Course outline

Basics of probability (2 weeks):

- Events, probability laws, etc.
- Independence
- Conditioning
- Expectations.
- Conditional Expectations.

Course outline continued:

Introductions ($\approx$ 1 week each) to:

- Markov Chains
- Poisson Processes
- Point Processes
- Birth and Death Processes
- Queuing Theory
- Brownian motion and diffusions
- Simulation

Student presentations (one week)

Models for coin tossing

Toss coin n times.

On trial k write down a 1 for heads and 0 for tails.

Typical outcome is $\omega = (\omega_1, \ldots, \omega_n)$ a sequence of zeros and ones.

Example: $n = 3$ gives 8 possible outcomes

$$\Omega = \{(0,0,0),(0,0,1),(0,1,0),(0,1,1),\\ (1,0,0),(1,0,1),(1,1,0),(1,1,1)\}\,.$$

General case: set of all possible outcomes is $\Omega = \{0,1\}^n$; $\text{card}(\Omega) = 2^n$.

Meaning of *random* not defined here. Interpretation of probability is usually long run limiting relative frequency (but then we deduce existence of long run limiting relative frequency from axioms of probability).

Probability measure: function P defined on the set of all subsets of Ω such that: with the following properties:

1. For each $A \subset \Omega$, $P(A) \in [0, 1]$.

2. If $A_1, \ldots, A_k$ are *pairwise disjoint* (meaning that for $i \neq j$ the intersection $A_i \cap A_j$ which we usually write as $A_i A_j$ is the empty set $\emptyset$) then

$$P(\cup_1^k A_j) = \sum_1^k P(A_j)$$

3. $P(\Omega) = 1$.

Probability modelling: select family of possible probability measures.

Make best match between mathematics, real world.

interpretation of probability: long run limiting relative frequency

Coin tossing problem: many possible probability measures on Ω.

For $n = 3$, Ω has 8 elements and $2^8 = 256$ subsets.

To specify P: specify 256 numbers. Generally impractical.

Instead: *model* by listing some assumptions about P.

Then deduce what P is, or how to calculate $P(A)$

Three approaches to modelling coin tossing:

1. Counting model:
$$P(A) = \frac{\text{number of elements in } A}{\text{number of elements in } \Omega} \qquad (1)$$
Disadvantage: no insight for other problems.

2. Equally likely elementary outcomes: if $A = \{\omega_1\}$ and $B = \{\omega_2\}$ are two singleton sets in Ω then $P(A) = P(B)$. If card$(\Omega) = m$, say $\Omega = (\omega_1, \ldots, \omega_m)$ then
$$\begin{aligned} P(\Omega) &= P(\cup_1^m \{\omega_j\}) \\ &= \sum_1^m P(\{\omega_j\}) \\ &= mP(\{(\omega_1\}) \end{aligned}$$
So $P(\{\omega_i\}) = 1/m$ and (1) holds.

Defect of models: infinite Ω not easily handled.

Toss coin till first head. Natural Ω is set of all sequences of k zeros followed by a one.

OR: $\Omega = \{0, 1, 2, \ldots\}$.

Can't assume all elements equally likely.

Third approach: model using **independence**:

Coin tossing example: $n = 3$.

Define $A = \{\omega : \omega_1 = 1, \omega_2 = 0, \omega_3 = 1\}$ and

$$A_1 = \{\omega : \omega_1 = 1\}$$
$$A_2 = \{\omega : \omega_2 = 0\}$$
$$A_3 = \{\omega : \omega_3 = 1\}\,.$$

Then $A = A_1 \cap A_2 \cap A_3$

Note $P(A) = 1/8$, $P(A_i) = 1/2$.

So: $P(A) = \prod P(A_i)$

General case: n tosses. $B_i \subset \{0, 1\}$; $i = 1, \ldots, n$

Define

$$A_i = \{\omega : \omega_i \in B_i\} \qquad A = \cap A_i.$$

It is possible to prove that

$$P(A) = \prod P(A_i)$$

Jargon to come later: random variables X_i defined by $X_i(\omega) = \omega_i$ are independent.

Basis of most common modelling tactic.

Assume

$$P(\{\omega : \omega_i = 1\}) = P(\{\omega : \omega_i = 0\}) = 1/2 \quad (2)$$

and for any set of events of form given above

$$P(A) = \prod P(A_i)\,. \quad (3)$$

Motivation: long run rel freq interpretation plus assume outcome of one toss of coin incapable of influencing outcome of another toss.

Advantages: generalizes to infinite Ω.

Toss coin infinite number of times:

$$\Omega = \{\omega = (\omega_1, \omega_2, \cdots)\}$$

is an uncountably infinite set. Model assumes for any n and any event of the form $A = \cap_1^n A_i$ with each $A_i = \{\omega : \omega_i \in B_i\}$ we have

$$P(A) = \prod_1^n P(A_i) \quad (4)$$

For a *fair* coin add the assumption that

$$P(\{\omega : \omega_i = 1\}) = 1/2\,. \quad (5)$$

Is $P(A)$ determined by these assumptions??

Consider $A = \{\omega \in \Omega : (\omega_1, \ldots, \omega_n) \in B\}$ where $B \subset \Omega_n = \{0, 1\}^n$. Our assumptions guarantee

$$P(A) = \frac{\text{number of elements in } B}{\text{number of elements in } \Omega_n}$$

In words, our model specifies that the first n of our infinite sequence of tosses behave like the equally likely outcomes model.

Define C_k to be the event *first head occurs after k consecutive tails*:

$$C_k = A_1^c \cap A_2^c \cdots \cap A_k^c \cap A_{k+1}$$

where $A_i = \{\omega : \omega_i = 1\}$; A^c means complement of A. Our assumption guarantees

$$\begin{aligned} P(C_k) &= P(A_1^c \cap A_2^c \cdots \cap A_k^c \cap A_{k+1}) \\ &= P(A_1^c) \cdots P(A_k^c) P(A_{k+1}) \\ &= 2^{-(k+1)} \end{aligned}$$

Complicated Events: examples

$$A_1 \equiv \{\omega : \lim_{n\to\infty} (\omega_1 + \cdots + \omega_n)/n \text{ exists } \}$$

$$A_2 \equiv \{\omega : \lim_{n\to\infty} (\omega_1 + \cdots + \omega_n)/n = 1/2\}$$

$$A_3 \equiv \{\omega : \lim_{n\to\infty} \sum_1^n (2\omega_k - 1)/k \text{ exists } \}$$

- Strong Law of Large Numbers: for our model $P(A_2) = 1$.

- In fact, $A_3 \subset A_2 \subset A_1$.

- If $P(A_2) = 1$ then $P(A_1) = 1$.

- In fact $P(A_3) = 1$ so $P(A_2) = P(A_1) = 1$.

Some mathematical questions to answer:

1. Do (4) and (5) determine $P(A)$ for every $A \subset \Omega$? [NO]

2. Do (4) and (5) determine $P(A_i)$ for $i = 1, 2, 3$? [YES]

3. Are (4) and (5) logically consistent? [YES]

Probability Definitions

Probability Space (or **Sample Space**): ordered triple $(\Omega, \mathcal{F}, P)$.

- Ω is a set (possible outcomes).

- $\mathcal{F}$ is a family of subsets (**events**) of Ω with the property that $\mathcal{F}$ is a σ-field (or Borel field or σ-algebra):

  1. The empty set $\emptyset$ and Ω are members of $\mathcal{F}$.

  2. $A \in \mathcal{F}$ implies $A^c = \{\omega \in \Omega : \omega \notin A\} \in \mathcal{F}$

  3. $A_1, A_2, \cdots$ all in $\mathcal{F}$ implies

$$A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}\,.$$

- P a function, domain $\mathcal{F}$, range a subset of $[0,1]$ satisfying:

  1. $P(\emptyset) = 0$ and $P(\Omega) = 1$.

  2. **Countable additivity**: $A_1, A_2, \cdots$ **pairwise disjoint** $(j \neq k \implies A_j A_k = \emptyset)$

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Axioms guarantee can compute probabilities by usual rules, including approximation without contradiction.

Consequences:

1. **Finite additivity** $A_1, \cdots, A_n$ pairwise disjoint:

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)\,.$$

2. For any event A $P(A^c) = 1 - P(A)$.

3. If $A_1 \subset A_2 \subset \cdots$ are events then

$$P(\bigcup_1^\infty A_i) = \lim_{n\to\infty} P(A_n)\,.$$

4. If $A_1 \supset A_2 \supset \cdots$ then

$$P(\bigcap_1^\infty A_i) = \lim_{n\to\infty} P(A_n)\,.$$

Most subtle point is σ-field, $\mathcal{F}$. Needed to avoid some contradictions which arise if you try to define $P(A)$ for every subset A of Ω when Ω is a set with uncountably many elements.

Events in Set Notation

Event that Y_n converges to 0 is

$$A \equiv \{\omega : \lim_{n\to\infty} Y_n(\omega) = 0\}$$

Not explicitly written in terms of simple events involving only a finite number of Ys.

Recall basic definition of limit: y_n converges to 0 if $\forall\epsilon > 0$ $\exists N$ such that $\forall n \geq N$ we have $|y_n| \leq \epsilon$.

Convert the definition in A into set theory notation:

- replace y_n by $Y_n(\omega)$,
- replace each *for every* by an intersection
- replace each *there exists* with a union.

We get

$$A = \bigcap_{\epsilon > 0} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{\omega : |Y_n(\omega)| \le \epsilon\}$$

Not obvious A is event because intersection over $\epsilon > 0$ is uncountable.

However, the intersection is countable. Let

$$B_\epsilon \equiv \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{\omega : |Y_n(\omega)| \le \epsilon\}$$

Notice that

$$\epsilon' < \epsilon \implies B_{\epsilon'} \subset B_\epsilon$$

This means that

$$\bigcap_{\epsilon > 0} B_\epsilon = \bigcap_{m=1}^{\infty} B_{1/m}$$

A is countable intersection of countable unions of countable intersections of events, so A is an event.

Random Variables:

Vector valued random variable: function X, domain Ω, range in $\mathbb{R}^p$ such that

$$P(X_1 \le x_1, \ldots, X_p \le x_p)$$

is defined for any constants $(x_1, \ldots, x_p)$. Notation: $X = (X_1, \ldots, X_p)$ and

$$X_1 \le x_1, \ldots, X_p \le x_p$$

is shorthand for an event:

$$\{\omega \in \Omega : X_1(\omega) \le x_1, \ldots, X_p(\omega) \le x_p\}$$

X function on Ω so X_1 function on Ω.

Independence

Events A and B **independent** if

$$P(AB) = P(A)P(B)\,.$$

Events A_i, $i = 1, \ldots, p$ are **independent** if

$$P(A_{i_1} \cdots A_{i_r}) = \prod_{j=1}^{r} P(A_{i_j})$$

for any set of distinct indices $i_1, \ldots, i_r$ between 1 and p.

Example: $p = 3$

$$\begin{aligned} P(A_1A_2A_3) &= P(A_1)P(A_2)P(A_3) \\ P(A_1A_2) &= P(A_1)P(A_2) \\ P(A_1A_3) &= P(A_1)P(A_3) \\ P(A_2A_3) &= P(A_2)P(A_3) \end{aligned}$$

Need all equations to be true for independence!

Example: Toss a coin twice. If A_1 is the event that the first toss is a Head, A_2 is the event that the second toss is a Head and A_3 is the event that the first toss and the second toss are different. then $P(A_i) = 1/2$ for each i and for $i \neq j$

$$P(A_i \cap A_j) = \frac{1}{4}$$

but

$$P(A_1 \cap A_2 \cap A_3) = 0 \neq P(A_1)P(A_2)P(A_3)\,.$$

Rvs $X_1, \ldots, X_p$ are **independent** if

$$P(X_1 \in A_1, \cdots, X_p \in A_p) = \prod P(X_i \in A_i)$$

for any choice of $A_1, \ldots, A_p$.

Theorem 1 *1. If X and Y are independent and discrete then*

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

for all x, y

2. If X and Y are discrete and

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

for **all** *x, y then X and Y are independent.*

Theorem 2 *If $X_1, \ldots, X_p$ are independent and $Y_i = g_i(X_i)$ then $Y_1, \ldots, Y_p$ are independent. Moreover, $(X_1, \ldots, X_q)$ and $(X_{q+1}, \ldots, X_p)$ are independent.*

Proof: The event

$$Y_1 \in A_1 \cap \cdots \cap Y_p \in A_p$$

is exactly the same as the event

$$X_1 \in g_1^{-1}(A_1) \cap \cdots \cap X_p \in g_p^{-1}(A_p)$$

so the first statement is easy.

The second statement is proved using a standard technique:

We must show

$$P\{(X_1, \ldots, X_q) \in A; (X_{q+1}, \ldots, X_p) \in B\} = P\{(X_1, \ldots, X_q) \in A\} P\{(X_{q+1}, \ldots, X_p) \in B\} \quad (6)$$

Proof studies the collections of A, B pairs for which (**??**) holds.

Conditional probability

Important modeling and computation technique:

Def'n: $P(A|B) = P(AB)/P(B)$ if $P(B) \neq 0$.

Def'n: For discrete rvs X, Y conditional pmf of Y given X is

$$\begin{aligned} f_{Y|X}(y|x) &= P(Y = y|X = x) \\ &= f_{X,Y}(x, y)/f_X(x) \\ &= f_{X,Y}(x, y)/\sum_t f_{X,Y}(x, t) \end{aligned}$$

IDEA: used as both computational tool and modelling tactic.

Specify joint distribution by specifying "marginal" and "conditional".

Modelling:

Assume $X \sim$ Poisson(λ).

Assume $Y|X \sim$Binomial(X, p).

Let $Z = X - Y$. Joint law of Y, Z?

$$
\begin{aligned}
& P(Y = y, Z = z) \\
&\quad = P(Y = y, X - Y = z) \\
&\quad = P(Y = y, X = z + y) \\
&\quad = P(Y = y | X = y + z) P(X = y + z) \\
&\quad = \binom{z+y}{y} p^y (1-p)^z e^{-\lambda} \lambda^{z+y} / (z+y)! \\
&\quad = \exp\{-p\lambda\} \frac{(p\lambda)^y}{y!} \exp\{(1-p)\lambda\} \frac{\{(1-p)\lambda\}^z}{z!}
\end{aligned}
$$

So: Y, Z independent Poissons.

Expected Value

Undergraduate definition of E: integral for absolutely continuous X, sum for discrete. But: $\exists$ rvs which are neither absolutely continuous nor discrete.

General definition of E.

A random variable X is **simple** if we can write

$$X(\omega) = \sum_1^n a_i 1(\omega \in A_i)$$

for some constants $a_1, \ldots, a_n$ and events A_i.

Def'n: For a simple rv X we define

$$E(X) = \sum a_i P(A_i)$$

For positive random variables which are not simple we extend our definition by approximation:

Def'n: If $X \geq 0$ (almost surely, $P(X \geq 0) = 1$) then

$$E(X) = \text{sup}\{E(Y) : 0 \leq Y \leq X, Y \text{ simple}\}$$

Def'n: We call X **integrable** if

$$E(|X|) < \infty\,.$$

In this case we define

$$E(X) = E(\max(X, 0)) - E(\max(-X, 0))$$

Facts: E is a linear, monotone, positive operator:

1. **Linear**: $E(aX+bY) = aE(X)+bE(Y)$ provided X and Y are integrable.

2. **Positive**: $P(X \geq 0) = 1$ implies $E(X) \geq 0$.

3. **Monotone**: $P(X \geq Y) = 1$ and X, Y integrable implies $E(X) \geq E(Y)$.

Major technical theorems:

Monotone Convergence: If $0 \le X_1 \le X_2 \le \cdots$ a.s. and $X = \lim X_n$ (which exists a.s.) then

$$E(X) = \lim_{n\to\infty} E(X_n)$$

Dominated Convergence: If $|X_n| \le Y_n$ and $\exists$ rv X st $X_n \to X$ a.s. and rv Y st $Y_n \to Y$ with $E(Y_n) \to E(Y) < \infty$ then

$$E(X_n) \to E(X)$$

Often used with all Y_n the same rv Y.

Fatou's Lemma: If $X_n \ge 0$ then

$$E(\liminf X_n) \le \liminf E(X_n)$$

Theorem: With this definition of E if X has density $f(x)$ (even in $\mathbb{R}^p$ say) and $Y = g(X)$ then

$$E(Y) = \int g(x)f(x)dx\,.$$

(This could be a multiple integral.)

Works even if X has density but Y doesn't.

If X has pmf f then

$$E(Y) = \sum_x g(x)f(x)\,.$$

Def'n: r^{th} moment (about origin) of a real rv X is $\mu'_r = E(X^r)$ (provided it exists). Generally use μ for $E(X)$. The r^{th} central moment is

$$\mu_r = E[(X-\mu)^r]$$

Call $\sigma^2 = \mu_2$ the variance.

Def'n: For an $\mathbb{R}^p$ valued rv X $\mu_X = E(X)$ is the vector whose i^{th} entry is $E(X_i)$ (provided all entries exist).

Def'n: The $(p \times p)$ variance covariance matrix of X is

$$Var(X) = E\left[(X-\mu)(X-\mu)^t\right]$$

which exists provided each component X_i has a finite second moment. More generally if $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ both have all components with finite second moments then

$$\text{Cov}(X,Y) = \text{E}\left[(X-\mu_X)(Y-\mu_Y)^T\right]$$

We have

$$\text{Cov}(AX+a, BY+b) = A\text{Cov}(X,Y)B^T$$

for general (conforming) matrices A, B and vectors a and b.

Moments and probabilities of rare events are closely connected.

Markov's inequality (one case is Chebyshev's inequality):

$$
\begin{aligned}
P(|X-\mu| \geq t) &= E[1(|X-\mu| \geq t)] \\
&\leq E\left[\frac{|X-\mu|^r}{t^r} 1(|X-\mu| \geq t)\right] \\
&\leq \frac{E[|X-\mu|^r]}{t^r}
\end{aligned}
$$

Intuition: if moments are small then large deviations from average are unlikely.

Moments and independence

Theorem: If $X_1, \ldots, X_p$ are independent and each X_i is integrable then $X = X_1 \cdots X_p$ is integrable and

$$E(X_1 \cdots X_p) = E(X_1) \cdots E(X_p)$$

Proof: Usual order: simple Xs first, then positive, then integrable.

Suppose each X_i is simple:

$$X_i = \sum_j x_{ij} 1(X_i = x_{ij})$$

where the x_{ij} are the possible values of X_i.

Then

$$
\begin{aligned}
E(X_1 \cdots X_p) & \\
&= \sum_{j_1 \cdots j_p} x_{1j_1} \cdots x_{pj_p} \times \\
&\qquad E(1(X_1 = x_{1j_1}) \cdots 1(X_p = x_{pj_p}) \\
&= \sum_{j_1 \cdots j_p} x_{1j_1} \cdots x_{pj_p} \times \\
&\qquad P(X_1 = x_{1j_1} \cdots X_p = x_{pj_p}) \\
&= \sum_{j_1 \cdots j_p} x_{1j_1} \cdots x_{pj_p} \times \\
&\qquad P(X_1 = x_{1j_1}) \cdots P(X_p = x_{pj_p}) \\
&= \left[\sum_{j_1} x_{1j_1} P(X_1 = x_{1j_1})\right] \times \cdots \times \\
&\qquad \left[\sum_{j_p} x_{pj_p} P(X_p = x_{pj_p})\right] \\
&= \prod E(X_i)
\end{aligned}
$$

General $X_i \geq 0$: $X_{i,n}$ is X_i rounded down to the nearest multiple of 2^{-n} (to a maximum of n). Each $X_{i,n}$ is simple and $X_{1,n}, \ldots, X_{p,n}$ are independent. Thus

$$\mathrm{E}(\prod X_{j,n}) = \prod \mathrm{E}(X_{j,n})$$

for each n. If

$$X_n^* = \prod X_{j,n}$$

then

$$0 \leq X_1^* \leq X_2^* \leq \cdots$$

and X_n^* converges to $X^* = \prod X_i$ so that

$$\mathrm{E}(X^*) = \lim \mathrm{E}(X_n^*)$$

by monotone convergence. Also by monotone convergence

$$\lim \prod \mathrm{E}(X_{j,n}) = \prod \mathrm{E}(X_j)) < \infty$$

This shows both that X^* is integrable and that

$$E(\prod X_j) = \prod \mathsf{E}(X_j)$$

The general case uses the fact that we can write each X_i as the difference of its positive and negative parts:

$$X_i = \max(X_i, 0) - \max(-X_i, 0)$$

Just expand out the product and use the previous case.

Multiple Integration: Lebesgue integrals over $\mathbb{R}^p$ defined using Lebesgue measure on $\mathbb{R}^p$.

Iterated integrals wrt Lebesgue measure on $\mathbb{R}^1$ give same answer.

Theorem[Tonelli]: If $f : \mathbb{R}^{p+q} \mapsto \mathbb{R}$ is Borel and $f \geq 0$ almost everywhere then for almost every $x \in \mathbb{R}^p$ the integral

$$g(x) \equiv \int f(x,y)dy$$

exists and

$$\int g(x)dx = \int f(x,y)dxdy$$

RHS denotes $p+q$ dimensional integral defined previously.

Theorem[Fubini] If $f : \mathbb{R}^{p+q} \mapsto \mathbb{R}$ is Borel and integrable then for almost every $x \in \mathbb{R}^p$ the integral

$$g(x) \equiv \int f(x, y) dy$$

exists and is finite. Moreover g is integrable and

$$\int g(x) dx = \int f(x, y) dx dy \, .$$

Results true for measures other than Lebesgue.

Conditional distributions, expectations

When X and Y are discrete we have

$$\mathsf{E}(Y|X=x) = \sum_y yP(Y=y|X=x)$$

for any x for which $P(X=x)$ is positive.

Defines a function of x.

This function evaluated at X gives rv which is ftn of X denoted

$$\mathsf{E}(Y|X).$$

Example: $Y|X = x \sim$ Binomial(x, p). Since mean of a Binomial(n, p) is np we find

$$\mathsf{E}(Y|X = x) = px$$

and

$$\mathsf{E}(Y|X) = pX$$

Notice you simply replace x by X.

Here are some properties of the function

$$\mathsf{E}(Y|X=x)$$

1) Suppose A is a function defined on the range of X. Then

$$\mathsf{E}(A(X)Y|X=x) = A(x)\mathsf{E}(Y|X=x)$$

and so

$$\mathsf{E}(A(X)Y|X) = A(X)\mathsf{E}(Y|X)$$

Second assertion follows from first. Note that if $Z = A(X)Y$ then Z is discrete and

$$P(Z=z) = \sum_{x,y} P(Y=y, X=x)1(z = A(x)y)$$

Also

$$\begin{aligned} &P(Z=z|X=x) \\ &\quad = \frac{\sum_y P(Y=y, X=x)1(z=A(x)y)}{P(X=x)} \\ &\quad = \sum_y P(Y=y|X=x)1(z=A(x)y) \end{aligned}$$

Thus

$$
\begin{aligned}
&\mathsf{E}(Z|X = x) \\
&\quad = \sum_z zP(Z = z|X = x) \\
&\quad = \sum_z \sum_y zP(Y = y|X = x)\mathsf{1}(z = A(x)y) \\
&\quad = \sum_z \sum_y A(x)yP(Y = y|X = x)\mathsf{1}(z = A(x)y) \\
&\quad = A(x)\sum_y yP(Y = y|X = x)\sum_z \mathsf{1}(z = A(x)y) \\
&\quad = A(x)\sum_y yP(Y = y|X = x)
\end{aligned}
$$

2) Repeated conditioning: if X, Y and Z discrete then

$$\mathsf{E}\{\mathsf{E}(Z|X,Y)|X\} = \mathsf{E}(Z|X)$$
$$\mathsf{E}\{\mathsf{E}(Y|X)\} = \mathsf{E}(Y)$$

3) Additivity

$$\mathsf{E}(Y+Z|X) = \mathsf{E}(Y|X) + \mathsf{E}(Z|X)$$

4) Putting the first two items together gives

$$\begin{aligned} &\mathsf{E}\{\mathsf{E}(A(X)Y|X)\} = \\ &\mathsf{E}\{A(X)\mathsf{E}(Y|X)\} = \mathsf{E}(A(X)Y) \end{aligned} \tag{7}$$

Definition of E$(Y|X)$ when X and Y are not assumed to discrete:

E$(Y|X)$ is rv which is measurable function of X satisfying(**??**).

Existence is measure theory problem.

Suppose X is discrete and $X^* = g(X)$ is a one to one transformation of X. Since $X = x$ is the same event as $X^* = g(x)$ we find

$$\mathsf{E}(Y|X = x) = \mathsf{E}(Y|X^* = g(x))$$

Let $h^*(u)$ denote the function $\mathsf{E}(Y|X^* = u)$ and $h(u) = \mathsf{E}(Y|X = u)$. Then

$$h(x) = h^*(g(x))$$

Thus

$$h(X) = h^*(g(X)) = h^*(X^*)$$

This just means

$$\mathsf{E}(Y|X) = \mathsf{E}(Y|X^*)$$

Interpretation.

Formula is “obvious”.

Example: Toss coin $n = 20$ times. Y is indicator of first toss is a heads. X is number of heads and X^* number of tails. Formula says:

$$\mathrm{E}(Y|X = 17) = \mathrm{E}(Y|X^* = 3)$$

In fact for a general k and n

$$\mathrm{E}(Y|X = k) = \frac{k}{n}$$

so

$$\mathrm{E}(Y|X) = \frac{X}{n}$$

At the same time

$$\mathrm{E}(Y|X^* = j) = \frac{n-j}{n}$$

so

$$\mathrm{E}(Y|X^*) = \frac{n-X^*}{n}$$

But of course $X = n - X^*$ so these are just two ways of describing the same random variable.

Another interpretation: Rv X partitions Ω into countable set of events of the form $X = x$.

Other rv X^* partitions Ω into the same events.

Then values of $\mathsf{E}(Y|X^* = x^*)$ are same as values of $\mathsf{E}(Y|X = x)$ but labelled differently.

To form $\mathsf{E}(Y|X)$ take value ω, compute $X(\omega)$ to determine member A of the partition we being conditionsed on, then write down corresponding $\mathsf{E}(Y|A)$.

Hence conditional expectation depends only on partition of Ω.

X not discrete: replace partition with σ-field. Suppose X and X^* 2 rvs such that $\mathcal{F}(X) = \mathcal{F}(X^*)$. Then:

- There is g Borel,one to one with one to one Borel inverse s.t. $X^* = g(X)$.

- $\text{E}(Y|X) = \text{E}(Y|X^*)$ almost surely.

In other words $\text{E}(Y|X)$ depends *only* on the σ-field generated by X. We write

$$\text{E}(Y|\mathcal{F}(X)) = \text{E}(Y|X)$$

Def'n: Suppose $\mathcal{G}$ is sub-σ-field of $\mathcal{F}$. X is $\mathcal{G}$ measurable if, for every Borel B

$$\{\omega : X(\omega) \in B\} \in \mathcal{G}\,.$$

Def'n: $\text{E}(Y|\mathcal{G})$ is any $\mathcal{G}$ measurable rv s.t. for every $\mathcal{G}$ measurable rv variable A we have

$$\text{E}(AY) = \text{E}\left\{A\text{E}(Y|\mathcal{G})\right\}\,.$$

Again existence is measure theory problem.

Markov Chains

Stochastic process: family $\{X_i; i \in I\}$ of rvs I the **index set**. Often $I \subset \mathbb{R}$, e.g. $[0, \infty)$, $[0, 1]$ $\mathbb{Z}$ or $\mathbb{N}$.

Continuous time: I is an interval

Discrete time: $I \subset \mathbb{Z}$.

Generally all X_n take values in **state space** S. In following S is a finite or countable set; each X_n is discrete.

Usually S is $\mathbb{Z}$, $\mathbb{N}$ or $\{0, \ldots, m\}$ for some finite m.

Markov Chain: stochastic process $X_n; n \in \mathbb{N}$. taking values in a finite or countable set S such that for every n and every event of the form

$$A = \{(X_0, \ldots, X_{n-1}) \in B \subset S^n\}$$

we have

$$P(X_{n+1} = j | X_n = i, A) = P(X_1 = j | X_0 = i) \tag{8}$$

Notation: $\mathbf{P}$ is the (possibly infinite) array with elements

$$P_{ij} = P(X_1 = j | X_0 = i)$$

indexed by $i, j \in S$.

$\mathbf{P}$ is the (one step) **transition matrix** of the Markov Chain.

WARNING: in (6) we require the condition to hold **only** when

$$P(X_n = i, A) > 0$$

Evidently the entries in $\mathbf{P}$ are non-negative and

$$\sum_j P_{ij} = 1$$

for all $i \in S$. Any such matrix is called **stochastic**.

We define powers of $\mathbf{P}$ by

$$(\mathbf{P}^n)_{ij} = \sum_k \left(\mathbf{P}^{n-1}\right)_{ik} P_{kj}$$

Notice that even if S is infinite these sums converge absolutely.

Chapman-Kolmogorov Equations

Condition on X_{l+n-1} to compute

$$P(X_{l+n} = j | X_l = i)$$

$$\begin{aligned}
P(X_{l+n} &= j | X_l = i) \\
&= \sum_k P(X_{l+n} = j, X_{l+n-1} = k | X_l = i) \\
&= \sum_k P(X_{l+n} = j | X_{l+n-1} = k, X_l = i) \\
&\qquad \times P(X_{l+n-1} = k | X_l = i) \\
&= \sum_k P(X_1 = j | X_0 = k) \\
&\qquad \times P(X_{l+n-1} = k | X_l = i) \\
&= \sum_k P(X_{l+n-1} = k | X_l = i) \mathbf{P}_{kj}
\end{aligned}$$

Now condition on X_{l+n-2} to get

$$P(X_{l+n} = j | X_l = i) = \sum_{k_1 k_2} \mathbf{P}_{k_1 k_2} \mathbf{P}_{k_2 j} P(X_{l+n-2} = k_1 | X_l = i)$$

Notice: sum over k_2 computes k_1, j entry in matrix $\mathbf{PP} = \mathbf{P}^2$.

$$P(X_{l+n} = j | X_l = i) = \sum_{k_1} (\mathbf{P}^2)_{k_1,j} P(X_{l+n-2} = k_1 | X_l = i)$$

We may now prove by induction on n that

$$P(X_{l+n} = j | X_l = i) = (\mathbf{P}^n)_{ij} .$$

This proves Chapman-Kolmogorov equations:

$$P(X_{l+m+n} = j | X_l = i) = \sum_k P(X_{l+m} = k | X_l = i) \times P(X_{l+m+n} = j | X_{l+m} = k)$$

These are simply a restatement of the identity

$$\mathbf{P}^{n+m} = \mathbf{P}^n \mathbf{P}^m .$$

Remark: It is important to notice that these probabilities depend on m and n but **not** on l. We say the chain has **stationary** transition probabilities. A more general definition of Markov chain than (6) is

$$P(X_{n+1} = j | X_n = i, A) \\ = P(X_{n+1} = j | X_n = i) \, .$$

Notice RHS now permitted to depend on n.

Define $\mathbf{P}^{n,m}$: matrix with i,jth entry

$$P(X_m = j | X_n = i)$$

for $m > n$. Then

$$\mathbf{P}^{r,s}\mathbf{P}^{s,t} = \mathbf{P}^{r,t}$$

Also called Chapman-Kolmogorov equations. This chain does not have stationary transitions.

Remark: The calculations above involve sums in which all terms are positive. They therefore apply even if the state space S is countably infinite.

Extensions of the Markov Property

Function $f(x_0, x_1, \ldots)$ defined on S^∞ = all infinite sequences of points in S.

Let B_n be the event

$$f(X_n, X_{n+1}, \ldots) \in C$$

for suitable C in range space of f. Then

$$P(B_n|X_n = x, A) = P(B_0|X_0 = x) \qquad (9)$$

for any event A of the form

$$\{(X_0, \ldots, X_{n-1}) \in D\}$$

Also

$$P(AB_n|X_n = x) = P(A|X_n = x)P(B_n|X_n = x) \qquad (10)$$

"Given the present the past and future are conditionally independent."

Proof of (7):

Special case:

$$B_n = \{(X_{n+1} = x_1, \cdots, X_{n+m} = x_m\}$$

LHS of (7) evaluated by repeated conditioning (cf. Chapman-Kolmogorov):

$$\mathbf{P}_{x,x_1}\mathbf{P}_{x_1,x_2}\cdots\mathbf{P}_{x_{m-1},x_m}$$

Same for RHS.

Events defined from $X_n, \ldots, X_{n+m}$: sum over appropriate vectors $x, x_1, \ldots, x_m$.

General case: monotone class techniques.

To prove (8) write

$$\begin{aligned} P(AB_n|X_n = x) \\ &= P(B_n|X_n = x, A)P(A|X_n = x) \\ &= P(B_n|X_n = x)P(A|X_n = x) \end{aligned}$$

using (7).

Classification of States

If an entry $\mathbf{P}_{ij}$ is 0 it is not possible to go from state i to state j in one step. It may be possible to make the transition in some larger number of steps, however. We say i **leads to** j (or j is accessible from i) if there is an integer $n \geq 0$ such that

$$P(X_n = j | X_0 = i) > 0\,.$$

We use the notation $i \rightsquigarrow j$. Define $\mathbf{P}^0$ to be identity matrix $\mathbf{I}$. Then $i \rightsquigarrow j$ if there is an $n \geq 0$ for which $(\mathbf{P}^n)_{ij} > 0$.

States i and j **communicate** if $i \rightsquigarrow j$ and $j \rightsquigarrow i$.

Write $i \leftrightarrow j$ if i and j communicate.

Communication is an equivalence relation: reflexive, symmetric, transitive relation on states of S.

More precisely:

Reflexive: for all i we have $i \leftrightarrow j$.

Symmetric: if $i \leftrightarrow j$ then $j \leftrightarrow i$.

Transitive: if $i \leftrightarrow j$ and $j \leftrightarrow k$ then $i \leftrightarrow k$.

Proof:

Reflexive: follows from inclusion of $n = 0$ in definition of leads to.

Symmetry is obvious.

Transitivity: suffices to check that $i \rightsquigarrow j$ and $j \rightsquigarrow k$ imply that $i \rightsquigarrow k$. But if $(\mathbf{P}^m)_{ij} > 0$ and $(\mathbf{P}^n)_{jk} > 0$ then

$$\begin{aligned}(\mathbf{P}^{m+n})_{ik} &= \sum_l (\mathbf{P}^m)_{il}(\mathbf{P}^n)_{lk} \\ &\geq (\mathbf{P}^m)_{ij}(\mathbf{P}^n)_{jk} \\ &> 0\end{aligned}$$

Any equivalence relation on a set partitions the set into **equivalence classes**; two elements are in the same equivalence class if and only if they are equivalent.

Communication partitions S into equivalence classes called **communicating classes**.

Example:

$$\mathbf{P} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Find communicating classes: start with say state 1, see where it leads.

- $1 \rightsquigarrow 2$, $1 \rightsquigarrow 3$ and $1 \rightsquigarrow 4$ in row 1.

- Row 4: $4 \rightsquigarrow 1$. So: (transitivity) 1, 2, 3 and 4 all in the same communicating class.

- Claim: none of these leads to 5, 6, 7 or 8.

  Suppose $i \in \{1,2,3,4\}$ and $j \in \{5,6,7,8\}$. Then $(\mathbf{P}^n)_{ij}$ is sum of products of $\mathbf{P}_{kl}$. Cannot be positive unless there is a sequence $i_0 = i, i_1, \ldots, i_n = j$ with $\mathbf{P}_{i_{k-1},i_k} > 0$ for $k = 1, \ldots, n$.

  Consider first k for which $i_k \in \{5,6,7,8\}$ Then $i_{k-1} \in \{1,2,3,4\}$ and so $\mathbf{P}_{i_{k-1},i_k} = 0$.

So: $\{1, 2, 3, 4\}$ is a communicating class.

- $5 \rightsquigarrow 1$, $5 \rightsquigarrow 2$, $5 \rightsquigarrow 3$ and $5 \rightsquigarrow 4$.

- None of these lead to any of $\{5, 6, 7, 8\}$ so $\{5\}$ must be communicating class.

- Similarly $\{6\}$ and $\{7, 8\}$ are communicating classes.

Note: states 5 and 6 have special property. Each time you are in either state you run a risk of going to one of the states 1, 2, 3 or 4. Eventually you will make such a transition and then never return to state 5 or 6.

States 5 and 6 are **transient**.

To make this precise define hitting times:

$$T_k = \mathsf{min}\{n > 0 : X_n = k\}$$

We define

$$f_k = P(T_k < \infty | X_0 = k)$$

State k is **transient** if $f_k < 1$ and **recurrent** if $f_k = 1$.

Let N_k be number of times chain is ever in state k.

Claims:

1. If $f_i < 1$ then N_k has a Geometric distribution:

$$P(N_k = r|X_0 = k) = f_k^{r-1}(1 - f_k)$$

for $r = 1, 2, \ldots$.

2. If $f_i = 1$ then

$$P(N_k = \infty|X_0 = k) = 1$$

Proof using **Strong Markov Property**:

Stopping time for the Markov chain is a random variable T taking values in $\{0, 1, \cdots\} \cup \{\infty\}$ such that for each finite k there is a function f_k such that

$$1(T = k) = f_k(X_0, \ldots, X_k)$$

Notice that T_k in theorem is a stopping time.

Standard shorthand notation: by

$$P^x(A)$$

we mean

$$P(A | X_0 = x) \, .$$

Similarly we define

$$\mathsf{E}^x(Y) = \mathsf{E}(Y | X_0 = x) \, .$$

strong Markov property:

Stopping time for the Markov chain is a random variable T taking values in $\{0, 1, \cdots\} \cup \{\infty\}$ such that for each finite k there is a function f_k such that

$$1(T = k) = f_k(X_0, \ldots, X_k)$$

Notice that T_k in theorem is a stopping time.

Standard shorthand notation: by

$$P^x(A)$$

we mean

$$P(A|X_0 = x).$$

Similarly we define

$$\mathrm{E}^x(Y) = \mathrm{E}(Y|X_0 = x).$$

Goal: explain and prove

$$\mathsf{E}(f(X_T,\ldots)|X_T,\ldots,X_0) = \mathsf{E}^{X_T}(f(X_0,\ldots))$$

Simpler claim:

$$P(X_{T+1}=j|X_T=i) = \mathbf{P}_{ij} = P^i(X_1=j)\,.$$

Notation: $A_k = \{X_k = i, T = k\}$

Notice: $A_k = \{X_T = k, T = k\}$:

$$\begin{aligned}
P(X_{T+1}=j|X_T=i) &= \frac{P(X_{T+1}=j, X_T=i)}{P(X_T=i)} \\
&= \frac{\sum_k P(X_{T+1}=j, X_T=i, T=k)}{\sum_k P(X_T=i, T=k)} \\
&= \frac{\sum_k P(X_{k+1}=j, A_k)}{\sum_k P(A_k)} \\
&= \frac{\sum_k P(X_{k+1}=j|A_k)P(A_k)}{\sum_k P(A_k)} \\
&= \frac{\sum_k P(X_1=j|X_0=i)P(A_k)}{\sum_k P(A_k)} \\
&= \mathbf{P}_{i,j}
\end{aligned}$$

Notice use of fact that $T = k$ is event defined in terms of $X_0, \ldots, X_k$.

Technical problems with proof:

• It might be that $P(T = \infty) > 0$. What are X_T and X_{T+1} on the event $T = \infty$.

Answer: condition also on $T < \infty$.

• Prove formula only for stopping times where $\{T < \infty\} \cap \{X_T = i\}$ has positive probability.

We will now fix up these technical details.

Suppose $f(x_0, x_1, \ldots)$ is a (measurable) function on $S^{\mathbb{N}}$. Put

$$Y_n = f(X_n, X_{n+1}, \ldots) \, .$$

Assume $\mathrm{E}(|Y_0| \, | X_0 = x) < \infty$ for all x. Claim:

$$\mathrm{E}(Y_n | X_n, A) = \mathrm{E}^{X_n}(Y_0) \qquad (11)$$

whenever A is any event defined in terms of $X_0, \ldots, X_n$.

Proof:

1 Family of f for which claim holds includes all indicators; see extension of Markov Property in previous lecture.

2 family of f for which claim is true is vector space (so if f, g in family then so is $af + bg$ for any constants a and b.

- So family of f for which claim is true includes all simple functions.

- family of f for which claim true is closed under monotone increasing limits (of non-negative f_n) by the Monotone Convergence theorem.

- So claim true for every non-negative integrable f.

- Claim follows for integrable f by linearity.

Aside on "measurable": what sorts of events can be defined in terms of a family $\{Y_i : i \in I\}$?

Natural: any event of form $(Y_{i_1},\ldots,Y_{i_k}) \in C$ is "defined in terms of the family" for any finite set $i_1,\ldots,i_k$ and any (Borel) set C in S^k.

For countable S: each singleton $(s_1,\ldots,s_k) \in S^k$ Borel. So every subset of S^k Borel.

Natural: if you can define each of a sequence of events A_n in terms of the Ys then the definition "there exists an n such that (definition of A_n) ..." defines $\cup A_n$.

Natural: if A is definable in terms of the Ys then A^c can be defined from the Ys by just inserting the phrase "It is not true that" in front of the definition of A.

So family of events definable in terms of the family $\{Y_i : i \in I\}$ is a σ-field which includes every event of the form $(Y_{i_1},\ldots,Y_{i_k}) \in C$. We call the smallest such σ-field, $\mathcal{F}(\{Y_i : i \in I\})$, the σ-field generated by the family $\{Y_i : i \in I\}$.

Using the Markov property:

Toss coin till I get a head. What is the expected number of tosses?

Define state to be 0 if toss is tail and 1 if toss is heads.

Define $X_0 = 0$.

Let $N = \min\{n > 0 : X_n = 1\}$. Want

$$\text{E}(N) = \text{E}^0(N)$$

Note: if $X_1 = 1$ then $N = 1$. If $X_1 = 0$ then $N = 1 + \min\{n > 0 : X_{n+1} = 1\}$.

In symbols:

$$N = \min\{n > 0 : X_n = 1\} = f(X_1, X_2, \cdots)$$

and

$$N = 1 + 1(X_1 = 0) f(X_2, X_3, \cdots)$$

Take expected values starting from 0:

$$\mathsf{E}^0(N) = 1 + \mathsf{E}^0\{1(X_1 = 0)f(X_2, X_3, \cdots)\}$$

Condition on X_1 and get

$$\mathsf{E}^0(N) = 1 + \mathsf{E}^0[\mathsf{E}\{1(X_1 = 0)f(X_2, \cdots)|X_1\}]$$

But

$$\begin{aligned}\mathsf{E}\{1(X_1 = 0)f(X_2, X_3, \cdots)|X_1\}& \\ &= 1(X_1 = 0)\mathsf{E}^{X_1}\{f(X_1, X_2, \cdots)\} \\ &= 1(X_1 = 0)\mathsf{E}^0\{f(X_1, X_2, \cdots)\} \\ &= 1(X_1 = 0)\mathsf{E}^0(N)\end{aligned}$$

so that

$$\mathsf{E}^0(N) = 1 + p\mathsf{E}^0\{N\}$$

where p is the probability of tails. Solve for $\mathsf{E}(N)$ to get

$$\mathsf{E}(N) = \frac{1}{1-p}$$

This is the formula for expected value of the sort of geometric which starts at 1 and has p being the probability of failure.

Initial Distributions

Meaning of unconditional expected values?

Markov property specifies only cond'l probs; no way to deduce marginal distributions.

For every dstbn π on S and transition matrix $\mathbf{P}$ there is a a stochastic process $X_0, X_1, \ldots$ with

$$P(X_0 = k) = \pi_k$$

and which is a Markov Chain with transition matrix $\mathbf{P}$.

Note Strong Markov Property proof used only conditional expectations.

Notation: π a probability on S. E^π and P^π are expected values and probabilities for chain with initial distribution π.

Summary of easily verified facts:

- For any sequence of states $i_0, \ldots, i_k$

  $$P(X_0 = i_0, \ldots, X_k = i_k) = \pi_{i_0} \mathbf{P}_{i_0 i_1} \cdots \mathbf{P}_{i_{k-1} i_k}$$

- For any event A:

  $$\mathbf{P}^{\pi}(A) = \sum_k \pi_k \mathbf{P}^k(A)$$

- For any bounded rv $Y = f(X_0, \ldots)$

  $$\mathsf{E}^{\pi}(Y) = \sum_k \pi_k \mathsf{E}^k(A)$$

Recurrence and Transience

Now consider a transient state k, that is, a state for which

$$f_k = P^k(T_k < \infty) < 1$$

Note that $T_k = \min\{n > 0 : X_n = k\}$ is a stopping time. Let N_k be the number of visits to state k. That is

$$N_k = \sum_{n=0}^{\infty} 1(X_n = k)$$

Notice that if we define the function

$$f(x_0, x_1, \ldots) = \sum_{n=0}^{\infty} 1(x_n = k)$$

then

$$N_k = f(X_0, X_1, \ldots)$$

Notice, also, that on the event $T_k < \infty$

$$N_k = 1 + f(X_{T_k}, X_{T_k+1}, \ldots)$$

and on the event $T_k = \infty$ we have

$$N_k = 1$$

In short:

$$N_k = 1 + f(X_{T_k}, X_{T_k+1}, \ldots) 1(T_k < \infty)$$

Hence

$$\begin{aligned}
&\mathbf{P}^k(N_k = r) \\
&\quad = \mathsf{E}^k \{P(N_k = r | \mathcal{F}_T)\} \\
&\quad = \mathsf{E}^k \Big[P \Big\{ 1 + f(X_{T_k}, X_{T_k+1}, \ldots) \\
&\qquad\qquad \times 1(T_k < \infty) = r | \mathcal{F}_T \}] \\
&\quad = \mathsf{E}^k [1(T_k < \infty) \\
&\qquad\qquad \times P^{X_{T_k}} \{ f(X_0, X_1, \ldots) = r - 1 \} \Big] \\
&\quad = \mathsf{E}^k \Big\{ 1(T_k < \infty) P^k(N_k = r - 1) \Big\} \\
&\quad = \mathsf{E}^k \{ 1(T_k < \infty) \} P^k(N_k = r - 1) \\
&\quad = f_k P^k(N_k = r - 1)
\end{aligned}$$

It is easily verified by induction, then, that

$$\mathbf{P}^k(N_k = r) = f_k^{r-1} P^k(N_k = 1)$$

But $N_k = 1$ if and only if $T_k = \infty$ so

$$\mathbf{P}^k(N_k = r) = f_k^{r-1}(1 - f_k)$$

so N_k has (chain starts from k) Geometric dist'n, mean $1/(1 - f_k)$. Argument also shows that if $f_k = 1$ then

$$P^k(N_k = 1) = P^k(N_k = 2) = \cdots$$

which can only happen if all these probabilities are 0. Thus if $f_k = 1$

$$P(N_k = \infty) = 1$$

Since

$$N_k = \sum_{n=0}^{\infty} 1(X_n = k)$$

$$\mathsf{E}^k(N_k) = \sum_{n=0}^{\infty} (\mathbf{P}^n)_{kk}$$

So: State k is transient if and only if

$$\sum_{n=0}^{\infty} (\mathbf{P}^n)_{kk} < \infty$$

and this sum is $1/(1 - f_k)$.

Proposition 1 *Recurrence (or transience) is a class property. That is, if i and j are in the same communicating class then i is recurrent (respectively transient) if and only if j is recurrent (respectively transient).*

Proof: Suppose i is recurrent and $i \leftrightarrow j$. There are integers m and n such that

$$(\mathbf{P}^m)_{ji} > 0 \quad \text{and} \quad (\mathbf{P}^n)_{ij} > 0$$

Then

$$\begin{aligned}
\sum_k (\mathbf{P}^k)_{jj} &\geq \sum_{k \geq 0} (\mathbf{P}^{m+k+n})_{jj} \\
&\geq \sum_{k \geq 0} (\mathbf{P}^m)_{ji} (\mathbf{P}^k)_{ii} (\mathbf{P}^n)_{ij} \\
&= (\mathbf{P}^m)_{ji} \left\{ \sum_{k \geq 0} (\mathbf{P}^k)_{ii} \right\} (\mathbf{P}^n)_{ij}
\end{aligned}$$

The middle term is infinite and the two outside terms positive so

$$\sum_k (\mathbf{P}^k)_{jj} = \infty$$

which shows j is recurrent.

A finite state space chain has at least one recurrent state:

If all states we transient we would have for each k $P(N_k < \infty) = 1$. This would mean $P(\forall k\, .N_k < \infty) = 1$. But for any ω there must be at least one k for which $N_k = \infty$ (the total of a finite list of finite numbers is finite).

Infinite state space chain may have all states transient:

The chain X_n satisfying $X_{n+1} = X_n + 1$ on the integers has all states transient.

More interesting example:

- Toss a coin repeatedly.

- Let X_n be X_0 plus the number of heads minus the number of tails in the first n tosses.

- Let p denote the probability of heads on an individual trial.

$X_n - X_0$ is a sum of n iid random variables Y_i where $P(Y_i = 1) = p$ and $P(Y_i = -1) = 1 - p$.

SLLN shows X_n/n converges almost surely to $2p - 1$. If $p \neq 1/2$ this is not 0.

In order for X_n/n to have a positive limit we must have $X_n \to \infty$ almost surely so all states are visited only finitely many times. That is, all states are transient. Similarly for $p < 1/2$ $X_n \to -\infty$ almost surely and all states are transient.

Now look at $p = 1/2$. The law of large numbers argument no long shows anything. I will show that all states are recurrent.

Proof: We evaluate $\sum_n (\mathbf{P}^n)_{00}$ and show the sum is infinite. If n is odd then $(p_n)_{00} = 0$ so we evaluate

$$\sum_m (\mathbf{P}^{2m})_{00}$$

Now

$$(\mathbf{P}^{2m})_{00} = \binom{2m}{m} 2^{-2m}$$

According to Stirling's approximation

$$\lim_{m\to\infty} \frac{m!}{m^{m+1/2}e^{-m}\sqrt{2\pi}} = 1$$

Hence

$$\lim_{m\to\infty} \sqrt{m}(\mathbf{P}^{2m})_{00} = \frac{1}{\sqrt{\pi}}$$

Since

$$\sum \frac{1}{\sqrt{m}} = \infty$$

we are done.

Mean return times

Compute expected times to return. For $x \in S$ let T_x denote the hitting time for x.

Suppose x recurrent in **irreducible** chain (only one communicating class).

Derive equations for expected values of different T_x.

Each T_x is a certain function f_x applied to $X_1, \ldots$. Setting $\mu_{ij} = \mathsf{E}^i(T_j)$ we find

$$\mu_{ij} = \sum_k \mathsf{E}^i(T_j 1(X_1 = k))$$

Note that if $X_1 = x$ then $T_x = 1$ so

$$\mathsf{E}^i(T_j 1(X_1 = j)) = \mathbf{P}_{ij}$$

For $k \neq j$

$$T_x = 1 + f_x(X_2, X_3, \ldots)$$

and, by conditioning on $X_1 = k$ we find

$$\mathsf{E}^i(T_j 1(X_1 = k)) = \mathbf{P}_{ik}\left\{1 + \mathsf{E}^k(T_j)\right\}$$

This gives

$$\mu_{ij} = 1 + \sum_{k \neq j} \mathbf{P}_{ik}\mu_{kj} \qquad (12)$$

Technically, I should check that the expectations in (10) are finite. All the random variables involved are non-negative, however, and the equation actually makes sense even if some terms are infinite. (To prove this you actually study

$$T_{x,n} = \min(T_x, n)$$

deriving an identity for a fixed n, letting $n \to \infty$ and applying the monotone convergence theorem.)

Here is a simple example:

$$\mathbf{P} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

The identity (10) becomes

$$\begin{aligned}
\mu_{1,1} &= 1 + \frac{1}{2}\mu_{2,1} + \frac{1}{2}\mu_{3,1} \\
\mu_{1,2} &= 1 + \frac{1}{2}\mu_{3,2} \\
\mu_{1,3} &= 1 + \frac{1}{2}\mu_{2,3} \\
\mu_{2,1} &= 1 + \frac{1}{2}\mu_{3,1} \\
\mu_{2,2} &= 1 + \frac{1}{2}\mu_{1,2} + \frac{1}{2}\mu_{3,2} \\
\mu_{2,3} &= 1 + \frac{1}{2}\mu_{1,3} \\
\mu_{3,1} &= 1 + \frac{1}{2}\mu_{2,1} \\
\mu_{3,2} &= 1 + \frac{1}{2}\mu_{1,2} \\
\mu_{3,3} &= 1 + \frac{1}{2}\mu_{1,3} + \frac{1}{2}\mu_{2,3}
\end{aligned}$$

Seventh and fourth show $\mu_{2,1} = \mu_{3,1}$. Similar calculations give $\mu_{ii} = 3$ and for $i \neq j$ $\mu_{i,j} = 2$.

Example: Coin tossing Markov Chain with $p = 1/2$ shows situation can be different when S is infinite. Equations above become:

$$m_{0,0} = 1 + \frac{1}{2} m_{1,0} + \frac{1}{2} m_{-1,0}$$
$$m_{1,0} = 1 + \frac{1}{2} m_{2,0}$$

and many more.

Some observations:

Have to go through 1 to get to 0 from 2 so

$$m_{2,0} = m_{2,1} + m_{1,0}$$

Symmetry (switching H and T):

$$m_{1,0} = m_{-1,0}$$

Transition probabilities are **homogeneous**:

$$m_{2,1} = m_{1,0}$$

Conclusion:

$$\begin{aligned} m_{0,0} &= 1 + m_{1,0} \\ &= 1 + 1 + \frac{1}{2} m_{2,0} \\ &= 2 + m_{1,0} \end{aligned}$$

Notice that there are **no** finite solutions!

Summary of the situation:

Every state is recurrent.

All the expected hitting times m_{ij} are infinite.

All entries $\mathbf{P}^n_{ij}$ converge to 0.

Jargon: The states in this chain are null recurrent.

Model: 2 state MC for weather: ‘Dry’ or ‘Wet’.

```
> p:= matrix(2,2,[[3/5,2/5],[1/5,4/5]]);
                      [3/5    2/5]
                 p := [          ]
                      [1/5    4/5]
> p2:=evalm(p*p):
> p4:=evalm(p2*p2):
> p8:=evalm(p4*p4):
> p16:=evalm(p8*p8):
```

This computes the powers (`evalm` understands matrix algebra).

Fact:

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

```
> evalf(evalm(p));
              [.6000000000    .4000000000]
              [                          ]
              [.2000000000    .8000000000]
> evalf(evalm(p2));
              [.4400000000    .5600000000]
              [                          ]
              [.2800000000    .7200000000]
> evalf(evalm(p4));
              [.3504000000    .6496000000]
              [                          ]
              [.3248000000    .6752000000]
> evalf(evalm(p8));
              [.3337702400    .6662297600]
              [                          ]
              [.3331148800    .6668851200]
> evalf(evalm(p16));
              [.3333336197    .6666663803]
              [                          ]
              [.3333331902    .6666668098]
```

Where did 1/3 and 2/3 come from?

Suppose we toss a coin $P(H) = \alpha_D$ and start the chain with Dry if we get heads and Wet if we get tails.

Then

$$P(X_0 = x) = \begin{cases} \alpha_D & x = \text{Dry} \\ \alpha_W = 1 - \alpha_D & x = \text{Wet} \end{cases}$$

and

$$\begin{aligned} P(X_1 = x) &= \sum_y P(X_1 = x | X_0 = y) P(X_0 = y) \\ &= \sum_y \alpha_y P_{y,x} \end{aligned}$$

Notice last line is a matrix multiplication of row vector α by matrix $\mathbf{P}$. A special α: if we put $\alpha_D = 1/3$ and $\alpha_W = 2/3$ then

$$\begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

So: if $P(X_0 = D) = 1/3$ then $P(X_1 = D) = 1/3$ and analogously for W. This means that X_0 and X_1 have the same distribution.

A probability vector α is called the initial distribution for the chain if

$$P(X_0 = i) = \alpha_i$$

A Markov Chain is **stationary** if

$$P(X_1 = i) = P(X_0 = i)$$

for all i

Finding stationary initial distributions. Consider $\mathbf{P}$ above. The equation

$$\alpha\mathbf{P} = \alpha$$

is really

$$\begin{aligned} \alpha_D &= 3\alpha_D/5 + \alpha_W/5 \\ \alpha_W &= 2\alpha_D/5 + 4\alpha_W/5 \end{aligned}$$

The first can be rearranged to

$$\alpha_W = 2\alpha_D.$$

So can the second. If α is probability vector then

$$\alpha_W + \alpha_D = 1$$

so we get

$$1 - \alpha_D = 2\alpha_D$$

leading to

$$\alpha_D = 1/3$$

Some more examples:

$$\mathbf{P} = \begin{bmatrix} 0 & 1/3 & 0 & 2/3 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 2/3 & 0 & 1/3 & 0 \end{bmatrix}$$

Set $\alpha\mathbf{P} = \alpha$ and get

$$\begin{aligned} \alpha_1 &= \alpha_2/3 + 2\alpha_4/3 \\ \alpha_2 &= \alpha_1/3 + 2\alpha_3/3 \\ \alpha_3 &= 2\alpha_2/3 + \alpha_4/3 \\ \alpha_4 &= 2\alpha_1/3 + \alpha_3/3 \\ 1 &= \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 \end{aligned}$$

First plus third gives

$$\alpha_1 + \alpha_3 = \alpha_2 + \alpha_4$$

so both sums 1/2. Continue algebra to get

$$(1/4, 1/4, 1/4, 1/4) .$$

```
p:=matrix([[0,1/3,0,2/3],[1/3,0,2/3,0],
          [0,2/3,0,1/3],[2/3,0,1/3,0]]);

               [ 0     1/3     0     2/3]
               [                        ]
               [1/3     0     2/3     0 ]
          p := [                        ]
               [ 0     2/3     0     1/3]
               [                        ]
               [2/3     0     1/3     0 ]
> p2:=evalm(p*p);
             [5/9     0     4/9     0 ]
             [                        ]
             [ 0     5/9     0     4/9]
        p2:= [                        ]
             [4/9     0     5/9     0 ]
             [                        ]
             [ 0     4/9     0     5/9]
> p4:=evalm(p2*p2):
> p8:=evalm(p4*p4):
> p16:=evalm(p8*p8):
> p17:=evalm(p8*p8*p):
```

```
> evalf(evalm(p16));
    [.5000000116 , 0 , .4999999884 , 0]
    [                                  ]
    [0 , .5000000116 , 0 , .4999999884]
    [                                  ]
    [.4999999884 , 0 , .5000000116 , 0]
    [                                  ]
    [0 , .4999999884 , 0 , .5000000116]
> evalf(evalm(p17));
    [0 , .4999999961 , 0 , .5000000039]
    [                                  ]
    [.4999999961 , 0 , .5000000039 , 0]
    [                                  ]
    [0 , .5000000039 , 0 , .4999999961]
    [                                  ]
    [.5000000039 , 0 , .4999999961 , 0]
```

```
> evalf(evalm((p16+p17)/2));
  [.2500, .2500, .2500, .2500]
  [                          ]
  [.2500, .2500, .2500, .2500]
  [                          ]
  [.2500, .2500, .2500, .2500]
  [                          ]
  [.2500, .2500, .2500, .2500]
```

$\mathbf{P}^n$ doesn't converges but$(\mathbf{P}^n + \mathbf{P}^{n+1})/2$ does.
Next example:

$$\P = \begin{bmatrix} \frac{2}{5} & \frac{3}{5} & 0 & 0 \\ \frac{1}{5} & \frac{4}{5} & 0 & 0 \\ 0 & 0 & \frac{2}{5} & \frac{3}{5} \\ 0 & 0 & \frac{1}{5} & \frac{4}{5} \end{bmatrix}$$

Solve $\alpha\mathbf{P} = \alpha$:

$$\begin{aligned} \alpha_1 &= \frac{2}{5}\alpha_1 + \frac{1}{5}\alpha_2 \\ \alpha_2 &= \frac{3}{5}\alpha_1 + \frac{4}{5}\alpha_2 \\ \alpha_3 &= \frac{2}{5}\alpha_3 + \frac{1}{5}\alpha_4 \\ \alpha_4 &= \frac{3}{5}\alpha_3 + \frac{4}{5}\alpha_4 \\ 1 &= \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 \end{aligned}$$

Second and fourth equations redundant. Get

$$\begin{aligned} \alpha_2 &= 3\alpha_1 \\ 3\alpha_3 &= \alpha_4 \\ 1 &= 4\alpha_1 + 4\alpha_3 \end{aligned}$$

Pick α_1 in $[0, 1/4]$; put $\alpha_3 = 1/4 - \alpha_1$.

$$\alpha = (\alpha_1, 3\alpha_1, 1/4 - \alpha_1, 3(1/4 - \alpha_1))$$

solves $\alpha\mathbf{P} = \alpha$. So solution is not unique.

```
> p:=matrix([[2/5,3/5,0,0],[1/5,4/5,0,0],
            [0,0,2/5,3/5],[0,0,1/5,4/5]]);
                  [2/5    3/5     0       0 ]
                  [                         ]
                  [1/5    4/5     0       0 ]
            p :=  [                         ]
                  [ 0      0     2/5    3/5]
                  [                         ]
                  [ 0      0     1/5    4/5]
> p2:=evalm(p*p):
> p4:=evalm(p2*p2):
> p8:=evalm(p4*p4):
> evalf(evalm(p8*p8));
           [.2500000000 , .7500000000 , 0 , 0]
           [                                 ]
           [.2500000000 , .7500000000 , 0 , 0]
           [                                 ]
           [0 , 0 , .2500000000 , .7500000000]
           [                                 ]
           [0 , 0 , .2500000000 , .7500000000]
```

Notice that rows converge but to two different vectors:

$$\alpha^{(1)} = (1/4, 3/4, 0, 0)$$

and

$$\alpha^{(2)} = (0, 0, 1/4, 3/4)$$

Solutions of $\alpha\mathbf{P} = \alpha$ revisited? Check that

$$\alpha^{(1)}\mathbf{P} = \alpha^{(1)}$$

and

$$\alpha^{(2)}\mathbf{P} = \alpha^{(2)}$$

If $\alpha = \lambda\alpha^{(1)} + (1-\lambda)\alpha^{(2)}$ $(0 \le \lambda \le 1)$ then

$$\alpha\mathbf{P} = \alpha$$

so again solution is not unique.

Last example:

```
> p:=matrix([[2/5,3/5,0],[1/5,4/5,0],
            [1/2,0,1/2]]);

                  [2/5    3/5     0 ]
                  [                 ]
             p := [1/5    4/5     0 ]
                  [                 ]
                  [1/2     0     1/2]
> p2:=evalm(p*p):
> p4:=evalm(p2*p2):
> p8:=evalm(p4*p4):
> evalf(evalm(p8*p8));
  [.2500000000 .7500000000        0       ]
  [                                       ]
  [.2500000000 .7500000000        0       ]
  [                                       ]
  [.2500152588 .7499694824 .00001525878906]
```

Interpretation of examples

- For some $\mathbf{P}$ all rows converge to some α. In this case this α is a stationary initial distribution.

- For some $\mathbf{P}$ the locations of zeros flip flop. $\mathbf{P}^n$ does not converge. Observation: average

$$\frac{\mathbf{P}+\mathbf{P}^2+\cdots+\mathbf{P}^n}{n}$$

  does converge.

- For some $\mathbf{P}$ some rows converge to one α and some to another. In this case the solution of $\alpha\mathbf{P}=\alpha$ is not unique.

Basic distinguishing features: pattern of 0s in matrix $\mathbf{P}$.

The ergodic theorem

Consider a finite state space chain. If x is a vector then the ith entry in $\mathbf{P}x$ is

$$\sum_j \mathbf{P}_{ij} x_j$$

Rows of $\mathbf{P}$ probability vectors, so a weighted average of the entries in x.

If weights strictly between 0, 1 and largest and smallest entries in x not same then $\sum_j \mathbf{P}_{ij} x_j$ strictly between largest and smallest entries in x. In fact

$$\begin{aligned}\sum_j \mathbf{P}_{ij} x_j - \min(x_k) &= \sum_j \mathbf{P}_{ij}\{x_j - \min(x_k)\} \\ &\ge \min_j\{p_{ij}\}(\max\{x_k\} - \min\{x_k\})\end{aligned}$$

and

$$\begin{aligned}&\max\{x_j\} - \sum_j \mathbf{P}_{ij} x_j \\ &\qquad \ge \min_j\{p_{ij}\}(\max\{x_k\} - \min\{x_k\})\end{aligned}$$

Now multiply $\mathbf{P}^r$ by $\mathbf{P}^m$.

ijth entry in $\mathbf{P}^{r+m}$ is a weighted average of the jth column of $\mathbf{P}^m$.

So, if all the entries in row i of $\mathbf{P}^r$ are positive and the jth column of $\mathbf{P}^m$ is not constant, the ith entry in the jth column of $\mathbf{P}^{r+m}$ must be strictly between the minimum and maximum entries of the jth column of $\mathbf{P}^m$.

In fact, fix a j.

$\overline{x}_m =$ maximum entry in column j of $\mathbf{P}^m$

$\underline{x}_m$ the minimum entry.

Suppose all entries of $\mathbf{P}^r$ are positive.

Let $\delta > 0$ be the smallest entry in $\mathbf{P}^r$. Our argument above shows that

$$\overline{x}_{m+r} \leq \overline{x}_m - \delta(\overline{x}_m - \underline{x}_m)$$

and

$$\underline{x}_{m+r} \geq \underline{x}_m + \delta(\overline{x}_m - \underline{x}_m)$$

Putting these together gives

$$(\overline{x}_{m+r} - \underline{x}_{m+r}) \leq (1 - 2\delta)(\overline{x}_m - \underline{x}_m)$$

In summary the column maximum decreases, the column minimum increases and the gap between the two decreases exponentially along the sequence $m, m + r, m + 2r, \ldots$.

This idea can be used to prove

Proposition 2 *Suppose X_n finite state space Markov Chain with stationary transition matrix $\mathbf{P}$. Assume that there is a power r such that all entries in $\mathbf{P}^r$ are positive. Then for $\mathbf{P}^k$ has all entries positive for all $k \geq r$ and $\mathbf{P}^n$ converges, as $n \to \infty$ to a matrix $\mathbf{P}^\infty$. Moreover,*

$$(\mathbf{P}^\infty)_{ij} = \pi_j$$

where π is the unique row vector satisfying

$$\pi = \pi \mathbf{P}$$

whose entries sum to 1.

Proof: First for $k > r$

$$(\mathbf{P}^k)_{ij} = \sum_k (\mathbf{P}^{k-r})_{ik} (\mathbf{P}^r)_{kj}$$

For each i there is a k for which $(\mathbf{P}^{k-r})_{ik} > 0$ and since $(\mathbf{P}^r)_{kj} > 0$ we see $(\mathbf{P}^k)_{ij} > 0$.

The argument before the proposition shows that

$$\lim_{j \to \infty} \mathbf{P}^{m+jk}$$

exists for each m and $k \geq r$. This proves $\mathbf{P}^n$ has a limit which we call $\mathbf{P}^\infty$. Since $\mathbf{P}^{n-1}$ also converges to $\mathbf{P}^\infty$ we find

$$\mathbf{P}^\infty = \mathbf{P}^\infty \mathbf{P}$$

Hence each row of $\mathbf{P}^\infty$ is a solution of $x\mathbf{P} = x$. The argument before the statement of the proposition shows all rows of $\mathbf{P}^\infty$ are equal. Let π be this common row.

Now if α is any vector whose entries sum to 1 then $\alpha \mathbf{P}^n$ converges to

$$\alpha \mathbf{P}^\infty = \pi$$

If α is any solution of $x = x\mathbf{P}$ we have by induction $\alpha \mathbf{P}^n = \alpha$ so $\alpha \mathbf{P}^\infty = \alpha$ so $\alpha = \pi$. That is exactly one vector whose entries sum to 1 satisfies $x = x\mathbf{P}$. •

Note conditions:

There is an r for which all entries in $\mathbf{P}^r$ are positive.

The chain has a finite state space.

Consider finite state space case: $\mathbf{P}^n$ need not have limit. Example:

$$\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Note $\mathbf{P}^{2n}$ is the identity while $\mathbf{P}^{2n+1} = \mathbf{P}$. Note, too, that

$$\frac{\mathbf{P}^0 + \cdots + \mathbf{P}^n}{n+1} \to \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Consider the equations $\pi = \pi\mathbf{P}$ with $\pi_1 + \pi_2 = 1$. We get

$$\pi_1 = \frac{1}{2}\pi_1 + \frac{1}{2}(1 - \pi_1) = \frac{1}{2}$$

so that the solution to $\pi = \pi\mathbf{P}$ is again unique.

Def'n: The period d of a state i is the greatest common divisor of

$$\{n : (\mathbf{P}^n)_{ii} > 0\}$$

Lemma 1 *If $i \leftrightarrow j$ then i and j have the same period.*

Def'n: A state is **aperiodic** if its period is 1.

Proof: I do the case $d = 1$. Fix i. Let

$$G = \{k : (\mathbf{P}^k)_{ii} > 0\}$$

If $k_1, k_2 \in G$ then $k_1 + k_2 \in G$.

This (and aperiodic) implies (number theory argument) that there is an r such that $k \geq r$ implies $k \in G$.

Now find m and n so that

$$(\mathbf{P}^m)_{ij} > 0 \text{ and } (\mathbf{P}^n)_{ji} > 0$$

For $k > r+m+n$ we see $(\mathbf{P}^k)_{jj} > 0$ so the gcd of the set of k such that $(\mathbf{P}^k)_{jj} > 0$ is 1. •

The case of period $d > 1$ can be dealt with by considering $\mathbf{P}^d$.

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

For this example $\{1,2,3\}$ is a class of period 3 states and $\{4,5\}$ a class of period 2 states.

$$\mathbf{P} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

has a single communicating class of period 2.

A chain is **aperiodic** if all its states are aperiodic.

Hitting Times

Start irreducible recurrent chain X_n in state i. Let T_j be first $n > 0$ such that $X_n = j$. Define

$$m_{ij} = \mathsf{E}(T_j | X_0 = i)$$

First step analysis:

$$\begin{aligned} m_{ij} &= 1 \cdot P(X_1 = j | X_0 = i) \\ &\qquad + \sum_{k \neq j} (1 + \mathsf{E}(T_j | X_0 = k)) P_{ik} \\ &= \sum_j P_{ij} + \sum_{k \neq j} P_{ik} m_{kj} \\ &= 1 + \sum_{k \neq j} P_{ik} m_{kj} \end{aligned}$$

Example

$$\mathbf{P} = \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix}$$

The equations are

$$
\begin{aligned}
m_{11} &= 1 + \frac{2}{5} m_{21} \\
m_{12} &= 1 + \frac{3}{5} m_{12} \\
m_{21} &= 1 + \frac{4}{5} m_{21} \\
m_{22} &= 1 + \frac{1}{5} m_{12}
\end{aligned}
$$

The second and third equations give immediately

$$
\begin{aligned}
m_{12} &= \frac{5}{2} \\
m_{21} &= 5
\end{aligned}
$$

Then plug in to the others to get

$$
\begin{aligned}
m_{11} &= 3 \\
m_{22} &= \frac{3}{2}
\end{aligned}
$$

Notice stationary initial distribution is

$$
\left(\frac{1}{m_{11}}, \frac{1}{m_{22}} \right)
$$

Consider fraction of time spent in state j:

$$\frac{1(X_0 = j) + \cdots + 1(X_n = j)}{n+1}$$

Imagine chain starts in chain i; take expected value.

$$\frac{\sum_{r=1}^{n} \mathbf{P}_{ij}^{r} + 1(i = j)}{n+1}$$

If rows of $\mathbf{P}^r$ converge to π then fraction converges to π_j; i.e. limiting fraction of time in state j is π_j.

Heuristic: start chain in i. Expect to return to i every m_{ii} time units. So are in state i about once every m_{ii} time units; i.e. limiting fraction of time in state i is $1/m_{ii}$.

Conclusion: for an irreducible recurrent finite state space Markov chain

$$\pi_i = \frac{1}{m_{ii}}.$$

Real proof: Renewal theorem or variant.

Idea: $S_1 < S_2 < \ldots$ are times of visits to i. Segment i:

$$X_{S_{i-1}+1}, \ldots, X_{S_i}.$$

Segments are iid by Strong Markov.

Number of visits to i by time S_k is exactly k.

Total elapsed time is $S_k = T_1 + \cdots + T_k$ where T_i are iid.

Fraction of time in state i by time S_k is

$$\frac{k}{S_k} \to \frac{1}{m_{ii}}$$

by SLLN. So if fraction converges to π_i must have

$$\pi_i = \frac{1}{m_{ii}}.$$

Summary of Theoretical Results:

For an irreducible aperiodic positive recurrent Markov Chain:

1. $\mathbf{P}^n$ converges to a stochastic matrix $\mathbf{P}^\infty$.

2. Each row of $\mathbf{P}^\infty$ is π the unique stationary initial distribution.

3. The stationary initial distribution is given by

$$\pi_i = 1/m_i$$

   where m_i is the mean return time to state i from state i.

If the state space is finite an irreducible chain is positive recurrent.

Ergodic Theorem

Notice slight of hand: I showed

$$\frac{\mathrm{E}\left\{\sum_{i=0}^{n} 1(X_i = k)\right\}}{n} \to \pi_k$$

but claimed

$$\frac{\sum_{i=0}^{n} 1(X_i = k)}{n} \to \pi_k$$

almost surely which is also true. This is a step in the proof of the ergodic theorem. For an irreducible positive recurrent Markov chain and any f on S such that $\mathrm{E}^\pi(f(X_0)) < \infty$:

$$\frac{\sum_0^n f(X_i)}{n} \to \sum \pi_j f(j)$$

almost surely. The limit works in other senses, too. You also get

$$\frac{\sum_0^n f(X_i, \ldots, X_{i+k})}{n} \to \mathrm{E}^\pi \left\{ f(X_0, \ldots, X_k) \right\}$$

E.g. fraction of transitions from i to j goes to

$$\pi_i \mathbf{P}_{ij}$$

For an irreducible positive recurrent chain of period d:

1. $\mathbf{P}^d$ has d communicating classes each of which forms an irreducible aperiodic positive recurrent chain.

2. $(\mathbf{P}^{n+1} + \cdots + \mathbf{P}^{n+d})/d$ has a limit $\mathbf{P}^\infty$.

3. Each row of $\mathbf{P}^\infty$ is π the unique stationary initial distribution.

4. Stationary initial distribution places probability $1/d$ on each of the communicating classes in 1.

For an irreducible null recurrent chain:

1. $\mathbf{P}^n$ converges to 0 (pointwise).

2. there is no stationary initial distribution.

For an irreducible transient chain:

1. $\mathbf{P}^n$ converges to 0 (pointwise).

2. there is no stationary initial distribution.

For a chain with more than 1 communicating class:

1. If $\mathcal{C}$ is a recurrent class the submatrix $\mathbf{P}_{\mathcal{C}}$ of $\mathbf{P}$ made by picking out rows i and columns j for which $i, j \in \mathcal{C}$ is a stochastic matrix. The corresponding entries in $\mathbf{P}^n$ are just $(\mathbf{P}_{\mathcal{C}})^n$ so you can apply the conclusions above.

2. For any transient or null recurrent class the corresponding columns in $\mathbf{P}^n$ converge to 0.

3. If there are multiple positive recurrent communicating classes then the stationary initial distribution is not unique.

Poisson Processes

Particles arriving over time at a particle detector.

Several ways to describe most common model.

Approach 1:

a) numbers of particles arriving in an interval has Poisson distribution,

b) mean proportional to length of interval,

c) numbers in several non-overlapping intervals independent.

For $s < t$, denote number of arrivals in $(s, t]$ by $N(s, t)$.

Jargon: $N(A)$ = number of points in A is a **counting process**.

Model is

1. $N(s,t)$ has a Poisson$(\lambda(t-s))$ distribution.

2. For $0 \le s_1 < t_1 \le s_2 < t_2 \cdots \le s_k < t_k$ the variables $N(s_i,t_i); i = 1,\ldots,k$ are independent.

Approach 2:

Let $0 < S_1 < S_2 < \cdots$ be the times at which the particles arrive.

Let $T_i = S_i - S_{i-1}$ with $S_0 = 0$ by convention. T_i are called **interarrival** times.

Then $T_1, T_2, \ldots$ are independent Exponential random variables with mean $1/\lambda$.

Note $P(T_i > x) = e^{-\lambda x}$ is called **survival function** of T_i.

Approaches are equivalent. Both are deductions of a model based on **local** behaviour of process.

Approach 3: Assume:

1. given all the points in $[0, t]$ the probability of 1 point in the interval $(t, t+h]$ is of the form

$$\lambda h + o(h)$$

2. given all the points in $[0, t]$ the probability of 2 or more points in interval $(t, t+h]$ is of the form

$$o(h)$$

Notation: given functions f and g we write

$$f(h) = g(h) + o(h)$$

provided

$$\lim_{h \to 0} \frac{f(h) - g(h)}{h} = 0$$

[Aside: if there is a constant M such that

$$\limsup_{h\to 0}\left|\frac{f(h)-g(h)}{h}\right| \le M$$

we say

$$f(h) = g(h) + O(h)$$

Notation due to Landau. Another form is

$$f(h) = g(h) + O(h)$$

means there is $\delta > 0$ and M s.t. for all $|h| < \delta$

$$|f(h) - g(h)| \le M|h|$$

Idea: $o(h)$ is tiny compared to h while $O(h)$ is (very) roughly the same size as h.]

Generalizations:

1. First (Poisson) model generalizes to $N(s,t]$ having a Poisson distribution with parameter $\Lambda(t) - \Lambda(s)$ for some non-decreasing non-negative function Λ (called **cumulative intensity**). Result called **inhomogeneous** Poisson process.

2. Exponential interarrival model generalizes to independent non-exponential interarrival times. Result is **renewal process** or **semi-Markov** process.

3. Infinitesimal probability model generalizes to other infinitesimal jump rates. Model specifies **infinitesimal generator**. Yields other **continuous time Markov Chains**.

Equivalence of Modelling Approaches

All 3 approaches to Poisson process are equivalent. I show: 3 implies 1, 1 implies 2 and 2 implies 3. First explain o, O.

Model 3 implies 1: Fix t, define $f_t(s)$ to be conditional probability of 0 points in $(t, t+s]$ given value of process on $[0, t]$.

Derive differential equation for f. Given process on $[0, t]$ and 0 points in $(t, t+s]$ probability of no points in $(t, t+s+h]$ is

$$f_{t+s}(h) = 1 - \lambda h + o(h)$$

Given the process on $[0, t]$ the probability of no points in $(t, t+s]$ is $f_t(s)$. Using $P(AB|C) = P(A|BC)P(B|C)$ gives

$$\begin{aligned} f_t(s+h) &= f_t(s) f_{t+s}(h) \\ &= f_t(s)(1 - \lambda h + o(h)) \end{aligned}$$

Now rearrange, divide by h to get

$$\frac{f_t(s+h) - f_t(s)}{h} = -\lambda f_t(s) + \frac{o(h)}{h}$$

Let $h \to 0$ and find

$$\frac{\partial f_t(s)}{\partial s} = -\lambda f_t(s)$$

Differential equation has solution

$$f_t(s) = f_t(0) \exp(-\lambda s) = \exp(-\lambda s) .$$

Things to notice:

- $f_t(s) = e^{-\lambda s}$ is survival function of exponential rv..

- We had suppressed dependence of $f_t(s)$ on $N(u); 0 \le u \le t$ but solution does not depend on condition.

- So: the event of getting 0 points in $(t, t+s]$ is independent of $N(u); 0 \le u \le t$.

- We used: $f_t(s)o(h) = o(h)$. Other rules:

$$
\begin{aligned}
o(h) + o(h) &= o(h) \\
O(h) + O(h) &= O(h) \\
O(h) + o(h) &= O(h) \\
o(h^r)O(h^s) &= o(h^{r+s}) \\
O(o(h)) &= o(h)
\end{aligned}
$$

General case: notation: $N(t) = N(0,t)$.

$N(t)$ is a non-decreasing function of t. Let

$$P_k(t) = P(N(t) = k)$$

Evaluate $P_k(t+h)$: condition on $N(s); 0 \le s < t$ and on $N(t) = j$.

Given $N(t) = j$ probability that $N(t+h) = k$ is conditional probability of $k-j$ points in $(t, t+h]$.

So, for $j \le k-2$:

$$P(N(t+h) = k | N(t) = j, N(s), 0 \le s < t) = o(h)$$

For $j = k-1$ we have

$$P(N(t+h) = k | N(t) = k-1, N(s), 0 \le s < t) = \lambda h + o(h)$$

For $j = k$ we have

$$P(N(t+h) = k | N(t) = k, N(s), 0 \le s < t) = 1 - \lambda h + o(h)$$

N is increasing so only consider $j \leq k$.

$$\begin{aligned} P_k(t+h) &= \sum_{j=0}^{k} P(N(t+h)=k|N(t)=j)P_j(t) \\ &= P_k(t)(1-\lambda h) + \lambda h P_{k-1}(t) + o(h) \end{aligned}$$

Rearrange, divide by h and let $h \to 0$ t get

$$P_k'(t) = -\lambda P_k(t) + \lambda P_{k-1}(t)$$

For $k = 0$ the term P_{k-1} is dropped and

$$P_0'(t) = \lambda P_0(t)$$

Using $P_0(0) = 1$ we get

$$P_0(t) = e^{-\lambda t}$$

Put this into the equation for $k = 1$ to get

$$P_1'(t) = -\lambda P_1(t) + \lambda e^{-\lambda t}$$

Multiply by $e^{\lambda t}$ to see

$$\left(e^{\lambda t} P_1(t)\right)' = \lambda$$

With $P_1(0) = 0$ we get

$$P_1(t) = \lambda t e^{-\lambda t}$$

For general k we have $P_k(0) = 0$ and

$$\left(e^{\lambda t} P_k(t)\right)' = \lambda e^{\lambda t} P_{k-1}(t)$$

Check by induction that

$$e^{\lambda t} P_k(t) = (\lambda t)^k / k!$$

Hence: $N(t)$ has Poisson(λt) distribution.

Similar ideas permit proof of

$$P(N(s,t) = k | N(u); 0 \le u \le s) = \frac{\{\lambda(t-s)\}^k e^{-\lambda}}{k!}$$

Now prove (by induction) N has independent Poisson increments.

Exponential Interarrival Times

If N is a Poisson Process we define $T_1, T_2, \ldots$ to be the times between 0 and the first point, the first point and the second and so on.

Fact: $T_1, T_2, \ldots$ are iid exponential rvs with mean $1/\lambda$.

We already did T_1 rigorously. The event $T_1 > t$ is exactly the event $N(t) = 0$. So

$$P(T_1 > t) = \exp(-\lambda t)$$

which is the survival function of an exponential rv.

Do case of T_1, T_2. Let t_1, t_2 be two positive numbers and $s_1 = t_1$, $s_2 = t_1 + t_2$. Event

$$\{t_1 < T_1 \le t_1 + \delta_1\} \cap \{t_2 < T_2 \le t_2 + \delta_2\} .$$

is almost the same as the intersection of four events:

$$\begin{aligned} N(0, t_1] &= 0 \\ N(t_1, t_1 + \delta_1] &= 1 \\ N(t_1 + \delta_1, t_1 + \delta_1 + t_2] &= 0 \\ N(s_2 + \delta_1, s_2 + \delta_1 + \delta_2] &= 1 \end{aligned}$$

which has probability

$$e^{-\lambda t_1} \times \lambda \delta_1 e^{-\lambda \delta_1} \times e^{-\lambda t_2} \times \lambda \delta_2 e^{-\lambda \delta_2}$$

Divide by $\delta_1 \delta_2$ and let δ_1 and δ_2 go to 0 to get joint density of T_1, T_2 is

$$\lambda^2 e^{-\lambda t_1} e^{-\lambda t_2}$$

which is the joint density of two independent exponential variates.

More rigor:

- Find joint density of $S_1, \ldots, S_k$.

- Use **change of variables** to find joint density of $T_1, \ldots, T_k$.

First step: Compute

$$P(0 < S_1 \le s_1 < S_2 \le s_2 \cdots < S_k \le s_k)$$

This is just the event of exactly 1 point in each interval $(s_{i-1}, s_i]$ for $i = 1, \ldots, k-1$ ($s_0 = 0$) and at least one point in $(s_{k-1}, s_k]$ which has probability

$$\prod_1^{k-1} \left\{ \lambda(s_i - s_{i-1}) e^{-\lambda(s_i - s_{i-1})} \right\} \left(1 - e^{-\lambda(s_k - s_{k-1})}\right)$$

Second step: write this in terms of joint cdf of $S_1, \ldots, S_k$. I do $k = 2$:

$$P(0 < S_1 \le s_1 < S_2 \le s_2) \\ = F_{S_1,S_2}(s_1, s_2) - F_{S_1,S_2}(s_1, s_1)$$

Notice tacit assumption $s_1 < s_2$.

Differentiate twice, that is, take

$$\frac{\partial^2}{\partial s_1 \partial s_2}$$

to get

$$f_{S_1,S_2}(s_1, s_2) = \\ \frac{\partial^2}{\partial s_1 \partial s_2} \lambda s_1 e^{-\lambda s_1} \left(1 - e^{-\lambda(s_2 - s_1)}\right)$$

Simplify to

$$\lambda^2 e^{-\lambda s_2}$$

Recall tacit assumption to get

$$f_{S_1,S_2}(s_1, s_2) = \lambda^2 e^{-\lambda s_2} 1(0 < s_1 < s_2)$$

That completes the first part.

Now compute the joint cdf of T_1, T_2 by

$$F_{T_1,T_2}(t_1,t_2) = P(S_1 < t_1, S_2 - S_1 < t_2)$$

This is

$$\begin{aligned} P(S_1 < t_1, & S_2 - S_1 < t_2) \\ &= \int_0^{t_1} \int_{s_1}^{s_1+t_2} \lambda^2 e^{-\lambda s_2} ds_2 ds_1 \\ &= \lambda \int_0^{t_1} \left(e^{-\lambda s_1} - e^{-\lambda(s_1+t_2)} \right) ds_1 \\ &= 1 - e^{-\lambda t_1} - e^{-\lambda t_2} + e^{-\lambda(t_1+t_2)} \end{aligned}$$

Differentiate twice to get

$$f_{T_1,T_2}(t_1,t_2) = \lambda e^{-\lambda t_1} \lambda e^{-\lambda t_2}$$

which is the joint density of two independent exponential random variables.

Summary so far:

Have shown:

Instantaneous rates model implies independent Poisson increments model implies independent exponential interarrivals.

Next: show independent exponential interarrivals implies the instantaneous rates model.

Suppose $T_1, \ldots$ iid exponential rvs with means $1/\lambda$. Define N_t by $N_t = k$ if and only if

$$T_1 + \cdots + T_k \le t \le T_1 + \cdots + T_{k+1}$$

Let A be event $N(s) = n(s); 0 < s \le t$. We are to show

$$P(N(t, t+h] = 1 | N(t) = k, A) = \lambda h + o(h)$$

and

$$P(N(t, t+h] \ge 2 | N(t) = k, A) = o(h)$$

If $n(s)$ is a possible trajectory consistent with $N(t) = k$ then n has jumps at points

$$t_1, t_1 + t_2, \ldots, s_k \equiv t_1 + \cdots + t_k < t$$

and at no other points in $(0, t]$.

So given $N(s) = n(s); 0 < s \le t$ with $n(t) = k$ we are essentially being given

$$T_1 = t_1, \ldots, T_k = t_k, T_{k+1} > t - s_k$$

and asked the conditional probabilty in the first case of the event B given by

$$t - s_k < T_{k+1} \le t - s_k + h < T_{k+2} + T_{k+1} .$$

Conditioning on $T_1, \ldots, T_k$ irrelevant (independence).

$$\begin{aligned}
P(N(t,t+h] = 1|N(t) = k, A)/h \\
&= P(B|T_{k+1} > t - s_k)/h \\
&= \frac{P(B)}{he^{-\lambda(t-s_k)}}
\end{aligned}$$

Numerator evaluated by integration:

$$P(B) = \int_{t-s_k}^{t-s_k+h} \int_{t-s_k+h-u_1}^{\infty} \lambda^2 e^{-\lambda(u_1+u_2)} du_2 du_1$$

Let $h \to 0$ to get the limit

$$P(N(t,t+h] = 1|N(t) = k, A)/h \to \lambda$$

as required.

The computation of

$$\lim_{h\to 0} P(N(t,t+h] \geq 2|N(t) = k, A)/h$$

is similar.

Properties of exponential rvs

Convolution: If X and Y independent rvs with densities f and g respectively and $Z = X + Y$ then

$$P(Z \le z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x)g(y)dydx$$

Differentiating wrt z we get

$$f_Z(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx$$

This integral is called the **convolution** of densities f and g.

If $T_1, \ldots, T_n$ iid Exponential(λ) then $S_n = T_1 + \cdots + T_n$ has a Gamma(n, λ) distribution. Density of S_n is

$$f_{S_n}(s) = \lambda(\lambda s)^{n-1} e^{-\lambda s}/(n-1)!$$

for $s > 0$.

Proof:

$$\begin{aligned} P(S_n > s) &= P(N(0,s] < n) \\ &= \sum_{j=0}^{n-1} (\lambda s)^j e^{-\lambda s}/j! \end{aligned}$$

Then

$$\begin{aligned} f_{S_n}(s) &= \frac{d}{ds} P(S_n \le s) \\ &= \frac{d}{ds}\{1 - P(S_n > s)\} \\ &= -\lambda \sum_{j=1}^{n-1} \left\{ j(\lambda s)^{j-1} - (\lambda s)^j \right\} \frac{e^{-\lambda s}}{j!} \\ &\quad + \lambda e^{-\lambda s} \\ &= \lambda e^{-\lambda s} \sum_{j=1}^{n-1} \left\{ \frac{(\lambda s)^j}{j!} - \frac{(\lambda s)^{j-1}}{(j-1)!} \right\} \\ &\quad + \lambda e^{-\lambda s} \end{aligned}$$

This telescopes to

$$f_{S_n}(s) = \lambda(\lambda s)^{n-1} e^{-\lambda s}/(n-1)!$$

Extreme Values: If $X_1, \ldots, X_n$ are independent exponential rvs with means $1/\lambda_1, \ldots, 1/\lambda_n$ then $Y = \min\{X_1, \ldots, X_n\}$ has an exponential distribution with mean

$$\frac{1}{\lambda_1 + \cdots + \lambda_n}$$

Proof:

$$\begin{aligned} P(Y > y) &= P(\forall k X_k > y) \\ &= \prod e^{-\lambda_k y} \\ &= e^{-\sum \lambda_k y} \end{aligned}$$

Memoryless Property: conditional distribution of $X - x$ given $X \geq x$ is exponential if X has an exponential distribution.

Proof:

$$P(X - x > y | X \geq x)$$
$$= \frac{P(X > x + y, X \geq x)}{P(X > x)}$$
$$= \frac{PX > x + y)}{P(X \geq x)}$$
$$= \frac{e^{-\lambda(x+y)}}{e^{-\lambda x}}$$
$$= e^{-\lambda y}$$

Hazard Rates

The hazard rate, or instantaneous failure rate for a positive random variable T with density f and cdf F is

$$r(t) = \lim_{\delta \to 0} \frac{P(t < T \leq t + \delta | T \geq t)}{\delta}$$

This is just

$$r(t) = \frac{f(t)}{1 - F(t)}$$

For an exponential random variable with mean $1/\lambda$ this is

$$h(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

The exponential distribution has constant failure rate.

Weibull random variables have density

$$f(t|\lambda,\alpha) = \lambda(\lambda t)^{\alpha-1} e^{-(\lambda t)^\alpha}$$

for $t > 0$. The corresponding survival function is

$$1 - F(t) = e^{-(\lambda t)^\alpha}$$

and the hazard rate is

$$r(t) = \lambda(\lambda t)^{\alpha-1};$$

increasing for $\alpha > 1$, decreasing for $\alpha < 1$. For $\alpha = 1$: exponential distribution.

Since

$$r(t) = \frac{dF(t)/dt}{1 - F(t)} = -\frac{d \log(1 - F(t))}{dt}$$

we can integrate to find

$$1 - F(t) = \exp\{-\int_0^t r(s)ds\}$$

so that r determines F and f.

Properties of Poisson Processes

1) If N_1 and N_2 are independent Poisson processes with rates λ_1 and λ_2, respectively, then $N = N_1 + N_2$ is a Poisson processes with rate $\lambda_1 + \lambda_2$.

2) Let$N$ be a Poisson process with rate λ.

• Suppose each point is marked with a label, say one of $L_1, \ldots, L_r$, independently of all other occurrences.

• Suppose p_i is the probability that a given point receives label L_i.

• Let N_i count the points with label i (so that $N = N_1 + \cdots + N_r$).

• Then $N_1, \ldots, N_r$ are independent Poisson processes with rates $p_i\lambda$.

3) Suppose $U_1, U_2, \ldots$ independent rvs, each uniformly distributed on $[0, T]$.

• Suppose M is a Poisson(λT) random variable independent of the U's.

• Let

$$N(t) = \sum_1^M 1(U_i \le t)$$

• Then N is a Poisson process on $[0, T]$ with rate λ.

4) Let N be Poisson process with rate λ.

• $S_1 < S_2 < \cdots$ times at which points arrive

• Given $N(T) = n$, $S_1, \ldots, S_n$ have same distribution as order statistics of sample of size n from uniform distribution on $[0, T]$.

5) Given $S_{n+1} = T$, $S_1, \ldots, S_n$ have same distribution as order statistics of sample of size n from uniform distribution on $[0, T]$.

Indications of some proofs:

1) $N_1, \ldots, N_r$ independent Poisson processes rates λ_i, $N = \sum N_i$. Let A_h be the event of 2 or more points in N in the time interval $(t, t+h]$, B_h, the event of exactly one point in N in the time interval $(t, t+h]$.

Let A_{ih} and B_{ih} be the corresponding events for N_i.

Let H_t denote the history of the processes up to time t; we condition on H_t. Technically, H_t is the σ-field generated by

$$\{N_i(s); 0 \le s \le t, i = 1, \ldots, r\}$$

We are given:

$$P(A_{ih}|H_t) = o(h)$$

and

$$P(B_{ih}|H_t) = \lambda_i h + o(h)\,.$$

Note that

$$A_h \subset \bigcup_{i=1}^{r} A_{ih} \cup \bigcup_{i \neq j} \Big(B_{ih} \cap B_{jh} \Big)$$

Since

$$\begin{aligned} P(B_{ih} \cap B_{jh}|H_t) &= P(B_{ih}|H_t)P(B_{jh}|H_t) \\ &= (\lambda_i h + o(h))(\lambda_j h + o(h)) \\ &= O(h^2) \\ &= o(h) \end{aligned}$$

and

$$P(A_{ih}|H_t) = o(h)$$

we have checked one of the two infinitesimal conditions for a Poisson process.

Next let C_h be the event of no points in N in the time interval $(t, t+h]$ and C_{ih} the same for N_i. Then

$$\begin{aligned} P(C_h|H_t) &= P(\cap C_{ih}|H_t) \\ &= \prod P(C_{ih}|H_t) \\ &= \prod(1 - \lambda_i h + o(h)) \\ &= 1 - (\sum \lambda_i)h + o(h) \end{aligned}$$

shows

$$\begin{aligned} P(B_h|H_t) &= 1 - P(C_h|H_t) - P(A_h|H_t) \\ &= (\sum \lambda_i)h + o(h) \end{aligned}$$

Hence N is a Poisson process with rate $\sum \lambda_i$.

2) The infinitesimal approach used for 1 can do part of this. See text for rest. Events defined as in **1)**: The event B_{ih} that there is one point in N_i in $(t, t+h]$ is the event, B_h that there is exactly one point in any of the r processes together with a subset of A_h where there are two or more points in N in $(t, t+h]$ but exactly one is labeled i. Since $P(A_h|H_t) = o(h)$

$$\begin{aligned} P(B_{ih}|H_t) &= p_i P(B_h|H_t) + o(h) \\ &= p_i(\lambda h + o(h)) + o(h) \\ &= p_i \lambda h + o(h) \end{aligned}$$

Similarly, A_{ih} is a subset of A_h so

$$P(A_{ih}|H_t) = o(h)$$

This shows each N_i is Poisson with rate λp_i. To get independence requires more work; see the homework for an easier algebraic method.

3) Fix $s < t$. Let $N(s,t)$ be number of points in $(s,t]$. Given $N = n$ conditional dist of $N(s,t)$ is Binomial(n,p) with $p = (s-t)/T$. So

$$
\begin{aligned}
P(N(s,t) = k) & \\
&= \sum_{n=k}^{\infty} P(N(s,t) = k, N = n) \\
&= \sum_{n=k}^{\infty} P(N(s,t) = k | N = n) P(N = n) \\
&= \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \frac{(\lambda T)^n}{n!} e^{-\lambda T} \\
&= \frac{e^{-\lambda T}}{k!} (\lambda T p)^k \sum_{n=k}^{\infty} \frac{(1-p)^{n-k} (\lambda T)^{n-k}}{(n-k)!} \\
&= \frac{e^{-\lambda T}}{k!} (\lambda T p)^k \sum_{m=0}^{\infty} (1-p)^m (\lambda T)^m / m! \\
&= \frac{e^{-\lambda T}}{k!} (\lambda T p)^k e^{-\lambda T(1-p)} \\
&= \frac{e^{-\lambda(s-t)} (\lambda(s-t))^k}{k!}
\end{aligned}
$$

4): Fix s_i, h_i for $i = 1, \ldots, n$ such that

$$0 < s_1 < s_1 + h_1 < s_2 < \cdots < s_n < s_n + h_n < T$$

Given $N(T) = n$ we compute the probability of the event

$$A = \bigcap_{i=1}^{n} \{s_i < S_i < s_i + h_i\}$$

Intersection of A,1 $N(T) = n$ is ($s_0 = h_0 = 0$):

$$B \equiv \bigcap_{i=1}^{n} \{N(s_{i-1} + h_{i-1}, s_i] = 0, N(s_i, s_i + h_i] = 1\} \cap \{N(s_n + h_n, T] = 0\}$$

whose probability is

$$\left(\prod \lambda h_i\right) e^{-\lambda T}$$

So

$$
\begin{aligned}
P(A|N(t) = n) &= \frac{P(A, N(T) = n)}{(N(T) = n)} \\
&= \frac{\lambda^n e^{-\lambda T} \prod h_i}{(\lambda T)^n e^{-\lambda T}/n!} \\
&= \frac{n! \prod h_i}{T^n}
\end{aligned}
$$

Divide by $\prod h_i$ and let all h_i go to 0 to get joint density of $S_1, \ldots, S_n$ is

$$\frac{n!}{T^n} 1(0 < s_1 < \cdots < s_n < T)$$

which is the density of order statistics from a Uniform$[0, T]$ sample of size n.

5) Replace the event $S_{n+1} = T$ with $T < S_{n+1} < T + h$. With A as before we want

$$P(A|T < S_{n+1} < T + h) = \frac{P(B, N(T, T+h] \geq 1)}{P(T < S_{n+1} < T + h)}$$

Note that B is independent of $\{N(T, T+h] \geq 1\}$ and that we have already found the limit

$$\frac{P(B)}{\prod h_i} \to \lambda^n e^{-\lambda T}$$

We are left to compute the limit of

$$\frac{P(N(T,T+h]\geq 1)}{P(T<S_{n+1}<T+h)}$$

The denominator is

$$F_{S_{n+1}}(t+h)-F_{S_{n+1}}(t)=f_{S_{n+1}}(t)h+o(h)$$

Thus

$$\frac{P(N(T,T+h]\geq 1)}{P(T<S_{n+1}<T+h)}=\frac{\lambda h+o(h)}{\frac{(\lambda T)^n}{n!}e^{-\lambda T}\lambda h+o(h)}$$
$$\rightarrow \frac{n!}{(\lambda T)^n e^{-\lambda T}}$$

This gives the conditional density of $S_1,\ldots,S_n$ given $S_{n+1}=T$ as in **4)**.

Inhomogeneous Poisson Processes

Hazard rate can be used to extend notion of Poisson Process.

Suppose $\lambda(t) \ge 0$ is a function of t.

Suppose N is a counting process such that

$P(N(t+h) = k+1|N(t) = k, H_t) = \lambda(t)h+o(h)$

and

$$P(N(t+h) \ge k+2|N(t) = k, H_t) = o(h)$$

Then:

a) N has independent increments and

b) $N(t+s)-N(t)$ has Poisson distribution with mean

$$\int_t^{t+s} \lambda(u)du$$

If we put

$$\Lambda(t) = \int_0^t \lambda(u)du$$

then mean of $N(t+s)-N(T)$ is $\Lambda(t+s)-\Lambda(t)$.

Jargon: λ is the **intensity** or **instantaneous intensity** and Λ the **cumulative intensity**.

Can use the model with Λ any non-decreasing right continuous function, possibly without a derivative. This allows ties.

Continuous Time Markov Chains

Consider a population of single celled organisms in a stable environment.

Fix short time interval, length h.

Each cell has some prob of dividing to produce 2, some other prob of dying.

We might suppose:

- Different organisms behave independently.
- Probability of division (for specified organism) is λh plus $o(h)$.
- Probability of death is μh plus $o(h)$.
- Prob an organism divides twice (or divides once and dies) in interval of length h is $o(h)$.

Notice tacit assumption: constants of proportionality do not depend on time (that is our interpretation of "stable environment").

Notice too that we have taken the constants not to depend on which organism we are talking about. We are really assuming that the organisms are all similar and live in similar environments.

$Y(t)$: total population at time t.

$\mathcal{H}_t$: history of the process up to time t.

We generally take

$$\mathcal{H}_t = \sigma\{Y(s); 0 \leq s \leq t\}$$

General definition of a **history** (alternative jargon **filtration**): any family of σ-fields indexed by t satisfying:

- $s < t$ implies $\mathcal{H}_s \subset \mathcal{H}_t$.
- $Y(t)$ is a $\mathcal{H}_t$ measurable random variable.
- $\mathcal{H}_t = \bigcap_{s>t} \mathcal{H}_s$.

The last assumption is a technical detail we will ignore from now on.

Condition on event $Y(t) = n$.

Then the probability of two or more divisions (either more than one division by a single organism or two or more organisms dividing) is $o(h)$ by our assumptions.

Similarly the probability of both a division and a death or of two or more deaths is $o(h)$.

So probability of exactly 1 division by any one of the n organisms is $n\lambda h + o(h)$.

Similarly probability of 1 death is $n\mu h + o(h)$.

We deduce:

$$
\begin{aligned}
P(Y(t+h) &= n+1 | Y(t) = n, \mathcal{H}_t) \\
&= n\lambda h + o(h) \\
P(Y(t+h) &= n-1 | Y(t) = n, \mathcal{H}_t) \\
&= n\mu h + o(h) \\
P(Y(t+h) &= n | Y(t) = n, \mathcal{H}_t) \\
&= 1 - n(\lambda+\mu)h + o(h) \\
P(Y(t+h) &\notin \{n-1, n, n+1\} | Y(t) = n, \mathcal{H}_t) \\
&= o(h)
\end{aligned}
$$

These equations lead to:

$$
\begin{aligned}
P(Y(t+s) = j | Y(s) = i, \mathcal{H}_s) & \\
&= P(Y(t+s) = j | Y(s) = i) \\
&= P(Y(t) = j | Y(0) = i)
\end{aligned}
$$

This is the **Markov Property**.

Def'n: A process $\{Y(t); t \geq 0\}$ taking values in S, a finite or countable state space is a Markov Chain if

$$\begin{aligned} P(Y(t+s) = j | Y(s) = i, \mathcal{H}_s) \\ &= P(Y(t+s) = j | Y(s) = i) \\ &\equiv \mathbf{P}_{ij}(s, s+t) \end{aligned}$$

Def'n: A Markov chain Y has **stationary transitions** if

$$\mathbf{P}_{ij}(s, s+t) = \mathbf{P}_{ij}(0, t) \equiv \mathbf{P}_{ij}(t)$$

From now on: our chains have stationary transitions.

Summary of Markov Process Results

- Chapman-Kolmogorov equations:

$$\mathbf{P}_{ik}(t+s) = \sum_j \mathbf{P}_{ij}(t)\mathbf{P}_{jk}(s)$$

- Exponential holding times: starting from state i time, T_i, until process leaves i has exponential distribution, rate denoted v_i.

- Sequence of states visited, $Y_0, Y_1, Y_2, \ldots$ is Markov chain – transition matrix has $\mathbf{P}_{ii} = 0$. Y sometimes called **skeleton**.

- **Communicating classes** defined for skeleton chain. Usually assume chain has 1 communicating class.

- Periodicity irrelevant because of continuity of exponential distribution.

- Instantaneous transition rates from i to j:

$$q_{ij} = v_i \mathbf{P}_{ij}$$

- Kolmogorov backward equations:

$$\mathbf{P}'_{ij}(t) = \sum_{k \neq i} q_{ik} \mathbf{P}_{kj}(t) - v_i \mathbf{P}_{ij}(t)$$

- Kolmogorov forward equations:

$$\mathbf{P}'_{ij}(t) = \sum_{k \neq j} q_{kj} \mathbf{P}_{ik}(t) - v_i \mathbf{P}_{ij}(t)$$

- For strongly recurrent chains with a single communicating class:

$$\mathbf{P}_{ij}(t) \rightarrow \pi_j$$

- Stationary initial probabilities π_i satisfy:

$$v_j \pi_j = \sum_{k \neq j} q_{kj} \pi_k$$

- Transition probabilities given by

$$\mathbf{P}(t) = e^{\mathbf{R}t}$$

where $\mathbf{R}$ has entries

$$\mathbf{R}_{ij} = \begin{cases} q_{ij} & i \neq j \\ -v_i & i = j \end{cases}$$

- Process is a **Birth and Death** process if

$$\mathbf{P}_{ij} = 0 \text{ if } |i-j| > 1$$

In this case we write λ_i for the instantaneous "birth" rate:

$$P(Y(t+h) = i+1 | Y_t = i) = \lambda_i h + o(h)$$

and μ_i for the instantaneous "death" rate:

$$P(Y(t+h) = i-1 | Y_t = i) = \mu_i h + o(h)$$

We have

$$q_{ij} = \begin{cases} 0 & |i-j| > 1 \\ \lambda_i & j = i+1 \\ \mu_i & j = i-1 \end{cases}$$

- If all $\mu_i = 0$ then process is a **pure birth** process. If all $\lambda_i = 0$ a **pure death** process.

- Birth and Death process have stationary distribution

$$\pi_n = \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n \left(1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n}\right)}$$

  Necessary condition for existence of π is

$$\sum_{n=1}^{\infty} \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} < \infty$$

Detailed development

Suppose X a Markov Chain with stationary transitions. Then

$$
\begin{aligned}
P(X(t+s) = k|X(0) = i) \\
&= \sum_j P(X(t+s)=k, X(t)=j|X(0)=i) \\
&= \sum_j P(X(t+s)=k|X(t)=j, X(0)=i) \\
&\qquad \times P(X(t)=j|X(0)=i) \\
&= \sum_j P(X(t+s)=k|X(t)=j) \\
&\qquad \times P(X(t)=j|X(0)=i) \\
&= \sum_j P(X(s)=k|X(0)=j) \\
&\qquad \times P(X(t)=j|X(0)=i)
\end{aligned}
$$

This shows

$$\mathbf{P}(t+s) = \mathbf{P}(t)\mathbf{P}(s) = \mathbf{P}(s)\mathbf{P}(t)$$

which is the Chapman-Kolmogorov equation.

Now consider the chain starting from i and let T_i be the first t for which $X(t) \neq i$. Then T_i is a stopping time.

[Technically:

$$\{T_i \leq t\} \in \mathcal{H}_t$$

for each t.] Then

$$\begin{aligned} & P(T_i > t + s | T_i > s, X(0) = i) \\ & \qquad = P(T_i > t + s | X(u) = i; 0 \leq u \leq s) \\ & \qquad = P(T_i > t | X(0) = i) \end{aligned}$$

by the Markov property. Note: we actually are asserting a generalization of the Markov property: If f is some function on the set of possible paths of X then

$$\begin{aligned} \mathsf{E}(f(X(s+\cdot)) | X(u) = x(u), 0 \leq u \leq s) & \\ = \mathsf{E}\left[f(X(\cdot)) | X(0) = x(s)\right] & \\ = \mathsf{E}^{x(s)}\left[f(X(\cdot))\right] & \end{aligned}$$

The formula requires some sophistication to appreciate. In it, f is a function which associates a sample path of X with a real number. For instance,

$$f(x(\cdot)) = \text{sup}\{t : x(u) = i, 0 \le u \le t\}$$

is such a functional. Jargon: **functional** is a function whose argument is itself a function and whose value is a scalar.

FACT: Strong Markov Property – for a stopping time T

$$\text{E}\left[f\left\{X(T+\cdot)\right\}|\mathcal{F}_T\right] = \text{E}^{X(T)}\left[f\left\{X(\cdot)\right\}\right]$$

with suitable fix on event $T = \infty$.

Conclusion: given $X(0) = i$, T_i has memoryless property so T_i has an exponential distribution. Let v_i be the rate parameter.

Embedded Chain: Skeleton

Let $T_1 < T_2 < \cdots$ be the stopping times at which transitions occur.

Then $X_n = X(T_n)$.

Sequence X_n is a Markov chain by the strong Markov property.

That $\mathbf{P}_{ii} = 0$ reflects fact that $P(X(T_{n+1}) = X(T_n)) = 0$ by design.

As before we say $i \rightsquigarrow j$ if $\mathbf{P}_{ij}(t) > 0$ for some t. It is fairly clear that $i \rightsquigarrow j$ for the $X(t)$ if and only if $i \rightsquigarrow j$ for the embedded chain X_n.

We say $i \leftrightarrow j$ if $i \rightsquigarrow j$ and $j \rightsquigarrow i$.

Now consider

$$P(X(t+h)=j|X(t)=i,\mathcal{H}_t)$$

Suppose the chain has made n transitions so far so that $T_n < t < T_{n+1}$. Then the event $X(t+h)=j$ is, except for possibilities of probability $o(h)$ the event that

$$t < T_{n+1} \le t+h \text{ and } X_{n+1}=j$$

The probability of this is

$$(v_i h + o(h))\mathbf{P}_{ij} = v_i \mathbf{P}_{ij} h + o(h)$$

Kolmogorov's Equations

The Chapman-Kolmogorov equations are

$$\mathbf{P}(t+h) = \mathbf{P}(t)\mathbf{P}(h)$$

Subtract $\mathbf{P}(t)$ from both sides, divide by h and let $h \to 0$. Remember that $\mathbf{P}(0)$ is the identity. We find

$$\frac{\mathbf{P}(t+h) - \mathbf{P}(t)}{h} = \frac{\mathbf{P}(t)(\mathbf{P}(h) - \mathbf{P}(0))}{h}$$

which gives

$$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{P}'(0)$$

The Chapman-Kolmogorov equations can also be written

$$\mathbf{P}(t+h) = \mathbf{P}(h)\mathbf{P}(t)$$

Now subtracting $\mathbf{P}(t)$ from both sides, dividing by h and letting $h \to 0$ gives

$$\mathbf{P}'(t) = \mathbf{P}'(0)\mathbf{P}(t)$$

Look at these equations in component form:

$$\mathbf{P}'(t) = \mathbf{P}'(0)\mathbf{P}(t)$$

becomes

$$\mathbf{P}'_{ij}(t) = \sum_k \mathbf{P}'_{ik}(0)\mathbf{P}_{kj}(t)$$

For $i \neq k$ our calculations of instantaneous transition rates gives

$$\mathbf{P}'_{ik}(0) = v_i \mathbf{P}_{ik}$$

For $i = k$ we have

$$P(X(h) = i | X(0) = i) = e^{-v_i h} + o(h)$$

($X(h) = i$ either means $T_i > h$ which has probability $e^{-v_i h}$ or there have been two or more transitions in $[0, h]$, a possibility of probability $o(h)$.) Thus

$$\mathbf{P}'_{ii}(0) = -v_i$$

Let $\mathbf{R}$ be the matrix with entries

$$\mathbf{R}_{ij} = \begin{cases} q_{ij} \equiv v_i \mathbf{P}_{ij} & i \neq j \\ -v_i & i = j \end{cases}$$

$\mathbf{R}$ is the **infinitesimal generator** of the chain.

Thus

$$\mathbf{P}'(t) = \mathbf{P}'(0)\mathbf{P}(t)$$

becomes

$$\begin{aligned} \mathbf{P}'_{ij}(t) &= \sum_k \mathbf{R}_{ik} \mathbf{P}_{kj}(t) \\ &= \sum_{k \neq i} q_{ik} \mathbf{P}_{kj}(t) - v_i \mathbf{P}_{ij}(t) \end{aligned}$$

Called **Kolmogorov's backward equations**.

On the other hand

$$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{P}'(0)$$

becomes

$$\begin{aligned}\mathbf{P}'_{ij}(t) &= \sum_k \mathbf{P}_{ik}(t)\mathbf{R}_{kj} \\ &= \sum_{k\neq j} q_{kj}\mathbf{P}_{ik}(t) - v_j\mathbf{P}_{ij}(t)\end{aligned}$$

These are **Kolmogorov's forward equations**.

Remark: When the state space is infinite the forward equations may not be justified. In deriving them we interchanged a limit with an infinite sum; the interchange is always justified for the backward equations but not for forward.

Example: $S = \{0, 1\}$. Then

$$\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and the chain is otherwise specified by v_0 and v_1. The matrix $\mathbf{R}$ is

$$\mathbf{R} = \begin{bmatrix} -v_0 & v_0 \\ v_1 & -v_1 \end{bmatrix}$$

The backward equations become

$$\begin{aligned}
\mathbf{P}'_{00}(t) &= v_0 \mathbf{P}_{10}(t) - v_0 \mathbf{P}_{00}(t) \\
\mathbf{P}'_{01}(t) &= v_0 \mathbf{P}_{11}(t) - v_0 \mathbf{P}_{01}(t) \\
\mathbf{P}'_{10}(t) &= v_1 \mathbf{P}_{00}(t) - v_1 \mathbf{P}_{10}(t) \\
\mathbf{P}'_{11}(t) &= v_1 \mathbf{P}_{01}(t) - v_1 \mathbf{P}_{11}(t)
\end{aligned}$$

while the forward equations are

$$\begin{aligned}
\mathbf{P}'_{00}(t) &= v_1 \mathbf{P}_{01}(t) - v_0 \mathbf{P}_{00}(t) \\
\mathbf{P}'_{01}(t) &= v_0 \mathbf{P}_{00}(t) - v_1 \mathbf{P}_{01}(t) \\
\mathbf{P}'_{10}(t) &= v_1 \mathbf{P}_{11}(t) - v_0 \mathbf{P}_{10}(t) \\
\mathbf{P}'_{11}(t) &= v_0 \mathbf{P}_{10}(t) - v_1 \mathbf{P}_{11}(t)
\end{aligned}$$

Add v_1 times first and v_0 times third backward equations to get

$$v_1 \mathbf{P}'_{00}(t) + v_0 \mathbf{P}'_{10}(t) - 0$$

so

$$v_1 \mathbf{P}_{00}(t) + v_0 \mathbf{P}_{10}(t) = c.$$

Put $t = 0$ to get $c = v_1$. This gives

$$\mathbf{P}_{10}(t) = \frac{v_1}{v_0} \{1 - \mathbf{P}_{00}(t)\}$$

Plug this back in to the first equation and get

$$\mathbf{P}'_{00}(t) = v_1 - (v_1 + v_0)\mathbf{P}_{00}(t)$$

Multiply by $e^{(v_1+v_0)t}$ and get

$$\left\{e^{(v_1+v_0)t}\mathbf{P}_{00}(t)\right\}' = v_1 e^{(v_1+v_0)t}$$

which can be integrated to get

$$\mathbf{P}_{00}(t) = \frac{v_1}{v_0 + v_1} + \frac{v_0}{v_0 + v_1} e^{-(v_1+v_0)t}$$

Alternative calculation:

$$\mathbf{R} = \begin{bmatrix} -v_0 & v_0 \\ v_1 & -v_1 \end{bmatrix}$$

can be written as

$$\mathbf{M}\mathbf{\Lambda}\mathbf{M}^{-1}$$

where

$$\mathbf{M} = \begin{bmatrix} 1 & v_0 \\ \\ 1 & -v_1 \end{bmatrix}$$

$$\mathbf{M}^{-1} = \begin{bmatrix} \frac{v_1}{v_0+v_1} & \frac{v_0}{v_0+v_1} \\ \\ \frac{1}{v_0+v_1} & \frac{-1}{v_0+v_1} \end{bmatrix}$$

and

$$\mathbf{\Lambda} = \begin{bmatrix} 0 & 0 \\ \\ 0 & -(v_0+v_1) \end{bmatrix}$$

Then

$$
\begin{aligned}
e^{\mathbf{R}t} &= \sum_0^\infty \mathbf{R}^n t^n / n! \\
&= \sum_0^\infty \left(\mathbf{M}\boldsymbol{\Lambda}\mathbf{M}^{-1}\right)^n \frac{t^n}{n!} \\
&= \mathbf{M}\left(\sum_0^\infty \boldsymbol{\Lambda}^n \frac{t^n}{n!}\right)\mathbf{M}^{-1}
\end{aligned}
$$

Now

$$
\sum_0^\infty \boldsymbol{\Lambda}^n \frac{t^n}{n!} = \begin{bmatrix} 1 & 0 \\ 0 & e^{-(v_0+v_1)t} \end{bmatrix}
$$

so we get

$$
\begin{aligned}
\mathbf{P}(t) = e^{\mathbf{R}t} &= \mathbf{M}\begin{bmatrix} 1 & 0 \\ 0 & e^{-(v_0+v_1)t} \end{bmatrix}\mathbf{M}^{-1} \\
&= \mathbf{P}^\infty - \frac{e^{-(v_0+v_1)t}}{v_0+v_1}\mathbf{R}
\end{aligned}
$$

where

$$
\mathbf{P}^\infty = \begin{bmatrix} \frac{v_1}{v_0+v_1} & \frac{v_0}{v_0+v_1} \\ \\ \frac{v_1}{v_0+v_1} & \frac{v_0}{v_0+v_1} \end{bmatrix}
$$

Notice: rows of $\mathbf{P}^\infty$ are a stationary initial distribution. If rows are π then

$$\mathbf{P}^\infty = \left[\begin{array}{c} 1 \\ 1 \end{array}\right] \pi \equiv \mathbf{1}\pi$$

so

$$\pi\mathbf{P}^\infty = (\pi\mathbf{1})\pi = \pi$$

Moreover

$$\pi\mathbf{R} = \mathbf{0}$$

Fact: $\pi_0 = v_1/(v_0 + v_1)$ is long run fraction of time in state 0.

Fact:

$$\frac{1}{T}\int_0^T f(X(t))dt \to \sum_j \pi_j f(j)$$

Ergodic Theorem in continuous time.

Potential Pathologies

Suppose that for each k you have a sequence

$$T_{k,1}, T_{k,2}, \cdots$$

such that all T_{ij} are independent exponential random variables and T_{ij} has rate parameter λ_j. We can use these times to make a Markov chain with state space $S = \{1, 2, \ldots\}$:

Start the chain in state 1. At time $T_{1,1}$ move to 2, $T_{1,2}$ time units later move to 3, etc. Chain progresses through states in order 1,2,.... We have

$$v_i = \lambda_i$$

and

$$\mathbf{P}_{ij} = \begin{cases} 0 & j \neq i+1 \\ 1 & j = i+1 \end{cases}$$

Does this define a process?

Depends on $\sum \lambda_i^{-1}$.

Case 1: if $\sum \lambda_i^{-1} = \infty$ then

$$P(\sum_1^\infty T_{1,j} = \infty) = 1$$

(converse to Borel Cantelli) and our construction defines a process $X(t)$ for all t.

Case 2: if $\sum \lambda_j^{-1} < \infty$ then for each k

$$P(\sum_{j=1}^\infty T_{kj} < \infty) = 1$$

In this case put $T_k = \sum_{j=1}^\infty T_{kj}$. Our definition above defines a process $X(t)$ for $0 \le t < T_1$. We put $X(T_1) = 1$ and then begin the process over with the set of holding times $T_{2,j}$. This defines X for $T_1 \le t < T_1 + T_2$. Again we put $X(T_2) = 1$ and continue the process.

Result: X is a Markov Chain with specified transition rates.

Problem: what if we put $X(T_1) = 2$ and continued?

What if we used probability vector $\alpha_1, \alpha_2, \ldots$ to pick a value for $X(T_1)$ and continued?

All yield Markov Processes with the same infinitesimal generator $\mathbf{R}$.

Point of all this: gives example of non-unique solution of differential equations!

Birth and Death Processes

Consider a population of $X(t) = i$ individuals. Suppose in next time interval $(t, t+h)$ probability of population increase of 1 (called a birth) is $\lambda_i h + o(h)$ and probability of decrease of 1 (death) is $\mu_i h + o(h)$.

Jargon: X is a birth and death process.

Special cases:

All $\mu_i = 0$; called a **pure birth** process.

All $\lambda_i = 0$ (0 is absorbing): **pure death** process.

$\lambda_n = n\lambda$ and $\mu_n = n\mu$ is a **linear** birth and death process.

$\lambda_n \equiv 1$, $\mu_n \equiv 0$: Poisson Process.

$\lambda_n = n\lambda + \theta$ and $\mu_n = n\mu$ is a **linear** birth and death process with immigration.

Applications:

1) cable strength: Cable consists of n fibres.

$X(t)$ is number which have *not* failed up to time t.

Pure death process: μ_i will be large for small i, small for large i.

2) Chain reactions. $X(t)$ is number of free neutrons in lump of uranium.

Births produced as sum of: spontaneous fission rate (problem — I think each fission produces 2 neutrons) plus rate of collision of neutron with nuclei.

Ignore: neutrons leaving sample and decay of free neutrons.

Get $\lambda_n = n\lambda + \theta$

(At least in early stages where decay has removed a negligible fraction of atoms).

Conditions for stationary initial distribution:

1) $v_n = \lambda_n + \mu_n$.

2) $P_{n,n+1} = \lambda_n / v_n = 1 - P_{n,n-1}$.

3)

$$v_n \pi_n = \lambda_{n-1} \pi_{n-1} + \mu_{n+1} \pi_{n+1}$$

4) Start at $n = 0$:

$$\lambda_0 \pi_0 = \mu_1 \pi_1$$

so $\pi_1 = (\lambda_0 / \mu_1) \pi_0$.

5) Now look at $n = 1$.

$$(\lambda_1 + \mu_1) \pi_1 = \lambda_0 \pi_0 + \mu_2 \pi_2$$

Solve for π_2 to get

$$\pi_2 = \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} \pi_0$$

And so on. Then use $\sum \pi_n = 1$.

Queuing Theory

Ingredients of Queuing Problem:

1: Queue input process.

2: Number of servers

3: Queue discipline: first come first serve? last in first out? pre-emptive priorities?

4: Service time distribution.

Example: Imagine customers arriving at a facility at times of a Poisson Process N with rate λ. This is the input process, denoted M (for Markov) in queuing literature.

Single server case:

Service distribution: exponential service times, rate μ.

Queue discipline: first come first serve.

$X(t)$ = number of customers in line at time t.

X is a Markov process called $M/M/1$ queue:

$$v_i = \lambda + \mu 1(i > 0)$$

$$\mathbf{P}_{ij} = \begin{cases} \frac{\mu}{\mu+\lambda} & j = i - 1 \geq 0 \\ \frac{\lambda}{\mu+\lambda} & j = i + 1, i > 0 \\ 1 & j = 1, i = 0 \\ 0 & \text{otherwise} \end{cases}$$

Example: $M/M/\infty$ queue:

Customers arrive according to PP rate λ. Each customer begins service immediately. $X(t)$ is number being served at time t. X is a birth and death process with

$$v_n = \lambda + n\mu$$

and

$$\mathbf{P}_{ij} = \begin{cases} \frac{i\mu}{i\mu+\lambda} & j = i-1 \geq 0 \\ \frac{\lambda}{i\mu+\lambda} & j = i+1 \\ 0 & \text{otherwise} \end{cases}$$

Stationary distributions?

For $M/M/1$ queue:

Solve

$$\{\lambda + \mu 1(n > 0)\}\pi_n = \mu\pi_{n+1} + \lambda 1(n > 0)\pi_{n-1}$$

Just use general birth and death process formulation:

$$\lambda_n = \lambda \quad \mu_n = \mu 1(n > 0)$$

so

$$\frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} = (\lambda/\mu)^n$$

and

$$\sum_{n=0}^{\infty} (\lambda/\mu)^n = 1/(1 - \lambda/\mu)$$

so

$$\pi_n = \frac{(\lambda/\mu)^n}{1 + 1/(1 - \lambda/\mu)}$$

Exists only if $\lambda < \mu$.

For $M/M/\infty$ queue:

$$\pi_n \propto \frac{\lambda^n}{\mu^n n!}$$

and

$$\sum_{n=0}^{\infty} \frac{\lambda^n}{\mu^n n!} = \exp(\lambda/\mu)$$

so

$$\pi_n = \exp(-\lambda/\mu)\frac{\lambda^n}{\mu^n n!}$$

Notice this exists for all $\lambda > 0$ and all $\mu > 0$.

Scope of Queuing Theory:

1) $M/M/k$ queues. $X(t)$ is number queued or in service.

Birth and Death process; death rate maxes out at $k\mu$.

Stationary distribution exists if $\lambda < k\mu$.

2) Same input / service processes as $M/M/k$ but customers not served leave. Question of interest: customers lost per time unit?

Take X to be number in service. ($0 \le X(t) \le k$).

Find stationary distribution.

Fraction of time spent in state k is π_k.

During time in state k lose customers at rate λ. So lost $\pi_k\lambda$ customers per unit time.

3) $G/M/1$ queue. General distribution of interarrival times for input. Input is a **renewal process**. Not Markov.

4) $M/G/1$ and others.

5) Networks: output of 1 queue is input of next; feedback ...

Quantities of potential interest:

Average fraction of time server idle.

Average time in system for customer.

Average wait to see server.

One example calculation: in $G/M/1$ queue.

Compute long run fraction time system is idle.

Idea: interarrival times are iid with cdf G.

Service rate μ.

Let X_n be number of customers in service / in line when nth customer arrives.

Claim X_n is a Markov chain.

(Example of general tactic: find simple process buried within process of interest.)

Notation: $T_1, T_2, \cdots$ iid interarrival times.

Given $X_n = i$ and $T_{n+1} = t$ number served between nth arrival and $n+1$st arrival is

$$\min\{\text{Poisson}(\mu t), i+1\}$$

So: if $X_n = i$ and the Poisson variable above is j then

$$X_{n+1} = i + 1 - \min\{j, i+1\}$$

Now to compute prob of j served must average over T_{n+1}:

$$P(j \text{ served}) = \int e^{-\mu t} \frac{(\mu t)^j}{j!} dG(t) \equiv a_j$$

for $j \le i+1$.

This gives:

$$P_{ik} = \begin{cases} a_{i+1-k} & 1 \le k \le i+1 \\ 1 - \sum_0^i a_j & k = 0 \\ 0 & \text{otherwise} \end{cases}$$

Computing stationary distribution?

Brownian Motion

For fair random walk Y_n = number of heads minus number of tails,

$$Y_n = U_1 + \cdots + U_n$$

where the U_i are independent and

$$P(U_i = 1) = P(U_i = -1) = \frac{1}{2}$$

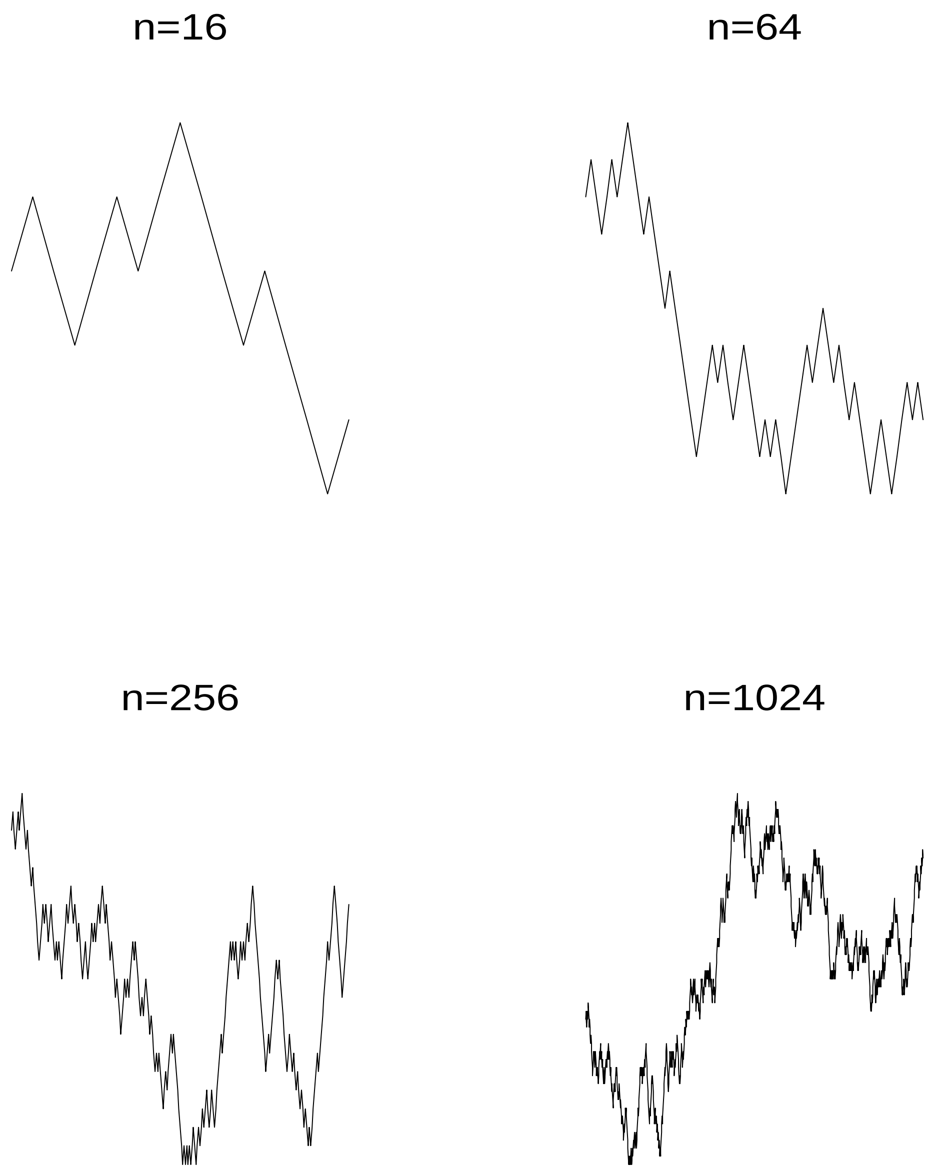
Notice:

$$\begin{aligned} \mathrm{E}(U_i) &= 0 \\ \mathrm{Var}(U_i) &= 1 \end{aligned}$$

Recall central limit theorem:

$$\frac{U_1 + \cdots + U_n}{\sqrt{n}} \Rightarrow N(0,1)$$

Now: rescale time axis so that n steps take 1 time unit and vertical axis so step size is $1/\sqrt{n}$.


n=16
n=64
n=256
n=1024

We now turn these pictures into a stochastic process:

For $\frac{k}{n} \le t < \frac{k+1}{n}$ we define

$$X_n(t) = \frac{U_1 + \cdots + U_k}{\sqrt{n}}$$

Notice:

$$\mathrm{E}(X_n(t)) = 0$$

and

$$\mathrm{Var}(X_n(t)) = \frac{k}{n}$$

As $n \to \infty$ with t fixed we see $k/n \to t$. Moreover:

$$\frac{U_1 + \cdots + U_k}{\sqrt{k}} = \sqrt{\frac{n}{k}} X_n(t)$$

converges to $N(0,1)$ by the central limit theorem. Thus

$$X_n(t) \Rightarrow N(0,t)$$

Also: $X_n(t+s)-X_n(t)$ is independent of $X_n(t)$ because the 2 rvs involve sums of different U_i.

Conclusions.

As $n \to \infty$ the processes X_n converge to a process X with the properties:

1. $X(t)$ has a $N(0,t)$ distribution.

2. X has independent increments: if

   $$0 = t_0 < t_1 < t_2 < \cdots < t_k$$

   then

   $$X(t_1) - X(t_0), \ldots, X(t_k) - X(t_{k-1})$$

   are independent .

3. The increments are **stationary**: for all s

   $$X(t+s) - X(s) \sim N(0,t)$$

4. $X(0) = 0$.

Def'n: Any process satisfying 1-4 above is a Brownian motion.

Properties of Brownian motion

- Suppose $t > s$. Then

$$
\begin{aligned}
\mathsf{E}(X(t)|X(s)) &= \mathsf{E}\left\{X(t) - X(s) + X(s)|X(s)\right\} \\
&= \mathsf{E}\left\{X(t) - X(s)|X(s)\right\} \\
&\qquad + \mathsf{E}\left\{X(s)|X(s)\right\} \\
&= 0 + X(s) = X(s)
\end{aligned}
$$

Notice the use of independent increments and of $\mathsf{E}(Y|Y) = Y$.

- Again if $t > s$:

$$
\begin{aligned}
&\mathsf{Var}\left\{X(t)|X(s)\right\} \\
&\quad = \mathsf{Var}\left\{X(t) - X(s) + X(s)|X(s)\right\} \\
&\quad = \mathsf{Var}\left\{X(t) - X(s)|X(s)\right\} \\
&\quad = \mathsf{Var}\left\{X(t) - X(s)\right\} \\
&\quad = t - s
\end{aligned}
$$

Suppose $t < s$. Then $X(s) = X(t) + \{X(t) - X(s)\}$ is a sum of two independent normal variables. Do following calculation:

$X \sim N(0, \sigma^2)$, and $Y \sim N(0, \tau^2)$ independent. $Z = X + Y$.

Compute conditional distribution of X given Z:

$$\begin{aligned} f_{X|Z}(x|z) &= \frac{f_{X,Z}(x,z)}{f_Z(z)} \\ &= \frac{f_{X,Y}(x, z-x)}{f_Z(z)} \\ &= \frac{f_X(x) f_Y(z-x)}{f_Z(z)} \end{aligned}$$

Now Z is $N(0, \gamma^2)$ where $\gamma^2 = \sigma^2 + \tau^2$ so

$$\begin{aligned} f_{X|Z}(x|z) &= \frac{\frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/(2\sigma^2)} \frac{1}{\tau\sqrt{2\pi}} e^{-(z-x)^2/(2\tau^2)}}{\frac{1}{\gamma\sqrt{2\pi}} e^{-z^2/(2\gamma^2)}} \\ &= \frac{\gamma}{\tau\sigma\sqrt{2\pi}} \exp\{-(x-a)^2/(2b^2)\} \end{aligned}$$

for suitable choices of a and b. To find them compare coefficients of x^2, x and 1.

Coefficient of x^2:

$$\frac{1}{b^2} = \frac{1}{\sigma^2} + \frac{1}{\tau^2}$$

so $b = \tau\sigma/\gamma$.

Coefficient of x:

$$\frac{a}{b^2} = \frac{z}{\tau^2}$$

so that

$$a = b^2 z/\tau^2 = \frac{\sigma^2}{\sigma^2 + \tau^2} z$$

Finally you should check that

$$\frac{a^2}{b^2} = \frac{z^2}{\tau^2} - \frac{z^2}{\gamma^2}$$

to make sure the coefficients of 1 work out as well.

Conclusion: given $Z = z$ the conditional distribution of X is $N(a, b^2)$ with a and b as above.

Application to Brownian motion:

- For $t < s$ let X be $X(t)$ and Y be $X(s) - X(t)$ so $Z = X + Y = X(s)$. Then $\sigma^2 = t$, $\tau^2 = s - t$ and $\gamma^2 = s$. Thus

  $$b^2 = \frac{(s-t)t}{s}$$

  and

  $$a = \frac{t}{s} X(s)$$

  SO:

  $$\mathsf{E}(X(t)|X(s)) = \frac{t}{s} X(s)$$

  and

  $$\mathsf{Var}(X(t)|X(s)) = \frac{(s-t)t}{s}$$

The Reflection Principle

Tossing a fair coin:

HTHHHTHTHHTHHHTTHTH	5 more heads than tails
THTTTHTHTTHTTTHHTHT	5 more tails than heads

Both sequences have the same probability.

So: for random walk starting at stopping time:

Any sequence with k more heads than tails in next m tosses is matched to sequence with k more tails than heads. Both sequences have same prob.

Suppose Y_n is a fair ($p = 1/2$) random walk. Define

$$M_n = \max\{Y_k, 0 \le k \le n\}$$

Compute $P(M_n \geq x)$? Trick: Compute

$$P(M_n \geq x, Y_n = y)$$

First: if $y \geq x$ then

$$\{M_n \geq x, Y_n = y\} = \{Y_n = y\}$$

Second: if $M_n \geq x$ then

$$T \equiv \min\{k : Y_k = x\} \leq n$$

Fix $y < x$. Consider a sequence of H's and T's which leads to say $T = k$ and $Y_n = y$.

Switch the results of tosses $k+1$ to n to get a sequence of H's and T's which has $T = k$ and $Y_n = x + (x - y) = 2x - y > x$. This proves

$$P(T = k, Y_n = y) = P(T = k, Y_n = 2x - y)$$

This is true for each k so

$$\begin{aligned} P(M_n \geq x, Y_n = y) &= P(M_n \geq x, Y_n = 2x - y) \\ &= P(Y_n = 2x - y) \end{aligned}$$

Finally, sum over all y to get

$$\begin{aligned} P(M_n \geq x) = \sum_{y \geq x} & P(Y_n = y) \\ &+ \sum_{y < x} P(Y_n = 2x - y) \end{aligned}$$

Make the substitution $k = 2x - y$ in the second sum to get

$$\begin{aligned} P(M_n \geq x) &= \sum_{y \geq x} P(Y_n = y) \\ &\quad + \sum_{k > x} P(Y_n = k) \\ &= 2 \sum_{k > x} P(Y_n = k) + P(Y_n = x) \end{aligned}$$

Brownian motion version:

$$M_t = \max\{X(s); 0 \le s \le t\}$$

$$T_x = \min\{s : X(s) = x\}$$

(called hitting time for level x). Then

$$\{T_x \le t\} = \{M_t \ge x\}$$

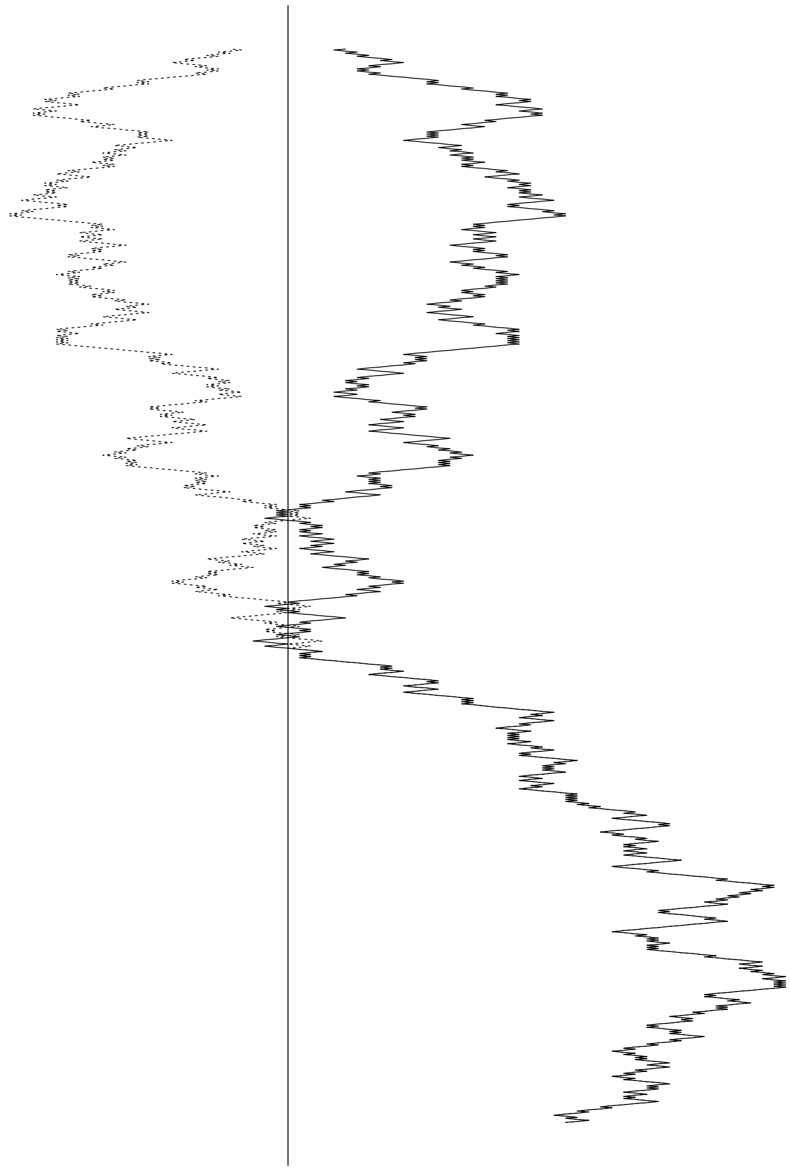
Any path with $T_x = s < t$ and $X(t) = y < x$ is matched to an equally likely path with $T_x = s < t$ and $X(t) = 2x - y > x$.

So for $y > x$

$$P(M_t \ge x, X(t) > y) = P(X(t) > y)$$

while for $y < x$

$$P(M_t \ge x, X(t) < y) = P(X(t) > 2x - y)$$



Let $y \to x$ to get

$$\begin{aligned} P(M_t \geq x, X(t) > x) &= P(M_t \geq x, X(t) < x) \\ &= P(X(t) > x) \end{aligned}$$

Adding these together gives

$$\begin{aligned} P(M_t > x) &= 2P(X(t) > x) \\ &= 2P(N(0,1) > x/\sqrt{t}) \end{aligned}$$

Hence M_t has the distribution of $|N(0,t)|$.

On the other hand in view of

$$\{T_x \le t\} = \{M_t \ge x\}$$

the density of T_x is

$$\frac{d}{dt} 2P(N(0,1) > x/\sqrt{t})$$

Use the chain rule to compute this. First

$$\frac{d}{dy} P(N(0,1) > y) = -\phi(y)$$

where ϕ is the standard normal density

$$\phi(y) = \frac{e^{-y^2/2}}{\sqrt{2\pi}}$$

because $P(N(0,1) > y)$ is 1 minus the standard normal cdf.

So

$$\frac{d}{dt} 2P(N(0,1) > x/\sqrt{t})$$
$$= -2\phi(x/\sqrt{t}) \frac{d}{dt}(x/\sqrt{t})$$
$$= \frac{x}{\sqrt{2\pi} t^{3/2}} \exp\{-x^2/(2t)\}$$

This density is called the **Inverse Gaussian** density. T_x is called a **first passage time**

NOTE: the preceding is a density when viewed as a function of the variable t.

Martingales

A stochastic process $M(t)$ indexed by either a discrete or continuous time parameter t is a **martingale** if:

$$\mathrm{E}\{M(t)|M(u); 0 \le u \le s\} = M(s)$$

whenever $s < t$.

Examples

- A fair random walk is a martingale.
- If $N(t)$ is a Poisson Process with rate λ then $N(t) - \lambda t$ is a martingale.
- Standard Brownian motion (defined above) is a martingale.

Note: Brownian motion with drift is a process of the form

$$X(t) = \sigma B(t) + \mu t$$

where B is **standard** Brownian motion, introduced earlier. X is a martingale if $\mu = 0$. We call μ the **drift**

- If $X(t)$ is a Brownian motion with drift then

$$Y(t) = e^{X(t)}$$

  is a geometric Brownian motion. For suitable μ and σ we can make $Y(t)$ a martingale.

- If a gambler makes a sequence of fair bets and M_n is the amount of money s/he has after n bets then M_n is a martingale – even if the bets made depend on the outcomes of previous bets, that is, even if the gambler plays a strategy.

Some evidence for some of the above:

Random walk: $U_1, U_2, \ldots$ iid with

$$P(U_i = 1) = P(U_i = -1) = 1/2$$

and $Y_k = U_1 + \cdots + U_k$ with $Y_0 = 0$. Then

$$\begin{aligned}
\mathsf{E}(Y_n|Y_0, \ldots, Y_k) & \\
&= \mathsf{E}(Y_n - Y_k + Y_k|Y_0, \ldots, Y_k) \\
&= \mathsf{E}(Y_n - Y_k|Y_0, \ldots, Y_k) + Y_k \\
&= \sum_{k+1}^{n} \mathsf{E}(U_j|U_1, \ldots, U_k) + Y_k \\
&= \sum_{k+1}^{n} \mathsf{E}(U_j) + Y_k \\
&= Y_k
\end{aligned}$$

Things to notice:

Y_k treated as constant given $Y_1, \ldots, Y_k$.

Knowing $Y_1, \ldots, Y_k$ is equivalent to knowing $U_1, \ldots, U_k$.

For $j > k$ we have U_j independent of $U_1, \ldots, U_k$ so conditional expectation is unconditional expectation.

Since Standard Brownian Motion is limit of such random walks we get martingale property for standard Brownian motion.

Poisson Process: $X(t) = N(t) - \lambda t$. Fix $t > s$.

$$
\begin{aligned}
\mathsf{E}(X(t)|X(u); 0 \le u \le s) & \\
&= \mathsf{E}(X(t) - X(s) + X(s)|\mathcal{H}_s) \\
&= \mathsf{E}(X(t) - X(s)|\mathcal{H}_s) + X(s) \\
&= \mathsf{E}(N(t) - N(s) - \lambda(t-s)|\mathcal{H}_s) + X(s) \\
&= \mathsf{E}(N(t) - N(s)) - \lambda(t-s) + X(s) \\
&= \lambda(t-s) - \lambda(t-s) + X(s) \\
&= X(s)
\end{aligned}
$$

Things to notice:

I used independent increments.

$\mathcal{H}_s$ is shorthand for the conditioning event.

Similar to random walk calculation.

Black Scholes

We model the price of a stock as

$$X(t) = x_0 e^{Y(t)}$$

where

$$Y(t) = \sigma B(t) + \mu t$$

is a Brownian motion with drift (B is standard Brownian motion).

If annual interest rates are $e^{\alpha} - 1$ we call α the instantaneous interest rate; if we invest \$1 at time 0 then at time t we would have $e^{\alpha t}$. In this sense an amount of money $x(t)$ to be paid at time t is worth only $e^{-\alpha t} x(t)$ at time 0 (because that much money at time 0 will grow to $x(t)$ by time t).

Present Value: If the stock price at time t is $X(t)$ per share then the present value of 1 share to be delivered at time t is

$$Z(t) = e^{-\alpha t} X(t)$$

With X as above we see

$$Z(t) = x_0 e^{\sigma B(t) + (\mu - \alpha)t}$$

Now we compute

$$\begin{aligned} \mathrm{E}\left\{Z(t) | Z(u); 0 \le u \le s\right\} \\ = \mathrm{E}\left\{Z(t) | B(u); 0 \le u \le s\right\} \end{aligned}$$

for $s < t$. Write

$$Z(t) = x_0 e^{\sigma B(s) + (\mu - \alpha)t} \times e^{\sigma(B(t) - B(s))}$$

Since B has independent increments we find

$$\begin{aligned} \mathrm{E}\left\{Z(t) | B(u); 0 \le u \le s\right\} \\ = x_0 e^{\sigma B(s) + (\mu - \alpha)t} \times \mathrm{E}\left[e^{\sigma\{B(t) - B(s)\}}\right] \end{aligned}$$

Note: $B(t) - B(s)$ is $N(0, t-s)$; the expected value needed is the moment generating function of this variable at σ.

Suppose $U \sim N(0,1)$. The Moment Generating Function of U is

$$M_U(r) = \mathsf{E}(e^{rU}) = e^{r^2/2}$$

Rewrite

$$\sigma\{B(t) - B(s)\} = \sigma(t-s)U$$

where $U \sim N(0,1)$ to see

$$\mathsf{E}\left[e^{\sigma\{B(t)-B(s)\}}\right] = e^{\sigma^2(t-s)/2}$$

Finally we get

$$\begin{aligned}\mathsf{E}\{Z(t)|Z(u); 0 \le u \le s\} \\ &= x_0 e^{\sigma B(s)+(\mu-\alpha)s} e^{(\mu-\alpha)(t-s)+\sigma^2(t-s)/2} \\ &= Z(s)\end{aligned}$$

provided

$$\mu + \sigma^2/2 = \alpha\,.$$

If this identity is satisfied then the present value of the stock price is a martingale.

Option Pricing

Suppose you can pay $\$c$ today for the right to pay K for a share of this stock at time t (regardless of the actual price at time t).

If, at time t, $X(t) > K$ you will**exercise** your **option** and buy the share making $X(t) - K$ dollars.

If $X(t) \le K$ you will not exercise your option; it becomes worthless.

The present value of this option is

$$e^{-\alpha t}(X(t) - K)_+ - c$$

where

$$z_+ = \begin{cases} z & z > 0 \\ 0 z \le 0 & \end{cases}$$

(Called **positive part** of z.)

In a fair market:

- The discounted share price $e^{-\alpha t}X(t)$ is a martingale.

- The expected present value of the option is 0.

So:

$$c = e^{-\alpha t}\mathsf{E}\left[\{X(t) - K\}_+\right]$$

Since

$$X(t) = x_0 e^{N(\mu t, \sigma^2 t)}$$

we are to compute

$$\mathsf{E}\left\{\left(x_0 e^{\sigma t^{1/2}U + \mu t} - K\right)_+\right\}$$

This is

$$\int_a^\infty \left(x_0 e^{bu+d} - K\right) e^{-u^2/2} du/\sqrt{2\pi}$$

where

$$\begin{aligned} a &= (\log(K/x_0) - \mu t)/(\sigma t^{1/2}) \\ b &= \sigma t^{1/2} \\ d &= \mu t \end{aligned}$$

Evidently

$$K \int_a^\infty e^{-u^2/2} du/\sqrt{2\pi} = KP(N(0,1) > a)$$

The other integral needed is

$$\begin{aligned} \int_a^\infty e^{-u^2/2+bu} du/\sqrt{2\pi} \\ &= \int_a^\infty \frac{e^{-(u-b)^2/2} e^{b^2/2}}{\sqrt{2\pi}} du \\ &= \int_{a-b}^\infty \frac{e^{-v^2/2} e^{b^2/2}}{\sqrt{2\pi}} dv \\ &= e^{b^2/2} P(N(0,1) > a-b) \end{aligned}$$

Introduce the notation

$$\Phi(v) = P(N(0,1) \le v) = P(N(0,1) > -v)$$

and do all the algebra to get

$$\begin{aligned} c &= \left\{ e^{-\alpha t} e^{b^2/2+d} x_0 \Phi(b-a) - K e^{-\alpha t} \Phi(-a) \right\} \\ &= x_0 e^{(\mu+\sigma^2/2-\alpha)t} \Phi(b-a) - K e^{-\alpha t} \Phi(-a) \\ &= x_0 \Phi(b-a) - K e^{-\alpha t} \Phi(-a) \end{aligned}$$

This is the Black-Scholes option pricing formula.

Modelling Traffic Loading on the Lion's Gate Bridge

Idea: want to know how strong bridge needs to be.

Compute: load x such that

Expected time to first exceedance of load x is 100 years.

Method uses:

1) modelling assumptions.

2) conservative modelling; to replace random variable of interest with stochastically larger quantity.

3) moment generating functions; Markov's inequality — compute upper bound on x.

Facts about the bridge:

Built 1936-38 for $6M.

3 spans: 614, 1550, 614 feet long.

Originally 2 lanes now 3.

Originally toll bridge built by developers.

See

`http://www.b-t.com/projects/liongate.htm`

at Buckland and Taylor web site for engineering info.

Begin with definition of process of interest.

Think of bridge as rectangle.

Co-ordinates: x runs from 0 to length L_B of bridge, y runs from 0 to width W_B of bridge.

Define:

$Z(x, y, t) =$ load on bridge at (x, y) at time t

General quantity of interest at time t: total load or other force on segment of bridge:

$$\int_{xy} Z(x, y, t) w(x, y) dx dy$$

Example w: load in strip across bridge between x_1 and x_2 feet out from south side on central span

$$W(t, x_1, x_2) = \int_{x_1}^{x_2} \int_0^{W_B} Z(x, y, t) dy dx$$

Quantity of concern to engineers:

$$M_T(L) \equiv \max_{t\in[0,T]} \max_{0\le x_1 \le L_B - L} W(t, x_1, x_1 + L)$$

First modelling assumption. Years $1, \ldots, T$ are iid.

So:

$$P(M_T(L) \le y) = P(M_1(L) \le y)^T$$

So: years to first exceedance of level y has geometric distribution with probability of success

$$P(M_1(L) > y)$$

Find y so this last is 1/100; expected value of geometric is 100.

Call this the 100 year return time load.

Next modelling consideration.

Two kinds of loads: static and dynamic.

Consider only static not dynamic loading.

Observation: static loading much higher when traffic stopped than not.

So: define N to be number of traffic stoppages in year.

Let $M_{1,n}(L)$ be worst load over segment of length L during nth of N stoppages.

Idea

$$P(M_1(L) > y) = P(\max_{1 \le n \le N} M_{1,n}(L) > y)$$

Treat $M_{1,n}(L)$ as iid *given* N.

Next: Evaluate $P(M_1(L) > y)$ by conditioning.

Shorten notation:

$$M = \max_{1 \le n \le N} X_n$$

where X_i iid, cdf F, survival ftn $S = 1 - F$.

$$\begin{aligned} P(M \le y) &= \mathsf{E}\{P(M \le y|N)\} \\ &= \mathsf{E}[\{1 - S(y)\}^N)] \\ &= \phi[\log\{1 - S(y)\}] \end{aligned}$$

where

$$\phi(t) = \mathsf{E}\left(e^{tN}\right)$$

is the moment generating function of N.

Comment: ϕ is monotone increasing.

So: if $S(y) \leq g(y)$ then

$$P(M > y) \leq 1 - \phi[\log\{1 - g(y)\}]$$

and solving

$$\phi[\log\{1 - g(y)\}] = 0.99$$

gives larger solution than

$$P(M \leq y) = 0.99$$

Remaining steps:

1) Model for N.

2) Model / upper bound for S.

Modelling N:

Simplest idea: Poisson process of accidents.

So N has Poisson(λ) dist for some λ.

Then

$$\phi(t) = \sum e^{-\lambda} \frac{\left(\lambda e^t\right)^n}{n!}$$

which is

$$\phi(t) = \exp\{\lambda(e^t - 1)\}$$

Criticisms:

No allowance for variation in traffic densities, weather, etc from year to year.

Potentially better assumption.

N is overdispersed Poisson, say, Negative Binomial:

$$P(N = k) = \binom{r+k-1}{k} p^r (1-p)^k \quad k = 0, 1, \cdots$$

This makes

$$\mathsf{E}(e^{tX}) = \frac{p^r}{\{1-(1-p)e^t\}^r}$$

Idea: for Poisson $\sigma = \sqrt{\mu}$.

For Negative Binomial $\mu = r(1-p)/p$ and

$$\sigma = \sqrt{r(1-p)/p^2} > \sqrt{1/p}\sqrt{\mu} > \sqrt{\mu}$$

Idea: use of overdispersed variable makes for longer tails relative to mean.

Now we need to model / bound the survival function S.

Stoppage lasts random time T.

During that time traffic builds up behind stoppage; cars jam together.

Worst section of length L found by sliding window along line of stopped cars to find maximum.

Notional model (not the way we did it):

Model vehicles arriving at end of queue.

Might use Poisson Process.

Each vehicle has random mass, length, distribution of load along length.

Random gaps between vehicle.

Just before traffic starts to move again:

Look for heaviest segement of length L in stoppage.

Problems:

1) different kinds of stoppage: # lanes, direction of flow, location on bridge, cars trickle past?

2) hard to deal with supremum over all segments of length L.

3) specify joint law of mass, length, distribution of load along single vehicle.

Digression to method we didn't use:

Model length of stoppage T with density g.

Model N, number of vehicles arriving at end of stoppage, given T as Poisson(λT).

Assume next vehicle arriving picked at random; joint density $h(w, l)$ of weight, length. W_i, L_i values for ith arrival.

Assume load distributed evenly along length of vehicle.

Final length of line at end of stoppage is

$$L_T \equiv \sum_{i=1}^{N} L_i.$$

Can compute mean, variance, generating function of L_T?

$$\begin{aligned} \mathrm{E}\{\exp(sL)\} &= \mathrm{E}\left[\mathrm{E}\{\exp(sL)|N\}\right] \\ &= \mathrm{E}\left([\mathrm{E}\{\exp(sL_i)\}]^N\right) \\ &= \phi_N[\log\{\phi_L(s)\}] \end{aligned}$$

Here each ϕ is a moment generating function.

This is a method of analysis for a compound Poisson Process. Can use the mgf of L to compute distribution of L by inversion of Laplace transform.

Problem: how to scan for maximum load?

Simplify problem: discretize and bound.

General idea discretization.

If h is small and $X(s)$ some process then

$$\sup_{0 \le t \le T-\tau} \int_t^{t+\tau} X(s)ds$$

is close to

$$\max_k \int_{kh}^{kh+\tau} X(s)ds$$

We took h to be 50 feet.

Considered $\tau = 50n$ feet.

Switch from thinking about length of stoppage in time to length of stoppage in multiples of h.

Let N_i be number of segments of length h building up on bridge during stoppage

Let $X_j; j = 1, \ldots, N_i$ be the loads on the consecturive segments.

So: our interest is in

$$\max\{X_r + \cdots X_{r+n-1}; 1 \le r \le N_i + 1 - n\}$$

Upper bound on survival function?

Define

$$U_r = X_r + \cdots X_{r+n-1}$$

Argue that

$$1 - P(\max_{1 \le r \le m} U_r > y) \le 1 - \prod_{1 \le r \le m} \{1 - P(U_r > y)\}$$

for any m.

Rationale: Values of U_r are *positive orthant dependent*.

(Large values of one U suggest large values of adjacent U.)

So now:

$$P(\max\{U_r; 1 \le r \le N_i\} > y) \\ \le 1 - \phi_{N_i}[\log\{1 - S_U(y)\}]$$

Final step we took.

Model law of X_i by considering possible loading patterns by cars, trucks, buses.

We took cars to be fixed length and weight.

Same for buses.

Trucks had fixed length, weight uniform on 12 to 40 tons.

Computed moment generating function of an X:

$$\phi_X(t) = \mathrm{E}(e^{tX})$$

Final step. Need to compute S_U.

Instead use Markov's inequality:

$$P(X \geq x) \leq \frac{\mathsf{E}\{g(X)\}}{g(x)}$$

for any increasing positive g.

Choose $g(\cdot) = \exp(h\cdot)$.

So:

$$S_U(y) \leq \frac{\mathsf{E}(e^{hU})}{\exp(hy)} = \frac{\{\phi_X(h)\}^n}{\exp(hy)}$$

where

$$\phi_X(h) = \mathsf{E}\{\exp(hX)\}$$

These can be assembled to give a bound depending on h.

Then: minimize over $h > 0$ to find good bound.