The Basics: Let there be a sample space, a set of events (with axioms), and a probability measure on the events (with axioms).

In practice, there is a basic countable set of rv’s that are IID, Markov, etc.

A sample point is then a collection of sample values, one for each rv.

There are often uncountable sets of rv’s, e.g., \{N(t); t \geq 0\}, but they can usually be defined in terms of a basic countable set.

For a sequence of IID rv’s, \(X_1, X_2, \ldots\) (Poisson and renewal processes), the laws of large numbers specify long term behavior.

The sample (time) average is \(S_n/n, S_n = X_1 + \cdots X_n\). It is a rv of mean \(\bar{X}\) and variance \(\sigma^2/n\).
The weak LLN: If $E[|X|] < \infty$, then
\[
\lim_{n \to \infty} \Pr\left\{ \left| \frac{S_n}{n} - \bar{X} \right| \geq \epsilon \right\} = 0 \quad \text{for every } \epsilon > 0.
\]
This says that $\Pr\left\{ \frac{S_n}{n} \leq x \right\}$ approaches a unit step at $\bar{X}$ as $n \to \infty$ (Convergence in probability and in distribution).

The strong LLN: If $E[|X|] < \infty$, then
\[
\lim_{n \to \infty} \frac{S_n}{n} = \bar{X} \quad \text{W.P.1}
\]
This says that, except for a set of sample points of zero probability, all sample sequences have a limiting sample path average equal to $\bar{X}$.

Also, essentially \( \lim_{n \to \infty} f\left(\frac{S_n}{n}\right) = f(\bar{X}) \) W.P.1.

There are many extensions of the weak law telling how fast the convergence is. The most useful result about convergence speed is the central limit theorem. If $\sigma_X^2 < \infty$, then
\[
\lim_{n \to \infty} \left[ \Pr\left\{ \frac{S_n - n\bar{X}}{\sqrt{n} \sigma} \leq y \right\} \right] = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) \, dx.
\]
Equivalently,
\[
\lim_{n \to \infty} \left[ \Pr\left\{ \frac{S_n}{n} - \bar{X} \leq \frac{y\sigma}{\sqrt{n}} \right\} \right] = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) \, dx.
\]
In other words, $S_n/n$ converges to $\bar{X}$ with $1/\sqrt{n}$ and becomes Gaussian as an extra benefit.
Arrival processes

Def: An arrival process is an increasing sequence of rv's, $0 < S_1 < S_2 < \cdots$. The interarrival times are $X_1 = S_1$ and $X_i = S_i - S_{i-1}$, $i \geq 1$.

An arrival process can model arrivals to a queue, departures from a queue, locations of breaks in an oil line, etc.

The process can be specified by the joint distribution of either the arrival epochs or the interarrival times.

The counting process, $\{N(t); t \geq 0\}$, for each $t$, is the number of arrivals up to and including $t$, i.e., $N(t) = \max\{n: S_n \leq t\}$. For every $n$, $t$,

$$\{S_n \leq t\} = \{N(t) \geq n\}$$

Note that $S_n = \min\{t: N(t) \geq n\}$, so that $\{N(t); t \geq 0\}$ specifies $\{S_n; n > 0\}$. 

5

6
Def: A renewal process is an arrival process for which the interarrival rv's are IID. A Poisson process is a renewal process for which the interarrival rv's are exponential.

Def: A memoryless rv is a nonnegative non-deterministic rv for which

\[ \Pr\{X > t+x\} = \Pr\{X > x\}\Pr\{X > t\} \quad \text{for all } x, t \geq 0. \]

This says that \( \Pr\{X > t+x \mid X > t\} = \Pr\{X > x\} \). If \( X \) is the time until an arrival, and the arrival has not happened by \( t \), the remaining distribution is the original distribution.

The exponential is the only memoryless rv.

Thm: Given a Poisson process of rate \( \lambda \), the interval from any given \( t > 0 \) until the first arrival after \( t \) is a rv \( Z_1 \) with \( F_{Z_1}(z) = 1 - \exp[-\lambda z] \). \( Z_1 \) is independent of all \( N(\tau) \) for \( \tau \leq t \).

\( Z_1 \) (and \( N(\tau) \) for \( \tau \leq t \)) are also independent of future interarrival intervals, say \( Z_2, Z_3, \ldots \). Also \( \{Z_1, Z_2, \ldots\} \) are the interarrival intervals of a PP starting at \( t \).

The corresponding counting process is \( \{\bar{N}(t, \tau) ; \tau \geq t\} \) where \( \bar{N}(t, \tau) = N(\tau) - N(t) \) has the same distribution as \( N(\tau - t) \).

This is called the stationary increment property.
The probability distributions

\( f_{S_1, \ldots, S_n}(s_1, \ldots, s_n) = n^s \exp(-\lambda s_n) \quad \text{for} \ 0 \leq s_1 \leq \cdots \leq s_n \)

The intermediate arrival epochs are equally likely to be anywhere (with \( s_1 < s_2 < \cdots \)). Integrating,

\[
f_{S_n}(t) = \frac{\lambda^n t^{n-1} \exp(-\lambda t)}{(n-1)!} \quad \text{Erlang}
\]

The probability of arrival \( n \) in \((t, t+\delta)\) is

\[
\Pr\{N(t) = n-1\} \lambda \delta = \delta f_{S_n}(t) + o(\delta)
\]

\[
\Pr\{N(t) = n-1\} = \frac{\lambda}{f_{S_n}(t)} \frac{(\lambda t)^{n-1} \exp(-\lambda t)}{(n-1)!} = \frac{(\lambda t)^{n-1} \exp(-\lambda t)}{n!} \quad \text{Poisson}
\]
Combining and splitting

If \( N_1(t), N_2(t), \ldots, N_k(t) \) are independent PP’s of rates \( \lambda_1, \ldots, \lambda_k \), then \( N(t) = \sum_i N_i(t) \) is a Poisson process of rate \( \sum_j \lambda_j \).

Two views: 1) Look at arrival epochs, as generated, from each process, then combine all arrivals into one Poisson process.

(2) Look at combined sequence of arrival epochs, then allocate each arrival to a sub-process by a sequence of IID rv’s with PMF \( \lambda_i / \sum_j \lambda_j \).

This is the workhorse of Poisson type queueing problems.

Conditional arrivals and order statistics

\[ f_{S^{(n)} | N(t)}(s^{(n)} | n) = \frac{n!}{t^n} \quad \text{for} \ 0 < s_1 < \cdots s_n < t \]

\[ \Pr\{S_1 > \tau \mid N(t)=n\} = \left[ \frac{t-\tau}{t} \right]^n \quad \text{for} \ 0 < \tau \leq t \]

\[ \Pr\{S_n < t - \tau \mid N(t)=n\} = \left[ \frac{t-\tau}{t} \right]^n \quad \text{for} \ 0 < \tau \leq t \]

The joint distribution of \( S_1, \ldots, S_n \) given \( N(t) = n \) is the same as the joint distribution of \( n \) uniform rv’s that have been ordered.
**Finite-state Markov chains**

An integer-time stochastic process \( \{X_n; n \geq 0\} \) is a Markov chain if for all \( n, i, j, k, \ldots \),

\[
Pr\{X_n = j \mid X_{n-1} = i, X_{n-2} = k \ldots X_0 = m\} = P_{ij},
\]

where \( P_{ij} \) depends only on \( i, j \) and \( p_{X_0}(m) \) is arbitrary.

A Markov chain is finite-state if the sample space for each \( X_i \) is a finite set, \( S \). The sample space \( S \) usually taken to be the integers \( 1, 2, \ldots, M \).

A Markov chain is completely described by \( \{P_{ij}; 1 \leq i, j \leq M\} \) plus the initial probabilities \( p_{X_0}(i) \).

The set of transition probabilities \( \{P_{ij}; 1 \leq i, j \leq M\} \), is usually viewed as the Markov chain with \( p_{X_0} \) viewed as a parameter.

A finite-state Markov chain can be described as a directed graph or as a matrix.

\[
[P] = \begin{bmatrix}
P_{11} & P_{12} & \cdots & P_{16} \\
P_{21} & P_{22} & \cdots & P_{26} \\
\vdots & \vdots & \ddots & \vdots \\
P_{61} & P_{62} & \cdots & P_{66}
\end{bmatrix}
\]

a) Graphical  

b) Matrix

An edge \((i, j)\) is put in the graph only if \( P_{ij} > 0 \), making it easy to understand connectivity.

The matrix is useful for algebraic and asymptotic issues.
Classification of states

An \((n\text{-step})\) walk is an ordered string of nodes (states), say \((i_0, i_1, \ldots, i_n)\), \(n \geq 1\), with a directed arc from \(i_{m-1}\) to \(i_m\) for each \(m\), \(1 \leq m \leq n\).

A path is a walk with no repeated nodes.

A cycle is a walk in which the last node is the same as the first and no other node is repeated.

A node \(j\) is accessible from \(i\), \((i \to j)\) if there is a walk from \(i\) to \(j\), i.e., if \(P_{ij}^n > 0\) for some \(n > 0\).

If \((i \to j)\) and \((j \to k)\) then \((i \to k)\).

Two states \(i, j\) communicate (denoted \(i \leftrightarrow j\)) if \((i \to j)\) and \((j \to i)\).

A class \(C\) of states is a non-empty set such that \((i \leftrightarrow j)\) for each \(i, j \in C\) but \(i \not\leftrightarrow j\) for each \(i \in C, j \notin C\).

\(S\) is partitioned into classes. The class \(C\) containing \(i\) is \(\{i\} \cup \{j : (i \leftrightarrow j)\}\).

For finite-state chains, a state \(i\) is transient if there is a \(j \in S\) such that \(i \to j\) but \(j \not\to i\). If \(i\) is not transient, it is recurrent.

All states in a class are transient or all are recurrent.

A finite-state Markov chain contains at least one recurrent class.
The period, \( d(i) \), of state \( i \) is \( \text{gcd}\{n : P^n_{ii} > 0\} \), i.e., returns to \( i \) can occur only at multiples of some largest \( d(i) \).

All states in the same class have the same period.

A recurrent class with period \( d > 1 \) can be partitioned into subclasses \( S_1, S_2, \ldots, S_d \). Transitions from each class go only to states in the next class (viewing \( S_1 \) as the next subclass to \( S_d \)).

An ergodic class is a recurrent aperiodic class. A Markov chain with only one class is ergodic if that class is ergodic.

\[ \text{Thm: \ For an ergodic finite-state Markov chain,} \lim_{n \to \infty} P^n_{ij} = \pi_j, \text{ i.e., the limit exists for all } i, j \text{ and is independent of } i. \{\pi_i; 1 \leq M\} \text{ satisfies} \sum_i \pi_i P_{ij} = \pi_j > 0 \text{ with} \sum_i \pi_i = 1. \]

A substep for this theorem is showing that for an ergodic \( M \) state Markov chain, \( P^n_{ij} > 0 \) for all \( i, j \) and all \( n \geq (M - 1)^2 + 1 \).

The reason why \( n \) must be so large to ensure that \( P^n_{ij} > 0 \) is indicated by the following chain where the smallest cycle has length \( M - 1 \).

Starting in state 2, the state at the next 4 steps is deterministic. For the next 4 steps, there are two possible choices then 3, etc.

A second substep is the special case of the theorem where \( P_{ij} > 0 \) for all \( i, j \).
Lemma 2: Let $[P] > 0$ be the transition matrix of a finite-state Markov chain and let $\alpha = \min_{i,j} P_{ij}$. Then for all states $j$ and all $n \geq 1$:

\[
\max_i P_{ij}^{n+1} - \min_i P_{ij}^{n+1} \leq \left( \max_{\ell} P_{\ell j}^n - \min_{\ell} P_{\ell j}^n \right) (1 - 2\alpha).
\]

This shows that $\lim_{n \to \infty} P_{\ell j}^n$ approaches a limit independent of $\ell$, and approaches it exponentially for $[P] > 0$. The theorem (for ergodic $[P]$) follows by looking at $\lim_{n \to \infty} P_{ij}^{nh}$ for $h = (M - 1)^2 + 1$.

An ergodic unichain is a Markov chain with one ergodic recurrent class plus, perhaps, a set of transient states. The theorem for ergodic chains extends to unichains:

Thm: For an ergodic finite-state unichain, $\lim_n P_{ij}^n = \pi_j$, i.e., the limit exists for all $i,j$ and is independent of $i$. \{\pi_i; 1 \leq M\} satisfies $\sum_i \pi_i P_{ij} = \pi_j$ with $\sum_i \pi_i = 1$. Also $\pi_i > 0$ for $i$ recurrent and $\pi_i = 0$ otherwise.

This can be restated in matrix form as $\lim_n [P^n] = \bar{e} \pi$ where $\bar{e} = (1, 1, \ldots, 1)^T$ and $\pi$ satisfies $\pi [P] = \pi$ and $\pi \bar{e} = 1$. 
We get more specific results by looking at the eigenvalues and eigenvectors of an arbitrary stochastic matrix (matrix of a Markov chain).

\( \lambda \) is an eigenvalue of \([P]\) iff \([P - \lambda I]\) is singular, iff det\([P - \lambda I] = 0\), iff \([P] = \lambda \nu\) for some \(\nu \neq 0\), and iff \(\pi[P] = \lambda \pi\) for some \(\pi \neq 0\).

\( e \) is always a right eigenvector of \([P]\) with eigenvalue 1, so there is always a left eigenvector \(\pi\).

det\([P - \lambda I]\) is an Mth degree polynomial in \(\lambda\). It has M roots, not necessarily distinct. The multiplicity of an eigenvalue is the number of roots of that value.

The multiplicity of \(\lambda = 1\) is equal to the number of recurrent classes.

For the special case where all M eigenvalues are distinct, the right eigenvectors are linearly independent and can be represented as the columns of an invertible matrix \([U]\). Thus

\[
[P][U] = [U][\Lambda]; \quad \quad [P] = [U][\Lambda][U^{-1}]
\]

The matrix \([U^{-1}]\) turns out to have rows equal to the left eigenvectors.

This can be further broken up by expanding \([\Lambda]\) as a sum of eigenvalues, getting

\[
[P] = \sum_{i=1}^{M} \lambda_i \bar{v}^{(i)} \bar{\pi}^{(i)}
\]

\[
[P^n] = [U][\Lambda^n][U^{-1}] = \sum_{i=1}^{M} \lambda_i^n \bar{v}^{(i)} \bar{\pi}^{(i)}
\]
Renewal processes

Thm: For a renewal process (RP) with mean inter-renewal interval $X > 0$,

$$\lim_{t \to \infty} \frac{N(t)}{t} = \frac{1}{X} \quad \text{W.P.1.}$$

This also holds if $X = \infty$.

In both cases, $\lim_{t \to \infty} N(t) = \infty$ with probability 1.

There is also the elementary renewal theorem, which says that

$$\lim_{t \to \infty} E \left[ \frac{N(t)}{t} \right] = \frac{1}{X}$$

Facts: All eigenvalues $\lambda$ satisfy $|\lambda| \leq 1$.

For each recurrent class $\mathcal{C}$, there is one $\lambda = 1$ with a left eigenvector equal to steady state on that recurrent class and zero elsewhere. The right eigenvector $\nu$ satisfies $\lim_n \Pr\{X_n \in \mathcal{C} \mid X_0 = i\} = \nu_i$.

For each recurrent periodic class of period $d$, there are $d$ eigenvalues equi-spaced on the unit circle. There are no other eigenvalues with $|\lambda| = 1$.

If the eigenvectors span $\mathbb{R}^M$, then $P^n_{ij}$ converges to $\pi_j$ as $\lambda^n$ for a unichain where $|\lambda_2|$ is the second largest magnitude eigenvalue.

If the eigenvectors do not span $\mathbb{R}^M$, then $[P^n] = [U][J][U^{-1}]$ where $[J]$ is a Jordan form.
The integral of $Y(t)$ over $t$ is a sum of terms $X_n^2/2$.

The time average value of $Y(t)$ is

$$\lim_{t \to \infty} \frac{\int_{\tau=0}^{t} Y(\tau) \, d\tau}{t} = \frac{\mathbb{E}[X^2]}{2\mathbb{E}[X]} \quad \text{W.P.1}$$

The time average duration is

$$\lim_{t \to \infty} \frac{\int_{\tau=0}^{t} X(\tau) \, d\tau}{t} = \frac{\mathbb{E}[X^2]}{\mathbb{E}[X]} \quad \text{W.P.1}$$

For PP, this is twice $\mathbb{E}[X]$. Big intervals contribute in two ways to duration.
Residual life and duration are examples of renewal reward functions.

In general $R(Z(t), X(t))$ specifies reward as function of location in the local renewal interval.

Thus reward over a renewal interval is

$$R_n = \int_{S_{n-1}}^{S_n} R(\tau - S_{n-1}, X_n) d\tau = \int_{z=0}^{X_n} R(z, X_n) dz$$

$$\lim_{t \to \infty} \frac{1}{t} \int_{\tau=0}^{t} R(\tau) d\tau = \frac{E[R_n]}{X} \quad \text{W.P.1}$$

This also works for ensemble averages.

Def: A stopping trial (or stopping time) $J$ for a sequence $\{X_n; n \geq 1\}$ of rv's is a positive integer-valued rv such that for each $n \geq 1$, the indicator rv $I\{J=n\}$ is a function of $\{X_1, X_2, \ldots, X_n\}$.

A possibly defective stopping trial is the same except that $J$ might be a defective rv. For many applications of stopping trials, it is not initially obvious whether $J$ is defective.

Theorem (Wald's equality) Let $\{X_n; n \geq 1\}$ be a sequence of IID rv's, each of mean $\overline{X}$. If $J$ is a stopping trial for $\{X_n; n \geq 1\}$ and if $E[J] < \infty$, then the sum $S_J = X_1 + X_2 + \cdots + X_J$ at the stopping trial $J$ satisfies

$$E[S_J] = \overline{X} E[J].$$
Wald: Let \( \{X_n; n \geq 1\} \) be IID rv's, each of mean \( \overline{X} \). If \( J \) is a stopping time for \( \{X_n; n \geq 1\} \), \( E[J] < \infty \), and \( S_J = X_1 + X_2 + \cdots + X_J \), then
\[
E[S_J] = \overline{X}E[J]
\]
In many applications, where \( X_n \) and \( S_n \) are nonnegative rv's, the restriction \( E[J] < \infty \) is not necessary.

For cases where \( X \) is positive or negative, it is necessary as shown by ‘stop when you’re ahead.’

Little’s theorem

This is little more than an accounting trick. Consider an queueing system with arrivals and departures where renewals occur on arrivals to an empty system.

Consider \( L(t) = A(t) - D(t) \) as a renewal reward function. Then \( L_n = \sum W_i \) also.
Let $\bar{L}$ be the time average number in system,

$$\bar{L} = \frac{1}{t} \lim_{t \to \infty} \int_0^t L(\tau) \, d\tau$$

$$\lambda = \lim_{t \to \infty} \frac{1}{t} A(t)$$

$$\bar{W} = \lim_{t \to \infty} \frac{1}{A(t)} \sum_{i=1}^{A(t)} W_i$$

$$= \lim_{t \to \infty} \frac{1}{A(t)} \lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^{A(t)} W_i$$

$$= \frac{\bar{L}}{\lambda}$$