Recitation 1: Probability Review and an Introduction to Stochastic Processes

Contact: Amer Al-Hiyasat, office 6C-433

hiyasat@mit.edu, OH: 11-12pm in 4-146, day before pset due.

Probability spaces and random variables

A probability space is a triple: $(\Omega, \mathcal{F}, \mathbb{P})$

Ω ↑ sample space, $\mathcal{F}$ ↑ event space, $\mathbb{P}$ ↑ Probability measure

Ω: The sample space: set of realizations of an experiment

e.g. For a dice, $\Omega = \{1, 2, \dots, 6\}$

e.g. For a darts board, Ω = Disc in $\mathbb{R}^2$.

$\mathcal{F}$: Event space: a set of subsets of Ω satisfying

σ-algebra:

i) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ (closed under unions)

ii) $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$, more generally $\{A_1, A_2, \dots\} \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ (closed under countable unions)

e.g. For a dice, the set of all subsets of Ω

e.g. For $\Omega = \mathbb{R}$, the set of open and closed intervals and countable unions thereof.

$\mathbb{P}$: Probability measure: a set function $\mathbb{P}: \mathcal{F} \to [0, 1]$ satisfying

i) $\mathbb{P}(A) \geq 0$, $\mathbb{P}(\phi) = 0$

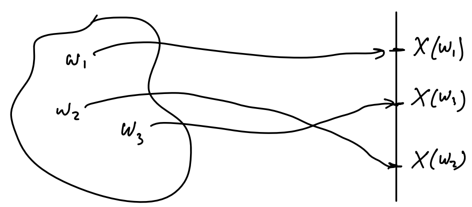
ii) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ if $A \cap B = 0$. More generally: $\mathbb{P}(\cup_i A_i) = \sum_i \mathbb{P}(A_i)$ for a countable sequence in $\mathcal{F}$.

ii) $\mathbb{P}(\Omega) = 1$.

A random variable is a measurable function of the sample space $X: \Omega \to \mathbb{R}$

↳ $X^{-1}(A) \in \mathcal{F} \;\; \forall A \subseteq \mathbb{R}$.

(more generally, $X: \Omega \to S$ for some measurable space E, though we typically have $E \subseteq \mathbb{R}$).



An expectation value is an integral of X w.r.t $\mathbb{P}$:

$$\langle X \rangle = \mathbb{E}[X] = \int X \, d\mathbb{P}$$

What does $\int d\mathbb{P}\, X$ mean? Roughly, sum over all outputs of X weighted by the probabilities of the pre-images. For discrete Ω this is just

$$\langle X \rangle = \sum_{\omega\in\Omega} X(\omega)\, \mathbb{P}(\omega)$$

For continuous Ω, we often need a probability density $d\mathbb{P} = p(x)\, dx$, defined to satisfy:

$$\mathbb{P}[X\in A] = \int_A dx\; p(x)$$

Things you should know/review for this class:

Independence; conditioning; Joint, conditional, and marginal probability densities; conditional expectation; Moments and cumulants; Characteristic functions, cumulant generating functions.

Stochastic Process

A stochastic process is a collection of random variables indexed by some index set T:

$$Y = \{X_t : t\in T\}$$

Often, t refers to time, though it could refer to space or something more complicated.

$\Rightarrow$ A stochastic process is discrete if T is countable. Often, $T=\mathbb{N}$, and T has an ordering.

e.g.
- # of letters arriving in a mailbox each day;
- value of a stock when the market closes each day;
- configurations of a set of spins on a lattice ($T=\mathbb{Z}^d$)

$\Rightarrow$ A continuous stochastic process has a continuous index set, typically $T=\mathbb{R}^+$, written $\{X(t)\}_{t\geq 0}$.

e.g.
- # of letters in a mailbox at any point in time;
- value of a stock during market hours;
- configurations of a field over space (a "Random field" with $T=\mathbb{R}^d$)

An equivalent interpretation of the definition given above is the following:

$\Rightarrow$ While a random variable assigns to each $\omega\in\Omega$ a number $X(\omega)\in\mathbb{R}$, a stochastic process assigns to each $\omega\in\Omega$ a function $T\longrightarrow\mathbb{R}$

Discrete: $\quad Y(\omega) = (x_1, x_2, \dots)$

Continuous: $\quad Y(\omega) = x(t), \quad$ where $x: \mathbb{R}^+ \to \mathbb{R}$.

The theory of stochastic processes is thus a theory of Random Functions that map $T \to \mathbb{R}$. A particular realization $Y(\omega)$ is called a sample path or a trajectory.

To specify a stochastic process requires us to specify the joint probability density of every possible sequence $\{(X(t_1), X(t_2), \ldots, X(t_p)) : t_1, \ldots, t_p \in T\}$. We write this

$$P(x_1, t_1 ; x_2, t_2 ; \cdots ; x_p, t_p) \qquad (*)$$

Physical observables are functions of a trajectory. To calculate expectation values of observables we ask:

Q: Can we assign a probability density to trajectories of a given length?

A: ⇒ In the discrete case, this is easy. The probability density of a path of length n is simply the joint density of the n random variables $p(\{x_i, t_i\})$.

Expectation values of functions of Y are calculated as

$$\langle f(Y) \rangle = \int d\mathbb{P}\, f(Y) = \int \prod_{i=1}^{n} dx_n \, p(\{x_i, t_i\}) \, f(\{x_i, t_i\})$$

prob. dens. of trajectory.

$p(\{x_i, t_i\})$ is a density w.r.t to the standard (Lebesgue) measure on $\mathbb{R}^n$.

⇒ In the continuous case, this is tricky. The probability weight of a trajectory $\{x(t) \mid t \in (0, T)\}$ is a functional of $x(t)$, written $\mathcal{P}[x(t)]$

Expectation values of functionals of Y are obtained as

$$\langle F[Y] \rangle = \int F[Y] \, d\mathbb{P} \stackrel{?}{=} \int \mathcal{D}[x(t)] \, \mathcal{P}[x(t)] \, F[x(t)]$$

$\mathcal{P}[x(t)]$ is a density w.r.t to the so called "path integral measure", which sums over all possible paths $x(t)$. Except in a few simple cases, this is not mathematically well defined, but the Physics is ahead of the math on this one. We will talk about these later in the course, but often densities of finite subsequences like $(*)$ will suffice for many calculations.

Markov Processes

A defining feature of a stochastic process is the dependence structure among the various $\{X_t\}$. A very important class of processes are the so called Markovian processes. A discrete stochastic process $Y = \{X_t : t = 0, 1, \dots\}$ is said to have the Markov property if

$$\mathbb{P}[X_{t+1} \in A \mid X_t, X_{t-1}, \dots, X_0] = \mathbb{P}[X_{t+1} \in A \mid X_t]$$

where $\mathbb{P}[\cdot \mid \cdot]$ denotes a conditional probability. This means that the future value of a stochastic process depends only on its value at present. Conditioning on further past values of X provides no additional information.

In terms of conditional probability densities:

$$p(x_{t+1} \mid x_t, \dots, x_0) = \underline{p(x_{t+1} \mid x_t)} \longrightarrow \text{transition probability density}$$

so that the joint density of a trajectory is found by chaining the transition probs:

$$p(x_1, x_2, \dots, x_t) = p(x_t \mid x_{t-1})\, p(x_{t-1} \mid x_{t-2}) \cdots p(x_1 \mid x_0).$$

Warning: this does not imply that X_t and X_{t-2} are independent: they are only conditionally independent given X_{t-1}. Indeed:

$$p(x_t, x_{t-2}) = \sum_{x_{t-1}} p(x_t \mid x_{t-1})\, p(x_{t-1} \mid x_{t-2})\, p(x_{t-2}) \neq p(x_t)\, p(x_{t-2}).$$

i.e. if I know x_{t-1}, telling me x_{t-2} gives me no more info about x_t. But if I don't know x_{t-1}, telling me x_{t-2} does give information.

The generalization to continuous-time stochastic processes is conceptually immediate but involves some technicalities which will be treated in lecture.

Examples

1) The Bernoulli Process.

$B = \{X_t \mid t \in 0, 1, \dots\}$, where the $\{X_t\}$ are i.i.d. with

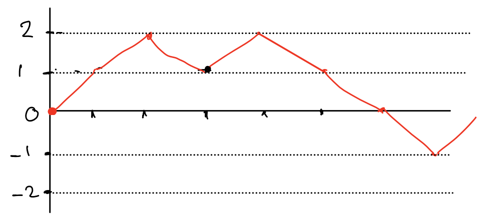
$$p(X_t = x) = \begin{cases} p & \text{for } x = 1 \\ 1-p & \text{for } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

i.e. a sequence of independent biased coin flips. This is a Markov process, but a trivial one because $p(x_t \mid x_{t-1}) = p(x_t)$.

2) The (discrete) Random Walk on $\mathbb{Z}$.

A Markov process $R = \{X_t \mid t \in 0, 1, \dots\}$ where

$$\mathbb{P}[X_t = x \mid X_{t-1}] = \begin{cases} 1/2 & \text{if } x = X_{t-1} \pm 1 \\ 0 & \text{otherwise} \end{cases}$$



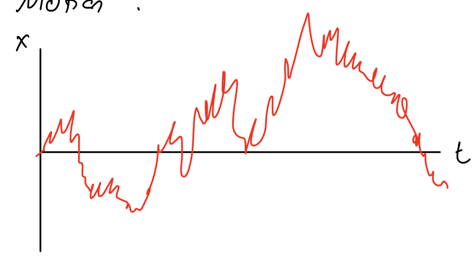
3) The Wiener Process, or "Standard Brownian Motion".

A process $\{W(t) : t \geq 0\}$ having

i) Continuous paths

ii) Stationary, independent increments

iii) $W(t) \sim N(0, t) \;\; \forall\, t \geq 0$



We write $W(t)$, but really it is $W: \mathbb{R}^+ \times \Omega \longrightarrow \mathbb{R}$, so that a sample path is $W(\omega, t)$

What does each of these mean?

i) Continuous paths

$$\mathbb{P}\{\omega \in \Omega : W(\cdot, \omega) \text{ is a continuous function}\} = 1.$$

ii) Independent increments: for any choice $0 \leq t_0 < t_1 < \dots < t_n < \infty$, the random variables $W(t_1) - W(t_0)$, $W(t_2) - W(t_1)$, ..., $W(t_n) - W(t_{n-1})$ are independent.

Stationary increments: $W(t_2) - W(t_1) \sim W(t_2 + s) - W(t_1 + s)$.

where $X \sim Y$ indicates that X and Y have the same distribution.

iii) $W(t) \sim N(0, t)$: $\mathbb{P}[W(0) = 0] = 1$. $\langle W(t)^2 \rangle = t$. $\langle W(t)^{n>2} \rangle_c = 0$.

What are some consequences of this?

a) Gaussian Increments

$$W(t) - W(s) \sim W(t-s) - W(0) = W(t-s) \sim \mathcal{N}(0, t-s)$$

b) A useful alternative definition: First note that for any finite set of times $t_1, \dots, t_n$, the random vector $(W(t_1), \dots, W(t_n)) \in \mathbb{R}^n$ has a joint normal distribution. This means that any linear combination is Gaussian:

$$\sum_{i=1}^{n} a_i W(t_i) \text{ is Gaussian } \forall \{a_i\} \in \mathbb{R}, \quad \{t_i\} \in \mathbb{R}^+.$$

Any stochastic process with this property is called a Gaussian process. Any such process is completely specified by its mean (zero for Wiener process) and covariance function

$$r(t_1, t_2) = \mathrm{Cov}(W(t_1), W(t_2)) \equiv \langle W(t_1) W(t_2) \rangle_c \equiv \langle W(t_1) W(t_2) \rangle - \langle W(t_1) \rangle \langle W(t_2) \rangle$$

Let us compute $r(t, t+s)$ for $s > 0$:

$$\begin{aligned} r(t, t+s) &= \langle W(t)\, W(t+s) \rangle_c \\ &= \langle W(t)\, (W(t+s) - W(t) + W(t)) \rangle_c \\ &= \langle W(t)^2 \rangle_c + \langle [W(t) - W(0)][W(t+s) - W(t)] \rangle_c \\ &= t \end{aligned}$$

where we have used the independence of increments. For general $t_1 \neq t_2$, we thus have

$$\begin{aligned} \langle W(t_1) W(t_2) \rangle_c &= \langle W(\min(t_1, t_2))\, W(\max(t_1, t_2)) \rangle_c \\ &= \langle W(\min(t_1, t_2))\, W(\min(t_1, t_2) + s) \rangle_c, \quad \text{where } s = |t_2 - t_1| \\ &= \boxed{\min(t_1, t_2)} \end{aligned}$$