Introduction to stochastic processes
Frans Willekens
19 October 2015

Overview

Actions of agents and interactions between agents cannot be predicted with certainty, even if we know a lot about an actor, his or her social network and the contextual factors that could trigger a need or desire to act. Decisions to act are made under uncertainty. Agent-based models (ABM) should account for the uncertainties and the impact of chance on decision outcomes. As a consequence, agent-based models should be probability models.

Random variables constitute the elementary building blocks of a probability model. Random variables may take on a finite number of values (discrete random variable) or an infinite number of values (continuous random variable). The likelihood of a value or range of values is expressed as probabilities. Each random variable is characterized by a probability distribution. A distinction is made between empirical (observed) distributions and theoretical distributions. The normal distribution, the exponential distribution, the binomial/multinomial distribution and the Poisson distribution are common probability distributions.

The waiting time to an action or interaction is a random variable, characterized by a waiting time distribution. The exponential distribution, the Gompertz distribution, the extreme value distribution and the gamma distribution are used regularly in demography. The outcome of an action is a random variable too. If the outcome is a continuous variable (e.g. reward), possible values are described by a probability density function. If the number of possible outcomes is finite, which is often the case in demography and social sciences, the random variable is discrete and the distribution of the likelihood of each value is the probability mass function. Models that combine waiting time to an action and outcome of that action are competing risks models. In the discrete choice literature, they are called horse race models.

You learn to view reality from a probabilistic perspective and to distinguish choice and chance. The 'facts' we observe are products of underlying causes and chance. They are realizations or manifestations of random processes. You learn to:

- Define random variables (different types)
- Assign probabilities to outcomes and determine probability distributions
- Approach microsimulation as sampling from distributions (empirical or theoretical), i.e. assign values to random variables (random number generation). The workhorse of microsimulation is the inverse distribution function (quantile distribution).
    - Popular examples of quantile functions: probit and logit models
- Describe the distribution function of a sum of random waiting times as a convolution of waiting time distributions. Many distributions we observe, e.g. age patterns of fertility and migration, are convolutions of several waiting time distributions.
- Approach a collection or sequence of random variables over a suitable index set as a stochastic process $\{Y_t ; t \in T\}$. Stochastic processes are characterized by:

- The range of possible values for the random variable (*state space*)
    - The index set T
    - The dependence relations among the random variables Y_t
- Model demographic processes as stochastic processes (accounting for individual-level uncertainties). Common stochastic processes:
    - Poisson process
    - Markov process
    - Random walk and Wiener process (elementary diffusion process)
- Associate utilities or rewards with outcomes (of random variables). The utility associated with an outcome is a random variable. The utility depends on attributes of the actor and attributes of the option or alternative, and random factors. Random factors follow a probability distribution.
- Approach decision-making under uncertainty as choosing the outcome of a random process that gives the highest reward (random utility discrete choice model). The highest reward depends on the distribution of possible outcomes of an action and the distribution of rewards associated with outcomes.
-
- Adopt a probabilistic perspective on social network dynamics. Social networks change as a result of formation and termination of ties between agents or actors (ego i and alter j). Network ties are states; the possible ties constitute the state space. The current state of a network is described by a state vector. The sequence of state vectors is a stochastic process.
    - Actors have attributes
    - Actors may change attributes and network ties (formation and termination of ties (outgoing ties))
    - Time between two changes (of attribute or network tie) is a random variable. The time between the formation and termination of a tie is a waiting time, characterized by a waiting time distribution.
    - Actors derive satisfaction (reward; utility) from a social network. Satisfaction is a random variable, which depends on (a) actor's attributes and the network ties, and (b) random noise, which follows a probability distribution.
    - Actors choose a network configuration that maximizes satisfaction
        - Snijders and Steglich (2013) assume that random noise follows a Type I extreme value distribution. Hence a logistic distribution describes the probabilities of given changes in attributes or network ties. Their model incorporates a random utility discrete choice model

Outline lecture

1. Introduction
    a. Deterministic vs stochastic
    b. Concept of random variable
    c. Indicator function of random variable
    d. Probability distribution of random variable
2. Random variables and their probability distributions
    a. Empirical distribution
    b. Theoretical distribution
        - Continuous random variable

        - Normal
        - Logistic
        - Exponential
        - Weibull
        - Gompertz
        - Gamma
    - Discrete
        - Bernoulli
        - Binomial
        - Geometric

3. Sum of random variables: convolution
4. Sequence of random variables: stochastic process
    a. Poisson process
        - Simple Poisson process
        - Compound Poisson process
        - Marked Poisson process
    b. Markov process
    c. Random walk and related stochastic processes
5. Discrete choice model
    a. Choice theory and decision rules
    b. Random utility theory
    c. Random utility discrete choice models
        - Closed-form expression of choice outcome
        - No closed-form expression of choice outcome: simulation
        - Relation between unobserved part of utility and the choice model
    d. Illustration: migration: choice of destination
    e. Random utility discrete choice models with a accumulation of evidence: the horse-race model (and its relation to theory of competing risks)

Textbooks on stochastic processes and stochastic modeling (for information only)

Taylor, H.M. and S. Karlin (2011) An introduction to stochastic modeling. Academic Press (Elsevier). 4th edition. A very good textbook.

Çinlar, E. (2013) Introduction to stochastic processes. Dover Publishing (Dover Books on Mathematics; reprint edition 2013). First edition published by Prentice-Hall in 1975. A very good book, but less accessible to social scientists as Taylor and Karlin. It is the text used by Cinlar in the course I took when I was a graduate student at Northwestern University in 1974.

Mode, C.J. (1985) Stochastic processes in demography and their computer implementation. Springer (Biomathematics) Advanced text.

Namboodiri, N.K. (1990) Demographic analysis: a stochastic approach. Academic Press. An introduction to demography from a probabilistic point of view.

Biswas, S. and V.K. Sehgal (1988) Stochastic processes in demography and applications. Wiley Eastern, New Delhi. New edition: S. Biswas and G.L. Sriwastav (2011) Published by New Central Book Agency, Calcutta. A good book, but for mathematicians and not for demographers.

> "This textbook attempts to fill a "gap in the research and pedagogy of Mathematical Demography and other related topics in Survival Analysis by focusing on a wide range of traditional as well as new inputs using a modern Stochastic Process and Renewal Theory oriented approach. Special topics such as Martingales theory, Cox's regression model, Parametric and Non-parametric techniques in Survival theory"

Courgeau, D. (2012) Probability and social sciences. Springer. A historical and philosophical description of the co-evolution of probability theory and population studies.

Useful websites

Online courses in probability theory:
http://education-portal.com/articles/List_of_Free_Online_Probability_Courses_and_Tutorials.html

http://ocw.mit.edu/courses/mathematics/18-175-theory-of-probability-fall-2008/

There are several courses on YouTube, e.g. the Harvard Statistics 110 course on probability theory by professor Joe Blitzstein (http://www.people.fas.harvard.edu/~blitz/Site/Home.html):

https://www.youtube.com/playlist?list=PL2SOU6wwxB0uwwH80KTQ6ht66KWxbzTIo

Gallery of distributions
http://www.itl.nist.gov/div898/handbook/eda/section3/eda366.htm
https://en.wikipedia.org/wiki/List_of_probability_distributions