A stochastic process is a collection of random variables \((X_t : t \in T)\) where \(T\) is some index set. \(X_t\) may have a discrete, continuous or mixed distribution; the sample space of each \(X_t\) is called the state space, often denoted by \(S\), i.e. \(X_t \in S\).

The dependence between the variables can be expressed through the equation:

\[
P(X_0 \in A_0, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i | X_0 \in A_0, \ldots, X_{i-1} \in A_{i-1}),
\]

where the index set is discrete with values \(0, 1, 2, \ldots, n\).

The theory of stochastic processes provides various specifications of conditional probabilities on the righthand side of the equation.

Note that, not all specifications are necessarily valid. Various types of regularity conditions are required.
The positivity condition

Consider a discrete multivariate random variable
\( x = (x_1, \ldots, x_m) \) with probability distribution
\( q(x) = P(x = x), \quad x \in S = \{ x : P(X_i = x) > 0 \} \).
This choice of \( S \) is called the minimal state space. Now consider the minimal state spaces for each of the components \( x_i \), i.e.
\[ S_i = \{ x : P(x_i = x) > 0 \}. \]
We say that \( q \) satisfies the positivity condition if
\[ S = S_1 \times S_2 \times \ldots \times S_m, \]
so that if \( x \in S_i \), \( i = 1(1)m \), we have
\[ x = (x_1, \ldots, x_m) \in S. \]

The conditional distribution of \( x_i \), given the values of all the other variables, which we denote by \( x_{-i} \), is
\[ q_i (x_i \mid x_{-i}) = \frac{P(x_i = x_i \mid x_{-i} = x_{-i})}{P(x_{-i} = x_{-i})} = \frac{q(x)}{\sum_{x_{-i} \in S} q(x)} \]

If \( q \) satisfies positivity condition, these \( q_i \) are well-defined.
The great advantage of a process satisfying positivity condition is that the knowledge of \( q_i \) suffices to determine \( q \). Note that \( q_i \)'s are univariate probability distribution.

**Proposition**

Suppose that \( q \) satisfies the positivity condition. Then

\[
\frac{q(x)}{q(x)} = \prod_{i=1}^{n} \frac{q_i(z_i \mid x_1, x_2, \ldots, x_{i-1}, z_{i+1}, \ldots, y_m)}{q_i(z_i \mid x_1, x_2, \ldots, z_{i-1}, y_{i+1}, \ldots, y_m)}
\]

**Proof:**

\[
q(x) = q_m(x_m \mid x_1, \ldots, x_{m-1}) \prod_{i=2}^{m} q_i(x_i \mid x_{i-1}, \ldots, x_1)
\]

\[
= q_m(x_m \mid x_1, \ldots, x_{m-1}) \prod_{i=2}^{m} \frac{q_i(x_i \mid x_{i-1}, \ldots, x_1)}{q_i(z_i \mid x_1, x_2, \ldots, z_{i-1}, y_{i+1}, \ldots, y_m)}
\]

\[
= \frac{q_m(x_m \mid x_1, \ldots, x_{m-1})}{q_m(y_m \mid x_1, \ldots, x_{m-1})} \cdot \frac{q_m(x_m \mid x_1, \ldots, x_{m-1})}{q_m(y_m \mid x_1, \ldots, x_{m-1})}
\]

\[
\frac{q_m(x_m \mid x_1, \ldots, x_{m-1})}{q_m(y_m \mid x_1, \ldots, x_{m-1})} \cdot \frac{q_m(x_m \mid x_1, \ldots, x_{m-1})}{q_m(y_m \mid x_1, \ldots, x_{m-1})}
\]

\[
= \frac{q_m(x_m \mid x_1, \ldots, x_{m-1})}{q_m(y_m \mid x_1, \ldots, x_{m-1})} \cdot \frac{q_m(x_m \mid x_1, \ldots, x_{m-1})}{q_m(y_m \mid x_1, \ldots, x_{m-1})}
\]

\[
= \frac{q_m(x_m \mid x_1, \ldots, x_{m-1})}{q_m(y_m \mid x_1, \ldots, x_{m-1})} \cdot \frac{q_m(x_m \mid x_1, \ldots, x_{m-1})}{q_m(y_m \mid x_1, \ldots, x_{m-1})}
\]

\[
= \frac{q_m(x_m \mid x_1, \ldots, x_{m-1})}{q_m(y_m \mid x_1, \ldots, x_{m-1})} \cdot \frac{q_m(x_m \mid x_1, \ldots, x_{m-1})}{q_m(y_m \mid x_1, \ldots, x_{m-1})}
\]
Continue the process for each $i$ "replacing" $x_i$ with $y_i$ by multiplying and dividing by $q_i$ ($y_i$: $x_1, \ldots, x_i-1, y_i, \ldots, x_m$) and regrouping them. Positivity assures if $q(x) > 0$ and $q(y) > 0$,

$q(x_1, \ldots, x_i, y_i, \ldots, x_m) > 0$ for $i = 1, \ldots, n$.

The consequence is that $q$ is uniquely determined by the conditional distribution, since

$$\sum_{x \in S} q(x) = 1.$$  

The Kolmogorov consistency condition

Consider a stochastic process $(x_i: i \geq 0, \ldots)$ having a distribution such that

$$P(x_i \in A_1, \ldots, x_m \in A_m) = P(x_i \in A_1, \ldots, x_m \in A_m)$$

for all $i, \ldots, i+m \in \{0, 1, \ldots, m\}$ and $A_1, A_2, \ldots, A_m$ events (measurable subsets).

This condition guarantees existence of a probability measure corresponding to the stochastic process.
Depending on the structure of the state space \( S \) and the index set \( T \) one has different classifications of stochastic processes. It is often referred to as time.

**Example:**

1. Forecasting the weather:

\[ x_i \] is the weather on \( i \)-th day. It can take 3 possible values (state 0 when it rains, state 1 when it snows, state 2 when it is clear).

\[ S = \{0, 1, 2\} \quad \text{and} \quad T = \mathbb{N} \quad \text{Time and State space both discrete} \]

2. \( x_i \) is the number of earthquakes in Japan in one time period \([0, +\infty)\) where \( 0 \) is beginning of recording. This is called a counting process.

\[ T = \mathbb{R}_+ \quad \text{and} \quad S = \mathbb{N} \quad \text{Time is continuous, state space discrete} \]

3. \( x_{ij} \) is the amount of acid rain at location \( j \) at time \( i \).

\[ T = \mathbb{R}_+ \times \mathbb{R}_+ \quad \text{and} \quad \alpha = (0, +\infty) \quad \text{so \([0, +\infty)\)} \]

This is called a random field. The state space is continuous.
(4) \( x \): is the thickness of an optical fiber at a distance \( t \) from the origin. Both state space and time are continuous with \( T = S = \mathbb{R}_+ \). Time is really a distance here.

\[ \text{Markov Chain} \]

We will first consider a stochastic process with discrete state space and discrete-time index set.

A discrete state, discrete-time stochastic process \( \{ X_n : n = 0, 1, 2, \ldots \} \) is called a Markov chain if

\[ P(X_{n+1} = j | X_n = i, X_{n-1}, \ldots, X_0 = i) = P_j(i) \]

for all states \( i, j \in \mathcal{S} \) and all \( n \geq 0 \).

Thus,

\[ P(X_0 \in A_0, \ldots, X_n \in A_n) = P(X_0 \in A_0) \cdot P(X_1 \in A_1 | X_0 \in A_0) \cdot \ldots \cdot P(X_n \in A_n | X_{n-1} \in A_{n-1}) \]
For any Markov chain \( \{X_n\} \),

\[
P(X_0, X_2 | X_1) = P(X_2 | X_1) P(X_0 | X_1)
\]

sets of past and future states of a Markov chain are conditionally independent given the present.

\[
P(X_0, X_2 | X_1) = \frac{P(X_0, X_1, X_2)}{P(X_1)}
\]

\[
= \frac{P(X_2 | X_0, X_1) P(X_1 | X_0)}{P(X_1)}
\]

\[
= P(X_2 | X_1) P(X_0 | X_1)
\]

for any Markov chain \( \{X_n\} \),

\[
P(X_3 | X_0, X_1) = P(X_3 | X_0)
\]

\[
P(X_n | \text{any set of previous states})
\]

\[
= P(X_n | \text{latest available})
\]

\[
P(X_0, X_2 | X_1) = \sum_{X_2} P(X_2, X_3 | X_0, X_1) P(X_2 | X_0, X_1)
\]

\[
= \sum_{X_2} \sum_{X_3} P(X_3 | X_0, X_1, X_2) P(X_2 | X_0, X_1)
\]

\[
= \sum_{X_2} \sum_{X_3} P(X_3 | X_2) P(X_2 | X_1) P(X_0 | X_1)
\]

\[
= \sum_{X_2} P(X_3, X_0 | X_1) P(X_2 | X_1)
\]
Coin flips: Let's assume coin flips are independent and it occurs with prob. $p_0$ in each occasion.

Define $X_n$ as the excess of $U$ over $P_t = p, p, \ldots$ after $n$ flips of a coin and take $S = \{-\infty, -1, 0, 1, \ldots\}.$

Note here that the choice of state space is not unique and $S$ may contain states that are not possible. (Recall: Minimal state space)

\[ P_{i+1}(n) = p, \quad P_{i-1}(n) = q = 1 - p, \quad P_{ij} = 0, i \neq j. \]

Note $P_{ij}(n)$ are independent of $n.$

**Definition:** The $P_{ij}(n)$'s of a Markov chain are referred to as the one-step transition probabilities. If the $P_{ij}(n)$'s are time invariant, so that $P_{ij}(n) = P_{ij},$ the chain is called "homogeneous" or said to have "stationary transition probabilities."

6. A basis of the transition process, the joint dist. of a first-order Markov chain can be written as

\[ P_r(x_1, x_2, \ldots, x_n | x_0) = P_{x_0} x_1 P_{x_2} \cdots P_{x_{n-1}} x_n P_{x_n} x_{n+1} \ldots P_{x_n} x_{n+1} \ldots P_{x_n} x_{n+k} \]
Homogeneous Markov Chain (HMC)

we will mainly focus on HMCs in the class.
We will deal with the

Transition Probability Matrix \( \mathbf{P} \): The \( s \times s \) matrix

\( p \) with \( \left( i,j \right) \)th element \( p_{ij} \) is called the

\( \mathbf{P} \) of the chain.

What will be the \( \mathbf{P} \) for the excess car loss problem?

Properties of \( \mathbf{P} \):

(i) \( p_{ij} \geq 0 \), for all \( i, j \in S \)
(ii) \( \sum_{j} p_{ij} = 1 \) \( \forall i \in S \)

In matrix notation,

(i) \( P \geq 0 \) (ii) \( P \mathbf{1} = \mathbf{1} \)

Any square matrix satisfying above condition
is called "stochastic" matrix.

The \( n \)-step transition probabilities

The \( n \)-step transition probability is defined as

\[ p_{ij}^{(n)} = \Pr \left( X_{n+1} = j \mid X_n = i \right) \]

It is independent of \( n \) because of time homogeneity.
Proposition:

The n-step transition matrix \( P^n \) of \((i, j)\)th element \( p_{ij}^{(n)} \) is given by:

\[
P_n = p^n, \quad n \geq 1
\]

where \( p^0 = I \) is the appropriate identity matrix.

\( p^n \) is a stochastic matrix, whose i-th row gives the conditional distribution of \( X_n \), given the initial state \( X_0 = i \).

Proof: Taking \( x_0 = i \) and \( x_n = j \) in

\[
P(x_1, \ldots, x_n \mid x_0) = p_{x_0x_1}(0) p_{x_1x_2}(1) \cdots p_{x_{n-1}x_n}(n-1)
\]

we get

\[
p_{ij}^{(n)} = \sum_{x_1} \cdots \sum_{x_{n-1}} p_{ix_1} p_{x_1x_2} \cdots p_{x_{n-1}j}
\]

[because it is HTMC]

Note that

\[
[p^n]_{ij} = \sum_{x_{n-1}} \cdots \sum_{x_2} \sum_{x_1} p_{x_1x_2} \cdots p_{x_{n-1}x_n} p_{x_0x_1} = \sum_{x_{n-1}} \cdots \sum_{x_2} \sum_{x_1} p_{x_0x_1} p_{x_1x_2} \cdots p_{x_{n-1}j}
\]

\[
= \sum_{x_{n-1}} \cdots \sum_{x_2} \sum_{x_1} p_{x_0x_1} p_{x_1x_2} \cdots p_{x_{n-1}j}
\]

\[
= \sum_{x_{n-1}} \cdots \sum_{x_2} \sum_{x_1} p_{x_0x_1} p_{x_1x_2} \cdots p_{x_{n-1}j}
\]

\[
= \sum_{x_{n-1}} \cdots \sum_{x_2} \sum_{x_1} p_{x_0x_1} p_{x_1x_2} \cdots p_{x_{n-1}j}
\]