MVE550 2019 Lecture 1
Introduction to stochastic processes
Course introduction

Petter Mostad

Chalmers University

November 6, 2019

Outline

- Introduction to stochastic processes
- Introduction to Bayesian inference
- Course structure and course content
- Review: Dobrow Chapter 1, appendices A, B, C, D
- Conditional probability and conditional expectation.

Stochastic (or probabilistic) models

- Example: A random variable X that is normally distributed with expectation 14.7 and standard deviation 2.3 models the braking distance of a car at a certain speed.
- Note: The prediction is a *probability distribution*
- An *equivalent* representation of the model: A computer program which em simulates predicted values (for example braking distances).
- Monte Carlo simulation:

$$\text{Frequency of computer output} \approx \text{Probability of output}$$

- Note: The prediction need not be a single number, it can be
  - A vector of numbers
  - An image (represented by numbers in a grid)
  - A 3D model of a building (represented by numbers at points)
  - An infinite sequence of numbers
  - A continuous function from $[0, 1]$ to real numbers.
  - ...

Stochastic processes

A stochastic process is a collection of random variables $\{X_t, t \in I\}$ where

- the *index set* I can be for example $\{0, 1, 2, \ldots\}$ or $[0, \infty)$.
- The random variables are defined on a *common state space* S.
- Example: $I = \{0, 1, \ldots\}$ and S is finite.
- Example: $I = \{0, 1, \ldots\}$ and $S = \{0, 1, \ldots\}$.
- Example: $I = \{0, 1, \ldots\}$ and $S = \mathbb{R}$.
- Example: $I = [0, \infty)$ and S is finite.
- Example: $I = [0, \infty)$ and $S = \mathbb{R}$.
- NOTE: The random variables are generally not independent!

Review: Random variables

A random variable X with state space S is a real-valued function on S together with a *probability* $\Pr(\cdot)$ on S. The probability $\Pr(\cdot)$ satisfies

- $0 \leq \Pr(A) \leq 1$ for all *measurable* subsets $A \subseteq S$.
- $\Pr(S) = 1$
- $\Pr(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \Pr(A_i)$ when the A_i are disjoint.

- These are the Kolmogorov axioms for probability.
- Measurable sets are called *events*.
- What is a *measurable* subset?

Measurable subsets

Let S be any set.

- A *sigma algebra* Ω on S is a set of subsets of S such that
  - Ω includes S
  - If $A \in \Omega$ then $A^c = S \setminus A \in \Omega$.
  - If $A_1, A_2, \ldots, \in \Omega$ then $\cup_{i=1}^{\infty} A_i \in \Omega$
- The *measurable sets* are those that are in an appropriately defined sigma-algebra.
- What you need to know for this course: When S is finite or countable, all subsets will be measurable. When S is some interval of real numbers, there will exist subsets that are not measurable, but we will not be concerned with them.

Inference

- We have shown some stochastic processes which we would like to use as scientific *models*.
- In some examples, stochastic models can be specified based on what is *reasonable*.
- In *most* real applications, stochastic models have parameters that need to be learned from *data*. This learning process is called *inference* (svenska: slutledning).
- In this course, we have added *Bayesian inference* to the material in Dobrow.
- More about Bayesian inference in Lecture 2.

Course components

- Lectures
- Exercise classes
- Three obligatory assignments
- Written exam

Learning material

- Dobrow (e-book)
- Lecture Notes (on Canvas course homepage)
- Insua et al: reference material (e-book)
- Overheads from lectures (on Canvas)
- Old exams with solutions (on Canvas)

Conditional probability

- Given events A and B, the *conditional probability* for A given B is

$$\Pr(A \mid B) = \frac{\Pr(A, B)}{\Pr(B)}$$

- Events A and B are *independent* if $\Pr(A, B) = \Pr(A)\Pr(B)$.
- Law of total probability: Let $B_1, \ldots, B_k$ be a sequence of events that *partitions* S. Then

$$\Pr(A) = \sum_{i=1}^{k} \Pr(A \cap B_i) = \sum_{i=1}^{k} \Pr(A \mid B_i) \Pr(B_i).$$

Conditional distributions

- Discrete case:

$$\Pr(Y = y \mid X = x) = \frac{\Pr(X = x, Y = y)}{\Pr(X = x)}$$

- Continuous case:

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)}$$

- Common notational convention (used in Lecture Notes): Use π as generic function:

$$\pi(y \mid x) = \frac{\pi(x, y)}{\pi(x)}$$

Conditional expectation

- Recall, the expectation of a discrete random variable is

$$\mathrm{E}(Y) = \sum_y y \Pr(Y = y)$$

and of a continuous random variable

$$\mathrm{E}(Y) = \int_y y f_Y(y)\, dy.$$

- The conditional expectation in the discrete case is

$$\mathrm{E}(Y \mid X = x) = \sum_y y \Pr(Y = y \mid X = x)$$

and in the continous case

$$\mathrm{E}(Y \mid X = x) = \int_y y f_{Y|X}(y \mid x)\, dy.$$

- Note that if we use the π notation and interpret integrals of discrete variables as sums, we can write both equations as

$$\mathrm{E}(Y \mid X = x) = \int_y y \pi(y \mid x)\, dy.$$

Law of total expectation

- If $A_1, \ldots, A_k$ are a sequence of events partitioning S then

$$\mathrm{E}(Y) = \sum_{i=1}^{k} \mathrm{E}(Y \mid A_i) \Pr(A_i)$$

- More generally, for a random variable X we have

$$\mathrm{E}(Y) = \int_x \mathrm{E}(Y \mid X = x)\, \pi(x)\, dx$$

- This can also be written as

$$\mathrm{E}(Y) = \mathrm{E}(\mathrm{E}(Y \mid X)).$$

Law of total variance

- Recall that, by definition,

$$\mathrm{Var}\,(Y) = \mathrm{E}\left((Y - \mathrm{E}\,(Y))^2\right) = \mathrm{E}\left(Y^2\right) - \mathrm{E}\,(Y)^2.$$

- If X is another random variable we get

$$\mathrm{Var}\,(Y) = \mathrm{E}\,(\mathrm{Var}\,(Y \mid X)) + \mathrm{Var}\,(\mathrm{E}\,(Y \mid X))$$