Independence of random variables:

We say that a collection of random variables \( \{ X_n \}_{n \in \mathbb{I}} \) is independent if for any finite subset \( J \subseteq \mathbb{I} \)

\[
P \left( \bigcap_{n \in J} X_n \in B_n \right) = \prod_{n \in J} P \left( X_n \in B_n \right)
\]

One important consequence of independence of random variables is that:

\[
\mathbb{E} \left( \prod_{n \in J} y_n(X_n) \right) = \prod_{n \in J} \mathbb{E} y_n(X_n)
\]

when \( \{ X_n \}_{n \in \mathbb{I}} \) are independent \( J \subseteq \mathbb{I} \)

The essence is that independence of random variables often allows a calculation involving multiple random variables to be expressed in terms of calculations just involving one random variable at a time.

Generating and characteristic functions:

For a discrete random variable \( X \) we define its

Probability generating function:
Characteristic function: (for any random variable $X$)

$$
\phi_X(k) = \mathbb{E} e^{ikX} = \begin{cases} \\
\sum_{n \in \mathbb{N}} e^{ikn} p(X=n) & \text{discrete } X \\
\int_{-\infty}^{\infty} e^{ikx} \rho_X(x) \, dx & \text{continuous } X \\
\end{cases}
$$

The utility of probability generating functions and characteristic functions is that they provide a one-to-one transformation of information about the random variable, meaning that the probability generating function and characteristic function have full information about the random variable. So what's the point?

These transforms sometimes provide more convenient ways to analyze random variables than the original probability distribution or probability density.

For example, moments can be easily computed from the characteristic function:

$$
\mathbb{E} X^n = \left( -i \frac{d}{dk} \right)^n \phi_X(k) \bigg|_{k=0}
$$

$$
= \left( -i \frac{d}{dk} \right)^n \mathbb{E} e^{ikX} \bigg|_{k=0}
$$
The more important application for this class is that probability generating functions and characteristic functions work very well with sums of independent random variables. (whereas doing this directly with probability distributions can be painful)

Consider

\[ Z = \sum_{j=1}^{n} X_j \]

\[ \mathcal{Z}(k) = \mathbb{E} e^{ikZ} = \mathbb{E} e^{ik \sum_{j=1}^{n} X_j} \]

\[ = \mathbb{E} \left( \prod_{j=1}^{n} e^{ikX_j} \right) \text{ independence} \]

\[ = \prod_{j=1}^{n} \mathbb{E} e^{ikX_j} \]

\[ = \prod_{j=1}^{n} \mathcal{X}_j(k) \]

(And if the \( X_j \) are all identically distributed (same probability distribution)), then

\[ \mathcal{Z}_\text{eq}(k) = \mathcal{X}_\text{eq}(k) \]
So how is this helpful? If you start with probability distributions for the independent random variables $X_j$ And you want to get the probability distribution for their sum $Z$ then you can proceed as follows:

1. Compute the characteristic function or probability generating function of the $X_j$
2. Obtain the characteristic function or probability generating function for $Z$ by the above relationships
3. Extract the probability distribution for $Z$ by inverting the transformation from characteristic function or probability generating function.

How does one do the inversion step?

For probability generating functions:

$$p(Z = n) = \left. \frac{1}{n!} \frac{d^n}{ds^n} G_Z(s) \right|_{s=0}$$

For characteristic functions, the inversion formula is essentially an inverse Fourier transform (if one is using continuously distributed random variables):

$$p_Z(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ikz} \phi_Z(k) \, dk$$

Or...just use tables that relate transforms to each other.
Mathematical Framework for Stochastic Processes

At an intuitive level, a stochastic process is just a random function mapping from a parameter domain $T$ to a state space $S$ (which in this class will be discrete)

Mathematically, this can be formalized in a number of ways:

1. Direct formulation is to just think of $\{X(t)\}_{t \in T}$

   As a collection of (usually not independent) random variables, indexed by the time parameter. This turns out to be the basis for a rigorous construction of stochastic processes known as the Kolmogorov extension theorem. This formulation works OK for discrete time but is usually not terribly useful in continuous time (due to uncountably many times being involved) except when the stochastic process is Gaussian (which we won't cover in this class).

2. Abstract formulation: Think of the stochastic process as a random variable in function space. This is an approach taken often in pure mathematics and sometimes theoretical physics (functional calculus) but not broadly useful in most applications.

3. "Monte Carlo" formulation: Think of the stochastic process as implicitly being a joint map from an underlying sample space and the parameter domain into state space:

\[ X: T \times \Omega \rightarrow S \]

\[ (t, \omega) \rightarrow X(t, \omega) \]
4. "Weak" formulation: Any observation (nice linear functional) of the stochastic process is a random variable:

\[ \mathbb{V}(2), \int_1^3 \mathbb{V}(s) \, ds, \mathbb{V}(s) \mid s \geq 2 \]

Notation: For discrete time stochastic processes, we will generally index time (parameter domain) by \textbf{epoch} \( n = 0, 1, 2, 3, \ldots \) And the stochastic process will be denoted

\[ \{ X_n \}_{n=0}^{\infty} = \{ X_n(\omega) \}_{n=0}^{\infty} \]

Simplest example of a stochastic process is one for which its value at different epochs are independent.

These are easy to work with, using the formulas above for independent random variables, but has rather limited applications:

- Coin or die tosses
- Amount of demand for a certain product on a given day
- Intervals between cars on a road
- Time between radioactive decays
- "Velocities" of particles undergoing Brownian motion

In many applications, the variables of interest have some sort of memory and so such a simple model is not so widely useful. However, one does want to use the nice mathematical properties of independent random variables in order to make the stochastic models tractable. The simplest kind of stochastic models that have memory are those for which the changes in the state of the random variable can be modeled in terms of independent random variables. This is the idea behind the broad category of stochastic processes known as \textbf{Markov processes} which we will concentrate on for the most part in this class.

Really the only stochastic processes for which I'm aware of a well-developed
mathematical theory are Markov processes and Gaussian processes. (There is research being done on other stochastic processes, but much more technical.)

**Finite state, discrete time (FSDT) Markov chain**

**Reading:** Lawler Ch. 1

**State space:** \[ S = \{1, 2, \ldots, M\} \]

**Parameter domain:** \[ T = \{0, 1, 2, \ldots\} \]

(and sometimes one extends to negative time, and that's fine as we'll discuss in a moment.)

There's two equivalent ways to formulate a FSDT Markov chain.

**Stochastic update rule:** (Resnick Sec. 2.1)

This is a mathematical encoding of the intuition described above.

Prescribe arbitrary updating functions: \[ f_n : S \times S \rightarrow S \]

We generate a sequence of independent, identically distributed random variables

\[ \{ Z_n \}_{n \geq 0} \quad Z_n \in S \]

Representing the unpredictable inputs into the system that appear at epoch \( n \).
Often, we will work with time-homogenous Markov chains for which the updating rule is independent of time:

\[
X_{n+1} = f_n(X_n, Z_n)
\]

This defines a random dynamical system, and requires initial data.

Generally the initial condition can be framed probabilistically, meaning that we prescribe the probability distribution for the initial state of the system:

\[
\phi_i = P(X_0 = i) \quad \text{for } i \in \{1, \ldots, M\}
\]

\[
\sum_{i=1}^{M} \phi_i = 1
\]

Often the state \(X_0\) is known precisely, in which case we can pose initial data of the special form:

\[
\phi_i = \delta_{i, i_x} = \begin{cases} 1 & \text{if } i = i_x \\ 0 & \text{if } i \neq i_x \end{cases}
\]

Kronecker delta function when \(X_0 = i_x\) is known.
A second, more classical formulation of Markov chains is to define the Markov property and follow through its consequences.

A discrete-time stochastic process is said to have the Markov property provided that:

\[
P(X_{n+1} = j \mid X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = P(X_{n+1} = j \mid X_n = i_n)
\]

for all \( i_0, i_1, \ldots, i_n \in S \), \( n \geq 0 \)

The English codification of this is: Given the current state of the system, the future states are independent of the past states.