4 Branching Processes

Organise by **generations**: Discrete time.

If \( P(\text{no offspring}) \neq 0 \) there is a probability that the process will die out.

Let \( X = \text{number of offspring of an individual} \)

\[
p(x) = P(X = x) = \text{“offspring prob. function”}
\]

**Assume:**

(i) \( p \) same for all individuals

(ii) individuals reproduce independently

(iii) process starts with a single individual at time 0.

Assumptions (i) and (ii) define the **Galton-Watson** discrete time branching process.

Two random variables of interest:

\[
Z_n = \text{number of individuals at time } n \quad (Z_0 = 1 \quad \text{by (iii)})
\]

\[
T_n = \text{total number born up to and including generation } n
\]

**e.g.**

\[
p(0) = r \quad p(1) = q \quad p(z) = p
\]

What is the probability that the second generation will contain 0 or 1 member?

\[
P(Z_2 = 0) = P(Z_1 = 0) + P(Z_1 = 1) \times P(Z_2 = 0 \mid Z_1 = 1) + P(Z_1 = 2) \times P(Z_2 = 0 \mid Z_1 = 2)
\]

\[
= r + qr + pr^2
\]

\[
P(Z_1 = 1) = P(Z_1 = 1) \times P(Z_2 = 1 \mid Z_1 = 1) + P(Z_1 = 2) \times P(Z_2 = 1 \mid Z_1 = 2)
\]

\[
= q^2 + p(rq + qr) = q^2 + 2pqr.
\]

**Note:** things can be complicated because

\[
Z_2 = X_1 + X_2 + \ldots X_{Z_1}
\]

with \( Z_1 \) a random variable.
4.1 Revision: Probability generating functions

Suppose a discrete random variable $X$ takes values in $\{0, 1, 2, \ldots\}$ and has probability function $p(x)$.

Then p.g.f. is

$$\Pi_X(s) = E(s^X) = \sum_{x=0}^{\infty} p(x)s^x$$

Note: $\Pi(0) = p(0)$

$$\Pi(1) = \sum p(x) = 1$$

pgf uniquely determines the distribution and vice versa.

4.2 Some important pgfs

<table>
<thead>
<tr>
<th>Distribution</th>
<th>pdf</th>
<th>Range</th>
<th>pgf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli($p$)</td>
<td>$p^xq^{1-x}$</td>
<td>$0, 1$</td>
<td>$q + ps$</td>
</tr>
<tr>
<td>Binomial($n, p$)</td>
<td>$\binom{n}{x}p^xq^{n-x}$</td>
<td>$0, 1, 2, \ldots, n$</td>
<td>$(q + ps)^n$</td>
</tr>
<tr>
<td>Poisson($\mu$)</td>
<td>$\frac{e^{-\mu}s^x}{x!}$</td>
<td>$0, 1, 2\ldots$</td>
<td>$e^{-\mu(1-s)}$</td>
</tr>
<tr>
<td>Geometric,$G_1(p)$</td>
<td>$q^{x-1}p$</td>
<td>$1, 2\ldots$</td>
<td>$\frac{ps}{1-qs}$</td>
</tr>
<tr>
<td>Negative Binomial</td>
<td>$\binom{x-1}{r-1}q^{x-r}p^r$</td>
<td>$r, r + 1, \ldots$</td>
<td>$\left(\frac{ps}{1-qs}\right)^r$</td>
</tr>
</tbody>
</table>

4.3 Calculating moments using pgfs

$$\Pi(s) = \sum_{x=0}^{\infty} p(x)s^x = E(s^X).$$

Then

$$\Pi'(s) = E(Xs^{X-1})$$

$$\Pi'(1) = E(X)$$
Likewise
\[
\Pi''(s) = E\left[X(X-1)s^{X-2}\right]
\]
\[
\Pi''(1) = E\left[X(X-1)\right] = E(X^2) - E(X).
\]
So
\[
\text{var}(X) = E(X^2) - E^2(X)
\]
\[
= \left[\Pi''(1) + E(X)\right] - \Pi'(1)^2
\]
\[
\text{var}(X) = \Pi''(1) + \Pi'(1) - \Pi'(1)^2
\]
\[
\mu = \Pi'(1); \quad \sigma^2 = \Pi''(1) + \mu - \mu^2
\]

4.4 Distribution of sums of independent rvs

$X, Y$ independent discrete rvs on $\{0, 1, 2, \ldots\}$, let $Z = X + Y$.

\[
\Pi_Z(s) = E(s^Z) = E(s^{X+Y})
\]
\[
= E(s^X)E(s^Y) \quad \text{(indep)}
\]
\[
= \Pi_X(s)\Pi_Y(s)
\]

In general:
If
\[
Z = \sum_{i=1}^{n} X_i
\]
with $X_i$ independent discrete rvs with pgfs $\Pi_i(s)$, $i = 1, \ldots, n$, then
\[
\Pi_Z(s) = \Pi_1(s)\Pi_2(s)\ldots\Pi_n(s).
\]

In particular, if $X_i$ are identically distributed with pgf $\Pi(s)$, then
\[
\Pi_Z(s) = [\Pi(s)]^n.
\]
e.g. Q: If $X_i \sim G_1(p)$, $i = 1, \ldots, n$ (number of trials up to and including the 1st success) are independent, find the pgf of $Z = \sum_{i=1}^{n} X_i$, and hence identify the distribution of $Z$.

A: Pgf of $G_1(p)$

$$\Pi_X(s) = \frac{ps}{1 - qs}.$$ 

So pgf of $Z$ is

$$\Pi_Z(s) = \left(\frac{ps}{1 - qs}\right)^n,$$

which is the pgf of a negative binomial distribution.

Intuitively:

Neg. bin. = number trials up to and including $n$th success 

= sum of $n$ sequences of trials each consisting of number of failures 

followed by a success.

= sum of $n$ geometrics.

4.5 Distribution of sums of a random number of independent rvs

Let 

$$Z = X_1 + X_2 + \ldots + X_N$$

$N$ is a rv on \{0, 1, 2, \ldots\} 

$X_i$ iid rvs on \{0, 1, 2, \ldots\} 

(Convention $Z = 0$ when $N = 0$).

$$\Pi_Z(s) = \sum_{z=0}^{\infty} P(Z = z) s^z$$

$$P(Z = z) = \sum_{n=0}^{\infty} P(Z = z \mid N = n) P(N = n) \quad \text{(Thm T.P.)}$$

$$\Pi_Z(s) = \sum_{n=0}^{\infty} P(N = n) \sum_{z=0}^{\infty} P(Z = z \mid N = n) s^z$$

$$= \sum_{n=0}^{\infty} P(N = n) [\Pi_X(s)]^n \quad \text{(since $X_i$ iid)}$$

$$= \Pi_N [\Pi_X(s)]$$

23
If $Z$ is the sum of $N$ independent discrete rvs $X_1, X_2, \ldots, X_N$, each with range \{0, 1, 2, \ldots\} each having pgf $\Pi_X(s)$ and where $N$ is a rv with range \{0, 1, 2, \ldots\} and pgf $\Pi_N(s)$ then

$$\Pi_Z(s) = \Pi_N[\Pi_X(s)]$$

and $Z$ has a compound distribution.

e.g. Q: Suppose $N \sim G_0(p)$ and each $X_i \sim Binomial(1, \theta)$ (independent).
Find the distribution of $Z = \sum_{i=1}^N X_i$.

A:

$$\Pi_N(s) = \frac{q}{1 - ps}$$
$$\Pi_X(s) = 1 - \theta + \theta s$$

So

$$\Pi_Z(s) = \frac{q}{1 - p(1 - \theta + \theta s)}$$
$$= \frac{q}{q + p\theta - p\theta s} = \frac{q/(q + p\theta)}{1 - (p\theta s/(q + p\theta))}$$
$$= \frac{1 - p\theta/(q + p\theta)}{1 - (p\theta/(q + p\theta)) s}$$

which is the pgf of $G_0\left(\frac{p\theta}{q + p\theta}\right)$ distribution.

Note: even in the cases where $Z = \sum_{i=1}^N X_i$ does not have a recognisable pgf, we can still use the resultant pgf to find properties (e.g. moments) of the distribution of $Z$.

We have probability generating function (pgf):

$$\Pi_X(s) = E(s^X).$$

Also: moment generating function (mgf):

$$\Phi_X(t) = E(e^{tX}),$$

Take transformation $s = e^t$:

$$\Pi_X(e^t) = E(e^{tX})$$

the mgf has many properties in common with the pgf but can be used for a wider class of distributions.
4.6 Branching processes and pgfs

Recall $Z_n = \text{number of individuals at time } n \ (Z_0 = 1)$, and $X_i = \text{number of offspring of individual } i$. We have

$$Z_2 = X_1 + X_2 + \ldots X_{Z_1}.$$ 

So,

$$\Pi_2(s) = \Pi_1 [\Pi_1(s)].$$

\textbf{e.g.} consider the branching process in which

$$p(0) = r \quad p(1) = q \quad p(2) = p.$$ 

So

$$\Pi(s) = \Pi_1(s) = \sum_{x=0}^{2} p(x) s^x = r + qs + ps^2,$$

and

$$\Pi_2(s) = \Pi_1[\Pi_1(s)] = r + q\Pi_1(s) + p\Pi_1(s)^2$$

$$= r + q(r + qs + ps^2) + p(r + qs + ps^2)^2$$

$$= r + qr + pr^2 + (q^2 + 2pqr)s + (pq + pq^2 + 2pq^2r)s^2 + 2p^2qs^3 + p^3s^4$$

Coefficients of $s^x \ (x = 0, 1, 2, \ldots, 4)$ give probability $Z_2 = x$.

What about the $n$th generation?

Let

$$Y_i = \text{number offspring of } i\text{th member of } (n - 1)\text{th generation}$$

Then

$$Z_n = Y_1 + Y_2 \ldots + Y_{Z_{n-1}},$$

so,

$$\Pi_n(s) = \Pi_{n-1}[\Pi(s)]$$

$$= \Pi_{n-2}[\Pi[\Pi(s)]]$$

$$= \ldots$$

$$= \Pi[\underbrace{[\ldots[\Pi(s)]\ldots]}_n]$$

25
writing this out explicitly can be complicated.

But sometimes we get lucky: e.g. \( X \sim \text{Binomial}(1, p) \), then \( \Pi_X(s) = q + ps \).

So,

\[
\begin{align*}
\Pi_2(s) & = q + p(q + ps) = q + pq + p^2s \\
\Pi_3(s) & = q + p(q + p(q + ps)) \\
& = q + pq + p^2q + p^3s \\
& \vdots \\
\Pi_n(s) & = q + pq + p^2q + \ldots + p^{n-1}q + p^n s.
\end{align*}
\]

Now \( \Pi_n(1) = 1 \), so

\[
\begin{align*}
(1 - p^n) & = q + pq + p^2q + \ldots + p^{n-1}q \\
\Pi_n(s) & = 1 - p^n + p^n s
\end{align*}
\]

This is the pgf of a \( \text{Binomial}(1, p^n) \) distribution.

\( \Rightarrow \) The distribution of the number of cases in the \( n \)th generation is Bernoulli with parameter \( p^n \). i.e:

\[
\begin{align*}
P(Z_n = 1) & = p^n \\
P(Z_n = 0) & = 1 - p^n.
\end{align*}
\]

### 4.7 Mean and Variance of size of \( n \)th generation of a branching process

**mean:** Let \( \mu = \mathbb{E}(X) \) and let \( \mu_n = \mathbb{E}(Z_n) \).

\[
\begin{align*}
\mu & = \Pi'(1) \\
\Pi_n(s) & = \Pi_{n-1}[\Pi(s)] \\
\Rightarrow \Pi_n'(s) & = \Pi_{n-1}'[\Pi(s)]\Pi'(s) \\
\Pi_n'(1) & = \Pi_{n-1}'[\Pi(1)]\Pi'(1) \\
& = \Pi_{n-1}'(1)\Pi'(1) \\
\Rightarrow \mu & = \mu_{n-1}\mu = \mu_{n-2}\mu^2 = \ldots = \mu^n.
\end{align*}
\]
Note: as $n \to \infty$

$$
\mu_n = \mu^n \to \begin{cases} 
\infty & \mu > 1 \\
1 & \mu = 1 \\
0 & \mu < 1 
\end{cases}
$$

so, at first sight, it looks as if the generation size will either increase unboundedly (if $\mu > 1$) or die out (if $\mu < 1$) - slightly more complicated...

**variance:** Let $\sigma^2 = \text{var}(X)$ and let $\sigma_n^2 = \text{var}(Z_n)$.

$$
\Pi'_n(s) = \Pi'_{n-1} [\Pi(s)] \Pi'(s) \\
\Pi''_n(s) = \Pi''_{n-1} [\Pi(s)] \Pi'(s)^2 + \Pi'_{n-1} [\Pi(s)] \Pi''(s) \tag{1}
$$

Now $\Pi(1) = 1$, $\Pi'(1) = \mu$, $\Pi''(1) = \sigma^2 - \mu + \mu^2$.

Also, since $\sigma_n^2 = \Pi''_n(1) + \mu_n - \mu_n^2$, we have

$$
\Pi''_n(1) = \sigma_n^2 - \mu_n + \mu_n^2 \\
\text{and } \Pi''_{n-1}(1) = \sigma_{n-1}^2 - \mu_n^{n-1} + \mu_n^{2n-2}.
$$

From (1),

$$
\Pi''(1) = \Pi''_{n-1}(1) \Pi'(1)^2 + \Pi'_{n-1}(1) \Pi''(1) \\
\sigma_n^2 - \mu_n + \mu_n^2 = (\sigma_{n-1}^2 - \mu_n^{n-1} + \mu_n^{2n-2}) \mu_n^2 + \mu_n^{n-1}(\sigma^2 - \mu + \mu^2) \\
\Rightarrow \sigma_n^2 = \mu_n \sigma_n^2 + \mu_n^{n-1} \sigma^2
$$

Leading to

$$
\sigma_n^2 = \mu_n^{n-1} \sigma^2 (1 + \mu + \mu^2 + \ldots + \mu^{n-1})
$$

So, we have

$$
\sigma_n^2 = \begin{cases} 
\mu_n^{n-1} \sigma^2 \frac{1 - \mu^n}{1 - \mu} & \mu \neq 1 \\
n \sigma^2 & \mu = 1
\end{cases}
$$

### 4.8 Total number of individuals

Let $T_n$ be the total number up to and including generation $n$. Then

$$
E(T_n) = E(Z_0 + Z_1 + Z_2 + \ldots + Z_n) \\
= 1 + E(Z_1) + E(Z_2) + \ldots + E(Z_n)
$$
\[ \lim_{n \to \infty} \frac{\mu^{n+1} - 1}{\mu - 1} = \begin{cases} \frac{\mu^{n+1} - 1}{\mu - 1} & \mu \neq 1 \\ n + 1 & \mu = 1 \end{cases} \]

\[ \lim_{n \to \infty} E(T_n) = \begin{cases} \infty & \mu \geq 1 \\ \frac{1}{1-\mu} & \mu < 1 \end{cases} \]

### 4.9 Probability of ultimate extinction

Necessary that \( P(X = 0) = p(0) \neq 0 \).

Let \( \theta_n = P(n\text{th generation contains 0 individuals}) = P(\text{extinction occurs by } n\text{th generation}) \)

\[ \theta_n = P(Z_n = 0) = \Pi_n(0) \]

Now \( P(\text{extinct by } n\text{th generation}) = P(\text{extinct by } (n - 1)\text{th}) + P(\text{extinct at } n) \).

So, \( \theta_n = \theta_{n-1} + P(\text{extinct at } n) \)

\[ \Rightarrow \theta_n \geq \theta_{n-1}. \]

Now,

\[ \Pi_n(s) = \Pi[\Pi_{n-1}(s)] \]
\[ \Pi_n(0) = \Pi[\Pi_{n-1}(0)] \]
\[ \theta_n = \Pi(\theta_{n-1}). \]

\( \theta_n \) is a non-decreasing sequence that is bounded above by 1 (it is a probability), hence, by the monotone convergence theorem \( \lim_{n \to \infty} \theta_n = \theta^* \) exists and \( \theta^* \leq 1 \).

Now \( \lim_{n \to \infty} \theta_n = \Pi(\lim_{n \to \infty} \theta_{n-1}) \), so \( \theta^* \) satisfies

\[ \theta = \Pi(\theta), \quad \theta \in [0, 1]. \]

Consider

\[ \Pi(\theta) = \sum_{x=0}^{\infty} p(x)\theta^x \]

\( \Pi(0) = p(0) \ (> 0) \), and \( \Pi(1) = 1 \), also \( \Pi'(1) > 0 \) and for \( \theta > 0, \Pi''(\theta) > 0 \), so \( \Pi(\theta) \) is a convex increasing function for \( \theta \in [0, 1] \) and so solutions of \( \theta = \Pi(\theta) \) are determined by slope of \( \Pi(\theta) \) at \( \theta = 1 \), i.e. by \( \Pi'(1) = \mu \).
So,

1. If $\mu < 1$ there is one solution at $\theta^* = 1$.
   $\Rightarrow$ extinction is certain.

2. If $\mu > 1$ there are two solutions: $\theta^* < 1$ and $\theta^* = 1$, as $\theta_n$ is increasing, we want the smaller solution.
   $\Rightarrow$ extinction is NOT certain.

3. If $\mu = 1$ solution is $\theta^* = 1$.
   $\Rightarrow$ extinction is certain.

Note: mean size of $n$th generation is $\mu^n$. So if extinction does not occur the size will increase without bound.

Summary:

$P(\text{ultimate extinction of branching process}) = \text{smallest positive solution of } \theta = \Pi(\theta)$

1. $\mu \leq 1 \Rightarrow \theta^* = 1 \Rightarrow$ ultimate extinction certain.

2. $\mu > 1 \Rightarrow \theta^* < 1 \Rightarrow$ ultimate extinction not certain

E.g.

$X \sim \text{Binomial}(3, \mu); \quad \theta = \Pi(\theta) \quad \Rightarrow \quad \theta = (q + p\theta)^3$.

i.e.

$p^3\theta^3 + 3p^2q\theta^2 + (3pq^2 - 1)\theta + q^3 = 0. \quad (2)$

Now $E(X) = \mu = 3p$ i.e. $\mu > 1$ when $p > 1/3$, and

$P(\text{extinction}) = \text{smallest solution of } \theta = \Pi(\theta)$.

Since we know $\theta = 1$ is a solution, we can factorise (2):

$(\theta - 1)(p^3\theta^2 + (3p^2q + p^3)\theta - q^3) = 0$

E.g. if $p = 1/2$ (i.e $p > 1/3$ satisfied), we know that $\theta^*$ satisfies

$\theta^2 + 4\theta - 1 = 0 \quad \Rightarrow \quad \theta^* = \sqrt{5} - 2 = 0.236$.
4.10 Generalizations of simple branching process

1: \( k \) individuals in generation 0

Let

\[ Z_{n,i} = \text{number individuals in nth generation descended from } i\text{th ancestor} \]

\[ S_n = Z_{n,1} + Z_{n,2} + \ldots + Z_{n,k} \]

Then,

\[ \Pi_{S_n} = \left[ \Pi_n(s) \right]^k. \]

2: Immigration: \( W_n \) immigrants arrive at \( n \)th generation and start to reproduce.

Let pgf for number of immigrants be \( \Psi(s) \).

\[ Z^*_n = \text{size of n}th \text{ generation } (n = 0, 1, 2, \ldots) \text{ with pgf } \Pi^*_n(s) \]

\[ Z^*_0 = 1 \]

\[ Z^*_1 = W_1 + Z_1 \]

\[ \Pi^*_1(s) = \Psi(s)\Pi(s) \]

\[ Z^*_2 = W_2 + Z_2 \]

\[ \Pi^*_2(s) = \Psi(s)\Pi^*_1(\Pi(s)), \]

as \( \Pi^*_1(\Pi(s)) \) is the pgf of the number of offspring of the \( Z^*_1 \) members of generation 1, each of these has offspring according to a distribution with pgf \( \Pi(s) \).

In general

\[ \Pi^*_n(s) = \Psi(s)\Pi^*_{n-1}[\Pi(s)] \]

\[ = \Psi(s)\Psi[\Pi(s)]\Pi^*_{n-2}[\Pi[\Pi(s)]] \]

\[ = \ldots \]

\[ = \Psi(s)\Psi[\Pi(s)]\ldots\Psi[\underbrace{\Pi[\Pi[\ldots[\Pi[\ldots[\Pi(s)]\ldots]}_{(n-1)\Pi's}]\ldots] \]

\[ = \Psi(s)[\Pi(s)]\ldots[\Pi[\Pi\ldots[\Pi(s)]\ldots]}_{n\Pi's}\ldots] \]

e.g. Suppose that the number of offspring, \( X \), has a Bernoulli distribution and the number of immigrants has a Poisson distribution.

1. Derive the pgf of size of \( n \)th generation \( \Pi^*_n(s) \) and
2. investigate its behaviour as \( n \rightarrow \infty \).

**A:** 1. \( X \sim \text{Binomial}(1, p), \ W_n \sim \text{Poisson}(\mu), \ n = 1, 2, \ldots \)

\[
\Pi(s) = q + ps \quad \Psi(s) = e^{-\mu(1-s)}.
\]

So,

\[
\Pi_1^*(s) = e^{-\mu(1-s)}(q + ps)
\]
\[
\Pi_2^*(s) = \Psi(s)\Pi_1^*(\Pi(s))
\]
\[
= e^{-\mu(1-s)}\Pi_1^*(q + ps)
\]
\[
= e^{-\mu(1-s)}(q + p(q + ps))e^{-\mu(1-q-ps)}
\]
\[
= e^{-\mu(1+p)(1-s)}(1 - p^2 + p^2 s)
\]
\[
\Pi_3^*(s) = \Psi(s)\Pi_2^*(\Pi(s))
\]
\[
= e^{-\mu(1-s)(1+p+p^2)}(1 - p^3 + p^3 s)
\]
\[
\vdots
\]
\[
\Pi^*_n(s) = e^{-\mu(1-s)(1+p+p^2+\ldots+p^{n-1})}(1 - p^n + p^n s)
\]
\[
= e^{-\mu(1-s)(1-p^n)/(1-p)}(1 - p^n + p^n s).
\]

2. As \( n \rightarrow \infty, \ p^n \rightarrow 0 \ (0 < p < 1) \), so

\[
\Pi^*_n(s) \rightarrow e^{-\mu(1-s)/(1-p)} \quad \text{as} \ n \rightarrow \infty.
\]

This is the pgf of a Poisson distribution with parameter \( \mu/(1 - p) \).

- Without immigration a branching process either becomes extinct or increases unboundedly.
- With immigration there is also the possibility that there is a limiting distribution for generation size.
5 Random Walks

Consider a particle at some position on a line, moving with the following transition probabilities:

- with prob $p$ it moves 1 unit to the right.
- with prob $q$ it moves 1 unit to the left.
- with prob $r$ it stays where it is.

Position at time $n$ is given by,

$$X_n = Z_1 + \ldots + Z_n \quad Z_n = \begin{cases} +1 \\ -1 \\ 0 \end{cases}$$

A random process $\{X_n; n = 0, 1, 2, \ldots \}$ is a random walk if, for $n \geq 1$

$$X_n = Z_1 + \ldots Z_n$$

where $\{Z_i\}, i = 1, 2, \ldots$ is a sequence of iid rvs. If the only possible values for $Z_i$ are $-1, 0, +1$ then the process is a simple random walk.

5.1 Random walks with barriers

Absorbing barriers

Flip fair coin: $p = q = \frac{1}{2}, r = 0$.

H $\rightarrow$ you win £1, T $\rightarrow$ you lose £1.

Let $Z_n$ = amount you win on $n$th flip.

Then $X_n$ = total amount you’ve won up to and including $n$th flip.

BUT, say you decide to stop playing if you lose £50 $\Rightarrow$ State space $= \{-50, -49, \ldots \}$

and $-50$ is an absorbing barrier (once entered cannot be left).

Reflecting barriers

A particle moves on a line between points $a$ and $b$ (integers with $b > a$), with the following transition probabilities:

$$
P(X_n = x + 1 \mid X_{n-1} = x) = \frac{2}{3} \quad \begin{cases} a < x < b \\ P(X_n = x - 1 \mid X_{n-1} = x) = \frac{1}{3} \end{cases}
$$
\[
P(X_n = a + 1 \mid X_{n-1} = a) = 1
\]
\[
P(X_n = b - 1 \mid X_{n-1} = b) = 1
\]

\(a\) and \(b\) are reflecting barriers.

Can also have
\[
P(X_n = a + 1 \mid X_{n-1} = a) = p
\]
\[
P(X_n = a \mid X_{n-1} = a) = 1 - p
\]
and similar for \(b\).

**Note:** random walks satisfy the Markov property.

i.e. the distribution of \(X_n\) is determined by the value of \(X_{n-1}\) (earlier history gives no extra info.)

A stochastic process in discrete time which has the Markov property is a Markov Chain.

\[
X \text{ a random walk } \Rightarrow X \text{ a Markov chain}
\]
\[
X \text{ a Markov chain } \not\Rightarrow X \text{ a random walk}
\]

Since the \(Z_i\) in a random walk are iid, the transition probabilities are independent of current position, i.e.

\[
P(X_n = a + 1 \mid X_{n-1} = a) = P(X_n = b + 1 \mid X_{n-1} = b).
\]

### 5.2 Gambler’s ruin

Two players \(A\) and \(B\).

\(A\) starts with \(£j\), \(B\) with \(£(a - j)\).

Play a series of indep. games until one or other is ruined.

\(Z_i = \) amount \(A\) wins in \(i\)th game = ±1.

\[
P(Z_i = 1) = p \quad P(Z_i = -1) = 1 - p = q.
\]

After \(n\) games \(A\) has \(X_n = X_{n-1} + Z_n\),

\[
0 < X_{n-1} < a.
\]

Stop if \(X_{n-1} = 0\) \(A\) loses

or \(X_{n-1} = a\) \(A\) wins.
Random walk with state space \{0, 1, \ldots, a\} and absorbing barriers at 0 and a.

What is the probability that A loses?

Let \( R_j \) = event A is ruined if he starts with \$j$

\[
q_j = P(R_j) \quad q_0 = 1 \quad q_n = 0.
\]

For \(0 < j < a\),

\[
P(R_j) = P(R_j | W)P(W) + P(R_j | \overline{W})P(\overline{W}),
\]

where \(W = \text{event that A wins first bet.}\)

Now \(P(W) = p, P(\overline{W}) = q\).

\[
P(R_j | W) = P(R_{j+1}) = q_{j+1}
\]

because, if he wins first bet he has \$j + 1\$.

So,

\[
q_j = q_{j+1}p + q_{j-1}q \quad j = 1, \ldots, (a - 1) \quad \text{RECURRENCE RELATION}
\]

To solve this, try \(q_k = cx^k\)

\[
\begin{align*}
cl^j &= pcx^{j+1} + qcx^{j-1} \\
    x &= ps^2 + q \quad \text{AUXILIARY/CHARACTERISTIC EQUATIONS} \\
    0 &= px^2 - x + q \\
    0 &= (px - q)(x - 1) \\
    \Rightarrow x &= q/p \quad x = 1
\end{align*}
\]

General solutions:

\textbf{case 1:} If the roots are distinct \((p \neq q)\)

\[
q_j = c_1 \left( \frac{q}{p} \right)^j + c_2.
\]

\textbf{case 2:} If the roots are equal \((p = q = \frac{1}{2})\)

\[
q_j = c_1 + c_2 j.
\]

\textbf{Particular solutions:} using boundary conditions \(q_0 = 1, q_a = 0\) gives
case 1: $p \neq q$

$$q_0 = c_1 + c_2 \quad q_a = c_1 \left( \frac{q}{p} \right)^a + c_2.$$  

Giving,

$$q_j = \frac{(q/p)^j - (q/p)^a}{1 - (q/p)^a}$$  
(check!)

case 2: $p = q = \frac{1}{2}$

$$q_0 = c_1 \quad q_a = c_1 + ac_2$$

So,

$$q_j = 1 - \frac{j}{a}$$

i.e. If $A$ begins with £$j$, the probability that $A$ is ruined is

$$q_j = \begin{cases} 
\frac{(q/p)^j - (q/p)^a}{1 - (q/p)^a} & p \neq q \\
1 - \frac{j}{a} & p = q = \frac{1}{2}
\end{cases}$$

5.2.1 $B$ with unlimited resources

e.g. casino

Case 1: $p \neq q$,

$$q_j = \frac{(q/p)^j - (q/p)^a}{1 - (q/p)^a}.$$  

(a) $p > q$: As $a \to \infty$, $q_j \to (q/p)^j$.

(b) $p < q$: As $a \to \infty$,

$$q_j = \frac{(p/q)^a - j}{(p/q)^a - 1} \to 1.$$  

case 2: $p = q = 1/2$.

As $a \to \infty$, $q_j = 1 - j/a \to 1$.

So: If $B$ has unlimited resources, $A$’s probability of ultimate ruin when beginning with £$j$ is

$$q_j = \begin{cases} 
1 & p \leq q \\
(q/p)^j & p > q
\end{cases}$$
5.2.2 Expected duration

Let $X =$ duration when $A$ starts with $\ell j$.

Let $E(X) = D_j$.

Let $Y = A$’s winnings on first bet. So,

$$P(Y = +1) = p \quad P(Y = -1) = q.$$ 

$$E(X) = E_Y[E(X | Y)]$$

$$= \sum_y E(X | Y = y)P(Y = y)$$

$$= E(X | Y = 1)p + E(X | Y = -1)q$$

Now

$$E(X | Y = 1) = 1 + D_{j+1}$$

$$E(X | Y = -1) = 1 + D_{j-1}$$

Hence, for $0 < j < a$

$$D_j = (1 + D_{j+1})p + (1 + D_{j-1})q$$

$$D_j = pD_{j+1} + qD_{j-1} + 1$$

–second-order, non-homogeneous recurrence relation - so, add a particular solution to the general solution of the corresponding homogeneous recurrence relation.

**case 1**: $p \neq q$ (one player has advantage)

General solution for

$$D_j = pD_{j+1} + qD_{j-1}.$$ 

As before $D_j = c_1 + c_2(q/p)^j$.

Now find a particular solution for $D_j = pD_{j+1} + qD_{j-1} + 1$, try $D_j = j/(q-p)$:

$$\frac{j}{q-p} = \frac{p(j+1)}{q-p} + \frac{q(j-1)}{q-p} + 1$$

$$j = pj + qj$$

So general solution to non-homogeneous problem is:

$$D_j = c_1 + c_2 \left( \frac{q}{p} \right)^j + \frac{j}{(q-p)}.$$
Find \( c_1 \) and \( c_2 \) from boundary conditions:

\[
0 = D_0 = c_1 + c_2 \\
0 = D_a = c_1 + c_2 \left( \frac{q}{p} \right)^a + \frac{a}{q-p} \Rightarrow c_2 \left[ 1 - \left( \frac{q}{p} \right)^a \right] = \frac{a}{q-p}.
\]

\[
\begin{align*}
  c_2 &= \frac{a}{(q-p)[1-(q/p)^a]} \\
  c_1 &= \frac{-a}{(q-p)[1-(q/p)^a]}
\end{align*}
\]

**case 2: \( p = q \).**

General solution for

\[
D_j = pD_{j+1} + qD_{j-1}.
\]

As before \( D_j = c_1 + c_2 j \). A particular solution is \( D_j = -j^2 \).

So general solution to non-homogeneous problem is:

\[
D_j = c_1 + c_2 j - j^2.
\]

Find \( c_1 \) and \( c_2 \) from boundary conditions:

\[
0 = D_0 = c_1 \quad 0 = D_a = -a^2 + 0 + c_2 a \Rightarrow c_1 = 0, c_2 = a.
\]

So,

\[
D_j = j(a - j).
\]

Note: this may not match your intuition.

e.g. One player starts with £1000 and the other with £1. They each place £1 bets on a fair coin, until one or other is ruined. What is the expected duration of the game?

We have

\[
a = 1001, \; j = 1, \; p = q = \frac{1}{2}
\]

Expected duration

\[
D_j = j(a - j) = 1(1001 - 1) = 1000 \text{ games!}
\]
5.3 Unrestricted random walks
(one without barriers)
Various questions of interest:

- what is the probability of return to the origin?
- is eventual return certain?
- how far from the origin is the particle likely to be after \( n \) steps?

Let \( R = \) event that particle eventually returns to the origin.
\( A = \) event that the first step is to the right.
\( \bar{A} = \) event that the first step is to the left.

\[
P(A) = p \quad P(\bar{A}) = q = 1 - p
\]

\[
P(R) = P(R | A)P(A) + P(R | \bar{A})P(\bar{A})
\]

Now: event \( R | A \) is the event of eventual ruin when a gambler with a starting amount of £1 is playing against a casino with unlimited funds, so

\[
P(R | A) = \begin{cases} 
  1 & p \leq q \\
  \frac{q}{p} & p > q 
\end{cases}
\]

Similarly,

\[
P(R | \bar{A}) = \begin{cases} 
  1 & p \geq q \\
  \frac{p}{q} & p < q 
\end{cases}
\]

(by replacing \( p \) with \( q \)).

So

\[
p < q : P(R) = 2p; \quad p = q : P(R) = 1; \quad p > q : P(R) = 2q.
\]

i.e. return to the origin is certain only when \( p = q \).

\( p = q \): the random walk is **symmetric** and in this case it is **recurrent** – return to origin is certain.

\( p \neq q \): return is **not** certain. There is a non-zero probability it will never return – the random walk is **transient**.
Note: same arguments apply to every state
⇒ all states are either recurrent or transient,
⇒ the random walk is recurrent or transient.

5.4 Distribution of $X_n$ – the position after $n$ steps

Suppose it has made $x$ steps to the right and $y$ to the left.
Then $x + y = n$, so $X_n = x - y = 2x - n$.

So $n$ even ⇒ $X_n$ even

$n$ odd ⇒ $X_n$ odd

In particular $P(X_n = k) = 0$ if $n$ and $k$ are not either both even or both odd.

Let

$$W_n = \text{number positive steps in first } n \text{ steps}$$

Then $W_n \sim \text{Binomial}(n, p)$.

$$P(W_n = x) = \binom{n}{x} p^x q^{n-x} \quad 0 \leq x \leq n$$

$$P(X_n = 2x - n) = \binom{n}{x} p^x q^{n-x}$$

$$P(X_n = k) = \binom{n}{(n + k)/2} p^{(n+k)/2} q^{(n-k)/2}$$

$X_n$ is sum of $n$ iid rvs, $Z_i$, so use CLT to see large $n$ behaviour:

CLT: $X_n$ is approx. $N(E(X_n), \text{var}(X_n))$ large $n$.

$$E(X_n) = \sum \text{E}(Z_i) = \sum [1 \times p + (-1) \times q] = n(p - q).$$

$$\text{var}(X_n) = \sum \text{var}(Z_i)$$

$$= \sum [\text{E}(Z_i^2) - (\text{E}(Z_i))^2]$$

$$= n[(1 \times p + 1 \times q) - (p - q)^2]$$

$$= 4npq.$$

So,

• If $p > q$ the particle drifts to the right as $n$ increases.

• this drift is faster, the larger $p$. 
• the variance increases with $n$.

• the variance is smaller the larger is $p$.

5.5 Return Probabilities

Recall, probability of return of a SRW (simple random walk) with $p + q = 1$ is 1 if symmetric ($p = q$), $< 1$ otherwise ($p \neq q$).

When does the return occur? Let,

\[ f_n = P(\text{first return occurs at } n) = P(X_n = 0 \text{ and } X_r \neq 0 \text{ for } 0 < r < n) \]
\[ u_n = P(\text{some return occurs at } n) = P(X_n = 0) \]

Since $X_0 = 0$: $u_0 = 1$

Define $f_0 = 0$ for convenience.

We also have $f_1 = u_1 = P(X_1 = 0)$.

We have already found $u_n$:

\[ u_n = P(X_n = 0) = \begin{cases} \binom{n}{n/2} p^{n/2} q^{n/2} & \text{n even} \\ 0 & \text{n odd} \end{cases} \]

Let

\[ R = \text{Event: return eventually occurs, } f = P(R) \]
\[ R_n = \text{Event: first return is at } n, \quad f_n = P(R_n) \]

Then $f = f_1 + f_2 + \ldots + f_n + \ldots$.

To decide if a RW is recurrent or not we could find the $f_n$. Easier to find a relationship between $f_n$ and $u_n$.

Let

\[ F(s) = \sum_{n=0}^{\infty} f_n s^n \quad \text{a pgf if } \sum f_n = 1: \text{true if recurrent} \]
\[ U(s) = \sum_{n=0}^{\infty} u_n s^n \quad \text{not a pgf because } \sum u_n \neq 1 \text{ in general} \]
For any random walk

\[
begin{align*}
    u_1 &= f_1 \\
    u_2 &= f_1 u_1 + f_2 = f_0 u_2 + f_1 u_1 + f_2 u_0 \\
    u_3 &= f_0 u_3 + f_1 u_2 + f_2 u_1 + f_3 u_0.
\end{align*}
\]

In general,

\[
    u_n = f_0 u_n + f_1 u_{n-1} + \ldots + f_{n-1} u_1 + f_n u_0 \quad n \geq 1. \tag{3}
\]

Now,

\[
    F(s) U(s) = \left( \sum_{r=0}^{\infty} f_r s^r \right) \left( \sum_{q=0}^{\infty} u_q s^q \right) = \sum_{n=0}^{\infty} (f_0 u_n + f_1 u_{n-1} + \ldots + f_n u_0) s^n = \sum_{n=1}^{\infty} u_n s^n \text{ from 3 and } f_0 u_0 = 0 = \sum_{n=0}^{\infty} u_n s^n - u_0 = U(s) - 1.
\]

That is

\[
    U(s) = 1 + F(s) U(s); \quad F(s) = 1 - \frac{1}{U(s)}, U(s) \neq 0.
\]

Let \( s \to 1 \) to give \( \sum f_n = 1 - 1/\sum u_n \),

So: \( \sum f_n = 1 \) iff \( \sum u_n = \infty \)

\( \Rightarrow \) a RW is recurrent iff sum of return probabilities is \( \infty \).