6.262: Discrete Stochastic Processes 3/28/11

Lecture 14: Review

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.

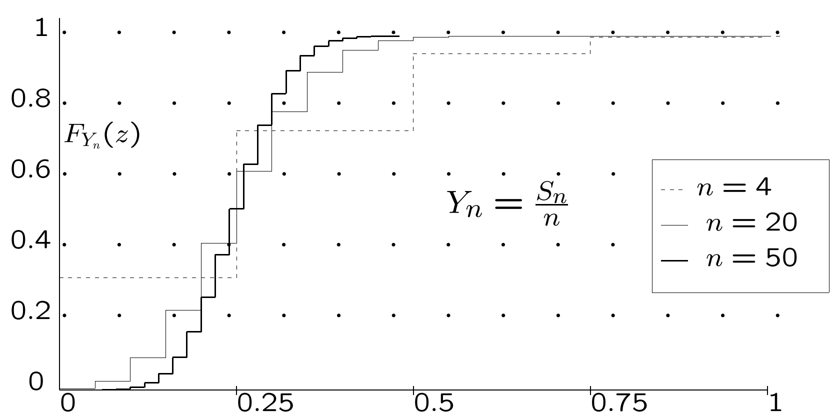
A sample point is then a collection of sample values, one for each rv.

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 $\overline{X}$ and variance $\sigma^2/n$.



**The weak LLN: If** $\mathsf{E}[|X|] < \infty$**, then**

$$\lim_{n\to\infty} \Pr\left\{\left|\frac{S_n}{n} - \overline{X}\right| \geq \epsilon\right\} = 0 \qquad \textbf{for every } \epsilon > 0.$$

**This says that** $\Pr\left\{\frac{S_n}{n} \leq x\right\}$ **approaches a unit step at** $\overline{X}$ **as** $n \to \infty$ **(Convergence in probability and in distribution).**

**The strong LLN: If** $\mathsf{E}[|X|] < \infty$**, then**

$$\lim_{n\to\infty} \frac{S_n}{n} = \overline{X} \qquad \textbf{W.P.1}$$

**This says that, except for a set of sample points of zero probability, all sample sequences have a limiting sample path average equal to** $\overline{X}$**.**

**Also, essentially** $\lim_{n\to\infty} f(S_n/n) = f(\overline{X})$ **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**

$$\lim_{n\to\infty} \left[\Pr\left\{\frac{S_n - n\overline{X}}{\sqrt{n}\,\sigma} \leq y\right\}\right] = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx.$$

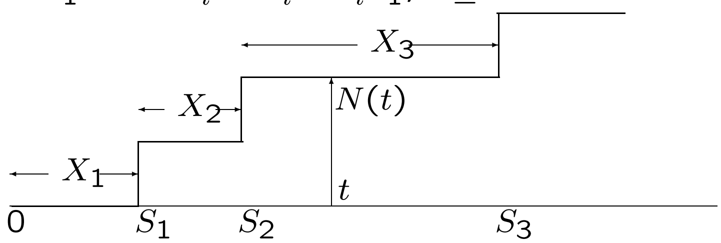
**Equivalently,**

$$\lim_{n\to\infty} \left[\Pr\left\{\frac{S_n}{n} - \overline{X} \leq \frac{y\sigma}{\sqrt{n}}\right\}\right] = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx.$$

**In other words,** $S_n/n$ **converges to** $\overline{X}$ **with** $1/\sqrt{n}$ **and becomes Gaussian as an extra benefit.**

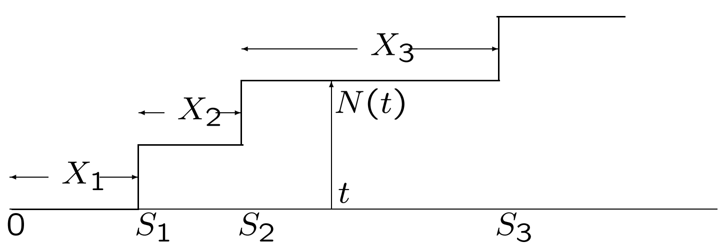
## Arrival processes

**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 $\{S_n;\, n > 0\}$.**

**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**

$$\Pr\{X > t+x\} = \Pr\{X > x\}\Pr\{X > t\} \quad \textbf{for all } x, t \geq 0.$$

**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** $\{\tilde{N}(t,\tau);\ \tau \geq t\}$ **where** $\tilde{N}(t,\tau) = N(\tau) - N(t)$ **has the same distribution as** $N(\tau - t)$**.**

**This is called the stationary increment property.**

**Def: The independent increment property for a counting process is that for all $0 < t_1 < t_2 < \cdots t_k$, the rv's $N(t_1), [\tilde{N}(t_1, t_2)], \ldots, [\tilde{N}(t_{n-1}, t_n)]$ are independent.**

**Thm: PP's have both the stationary and independent increment properties.**

**PP's can be defined by the stationary and independent increment properties plus either the Poisson PMF for $N(t)$ or**

$$\begin{aligned} \Pr\Big\{\tilde{N}(t, t{+}\delta) = 1\Big\} &= \lambda\delta + o(\delta) \\ \Pr\Big\{\tilde{N}(t, t{+}\delta) > 1\Big\} &= o(\delta). \end{aligned}$$



## The probability distributions

$f_{S_1,\ldots,S_n}(s_1, \ldots, s_n) = \lambda^n \exp(-\lambda s_n)$ **for** $0 \le s_1 \le \cdots \le s_n$

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

$$f_{S_n}(t) = \frac{\lambda^n t^{n-1} \exp(-\lambda t)}{(n-1)!} \qquad \textbf{Erlang}$$

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

$$\begin{aligned} \Pr\{N(t) = n{-}1\}\, \lambda\delta &= \delta f_{S_n}(t) + o(\delta) \\ \Pr\{N(t) = n{-}1\} &= \frac{f_{S_n}(t)}{\lambda} \\ &= \frac{(\lambda t)^{n-1} \exp(-\lambda t)}{(n-1)!} \\ p_{N(t)}(n) &= \frac{(\lambda t)^n \exp(-\lambda t)}{n!} \qquad \textbf{Poisson} \end{aligned}$$

## 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

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

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

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

**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 $n, i, j, k, \dots$,**

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

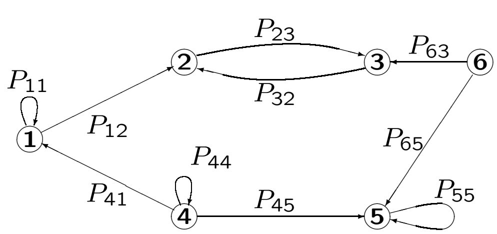
**where $P_{ij}$ depends only on $i, j$ and $\mathsf{p}_{X_0}(m)$ is arbitrary. A Markov chain is finite-state if the sample space for each $X_i$ is a finite set, $\mathcal{S}$. The sample space $\mathcal{S}$ usually taken to be the integers $1, 2, \dots, \mathsf{M}$.**

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

**The set of transition probabilities $\{P_{ij};\, 1 \leq i, j \leq \mathsf{M}\}$, is usually viewed as the Markov chain with $\mathsf{p}_{X_0}$ viewed as a parameter.**



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




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

**a) Graphical**      **b) Matrix**

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

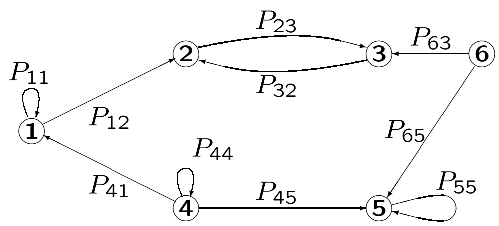
**The matrix is useful for algebraic and asymptotic issues.**

## Classification of states

**An ($n$-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.**




**Walk: (4, 4, 1, 2, 3, 2)**
**Walk: (4, 1, 2, 3)**
**Path: (4, 1, 2, 3)**
**Path: (6, 3, 2)**
**Cycle: (2, 3, 2)**
**Cycle: (5, 5)**

**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 $\mathcal{C}$ of states is a non-empty set such that $(i \leftrightarrow j)$ for each $i, j \in \mathcal{C}$ but $i \not\leftrightarrow j)$ for each $i \in \mathcal{C}$, $j \notin \mathcal{C}$.**

**$\mathcal{S}$ is partitioned into classes. The class $\mathcal{C}$ containing $i$ is $\{i\} \bigcup \{j : (i \leftrightarrow j)\}$.**

**For finite-state chains, a state $i$ is transient if there is a $j \in \mathcal{S}$ 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 $\gcd\{n : P_{ii}^n > 0\}$, 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 $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_d$. Transitions from each class go only to states in the next class (viewing $\mathcal{S}_1$ as the next subclass to $\mathcal{S}_d$).

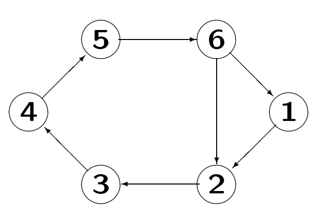
An ergodic class is a recurrent aperiodic class. A Markov chain with only one class is ergodic if that class is ergodic.

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



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

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




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 $n \geq 1$:**

$$\max_i P_{ij}^{n+1} - \min_i P_{ij}^{n+1} \leq \left(\max_\ell P_{\ell j}^n - \min_\ell P_{\ell j}^n\right)(1-2\alpha).$$
$$\left(\max_\ell P_{\ell j}^n - \min_\ell P_{\ell j}^n\right) \leq (1-2\alpha)^n.$$
$$\lim_{n\to\infty} \max_\ell P_{\ell j}^n = \lim_{n\to\infty} \min_\ell P_{\ell j}^n > 0.$$

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



**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 $i$. $\{\pi_i; 1 \leq \mathsf{M}\}$ 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 $\lim_n [P^n] = \vec{e}\boldsymbol{\pi}$ where $\vec{e} = (1, 1, \ldots, 1)^\mathsf{T}$ and $\boldsymbol{\pi}$ satisfies $\boldsymbol{\pi}[P] = \boldsymbol{\pi}$ and $\boldsymbol{\pi}\vec{e} = 1$.**

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 $[P]\boldsymbol{\nu} = \lambda\boldsymbol{\nu}$ for some $\boldsymbol{\nu} \neq 0$, and iff $\boldsymbol{\pi}[P] = \lambda\boldsymbol{\pi}$ for some $\boldsymbol{\pi} \neq 0$.

$\vec{e}$ is always a right eigenvector of $[P]$ with eigenvalue 1, so there is always a left eigenvector $\boldsymbol{\pi}$.

$\det[P - \lambda I]$ 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

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

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

$$[P] = \sum_{i=1}^{\mathsf{M}} \lambda_i \, \vec{\nu}^{(i)} \, \vec{\pi}^{(i)}$$

$$[P^n] \; = \; [U][\Lambda^n][U^{-1}] \; = \; \sum_{i=1}^{\mathsf{M}} \lambda_i^n \vec{\nu}^{(i)} \vec{\pi}^{(i)}$$

**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 $\boldsymbol{\nu}$ 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 $\mathbb{R}^{\mathsf{M}}$, then $P_{ij}^n$ converges to $\pi_j$ as $\lambda_2^n$ for a unichain where $|\lambda_2|$ is the is the second largest magnitude eigenvalue.**

**If the eigenvectors do not span $\mathbb{R}^{\mathsf{M}}$, then $[P^n] = [U][J][U^{-1}]$ where $[J]$ is a Jordan form.**



## Renewal processes

**Thm: For a renewal process (RP) with mean inter-renewal interval $\overline{X} > 0$,**

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\overline{X}} \qquad \textbf{W.P.1.}$$

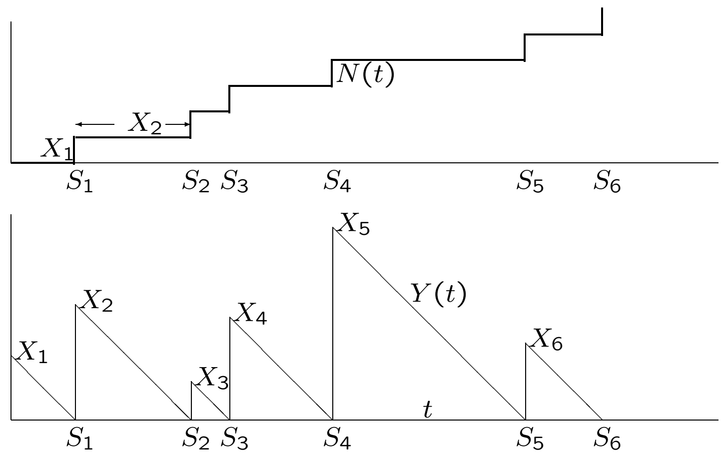
**This also holds if $\overline{X} = \infty$.**

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

**There is also the elementary renewal theorem, which says that**

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

## Residual life




**The integral of $Y(t)$ over $t$ is a sum of terms $X_n^2/2$.**



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

$$\lim_{t\to\infty} \frac{\int_{\tau=0}^{t} Y(\tau)\, d\tau}{t} = \frac{\mathsf{E}\left[X^2\right]}{2\mathsf{E}\left[X\right]} \qquad \textbf{W.P.1}$$

**The time average duration is**

$$\lim_{t\to\infty} \frac{\int_{\tau=0}^{t} X(\tau)\, d\tau}{t} = \frac{\mathsf{E}\left[X^2\right]}{\mathsf{E}\left[X\right]} \qquad \textbf{W.P.1}$$

**For PP, this is twice $\mathsf{E}[X]$. Big intervals contribute in two ways to duration.**

Residual life and duration are examples of renewal reward functions.

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

Thus reward over a renewal interval is

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

$$\lim_{t\to\infty} \frac{1}{t} \int_{\tau=0}^{t} R(\tau)\, d\tau = \frac{\mathsf{E}[R_n]}{\overline{X}} \qquad \textbf{W.P.1}$$

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 $\mathbb{I}_{\{J=n\}}$ is a function of $\{X_1, X_2, \dots, X_n\}$.

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 $\mathsf{E}[J] < \infty$, then the sum $S_J = X_1 + X_2 + \cdots + X_J$ at the stopping trial $J$ satisfies

$$\mathsf{E}[S_J] = \overline{X}\mathsf{E}[J].$$

**Wald: Let $\{X_n; n \geq 1\}$ be IID rv's, each of mean $\overline{X}$. If $J$ is a stopping time for $\{X_n; n \geq 1\}$, $\mathsf{E}[J] < \infty$, and $S_J = X_1 + X_2 + \cdots + X_J$, then**

$$\mathsf{E}[S_J] = \overline{X}\mathsf{E}[J]$$

**In many applications, where $X_n$ and $S_n$ are nonnegative rv's , the restriction $\mathsf{E}[J] < \infty$ is not necessary.**

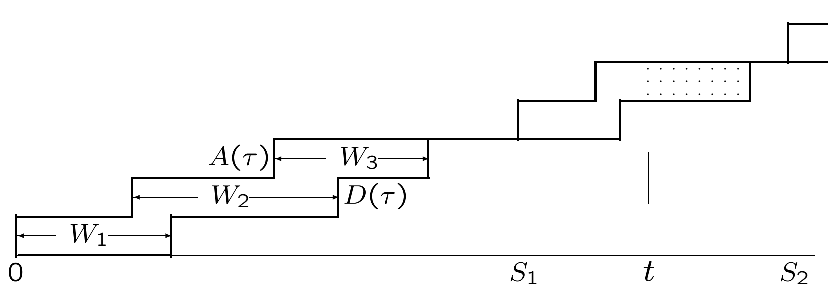
**For cases where $X$ is positive or negative, it is necessary as shown by 'stop when you're ahead.'**



## 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 $\overline{L}$ be the time average number in system,**

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

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

$$\begin{aligned}
\overline{W} &= \lim_{t\to\infty} \frac{1}{A(t)} \sum_{i=1}^{A(t)} W_i \\
&= \lim_{t\to\infty} \frac{t}{A(t)} \lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{A(t)} W_i \\
&= \overline{L}/\lambda
\end{aligned}$$

6.262 Discrete Stochastic Processes
Spring 2011