Lecture 12: Issues in Modelling

Slide 1

ISSUES IN MODELLING

The major issues in population studies are as follows:

- growth of population;
- changes in composition of population; and
- demographic processes such as nuptiality (i.e., marriage), fertility, mortality and migration.

Mathematical and statistical models have been used in studies of all the above issues. Among them as compared to nuptiality more attention has been paid to modelling of fertility and mortality.

PREDICTING GROWTH OF POPULATIONS AND SUBPOPULATIONS

Mathematical models have been used commonly to predict the size and composition of population of a country or a geographical region. For this purpose, various functions such as linear function, geometric or exponential growth function, modified exponential function, logistic curve, Makeham curve, Gompertz curve, polynomials, hyperbolic functions and autoregressive series have been used (Misra, 1980). Among them logistic curve has found more support than any other function on empirical grounds as well as the logic that the "population increase is proportional to the absolute population size already attained and the amount still left until the maximum, where the population becomes stationary" (UN, 1973).

Slide 2

Assuming that r is the maximum rate of increase of the population,

$$\frac{K-N}{K}$$

the fraction by which the actual population (N) remains below the maximum (K), then the increase of population per unit of time is

$$\frac{dN}{dT} = rN\frac{K-N}{K}$$

The total population will then be

$$N = \frac{K}{1 + be^{-rt}}$$

Figure 4.1 shows the various forms of logistic growth curve for different values of t and r (Weisstein, Eric W., 2009). In a typical logistic growth model, showing growth of population, N increases continuously from zero to a saturating level in a certain manner: initially when N is small the growth rate is also small but it continues to increase till a maximum point of growth rate is reached after which although N continues to rise but the rate of growth becomes smaller and smaller, ultimately reaching zero. Is this not a realistic assumption for population growth? Almost same thing can be said about degree of urbanization, spread of literacy and many other things in population studies. All these processes can be modelled using logistic curve.

Slide 3

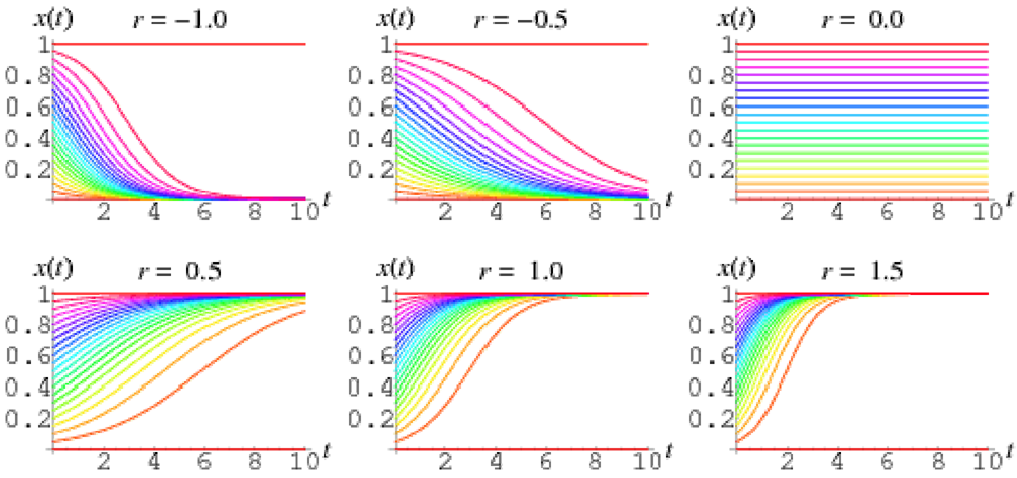

FIGURE 4.1: FORMS OF LOGISTIC GROWTH CURVE

In the above equation r is also called the Malthusian parameter (being rate of maximum population growth) and K is called the carrying capacity (i.e., the maximum sustainable population). Dividing both sides by K and defining x as N/K gives the differential equation

$$\frac{dx}{dt} = r x (1 - x),$$

Slide 4

Growth rate of a population following logistic curve has a definite pattern: rising from an extremely low level to a maximum level sometime and declining after that, gradually reaching zero. Logistic model has also been found to be of immense use in predicting subpopulations (Leach, 1981) because it provides a working model of the "mechanism of self correction". UN Manual XIII (UN, 1974) showed that a constant urban-rural growth difference (URGD) leads to logistic growth of the degree of urbanization (i.e., percent urban).

Subsequently, URGD method was used for projections of ratio of urban population to total population, and even other types of ratios such as ratios of populations of cities to total urban population, ratios of labour force to total population and school enrolment rates. Thus urban population can be predicted by multiplying projections of total population by projected ratios of urban population to total population. Projections of total population are obtained using component method (Smith, 1992).

ESTIMATION OF MORTALITY AND FERTILITY

In 1950's when the studies of population started in both developed and developing countries lack of data on the vital rates was a big problem for most countries. This was the best period for development of mathematical models in population studies. Mathematicians and statisticians developed models to estimate birth rate, death rate, total fertility rate and life expectancy using incomplete and unreliable data (Shryock et al., 1971; Roy, 1987). The main issue was: can we get working estimates of fertility and mortality using census growth rate, age distribution of population, and open and closed birth intervals found from survey data? Mathematical models helped in this greatly. For example, when reliable data on total fertility rate (TFR) did not exist but data on parity by age was available from surveys, using an empirical relationship Coale and Demeny suggested the following approximate formula:

$$TFR = \frac{P_3^2}{P_2}$$

Slide 5

Subsequently, Brass showed that if the fertility pattern can be described by a Gompertz function of the proportion experienced by each age, then the following is a better estimate of TFR than the above (Brass, 1979).

$$P_2\left(\frac{P_4}{P_3}\right)^4$$

With improvement in data sources and development of new sources such as Sample Registration Scheme need for such models declined.

Knud (1983) compared cancer mortality between sexes, cohorts and cities by using Poisson distribution for number of deaths at a particular age and the mortality rate (defined as chance of survival to age x) as follows:

$$l_x = b^{x^k}$$

b and k are two parameters for which maximum likelihood estimates were obtained.

Gompertz model has been commonly applied for studying mortality. Using an age-dependent shape parameter, Weon (2004) used a Weibull model for mortality rate μ(t), i.e., ratio of density (f(t)) and survival functions (S(t) = 1 – F(t)) for estimating maximum longevity, as follows:

$$S(t) = \exp(-t/\alpha)^{\beta(t)})$$

$$\mu(t) = (t/\alpha)^{\beta(t)} * [\frac{\beta(t)}{t} + \ln(t/\alpha) * \frac{d\beta(t)}{dt}]$$

Slide 6

He assumed that the typical human survival curves shows: a rapid decrease in survival in the first few years of life, a relatively steady decrease, and then an abrupt decrease near death.

In absence of data on measures of fertility they were derived from surveys. In India demographers derived parity progression ratios from survey data and used exponential model, displaced exponential model, Poisson model and life table approach for birth intervals (Pathak, 1989; Krishnamoorthy, 1989; Pathak and Ram, 1989).

PREDICTION AND ESTIMATION OF MIGRATION

Studies of migration aim at estimation, prediction and explanation of migration. In this context, for the first time Ravenstein published an article in *Journal of the Statistical Society* in 1885 in which he showed that, as believed by William Farr and many others, migration is not without laws. He developed some well known laws of migration, such as follows, which are still difficult to refute:

- The great body of migrants travel short distances.
- Women outnumber men in short distance migration.
- Migrants move from agricultural areas (places of dispersion) to industrial cities (places of absorption) followed by migration from centres of industrial cities to suburban areas and from remote areas to places of dispersion.
- Each migration current has a counter current with similar characteristics.
- The major causes of migration are economic.

Slide 7

Four years later in 1889 Ravenstein published another article in the *Journal of Royal Statistical Society*. This article was based on the experience of North America and Europe (the first article was based on the experience of UK). In this article he said that people travel long distances to occupy unsettled land. Nearly 100 years later, in 1940 Samuel A. Stouffer published an article in *American Sociological Review*. He showed that the number of migrants from place i to place j is inversely proportional to intervening opportunities. The model was confirmed by 1n 1975 by Wadycki who found it to be quite an accurate description of migration. Zipf (1946) developed and validated a gravity model to represent that volume of migration between two cities (M_{ij}) is directly proportion to populations of the cities (P_i and P_j) and inversely proportional to distance separating the two cities (d_{ij}) He assumed that income and unemployment are uniformly distributed over the areas.

$$M_{ij} = K (P_i * P_j)/D$$

where K is a constant to be found out from the empirical data. To some sociologists the model may look funny or absurd, but Zipf found that the model fitted very well for all modes of transportation.

After all, if the logic used in the model is sound that accurate data are available why would mathematical equations not produce a good fit to empirical data?

Slide 8

In 1975 Using a similar argument Dorigo and Tobler (1983) expressed:

$$M_{ij} = k(U_i W_i / U_j W_j) L_i L_j / d_{ij}$$

Also

$$R_i - E_j = k(U_i W_j / U_j W_i) L_i L_j$$

where M_{ij} refers to migration from place i to place j. R refers to rejecting, repelling, or repulsing, “push” away factors. E refers to enticing or “pull” toward factors. U, W and L refer to unemployment rate, wage rate and number of people in the labour market.

One can change the variables in the model, delete or add some, transform variables to mathematical or trigonometric form, but as long as the logic developed by Zipf in suggesting gravity models is correct the above types of models would remain relevant.

Slide 9

ECONOMIC-DEMOGRAPHIC MODELS

Coale and Hoover (1958) pioneered the modelling of demographic and economic variables through which they showed the short term and long term effects of population growth on economic development. Their model paved the way for large scale computer simulation models in population and development (Arthur and McNicoll, 1975).