4
Nonparametric Inference—Genesis

4.1 Introduction

In doing statistical inference, if you do not have to make assumptions about the underlying population being sampled, then it is best to not do so. Record-breaking data often arise in industrial settings, and therefore it seems natural to use the exponential, extreme value, or Weibull distributions as models for the underlying populations of interest. However, in general, there is no guarantee that the correct parametric model is being used in order to make inferences from the data. So, there was a real need to develop nonparametric inference from such data.

Nonparametric inference from record values saw its birth in the paper by Foster and Stuart (1954). They proposed two very simple distribution-free statistics to test for randomness in a series of observations using upper and lower records. However, after that, work on all inference from records apparently came to a standstill, only to resume some twenty years later with work on prediction of future records. General nonparametric inference did not appear until 1988 when Samaniego and Whitaker published a paper on nonparametric maximum likelihood estimation from such data. They developed and studied the nonparametric MLE of the underlying distribution function and laid the foundation for further nonparametric work in this field. After that, there were several papers on smooth function estimation, and even on Bayesian nonparametric inference, from record-breaking data. Some of these results are presented in the following sections and chapters.

4.2 The Work of Foster and Stuart

The statistics proposed by Foster and Stuart (1954) are linear functions of the upper records (or of the lower records) obtained from a series of observations. To test for a trend in location, the proposed statistic is the difference of the number of upper records and the number of lower
records in a series. Similarly, to test for trend in dispersion, the proposed statistic is the sum of the number of upper and lower records in the series. It is intuitively clear that if there is a trend in location, there will tend to be more records in one direction than the other, and if there is an increasing (decreasing) trend in dispersion, there will be too many (or too few) records. Under the null hypothesis, Foster and Stuart provided a joint distribution of the statistics and showed that the statistics are uncorrelated and asymptotically normally distributed. We discuss their test statistics next.

Suppose as before that $Y_1, Y_2, \ldots, Y_n$ is a random sample from a continuous distribution function $F$ with density $f$. Define the following scores:

$$u_r = \begin{cases} 1, & \text{if the } r\text{th observation is an upper record} \\ 0, & \text{otherwise} \end{cases}$$

and

$$l_r = \begin{cases} 1, & \text{if the } r\text{th observation is a lower record} \\ 0, & \text{otherwise} \end{cases}$$

Furthermore, let

$$s_r = u_r + l_r$$

and

$$d_r = u_r - l_r.$$  

Then the test statistics are

$$s = \sum_{r=2}^{n} s_r, \quad \text{with } 0 \leq s \leq n - 1,$$

and

$$d = \sum_{r=2}^{n} d_r, \quad \text{with } -(n - 1) \leq d \leq n - 1. \quad (4.2.1)$$

Under the null hypothesis that the data are i.i.d., every ordering of a series of $r$ observations is equally probable. Hence,