

Lecture 1: Introduction and Probability review

Outline:

- Probability in the real world
- Probability as a branch of mathematics
- Discrete stochastic processes
- Processes to be studied
- When, where, and how is this useful?
- The axioms of probability theory
- Independent events and experiments
- Random variables

1

Probability in the real world

Games of chance started in ancient civilizations. Along with a propensity to gamble, people have an intuitive sense of likelihood and average behavior.

Games of chance are essentially repeatable, and experimental verification of likelihoods is essentially possible.

Most of life's decisions involve uncertainty. Wise people learn to associate some sense of likelihood with uncertain possibilities.

Probability is most useful when repeatability under essentially the same conditions occurs.

2

Essentially – essentially – essentially ??? In trying to be precise, many problems emerge.

In flipping a coin, the outcome depends on initial velocity and orientation, the coin surfaces, and the ground surface. Nothing is random here. Subsequent tosses are also related through the coin and the flipper.

Important questions involving uncertainty are far harder to make sense of. What is the probability of another catastrophic oil spill in the coming year? What is the probability that Google stock will double in 5 years?

3

Probability as a branch of mathematics

Despite discomfort with what probability ‘means,’ people have felt comfortable using combinatorics and symmetry to create probabilities for events in all areas of science and life.

Going one step further, standard models are created where events have probabilities and there are sensible rules for working with these probabilities.

Students are given a well-specified model and calculate various quantities. Heads and tails are equiprobable and subsequent tosses are ‘independent.’

Everyone is happy. Students compute; professors write papers; business and government leaders obtain questionable models and data on which they can blame failures.

4

The use of probability models has 2 major problems:

First, how do you make a probability model for a real world problem?

Partial answer: Learn about estimation and decisions within standard models. Then learn a great deal about the real-world problem. Then use common sense and tread lightly.

Better answer: Try oversimplified models first. Use the mathematics of those simple models to help understand the real problem. Then in multiple stages, add to and modify the models to understand the original problem better.

Usually no model is perfect (look at coin tossing).

Alfred North Whitehead: Seek simplicity and distrust it.

5

Second problem: How do you make a probability model that has no hidden paradoxes?

Everyone's answer: Follow Kolmogorov's axioms of probability.

Kolmogorov did this in 1933, finally putting probability on a firm mathematical foundation and opening the field to steady progress.

These axioms essentially say that probability theory is a branch of measure theory.

These axioms are needed here to avoid paradoxes, but for the topics treated, measure theory is not needed and will not be used.

6

Discrete stochastic processes

A stochastic process is a special type of probability model in which the sample points represent functions in time.

It often can be viewed as a sequence of random variables evolving in time. Often there is a continuum of random variables, one for each real valued instant of time.

A discrete stochastic process is a stochastic process where either the random variables are discrete in time or the set of possible sample values is discrete.

It is not important to define which stochastic processes are discrete precisely.

7

Processes to be studied

Counting processes — Each sample point is a sequence of ‘arrival’ times. Special cases are Poisson processes (chap. 2) and Renewal processes (chap. 4).

Markov processes — The future state depends on the past only through the present. Special cases are Finite Markov chains (chap. 3), countable Markov chains (chap. 5) and Markov processes with countable state spaces (chap. 6).

Random Walks and martingales (chap. 7)

We will study various mixtures of these, particularly standard models for many applications. See table of contents (or text itself) for more detail.

8

When, where, and how is this useful?

Broad answer: Probability and stochastic processes are an important adjunct to rational thought about all human and scientific endeavor.

Narrow answer: Probability and stochastic processes are essential components of the following areas:

Communication systems and networks; computer systems; Queueing in all areas; risk management in all areas; catastrophe management; failures in all types of systems; operations research; biology; medicine; optical systems; control systems; etc.

9

The axioms of probability theory

Probability models have 3 components: a sample space Ω , which is an arbitrary set of sample points; a collection of events, each of which is a subset of Ω ; and a probability measure, which assigns a probability (a number in $[0, 1]$) to each event. The collection of events satisfies the following axioms:

1. Ω is an event.
2. If A_1, A_2, \dots , are events, then $\bigcup_{n=1}^{\infty} A_n$ is an event.
3. If A is an event, the complement A^c is an event.

Not all subsets need be events. Usually each sample point is taken to be a singleton event. Then non-events are weird, often necessary, but usually ignorable.

10

The empty set ϕ is Ω^c , so is an event.

If all sample points are singleton events, then all finite and countable sets are events (i.e., they are finite and countable unions of singleton sets).

From deMorgan's law,

$$[\bigcup_n A_n]^c = \bigcap_n A_n^c.$$

so countable intersections of events are events. All combinations of intersections and unions of events are also events.

11

The probability measure on events satisfies the following axioms:

1. $\Pr\{\Omega\} = 1$.
2. If A is an event, then $\Pr\{A\} \geq 0$.
3. If A_1, A_2, \dots are disjoint events, then

$$\Pr\left\{\bigcup_{n=1}^{\infty} A_n\right\} = \sum_{n=1}^{\infty} \Pr\{A_n\} = \lim_{m \rightarrow \infty} \sum_{n=1}^m \Pr\{A_n\}$$

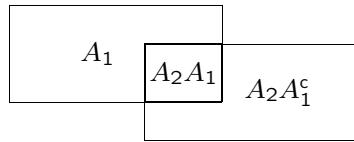
It's surprising that this is all that is needed to avoid paradoxes. A few simple consequences are

$$\begin{aligned}\Pr\{\phi\} &= 0 \\ \Pr\{A^c\} &= 1 - \Pr\{A\} \\ \Pr\{A\} &\leq \Pr\{B\} \leq 1 \quad \text{for } A \subseteq B\end{aligned}$$

12

Another consequence is the union bound,

$$\Pr\left\{\bigcup_n A_n\right\} \leq \sum_n \Pr\{A_n\}; \quad \text{finite or countable } n$$



$$\begin{aligned} A_1 \cup A_2 &= A_1 \cup A_2 A_1^c \\ \Pr\{A_1 \cup A_2\} &= \Pr\{A_1\} + \Pr\{A_2 A_1^c\} \\ &\leq \Pr\{A_1\} + \Pr\{A_2\} \end{aligned}$$

These axioms probably look ho-hum, and we ignore them much of the time. They are often needed for infinite sums and limits.

As in elementary probability courses, we emphasize random variables and expectations. The axioms, however, say that events and probabilities of events are the fundamental quantities.

13

Independent events and experiments

Two events A_1 and A_2 are independent if $\Pr\{A_1 A_2\} = \Pr\{A_1\} \Pr\{A_2\}$.

Given two probability models, a combined model can be defined in which, first, the sample space Ω is the Cartesian product $\Omega_1 \times \Omega_2$, and, second, for every event A in model 1 and B in model 2, $\Pr\{AB\} = \Pr\{A\} \Pr\{B\}$.

The two original models are then said to be independent in the combined model. We won't try to develop notation for this.

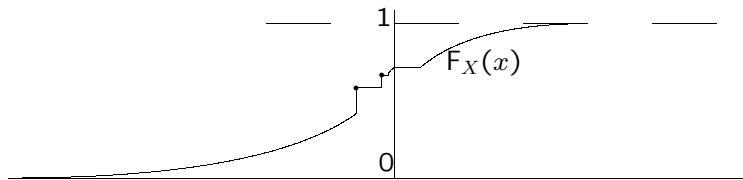
If the axioms are satisfied in each separately, they can be satisfied in the combined model, so complex models can be formed from simpler models.

14

Random variables (rv's)

Def: A rv X (or $X(\omega)$) is a function from Ω to \mathbb{R} . This function must satisfy the constraint that $\{\omega : X(\omega) \leq a\}$ is an event for all $a \in \mathbb{R}$. Also, if X_1, X_2, \dots, X_n are each rv's, then $\{\omega : X_1(\omega) \leq a_1; \dots, X_n(\omega) \leq a_n\}$ is an event for all a_1, \dots, a_n each in \mathbb{R} .

Every rv X has a distribution function $F_X(x) = \Pr\{X \leq x\}$. It's a non-decreasing function from 0 to 1.



15

If X maps only into a finite or countable set of values, it is discrete and has a probability mass function (PMF) where $p_X(x) = \Pr\{X = x\}$.

If $dF_X(x)/dx$ exists and is finite for all x , then X is continuous and has a density, $f_X(x) = dF_X(x)/dx$.

If X has discrete and continuous components, it's sometimes useful to view it as a density with impulses.

In general, $F_X(x) = \Pr\{X \leq x\}$ always exists. Because $X = x$ is included in $X \leq x$, we see that if $F_X(x)$ has a jump at x , then $F_X(x)$ is the value at the top of the jump.

16

Theoretical nit-pick: $F_X(x)$ must be continuous from the right, i.e., $\lim_{k \rightarrow \infty} F_X(x + 1/k) = F_X(x)$.

This seems obvious, since for a discontinuity at x , $F_X(x)$ is the value at the top (right) of the jump.

Proof: Let $A_k = \{\omega : X(\omega) > x + \frac{1}{k}\}$. Then $A_{k-1} \subseteq A_k$ for each $k > 1$.

$$\begin{aligned} \{\omega : X(\omega) > x\} &= \bigcup_{k=1}^{\infty} A_k \\ \Pr\{X > x\} &= \Pr\left\{\bigcup_k A_k\right\} = \lim_k \Pr\{A_k\} = \lim_k \Pr\left\{X > x + \frac{1}{k}\right\} \end{aligned}$$

Center step: let $B_1 = A_1$; $B_k = A_k - A_{k-1}$ for $k > 1$.
Then $\{B_k; k \geq 1\}$ are disjoint.

$$\begin{aligned} \Pr\left\{\bigcup_k A_k\right\} &= \Pr\left\{\bigcup_k B_k\right\} = \sum_{k=1}^{\infty} \Pr\{B_k\} \\ &= \lim_{k \rightarrow \infty} \sum_{m=1}^k B_m = \lim_{k \rightarrow \infty} A_k \end{aligned}$$

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Lecture 2: More review; the Bernoulli process**Outline:**

- **Expectations**
- **Indicator random variables**
- **Multiple random variables**
- **IID random variables**
- **Laws of large numbers in pictures**
- **The Bernoulli process**
- **Central limit theorem for Bernoulli**

1

Expectations

The distribution function of a rv X often contains more detail than necessary. The expectation $\bar{X} = E[X]$ is sometimes all that is needed.

$$E[X] = \sum_i a_i p_X(a_i) \quad \text{for discrete } X$$

$$E[X] = \int x f_X(x) dx \quad \text{for continuous } X$$

$$E[X] = \int F_X^c(x) dx \quad \text{for arbitrary nonneg } X$$

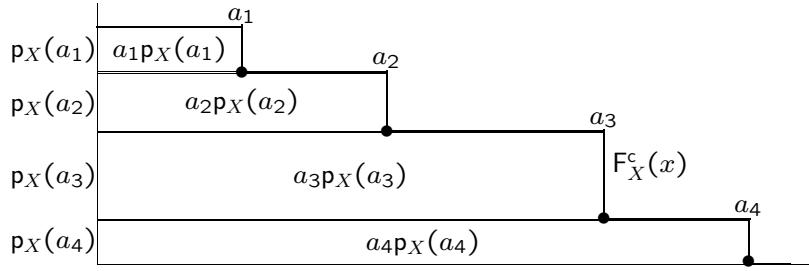
$$E[X] = \int_{-\infty}^0 F_X(x) dx + \int_0^\infty F_X^c(x) dx \quad \text{for arbitrary } X.$$

Almost as important is the standard deviation,

$$\sigma_X = \sqrt{E[(X - \bar{X})^2]}$$

2

Why is $E[X] = \int F_X^c(x) dx$ for arbitrary nonneg X ?
Look at discrete case. Then $\int F_X^c(x) dx = \sum_i a_i p_X(a_i)$.



If X has a density, the same argument applies to every Riemann sum for $\int_x x f_X(s) dx$ and thus to the limit.

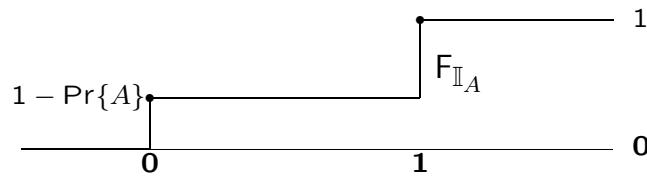
It is simpler and more fundamental to take $\int F_X^c(x) dx$ as the general definition of $E[X]$. This is also useful in solving problems

3

Indicator random variables

For every event A in a probability model, an indicator rv \mathbb{I}_A is defined where $\mathbb{I}_A(\omega) = 1$ for $\omega \in A$ and $\mathbb{I}_A(\omega) = 0$ otherwise. Note that \mathbb{I}_A is a binary rv.

$$p_{\mathbb{I}_A}(0) = 1 - \Pr\{A\}; \quad p_{\mathbb{I}_A}(1) = \Pr\{A\}.$$



$$E[\mathbb{I}_A] = \Pr\{A\} \quad \sigma_{\mathbb{I}_A} = \sqrt{\Pr\{A\}(1 - \Pr\{A\})}$$

Theorems about rv's can thus be applied to events.

4

Multiple random variables

Is a random variable (rv) X specified by its distribution function $F_X(x)$?

No, the relationship between rv's is important.

$$F_{XY}(x, y) = \Pr\left\{\{\omega : X(\omega) \leq x\} \cap \{\omega : Y(\omega) \leq y\}\right\}$$

The rv's X_1, \dots, X_n are independent if

$$F_{\vec{X}}(x_1, \dots, x_n) = \prod_{m=1}^n F_{X_m}(x_m) \quad \text{for all } x_1, \dots, x_n$$

This product form carries over for PMF's and PDF's.

5

For discrete rv's, independence is more intuitive when stated in terms of conditional probabilities.

$$p_{X|Y}(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)}$$

Then X and Y are independent if $p_{X|Y}(x|y) = p_X(x)$ for all sample points x and y . This essentially works for densities, but then $\Pr\{Y = y\} = 0$ (see notes). This is not very useful for distribution functions.

NitPick: If X_1, \dots, X_n are independent, then all subsets of X_1, \dots, X_n are independent. (This isn't always true for independent events).

6

IID random variables

The random variables X_1, \dots, X_n are independent and identically distributed (IID) if for all x_1, \dots, x_n

$$F_{\vec{X}}(x_1, \dots, x_n) = \prod_{k=1}^n F_X(x_k)$$

This product form works for PMF's and PDF's also.

Consider a probability model in which \mathbb{R} is the sample space and X is a rv.

We can always create an extended model in which \mathbb{R}^n is the sample space and X_1, X_2, \dots, X_n are IID rv's. We can further visualize $n \rightarrow \infty$ where X_1, X_2, \dots is a stochastic process of IID variables.

7

We study the sample average, $S_n/n = (X_1 + \dots + X_n)/n$. The laws of large numbers say that S_n/n 'essentially becomes deterministic' as $n \rightarrow \infty$.

If the extended model corresponds to repeated experiments in the real world, then S_n/n corresponds to the arithmetic average in the real world.

If X is the indicator rv for event A , then the sample average is the relative frequency of A .

Models can have two types of difficulties. In one, a sequence of real-world experiments are not sufficiently similar and isolated to correspond to the IID extended model. In the other, the IID extension is OK but the basic model is not.

We learn about these problems here through study of the models.

8

Science, symmetry, analogies, earlier models, etc.
are all used to model real-world situations.

Trivial example: Roll a white die and a red die.
There are 36 sample outcomes, $(i, j), 1 \leq i, j \leq 6$,
taken as equiprobable by symmetry.

Roll 2 indistinguishable white dice. The white and
red outcomes (i, j) and (j, i) for $i \neq j$ are now in-
distinguishable. There are now 21 ‘finest grain’
outcomes, but no sane person would use these as
sample points.

The appropriate sample space is the ‘white/red’
sample space with an ‘off-line’ recognition of what
is distinguishable.

Neither the axioms nor experimentation motivate
this model, i.e., modeling requires judgement and
common sense.

9

Comparing models for similar situations and analyzing limited and defective models helps in clarifying fuzziness in a situation of interest.

Ultimately, as in all of science, some experimentation is needed.

The outcome of an experiment is a sample point, not a probability.

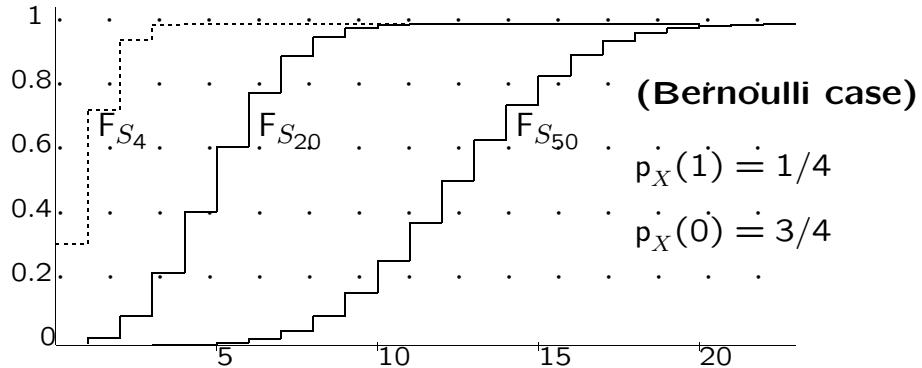
Experimentation with probability requires multiple trials. The outcome is modeled as a sample point in an extended version of the original model.

Experimental tests of an original model come from the laws of large numbers in the context of an extended model.

10

Laws of large numbers in pictures

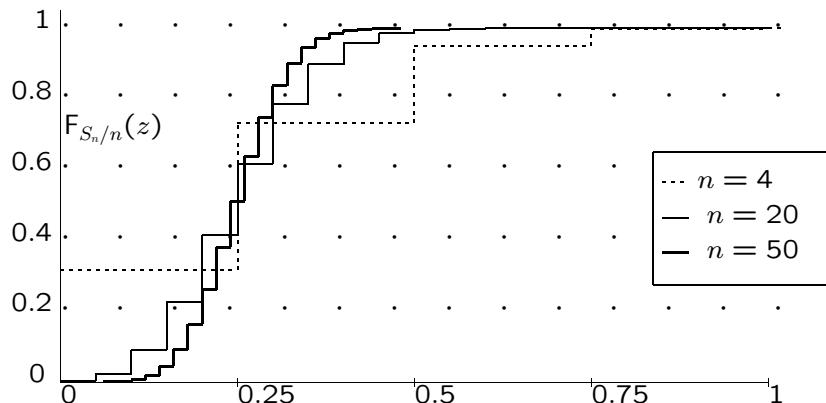
Let X_1, X_2, \dots, X_n be IID rv's with mean \bar{X} , variance σ^2 . Let $S_n = X_1 + \dots + X_n$. Then $\sigma_{S_n}^2 = n\sigma^2$.



The center of the distribution varies with n and the spread (σ_{S_n}) varies with \sqrt{n} .

11

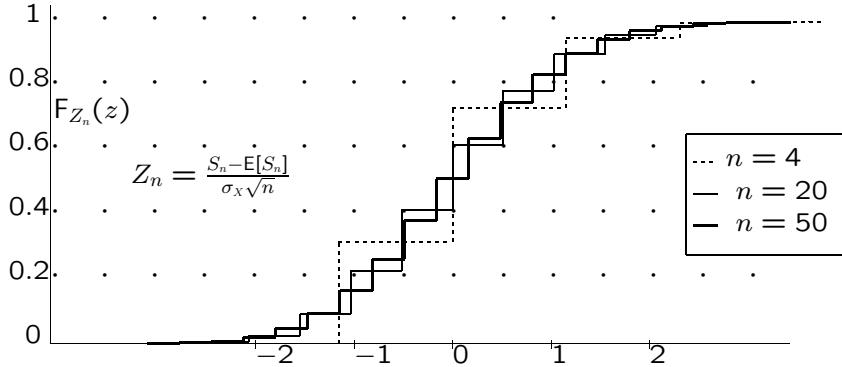
The sample average is S_n/n , which is a rv of mean \bar{X} and variance σ^2/n .



The center of the distribution is \bar{X} and the spread decreases with $1/\sqrt{n}$.

12

Note that $S_n - n\bar{X}$ is a zero mean rv with variance $n\sigma^2$. Thus $\frac{S_n - n\bar{X}}{\sqrt{n}\sigma}$ is zero mean, unit variance.



Central limit theorem:

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

13

The Bernoulli process

$$S_n = Y_1 + \cdots + Y_n \quad p_Y(1) = p > 0, \quad p_Y(0) = 1 - p = q > 0$$

The n -tuple of k 1's followed by $n-k$ 0's has probability $p^k q^{n-k}$.

Each n tuple with k ones has this same probability.
For $p < 1/2$, $p^k q^{n-k}$ is largest at $k = 0$ and decreasing in k to $k = n$.

There are $\binom{n}{k}$ n -tuples with k 1's. This is increasing in k for $k < n/2$ and then decreasing. Altogether,

$$p_{S_n}(k) = \binom{n}{k} p^k q^{n-k}$$

14

$$p_{S_n}(k) = \binom{n}{k} p^k q^{n-k}$$

To understand how this varies with k , consider

$$\begin{aligned} \frac{p_{S_n}(k+1)}{p_{S_n}(k)} &= \frac{n!}{(k+1)!(n-k-1)!} \frac{k!(n-k)!}{n!} \frac{p^{k+1}q^{n-k-1}}{p^k q^{n-k}} \\ &= \frac{n-k}{k+1} \frac{p}{q} \end{aligned}$$

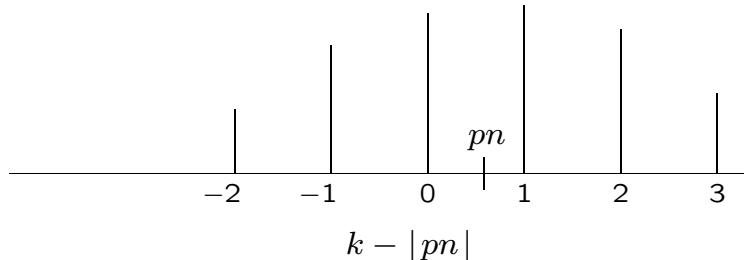
This is strictly decreasing in k . It also satisfies

$$\frac{p_{S_n}(k+1)}{p_{S_n}(k)} \begin{cases} < 1 & \text{for } k \geq pn \\ \approx 1 & \text{for } k < pn < k+1 \\ > 1 & \text{for } k+1 \leq pn \end{cases}$$

15

$$\frac{p_{S_n}(k+1)}{p_{S_n}(k)} = \frac{n-k}{k+1} \frac{p}{q} \quad (1)$$

$$\frac{p_{S_n}(k+1)}{p_{S_n}(k)} \begin{cases} < 1 & \text{for } k \geq pn \\ \approx 1 & \text{for } k < pn < k+1 \\ > 1 & \text{for } k+1 \leq pn \end{cases}$$



In other words, $p_{S_n}(k)$, for fixed n , is increasing with k for $k < pn$ and decreasing for $k > pn$.

16

CLT for Bernoulli process

$$\frac{p_{S_n}(k+1)}{p_{S_n}(k)} = \frac{n-k}{k+1} \frac{p}{q}$$

We now use this equation for large n where k is relatively close to pn . To simplify the algebra, assume pn is integer and look at $k = pn + i$ for relatively small i . Then

$$\begin{aligned} \frac{p_{S_n}(pn+i+1)}{p_{S_n}(pn+i)} &= \frac{n-pn-i}{pn+i+1} \frac{p}{q} = \frac{nq-i}{pn+i+1} \frac{p}{q} \\ &= \frac{1 - \frac{i}{nq}}{1 + \frac{i+1}{np}} \\ \ln \left[\frac{p_{S_n}(pn+i+1)}{p_{S_n}(pn+i)} \right] &= \ln \left[1 - \frac{i}{nq} \right] - \ln \left[1 + \frac{i+1}{np} \right] \end{aligned}$$

17

Recall that $\ln(1+x) \approx x - x^2/2 + \dots$ for $|x| \ll 1$.

$$\begin{aligned} \ln \left[\frac{p_{S_n}(pn+i+1)}{p_{S_n}(pn+i)} \right] &= \ln \left[1 - \frac{i}{nq} \right] - \ln \left[1 + \frac{i+1}{np} \right] \\ &= -\frac{i}{nq} - \frac{i}{np} - \frac{1}{np} + \dots \\ &= -\frac{i}{npq} - \frac{1}{np} + \dots \end{aligned}$$

where we have used $1/p + 1/q = 1/pq$ and the neglected terms are of order i^2/n^2 .

This says that these log of unit increment terms are essentially linear in i . We now have to combine these unit incremental terms.

18

$$\ln \left[\frac{p_{S_n}(pn+i+1)}{p_{S_n}(pn+i)} \right] = -\frac{i}{npq} - \frac{1}{np} + \dots$$

Expressing an increment of j terms as a telescoping sum of j unit increments,

$$\begin{aligned} \ln \left[\frac{p_{S_n}(pn+j)}{p_{S_n}(pn)} \right] &= \sum_{i=0}^{j-1} \ln \left[\frac{p_{S_n}(pn+i+1)}{p_{S_n}(pn+i)} \right] \\ &= \sum_{i=0}^{j-1} -\frac{i}{npq} - \frac{1}{np} + \dots \\ &= -\frac{j(j-1)}{2npq} - \frac{j}{np} + \dots \approx \frac{-j^2}{2npq} \end{aligned}$$

where we have used the fact that $1+2+\dots+j-1 = j((j-1)/2)$. We have also ignored terms linear in j since they are of the same order as a unit increment in j .

19

Finally,

$$\begin{aligned} \ln \left[\frac{p_{S_n}(pn+j)}{p_{S_n}(pn)} \right] &\approx \frac{-j^2}{2npq} \\ p_{S_n}(pn+j) &\approx p_{S_n}(pn) \exp \left[\frac{-j^2}{2npq} \right] \end{aligned}$$

This applies for j both positive and negative, and is a quantized version of a Gaussian distribution, with the unknown scaling constant $p_{S_n}(pn)$. Choosing this to get a PMF,

$$p_{S_n}(pn+j) \approx \frac{1}{\sqrt{2\pi npq}} \exp \left[\frac{-j^2}{2npq} \right],$$

which is the discrete PMF form of the central limit theorem. See Section 1.5.3 for a different approach.

20

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Lecture 3: Laws of large numbers, convergence

Outline:

- Review of probability models
- Markov, Chebychev, Chernoff bounds
- Weak law of large numbers and convergence
- Central limit theorem and convergence
- Convergence with probability 1

1

Review of probability models

Probability models are natural for real-world situations that are repeatable, using trials that

- have the same initial conditions
- are essentially isolated from each other
- have a fixed set of possible outcomes
- have essentially ‘random’ individual outcomes.

For any model, an extended model for a sequence or an n -tuple of IID repetitions is well-defined.

Relative frequencies and sample averages (in the extended model) ‘become deterministic’ and can be compared with real-world relative frequencies and sample averages in the repeated experiment.

2

The laws of large numbers (LLN's) specify what ‘become deterministic’ means.

They only operate within the extended model, but provide our only truly experimental way to compare the model with repeated trials of the real-world experiment.

Probability theory provides many many consistency checks and ways to avoid constant experimentation.

Common sense, knowledge of the real-world system, focus on critical issues, etc. often make repeated trials unnecessary.

The determinism in large numbers of trials underlies much of the value of probability.

3

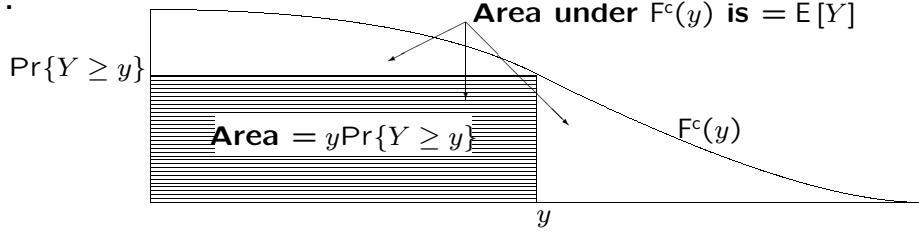
Markov, Chebychev, Chernoff bounds

Inequalities, or bounds, play an unusually large role in probability. Part of the reason is their frequent use in limit theorems and part is an inherent imprecision in probability applications.

One of the simplest and most useful bounds is the Markov inequality: If Y is a non-negative rv with an expectation $E[Y]$, then for any real $y > 0$,

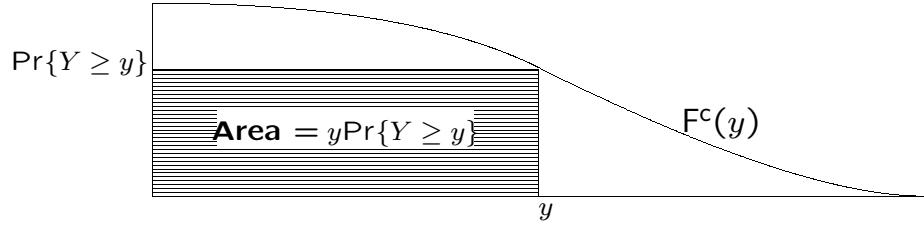
$$\Pr\{Y \geq y\} \leq \frac{E[Y]}{y}$$

Pf:



4

Markov inequality: $\Pr\{Y \geq y\} \leq \frac{\mathbb{E}[Y]}{y}$



Note that the Markov bound is usually very loose. It is tight (satisfied with equality) if Y is binary with possible values 0 and y .

The Markov bound decreases very slowly (as $1/y$) with increasing y .

5

The Chebyshev inequality: If Z has a mean $\mathbb{E}[Z] = \bar{Z}$ and a variance, σ_Z^2 , then for any $\epsilon > 0$,

$$\Pr\{|Z - \bar{Z}| \geq \epsilon\} \leq \frac{\sigma_Z^2}{\epsilon^2} \quad (1)$$

Pf: Let $Y = (Z - \bar{Z})^2$. Then $\mathbb{E}[Y] = \sigma_Z^2$ and for any $y > 0$,

$$\Pr\{Y \geq y\} \leq \sigma_Z^2/y; \quad \Pr\{\sqrt{Y} \geq \sqrt{y}\} \leq \sigma_Z^2/y$$

Now $\sqrt{Y} = |Z - \bar{Z}|$. Setting $\epsilon = \sqrt{y}$ yields (1).

Chebychev requires a variance, but decreases as $1/\epsilon^2$ with increasing distance ϵ from the mean.

6

The Chernoff bound: For any $z > 0$ and any $r > 0$ such that the moment generating function $g_Z(r) = E[e^{rZ}]$ exists,

$$\Pr\{Z \geq z\} \leq g_Z(r) \exp(-rz) \quad (2)$$

Pf: Let $Y = e^{rZ}$. Then $E[Y] = g_Z(r)$. For any $y > 0$, Markov says,

$$\Pr\{Y \geq y\} \leq g_Z(r)/y; \quad \Pr\{e^{rZ} \geq e^{rz}\} \leq g_Z(r)/e^{rz},$$

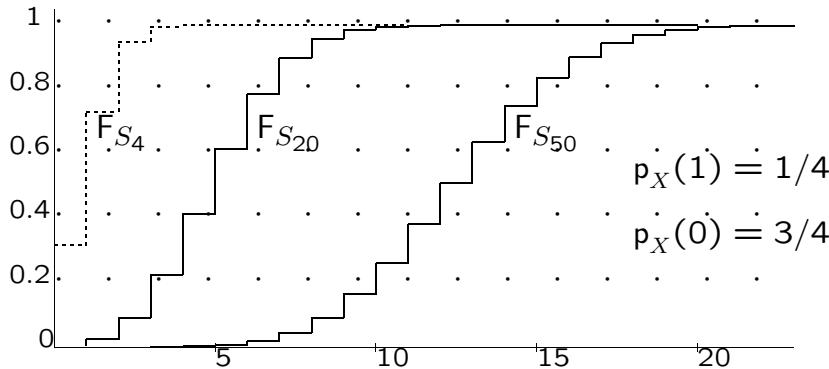
which is equivalent to (2).

This decreases exponentially with z and is useful in studying large deviations from the mean.

7

The weak law of large numbers and convergence

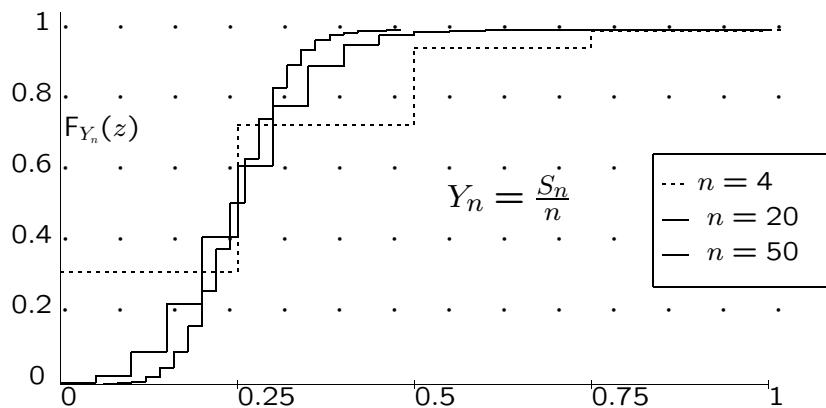
Let X_1, X_2, \dots, X_n be IID rv's with mean \bar{X} , variance σ^2 . Let $S_n = X_1 + \dots + X_n$. Then $\sigma_{S_n}^2 = n\sigma^2$.



The mean of the distribution varies with n and the standard deviation varies with \sqrt{n} .

8

The sample average is S_n/n , which is a rv of mean \bar{X} and variance σ^2/n .



The mean of the distribution is \bar{X} and the standard deviation decreases with $1/\sqrt{n}$.

9

$$\text{VAR} \left(\frac{S_n}{n} \right) = E \left[\left(\frac{S_n}{n} - \bar{X} \right)^2 \right] = \frac{\sigma^2}{n}. \quad (3)$$

$$\lim_{n \rightarrow \infty} E \left[\left(\frac{S_n}{n} - \bar{X} \right)^2 \right] = 0. \quad (4)$$

Note that (3) says more than (4), since it says the convergence is as $1/n$ and in fact it gives the variance explicitly. But (4) establishes a standard form of convergence of rv's called convergence in mean square.

Def: A sequence of rv's, Y_1, Y_2, \dots converges in mean square to a rv Y if

$$\lim_{n \rightarrow \infty} E \left[(Y_n - Y)^2 \right] = 0$$

10

The fact that S_n/n converges in mean square to \bar{X} doesn't tell us directly what might be more interesting: what is the probability that $|S_n/n - \bar{X}|$ exceeds ϵ as a function of ϵ and n ?

Applying Chebyshev to (3), however,

$$\Pr\left\{\left|\frac{S_n}{n} - \bar{X}\right| \geq \epsilon\right\} \leq \frac{\sigma^2}{n\epsilon^2} \quad \text{for every } \epsilon > 0 \quad (5)$$

One can get an arbitrary accuracy of ϵ between sample average and mean with probability $1 - \sigma^2/n\epsilon^2$, which can be made as close to 1 as we wish, by increasing n .

This gives us the weak law of large numbers (WLLN):

$$\lim_{n \rightarrow \infty} \Pr\left\{\left|\frac{S_n}{n} - \bar{X}\right| \geq \epsilon\right\} = 0 \quad \text{for every } \epsilon > 0.$$

11

WLLN: $\lim_{n \rightarrow \infty} \Pr\left\{\left|\frac{S_n}{n} - \bar{X}\right| \geq \epsilon\right\} = 0 \quad \text{for every } \epsilon > 0.$

We have proven this under the assumption that $S_n = \sum_{n=1}^n X_n$ where X_1, X_2, \dots , are IID with finite variance.

An equivalent statement (following from the definition of a limit of real numbers) is that for every $\delta > 0$,

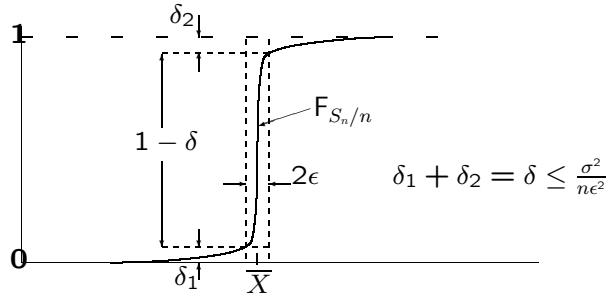
$$\Pr\left\{\left|\frac{S_n}{n} - \bar{X}\right| \geq \epsilon\right\} \leq \delta \quad \text{for all large enough } n. \quad (6)$$

Note that (6) tells us less about the speed of convergence than

$$\Pr\left\{\left|\frac{S_n}{n} - \bar{X}\right| \geq \epsilon\right\} \leq \frac{\sigma^2}{n\epsilon^2}$$

But (6) holds without a variance (if $E[|X|] < \infty$.)

12



What this says is that $\Pr\left\{\frac{S_n}{n} \leq x\right\}$ is approaching a unit step at \bar{X} as $n \rightarrow \infty$. For any fixed ϵ , δ goes to 0 as $n \rightarrow \infty$. If $\sigma_X < \infty$, then $\delta \rightarrow 0$ at least as $\sigma^2/n\epsilon^2$. Otherwise it might go to 0 more slowly.

Def: A sequence of rv's, Y_1, Y_2, \dots converges in probability to a rv Y if for every $\epsilon > 0, \delta > 0$,

$$\Pr\{|Y_n - Y| \geq \epsilon\} \leq \delta \quad \text{for all large enough } n$$

This means that $\{S_n/n; n \geq 1\}$ converges to \bar{X} in probability if $E[|X|] < \infty$.

13

Review: We saw that if σ_X exists and X_1, X_2, \dots are IID, then $\sigma_{S_n/n} = \sigma_X/\sqrt{n}$.

Thus S_n/n converges to \bar{X} in mean square. Chebychev then shows that S_n/n converges to \bar{X} in probability.

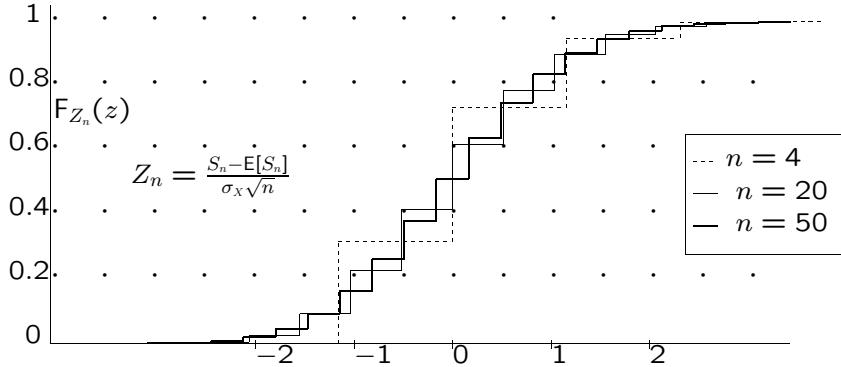
In the same way, if $\{Y_n; n \geq 1\}$ converges to Y in mean square, Chebychev show that it converges in probability.

That is, mean square convergence implies convergence in probability. The reverse is not true, since a variance is not required for the WLLN.

Finally, convergence in probability means that the distribution of $Y_n - Y$ approaches a unit step at 0.

14

Recall that $S_n - n\bar{X}$ is a zero mean rv with variance $n\sigma^2$. Thus $\frac{S_n - n\bar{X}}{\sqrt{n}\sigma}$ is zero mean, unit variance.



Central limit theorem:

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

15

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

Not only does $(S_n - n\bar{X})/\sqrt{n}\sigma_X$ have mean 0, variance 1 for all n , but it also becomes normal Gaussian.

We saw this for the Bernoulli case, but the general case is messy and the proof (by Fourier transforms) is not insightful.

The CLT applies to F_{S_n} , not to the PMF or PDF.

Def: A sequence Z_1, Z_2, \dots of rv's converges in distribution to Z if $\lim_{n \rightarrow \infty} F_{Z_n}(z) = F_Z(z)$ for all z where $F_Z(z)$ is continuous.

The CLT says that $(S_n - n\bar{X})/\sqrt{n}\sigma_X$ converges in distribution to Φ .

16

Convergence in distribution is almost a misnomer, since the rv's themselves do not necessarily become close to each other in any ordinary sense.

For example any sequence of IID rv's converge in distribution since they have the same distribution to start with.

Thm: Convergence in probability implies convergence in distribution.

Pf: Convergence of $\{Y_n; n \geq 1\}$ in probability means convergence to a unit step.

Thus convergence in mean square implies convergence in probability implies convergence in distribution.

17

Paradox: The CLT says something very strong about how S_n/n converges to \bar{X} , but convergence in distribution is a very weak form of convergence.

Resolution: The rv's that converge in distribution in the CLT are $(S_n - n\bar{X})/\sqrt{n}\sigma_X$. Those that converge in probability to 0 are $(S_n - n\bar{X})/n$, a squashed version of $(S_n - n\bar{X})/\sqrt{n}\sigma_X$.

The CLT, for $0 < \sigma_X < \infty$, for example, says that $\lim_{n \rightarrow \infty} \Pr\{(S_n - n\bar{X})/n \leq 0\} = 1/2$. This can not be deduced from the WLLN.

18

Convergence with probability 1

A rv is a far more complicated thing than a number. Thus it is not surprising that there are many types of convergence of a sequence of rv's.

A very important type is convergence with probability 1 (WP1). We introduce convergence WP1 here and discuss it more in Chap. 4.

The definition is deceptively simple.

Def: A sequence Z_1, Z_2, \dots , of rv's converges WP1 to a rv Z if

$$\Pr\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} Z_n(\omega) = Z(\omega)\right\} = 1$$

19

$$\Pr\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} Z_n(\omega) = Z(\omega)\right\} = 1$$

In order to parse this, note that each sample point maps into a sequence of real numbers, $Z_1(\omega), Z_2(\omega), \dots$.

Some of those sequences of real numbers have a limit, and in some cases, that limit is $Z(\omega)$. Convergence WP1 means that the set ω for which $Z_1(\omega), Z_2(\omega)$ has a limit, and that limit is $Z(\omega)$, is an event and that the probability of that event is 1.

One small piece of complexity that can be avoided here is looking at the sequence $\{Y_i = Z_i - Z; i \geq 1\}$ and asking if that sequence converges to 0 WP1.

20

The strong law of large numbers (**SLLN**) says the following: Let X_1, X_2, \dots be IID rv's with $E[|X|] < \infty$. Then $\{S_n/n; n \geq 1\}$ converges to \bar{X} WP1. In other words, all sample paths of $\{S_n/n; n \geq 1\}$ converge to \bar{X} except for a set of probability 0.

These are the same conditions under which the **WLLN** holds. We will see, when we study renewal processes, that the **SLLN** is considerably easier to work with than the **WLLN**.

It will take some investment of time to feel at home with the **SLLN**, (and in particular to have a real sense about these sets of probability 1) and we put that off until chap 4.

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Lecture 4: Poisson (the perfect arrival process)**Outline:**

- Review: Convergence and sequences of IID rv's.
- Arrival processes
- Poisson processes and exponential rv's.
- Stationary and independent increments
- The Erlang and Poisson distributions
- Alternate definitions of Poisson process
- Relation to Bernoulli process

1

Review: Convergence and sequences of IID rv's.

Def: A sequence Z_1, Z_2, \dots , of random variables, converges in distribution to a random variable Z if $\lim_{n \rightarrow \infty} F_{Z_n}(z) = F_Z(z)$ at each z for which $F_Z(z)$ is continuous.

Example: (CLT) If X_1, X_2, \dots are IID with variance σ^2 , $S_n = \sum_{i=1}^n X_i$, and $Z_n = (S_n - n\bar{X})/\sigma\sqrt{n}$, then Z_1, Z_2, \dots converges in distribution to $\mathcal{N}(0, 1)$.

Example: If X_1, X_2, \dots , are IID with mean \bar{X} and $S_n = \sum_{i=1}^n X_i$, then $\{S_n/n; n \geq 1\}$ converges in distribution to the deterministic rv \bar{X} .

2

Def: A sequence Z_1, Z_2, \dots , of random variables converges in probability to a random variable Z if $\lim_{n \rightarrow \infty} \Pr\{|Z_n - Z| > \epsilon\} = 0$ for every $\epsilon > 0$ (alternatively, if for every $\epsilon > 0, \delta > 0$, $\Pr\{|Z_n - Z| > \epsilon\} \leq \delta$ for all large enough n .)

Example: (WLLN) If X_1, X_2, \dots , are IID with mean \bar{X} and $S_n = \sum_{i=1}^n X_i$, then $\{S_n/n; n \geq 1\}$ converges in probability to the deterministic rv \bar{X} . (see Thms. 1.5.1 and 1.5.3 of text)

Def: A sequence Z_1, Z_2, \dots , of rv's converges in mean square to a rv Z if $\lim_{n \rightarrow \infty} E[|Z_n - Z|^2] = 0$.

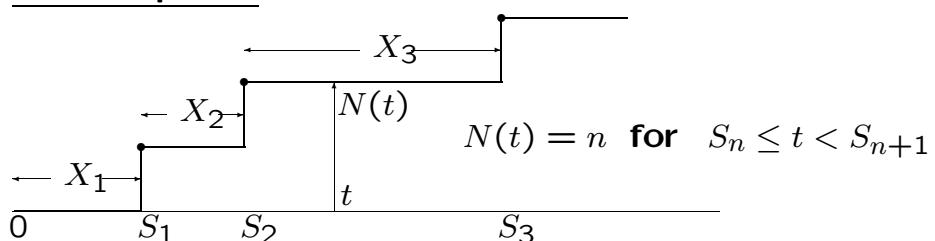
Convergence in mean square implies convergence in probability implies convergence in distribution.

3

Arrival processes

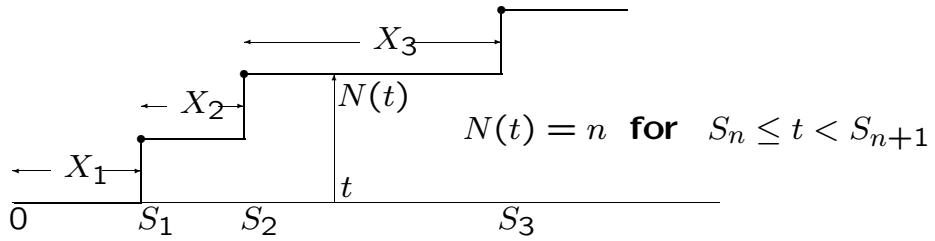
Def: An arrival process is a sequence of increasing rv's $0 < S_1 < S_2 < \dots$ where $S_{i-1} < S_i$ means that $S_i - S_{i-1} = X_i$ is a positive rv, i.e., $F_{X_i}(0) = 0$.

The differences $X_i = S_i - S_{i-1}$ for $i \geq 2$ and $X_1 = S_1$ are called interarrival times and the S_i are called arrival epochs.



For each $t > 0$, $N(t)$ is the number of arrivals in $(0, t]$. We call $\{N(t); t > 0\}$ an arrival counting process.

4

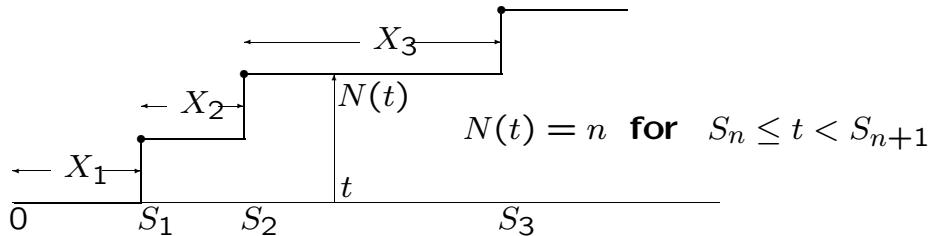


A sample path or sample function of the process is a sequence of sample values, $S_1 = s_1, S_2 = s_2, \dots$

Each sample path corresponds to a particular stair case function and the process can be viewed as the ensemble (with joint probability distributions) of such stair case functions.

The figure shows how the arrival epochs, interarrival times, and counting variables are interrelated for a generic stair case function.

5



It can also be seen that any sample path can be specified by the sample values of $N(t)$ for all t , by S_i for all i , or by X_i for all i , so that essentially an arrival process can be specified by the counting process, the interarrival times, or the arrival epochs.

The major relation we need to relate the counting process $\{N(t); t > 0\}$ to the arrival process is

$$\{S_n \leq t\} = \{N(t) \geq n\}; \quad \text{for all } n \geq 1, t > 0.$$

If $S_n = \tau$ for some $\tau \leq t$, then $N(\tau) = n$ and $N(t) \geq n$.

6

Although stochastic processes are usually defined by a sequence of rv's or a family of rv's indexed by the reals, we represent arrival processes by arrival epochs, interarrival intervals, or counting variables, which ever is convenient at the moment.

The general class of arrival processes is too complicated to make much further progress. We simplify as follows:

Def: A renewal process is an arrival process for which the interarrival intervals X_1, X_2, \dots are IID.

Def: A Poisson process is a renewal process for which each X_i has an exponential distribution, $F_X(x) = 1 - \exp(-\lambda x)$ for $x \geq 0$, where λ is a fixed parameter called the rate.

7

Poisson processes and exponential rv's.

The remarkable simplicity of Poisson processes is closely related to the 'memoryless' property of the exponential rv.

Def: A rv X is memoryless if X is positive and, for all real $t > 0$ and $x > 0$,

$$\Pr\{X > t + x\} = \Pr\{X > t\} \Pr\{X > x\}. \quad (1)$$

Since the interarrival interval for a Poisson process is exponential, i.e., $\Pr\{X > x\} = \exp(-\lambda x)$ for $x > 0$.

$$\exp(-\lambda(t + x)) = \exp(-\lambda t) \exp(-\lambda x).$$

Thus X is memoryless.

Thm: A rv X is memoryless if and only if it is exponential. (see text)

8

The reason for the word ‘memoryless’ is more apparent when using conditional probabilities,

$$\Pr\{X > t + x \mid X > t\} = \Pr\{X > x\}$$

If people in a checkout line have exponential service, and you have waited 15 minute for the person in front, what is his or her remaining service time?

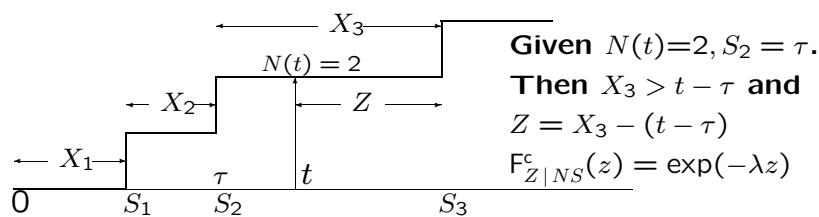
Exponential, same as when service started. The remaining service time ‘doesn’t remember’ the elapsed time. Has your time waiting been wasted?

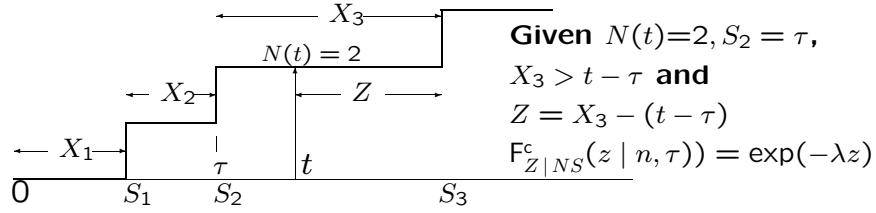
Why do you move to another line if someone takes a long time?

9

Thm: For a Poisson process of rate λ , and any given $t > 0$, the interval Z from t to the next arrival after t has distribution $F_Z^c(z) = \exp(-\lambda z)$ for all $z > 0$. The rv Z is independent of $N(t)$, and, given $N(t) = n$, Z is independent of S_1, \dots, S_n and of $\{N(\tau); 0 < \tau < t\}$. (Thm 2.2.1 in text, but not stated very well there).

Idea of proof: Condition on $N(t) = n$ and $S_n = \tau$, i.e., the number n of arrivals in $(0, t]$ and the time, τ of the most recent arrival in $(0, t]$.





The conditional distribution of Z does not vary with the conditioning values, $N(t) = n$ and $S_n = \tau$, so Z is stat. independent of $N(t)$ and $S_{N(t)}$ (see text).

The rv $S_{N(t)}$ is the time of the last arrival up to and including t . A given sample point ω maps into $N(t)(\omega) = n$, say, and then into $S_n = \tau$. We find the distribution function of $S_{N(t)}$ much later, but don't need it here.

11

This theorem essentially extends the idea of memorylessness to the entire Poisson process. That is, starting at any $t > 0$, the interval Z to the next arrival is an exp rv of rate λ . Z is independent of everything before t .

Subsequent interarrival times are independent of Z and of the past. Thus the interarrival process starting at t with first interarrival Z , and continuing with subsequent interarrivals is a Poisson process.

The counting process corresponding to this interarrival process is $N(t') - N(t)$ for $t' > t$. This is a Poisson process shifted to start at time t , i.e., for each t' , $N(t') - N(t)$ has the same distribution as $N(t' - t)$. Same for joint distributions.

This new process is independent of $\{N(\tau); 0 < \tau \leq t\}$

12

Stationary and independent increments

Def: A counting process $\{N(t); t > 0\}$ has the stationary increment property if $N(t') - N(t)$ has the same distribution as $N(t' - t)$ for all $t' > t > 0$.

Stationary increments means that the distribution of the number of arrivals in the increment $(t, t']$ is a function only of $t' - t$. The distribution depends on the length of the interval, but not the starting time.

Let $\widetilde{N}(t, t') = N(t') - N(t)$, i.e., $\widetilde{N}(t, t')$ is the number of arrivals in the increment $(t, t']$. Thus stationary increments means that $\widetilde{N}(t, t')$ has the same distribution as $N(t' - t)$.

Poisson processes have the stationary increment property.

13

Def: A counting process $\{N(t); t > 0\}$ has the independent increment property if, for every t_1, t_2, \dots, t_n , the rv's $N(t_1), \widetilde{N}(t_1, t_2), \dots, \widetilde{N}(t_{n-1}, t_n)$ are independent.

This implies that the number of arrivals in each of a set of non-overlapping intervals are independent rv's.

For a Poisson process, we have seen that $\widetilde{N}(t_{i-1}, t_i)$ is independent of $\{N(\tau); \tau \leq t_{i-1}\}$, so Poisson processes have independent increments.

Thm: Poisson processes have stationary and independent increments.

14

The Erlang and Poisson distributions

For a Poisson process of rate λ , the PDF of arrival epoch S_2 can be found by convolving the density of X_1 and X_2 . Thus

$$\begin{aligned} f_{S_2}(t) &= \int_0^t [\lambda \exp(-\lambda x)][\lambda \exp(-\lambda(t-x))]dx \\ &= \lambda^2 t \exp(-\lambda t) \end{aligned}$$

Using iteration and convolving $f_{S_{n-1}}(t)$ with $\lambda \exp(-\lambda t)$ for each n ,

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

This is called the Erlang density.

15

Stopping to smell the roses while doing all this computation will be very helpful.

The joint density of X_1, \dots, X_n is

$$\begin{aligned} f_{\vec{X}^n}(x_1, \dots, x_n) &= \lambda^n \exp(-\lambda x_1 - \lambda x_2 - \dots - \lambda x_n) \\ &= \lambda^n \exp(-\lambda s_n) \quad \text{where } s_n = \sum_{i=1}^n x_i \\ f_{\vec{S}^n}(s_1, \dots, s_n) &= \lambda^n \exp(-\lambda s_n) \end{aligned}$$

This says that the joint density is uniform over the region where $s_1 < s_2 < \dots < s_n$. Given that the n th arrival is at s_n , the other $n-1$ arrivals are uniformly distributed in $(0, s_n)$, subject to the ordering.

Integrating (or looking ahead), we get the Erlang marginal density.

16

Thm: For a Poisson process of rate λ , the PMF for $N(t)$ is the Poisson PMF,

$$p_{N(t)}(n) = \frac{(\lambda t)^n \exp(-\lambda t)}{n!}$$

Pf: We will calculate $\Pr\{t < S_{n+1} \leq t + \delta\}$ in two ways and go to the limit $\delta \rightarrow 0$. First we use the density for S_{n+1} to get

$$\Pr\{t < S_{n+1} \leq t + \delta\} = f_{S_{n+1}}(t)(\delta + o(\delta))$$

where $\lim_{\delta \rightarrow 0} \frac{o(\delta)}{\delta} = 0$. Next we use the independent increment property over $(0, t]$ and $(t, t + \delta]$.

$$\Pr\{t < S_{n+1} \leq t + \delta\} = p_{N(t)}(n)(\lambda \delta + o(\delta)) + o(\delta)$$

Equating and going to the limit, $p_{N(t)}(n) = f_{S_{n+1}}(t)/\lambda$.

17

$$p_{N(t)}(n) = \frac{(\lambda t)^n \exp(-\lambda t)}{n!}; \quad \text{Poisson PMF}$$

Note that the Poisson PMF is a function of λt and not of λ or t separately. This is the probability of n arrivals in an interval of size t with rate λ .

If we measure length in a different system of units, λ will change accordingly, so the Poisson PMF has to be a function of λt only.

Note also that $N(t) = N(t_1) + \tilde{N}(t_1, t)$ for any $0 < t_1 < t$. Thus $N(t)$ is the sum of 2 independent rv's, one with the Poisson distribution for t_1 and the other for $t - t_1$.

This extends to any k disjoint intervals, which is one reason the Poisson counting process is so easy to work with.

18

Alternate definitions of Poisson process

Is it true that any arrival process for which $N(t)$ has the Poisson PMF for a given λ and for all t is a Poisson process of rate λ ?

As usual, the marginal PMF's alone are not enough. The joint distributions must also be those of the Poisson process (see text).

Thm: If an arrival process has the stationary and independent increment properties and if $N(t)$ has the Poisson PMF for given λ and all $t > 0$, then the process is Poisson.

VHW Pf: The stationary and independent increment properties show that the joint distribution of arrivals over any given set of disjoint intervals is that of a Poisson process. Clearly this is enough.

19

Is it true that any arrival process with the stationary and independent increment properties is a Poisson process?

These properties capture much of our intuition about Poisson processes, but they allow bulk processes to sneak through, i.e., processes in which simultaneous arrivals are possible. Poisson processes satisfy the following condition for very small increments:

$$\Pr\{\widetilde{N}(t, t+\delta) = n\} = \begin{cases} 1 - \lambda\delta + o(\delta) & \text{for } n = 0 \\ \lambda\delta & \text{for } n = 1 \\ o(\delta) & \text{for } n \geq 2 \end{cases}$$

Thm: If an arrival process has the stationary and independent increment properties and satisfies the above incremental condition, then it is a Poisson process.

20

Relation to Bernoulli process

Bernoulli processes are often viewed as the discrete time version of Poisson processes. There is a confusing feature here that must be cleared up first

The binary rv's Y_1, Y_2, \dots of the Bernoulli process have no direct analogs in the Poisson process.

When we view a Bernoulli process as an arrival process, an arrival occurs at discrete time n if and only if $Y_n = 1$. Thus $S_n = Y_1 + \dots + Y_n$ is the number of arrivals up to and including time n . Thus $\{S_n; n \geq 1\}$ is analogous to the Poisson counting process $\{N(t); t > 0\}$.

21

The interarrival intervals, X_1, X_2, \dots for a Bernoulli arrival process are the intervals between successive 1's in the binary stream. Thus $X_1 = k$ if $Y_i = 0$ for $1 \leq i \leq k - 1$ and $Y_k = 1$. Thus $p_{X_1}(k) = p(1 - p)^{k-1}$ for all $k \geq 1$. Subsequent interarrival intervals are IID with X_1 .

Thus, the interarrival intervals for the Bernoulli counting process are geometrically distributed.

The Bernoulli counting process is defined only at integer times, but if we consider the arrivals within integer intervals, we see that the stationary and independent increment properties are satisfied (over those integer values).

22

We can clearly extend the definition of a Bernoulli counting process to a shrunken Bernoulli counting process where changes occur at intervals of δ rather than unit intervals.

Consider a sequence of shrinking Bernoulli processes, holding λ/δ constant but shrinking λ and δ . The geometric interarrival interval becomes exponential and the Bernoulli counting process converges to the Poisson counting process. (see text).

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Lecture 5: Poisson combining and splitting etc.

Outline:

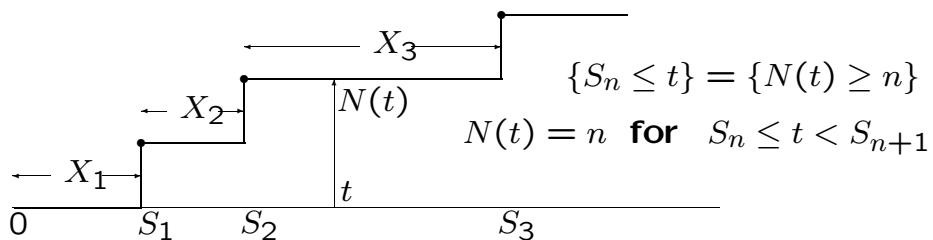
- Review of Poisson processes
- Combining independent Poisson processes
- Splitting a Poisson process
- Non-homogeneous Poisson processes
- Conditional arrival densities

1

Review of Poisson processes

A Poisson process is an arrival process with IID exponentially-distributed interarrival times.

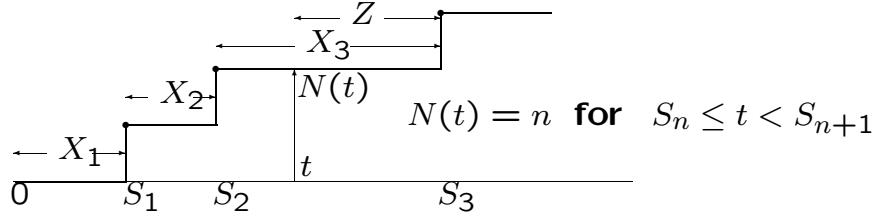
It can be represented by its arrival epochs, S_1, S_2, \dots , or by its interarrival times, X_1, X_2, \dots or by its counting process, $\{N(t); t > 0\}$, where



2

The interarrival times X_i of a Poisson process are memoryless, i.e., for $x, t > 0$,

$$\Pr\{X_i > t + x \mid X_i > t\} = \Pr\{X_i > x\} = \exp(-\lambda x)$$



Given $N(t) = n$ and $S_n = \tau$, we have $X_{n+1} > t - \tau$.

The interval $Z = X_{n+1} - (t - \tau)$.

$$\Pr\{Z > z \mid N(t), S_{N(t)}\} = \exp(-\lambda z)$$

Z is independent of $\{N(\tau); \tau \leq t\}$; Z is the first interarrival of the Poisson process $\{N(t') - N(t); t' > t\}$; $\{N(t') - N(t); t' > t\}$ is independent of $\{N(\tau); \tau \leq t\}$.

3

For any set of times, $0 < t_1 < t_2 < \dots < t_k$, the Poisson process increments, $\{N(t); 0 < t \leq t_1\}, \{\tilde{N}(t_1, t); t_1 < t \leq t_2\}, \dots, \{\tilde{N}(t_{k-1}, t); t_{k-1} < t \leq t_k\}$ are stationary and independent Poisson counting processes (over their given intervals). Also,

$$p_{N(t)}(n) = \frac{(\lambda t)^n \exp(-\lambda t)}{n!}$$

This is a function only of the mean λt . By the stationary and ind. inc. property, we know that $N(t_1)$ and $\tilde{N}(t_1, t)$ are independent. They are also Poisson and their sum, $N(t)$, is Poisson. In general, sums of independent Poisson rv's are Poisson, with the means adding.

4

Alternate definitions of Poisson process: (i.e., alternate conditions which suffice to show that an arrival process is Poisson).

Thm: If an arrival process has the stationary and independent increment properties and if $N(t)$ has the Poisson PMF for given λ and all $t > 0$, then the process is Poisson.

Thm: If an arrival process has the stationary and independent increment properties and satisfies the following incremental condition, then the process is Poisson.

$$\Pr\{\widetilde{N}(t, t+\delta) = n\} = \begin{cases} 1 - \lambda\delta + o(\delta) & \text{for } n = 0 \\ \lambda\delta + o(\delta) & \text{for } n = 1 \\ o(\delta) & \text{for } n \geq 2 \end{cases}$$

Combining independent Poisson processes

Two Poisson processes $\{N_1(t); t > 0\}$ and $\{N_2(t); t > 0\}$ are independent if for all t_1, \dots, t_n , the rv's $N_1(t_1), \dots, N_1(t_n)$ are independent of $N_2(t_1), \dots, N_2(t_n)$.

Thm: if $\{N_1(t); t > 0\}$ and $\{N_2(t); t > 0\}$ are independent Poisson processes of rates λ_1 and λ_2 and $N(t) = N_1(t) + N_2(t)$ for all $t > 0$, then $\{N(t); t > 0\}$ is a Poisson process of rate $\lambda = \lambda_1 + \lambda_2$.

The idea is that in any increment $(t, t+\delta]$,

$$\begin{aligned}
 \widetilde{N}(t, t+\delta) &= \widetilde{N}_1(t, t+\delta) + \widetilde{N}_2(t, t+\delta) \\
 p_{\widetilde{N}(t, t+\delta)}(1) &= p_{\widetilde{N}_1(t, t+\delta)}(1)p_{\widetilde{N}_2(t, t+\delta)}(0) \\
 &\quad + p_{\widetilde{N}_1(t, t+\delta)}(0)p_{\widetilde{N}_2(t, t+\delta)}(1) \\
 &= [\delta\lambda_1 + o(\delta)][1 - \delta\lambda_2 + o(\delta)] \\
 &\quad + [1 - \delta\lambda_1 + o(\delta)][\delta\lambda_2 + o(\delta)] \\
 &= \delta(\lambda_1 + \lambda_2) + o(\delta)
 \end{aligned}$$

7

Similarly, $\widetilde{N}(t, t+\delta) = 0$ **if both** $\widetilde{N}_1(t, t+\delta) = 0$ **and** $\widetilde{N}_2(t, t+\delta) = 0$. **Thus**

$$\Pr\{\widetilde{N}(t, t+\delta) = 0\} = [1 - (\lambda_1 + \lambda_2)\delta + o(\delta)]$$

It is much cleaner analytically to use the Poisson distribution directly. Since $\widetilde{N}_1(t, t+\delta)$ and $\widetilde{N}_2(t, t+\delta)$ are independent and Poisson,

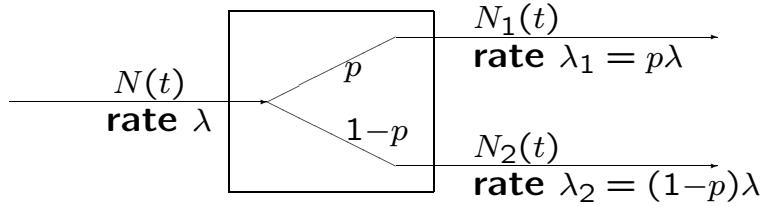
$$\widetilde{N}(t, t+\delta) = \widetilde{N}_1(t, t+\delta) + \widetilde{N}_2(t, t+\delta)$$

is a Poisson rv with mean $\lambda\delta$.

The sum of many small independent arrival processes tends to be close to Poisson even if the small processes are not. In a sense, the independence between the processes overcomes the dependence between successive arrivals in each process.

8

Splitting a Poisson process



Each arrival is switched to $\{N_1(t); t > 0\}$ with probability p and otherwise goes to $\{N_2(t); t > 0\}$. View the switch as a Bernoulli process independent of $\{N(t); t > 0\}$. A p biased coin is flipped independently at each arrival.

Each new process clearly has the stationary and independent increment property and each satisfies the small increment property. Thus each is Poisson.

9

The small increment property doesn't make it clear that the split processes are independent. For independence, both processes must sometimes have arrivals in the same small increment. Independence is hidden in the $o(\delta)$ terms. See text.

Combining and splitting are often done together. First one views separate independent Poisson processes as a combined process. Then it is split again with binary choices between processes.

Example: Consider a last-come first-serve queue with Poisson arrivals, rate λ and independent exponential services, rate μ . A new arrival starts service immediately, but is interrupted if a new arrival occurs before service completion.

10

View services as a Poisson process. We can either ignore this process when there is nothing to serve, or visualize a low priority task that is served at rate μ when there is nothing else to do. The arrival process plus the service process is Poisson, rate $\lambda + \mu$.

The probability an arrival completes service before being interrupted is $\mu/(\lambda + \mu)$.

Given that you are interrupted, what is the probability of no further interruption? Two services (the interrupting job and you) must finish before the next interruption, so Answer: $\frac{\mu^2}{(\lambda+\mu)^2}$.

11

Non-homogeneous Poisson processes

Consider optical transmission, where an optical stream of photons is modulated by variable power. The photon stream is reasonably modelled as a Poisson process, and the modulation converts the steady photon rate into a variable rate, say $\lambda(t)$.

We model the number of photons in any interval $(t, t+\delta]$ as a Poisson random variable whose rate parameter over $(t, t+\delta]$ is the average photon rate over $(t, t+\delta]$ times δ .

In the small increment model, we have

$$\Pr\{\widetilde{N}(t, t+\delta) = n\} = \begin{cases} 1 - \delta\lambda(t) + o(\delta) & \text{for } n = 0 \\ \lambda(t)\delta + o(\delta) & \text{for } n = 1 \\ o(\delta) & \text{for } n \geq 2 \end{cases}$$

12

We can use this small increment model to see that the number of arrivals in each increment is Poisson with

$$\Pr\{\tilde{N}(t, \tau) = n\} = \frac{[\tilde{m}(t, \tau)]^n \exp[-\tilde{m}(t, \tau)]}{n!}$$

where

$$\tilde{m}(t, \tau) = \int_t^\tau \lambda(u) du.$$

Combining and splitting non-homogeneous processes still works as in the homogeneous case, but the independent exponential interarrivals doesn't work.

We now return to homogeneous Poisson processes.

13

Conditional arrival densities

There are many interesting and useful results about the increment $(0, t]$ of a Poisson process conditional on $N(t)$. First condition on $N(t) = 1$.

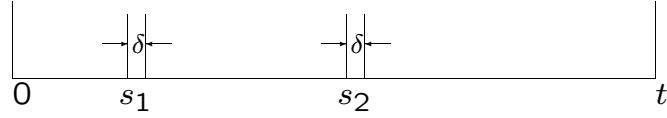


$$\begin{aligned} f_{S_1|N(t)}(s_1|1) &= \lim_{\delta \rightarrow 0} \frac{p_{N(s_1)}(0) p_{\tilde{N}(s_1, s_1 + \delta)}(1) p_{\tilde{N}(s_1 + \delta, t)}(0)}{\delta p_{N(t)}(1)} \\ &= \frac{e^{-\lambda s_1} \lambda \delta e^{-\lambda \delta} e^{-\lambda(t-s_1-\delta)}}{\delta \lambda t e^{-\lambda t}} = \frac{1}{t} \end{aligned}$$

The important point is that this does not depend on s_1 , i.e., it is uniform over $(0, t]$.

14

Next consider $N(t) = 2$.



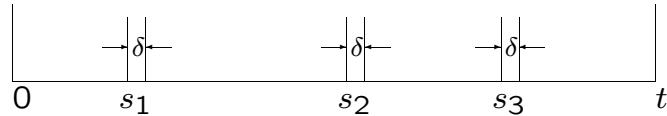
$$f(s_1 s_2 | 2) = \lim_{\delta} \frac{e^{-\lambda s_1} \lambda \delta e^{-\lambda \delta} e^{-\lambda(s_2 - s_1 - \delta)} \lambda \delta e^{-\lambda \delta} e^{-\lambda(t - s_2 - \delta)}}{\delta^2 p_{N(t)}(2)}$$

$$= \frac{2}{t^2}$$

Again, this does not depend on s_1 , s_2 , or λ for $0 < s_1 < s_2 < t$, i.e., it is uniform over the given region of s_1, s_2 .

We can do the same thing for $N(t) = n$ for arbitrary n . Note that the exponents above always sum to λt .

15



$$f_{\vec{s}^{(n)} | N(t)}(\vec{s}^{(n)} | n) = \lim_{\delta \rightarrow 0} \frac{(\delta \lambda)^n \exp(-\lambda t)}{\delta^n p_{N(t)}(n)}$$

$$= \frac{n!}{t^n}$$

This is ‘uniform’ over $0 < s_1 < \dots < s_n < t$.

This is a uniform n dimensional probability density over the volume $t^n/n!$ corresponding to the constraint region $0 < s_1 < \dots < s_n < t$.

How did this derivation ‘know’ that the volume of s_1, \dots, s_n over $0 < s_1 < \dots < s_n < t$ is $n!/t^n$?

16

To see why $n!$ appears in this uniform density, let U_1, \dots, U_n be n IID rv's, each uniform over $(0, t]$. Let S_1, \dots, S_n be the order statistics for \vec{U}^n , i.e.,

$$S_1 = \min(U_1, \dots, U_n), \dots, S_k = k^{\text{th}} \text{ smallest}, \dots$$

The region of volume t^n where the density of \vec{U}^n is nonzero partitions into $n!$ regions, one in which $u_1 < u_2 < \dots < u_n$ and one for each other ordering of u_1, \dots, u_n . From symmetry, each volume is the same, and thus each is $t^n/n!$.

The region where \vec{S}^n is nonzero is one of these partitions, and thus also has volume $t^n/n!$.

Since \vec{S}^n has the same density, whether it is the conditional density of n arrival epochs given $N(t) = n$ or the order statistics of n uniform rv's, we can use results about either for the other.

17

As an example of using order statistics, consider finding the distribution function of S_1 conditional on $N(t) = n$. Viewing S_1 as the minimum of U_1, \dots, U_n , we have

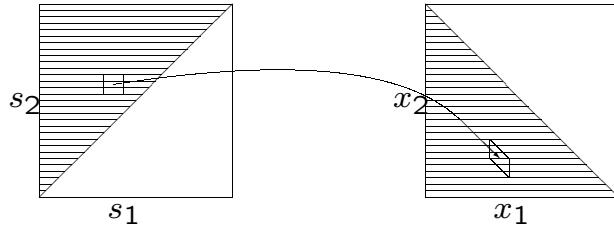
$$\begin{aligned} \Pr\left\{\min_{1 \leq i \leq n} U_i > s_1\right\} &= \prod_{i=1}^n \Pr\{U_i > s\} \\ &= \prod_{i=1}^n \left(1 - \frac{s_1}{t}\right) = \left(1 - \frac{s_1}{t}\right)^n \\ F_{S_1|N(t)}^c(s_1|n) &= \left(1 - \frac{s_1}{t}\right)^n \quad \text{for } s_1 \leq t \end{aligned}$$

$E[S_1 | N(t) = n]$ can be found by integration,

$$E[S_1 | N(t) = n] = \frac{t}{n+1}$$

18

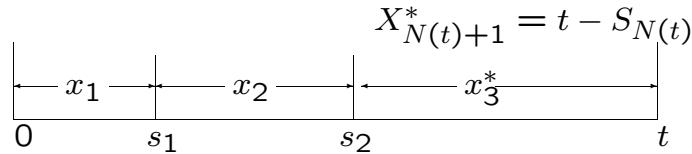
Return to $N(t) = 2$ and look at $f_{X_1 X_2 | N(t)}(x_1, x_2 | 2)$.



Note that the area in the s_1, s_2 space where $0 < s_1 < s_2 < t$ is $t^2/2$, explaining why the density is $2/t^2$.

Note that a δ^2 box in the s_1, s_2 space maps into a parallelepiped of the same area in the x_1, x_2 space. The area in the x_1, x_2 space where $0 < x_1 + x_2 < t$ is again $t^2/2$ and $f(x_1, x_2 | N(t)=2) = 2/t^2$.

19



$$f(x_1, x_2 | N(t)) = 2/t^2 \quad \text{for } x_1, x_2, x_3^* > 0; x_1 + x_2 + x_3^* = t$$

From symmetry, any 2 of the variables X_1, X_2, X_3^* can replace X_1, X_2 above.

Also, since $X_1 = S_1$, X_1 has density and expected value

$$\begin{aligned} F_{X_1 | N(t)}^c(x_1 | 2) &= \left(1 - \frac{x_1}{t}\right)^2 \\ E[X_1 | N(t)=2] &= \frac{t}{3}. \end{aligned}$$

we see that X_2 and X_3^* can be substituted for X_1 in the above formulas. As a sanity check, note that $E[X_1 + X_2 + X_3^*] = t$

20

This extends to $N(t) = n$ for arbitrary n .

$$\begin{aligned} f_{X_1|N(t)}(x_1|n) &= \left(1 - \frac{x_1}{t}\right)^n \\ E[X_1 | N(t)=n] &= \frac{t}{n+1}. \end{aligned}$$

This relation also applies to $X_{n+1}^* = t - S_{N(t)}$

$$\begin{aligned} f_{X_{n+1}^*|N(t)}(x|n) &= \left(1 - \frac{x}{t}\right)^n \\ E[X_{n+1}^* | N(t)=n] &= \frac{t}{n+1}. \end{aligned}$$

$$\begin{aligned} E[X_{N(t)+1}^*] &= \sum_{n=0}^{\infty} \frac{t}{n+1} \frac{(\lambda t)^n e^{-\lambda t}}{n!} \\ &= \sum_{n=0}^{\infty} \frac{(\lambda t)^{n+1} e^{-\lambda t}}{\lambda(n+1)!} = \frac{1 - e^{-\lambda t}}{\lambda} \end{aligned}$$

21

Paradox: The mean interarrival time for a Poisson process is $1/\lambda$. But the mean time from any given t to the next arrival is $1/\lambda$ and the mean time back to the previous arrival is $(1/\lambda)(1 - e^{-\lambda t})$. Thus the mean length of the interval containing t is $(1/\lambda)(2 - e^{-\lambda t})$.

This paradox will become clearer when we study renewals. A temporary half-intuitive explanation is to first choose a sample path for a Poisson process and then choose a uniform random value for t over some large interval far from 0. The larger interarrival intervals occupy proportionally more of the overall interval than the smaller, so t is biased to lie in one of those larger intervals.

22

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

6.262: Discrete Stochastic Processes 2/21/11

Lecture 6: From Poisson to Markov

Outline:

- Joint conditional densities for Poisson
- Definition of finite-state Markov chains
- Classification of states
- Periodic states and classes
- Ergodic Markov chains

Recall (Section 2.2.2) that the joint density of interarrivals X_1, \dots, X_n and arrival epoch S_{n+1} is

$$f_{X_1 \dots X_n S_{n+1}}(x_1, \dots, x_n, s_{n+1}) = \lambda^{n+1} \exp(-\lambda s_{n+1})$$

Conditional on $S_{n+1} = t$ (which is Erlang),

$$f_{X_1 \dots X_n | S_{n+1}}(x_1, \dots, x_n | t) = \frac{\lambda^{n+1} e^{-\lambda t}}{\left[\frac{\lambda^{n+1} t^n e^{-\lambda t}}{n!} \right]} = \frac{n!}{t^n} \quad (1)$$

Similarly (from Eqn. 2.43, text)

$$f_{X_1 \dots X_n | N(t)}(x_1, \dots, x_n | n) = \frac{\lambda^{n+1} e^{-\lambda t}}{\left[\frac{\lambda^{n+1} t^n e^{-\lambda t}}{n!} \right]} = \frac{n!}{t^n} \quad (2)$$

Both equations are for $0 < x_1, \dots, x_n$ and $\sum x_k < t$.

Both say the conditional density is uniform over the constraint region.

Why are the two equations the same? If we condition X_1, \dots, X_n on both $N(t) = n$ and $S_{n+1} = t_1$ for any $t_1 > t$, (2) is unchanged.

By going to the limit $t_1 \rightarrow t$, we get the first equation. The result, $n!/t^n$, is thus the density conditional on n arrivals in the open interval $(0, t)$ and is unaffected by future arrivals.

This density, and its constraint region, is symmetric in the arguments x_1, \dots, x_n . More formally, the constraint region (and trivially the density in the constraint region) is unchanged by any permutation of x_1, \dots, x_n .

Thus the marginal distribution, $F_{X_k|N(t)}(x_k|n)$ is the same for $1 \leq k \leq n$. From analyzing $S_1 = X_1$, we then know that $F_{X_k|N(t)}^c(x_k|n) = (t - x_n)^n/t^n$ for $1 \leq k \leq n$.

For the constraint $S_{n+1} = t$, we have analyzed X_1, \dots, X_n , but have not considered X_{n+1} , the final interarrival interval before t .

The reason is that $\sum_{k=1}^{n+1} X_k = S_{n+1} = t$, so that these variables do not have an $n+1$ dimensional density.

The same uniform density as before applies to each subset of n of the $n+1$ variables, and the constraint is symmetric over all $n+1$ variables.

This also applies to the constraint $N(t) = n$, using $X_{n+1}^* = t - S_n$

Definition of finite-state Markov chains

Markov chains are examples of integer-time stochastic processes, $\{X_n; n \geq 0\}$ where each X_n is a rv.

A finite-state Markov chain is a Markov chain in which the sample space for each rv X_n is a fixed finite set, usually taken to be $\{1, 2, \dots, M\}$.

Any discrete integer-time process is characterized by $\Pr\{X_n = j | X_{n-1} = i, X_{n-2} = k, \dots, X_0 = m\}$ for $n \geq 0$ and all i, j, k, \dots, m , each in the sample space.

For a finite-state Markov chain, these probabilities are restricted to be

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

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

The definition first says that X_n depends on the past only through X_{n-1} , and second says that the probabilities don't depend on n for $n \geq 1$.

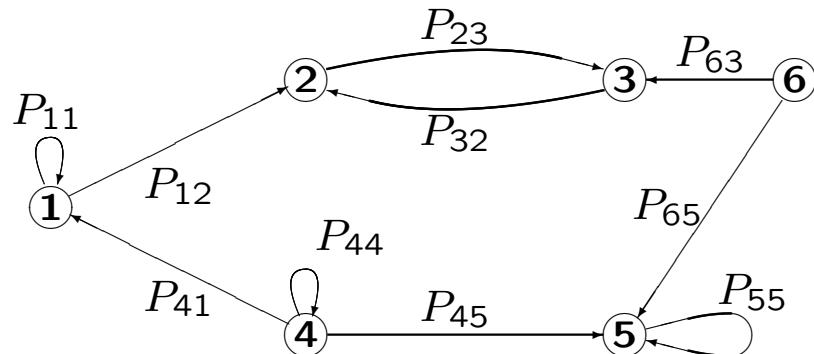
Some people call this a homogeneous Markov chain and allow P_{ij} to vary with n in general.

The rv's $\{X_n; n \geq 0\}$ are dependent, but in only a very simple way. X_n is called the state at time n and characterizes everything from the past that is relevant for the future.

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

We often take the initial state to be a fixed value, and often view the Markov chain as just the set $\{P_{ij}; 1 \leq i, j \leq M\}$, with the initial state viewed as a parameter.

Sometimes we visualize $\{P_{ij}\}$ in terms of a directed graph and sometimes as a matrix.



a) Graphical

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

b) Matrix

The graph emphasizes the possible and impossible (an edge from i to j explicitly means that $P_{ij} > 0$).

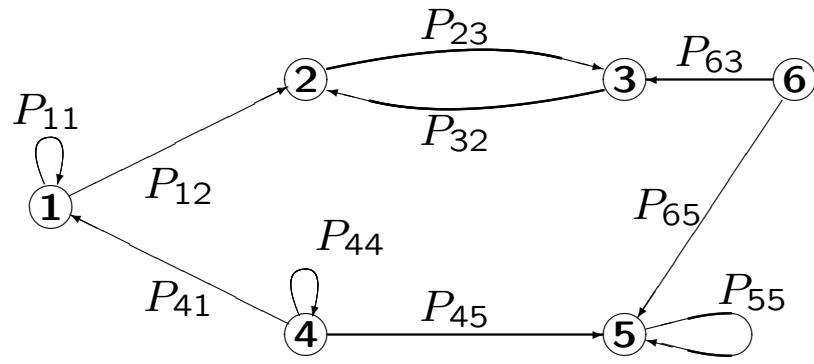
The matrix is useful for algebraic and asymptotic issues.

Classification of states

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

Def: A path is a walk with no repeated nodes.

Def: 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)

It doesn't make any difference whether you regard (2, 3, 2) and (3, 2, 3) as the same or different cycles.

Def: A state (node) j is accessible from i ($i \rightarrow j$) if a walk exists from i to j .

Let $P_{ij}^n = \Pr\{X_n = j \mid X_0 = i\}$. Then if i, k, j is a walk, $P_{ik} > 0$ and $P_{kj} > 0$, so $P_{ij}^2 \geq P_{ik}P_{kj} > 0$.

Similarly, if there is an n -step walk starting at i and ending at j , then $P_{ij}^n > 0$.

Thus if $i \rightarrow j$, there is some n for which $P_{ij}^n > 0$. To the contrary, if j is not accessible from i ($i \not\rightarrow j$), then $P_{ij}^n = 0$ for all $n \geq 1$.

$i \rightarrow j$ means that, starting in i , entry to j is possible, perhaps with multiple steps. $i \not\rightarrow j$ means there is no possibility of ever reaching j from i .

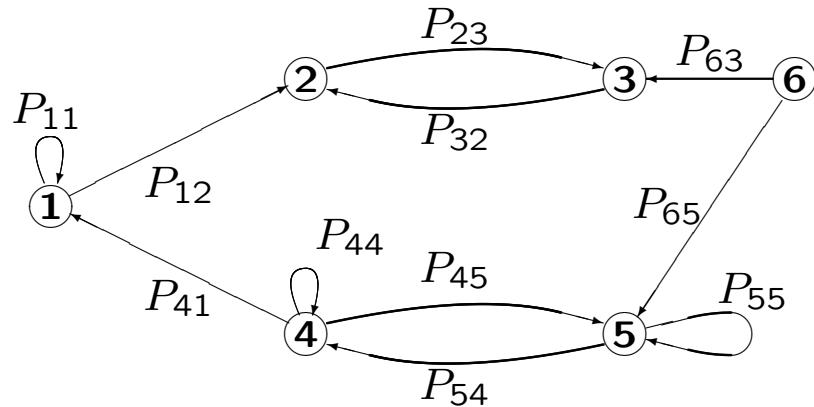
If $i \rightarrow j$ and $j \rightarrow k$, then $i \rightarrow k$. (Concatenate a walk from i to j with a walk from j to k .)

Def: States i and j communicate ($i \leftrightarrow j$) if $i \rightarrow j$ and $j \rightarrow i$.

Note that if $(i \leftrightarrow j)$ and $(j \leftrightarrow k)$, then $(i \leftrightarrow k)$.

Note that if $(i \leftrightarrow j)$, then there is a cycle that contains both i and j .

Def: A class \mathcal{C} of states is a non-empty set of states such that each $i \in \mathcal{C}$ communicates with every other $j \in \mathcal{C}$ and communicates with no $j \notin \mathcal{C}$.



$$\mathcal{C}_1 = \{2, 3\}$$

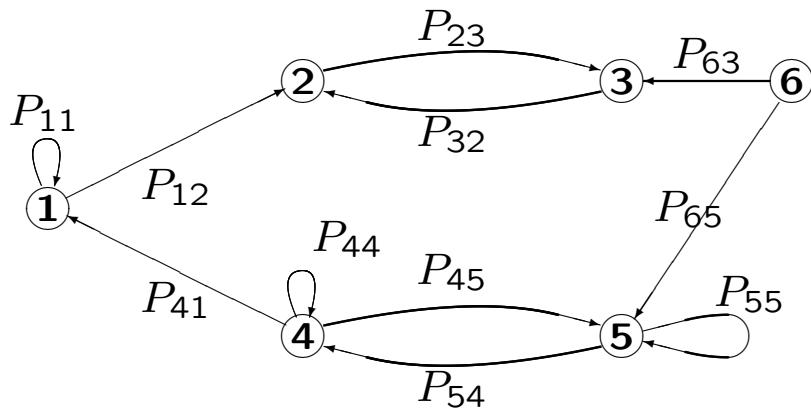
$$\mathcal{C}_2 = \{4, 5\}$$

$$\mathcal{C}_3 = \{1\}$$

$$\mathcal{C}_4 = \{6\}$$

Why is $\{6\}$ a class

Def: A state i is recurrent if $j \rightarrow i$ for all j such that $i \rightarrow j$. (i.e., if no state from which there is no return can be entered.) If a state is not recurrent, it is transient.



2 and 3 are recurrent
 4 and 5 are transient
 $4 \rightarrow 1, 5 \rightarrow 1, 1 \not\rightarrow 4, 5$
 6 and 1 also transient

Thm: The states in a class are all recurrent or all transient.

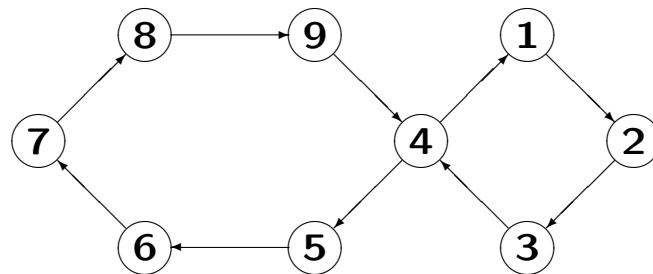
Pf: Assume i recurrent and let $S_i = \{j : i \rightarrow j\}$. By recurrence, $j \rightarrow i$ for all $j \in S_i$. Thus $i \leftrightarrow j$ if and only if $j \in S_i$, so S_i is a class. Finally, if $j \in S_i$, then $j \rightarrow k$ implies $i \rightarrow k$ and $k \rightarrow i \rightarrow j$, so j is recurrent.

Periodic states and classes

Def: The period, $d(i)$, of state i is defined as

$$d(i) = \gcd\{n : P_{ii}^n > 0\}$$

If $d(i) = 1$, i is aperiodic. If $d(i) > 1$, i is periodic with period $d(i)$.



For example, $P_{44}^n > 0$ for
 $n = 4, 6, 8, 10$; $d(4) = 2$

For state 7, $P_{77}^n > 0$ for
 $n = 6, 10, 12, 14$; $d(7) = 2$

Thm: All states in the same class have the same period.

See text for proof. It is not very instructive.

A periodic class of states with period $d > 1$ can be partitioned into subclasses $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_d$ so that for $1 \leq \ell < d$, and all $i \in \mathcal{S}_\ell$, $P_{ij} > 0$ only for $j \in \mathcal{S}_{\ell+1}$. For $i \in \mathcal{S}_d$, $P_{ij} > 0$ only for $j \in \mathcal{S}_1$. (see text)

In other words, starting in a given subclass, the state cycles through the d subclasses.

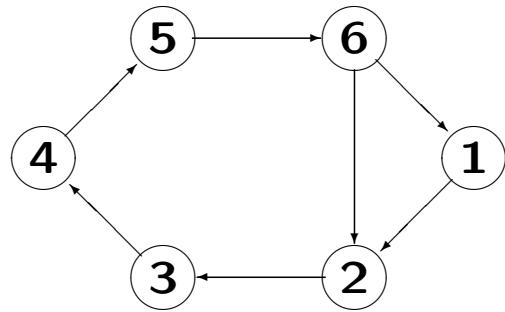
Ergodic Markov chains

The most fundamental and interesting classes of states are those that are recurrent and aperiodic. These are called ergodic. A Markov chain with a single class that is ergodic is an ergodic Markov chain.

Ergodic Markov chains gradually lose their memory of where they started, i.e., P_{ij}^n goes to a limit $\pi_j > 0$ as $n \rightarrow \infty$, and this limit does not depend on the starting state i .

This result is also basic to arbitrary finite-state Markov chains, so we look at it carefully and prove it next lecture.

A first step in showing that $P_{ij}^n \rightarrow \pi_j$ is the much weaker statement that $P_{ij}^n > 0$ for all large enough n . This is more a combinatorial issue than probabilistic, as indicated below.



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.

This hints at the following theorem:

Thm: For an ergodic M state Markov chain, $P_{ij}^n > 0$ for all i, j , and all $n \geq (M - 1)^2 + 1$.

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes
Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Lecture 7: Finite-state Markov Chains; the matrix approach

Outline:

- **The transition matrix and its powers**
- **Convergence of $[P^n] > 0$**
- **Ergodic Markov chains**
- **Ergodic unichains**
- **Other finite-state Markov chains**

Recall that the state X_n of a Markov chain at step n depends on the past only through the previous step, i.e.,

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

This implies that the joint probability of X_0, X_1, \dots, X_n can be expressed as a function of $p_{X_0}(x_0)$ and of the transition probabilities, $\{P_{ij}; 1 \leq i, j \leq M\}$.

The transition probabilities are conveniently represented in terms of a transition matrix,

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

If we condition only on the state at time 0, and define $P_{ij}^n = \Pr\{X_n = j \mid X_0 = i\}$, then, starting with $n = 2$, we have

$$\begin{aligned} P_{ij}^2 &= \sum_k \Pr\{X_2 = j \mid X_1 = k, X_0 = i\} \Pr\{X_1 = k \mid X_0 = i\} \\ &= \sum_k P_{ik} P_{kj} \end{aligned}$$

Note that $\sum_k P_{ik} P_{kj}$ is the i, j term of the product of the transition matrix $[P]$ with itself, which is $[P^2]$.

Thus the 2-step transition probabilities $\{P_{ij}^2; 1 \leq i, j \leq M\}$ are simply the elements of $[P^2]$.

Iterating to find P_{ij}^n for successively larger n ,

$$\begin{aligned} P_{ij}^n &= \sum_k \Pr\{X_n=j|X_{n-1}=k\} \Pr\{X_{n-1}=k|X_0=i\} \\ &= \sum_k P_{ik}^{n-1} P_{kj} \end{aligned}$$

Thus P_{ij}^n is the i, j element of $[P^n]$, i.e., the matrix $[P]$ to the n^{th} power.

Computational hack: To find high powers of $[P]$, calculate $[P^2]$, $[P^4]$, $[P^8]$, etc. and then multiply the required powers of 2.

Chapman-Kolmogorov eqns: Since $[P^{m+n}] = [P^m][P^n]$,

$$P_{ij}^{m+n} = \sum_k P_{ik}^m P_{kj}^n$$

Convergence of $[P^n] > 0$

An important question for Markov chains, and one that effects almost everything else, is whether the effect of the initial state dies out with time, i.e., whether $\lim_{n \rightarrow \infty} P_{ij}^n = \pi_j$ for all i and j , where π_j is a function only of j and not of i or n .

If this limit exists, we can multiply both sides by P_{jk} and sum over j , getting

$$\lim_{n \rightarrow \infty} \sum_j P_{ij}^n P_{jk} = \sum_j \pi_j P_{jk}$$

The left side is $\lim_{n \rightarrow \infty} P_{ik}^{n+1} = \pi_k$. Thus if this limit exists, the vector $\vec{\pi}$ must satisfy $\pi_k = \sum_j \pi_j P_{jk}$ for each k .

In matrix terms, does $\lim_{n \rightarrow \infty} [P^n]$ exist, and is each row the same vector, $\vec{\pi}$? If so, then $\vec{\pi}$ must satisfy the matrix equation $\vec{\pi} = \vec{\pi}[P]$.

Def: A probability vector is a vector $\vec{\pi} = (\pi_1, \dots, \pi_M)$ for which each π_i is nonnegative and $\sum_i \pi_i = 1$. A probability vector $\vec{\pi}$ is called a steady-state vector for the transition matrix $[P]$ if $\vec{\pi} = \vec{\pi}[P]$.

One would now think that we have reduced the question of whether $\lim_{n \rightarrow \infty} [P^n]$ exists to the study of the steady-state equation $\vec{\pi} = \vec{\pi}[P]$.

Surprisingly, studying $\lim_{n \rightarrow \infty} [P^n]$ is relatively simple, whereas understanding the set of solutions to $\vec{\pi} = \vec{\pi}[P]$ is more complicated. We will find that $\vec{\pi} = \vec{\pi}[P]$ always has one (and often more) probability vector solutions, but this doesn't imply that $\lim_{n \rightarrow \infty} [P^n]$ exists.

Ergodic Markov chains

Another way to express that $\lim_{n \rightarrow \infty} [P^n]$ converges to a matrix of equal rows $\vec{\pi}$ is the statement that, for each column j , $\lim_{n \rightarrow \infty} P_{ij}^n = \pi_j$ for each i .

The following theorem demonstrates this type of convergence, and some stronger results, for ergodic Markov chains.

Thm: Let an ergodic finite-state Markov chain have transition matrix $[P]$. Then for each j , $\max_i P_{ij}^n$ is nonincreasing in n , $\min_i P_{ij}^n$ is nondecreasing in n , and

$$\lim_{n \rightarrow \infty} \max_i P_{ij}^n = \lim_{n \rightarrow \infty} \min_i P_{ij}^n \doteq \pi_j > 0$$

with exponential convergence in n

The key to this theorem is the pair of statements that $\max_i P_{ij}^n$ is nonincreasing in n and $\min_i P_{ij}^n$ is non-decreasing in n .

It turns out, with an almost trivial proof, that these statements are true for all Markov chains, so we first establish this as a lemma.

Lemma 1: Let $[P]$ be the transition matrix of an arbitrary finite-state Markov chain. Then for each j , $\max_i P_{ij}^n$ is nonincreasing in n and $\min_i P_{ij}^n$ is non-decreasing in n .

Example 1: Consider the 2-state chain with $P_{12} = P_{21} = 1$. Then P_{12}^n alternates between 1 and 0 for increasing n and P_{22}^n alternates between 0 and 1. The maximum of P_{12}^n and P_{22}^n is 1, which is nonincreasing, and the minimum is 0.

Lemma 1: Let $[P]$ be the transition matrix of an arbitrary finite-state Markov chain. Then for each j , $\max_i P_{ij}^n$ is nonincreasing in n and $\min_i P_{ij}^n$ is non-decreasing in n .

Example 2: Consider the 2-state ergodic chain with $P_{12} = P_{21} = 3/4$. Then $P_{12}^n = \frac{3}{4}, \frac{3}{8}, \frac{9}{16}, \dots$ for increasing n and $P_{22}^n = \frac{1}{4}, \frac{5}{8}, \frac{7}{16}, \dots$

Each sequence oscillates while approaching $1/2$, but $\max(P_{12}^n, P_{22}^n) = \frac{3}{4}, \frac{5}{8}, \frac{9}{16}, \dots$ which is decreasing toward $1/2$. Similarly the minimum approaches $1/2$ from below, $\min(P_{12}^n, P_{22}^n) = \frac{1}{4}, \frac{3}{8}, \frac{7}{16}, \dots$

Lemma 1: Let $[P]$ be the transition matrix of an arbitrary finite-state Markov chain. Then for each j , $\max_i P_{ij}^n$ is nonincreasing in n and $\min_i P_{ij}^n$ is non-decreasing in n .

Proof: For any states i, j and any step n ,

$$\begin{aligned} P_{ij}^{n+1} &= \sum_k P_{ik} P_{kj}^n \\ &\leq \sum_k P_{ik} \max_\ell P_{\ell j}^n \\ &= \max_\ell P_{\ell j}^n \end{aligned}$$

Since this holds for all states i , it holds for the maximizing i , so $\max_i P_{ij}^{n+1} \leq \max_\ell P_{\ell j}^n$. Replacing maxima with minima and reversing inequalities,

$$\min_i P_{ij}^{n+1} \geq \min_\ell P_{\ell j}^n.$$

Before completing the proof of the theorem, we specialize the theorem to the case where $[P] > 0$, i.e., 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 $n \geq 1$:

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

$$\left(\max_\ell P_{\ell j}^n - \min_\ell P_{\ell j}^n \right) \leq (1 - 2\alpha)^n.$$

$$\lim_{n \rightarrow \infty} \max_\ell P_{\ell j}^n = \lim_{n \rightarrow \infty} \min_\ell P_{\ell j}^n > 0.$$

Note that Lemma 1 implies that $\lim_{n \rightarrow \infty} \max_\ell P_{\ell j}^n$ must exist since this is the limit of a decreasing non-negative sequence. This lemma then shows the maxima and minima both have the same limit.

Proof of lemma 2: We tighten the proof of lemma 1 slightly to make use of the positive elements. For a given n and j , let ℓ_{\min} be a state that minimizes P_{ij}^n over i . Then

$$\begin{aligned}
 P_{ij}^{n+1} &= \sum_k P_{ik} P_{kj}^n \\
 &\leq \sum_{k \neq \ell_{\min}} P_{ik} \max_{\ell} P_{\ell j}^n + P_{i\ell_{\min}} \min_{\ell} P_{\ell j}^n \\
 &= (1 - P_{i\ell_{\min}}) \max_{\ell} P_{\ell j}^n + P_{i\ell_{\min}} \min_{\ell} P_{\ell j}^n \\
 &= \max_{\ell} P_{\ell j}^n - P_{i\ell_{\min}} \left(\max_{\ell} P_{\ell j}^n - \min_{\ell} P_{\ell j}^n \right) \\
 &\leq \max_{\ell} P_{\ell j}^n - \alpha \left(\max_{\ell} P_{\ell j}^n - \min_{\ell} P_{\ell j}^n \right) \\
 \max_i P_{ij}^{n+1} &\leq \max_{\ell} P_{\ell j}^n - \alpha \left(\max_{\ell} P_{\ell j}^n - \min_{\ell} P_{\ell j}^n \right)
 \end{aligned}$$

We have shown that

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

Interchanging max with min and \leq with \geq , we get

$$\min_i P_{ij}^{n+1} \geq \min_\ell P_{\ell j}^n + \alpha \left(\max_\ell P_{\ell j}^n - \min_\ell P_{\ell j}^n \right).$$

Subtracting these equations,

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

Since $\min_\ell P_{\ell j} \geq \alpha$ and $\max_\ell P_{\ell j} \leq 1 - \alpha$,

$$\max_\ell P_{\ell j} - \min_\ell P_{\ell j} \leq 1 - 2\alpha$$

Iterating on n ,

$$\max_\ell P_{\ell j}^n - \min_\ell P_{\ell j}^n \leq (1 - 2\alpha)^n$$

Finally, we can get back to arbitrary finite-state ergodic chains with transition matrix $[P]$.

We have shown that $[P^h]$ is positive for $h = (M - 1)^2 + 1$, so we can apply Lemma 2 to $[P^h]$, with $\alpha = \min_{ij} P_{ij}^h$.

We don't much care about the value of α , but only that it is positive. Then

$$\lim_{m \rightarrow \infty} \max_{\ell} P_{\ell j}^{hm} = \min_{\ell} P_{\ell j}^{hm} = \pi_{\ell} > 0$$

To show that the limit applies for all n rather than just multiples of h , we use Lemma 1, showing that $\max_{\ell} P_{\ell j}^n$ is non-increasing in n , so it must have the same limit as $\max_{\ell} P_{\ell j}^{hm}$. The same argument applies for the minima. QED

Ergodic unichains

We have now seen that for ergodic chains, $\lim_{n \rightarrow \infty} P_{ij}^n = \pi_j$ for all i where $\vec{\pi}$ is a probability vector. The resulting vector $\vec{\pi}$ is also a steady-state vector and is the unique probability vector solution to $\vec{\pi}[P] = \vec{\pi}$ (see Thm 3.3.1).

It is fairly easy to extend this result to a more general class called ergodic unichains. These are chains containing a single ergodic class along with an arbitrary set of transient states.

If a state is in a singleton transient class, then there is a fixed probability, say α , of leaving the class at each step, and the probability of remaining in the class for more than n steps is $(1 - \alpha)^n$.

The probability of remaining in an arbitrary set of transient states also decays to 0 exponentially with n . Essentially each transient state has at least one path to a recurrent state, and one of those paths must be taken eventually.

For an ergodic unichain, the ergodic class is eventually entered, and then steady state in that class is reached.

For every state j then,

$$\lim_{n \rightarrow \infty} \max_i P_{ij}^n = \lim_{n \rightarrow \infty} \min_i P_{ij}^n = \pi_j$$

The difference here is that $\pi_j = 0$ for each transient state and $\pi_j > 0$ for each recurrent state.

Other finite-state Markov chains

First consider a Markov chain with several ergodic classes, $\mathcal{C}_1, \dots, \mathcal{C}_m$. The classes don't communicate and should be considered separately.

If one insists on analyzing the entire chain, $[P]$ will have m independent steady state vectors, one nonzero on each class. $[P^n]$ will then converge, but the rows will not all be the same.

There will be m sets of rows, one for each class, and the row for class k will be nonzero only for the elements of that class.

Next consider a periodic recurrent chain of period d . This can be separated into d subclasses with a cyclic rotation between them.

If we look at $[P^d]$, we see that each subclass becomes an ergodic class, say $\mathcal{C}_1, \dots, \mathcal{C}_d$. Thus $\lim_{n \rightarrow \infty} [P^{nd}]$ exists.

A steady state is reached within each subclass, but the chain rotates from one subclass to another.

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes
Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Lecture 8: Markov eigenvalues and eigenvectors

Outline:

- Review of ergodic unichains
- Review of basic linear algebra facts
- Markov chains with 2 states
- Distinct eigenvalues for $M > 2$ states
- M states and M independent eigenvectors
- The Jordan form

1

Recall that for an ergodic finite-state Markov chain, the transition probabilities reach a limit in the sense that $\lim_{n \rightarrow \infty} P_{ij}^n = \pi_j$ where $\vec{\pi} = (\pi_1, \dots, \pi_M)$ is a strictly positive probability vector.

Multiplying both sides by P_{jk} and summing over j ,

$$\pi_k = \lim_{n \rightarrow \infty} \sum_j P_{ij}^n P_{jk} = \sum_j \pi_j P_{jk}$$

Thus $\vec{\pi}$ is a steady-state vector for the Markov chain, i.e., $\vec{\pi} = \vec{\pi}[P]$ and $\vec{\pi} \geq 0$.

In matrix terms, $\lim_{n \rightarrow \infty} [P^n] = \vec{e}\vec{\pi}$ where $\vec{e} = (1, 1, \dots, 1)^\top$ is a column vector and $\vec{\pi}$ is a row vector.

2

The same result almost holds for ergodic unichains, i.e., one ergodic class plus an arbitrary set of transient states.

The sole difference is that the steady-state vector is positive for all ergodic states and 0 for all transient states.

$$[P] = \left[\begin{array}{c|c} [P_T] & [P_{TR}] \\ \hline [0] & [P_R] \end{array} \right] \quad \text{where} \quad [P_T] = \left[\begin{array}{ccc} P_{11} & \dots & P_{1t} \\ \dots & \dots & \dots \\ P_{t1} & \dots & P_{tt} \end{array} \right]$$

The idea is that each transient state eventually has a transition (via $[P_{TR}]$) to a recurrent state, and the class of recurrent states lead to steady state as before.

3

Review of basic linear algebra facts

Def: A complex number λ is an eigenvalue of a real square matrix $[A]$, and a complex vector $\vec{v} \neq 0$ is a right eigenvector of $[A]$, if $\lambda\vec{v} = [A]\vec{v}$.

For every stochastic matrix (the transition matrix of a finite-state Markov chain $[P]$), we have $\sum_j P_{ij} = 1$ and thus $[P]\vec{e} = \vec{e}$.

Thus $\lambda = 1$ is an eigenvalue of an arbitrary stochastic matrix $[P]$ with right eigenvector \vec{e} .

An equivalent way to express the eigenvalue/eigenvector equation is that $[P - \lambda I]\vec{v} = 0$ where I is the identity matrix.

4

Def: A square matrix $[A]$ is singular if there is a vector $\vec{v} \neq 0$ such that $[A]\vec{v} = 0$.

Thus λ is an eigenvalue of $[P]$ if and only if (iff) $[P - \lambda I]$ is singular for some $\vec{v} \neq 0$.

Let $\vec{a}_1, \dots, \vec{a}_M$ be the columns of $[A]$. Then $[A]$ is singular iff $\vec{a}_1, \dots, \vec{a}_M$ are linearly dependent.

The square matrix $[A]$ is singular iff the rows of $[A]$ are linearly dependent and iff the determinant $\det[A]$ of $[A]$ is 0.

Summary: λ is an eigenvalue of $[P]$ iff $[P - \lambda I]$ is singular, iff $\det[P - \lambda I] = 0$, iff $[P]\vec{v} = \lambda\vec{v}$ for some $\vec{v} \neq 0$, and iff $\vec{u}[P] = \lambda\vec{u}$ for some $\vec{u} \neq 0$.

5

For every stochastic matrix $[P]$, $[P]\vec{e} = \vec{e}$ and thus $[P - I]$ is singular and there is a row vector $\pi \neq 0$ such that $\vec{\pi}[P] = \vec{\pi}$.

This does not show that there is a probability vector $\vec{\pi}$ such that $\vec{\pi}[P] = \vec{\pi}$, but we already know there is such a probability vector (i.e., a steady-state vector) if $[P]$ is the matrix of an ergodic unichain.

We show later that there is a steady-state vector π for all Markov chains.

6

The determinant of an M by M matrix can be determined as

$$\det A = \sum_{\mu} \pm \prod_{i=1}^M A_{i,\mu(i)}$$

where the sum is over all permutations μ of the integers $1, \dots, M$. Plus is used for each even permutation and minus for each odd.

The important facet of this formula for us is that $\det[P - \lambda I]$ must be a polynomial in λ of degree M .

Thus there are M roots of the equation $\det[P - \lambda I] = 0$, and consequently M eigenvalues of $[P]$.

Some of these M eigenvalues might be the same, and if k of these roots are equal to λ , the eigenvalue λ is said to have algebraic multiplicity k .

7

Markov chains with 2 states

$$\begin{array}{ll} \pi_1 P_{11} + \pi_2 P_{21} = \lambda \pi_1 & P_{11} \nu_1 + P_{12} \nu_2 = \lambda \nu_1 \\ \pi_1 P_{12} + \pi_2 P_{22} = \lambda \pi_2 & P_{21} \nu_1 + P_{22} \nu_2 = \lambda \nu_2 \end{array} .$$

left eigenvector right eigenvector

$$\det[P - \lambda I] = (P_{11} - \lambda)(P_{22} - \lambda) - P_{12}P_{21}$$

$$\lambda_1 = 1; \quad \lambda_2 = 1 - P_{12} - P_{21}$$

If $P_{12} = P_{21} = 0$ (the chain has 2 recurrent classes), then $\lambda = 1$ has multiplicity 2. Otherwise $\lambda = 1$ has multiplicity 1.

If $P_{12} = P_{21} = 1$ (the chain is periodic), then $\lambda_2 = -1$. Otherwise $|\lambda_2| < 1$.

8

$$\begin{array}{ll} \pi_1 P_{11} + \pi_2 P_{21} = \lambda \pi_1 & P_{11} \nu_1 + P_{12} \nu_2 = \lambda \nu_1 \\ \pi_1 P_{12} + \pi_2 P_{22} = \lambda \pi_2 & P_{21} \nu_1 + P_{22} \nu_2 = \lambda \nu_2 \end{array}.$$

$$\lambda_1 = 1; \quad \lambda_2 = 1 - P_{12} - P_{21}$$

Assume throughout that either $P_{12} > 0$ or $P_{21} > 0$. Then

$$\begin{array}{ll} \vec{\pi}^{(1)} = \left(\frac{P_{21}}{P_{12}+P_{21}}, \frac{P_{12}}{P_{12}+P_{21}} \right) & \vec{\nu}^{(1)} = (1, 1)^c \\ \vec{\pi}^{(2)} = (1, -1) & \vec{\nu}^{(2)} = \left(\frac{P_{12}}{P_{12}+P_{21}}, \frac{-P_{21}}{P_{12}+P_{21}} \right)^c \end{array}$$

Note that $\vec{\pi}^{(i)} \vec{\nu}^{(j)} = \delta_{ij}$. In general, if $\vec{\pi}^{(i)}[P] = \lambda_i \vec{\pi}^{(i)}$ and $[P] \vec{\nu}^{(i)} = \lambda_i \vec{\nu}^{(i)}$ for $i = 1, \dots, M$, then $\vec{\pi}^{(i)} \vec{\nu}^{(j)} = 0$ if $\lambda_i \neq \lambda_j$. To see this,

$$\lambda_i \vec{\pi}^{(i)} \vec{\nu}^{(j)} = \vec{\pi}^{(i)}[P] \vec{\nu}^{(j)} = \vec{\pi}^{(i)}(\lambda_j \vec{\nu}^{(j)}) = \lambda_j \vec{\pi}^{(i)} \vec{\nu}^{(j)}$$

so if $\lambda_i \neq \lambda_j$, then $\vec{\pi}_i \vec{\nu}_j = 0$. Normalization (of either $\vec{\pi}_i$ or $\vec{\nu}_i$) can make $\vec{\pi}_i \vec{\nu}_i = 1$ for each i .

Note that the equations

$$P_{11} \nu_1^{(i)} + P_{12} \nu_2^{(i)} = \lambda_i \nu_1^{(i)}; \quad P_{21} \nu_1^{(i)} + P_{22} \nu_2^{(i)} = \lambda_i \nu_2^{(i)}$$

can be rewritten in matrix form as

$$[P][U] = [U][\Lambda] \quad \text{where}$$

$$[\Lambda] = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \quad \text{and} \quad [U] = \begin{bmatrix} \nu_1^{(1)} & \nu_1^{(2)} \\ \nu_2^{(1)} & \nu_2^{(2)} \end{bmatrix},$$

Since $\vec{\pi}^{(i)} \vec{\nu}^{(j)} = \delta_{ij}$, we see that

$$\begin{bmatrix} \pi_1^{(1)} & \pi_2^{(1)} \\ \pi_1^{(2)} & \pi_2^{(2)} \end{bmatrix} \begin{bmatrix} \nu_1^{(1)} & \nu_1^{(2)} \\ \nu_2^{(1)} & \nu_2^{(2)} \end{bmatrix} = [I],$$

so $[U]$ is invertible and $[U^{-1}]$ has $\vec{\pi}^{(1)}$ and $\vec{\pi}^{(2)}$ as rows. Thus $[P] = [U][\Lambda][U^{-1}]$ and

$$[P^2] = [U][\Lambda][U^{-1}][U][\Lambda][U^{-1}] = [U][\Lambda^2][U^{-1}]$$

Similarly, for any $n \geq 2$,

$$[P^n] = [U][\Lambda^n][U^{-1}] \quad (1)$$

Eq. 3.29 in text has a typo and should be (1) above.

We can solve (1) in general (if all M eigenvalues are distinct) as easily as for M = 2.

Break $[\Lambda]^n$ into M terms,

$$[\Lambda]^n = [\Lambda_1^n] + \cdots + [\Lambda_M^n] \quad \text{where}$$

$[\Lambda_i^n]$ has λ_i^n in position (i, i) and has zeros elsewhere.
Then

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

11

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

$$\begin{aligned} \vec{\pi}^{(1)} &= \left(\frac{P_{21}}{P_{12}+P_{21}}, \frac{P_{12}}{P_{12}+P_{21}} \right) & \vec{\nu}^{(1)} &= (1, 1)^c \\ \vec{\pi}^{(2)} &= (1, -1) & \vec{\nu}^{(2)} &= \left(\frac{P_{12}}{P_{12}+P_{21}}, \frac{-P_{21}}{P_{12}+P_{21}} \right)^c \end{aligned}$$

The steady-state vector is $\vec{\pi} = \vec{\pi}^{(1)}$ and

$$\vec{\nu}^{(1)} \vec{\pi} = \begin{bmatrix} \pi_1 & \pi_2 \\ \pi_1 & \pi_2 \end{bmatrix} \quad \vec{\nu}^{(2)} \vec{\pi}^{(2)} = \begin{bmatrix} \pi_2 & -\pi_2 \\ -\pi_1 & \pi_1 \end{bmatrix}$$

$$[P^n] = \begin{bmatrix} \pi_1 + \pi_2 \lambda_2^n & \pi_2 - \pi_2 \lambda_2^n \\ \pi_1 - \pi_1 \lambda_2^n & \pi_2 + \pi_1 \lambda_2^n \end{bmatrix}$$

We see that $[P^n]$ converges to $\vec{e}\vec{\pi}$, and the rate of convergence is λ_2 . This solution is exact. It essentially extends to arbitrary finite M.

12

Distinct eigenvalues for $M > 2$ states

Recall that, for an M state Markov chain, $\det[P - \lambda I]$ is a polynomial of degree M in λ . It thus has M roots (eigenvalues), which we assume here to be distinct.

Each eigenvalue λ_i has a right eigenvector $\vec{v}^{(i)}$ and a left eigenvector $\vec{\pi}^{(i)}$. Also $\vec{\pi}^{(i)}\vec{v}^{(j)} = 0$ for each $i, j \neq i$.

By scaling $\vec{v}^{(i)}$ or $\vec{\pi}^{(i)}$, we can satisfy $\vec{\pi}^{(i)}\vec{v}^{(i)} = 1$.

Let $[U]$ be the matrix with columns $\vec{v}^{(1)}$ to $\vec{v}^{(M)}$ and let $[V]$ have rows $\vec{\pi}^{(1)}$ to $\vec{\pi}^{(M)}$.

Then $[V][U] = I$, so $[V] = [U^{-1}]$. Thus the eigenvectors $\vec{v}^{(1)}$ to $\vec{v}^{(M)}$ are linearly independent and span M space. Same with $\vec{\pi}^{(1)}$ to $\vec{\pi}^{(M)}$.

13

Putting the right eigenvector equations together, $[P][U] = [U][\Lambda]$. Postmultiplying by $[U^{-1}]$, this becomes

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

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

Breaking $[\Lambda^n]$ into a sum of M terms as before,

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

Since each row of $[P]$ sums to 1, \vec{e} is a right eigenvector of eigenvalue 1.

Thm: The left eigenvector $\vec{\pi}$ of eigenvalue 1 is a steady-state vector if it is normalized to $\vec{\pi}\vec{e} = 1$.

14

Thm: The left eigenvector $\vec{\pi}$ of eigenvalue 1 is a steady-state vector if it is normalized to $\vec{\pi}\vec{e} = 1$.

Pf: There must be a left eigenvector $\vec{\pi}$ for eigenvalue 1. For every j , $1 \leq j \leq M$, $\pi_j = \sum_k \pi_k P_{kj}$. Taking magnitudes,

$$|\pi_j| \leq \sum_k |\pi_k| P_{kj} \quad (2)$$

with equality iff $\pi_j = |\pi_j| e^{i\phi}$ for all j and some ϕ . Summing over j , $\sum_j |\pi_j| \leq \sum_k |\pi_k|$. This is satisfied with equality, so (2) is satisfied with equality for each j .

Thus $(|\pi_1|, |\pi_2|, \dots, |\pi_M|)$ is a nonnegative vector satisfying the steady-state vector equation. Normalizing to $\sum_j |\pi_j| = 1$, we have a steady-state vector.

15

Thm: Every eigenvalue λ_ℓ satisfies $|\lambda_\ell| \leq 1$.

Pf: We have seen that if $\vec{\pi}^{(\ell)}$ is a left eigenvector of $[P]$ with eigenvalue λ_ℓ , then it is also a left eigenvector of $[P^n]$ with eigenvalue λ_ℓ^n . Thus

$$\begin{aligned} \lambda_\ell^n \pi_j^{(\ell)} &= \sum_i \pi_i^{(\ell)} P_{ij}^n \quad \text{for all } j. \\ |\lambda_\ell^n| |\pi_j^{(\ell)}| &\leq \sum_i |\pi_i^{(\ell)}| P_{ij}^n \quad \text{for all } j. \end{aligned}$$

Let β be the largest of $|\pi_j^{(\ell)}|$ over j . For that maximizing j ,

$$|\lambda_\ell^n| \beta \leq \sum_i \beta P_{ij}^n \leq \beta M$$

Thus $|\lambda_\ell^n| \leq M$ for all n , so $|\lambda_\ell| \leq 1$.

16

These two theorems are valid for all finite-state Markov chains. For the case with M distinct eigenvalues, we have

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

If the chain is an ergodic unichain, then one eigenvalue is 1 and the rest are strictly less than 1 in magnitude.

Thus the rate at which $[P^n]$ approaches $\vec{\pi}$ is determined by the second largest eigenvalue.

If $[P]$ is a periodic unichain with period d , then there are d eigenvalues equally spaced around the unit circle and $[P^n]$ does not converge.

17

M states and M independent eigenvectors

Next assume that one or more eigenvalues have multiplicity greater than 1, but that if an eigenvalue has multiplicity k , then it has k linearly independent eigenvectors.

We can choose the left eigenvectors of a given eigenvalue to be orthonormal to the right eigenvectors of that eigenvalue.

After doing this and defining $[U]$ as the matrix with columns $\vec{\nu}^{(1)}, \dots, \vec{\nu}^{(M)}$, we see $[U]$ is invertible and that $[U^{-1}]$ is the matrix with rows $\vec{\pi}^{(1)}, \dots, \vec{\pi}^{(M)}$. We then again have

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

18

Example: Consider a Markov chain consisting of ℓ ergodic sets of states.

Then each ergodic set will have an eigenvalue equal to 1 with a right eigenvector equal to 1 on the states of that set and 0 elsewhere.

There will also be a ‘steady-state’ vector, nonzero only on that set of states.

Then $[P^n]$ will converge to a block diagonal matrix where for each ergodic set, the rows within that set are the same.

19

The Jordan form

Unfortunately, it is possible that an eigenvalue of algebraic multiplicity $k \geq 2$ has fewer than k linearly independent eigenvectors.

The decomposition $[P] = [U][\Lambda][U^{-1}]$ can be replaced in this case by a Jordan form, $[P] = [U][J][U^{-1}]$ where $[J]$ has the form

$$[J] = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & \lambda_2 & 1 & 0 \\ 0 & 0 & 0 & \lambda_2 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{bmatrix}.$$

The eigenvalues are on the main diagonal and ones are on the next diagonal up where needed for deficient eigenvectors.

20

Example:

$$[P] = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \end{bmatrix}.$$

The eigenvalues are 1 and 1/2, with algebraic multiplicity 2 for $\lambda = 1/2$.

There is only one eigenvector (subject to a scaling constant) for the eigenvalue 1/2. $[P^n]$ approaches steady-state as $n(1/2)^n$.

Fortunately, if $[P]$ is stochastic, the eigenvalue 1 always has as many linearly independent eigenvectors as its algebraic multiplicity.

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Lecture 9: Markov rewards and dynamic prog.**Outline:**

- Review plus of eigenvalues and eigenvectors
- Rewards for Markov chains
- Expected first-passage-times
- Aggregate rewards with a final reward
- Dynamic programming
- Dynamic programming algorithm

1

The determinant of an M by M matrix is given by

$$\det A = \sum_{\mu} \pm \prod_{i=1}^M A_{i,\mu(i)} \quad (1)$$

where the sum is over all permutations μ of the integers $1, \dots, M$. If $[A_T]$ is t by t , with

$$[A] = \left[\begin{array}{c|c} [A_T] & [A_{T\mathcal{R}}] \\ \hline [0] & [A_{\mathcal{R}}] \end{array} \right] \quad \det[A] = \det[A_T] \det[A_{\mathcal{R}}]$$

The reason for this is that for the product in (1) to be nonzero, $\mu(i) > t$ whenever $i > t$, and thus $\mu(i) \leq t$ when $i \leq t$. Thus the permutations can be factored into those over 1 to t and those over $t+1$ to M .

$$\det[P - \lambda I] = \det[P_T - \lambda I_t] \det[P_{\mathcal{R}} - \lambda I_r]$$

2

$$\det[P - \lambda I] = \det[P_T - \lambda I_t] \det[P_{\mathcal{R}} - \lambda I_r]$$

The eigenvalues of $[P]$ are the t eigenvalues of $[P_T]$ and the r eigenvalues of $[P_{\mathcal{R}}]$.

If $\vec{\pi}$ is a left eigenvector of $[P_{\mathcal{R}}]$, then $(0, \dots, 0, \pi_1, \dots, \pi_r)$ is a left eigenvector of $[P]$, i.e.,

$$(\vec{0} | \vec{\pi}) \left[\begin{array}{c|c} [P_T] & [P_{T\mathcal{R}}] \\ \hline [0] & [P_{\mathcal{R}}] \end{array} \right] = \lambda (\vec{0} | \vec{\pi})$$

The left eigenvectors of $[P_T]$ are more complicated but not very interesting.

3

Next, assume

$$[P] = \left[\begin{array}{c|c|c} [P_T] & [P_{T\mathcal{R}}] & [P_{T\mathcal{R}'}] \\ \hline [0] & [P_{\mathcal{R}}] & [0] \\ \hline [0] & [0] & [P_{\mathcal{R}'}] \end{array} \right]$$

In the same way as before,

$$\det[P - \lambda I] = \det[P_T - \lambda I_t] \det[P_{\mathcal{R}} - \lambda I_r] \det[P_{\mathcal{R}'} - \lambda I_{r'}]$$

The eigenvalues of $[P]$ are comprised of the t from $[P_T]$, the r from $[P_{\mathcal{R}}]$, and the r' from $[P_{\mathcal{R}'}]$.

If $\vec{\pi}$ is a left eigenvector of $[P_{\mathcal{R}}]$, then $(\vec{0}, \vec{\pi}, \vec{0})$ is a left eigenvector of $[P]$. If $\vec{\pi}$ is a left eigenvector of $[P_{\mathcal{R}'}]$, then $(\vec{0}, \vec{0}, \vec{\pi})$ is a left eigenvector of $[P]$.

4

Rewards for Markov chains

Suppose that each state i of a Markov chain is associated with a given reward, r_i .

Letting the rv X_n be the state at time n , the (random) reward at time n is the rv $R(X_n)$ that maps $X_n = i$ into r_i for each i .

We will be interested only in expected rewards, so that, for example, the expected reward at time n , given that $X_0 = i$ is $E[R(X_n)|X_0 = i] = \sum_j r_j P_{ij}^n$.

The expected aggregate reward over the n steps from m to $m+n-1$, conditional on $X_m = i$ is then

$$\begin{aligned} v_i(n) &= E[R(X_m) + \dots + R(X_{m+n-1})|X_m = i] \\ &= r_i + \sum_j P_{ij}r_j + \dots + \sum_j P_{ij}^{n-1}r_j \end{aligned}$$

5

$$v_i(n) = r_i + \sum_j P_{ij}r_j + \dots + \sum_j P_{ij}^{n-1}r_j$$

If the Markov chain is an ergodic unichain, then successive terms of this expression tend to a steady state gain per step,

$$g = \sum_j \pi_j r_j,$$

which is independent of the starting state. Thus $v_i(n)$ can be viewed as a transient in i plus ng .

The transient is important, and is particularly important if $g = 0$.

6

Expected first-passage-time

Suppose, for some arbitrary unichain, that we want to find the expected number of steps, starting from a given state i until some given recurrent state, say 1, is first entered. Assume $i \neq 1$.

This can be viewed as a reward problem by assigning one unit of reward to each successive state until state 1 is entered.

Modify the Markov chain by changing the transition probabilities from state 1 to $P_{11} = 1$. We set $r_1 = 0$, so the reward stops when state 1 is entered.

For each sample path starting from state $i \neq 1$, the probability of the initial segment until 1 is entered is unchanged, so the expected first-passage-time is unchanged.

7

The modified Markov chain is now an ergodic unichain with a single recurrent state, i.e., state 1 is a trapping state.

Let $r_i = 1$ for $i \neq 1$ and let $r_1 = 0$.

Thus if state 1 is first entered at time ℓ , then the aggregate reward, from 0 to n , is ℓ for all $n \geq \ell$.

The expected first passage time, starting in state i , is $v_i = \lim_{n \rightarrow \infty} v_i(n)$.

There is a sneaky way to calculate this for all i .

For each $i \neq 1$, assume that $X_0 = i$. There is then a unit reward at time 0. In addition, given that $X_1 = j$, the remaining expected reward is v_j . Thus $v_i = 1 + \sum_j P_{ij}v_j$ for $i \neq 1$, with $v_1 = 0$.

8

The expected first-passage-time to state 1 from state $i \neq 1$ is then

$$v_i = 1 + \sum_j P_{ij} v_j \quad \text{with } v_1 = 0$$

This can be expressed in vector form as

$\vec{v} = \vec{r} + [P]\vec{v}$ where $\vec{r} = (0, 1, 1, \dots, 1)$, and $v_1 = 0$ and $P_{11} = 1$.

Note that if \vec{v} satisfies $\vec{v} = \vec{r} + [P]\vec{v}$, then $\vec{v} + \alpha\vec{e}$ also satisfies it, so that $v_1 = 0$ is necessary to resolve the ambiguity.

Also, since $[P]$ has 1 as a simple eigenvalue, this equation has a unique solution with $v_1 = 0$.

9

Aggregate rewards with a final reward

There are many situations in which we are interested in the aggregate reward over n steps, say time m to $m+n-1$, followed by a special final reward u_j for $X_{m+n} = j$.

The flexibility of assigning such a final reward will be particularly valuable in dynamic programming.

The aggregate expected reward, including this final reward, is then

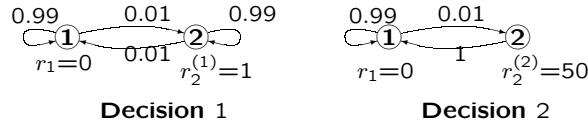
$$v_i(n, \vec{u}) = r_i + \sum_j P_{ij} r_j + \dots + \sum_j P_{ij}^{n-1} r_j + \sum_j P_{ij}^n u_j$$

In vector form,

$$\vec{v}(n, \vec{u}) = \vec{r} + [P]\vec{r} + \dots + [P^{n-1}]\vec{r} + [P^n]\vec{u}$$

Dynamic programming

Consider a discrete-time situation with a finite set of states, $1, 2, \dots, M$, where at each time ℓ , a decision maker can observe the state, say $X_\ell = j$ and choose one of a finite set of alternatives. Each alternative k consists of a current reward $r_j^{(k)}$ and a set of transition probabilities $\{P_{j\ell}^{(k)}; 1 \leq \ell \leq M\}$ for going to the next state.



For this example, decision 2 seeks instant gratification, whereas decision 1 seeks long term gratification.

11

Assume that this process of random transitions combined with decisions based on the current state starts at time m in some given state and continues until time $m + n - 1$.

After the n th decision, made at time $m + n - 1$, there is a final transition based on that decision.

At time $m + n$, there is a final reward, (u_1, \dots, u_M) based on the final state.

12

The objective of dynamic programming is both to determine the optimal decision at each time and to determine the expected reward for each starting state and for each number n of steps.

As one might suspect, it is best to start with a single step ($n = 1$) and then proceed to successively more steps.

Surprisingly, this is best thought of as starting at the end and working back to the beginning.

The algorithm to follow is due to Richard Bellman. Its simplicity is a good example of looking at an important problem at the right time. Given the formulation, anyone could develop the algorithm.

13

The dynamic programming algorithm

As suggested, we first consider the optimal expected aggregate reward over a single time period.

That is, starting at an arbitrary time m in a given state i , we make a decision, say decision k at time m .

This provides a reward $r_i^{(k)}$ at time m . Then the selected transition probabilities, P_{ij}^k lead to a final expected reward $\sum_j u_j P_{ij}^k$ at time $m + 1$.

The decision k is chosen to maximize the corresponding aggregate reward, i.e.,

$$v_i^*(1) = \max_k \left(r_i(k) + \sum_j u_j P_{ij}^k \right)$$

14

Next consider $v_i^*(2, \vec{u})$, i.e., the maximal expected aggregate reward starting at $X_m = i$ with decisions made at times m and $m + 1$ and a final reward at time $m + 2$.

The key to dynamic programming is that an optimal decision at time $m + 1$ can be selected based only on the state j at time $m + 1$.

This decision (given $X_{m+1} = j$) is optimal independent of the decision at time m .

That is, whatever decision is made at time m , the maximal expected reward at times $m + 1$ and $m + 2$, given $X_{m+1} = j$, is $\max_k \left(r_j^{(k)} + \sum_\ell P_{j\ell}^{(k)} u_\ell \right)$. This is $v_j^*(1, \vec{u})$, as just found.

15

We have just seen that

$$v_j^*(1, \vec{u}) = \max_k \left(r_j^{(k)} + \sum_\ell P_{j\ell}^{(k)} u_\ell \right)$$

is the maximum expected aggregate reward over times $m + 1$ and $m + 2$, conditional $X_{m+1} = j$.

Thus the maximum expected aggregate reward over $m, m + 1, m + 2$, conditional on $X_m = i$, is

$$v_i^*(2, \vec{u}) = \max_k \left(r_i^{(k)} + \sum_j P_{ij}^{(k)} v_j^*(1, \vec{u}) \right)$$

16

This same procedure can be used to find the optimal policy and optimal expected reward for $n = 3$,

$$v_i^*(3, \vec{u}) = \max_k \left(r_i^{(k)} + \sum_j P_{ij}^{(k)} v_j^*(2, \vec{u}) \right)$$

This solution shows how to choose the optimal decision at time m and finds the optimal aggregate expected reward, but is based on first finding the optimal solution for $n = 2$. In general,

$$v_i^*(n, \vec{u}) = \max_k \left(r_i^{(k)} + \sum_j P_{ij}^{(k)} v_j^*(n-1, \vec{u}) \right)$$

For any given n , then, the algorithm calculates $v_j^*(m, \vec{u})$ for all states j and all $m \leq n$, starting at $m = 1$.

This is the dynamic programming algorithm.

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Lecture 10: Renewals and the SLLN

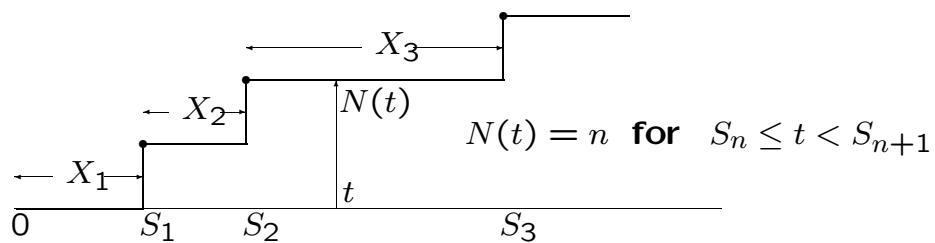
Outline:

- Arrival processes and renewal processes
- Convergence WP1 and the SLLN
- Proof of convergence WP1 theorem
- The strong law with a 4th moment
- SLLN and WLLN
- Strange aspects of the SLLN

1

Renewal processes

Recall that an arrival process can be specified by its arrival epochs, $\{S_1, S_2, \dots\}$, or by its interarrival intervals, $\{X_1, X_2, \dots\}$, or by its counting process, $\{N(t); t > 0\}$.



Def: A renewal process is an arrival process for which the interarrival intervals X_1, X_2, \dots are IID.

2

Renewal processes are characterized by ‘starting over’ at each renewal.

We will see later exactly what this means, but intuitively it means that complex processes can be broken into discrete time intervals.

Renewal theory treats the gross characteristics of this (how many intervals occur per unit time, laws of large numbers about long term behavior, etc.)

The local characteristics can then be studied without worrying about long term interactions.

3

Example: Markov chains, for any given recurrent state, have renewals on successive visits to that state.

It is almost obvious that the intervals between visits to a given recurrent state are independent, but we look at this carefully later.

A non-obvious result that will arise from studying renewals is that the expected recurrence time between visits to state i is π_i . We could have derived that from Markov chains directly, but using renewals is cleaner.

The whole theory for Markov chains with a countably infinite state space will come from renewal processes.

4

Example: G/G/m queues (queues with a general IID interarrival distribution, a general IID service interval, and m servers)

The queue is assumed empty at $t < 0$ and an arrival is assumed at $t = 0$.

A very complicated interaction goes on between arrivals and departures, until finally the queue empties out.

After some interval depending on this busy period, a new arrival occurs. This can be taken as a renewal.

5

Proof of theorem on convergence WP1

Thm: let $\{Y_n; n \geq 1\}$ be rv's satisfying $\sum_{n=1}^{\infty} E[|Y_n|] < \infty$. Then $\{Y_n; n \geq 1\}$ converges to 0 WP1.

That is, we want to prove that $\{\omega \in \Omega : \lim_{n \rightarrow \infty} Y_n(\omega) = 0\} = 1$.

Proof: For any $\alpha > 0$ and any finite integer $m \geq 1$, Markov says that

$$\begin{aligned} \Pr \left\{ \sum_{n=1}^m |Y_n| > \alpha \right\} &\leq \frac{E \left[\sum_{n=1}^m |Y_n| \right]}{\alpha} = \frac{\sum_{n=1}^m E[|Y_n|]}{\alpha} \\ &\leq \frac{\sum_{n=1}^{\infty} E[|Y_n|]}{\alpha}. \end{aligned}$$

$$\Pr\left\{\sum_{n=1}^m |Y_n| > \alpha\right\} \leq \frac{\sum_{n=1}^{\infty} E[|Y_n|]}{\alpha}.$$

Let $A_m = \{\omega : \sum_{n=1}^m |Y_n(\omega)| > \alpha\}$, **so that**

$$\Pr\left\{\sum_{n=1}^m |Y_n| > \alpha\right\} = \Pr\{A_m\} \leq \frac{\sum_{n=1}^{\infty} E[|Y_n|]}{\alpha}. \quad (1)$$

Since $|Y_n(\omega)| \geq 0$, we have $A_m \subseteq A_{m+1}$ for $m \geq 1$. Thus the left side of (1), as a function of m , is a nondecreasing bounded sequence of real numbers. Thus

$$\lim_{m \rightarrow \infty} \Pr\left\{\sum_{n=1}^m |Y_n| > \alpha\right\} = \lim_{m \rightarrow \infty} \Pr\{A_m\} \leq \frac{\sum_{n=1}^{\infty} E[|Y_n|]}{\alpha}.$$

Also, by property (9) (nesting) of the probability axioms,

$$\Pr\left\{\bigcup_{m=1}^{\infty} A_m\right\} = \lim_{m \rightarrow \infty} \Pr\{A_m\}$$

8

Summarizing, $A_m = \{\omega : \sum_{n=1}^m |Y_n(\omega)| > \alpha\}$ **and**

$$\Pr\left\{\bigcup_{m=1}^{\infty} A_m\right\} \leq \frac{\sum_{n=1}^{\infty} E[|Y_n|]}{\alpha}.$$

For given ω , $\sum_{n=1}^m |Y_n(\omega)|$ **is a sequence (in m) of real numbers that is nondecreasing in m .** Either all these real numbers are upper bounded by α , in which case

$$\sum_{n=1}^{\infty} |Y_n(\omega)| \leq \alpha; \quad \omega \notin \bigcup_{m=1}^{\infty} A_m$$

or $\sum_{n=1}^m |Y_n(\omega)| > \alpha$ **for some m ,** and

$$\sum_{n=1}^{\infty} |Y_n(\omega)| > \alpha; \quad \omega \in \bigcup_{m=1}^{\infty} A_m$$

Thus

$$\Pr\left\{\omega : \sum_{n=1}^{\infty} |Y_n(\omega)| > \alpha\right\} \leq \frac{\sum_{n=1}^{\infty} E[|Y_n|]}{\alpha}.$$

9

For $\alpha > \sum_n E[|Y_n|]$

$$\Pr\left\{\omega : \sum_{n=1}^{\infty} |Y_n(\omega)| \leq \alpha\right\} > 1 - \frac{\sum_{n=1}^{\infty} E[|Y_n|]}{\alpha}.$$

If $\sum_n |Y_n(\omega)| \leq \alpha$ **for a given** ω , **then** $\sum_n |Y_n(\omega)|$ **converges and** $\lim_{m \rightarrow \infty} \sum_{n=m}^{\infty} |Y_n(\omega)| = 0$. **Thus**

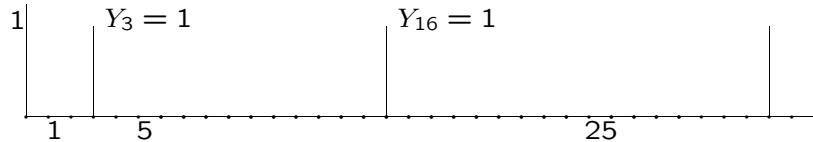
$$\Pr\left\{\omega : \sum_{n=1}^{\infty} |Y_n(\omega)| \leq \alpha\right\} \implies \lim_{n \rightarrow \infty} |Y_n(\omega)| = 0$$

$$\Pr\left\{\omega : \lim_{n \rightarrow \infty} |Y_n(\omega)| = 0\right\} > 1 - \frac{\sum_{n=1}^{\infty} E[|Y_n|]}{\alpha}.$$

Let $\alpha \rightarrow \infty$. **Then** $\Pr\{\omega : \lim_{n \rightarrow \infty} |Y_n(\omega)| = 0\} = 1$. \square

10

Example: (convergence in probability but not WP1). $\{Y_n; n \geq 1\}$. **For each** $j \geq 0$, $Y_n = 1$ **for an equiprobable choice of** $n \in [5^j, 5^{j+1})$ **and** $Y_n = 0$ **otherwise.**



For every ω, j , $Y_n(\omega)$ **is 1 for some** $n \in [5^j, 5^{j+1})$ **and is 0 elsewhere.** **Thus** $Y_n(\omega)$ **does not converge for any** ω ; **i.e.,** $\{Y_n; n \geq 1\}$ **doesn't converge WP1.**

Note that $E[|Y_n|] = 1/(5^{j+1} - 5^j)$ **for** $5^j \leq n < 5^{j+1}$ **and thus** $\sum_n E[|Y_n|] = \infty$, **so the theorem doesn't apply.**

However, $\lim_{n \rightarrow \infty} \Pr\{|Y_n| > \epsilon\} = 0$ **for all** $\epsilon > 0$, **so** $\{Y_n; n \geq 1\}$ **converges in probability.**

11

The strong law of large numbers

Now that we have a convenient property implying convergence WP1, we use this property to prove the strong law of large numbers.

Theorem: Let $\{X_n; n \geq 1\}$ be a sequence of IID rv's satisfying $E[|X|] < \infty$. For each $n \geq 1$, let $S_n = \sum_{m=1}^{\infty} X_m$. Then

$$\Pr\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} \frac{S_n(\omega)}{n} = \bar{X}\right\} = 1$$

This is stated more tersely as $\Pr\{\lim_n S_n/n = \bar{X}\} = 1$, more tersely yet as $\lim_n S_n/n = \bar{X}$ WP1 and still more tersely as $S_n/n \xrightarrow{\text{WP1}} \bar{X}$. The meaning of the theorem is complicated and the terse forms sometimes conceal this meaning.

12

Discussion: The strong law (and convergence WP1) is quite different from the other forms of convergence. It focusses directly on sample paths from $n = 1$ to ∞ .

This makes it more difficult to talk about the rate of convergence as $n \rightarrow \infty$.

It is connected directly to the standard notion of convergence of a sequence of numbers applied to the sample paths. The power of this will be more apparent when looking at renewal processes.

Most of the heavy lifting with the SLLN has been done via the analysis of convergence WP1.

13

Proof of SLLN assuming $\bar{X} = 0$ and $E[X^4] < \infty$:

$$\begin{aligned} E[S_n^4] &= E\left(\sum_{i=1}^n X_i\right) \left(\sum_{j=1}^n X_j\right) \left(\sum_{k=1}^n X_k\right) \left(\sum_{\ell=1}^n X_{\ell}\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{\ell=1}^n E[X_i X_j X_k X_{\ell}], \end{aligned}$$

Consider any given term with $i = 1$ and $j, k, \ell > 1$.

Note that X_1 (the first rv of the n -tuple (X_1, \dots, X_n)) is independent of X_j, X_k, X_{ℓ} for each $j, k, \ell > 1$.

Thus $E[X_1 X_j X_k X_{\ell}] = 0$ for terms with $j, k, \ell > 1$.

Similarly, all terms in which any one of i, j, k, ℓ is different from the rest is 0.

There are then two kinds of nonzero terms.

14

In the first, there are n terms (one for each of the n choices for i) in which $i = j = k = \ell$.

For the second kind of nonzero term, there are $n(n-1)$ terms for which $i = j$ and $k = \ell$. There are also $n(n-1)$ terms with $i = k$ and $j = \ell$, and $n(n-1)$ more with $i = \ell$ and $j = k$.

$$E[S_n^4] = nE[X^4] + 3n(n-1)(E[X^2])^2$$

Now $E[X^4]$ is the second moment of the rv X^2 , so $(E[X^2])^2 \leq E[X^4]$. Thus

$$E[S_n^4] = [n + 3n(n-1)]E[X^4] \leq 3n^2E[X^4]$$

$$\sum_{n=1}^{\infty} \frac{E[S_n^4]}{n^4} \leq 3E[X^4] \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$$

15

From the theorem on convergence WP1,

$$\Pr\left\{\omega : \lim_{n \rightarrow \infty} S_n^4(\omega)/n^4 = 0\right\} = 1$$

For every ω such that $\lim_{n \rightarrow \infty} S_n^4(\omega)/n^4 = 0$, we see that $\lim_{n \rightarrow \infty} |S_n/n| = 0$. Thus,

$$\Pr\left\{\omega : \lim_{n \rightarrow \infty} \frac{S_n(\omega)}{n} = 0\right\} = 1 \quad \square$$

The ability to go, on a sample path basis, from $\{S_n^4(\omega)/n^4; n \geq 1\}$ to $\{|S_n(\omega)/n|; n \geq 1\}$ is the key to much of the usefulness of the strong law.

16

Example: Consider the Bernoulli process with $p_X(1) = p$. Then $E[X] = p$ and, according to the theorem, the set of sample paths $\{\omega : \lim_n S_n/n = p\}$ has probability 1.

Consider another Bernoulli process with $p_X(1) = p' \neq p$. Now $\Pr\{\omega : \lim_n S_n/n = p\} = 0$ and, with no change of events, $\Pr\{\omega : \lim_n S_n/n = p'\} = 1$.

There are uncountably many choices for p , so there are uncountably many events, each of probability 1 for its own p .

This partitions the sample space into an uncountable collection of events, each with probability 1 for its own p , plus events with no convergence.

There is nothing wrong here, but these events are peculiar and must be treated carefully.

17

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Lecture 11 : Renewals: strong law and rewards**Outline:**

- Review of convergence WP1
- Review of SLLN
- The strong law for renewal processes
- Central limit theorem (CLT) for Renewals
- Time average residual life

1

Theorem 1: let $\{Y_n; n \geq 1\}$ be rv's satisfying $\sum_{n=1}^{\infty} E[|Y_n|] < \infty$. Then $\Pr\{\omega : \lim_{n \rightarrow \infty} Y_n(\omega) = 0\} = 1$. (conv. WP1)

Note 1:

$$\lim_{m \rightarrow \infty} \sum_{n=1}^m E[|Y_n|] = \sum_{n=1}^{\infty} E[|Y_n|] < \infty$$

means that the limit exists, and thus that

$$\lim_{m \rightarrow \infty} \sum_{n=m}^{\infty} E[|Y_n|] = 0 \quad (1)$$

This is a stronger requirement than $\lim_n E[|Y_n|] = 0$.

The requirement $\lim_n E[|Y_n|] = 0$ implies that $\{Y_n; n \geq 1\}$ converges to 0 in probability (see problem set).

The stronger requirement in (1) lets us say something about entire sample paths.

2

Review of SLLN

Note 2: The usefulness of Theorem 1 depends on how the rv's Y_1, Y_2, \dots are chosen.

For SLLN (assuming $\bar{X} = 0$ and $E[X^4] < \infty$), we chose $Y_n = S_n^4/n^4$ where $S_n = X_1 + \dots + X_n$.

Since $E[S_n^4] \sim n^2$, we saw that $\sum_n E[S_n^4]/n^4 < \infty$.

Thus $\Pr\{\lim_{n \rightarrow \infty} S_n^4/n^4 = 0\} = 1$.

If $S_n^4(\omega)/n^4 = a_n$, then $|S_n(\omega)|/n = a_n^{1/4}$. If $\lim_{n \rightarrow \infty} a_n = 0$, then $\lim_{n \rightarrow \infty} a_n^{1/4} = 0$. Thus

$$\Pr\left\{\lim_{n \rightarrow \infty} S_n/n = 0\right\} = 1.$$

This is the SLLN for $\bar{X} = 0$ and $E[X^4] < \infty$.

3

This all becomes slightly more general if $\{Z_n; n \geq 1\}$ is said to converge to a constant α **WP1** if $\Pr\{\lim_{n \rightarrow \infty} Z_n(\omega) = \alpha\} = 1$.

Note that $\{Z_n; n \geq 1\}$ converges to α if and only if $\{Z_n - \alpha; n \geq 1\}$ converges to 0.

Similarly, if $\{X_n; n \geq 1\}$ is an IID sequence with $\bar{X} = \alpha$, then $\{X_n - \alpha; n \geq 1\}$ is IID with mean 0.

For renewal processes, the inter renewal intervals are positive, so it is more convenient to leave the mean in.

4

The strong law for renewal processes

A major factor in making the SLLN and convergence WP1 easy to work with was illustrated in going from S_n^4/n^4 to $|S_n|/n$ above.

The following theorem generalizes this.

Theorem 2: Assume that $\{Z_n; n \geq 1\}$ converges to α WP1 and assume that $f(x)$ is a real valued function of a real variable that is continuous at $x = \alpha$. Then $\{f(Z_n); n \geq 1\}$ converges WP1 to $f(\alpha)$.

Example 1: If $f(x) = x + \beta$ for some constant β , and Z_n converges to α , then $U_n = Z_n + \beta$ converges to $\alpha + \beta$, corresponding to a trivial change of mean.

Example 2: If $f(x) = x^{1/4}$ for $x \geq 0$, and $Z_n \geq 0$ converges to 0 WP1, then $f(Z_n)$ converges to $f(0)$.

5

Thm: $\lim_n Z_n = \alpha$ WP1 and $f(x)$ continuous at α implies $\lim_n f(Z_n) = f(\alpha)$ WP1.

Pf: For each ω such that $\lim_n Z_n(\omega) = \alpha$, we use the result for a sequence of numbers that says $\lim_n f(Z_n(\omega)) = f(\alpha)$.

For renewal processes, each inter-renewal interval X_i is positive and (assuming that $E[X]$ exists), $E[X] > 0$ (see Exercise 4.2a). Thus $E[S_n/n] > 0$.

The SLLN then applies, and

$$\Pr\left\{\omega : \lim_{n \rightarrow \infty} S_n(\omega)/n = \bar{X}\right\} = 1.$$

Using $f(x) = 1/x$,

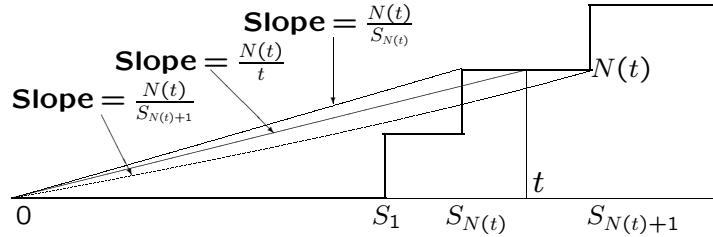
$$\Pr\left\{\omega : \lim_{n \rightarrow \infty} \frac{n}{S_n(\omega)} = \frac{1}{\bar{X}}\right\} = 1.$$

6

$$\Pr\left\{\omega : \lim_{n \rightarrow \infty} \frac{n}{S_n(\omega)} = \frac{1}{\bar{X}}\right\} = 1.$$

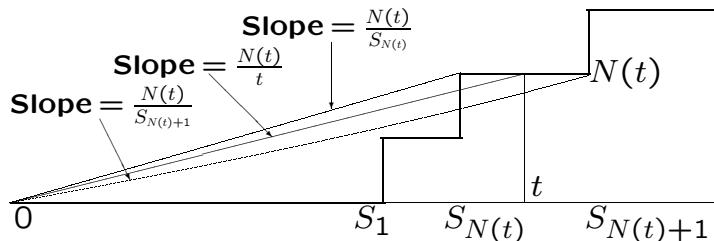
This implies the strong law for renewal processes:
Theorem 3: (strong law for renewal processes) For a renewal process with $\bar{X} < \infty$,

$$\Pr\left\{\omega : \lim_{t \rightarrow \infty} N(t)/t = 1/\bar{X}\right\} = 1.$$



Note that $N(t)/t \leq N(t)/S_{N(t)}$, and that $N(t)/S_{N(t)}$ goes through the same set of values as n/S_n

7

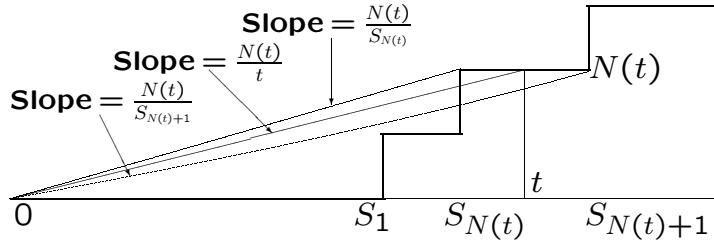


Note that $N(t)/t \leq N(t)/S_{N(t)}$, and that $N(t)/S_{N(t)}$ goes through the same set of values as n/S_n , i.e.,

$$\lim_{t \rightarrow \infty} \frac{N(t)}{S_{N(t)}} = \lim_{n \rightarrow \infty} \frac{n}{S_n} = \frac{1}{\bar{X}} \quad \text{WP1.}$$

This assumes that $\lim_{t \rightarrow \infty} N(t) = \infty$ WP1, which is demonstrated in the text.

8



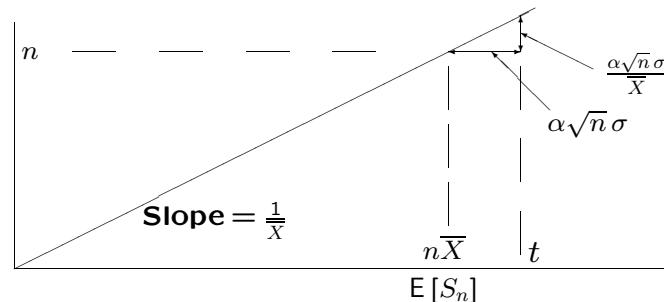
Note that $N(t)/t \geq N(t)/(S_{N(t)+1})$, and that $N(t)/(S_{N(t)+1})$ goes through the same set of values as $n/(S_{n+1})$, i.e.,

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{N(t)}{S_{N(t)+1}} &= \lim_{n \rightarrow \infty} \frac{n}{S_{n+1}} \\ &= \lim_{n \rightarrow \infty} \frac{n+1}{S_{n+1}} \frac{n}{n+1} = \frac{1}{\bar{X}} \quad \text{WP1.} \end{aligned}$$

Since $N(t)/t$ is between these two quantities with the same limit, $\lim_n N(t)/t = 1/\bar{X}$ WP1.

9

Central limit theorem (CLT) for Renewals

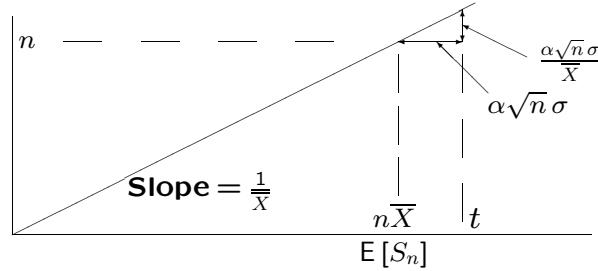


$$\Pr\{S_n \leq t\} \approx \Phi(\alpha); \quad t = n\bar{X} + \alpha\sigma\sqrt{n}$$

$$\Pr\{N(t) \geq n\} \approx \Phi(\alpha); \quad n = \frac{t}{\bar{X}} - \frac{\alpha\sigma\sqrt{n}}{\bar{X}} \approx \frac{t}{\bar{X}} - \frac{\alpha\sigma\sqrt{t}}{\bar{X}^{3/2}}$$

$$\Pr\left\{N(t) \geq \frac{t}{\bar{X}} - \frac{\alpha\sigma\sqrt{t}}{\bar{X}^{3/2}}\right\} \approx \Phi(\alpha);$$

10



$$\Pr\left\{N(t) \geq \frac{t}{\bar{X}} - \frac{\alpha\sigma\sqrt{t}}{\bar{X}^{3/2}}\right\} \approx \Phi(\alpha);$$

$$\Pr\left\{\frac{N(t) - t/\bar{X}}{\sigma\sqrt{t}\bar{X}^{-3/2}} \geq -\alpha\right\} \approx \Phi(\alpha)$$

$$\Pr\left\{\frac{N(t) - t/\bar{X}}{\sigma\sqrt{t}\bar{X}^{-3/2}} \leq -\alpha\right\} \approx 1 - \Phi(\alpha) = \Phi(-\alpha)$$

This is the CLT for renewal processes. $N(t)$ tends to Gaussian with mean t/\bar{X} and s.d. $\sigma\sqrt{t}\bar{X}^{-3/2}$.

11

Time-average Residual life

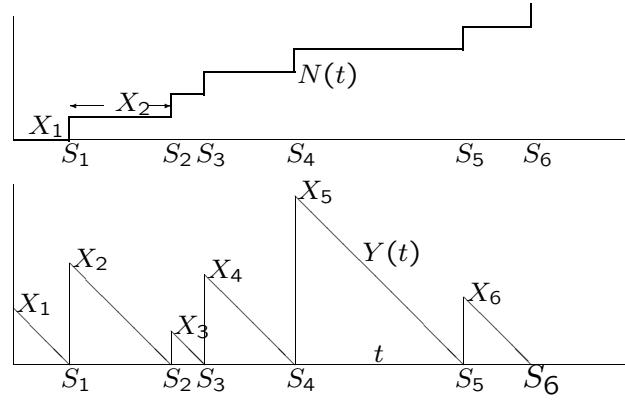
Def: The residual life $Y(t)$ of a renewal process at time t is the remaining time until the next renewal, i.e., $Y(t) = S_{N(t)+1} - t$.

It's how long you have to wait for a bus (if bus arrivals were renewal processes).

We can view residual life as a reward function on a renewal process. The sample reward at t is a function of the sample path of renewals.

The residual life, as a function of t , is a random process, and we can look at its time-average value, $\left[\int_0^t Y(\tau) d\tau\right]/t$.

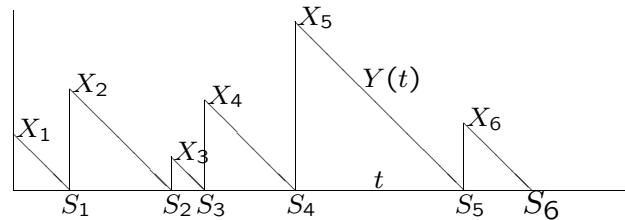
12



Note that a residual-life sample function is a sequence of isosceles triangles, one starting at each arrival epoch. The time average for a given sample function is

$$\frac{1}{t} \int_0^t y(\tau) d\tau = \frac{1}{2t} \sum_{i=1}^{n(t)} x_i^2 + \frac{1}{t} \int_{\tau=s_{n(t)}}^t y(\tau) d\tau$$

13



$$\frac{1}{2t} \sum_{n=1}^{N(t)} X_n^2 \leq \frac{1}{t} \int_0^t Y(\tau) d\tau \leq \frac{1}{2t} \sum_{n=1}^{N(t)+1} X_n^2$$

$$\lim_{t \rightarrow \infty} \frac{\sum_{n=1}^{N(t)} X_n^2}{2t} = \lim_{t \rightarrow \infty} \frac{\sum_{n=1}^{N(t)} X_n^2}{N(t)} \frac{N(t)}{2t} = \frac{\mathbb{E}[X^2]}{2\mathbb{E}[X]} \quad \text{WP1}$$

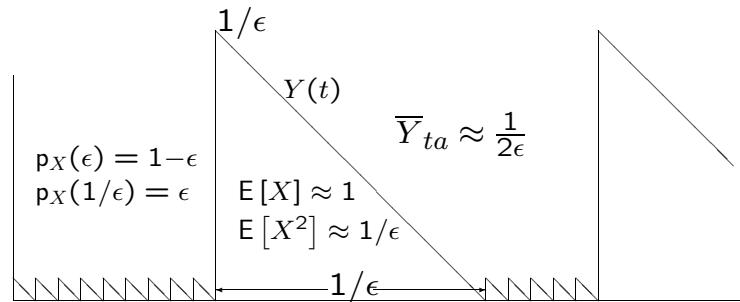
$$\lim_{t \rightarrow \infty} \frac{\int_0^t Y(\tau) d\tau}{t} = \frac{\mathbb{E}[X^2]}{2\mathbb{E}[X]} \quad \text{W.P.1}$$

14

Time average residual life $\bar{Y}_{ta} = \frac{\mathbb{E}[X^2]}{2\mathbb{E}[X]}$.

If X is almost deterministic, $\bar{Y}_{ta} \approx \mathbb{E}[X]/2$.

If X exponential, $\bar{Y}_{ta} = \mathbb{E}[X]$.



The expected duration between long intervals is ≈ 1 .

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Lecture 12 : Renewal rewards, stopping trials, and Wald's equality

Outline:

- Review strong law for renewals
- Review of residual life
- Time-averages for renewal rewards
- Stopping trials for stochastic processes
- Wald's equality
- Stop when you're ahead

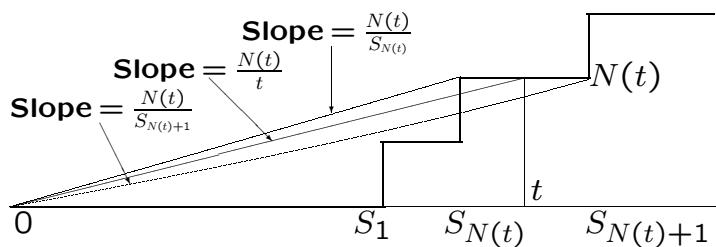
1

Theorem: If $\{Z_n; n \geq 1\}$ converges to α WP1, (i.e., $\Pr\{\omega : \lim_n (Z_n(\omega) - \alpha) = 0\} = 1$), and $f(x)$ is continuous at α . Then $\Pr\{\omega : \lim_n f(Z_n(\omega)) = \alpha\} = 1$.

For a renewal process with inter-renewals X_i , $0 < \bar{X} < \infty$, $\Pr\{\omega : \lim_n (\frac{1}{n} S_n(\omega) - \bar{X}) = 0\} = 1$)

$$\Pr\left\{\omega : \lim_{n \rightarrow \infty} \frac{n}{S_n(\omega)} = \frac{1}{\bar{X}}\right\} = 1.$$

For renewal processes, n/S_n and $N(t)/t$ are related by



2

The strong law for renewal processes follows from this relation between n/S_n and $N(t)/t$.

Theorem: For a renewal process with $\bar{X} < \infty$,

$$\Pr\left\{\omega : \lim_{t \rightarrow \infty} N(t, \omega)/t = 1/\bar{X}\right\} = 1.$$

This says that the rate of renewals over the infinite time horizon (i.e., $\lim_t N(t)/t$) is $1/\bar{X}$ WP1.

This also implies the weak law for renewals,

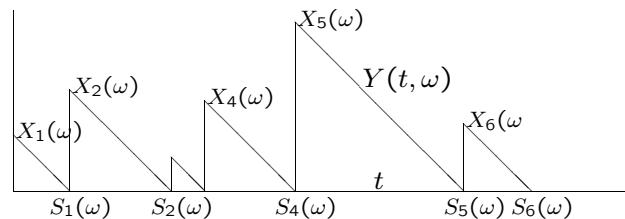
$$\lim_{t \rightarrow \infty} \Pr\left\{\left|\frac{N(t)}{t} - \frac{1}{\bar{X}}\right| > \epsilon\right\} = 0 \quad \text{for all } \epsilon > 0$$

3

Review of residual life

Def: The residual life $Y(t)$ of a renewal process at time t is the remaining time until the next renewal, i.e., $Y(t) = S_{N(t)+1} - t$.

Residual life is a random process; for each sample point ω , $Y(t, \omega)$ is a sample function.



$$\sum_{n=1}^{N(t,\omega)} \frac{X_i^2(\omega)}{2t} \leq \frac{1}{t} \int_0^t Y(t, \omega) dt \leq \sum_{n=1}^{N(t,\omega)+1} \frac{X_i^2(\omega)}{2t}$$

4

$$\sum_{i=1}^{N(t,\omega)} \frac{X_i^2(\omega)}{2t} \leq \frac{1}{t} \int_0^t Y(t, \omega) dt \leq \sum_{i=1}^{N(t,\omega)+1} \frac{X_i^2(\omega)}{2t}$$

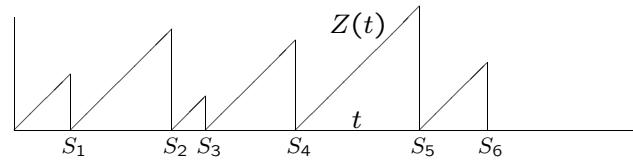
Going to the limit $t \rightarrow \infty$

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t Y(t, \omega) dt &= \lim_{t \rightarrow \infty} \sum_{n=1}^{N(t,\omega)} \frac{X_n^2(\omega)}{2N(t,\omega)} \frac{N(t,\omega)}{t} \\ &= \frac{\mathbb{E}[X^2]}{2\mathbb{E}[X]} \end{aligned}$$

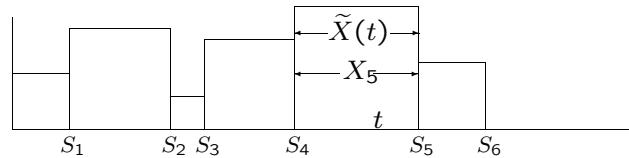
This is infinite if $\mathbb{E}[X^2] = \infty$. **Think of example where** $p_X(\epsilon) = 1 - \epsilon$, $p_X(1/\epsilon) = \epsilon$.

5

Similar examples: **Age** $Z(t) = t - S_{N(t)}$ **and duration,** $\tilde{X}(t) = S_{N(t)+1} - S_{N(t)}$.



$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \tilde{X}(\tau) d\tau = \frac{\mathbb{E}[X^2]}{2\mathbb{E}[X]} \quad \text{WP1.}$$



$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \tilde{X}(\tau) d\tau = \frac{\mathbb{E}[X^2]}{\mathbb{E}[X]} \quad \text{WP1.}$$

6

Time-averages for renewal rewards

Residual life, age, and duration are examples of assigning rewards to renewal processes.

The reward $R(t)$ at any time t is restricted to be a function of the inter-renewal period containing t .

In simplest form, $R(t)$ is restricted to be a function $\mathcal{R}(Z(t), \tilde{X}(t))$.

The time-average for a sample path of $R(t)$ is found by analogy to residual life. Start with the n th inter-renewal interval.

$$R_n(\omega) = \int_{S_{n-1}(\omega)}^{S_n(\omega)} R(t, \omega) dt$$

Interval 1 goes from 0 to S_1 , with $Z(t) = t$. For interval n , $Z(t) = t - S_{n-1}$, i.e., $S_{N(t)} = S_{n-1}$.

7

$$\begin{aligned} R_n &= \int_{S_{n-1}}^{S_n} R(t) dt \\ &= \int_{S_{n-1}}^{S_n} \mathcal{R}(Z(t), \tilde{X}(t)) dt \\ &= \int_{S_{n-1}}^{S_n} \mathcal{R}(t - S_{n-1}, X_n) dt \\ &= \int_0^{X_n} \mathcal{R}(z, X_n) dz \end{aligned}$$

This is a function only of the rv X_n . Thus

$$\mathbb{E}[R_n] = \int_{x=0}^{\infty} \int_{z=0}^x \mathcal{R}(z, x) dz dF_X(x).$$

Assuming that this expectation exists,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_{\tau=0}^t R(\tau) d\tau = \frac{\mathbb{E}[R_n]}{\bar{X}} \quad \text{WP1}$$

8

Example: Suppose we want to find the k th moment of the age.

Then $\mathcal{R}(Z(t), \widetilde{X}(t)) = Z^k(t)$. Thus

$$\begin{aligned}\mathbb{E}[R_n] &= \int_{x=0}^{\infty} \int_{z=0}^x z^k dz d\mathbb{F}_X(x) \\ &= \int_0^{\infty} \frac{x^{k+1}}{k+1} d\mathbb{F}_X(x) = \frac{1}{k} \mathbb{E}[X^{k+1}] \\ \lim_{t \rightarrow \infty} \frac{1}{t} \int_{\tau=0}^t R(\tau) d\tau &= \frac{\mathbb{E}[X^{k+1}]}{(k+1)\bar{X}} \quad \text{WP1}\end{aligned}$$

9

Stopping trials for stochastic processes

It is often important to analyze the initial segment of a stochastic process, but rather than investigating the interval $(0, t]$ for a fixed t , we want to investigate $(0, t]$ where t is selected by the sample path up until t .

It is somewhat tricky to formalize this, since t becomes a rv which is a function of $\{X(\tau); \tau \leq t\}$. This approach seems circular, so we have to be careful.

We consider only discrete-time processes $\{X_i; i \geq 1\}$.

10

Let J be a positive integer rv that describes when a sequence X_1, X_2, \dots , is to be stopped.

At trial 1, $X_1(\omega)$ is observed and a decision is made, based on $X_1(\omega)$, whether or not to stop. If we stop, $J(\omega) = 1$

At trial 2 (if $J(\omega) \neq 1$), $X_2(\omega)$ is observed and a decision is made, based on $X_1(\omega), X_2(\omega)$, whether or not to stop. If we stop, $J(\omega) = 2$.

At trial 3 (if $J(\omega) \neq 1, 2$), $X_3(\omega)$ is observed and a decision is made, based on $X_1(\omega), X_2(\omega), X_3(\omega)$, whether or not to stop. If we stop, $J(\omega) = 3$, etc.

At each trial n (if stopping has not yet occurred), X_n is observed and a decision (based on $X_1 \dots, X_n$) is made; if we stop, then $J(\omega) = n$.

11

Def: A stopping trial (or stopping time) J for $\{X_n; n \geq 1\}$, 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 defective.

We visualize ‘conducting’ successive trials X_1, X_2, \dots , until some n at which the event $\{J = n\}$ occurs; further trials then cease. It is simpler conceptually to visualize stopping the observation of trials after the stopping trial, but continuing to conduct trials.

Since J is a (possibly defective) rv, the events $\{J = 1\}, \{J = 2\}, \dots$ are disjoint.

12

Example 1: A gambler goes to a casino and gambles until broke.

Example 2: Flip a coin until 10 successive heads appear.

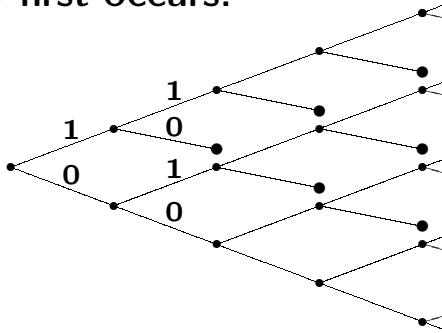
Example 3: Test an hypothesis with repeated trials until one or the other hypothesis is sufficiently probable a posteriori.

Example 4: Observe successive renewals in a renewal process until $S_n \geq 100$.

13

Suppose the rv's X_i in a process $\{X_n; n \geq 1\}$ have a finite number of possible sample values. Then any (possibly defective) stopping trial J can be represented as a rooted tree where the trial at which each sample path stops is represented by a terminal node.

Example: X is binary and stopping occurs when the pattern $(1, 0)$ first occurs.



14

Wald's equality

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

$$E[S_J] = \bar{X}E[J]$$

Prf:

$$S_J = X_1 \mathbb{I}_{J \geq 1} + X_2 \mathbb{I}_{J \geq 2} + \dots + X_n \mathbb{I}_{J \geq n} + \dots$$

$$E[S_J] = E\left[\sum_n X_n \mathbb{I}_{J \geq n}\right] = \sum_n E[X_n \mathbb{I}_{J \geq n}]$$

The essence of the proof is to show that X_n and $\mathbb{I}_{J \geq n}$ are independent.

15

To show that X_n and $\mathbb{I}_{J \geq n}$ are independent, note that $\mathbb{I}_{J \geq n} = 1 - \mathbb{I}_{J < n}$. Also $\mathbb{I}_{J < n}$ is a function of X_1, \dots, X_{n-1} . Since the X_i are IID, X_n is independent of X_1, \dots, X_{n-1} , and thus $\mathbb{I}_{J < n}$, and thus of $\mathbb{I}_{J \geq n}$.

This is surprising, since X_n is certainly not independent of $\mathbb{I}_{J=n}$, nor of $\mathbb{I}_{J=n+1}$, etc.

The resolution of this ‘paradox’ is that, given that $J \geq n$ (i.e., that stopping has not occurred before trial n), the trial at which stopping occurs depends on X_n , but whether or not $J \geq n$ occurs depends only on X_1, \dots, X_{n-1} .

Now we can finish the proof.

16

$$\begin{aligned}
E[S_J] &= \sum_n E[X_n \mathbb{I}_{J \geq n}] \\
&= \sum_n E[X_n] E[\mathbb{I}_{J \geq n}] \\
&= \bar{X} \sum_n E[\mathbb{I}_{J \geq n}] \\
&= \bar{X} \sum_n \Pr\{J \geq n\} = \bar{X} E[J]
\end{aligned}$$

In many applications, this gives us one equation in two quantities neither of which is known. Frequently, $E[S_J]$ is easy to find and this solves for $E[J]$.

The following example shows, among other things, why $E[J] < \infty$ is required for Wald's equality.

17

Stop when you're ahead

Consider tossing a coin with probability of heads equal to p . \$1 is bet on each toss and you win on heads, lose on tails. You stop when your winnings reach \$1.

If $p > 1/2$, your winnings (in the absence of stopping) would grow without bound, passing through 1, so J must be a rv. $S_J = 1$ WP1, so $E[S_J] = 1$. Thus, Wald says that $E[J] = 1/\bar{X} = \frac{1}{2p-1}$. Let's verify this in another way.

Note that $J = 1$ with probability p . If $J > 1$, i.e., if $S_1 = -1$, then the only way to reach $S_n = 1$ is to go from $S_1 = -1$ to $S_m = 0$ for some m (requiring \bar{J} steps on average); \bar{J} more steps on average then gets to 1. Thus $\bar{J} = 1 + (1-p)2\bar{J} = \frac{1}{2p-1}$.

18

Next consider $p < 1/2$. It is still possible to win and stop (for example, $J = 1$ with probability p and $J = 3$ with probability $p^2(1-p)$). It is also possible to head South forever.

Let $\theta = \Pr\{J < \infty\}$. Note that $\Pr\{J = 1\} = p$. Given that $J > 1$, i.e., that $S_1 = -1$, the event $\{J < \infty\}$ requires that $S_m - S_1 = 1$ for some m , and then $S_n - S_m = 1$ for some $n > m$. Each of these are independent events of probability θ , so

$$\theta = p + (1-p)\theta^2$$

There are two solutions, $\theta = p/(1-p)$ and $\theta = 1$, which is impossible. Thus J is defective and Wald's equation is inapplicable.

19

Finally consider $p = 1/2$. In the limit as p approaches $1/2$ from below, $\Pr\{J < \infty\} = 1$. We find other more convincing ways to see this later. However, as p approaches $1/2$ from above, we see that $E[J] = \infty$.

Wald's equality does not hold here, since $E[J] = \infty$, and in fact does not make sense since $\bar{X} = 0$.

However, you make your \$1 with probability 1 in a fair game and can continue to repeat the same feat.

It takes an infinite time, however, and requires access to an infinite capital.

20

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Lecture 13 : Little, M/G/1, ensemble averages**Outline:**

- Review Wald's equality
- The elementary renewal theorem
- Generalized stopping trials
- The G/G/1 queue
- Little's theorem
- Pollaczek-Khinchin result for M/G/1

1

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 \bar{X} . If J is a stopping trial for $\{X_n; n \geq 1\}$ and if $E[J] < \infty$, then the sum $S_J = X_1 + X_2 + \dots + X_J$ at the stopping trial J satisfies

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

2

The elementary renewal theorem

Wald's equality is useful for determining $E[N(t)]$ as a function of t for a renewal counting process. We have the strong and weak laws for $N(t)$ as $t \rightarrow \infty$, but often it is useful to be explicit for finite t .

For a given t , let J be the smallest n for which $S_n > t$. Then J is a stopping trial for the inter-arrivals $\{X_i; i \geq 1\}$. That is, we stop at trial n if $S_n > t$ and $S_{n-1} \leq t$, and this is determined by X_1, \dots, X_n .

Note $N(t)$ is the number of arrivals that have occurred up to and including t , so $N(t) + 1 = J$ is the number of the first arrival after J . Since $E[N(t)]$ is finite, $E[J] < \infty$. From Wald,

$$E[S_{N(t)+1}] = \bar{X}E[J] = \bar{X}(E[N(t)] + 1)$$

3

$$E[S_{N(t)+1}] = \bar{X}E[J] = \bar{X}(E[N(t)] + 1)$$

Wald's equality relates two unknown quantities, $E[N(t)]$ and $E[S_{N(t)+1}]$. Since $E[S_{N(t)+1}] > t$, we get a simple bound from this.

$$E[N(t)] = \frac{E[S_{N(t)+1}]}{\bar{X}} - 1 > \frac{t}{\bar{X}} - 1$$

Thm (Elementary renewal thm): Let \bar{X} be mean inter-renewal of a renewal counting process $\{N(t); t > 0\}$. Then $\lim_{t \rightarrow \infty} E[N(t)/t] = 1/\bar{X}$.

Pf: Need an upper bound on $E[N(t)]$. Truncate X to $X \leq b$, carry out bound, let b grow with t .

We can view this as convergence in the mean (one more type of convergence).

4

Generalized stopping trials

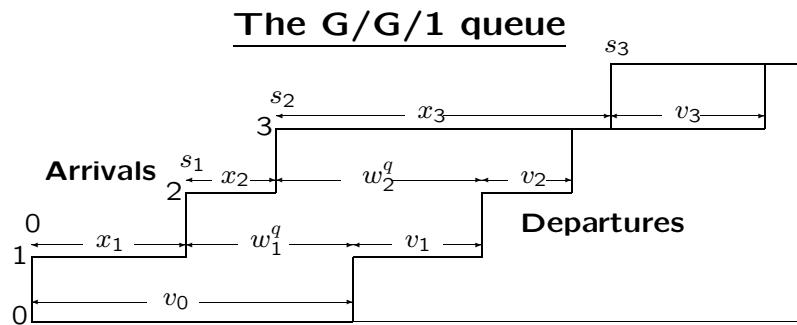
Def: A generalized stopping trial J for a sequence of pairs of rv's $(X_1, V_1), (X_2, V_2) \dots$, is a positive integer rv such that, for each $n \geq 1$, $\mathbb{I}_{\{J=n\}}$ is a function of $X_1, V_1, X_2, V_2, \dots, X_n, V_n$.

It follows that $\mathbb{I}_{\{J < n\}} = 1 - \mathbb{I}_{\{J \geq n\}}$ is a function of $X_1, V_1, X_2, V_2, \dots, X_{n-1}, V_{n-1}$.

Wald's equality, $E[S_n] = \bar{X}E[J]$, where $S_n = X_1 + \dots + X_n$ still holds (by the same proof) if the X_i are IID and each X_i is independent of $(X_1, V_1, \dots, X_{i-1}, V_{i-1})$.

Also each V_i can be replaced by a vector of rv's.

5



Consider the first arrival (s_3 above) that starts a new busy period as a generalized stopping trial.

The sequence of paired rv's is $(X_1, V_0), (X_2, V_1), \dots$
 'Stopping' at $J = 3$ is $f(X_1, V_0, X_2, V_1, X_3, V_2)$.

Wald's equality holds. Also new arrivals $(X_{J+1}, X_{J+2}, \dots)$ and services (V_J, V_{J+1}, \dots) are independent of the old.

6

The stopping rule J here is the index of the first arrival in a new busy period. The arrivals and departures in the new busy period are independent and identically distributed to those in the old.

Thus the intervals between new busy periods form a renewal process.

We then have one renewal process embedded in another. Call one the arrival process and the other the renewal process. The renewal process embodies both arrivals and services.

This analysis applies also to G/G/m and to many other queuing systems.

7

Little's theorem

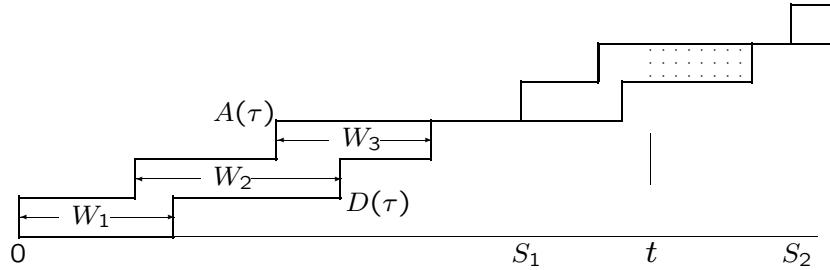
Consider a queueing system where the arrival process is a renewal process. Assume an arrival at time 0.

The service process can be almost anything, but assume a G/G/1 queue to be specific.

Assume the system empties out eventually WP1 and that it restarts on the next arrival.

We have seen that intervals between restartings form a renewal process for the G/G/1 queue, and for an even broader class of queues.

8



Let $L(\tau) = A(\tau) - D(\tau)$. This depends on the departure process, but its value is a function of the inter-arrivals and service times within the current inter-renewal period.

Thus it can be viewed as a generalized renewal-reward function.

The total reward within an inter-renewal period is then the integral of $L(\tau)$ over that period (i.e., R_n).

9

In each inter-renewal period,

$$R_n = \int L(\tau) d\tau = \sum_i W_i,$$

where the sum is over the arrivals in that inter-renewal period. The time averages are then

$$\begin{aligned} L_{ta} &= \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^{A(t)} W_i \\ &= \lim_{t \rightarrow \infty} \frac{\sum_{i=1}^{A(t)} W_i}{A(t)} \lim_{t \rightarrow \infty} \frac{A(t)}{t} \\ &= W_{ta} \lambda \end{aligned}$$

where λ is arrival rate.

This is Little's theorem. The time-average number in the system equals λ times the time-average customer wait WP1.

In essence, Little's theorem is an accounting identity. Over a busy period and the following idle period (i.e., an inter-renewal period), $\int L(\tau) d\tau = \sum_i W_i$.

To turn this simple result into mathematics, we need renewal theory, which essentially allows us to go to the limit of many renewals.

The question is not whether $L = \lambda W$, but whether these quantities exist as sensible time averages WP1 or as limiting ensemble averages.

11

The Pollaczek-Khinchin formula for M/G/1 queues

Let X_1, X_2, \dots be IID exponential arrivals at rate λ and let V_1, V_2, \dots be IID service times with first and second moments $\bar{V} = E[V]$ and $\bar{V^2} = E[V^2]$. Then the Pollaczek-Khinchin formula gives the expected delay in queue (between arrival and entering service) as

$$\bar{W}^q = \frac{\lambda \bar{V^2}}{2(1 - \rho)} \quad \text{where } \rho = \lambda \bar{V}$$

The expected total delay, total number in system and number in queue are then

$$\bar{W} = \frac{\lambda \bar{V^2}}{2(1 - \rho)} + \bar{V}; \quad \bar{N} = \frac{\lambda^2 \bar{V^2}}{2(1 - \rho)} + \rho; \quad \bar{N}^q = \frac{\lambda^2 \bar{V^2}}{2(1 - \rho)}$$

12

Examples: For deterministic service, $\overline{V^2} = (\overline{V})^2$, so

$$\overline{W}^q = \frac{\rho \overline{V}}{2(1 - \rho)} \quad \text{for M/D/1}$$

For exponential inter-arrivals M/M/1, $\overline{V^2} = 2(\overline{V})^2$ so

$$\overline{W}^q = \frac{\rho \overline{V}}{1 - \rho} \quad \text{for M/M/1.}$$

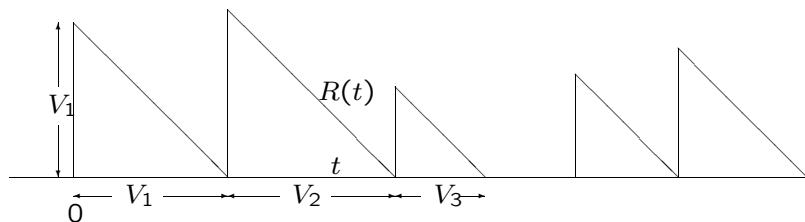
For $p_V(\epsilon) = 1 - \epsilon$, $p_V(1/\epsilon) = \epsilon$,

$$\overline{W}^q \approx \frac{\rho}{\epsilon(1 - \rho)}$$

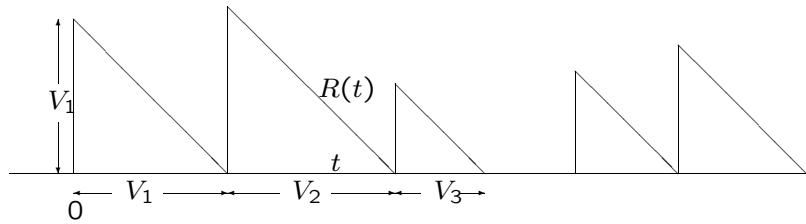
13

Why does \overline{W}^q go up with \overline{V}^2 ? Look at the time-average wait, $E[R(t)]$, for the customer in service to finish service.

Have you ever noticed, when entering a line for service that the customer being served often takes much longer than anyone else?



14



$$\begin{aligned}
 E[R] &= \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau R(t) dt = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \sum_{i=1}^{A(\tau)} \frac{V_i^2}{2} \\
 &= \frac{\lambda \bar{V}^2}{2} \quad (\text{time until end of current service}) \\
 \bar{W}^q &= E[R] + \bar{N}^q \bar{V} \quad (\text{add service of queue}) \\
 &= \frac{\lambda \bar{V}^2}{2(1 - \lambda \bar{V})} \quad (\text{a little Little})
 \end{aligned}$$

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

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.

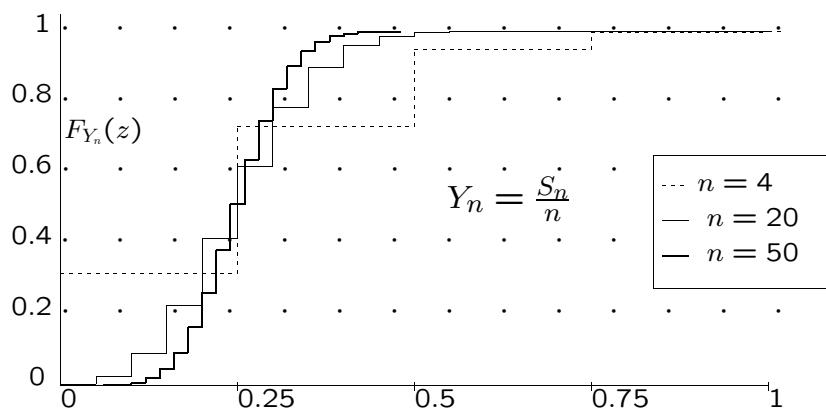
A sample point is then a collection of sample values, one for each rv.

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.

1

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

The sample (time) average is S_n/n , $S_n = X_1 + \dots + X_n$. It is a rv of mean \bar{X} and variance σ^2/n .



2

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

$$\lim_{n \rightarrow \infty} \Pr\left\{\left|\frac{S_n}{n} - \bar{X}\right| \geq \epsilon\right\} = 0 \quad \text{for every } \epsilon > 0.$$

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

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

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \bar{X} \quad \text{W.P.1}$$

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 $\lim_{n \rightarrow \infty} f(S_n/n) = f(\bar{X})$ **W.P.1**.

3

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

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

Equivalently,

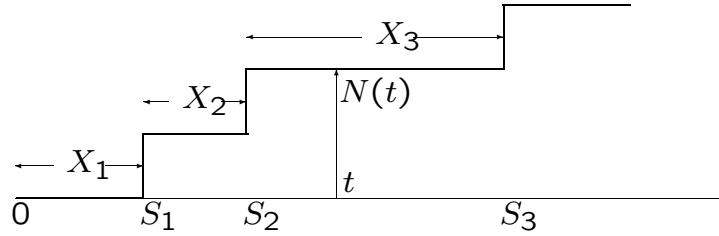
$$\lim_{n \rightarrow \infty} \left[\Pr\left\{ \frac{S_n}{n} - \bar{X} \leq \frac{y\sigma}{\sqrt{n}} \right\} \right] = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx.$$

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

4

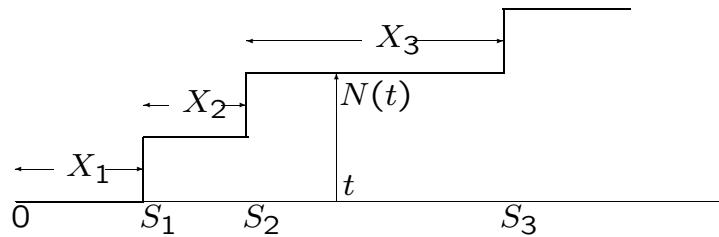
Arrival processes

Def: An arrival process is an increasing sequence of rv's, $0 < S_1 < S_2 < \dots$. 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.

5



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

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 $\Pr\{X > t+x | X > t\} = \Pr\{X > x\}$. 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.

7

Thm: Given a Poisson process of rate λ , 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, \dots . Also $\{Z_1, Z_2, \dots\}$ 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.

8

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

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

The probability distributions

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

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

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

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

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

Combining and splitting

If $N_1(t), N_2(t), \dots, N_k(t)$ are independent PP's of rates $\lambda_1, \dots, \lambda_k$, **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.

11

Conditional arrivals and order statistics

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

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

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

The joint distribution of S_1, \dots, S_n given $N(t) = n$ is the same as the joint distribution of n uniform rv's that have been ordered.

12

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 ,

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

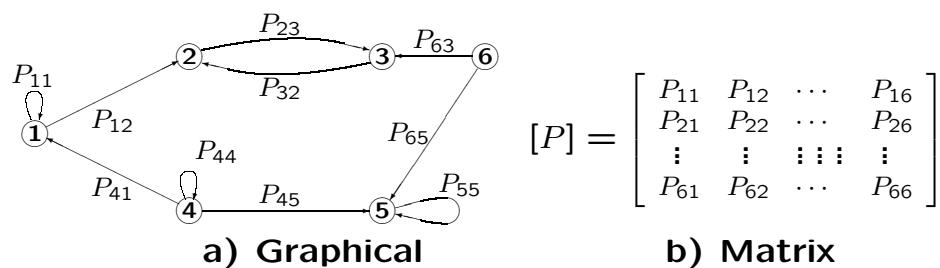
where P_{ij} depends only on i, j and $p_{X_0}(m)$ is arbitrary. A Markov chain is finite-state if the sample space for each X_i is a finite set, \mathcal{S} . The sample space \mathcal{S} usually taken to be the integers $1, 2, \dots, M$.

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.

13

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

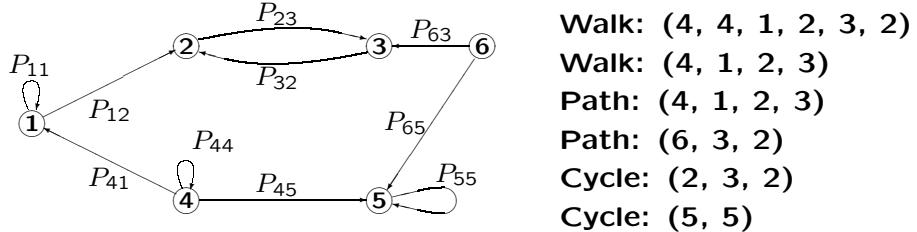
14

Classification of states

An (n -step) walk is an ordered string of nodes (states), say (i_0, i_1, \dots, 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 , $(i \rightarrow j)$ if there is a walk from i to j , i.e., if $P_{ij}^n > 0$ for some $n > 0$.

15

If $(i \rightarrow j)$ and $(j \rightarrow k)$ then $(i \rightarrow k)$.

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

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 $\{i\} \cup \{j : (i \leftrightarrow j)\}$.

For finite-state chains, a state i is transient if there is a $j \in \mathcal{S}$ such that $i \rightarrow j$ but $j \not\rightarrow i$. 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.

16

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 S_1, S_2, \dots, S_d . Transitions from each class go only to states in the next class (viewing S_1 as the next subclass to S_d).

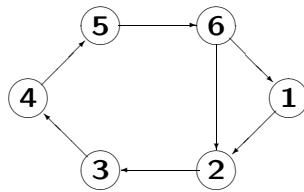
An ergodic class is a recurrent aperiodic class. A Markov chain with only one class is ergodic if that class is ergodic.

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 . $\{\pi_i ; 1 \leq M\}$ satisfies $\sum_i \pi_i P_{ij} = \pi_j > 0$ with $\sum_i \pi_i = 1$.

17

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 $n \geq (M - 1)^2 + 1$.

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 $M - 1$.



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 .

18

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 $n \geq 1$:

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

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

19

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 . $\{\pi_i; 1 \leq \mathbf{M}\}$ 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 $\lim_n [P^n] = \vec{e}\pi$ where $\vec{e} = (1, 1, \dots, 1)^\top$ and π satisfies $\pi[P] = \pi$ and $\pi\vec{e} = 1$.

20

We get more specific results by looking at the eigenvalues and eigenvectors of an arbitrary stochastic matrix (matrix of a Markov chain).

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

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

$\det[P - \lambda I]$ is an M th degree polynomial in λ . 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.

21

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

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

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

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

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

22

Facts: All eigenvalues λ 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 ν 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_{ij}^n converges to π_j as λ_2^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.

23

Renewal processes

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

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

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

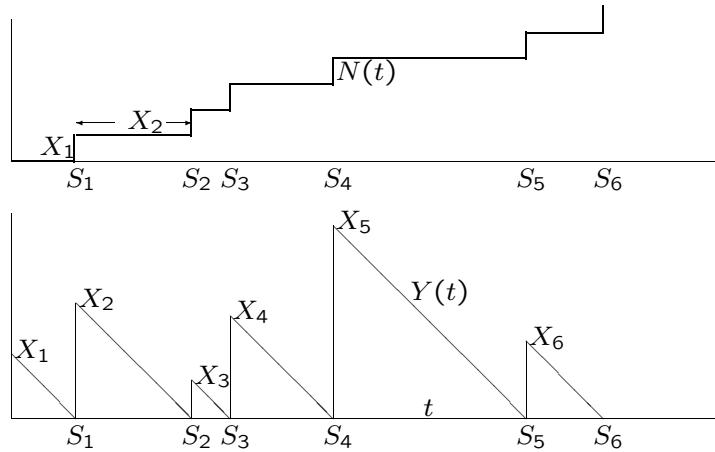
In both cases, $\lim_{t \rightarrow \infty} N(t) = \infty$ with probability 1.

There is also the elementary renewal theorem, which says that

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

24

Residual life



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

25

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

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

The time average duration is

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

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

26

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

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

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

This also works for ensemble averages.

27

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 \bar{X} . If J is a stopping trial for $\{X_n; n \geq 1\}$ and if $\mathbb{E}[J] < \infty$, then the sum $S_J = X_1 + X_2 + \dots + X_J$ at the stopping trial J satisfies

$$\mathbb{E}[S_J] = \bar{X}\mathbb{E}[J].$$

28

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

$$E[S_J] = \bar{X}E[J]$$

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

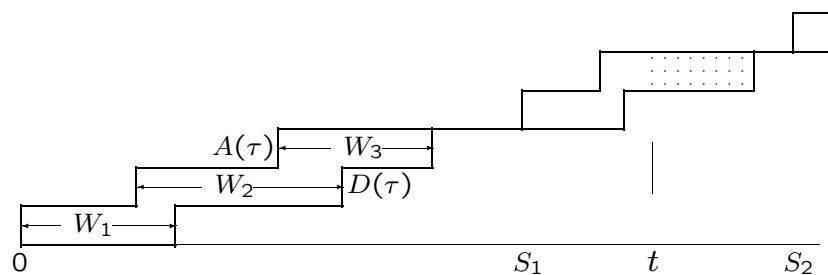
For cases where X is positive or negative, it is necessary as shown by 'stop when you're ahead.'

29

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.



30

Let \bar{L} be the time average number in system,

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

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

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

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

6.262: Discrete Stochastic Processes 3/30/11

Reminder: Quiz, 4/4/11, 7 - 9:30pm, Room 32-141

Sections of notes not covered: 1.5.3-4, 2.4, 3.5.3, 3.6, 4.6-8

For text with most errors corrected, see

<http://www.rle.mit.edu/rgallager/notes.htm>

Lecture 15: The last(?) renewal

Outline:

- Review sample-path averages and Wald
- Little's theorem
- Markov chains and renewal processes
- Expected number of renewals, $m(t) = E[N(t)]$
- Elementary renewal and Blackwell thms
- Delayed renewal processes

1

One of the main reasons why the concept of convergence WP1 is so important is the following:

Thm: Assume that $\{Z_n; n \geq 1\}$ converges to α WP1 and assume that $f(x)$ is a real valued function of a real variable that is continuous at $x = \alpha$. Then $\{f(Z_n); n \geq 1\}$ converges WP1 to $f(\alpha)$.

For a renewal process with interarrivals $\{X_n; n \geq 1\}$ where $E[X] < \infty$, the arrival epochs satisfy $S_n/n \rightarrow E[X]$ WP1 and thus $n/S_n \rightarrow 1/\bar{X}$ WP1. The strong law for renewals follows.

Thm: $\Pr\left\{\lim_{t \rightarrow \infty} \frac{N(t)}{t} = \frac{1}{\bar{X}}\right\} = 1$.

2

The strong law for renewals also holds if $\bar{X} = \infty$. In this case, since X is a rv and S_n is a rv for all n , $N(t)$ grows without bound as $t \rightarrow \infty$, but $N(t)/t \rightarrow 0$.

Since $N(t)/t$ converges WP1 to $1/\bar{X}$, it also must converge in probability, i.e.,

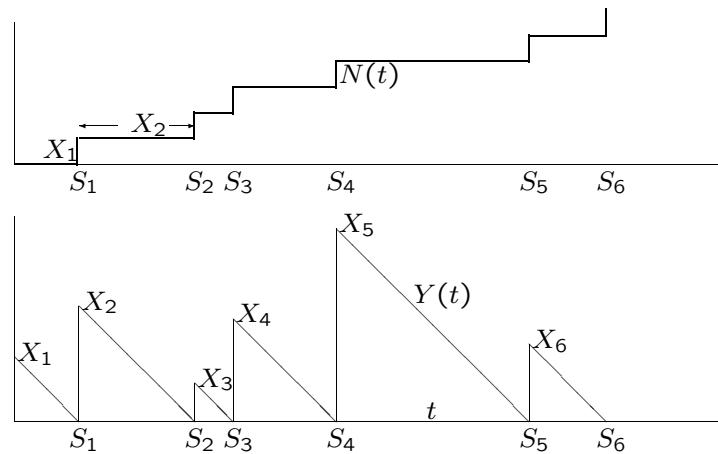
$$\lim_{t \rightarrow \infty} \Pr \left\{ \left| \frac{N(t)}{t} - \frac{1}{\bar{X}} \right| > \epsilon \right\} = 0 \quad \text{for all } \epsilon > 0$$

This is similar to the elementary renewal theorem, which says that

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

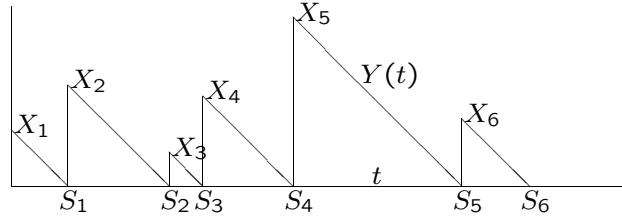
3

Residual life



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

4



$$\frac{1}{2t} \sum_{n=1}^{N(t)} X_n^2 \leq \frac{1}{t} \int_0^t Y(\tau) d\tau \leq \frac{1}{2t} \sum_{n=1}^{N(t)+1} X_n^2$$

$$\lim_{t \rightarrow \infty} \frac{\sum_{n=1}^{N(t)} X_n^2}{2t} = \lim_{t \rightarrow \infty} \frac{\sum_{n=1}^{N(t)} X_n^2}{N(t)} \frac{N(t)}{2t} = \frac{\mathbb{E}[X^2]}{2\mathbb{E}[X]} \quad \text{WP1}$$

Why is this true? It is an abbreviation for a sample-path result.

5

For the sample point ω , if the limits exist, we have

$$\lim_{t \rightarrow \infty} \frac{\sum_{n=1}^{N(t,\omega)} X_n^2(\omega)}{2t} = \lim_{t \rightarrow \infty} \frac{\sum_{n=1}^{N(t,\omega)} X_n^2(\omega)}{N(t,\omega)} \frac{N(t,\omega)}{2t}$$

For the given ω and a given t , the RHS above is the product of 2 numbers, and as t increases, we are looking at the limit of a product of numerical functions of t .

For those ω in a set of probability 1, both those functions converge to finite values as $t \rightarrow \infty$. Thus the limit of the product is the product of the limits.

This is a good example of why the strong law, dealing with sample paths, is so powerful.

6

Residual life and duration are examples of renewal reward functions.

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

Thus reward over a renewal interval is

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

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

This also works for ensemble averages.

7

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 \bar{X} . If J is a stopping trial for $\{X_n; n \geq 1\}$ and if $E[J] < \infty$, then the sum $S_J = X_1 + X_2 + \dots + X_J$ at the stopping trial J satisfies

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

8

Wald: Let $\{X_n; n \geq 1\}$ be IID rv's, each of mean \bar{X} .

If J is a stopping time for $\{X_n; n \geq 1\}$, $E[J] < \infty$, and $S_J = X_1 + X_2 + \dots + X_J$, then

$$E[S_J] = \bar{X}E[J]$$

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

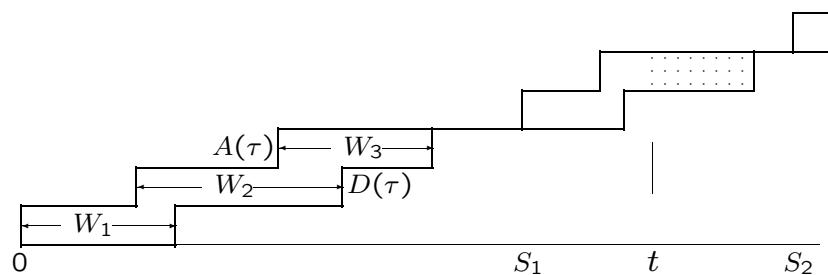
For cases where X is positive or negative, it is necessary as shown by 'stop when you're ahead.'

9

Little's theorem

This is an accounting trick plus some intricate handling of limits. 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 $L_n = \sum W_i$ over each busy period.



10

Let \bar{L} be the time average number in system,

$$\bar{L} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t L(\tau) d\tau = \lim_{t \rightarrow \infty} \frac{\sum_{i=0}^{N(t)} W_i}{t}$$

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

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

This is the same use of sample path limits as before.

Markov chains and renewal processes

For any finite-state ergodic Markov chain $\{X_n; n \geq 0\}$ with $X_0 = i$, there is a renewal counting process $\{N_i(t); t \geq 1\}$ where $N_i(t)$ is the number of visits to state i from time 1 to t . Let Y_1, Y_2, \dots be the inter-renewal periods. By the elementary renewal thm,

$$\lim_{t \rightarrow \infty} \frac{E[N_i(t)]}{t} = \frac{1}{\bar{Y}}$$

$$P_{ii}^t = \Pr\{N_i(t) - N_i(t-1) = 1\} = E[N_i(t) - N_i(t-1)]$$

$$\sum_{n=1}^t P_{ii}^n = E[N_i(t)]$$

But since $P_{ii}^t \rightarrow \pi_i$ exponentially,

$$\pi_i = \lim_{t \rightarrow \infty} \frac{\sum_{n=1}^t P_{ii}^t}{t} = \frac{E[N_i(t)]}{t} = \frac{1}{\bar{Y}}$$

Thus the mean recurrence time of state i is $1/\pi_i$.

Expected number of renewals, $m(t) = E[N(t)]$

The elementary renewal theorem says

$$\lim_{t \rightarrow \infty} E[N(t)] / t = 1/\bar{X}$$

For finite t , $m(t)$ can be very messy. Suppose the interarrival interval X is 1 or $\sqrt{2}$. As t increases, the points at which t can increase get increasingly dense, and $m(t)$ is non-decreasing but otherwise ugly.

Some progress can be made by expressing $m(t)$ in terms of its values at smaller t by the ‘renewal equation.’

$$\begin{aligned} m(t) &= F_X(t) + \int_0^t m(t-x) dF_X(x); & m(0) = 0 \\ &= \int_0^t [1 + m(t-x)] f_X(x) dx & \text{if } f_X(x) \text{ exists} \end{aligned}$$

13

The renewal equation is linear in the function $m(t)$ and looks like equations in linear systems courses. It can be solved if $f_X(x)$ has a rational Laplace transform. The solution has the form

$$m(t) = \frac{t}{\bar{X}} + \frac{\sigma^2}{2\bar{X}^2} - \frac{1}{2} + \epsilon(t) \quad \text{for } t \geq 0,$$

where $\lim_{t \rightarrow \infty} \epsilon(t) = 0$.

The most significant term for large t is t/\bar{X} , consistent with the elementary renewal thm. The next two terms say the initial transient never quite dies away.

Heavy tailed distribution pick up extra renewals initially (recall $p_X(\epsilon) = 1 - \epsilon$, $p_X(1/\epsilon) = \epsilon$).

14

Blackwell's theorem

Blackwell's theorem essentially says that the expected renewal rate for large t is $1/\bar{X}$.

It cannot quite say this, since if X is discrete, then S_n is discrete for all n . Thus suggests that $m(t) = E[N(t)]$ does not have a derivative.

Fundamentally, there are two kinds of distribution functions — arithmetic and non-arithmetic.

A rv X has an arithmetic distribution if its set of possible sample values are integer multiples of some number, say λ . The largest such choice of λ is the span of the distribution.

15

If X is arithmetic with span $\lambda > 0$, then every S_n must be arithmetic with a span either λ or an integer multiple of λ .

Thus $N(t)$ can increase only at multiples of λ .

For a non-arithmetic discrete distribution (example: $f_X(1)=1/2, f_X(\pi)=1/2$), the points at which $N(t)$ can increase become dense as $t \rightarrow \infty$.

Blackwell's thm:

$$\lim_{t \rightarrow \infty} [m(t+\lambda) - m(t)] = \frac{\lambda}{\bar{X}} \quad \text{Arith. } X, \text{ span } \lambda$$

$$\lim_{t \rightarrow \infty} [m(t+\delta) - m(t)] = \frac{\delta}{\bar{X}} \quad \text{Non-Arith. } X, \text{ any } \delta > 0$$

16

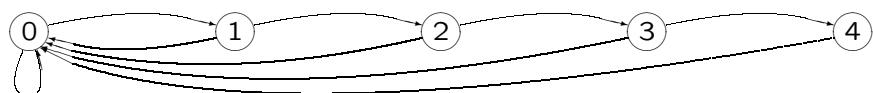
Blackwell's theorem uses difficult analysis and doesn't lead to much insight. If Laplace techniques work, then it follows from the solution there.

The hard case is non-arithmetic but discrete distributions.

The arithmetic case with a finite set of values is easy. We model the renewal process as returns to a given state in a Markov chain. Choose $\lambda = 1$ for simplicity.

17

For any renewal process with inter-renewals at a finite set of integers times, there is a corresponding Markov chain modeling returns to state 0.



The transition probabilities can be seen to be

$$P_{i,i+1} = \frac{1 - p_X(0) - p_X(1) - \cdots - p_X(i)}{1 - p_X(0) - p_X(1) - \cdots - p_X(i-1)}$$

Assuming that the chain is aperiodic, we know that $\lim_{n \rightarrow \infty} P_{00}^n = \pi_0$. As seen before, $\pi_0 = 1/\bar{X}$.

Moral of story: When doing renewals, think Markov, and when doing Markov, think renewals.

18

Delayed renewal processes

A delayed renewal process is a modification of a renewal process for which the first inter-renewal interval X_1 has a different distribution than the others. They are still all independent.

The bottom line here is that all the limit theorems remain unchanged, even if $E[X_1] = \infty$.

When modelling returns to a given state for a Markov chain, this lets us start in one state and count visits to another state.

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Lecture 16: Renewals and Countable-state Markov

Outline:

- Review major renewal theorems
- Age and duration at given t
- Countable-state Markov chains

1

Sample-path time average (strong law for renewals)

$$\Pr \left\{ \lim_{t \rightarrow \infty} \frac{N(t)}{t} = \frac{1}{\bar{X}} \right\} = 1.$$

Ensemble & time average (elementary renewal thm)

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

Ensemble average (Blackwell's thm); $m(t) = E[N(t)]$

$$\lim_{t \rightarrow \infty} [m(t+\lambda) - m(t)] = \frac{\lambda}{\bar{X}} \quad \text{Arith. } X, \text{ span } \lambda$$

$$\lim_{t \rightarrow \infty} [m(t+\delta) - m(t)] = \frac{\delta}{\bar{X}} \quad \text{Non-Arith. } X, \text{ any } \delta > 0$$

2

$$\lim_{t \rightarrow \infty} [m(t+\lambda) - m(t)] = \frac{\lambda}{\bar{X}} \quad \text{Arith. } X, \text{ span } \lambda$$

can be rewritten as

$$\lim_{n \rightarrow \infty} \Pr\{\text{renewal at } n\lambda\} = \frac{\lambda}{\bar{X}} \quad \text{Arith. } X, \text{ span } \lambda$$

If we model an arithmetic renewal process as a Markov chain starting in the renewal state 0, this essentially says $P_{00}^n \rightarrow \pi_0$.

$$\lim_{t \rightarrow \infty} [m(t+\delta) - m(t)] = \frac{\delta}{\bar{X}} \quad \text{Non-Arith. } X, \text{ any } \delta > 0$$

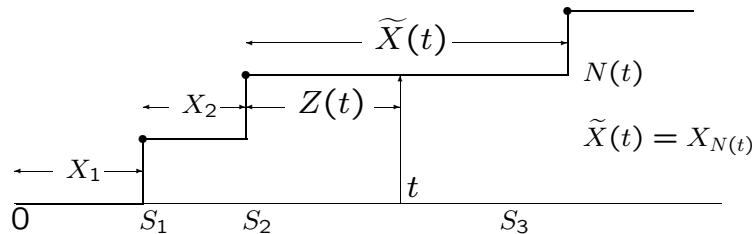
This is the best one could hope for. Note that

$$\lim_{\delta \rightarrow 0} \lim_{t \rightarrow \infty} \frac{m(t+\delta) - m(t)}{\delta} = \frac{1}{\bar{X}}$$

but the order of the limits can't be interchanged.

3

Age and duration at given t



Assume an arithmetic renewal process of span 1.

For integer t , $Z(t) = i \geq 0$ and $\tilde{X}(t) = k > i$ iff there are successive arrivals at $t - i$ and $t - i + k$.

Let $q_j = \Pr\{\text{arrival at time } j\} = \sum_{n \geq 1} p_{S_n}(j)$ and let $q_0 = 1$ (nominal arrival at time 0). Then

$$p_{Z(t), \tilde{X}(t)}(i, k) = q_{t-i} p_X(k) \quad \text{for } 0 \leq i \leq t; k > i$$

4

$$p_{Z(t), \tilde{X}(t)}(i, k) = q_{t-i} p_X(k) \quad \text{for } 0 \leq i \leq t; k > i$$

Note that

$$q_i = \Pr\{\text{arrival at } j\} = E[\text{arrival at } j] = m(i) - m(i-1),$$

so by Blackwell, $\lim_{j \rightarrow \infty} 1/\bar{X}$.

$$\lim_{t \rightarrow \infty} p_{Z(t), \tilde{X}(t)}(i, k) = \frac{p_X(k)}{\bar{X}} \quad \text{for } k > i \geq 0.$$

$$\lim_{t \rightarrow \infty} p_{Z(t)}(i) = \frac{\sum_{k=i+1}^{\infty} p_X(k)}{\bar{X}} = \frac{F_X^c(i)}{\bar{X}} \quad \text{for } i \geq 0.$$

$$\lim_{t \rightarrow \infty} p_{\tilde{X}(t)}(k) = \frac{\sum_{i=0}^{k-1} p_X(k)}{\bar{X}} = \frac{k p_X(k)}{\bar{X}} \quad \text{for } k \geq 1.$$

5

Now look at asymptotic expected duration:

$$\lim_{t \rightarrow \infty} E[\tilde{X}(t)] = \sum_{k=1}^{\infty} k \cdot k p_X(k)/\bar{X} = E[X^2]/\bar{X}$$

This is the same as the sample-path average, but now we can look at the finite t case. More important, we get a different interpretation.

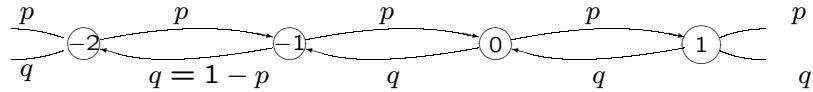
For a given $\tilde{X} = k$, there are k equiprobable choices for age; for each choice, the joint Z, \tilde{X} PMF is $p_X(k)/\bar{X}$. Thus large durations are enhanced relative to inter-renewals.

The expected age (after some work) is $E[X^2]/2\bar{X} - \frac{1}{2}$. This is at integer values of large t . The age increases linearly with slope 1 to the next integer value and then drops by 1.

6

Countable -state Markov chains

The biggest change from finite-state Markov chains to countable-state chains is the concept of a recurrent class. Example:

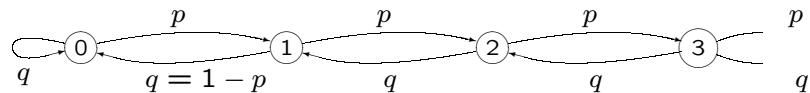


This Markov chain models a Bernoulli ± 1 process. The state at time n is $S_n = X_1 + X_2 + \dots + X_n$. The state S_n at time n is $j = 2k - n$ where k is the number of positive transitions in the n trials.

All states communicate and have a period $d = 2$; $\sigma_{S_n}^2 = n[1 - (p - q)^2]$. $P_{0,j}^n$ approaches 0 at least as $1/\sqrt{n}$ for every j .

7

Another example (called a birth-death chain)



In this case, if $p > 1/2$, the state drifts to the right and $P_{0,j}^n$ approaches 0 for all j . If $p < 1/2$, it drifts to the left and keeps bumping state 0.

A truncated version of this was analyzed in the homework. With $p > 1/2$, the steady-state increases to the right, with $p < 1/2$, it increases to the left, and at $p = 1/2$ it is uniform.

As the truncation point increases, the ‘steady-state’ remains positive only for $p < 1/2$.

8

We want to define recurrent to mean that, given $X_0 = i$, there is a future return to state i WP1. We will see that the birth-death chain above is recurrent if $p < 1/2$ and not recurrent if $p > 1/2$. The case $p = 1/2$ is strange and will be called null-recurrent.

We can use renewal theory to study recurrent chains, but first must understand first-passage-times.

Def: The first-passage-time probability, $f_{ij}(n)$, is

$$f_{ij}(n) = \Pr\{X_n=j, X_{n-1} \neq j, X_{n-2} \neq j, \dots, X_1 \neq j | X_0=i\}.$$

It's the probability, given $X_0 = i$, that n is the first epoch at which $X_n = j$. Then

$$f_{ij}(n) = \sum_{k \neq j} P_{ik} f_{kj}(n-1); \quad n > 1; \quad f_{ij}(1) = P_{ij}.$$

9

$$f_{ij}(n) = \sum_{k \neq j} P_{ik} f_{kj}(n-1); \quad n > 1; \quad f_{ij}(1) = P_{ij}.$$

Recall that Chapman-Kolmogorov says

$$P_{ij}^n = \sum_k P_{ik} P_{kj}^{n-1},$$

so the difference between $f_{ij}(n)$ and P_{ij}^n is only in cutting off the outputs from j (as before in finding expected first-passage-times).

Let $F_{ij}(n) = \sum_{m \leq n} f_{ij}(m)$ be the probability of reaching j by time n or before. If $\lim_{n \rightarrow \infty} F_{ij}(n) = 1$, there is a rv T_{ij} with distribution function F_{ij} that is the first-passage-time rv.

10

We can also express $F_{ij}(n)$ as

$$F_{ij}(n) = P_{ij} + \sum_{k \neq j} P_{ik} F_{kj}(n-1); \quad n > 1; \quad F_{ij}(1) = P_{ij}$$

Since $F_{ij}(n)$ is nondecreasing in n , the limit $F_{ij}(\infty)$ must exist and satisfy

$$F_{ij}(\infty) = P_{ij} + \sum_{k \neq j} P_{ik} F_{kj}(\infty).$$

Unfortunately, choosing $F_{ij}(\infty) = 1$ for all i, j satisfies these equations. The correct solution turns out to be the smallest set of $F_{ij}(\infty)$ that satisfies these equations.

If $F_{jj}(\infty) = 1$, then an eventual return from state j occurs with probability 1 and the sequence of returns is the sequence of renewal epochs in a renewal process.

11

If $F_{jj}(\infty) = 1$, then there is a rv T_{jj} with the distribution function $F_{jj}(n)$ and j is recurrent. The renewal process of returns to j then has inter-renewal intervals with the distribution function $F_{jj}(n)$.

From renewal theory, the following are equivalent:

- 1) state j is recurrent.
- 2) $\lim_{t \rightarrow \infty} N_{jj}(t) = \infty$ with probability 1.
- 3) $\lim_{t \rightarrow \infty} E[N_{jj}(t)] = \infty$.
- 4) $\lim_{t \rightarrow \infty} \sum_{1 \leq n \leq t} P_{jj}^n = \infty$.

None of these imply that $E[T_{jj}] < \infty$.

12

Two states are in the same class if they communicate (same as for finite-state chains).

If states i and j are in the same class then either both are recurrent or both transient (not recurrent).

Pf: If j is recurrent, then $\sum_n P_{jj}^n = \infty$. Then

$$\sum_{n=1}^{\infty} P_{ii}^n \geq \sum_{k=1}^{\infty} P_{ij}^m P_{jj}^k P_{jk}^{\ell} = \infty$$

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

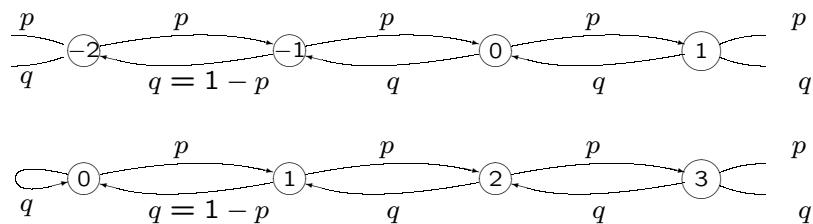
By the same kind of argument, if i, j are recurrent, then $F_{ij}(\infty) = 1$.

13

If a state j is recurrent, then T_{jj} might or might not have a finite expectation.

Def: If $E[T_{jj}] < \infty$, j is positive recurrent. If T_{jj} is a rv and $E[T_{jj}] = \infty$, then j is null recurrent. Otherwise j is transient.

For $p = 1/2$, each state in each of the following is null recurrent.



14

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Lecture 17: Countable-state Markov chains

Outline:

- Strong law proofs
- Positive-recurrence and null-recurrence
- Steady-state for positive-recurrent chains
- Birth-death Markov chains
- Reversibility

1

Let $\{Y_i; i \geq 1\}$ be the IID service times for a $(G/G/\infty)$ queue and let $\{N(t); t > 0\}$ be the renewal process with interarrivals $\{X_i; i \geq 1\}$. Consider the following plausibility argument for $\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^{N(t,\omega)} Y_i(\omega)$.

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^{N(t,\omega)} Y_i(\omega) &= \lim_{t \rightarrow \infty} \left[\frac{N(t,\omega)}{t} \frac{\sum_{i=1}^{N(t,\omega)} Y_i(\omega)}{N(t,\omega)} \right] \quad (1) \\ &= \lim_{t \rightarrow \infty} \frac{N(t,\omega)}{t} \lim_{t \rightarrow \infty} \frac{\sum_{i=1}^{N(t,\omega)} Y_i(\omega)}{N(t,\omega)} \quad (2) \\ &= \lim_{t \rightarrow \infty} \frac{N(t,\omega)}{t} \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n Y_i(\omega)}{n} \quad (3) \\ &= \frac{1}{\bar{X}} \bar{Y} \quad \text{WP1} \quad (4) \end{aligned}$$

This assumes $\bar{X} < \infty$, $\bar{Y} < \infty$.

2

To do this carefully, work from bottom up.

Let $A_1 = \{\omega : \lim_{t \rightarrow \infty} N(t, \omega)/t = 1/\bar{X}\}$. By the strong law for renewal processes $\Pr\{A_1\} = 1$.

Let $A_2 = \{\omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n Y_i(\omega) = \bar{Y}\}$. By the SLLN, $\Pr\{A_2\} = 1$. Thus (3) = (4) for $\omega \in A_1 A_2$ and $\Pr\{A_1 A_2\} = 1$.

Assume $\omega \in A_2$, and $\epsilon > 0$. Then $\exists m(\epsilon, \omega)$ such that $|\frac{1}{n} \sum_{i=1}^n Y_i(\omega) - \bar{Y}| < \epsilon$ for all $n \geq m(\epsilon, \omega)$. If $\omega \in A_1$ also, then $\lim_{t \rightarrow \infty} N(t, \omega) = \infty$, so $\exists t(\epsilon, \omega)$ such that $N(t, \omega) \geq m(\epsilon, \omega)$ for all $t \geq t(\epsilon, \omega)$.

$$\left| \frac{\sum_{i=1}^{N(t, \omega)} Y_i(\omega)}{N(t, \omega)} - \bar{Y} \right| < \epsilon \quad \text{for all } t \geq t(\epsilon, \omega)$$

Since ϵ is arbitrary, (2) = (3) = (4) for $\omega \in A_1 A_2$.

3

Finally, can we interchange the limit of a product of two functions (say $f(t)g(t)$) with the product of the limits? If the two functions each have finite limits (as the functions of interest do for $\omega \in A_1 A_2$), the answer is yes, establishing (1) = (4).

To see this, assume $\lim_t f(t) = a$ and $\lim_t g(t) = b$. Then

$$\begin{aligned} f(t)g(t) - ab &= (f(t) - a)(g(t) - b) + a(g(t) - b) + b(f(t) - a) \\ |f(t)g(t) - ab| &\leq |f(t) - a||g(t) - b| + |a||g(t) - b| + |b||f(t) - a| \end{aligned}$$

For any $\epsilon > 0$, choose $t(\epsilon)$ such that $|f(t) - a| \leq \epsilon$ for $t \geq t(\epsilon)$ and $|g(t) - b| \leq \epsilon$ for $t \geq t(\epsilon)$. Then

$$|f(t)g(t) - ab| \leq \epsilon^2 + \epsilon|a| + \epsilon|b| \quad \text{for } t \geq t(\epsilon).$$

Thus $\lim_t f(t)g(t) = \lim_t f(t) \lim_t g(t)$.

4

Review - Countable-state chains

Two states are in the same class if they communicate (same as for finite-state chains).

Thm: All states in the same class are recurrent or all are transient.

Pf: Assume j is recurrent; then $\sum_n P_{jj}^n = \infty$. For any i such that $j \leftrightarrow i$, $P_{ij}^m > 0$ for some m and P_{ji}^ℓ for some ℓ . Then (recalling $\lim_t E[N_{ii}(t)] = \sum_n P_{ii}^n$)

$$\sum_{n=1}^{\infty} P_{ii}^n \geq \sum_{k=n-m-\ell}^{\infty} P_{ij}^m P_{jj}^k P_{ji}^\ell = \infty$$

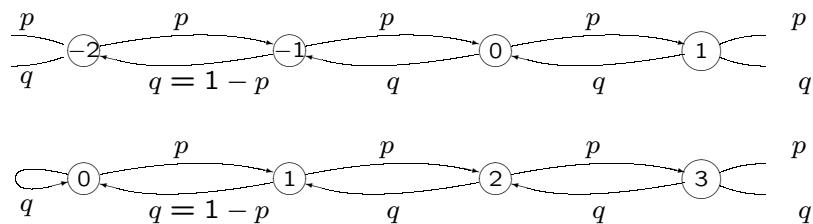
By the same kind of argument, if $i \leftrightarrow j$ are recurrent, then $\sum_{n=1}^{\infty} P_{ij}^n = \infty$ (so also $\lim_t E[N_{ij}^t] = \infty$).

5

If a state j is recurrent, then the recurrence time T_{jj} might or might not have a finite expectation.

Def: If $E[T_{jj}] < \infty$, j is positive-recurrent. If T_{jj} is a rv and $E[T_{jj}] = \infty$, then j is null-recurrent. Otherwise j is transient.

For $p = 1/2$, each state in each of the following is null recurrent.



6

Positive-recurrence and null-recurrence

Suppose $i \leftrightarrow j$ are recurrent. Consider the renewal process of returns to j with $X_0 = j$. Consider rewards $R(t) = 1$ whenever $X(t) = i$. By the renewal-reward thm (4.4.1),

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t R(\tau) d\tau = \frac{\mathbb{E}[R_n]}{\bar{T}_{jj}} \quad \text{WP1},$$

where $\mathbb{E}[R_n]$ is the expected number of visits to i within a recurrence of j . The left side is $\lim_{t \rightarrow \infty} \frac{1}{t} N_{ji}(t)$, which is $1/\bar{T}_{ii}$. Thus

$$\frac{1}{\bar{T}_{ii}} = \frac{\mathbb{E}[R_n]}{\bar{T}_{jj}}$$

Since there must be a path from j to i , $\mathbb{E}[R_n] > 0$.

Thm: For $i \leftrightarrow j$ recurrent, either both are positive-recurrent or both null-recurrent.

7

Steady-state for positive-recurrent chains

We define steady-state probabilities for countable-state Markov chains in the same way as for finite-state chains, namely,

Def: $\{\pi_i; i \geq 0\}$ is a steady-state distribution if

$$\pi_j \geq 0; \quad \pi_j = \sum_i \pi_i P_{ij} \quad \text{for all } j \geq 0 \quad \text{and} \quad \sum_j \pi_j = 1$$

Def: An irreducible Markov chain is a Markov chain in which all pairs of states communicate.

For finite-state chains, irreducible means recurrent. Here it can be positive-recurrent, null-recurrent, or transient.

8

If steady-state π exists and if $\Pr\{X_0 = i\} = \pi_i$ for each i , then $p_{X_1}(j) = \sum_i \pi_i P_{ij} = \pi_j$. Iterating, $p_{X_n}(j) = \pi_j$, so steady-state is preserved. Let $\widetilde{N}_j(t)$ be number of visits to j in $(0, t]$ starting in steady state. Then

$$E[\widetilde{N}_j(t)] = \sum_{k=1}^n \Pr\{X_k = j\} = n\pi_j$$

Awkward thing about renewals and Markov: $\widetilde{N}_j(t)$ works for some things and $N_{jj}(t)$ works for others. Here is a useful hack:

$N_{ij}(t)$ is 1 for first visit to j (if any) plus $N_{ij}(t) - 1$ for subsequent recurrences j to j . Thus

$$\begin{aligned} E[N_{ij}(t)] &\leq 1 + E[N_{jj}(t)] \\ E[\widetilde{N}_j(t)] &= \sum_i \pi_i E[N_{ij}(t)] \leq 1 + E[N_{jj}(t)] \end{aligned}$$

9

Major theorem: For an irreducible Markov chain, the steady-state equations have a solution if and only if the states are positive-recurrent. If a solution exists, then $\pi_i = 1/\bar{T}_{ii} > 0$ for all i .

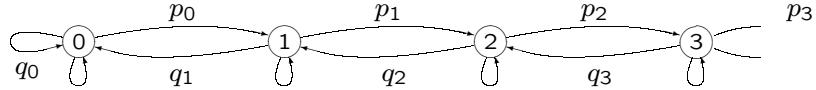
Pf: (only if; assume π exists, show positive-recur.)
For each j and t ,

$$\begin{aligned} \pi_j &= \frac{E[\widetilde{N}_j(t)]}{t} \leq \frac{1}{t} + \frac{E[N_{jj}(t)]}{t} \\ &\leq \lim_{t \rightarrow \infty} \frac{E[N_{jj}(t)]}{t} = \frac{1}{\bar{T}_{jj}} \end{aligned}$$

Since $\sum_j \pi_j = 1$, some $\pi_j > 0$. Thus $\lim_{t \rightarrow \infty} E[N_{jj}(t)]/t > 0$ for that j , so j is positive-recurrent. Thus all states are positive-recurrent. See text to show that ' \leq ' above is equality.

10

Birth-death Markov chains



For any state i and any sample path, the number of $i \rightarrow i+1$ transitions is within 1 of the number of $i+1 \rightarrow j$ transitions; in the limit as the length of the sample path $\rightarrow \infty$,

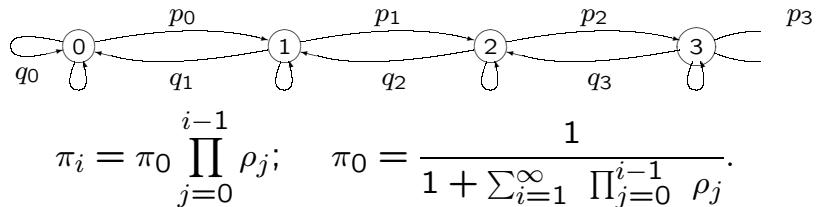
$$\pi_i p_i = \pi_{i+1} q_{i+1}; \quad \pi_{i+1} = \frac{\pi_i p_i}{q_{i+1}}$$

Letting $\rho_i = p_i/q_{i+1}$, this becomes

$$\pi_i = \pi_0 \prod_{j=0}^{i-1} \rho_j; \quad \pi_0 = \frac{1}{1 + \sum_{i=1}^{\infty} \prod_{j=0}^{i-1} \rho_j}.$$

This agrees with the steady-state equations.

11



This solution is a function only of ρ_0, ρ_1, \dots and doesn't depend on size of self loops.

The expression for π_0 converges (making the chain positive recurrent) (essentially) if the ρ_i are asymptotically less than 1.

Methodology: We could check renewal results carefully to see if finding π_i by up/down counting is justified. Using the major theorem is easier.

Birth-death chains are particularly useful in queuing where births are arrivals and deaths departures.

12

Reversibility

$$\Pr\{X_{n+k}, \dots, X_{n+1} | X_n, \dots, X_0\} = \Pr\{X_{n+k}, \dots, X_{n+1} | X_n\}$$

For any A^+ defined on X_{n+1} up and A^- defined on X_{n-1} down,

$$\Pr\{A^+ | X_n, A^-\} = \Pr\{A^+ | X_n\}$$

$$\Pr\{A^+, A^- | X_n\} = \Pr\{A^+ | X_n\} \Pr\{A^- | X_n\}.$$

$$\Pr\{A^- | X_n, A^+\} = \Pr\{A^- | X_n\}.$$

$$\Pr\{X_{n-1} | X_n, X_{n+1}, \dots, X_{n+k}\} = \Pr\{X_{n-1} | X_n\}.$$

13

By Bayes,

$$\Pr\{X_{n-1} | X_n\} = \frac{\Pr\{X_n | X_{n-1}\} \Pr\{X_{n-1}\}}{\Pr\{X_n\}}.$$

If the forward chain is in steady state, then

$$\Pr\{X_{n-1} = j | X_n = i\} = P_{ji} \pi_j / \pi_i.$$

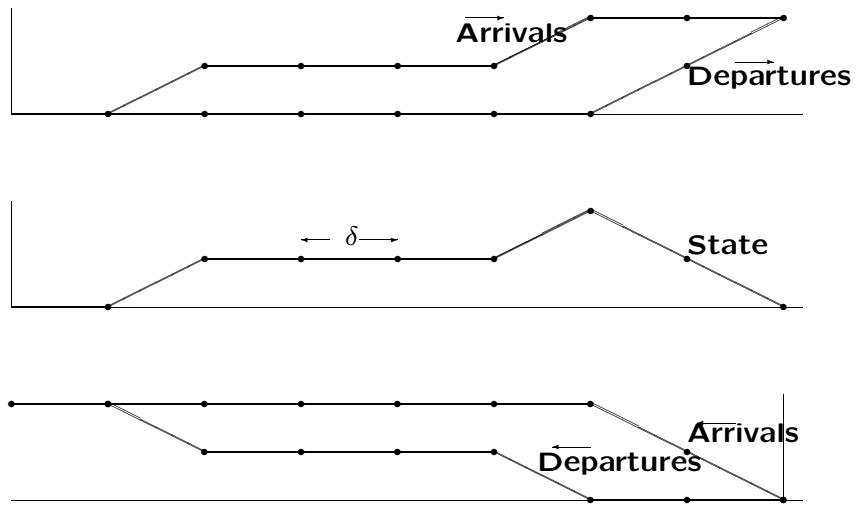
Aside from the homogeneity involved in starting at time 0, this says that a Markov chain run backwards is still Markov. If we think of the chain as starting in steady state at time $-\infty$, these are the equations of a (homogeneous) Markov chain. Denoting $\Pr\{X_{n-1} = j | X_n = i\}$ as the backward transition probabilities P_{ji}^* , forward/ backward are related by

$$\pi_i P_{ij}^* = \pi_j P_{ji}.$$

Def: A chain is reversible if $P_{ij}^* = P_{ij}$ for all i, j .

14

Thm: A birth/death Markov chain is reversible if it has a steady-state distribution.



MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

L18: Countable state Markov chains and processes

Outline:

- Review - Reversibility
- Sample-time M/M/1 queue
- Branching processes
- Markov processes with countable state spaces
- The M/M/1 queue

1

For any Markov chain,

$$\Pr\{X_{n+k}, \dots, X_{n+1} | X_n, \dots, X_0\} = \Pr\{X_{n+k}, \dots, X_{n+1} | X_n\}$$

For any A^+ defined on X_{n+1} up and A^- defined on X_{n-1} down,

$$\Pr\{A^+ | X_n, A^-\} = \Pr\{A^+ | X_n\}$$

$$\Pr\{A^+, A^- | X_n\} = \Pr\{A^+ | X_n\} \Pr\{A^- | X_n\}.$$

$$\Pr\{A^- | X_n, A^+\} = \Pr\{A^- | X_n\}.$$

$$\Pr\{X_{n-1} | X_n, X_{n+1}, \dots, X_{n+k}\} = \Pr\{X_{n-1} | X_n\}.$$

The Markov condition works in both directions, but need steady state in forward chain for homogeneity in backward chain.

2

For a positive-recurrent Markov chain in steady-state, the backward probabilities are

$$\Pr\{X_{n-1} = j \mid X_n = i\} = P_{ji}\pi_j/\pi_i.$$

Denote $\Pr\{X_{n-1} = j \mid X_n = i\}$ as the backward transition probabilities. Then

$$\pi_i P_{ij}^* = \pi_j P_{ji} = \Pr\{X_n = i, X_{n-1} = j\}.$$

Def: A chain is reversible if $P_{ij}^* = P_{ij}$ for all i, j .

If chain is reversible, then $\pi_i P_{ij} = \pi_j P_{ji}$ for all i, j , i.e., if $\Pr\{X_n = i, X_{n-1} = j\} = \Pr\{X_n = j, X_{n-1} = i\}$. In other words, reversibility means that the long-term fraction of i to j transitions is the same as the long-term fraction of j to i transitions.

All positive-recurrent birth-death chains are reversible.

3

More general example: Suppose the non-zero transitions of a positive-recurrent Markov chain form a tree. Then the number of times a transition is crossed in one direction differs by at most one from the number of transitions in the other direction, so the chain is reversible.

Note that a birth-death chain is a very skinny tree.

The following theorem is a great time-saver and is sometimes called the guessing theorem.

Thm: For a Markov chain $\{P_{ij}; i, j \geq 0\}$, if a set of numbers $\pi_i > 0, \sum_i \pi_i = 1$ exist such that $\pi_i P_{ij} = \pi_j P_{ji}$ for all i, j , then the chain is positive-recurrent and reversible and $\{\pi_i; i \geq 0\}$ is the set of steady-state probabilities.

4

Thm: $\pi_i P_{ij} = \pi_j P_{ji}$ for all i, j implies reversibility with $\{\pi_i\}$ steady-state.

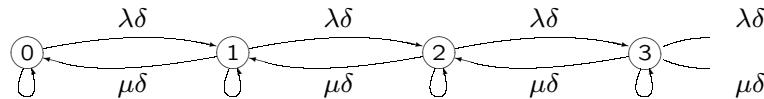
Pf: Sum over i to get $\sum_i \pi_i P_{ij} = \pi_j \sum_i P_{ji} = \pi_j$. These (along with $\sum_i \pi_i = 1$ and $\pi_i \geq 0$) are the steady state equations and have a unique, positive solution.

Sanity checks for reversibility: 1) If $P_{ij} > 0$ then $P_{ji} > 0$. 2) If periodic, period is 2. 3) $P_{ij}P_{jk}P_{ki} = P_{ik}P_{kj}P_{ji}$.

Generalization of guessing thm to non-reversible chains: If $\exists \{\pi_i \geq 0; i \geq 0\}$ with $\sum_i \pi_i = 1$ and \exists transition probabilities $\{P_{ij}^*\}$ such that $\pi_i P_{ij} = \pi_j P_{ji}^*$ for all i, j , then $\{\pi_i; i \geq 0\}$ are steady-state probabilities and $\{P_{ij}^*\}$ are the backward probabilities.

5

Suppose we sample the state of an M/M/1 queue at some time increment δ so small that we can ignore more than one arrival or departure in an increment. The rate of arrivals is λ and that of departures is $\mu > \lambda$.

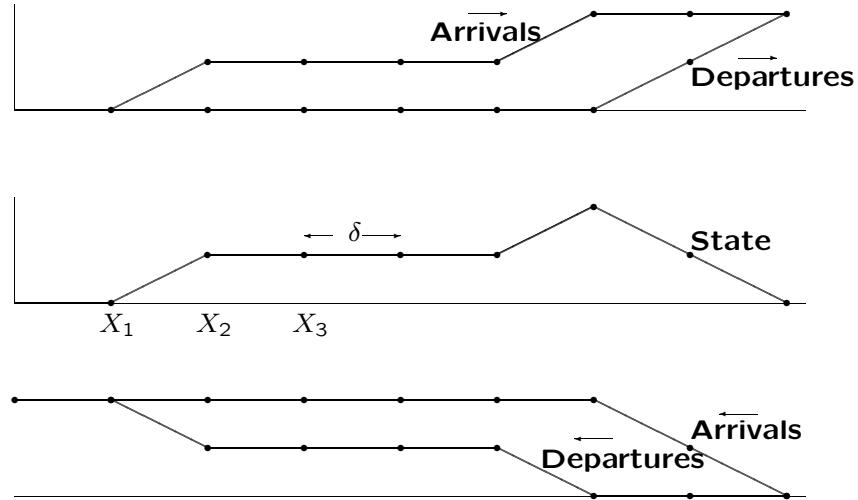


Either from the guessing theorem or the general result for birth/death chains, we see that $\pi_{n-1}\lambda\delta = \pi_n\mu\delta$ so, with $\rho = \lambda/\mu$,

$$\pi_n = \rho\pi_{n-1}; \quad \pi_n = \rho^n\pi_0; \quad \pi_n = (1 - \rho)\rho^n$$

Curiously, this does not depend on δ (so long as $(\lambda + \mu)\delta \leq 1$), so these are the steady state probabilities as $\delta \rightarrow 0$.

6



In the original (right-moving) chain, the state increases on arrivals and decreases on departures.

Each sample path corresponds to both a right and left moving chain, each M/M/1

7

Burke's thm: Given an M/M/1 sample-time Markov chain in steady state, first, the departure process is Bernoulli at rate λ . Second, the state at $n\delta$ is independent of departures prior to $n\delta$.

When we look at a sample path from right to left, each departure becomes an arrival and vice-versa. The right to left Markov chain is M/M/1.

Thus everything we know about the M/M/1 sample-time chain has a corresponding statement with time reversed and arrival-departure switched.

8

Branching processes

A branching process is a very simple model for studying how organisms procreate or die away. It is a simplified model of photons in a photomultiplier, cancer cells, insects, etc.

Let X_n be the number of elements in generation n . For each element k , $1 \leq k \leq X_n$, let $Y_{k,n}$ be the number of offspring of that element. Then

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{k,n}$$

The nonnegative integer rv's $Y_{k,n}$ are IID over both n and k .

The initial generation X_0 can be an arbitrary positive rv, but is usually taken to be 1.

9

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{k,n}$$

Examples: If $Y_{k,n}$ is deterministic and $Y_{k,n} = 1$, then $X_n = X_{n-1} = X_0$ for all $n \geq 1$.

If $Y_{k,n} = 2$, then $X_n = 2X_{n-1} = 2^n X_0$ for all $n \geq 1$.

If $p_Y(0) = 1/2$ and $p_Y(2) = 1/2$, then $\{X_n; n \geq 0\}$ is a rather peculiar Markov chain. It can grow explosively, or it can die out. If it dies out, it stays dead, so state 0 is a trapping state.

The state 0 is a trapping state in general. The even numbered states all communicate (but, as we will see, are all transient), and each odd numbered state does not communicate with any other state.

10

Let's find the probability (for the general case) that the process dies out. Let

$$p_Y(k) = p_k \quad \text{and} \quad \Pr\{X_n=j \mid X_{n-1}=i\} = P_{ij}.$$

Let $F_{ij}(n)$ be the probability that state j is reached on or before step n starting from state i . Then

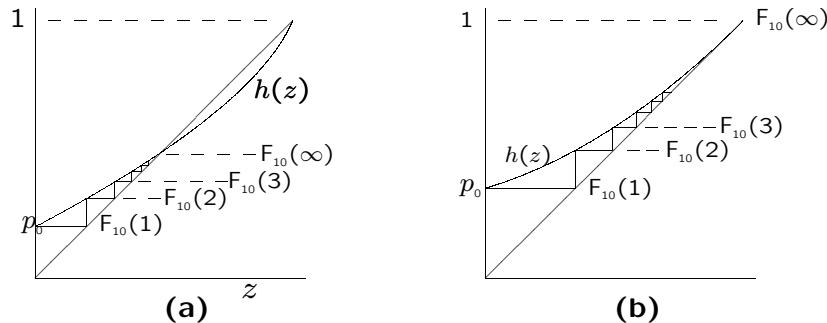
$$F_{ij}(n) = P_{ij} + \sum_{k \neq j} P_{ik} F_{kj}(n-1), \quad n > 1; \quad F_{ij}(1) = P_{ij}.$$

$$\begin{aligned} F_{10}(n) &= p_0 + \sum_{k=1}^{\infty} p_k [F_{10}(n-1)]^k \\ &= \sum_{k=0}^{\infty} p_k [F_{10}(n-1)]^k. \end{aligned}$$

Let $h(z) = \sum_k p_k z^k$. Then $F_{10}(n) = h(F_{10}(n-1))$.

11

Let $h(z) = \sum_k p_k z^k$. Then $F_{10}(n) = h(F_{10}(n-1))$.



We see that $F_{10}(\infty) < 1$ in case (a) and $F_{10}(\infty) = 1$ in case (b). For case (a), $h'(z)|_{z=1} = \bar{Y} > 1$ and in case (b), $h'(z)|_{z=1} = \bar{Y} \leq 1$.

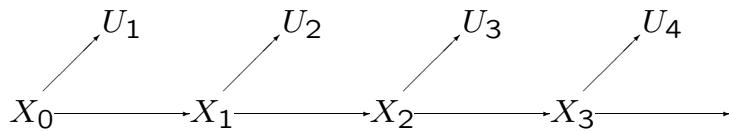
For case a), the process explodes (with probability $1 - F_{10}(\infty)$) or dies out (with probability $F_{10}(\infty)$).

12

Markov processes

A countable-state Markov process can be viewed as an extension of a countable-state Markov chain. Along with each step in the chain, there is an exponential holding time U_i before the next step into state X_i .

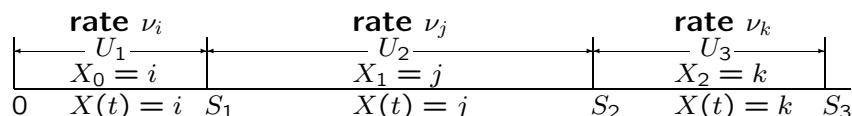
The rate of each exponential holding time U_i is determined by X_{i-1} but is otherwise independent of other holding times and other states. The dependence is as illustrated below.



Each rv U_n , conditional on X_{n-1} , is independent of all other states and holding times.

13

The evolution in time of a Markov process can be visualized by



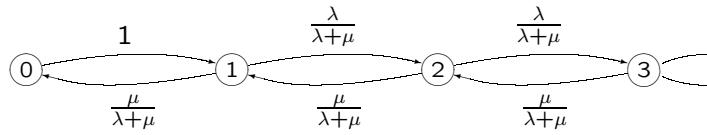
We will usually assume that the embedded Markov chain for a Markov process has no self-transitions, since these are ‘hidden’ in a sample path of the process.

The Markov process is taken to be $\{X(t); t \geq 0\}$. Thus a sample path of $X_n; n \geq 0$ and $\{U_n; n \geq 1\}$ specifies $\{X(t); t \geq 0\}$ and vice-versa.

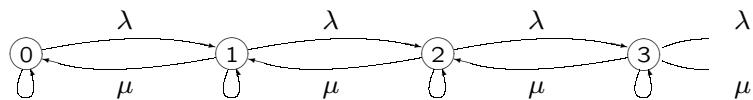
$$\begin{aligned} \Pr\{X(t)=j \mid X(\tau)=i, \{X(s); s < \tau\}\} &= \\ &= \Pr\{X(t-\tau)=j \mid X(0)=i\}. \end{aligned}$$

14

The M/M/1 queue



This diagram gives the embedded Markov chain for the M/M/1 Markov process. The process itself can be represented by



This corresponds to the rate of transitions given a particular state.

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

L19: Countable-state Markov processes

Outline:

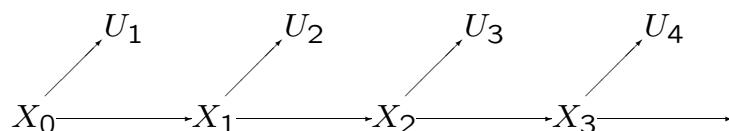
- Review - Markov processes
- Sampled-time approximation to MP's
- Renewals for Markov processes
- Steady-state for irreducible MP's

1

Markov processes

A countable-state Markov process can be defined as an extension of a countable-state Markov chain. Along with each step, say from X_{n-1} to X_n , in the embedded Markov chain, there is an exponential holding time U_n before X_n is entered.

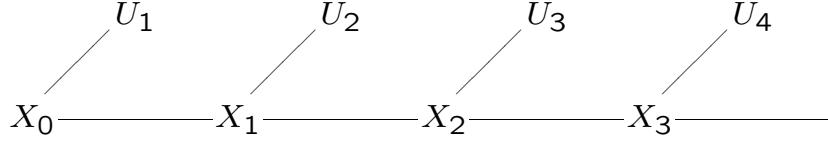
The rate of each exponential holding time U_n is determined by X_{n-1} but is otherwise independent of other holding times and other states. The dependence is as illustrated below.



Each rv U_n , conditional on X_{n-1} , is independent of all other states and holding times.

2

In a directed tree of dependencies, each rv, conditional on its parent, is statistically independent of all earlier rv's. But the direction in the tree is not needed.



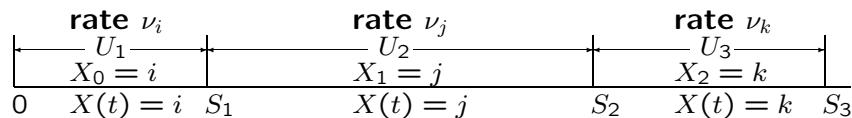
For example,

$$\begin{aligned}\Pr\{X_0 X_1 X_2 U_2\} &= \Pr\{X_0\} \Pr\{X_1|X_0\} \Pr\{X_2|X_1\} \Pr\{U_2|X_1\} \\ &= \Pr\{X_1\} \Pr\{X_0|X_1\} \Pr\{X_2|X_1\} \Pr\{U_2|X_1\}\end{aligned}$$

Conditioning on any node breaks the tree into independent subtrees. Given X_2 , (X_0, X_1, U_1, U_2) and (U_3) and (X_3, U_4) are statistically independent.

3

The evolution in time of a Markov process can be visualized by



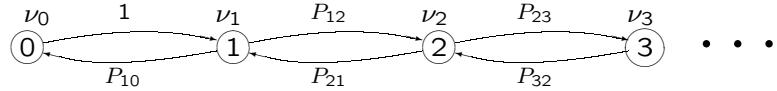
We usually assume that the embedded Markov chain for a Markov process has no self-transitions, since these are hidden in a sample path of the process.

The Markov process is taken to be $\{X(t); t \geq 0\}$. Thus a sample path of $X_n; n \geq 0$ and $\{U_n; n \geq 1\}$ specifies $\{X(t); t \geq 0\}$ and vice-versa.

$$\begin{aligned}\Pr\{X(t)=j | X(\tau)=i, \{X(s); s < \tau\}\} &= \\ &= \Pr\{X(t-\tau)=j | X(0)=i\}.\end{aligned}$$

4

We can represent a Markov process by a graph for the embedded Markov chain with rates given on the nodes:



Ultimately, we are usually interested in the state as a function of time, namely the process $\{X(t); t \geq 0\}$. This is usually called the Markov process itself.

$$X(t) = X_n \quad \text{for } t \in [S_n, S_{n+1})$$

Self transitions don't change $X(t)$.

5

We can visualize a transition from one state to another by first choosing the state (via $\{P_{ij}\}$) then choosing the transition time (exponential with ν_i).

Equivalently, choose the transition time first, then the state (they are independent).

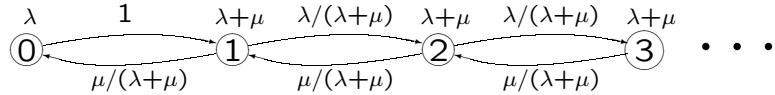
Equivalently, visualize a Poisson process for each state pair i, j with a rate $q_{ij} = \nu_i P_{ij}$. On entry to state i , the next state is the j with the next Poisson arrival according to q_{ij} .

What is the conditional distribution of U_1 given $X_0 = i$ and $X_1 = j$?

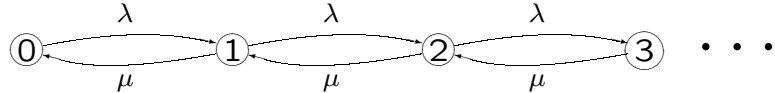
$$\nu_i = \sum_j q_{ij}; \quad P_{ij} = q_{ij}/\nu_i : \quad [q] \text{ specifies } [P], \vec{\nu}.$$

6

It is often more insightful to use q_{ij} in a Markov process graph.



An M/M/1 queue using $[P]$ and $\vec{\nu}$



The same M/M/1 queue using $[q]$.

Both these graphs contain the same information. The latter corresponds more closely to our real-world interpretation of an M/M/1 queue.

7

Sampled-time approximation to MP's

Suppose we quantize time to δ increments and view all Poisson processes in a MP as Bernoulli with $P_{ij}(\delta) = \delta q_{ij}$.

Since shrinking Bernoulli goes to Poisson, we would conjecture that the limiting Markov chain as $\delta \rightarrow 0$ goes to a MP in the sense that $X(t) \approx X'(\delta n)$.

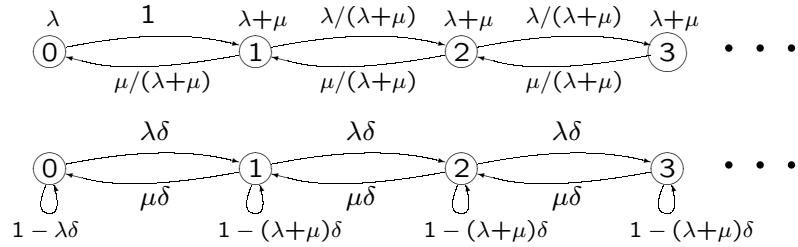
It is necessary to put self-transitions into a sampled-time approximation to model increments where nothing happens.

$$P_{ii} = 1 - \delta \nu_i; \quad P_{ij} = \delta q_{ij} \quad j \neq i$$

This requires $\delta \leq \frac{1}{\max \nu_i}$ and is only possible when the holding-time rates are bounded.

8

The embedded-chain model and sampled-time model of an M/M/1 queue:



Steady state for the embedded chain, is $\pi_0 = (1 - \rho)/2$ and $\pi_i = \frac{1}{2}(1 - \rho)^2 \rho^{n-1}$ for $i > 1$ where $\rho = \lambda/\mu$. The fraction of transitions going into state i is π_i .

Steady state for sampled-time does not depend on δ and is $\pi'_i = (1 - \rho)\rho^i$ where $\rho = \lambda/\mu$. This is the fraction of time in state i .

9

Renewals for Markov processes

Def: An irreducible MP is a MP for which the embedded Markov chain is irreducible (i.e., all states are in the same class).

We saw that irreducible Markov chains could be transient - the state simply wanders off with high probability, never to return.

We will see that irreducible MP's can have even more bizarre behavior such as infinitely many transitions in a finite time or a transition rate decaying to 0.

Review: An irreducible countable-state Markov chain is positive recurrent iff the steady-state equations,

$$\pi_j = \sum_i \pi_i P_{ij} \text{ for all } j; \quad \pi_j \geq 0 \text{ for all } j; \quad \sum_j \pi_j = 1$$

have a solution. If there is a solution, it is unique and $\pi_i > 0$ for all i . Also, the number of visits, $N_{ij}(n)$, in the first n transitions to j given $X_0 = i$ satisfies

$$\lim_{n \rightarrow \infty} \frac{1}{n} N_{ij}(n) = \pi_j \quad \text{WP1}$$

We guess that for an MP, the fraction of time in state j should be

$$p_j = \frac{\pi_j / \nu_j}{\sum_i \pi_i / \nu_i}$$

11

Thm: Let $M_i(t)$ be the number of transitions in $(0, t]$ for a MP starting in state i . Then $\lim_{t \rightarrow \infty} M_i(t) = \infty$ WP1.

Essentially, given any state, a transition must occur within finite time. Then another, etc. See text.

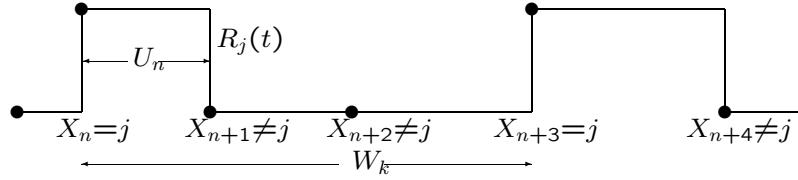
Thm: Let $M_{ij}(t)$ be the number of transitions to j in $(0, t]$ starting in state i . If the embedded chain is recurrent, then $M_{ij}(t)$ is a delayed renewal process.

Essentially, transitions keep occurring so renewals into state j must keep occurring.

12

Steady-state for irreducible MP's

Let $p_j(i)$ be the time-average fraction of time in state j for the delayed RP $\{M_{ij}(t); t > 0\}$:



From the (delayed) renewal reward theorem,

$$p_j(i) = \lim_{t \rightarrow \infty} \frac{\int_0^t R_j(\tau) d\tau}{t} = \frac{\bar{U}(j)}{\bar{W}(j)} = \frac{1}{\nu_j \bar{W}(j)} \quad \text{WP1.}$$

This relates the time-average state probabilities (WP1) to the mean recurrence times. Also $p_j(i)$ is independent of the starting state i .

13

If we can find $\bar{W}(j)$, we will also know p_j . Since $M_{ij}(t)$ is a (delayed) renewal process, the strong law for renewals says

$$\lim_{t \rightarrow \infty} M_{ij}(t)/t = 1/\bar{W}(j) \quad \text{WP1}$$

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{M_{ij}(t)}{M_i(t)} &= \lim_{t \rightarrow \infty} \frac{N_{ij}(M_i(t))}{M_i(t)} \\ &= \lim_{n \rightarrow \infty} \frac{N_{ij}(n)}{n} = \pi_j \quad \text{WP1.} \end{aligned}$$

$$\begin{aligned} \frac{1}{\bar{W}(j)} &= \lim_{t \rightarrow \infty} \frac{M_{ij}(t)}{t} = \lim_{t \rightarrow \infty} \frac{M_{ij}(t)}{M_i(t)} \frac{M_i(t)}{t} \\ &= \pi_j \lim_{t \rightarrow \infty} \frac{M_i(t)}{t} = p_j \nu_j \end{aligned}$$

This shows that $\lim_t M_i(t)/t$ is independent of i .

14

$$p_j = \frac{1}{\nu_j \bar{W}(j)} = \frac{\pi_j}{\nu_j} \lim_{t \rightarrow \infty} \frac{M_i(t)}{t} \quad \text{WP1.}$$

Thm: If the embedded chain is positive recurrent, then

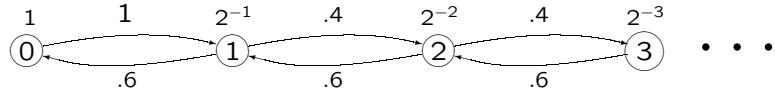
$$p_j = \frac{\pi_j / \nu_j}{\sum_k \pi_k / \nu_k}; \quad \lim_{t \rightarrow \infty} \frac{M_i(t)}{t} = \frac{1}{\sum_k \pi_k / \nu_k} \quad \text{WP1}$$

If $\sum_k \pi_k / \nu_k < \infty$, this is almost obvious except for mathematical details. We can interpret $\lim_t M_i(t)/t$ as the transition rate of the process, and it must have the given value so that $\sum_j p_j = 1$.

It is possible to have $\sum_k \pi_k / \nu_k = \infty$. This suggests that the rate of transitions is 0.

15

Case where $\sum_k \pi_k / \nu_k = \infty$



This can be viewed as a queue where the server becomes increasingly rattled and the customers increasingly discouraged as the state increases.

We have $\pi_j = (1 - \rho)\rho^j$ for $\rho = 2/3$. Thus

$$\pi_j / \nu_j = 2^j (1 - \rho)\rho^j = (1 - \rho)(4/3)^j$$

By truncating the chain, it can be verified that the service rate approaches 0 as more states are added.

16

Again assume the typical case of a positive recurrent embedded chain with $\sum_i \pi_i / \nu_i < \infty$. Then

$$p_j = \frac{\pi_j / \nu_j}{\sum_k \pi_k / \nu_k} \quad (1)$$

We can solve these directly using the steady-state embedded equations:

$$\begin{aligned} \pi_j &= \sum_i \pi_i P_{ij}; \quad \pi_i > 0; \quad \sum_i \pi_i = 1 \\ p_j \nu_j &= \sum_i p_i q_{ij}; \quad p_j > 0; \quad \sum_j p_j = 1 \end{aligned} \quad (2)$$

$$\pi_j = \frac{p_j \nu_j}{\sum_i p_i \nu_i} \quad (3)$$

Thm: If embedded chain is positive recurrent and $\sum_i \pi_i / \nu_i < \infty$, then (2) has unique solution, $\{p_j\}$ and $\{\pi_j\}$ are related by (1) and (3), and

$$\sum_i \pi_i / \nu_i = (\sum_i p_j \nu_j)^{-1}$$

17

We can go the opposite way also. If

$$p_j \nu_j = \sum_i p_i q_{ij}; \quad p_j > 0; \quad \sum_j p_j = 1$$

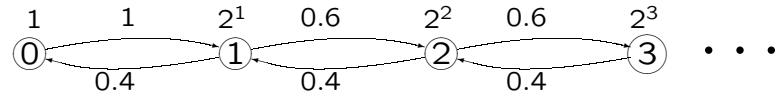
and if $\sum_j p_j \nu_j < \infty$, then $\pi_j = p_j \nu_j / (\sum_j p_j \nu_j)$ gives the steady-state equations for the embedded chain and the embedded chain is positive recurrent.

If ν_j is bounded over j , then $\sum_j p_j \nu_j < \infty$. Also the sampled-time chain exists and has the same steady-state solution.

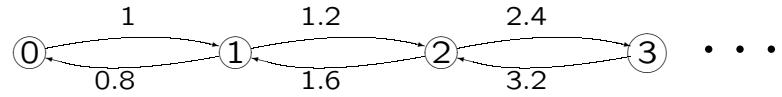
For a birth/death process, we also have $p_i q_{i,1+1} = p_{i+1} q_{i+1,i}$.

If $\sum_j p_j \nu_j = \infty$, then $\pi_j = 0$ for all j and the embedded chain is transient or null-recurrent. In the transient case, there can be infinitely many transitions in finite time, so the notion of steady-state doesn't make much sense.

18



Imbedded chain for hyperactive birth/death



Same process in terms of $\{q_{ij}\}$

There is a nice solution for p_j , but the imbedded chain is transient.

These chains are called irregular. The expected number of transitions per unit time is infinite, and they don't make much sense.

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

L20: Markov processes and Random Walks

Outline:

- Review - Steady state for MP
- Reversibility for Markov processes
- Random walks
- Queueing delay in a G/G/1 queue
- Detection, decisions, & Hypothesis testing

1

If the embedded chain of a MP is positive recurrent, then

$$p_j = \frac{\pi_j/\nu_j}{\sum_k \pi_k/\nu_k}; \quad \lim_{t \rightarrow \infty} \frac{M_i(t)}{t} = \frac{1}{\sum_k \pi_k/\nu_k} \quad \text{WP1}$$

where $M_i(t)$ is the sample-path average rate at which transitions occur WP1 and p_j is the sample-path average fraction of time in state j WP1, independent of starting state.

If $\sum_k \pi_k/\nu_k = \infty$, the transition rate $M_i(t)/t \rightarrow 0$ and the process has no meaningful steady state. Otherwise the steady state uniquely satisfies

$$p_j \nu_j = \sum_i p_i q_{ij}; \quad p_j > 0; \quad \text{all } j; \quad \sum_j p_j = 1$$

This says that rate in equals rate out for each state.
For birth/death, $p_j q_{j,j+1} = p_{j+1} q_{j+1,j}$.

2

For an irreducible process, if there is a solution to the equations

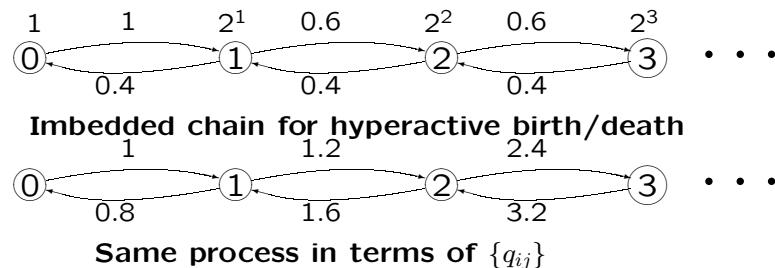
$$p_j \nu_j = \sum_i p_i q_{ij}; \quad p_j > 0; \quad \text{all } j; \quad \sum_j p_j = 1$$

and if $\sum_i \nu_i p_i < \infty$, then the embedded chain is positive recurrent and

$$\pi_j = \frac{p_j \nu_j}{\sum_i p_i \nu_i}; \quad \sum_i \pi_i / \nu_i = (\sum_i p_j \nu_j)^{-1}$$

If $\sum_i \nu_i p_i = \infty$, then each $\pi_j = 0$, the embedded chain is either transient or null-recurrent, and the notion of steady-state makes no sense.

3



Using $p_j q_{j,j+1} = p_{j+1} q_{j+1,j}$, we see that $p_{j+1} = \frac{3}{4} p_j$, so

$$p_j = (1/4) (3/4)^j \quad \text{and} \quad \sum_j p_j \nu_j = \infty.$$

If we truncate this process to k states, then

$$\begin{aligned} p_j &= \frac{1}{4} \left(1 - \left(\frac{3}{4} \right)^k \right) \left(\frac{3}{4} \right)^j; \quad \pi_j = \frac{1}{3} \left(1 - \left(\frac{2}{3} \right)^k \right) \left(\frac{2}{3} \right)^{k-j} \\ \sum_j p_j \nu_j &= \frac{1}{2} \left(1 - \left(\frac{3}{4} \right)^k \right) \left(\left(\frac{3}{2} \right)^k - 1 \right) \rightarrow \infty \end{aligned}$$

4

Reversibility for Markov processes

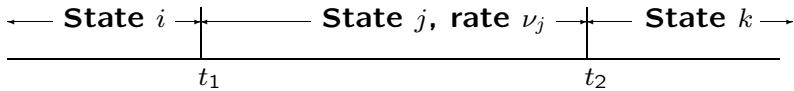
For any Markov chain in steady state, the backward transition probabilities P_{ij}^* are defined as

$$\pi_i P_{ij}^* = \pi_j P_{ji}$$

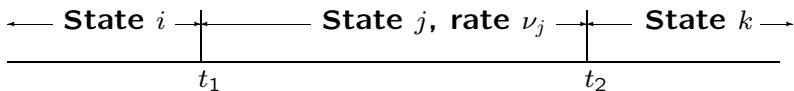
There is nothing mysterious here, just

$$\begin{aligned}\Pr\{X_n = j, X_{n+1} = i\} &= \Pr\{X_{n+1} = i\} \Pr\{X_n = j | X_{n+1} = i\} \\ &= \Pr\{X_n = j\} \Pr\{X_{n+1} = i | X_n = j\}\end{aligned}$$

This also holds for the embedded chain of a Markov process.



5



Moving right, after entering state j , the exit rate is ν_j , i.e., we exit in each δ with probability $\nu_j \delta$. The same holds moving left.

That is, a Poisson process is clearly reversible from the incremental definition.

Thus $\{\pi_i\}$ and $\{\nu_i\}$ are the same going left as going right

6

Note that the probability of having a (right) transition from state j to k in $(t, t+\delta)$ is $p_j q_{jk} \delta$. Similarly, if q_{kj}^* is the left-going process transition rate, the probability of having the same transition is $p_k q_{kj}^*$. Thus

$$p_j q_{jk} = p_k q_{kj}^*$$

By fiddling equations, $q_{kj}^* = \nu_k P_{kj}^*$.

Def: A MP is reversible if $q_{ij}^* = q_{ji}$ for all i, j

Assuming positive recurrence and $\sum_i \pi_i / \nu_i < \infty$, the MP process is reversible if and only if the embedded chain is.

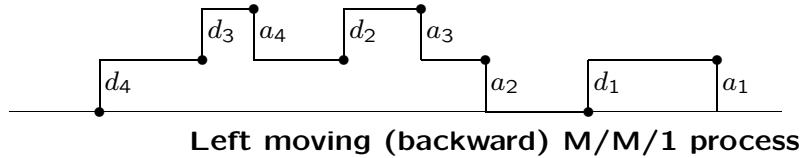
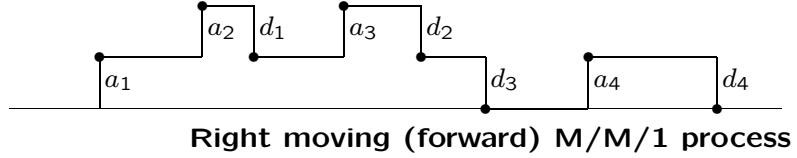
7

The guessing theorem: Suppose a MP is irreducible and $\{p_i\}$ is a set of probabilities that satisfies $p_i q_{ij} = p_j q_{ji}$ for all i, j and satisfies $\sum_i p_i \nu_i < \infty$.

Then (1), $p_i > 0$ for all i , (2), p_i is the sample-path fraction of time in state i WP1, (3), the process is reversible, and (4), the embedded chain is positive recurrent.

Useful application: All birth/death processes with $\sum_j p_j \nu_j < \infty$ are reversible. Similarly, if the Markov graph is a tree with $\sum_j p_j \nu_j < \infty$, the process is reversible.

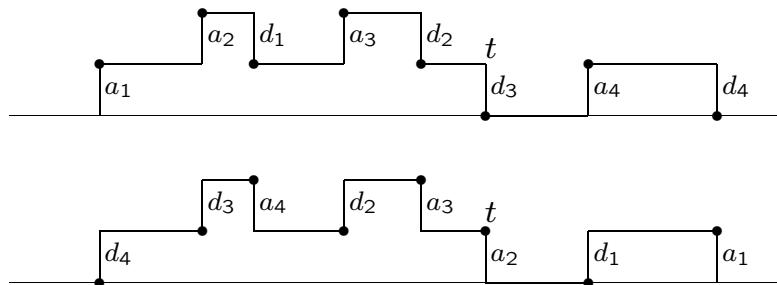
8



Burke's thm: Given an M/M/1 queue in steady-state with (arrival rate) $\lambda < \mu$ (departure rate),

- (1) Departure process is Poisson with rate λ
- (2) State $X(t)$ is independent of departures before t
- (3) For FCFS, a customer's arrival time, given its departure at t , is independent of departures before t .

9

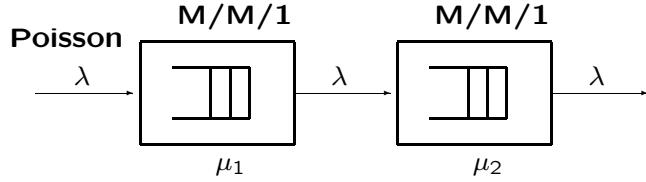


A departure at t in (right-moving) sample path is an arrival in the M/M/1 (left-moving) sample path.

For FCFS left-moving process, departure time of arrival at t depends on arrivals (and their service req.) to the right of t ; independent of arrivals to left.

For corresponding (right-moving) process, the arrival time of that departure is independent of departures before t .

10



Consider tandem $M/M/1$ queues. Departures from first are Poisson with rate λ . Assume service times at rates μ_1 and μ_2 , independent from queue to queue and independent of arrivals at each.

Arrivals at queue 2 are Poisson at rate λ by Burke and are independent of service times at 2. Thus the second queue is $M/M/1$.

The states of the two systems are independent and the time of a customer in system 1 is independent of that in 2.

Random walks

Def: Let $\{X_i; i \geq 1\}$ be a sequence of IID rv's, and let $S_n = X_1 + X_2 + \dots + X_n$ for $n \geq 1$. The integer-time stochastic process $\{S_n; n \geq 1\}$ is called a random walk, or, specifically, the random walk based on $\{X_i; i \geq 1\}$.

We are used to sums of IID rv's, but here the interest is in the process. We ask such questions as:

- 1) Threshold crossing: For given $\alpha > 0$, what is the probability that $S_n \geq \alpha$ for at least one $n \geq 1$; what is the smallest n for which this crossing happens; and what is the overshoot $S_n - \alpha$?
- 2) Two thresholds: For given $\alpha > 0, \beta < 0$, what is the probability that $\{S_n; n \geq 1\}$ crosses α before it crosses β , and what is the n at which the first such crossing occurs?

These threshold-crossing problems are important in studying overflow in queues, errors in digital communication systems, hypothesis testing, ruin and other catastrophes, etc.

In many of the important applications, the relevant probabilities are very small and the problems are known as large deviation problems.

Moment generating functions and their use in upper bounds on these small probabilities are important here.

We start with a brief discussion of 3 simple cases: simple random walks, integer random walks, and renewal processes.

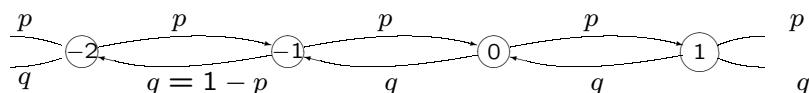
13

Simple random walks

A random walk (RW) $\{S_n; n \geq 1\}$, $S_n = X_1 + \dots + X_n$ is simple if X_n is binary with $p_X(1) = p$, $p_X(-1) = q = 1-p$. This is just a scaling variation on a Bernoulli process. The probability that $X_i = 1$ for m out of n trials is

$$\Pr\{S_n = 2m - n\} = \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m}.$$

Viewed as a Markov chain,



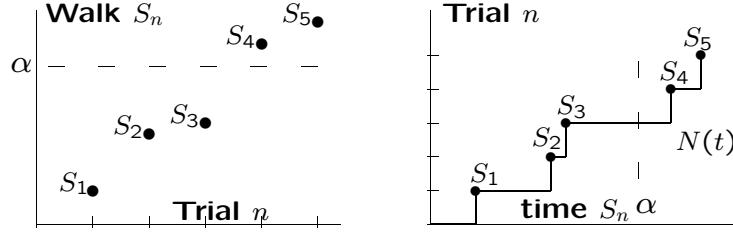
As in ‘stop when you’re ahead’

$$\Pr\left\{\bigcup_{n=1}^{\infty} \{S_n \geq k\}\right\} = \left(\frac{p}{1-p}\right)^k \quad \text{if } p \leq 1/2.$$

14

Integer RW's (where X is an integer rv) are similar. An integer RW can also be modeled as a Markov chain, but there might be an overshoot when crossing a threshold and the analysis is much harder.

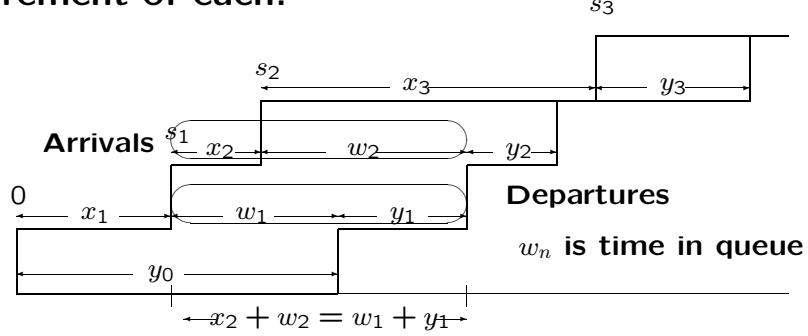
Renewal processes are also special cases of random walks where X is a positive rv. When sketching sample paths, the axes are usually reversed from RP to RW.



15

Queueing delay in a G/G/1 queue

Let $\{X_i; i \geq 1\}$ be the (IID) interarrival intervals of a G/G/1 queue and let $\{Y_i; i \geq 1\}$ be the (IID) service requirement of each.



If arrival n is queued (e.g., arrival 2 above), then

$$x_n + w_n = y_{n-1} + w_{n-1}$$

If arrival n sees an empty queue, then $w_n = 0$.

16

$$w_n = y_{n-1} - x_n + w_{n-1} \text{ if } w_{n-1} + y_{n-1} \geq x_n \text{ else } w_n = 0$$

$$w_n = \max[w_{n-1} + y_{n-1} - x_n, 0]$$

Since this is true for all sample paths,

$$W_n = \max[W_{n-1} + Y_{n-1} - X_n, 0]$$

Define $U_n = Y_{n-1} - X_n$. **Then**

$$W_n = \max[W_{n-1} + U_n, 0]$$

Without the max, $\{W_n; n \geq 1\}$ **would be a random walk based on** $\{U_i; i \geq 1\}$.

With the max, $\{W_n; n \geq 1\}$ **is like a random walk, but it resets to 0 every time it goes negative.** The text restates this in an alternative manner.

Detection, decisions, & Hypothesis testing

These are different names for the same thing. Given observations, a decision must be made between a set of alternatives.

Here we consider only binary decisions, i.e., a choice between two hypotheses.

Consider a sample space containing a rv H (the hypothesis) with 2 possible values, $H = 0$ and $H = 1$. The PMF for H , $p_H(0) = p_0$, $p_H(1) = p_1$, is called the a priori probabilities of H .

Assume n observations, Y_1, \dots, Y_n are made. These are IID conditional on $H = 0$ and IID conditional on $H = 1$. Assume a pdf

$$f_{\vec{Y}|H}(\vec{y} | \ell) = \prod_{i=1}^n f_{Y|H}(y_i | \ell).$$

By Baye's law,

$$\Pr\{H=\ell \mid \vec{y}\} = \frac{p_\ell f_{\vec{Y}|H}(\vec{y} \mid \ell)}{p_0 f_{\vec{Y}|H}(\vec{y} \mid 0) + p_1 f_{\vec{Y}|H}(\vec{y} \mid 1)}.$$

Comparing $\Pr\{H=0 \mid \vec{y}\}$ **and** $\Pr\{H=1 \mid \vec{y}\}$,

$$\frac{\Pr\{H=0 \mid \vec{y}\}}{\Pr\{H=1 \mid \vec{y}\}} = \frac{p_0 f_{\vec{Y}|H}(\vec{y} \mid 0)}{p_1 f_{\vec{Y}|H}(\vec{y} \mid 1)}.$$

The probability that $H = \ell$ is the correct hypothesis, given the observation, is $\Pr\{H=\ell \mid \vec{Y}\}$. Thus we maximize the a posteriori probability of choosing correctly by choosing the maximum over ℓ of $\Pr\{H=\ell \mid \vec{Y}\}$. This is called the MAP rule (maximum a posteriori probability). It requires knowing p_0 and p_1 .

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

L21: Hypothesis testing and Random Walks**Outline:**

- **Random walks**
- **Detection, decisions, & Hypothesis testing**
- **Threshold tests and the error curve**
- **Thresholds for random walks and Chernoff**

1

Random walks

Def: Let $\{X_i; i \geq 1\}$ be a sequence of IID rv's, and let $S_n = X_1 + X_2 + \dots + X_n$ for $n \geq 1$. The integer-time stochastic process $\{S_n; n \geq 1\}$ is called a **random walk**, or, specifically, the **random walk based on** $\{X_i; i \geq 1\}$.

Our focus will be on threshold-crossing problems. For example, if X is binary with $p_X(1) = p$, $p_X(-1) = q = 1 - p$, then

$$\Pr \left\{ \bigcup_{n=1}^{\infty} \{S_n \geq k\} \right\} = \left(\frac{p}{1-p} \right)^k \quad \text{if } p \leq 1/2.$$

2

Detection, decisions, & Hypothesis testing

The model here contains a discrete, usually binary, rv H called the hypothesis rv. The sample values of H , say 0 and 1, are called the alternative hypotheses and have marginal probabilities, called a priori probabilities $p_0 = \Pr\{H = 0\}$ and $p_1 = \Pr\{H = 1\}$.

Among arbitrarily many other rv's, there is a sequence $\vec{Y}^{(m)} = (Y_1, Y_2, \dots, Y_m)$ of rv's called the observation. We usually assume that Y_1, Y_2, \dots , are IID conditional on $H = 0$ and IID conditional on $H = 1$. Thus, if the Y_n are continuous,

$$f_{\vec{Y}^{(m)}|H}(\vec{y} | \ell) = \prod_{n=1}^m f_{Y|H}(y_n | \ell).$$

3

Assume that, on the basis of observing a sample value \vec{y} of \vec{Y} , we must make a decision about H , i.e., choose $H = 0$ or $H = 1$, i.e., detect whether or not H is 1.

Decisions in probability theory, as in real life, are not necessarily correct, so we need a criterion for making a choice.

We might maximize the probability of choosing correctly, for example, or, given a cost for the wrong choice, might minimize the expected cost.

Note that the probability experiment here includes not only the experiment of gathering data (i.e., measuring the sample value \vec{y} of \vec{Y}) but also the sample value of the hypothesis.

4

From Bayes', recognizing that $f(\vec{y}) = p_0 f(\vec{y}|0) + p_1 f(\vec{y}|1)$

$$\Pr\{H=\ell \mid \vec{y}\} = \frac{p_\ell f_{\vec{Y}|H}(\vec{y} \mid \ell)}{p_0 f_{\vec{Y}|H}(\vec{y} \mid 0) + p_1 f_{\vec{Y}|H}(\vec{y} \mid 1)}.$$

Comparing $\Pr\{H=0 \mid \vec{y}\}$ **and** $\Pr\{H=1 \mid \vec{y}\}$,

$$\frac{\Pr\{H=0 \mid \vec{y}\}}{\Pr\{H=1 \mid \vec{y}\}} = \frac{p_0 f_{\vec{Y}|H}(\vec{y} \mid 0)}{p_1 f_{\vec{Y}|H}(\vec{y} \mid 1)}.$$

The probability that $H = \ell$ is the correct hypothesis, given the observation, is $\Pr\{H=\ell \mid \vec{Y}\}$. Thus we maximize the a posteriori probability of choosing correctly by choosing the maximum over ℓ of $\Pr\{H=\ell \mid \vec{Y}\}$.

This is called the **MAP rule (maximum a posteriori probability)**. It requires knowing p_0 and p_1 .

5

The MAP rule (and other decision rules) are clearer if we define the likelihood ratio,

$$\Lambda(\vec{y}) = \frac{f_{\vec{Y}|H}(\vec{y} \mid 0)}{f_{\vec{Y}|H}(\vec{y} \mid 1)}.$$

The MAP rule is then

$$\Lambda(\vec{y}) \quad \begin{cases} > p_1/p_0 & ; \quad \text{select } \hat{h}=0 \\ \leq p_1/p_0 & ; \quad \text{select } \hat{h}=1. \end{cases}$$

Many decision rules, including the most common and the most sensible, are rules that compare $\Lambda(\vec{y})$ to a fixed threshold, say η , independent of \vec{y} . Such decision rules vary only in the way that η is chosen.

Example: For maximum likelihood, the threshold is 1 (this is MAP for $p_0 = p_1$, but it is also used in other ways).

6

Back to random walks: Note that the logarithm of the threshold ratio is given by

$$\ln \Lambda(\vec{y}^{(m)}) = \sum_{n=1}^m \Lambda(y_n); \quad \Lambda(y_n) = \ln \left(\frac{f_{Y|H}(y_n|0)}{f_{Y|H}(y_n|1)} \right)$$

Note that $\Lambda(y_n)$ is a real-valued function of y_n , and is the same function for each n . Thus, since Y_1, Y_2, \dots , are IID rv's conditional on $H = 0$ (or $H = 1$), $\Lambda(Y_1), \Lambda(Y_2)$, are also IID conditional on $H = 0$ (or $H = 1$).

It follows that $\ln \Lambda(\vec{y}^{(m)})$, conditional on $H = 0$ (or $H = 1$) is a sum of m IID rv's and $\{\ln \Lambda(\vec{y}^{(m)}) ; m \geq 1\}$ is a random walk conditional on $H = 0$ (or $H = 1$). The two random walks contain the same sequence of sample values but different probability measures.

Later we look at sequential detection, where observations are made until a threshold is passed.

7

Threshold tests and the error curve

A general hypothesis testing rule (a test) consists of mapping each sample sequence \vec{y} into either 0 or 1. Thus a test can be viewed as the set A of sample sequences mapped into hypothesis 1. The error probability, given $H = 0$ or $H = 1$, using test A , is given by

$$q_0(A) = \Pr\{Y \in A \mid H = 0\}; \quad q_1(A) = \Pr\{Y \in A^c \mid H = 1\}$$

With a priori probabilities p_0, p_1 and $\eta = p_1/p_0$,

$$\Pr\{\mathbf{e}(A)\} = p_0 q_0(A) + p_1 q_1(A) = p_0 [q_0(A) + \eta q_1(A)]$$

For the threshold test based on η ,

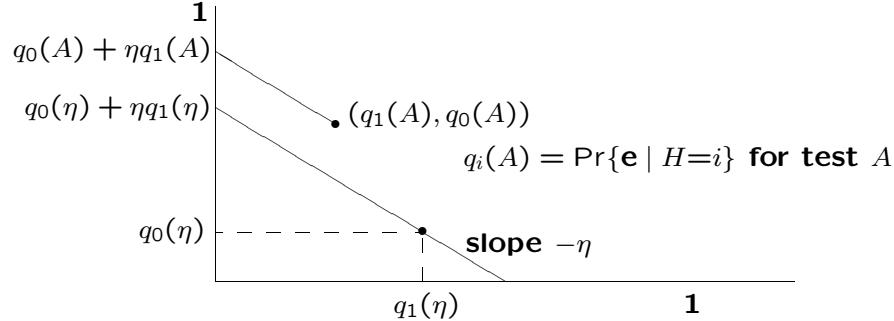
$$\Pr\{\mathbf{e}(\eta)\} = p_0 q_0(\eta) + p_1 q_1(\eta) = p_0 [q_0(\eta) + \eta q_1(\eta)]$$

$$q_0(\eta) + \eta q_1(\eta) \leq q_0(A) + \eta q_1(A); \quad \text{by MAP}$$

8

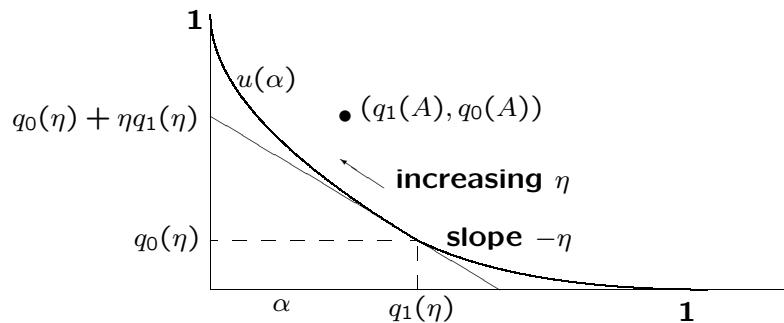
$$q_0(\eta) + \eta q_1(\eta) \leq q_0(A) + \eta q_1(A); \quad \text{by MAP}$$

Note that the point $(q_0(A), q_1(A))$ does not depend on p_0 : the a priori probabilities were simply used to prove the above inequality.



For every A and every η , $(q_0(A), q_1(A))$ lies NorthEast of the line of slope $-\eta$ through $(q_0(\eta), q_1(\eta))$. Thus $(q_0(A), q_1(A))$ is NE of the upper envelope of these straight lines.

9



If the vertical axis of the error curve is inverted, it is called a receiver operating curve (ROC) which is a staple of radar system design.

The Neyman-Pearson test is a test that chooses A to minimize $q_1(A)$ for a given constraint on $q_0(A)$. Typically this is a threshold test, but sometimes, especially if Y is discrete, it is a randomized threshold test.

10

Thresholds for random walks and Chernoff bounds

The Chernoff bound says that for any real b and any r such that $g_Z(r) = E[e^{rZ}]$ exists,

$$\begin{aligned}\Pr\{Z \geq b\} &\leq g_Z(r) \exp(-rb); & \text{for } b > \bar{Z}, r > 0 \\ \Pr\{Z \leq b\} &\leq g_Z(r) \exp(-rb); & \text{for } b < \bar{Z}, r < 0\end{aligned}$$

This is most useful when applied to a sum, $S_n = X_1 + \dots + X_n$ of IID rv's. If $g_X(r) = E[e^{rX}]$ exists, then

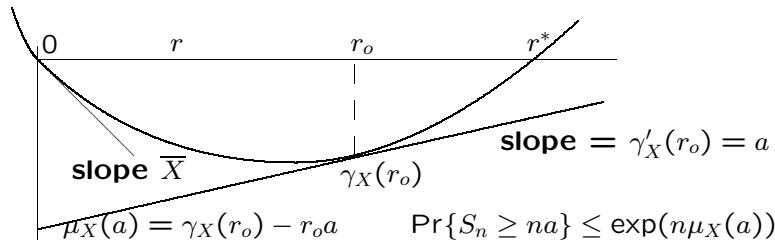
$$E[e^{rS_n}] = E\left[\prod_{i=1}^n e^{rX_i}\right] = g_X^n(r)$$

$$\begin{aligned}\Pr\{S_n \geq na\} &\leq g_X^n(r) \exp(-rna); & \text{for } a > \bar{X}, r > 0 \\ \Pr\{S_n \leq na\} &\leq g_X^n(r) \exp(-rna); & \text{for } a < \bar{X}, r < 0\end{aligned}$$

11

This is easier to interpret and work with if expressed in terms of the semi-invariant MGF, $\gamma_X(r) = \ln g_X(r)$. Then $g_X^n(r) = e^{n\gamma_X(r)}$ and

$$\begin{aligned}\Pr\{S_n \geq na\} &\leq \exp(n[\gamma_X(r) - ra]); & \text{for } a > \bar{X}, r > 0 \\ \Pr\{S_n \leq na\} &\leq \exp(n[\gamma_X(r) - ra]); & \text{for } a < \bar{X}, r < 0\end{aligned}$$



The Chernoff bound, optimized over r , is essentially exponentially tight; i.e., $\Pr\{S_n \geq na\} \geq \exp(n(\mu_X(a) - \epsilon))$ for large enough n .

12

In looking at threshold problems, we want to find the probability that $\Pr\{S_n \geq \alpha\}$ for any n . Thus we want a bound that focuses on variable n for a fixed α , i.e., on when the threshold is crossed if it is crossed.

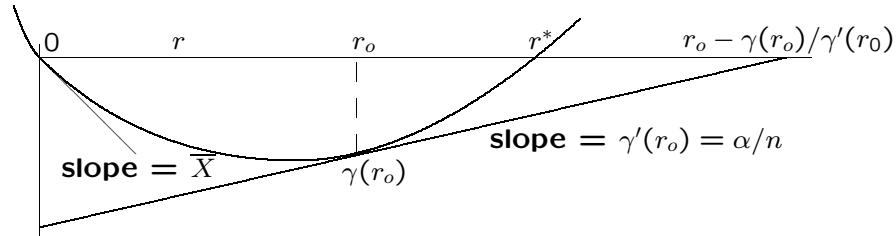
We want a bound of the form $\Pr\{S_n \geq \alpha\} \leq \exp \alpha f(n)$

Start with the bound $\Pr\{S_n \geq na\} \leq \exp(n[\gamma_X(r_0) - r_0 a])$, with $\alpha = an$ and r_0 such that $\gamma'_X(r_0) = \alpha/n$. Substituting $\alpha/\gamma'_X(r_0)$ for n ,

$$\Pr\{S_n \geq \alpha\} \leq \exp\left(\alpha \left[\frac{\gamma_X(r_0)}{\gamma'_X(r_0)} - r_0 \right]\right)$$

13

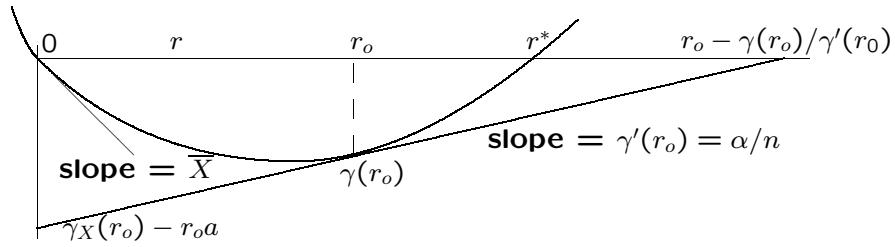
$$\Pr\{S_n \geq \alpha\} \leq \exp\left(\alpha \left[\frac{\gamma_X(r_0)}{\gamma'_X(r_0)} - r_0 \right]\right)$$



When n is very large, the slope $\gamma'_X(r_0)$ is close to 0 and the horizontal intercept (the negative exponent) is very large. As n decreases, the intercept decreases to r^* and then increases again.

Thus $\Pr\{\cup_n S_n \geq \alpha\} \approx \exp(-\alpha r^*)$, where the nature of the approximation remains to be explained.

14



Example: $p_X(1) = p$, $p_X(-1) = 1-p$; $p < 1/2$. Then $g_X(r) = pe^r + (1-p)e^{-r}$; $\gamma_X(r) = \ln[pe^r + (1-p)e^{-r}]$

Since $\gamma_X(r^*) = 0$, we have $pe^{r^*} + (1-p)e^{-r^*} = 1$. Letting $z = e^{r^*}$, this is $pz + (1-p)/z = 1$ so z is either 1 or $(1-p)/p$. Thus $r^* = \ln(1-p)/p$ and

$$\Pr\left\{\bigcup_n S_n \geq \alpha\right\} \approx \exp(-\alpha r^*) = \left(\frac{1-p}{p}\right)^{-\alpha}$$

which is exact for α integer. The bound for individual n is the exponent in the Gaussian approximation.

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

L22: Random Walks and thresholds

Outline:

- Review of Chernoff bounds
- Wald's identity with 2 thresholds
- The Kingman bound for G/G/1
- Large deviations for hypothesis tests
- Sequential detection
- Tilted probabilities and proof of Wald's id.

1

Let a rv Z have an MGF $g_Z(r)$ for $0 \leq r < r_+$ and mean $\bar{Z} < 0$. By the Chernoff bound, for any $\alpha > 0$ and any $r \in (0, r_+)$,

$$\Pr\{Z \geq \alpha\} \leq g_Z(r) \exp(-r\alpha) = \exp(\gamma_Z(r) - r\alpha)$$

where $\gamma_Z(r) = \ln g_Z(r)$. If Z is a sum $S_n = X_1 + \dots + X_n$, of IID rv's, then $\gamma_{S_n}(r) = n\gamma_X(r)$.

$$\Pr\{S_n \geq na\} \leq \min_r (\exp[n(\gamma_X(r) - ra)]).$$

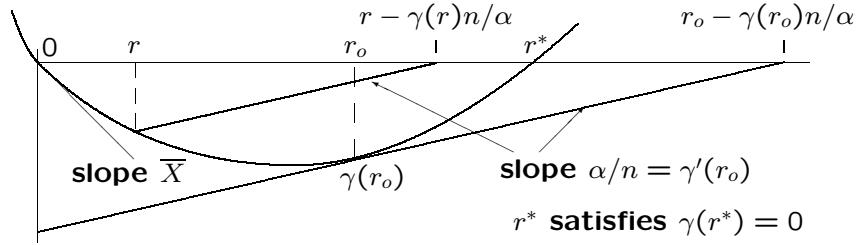
This is exponential in n for fixed a (i.e., $\gamma'(r) = a$). We are now interested in threshold crossings, i.e., $\Pr\{\cup_n(S_n \geq \alpha)\}$. As a preliminary step, we study how $\Pr\{S_n \geq \alpha\}$ varies with n for fixed α .

$$\Pr\{S_n \geq \alpha\} \leq \min_r (\exp[n\gamma_X(r) - r\alpha]).$$

Here the minimizing r varies with n (i.e., $\gamma'(r) = \alpha/n$).

2

$$\Pr\{S_n \geq \alpha\} \leq \min_{0 < r < r_+} \exp\left(-\alpha \left[r - \frac{n}{\alpha} \gamma_X(r)\right]\right)$$



When n is very large, the slope $\frac{\alpha}{n} = \gamma'_X(r_0)$ is close to 0 and the horizontal intercept (the negative exponent) is very large. As n decreases, the intercept decreases to r^* and then increases again.

Thus $\Pr\{\cup_n\{S_n \geq \alpha\}\} \approx \exp(-\alpha r^*)$, where the nature of the approximation will be explained in terms of the Wald identity.

3

Wald's identity with 2 thresholds

Consider a random walk $\{S_n; n \geq 1\}$ with $S_n = X_1 + \dots + X_n$ and assume that X is not identically zero and has a semi-invariant MGF $\gamma(r)$ for $r \in (r_-, r_+)$ with $r_- < 0 < r_+$. Let $\alpha > 0$ and $\beta < 0$ be two thresholds. Let J be the smallest n for which either $S_n \geq \alpha$ or $S_n \leq \beta$.

Note that J is a stopping trial, i.e., $\mathbb{I}_{J=n}$ is a function of S_1, \dots, S_n and J is a rv. The fact that J is a rv is proved in Lemma 7.5.1, but is almost obvious.

Wald's identity now says that for any r , $r_- < r < r_+$,

$$E[\exp(rS_J - J\gamma(r))] = 1.$$

If we replace J by a fixed step n , this just says that $E[\exp(rS_n)] = \exp(n\gamma(r))$, so this is not totally implausible.

4

$$E[\exp(rS_J - J\gamma(r))] = 1 \quad (\text{Wald's identity}).$$

Before justifying this, we use it to bound the probability of crossing a threshold.

Corollary: Assume further that $\bar{X} < 0$ and that $r^* > 0$ exists such that $\gamma(r^*) = 0$. Then

$$\Pr\{S_J \geq \alpha\} \leq \exp(-r^*\alpha).$$

Wald's id. at r^* is $E[\exp(r^*S_J)] = 1$. Since $\exp(r^*S_J) \geq 0$,

$$\Pr\{S_J \geq \alpha\} E[\exp(r^*S_J) | S_J \geq \alpha] \leq E[\exp(r^*S_J)] = 1.$$

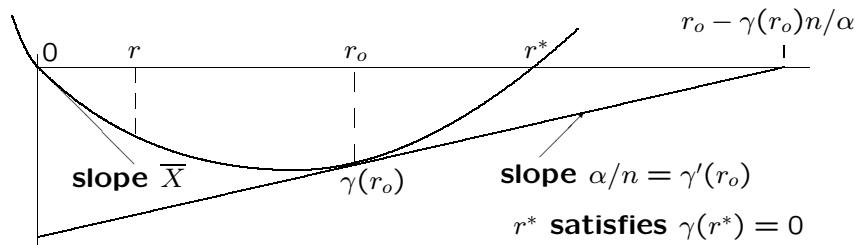
For $S_J \geq \alpha$, we have $\exp(r^*S_J) \geq \exp(r^*\alpha)$. Thus

$$\Pr\{S_J \geq \alpha\} \exp(r^*\alpha) \leq 1.$$

This is valid for all choices of $\beta < 0$, so it turns out to be valid without a lower threshold, i.e., $\Pr\{\bigcup_n\{S_n \geq \alpha\}\} \leq \exp(-r^*\alpha)$.

5

We saw before that $\Pr\{S_n \geq \alpha\} \leq \exp(-\alpha r^*)$ for all n , but this corollary makes the stronger and cleaner statement that $\Pr\{\bigcup_{n \geq 1}\{S_n \geq \alpha\}\} \leq \exp(-r^*\alpha)$



The Chernoff bound has the advantage of showing that the n for which the probability of threshold crossing is essentially highest is $n = \alpha/\gamma'(r^*)$.

6

The Kingman bound for G/G/1

The corollary can be applied to the queueing time W_i for the i th arrival to a G/G/1 system.

We let $U_i = X_i - Y_{i-1}$, i.e., U_i is the difference between the i th interarrival time and the previous service time.

Recall that we showed that $\{U_i; i \geq 1\}$ is a modification of a random walk. The text shows that it is a random walk looking backward.

Letting $\gamma(r)$ be the semi-invariant MGF of each U_i , then the Kingman bound (the corollary to the Wald identity for the G/G/1 queue) says that for all $n \geq 1$,

$$\Pr\{W_n \geq \alpha\} \leq \Pr\{W \geq \alpha\} \leq \exp(-r^* \alpha); \quad \text{for all } \alpha > 0.$$

7

Large deviations for hypothesis tests

Let $\vec{Y} = (Y_1, \dots, Y_n)$ be IID conditional on \mathbf{H}_0 and also IID conditional on \mathbf{H}_1 . Then

$$\ln(\Lambda(\vec{y})) = \ln \frac{f(\vec{y} | \mathbf{H}_0)}{f(\vec{y} | \mathbf{H}_1)} = \sum_{i=1}^n \ln \frac{f(y_i | \mathbf{H}_0)}{f(y_i | \mathbf{H}_1)}$$

$$\text{Define } z_i \text{ by } z_i = \ln \frac{f(y_i | \mathbf{H}_0)}{f(y_i | \mathbf{H}_1)}$$

A threshold test compares $\sum_{i=1}^n z_i$ with $\ln(\eta) = \ln(p_1/p_0)$.

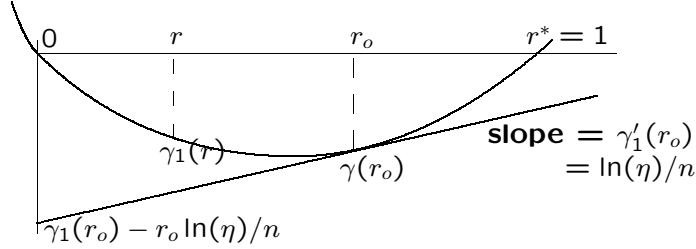
Conditional on \mathbf{H}_1 , make error if $\sum_i Z_i^1 > \ln(\eta)$ where Z_i^1 , $1 \leq i \leq n$, are IID conditional on \mathbf{H}_1 .

8

Exponential bound for $\sum_i Z_i^1$

$$\begin{aligned}\gamma_1(r) &= \ln \left\{ \int f(y | \mathbf{H}_1) \exp \left[r \ln \frac{f(y | \mathbf{H}_0)}{f(y | \mathbf{H}_1)} \right] dy \right\} \\ &= \ln \left\{ \int f^{1-r}(y | \mathbf{H}_1) f^r(y | \mathbf{H}_0) dy \right\}\end{aligned}$$

At $r = 1$, this is $\ln(\int f(y | \mathbf{H}_0) dy) = 0$.



$$q_1(\eta) \leq \exp n [\gamma_1(r_0) - r_0 \ln(\eta)/n]$$

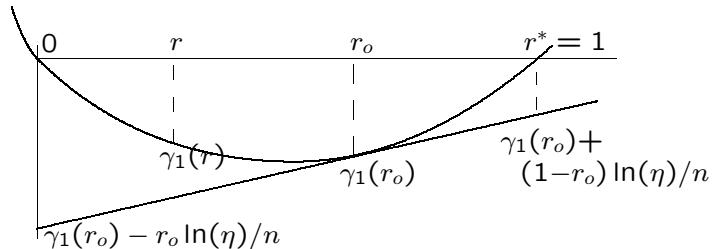
where $q_\ell(\eta) = \Pr\{e | H = \ell\}$

9

Exponential bound for $\sum_i Z_i^0$

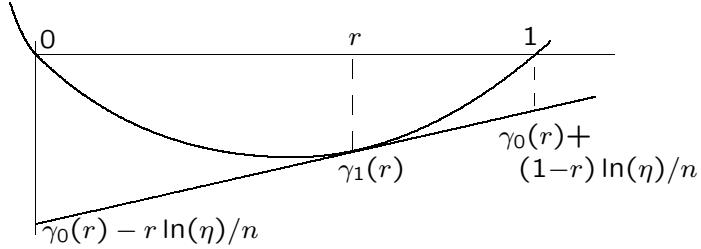
$$\begin{aligned}\gamma_0(s) &= \ln \left\{ \int f(y | 0) \exp \left[s \ln \frac{f(y | \mathbf{H}_0)}{f(y | \mathbf{H}_1)} \right] dy \right\} \\ &= \ln \left\{ \int f^{-s}(y | \mathbf{H}_1) f^{1+s}(y | \mathbf{H}_0) dy \right\}\end{aligned}$$

At $s = -1$, this is $\ln(\int f(y | \mathbf{H}_1) dy) = 0$. Note: $\gamma_0(s) = \gamma_1(r-1)$.



$$q_0(\eta) \leq \exp n [\gamma_1(r_0) + (1-r_0) \ln(\eta)/n]$$

10



These are the exponents for the two kinds of errors. This can be viewed as a large deviation form of Neyman Pearson. Choose one exponent and the other is given by the inverted see-saw above.

The a priori probabilities are usually not the essential characteristic here, but the bound for MAP is optimized at r such that $\ln(\eta)/n - \gamma'_0(r)$

11

Sequential detection

This large-deviation hypothesis-testing problem screams out for a variable number of trials.

We have two coupled random walks, one based on H_0 and one on H_1 .

We use two thresholds, $\alpha > 0$ and $\beta < 0$. Note that $E[Z | H_0] < 0$ and $E[Z | H_1] > 0$.

Thus crossing α is a rare event given the random walk with H_0 and crossing β is rare given H_1 .

Since $r^* = 1$ for the H_0 walk, $\Pr\{e | H_0\} \leq e^{-\alpha}$.

This is not surprising; for the simple RW with $p_1 = 1/2$, $\sum_i Z_i = \alpha$ means that

$$\ln[\Pr\{e | H_1\} / \Pr\{e | H_0\}] = \alpha$$

Also, $\Pr\{e | H_1\} \leq e^\beta$.

12

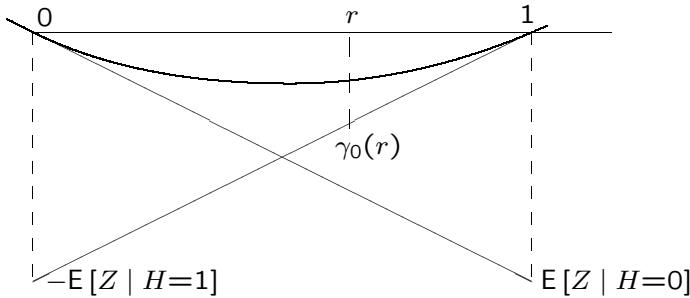
The coupling between errors given H_1 and errors given H_0 is weaker here than for fixed n .

Increasing α lowers $\Pr\{e | H_0\}$ exponentially and increases $E[J | H_1] \approx \alpha/E[Z | H_1]$ (from Wald's equality since $\alpha \approx E[S_J | H=1]$). Thus

$$\Pr\{e | H=0\} \sim \exp(-E[J | H=1] E[Z | H=1])$$

In other words, $\Pr\{e | H=0\}$ is essentially exponential in the expected number of trials given $H=1$. The exponent is $E[Z | H=1]$, illustrated below.

Similarly, $\Pr\{e | H=1\} \sim \exp(E[J | H=0] E[Z | H=0])$.



13

Tilted probabilities

Let $\{X_n; n \geq 1\}$ be a sequence of IID discrete rv's with a MGF at some given r . Given the PMF of X , define a tilted PMF (for X) as

$$q_{X,r}(x) = p_X(x) \exp[rx - \gamma(r)].$$

Summing over x , $\sum q_{X,r}(x) = g_X(r)e^{-\gamma_X(r)} = 1$. We view $q_{X,r}(x)$ as the PMF on X in a new probability space with this given relationship to the old space.

We can then use all the laws of probability in this new measure. In this new measure, $\{X_n; n \geq 1\}$ are taken to be IID. The mean of X in this new space is

$$\begin{aligned} E_r[X] &= \sum_x x q_{X,r}(x) = \sum_x x p_X(x) \exp[rx - \gamma(r)] \\ &= \frac{1}{g_X(r)} \sum_x \frac{d}{dr} p_X(x) \exp[rx] \\ &= \frac{g'_X(r)}{g_X(r)} = \gamma'(r). \end{aligned}$$

14

The joint tilted PMF for $\vec{X}^n = (X_1, \dots, X_n)$ is then

$$q_{\vec{X}^n, r}(x_1, \dots, x_n) = p_{\vec{X}^n}(x_1, \dots, x_n) \exp\left(\sum_{i=1}^n [rx_i - \gamma(r)]\right).$$

Let $A(s_n)$ be the set of n -tuples such that $x_1 + \dots + x_n = s_n$.

Then (in the original space) $p_{S_n}(s_n) = \Pr\{S_n = s_n\} = \Pr\{A(s_n)\}$.

Also, for each $\vec{x}^n \in A(s_n)$,

$$\begin{aligned} q_{\vec{X}^n, r}(x_1, \dots, x_n) &= p_{\vec{X}^n}(x_1, \dots, x_n) \exp(rs_n - n\gamma(r)] \\ q_{S_n, r}(s_n) &= p_{S_n}(s_n) \exp[rs_n - n\gamma(r)], \end{aligned}$$

where we have summed over $A(s_n)$. This is the key to much of large deviation theory. For $r > 0$, it tilts the probability measure on S_n toward large values, and the laws of large numbers can be used on this tilted measure.

Proof of Wald's identity

The stopping time J for the 2 threshold RW is a rv (from Lemma 7.5.1) and it is also a rv for the tilted probability measure. Let $\mathcal{T}_n = \{\vec{x}_n : s_n \notin (\beta, \alpha); s_i \in (\beta, \alpha); 1 \leq i < n\}$.

That is, \mathcal{T}_n is the set of n tuples for which stopping occurs on trial n . Letting $q_{J, r(n)}$ be the PMF of J in the tilted probability measure,

$$\begin{aligned} q_{J, r}(n) &= \sum_{\vec{x}^n \in \mathcal{T}_n} q_{\vec{X}^n, r}(\vec{x}^n) = \sum_{\vec{x}^n \in \mathcal{T}_n} p_{\vec{X}^n}(\vec{x}^n) \exp[rs_n - n\gamma(r)] \\ &= E[\exp[rs_n - n\gamma(r) | J=n] \Pr\{J = n\}]. \end{aligned}$$

Summing over n completes the proof.

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

L23: Martingales, plain, sub, and super

Outline:

- Review of Wald and sequential tests
- Wald's identity with zero-mean rv's
- Martingales
- Simple Examples of martingales
- Sub and super martingales

1

Thm: (Wald) Let $\{X_i; i \geq 1\}$ be IID with a semi-invariant MGF $\gamma(r) = \ln(E[\exp(rX)])$ that exists for $(r_- < 0 < r_+)$. Let $\{S_n; n \geq 1\}$ be the RW with $S_n = X_1 + \dots + X_n$. If J is the trial at which S_n first crosses $\alpha > 0$ or $\beta < 0$,

$$E[\exp(rS_J - J\gamma(r))] = 1 \quad \text{for } r \in (r_-, r_+)$$

Corollary: If $\bar{X} < 0$ and $\gamma(r^*) = 0$ for $0 < r^*$, then

$$\Pr\{S_J \geq \alpha\} \leq \exp(-\alpha r^*)$$

Pf: The Wald identity says $E[\exp(r^*S_J)] = 1$, so this follows from the Markov inequality.

This is valid for all lower thresholds and also for no lower threshold, where it is better stated as

$$\Pr\left\{\bigcup_n \{S_n \geq \alpha\}\right\} \leq \exp(-r^*\alpha)$$

This is stronger (for the case of threshold crossing) than the Chernoff bound, which says that $\Pr\{S_n \geq \alpha\} \leq \exp -r^*\alpha$ for all n .

2

Review of hypothesis testing: View a binary hypothesis as a binary rv H with $p_H(0) = p_0$ and $p_H(1) = p_1$.

We observe $\{Y_n; n \geq 1\}$, which, conditional on $H = \ell$ is IID with density $f_{Y|H}(y|\ell)$. Define the likelihood ratio

$$\Lambda(\vec{y}^n) = \prod_{i=1}^n \frac{f_{Y_i|H}(y_i|0)}{f_{Y_i|H}(y_i|1)}$$

$$\frac{\Pr\{H=0 | \vec{y}^n\}}{\Pr\{H=1 | \vec{y}^n\}} = \frac{p_0 f_{\vec{Y}^n|H}(y^n | 0)}{p_1 f_{\vec{Y}^n|H}(y^n | 1)} = \frac{p_0}{p_1} \Lambda(\vec{y}^n).$$

MAP rule: $\Lambda(\vec{y}^n) \begin{cases} > p_1/p_0 & ; & \text{select } \hat{h}=0 \\ \leq p_1/p_0 & ; & \text{select } \hat{h}=1. \end{cases}$

3

Define the log likelihood ratio as

$$LLR = \ln[\Lambda(\vec{y}^n)] = \sum_{i=1}^n \ln \frac{f_{Y_i|H}(y_i|0)}{f_{Y_i|H}(y_i|1)}$$

$$s_n = \sum_{i=1}^n z_i \quad \text{where } z_i = \ln \frac{f_{Y_i|H}(y_i|0)}{f_{Y_i|H}(y_i|1)}$$

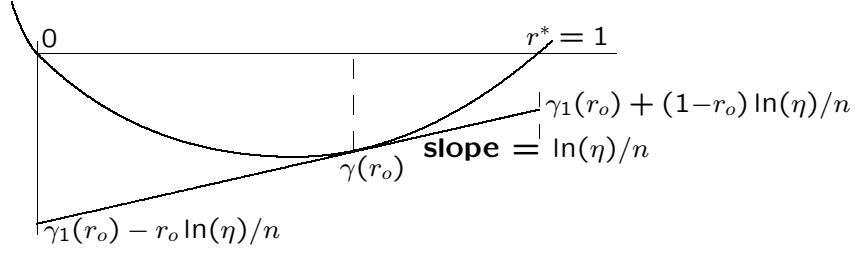
Conditional on $H = 1$, $\{S_n; n \geq 1\}$ is a RW with $S_n = Z_1 + \dots + Z_n$, where each Z_i is a function of Y_i . The Z_i , given $H = 1$ are then IID.

$$\begin{aligned} \gamma_1(r) &= \ln \left\{ \int f_{Y_i|H}(y_i|1) \exp \left[r \ln \frac{f_{Y_i|H}(y_i|0)}{f_{Y_i|H}(y_i|1)} \right] dy \right\} \\ &= \ln \left\{ \int f_{Y_i|H}^{1-r}(y_i|1) f_{Y_i|H}^r(y_i|0) dy \right\} \end{aligned}$$

Note that $\gamma_1(1) = 0$, so $r^* = 1$.

4

For fixed n , a threshold rule says choose $\hat{H} = 0$ if $S_n \geq \ln \eta$. Thus, given $H = 1$, an error occurs if $S_n \geq \ln \eta$. From the Chernoff bound,



$$\Pr\{e | H=1\} \leq \exp(n\gamma_1(r_o) - r_o \ln \eta)$$

Given $H = 0$, a similar argument shows that

$$\Pr\{e | H=0\} \leq \exp(n\gamma_1(r_o) + (1 - r_o) \ln \eta)$$

A better strategy is sequential decisions. For the same pair of RW's, continue trials until either $S_n \geq \alpha$ or $S_n \leq \beta$ where $\alpha > 0$ and $\beta < 0$.

5

Given $H = 1$, $\{S_n; n \geq 1\}$ is a random walk. Choose some $\alpha > 0$ and $\beta < 0$ and let J be a stopping time, stopping when first $S_n \geq \alpha$ or $S_n \leq \beta$.

If $S_J \geq \alpha$, decide $\hat{H} = 0$ and if $S_J \leq \beta$, decide $\hat{H} = 1$. Conditional on $H = 1$, an error is made if $S_J \geq \alpha$. Then

$$\Pr\{e | H=1\} = \Pr\{S_J \geq \alpha | H=1\} \leq \exp[-\alpha r^*]$$

where r^* is the root of $\gamma(r) = \ln E[\exp(rZ) | H = 1]$, i.e., $r^* = 1$.

$$\begin{aligned}\gamma(r) &= \ln \int_y f_{Y|H}(y|1) \exp \left[r \ln \left(\frac{f_{Y|H}(y|0)}{f_{Y|H}(y|1)} \right) \right] dy \\ &= \ln \int_y [f_{Y|H}(y|1)]^{1-r} [f_{Y|H}(y|0)]^r dy\end{aligned}$$

6

Choose apriori's $p_0 = p_1$. **Then at the end of trial n**

$$\frac{\Pr\{H=0 \mid S_n\}}{\Pr\{H=1 \mid S_n\}} = \exp(S_n); \quad \frac{1 - \Pr\{H=1 \mid S_n\}}{\Pr\{H=1 \mid S_n\}} = \exp(-S_n)$$

$$\Pr\{H=1 \mid S_n\} = \frac{\exp(-S_n)}{1 + \exp(-S_n)}$$

This is the probability of error if a decision $\hat{h} = 0$ is made at the end of trial n . Thus deciding $\hat{h} = 0$ on crossing α guarantees that $\Pr\{e \mid H=1\} \leq \exp -\alpha$.

As we saw last time, the cost of choosing α to be large is many trials under $H = 0$. In particular, the stopping time J satisfies

$$\mathbb{E}[J \mid H=0] = \frac{\mathbb{E}[S_J \mid H=0]}{\mathbb{E}[Z \mid H=0]} \approx \frac{\alpha + \mathbb{E}[\text{overshoot} \mid H=0]}{\mathbb{E}[Z \mid H=0]}$$

7

Wald's identity with zero-mean rv's

If we take the first 2 derivatives of Wald's identity at $r = 0$, we get Wald's equality and a useful result for zero-mean rv's.

$$\frac{d}{dr} \mathbb{E}[\exp(rS_J - J\gamma(r))] = \mathbb{E}[[S_J - J\gamma'(r)] \exp(rS_J - J\gamma(r))]$$

$$\left. \frac{d}{dr} \mathbb{E}[\exp(rS_J - J\gamma(r))] \right|_{r=0} = \mathbb{E}[S_J - J\bar{X}] = 0; \quad (\text{Wald eq.})$$

$$\left. \frac{d^2}{dr^2} \mathbb{E}[\exp(rS_J - J\gamma(r))] \right|_{r=0} = \mathbb{E}[S_J^2 - \sigma_X^2 \bar{J}] = 0; \quad \text{if } \bar{X} = 0$$

For zero-mean simple RW with threshold at $\alpha > 0$ and $\beta < 0$, we have $\bar{J} = -\beta\alpha$

8

Martingales

A sequence $\{Z_n; n \geq 1\}$ of rv's is a martingale if $E[|Z_n|] < \infty$ for all $n \geq 1$ and

$$E[Z_n | Z_{n-1}, Z_{n-2}, \dots, Z_1] = Z_{n-1} \quad (1)$$

The condition $E[|Z_n|] < \infty$ is almost a mathematical fine point, and we mostly ignore it here. The condition (1) appears to be a very weak condition, but it leads to surprising applications. In times of doubt, write (1) as

$$E[Z_n | Z_{n-1}=z_{n-1}, \dots, Z_1=z_1] = z_{n-1}$$

for all sample values $z_{n-1}, z_{n-2}, \dots, z_1$

Lemma: For a martingale, $\{Z_n; n \geq 1\}$, and for $n > i \geq 1$,

$$E[Z_n | Z_i, Z_{i-1}, \dots, Z_1] = Z_i$$

Pf: To start, we show that $E[Z_3 | Z_1] = Z_1$. Recall the meaning of $E[X] = E[E[X|Y]]$. Then

$$E[Z_3 | Z_1] = E[E[Z_3 | Z_2, Z_1] | Z_1]$$

9

$$E[Z_3 | Z_1] = E[E[Z_3 | Z_2, Z_1] | Z_1] = E[Z_2 | Z_1] = Z_1$$

In the same way,

$$\begin{aligned} E[Z_{i+2} | Z_i, \dots, Z_1] &= E[E[Z_{i+2} | Z_{i+1}, \dots, Z_1] | Z_i, \dots, Z_1] \\ &= E[Z_{i+1} | Z_i, \dots, Z_1] = Z_i \end{aligned}$$

After more of the same, $E[Z_n | Z_i, \dots, Z_1] = Z_i$.

The most important special case is $E[Z_n | Z_1] = Z_1$, and thus $E[Z_n] = E[Z_1]$.

Simple Examples of martingales

1) Zero-mean random walk: Let $Z_n = X_1 + \cdots + X_n$ where $\{X_i; i \geq 1\}$ are IID and zero mean.

$$\begin{aligned}\mathbb{E}[Z_n | Z_{n-1}, \dots, Z_1] &= \mathbb{E}[X_n + Z_{n-1} | Z_{n-1}, \dots, Z_1] \\ &= \mathbb{E}[X_n] + Z_{n-1} = Z_{n-1}.\end{aligned}$$

2) Sums of ‘arbitrary’ dependent rv’s: Suppose $\{X_i; i \geq 1\}$ satisfy $\mathbb{E}[X_i | X_{i-1}, X_{i-2}, \dots, X_1] = 0$. Then $\{Z_n; n \geq 1\}$ where $Z_n = X_1 + \cdots + X_n$ is a martingale.

This can be taken as an alternate definition of a martingale. We can either start with the sums Z_n or with the differences between successive sums.

11

3) Let $X_i = U_i Y_i$ where $\{U_i; i \geq 1\}$ are IID, equiprobable ± 1 . The Y_i are non-negative and independent of the U_i but otherwise arbitrary. Then

$$\mathbb{E}[X_n | X_{n-1}, \dots, X_1] = 0$$

Thus $\{Z_n; n \geq 1\}$ where $Z_n = X_1 + \cdots + X_n$ is a martingale.

4) Product form martingales. Suppose $\{X_i; i \geq 1\}$ is a sequence of IID unit-mean rv’s. Then $\{Z_n; n \geq 1\}$ where $Z_n = X_1 X_2 \cdots X_n$ is a martingale.

$$\begin{aligned}\mathbb{E}[Z_n | Z_{n-1}, \dots, Z_1] &= \mathbb{E}[X_n Z_{n-1} | Z_{n-1}, \dots, Z_1] \\ &= \mathbb{E}[X_n] \mathbb{E}[Z_{n-1} | Z_{n-1}, \dots, Z_1] \\ &= \mathbb{E}[Z_{n-1} | Z_{n-1}] = Z_{n-1}.\end{aligned}$$

12

5) Special case of product form martingale: let X_i be IID and equiprobably 2 or 0.

$$\Pr\{Z_n = 2^n\} = 2^{-n}; \quad \Pr\{Z_n = 0\} = 1 - 2^{-n}; \quad E[Z_n] = 1.$$

Thus $\lim_n Z_n = 0$ **WP1** but $E[Z_n] = 1$ for all n

6) Recall the branching process where X_n is the number of elements in gen n and $X_{n+1} = \sum_{i=1}^{X_n} Y_{i,n}$ where the $Y_{i,n}$ are IID.

Let $Z_n = X_n/\bar{Y}^n$, i.e., $\{Z_n; n \geq 1\}$ is a scaled down branching process.

$$E[Z_n | Z_{n-1}, \dots, Z_1] = E\left[\frac{X_n}{\bar{Y}^n} | X_{n-1}, \dots, X_1\right] = \frac{\bar{Y}X_{n-1}}{\bar{Y}^n} = Z_{n-1}.$$

Thus this is a martingale.

13

Submartingales and supermartingales

These are sequences $\{Z_n; n \geq 1\}$ with $E[|Z_n|] < \infty$ like martingales, but with inequalities instead of equalities.

For all $n \geq 1$,

$$\begin{aligned} E[Z_n | Z_{n-1}, \dots, Z_1] &\geq Z_{n-1} && \text{submartingale} \\ E[Z_n | Z_{n-1}, \dots, Z_1] &\leq Z_{n-1} && \text{supermartingale} \end{aligned}$$

We refer only to submartingales in what follows, since the supermartingale case results from replacing Z_n with $-Z_n$.

For submartingales,

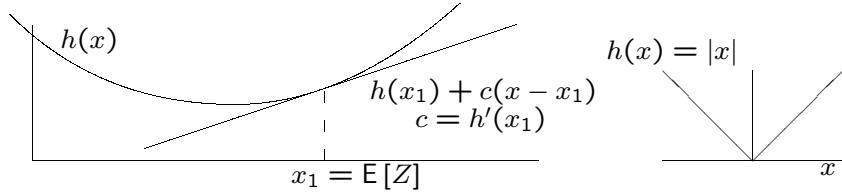
$$E[Z_n | Z_i, \dots, Z_1] \geq Z_i \quad \text{for all } n > i > 0$$

$$E[Z_n] \geq E[Z_i] \quad \text{for all } n > i > 0$$

14

Convex functions

A function $h(x)$, $\mathbb{R} \rightarrow \mathbb{R}$, is convex if each tangent to the curve lies on or below the curve. The condition $h''(x) \geq 0$ is sufficient but not necessary.



Lemma (Jensen's inequality): If h is convex and Z is a rv with finite expectation, then

$$h(\mathbb{E}[Z]) \leq \mathbb{E}[h(Z)]$$

15

Jensen's inequality can be used to prove the following theorem. See Section 7.7 for a proof.

If $\{Z_n; n \geq 1\}$ is a martingale or submartingale, if h is convex, and if $\mathbb{E}[|h(Z_n)|] < \infty$ for all n , then $\{h(Z_n); n \geq 1\}$ is a submartingale.

For example, if $\{Z_n; n \geq 1\}$ is a martingale, then essentially $\{|Z_n|; n \geq 1\}$, $\{Z_n^2; n \geq 1\}$ and $\{e^{rZ_n}; n \geq 1\}$ are submartingales.

16

Stopped martingales

The definition of a stopping time for a stochastic process $\{Z_n; n \geq 1\}$ applies to any process. That is, J must be a rv and $\{J = n\}$ must be specified by $\{Z_1, \dots, Z_n\}$.

This can be extended to possibly defective stopping times if J is possibly defective (consider a random walk with a single threshold).

A stopped process $\{Z_n^*; n \geq 1\}$ for a possibly defective stopping time J on a process $\{Z_n; n \geq 1\}$ satisfies $Z_n^* = Z_n$ if $n \leq J$ and $Z_n^* = Z_J$ if $n > J$.

17

For example, a given gambling strategy, where Z_n is the net worth at time n , could be modified to stop when Z_n reaches some given value. Then Z_n^* would remain at that value forever after, while Z_n follows the original strategy.

Theorem: If $\{Z_n; n \geq 1\}$ is a martingale (submartingale) and J is a possibly defective stopping rule for it, then the stopped process $\{Z_n^*; n \geq 1\}$ is a martingale (submartingale).

Pf: Obvious??? The intuition here is that before stopping occurs, $Z_n^* = Z_n$, so Z_n^* satisfies the martingale (subm.) condition. Afterwards, Z_n^* is constant, so it again satisfies the martingale (subm) condition. Section 7.8 does this carefully.

18

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

L24: Martingales: stopping and converging**Outline:**

- Review of martingales
- Stopped martingales
- The Kolmogorov submartingale inequality
- SLLN for IID rv's with a variance
- The martingale convergence theorem

1

A sequence $\{Z_n; n \geq 1\}$ of rv's is a martingale if

$$\mathbb{E}[Z_n | Z_{n-1}, Z_{n-2}, \dots, Z_1] = Z_{n-1}; \quad \mathbb{E}[|Z_n|] < \infty \quad (1)$$

for all $n \geq 1$.

Recall that $\mathbb{E}[Z_n | Z_{n-1}, \dots, Z_1]$ is a rv that maps each sample point ω to the conditional expectation of Z_n conditional on $Z_1(\omega), \dots, Z_{n-1}(\omega)$. For a martingale, this expectation must be the rv Z_{n-1} .

Lemma: For a martingale, $\{Z_n; n \geq 1\}$, and for $n > i \geq 1$,

$$\mathbb{E}[Z_n | Z_i, Z_{i-1}, \dots, Z_1] = Z_i; \quad \mathbb{E}[Z_n] = \mathbb{E}[Z_i]$$

Can you figure out why $\mathbb{E}[Z_n | Z_m, \dots, Z_1] = Z_n$ for $m \geq n$?

Note that $\mathbb{E}[Z_n | Z_1] = Z_1$ and $\mathbb{E}[Z_n] = \mathbb{E}[Z_1]$ for all $n > 1$.

2

Simple Examples of martingales

- 1) **Zero-mean RW:** If $Z_n = \sum_{i=1}^n X_i$ where $\{X_i; i \geq 1\}$ are IID and zero mean, then $\{Z_n; n \geq 1\}$ is a martingale.
- 2) If $Z_n = \sum_{i=1}^n X_i$ where $E[X_i | X_{i-1}, \dots, X_1] = 0$ for each $i \geq 1$, then $\{Z_n; n \geq 1\}$ is a martingale.
- 3) Let $X_i = U_i Y_i$ where $\{U_i; i \geq 1\}$ are IID, equiprobable ± 1 . The Y_i are independent of the U_i . Then $\{Z_n; n \geq 1\}$, where $Z_n = X_1 + \dots + X_n$, is a martingale.
- 4) **Product form martingales:** Let $\{X_i; i \geq 1\}$ be a sequence of IID unit-mean rv's. Then $\{Z_n; n \geq 1\}$, where $Z_n = X_1 X_2 \cdots X_n$, is a martingale.

If $p_X(0) = p_X(2) = 1/2$, then $p_{Z_n}(2^n) = 2^{-n}$ and $p_{Z_n}(0) = 1 - 2^{-n}$. Thus, $\lim_{n \rightarrow \infty} Z_n = 0$ WP1 and $\lim_{n \rightarrow \infty} E[Z_n] = 1$.

3

Sub- and supermartingales are sequences $\{Z_n; n \geq 1\}$ with $E[|Z_n|] < \infty$ for which inequalities replace the equalities of martingales. For all $n \geq 1$,

$$\begin{aligned} E[Z_n | Z_{n-1}, \dots, Z_1] &\geq Z_{n-1} && \text{submartingale} \\ E[Z_n | Z_{n-1}, \dots, Z_1] &\leq Z_{n-1} && \text{supermartingale} \end{aligned}$$

If $\{Z_n; n \geq 1\}$ is a submartingale, then $\{-Z_n; n \geq 1\}$ is a supermartingale and vice-versa, so we consider only submartingales. For submartingales,

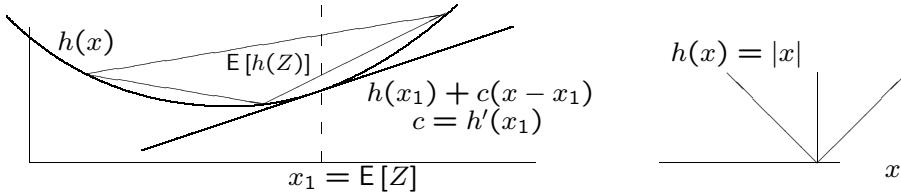
$$E[Z_n | Z_i, \dots, Z_1] \geq Z_i \quad \text{for all } n > i > 0$$

$$E[Z_n] \geq E[Z_i] \quad \text{for all } n > i > 0$$

4

Convex functions

A function $h(x)$, $\mathbb{R} \rightarrow \mathbb{R}$, is convex if each tangent to the curve lies on or below the curve. The condition $h''(x) \geq 0$ is sufficient but not necessary.



Lemma (Jensen's inequality): If h is convex and Z is a rv with finite expectation, then

$$h(E[Z]) \leq E[h(Z)]$$

5

Jensen's inequality leads to the following theorem. See proof in text.

Thm 7.8.1: If $\{Z_n; n \geq 1\}$ is a martingale or submartingale, if h is convex, and if $E[|h(Z_n)|] < \infty$ for all n , then $\{h(Z_n); n \geq 1\}$ is a submartingale.

For example, if $\{Z_n; n \geq 1\}$ is a martingale, then $\{|Z_n|; n \geq 1\}$, $\{Z_n^2; n \geq 1\}$ and $\{e^{rZ_n}; n \geq 1\}$ are submartingales if the marginal expected values exist.

6

Stopped martingales

The definition of a stopping trial J for a stochastic process $\{Z_n; n \geq 1\}$ applies to any process. That is, J must be a rv and $\{J = n\}$ must be specified by $\{Z_1, \dots, Z_n\}$.

A possibly defective rv J is a mapping from Ω to the extended reals \mathbb{R}^+ where $\{J = \infty\}$ and $\{J = -\infty\}$ might have positive probability. The other provisos of rv's still hold.

A possibly defective stopping trial is thus a stopping rule in which stopping may never happen (such as RW's with a single threshold).

A stopped process $\{Z_n^*; n \geq 1\}$ for a possibly defective stopping time J on a process $\{Z_n; n \geq 1\}$ satisfies $Z_n^* = Z_n$ if $n \leq J$ and $Z_n^* = Z_J$ if $n > J$.

7

For example, a given gambling strategy, where Z_n is the net worth at time n , could be modified to stop when Z_n reaches some given value. Then Z_n^* would remain at that value forever after, while Z_n follows the original strategy.

Theorem: If J is a possibly defective stopping rule for a martingale (submartingale), $\{Z_n; n \geq 1\}$, then the stopped process $\{Z_n^*; n \geq 1\}$ is a martingale (submartingale).

Pf: Obvious??? The intuition here is that before stopping occurs, $Z_n^* = Z_n$, so Z_n^* satisfies the martingale (subm.) condition. Afterwards, Z_n^* is constant, so it again satisfies the martingale (subm) condition.

8

Proof that $\{Z_n^*; n \geq 1\}$ is a martingale: Note that

$$Z_n^* = \sum_{m=1}^{n-1} Z_m \mathbb{I}_{\{J=m\}} + Z_n \mathbb{I}_{\{J \geq n\}}$$

Thus $|Z_n^*| \leq \sum_{m < n} |Z_m| + |Z_n|$. Thus means that $E[|Z_n^*|] < \infty$ since it is bounded by the sum of n finite numbers.

Next, let $\vec{Z}^{(n-1)}$ denote Z_{n-1}, \dots, Z_1 and consider

$$\begin{aligned} E[Z_n^* | \vec{Z}^{(n-1)}] &= \sum_{m < n} E[Z_m \mathbb{I}_{\{J=m\}} | \vec{Z}^{(n-1)}] + E[Z_n \mathbb{I}_{\{J \geq n\}} | \vec{Z}^{(n-1)}] \\ E[Z_m \mathbb{I}_{\{J=m\}} | \vec{Z}^{(n-1)} = \vec{z}^{(n-1)}] &= \begin{cases} z_m; & \text{if } J = m \\ 0; & \text{if } J \neq m. \end{cases} \quad \text{for } m < n \\ E[Z_m \mathbb{I}_{\{J=m\}} | \vec{Z}^{(n-1)}] &= Z_m \mathbb{I}_{J=m}. \\ E[Z_n \mathbb{I}_{\{J=n\}} | \vec{Z}^{(n-1)}] &= Z_{n-1} \mathbb{I}_{\{J \geq n\}} \\ E[Z_n^* | \vec{Z}^{(n-1)}] &= \sum_{m < n} Z_m \mathbb{I}_{\{J=m\}} + Z_{n-1} \mathbb{I}_{\{J \geq n\}} \\ &= \sum_{m < n-1} Z_m \mathbb{I}_{\{J=m\}} + Z_{n-1} [\mathbb{I}_{\{J=n-1\}} + \mathbb{I}_{\{J \geq n\}}] \\ &= Z_{n-1}^* \end{aligned}$$

9

This shows that $E[Z_n^* | \vec{Z}^{(n-1)}] = Z_{n-1}^*$. To show that $\{Z_n^*; n \geq 1\}$ is a martingale, though, we must show that $E[Z_n^* | \vec{Z}^{*(n-1)}] = Z_{n-1}^*$. However, $\vec{Z}^{*(n-1)}$ is a function of $\vec{Z}^{(n-1)}$.

For every sample point $\vec{z}^{(n-1)}$ of $\vec{Z}^{(n-1)}$ leading to a given $\vec{z}^{*(n-1)}$ of $\vec{Z}^{*(n-1)}$, we have

$$E[Z_n^* | \vec{Z}^{(n-1)} = \vec{z}^{(n-1)}] = z_{n-1}^*$$

and thus

$$E[Z_n^* | \vec{Z}^{*(n-1)} = \vec{z}^{*(n-1)}] = z_{n-1}^*.$$

QED

A consequence of the theorem, under the same assumptions, is that

$$\begin{aligned} E[Z_1] &\leq E[Z_n^*] \leq E[Z_n] && \text{(submartingale)} \\ E[Z_1] &= E[Z_n^*] = E[Z_n] && \text{(martingale)} \end{aligned}$$

This is also almost intuitively obvious and proved in Section 7.8.

Recall the generating function product martingale for a random walk. That is, let $\{X_n; n \geq 1\}$ be IID and $\{S_n; n \geq 1\}$ be a random walk where $S_n = X_1 + \cdots + X_n$.

Then for r such that $\gamma(r)$ exists, let $Z_n = \exp[rS_n - n\gamma(r)]$. Then $\{Z_n; n \geq 1\}$ is a martingale and $E[Z_n] = 1$ for all $n \geq 1$.

11

For r such that $\gamma(r)$ exists, let $Z_n = \exp[rS_n - n\gamma(r)]$. Then $\{Z_n; n \geq 1\}$ is a martingale and $E[Z_n] = 1$ for all $n \geq 1$.

Let J be the nondefective stopping time that stops on crossing either $\alpha > 0$ or $\beta < 0$. Then $E[Z_n^*] = 1$ for all $n \geq 1$.

Also, $\lim_{n \rightarrow \infty} Z_n^* = Z_J$ **WP1 and**

$$E[Z_J] = E[\exp[rS_J - J\gamma(r)]] = 1$$

This is Wald's identity in a more general form. The connection of $\lim_n Z_n^*$ to Z_J needs more care (see Section 7.8), but this shows the power of martingales.

12

Kolmogorov's submartingale inequality

Thm: Let $\{Z_n; n \geq 1\}$ be a non-negative submartingale. Then for any positive integer m and any $a > 0$,

$$\Pr\left\{\max_{1 \leq i \leq m} Z_i \geq a\right\} \leq \frac{\mathbb{E}[Z_m]}{a}. \quad (2)$$

If we replace the max with Z_m , this is the lowly but useful Markov inequality.

Proof: Let J be the stopping time defined as the smallest $n \leq m$ such that $Z_n \geq a$.

14

If $Z_n \geq a$ for some $n \leq m$, then J is the smallest n for which $Z_n \geq a$.

If $Z_n < a$ for all $n \leq m$, then $J = m$. Thus the process must stop by time m , and $Z_J \geq a$ iff $Z_n \geq a$ for some $n \leq m$. Thus

$$\Pr\left\{\max_{1 \leq n \leq m} Z_n \geq a\right\} = \Pr\{Z_J \geq a\} \leq \frac{\mathbb{E}[Z_J]}{a}.$$

Since the process must be stopped by time m , we have $Z_J = Z_m^*$.

$\mathbb{E}[Z_m^*] \leq \mathbb{E}[Z_m]$, so the right hand side above is less than or equal to $\mathbb{E}[Z_m]/a$, completing the proof.

15

The Kolmogorov submartingale inequality is a strengthening of the Markov inequality. The Chebyshev inequality is strengthened in the same way.

Let $\{Z_n; n \geq 1\}$ be a martingale with $E[Z_n^2] < \infty$ for all $n \geq 1$. Then

$$\Pr \left\{ \max_{1 \leq n \leq m} |Z_n| \geq b \right\} \leq \frac{E[Z_m^2]}{b^2}; \text{ for all integer } m \geq 2, \text{ all } b > 0.$$

Let $\{S_n; n \geq 1\}$ be a RW with $S_n = X_1 + \dots + X_n$ where each X_i has mean \bar{X} and variance σ^2 . Then for any positive integer m and any $\epsilon > 0$,

$$\Pr \left\{ \max_{1 \leq n \leq m} |S_n - n\bar{X}| \geq m\epsilon \right\} \leq \frac{\sigma^2}{m\epsilon^2}.$$

16

SLLN for IID rv's with a variance

Thm: Let $\{X_i; i \geq 1\}$ be a sequence of IID random variables with mean \bar{X} and standard deviation $\sigma < \infty$. Let $S_n = X_1 + \dots + X_n$. Then for any $\epsilon > 0$,

$$\Pr \left\{ \lim_{n \rightarrow \infty} \frac{S_n}{n} = \bar{X} \right\} = 1$$

Idea of proof:

$$\Pr \left\{ \bigcup_{m=k}^{\infty} \left\{ \max_{1 \leq n \leq 2^m} |S_n - n\bar{X}| \geq 2^m\epsilon \right\} \right\} \leq \sum_{m=k}^{\infty} \frac{\sigma^2}{2^{m\epsilon^2}} = \frac{2\sigma^2}{2^k\epsilon^2}$$

Then lower bound the left term to

$$\Pr \left\{ \bigcup_{m=k}^{\infty} \left\{ \max_{2^{m-1} \leq n \leq 2^m} |S_n - n\bar{X}| \geq 2n\epsilon \right\} \right\}$$

17

The martingale convergence theorem

Thm: Let $\{Z_n; n \geq 1\}$ be a martingale and assume that there is some finite M such that $E[|Z_n|] \leq M$ for all n . Then there is a random variable Z such that, $\lim_{n \rightarrow \infty} Z_n = Z$ WP1.

The text proves the theorem with the additional constraint that $E[Z_n^2]$ is bounded. Either bounded $E[Z_n^2]$ or bounded $E[|Z_n|]$ is a very strong constraint, but the theorem is still very powerful.

For a branching process $\{X_n; n \geq 1\}$ where the number Y of offspring of an element has $\bar{Y} > 1$, we saw that $\{X_n/\bar{Y}^n; n \geq 1\}$ is a martingale satisfying the constraint, so $X^n/\bar{Y}^n \rightarrow Z$ WP1.

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

L25: Putting it all together?

Outline:

- Martingales
- Markov chains (Countable or finite state)
- Countable state Markov processes
- Random walks

1

Martingales

A sequence $\{Z_n; n \geq 1\}$ is a martingale if for all $n > 1$,

$$\mathbb{E}[Z_n | Z_{n-1}, Z_{n-2}, \dots, Z_1] = Z_{n-1}; \quad \mathbb{E}[|Z_n|] < \infty$$

Lemma: For a martingale, $\{Z_n; n \geq 1\}$, and for $n > i \geq 1$,

$$\mathbb{E}[Z_n | Z_i, Z_{i-1}, \dots, Z_1] = Z_i; \quad \mathbb{E}[Z_n] = \mathbb{E}[Z_i]$$

The increments $X_n = Z_n - Z_{n-1}$ satisfy $\mathbb{E}[X_n | X_{n-1}, \dots, X_1] = 0$ iff $\{Z_n; n \geq 1\}$ is a martingale. A good special case is $X_n = U_n Y_n$ where U_n are IID, $\mathbb{P}_U(1) = \mathbb{P}_U(-1) = 1/2$.

Examples: Zero mean RW and product of unit-mean IID rv's.

2

Submartingales

$\{Z_n; n \geq 1\}$ is a submartingale if for all $n \geq 1$,

$$\mathbb{E}[Z_{n+1} | Z_n, \dots, Z_1] \geq Z_n; \quad \mathbb{E}[Z_n] < \infty$$

Lemma: If $\{Z_n; n \geq 1\}$ is a martingale, then for $n > i > 0$,

$$\mathbb{E}[Z_n | Z_i, \dots, Z_1] \geq Z_i; \quad \mathbb{E}[Z_n] \geq \mathbb{E}[Z_i]$$

If $h(x)$ is convex, then Jensen's inequality says $\mathbb{E}[h(X)] \geq h(\mathbb{E}[X])$. If $\{Z_n; n \geq 1\}$ is a submartingale (including martingale), h is convex, and $\mathbb{E}[|h(X)|] < \infty$, then $\{h(Z_n); n \geq 1\}$ is a submartingale.

3

Stopped (sub)martingales

A stopped process $\{Z_n^*; n \geq 1\}$ for a possibly defective stopping time J on a process $\{Z_n; n \geq 1\}$ satisfies $Z_n^* = Z_n$ if $n \leq J$ and $Z_n^* = Z_J$ if $n > J$.

Theorem: The stopped process, $\{Z_n^*; n \geq 1\}$, for a (sub) martingale with a (possibly defective) stopping rule is a (sub)martingale and satisfies

$$\begin{aligned} \mathbb{E}[Z_1] &\leq \mathbb{E}[Z_n^*] \leq \mathbb{E}[Z_n] && \text{(submartingale)} \\ \mathbb{E}[Z_1] &= \mathbb{E}[Z_n^*] = \mathbb{E}[Z_n] && \text{(martingale)} \end{aligned}$$

For the product martingale $Z_n = \exp[rS_n - n\gamma(r)]$, and a stopping rule J such as threshold crossing, we get a stopped martingale. Subject to some small mathematical nits, this leads to Wald's identity,

$$\mathbb{E}[Z_J] = \mathbb{E}[\exp[rS_J - J\gamma(r)]] = 1$$

4

Kolmogorov's submartingale inequality

Thm: Let $\{Z_n; n \geq 1\}$ be a non-negative submartingale. Then for any positive integer m and any $a > 0$,

$$\Pr \left\{ \max_{1 \leq i \leq m} Z_i \geq a \right\} \leq \frac{\mathbb{E}[Z_m]}{a}.$$

This is Kolmogorov's strengthening of the Markov inequality.

For non-negative martingales, we can go to the limit $m \rightarrow \infty$ since $\mathbb{E}[Z_m] = \mathbb{E}[Z_1]$:

$$\Pr \left\{ \sup_{n \geq 1} Z_n \geq a \right\} \leq \frac{\mathbb{E}[Z_1]}{a}$$

5

Kolmogorov version of Chebyshev: Let $\{Z_n; n \geq 1\}$ be a martingale (or submartingale) with $\mathbb{E}[Z_n^2] < \infty$ for all $n \geq 1$. Then

$$\Pr \left\{ \max_{1 \leq n \leq m} |Z_n| \geq b \right\} \leq \frac{\mathbb{E}[Z_m^2]}{b^2}; \quad m \geq 1, \quad b > 0.$$

It is often more useful to maximize only over half the interval and take a union over different intervals, e.g.,

$$\Pr \left\{ \bigcup_{j \geq k} \left\{ \max_{2^{j-1} < n \leq 2^j} |Z_n| \geq b_j \right\} \right\} \leq \sum_{j=k}^{\infty} \frac{\mathbb{E}[Z_{2^j}]^2}{b_j^2}$$

For the zero-mean RW $\{S_n; n \geq 1\}$ where $S_n = X_1 + \dots + X_n$ with $\bar{X} = 0$ and $\mathbb{E}[X^2] = \sigma^2$, $\mathbb{E}[Z_{2^j}^2] = 2^j \sigma^2$.

$$\Pr \left\{ \bigcup_{j \geq k} \left\{ \max_{2^{j-1} < n \leq 2^j} |S_n| \geq \left(\frac{3}{2}\right)^j \right\} \right\} \leq \sum_{j=k}^{\infty} \left(\frac{8}{9}\right)^j \sigma^2 = \left(\frac{8}{9}\right)^k 9\sigma^2$$

where $b_j = (3/2)^j$.

6

SLLN for $\{S_n; n \geq 1\}$ where $S_n = X_1 + \dots + X_n$ and $\{X_n; n \geq 1\}$ are IID with $\bar{X} = 0, E[X^2] = \sigma^2 < \infty$. Then

$$\Pr\left\{\omega : \lim_{n \rightarrow \infty} \frac{S_n}{n} = 0\right\} = 1.$$

Proof:

$$\Pr\left\{\bigcup_{j \geq k} \left\{ \max_{2^{j-1} < n \leq 2^j} |S_n| \geq \left(\frac{3}{2}\right)^j \right\}\right\} \leq \left(\frac{8}{9}\right)^k 9\sigma^2$$

Lower bounding the left side,

$$\Pr\left\{\bigcup_{j \geq k} \left\{ \max_{2^{j-1} < n \leq 2^j} \frac{|S_n|}{n} \geq 2 \left(\frac{3}{4}\right)^j \right\}\right\} \leq \left(\frac{8}{9}\right)^k 9\sigma^2$$

Any sample sequence $\{S_n(\omega); n \geq 1\}$ that is not contained in the union on the left must satisfy $\lim_{n \rightarrow \infty} |S_n|/n = 0$, and thus $\Pr\{\omega : \lim_n |S_n|/n = 0\} > 1 - (8/9)^k 9\sigma^2$. Since this is true for all k , the theorem is proved. It applies to other martingales also.

7

Markov chains (Countable or finite state)

Def: The first passage time T_{ij} from state i to j is the smallest n , given $X_0 = i$, at which $X_n = j$. T_{ij} is a possibly defective rv with PMF $f_{ij}(n)$ and dist. fcn. $F_{ij}(n)$.

$$f_{ij}(n) = \sum_{k \neq j} P_{ik} f_{kj}(n-1); \quad F_{ij}(n) = \sum_{m=1}^n f_{ij}(m); \quad n > 1.$$

Def: State j is recurrent if T_{jj} is non-defective and transient otherwise. If recurrent, it is positive recurrent if $E[T_{jj}] < \infty$ and null recurrent otherwise.

For each recurrent j there is a integer renewal counting process $\{N_{jj}(t); t > 0\}$ of visits to j starting in j . It has interrenewal distribution $F_{jj}(n)$.

There is a delayed renewal counting process $\{N_{ij}(t); t > 0\}$ for visits to j starting in i .

8

Thm: All states in a class are positive recurrent, or all are null recurrent, or all are transient.

Def: A chain is irreducible if all state pairs communicate.

It is called irreducible because the problem of getting from one or another transient class to an 'irreducible class' is largely separable from the analysis of that 'irreducible class' which then becomes the entire chain.

An irreducible class can be positive recurrent, null recurrent, or transient.

9

Thm: For an irreducible Markov chain, if the 'steady state' equations

$$\pi_j = \sum_i \pi_i P_{ij} \text{ and } \pi_j \geq 0 \text{ for all } j; \sum_j \pi_j = 1$$

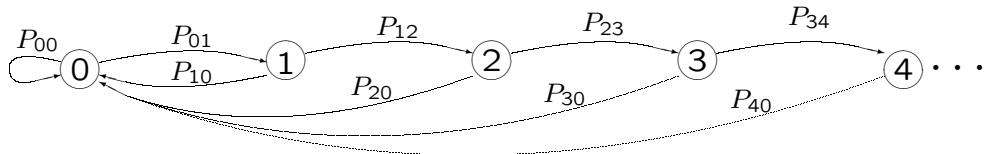
has a solution, then the solution is unique, $\pi_j = 1/T_{jj} > 0$ for all j , and the states are positive recurrent. Also if the states are positive recurrent, then the steady state equations have a solution.

This is an infinite set of equations, so not necessarily computer solvable.

The counting processes under positive recurrence must satisfy

$$\lim_{n \rightarrow \infty} \frac{N_{ij}(n)}{n} = \pi_j \quad \mathbf{WP1}$$

10



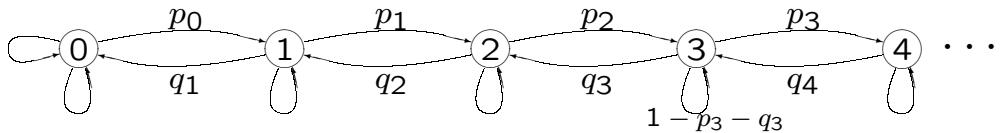
Markov model of age of renewal process

$$\pi_n = \pi_0 P_{01} P_{12} \cdots P_{n-1,n} = \Pr\{Z > n\}$$

$$1 = \sum_i \pi_i = \pi_0 \sum_{i=0}^{\infty} \Pr\{Z > n\} = \pi_0 \bar{Z}$$

This is a nice chain for examples about null-recurrence.

11



A birth-death Markov chain

This is another very useful model for examples about recurrence and for models of sampled-time queueing systems.

Note that the steady state equations are almost trivial.

$$\pi_i p_i = \pi_{i+1} q_{i+1}; \quad \frac{\pi_{i+1}}{\pi_i} = \rho_i$$

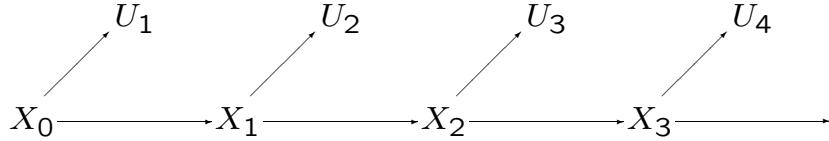
where $\rho_i = p_i/q_{i+1}$.

12

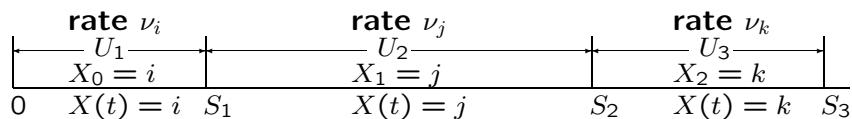
Markov processes

A Markov process $\{X(t); t \geq 0\}$ is a combination of a countable state Markov chain $\{X_n; n \geq 1\}$ along with an exponential holding time U_n for each state.

$$\Pr\{U_n \leq \tau \mid X_n = j, \text{past}\} = 1 - \exp(-\nu_j \tau)$$



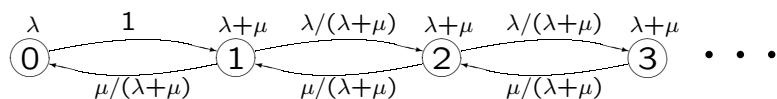
A Markov process is specified by the embedded transition probabilities P_{ij} and the rates ν_i .



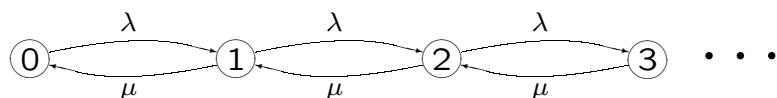
Def: Transition rates are given by $q_{ij} = \nu_i P_{ij}$.

13

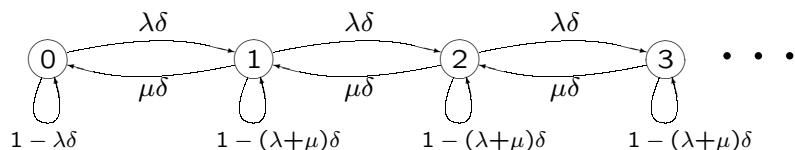
Three ways to represent a Markov process.



An M/M/1 queue using [P] and ν

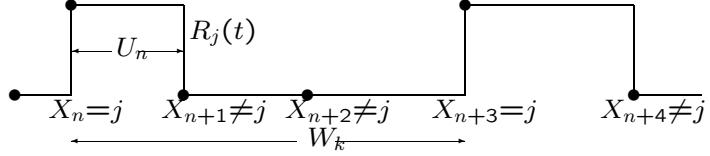


The same M/M/1 queue using [q].



The same M/M/1 queue in sampled time.

14



From the (delayed) renewal reward theorem,

$$p_j = \lim_{t \rightarrow \infty} \frac{\int_0^t R_j(\tau) d\tau}{t} = \frac{\overline{U}(j)}{\overline{W}(j)} = \frac{1}{\nu_j \overline{W}(j)} \quad \text{W.P.1.}$$

We can also assign unit reward for each transition in the renewal interval from state j to j . Let $M_i(t)$ be number of transitions in $(0, t]$ given that $X(0) = i$.

$$\overline{M}_i = \lim_{t \rightarrow \infty} \frac{M_i(t)}{t} = \frac{1}{\pi_j \overline{W}_j} \quad \text{W.P.1.}$$

$$p_j = \frac{\pi_j}{\nu_j} \overline{M}_i$$

15

$$p_j = \frac{1}{\nu_j \overline{W}(j)}; \quad \overline{M} = \frac{1}{\pi_j \overline{W}_j}; \quad p_j = \frac{\pi_j}{\nu_j} \overline{M}$$

If $0 < \overline{M} < \infty$, then each $p_j > 0$ and $\sum_j p_j = 1$.

$$\overline{M} = \frac{1}{\sum_i \pi_i / \nu_i}; \quad p_j = \frac{\pi_j / \nu_j}{\sum_i \pi_i / \nu_i}$$

Similarly, since $\sum_i \pi_i = 1$,

$$\overline{M} = \sum_i p_i \nu_i; \quad \pi_j = \frac{p_j \nu_j}{\sum_i p_i \nu_i}$$

A sampled time MP exists if $\nu_i \leq A$ for some A and all i . The steady state probabilities are the time average probabilities, $\{p_i\}$, which satisfy

$$\nu_j p_j = \sum_i \nu_i p_j; \quad p_j \geq 0; \quad \sum_i p_i = 1$$

16

The strange cases occur when \bar{M} is 0 or infinite.

$\bar{M} = 0$ for the rattled server and discouraged customer queue. Here the embedded chain is positive recurrent, but all p_j are zero and the sampled-time chain is null recurrent.

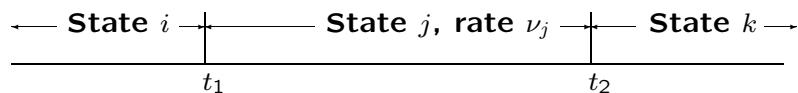
The case $\bar{M} = \infty$ is not possible for a positive recurrent embedded chain, but is possible when the equations $\nu_j p_j = \sum_i \nu_i p_i$ have a solution. These processes are irregular, and allow an infinite number of transitions in finite time.

17

Reversibility for MP's

For any (embedded) Markov chain, the backward transition probabilities are defined as

$$\pi_i P_{ij}^* = \pi_j P_{ji}$$



Moving right, after entering state j , the exit rate is ν_j , i.e., we exit in each δ with probability $\nu_j \delta$. The same holds moving left.

Thus $\{\pi_i\}$ and $\{\nu_i\}$ are the same going left as going right

18

Note that the probability of having a (right) transition from state j to k in $(t, t+\delta)$ is $p_j q_{jk} \delta$. Similarly, for the left going process, if q_{kj}^* is the process transition rate, the probability of having the same transition is $p_k q_{kj}^*$. Thus

$$p_j q_{jk} = p_k q_{kj}^*$$

By fiddling equations, $q_{kj}^* = \nu_k P_{kj}^*$.

Def: A MP is reversible if $q_{ij}^* = q_{ji}$ for all i, j

Assuming positive recurrence and $\sum_i \pi_i / \nu_i < \infty$, the MP process is reversible if and only if the embedded chain is.

19

The guessing theorem: Suppose a MP is irreducible and $\{p_i\}$ is a set of probabilities and satisfies $p_i q_{ij} = p_j q_{ji}$ for all i, j and satisfies $\sum_i p_i \nu_i < \infty$.

Then $p_i > 0$ for all i , p_i is the steady state time average probability of state i , the process is reversible, and the embedded chain is positive recurrent.

Useful application: All birth/death processes are reversible (if $\sum_j p_j \nu_j < \infty$)

20

Random Walks

Def: A random walk is a sequence $\{S_n; n \geq 1\}$ of successive sums $S_n = X_1 + \dots + X_n$ of IID rv's X_i .

We are interested in exponential bounds on S_n for large n (**Chernoff bounds**)

We are also interested in threshold crossings: for 2 thresholds, $\alpha > 0$ and $\beta < 0$, what is the stopping time when S_n first crosses α or β , what is the probability of crossing each, and what is the distribution of the overshoot.

21

Let a rv Z have an MGF $g_Z(r)$ for $0 \leq r < r_+$ and mean $\bar{Z} < 0$. By the Chernoff bound, for any $\alpha > 0$ and any $r \in (0, r_+)$,

$$\Pr\{Z \geq \alpha\} \leq g_Z(r) \exp(-r\alpha) = \exp(\gamma_Z(r) - r\alpha)$$

where $\gamma_Z(r) = \ln g_Z(r)$. If Z is a sum $S_n = X_1 + \dots + X_n$, of IID rv's, then $\gamma_{S_n}(r) = n\gamma_X(r)$.

$$\Pr\{S_n \geq na\} \leq \min_r (\exp[n(\gamma_X(r) - ra)]) .$$

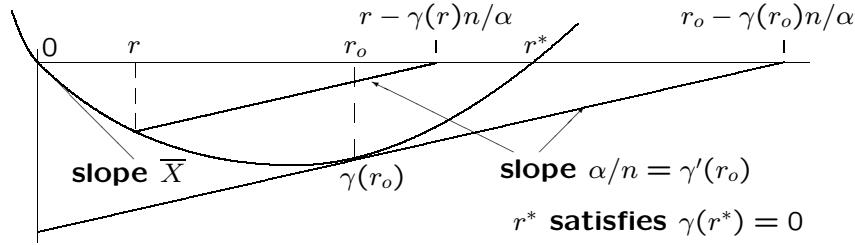
This is exponential in n for fixed a (i.e., $\gamma'(r) = a$). We are now interested in threshold crossings, i.e., $\Pr\{\cup_n(S_n \geq \alpha)\}$. As a preliminary step, we study how $\Pr\{S_n \geq \alpha\}$ varies with n for fixed α .

$$\Pr\{S_n \geq \alpha\} \leq \min_r (\exp[n\gamma_X(r) - r\alpha]) .$$

Here the minimizing r varies with n (i.e., $\gamma'(r) = \alpha/n$).

22

$$\Pr\{S_n \geq \alpha\} \leq \min_{0 < r < r_+} \exp\left(-\alpha \left[r - \frac{n}{\alpha} \gamma_X(r)\right]\right)$$



When n is very large, the slope $\frac{\alpha}{n} = \gamma'_X(r_0)$ is close to 0 and the horizontal intercept (the negative exponent) is very large. As n decreases, the intercept decreases to r^* and then increases again.

Thus $\Pr\{\cup_n\{S_n \geq \alpha\}\} \approx \exp(-\alpha r^*)$, where the nature of the approximation will be explained in terms of the Wald identity.

23

Thm: (Wald) Let $\gamma(r) = \ln(E[\exp(rX)])$ exist over (r_-, r_+) containing 0. Let J be trial at which S_n first exceeds $\alpha > 0$ or $\beta < 0$. Then

$$E[\exp(rS_J - J\gamma(r))] = 1 \quad \text{for } r \in (r_-, r_+)$$

More generally theorem holds if stopping time is a rv under both the given probability and the tilted probability.

The proof simply sums the probabilities of the stopping nodes under both the probability measure and the tilted probability measure.

$$E[S_J] = E[J] \bar{X}$$

$$E[S_J^2] = E[J] \sigma_X^2 \quad \text{if } \bar{X} = 0$$

24

These bounds are all exponentially tight.

If any ϵ is added to any such exponent, the upper bound becomes a lower bound at sufficiently large α with fixed α/n .

The slack in the bounds come partly from the overshoot and partly from the lower threshold.

The lower threshold is unimportant if both thresholds are far from 0.

The overshoot is similar to residual life for renewal processes. It doesn't exist for simple random walks and is easy to calculate if the positive part of the density of X is exponential.

25

Wald's identity for 2 thresholds

Let $\{X_i; i \geq 1\}$ be IID with $\gamma(r) = \ln(E[\exp(rX)])$ for $r \in (r_- < 0 < r_+)$. Let $\{S_n; n \geq 1\}$ be the RW with $S_n = X_1 + \dots + X_n$. If J is the trial at which S_n first crosses $\alpha > 0$ or $\beta < 0$, then

$$E[\exp(rS_J - J\gamma(r))] = 1 \quad \text{for } r \in (r_-, r_+)$$

Corollary: If $\bar{X} < 0$ and $r^* > 0$ satisfies $\gamma(r^*) = 0$, then

$$\Pr\{S_J \geq \alpha\} \leq \exp(-\alpha r^*)$$

26

Review of hypothesis testing: View a binary hypothesis as a binary rv H with $p_H(0) = p_0$ and $p_H(1) = p_1$.

We observe $\{Y_n; n \geq 1\}$, which, conditional on $H = \ell$ is IID with density $f_{Y|H}(Y|\ell)$. Define the log likelihood ratio

$$LLR(\vec{Y}^n) = \sum_{i=1}^n \ln \frac{f_{Y_i|H}(Y_i|0)}{f_{Y_i|H}(Y_i|1)}$$

$$\ln \frac{\Pr\{H=0 | \vec{y}^n\}}{\Pr\{H=1 | \vec{y}^n\}} = \ln \frac{p_0 f_{Y^n|H}(\vec{y}^n | 0)}{p_1 f_{Y^n|H}(\vec{y}^n | 1)} = \ln \frac{p_0}{p_1} + LLR(\vec{y}^n).$$

MAP rule: $LLR(\vec{y}^n) \begin{cases} > \ln p_1/p_0 & ; & \text{select } \hat{h}=0 \\ \leq \ln p_1/p_0 & ; & \text{select } \hat{h}=1. \end{cases}$

27

$$S_n = \sum_{i=1}^n Z_i \quad \text{where } Z_i = \ln \frac{f_{Y_i|H}(Y_i|0)}{f_{Y_i|H}(Y_i|1)}$$

Conditional on $H = 1$, $\{S_n; n \geq 1\}$ is a RW; $S_n = Z_1 + \dots + Z_n$ with the density $f_{Y|H}(y|1)$ where Z is a function of Y . $\gamma_1(r) = \ln E[e^r Z | H = 1]$ and $r^* = 1$ for $\gamma_1(r)$.

Conditional on $H = 0$, $\{S_n; n \geq 1\}$ is a RW; $S_n = Z_1 + \dots + Z_n$ with the density $f_{Y|H}(y|0)$ where Z is the same function of Y .

Under $H = 1$, the RW has negative drift. With thresholds at $\alpha > 0$ and $\beta < 0$, $\Pr\{S_J \geq \alpha\} \leq e^{-\alpha}$. This is the probability of error given $H = 1$ and is essentially the probability of error conditional on any sample path crossing α .

Under $H = 0$, the RW has positive drift and $\Pr\{S_J \leq \beta\} \leq e^\beta$. This also essentially holds for each sample path.

28

MIT OpenCourseWare
<http://ocw.mit.edu>

6.262 Discrete Stochastic Processes

Spring 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.