Mathematics for Informatics 4a

José Figueroa-O'Farrill

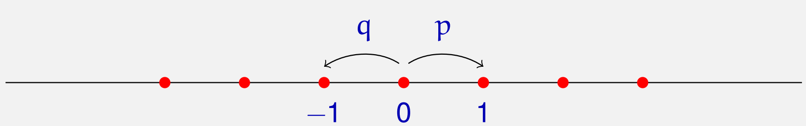
Lecture 17
21 March 2012

The story of the film so far...

- (Temporally homogeneous) **Markov chains** $\{X_0, X_1, \dots\}$ are characterised by an stochastic **transition matrix** P, with entries $p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$ for all n
- The probability distribution π_m at time m obeys $\pi_{m+n} = \pi_m P^n$ for all $m, n \geqslant 0$
- π is a **steady-state distribution** if $\pi P = \pi$
- Finite-state Markov chains always have steady state distributions.
- A (finite-state) Markov chain is **regular** if it has a unique steady state distribution to which all distributions converge
- A (finite-state) Markov chain is regular iff for some n, P^n has no zero entries.
- Examples of Markov chains are given by **random walks**
- Google's PageRank is the steady-state distribution of a random walk on the world wide web.

Random walk revisited

Let us consider again the random walk on the integers:



The jumps J_i are independent random variables with

$$\mathbb{P}(J_i = 1) = p \qquad \mathbb{P}(J_i = -1) = q = 1 - p$$

Starting at 0, $X_n = \sum_{i=1}^n J_i$ is the position after n steps. Let

$$T_r = \begin{cases} \text{number of steps until we visit } r \text{ for the first time}, & r \neq 0 \\ \text{number of steps until we revisit } 0, & r = 0. \end{cases}$$

Question: How the T_r are distributed? i.e., $\mathbb{P}(T_r = n) = ?$

Probability generating functions

To answer this question we introduce some more technology.

Definition

Let X be a d.r.v. taking values in $\{0, 1, 2, \dots\}$. The **probability generating function** $G_X(s)$ of X is the power series

$$G_X(s) = \sum_{n=0}^{\infty} \mathbb{P}(X = n) s^n$$

which agrees with $\mathbb{E}(s^X) = \sum_x p(x) s^x$.

Basic properties:

- $G_X(1) = \sum_x p(x) = 1$
- $G'_X(1) = \sum_x x p(x) = \mathbb{E}(X)$
- $G_X(e^t) = M_X(t)$, the moment generating function

Examples

1. Let X be binomial with parameters (n, p), so $p(r) = \binom{n}{r} p^r q^{n-r}$, for $0 \leqslant r \leqslant n$ and with $q = 1 - p$. Then

$$G_X(s) = \sum_{r=0}^{n} p(r) s^r = \sum_{r=0}^{n} \binom{n}{r} p^r q^{n-r} s^r = (q + ps)^n$$

using the binomial theorem.

2. Let X be geometrically distributed with parameter p, so that $p(k) = q^{k-1} p$ for $k \geqslant 1$ and again $q = 1 - p$. Then

$$G_X(s) = \sum_{k=1}^{\infty} p(k) s^k = \sum_{k=1}^{\infty} q^{k-1} p s^k = ps \sum_{n=0}^{\infty} (qs)^n = \frac{ps}{1 - qs} ,$$

for $|s| < \frac{1}{q}$.

The $\mathbb{P}(X = n)$ are obtained by expanding $G_X(s)$ in powers of s.

Behaviour under independence

Theorem

Let X, Y be independent d.r.v.s with probability generating functions $G_X(s)$ and $G_Y(s)$. Then

$$G_{X+Y}(s) = G_X(s) G_Y(s)$$

Proof is *mutatis mutandis* as for moment generating functions.

Example

Let $X = \sum_{k=1}^{n} I_k$, where I_k are independent Bernoulli trials with success probability p. Then $G_{I_k}(s) = q + ps$, with $q = 1 - p$, and

$$G_X(s) = \prod_{k=1}^{n} G_{I_k}(s) = \prod_{k=1}^{n} (q + ps) = (q + ps)^n ,$$

whence X is binomial with parameters (n, p), as expected.

Conditional expectation I

Definition

Let X, Y be random variables with joint distribution $p_{X,Y}(x, y)$. Then the **conditional distribution of X given Y** is

$$p(x \mid y) = \mathbb{P}(X = x \mid Y = y) = \frac{\mathbb{P}(\{X = x\} \cap \{Y = y\})}{\mathbb{P}(\{Y = y\})} = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

It follows that the marginal distribution

$$p_X(x) = \sum_y p_{X,Y}(x, y) = \sum_y p(x \mid y) p_Y(y)$$

so that

$$\mathbb{E}(X) = \sum_x x p_X(x) = \sum_x \sum_y x p(x \mid y) p_Y(y)$$

Conditional expectation II

Interchanging the order of the sums,

$$\mathbb{E}(X) = \sum_y \sum_x x p(x \mid y) p_Y(y) = \sum_y \mathbb{E}(X \mid Y = y) p_Y(y)$$

which defines the **conditional expectation of X given Y**:

$$\mathbb{E}(X \mid Y = y) = \sum_x x p(x \mid y)$$

This defines a random variable $\mathbb{E}(X \mid Y)$, which is a function of Y, whose value at y is $\mathbb{E}(X \mid Y = y)$. Thus we have

$$\mathbb{E}(X) = \mathbb{E}\left(\mathbb{E}(X|Y)\right)$$

and similarly for any function $Z = h(X)$,

$$\mathbb{E}(Z) = \mathbb{E}(\mathbb{E}(Z|Y)) \qquad \text{where} \qquad \mathbb{E}(Z|Y = y) = \sum_x h(x) p(x \mid y)$$

Example (Random sums)

Let $X_1, X_2, \dots$ be i.i.d. and let N be an $\mathbb{N}$-valued d.r.v. independent from the X_i. Let $T = \sum_{r=0}^{N} X_r$. *What is* $G_T(s)$*?*
We calculate this by conditioning on N:

$$\mathbb{E}(s^T) = \sum_n \mathbb{E}(s^T \mid N = n)\mathbb{P}(N = n)$$

By independence,

$$\mathbb{E}(s^T \mid N = n) = \mathbb{E}(s^{X_1+\cdots+X_n}) = \mathbb{E}(s^{X_1})\dots\mathbb{E}(s^{X_n}) = (G_X(s))^n$$

where $G_X(s)$ is the p.g.f. of any of the X_i. Hence

$$G_T(s) = \sum_n G_X(s)^n \mathbb{P}(N = n) = \mathbb{E}(G_X(s)^N) = G_N(G_X(s))$$

In particular, $\mathbb{E}(T) = G_T'(1) = G_N'(G_X(1))G_X'(1) = \mathbb{E}(N)\mathbb{E}(X)$

Example (Gambler's ruin – revisited)

A gambler starts with £k and makes a number of independent £1 bets with even odds. The gambler stops when she has either £0 or £N. Let T_k be the length of the game. *What is* $\mathbb{E}(T_k)$*?*
Conditioning on the result of the first bet, and letting $\tau_k = \mathbb{E}(T_k)$,

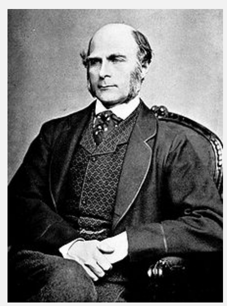
$$\begin{aligned}\tau_k &= \mathbb{E}(T_k|\text{win})\mathbb{P}(\text{win}) + \mathbb{E}(T_k|\text{lose})\mathbb{P}(\text{lose}) \\ &= \tfrac{1}{2}(1 + \tau_{k+1}) + \tfrac{1}{2}(1 + \tau_{k-1}) \\ &= 1 + \tfrac{1}{2}(\tau_{k+1} + \tau_{k-1}) \qquad \text{for } 0 < k < n\end{aligned}$$

whereas $\tau_0 = \tau_N = 0$. τ_k is quadratic in k with zeroes at 0 and N, so $\tau_k = ck(N-k)$ for some constant c. Plugging it into the equation for $k = 1$, we see that $c = 1$ and hence

$$\mathbb{E}(T_k) = k(N - k)$$

The Galton–Watson problem I

In 1873, Francis Galton posed a problem out of his concern in the decay of families of "men of note". In more modern language, a similar problem is the following.
A population of individuals reproduces itself in generations. Let X_n denote the size of the population in the nth generation. There are two rules:

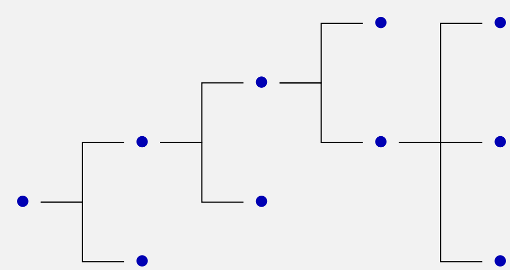


1. each member of a generation produces a family (maybe of size 0) in the next generation
2. family sizes of all individuals are i.i.d. random variables

If we assume that $X_0 = 1$*, what is the probability that* $X_n = 0$ *for some* n*?* i.e., will the family become extinct?

The Galton–Watson problem II

The problem was (partially) solved by the Reverend Henry Watson, a mathematician, who together with Galton wrote *On the probability of extinction of families* in 1874. It gave rise to a class of problems known as **branching processes**.



The Galton–Watson problem III

The population at the nth generation is a random sum of random variables:

$$X_n = \sum_{j=1}^{X_{n-1}} \xi_j^{(n-1)}$$

where $\xi_j^{(n-1)}$ is the size of the family of the jth individual of the $(n-1)$st generation. They are i.i.d. with p.g.f. $G(s)$. Let us write $G_n(s)$ for the p.g.f. of X_n. Then by the ▸ random sums example,

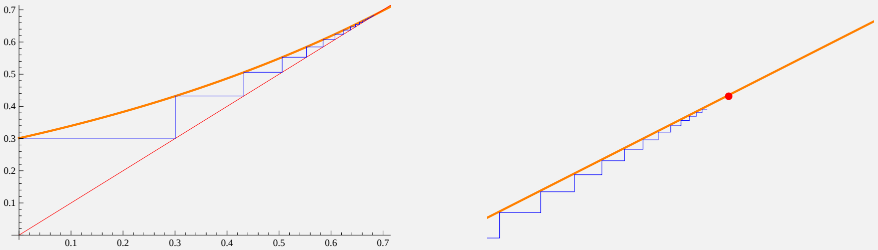
$$G_n(s) = G_{n-1}(G(s)) = G_{n-2}(G(G(s))) = \cdots = G^n(s)$$

i.e., the nth **iterate** of G.
G_n is the p.g.f. of X_n, whence

$$G_n(s) = \sum_{j=0}^{\infty} \mathbb{P}(X_n = j)s^j \implies \mathbb{P}(X_n = 0) = G_n(0) = G^n(0)$$

The Galton–Watson problem IV

We are interested in the large n limit, call it z. If z exists, it obeys $G(z) = z$. Formally, if $G^n(0) \to z$ as $n \to \infty$, applying G again to both sides, we have $G^{n+1}(0) \to G(z)$, but $G^{n+1}(0) \to z$, hence $G(z) = z$. We can also see this graphically:



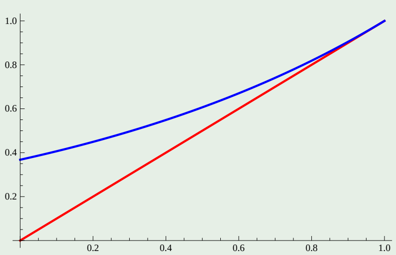
- There is always one solution: $z = 1$, namely *extinction*!
- Watson concluded (incorrectly) that extinction was inevitable.
- Luckily (?) that's not always the case.

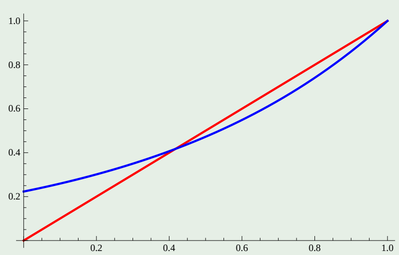
Example (Extinction and survival for Poisson branching)

Suppose that the family sizes are Poisson distributed, so that

$$G(s) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}s^k = e^{-\lambda}e^{\lambda s} = e^{\lambda(s-1)}$$

We must solve the equation $e^{\lambda(z-1)} = z$ for $0 \leqslant z \leqslant 1$. For $\lambda \leqslant 1$ the only solution is $z = 1$, so the family will be extinct with probability 1, but for $\lambda > 1$ there is a nonzero probability of survival:





Example (Extinction and survival for "geometric" branching)

Suppose that the family sizes are distributed by a geometric distribution $p(k) = q^k p$ for $k \geqslant 0$ and $q = 1 - p$. Then

$$G(s) = \sum_{k=0}^{\infty} q^k p s^k = \frac{p}{1 - qs}$$

We must solve the equation $\frac{p}{1-qz} = z$ for $0 \leqslant z \leqslant 1$. It has two roots (for $q \neq 0$, otherwise $z = 1$)

$$z = \frac{1 \pm \sqrt{1 - 4pq}}{2q} = \frac{1 \pm \sqrt{(2p-1)^2}}{2(1-p)}$$

so one root is always 1 (extinction) and the other is $\frac{p}{1-p}$, which is < 1 only for $p < \frac{1}{2}$. So if $p \geqslant \frac{1}{2}$, extinction is inevitable, but if $p < \frac{1}{2}$ there is a chance of survival.

Hitting times for random walks I

- Recall our motivating example: the ▸ one-dimensional random walk
- Let $r > 0$ and let T_r be the number of steps until we visit r for the first time, starting at 0.
- Let $T_{k,k+1}$ be the number of steps needed to reach $k+1$ having reached k. Then $T_{0,1} = T_1$ and the $T_{k,k+1}$ are i.i.d.
- $T_r = T_{0,1} + T_{1,2} + \cdots + T_{r-1,r}$ and by independence

$$\mathbb{E}(s^{T_r}) = \mathbb{E}(s^{T_1})^r$$

- Conditioning on the first jump,

$$\begin{aligned}\mathbb{E}(s^{T_1}) &= \mathbb{E}(s^{T_1} \mid J_1 = 1)\mathbb{P}(J_1 = 1) + \mathbb{E}(s^{T_1} \mid J_1 = -1)\mathbb{P}(J_1 = -1)\\ &= sp + s\mathbb{E}(s^{T_{-1,0}+T_{0,1}})q\\ &= sp + sq\mathbb{E}(s^{T_1})^2\end{aligned}$$

which we solve for $\mathbb{E}(s^{T_1})$

Hitting times for random walks II

- Let $\mathbb{E}(s^{T_1}) = x$ and we must solve $x = sp + sqx^2$
- Assuming that $q \neq 0$, there are two solutions:

$$x = \frac{1 \pm \sqrt{1 - 4pqs^2}}{2sq}$$

but only one has a power series expansion around $s = 0$:

$$\mathbb{E}(s^{T_1}) = \frac{1 - \sqrt{1 - 4pqs^2}}{2sq}$$

- Hence for $r > 0$,

$$\mathbb{E}(s^{T_r}) = \left(\frac{1 - \sqrt{1 - 4pqs^2}}{2sq}\right)^r$$

and for $r < 0$ we simply replace $p \leftrightarrow q$

Hitting times for random walks III

- How about $\mathbb{E}(s^{T_0})$?
- We condition on the first jump:

$$\begin{aligned}\mathbb{E}(s^{T_0}) &= \mathbb{E}(s^{T_0} \mid J_1 = 1)\mathbb{P}(J_1 = 1) + \mathbb{E}(s^{T_0} \mid J_1 = -1)\mathbb{P}(J_1 = -1)\\ &= s\mathbb{E}(s^{T_{1,0}})p + s\mathbb{E}(s^{T_{-1,0}})q\\ &= sp\mathbb{E}(s^{T_{-1}}) + sq\mathbb{E}(s^{T_1})\\ &= sp\left(\frac{1 - \sqrt{1 - 4pqs^2}}{2sp}\right) + sq\left(\frac{1 - \sqrt{1 - 4pqs^2}}{2sq}\right)\\ &= 1 - \sqrt{1 - 4pqs^2}\end{aligned}$$

$$\therefore \mathbb{E}(T_0) = \frac{4pq}{\sqrt{1 - 4pq}} \longrightarrow \infty \qquad \text{if } p = q$$

Summary

- We introduced the **probability generating function** $G_X(s) = \mathbb{E}(s^X)$ of an $\mathbb{N}$-valued d.r.v.
- If X, Y are independent, then $G_{X+Y}(s) = G_X(s)G_Y(s)$
- We defined the **conditional distribution of** X **given** Y: $p(x \mid y) = \mathbb{P}(X = x \mid Y = y)$
- and the **conditional expectation of** X **given** Y, $\mathbb{E}(X \mid Y)$, a d.r.v. and a function of Y: $\mathbb{E}(X \mid Y = y) = \sum_x xp(x \mid y)$
- $\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X \mid Y))$
- We looked at **random sums** of random variables
- We introduced **branching processes** and looked at the Galton–Watson problem of extinction of family names
- We revisited the one-dimensional random walk and calculated the p.g.f.s for the hitting times