Probability and Stochastic Processes I - Lecture 10

Michael Evans
University of Toronto
http://www.utstat.utoronto.ca/mikevans/stac62/STAC622023.html

2023

II.6 Stochastic Processes

Definition II.6.1 The set $\{(t, X_t) : t \in T\}$, where X_t is a random variable defined with respect to probability model $(\Omega, \mathcal{A}, P)$ for each $t \in T$, is called a *stochastic process* (sometimes a *random process*). The set T is called the *index set* of the process.

Example II.6.1 - a single random variable X_1 with $T = \{1\}$ is equivalent to the stochastic process $\{(1, X_1)\}$

- a random vector $\mathbf{X} = (X_1, \ldots, X_k)'$ is equivalent to the stochastic process $\{(t, X_t) : t \in T\}$ where $T = \{1, \ldots, k\}$ ■

note - T can be a very general set like the nodes of a graph and can be an infinite set

Example II.6.2 *Tossing a coin until the first head is observed.*

- suppose a coin is tossed in "independent" tosses until the first head is observed and the number of that toss is recorded

- what is Ω?

- denote a head by 1 and a tail by 0

- it is most convenient to put

$$\begin{aligned} \Omega &= \{(\omega_1, \omega_2, \ldots) : \omega_i \in \{0,1\}\} = X_{i=1}^{\infty}\{0,1\} \\ &= \text{ the set of all sequences of 0,s and 1's} \end{aligned}$$

- then define $Y : \Omega \to R$ by $Y(\omega) = i$ when $\omega_i = 1$ and $\omega_j = 0$ when $j < i$

- is Y well defined as a r.v.?

- let $p =$ probability of a head on a single toss

- if $p = 0$ then Y is not defined

- if $p > 0$, then Y is defined since

$$\begin{aligned} P(\text{"an infinite sequence of tails"}) &= \lim_{n\to\infty} P(\text{"first } n \text{ are tails"}) \\ &= \lim_{n\to\infty} (1-p)^n = 0 \end{aligned}$$

- then putting $A_i = \{\omega : \omega_1 = 0, \ldots, \omega_{i-1} = 0, \omega_i = 1\}$ and using independence

$$\begin{aligned} & p_Y(i) = P_Y(\{i\}) = P(A_i) \\ = \;& (1-p)^{i-1} p P(\{\omega : \omega_{i+1} \in \{0,1\}, \omega_{i+2} \in \{0,1\}, \ldots\}) \\ = \;& (1-p)^{i-1} p \end{aligned}$$

and $Y \sim$ geometric(p)

Exercise II.6.1 Prove that p_Y defines a probability distribution.

- the point is that we need an infinite dimensional Ω

- if we define $X_i(\omega) = \omega_i$, then $\{(t, X_t) : t \in \mathbb{N}\}$ is a stochastic process, called a Bernoulli(p) process ■

note - the main reason we need to consider stochastic processes is that in many applications the dependence on the index is important

Example II.6.3

- suppose $\mathbf{X} = (X_1, \ldots, X_k)'$ gives various measurements of an individual's face (for facial recognition purposes)

- then the order in which the measurements is listed doesn't matter as long as we are consistent so there is no real dependence on the index

- but now suppose $\mathbf{X}$ contains the measurements from a physical endurance test taken on an individual one month apart where i refers to month i and then the order definitely matters and we would take this into account when analyzing data

- in many contexts $t \in T \subset R^1$ refers to time as for example taking $T = \{0, 1, 2, \ldots\} = \mathbb{N}_0$ and X_t is the closing price of a stock on trading day t

- sometimes $T = \mathbb{Z}$ and negative times index the past

- now some stocks trade by nanoseconds so it makes sense in such a case to take, as an approximation, $T = [0, \infty)$ or $T = R^1$ which is continuous time as opposed to discrete time

- stochastic processes where t is time are often referred to as *time series*

- sometimes t is multidimensional as $t = (t_1, t_2) =$ (longitude, latitude) of a point on earth and $X_{(t_1,t_2)}$ measures some characteristic of that location such as its mean temperature over a particular year and such processes are referred to as *spatial processes*

- in fact to study global warming we want $t = (t_1, t_2, t_3) =$ (longitude, latitude, time) for a large grid covering the earth and for years extending into the past and future ■

- think of a realized value of a stochastic process as a function $X_\cdot(\omega) : T \longrightarrow R^1$, called the *sample function*, with value $X_t(\omega)$ at index t

- so a stochastic processes is in effect a probability measure on functions from T into R^1

- a stochastic process is a generalization of random variable where we start with $(\Omega, \mathcal{A}, P)$ and get P_X as follows

- $X : \Omega \longrightarrow R^T$ = set of functions with domain T mapping into R^1
$= \{g : g : T \longrightarrow R^1\}$

- $\mathcal{B}^T$ = smallest σ-algebra on R^T containing all sets of the form

$$\{g : g(t_1) \in (a_1, b_1], \ldots, g(t_n) \in (a_n, b_n]\}$$

for any $\{t_1, \ldots, t_n\} \subset T$ and intervals $(a_1, b_1], \ldots, (a_n, b_n]$

- then require $X^{-1}B \in \mathcal{A}$ for any $B \in \mathcal{B}^T$ which implies

$$P_X(B) = P(X^{-1}B)$$

Example II.6.4

- suppose $T = [0, \infty)$, $\omega \sim N(0, 1)$ and define $X_t(\omega) = \omega t$ so X gives a ray from the origin with $N(0, 1)$ distributed slope

- suppose $T = [0, 1]$, $\omega \sim \text{Uniform}(0, 10)$ and define $X_t(\omega) = \cos(\omega t)$ so X gives a cosinusoid with random frequency

- the definition of a stochastic process immediately implies that the finite dimensional distributions are consistent in the sense that the distribution of $(X_{s_1}, \ldots, X_{s_m})$ can be obtained from that of $(X_{t_1}, \ldots, X_{t_n})$ by marginalizing whenever $\{s_1, \ldots, s_m\} \subset \{t_1, \ldots, t_n\}$

- as a general way of defining stochastic processes we have the following

Proposition II.6.1 (*Kolmogorov's Consistency Theorem*)
Suppose $T \subset R$ and a probability model $(R^n, B^n, P_{(t_1,\ldots,t_n)})$ is given for each $\{t_1, \ldots, t_n\} \subset T$. If the probability models are consistent, then there exists probability model $(\Omega, \mathcal{A}, P)$ and random variables X_t such that $\{(t, X_t) : t \in T\}$ is a stochastic process with $P_{(X_{t_1},\ldots,X_{t_n})} = P_{(t_1,\ldots,t_n)}$.

Example II.6.5 *Bernoulli(p) process*

- let $T = \{1, 2, \ldots\}$ and $P_{(t_1,\ldots,t_n)}$ be the discrete probability measure concentrated on $\{0, 1\}^n$ given by the probability function

$$p_{(t_1,\ldots,t_n)}(x_1, \ldots, x_n) = \begin{cases} \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} & \text{if } (x_1, \ldots, x_n) \in \{0,1\}^n \\ 0 & \text{otherwise} \end{cases}$$

- these distributions are clearly consistent and so KCT indicates that this provides a valid definition of a s.p. $\{(t, X_t) : t \in T\}$ ■

Definition II.6.2 A s.p. $\{(t, X_t) : t \in T\}$ is a *Gaussian process* whenever

$$(X_{t_1}, \ldots, X_{t_n}) \sim N_n(\mu(t_1, \ldots, t_n), \Sigma(t_1, \ldots, t_n))$$

where $\mu(t_1, \ldots, t_n) \in R^n, \Sigma(t_1, \ldots, t_n) \in R^{n \times n}$ p.d. for every $\{t_1, \ldots, t_n\} \subset T$.■

- does a Gaussian process exist?

Example II.6.6 *Gaussian white noise process*

- specify $\sigma^2 : T \to (0, \infty)$ and for $\{t_1, \ldots, t_n\} \subset T$

$$\begin{aligned} \mu(t_1, \ldots, t_n) &= (0, \ldots, 0)' \\ \Sigma(t_1, \ldots, t_n) &= \text{diag}(\sigma^2(t_1), \ldots, \sigma^2(t_n)) \end{aligned}$$

- then putting $(X_{t_1}, \ldots, X_{t_n}) \sim N_n(\mathbf{0}, \text{diag}(\sigma^2(t_1), \ldots, \sigma^2(t_n)))$ defines a valid s.p. (**Exercise II.6.2**) ■