2.10 Appendix 1. Maximum Likelihood

The estimation procedures used in this book are based primarily upon the idea of minimizing the variance of the estimate about the true value. Alternatives exist. For example, given a set of observations with known joint probability density, one can use a principle of "maximum likelihood." This very general and powerful principle attempts to find those estimated parameters which render the actual observations the most likely to have occurred. By way of motivation, consider the simple case of uncorrelated jointly normal stationary time series, x_i, where,

$$\langle x_i \rangle = m, \ \langle (x_i - m)(x_j - m) \rangle = \sigma^2 \delta_{ij}.$$

The corresponding joint probability density for $\mathbf{x} = [x_1, x_2, ..., x_N]^T$ can be written,

$$\begin{aligned} p_{\mathbf{x}}(\mathbf{X}) &= \frac{1}{(2\pi)^{N/2}\sigma^N} \times \\ & \exp\left\{ -\frac{1}{2\sigma^2} \left[(X_1 - m)^2 + (X_2 - m)^2 + ... + (X_N - m)^2 \right] \right\}. \end{aligned} \tag{2.450}$$

Substitution of the observed values, $X_1 = x_1, X_2 = x_2, ...$ into Eq. (2.450) permits evaluation of the probability that these particular values occurred. Denote the corresponding probability density as L. One can demand those values of m, σ rendering the value of L as large as possible. L will be a maximum if $\log(L)$ is as large as possible: that is we seek to maximize,

$$\log(L) = -\frac{1}{2\sigma^2} \left[(x_1 - m)^2 + (x_2 - m)^2 + ... + (x_N - m)^2 \right] + N \log(\sigma) + \frac{N}{2} \log(2\pi),$$

with respect to m, σ. Setting the corresponding partial derivatives to zero and solving produces,

$$\tilde{m} = \frac{1}{N} \sum_{i=1}^{M} x_i, \ \tilde{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \tilde{m})^2.$$

That is, the usual sample mean, and biassed sample variance maximize the probability of the observed data actually occurring. A similar calculation is readily carried out using correlated normal, or any random variables with a different probability density.

Likelihood estimation, and its close cousin, Bayesian methods, are general powerful estimation methods which can be used as an alternative to almost everything covered in this book.[59] Some will prefer that route, but the methods used here are more intuitive, and adequate for a very wide range of problems.