ON THE STATISTICAL THEORY OF ANALYTIC GRADUATION

JAN M. HOEM
THE CENTRAL BUREAU OF STATISTICS OF NORWAY

1. Introduction; motivating examples

1.1. *Example* 1. *Rates of mortality*. Standard methods for the investigation of human mortality will produce statistics such as those given in extract in Table I. The mortality rate at age x is interpreted as a measure of the mortality risk for women born in the year 1968 minus x, and the corresponding number "exposed to risk" in Column 2 is used as a measure of the accuracy of this rate. (This will be made clearer later on.) Unless the population is substantially larger than the one producing these data. the diagram of the sequence of rates, plotted against age, will have a rather rugged appearance. Figure 1, based on the same data as Table I, shows the typical form of such diagrams. There seems to be a universal conviction, however, that "real mortality" would be portrayed by a smooth curve, and that any irregularities of curves of observed mortality rates are due to accidental circumstances. The observed rates are then regarded as "raw" or primary estimates of the underlying "real" rates, and graduation is employed to get a smoother curve.

A number of techniques have been developed to graduate age specific mortality rates, as can be seen from any text on the subject. (See, for instance, [55]; [59], pp. 145–197; [83], pp. 216–237, 243–244, and 251–252.) Most of these methods have been developed by intuitive arguments, at least initially, but investigations of statistical properties of some of them have also appeared [1]; [2]; [43]; [44]; [46]; [61]; [69]; [71]; [76]; [83], p. 252. One class of such methods consists in fitting a parametric function to the observed rates. We shall call this the class of *analytic graduation methods*.

Quite a number of functions have been suggested for analytic graduation of mortality rates [45], pp. 236–238; [67], pp. 453–454; [79], pp. 79–85; [83], pp. 56–60 and 243–244. By far the most commonly used for the adult ages is the Gompertz-Makeham formula

$$g_x(\alpha, \beta, c) = \alpha + \beta c^x \qquad \text{for} \quad \beta > 0, c > 1, \alpha > -\beta c^{x_{\min}}, \tag{1.1}$$

where x represents age attained. We have fitted this function to our data in

The present paper was written while the author visited the University of California, Berkeley, on a Ford Foundation fellowship.
This research was partially supported by National Science Foundation Grant GP 15283.