algorithm (2.3).